



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

ΙΔΡΥΘΕΝ ΤΟ 1837

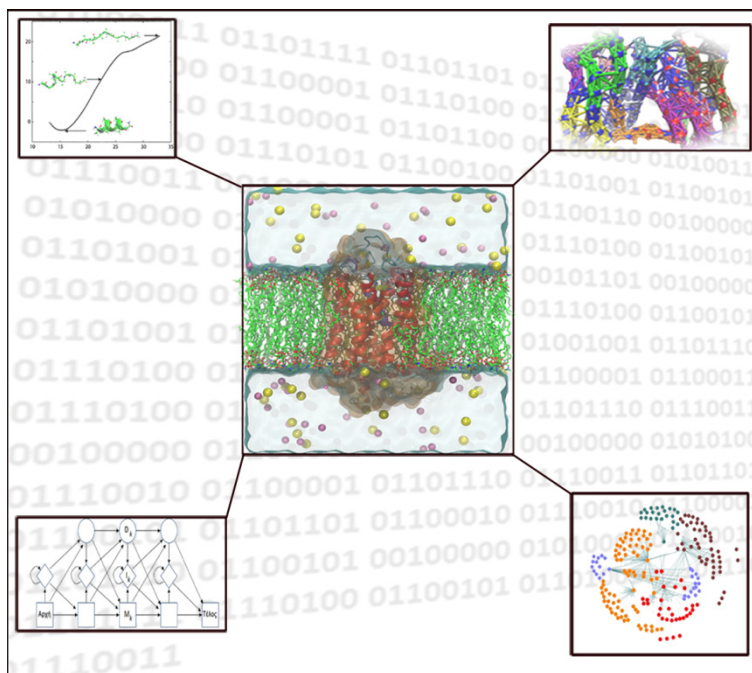
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΒΙΟΛΟΓΙΑΣ ΚΥΤΤΑΡΟΥ & ΒΙΟΦΥΣΙΚΗΣ

Δ Ι Δ Α Κ Τ Ο Ρ Ι Κ Η Δ Ι Α Τ Ρ Ι Β Η

«Υπολογιστικές Μελέτες Αλληλεπιδράσεων
Πρωτεϊνών – Πρωτεϊνών σε Διαμεμβρανικές
Πρωτεΐνες»



Μπαλτούμας Φώτιος

ΒΙΟΛΟΓΟΣ, Msc

ΑΘΗΝΑ 2019

$$\sim \Pi \sim$$



HELLENIC REPUBLIC

**National and Kapodistrian
University of Athens**

— EST. 1837 —

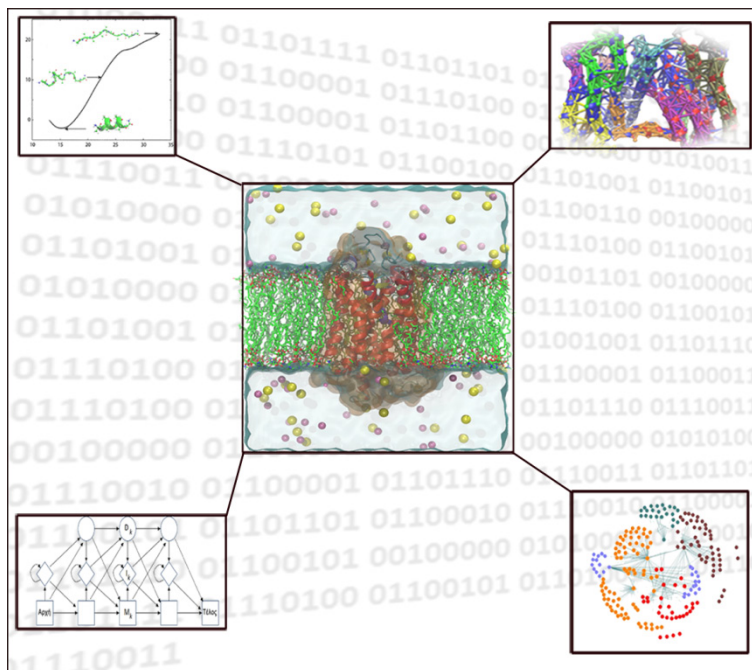
SCHOOL OF SCIENCES

DEPARTMENT OF BIOLOGY

SECTION OF CELL BIOLOGY & BIOPHYSICS

P h D T H E S I S

«Computational Studies of Protein – Protein Interactions in Transmembrane Proteins»



Baltoumas Fotis

Biologist, Msc

ATHENS 2019

Το κείμενο της διδακτορικής διατριβής δεν αποτελεί προϊόν λογοκλοπής.

Η έγκριση της διδακτορικής διατριβής από το Τμήμα Βιολογίας της Σχολής Θετικών Επιστημών του ΕΚΠΑ δεν υποδηλώνει αποδοχή των απόψεων του συγγραφέα (Ν. 5343/1932, άρθρο 202).

Τριμελής Συμβουλευτική Επιτροπή

- | | |
|---|---|
| 1. Ομοτ. Καθηγητής Σταύρος Ι. Χαμόδρακας (επιβλέπων διδακτορικής διατριβής) | Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 2. Επικ. Καθηγήτρια Βασιλική Α. Οικονομίδου | Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 3. Καθηγητής Παντελής Γ. Μπάγκος | Εργαστήριο Μοριακής & Υπολογιστικής Βιολογίας και Γενετικής, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |

Επταμελής Εξεταστική Επιτροπή

- | | |
|---|---|
| 1. Ομοτ. Καθηγητής Σταύρος Ι. Χαμόδρακας (επιβλέπων διδακτορικής διατριβής) | Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 2. Επικ. Καθηγήτρια Βασιλική Α. Οικονομίδου | Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 3. Καθηγητής Παντελής Γ. Μπάγκος | Εργαστήριο Μοριακής & Υπολογιστικής Βιολογίας και Γενετικής, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας |
| 4. Καθηγητής Κ.Ε. Βοργιάς | Τομέας Βιοχημείας & Μοριακής Βιολογίας, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 5. Αναπλ. Καθ. Δ. Ι. Στραβοπόδης | Τομέας Βιολογίας Κυττάρου & Βιοφυσικής, Τμήμα Βιολογίας, Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών |
| 6. Καθ. Η. Ηλιόπουλος | Τμήμα Βιοτεχνολογίας, Γεωπονικό Πανεπιστήμιο Αθηνών |
| 7. Ερευνητής Β' Δρ. Γεώργιος Α. Παυλόπουλος | Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών «Αλέξανδρος Φλέμινγκ» |

Περίληψη

Οι βιολογικές μεμβράνες αποτελούν απαραίτητα συστατικά όλων των κυττάρων. Είναι συνεχείς διπλοστιβάδες αποτελούμενες από διάφορα είδη λιπιδίων, στην επιφάνεια και στο εσωτερικό των οποίων ενσωματώνονται μεμβρανικές πρωτεΐνες. Οι τελευταίες διακρίνονται, ανάλογα με την τοπολογία τους σε σχέση με τη μεμβράνη, σε διαμεμβρανικές, περιφερειακές και αγκυροβολημένες πρωτεΐνες. Οι διαμεμβρανικές πρωτεΐνες υπολογίζεται ότι συνιστούν το 25-30% των γνωστών πρωτεωμάτων και ρυθμίζουν μεγάλο εύρος λειτουργιών στο κύτταρο, από τη μεταγωγή σήματος και τη μεταφορά ουσιών ως τη μηχανική στήριξη του κυττάρου, τη ρύθμιση της γονιδιακής έκφρασης, τον κυτταρικό πολλαπλασιασμό και τον κυτταρικό θάνατο. Έτσι, έχουν εμπλακεί σε μεγάλο αριθμό ασθενειών και αποτελούν κρίσιμους στόχους για το σχεδιασμό φαρμάκων.

Οι βιολογικές διεργασίες σπάνια επιτελούνται από μία μόνο πρωτεΐνη. Αντίθετα, το σύνολο σχεδόν των λειτουργιών του κυττάρου πραγματοποιούνται από ομάδες πρωτεϊνών, οι οποίες αλληλεπιδρούν μεταξύ τους και λειτουργούν συντονισμένα. Στην περίπτωση των διαμεμβρανικών πρωτεϊνών, συνεχώς αυξανόμενος όγκος δεδομένων προτείνουν ότι ο σχηματισμός υπερμοριακών συμπλόκων στο επίπεδο της μεμβράνης, τόσο ανάμεσα σε δύο ή περισσότερες διαμεμβρανικές πρωτεΐνες όσο και ανάμεσα σε διαμεμβρανικές και μη διαμεμβρανικές πρωτεΐνες, διαδραματίζει σημαντικό ρόλο στη λειτουργία τους. Επιπρόσθετα, πολλά υπερμοριακά συμπλέγματα διαμεμβρανικών πρωτεϊνών έχουν συσχετιστεί με μεγάλη ποικιλία ασθενειών. Τέλος, πειραματικά δεδομένα υποδεικνύουν ότι η φύση αυτών των αλληλεπιδράσεων δεν εξαρτάται αποκλειστικά από τις ίδιες τις πρωτεΐνες αλλά επηρεάζεται, σε μεγάλο βαθμό, και από τη σύσταση και τις ιδιότητες του μεμβρανικού περιβάλλοντος γύρω από αυτές.

Παρά τη σημασία τους για το κύτταρο, η πειραματική μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στις διαμεμβρανικές πρωτεΐνες παρουσιάζει προβλήματα. Η δομική ανάλυση των αλληλεπιδράσεων στα υπερμοριακά σύμπλοκα των διαμεμβρανικών πρωτεϊνών δυσχεραίνεται από τη δυσκολία στον προσδιορισμό της δομής των τελευταίων, λόγω της εγγενούς υδρόφοβης φύσης τους. Επιπρόσθετα, τα υπάρχοντα δεδομένα πάνω στις

αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών περιορίζονται, σε μεγάλο βαθμό, σε συγκεκριμένες κατηγορίες πρωτεϊνών που εντοπίζονται στην κυτταρική μεμβράνη, ενώ πολύ λίγα είναι γνωστά πάνω στις διαμεμβρανικές πρωτεΐνες άλλων υποκυτταρικών διαμερισμάτων. Τέλος, παρ'ότι είναι γενικά αποδεκτό ότι η λιπιδική διπλοστιβάδα επηρεάζει την πρωτεϊνική δομή και τις αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών, τα χαρακτηριστικά αυτών των επιρροών έχουν μελετηθεί ελάχιστα.

Αντικείμενο της παρούσας διδακτορικής διατριβής είναι η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών σε διαμεμβρανικές πρωτεΐνες με τη χρήση υπολογιστικών μεθόδων. Στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε μελέτη αυτών των αλληλεπιδράσεων τόσο σε επίπεδο δομής, μέσα από μεθόδους Μοριακής Προτυποποίησης, προσομοιώσεις Μοριακής Δυναμικής και υπολογισμούς Ελεύθερης Ενέργειας, όσο και σε επίπεδο συστήματος, με την εφαρμογή μεθόδων από το πεδίο της Θεωρίας Δικτύων. Παράλληλα με τη μελέτη των ίδιων των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών, διερευνήθηκαν και τα χαρακτηριστικά του μεμβρανικού περιβάλλοντος, με στόχο την αποσαφήνιση του μηχανισμού με τον οποίο η λιπιδική διπλοστιβάδα επηρεάζει τη δομή και τις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών. Οι παραπάνω μελέτες πραγματοποιήθηκαν τόσο για πρωτεΐνες της κυτταρικής μεμβράνης όσο και για πρωτεΐνες που εντοπίζονται σε άλλα υποκυτταρικά διαμερίσματα, όπως οι Εξωτερικές Μέμβρανες των αρνητικών κατά Gram βακτηρίων και ο πυρηνικός φάκελος. Τέλος, στα πλαίσια της διδακτορικής διατριβής αναπτύχθηκαν μια σειρά από υπολογιστικά εργαλεία για τη μελέτη της δομής, της λειτουργίας και των αλληλεπιδράσεων των βιολογικών μεμβρανών και των διαμεμβρανικών πρωτεϊνών.

Αλληλεπιδράσεις ολιγομερισμού σε Συζευγμένους με G-πρωτεΐνες υποδοχείς: Οι Συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs) αποτελούν τη μεγαλύτερη και πιο ποικιλόμορφη υπερικογένεια διαμεμβρανικών πρωτεϊνών στους ευκαρυωτικούς οργανισμούς. Ρυθμίζουν την πλειονότητα των κυτταρικών αποκρίσεων σε εξωκυτταρικά ερεθίσματα και έχουν εμπλακεί σε μεγάλο αριθμό ασθενειών όπως πολλά νευρολογικά και μεταβολικά σύνδρομα, διάφορους τύπους καρκίνου και τη μόλυνση από τον ιό HIV. Έτσι, οι GPCRs αποτελούν σήμερα στόχο για πάνω από 40% των φαρμάκων στην αγορά. Όλοι οι

GPCRs χαρακτηρίζονται από μια συντηρημένη, κοινή τοπολογία και αποτελούνται από 7 α-ελικοειδή διαμεμβρανικά τμήματα, με εξωκυτταρικό αμινοτελικό άκρο και ενδοκυτταρικό καρβοξυτελικό άκρο. Παρότι στο παρελθόν θεωρούταν ότι οι GPCRs δρουν αποκλειστικά ως μονομερή, πλέον είναι γνωστό ότι η πλειονότητα των υποδοχέων σχηματίζουν λειτουργικά διμερή και ανώτερης τάξης ολιγομερή. Για πολλούς GPCRs ο ολιγομερισμός είναι απαραίτητο στοιχείο για την ορθή λειτουργία τους, ενώ σε άλλες περιπτώσεις μπορεί να οδηγήσει σε τροποποίηση των σηματοδοτικών μονοπατιών τους. Επιπρόσθετα, πολλά διμερή υποδοχέων έχουν συσχετιστεί με διάφορες ασθένειες, με αποτέλεσμα να θεωρούνται πιθανοί στόχοι για το σχεδιασμό νέων, πιο αποτελεσματικών φαρμάκων. Ωστόσο, τα δεδομένα πάνω στη δομική φύση των αλληλεπιδράσεων ολιγομερισμού των GPCRs είναι αποσπασματικά. Επιπρόσθετα, πολύ λίγα είναι γνωστά πάνω στους μηχανισμούς που ρυθμίζουν το σχηματισμό και την ειδικότητα αυτών των αλληλεπιδράσεων.

Για τη μελέτη των αλληλεπιδράσεων ολιγομερισμού των GPCRs πραγματοποιήθηκαν εκτεταμένες προσομοιώσεις Μοριακής Δυναμικής, οι οποίες συνδυάστηκαν με μια σειρά από ενεργειακούς υπολογισμούς αλλά και μεθόδους ανάλυσης προερχόμενες από τη Θεωρία Δικτύων. Παράλληλα, διερευνήθηκε η επίδραση του μεμβρανικού περιβάλλοντος πάνω στις αλληλεπιδράσεις ολιγομερισμού, μέσα από την ενεργειακή ανάλυση των αλληλεπιδράσεων των GPCRs με το λιπιδικό περιβάλλον τους. Τέλος, διερευνήθηκε η επίδραση της χοληστερόλης, ενός βασικού συστατικού των ευκαρυωτικών μεμβρανών, πάνω στην ικανότητα των GPCRs να σχηματίζουν αλληλεπιδράσεις ολιγομερισμού, μέσα από προσομοιώσεις σε μεμβράνες με διαφορετικές συγκεντρώσεις στερολών. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι οι GPCRs σχηματίζουν αλληλεπιδράσεις ολιγομερισμού χρησιμοποιώντας συγκεκριμένα διαμεμβρανικά τμήματα, με τα πιο σημαντικά αλληλεπιδρώντα κατάλοιπα να εντοπίζονται κοντά στα άκρα των διαμεμβρανικών τμημάτων. Η ανάλυση των αποτελεσμάτων των προσομοιώσεων, μέσα από την εφαρμογή της Θεωρίας Δικτύων για την κατασκευή δομικών δικτύων αλληλεπιδράσεων, υπέδειξε την ύπαρξη συσχετίσεων ανάμεσα στις επιφάνειες αλληλεπίδρασης των ολιγομερών και σε περιοχές των GPCRs με λειτουργικό ρόλο, αναδεικνύοντας έτσι τα χαρακτηριστικά του μηχανισμού με τον οποίο ο ολιγομερισμός πιθανώς επηρεάζει την λειτουργία των υποδοχέων. Παράλληλα, η ενεργειακή ανάλυση των

αλληλεπιδράσεων ανάμεσα στους GPCRs και τη λιπιδική διπλοστιβάδα έδειξε ότι ο σχηματισμός υπερμοριακών συμπλόκων σχετίζεται με το φαινόμενο της υδροφοβικής ασυμφωνίας ανάμεσα στο μήκος των διαμεμβρανικών τμημάτων και το πάχος της λιπιδικής διπλοστιβάδας, καθώς και ότι οι ελαστικές ιδιότητες της μεμβράνης γύρω από τα διαμεμβρανικά τμήματα των υποδοχέων επηρεάζουν την ικανότητά τους να σχηματίζουν αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών. Τέλος, τα αποτελέσματα των προσομοιώσεων παρουσία της χοληστερόλης έδειξαν ότι η τελευταία επηρεάζει τόσο την ικανότητα των GPCRs να σχηματίζουν αλληλεπιδράσεις ολιγομερισμού όσο και το είδος αυτών των αλληλεπιδράσεων.

Εξωτερικές μεμβράνες των αρνητικών κατά Gram βακτηρίων: Τα αρνητικά κατά Gram βακτήρια περιβάλλονται από ένα σύστημα διπλών μεμβρανών, μιας Εσωτερικής και μιας Εξωτερικής Μεμβράνης. Η Εσωτερική Μεμβράνη είναι συμμετρική και παρουσιάζει δομή και σύσταση παρόμοια με αυτή των ευκαρυωτικών οργανισμών και των θετικών κατά Gram βακτηρίων. Αντίθετα, η Εξωτερική Μεμβράνη είναι μη συμμετρική, αποτελούμενη από τυπικά λιπίδια στην εσωτερική στιβάδα της και μια ειδική κατηγορία λιπογλυκανών, τους λιποπολυσακχαρίτες, στην εξωτερική στιβάδα της. Η δομή και η συνοχή της Εξωτερικής Μεμβράνης είναι κρίσιμη για την ορθή λειτουργία των αρνητικών κατά Gram βακτηρίων, καθώς το στρώμα των λιποπολυσακχαριτών ρυθμίζει μεγάλο μέρος των βακτηριακών λειτουργιών και προστατεύει το βακτηριακό κύτταρο από αντιβιοτικά και άλλες βλαβερές για αυτό ουσίες. Επιπρόσθετα, τα μόρια των λιποπολυσακχαριτών μπορούν να προκαλέσουν τοξικές αποκρίσεις στον άνθρωπο και έχουν συσχετιστεί με μεγάλο αριθμό ασθενειών. Παρ'όλα αυτά, η πειραματική μελέτη των Εξωτερικών Μεμβρανών είναι δύσκολη. Υπολογιστικές τεχνικές, όπως οι προσομοιώσεις Μοριακής Δυναμικής, χρησιμοποιούνται συχνά στη μελέτη των Εξωτερικών Μεμβρανών και των ιδιοτήτων τους. Ωστόσο, η μοριακή προτυποποίηση των Εξωτερικών Μεμβρανών για την διεξαγωγή προσομοιώσεων είναι μια δύσκολη, περίπλοκη και χρονοβόρα διαδικασία, κυρίως λόγω της αυξημένης πολυπλοκότητας των λιποπολυσακχαριτών σε σχέση με τα απλά λιπίδια.

Για την επίλυση του παραπάνω προβλήματος αναπτύχθηκε η υπολογιστική μέθοδος GNOMM, ένα εργαλείο για την αυτοματοποιημένη μοριακή προτυποποίηση Εξωτερικών

Μεμβρανών και την ενσωμάτωση διαμεμβρανικών πρωτεϊνών σε αυτές. Το GNOMM υποστηρίζει την κατασκευή μεμβρανικών συστημάτων σε τέσσερα διαφορετικά επίπεδα διακριτικότητας και παρέχει στο χρήστη μεγάλο αριθμό λιπιδικών μορίων, συμπεριλαμβανομένων μιας σειράς μορίων λιποπολυσακχαριτών από τέσσερα είδη αρνητικών κατά Gram βακτηρίων. Τα αποτελέσματα του GNOMM αξιολογήθηκαν μέσα από προσομοιώσεις Μοριακής Δυναμικής, στις οποίες μελετήθηκαν τόσο οι ιδιότητες της μεμβράνης όσο και οι αλληλεπιδράσεις ανάμεσα σε πρωτεΐνες και λιποπολυσακχαρίτες. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι οι μεμβράνες που προτυποποιούνται από το GNOMM αναπαράγουν πειραματικά προσδιορισμένες μετρήσεις για τους λιποπολυσακχαρίτες. Παράλληλα, τα μόρια των λιποπολυσακχαριτών βρέθηκαν να σχηματίζουν αλληλεπιδράσεις με πρωτεΐνες της Εξωτερικής Μεμβράνης οι οποίες είναι σύμφωνες με προηγούμενα βιοχημικά και κρυσταλλογραφικά δεδομένα. Στο σύνολό τους, τα αποτελέσματα υπέδειξαν ότι η μέθοδος GNOMM είναι ικανή για την προετοιμασία θερμοδυναμικά σταθερών και βιολογικά σημαντικών Εξωτερικών Μεμβρανών και μπορεί να αποτελέσει χρήσιμο εργαλείο στις προσομοιώσεις Μοριακής Δυναμικής αυτών των συστημάτων.

Αλληλεπιδράσεις ολιγομερισμού σε διαμεμβρανικά β-βαρέλια: Οι διαμεμβρανικές πρωτεΐνες της Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων ανήκουν σε μια ξεχωριστή κατηγορία πρωτεϊνών, τα διαμεμβρανικά β-βαρέλια. Τα διαμεμβρανικά β-βαρέλια υιοθετούν δομή β-κλώνων για τα διαμεμβρανικά τμήματά τους, τα οποία σχηματίζουν μια β-πτυχωτή επιφάνεια που περιελίσσεται και λαμβάνει τη δομή βαρελιού. Τα β-βαρέλια ρυθμίζουν μεγάλο εύρος των λειτουργιών του βακτηριακού κυττάρου όπως η μεταφορά ουσιών, η βακτηριακή σύζευξη, η μεταγωγή σήματος καθώς και εξειδικευμένες ενζυμικές λειτουργίες. Κατά συνέπεια, αποτελούν αντικείμενο μελέτης πάνω στη Βιολογία των αρνητικών κατά Gram βακτηρίων και χρησιμοποιούνται συχνά ως στόχοι για το σχεδιασμό αντιβιοτικών και άλλων αντιμικροβιακών φαρμάκων. Πολλά β-βαρέλια έχουν βρεθεί να σχηματίζουν ολιγομερή τα οποία διαδραματίζουν σημαντικό ρόλο στη λειτουργία τους. Ωστόσο, ελάχιστα είναι γνωστά πάνω στους παράγοντες που επηρεάζουν το σχηματισμό αυτών των ολιγομερών.

Στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε μελέτη των αλληλεπιδράσεων ολιγομερισμού των β-βαρελίων, μέσα από προσομοιώσεις Μοριακής Δυναμικής. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι τα ολιγομερή των β-βαρελίων υιοθετούν δομές τριμερών με συντηρημένα χαρακτηριστικά στις αλληλεπιδράσεις τους. Προκειμένου να διερευνηθεί η πιθανή επίδραση των στοιχείων της Εξωτερικής Μembrάνης στη δομή αυτών των τριμερών, πραγματοποιήθηκαν εξειδικευμένες προσομοιώσεις Μοριακής Δυναμικής για τη μελέτη της σταθερότητας των αλληλεπιδράσεων ολιγομερισμού σε περιβάλλον τυπικής λιπιδικής διπλοστιβάδας, καθώς και σε περιβάλλον Εξωτερικής Μembrάνης με λιποπολυσακχαρίτες. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι η σταθερότητα των αλληλεπιδράσεων ολιγομερισμού στη δομή του τριμερούς ενισχύεται από τις αλληλεπιδράσεις των β-βαρελίων με λιποπολυσακχαρίτες. Στο σύνολό τους, τα αποτελέσματα των προσομοιώσεων Μοριακής Δυναμικής δείχνουν ότι τα διαμεμβρανικά β-βαρέλια υιοθετούν συντηρημένα χαρακτηριστικά στις αλληλεπιδράσεις ολιγομερισμού τους, ενώ οι λιποπολυσακχαρίτες διαδραματίζουν σημαντικό ρόλο στη δομή και τη σταθερότητα αυτών των αλληλεπιδράσεων.

O διαμεμβρανικός μεταφορέας SLC11A1: O μεταφορέας δισθενών μεταλλικών ιόντων SLC11A1 είναι μια διαμεμβρανική πρωτεΐνη που ανήκει στην οικογένεια μεταφορέων SLC11, γνωστή και με το όνομα «οικογένεια πρωτεϊνών των μακροφάγων που σχετίζονται με τη φυσική αντίσταση» (NRRAMPs). O SLC11A1 εκφράζεται κυρίως στα μακροφάγα και, σε μικρότερο βαθμό, στα κύτταρα του ήπατος και είναι υπεύθυνος για τη μεταφορά ιόντων Fe^{2+} και Mn^{2+} στο εσωτερικό του κυττάρου. Η ορθή λειτουργία του είναι ιδιαίτερα σημαντική για τη φυσιολογία του οργανισμού. Μεταλλαγές που οδηγούν σε απώλεια της ικανότητας του SLC11A1 στη μεταφορά σιδήρου έχουν βρεθεί να προκαλούν αιμολυτική αναιμία, ενώ διάφοροι πολυμορφισμοί του έχουν συσχετιστεί με τη μόλυνση του οργανισμού από ενδοκυτταρικά παράσιτα, προκαλώντας ασθένειες όπως η φυματίωση, η ασθένεια του Crohn και η ρευματοειδής αρθρίτιδα στον άνθρωπο, και η παραφυματίωση σε διάφορα ζώα. Παρά τη βιολογική και κλινική σημασία του SLC11A1, ωστόσο, ελάχιστα είναι γνωστά τόσο για τη δομή του όσο και για τον τρόπο λειτουργίας του.

Η μελέτη της δομής και των αλληλεπιδράσεων του μεταφορέα SLC11A1 πραγματοποιήθηκε μέσα από μεθόδους μοριακής προτυποποίησης, προσομοιώσεις Μοριακής Δυναμικής και ενεργειακούς υπολογισμούς. Η δομή του μεταφορέα προτυποποιήθηκε σε διαφορετικές στερεοδιατάξεις, η σταθερότητα των οποίων ελέγχθηκε μέσα από προσομοιώσεις Μοριακής Δυναμικής. Η μελέτη των αποτελεσμάτων των προσομοιώσεων αποκάλυψε μια σειρά από βιολογικά σημαντικές δομικές μεταβολές, οι οποίες σχετίζονται με τη μετάβαση του SLC11A1 από τη μία στερεοδιάταξη στην άλλη, καθώς επίσης και με το μηχανισμό με τον οποίο ο μεταφορέας αναγνωρίζει και μεταφέρει ιόντα κατά μήκος της μεμβράνης. Επιπρόσθετα, τα αποτελέσματα των ενεργειακών υπολογισμών υπέδειξαν μια σειρά από κατάλοιπα που συμμετέχουν στην αναγνώριση και πρόσδεση των ιόντων από τον SLC11A1, καθώς και στην επιλεκτικότητα του μεταφορέα απέναντι στο υπόστρωμά του. Στο σύνολό τους, τα παραπάνω αποτελέσματα προτείνουν ένα μοντέλο για την τρισδιάστατη δομή του μεταφορέα SLC11A1 και, παράλληλα, παρουσιάζουν τα δομικά γνωρίσματα του μηχανισμού μέσω του οποίου ο SLC11A1 επιτελεί τη λειτουργία του.

Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs): Οι υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs) αποτελούν μια από τις πιο σημαντικές κατηγορίες υποδοχέων στα Μετάζωα. Είναι διαμεμβρανικές πρωτεΐνες με ένα διαμεμβρανικό τμήμα, ένα εξωκυτταρικό αμινοτελικό άκρο το οποίο περιέχει αυτοτελείς δομικές περιοχές για την αναγνώριση διάφορων ορμονών, και ένα ενδοκυτταρικό καρβοξυτελικό άκρο, το οποίο περιέχει μια αυτοτελή δομική περιοχή με ενζυμική ενεργότητα τυροσινικής κινάσης. Οι RTKs αναγνωρίζουν πεπτιδικές ορμόνες όπως η ινσουλίνη και διάφοροι αυξητικοί παράγοντες και ρυθμίζουν μια σειρά από σημαντικές λειτουργίες στο κύτταρο, συμπεριλαμβανομένων της γονιδιακής έκφρασης, της κυτταρικής αύξησης, της απόπτωσης και του προγραμματισμένου κυτταρικού θανάτου. Έτσι, έχουν συσχετιστεί με διάφορες ασθένειες, στις οποίες συμπεριλαμβάνονται πολλοί τύποι καρκίνου, νευρολογικά σύνδρομα και μεταβολικές παθήσεις. Παρά τη σημασία τους στη μεταγωγή σήματος, οι RTKs έχουν μελετηθεί πειραματικά κυρίως για τον άνθρωπο και για οργανισμούς μοντέλα όπως ο ποντικός, ενώ ελάχιστα είναι γνωστά για την παρουσία και τη λειτουργία τους σε άλλα πρωτεύοντα.

Με δεδομένη τη σημασία των RTKs στους μηχανισμούς μεταγωγής σήματος του κυττάρου αναπτύχθηκε η μέθοδος RTK-PRED, ένας αλγόριθμος για την πρόγνωση και το σχολιασμό των RTKs με βάση την αμινοξική αλληλουχία τους. Η μέθοδος βασίζεται στη χρήση προφίλ Hidden Markov Models (pHMMs) και πραγματοποιεί τόσο την αναγνώριση πιθανών RTKs όσο και τον πλήρη σχολιασμό της τοπολογίας τους, ενώ επιπρόσθετα επιχειρεί το λειτουργικό σχολιασμό των υποδοχέων με βάση το είδος των εξωκυτταρικών περιοχών τους. Η μέθοδος RTK-PRED παρουσιάζει υψηλές τιμές ευαισθησίας (100%), ειδικότητας (99.70%) και αξιοπιστίας (99.72%) και, κατά συνέπεια, αποτελεί ένα αξιόπιστο εργαλείο για την αυτοματοποιημένη ανίχνευση πιθανών RTKs σε νέα, υπό μελέτη πρωτεώματα. Σημειώνεται πως η μέθοδος RTK-PRED είναι η μοναδική, μέχρι στιγμής, διαθέσιμη μέθοδος για την αυτοματοποιημένη πρόγνωση RTKs από την αμινοξική αλληλουχία.

Πρωτεΐνες του πυρηνικού φακέλου: Ο πυρήνας των ευκαρυωτικών κυττάρων αποτελεί ένα από τα κυριότερα υποκυτταρικά διαμερίσματά τους καθώς προστατεύει, οργανώνει και ρυθμίζει το γονιδίωμα, συμβάλλει στη διαδικασία της αντιγραφής του DNA, στη ρύθμιση της μεταγραφής των γονιδίων σε mRNA αλλά και στην πρωτεϊνσύνθεση, καθώς επίσης και στην οργάνωση του κυτταροσκελετού στη μεσόφαση αλλά και στις διεργασίες της κυτταρικής διαίρεσης. Ο πυρήνας περιβάλλεται από ένα ιδιαίτερο σύστημα διπλής μεμβράνης, γνωστό σαν «πυρηνικός φάκελος». Ο τελευταίος διαχωρίζει το κυτταρόπλασμα και το εσωτερικό του πυρήνα, και ελέγχει όλες τις διαδικασίες ανταλλαγής μακρομορίων μεταξύ των δύο. Επιπρόσθετα, συνεισφέρει στον καθορισμό της χωρικής οργάνωσης του πυρήνα. Οι μεμβράνες του πυρηνικού φακέλου περιλαμβάνουν μεγάλο αριθμό μεμβρανικών πρωτεϊνών, οι οποίες συμμετέχουν σε μεγάλο αριθμό αλληλεπιδράσεων. Αυτές περιλαμβάνουν τόσο πρωτεΐνες που ρυθμίζουν τη μεταφορά ουσιών και τη στήριξη του πυρήνα και σχηματίζουν υπερμοριακά συμπλέγματα όπως τα Σύμπλοκα Πυρηνικών Πόρων και η πυρηνική λάμινα, όσο και μεγάλο αριθμό άλλων πρωτεϊνών που σχηματίζουν μεγάλο αριθμό αλληλεπιδράσεων και σχετίζονται με πολλές βιολογικές διεργασίες. Για την ακρίβεια, υπολογίζεται ότι η πολυπλοκότητα των πρωτεϊνικών αλληλεπιδράσεων του πυρηνικού φακέλου είναι εφάμιλλη με αυτή της κυτταρικής μεμβράνης. Ωστόσο, σε σύγκριση με άλλα

υποκυτταρικά διαμερίσματα, όπως η κυτταρική μεμβράνη ή οι μεμβράνες των οργανιδίων, ο πυρηνικός φάκελος παραμένει ελάχιστα μελετημένος.

Πραγματοποιήθηκε αρχικά η συστηματική μελέτη του συνόλου των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στις πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου, μέσα από την κατασκευή και την ανάλυση του δικτύου αλληλεπιδράσεων αυτών των πρωτεϊνών. Το δίκτυο αλληλεπιδράσεων παρουσιάζει τα χαρακτηριστικά των πραγματικών βιολογικών δικτύων και αναδεικνύει το σημαντικό ρόλο των πρωτεϊνών του πυρηνικού φακέλου στο ευκαρυωτικό κύτταρο, τόσο στη ρύθμιση των διεργασιών του πυρήνα όσο και στη ρύθμιση μεγάλου αριθμού άλλων κυτταρικών λειτουργιών. Παράλληλα με την ανάλυση των αλληλεπιδράσεων του ανθρώπινου πυρηνικού φακέλου, σχεδιάστηκε και υλοποιήθηκε η βάση δεδομένων NucEnnDB, μια δημόσια διαθέσιμη πηγή πληροφοριών για τις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις τους. Η βάση περιέχει 2863 καλά σχολιασμένες πρωτεΐνες με γνωστή παρουσία στον πυρηνικό φάκελο, οι οποίες προέρχονται από 394 οργανισμούς και συμμετέχουν σε πάνω από 20000 αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών. Επιπρόσθετα, η βάση παρέχει στο χρήστη μια σειρά από υπολογιστικά εργαλεία, μέσω των οποίων μπορεί να πραγματοποιηθεί η μελέτη των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου, με την αυτοματοποιημένη κατασκευή και ανάλυση δικτύων αλληλεπιδράσεων, καθώς και ο λειτουργικός σχολιασμός αυτών των δικτύων με όρους γονιδιακής οντολογίας, μεταβολικών μονοπατιών και συσχετίσεων με ασθένειες. Η βάση NucEnnDB αποτελεί τη μοναδική, μέχρι στιγμής, διαθέσιμη βάση δεδομένων για τις πρωτεΐνες του πυρηνικού φακέλου.

Abstract

Biological membranes are an integral component of all cells. They are bilayer mixtures composed by various types of lipids and a large number of membrane proteins. The latter can be classified to transmembrane, peripheral and lipid anchored proteins, based on their topology with respect to the membrane. Transmembrane proteins constitute approximately 25-30% of all known proteomes and have been found to regulate a wide range of cellular processes, ranging from signal transduction and substrate transport to maintaining cell integrity, the regulation of gene expression, cell growth and cell death. As such, transmembrane proteins have been implicated in a large number of diseases and are currently prime targets in drug design.

Proteins rarely act alone in the cell. Instead, the wide majority of biological processes are conducted by multiple proteins that interact with one another and operate in a coordinated fashion. In the case of transmembrane proteins, accumulating evidence suggests that the formation of supramolecular protein – protein complexes at the membrane, both between two or more transmembrane proteins and between transmembrane and non-transmembrane proteins, is an integral part of their canonical function. At the same time, a large number of transmembrane protein – protein complexes have been implicated in various diseases. Finally, experimental evidence suggests that the nature of transmembrane protein – protein interactions may not depend exclusively on the proteins themselves, but may also be influenced, to a significant extent, by the composition and properties of the membrane environment.

Despite their importance in the cell, the experimental study of transmembrane protein – protein interactions is not straightforward. The structural analysis of intermolecular contacts is hindered by the difficulties in determining the three dimensional structure of transmembrane proteins, due to their inherent hydrophobic nature. At the same time, existing evidence on protein – protein interactions in transmembrane proteins is mostly limited to specific protein classes found in the plasma membrane, while other cell components remain largely unexplored. Finally, although it is generally accepted that the lipid bilayer may influence the

structure, function and interactions of transmembrane proteins, the structural determinants of these influences remain elusive.

The goal of this dissertation was the computational study of protein – protein interactions in biological membranes and transmembrane proteins. Towards this end, an extensive study of protein – protein interactions was conducted for several transmembrane proteins, both at the structural level, through Molecular Modeling, Molecular Dynamics simulations and Free Energy calculations, and in a system-wide approach, through the application of concepts from Network Theory. Alongside protein – protein interactions, the structural and dynamic aspects of the membrane environment were also investigated, in order to identify and evaluate the structural determinants that govern the lipid bilayer's influences upon transmembrane protein structure and biomolecular interactions. The aforementioned studies were conducted both for transmembrane proteins located at the plasma membrane and for proteins found in other cell components, such as the Outer Membranes of Gram-negative bacteria and the Nuclear Envelope. Finally, during the course of this study, a number of publicly available, computational tools were developed, which can further aid in the study of biological membranes, transmembrane proteins and their interactions.

Oligomerization interactions of G-protein coupled receptors (GPCRs): G-protein coupled receptors (GPCRs) are one of the largest and most diverse superfamilies of transmembrane proteins in eukaryotic organisms. They regulate the majority of cell responses to extracellular stimuli and have been implicated in a wide range of diseases, including several neurological and metabolic syndromes, various types of cancer and HIV infection. As a result, GPCRs are targets for more than 40% of pharmaceuticals on the market. All GPCRs share a common, conserved topology and are composed by seven α -helical transmembrane segments, an extracellular N-terminus and a cytoplasmic C-terminus. Although GPCRs have traditionally been considered to act as monomers, accumulating evidence suggests they may also form dimers and higher-order oligomers. Oligomerization has been found to be an important part of the canonical functionality for some GPCRs, while in other cases the formation of oligomers has been found to lead to alternative signaling pathways for many GPCRs. Finally, an increasing number of GPCR oligomers have been implicated in various pathological conditions and are

considered to be targets for designing novel, more potent drugs. However, the available structural evidence on GPCR oligomerization is sparse and controversial. Furthermore, very little is known on the structural determinants that govern the formation and specificity of GPCR oligomeric interactions.

To investigate the structure and dynamics of GPCR oligomerization, extensive Molecular Dynamics simulations were performed on a large number of GPCR dimers, the results of which were combined with Free Energy calculations and analysis schemes utilizing Network Theory concepts. At the same time, the membrane's influences upon the structure and interactions of GPCR oligomers were investigated through simulations and Free Energy calculations on the dynamics of protein-lipid interactions and membrane elastic properties. Finally, the influence of cholesterol, an important component of eukaryotic plasma membranes, upon the formation of GPCR oligomeric interactions was evaluated through simulations in membrane environments with different cholesterol content values. Simulation results indicated that GPCRs form supramolecular complexes using specific transmembrane segments, with the most important residues of the interface area located near the ends of these segments. The application of Network Theory to the simulation results led to the construction and analysis of structural networks, which revealed the existence of structural and dynamic correlations of motion between the oligomerization interface areas and GPCR regions with functional relevance, thus revealing structural aspects of a potential mechanism through which oligomerization may affect GPCR function. At the same time, the Free Energy analysis of protein-lipid interactions showed that GPCR oligomerization can be directly correlated to the energetics of hydrophobic mismatch between transmembrane segments and the membrane's hydrophobic thickness, as well as that the membrane's elastic properties may influence the ability of specific transmembrane segments in forming protein – protein interactions. Finally, simulation results in cholesterol-rich membranes indicated that the presence and concentration of cholesterol in the membrane may influence both the ability of GPCRs towards forming protein-protein interactions and the structural features of these interactions.

The Outer Membranes of Gram-negative bacteria: The cell envelope of Gram-negative bacteria consists of two separate lipid bilayers, an Inner and an Outer Membrane. The Inner

Membrane is symmetric, very similar to the plasma membranes of other organisms and contains standard lipid molecules. Conversely, the Outer Membrane is highly non-symmetric, containing standard phospholipids in the inner leaflet and mostly lipopolysaccharides, a special class of lipoglycans, in the outer leaflet. The structure and integrity of Outer Membranes is crucial for the canonical functionality of Gram-negative bacteria, as the lipopolysaccharide layer can regulate several bacterial functions and can protect the bacterial cell from antibiotics and other harmful substances. At the same time, various lipopolysaccharide molecules have been found to induce toxic responses in human hosts and have been implicated in several diseases. Despite its importance, the experimental study of the Outer Membrane and its components is challenging. Computational techniques such as Molecular Dynamics simulations are often used to investigate the structural and dynamic properties of membranes. However, modeling and preparation of Outer Membrane systems for simulations is not straightforward, mostly due to the increased complexity of lipopolysaccharides compared to standard lipids.

To help solve this problem, the Gram-Negative Outer Membrane Modeler (GNOMM) was developed. GNOMM is a pipeline designed for the automated modeling of Outer Membrane simulation systems, as well as the insertion of transmembrane proteins in them. The method supports the construction of membrane systems in four different levels of resolution and currently offers a large number of lipids to construct these membranes, including several lipopolysaccharides from four Gram-negative species. Simulation systems designed by GNOMM were evaluated by performing several Molecular Dynamics simulations, the results of which were analyzed and evaluated both in terms of membrane proteins and in terms of protein-lipopolysaccharide interactions. Simulation results showed that the membranes designed by GNOMM can reproduce available experimental measurements on lipopolysaccharides with a significant level of accuracy. At the same time, lipopolysaccharides were found to form protein-lipid interactions reflecting previous biochemical and crystallographic evidence. Overall, these results indicate that GNOMM is capable of designing thermodynamically valid and biologically relevant Outer Membrane simulation systems that can be of use in conducting Molecular Dynamics simulations for Gram-negative bacterial components.

Oligomerization interactions of transmembrane β -barrels: Transmembrane proteins of the Gram-negative Outer Membranes fall under a special category of proteins, known as transmembrane β -barrels. Transmembrane β -barrels adopt a β -strand structure for their transmembrane segments, which constitute a β -sheet that twists to form a barrel structure. Transmembrane β -barrels act mainly as porins and have been found to regulate several bacterial functions, including substrate transport, bacterial conjugation, signal transduction and selected enzyme activities. As such, they are actively studied to determine the physiology of Gram-negative bacteria and are often used as targets for antibiotics and other antimicrobial drugs. Many β -barrels have been found to form oligomers that seem to regulate several aspects of their structure, function and biomolecular interactions. However, very little is known on the determinants that influence the formation of these oligomers.

During the course of this dissertation, a study was performed to investigate the oligomerization properties of transmembrane β -barrels through Molecular Dynamics simulations. Simulation results showed that β -barrel oligomers adopt a trimeric organization with conserved structural features. In order to investigate the potential influence of the Outer Membrane's environment upon this organization, additional simulations and specialized Free Energy calculations were also performed for the trimers, both in standard phospholipid bilayers and in Outer Membrane models containing various types of lipopolysaccharides. Their results indicated that the stability of β -barrel trimers is enhanced by interactions between the β -barrels and adjacent lipopolysaccharides. Overall, the aforementioned results show that transmembrane β -barrels follow a conserved configuration in their oligomerization interactions, and that lipopolysaccharides are important for maintaining the structure and stability of these interactions.

The SLC11A1 transmembrane transporter: SLC11A1 is a transporter of divalent metal ions, belonging to the class of SLC11 transmembrane transporters, also known as Natural Resistance-Associated Membrane Proteins (RAMPs). SLC11A1 is primarily expressed in macrophages and, to a lesser extent, hepatic cells, and is responsible for the transportation of Fe^{2+} and Mn^{2+} divalent ions inside the cell. The transporter's canonical function is highly important for the organism's health. Mutations affecting or abolishing the functionality of

SLC11A1 in iron transport have been found to induce hemolytic anemia, while several polymorphisms of the transporter have been implicated in the host's increased susceptibility to infection by intracellular parasites, causing diseases such as tuberculosis, Crohn's disease and rheumatoid arthritis in humans and paratuberculosis in several animal species. However, despite its biological and clinical relevance, very little is known on SLC11A1 regarding its structure and the mechanisms that mediate its function.

The structure and biomolecular interactions of SLC11A1 were investigated through Molecular Modeling, Molecular Dynamics simulations and Free Energy calculations. Models were constructed for the transporter in distinct, functionally relevant conformations, the stability of which was validated through simulations. The analysis of simulation results revealed a number of biologically relevant structural transitions that can be used to describe the transition of SLC11A1 between different conformations, as well as the mechanism through which iron/manganese transport is conducted through the membrane. At the same time, Free Energy calculations revealed the nature of protein-ion interactions and pinpointed a number of residues that mediate SLC11A1's recognition, selectivity and binding of its substrate. Overall, these results propose a model for the transmembrane structure of SLC11A1 and pinpoint the structural features of the mechanism through which the transporter performs its function.

Receptor Tyrosine Kinases (RTKs): Receptor tyrosine kinases (RTKs) are an important class of transmembrane receptors in Metazoa. They are single-spanning transmembrane proteins, with an extracellular N-terminus containing various domains that recognize peptide ligands and a cytoplasmic C-terminus containing a tyrosine kinase domain. RTKs recognize peptide hormones such as insulin and various growth factors and regulate several cell functions including gene expression, cell growth, apoptosis and programmed cell death. As such, they have been implicated in various diseases, including several types of cancer, neurological syndromes and metabolic deficiencies. Despite their importance inside the cell, RTKs have only been experimentally studied in humans and selected model mammalian species. Conversely, very little is known regarding the existence and function of RTKs in the proteomes of other organisms.

Given the importance of RTKs in signal transduction mechanisms the RTK-PRED method was designed, namely, an algorithm for the detection and annotation of RTKs based on their amino acid sequence. The method is based on the use of profile Hidden Markov Models (pHMMs) and performs the prediction of putative RTKs and a complete annotation of their topology, as well as a functional annotation of the receptors based on their extracellular domains. RTK-PRED produces high values for sensitivity (100%), specificity (99.70%) and accuracy (99.72%) and is, therefore, as suitable method for the automated detection of putative RTKs in novel, unstudied genomes and proteomes. It should be noted that RTK-PRED is, currently, the only publicly available method for the prediction of RTKs from their sequence.

The proteins of the Nuclear Envelope: The nucleus is one of the most important subcellular components of eukaryotic cells. It controls and regulates the structure and expression of genome components, it participates in DNA replication and transcription as well as protein synthesis, it regulates the organization of the cytoskeleton and its components and actively controls the cell cycle and division. The nucleus is surrounded by a double membrane system called the Nuclear Envelope. The latter acts as a barrier between the cytoplasm and nucleoplasm, controls the transportation of all substances and macromolecules between the two and participates in organizing the nucleus's interior components. The nuclear envelope's membranes contain a large number of membrane proteins that participate in numerous protein – protein interactions. These include both proteins that form macromolecular assemblies responsible for substrate transport and nucleus organization, such as the Nuclear Pore Complexes and Nuclear Lamina, and a large number of other membrane proteins that participate in several biological processes. In fact, it has been estimated that the complexity of the nuclear envelope's proteome and its interactions rivals that of the plasma membrane. However, while other cellular components, such as the plasma membrane or mitochondria have been extensively studied, the proteomics and interactions of the nuclear envelope remain largely unexplored.

A systematic study of protein – protein interactions in the nuclear envelope was initially performed, by constructing and analyzing the protein – protein interaction network of the human nuclear envelope, using concepts from Network Theory. The envelope's interaction

network adopts features commonly observed in real biological networks and showcases the importance of nuclear envelope proteins in the eukaryotic cell, both through regulating the nucleus's functionality and through participating in a large number of other biological processes. Alongside the analysis of the human nuclear envelope NucEnvDB, a publicly available database on nuclear envelope proteins and their interactions was designed and constructed. NucEnvDB contains records on 2863 manually annotated proteins with known presence in the nuclear envelope of 394 different organisms, that participate in more than 20000 experimentally validated protein – protein interactions. At the same time, the database features a number of computational tools, enabling the study of nuclear envelope proteins through the automated creation and analysis of protein – protein interaction networks, as well as the functional annotation of their results with Gene Ontology terms, metabolic pathway associations and implications with diseases. NucEnvDB is, currently, the only publicly available source of information dedicated on the nuclear envelope, its proteins and their interactions.

Πρόλογος

Η παρούσα διδακτορική διατριβή πραγματοποιήθηκε στον Τομέα Βιολογίας Κυττάρου και Βιοφυσικής, του Τμήματος Βιολογίας του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών, υπό την επίβλεψη του **Ομότιμου Καθηγητή** **κου Σταύρου Χαμόδρακα**, και με μέλη της Τριμελούς Συμβουλευτικής Επιτροπής την **Επίκουρη Καθηγήτρια** **κα Βασιλική Οικονομίδου** και τον **Καθηγητή** **κο Παντελή Μπάγκο**.

Αρχικά, οφείλω ένα μεγάλο ευχαριστώ στον **Ομοτ. Καθηγητή** **κο Σταύρο Χαμόδρακα** για την εμπιστοσύνη που έχει δείξει απέναντί μου όλα αυτά τα χρόνια. Τον ευχαριστώ για την ιδιαίτερη τιμή να με εντάξει στο δυναμικό του Εργαστηρίου Βιοφυσικής και Βιοπληροφορικής, αρχικά σαν προπτυχιακό φοιτητή στα πλαίσια της διπλωματικής εργασίας μου, στη συνέχεια σαν μεταπτυχιακό φοιτητή του ΠΜΣ «Βιοπληροφορική» και, εσχάτως, σαν υποψήφιο διδάκτορα. Θα ήθελα να τον ευχαριστήσω επίσης γιατί τόσο μέσα από τη διδασκαλία του όσο και μέσα από την προσωπική επαφή μαζί του με δίδαξε τον τρόπο με τον οποίο πρέπει να γίνεται η έρευνα και μου έδωσε σημαντικά εφόδια για να αντιλαμβάνομαι και να επεξεργάζομαι τα διάφορα ερωτήματα που εγείρονται κατά την διάρκεια μιας επιστημονικής μελέτης. Τον ευχαριστώ θερμά, τέλος, διότι με το να με συμπεριλάβει στην ομάδα του εργαστηρίου, μου έδωσε την ευκαιρία και την τιμή να γνωρίσω και να συνεργαστώ με αρκετούς αξιόλογους ανθρώπους.

Εξίσου θερμά θέλω να ευχαριστήσω την **Επικ. Καθηγήτρια** **κα Βασιλική Οικονομίδου**, με την οποία είχα την ευκαιρία να συνεργαστώ όλα αυτά τα χρόνια, τόσο στα πλαίσια της διδακτορικής διατριβής όσο και στα γενικότερα πλαίσια του ερευνητικού έργου του Εργαστηρίου Βιοφυσικής και Βιοπληροφορικής. Την ευχαριστώ επίσης για την ηθική και υλικοτεχνική υποστήριξη της διατριβής μου, αλλά και για την πολύτιμη βοήθειά της, σε καθημερινή βάση, πάνω σε όλα τα τεχνικά και διαδικαστικά ζητήματα που προέκυψαν κατά τη διάρκεια της διατριβής.

Ένα μεγάλο ευχαριστώ οφείλω, τέλος, στον **Καθηγητή** **κο Παντελή Μπάγκο**, ο οποίος συνέβαλε σημαντικά στην απρόσκοπτη διεξαγωγή της διατριβής. Παρ'ότι δεν είχαμε συχνή επαφή λόγω απόστασης, η στήριξη και οι υποδείξεις του ήταν καθοριστικές για την πορεία του ερευνητικού έργου της διατριβής. Τον ευχαριστώ, τέλος, για τα πολύτιμα μαθήματά του στα

πλαίσια του ΠΜΣ «Βιοπληροφορική», τα οποία εμπλούτισαν την επιστημονική κατάρτισή μου στα πεδία της Βιοπληροφορικής και της Υπολογιστικής Βιολογίας.

Ιδιαίτερες ευχαριστίες εκφράζω και προς τα υπόλοιπα μέλη της Επταμελούς Εξεταστικής Επιτροπής, τον **Καθηγητή κο Κωνσταντίνο Βοργιά**, τον **Αναπληρωτή Καθηγητή κο Δημήτριο Στραβοπόδη**, τον **Καθηγητή κο Ηλία Ηλιόπουλο** και τον **Ερευνητή Β' Δρ. Γεώργιο Παυλόπουλο**. Τους ευχαριστώ για την τιμή που μου έκαναν με το να συμμετάσχουν στην Εξεταστική Επιτροπή, καθώς και για την πολύτιμη διδασκαλία τους, την οποία είχα την ευκαιρία να παρακολουθήσω, είτε σαν προπτυχιακός φοιτητής του Τμήματος Βιολογίας είτε σαν μεταπτυχιακός φοιτητής του ΠΜΣ «Βιοπληροφορική».

Επιθυμώ επίσης να ευχαριστήσω θερμά τα υπόλοιπα μέλη του **Τομέα Βιολογίας Κυττάρου & Βιοφυσικής**, την **Καθηγήτρια κα Ισιδώρα Παपाσιδέρη**, τον **Καθηγητή κο Ιωάννη Τρουγκάκο**, την **Επικ. Καθηγήτρια κα Μαριάννα Αντωνέλου**, καθώς και όλα τα **μέλη ΕΔΙΠ**, τους **μεταδιδακτορικούς ερευνητές** και τους **υποψήφιους διδάκτορες** του Τομέα, για το πνεύμα συνεργασίας και κατανόησης μεταξύ μας. Ιδιαίτερα ευχαριστώ τον **Δρ. Αθανάσιο Βελέτζα**, για τη συνεργασία μας και τη βοήθειά του σε πολλά τεχνικά ζητήματα. Ευχαριστώ επίσης την **κα Μαρίνα Αρχοντάκη**, για τη γραμματειακή υποστήριξη του Τομέα κατά τη διεξαγωγή της διδακτορικής διατριβής.

Ιδιαίτερες ευχαριστίες οφείλω προς τον **Καθηγητή κο Κωνσταντίνο Βοργιά** και τον **Καθηγητή κο Ιωάννη Τρουγκάκο**, οι οποίοι μου έδωσαν την ευκαιρία να εργαστώ στην γραμματειακή και τεχνική υποστήριξη του ΠΜΣ «Βιοπληροφορική», στηρίζοντας με αυτόν τον τρόπο και τη διεξαγωγή της διδακτορικής διατριβής μου.

Η παρουσία μου στο Εργαστήριο Βιοφυσικής και Βιοπληροφορικής μου έδωσε την ευκαιρία να γνωριστώ και να συνεργαστώ με πλήθος αξιόλογων ανθρώπων, τους οποίους πλέον θεωρώ καλούς φίλους μου. Πρώτα και κύρια, θέλω να ευχαριστήσω την **υποψήφια διδάκτορα**, συμφοιτήτρια, συνεργάτιδα και, κυρίως, φίλη μου, **κα Κατερίνα Νάστου**. Από την αρχή κιόλας της διδακτορικής διατριβής βρίσκεται δίπλα μου, έχουμε μοιραστεί άπειρες ώρες στο εργαστήριο, με έχει βοηθήσει σε επιστημονικά και μη ζητήματα και είχαμε επικοινωνία και υποδειγματική συνεργασία. Η καθημερινή επαφή μαζί της, δημιούργησε ένα ευχάριστο κλίμα μέσα στο οποίο πραγματοποιήθηκε αυτή η διδακτορική διατριβή και θέλω να την

ευχαριστήσω θερμά διότι πραγματικά συνετέλεσε στην ολοκλήρωση της. Εξίσου θερμά ευχαριστώ τη **Δρ. Παρασκευή Τσιολάκη**, για την πολύτιμη συνεργασία της σε επιστημονικό και όχι μόνο επίπεδο αλλά και για την εξίσου πολύτιμη βοήθειά της, σε όλα τα θέματα και τα προβλήματα της καθημερινότητας του εργαστηρίου. Ευχαριστώ από καρδιάς τους **Δρ. Ζωή Λίτου** και **Δρ. Νικόλαο Παπανδρέου**, οι οποίοι αποτελούν τη ζωντανή ψυχή του Εργαστηρίου Βιοφυσικής και Βιοπληροφορικής. Τους ευχαριστώ για την πολύτιμη βοήθεια και τη στήριξή τους όλα αυτά τα χρόνια, τόσο στη διεξαγωγή της διδακτορικής διατριβής όσο και στην καθημερινότητά μου. Θέλω να ευχαριστήσω επίσης τα παρελθόντα μέλη του Εργαστηρίου **Δρ. Μαργαρίτα Θεοδωροπούλου, Δρ. Γεώργιο Τσαούση και Δρ. Νικόλαο Λούρο** για την άψογη συνεργασία μας. Ιδιαίτερα ευχαριστώ τη **Μαργαρίτα**, με την οποία συνεργαστήκαμε ενεργά, τόσο σε προπτυχιακό επίπεδο όσο και στα πρώτα στάδια της διδακτορικής διατριβής μου, με τα αποτελέσματα αυτής της συνεργασίας να διαμορφώνουν, σε μεγάλο βαθμό, και την πορεία που ακολούθησε αυτή η διατριβή στη συνέχεια. Ευχαριστίες οφείλω και στις υποψ. Διδάκτορες **Γεωργία Νάση** και **Αυγή Αποστολάκου**, για την καθημερινή συνεργασία και αλληλεπίδρασή μας. Τέλος, θέλω να ευχαριστήσω όλους τους φοιτητές, προπτυχιακούς και μεταπτυχιακούς, με τους οποίους είχα την ευκαιρία να συνεργαστώ, συγκεκριμένα τους **Μαρίνα Γαβριηλίδου, Γεωργία Σταυρουλάκη, Δημήτρη Σοφρά, Γιώργο Σπαννογιάννη, Ανδρέα Ζαμάνο, Μανώλη Οικονόμου, Κατερίνα Ρίζου και Γιώργο Φίλη**. Συνολικά, ευχαριστώ όλα τα μέλη του εργαστηρίου για τη συνεργασία τους, τη φιλική διάθεσή τους αλλά και την ανεκτικότητα που επέδειξαν απέναντι στις ιδιοτροπίες μου.

Τέλος, και πάνω από όλα, θέλω να ευχαριστήσω **την οικογένειά μου**, για την αγάπη και τη συμπαράσταση που μου προσέφεραν και συνεχίζουν να μου προσφέρουν, αλλά και επειδή συνέβαλαν στο να γίνω ο άνθρωπος που είμαι σήμερα. Ιδιαίτερα θέλω να ευχαριστήσω τον αδερφό μου, **Βασίλη**, για όλη τη βοήθεια και τη στήριξη που μου έχει προσφέρει όλα αυτά τα χρόνια.

Με εκτίμηση,

Φώτης

Μέρος της παρούσας έρευνας έχει συγχρηματοδοτηθεί από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Ταμείο Περιφερειακής Ανάπτυξης - ΕΤΠΑ) και από εθνικούς πόρους μέσω του Επιχειρησιακού Προγράμματος «Ανταγωνιστικότητα & Επιχειρηματικότητα» του Εθνικού Στρατηγικού Πλαισίου Αναφοράς (ΕΣΠΑ) (Κωδικός Έργου 09ΣΥΝ-13-999).

Η παρούσα έρευνα έχει υποστηριχθεί από το Εθνικό Δίκτυο Έρευνας και Τεχνολογίας (ΕΔΕΤ), μέσα από την παροχή υπολογιστικού χρόνου στο υπολογιστικό σύστημα υψηλών επιδόσεων ARIS (Κωδικοί Έργων: PR001025-M.D.S.B.M.S., PR002041-S.C.S.M.P. και PR004006-BioMemPro-MD).

Περιεχόμενα

Περίληψη.....	VII
Abstract	XVI
Πρόλογος.....	XXV
Περιεχόμενα.....	XXIX
Ευρετήριο Πινάκων.....	XXXIII
Ευρετήριο Εικόνων.....	XXXIV
ΚΕΦΑΛΑΙΟ Ι - ΕΙΣΑΓΩΓΗ	1
1.1 Βιολογικές Μembrάνες	1
1.2 Τα λιπίδια των μεμβρανών	3
1.2.1 Φωσφολιπίδια	4
1.2.2 Σφινγολιπίδια.....	7
1.2.3 Στερόλες.....	8
1.2.4 Λιποπολυσακχαρίτες	9
1.2.5 Η οργάνωση των λιπιδίων στη μεμβράνη	14
1.3 Οι μεμβρανικές πρωτεΐνες και οι κατηγορίες τους	17
1.3.1 Διαμεμβρανικές πρωτεΐνες	18
1.3.1.1 Οι α-ελικοειδείς διαμεμβρανικές πρωτεΐνες	19
1.3.1.2 Τα διαμεμβρανικά β-βαρέλια.....	23
1.3.2 Προσδιορισμός της δομής των διαμεμβρανικών πρωτεϊνών	27
1.4 Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών	30
1.5 Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών σε διαμεμβρανικές πρωτεΐνες.....	34
1.5.1 Συζευγμένοι με G-πρωτεΐνες Υποδοχείς (GPCRs).....	34
1.5.1.1 Ταξινόμηση των GPCRs.....	36
1.5.1.2 Η τοπολογία και η δομή των GPCRs	37
1.5.1.3 Οι ετεροτρίμερες G-πρωτεΐνες	41
1.5.1.4 Αλληλεπιδράσεις ολιγομερισμού στους GPCRs	43
1.5.2 Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs).....	46
1.5.2.1 Δομή και ταξινόμηση των RTKs	48
1.5.2.2 Ο μηχανισμός λειτουργίας των RTKs.....	51
1.5.3 Αλληλεπιδράσεις ολιγομερισμού στα διαμεμβρανικά β-βαρέλια	54
1.5.4 Οι πρωτεΐνες του πυρηνικού φακέλου	57

1.5.4.1 Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο	60
1.5.5 Οι διαμεμβρανικοί μεταφορείς SLC11	61
1.6 Οι επιρροές της μεμβράνης στις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών.....	65
1.7 Στόχοι της διδακτορικής διατριβής.....	69
ΚΕΦΑΛΑΙΟ II – ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	71
2.1 Προσομοιώσεις Μοριακής Δυναμικής	71
2.1.1 Μέθοδοι ελαχιστοποίησης ενέργειας	72
2.1.2 Ο βασικός αλγόριθμος των προσομοιώσεων Μοριακής Δυναμικής.....	74
2.1.3 Υλοποίηση του αλγορίθμου των προσομοιώσεων Μοριακής Δυναμικής στη μονάδα του χρόνου	76
2.1.4 Συνθήκες περιοδικών ορίων.....	77
2.1.5 Προτυποποίηση θερμοδυναμικών κατανομών	79
2.1.5.1 Αλγόριθμοι ελέγχου της θερμοκρασίας	80
2.1.5.2 Αλγόριθμοι ελέγχου της πίεσης	85
2.1.6 Συναρτήσεις επιμερισμού και θερμοδυναμικές ιδιότητες	89
2.1.7 Δυναμικά Πεδία.....	91
2.1.7.1 Ομοιοπολικές αλληλεπιδράσεις.....	93
2.1.7.2 Μη-ομοιοπολικές αλληλεπιδράσεις.....	94
2.1.8 Δυναμικά πεδία και επίπεδα διακριτικότητας.....	98
2.1.8.1 Το δυναμικό πεδίο πλήρους ατομικής διακριτικότητας CHARMM	100
2.1.8.2 Το δυναμικό πεδίο ενοποιημένων ατόμων GROMOS.....	101
2.1.8.3 Το δυναμικό πεδίο αδρής διακριτικότητας MARTINI	101
2.1.8.4 Το υβριδικό δυναμικό πεδίο PACE	106
2.1.9 Προσομοιώσεις Μοριακής Δυναμικής για διαμεμβρανικές πρωτεΐνες	107
2.1.10 Προσδιορισμός ιδιοτήτων σε αποτελέσματα προσομοιώσεων Μοριακής Δυναμικής.....	114
2.1.11 Καθοδηγούμενη Μοριακή Δυναμική	119
2.1.12 Υπολογισμοί Ελεύθερης Ενέργειας	120
2.1.12.1 Το Δυναμικό Μέσης Δύναμης	122
2.1.12.2 Διαταραχή Ελεύθερης Ενέργειας	128
2.1.12.3 Υπολογιστική Σάρωση Αλανίνης.....	131
2.1.12.4 Ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας.....	132
2.1.13 Προγράμματα προσομοιώσεων Μοριακής Δυναμικής	134
2.2 Δίκτυα Αλληλεπιδράσεων.....	138

2.2.1 Θεωρία Γράφων.....	139
2.2.2 Κατηγορίες Γράφων.....	140
2.2.3 Ιδιότητες δικτύων αλληλεπιδράσεων	141
2.2.4 Αλγόριθμοι ομαδοποίησης.....	146
2.2.5 Ανάλυση λειτουργικού εμπλουτισμού	148
2.2.6 Ανάλυση δυναμικών δικτύων αλληλεπιδράσεων σε προσομοιώσεις Μοριακής Δυναμικής	149
2.2.7 Προγράμματα οπτικοποίησης και ανάλυσης δικτύων αλληλεπιδράσεων.....	150
2.2.8 Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων	151
2.3 Κρυφά Μοντέλα Markov	154
2.3.1 Αλυσίδες Markov και Κρυφά Μοντέλα Markov (HMMs).....	154
2.3.2 Προφίλ Hidden Markov Models (pHMMs)	158
2.3.3 Κατασκευή και χρήση των pHMMs στην Υπολογιστική Βιολογία.....	160
2.3.3.1 Το πακέτο λογισμικού HMMER	161
2.3.3.2 Η βάση δεδομένων Pfam.....	162
2.3.3.3 Η διαδικασία εκπαίδευσης ενός pHMM	163
2.4 Αλληλεπιδράσεις ολιγομερισμού σε Συζευγμένους με G-πρωτεΐνες Υποδοχείς.....	166
2.4.1 Μελέτη των αλληλεπιδράσεων ολιγομερισμού των GPCRs με προσομοιώσεις Μοριακής Δυναμικής	166
2.4.2 Θερμοδυναμική ανάλυση των αλληλεπιδράσεων ολιγομερισμού	169
2.4.3 Μελέτη του αυθόρμητου διμερισμού των υποδοχέων A2A και mGluR5.....	171
2.5 Εξωτερικές Μεμβράνες των αρνητικών κατά Gram βακτηρίων	172
2.5.1 Κατασκευή μεθόδου για την προτυποποίηση εξωτερικών μεμβρανών: η μέθοδος προτυποποίησης GNOMM	172
2.5.2 Εφαρμογή του GNOMM σε προσομοιώσεις Εξωτερικών Μεμβρανών	179
2.5.3 Μελέτη των αλληλεπιδράσεων ολιγομερισμού σε διαμεμβρανικά β-βαρέλια	180
2.6 Ο διαμεμβρανικός μεταφορέας SLC11A1	182
2.7 Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs)	182
2.7.1 Κατασκευή μεθόδου πρόγνωσης και σχολιασμού των RTKs: ο αλγόριθμος RTK-PRED	182
2.7.2 Αξιολόγηση της μεθόδου RTK-PRED.....	187
2.8 Μεμβρανικές πρωτεΐνες του Πυρηνικού Φακέλου.....	189
2.8.1 Κατασκευή δικτύου αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου στον άνθρωπο	189

2.8.2 Κατασκευή βάσης δεδομένων για τις πρωτεΐνες του πυρηνικού φακέλου: η βάση δεδομένων NucEnvDB.....	192
ΚΕΦΑΛΑΙΟ III – ΑΠΟΤΕΛΕΣΜΑΤΑ-ΣΥΖΗΤΗΣΗ.....	195
3.1 Αλληλεπιδράσεις ολιγομερισμού των GPCRs.....	196
3.1.1 Δομική και δυναμική ανάλυση των αλληλεπιδράσεων ολιγομερισμού των GPCRs	196
3.1.2 Ενεργειακή ανάλυση και θερμοδυναμική σταθερότητα των αλληλεπιδράσεων ολιγομερισμού των GPCRs	206
3.1.3 Η επίδραση της χοληστερόλης στον αυθόρμητο σχηματισμό διμερών GPCRs.....	212
3.2 Εξωτερικές Μembrάνες των αρνητικών κατά Gram βακτηρίων	221
3.2.1 Η μέθοδος προτυποποίησης Εξωτερικών Μembrανών GNOMM.....	221
3.2.2 Μελέτη των αλληλεπιδράσεων ολιγομερισμού σε διαμεμβρανικά β-βαρέλια	230
3.3 Ο μεταφορέας δισθενών μεταλλικών ιόντων SLC11A1.....	237
3.4 Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs)	244
3.4.1 Ο αλγόριθμος πρόγνωσης RTK-PRED	244
3.4.2 Εντοπισμός πιθανών RTKs σε ευκαρυωτικά πρωτεώματα αναφοράς.....	247
3.5 Πρωτεΐνες του Πυρηνικού Φακέλου	251
3.5.1 Ανάλυση του δικτύου αλληλεπιδράσεων των πρωτεϊνών του ανθρώπινου Πυρηνικού Φακέλου	251
3.5.2 Η βάση δεδομένων NucEnvDB	266
ΚΕΦΑΛΑΙΟ IV – ΣΥΜΠΕΡΑΣΜΑΤΑ	277
4.1 Μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs.....	277
4.2 Μελέτη της Εξωτερικής Μembrάνης των αρνητικών κατά Gram βακτηρίων	280
4.3 Μελέτη της δομής του διαμεμβρανικού μεταφορέα SLC11A1.....	283
4.4 Κατασκευή μεθόδου πρόγνωσης και σχολιασμού των RTKs	284
4.5 Μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο	284
4.6 Μελλοντικές κατευθύνσεις.....	285
ΒΙΒΛΙΟΓΡΑΦΙΑ	289
ΠΑΡΑΡΤΗΜΑΤΑ	307
Παράρτημα I – Σύντομο Βιογραφικό	307
Παράρτημα II – Λίστα Υπολογιστικών Εργαλείων και Δημοσιεύσεων.....	309
Παράρτημα III – Δημοσιεύσεις των αποτελεσμάτων της διατριβής σε επιστημονικά περιοδικά	311

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 1 Παραδείγματα ασθενειών με τις οποίες έχουν συσχετιστεί οι λιποπολυσακχαρίτες διάφορων αρνητικών κατά Gram βακτηρίων.13

Πίνακας 2 Παραδείγματα υπολογιστικών μεθόδων πρόγνωσης της τοπολογίας των διαμεμβρανικών πρωτεϊνών.30

Πίνακας 3 Παραδείγματα ολιγομερών GPCRs τα οποία έχουν συσχετιστεί με ασθένειες.44

Πίνακας 4 Ταξινόμηση των RTKs με βάση το σύστημα της IUPHAR.51

Πίνακας 5 Παράμετροι προσομοιώσεων Μοριακής Δυναμικής για τα δυναμικά πεδία CHARMM, GROMOS, MARTINI και PACE, οι οποίες χρησιμοποιήθηκαν στα πλαίσια της παρούσας διδακτορικής διατριβής.112

Πίνακας 6 Τα διαθέσιμα λιπίδια του GNOMM177

Πίνακας 7 Η βιβλιοθήκη των rHMMs της Pfam τα οποία χρησιμοποιούνται στη μέθοδο RTK-PRED186

Πίνακας 8 Οι υποκυτταρικές θέσεις που ορίζονται από τη UniProt για τον πυρηνικό φάκελο.190

Πίνακας 9 Σύνοψη των αποτελεσμάτων των προσομοιώσεων Μοριακής Δυναμικής (MD) για τα διμερή GPCRs.198

Πίνακας 10 Αποτελέσματα προσομοιώσεων US για τα διμερή των β1AR, μOR και Ροδοψίνης.209

Πίνακας 11 Μετρήσεις του συντελεστή πλευρικής διάχυσης (D_{lat}) στις προσομοιώσεις διμερισμού των A2A και mGluR5.218

Πίνακας 12 Μετρήσεις της επιφάνειας ανά λιπίδιο και του πάχους της λιπιδικής διπλοστιβάδας από τα αποτελέσματα των προσομοιώσεων για τις ομοιογενείς μεμβράνες Λιπιδίου A.224

Πίνακας 13 Τιμές επιφάνειας ανά λιπίδιο σε $\text{\AA}^2$ για τα λιπίδια των μικτών μεμβρανών.226

Πίνακας 14 Αποτελέσματα ενεργειακών μετρήσεων (FEP) για τον μεταφορέα SLC11A1 με ιόντα Fe^{2+} και Mn^{2+}240

Πίνακας 15 Αξιολόγηση της μεθόδου RTK-PRED ως προς τον εντοπισμό RTKs σε ένα σύνολο ελέγχου 716 πρωτεϊνών.244

Πίνακας 16 Αξιολόγηση της ικανότητας του RTK-PRED στην κατηγοριοποίηση των RTKs στο σύστημα της IUPHAR, με βάση ένα σύνολο ελέγχου 269 καλά σχολιασμένων RTKs.245

Πίνακας 17 Αποτελέσματα από την εφαρμογή της μεθόδου RTK-PRED σε 37 πρωτεώματα αναφοράς.248

Πίνακας 18 Λειτουργική κατηγοριοποίηση των πιθανών RTKs από τα πρωτεώματα αναφοράς.249

Πίνακας 19 Ανάλυση των ιδιοτήτων του δικτύου με βάση τη Θεωρία Γράφων.255

Πίνακας 20 Λίστα με τους 10 κυριότερους κεντρικούς κόμβους (hubs) και τους 10 κυριότερους κόμβους-στενωπούς (bottlenecks) του δικτύου.256

Πίνακας 21 Λειτουργικός σχολιασμός των ομάδων του δικτύου με όρους Γονιδιακής Οντολογίας.262

ΕΥΡΕΤΗΡΙΟ ΕΙΚΟΝΩΝ

Εικόνα 1 Οι βιολογικές μεμβράνες των κυττάρων.....	2
Εικόνα 2 Χαρακτηριστικά παραδείγματα φωσφολιπιδίων.	5
Εικόνα 3 Χαρακτηριστικά παραδείγματα σφιγγολιπιδίων.	7
Εικόνα 4 Οι στερόλες των βιολογικών μεμβρανών..	8
Εικόνα 5 Οι μεμβράνες των αρνητικών κατά Gram βακτηρίων.	10
Εικόνα 6 Δομές του Λιπιδίου Α τεσσάρων αρνητικών κατά Gram βακτηρίων.	11
Εικόνα 7 Συμπεριφορά φάσης των λιπιδίων..	15
Εικόνα 8 Οι κατηγορίες των μεμβρανικών πρωτεϊνών.....	17
Εικόνα 9 Οι τέσσερις τύποι διαμεμβρανικών πρωτεϊνών με ένα διαμεμβρανικό τμήμα.....	20
Εικόνα 10 Παραδείγματα α-ελικοειδών διαμεμβρανικών πρωτεϊνών με περισσότερα του ενός διαμεμβρανικά τμήματα..	21
Εικόνα 11 Παραδείγματα διαμεμβρανικών β-βαρελίων.....	24
Εικόνα 12 Γεωμετρικά κριτήρια της δομής των β-βαρελίων.	25
Εικόνα 13 Η μέθοδος της λιπιδικής μεσόφασης για την κρυστάλλωση διαμεμβρανικών πρωτεϊνών.....	29
Εικόνα 14 Τα σηματοδοτικά μονοπάτια των GPCRs.	35
Εικόνα 15 Η συντηρημένη τρισδιάστατη δομή των GPCRs.	38
Εικόνα 16 Στερεοδιαταξές αλλαγές κατά την ενεργοποίηση των GPCRs.....	40
Εικόνα 17 Η δομή των ετεροτριμερών G-πρωτεϊνών.....	42
Εικόνα 18 Η τρισδιάστατη δομή των RTKs.....	48
Εικόνα 19 Το σύστημα ταξινόμησης των RTKs.	50
Εικόνα 20 Σηματοδοτικά μονοπάτια των RTKs..	53

Εικόνα 21 Παραδείγματα ολιγομερισμού στα διαμεμβρανικά β-βαρέλια.	55
Εικόνα 22 Η δομή του πυρηνικού φακέλου ...	58
Εικόνα 23 Οι διαμεμβρανικοί μεταφορείς SLC11.	63
Εικόνα 24 Προσαρμογές στο φαινόμενο της υδροφοβικής ασυμφωνίας.	67
Εικόνα 25 Η πορεία της ελαχιστοποίησης ενέργειας.	73
Εικόνα 26 Συνθήκες περιοδικών ορίων.	78
Εικόνα 27 Διαγραμματική απεικόνιση των θερμοδυναμικών κατανομών.	80
Εικόνα 28 Οι τρεις διαφορετικοί τρόποι ελέγχου της πίεσης.....	86
Εικόνα 29 Οι ομοιοπολικές αλληλεπιδράσεις στη συνάρτηση δυναμικής ενέργειας.	92
Εικόνα 30 Το δυναμικό Lennard-Jones.	95
Εικόνα 31 Κατηγορίες δυναμικών πεδίων.....	99
Εικόνα 32 Το δυναμικό πεδίο MARTINI.	103
Εικόνα 33 Προσανατολισμός μιας διαμεμβρανικής πρωτεΐνης στο επίπεδο της μεμβράνης.....	108
Εικόνα 34 Η δομή ενός συστήματος προσομοιώσεων Μοριακής Δυναμικής.	110
Εικόνα 35 Το Δυναμικό Μέσης Δύναμης (PMF)..	123
Εικόνα 36 Η μέθοδος της Δειγματοληψίας Ομπρέλας (Umbrella Sampling).....	126
Εικόνα 37 Ο θερμοδυναμικός κύκλος του υπολογισμού της ενέργειας πρόσδεσης σε ένα σύμπλοκο υποδοχέα-προσδέτη με τη μέθοδο FEP.	129
Εικόνα 38 Το δίκτυο αλληλεπιδράσεων πρωτεϊνών–πρωτεϊνών της <i>Escherichia coli</i> . .	138
Εικόνα 39 Οι τέσσερις βασικές κατηγορίες των γράφων.....	139
Εικόνα 40 Σύγκριση του μοντέλου ανεξαρτησίας με τις αλυσίδες Markov και τα HMMs.	155

Εικόνα 41 Σχηματική αναπαράσταση ενός τυπικού rHMM..	159
Εικόνα 42 Η τυπική διαδικασία κατασκευής ενός rHMM.....	164
Εικόνα 43 Η πορεία εργασίας στις προσομοιώσεις Μοριακής Δυναμικής των διμερών GPCRs.	167
Εικόνα 44 Υπολογισμοί του Δυναμικού Μέσης Δύναμης σε διμερή GPCRs.....	170
Εικόνα 45 Η πορεία εργασίας του GNOMM..	173
Εικόνα 46 Προτυποποίηση της μεμβράνης γύρω από ένα β-βαρέλι με το GNOMM..	175
Εικόνα 47 Η πορεία εργασίας του RTK-PRED...	183
Εικόνα 48 Χαρακτηριστικά παραδείγματα διμερών GPCRs.	197
Εικόνα 49 Αποτελέσματα των προσομοιώσεων Μοριακής Δυναμικής για τα διμερή των GPCRs.	200
Εικόνα 50 Απεικόνιση των σημαντικών καταλοίπων αλληλεπίδρασης στα διμερή των GPCRs..	202
Εικόνα 51 Σημαντικές κοινότητες στα δομικά δίκτυα αλληλεπιδράσεων των διμερών GPCRs.	204
Εικόνα 52 Δομές διμερών των GPCRs που μελετήθηκαν με ενεργειακούς υπολογισμούς..	208
Εικόνα 53 Θερμοδυναμική ανάλυση των αλληλεπιδράσεων διμερισμού των GPCRs..	211
Εικόνα 54 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ομοδιμερισμό του υποδοχέα A2A.	214
Εικόνα 55 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ομοδιμερισμό του υποδοχέα mGluR5.	215
Εικόνα 56 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ετεροδιμερισμό των υποδοχέων A2A και mGluR5.	217
Εικόνα 57 Η ιστοσελίδα της μεθόδου GNOMM.	221
Εικόνα 58 Η σελίδα προτυποποίησης του συστήματος προσομοίωσης.	222
Εικόνα 59 Προσομοιώσεις Μοριακής Δυναμικής σε μεμβράνες Λιπιδίου A.	223

Εικόνα 60 Αποτελέσματα προσομοιώσεων των μικτών μεμβρανών.	225
Εικόνα 61 Προσομοιώσεις Μοριακής Δυναμικής για το β-βαρέλι OmpA.....	227
Εικόνα 62 Προσομοιώσεις Μοριακής Δυναμικής για το β-βαρέλι OmpF.	228
Εικόνα 63 Προσομοιώσεις Μοριακής Δυναμικής των ομοτριμερών β-βαρελίων..	232
Εικόνα 64 Υπολογισμοί του Δυναμικού Μέσης Δύναμης για την OmpF σε διαφορετικές μεμβράνες.	235
Εικόνα 65 Η προτεινόμενη δομή του μεταφορέα SLC11A1.....	238
Εικόνα 66 Προσομοιώσεις Μοριακής Δυναμικής του SLC11A1.	240
Εικόνα 67 Η ιστοσελίδα της μεθόδου RTK-PRED.....	246
Εικόνα 68 Οι πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου.	251
Εικόνα 69 Το δίκτυο αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου.....	254
Εικόνα 70 Λειτουργικός σχολιασμός του δικτύου αλληλεπιδράσεων..	257
Εικόνα 71 Ομαδοποίηση των στοιχείων του δικτύου με βάση την υποκυτταρική θέση των πρωτεϊνών.	259
Εικόνα 72 Ομαδοποίηση των στοιχείων του δικτύου με τον αλγόριθμο MCL.	261
Εικόνα 73 Αναζήτηση πληροφοριών στη βάση δεδομένων NucEnnDB.	267
Εικόνα 74 Η βάση δεδομένων NucEnnDB....	268
Εικόνα 75 Η βάση δεδομένων NucEnnDB (2)..	269
Εικόνα 76 Αναζήτηση ομολογίας στη βάση NucEnnDB.	271
Εικόνα 77 Κατασκευή δικτύου αλληλεπιδράσεων μέσω της NucEnnDB.	272
Εικόνα 78 Αποτελέσματα της ανάλυσης δικτύου.	273
Εικόνα 79 Ανάλυση λειτουργικού εμπλουτισμού μέσω της NucEnnDB.....	274

ΚΕΦΑΛΑΙΟ I

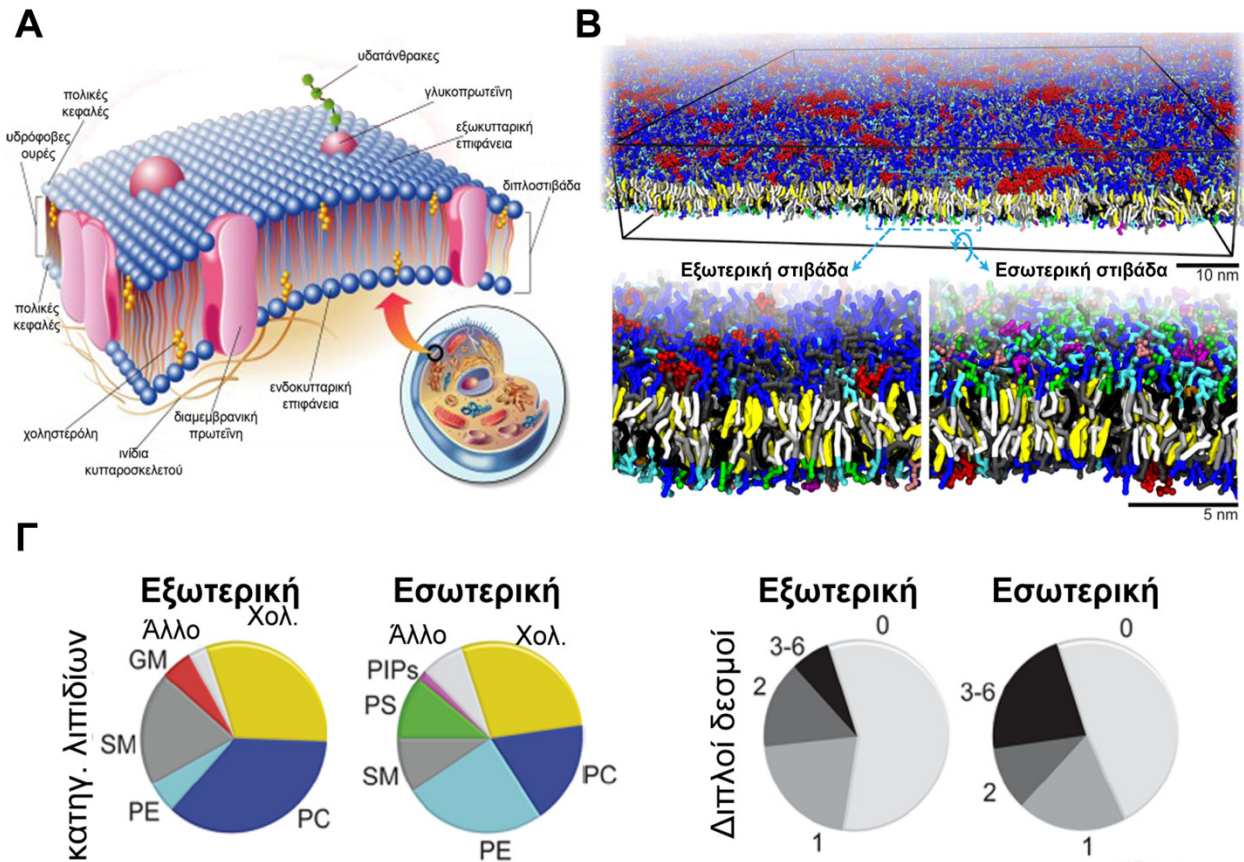
ΕΙΣΑΓΩΓΗ

1.1 Βιολογικές Μembrάνες

Οι βιολογικές μεμβράνες είναι βασικά δομικά στοιχεία όλων των κυττάρων. Λειτουργούν τόσο ως μηχανισμοί απομόνωσης και προστασίας του κυττάρου, όσο και ως εξειδικευμένα όργανα επικοινωνίας και αλληλεπίδρασης του κυττάρου με το περιβάλλον του. Η ισχύουσα, αποδεκτή απεικόνιση της δομής των μεμβρανών είναι το μοντέλο του «ρευστού μωσαϊκού» (Εικόνα 1). Σύμφωνα με αυτό το μοντέλο όλες οι μεμβράνες είναι συνεχείς, ρευστές διπλοστιβάδες αποτελούμενες από λιπίδια, στην επιφάνεια ή το εσωτερικό των οποίων εντοπίζονται διάφορων ειδών πρωτεΐνες (Singer & Nicolson, 1972). Η πιο κοινή βιολογική μεμβράνη είναι η κυτταρική ή πλασματική μεμβράνη, η οποία περιβάλλει το σύνολο του κυττάρου και εντοπίζεται τόσο στους ευκαρυωτικούς όσο και στους προκαρυωτικούς οργανισμούς. Η κυτταρική μεμβράνη παρέχει στο κύτταρο τη δομή και τη συνοχή του, ελέγχει τη μεταφορά ουσιών από και προς το κύτταρο, διατηρεί σταθερό το δυναμικό του κυττάρου, ρυθμίζει τη μεταγωγή σήματος και συμβάλλει στη διακυτταρική επικοινωνία (Ingolfsson et al., 2014). Εκτός όμως από την κυτταρική μεμβράνη, όμως, υπάρχουν πολλές άλλες κατηγορίες βιολογικών μεμβρανών, οι οποίες επιτελούν εξίσου σημαντικούς ρόλους (Alberts et al., 2002). Έτσι, τα ευκαρυωτικά κύτταρα διαθέτουν πλήθος διαφορετικών μεμβρανών, όπως οι μεμβράνες των μιτοχονδρίων και των χλωροπλαστών, το Ενδοπλασματικό Δίκτυο, το σύστημα Golgi και ο πυρηνικός φάκελος, οι οποίες συμβάλλουν στη διαμερισματοποίηση του κυττάρου και, παράλληλα, ρυθμίζουν μεγάλο εύρος των λειτουργιών του (van Meer et al., 2008). Μια άλλη κατηγορία μεμβρανών με ξεχωριστό ρόλο είναι οι Εξωτερικές Μembrάνες των αρνητικών κατά Gram βακτηρίων, οι οποίες αποτελούν ένα επιπλέον μεμβρανικό στρώμα γύρω από τα

Κεφάλαιο I - Εισαγωγή

κύτταρά τους και τους προσδίδουν μια σειρά από επιπρόσθετες ιδιότητες, όπως είναι η προστασία από αντιμικροβιακούς παράγοντες (Bos et al., 2007). Τέλος, μια ξεχωριστή κατηγορία μεμβρανικών σχηματισμών είναι τα εξωκυτταρικά κυστίδια, τα οποία παράγονται από τα κύτταρα και χρησιμοποιούνται για τη μεταφορά ουσιών ανάμεσα σε διαφορετικούς ιστούς του σώματος (Alberts et al., 2002).



Εικόνα 1 Οι βιολογικές μεμβράνες των κυττάρων. (Α) Το μοντέλο του «ρευστού μωσαϊκού» των μεμβρανών. Διακρίνονται τα φωσfolιπίδια, οι στερόλες και οι μεμβρανικές πρωτεΐνες της διπλοστιβάδας. (Β) Μοριακό μοντέλο μιας «τυπικής» κυτταρικής μεμβράνης, βασισμένο σε πειραματικά δεδομένα. Οι διαφορετικοί τύποι λιπιδίων απεικονίζονται με διαφορετικό χρώμα. (Γ) Η λιπιδική σύσταση μιας τυπικής κυτταρικής μεμβράνης. Δίνονται οι κατανομές των λιπιδίων στην εξωτερική και την εσωτερική στιβάδα με βάση τον τύπο του λιπιδίου (αριστερά) και τον αριθμό των διπλών δεσμών στις λιπιδικές ουρές (δεξιά). Προσαρμογή από Ingolfsson et al., 2014.

1.2 Τα λιπίδια των μεμβρανών

Το βασικό συστατικό μιας μεμβράνης είναι τα λιπίδια που την αποτελούν. Τα λιπίδια μπορούν να είναι διαφόρων ειδών, με όλες τις κατηγορίες να έχουν σαν κοινό χαρακτηριστικό την ύπαρξη μιας υδρόφιλης (πολικής ή φορτισμένης) «κεφαλής» και ενός ή περισσότερων υδρόφοβων τμημάτων («ουρές»). Αυτή η διττή φύση των λιπιδίων είναι υπεύθυνη για τη βιογένεση και τη δομή των μεμβρανών (van Meer et al., 2008). Κατά την επαφή τους με τον υδατικό διαλύτη, τα λιπίδια συσσωματώνονται σχηματίζοντας τη διπλοστιβάδα της μεμβράνης, έτσι ώστε οι πολικές κεφαλές τους να έρχονται σε επαφή με το διαλύτη ενώ οι υδρόφοβες ουρές τους να θάβονται στον πυρήνα της διπλοστιβάδας. Αυτή η συσσωμάτωση και οργάνωση των λιπιδίων προσδίδει στις μεμβράνες την ιδιαίτερη δομή τους και καθορίζει σε μεγάλο βαθμό τόσο τα χαρακτηριστικά της λιπιδικής διπλοστιβάδας, όσο και τα χαρακτηριστικά των μεμβρανικών πρωτεϊνών (Alberts et al., 2002; Λέκκα et al., 2015).

Η λιπιδική σύσταση της μεμβράνης είναι ο βασικός παράγοντας που καθορίζει ιδιότητες όπως η ελαστικότητα της διπλοστιβάδας, η διαπερατότητά της από ουσίες και η πλευρική διάχυση των συστατικών της. Οι πραγματικές βιολογικές μεμβράνες σπάνια αποτελούνται από ένα ή λίγα είδη λιπιδίων. Αντίθετα, είναι υψηλά ετερογενείς δομές, με τα υπάρχοντα δεδομένα να υποδεικνύουν ότι μια «τυπική» κυτταρική μεμβράνη μπορεί να διαθέτει πάνω από 60 διαφορετικά λιπίδια, σε διάφορες αναλογίες (Εικόνα 1). Είναι σημαντικό να σημειωθεί ότι, παρ'όλο που η γενική δομή της λιπιδικής διπλοστιβάδας είναι κοινή, δεν διαθέτουν όλες οι μεμβράνες την ίδια σύσταση. Αντίθετα, έχει δειχθεί πειραματικά ότι οι πλασματικές μεμβράνες διαφορετικών κυτταρικών τύπων παρουσιάζουν αντίστοιχες διαφορές στη λιπιδική τους σύσταση (Bos et al., 2007). Στην περίπτωση των προκαρυωτικών οργανισμών, οι πλασματικές μεμβράνες διαφορετικών ειδών διαφοροποιούνται ως προς τη σύστασή τους, περιέχοντας λιπίδια τα οποία συμβάλλουν στη βέλτιστη δυνατή προσαρμογή του κυττάρου στις συνθήκες στις οποίες ευδοκιμεί (Sohlenkamp & Geiger, 2016). Ακόμη μεγαλύτερη διαφοροποίηση παρατηρείται στην περίπτωση των ευκαρυωτικών οργανισμών, όπου η σύσταση της κυτταρικής μεμβράνης είναι διαφορετική από τη σύσταση των μεμβρανών των οργανιδίων, τα οποία συχνά περιέχουν λιπίδια που εντοπίζονται αποκλειστικά σε αυτά. Η σύσταση των δύο στρωμάτων της διπλοστιβάδας διαφοροποιείται επίσης, με

Κεφάλαιο I - Εισαγωγή

κάποια λιπίδια να εμφανίζονται και στις δύο στιβάδες της μεμβράνης, ενώ κάποια άλλα να εντοπίζονται αποκλειστικά στην κυτοπλασματική ή αποκλειστικά στην εξωκυτταρική όψη της (Ingolfsson et al., 2014; van Meer et al., 2008).

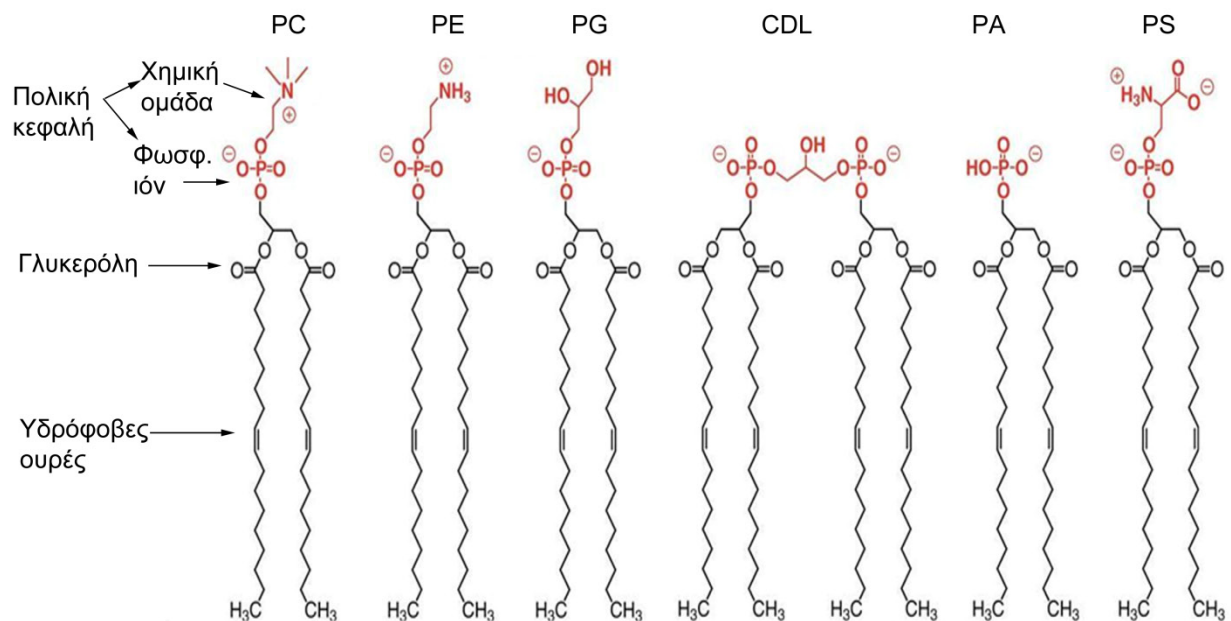
1.2.1 Φωσφολιπίδια

Η πιο άφθονη κατηγορία λιπιδίων στις μεμβράνες είναι τα *γλυκερο-φωσφολιπίδια*, ή απλά *φωσφολιπίδια*. Τα φωσφολιπίδια είναι αμφιπαθή μόρια, αποτελούμενα από μια πολική κεφαλή με ένα φωσφορικό ιόν (PO_4^{2-}) και δύο λιπαρά οξέα, που δομούν τις υδρόφοβες ουρές. Οι ουρές και το φωσφορικό ιόν συνδέονται ομοιοπολικά σε ένα μόριο γλυκερόλης, το οποίο λειτουργεί σαν σύνδεσμος («linker») ανάμεσα στο πολικό και το υδρόφοβο τμήμα του λιπιδίου. Στο ελεύθερο οξυγόνο του φωσφορικού ιόντος προσδένεται συχνά μια επιπρόσθετη χημική ομάδα (Alberts et al., 2002).

Τα φωσφολιπίδια διακρίνονται μεταξύ τους με βάση δύο κύρια χαρακτηριστικά, το είδος της πολικής κεφαλής και το είδος των υδρόφοβων ουρών (Bos et al., 2007). Η πολική κεφαλή καθορίζει την κατηγορία του φωσφολιπιδίου. Οι κύριες κατηγορίες των φωσφολιπιδίων διακρίνονται από το είδος της χημικής ομάδας που προσδένεται στο φωσφορικό ιόν (χολίνη, αιθανολαμίνη, γλυκερόλη, σερίνη κλπ), και η οποία καθορίζει ιδιότητες όπως το συνολικό φορτίο της πολικής κεφαλής και την παρουσία ή απουσία δοτών ή δεκτών για δεσμούς υδρογόνου και γέφυρες άλατος (Εικόνα 2). Η προσθήκη θετικά φορτισμένων ομάδων όπως η χολίνη ή η αιθανολαμίνη στο φωσφορικό ιόν οδηγούν σε πολικές αλλά αφόρτιστες λιπιδικές κεφαλές. Αντίθετα, η προσθήκη αφόρτιστων ομάδων όπως η γλυκερόλη και η σερίνη διατηρούν το αρνητικό φορτίο των κεφαλών. Σε άλλες περιπτώσεις, στη θέση της χημικής ομάδας προσδένονται σάκχαρα (είτε μονοσακχαρίτες είτε αλυσίδες ολιγο- ή πολυσακχαριτών), οδηγώντας στο σχηματισμό γλυκολιπιδίων (γλυκοφωσφολιπίδια). Στην περίπτωση των φωσφατιδικών οξέων, τέλος, δεν προσδένεται κάποια χημική ομάδα, και έτσι η πολική κεφαλή αποτελείται αποκλειστικά από το φωσφορικό ιόν, με το φορτίο του να μεταβάλλεται μέσω πρωτονίωσης ανάλογα με τις περιβαλλοντικές συνθήκες. Η αναλογία αφόρτιστων και φορτισμένων πολικών κεφαλών σε κάθε πλευρά της διπλοστιβάδας

Κεφάλαιο I - Εισαγωγή

διαδραματίζει σημαντικό ρόλο στη ρύθμιση του ηλεκτρικού δυναμικού της μεμβράνης, με κάποιους λιπιδικούς τύπους να εμφανίζονται αποκλειστικά στην εσωτερική ή την εξωτερική στιβάδα (Bos et al., 2007). Χαρακτηριστικό παράδειγμα είναι τα φωσφατιδυλ-ινοσιτίδια (PIs), τα οποία περιέχουν σάκχαρα – παράγωγα της ινοσιτόλης, είναι αρνητικά φορτισμένα και εντοπίζονται αποκλειστικά στην κυτοπλασματική όψη των μεμβρανών (Ingolfsson et al., 2014).



Εικόνα 2 Χαρακτηριστικά παραδείγματα φωσφολιπιδίων. Απεικονίζονται οι χημικές δομές φωσφολιπιδίων από τις παρακάτω κατηγορίες: φωσφατιδυλ-χολίνη (PC), φωσφατιδυλ-αιθανολαμίνη (PE), φωσφατιδυλ-γλυκερόλη (PG), φωσφατιδυλ-σερίνη (PS), φωσφατιδικό οξύ (PA) και καρδιολιπίνη (CDL). Σημαίνονται τα διαφορετικά τμήματα των λιπιδίων (πολική κεφαλή, γλυκερόλη, υδρόφοβες ουρές).

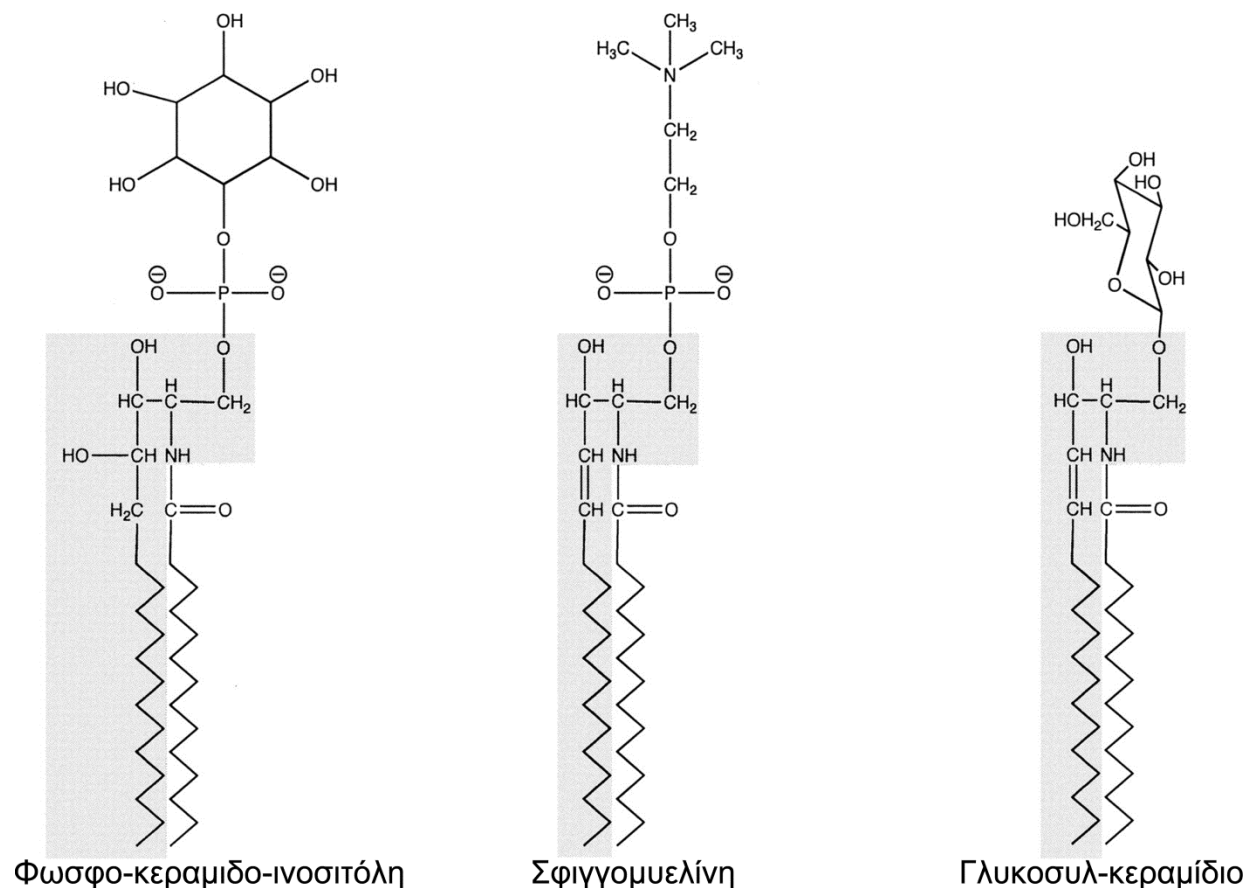
Το μήκος των υδρόφοβων ουρών καθορίζει το πάχος του υδρόφοβου πυρήνα της διπλοστιβάδας, ενώ η παρουσία ή απουσία διπλών δεσμών καθορίζει τη γεωμετρία των λιπιδικών ουρών, η οποία ρυθμίζει σε μεγάλο βαθμό τη ρευστότητα της διπλοστιβάδας. Γενικά, τα κορεσμένα φωσφολιπίδια διευθετούν τις υδρόφοβες ουρές τους κατακόρυφα στο επίπεδο της διπλοστιβάδας, οδηγώντας σε σφιχτό πακετάρισμα των λιπιδικών μορίων. Αυτή η οργάνωση μπορεί να μειώσει σημαντικά την ικανότητα πλευρικής διάχυσης τόσο των ίδιων των λιπιδίων, όσο και των υπόλοιπων συστατικών της μεμβράνης. Αντίθετα, η παρουσία ενός

Κεφάλαιο I - Εισαγωγή

ή περισσότερων διπλών δεσμών (μονοακόρεστα ή πολυακόρεστα φωσφολιπίδια) δημιουργεί κάμψεις στη δομή των ουρών, μειώνοντας το πακετάρισμα των λιπιδίων και αυξάνοντας την πλευρική διάχυσή τους.

Πρέπει να σημειωθεί ότι σε ένα λιπίδιο δεν είναι απαραίτητο οι υδρόφοβες ουρές να είναι πανομοιότυπες. Για παράδειγμα, η 1-παλμιτοϋλ-2-ολεύλ-φωσφατιδυλ-χολίνη (POPC), ένα από τα πιο κοινά φωσφολιπίδια, διαθέτει μια κορεσμένη ουρά παλμιτικού οξέως (16 άνθρακες) και μια μονοακόρεστη ουρά ολεϊκού οξέος (18 άνθρακες, 1 διπλός δεσμός). Επιπρόσθετα, πρέπει να σημειωθεί ότι δεν διαθέτουν όλα τα φωσφολιπίδια δύο υδρόφοβες ουρές. Χαρακτηριστικό παράδειγμα αποτελούν τα λυσοφωσφολιπίδια, τα οποία διαθέτουν μόνο μία υδρόφοβη ουρά και εντοπίζονται κυρίως στις μεμβράνες των λυσοσωμάτων, καθώς και σε εξειδικευμένα οργανίδια όπως τα μελανοσώματα των μελανοκυττάρων (Birgbauer & Chun, 2006). Ένα δεύτερο παράδειγμα είναι οι καρδιολιπίνες (CDLs), οι οποίες συντίθενται στα μιτοχόνδρια των ευκαρυωτικών κυττάρων, με την ομοιοπολική σύνδεση δύο φωσφατιδικών οξέων μέσω των φωσφορικών ιόντων τους (Schlame et al., 2000). Έτσι, ένα μόριο CDL είναι ουσιαστικά ένα ομοιοπολικά συνδεδεμένο διμερές λιπιδίων με τέσσερις υδρόφοβες ουρές, δύο μόρια γλυκερόλης και δύο φωσφορικά ιόντα (Εικόνα 2). Οι CDLs ανακαλύφθηκαν αρχικά στις εξωτερικές μεμβράνες των μιτοχονδρίων στα καρδιακά κύτταρα και, μέχρι πρόσφατα, θεωρούταν ότι εντοπίζονται μόνο σε αυτά. Ωστόσο, πλέον είναι γνωστό ότι υπάρχουν σε όλα τα μιτοχόνδρια, καθώς επίσης και στις μεμβράνες των αρνητικών κατά Gram βακτηρίων (Schlame et al., 2000; Sohlenkamp & Geiger, 2016).

1.2.2 Σφιγγολιπίδια



Εικόνα 3 Χαρακτηριστικά παραδείγματα σφιγγολιπιδίων. Παρουσιάζονται οι χημικές δομές τριών κοινών σφιγγολιπιδίων (φωσφο-κεραμιδο-ινοσιτόλη, σφιγγομυελίνη και γλυκοσυλ-κεραμίδιο). Και στα τρία μόρια η σφιγγοσίνη σημαίνεται με γκρίζο χρώμα.

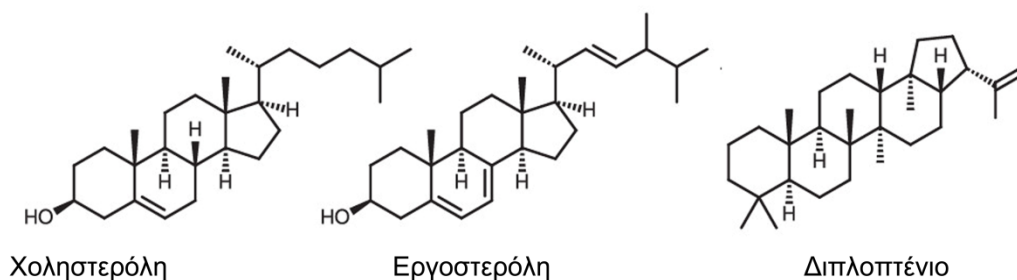
Η δεύτερη πιο άφθονη κατηγορία μεμβρανικών λιπιδίων είναι τα *σφιγγολιπίδια* (Εικόνα 3). Κύριο δομικό στοιχείο των σφιγγολιπιδίων είναι η σφιγγοσίνη (2-αμινο-4-τρανς-οκταδεκενο-1,4-διόλη), μια αμινική αλκοόλη. Τα σφιγγολιπίδια δομούνται από την πρόσδεση ενός λιπαρού οξέος στην αμινομάδα της σφιγγοσίνης, με το λιπαρό οξύ και την επιμήκη υδρογονανθρακική αλυσίδα της σφιγγοσίνης να αποτελούν τις υδρόφοβες ουρές του λιπιδίου, ενώ στην υδροξυλική ρίζα της σφιγγοσίνης μπορούν να προσδεθούν διάφορες χημικές ομάδες οι οποίες λειτουργούν σαν πολικές κεφαλές, αντίστοιχα με τα φωσfolιπίδια (Ingolfsson et al., 2014; van Meer et al., 2008). Επιπρόσθετα, πολλά σφιγγολιπίδια συνδέονται μέσω

Κεφάλαιο I - Εισαγωγή

γλυκοσιδικών δεσμών με σάκχαρα, οδηγώντας στο σχηματισμό γλυκολιπιδίων (γλυκοσφιγγολιπίδια). Τα σφιγγολιπίδια και ιδίως τα γλυκοσφιγγολιπίδια επιτελούν μια σειρά από ιδιαίτερα σημαντικές λειτουργίες για το κύτταρο. Η κύρια λειτουργία τους είναι η προστασία του κυττάρου από βλαβερές εξωκυτταρικές ουσίες μέσω της συσσωμάτωσής τους στην εξωτερική στιβάδα της κυτταρικής μεμβράνης, δημιουργώντας ένα ανθεκτικό και αδιαπέραστο προστατευτικό στρώμα. Επιπρόσθετα, κάποια σφιγγολιπίδια έχουν βρεθεί να ρυθμίζουν πιο εξειδικευμένες λειτουργίες, όπως η κυτταρική αναγνώριση και η μεταγωγή σήματος (Bartke & Hannun, 2009).

1.2.3 Στερόλες

Μετά τα φωσφολιπίδια και τα σφιγγολιπίδια, οι μεμβρανικές στερόλες (Εικόνα 4) αποτελούν ένα από τα πιο κρίσιμα συστατικά της λιπιδικής διπλοστιβάδας. Η πιο γνωστή στερόλη είναι η χοληστερόλη, η οποία εντοπίζεται σε όλους τους ζωικούς οργανισμούς και αποτελεί απαραίτητο συστατικό των μεμβρανών τους, ενώ ταυτόχρονα λειτουργεί σαν πρόδρομο μόριο στη σύνθεση ουσιών όπως οι στεροειδείς ορμόνες και η Βιταμίνη D (Bos et al., 2007). Αντίστοιχα μόρια υπάρχουν στα φυτά (σιτοστερόλη, λανοστερόλη κλπ) και τους μύκητες (εργοστερόλη), ενώ στις μεμβράνες των προκαρυωτικών οργανισμών έχει βρεθεί μια ομάδα στεροειδών μορίων, τα οπανοειδή, τα οποία φαίνονται να επιτελούν παρόμοια λειτουργία (Melo et al., 2015).



Εικόνα 4 Οι στερόλες των βιολογικών μεμβρανών. Η χοληστερόλη και τα παράγωγά της, όπως η εργοστερόλη, αποτελούν τις στερόλες των ευκαρυωτικών μεμβρανών, ενώ τα οπανοειδή όπως το διπλοπτένιο επιτελούν ανάλογο ρόλο στις μεμβράνες των βακτηρίων.

Κεφάλαιο I - Εισαγωγή

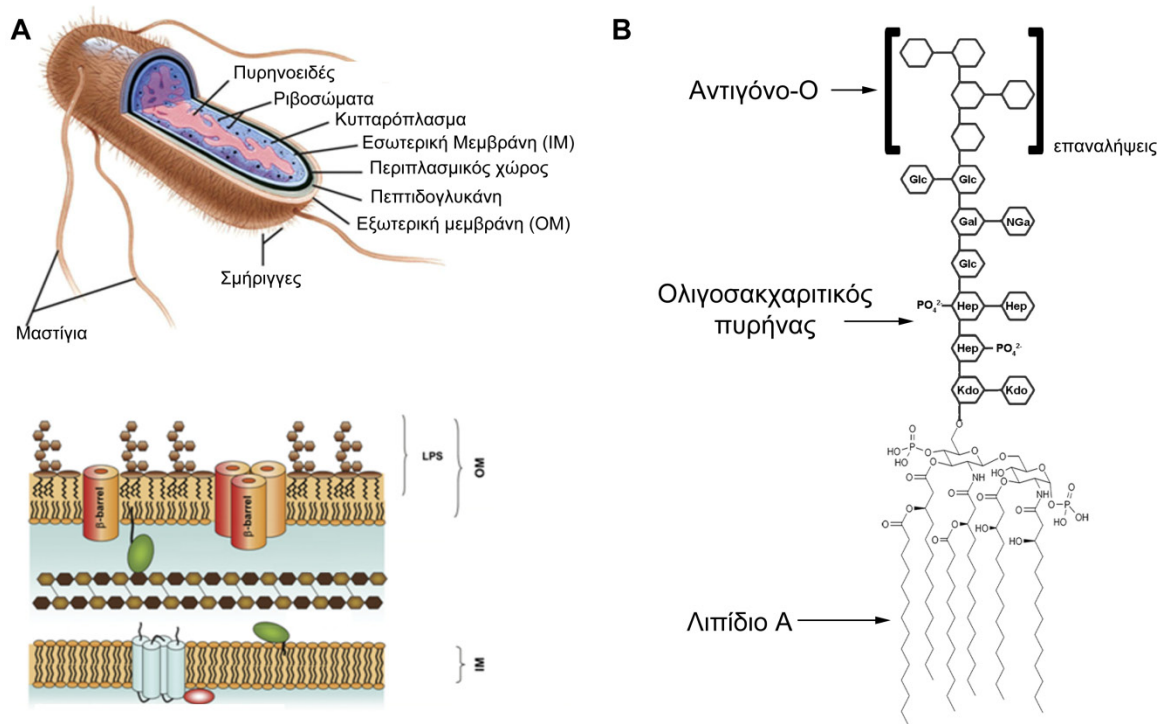
Οι στερόλες διαδραματίζουν σημαντικό ρόλο στη ρύθμιση ιδιοτήτων της μεμβράνης όπως η ρευστότητα, η διαπερατότητα από διάφορες ουσίες και η κινητικότητα των πρωτεϊνών. Στα ζωικά κύτταρα η χοληστερόλη αποτελεί το 10-50% των λιπιδίων της μεμβράνης, με την περιεκτικότητα να μεταβάλλεται ανάλογα με τον κυτταρικό τύπο (van Meer et al., 2008). Αντίστοιχες τιμές έχουν αναφερθεί και για την παρουσία της εργοστερόλης στις μεμβράνες των μυκήτων. Τα μόρια των στερολών διατάσσονται στο εσωτερικό της διπλοστιβάδας και αλληλεπιδρούν με τις υδρόφοβες ουρές των λιπιδίων, ενώ οι πολικές υδροξυλομάδες τους βρίσκονται κοντά στις πολικές κεφαλές των λιπιδίων (Ohno-Rekila et al., 2002). Επιπρόσθετα, τα μόρια των στερολών μπορούν να αλληλεπιδράσουν και με τις μεμβρανικές πρωτεΐνες, χρησιμοποιώντας την υδροξυλομάδα τους για το σχηματισμό δεσμών υδρογόνου με πολικά αμινοξικά κατάλοιπα ή τους δακτυλίους τους για το σχηματισμό αλληλεπιδράσεων π-ηλεκτρονίων (π-stacking) με αρωματικά κατάλοιπα.

1.2.4 Λιποπολυσακχαρίτες

Τα αρνητικά κατά Gram βακτήρια, σε αντίθεση με τα θετικά κατά Gram βακτήρια και τους ευκαρυωτικούς οργανισμούς, περιβάλλονται από ένα σύστημα δύο κυτταρικών μεμβρανών, μιας Εσωτερικής και μιας Εξωτερικής Μεμβράνης (Εικόνα 5) (Alberts et al., 2002). Η Εσωτερική Μεμβράνη παρουσιάζει πολλές ομοιότητες με την κυτταρική μεμβράνη των θετικών κατά Gram βακτηρίων και των ευκαρυωτικών κυττάρων (Sohlenkamp & Geiger, 2016): είναι συμμετρική και αποτελείται από τυπικά φωσφολιπίδια, σφιγγολιπίδια και στερόλες (οπανοειδή), ενώ οι πρωτεΐνες της είναι κυρίως α-ελικοειδείς διαμεμβρανικές πρωτεΐνες (βλ. Ενότητα 1.3.1.1). Ανάμεσα στην Εσωτερική και την Εξωτερική Μεμβράνη παρεμβάλλεται ο περιπλαστικός χώρος, ένα υποκυτταρικό διαμέρισμα με υψηλό ιξώδες, το οποίο περιέχει υδατικό διαλύτη, ένα λεπτό στρώμα πεπτιδογλυκάνης (το βασικό συστατικό του κυτταρικού τοιχώματος των βακτηρίων) και πρωτεΐνες (περιπλαστικές πρωτεΐνες) οι οποίες βοηθούν στην επικοινωνία των δύο μεμβρανών μεταξύ τους (Alberts et al., 2002). Η Εξωτερική Μεμβράνη, τέλος, είναι μια μη συμμετρική λιπιδική διπλοστιβάδα. Η εσωτερική στιβάδα της αποτελείται από τυπικά φωσφολιπίδια, όπως φωσφατιδυλ-αιθανολαμίνες και φωσφατιδυλ-γλυκερόλες,

Κεφάλαιο I - Εισαγωγή

ωστόσο, συχνά περιέχει και πιο σπάνιους τύπους λιπιδίων, όπως οι καρδιολιπίνες. Από την άλλη, η εξωτερική στιβάδα αποτελείται από μια ειδική κατηγορία λιποπολυσακχαρών, τους λιποπολυσακχαρίτες (Bos et al., 2007), ενώ οι πρωτεΐνες της είναι κατά βάση διαμεμβρανικά β-βαρέλια (βλ. Ενότητα 1.3.1.2).

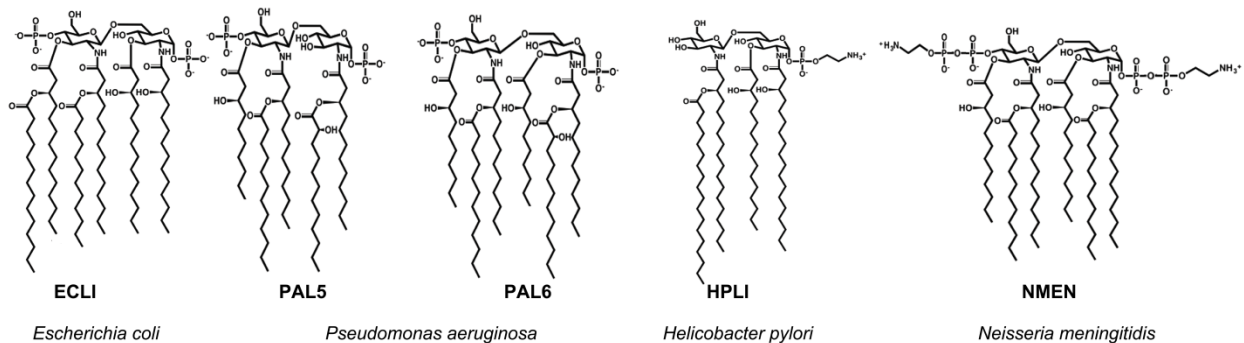


Εικόνα 5 Οι μεμβράνες των αρνητικών κατά Gram βακτηρίων. (A) Διαγραμματική απεικόνιση της Εσωτερικής και της Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων (προσαρμογή από Alberts et al., 2002). (B) Η δομή ενός λιποπολυσακχαρίτη. Διακρίνονται τα τρία τμήματα του μορίου (Λιπίδιο A, ολιγοσακχαριτικός πυρήνας και Αντιγόνο-O). Το Λιπίδιο A παρουσιάζεται σαν χημική δομή ενώ τα σάκχαρα του πυρήνα και του Αντιγόνου-O παρουσιάζονται διαγραμματικά, λόγω της πολύπλοκης δομής τους.

Πρέπει να σημειωθεί ότι η παρουσία των λιποπολυσακχαριτών στην Εξωτερική Μεμβράνη αποτελεί αποκλειστικά χαρακτηριστικό των αρνητικών κατά Gram βακτηρίων. Αντίθετα, οι εξωτερικές μεμβράνες των μιτοχονδρίων και των χλωροπλαστών, παρότι πιστεύεται ότι έχουν προκύψει εξελικτικά από τα αρνητικά κατά Gram βακτήρια σύμφωνα με τη θεωρία της ενδοσυμβίωσης (Chapman & Margulis, 1998), και παρότι διαθέτουν διαμεμβρανικά β-βαρέλια, δεν περιέχουν λιποπολυσακχαρίτες αλλά μόνο τυπικά λιπίδια (Bos

Κεφάλαιο I - Εισαγωγή

et al., 2007). Σε κάθε περίπτωση, η Εξωτερική Μembrάνη δρα σαν προστατευτικό στρώμα για τα βακτήρια, ελέγχοντας τη ροή ουσιών από και προς το εσωτερικό του κυττάρου, προφυλάσσοντας το βακτήριο από βλαβερές για αυτό ουσίες και βοηθώντας το να επιβιώσει σε αντίξοες συνθήκες (James et al., 2009; Nikaido, 2003).



Εικόνα 6 Δομές του Λιπιδίου Α τεσσάρων αρνητικών κατά Gram βακτηρίων. Παρατίθενται οι χημικές δομές του Λιπιδίου Α των βακτηρίων *Escherichia coli* (ECLI), *Pseudomonas aeruginosa* (PAL5 και PAL6), *Helicobacter pylori* (HPLI) και *Neisseria meningitidis* (NMEN).

Οι λιποπολυσακχαρίτες είναι μια ειδική κατηγορία σύνθετων λιπιδικών μορίων, τα οποία αποτελούνται από τρία διακριτά συστατικά (Εικόνα 5): ένα λιπιδικό μόριο (Λιπίδιο Α), ένα συντηρημένο ολιγοσακχαριτικό πυρήνα και μια υψηλά μεταβλητή πολυσακχαριτική αλυσίδα, γνωστή σαν Αντιγόνο-Ο (Raetz & Whitfield, 2002; Snyder et al., 1999). Το λιπίδιο Α (Εικόνα 6), γνωστό και σαν «Ενδοτοξίνη», αποτελείται από ένα διμερές φωσφορυλιωμένων γλυκοζαμινών το οποίο λειτουργεί σαν πολική κεφαλή, πάνω στις οποίες προσδένονται ομοιοπολικά μια σειρά από λιπαρά οξέα, τα οποία λειτουργούν σαν λιπιδικές ουρές. Ο αριθμός, το μήκος και η οργάνωση των λιπιδικών ουρών αποτελούν το βασικό χαρακτηριστικό που διαφοροποιεί το Λιπίδιο Α ανάμεσα σε διαφορετικά είδη βακτηρίων (Kim et al., 2016). Ένα δεύτερο γνώρισμα του Λιπιδίου Α είναι η φωσφορυλίωση του ζεύγους των γλυκοζαμινών, με τον αριθμό των φωσφορικών ομάδων να διαφοροποιείται επίσης ανάμεσα σε κάποια βακτήρια. Η παρουσία πολλαπλών φωσφορικών ομάδων οδηγεί σε υψηλά αρνητικό φορτίο. Έτσι, τα Λιπίδια Α απαιτούν την παρουσία δισθενών μεταλλικών ιόντων (Ca^{2+} ή Mg^{2+}) για να σταθεροποιηθούν στην εξωτερική στιβάδα της μεμβράνης. Η παρουσία των δισθενών ιόντων

Κεφάλαιο I - Εισαγωγή

φαίνεται να είναι καθοριστική για τη σωστή οργάνωση και σταθερότητα της Εξωτερικής Μembrάνης, καθώς η απουσία τους ή η αντικατάστασή τους από μονοσθενή ιόντα Na^+ έχει βρεθεί να αποσταθεροποιεί το στρώμα των λιποπολυσακχαριτών, οδηγώντας στη λύση της διπλοστιβάδας (Pontes et al., 2012; Snyder et al., 1999).

Ο ολιγοσακχαρικός πυρήνας (Εικόνα 5) είναι ομοιοπολικά συνδεδεμένος στο Λιπίδιο A. Δομικά διακρίνεται σε έναν εσωτερικό και έναν εξωτερικό πυρήνα. Ο εσωτερικός πυρήνας αποτελείται κυρίως από «σπάνιους» μονοσακχαρίτες, όπως το 3-δεοξυ-D-μαννο-2-οκτυλοσονικό οξύ (Kdo) και η L-γλυκερο-D-μαννο-επτόζη (Hep), ενώ ο εξωτερικός πυρήνας περιέχει πιο κοινά σάκχαρα όπως η D-γλυκόζη (Glc), η D-γαλακτόζη (Gal) και η N-ακετυλο-D-γλυκοζαμίνη (NGa). Σε αρκετές περιπτώσεις κάποιοι μονοσακχαρίτες (κυρίως η Hep) φωσφορυλιώνονται, προσδίδοντας επιπλέον αρνητικά φορτία τα οποία απαιτούν την παρουσία δισθενών κατιόντων για να εξουδετερωθούν, αντίστοιχα με το Λιπίδιο A (Snyder et al., 1999). Η βασική σύνθεση του πυρήνα, όπως περιγράφηκε, είναι υψηλά συντηρημένη ανάμεσα στα βακτήρια. Η παρουσία και σωστή οργάνωση του ολιγοσακχαρικού πυρήνα είναι απαραίτητη για την ορθή λειτουργία της Εξωτερικής Μembrάνης σαν προστατευτικό στρώμα. Για παράδειγμα, μελέτες με μεταλλαγμένα στελέχη του βακτηρίου *E. coli*, τα οποία συνθέτουν λιποπολυσακχαρίτες με ελλιπή ή τμηματικό ολιγοσακχαρικό πυρήνα (γνωστά σαν «αδρόι λιποπολυσακχαρίτες») έδειξαν ότι οι Εξωτερικές Μembrάνες τους ήταν πιο ευάλωτες σε αντιμικροβιακά πεπτίδια, οδηγώντας στη λύση του βακτηριακού κυττάρου (Bello et al., 2015). Αντίθετα, τα βακτήρια που συνθέτουν λιποπολυσακχαρίτες με πλήρη πυρήνα («λείοι λιποπολυσακχαρίτες») είναι πιο ανθεκτικά, εμποδίζοντας την είσοδο των πεπτιδίων.

Το τελευταίο συστατικό των λιποπολυσακχαριτών είναι το Αντιγόνο-O. Το Αντιγόνο-O είναι μια πολυσακχαρική αλυσίδα αποτελούμενη από επαναλήψεις γλυκάνης, η οποία προσδένεται ομοιοπολικά στον ολιγοσακχαρικό πυρήνα. Ονομάζεται «αντιγόνο» καθώς, σε αρκετές περιπτώσεις, έχει βρεθεί να αποτελεί στόχο για αντισώματα και να επάγει ανοσολογικές αποκρίσεις στους ξενιστές των βακτηρίων. Η σύνθεσή του είναι υψηλά μεταβλητή, τόσο ως προς τον αριθμό των επαναλήψεων όσο και ως προς τους μονοσακχαρίτες που τις αποτελούν (Snyder et al., 1999). Σε αντίθεση με το Λιπίδιο A και τον ολιγοσακχαρικό πυρήνα, το Αντιγόνο-O διαφοροποιείται έντονα όχι μόνο ανάμεσα σε διαφορετικά είδη

Κεφάλαιο I - Εισαγωγή

βακτηρίων αλλά και ανάμεσα σε άτομα του ίδιου είδους ή ακόμη και σε λιποπολυσακχαρίτες του ίδιου κυττάρου. Έτσι, μόνο για το στέλεχος K12 του βακτηρίου *E. coli* είναι γνωστά πάνω από 100 διαφορετικά Αντιγόνα-Ο, με τη σύνθεσή τους να εξαρτάται από τις συνθήκες του περιβάλλοντος (Wu et al., 2013). Αντίθετα με την *E. coli*, η παρουσία και η σύνθεση Αντιγόνων-Ο είναι άγνωστη για την πλειονότητα των αρνητικών κατά Gram βακτηρίων. Επιπρόσθετα, μεγάλος αριθμός βακτηρίων, συμπεριλαμβανομένης και της ίδιας της *E. coli*, έχουν βρεθεί να παράγουν λιποπολυσακχαρίτες που διαθέτουν μόνο το λιπίδιο Α και τον ολιγοσακχαρικό πυρήνα (γνωστοί και σαν «λείοι λιποπολυσακχαρίτες»), υποδεικνύοντας ότι η παρουσία του Αντιγόνου-Ο δεν είναι απαραίτητη για την οργάνωση των λιποπολυσακχαριτών (Bos et al., 2007).

Πίνακας 1 Παραδείγματα ασθενειών με τις οποίες έχουν συσχετιστεί οι λιποπολυσακχαρίτες διάφορων αρνητικών κατά Gram βακτηρίων.

Είδος Βακτηρίου	Ασθένειες
<i>Escherichia coli</i>	Τροφική δηλητηρίαση Λοιμώξεις του ουροποιητικού συστήματος
<i>Pseudomonas aeruginosa</i>	Σηψαιμία Επιδείνωση της κυστικής ίνωσης
<i>Helicobacter pylori</i>	Πνευμονία Γαστρίτιδα Έλκη στομάχου Γαστρικό καρκίνωμα
<i>Vibrio cholerae</i>	Χολέρα
<i>Neisseria meningitidis</i>	Μηνιγγίτιδα Σηψαιμία
<i>Neisseria gonorrhoeae</i>	Γονόρροια
<i>Yersinia pestis</i>	Σηπτική αρθρίτιδα Σηψαιμική Πανώλη

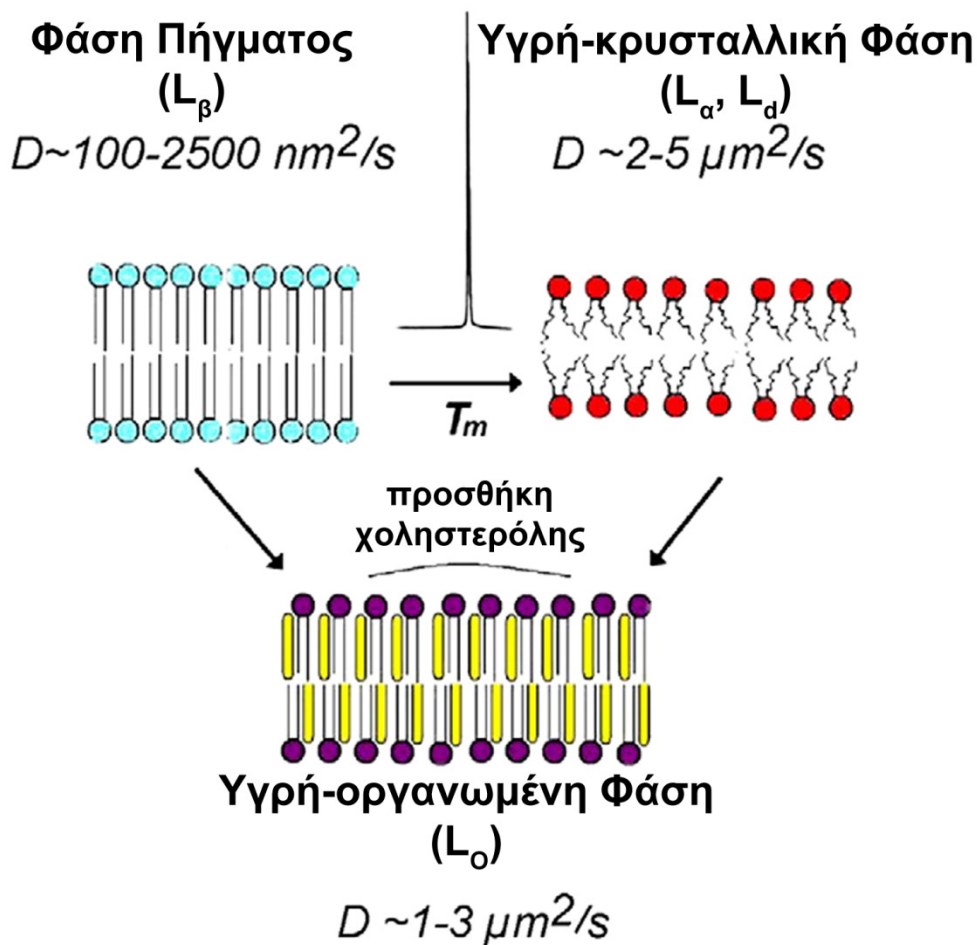
Η οργάνωση των λιποπολυσακχαριτών στην εξωκυτταρική στιβάδα της Εξωτερικής Μembrάνης είναι κρίσιμη για την ορθή λειτουργία των αρνητικών κατά Gram βακτηρίων. Χαρακτηριστικά της δομής και της λειτουργίας πολλών β-βαρελίων έχουν βρεθεί να εξαρτώνται από τις αλληλεπιδράσεις τους με παρακείμενα μόρια λιποπολυσακχαριτών (Ceccarelli et al., 2004; Ferguson et al., 2000; Klebba et al., 1990). Επιπρόσθετα, το οργανωμένο στρώμα των λιποπολυσακχαριτών έχει βρεθεί να δρα προστατευτικά για τα

Κεφάλαιο I - Εισαγωγή

βακτήρια, εμποδίζοντας την είσοδο διάφορων βλαβερών για τα βακτήρια ουσιών (π.χ. αντιβιοτικά) και τη δράση αντιμικροβιακών πεπτιδίων (Nikaido, 2003). Αντίθετα, η παρουσία ένθετων φωσφολιπιδίων στην εξωτερική στιβάδα της μεμβράνης, ακόμη και σε μικρές ποσότητες, φαίνεται να διαταράσσει αυτήν την οργάνωση, αποδυναμώνοντας τη διπλοστιβάδα (Hsu et al., 2016; Nikaido, 2003). Τέλος, διάφοροι λιποπολυσακχαρίτες έχουν βρεθεί να συμβάλλουν στην παθογένεια των βακτηρίων απέναντι στους ξενιστές τους. Συγκεκριμένα, το Λιπίδιο A των Λιποπολυσακχαριτών μπορεί να αλληλεπιδράσει με τους υποδοχείς Toll-Like του ανθρώπου και να οδηγήσει σε τοξικές αποκρίσεις, όπως ο πυρετός και το σηπτικό σοκ (Raetz & Whitfield, 2002). Έτσι, οι λιποπολυσακχαρίτες διάφορων βακτηρίων έχουν εμπλακεί σε μεγάλο εύρος ασθενειών (Πίνακας 1), που κυμαίνονται από απλές περιπτώσεις τροφικής δηλητηρίασης και λοιμώξεις του ουρογεννητικού συστήματος έως σοβαρές ασθένειες όπως η σηψαιμία, ο καρκίνος του στομάχου, η βλεννόρροια, η μηνιγγίτιδα και η σηψαιμική πανώλη (Griffiss et al., 1988; Kim et al., 2016; Raetz & Whitfield, 2002).

1.2.5 Η οργάνωση των λιπιδίων στη μεμβράνη

Όπως έχει ήδη αναφερθεί, η δομή και οι ιδιότητες της λιπιδικής διπλοστιβάδας εξαρτώνται άμεσα από τη δομή και τις ιδιότητες των λιπιδίων που την αποτελούν. Εκτός από τα είδη των λιπιδίων όμως, σημαντικό ρόλο στην οργάνωση της μεμβράνης διαδραματίζουν τόσο η συγκέντρωση του κάθε είδους λιπιδίου όσο και περιβαλλοντικοί παράγοντες όπως η θερμοκρασία, η πίεση, η ιοντική συγκέντρωση και το pH. Ο συνδυασμός όλων των παραπάνω παραγόντων έχει σαν αποτέλεσμα την ύπαρξη διαφορετικών καταστάσεων («φάσεις») για τη λιπιδική διπλοστιβάδα (Εικόνα 7), οι οποίες χαρακτηρίζονται από σημαντικές διαφορές στην οργάνωση των λιπιδίων (Λέκκα et al., 2015). Η ικανότητα των λιπιδικών μορίων να προσαρμόζουν τη ρευστότητα της διπλοστιβάδας σε αυτές τις καταστάσεις είναι γνωστή σαν «συμπεριφορά φάσης» (phase behavior). Μερικές από τις κυριότερες κατηγορίες φάσεων που υιοθετούνται από τις λιπιδικές διπλοστιβάδες είναι οι παρακάτω (Giocondi et al., 2010):



Εικόνα 7 Συμπεριφορά φάσης των λιπιδίων. Διαγραμματική απεικόνιση των τριών κύριων φάσεων των λιπιδίων στη διπλοστιβάδα. Τα λιπίδια σε μια μεμβράνη βρίσκονται σε δύο διακριτές φάσεις (φάση πήγματος και υγρή-κρυσταλλική φάση), ανάλογα με τη θερμοκρασία τήξης (T_m). Οι δύο φάσεις χαρακτηρίζονται από διαφορετική οργάνωση των λιπιδικών ουρών, καθώς και από διαφορετικές τιμές του συντελεστή διάχυσης (D). Η προσθήκη χοληστερόλης μπορεί να ωθήσει τα λιπίδια και στις δύο φάσεις να μεταβούν σε μια ενδιάμεση κατάσταση (υγρή-οργανωμένη φάση). Προσαρμογή από Giocondi et al., 2010.

1. Φάση πήγματος (gel, L_β): αντιστοιχεί σε υψηλά οργανωμένες διπλοστιβάδες με σημαντικά χαμηλή ρευστότητα και υψηλό ιξώδες. Οι μεμβράνες σε φάση πήγματος βρίσκονται σε ημι-στερεή κατάσταση και διαθέτουν πυκνά πακεταρισμένα λιπίδια με εκτεταμένες υδρόφοβες ουρές και μειωμένη πλευρική διάχυση.

2. Υγρή-αποδιοργανωμένη ή υγρή-κρυσταλλική φάση (liquid-disordered, L_α ή liquid-crystalline, L_d): αντιστοιχεί σε διπλοστιβάδες αυξημένης ρευστότητας. Οι μεμβράνες στην

Κεφάλαιο I - Εισαγωγή

υγρή-κρυσταλλική φάση διαθέτουν χαλαρά πακεταρισμένα λιπίδια με ελεύθερη κίνηση των υδρόφοβων ουρών τους, τα οποία διαθέτουν υψηλή πλευρική διάχυση.

3. Υγρή-οργανωμένη φάση (liquid-ordered, L_o): αποτελεί μια ενδιάμεση κατάσταση των φάσεων L_β και L_α . Οι μεμβράνες στην υγρή-οργανωμένη φάση δημιουργούν λιπιδικές «σχεδίες» (rafts), σχηματισμούς με πυκνά οργανωμένα λιπίδια οι οποίοι όμως έχουν ταυτόχρονα και σχετικά υψηλή πλευρική διάχυση.

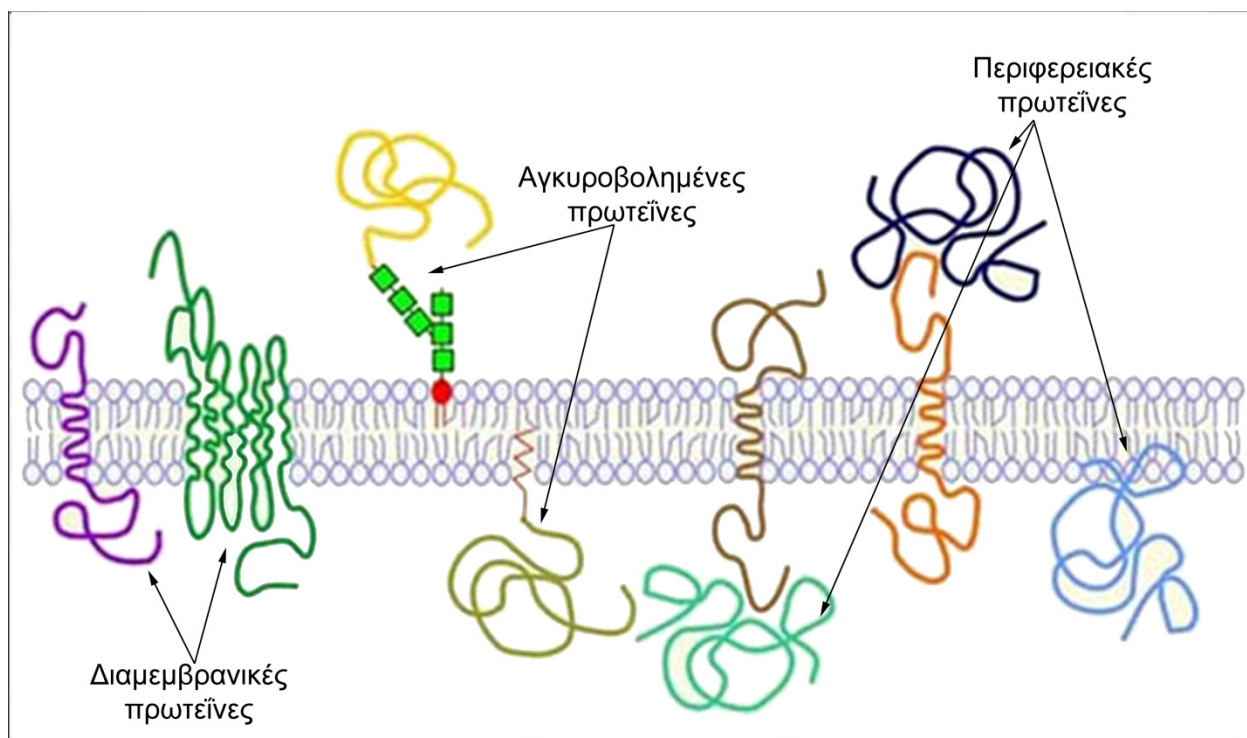
Οι μεμβράνες μεταπίπτουν από τη μία φάση στην άλλη, σαν αποτέλεσμα τόσο της λιπιδικής σύστασης όσο και περιβαλλοντικών παραγόντων όπως η θερμοκρασία (Εικόνα 7). Το φαινόμενο αλλαγής της φάσης που εξαρτάται από τη θερμοκρασία ονομάζεται θερμοτροπική μετάπτωση φάσης (thermal phase transition). Η θερμοκρασία μετάπτωσης ή σημείο τήξης (T_m) από την φάση υψηλού ιξώδους σε ρευστή μορφή είναι χαρακτηριστική για κάθε λιπίδιο (Λέκκα et al., 2015). Εξίσου σημαντική με τη θερμοκρασία είναι και η λιπιδική σύσταση, τόσο ως προς το είδος των λιπιδίων όσο και ως προς τη συγκέντρωσή τους. Γενικά, μεμβράνες αποτελούμενες αποκλειστικά από κορεσμένα λιπίδια είναι υψηλά οργανωμένες, και απαιτούν υψηλές θερμοκρασίες για να μεταβούν από τη φάση πήγματος στην υγρή-κρυσταλλική φάση. Τέτοιες μεμβράνες απαντώνται συχνά σε θερμόφιλα βακτήρια, τα οποία ευδοκιμούν σε υψηλές θερμοκρασίες. Αντίθετα, μεμβράνες που αποτελούνται από μείγμα κορεσμένων-ακόρεστων λιπιδίων είναι πιο χαλαρά συνδεδεμένες, διαθέτουν αυξημένη ρευστότητα και βρίσκονται στην υγρή-κρυσταλλική φάση σε χαμηλότερες τιμές θερμοκρασίας. Οι κυτταρικές μεμβράνες της πλειονότητας των ευκαρυωτικών κυττάρων, καθώς και των μεσόφιλων βακτηρίων, ανήκουν σε αυτήν την κατηγορία (Alberts et al., 2002). Τέλος, ένας καθοριστικός παράγοντας στη ρευστότητα των μεμβρανών και στη μετάβαση ανάμεσα σε διαφορετικές φάσεις είναι η παρουσία στερολών όπως η χοληστερόλη. Η προσθήκη χοληστερόλης σε υψηλά οργανωμένες μεμβράνες που βρίσκονται σε φάση πήγματος αυξάνει τη ρευστότητά τους, ενώ η προσθήκη της σε αποδιοργανωμένες μεμβράνες λειτουργεί αντίστροφα, μειώνοντας τη ρευστότητα των λιπιδίων και αυξάνοντας την οργάνωσή τους (Εικόνα 7). Καθώς η συγκέντρωση της χοληστερόλης στη μεμβράνη των ευκαρυωτικών κυττάρων μπορεί να ρυθμιστεί μεταβολικά, η παραπάνω ιδιότητά της μπορεί να χρησιμοποιηθεί από τα κύτταρα

Κεφάλαιο I - Εισαγωγή

σαν μηχανισμός διατήρησης της ρευστότητας των μεμβρανών τους, ανάλογα με τις περιβαλλοντικές συνθήκες (Alberts et al., 2002).

1.3 Οι μεμβρανικές πρωτεΐνες και οι κατηγορίες τους

Οι μεμβρανικές πρωτεΐνες επιτελούν μια σειρά από λειτουργίες απαραίτητες για τη ζωή του κυττάρου, από την κυτταρική αναγνώριση και τη μεταγωγή σήματος ως τη μεταφορά ουσιών διαμέσου της μεμβράνης, την κυτταρική αύξηση και τον κυτταρικό θάνατο (Alberts et al., 2002). Αυτές οι λειτουργίες είναι πολύ σημαντικές για την επιβίωση των οργανισμών καθώς πιθανή αλλοίωση τέτοιων πρωτεϊνών μπορεί να οδηγήσει σε διάφορες ασθένειες. Από την άλλη, οι πρωτεΐνες είναι δυνατόν να αποτελέσουν και οι ίδιες στόχο φαρμάκων, προκειμένου να ανασταλεί ή να ενισχυθεί η λειτουργία τους κατά περίπτωση (von Heijne, 2007).



Εικόνα 8 Οι κατηγορίες των μεμβρανικών πρωτεϊνών. Με κριτήριο την τοπολογία τους, οι μεμβρανικές πρωτεΐνες διακρίνονται σε διαμεμβρανικές, αγκυροβολημένες και περιφερειακές μεμβρανικές πρωτεΐνες.

Κεφάλαιο I - Εισαγωγή

Με βάση την τοπολογία τους σε σχέση με το επίπεδο της μεμβράνης (Εικόνα 8), οι μεμβρανικές πρωτεΐνες μπορούν να διακριθούν σε τρεις μεγάλες κατηγορίες, τις διαμεμβρανικές πρωτεΐνες, τις αγκυροβολημένες πρωτεΐνες και τις περιφερειακές μεμβρανικές πρωτεΐνες (Alberts et al., 2002). Οι διαμεμβρανικές πρωτεΐνες βυθίζονται στο εσωτερικό της λιπιδικής διπλοστιβάδας μέσω των διαμεμβρανικών τμημάτων τους και τη διαπερνούν, μία ή περισσότερες φορές (von Heijne, 2007). Οι αγκυροβολημένες πρωτεΐνες προσδένονται ομοιοπολικά στην επιφάνεια των μεμβρανών μέσω μετα-μεταφραστικών τροποποιήσεων, που μπορεί να περιλαμβάνουν την προσθήκη λιπιδικών ομάδων (π.χ. παλμιτοϋλίωση κυστεϊνών) ή τη σύνδεση της πρωτεΐνης σε κάποιο γλυκολιπίδιο μέσω γλυκοζυλίωσης. Οι περιφερειακές μεμβρανικές πρωτεΐνες προσκολλούνται μη ομοιοπολικά στην επιφάνεια της μεμβράνης είτε μόνιμα είτε παροδικά, σχηματίζοντας ασθενείς αλληλεπιδράσεις με τα λιπίδια της διπλοστιβάδας, με άλλες μεμβρανικές πρωτεΐνες ή με το συνδυασμό των παραπάνω (Lomize et al., 2012). Τέλος, τα τελευταία χρόνια έχει αναγνωριστεί μία επιπλέον κατηγορία μεμβρανικών πρωτεϊνών, οι «μονοτοπικές» πρωτεΐνες, οι οποίες βρίσκονται μόνιμα προσκολλημένες στη μεμβράνη και τη διαπερνούν μερικώς, με το να εισέρχονται στη μία στιβάδα της χωρίς να εξέρχονται από την άλλη (Nina et al., 2000). Ωστόσο, τα χαρακτηριστικά των μονοτοπικών πρωτεϊνών δεν έχουν αποσαφηνιστεί πλήρως, ενώ κάποια συστήματα ταξινόμησης των μεμβρανικών πρωτεϊνών τις ομαδοποιούν μαζί με τις περιφερειακές πρωτεΐνες (Lomize et al., 2012).

1.3.1 Διαμεμβρανικές πρωτεΐνες

Οι διαμεμβρανικές πρωτεΐνες διαπερνούν, μέσω των διαμεμβρανικών τμημάτων τους, τη λιπιδική διπλοστιβάδα. Αποτελούν περίπου το 20-30% των πρωτεωμάτων και συμμετέχουν σε μεγάλο εύρος κυτταρικών λειτουργιών όπως η μεταγωγή σήματος, η μεταφορά ουσιών, η ενζυμική κατάλυση και η μηχανική στήριξη του κυττάρου (Krogh et al., 2001; Pasquier et al., 2001; von Heijne, 2007). Επιπλέον, συμβάλλουν στη ρύθμιση της λιπιδικής σύστασης της διπλοστιβάδας και στη συγκρότηση και διατήρηση του σχήματος των μεμβρανών και, κατ' επέκταση, του ίδιου του κυττάρου (Dobson et al., 2015; Krogh et al., 2001). Έτσι, οι

Κεφάλαιο I - Εισαγωγή

διαμεμβρανικές πρωτεΐνες αποτελούν σήμερα στόχο για πάνω από 50% των φαρμάκων στην αγορά. Οι διαμεμβρανικές πρωτεΐνες διακρίνονται, με βάση τη δευτεροταγή δομή των διαμεμβρανικών τμημάτων, σε δύο μεγάλες κατηγορίες, τις α-ελικοειδείς διαμεμβρανικές πρωτεΐνες και τα διαμεμβρανικά β-βαρέλια. Οι α-ελικοειδείς διαμεμβρανικές πρωτεΐνες διαθέτουν διαμεμβρανικά τμήματα με δομή α-έλικας. Αποτελούν την πλειονότητα των διαμεμβρανικών πρωτεϊνών και εντοπίζονται σε όλες τις μεμβράνες των οργανισμών, τόσο στην κυτταρική μεμβράνη των κυττάρων όσο και στις μεμβράνες των οργανιδίων. Από την άλλη, τα διαμεμβρανικά β-βαρέλια δομούνται από β-πτυχωτές επιφάνειες αντιπαράλληλων β-κλώνων, οι οποίες περιελίσσονται και οδηγούν στο σχηματισμό ενός βαρελιού. Αποτελούν μια ειδική κατηγορία διαμεμβρανικών πρωτεϊνών, οι οποίες εντοπίζονται αποκλειστικά στις εξωτερικές μεμβράνες των αρνητικών κατά Gram βακτηρίων (βλ. Ενότητα 1.2.4), των μιτοχονδρίων και των χλωροπλαστών (Tusnady et al., 2004).

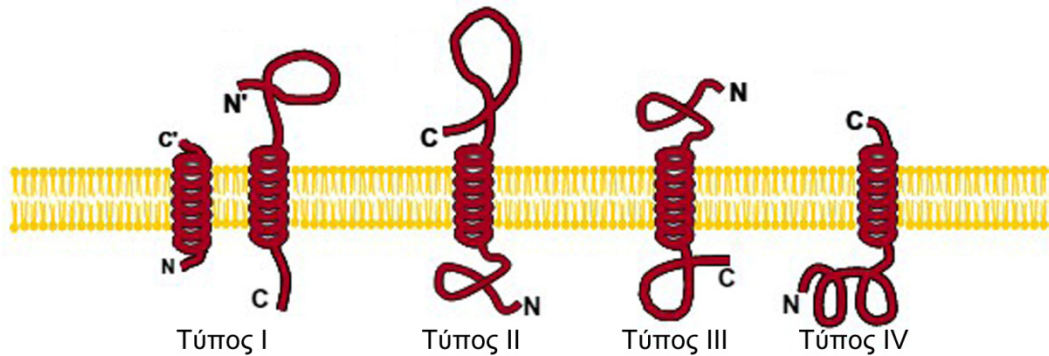
1.3.1.1 Οι α-ελικοειδείς διαμεμβρανικές πρωτεΐνες

Ως α-ελικοειδείς διαμεμβρανικές πρωτεΐνες χαρακτηρίζονται οι διαμεμβρανικές πρωτεΐνες με δομή α-έλικας στα διαμεμβρανικά τμήματά τους. Οι αλληλουχίες των διαμεμβρανικών τμημάτων τους είναι κατά βάση υδρόφοβες, με μήκος που συνήθως κυμαίνεται από 15-40 κατάλοιπα (Krogh et al., 2001). Οι α-ελικοειδείς πρωτεΐνες γενικά διακρίνονται σε δύο μεγάλες κατηγορίες με κριτήριο τον αριθμό των διαμεμβρανικών τμημάτων τους, ο οποίος αντιστοιχεί και στο πόσες φορές διαπερνούν τη μεμβράνη (Εικόνα 8). Έτσι, χωρίζονται σε πρωτεΐνες με ένα διαμεμβρανικό τμήμα («single-spanning» ή «single-pass»), οι οποίες διαπερνούν τη λιπιδική διπλοστιβάδα μία φορά και σε πρωτεΐνες με περισσότερα του ενός διαμεμβρανικά τμήματα («multi-spanning» ή «multi-pass»), οι οποίες διαπερνούν τη λιπιδική διπλοστιβάδα περισσότερες φορές (Tsaousis et al., 2017b).

Οι πρωτεΐνες με ένα διαμεμβρανικό τμήμα αποτελούν μια μεγάλη και αρκετά ετερογενή κατηγορία πρωτεϊνών, περιλαμβάνοντας πολλούς υποδοχείς, διάφορα ένζυμα, υπομονάδες μεγαλύτερων υπερμοριακών συμπλόκων, δομικές πρωτεΐνες και διάφορους ρυθμιστικούς παράγοντες (Pogozheva & Lomize, 2018). Χαρακτηριστικό παράδειγμα

Κεφάλαιο I - Εισαγωγή

πρωτεϊνών αυτού του τύπου είναι οι υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs), οι οποίοι ρυθμίζουν πλήθος κυτταρικών αποκρίσεων και έχουν εμπλακεί σε διάφορες ασθένειες (Lemmon & Schlessinger, 2010). Με βάση την τοπολογία τους στη μεμβράνη, το μήκος των N- και C-άκρων τους και την παρουσία ή απουσία σηματοδοτικού πεπτιδίου («signal peptide»), μπορούν να διακριθούν σε τέσσερις υποκατηγορίες (Εικόνα 9), τους Τύπους I-IV (Chou & Shen, 2007):



Εικόνα 9 Οι τέσσερις τύποι διαμεμβρανικών πρωτεϊνών με ένα διαμεμβρανικό τμήμα. Προσαρμογή από Chou et al., 2007.

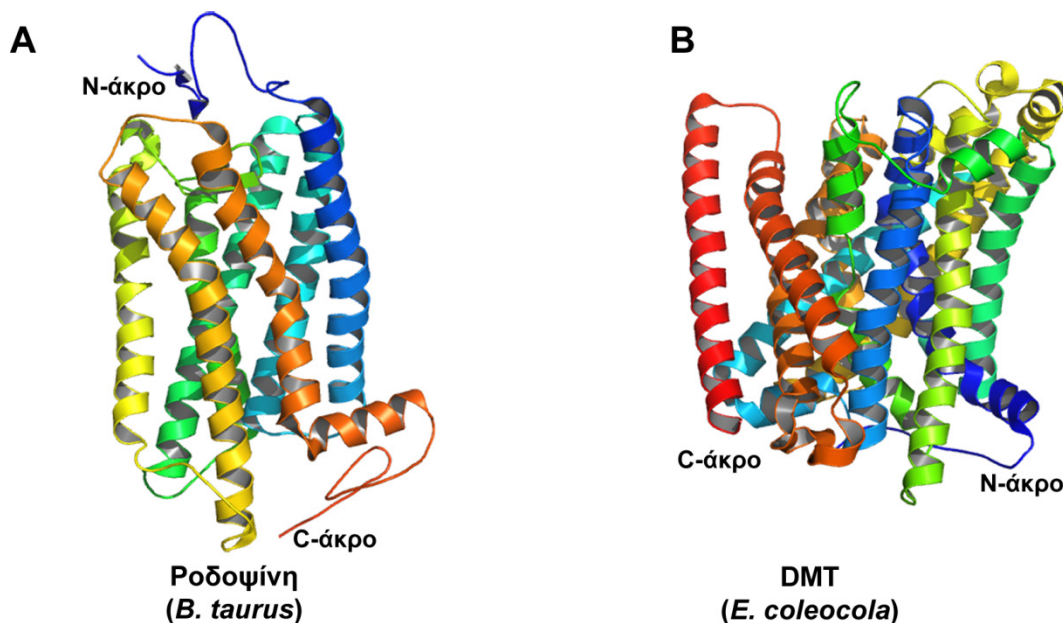
1. Πρωτεΐνες Τύπου I: Διαπερνούν τη διπλοστιβάδα έχοντας το N-άκρο στην εξωκυτταρική όψη ή στον αυλό του ενδοπλασματικού δικτύου (ΕΔ) και το C-άκρο στην κυτοπλασματική όψη. Διαθέτουν σηματοδοτικό πεπτίδιο, το οποίο κατευθύνει τις πρωτεΐνες στο ΕΔ κατά την πρωτεϊνοσύνθεση και αποκόπτεται κατά την είσοδό τους σε αυτό. Το μήκος του N-άκρου είναι, τυπικά, μεγαλύτερο από αυτό του C-άκρου, χωρίς ωστόσο αυτό να ισχύει για όλες τις περιπτώσεις.

2. Πρωτεΐνες Τύπου II: Διαπερνούν τη διπλοστιβάδα έχοντας το N-άκρο στην κυτοπλασματική όψη και το C-άκρο στην εξωκυτταρική όψη ή τον αυλό του ΕΔ. Δεν διαθέτουν σηματοδοτικό πεπτίδιο αλλά ένα πεπτίδιο αγκυροβόλησης, το οποίο κατευθύνει την είσοδό τους στο ΕΔ και παραμένει ανέπαφο στην πρωτεΐνη.

3. Πρωτεΐνες Τύπου III: Διαπερνούν τη διπλοστιβάδα έχοντας την ίδια τοπολογία με τις πρωτεΐνες Τύπου I. Ωστόσο, διαθέτουν πεπτίδιο αγκυροβόλησης αντί για σηματοδοτικό πεπτίδιο. Επιπρόσθετα, το μήκος του C-άκρου είναι, τυπικά, πολύ μεγαλύτερο από το μήκος του N-άκρου.

Κεφάλαιο I - Εισαγωγή

4. Πρωτεΐνες Τύπου IV: Διαπερνούν τη διπλοστιβάδα έχοντας τοπολογία παρόμοια με τις Πρωτεΐνες Τύπου II, ωστόσο είναι άγνωστος ο μηχανισμός με τον οποίο εισέρχονται στη μεμβράνη. Το μήκος του εξωκυτταρικού C-άκρου είναι πολύ μικρό, με το καρβοξυτελικό κατάλοιπο να βρίσκεται σχεδόν στο όριο της λιπιδικής διπλοστιβάδας. Οι πρωτεΐνες Τύπου IV είναι ελάχιστα μελετημένες και πολύ λίγα είναι γνωστά για τη δομή και τη λειτουργία τους.



Εικόνα 10 Παραδείγματα α-ελικοειδών διαμεμβρανικών πρωτεϊνών με περισσότερα του ενός διαμεμβρανικά τμήματα. (A) Η Ροδοψίνη του *Bos taurus*, ένας GPCR με 7 διαμεμβρανικά τμήματα, εξωκυτταρικό N-άκρο και κυτοπλασματικό C-άκρο. (B) Ο μεταφορέας δισθενών μεταλλικών ιόντων (DMT) του *Eremococcus coleocola*, ένας διαμεμβρανικός μεταφορέας με 12 διαμεμβρανικά τμήματα και κυτοπλασματικό N- και C-άκρο.

Οι πρωτεΐνες με περισσότερα από ένα διαμεμβρανικά τμήματα (Εικόνα 10) αποτελούν την πλειονότητα των διαμεμβρανικών πρωτεϊνών. Ο αριθμός των διαμεμβρανικών τμημάτων τους μπορεί να ποικίλλει από 2 ως 14 ή και ακόμη περισσότερα (Krogh et al., 2001; Tusnady et al., 2004). Διαίρούνται σε μεγάλο αριθμό υπερικογενειών και οικογενειών, τα μέλη των οποίων παρουσιάζουν ομοιότητα στην τοπολογία, τη δομή και τις λειτουργίες τους, ακόμη και σε περιπτώσεις που η ομοιότητα της αμινοξικής αλληλουχίας τους είναι πολύ χαμηλή. Τα μέλη των πολυτοπικών πρωτεϊνών περιλαμβάνουν υποδοχείς, μεταφορείς ουσιών, ιοντικά κανάλια, διάφορα ένζυμα και πολλές άλλες κατηγορίες πρωτεϊνών, οι οποίες ρυθμίζουν

Κεφάλαιο I - Εισαγωγή

πλήθος κυτταρικών αποκρίσεων (Tsaousis et al., 2017b). Χαρακτηριστικό παράδειγμα αποτελούν οι συζευγμένοι με G-πρωτεΐνες υποδοχείς (GPCRs), μια μεγάλη και αρκετά ετερογενής υπερικογένεια υποδοχέων, οι οποίοι όμως διαθέτουν κοινά τοπολογικά, δομικά και λειτουργικά χαρακτηριστικά (Kristiansen, 2004; Rosenbaum et al., 2009).

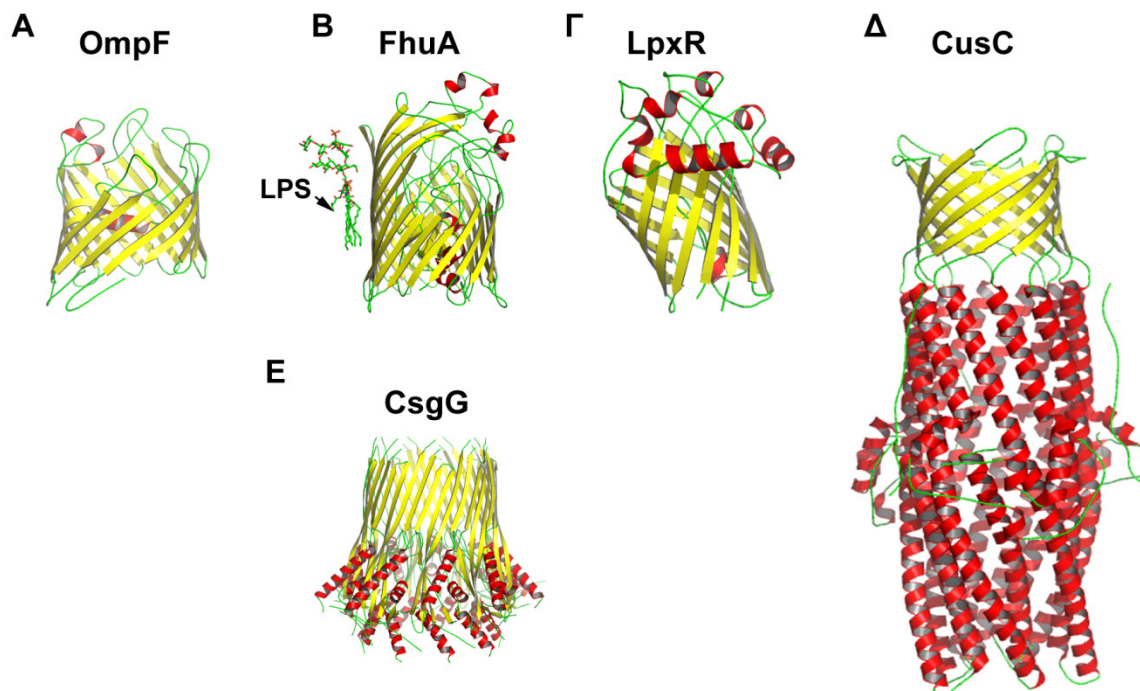
Τα διαμεμβρανικά τμήματα των α-ελικοειδών πρωτεϊνών έχουν μήκος που κυμαίνεται από 15 ως 40 κατάλοιπα και αποτελούνται, κατά βάση, από υδρόφοβα αμινοξικά κατάλοιπα (Krogh et al., 2001). Στα άκρα τους εντοπίζονται συχνά αρωματικά κατάλοιπα όπως τρυπτοφάνη και τυροσίνη, αλλά και θετικά φορτισμένα κατάλοιπα (Λυσίνη, Αργινίνη) τα οποία αλληλεπιδρούν με τα φωσφορικά ιόντα των φωσφολιπιδίων μέσα από γέφυρες άλατος (de Planque et al., 2002). Η α-ελικοειδής δομή πολλών διαμεμβρανικών τμημάτων παρουσιάζει συχνά κάμψεις λόγω της παρουσίας καταλοίπων προλίνης, οι οποίες σε αρκετές περιπτώσεις έχουν βρεθεί να έχουν λειτουργικό ρόλο (von Heijne, 1991). Οι βρόχοι ανάμεσα στα διαμεμβρανικά τμήματα έχουν τυπικά μήκος 5-20 καταλοίπων, ωστόσο, σε αρκετές πρωτεΐνες έχουν παρατηρηθεί μεγάλοι σε μέγεθος βρόχοι οι οποίοι σχηματίζουν αυτοτελείς δομικές περιοχές. Οι βρόχοι αποτελούνται κατά βάση από πολικά και φορτισμένα κατάλοιπα. Ωστόσο, σε αρκετές περιπτώσεις τμήματα των βρόχων έχουν βρεθεί να αποτελούνται από υδρόφοβα κατάλοιπα και να βυθίζονται μερικώς στο εσωτερικό της μεμβράνης. Η στατιστική ανάλυση μιας σειράς διαμεμβρανικών πρωτεϊνών με γνωστή τοπολογία έχει υποδείξει ότι οι ενδοκυτταρικοί βρόχοι παρουσιάζουν στατιστικά σημαντική υψηλότερη παρουσία θετικά φορτισμένων καταλοίπων σε σχέση με τους εξωκυτταρικούς βρόχους (Wallin & von Heijne, 1998). Το παραπάνω χαρακτηριστικό, γνωστό και σαν «positive inside rule», μπορεί να χρησιμοποιηθεί στην εκπαίδευση υπολογιστικών μεθόδων για την πρόγνωση της τοπολογίας των α-ελικοειδών διαμεμβρανικών πρωτεϊνών. Ωστόσο, έχουν βρεθεί πολλές διαμεμβρανικές πρωτεΐνες οι οποίες δεν φαίνονται να ακολουθούν το παραπάνω κριτήριο, μειώνοντας έτσι την αξιοπιστία του. Αρκετές α-ελικοειδείς διαμεμβρανικές πρωτεΐνες διαθέτουν σηματοδοτικό πεπτίδιο, το οποίο αποκόπτεται κατά την είσοδό τους στο ΕΔ. Η αλληλουχία του σηματοδοτικού πεπτιδίου παρουσιάζει χαρακτηριστικά παρόμοια με αυτά των διαμεμβρανικών α-ελίκων γεγονός που οδηγεί συχνά σε λανθασμένη πρόγνωση της διαμεμβρανικής τοπολογίας των πρωτεϊνών. Τέλος, πολλές διαμεμβρανικές πρωτεΐνες

Κεφάλαιο I - Εισαγωγή

διαθέτουν μια σειρά από αμφιπαθικές α-έλικες είτε στους βρόχους είτε στο N- ή το C-άκρο τους, οι οποίες αλληλεπιδρούν με την επιφάνεια της μεμβράνης (Tsaousis et al., 2017b).

1.3.1.2 Τα διαμεμβρανικά β-βαρέλια

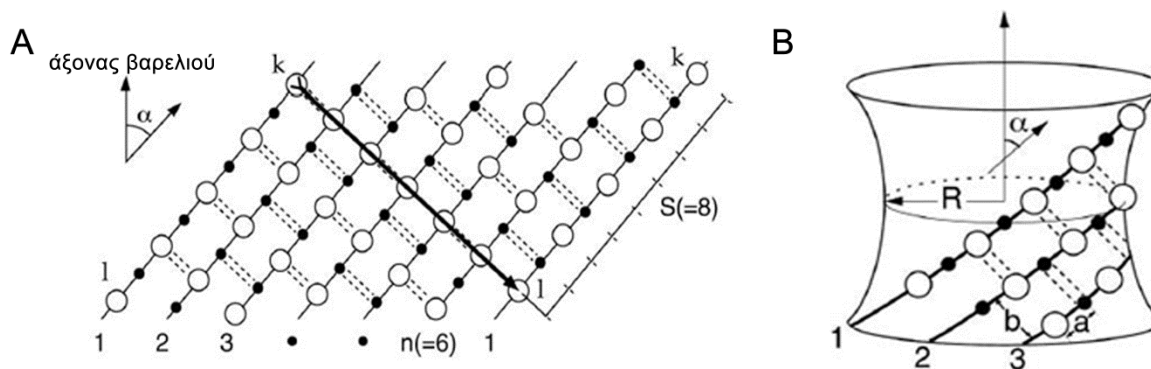
Η δεύτερη μεγάλη κατηγορία διαμεμβρανικών πρωτεϊνών είναι τα διαμεμβρανικά β-βαρέλια. Σε αντίθεση με τις α-ελικοειδείς διαμεμβρανικές πρωτεΐνες, οι οποίες εντοπίζονται στις μεμβράνες όλων των οργανισμών, τα διαμεμβρανικά β-βαρέλια εκφράζονται αποκλειστικά στις Εξωτερικές Μembrάνες των αρνητικών κατά Gram βακτηρίων, των χλωροπλαστών και των μιτοχονδρίων. Λόγω της αποκλειστικής παρουσίας τους στην Εξωτερική Μembrάνη, τα διαμεμβρανικά β-βαρέλια χαρακτηρίζονται, γενικά, ως «Πρωτεΐνες της Εξωτερικής Μembrάνης» (Outer Membrane Proteins, OMPs). Η δομή των διαμεμβρανικών β-βαρελίων, όπως υποδηλώνεται και από το όνομά τους, αποτελείται από διαδοχικούς β-κλώνους, οι οποίοι σχηματίζουν μια β-πτυχωτή επιφάνεια που περιελίσσεται (Εικόνα 11), δημιουργώντας μια δομή «βαρελιού» που δρα σαν διάυλος ή πόρος και διαπερνά την Εξωτερική Μembrάνη (Koeblnik et al., 2000). Μέσω αυτού του διαύλου διέρχεται ο υδατικός διαλύτης κατά μήκος της διπλοστιβάδας, καθώς επίσης και διάφορα είδη ουσιών, είτε παθητικά είτε ελεγχόμενα από το ίδιο το β-βαρέλι (Schulz, 2003). Επιπρόσθετα, τα διαμεμβρανικά β-βαρέλια διαθέτουν μια σειρά από συντηρημένα χαρακτηριστικά στην αλληλουχία και τη δομή τους, τα οποία έχουν χρησιμοποιηθεί σε μεγάλο βαθμό για το σχεδιασμό μεθόδων πρόγνωσης της δομής και της τοπολογίας τους (Tsaousis et al., 2017a):



Εικόνα 11 Παραδείγματα διαμεμβρανικών β-βαρελίων. (Α) Η πορίνη OmpF είναι ένα «τυπικό» διαμεμβρανικό β-βαρέλι, το οποίο δρα σαν διάυλος ιόντων χωρίς ειδικότητα. (Β) Ο υποδοχέας ιόντων σιδήρου FhuA συμμετέχει στη μεταγωγή σήματος των βακτηρίων. Στη δομή παρατηρούνται αλληλεπιδράσεις του β-βαρελίου με ένα λιποπολυσακχαρίτη. (Γ) Η αφυδρογονάση του Λιπιδίου Α (LpxR) είναι ένα β-βαρέλι με ενζυμική ενεργότητα. (Δ) Η πρωτεΐνη CusC λειτουργεί σαν αντλία ιόντων, καταλύοντας την ενεργό μεταφορά του ασβεστίου. (Ε) Η πρωτεΐνη CsgG μεταφέρει τις πρωτεΐνες CsgA και CsgB στο εξωτερικό του βακτηρίου για τη βιοσύνθεση των ινιδίων curli στην *E. coli*.

1. Οι διαμεμβρανικοί β-κλώνοι είναι αμφιπαθικοί, παρουσιάζοντας την εναλλαγή πολικών/φορτισμένων και υδρόφοβων καταλοίπων. Μέσω αυτής της εναλλαγής, η εξωτερική επιφάνεια του β-βαρελίου, η οποία έρχεται σε επαφή με τον πυρήνα της μεμβράνης είναι υδρόφοβη, ενώ η εσωτερική επιφάνειά του, η οποία σχηματίζει τον πόρο και έρχεται σε επαφή με το διαλύτη, είναι πολική.

2. Παρατηρείται υψηλή παρουσία αρωματικών καταλοίπων στα όρια των διαμεμβρανικών τμημάτων, τα οποία εντοπίζονται στην περιοχή επαφής της μεμβράνης με το διαλύτη. Αυτά τα κατάλοιπα σχηματίζουν «αρωματικές ζώνες» στην περιφέρεια του βαρελίου και πιστεύεται πως αλληλεπιδρούν με συγκεκριμένα τμήματα των λιπιδίων της μεμβράνης, όπως οι πολικές κεφαλές των φωσφολιπιδίων αλλά και τα σάκχαρα στη δομή των λιποπολυσακχαριτών.



Εικόνα 12 Γεωμετρικά κριτήρια της δομής των β-βαρελίων. (Α) Διαγραμματική απεικόνιση της μεθόδου υπολογισμού του αριθμού κάμψης S . (Β) Διαγραμματική απεικόνιση του πόρου ενός β-βαρελίου, στην οποία απεικονίζονται η γωνία α των β-κλώνων σε σχέση με τον κατακόρυφο άξονα του βαρελίου, καθώς και η ακτίνα R του πόρου. Προσαρμογή από Chothia et al., 1997.

3. Το N- και το C-άκρο των β-βαρελίων εντοπίζονται στον περιπλαστικό χώρο, με μήκος που μπορεί σε κάποιες περιπτώσεις να ξεπερνά τα 100 αμινοξικά κατάλοιπα.

4. Λόγω της ταυτόχρονης παρουσίας του N- και του C-άκρου στον περιπλαστικό χώρο, τα διαμεμβρανικά β-βαρέλια έχουν σχεδόν πάντα άρτιο αριθμό διαμεμβρανικών β-κλώνων. Το μήκος τους ποικίλλει από 6 ως 22 αμινοξικά κατάλοιπα και εξαρτάται από τη γωνία του κάθε β-κλώνου με τον κατακόρυφο άξονα του βαρελίου. Ωστόσο, σε κάποιες περιπτώσεις οι β-κλώνοι θάβονται μερικώς στη μεμβράνη, με το υπόλοιπο τμήμα τους να εξέχει στον εξωκυτταρικό χώρο και να σχηματίζει ευέλικτες β-φουρκέτες (β-hairpins).

5. Οι διαμεμβρανικοί β-κλώνοι του βαρελίου είναι πάντα αντιπαράλληλοι και συνδέονται γειτονικά μεταξύ τους μέσα από ένα πυκνό δίκτυο υδρογονικών δεσμών.

6. Οι βρόχοι που συνδέουν τα διαμεμβρανικά τμήματα μεταξύ τους είναι, συνήθως, μικρότεροι στον περιπλαστικό χώρο και μεγαλύτεροι στον εξωκυτταρικό χώρο. Ωστόσο, υπάρχουν εξαιρέσεις στον συγκεκριμένο κανόνα, όπως η πρωτεΐνη CusC (Εικόνα 11), η οποία έχει μεγάλες περιπλαστικές περιοχές με δομή α-έλικας και μικρούς εξωκυτταρικούς βρόχους.

7. Τα διαμεμβρανικά β-βαρέλια, σε σύγκριση με τις σφαιρικές υδατοδιαλυτές πρωτεΐνες που διαθέτουν δομή β-βαρελίου, παρουσιάζουν σημαντικά μεγαλύτερη ετερογένεια στην αμινοξική ακολουθία τους.

Κεφάλαιο I - Εισαγωγή

Τα β-βαρέλια γενικά κατηγοριοποιούνται ως προς τη δομή τους με ένα σύστημα γνωστό ως σύστημα « n,S » (Liu, 1998) το οποίο χρησιμοποιεί δύο χαρακτηριστικά, τον αριθμό των β-κλώνων (n) και τον αριθμό κάμψης (shear number, S). Ο αριθμός κάμψης S αποτελεί ουσιαστικά ένα δείκτη του πόσο στραμμένη είναι η β-πτυχωτή επιφάνεια του βαρελιού σε σχέση με τον κατακόρυφο άξονά του. Η τιμή του S υπολογίζεται ως εξής (Εικόνα 12): αν θεωρήσουμε ως σημείο έναρξης ένα κατάλοιπο l στον πρώτο β-κλώνο και μετακινηθούμε κατακόρυφα προς το β-κλώνο, ακολουθώντας διαδοχικά τα κατάλοιπα που συνδέονται μεταξύ τους με υδρογονικό δεσμό, τελικά καταλήγουμε ξανά στον πρώτο β-κλώνο λόγω της στρέψης του βαρελιού. Ωστόσο, το κατάλοιπο του πρώτου κλώνου στο οποίο καταλήγουμε δεν είναι το κατάλοιπο l αλλά το κατάλοιπο k , το οποίο απέχει από το κατάλοιπο l απόσταση ίση με S . Αυτή η απόσταση αποτελεί, ουσιαστικά, την τιμή του αριθμού κάμψης S (Chothia et al., 1997). Η αρχιτεκτονική ενός οποιουδήποτε β-βαρελιού μπορεί να περιγραφεί απλά με το συνδυασμό των τιμών n και S (π.χ. $n=16$, $S=20$). Τα διαμεμβρανικά β-βαρέλια, όπως αναφέρθηκε ήδη, διαθέτουν άρτιο αριθμό β-κλώνων, που τυπικά κυμαίνεται σε τιμές $n=6-24$, ενώ οι τιμές του S έχουν βρεθεί, με βάση τις διαθέσιμες τρισδιάστατες δομές, να κυμαίνονται από $S=8-30$. Πέρα από το κριτήριο « n,S », μια ακόμα ιδιότητα που προσδιορίζεται συχνά στα β-βαρέλια είναι η μέση γωνία κλίσης των β-κλώνων (α) σε σχέση με τον κατακόρυφο άξονα του βαρελιού (Εικόνα 12). Τέλος, ένα επιπλέον κριτήριο για την περιγραφή των διαμεμβρανικών β-βαρελίων είναι η τιμή της ακτίνας του πόρου (R), η οποία σε μεγάλο βαθμό σχετίζεται με τη λειτουργία του κάθε βαρελιού και αποτελεί συχνά ένδειξη για το μέγεθος του υποστρώματος που μεταφέρεται μέσω του πόρου (Schulz, 2003).

Τα διαμεμβρανικά β-βαρέλια λειτουργούν κυρίως σαν πορίνες και εμπλέκονται με τη μεταφορά ουσιών από και προς το εσωτερικό του κυττάρου, χωρίς ωστόσο αυτός να είναι ο μοναδικός μηχανισμός λειτουργίας τους (Schulz, 2003). Συγκεκριμένα, τα διαμεμβρανικά β-βαρέλια έχουν εμπλακεί σε μεγάλο αριθμό λειτουργιών (Εικόνα 11), όπως η μεταφορά ουσιών και θρεπτικών ουσιών, η μεταγωγή σήματος, η βακτηριακή σύζευξη, η σύνθεση και ρύθμιση μηχανισμών για την κίνηση του βακτηριακού κυττάρου (μαστίγια, σμήριγγες κλπ) αλλά και η ενζυμική δραστηριότητα (Koebnik et al., 2000; Schulz, 2003). Επιπρόσθετα, πολλά β-βαρέλια έχουν συσχετιστεί με τη ρύθμιση της άμυνας των αρνητικών κατά Gram βακτηρίων σε

Κεφάλαιο I - Εισαγωγή

εξωτερικούς παράγοντες (π.χ. αντιβιοτικά, χολικά άλατα κλπ), σε συνεργασία με άλλα συστατικά της Εξωτερικής Μεμβράνης όπως οι λιποπολυσακχαρίτες (Akira et al., 2006; Arunmanee et al., 2016). Έτσι, σήμερα τα διαμεμβρανικά β-βαρέλια αποτελούν συχνά στόχο για τη μελέτη της παθολογίας των αρνητικών κατά Gram βακτηρίων και το σχεδιασμό αντιμικροβιακών φαρμάκων (James et al., 2009; Ziervogel & Roux, 2013).

1.3.2 Προσδιορισμός της δομής των διαμεμβρανικών πρωτεϊνών

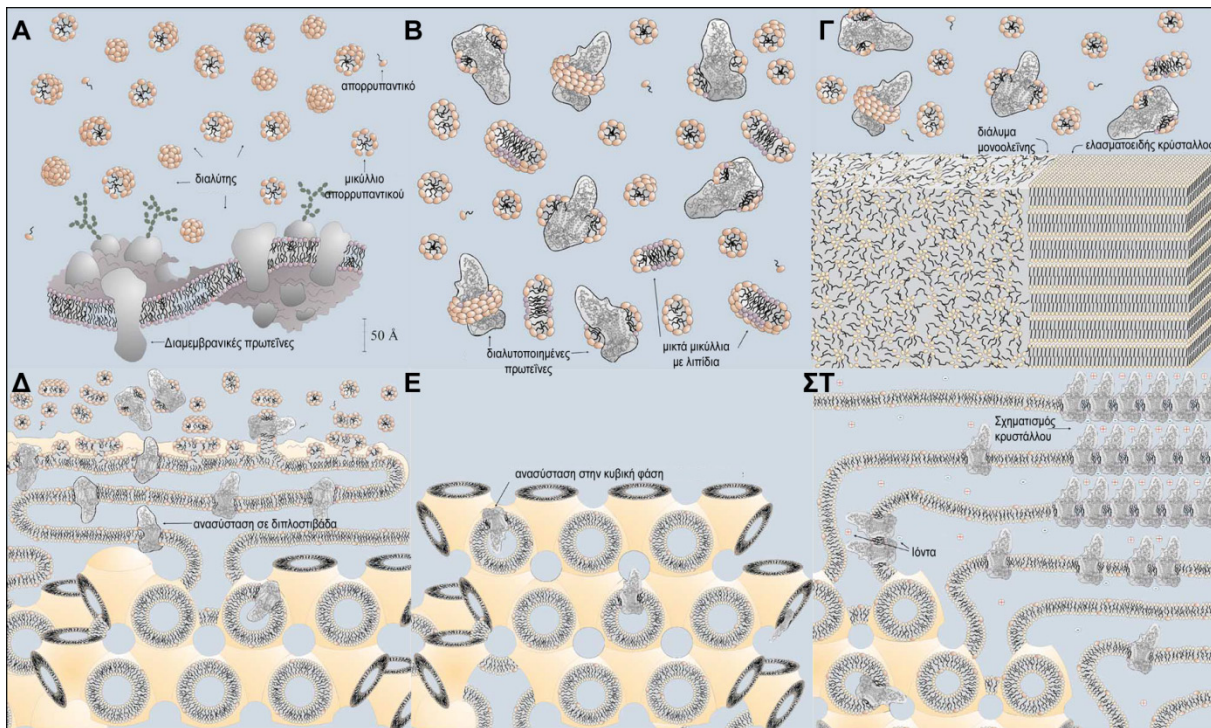
Η μελέτη της δομής και της τοπολογίας των διαμεμβρανικών πρωτεϊνών είναι ιδιαίτερα σημαντική για την κατανόηση της βιολογικής τους λειτουργίας και για την κατανόηση των μηχανισμών που αφορούν την τοπογένεση τους στις μεμβράνες. Η γνώση της τρισδιάστατης δομής μιας πρωτεΐνης σε ατομική διακριτικότητα είναι ένα αποφασιστικό βήμα στην προσπάθεια κατανόησης της βιολογικής της λειτουργίας. Παρά τη σημασία τους, ωστόσο, ο αριθμός των διαθέσιμων δομών για τις διαμεμβρανικές πρωτεΐνες, σε σύγκριση με τον αριθμό των αντίστοιχων δομών των σφαιρικών υδατοδιαλυτών πρωτεϊνών, παραμένει μικρός. Συγκεκριμένα, στο σύνολο των διαθέσιμων δομών της βάσης δεδομένων πρωτεϊνικών δομών Protein Data Bank (PDB), οι οποίες αθροίζονται πλέον στις 150861 (Απρίλιος 2019) μόλις το ~2 % (2269) αφορούν διαμεμβρανικές πρωτεΐνες. Από αυτές τις δομές, οι 2005 αφορούν τις α-ελικοειδείς διαμεμβρανικές πρωτεΐνες, ενώ ο αριθμός των δομών για τα διαμεμβρανικά β-βαρέλια είναι σημαντικά μικρότερος (264).

Τα βασικά προβλήματα που απαντώνται στην προσπάθεια επίλυσης της δομής μιας διαμεμβρανικής πρωτεΐνης είναι συνυφασμένα με τον κατά βάση υδρόφοβο χαρακτήρα της. Έτσι, οι προσπάθειες αποδιάταξης της μεμβράνης με συμβατικά απορρυπαντικά έχουν σε πολλές περιπτώσεις ως συνέπεια την αδυναμία του περαιτέρω καθαρισμού της πρωτεΐνης, με αποτέλεσμα τελικά να μην είναι δυνατή η κρυστάλλωση της (von Heijne, 2007). Ένα δεύτερο πρόβλημα είναι η θερμοδυναμική και πρωτεολυτική αστάθεια πολλών διαμεμβρανικών πρωτεϊνών, οι οποίες συχνά διαθέτουν κακή θερμοδυναμική σταθερότητα ή είναι ευάλωτες σε πρωτεολυτικά ένζυμα. Λύσεις στα παραπάνω προβλήματα έχουν αποτελέσει η χρήση ειδικών απορρυπαντικών για την απομόνωση των υποδοχέων (π.χ. χοληστερικά άλατα), η εφαρμογή

Κεφάλαιο I - Εισαγωγή

μεθόδων όπως η κρυστάλλωση μέσω λιπιδικής μεσόφασης (Caffrey, 2003) και η εισαγωγή ειδικών μεταλλαγών που αυξάνουν την ανθεκτικότητα των πρωτεϊνών στην πρωτεόλυση. Επιπρόσθετα, τα τελευταία χρόνια παρατηρείται αύξηση στον αριθμό των τρισδιάστατων δομών διαμεμβρανικών πρωτεϊνών οι οποίες προσδιορίζονται με άλλες μεθόδους εκτός της κρυσταλλογραφίας ακτίνων-Χ, όπως η κρυο-ηλεκτρονική μικροσκοπία, με ποιότητα συχνά εφάμιλλη των κρυσταλλικών δομών. Έτσι, πλέον ο αριθμός των τρισδιάστατων δομών των διαμεμβρανικών πρωτεϊνών έχει αυξηθεί σημαντικά. Ωστόσο, η διαδικασία προσδιορισμού της δομής μιας διαμεμβρανικής πρωτεΐνης παραμένει μια δύσκολη διαδικασία.

Λόγω της δυσκολίας στον πειραματικό προσδιορισμό της δομής των διαμεμβρανικών πρωτεϊνών, έχουν αναπτυχθεί διάφορες υπολογιστικές μέθοδοι για την πρόγνωση της δομής και της τοπολογίας τους (Πίνακας 2). Η πλειονότητα αυτών των μεθόδων βασίζεται σε μεθόδους μηχανικής μάθησης όπως τα Κρυφά Μοντέλα Markov (Hidden Markov Models, HMMs) και τα Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks, ANNs) και εφαρμόζονται στην πρόγνωση τόσο της τοπολογίας των διαμεμβρανικών πρωτεϊνών όσο και στον εντοπισμό χαρακτηριστικών όπως η ύπαρξη και η θέση του σηματοδοτικού πεπτιδίου (Bagos et al., 2004; Hayat & Elofsson, 2012; Kall et al., 2004; Petersen et al., 2011; Tsaousis et al., 2014; Tsirigos et al., 2015; Viklund & Elofsson, 2008; Viklund et al., 2008). Η τρισδιάστατη δομή των διαμεμβρανικών πρωτεϊνών μελετάται επίσης υπολογιστικά, μέσα από μεθόδους όπως η ομόλογη προτυποποίηση (Homology Modeling). Τέλος, η δυναμική συμπεριφορά των διαμεμβρανικών πρωτεϊνών και οι αλληλεπιδράσεις τους με τα στοιχεία της διπλοστιβάδας μελετώνται συχνά μέσα από μοριακές προσομοιώσεις όπως η μηχανική Monte Carlo και η Μοριακή Δυναμική (Molecular Dynamics).



Εικόνα 13 Η μέθοδος της λιπιδικής μεσόφασης για την κρυστάλλωση διαμεμβρανικών πρωτεϊνών. (Α) Πραγματοποιείται διαλυτοποίηση του κυτταρικού δείγματος σε υδατικό διάλυμα το οποίο περιέχει μικύλλια απορρυπαντικών. (Β) Κατά την αποδιάταξη του δείγματος οι πρωτεΐνες «παγιδεύονται» σε μικύλλια του απορρυπαντικού, τα μόρια του οποίου προσδένονται στην υδρόφοβη επιφάνεια των διαμεμβρανικών τμημάτων, προστατεύοντάς τα από το διαλύτη. Έτσι, επιτυγχάνεται η διαλυτοποίηση των πρωτεϊνών και αποτρέπεται ο σχηματισμός ιζήματος. Σε μικύλλια μπορούν να παγιδευτούν, εκτός από τις πρωτεΐνες, και λιπίδια της μεμβράνης. (Γ) Το υδατικό διάλυμα που περιέχει τις διαλυτοποιημένες πρωτεΐνες εισάγεται σε μη-πολικό διάλυμα συνθετικών λιπιδίων (π.χ. μονοολεΐνη). Η εισαγωγή του υδατικού διαλύματος στο λιπιδικό διάλυμα ωθεί την αυτοοργάνωση των λιπιδίων σε δομές, η πρώτη φάση των οποίων είναι ένας ελασματοειδής κρύσταλλος αποτελούμενος από λιπιδικές διπλοστιβάδες. (Δ) Πραγματοποιείται ανασύσταση των πρωτεϊνών στη διπλοστιβάδα μονοολεΐνης και, παράλληλα, ξεκινά ο σχηματισμός της κυβικής φάσης. (Ε) Ολοκληρώνεται ο σχηματισμός της κυβικής φάσης, με τις διαμεμβρανικές πρωτεΐνες να έχουν ανασυσταθεί στον κύβο των λιπιδίων. (ΣΤ) Η σταδιακή προσθήκη ιοντικού διαλύματος έχει σαν αποτέλεσμα τη σταδιακή αποδιάταξη της κυβικής φάσης, το οργανωμένο πακετάρισμα των διαμεμβρανικών πρωτεϊνών και τη δημιουργία του κρυστάλλου. Προσαρμογή από Caffrey, 2003.

Κεφάλαιο I - Εισαγωγή

Πίνακας 2 Παραδείγματα υπολογιστικών μεθόδων πρόγνωσης της τοπολογίας των διαμεμβρανικών πρωτεϊνών.

Μέθοδος	Περιγραφή
TMHMM	Πρόγνωση τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών με τη χρήση HMMs (Krogh et al., 2001).
Phobius	Συνδυασμένη πρόγνωση τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών και σηματοδοτικού πεπτιδίου με τη χρήση HMMs (Kall et al., 2004).
OCTOPUS	Πρόγνωση τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών με τη χρήση ANNs (Viklund & Elofsson, 2008).
SPOCTOPUS	Συνδυασμένη πρόγνωση τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών και σηματοδοτικού πεπτιδίου με τη χρήση ANNs (Viklund et al., 2008).
TOPCONS	Συναινετικός αλγόριθμος πρόγνωσης τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών, ο οποίος συνδυάζει τα αποτελέσματα 5 διαφορετικών μεθόδων πρόγνωσης (Tsirigos et al., 2015).
HMMpTM	Συνδυασμένη πρόγνωση τοπολογίας α-ελικοειδών διαμεμβρανικών πρωτεϊνών και μετα-μεταφραστικών τροποποιήσεων με τη χρήση HMMs (Tsaousis et al., 2014).
PRED-TMBB	Πρόγνωση τοπολογίας διαμεμβρανικών β-βαρελιών με τη χρήση HMMs (Bagos et al., 2004).
BOCTOPUS	Πρόγνωση τοπολογίας διαμεμβρανικών β-βαρελιών με τη συνδυασμένη χρήση HMMs και Support Vector Machines (Hayat & Elofsson, 2012).

1.4 Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών

Οι βιολογικές διεργασίες σπάνια επιτελούνται από μία μόνο πρωτεΐνη. Αντίθετα, το σύνολο σχεδόν των λειτουργιών του κυττάρου, από την ενζυμική κατάλυση και τη μηχανική στήριξη ως τους πολύπλοκους μηχανισμούς της μεταγωγής σήματος και της ρύθμισης της γονιδιακής έκφρασης, πραγματοποιούνται από ομάδες πρωτεϊνών, οι οποίες λειτουργούν συντονισμένα (De Las Rivas & Fontanillo, 2010). Με τον όρο «αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών» αναφερόμαστε στις επαφές δύο ή περισσότερων πρωτεϊνών, μέσω των οποίων επιτελείται μια βιολογική λειτουργία ή οργανώνεται κάποιος δομικός σχηματισμός με λειτουργική σημασία. Η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών αποτελεί ζωτικό κομμάτι των βιολογικών επιστημών (Taylor et al., 2009). Η εύρεση του μηχανισμού αλληλεπίδρασης μιας πρωτεΐνης συμβάλλει σημαντικά στον χαρακτηρισμό του τρόπου λειτουργίας της, ενώ η μελέτη του συνόλου των πρωτεϊνών που αλληλεπιδρούν και, συντονισμένα, επιτελούν κάποια βιολογική διεργασία, αποκαλύπτει τους μηχανισμούς που διέπουν αυτή τη λειτουργία. Τέλος, η μελέτη των μηχανισμών που διαταράσσουν αυτές τις

Κεφάλαιο I - Εισαγωγή

αλληλεπιδράσεις είναι ζωτικής σημασίας στη βιοϊατρική έρευνα, τόσο για την αποσαφήνιση των δομικών και των λειτουργικών χαρακτηριστικών διαφόρων ασθενειών όσο και για το σχεδιασμό νέων, πιο αποτελεσματικών φαρμάκων (Titz et al., 2004).

Σε μοριακό επίπεδο οι αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών περιλαμβάνουν το σύνολο των διαμοριακών επαφών ανάμεσα σε δύο ή περισσότερες πολυπεπτιδικές αλυσίδες. Αυτές περιλαμβάνουν κατά κύριο λόγο τις ασθενείς αλληλεπιδράσεις ανάμεσα στα κατάλοιπα των πρωτεϊνών (Χαμόδρακας, 1993), δηλαδή τις ηλεκτροστατικές αλληλεπιδράσεις, τους δεσμούς υδρογόνου, τις αλληλεπιδράσεις van der Waals και τις υδρόφοβες αλληλεπιδράσεις (Alberts et al., 2002; Janin & Chothia, 1990). Επιπρόσθετα, στις αλληλεπιδράσεις ανάμεσα στις πρωτεΐνες μπορούν να συμμετέχουν και μη πρωτεϊνικά μόρια (λιπίδια, σάκχαρα, ιόντα κλπ), τα οποία συχνά ρυθμίζουν τη λειτουργία του συμπλόκου (π.χ. η αίμη της αιμοσφαιρίνης). Η αλληλεπίδραση ανάμεσα σε δυο πρωτεΐνες μπορεί να είναι ειδική, με την έννοια ότι μία πρωτεΐνη μπορεί να αναγνωρίζει και να αλληλεπιδρά επιλεκτικά μόνο με μία άλλη συγκεκριμένη πρωτεΐνη ή με μια περιορισμένη ομάδα πρωτεϊνών με κοινά χαρακτηριστικά (Kastritis & Bonvin, 2013). Χαρακτηριστικό παράδειγμα ειδικών αλληλεπιδράσεων είναι οι αλληλεπιδράσεις ανάμεσα στα αντιγόνα και τα αντισώματα (Sela-Culang et al., 2013), οι οποίες έχουν χαρακτηριστική συμπληρωματικότητα «κλειδιού-κλειδαριάς», καθώς και οι αλληλεπιδράσεις των ενζύμων (π.χ. πρωτεϊνικές κινάσες), τα οποία μπορούν να δράσουν μόνο σε υποστρώματα με συγκεκριμένα χαρακτηριστικά (Alberts et al., 2002). Από την άλλη, μεγάλος αριθμός πρωτεϊνικών αλληλεπιδράσεων είναι μη ειδικές και εξαρτώνται από χαρακτηριστικά όπως η ταυτόχρονη έκφραση των πρωτεϊνών στον ίδιο κυτταρικό τύπο. Χαρακτηριστικό παράδειγμα μη ειδικών αλληλεπιδράσεων είναι οι αλληλεπιδράσεις των G-πρωτεϊνών (Oldham & Hamm, 2008), οι οποίες παρά το μικρό αριθμό τους έρχονται σε επαφή με μεγάλο αριθμό υποδοχέων (GPCRs) και ενδοκυτταρικών πρωτεϊνών και ρυθμίζουν, μέσω αυτών, μεγάλο αριθμό σηματοδοτικών μονοπατιών (Kastritis & Bonvin, 2013).

Πέρα από τις διαμοριακές επαφές, τα πρωτεϊνικά σύμπλοκα μπορούν να χαρακτηριστούν με βάση την ομοιότητα των υπομονάδων τους. Έτσι, ένα πρωτεϊνικό σύμπλοκο μπορεί να αποτελείται από δύο ή περισσότερα αντίγραφα της ίδιας πρωτεΐνης (ομομερές), ή από διαφορετικές πρωτεΐνες (ετερομερές). Τέλος, ένας σημαντικός παράγοντας

Κεφάλαιο I - Εισαγωγή

στον καθορισμό των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών σε ένα σύμπλοκο είναι ο χρόνος ζωής της μεταξύ τους αλληλεπίδρασης. Τα σύμπλοκα πρωτεϊνών – πρωτεϊνών μπορούν να είναι μόνιμα ή παροδικά, με το χρόνο ζωής του συμπλόκου να σχετίζεται άμεσα με τη λειτουργία του. Παραδείγματα μόνιμων συμπλόκων περιλαμβάνουν την αιμοσφαιρίνη, ένα υποχρεωτικό ετεροτετραμερές σφαιρινών α και β, τις ριβοσωμικές υπομονάδες, αποτελούμενες από μεγάλο αριθμό πρωτεϊνών αλλά και μορίων rRNA καθώς και σύμπλοκα που συμμετέχουν στη μεταφορά ουσιών, όπως πολλά ιοντικά κανάλια, τα οποία αποτελούνται από πολλαπλές υπομονάδες. Από την άλλη, οι αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών σε διεργασίες όπως η μεταγωγή σήματος είναι συνήθως παροδικές, και διαρκούν μόνο όσο χρειάζεται για να επιτευχθεί η μεταφορά της πληροφορίας από το εξωτερικό στο εσωτερικό του κυττάρου και να πραγματοποιηθεί η αντίστοιχη κυτταρική απόκριση (Kastritis & Bonvin, 2013).

Η πειραματική μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών πραγματοποιείται με διάφορες μεθόδους, οι οποίες μπορούν να διακριθούν σε μεθόδους «υψηλής απόδοσης» («high throughput») και σε μεθόδους «χαμηλής απόδοσης» («low throughput»). Οι μέθοδοι υψηλής απόδοσης εφαρμόζονται σε πειράματα μεγάλης κλίμακας, και έχουν σαν στόχο τη μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών σε επίπεδο συστήματος, επιχειρώντας τον προσδιορισμό και τη μελέτη του μεγαλύτερου δυνατού αριθμού πρωτεϊνικών αλληλεπιδράσεων σε έναν οργανισμό ή έναν κυτταρικό ιστό. Χαρακτηριστικά παραδείγματα τέτοιων μεθόδων αποτελούν η Φασματομετρία Μάζας (Abu-Farha et al., 2008), η μέθοδος διϋβριδισμού στο ζυμομύκητα (Yeast-Two hybrid) (Walhout & Vidal, 2001), οι πρωτεϊνικές μικροσυστοιχίες (Melton, 2004), αλλά και προγράμματα δομικής γονιδιωματικής (Structural Genomics) τα οποία προσανατολίζονται στον προσδιορισμό της δομής για μεγάλο αριθμό πρωτεϊνικών συμπλόκων (Teichmann et al., 2001). Οι χιλιάδες αλληλεπιδράσεις που προκύπτουν από τις μεθόδους υψηλής απόδοσης επιτρέπουν τη δημιουργία ενός «σκελετού» του δικτύου των πρωτεϊνικών αλληλεπιδράσεων που λαμβάνουν χώρα στο υπό μελέτη σύστημα (Titz et al., 2004). Ωστόσο, τα αποτελέσματά τους χαρακτηρίζονται συχνά από χαμηλής ποιότητας δεδομένα, τα οποία χρειάζεται να ελεγχθούν με πιο εξειδικευμένες πειραματικές μεθόδους για την επιβεβαίωσή τους. Συγκεκριμένα, οι

Κεφάλαιο I - Εισαγωγή

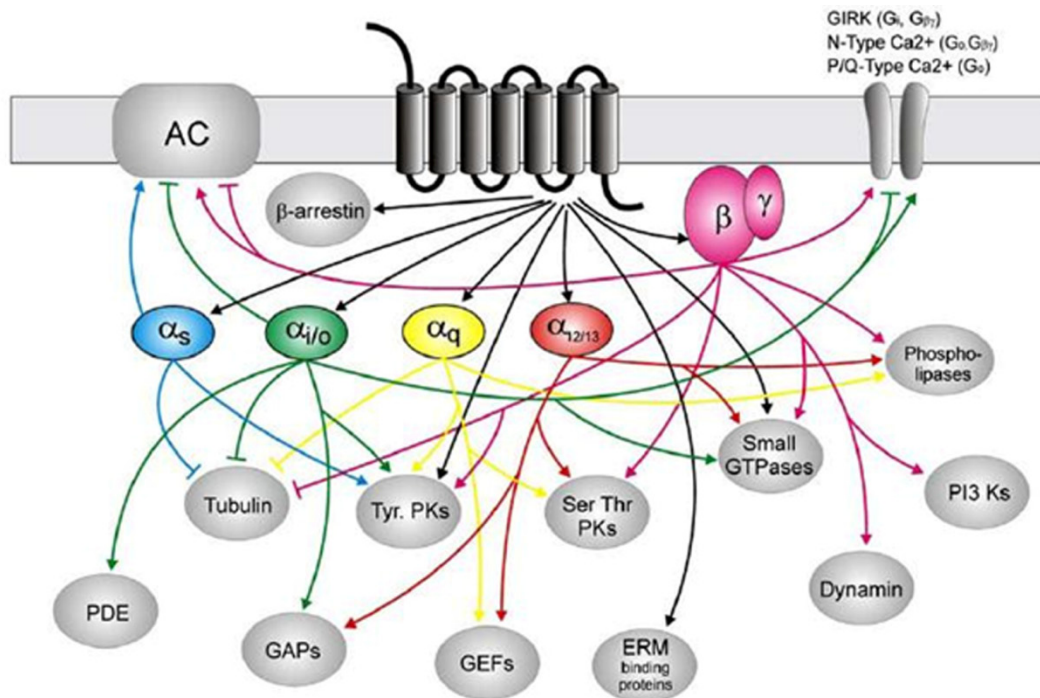
μέθοδοι υψηλής απόδοσης έχουν δεχθεί ιδιαίτερα σφοδρή κριτική ως προς την αποτελεσματικότητά τους, καθώς χαρακτηρίζονται από πολύ μεγάλο ποσοστό ψευδώς θετικών αλληλεπιδράσεων το οποίο συχνά ανέρχεται στο 70-90% των παρατηρήσεων (Qi et al., 2006; Rual et al., 2005). Από την άλλη, οι μέθοδοι χαμηλής απόδοσης εφαρμόζονται σε πειράματα μικρής κλίμακας και έχουν σαν στόχο τη λεπτομερή μελέτη των αλληλεπιδράσεων ανάμεσα στις πρωτεΐνες ενός συμπλόκου. Οι μέθοδοι μικρής κλίμακας ουσιαστικά περιλαμβάνουν την πλειονότητα των βιοχημικών πειραμάτων, από τις μελέτες συνανοσοκατακρήμνισης και τα πειράματα μεταλλαξιγένεσης έως τις μελέτες μεταφοράς ενέργειας καθώς και βιοφυσικές τεχνικές όπως η κρυσταλλογραφία ακτίνων-Χ, ο πυρηνικός μαγνητικός συντονισμός (NMR) και η κρυο-ηλεκτρονική μικροσκοπία (Alberts et al., 2002; Rosenberg, 2005).

Πολύ συχνά η πειραματική μελέτη των πρωτεϊνικών αλληλεπιδράσεων είναι δύσκολη, τόσο λόγω του κόστους όσο και λόγω της δυσκολίας στη διαχείριση των πρωτεϊνών. Έτσι, η υπολογιστική μελέτη των πρωτεϊνικών αλληλεπιδράσεων έχει γνωρίσει ιδιαίτερη άνθιση, με τη χρήση μεθόδων από τα πεδία της Βιοπληροφορικής, της Υπολογιστικής Βιολογίας και της Θεωρητικής Βιοφυσικής. Οι διαμοριακές επαφές πρωτεϊνών – πρωτεϊνών μελετούνται τόσο σε επίπεδο αλληλουχίας, με την ανάπτυξη προγνωστικών μεθόδων που βασίζονται σε μεθόδους μηχανικής μάθησης όπως τα τεχνητά νευρωνικά δίκτυα και τα Κρυφά Μοντέλα Markov, φυλογενετικές αναλύσεις και την εξελικτική ιστορία των αλληλουχιών, όσο και σε επίπεδο δομής, με μεθόδους μοριακής προτυποποίησης, αγκυροβόλησης πρωτεϊνών–πρωτεϊνών και βιομοριακές προσομοιώσεις όπως η Μοριακή Δυναμική (βλ. Ενότητα 2.1). Τέλος, η θεωρία δικτύων γίνεται όλο και πιο συχνή τα τελευταία χρόνια στο χώρο της Βιολογίας Συστημάτων μέσω των βιολογικών δικτύων αλληλεπιδράσεων (βλ. Ενότητα 2.2), τα οποία μπορούν να οπτικοποιήσουν και να αναλύσουν μεγάλο όγκο δεδομένων για την περιγραφή αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών σε επίπεδο συστήματος (Rajagopala et al., 2014), επιτρέποντας έτσι τη συνολική περιγραφή των αλληλεπιδράσεων ενός κυττάρου, σε αντιστοιχία με τις πειραματικές μεθόδους υψηλής απόδοσης (Pavlopoulos et al., 2011).

1.5 Αλληλεπιδράσεις πρωτεϊνών–πρωτεϊνών σε διαμεμβρανικές πρωτεΐνες

1.5.1 Συζευγμένοι με G-πρωτεΐνες Υποδοχείς (GPCRs)

Οι συζευγμένοι με G-πρωτεΐνες υποδοχείς (G-protein coupled receptors, GPCRs) αποτελούν τη μεγαλύτερη και πιο ποικιλόμορφη υπερικογένεια διαμεμβρανικών υποδοχέων στους ευκαρυωτικούς οργανισμούς (Rosenbaum et al., 2009). Είναι διαμεμβρανικές πρωτεΐνες, οι οποίες εντοπίζονται κυρίως στην κυτταρική μεμβράνη των ευκαρυωτικών κυττάρων. Μέχρι σήμερα είναι γνωστά πάνω από 800 γονίδια για GPCRs στο ανθρώπινο γονιδίωμα (περίπου 2% του γονιδιώματος) τα οποία, μέσω μηχανισμών εναλλακτικού ματίσματος, μπορούν να εκφράσουν πάνω από 2000 διαφορετικούς υποδοχείς (Almen et al., 2009). Αντίστοιχος αριθμός υποδοχέων έχουν βρεθεί και σε θηλαστικά όπως ο ποντικός, ενώ σε άλλους οργανισμούς όπως η *Drosophila melanogaster* ή ο *Caenorhabditis elegans* το ποσοστό των GPCRs σε σχέση με το συνολικό γονιδίωμα είναι ακόμα μεγαλύτερο (Fredriksson & Schioth, 2005). Τα κύρια μονοπάτια μεταγωγής σήματος των GPCRs επιτελούνται μέσω των αλληλεπιδράσεων με τις ετεροτριμερείς G-πρωτεΐνες οι οποίες δρουν σαν «μοριακοί διακόπτες» για τη μεταγωγή μηνυμάτων από τον εξωκυττάριο χώρο στο εσωτερικό του κυττάρου. Ενεργοποιούμενοι από προσδέτες με ιδιότητα αγωνιστή, οι GPCRs δρουν ως παράγοντες ανταλλαγής νουκλεοτιδίων γουανίνης (Guanine nucleotide Exchange Factors, GEFs) για τις G-πρωτεΐνες, που στη συνέχεια θα μεταδώσουν το σήμα ενδοκυτταρικά μέσω της ενεργοποίησης πρωτεϊνών-εκτελεστών (effectors), όπως ιοντικά κανάλια, ένζυμα και διάφοροι ρυθμιστικοί παράγοντες (Oldham & Hamm, 2008). Πέρα από αυτό το μοντέλο όμως, έχει βρεθεί ότι οι GPCRs μπορούν να συμμετέχουν σε ποικίλες πρωτεϊνικές αλληλεπιδράσεις με άλλα μόρια όπως οι κινάσες τύπου GRK (G-protein coupled Receptor Kinases) και οι β-αρρεστίνες, που παράγουν αποτελέσματα σε συνδυασμό με τη δράση των G-πρωτεϊνών ή ακόμα και ανεξάρτητα από αυτές (Rosenbaum et al., 2009).



Εικόνα 14 Τα σηματοδοτικά μονοπάτια των GPCRs. Ένας υποδοχέας μπορεί να αλληλεπιδράσει με διάφορες G-πρωτεΐνες, αποτελούμενες από την υπομονάδα Gα (G_{α_s} , $G_{\alpha_{i/o}}$, G_{α_q} , $G_{\alpha_{12/13}}$) και το ετεροδιμερές Gβγ. Μέσω των G-πρωτεϊνών οι GPCRs ρυθμίζουν τη λειτουργία μεγάλου αριθμού πρωτεϊνών – τελεστών που περιλαμβάνουν ένζυμα (π.χ. αδενυλική κυκλάση, κινάσες Ser/Thr), ρυθμιστικούς παράγοντες (π.χ. GEFs), ιοντικά κανάλια και στοιχεία του κυτταροσκελετού. Επιπρόσθετα, οι GPCRs μπορούν να δράσουν μέσω εναλλακτικών μονοπατιών, όπως το μονοπάτι των β-αρρεσινών. Η ίδια G-πρωτεΐνη μπορεί να ρυθμίσει τη λειτουργία περισσότερων του ενός τελεστών και, αντίστροφα, ένας τελεστής μπορεί να ρυθμιστεί από περισσότερες από μία G-πρωτεΐνες. Προσαρμογή από Woehler & Ponimaskin, 2009.

Οι GPCRs αναγνωρίζουν και προσδένουν ένα ευρύ σύνολο σηματοδοτικών μορίων (προσδέτες), τα οποία περιλαμβάνουν ορμόνες και νευροδιαβιβαστές, αμίνες, πεπτίδια, αμινοξέα, νουκλεοτίδια, φωσfolιπίδια, λιπαρά οξέα, και ιόντα καθώς και διάφορους εξωγενείς προσδέτες, όπως παράγοντες οσμής και γεύσης αλλά και ερεθίσματα μη χημικής φύσεως, όπως η φωτεινή ακτινοβολία (Kristiansen, 2004). Τα σηματοδοτικά μονοπάτια τους ρυθμίζουν μεγάλη ποικιλία κυτταρικών αποκρίσεων και λειτουργιών του οργανισμού, ανάμεσα στις οποίες συμπεριλαμβάνονται η ομοιόσταση, η κυτταρική αύξηση και διαίρεση, η απόπτωση και ο προγραμματισμένος κυτταρικός θάνατος, η μετάδοση του νευρικού παλμού, η αναγνώριση ορμονών, η γλυκόλυση και γλυκονεογένεση, καθώς και οι αισθήσεις της όρασης, της όσφρησης και της γεύσης. Έτσι, οι GPCRs έχουν εμπλακεί σε πλήθος ασθενειών,

Κεφάλαιο I - Εισαγωγή

ανάμεσα στα οποία περιλαμβάνονται νευροεκφυλιστικές ασθένειες, μεταβολικά σύνδρομα, το άσθμα, η καρδιακή ανεπάρκεια, διάφοροι τύποι καρκίνου και η μόλυνση από ιούς όπως ο HIV και οι ερπητοϊοί. Κατά συνέπεια, εκτιμάται ότι οι GPCRs αποτελούν στόχο για πάνω από 40% των φαρμάκων που υπάρχουν στην αγορά (Cherezon et al., 2010).

1.5.1.1 Ταξινόμηση των GPCRs

Στην προσπάθεια κατάταξης των GPCRs έχουν χρησιμοποιηθεί διάφορα συστήματα ταξινόμησης. Κάποια ταξινομούν τους υποδοχείς με βάση τη δομή τους, άλλα με βάση την εξελικτική τους καταγωγή και κάποια τους κατατάσσουν σε ομάδες ανάλογα με τους προσδέτες με τους οποίους προσδένονται, κανένα όμως δεν έχει τύχει πλήρους αποδοχής από την επιστημονική κοινότητα. Το πιο δημοφιλές σύστημα κατατάσσει τους υποδοχείς σε έξι κλάσεις ή Οικογένειες (A, B, C, D, E και F) στο ανώτερο επίπεδο, βασιζόμενο κυρίως στην ομοιότητα των αμινοξικών ακολουθιών, και προχωρεί περαιτέρω στην κατάταξή τους σε υποοικογένειες (Kolakowski, 1994). Πιο αναλυτικά:

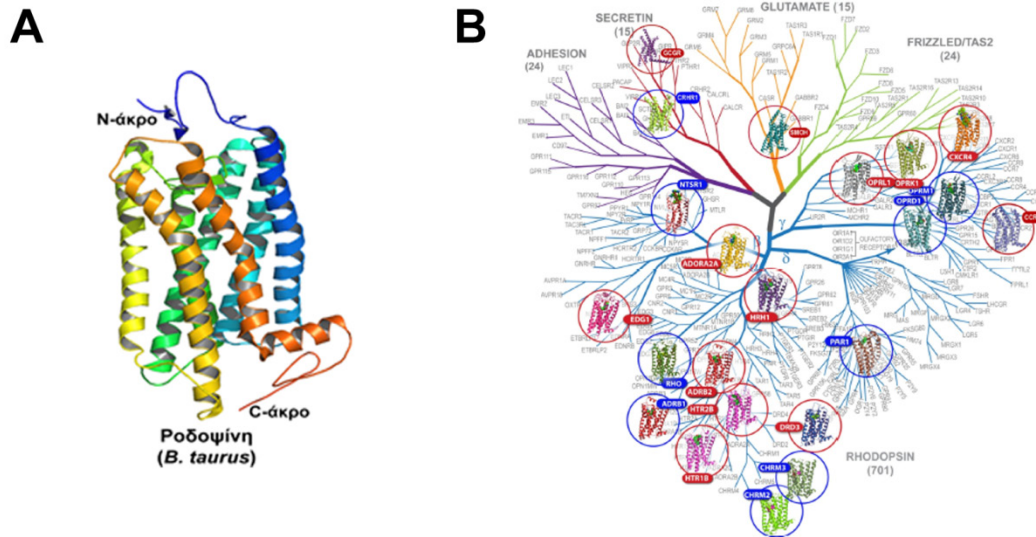
1. Κλάση A (ομοιάζοντες με Ροδοψίνη). Σε αυτήν περιλαμβάνονται υποδοχείς με μεγάλη ετερογένεια προσδετών, όπως βιογενείς αμίνες, γλυκοπρωτεΐνες, πεπτίδια, νουκλεοτίδια κλπ, καθώς και το σύνολο των οσφρητικών υποδοχέων. Η ομαδοποίηση των υποδοχέων γίνεται με βάση τη δομή τους, που μοιάζει με αυτή της Ροδοψίνης.
2. Κλάση B (ομοιάζοντες με σεκρετίνη). Υποδοχείς που προσδέουν πεπτιδικές ορμόνες όπως η σεκρετίνη, η γλυκαγόνη και η καλσιτονίνη, καθώς και υποδοχείς προσκόλλησης (Adhesion), οι οποίοι συμμετέχουν δομικά στις φυσικές επαφές των κυττάρων μεταξύ τους.
3. Κλάση C (Μεταβοτροπικοί υποδοχείς του γλουταμικού). Υποδοχείς του γλουταμικού και άλλων φορτισμένων αμινοξέων, του γ-αμινοβουτυρικού οξέος, υποδοχείς ιόντων Ca^{2+} και οι υποδοχείς γεύσης Τύπου I.
4. Κλάση D. Υποδοχείς φερομονών σε διάφορα είδη μυκήτων.
5. Κλάση E. Υποδοχείς του cAMP, οι οποίοι έχουν βρεθεί αποκλειστικά σε πρωτόζωα του γένους *Dictyostellium*.

Κεφάλαιο I - Εισαγωγή

6. Κλάση F (Frizzled / Smoothened). Οι υποδοχείς των σηματοδοτικών μονοπατιών Frizzled/Smoothened, καθώς και οι υποδοχείς γεύσης Τύπου II. Πρόσφατα, ωστόσο, προτάθηκε οι υποδοχείς γεύσης Τύπου II να διαχωριστούν από τους υπόλοιπους GPCRs της Κλάσης F σε μια ξεχωριστή Κλάση T2 (Taste-2), λόγω της φυλογενετικής διαφοροποίησής τους από τους υποδοχείς Frizzled και Smoothened (Munk et al., 2016).

1.5.1.2 Η τοπολογία και η δομή των GPCRs

Όλα τα μέλη της υπερικογένειας των GPCRs μοιράζονται μια κοινή τοπολογία και διαθέτουν 7 διαμεμβρανικές έλικες, τρεις ενδο- και τρεις εξωκυτταρικούς βρόχους, μια εξωκυτταρική περιοχή – αμινοτελικό άκρο και μια ενδοκυτταρική περιοχή – καρβοξυτελικό άκρο (Rosenbaum et al., 2009). Τα διαμεμβρανικά τμήματα των διάφορων υποδοχέων μοιράζονται μεγαλύτερο βαθμό συντήρησης της αμινοξικής τους ακολουθίας, ενώ τα εξωκυττάρια και ενδοκυττάρια τμήματα εμφανίζουν εκτενή μεταβλητότητα σε μέγεθος και πολυπλοκότητα. Οι εξωκυττάρια και διαμεμβρανικές περιοχές του υποδοχέα τυπικά εμπλέκονται στην αναγνώριση και πρόσδεση των διάφορων μορίων-προσδετών, ενώ τα ενδοκυττάρια τμήματα είναι σημαντικά για τη μεταγωγή σήματος και την ανατροφοδοτούμενη διαφοροποίηση της λειτουργίας του υποδοχέα (Kristiansen, 2004). Η κοινή τοπολογία που περιγράφηκε επιβεβαιώνεται από τα κρυσταλλογραφικά δεδομένα που έχουν κατατεθεί στη βάση δεδομένων Protein Data Bank (PDB), και είναι πλέον ευρύτερα αποδεκτή. Η πρώτη δομή GPCR που προσδιορίστηκε κρυσταλλογραφικά ήταν αυτή της Ροδοψίνης του βοδιού (Palczewski et al., 2000), ενώ πλέον υπάρχουν διαθέσιμες δομές για πάνω από 35 GPCRs, προερχόμενους από όλες τις Κλάσεις των GPCRs των Μεταζώων (Stevens et al., 2013). Οι δομές των περισσότερων από τους παραπάνω υποδοχείς έχουν λυθεί επανειλημμένα, σε διαφορετική διακριτικότητα και σε σύμπλοκα με διάφορα μόρια (Munk et al., 2019).



Εικόνα 15 Η συντηρημένη δομή των GPCRs. (Α) Η κρυσταλλική δομή της Ροδοψίνης του *B. taurus*, του πρώτου GPCR με πειραματικά προσδιορισμένη δομή. (Β) Η πρόοδος της κρυσταλλογραφίας ακτινών-Χ στο πεδίο των GPCRs. Απεικονίζονται οι διαθέσιμες δομές των GPCRs από τις οικογένειες των Μεταζώων, πάνω στο φυλογενετικό δέντρο των υποδοχέων. Προσαρμογή από Stevens et al, 2013.

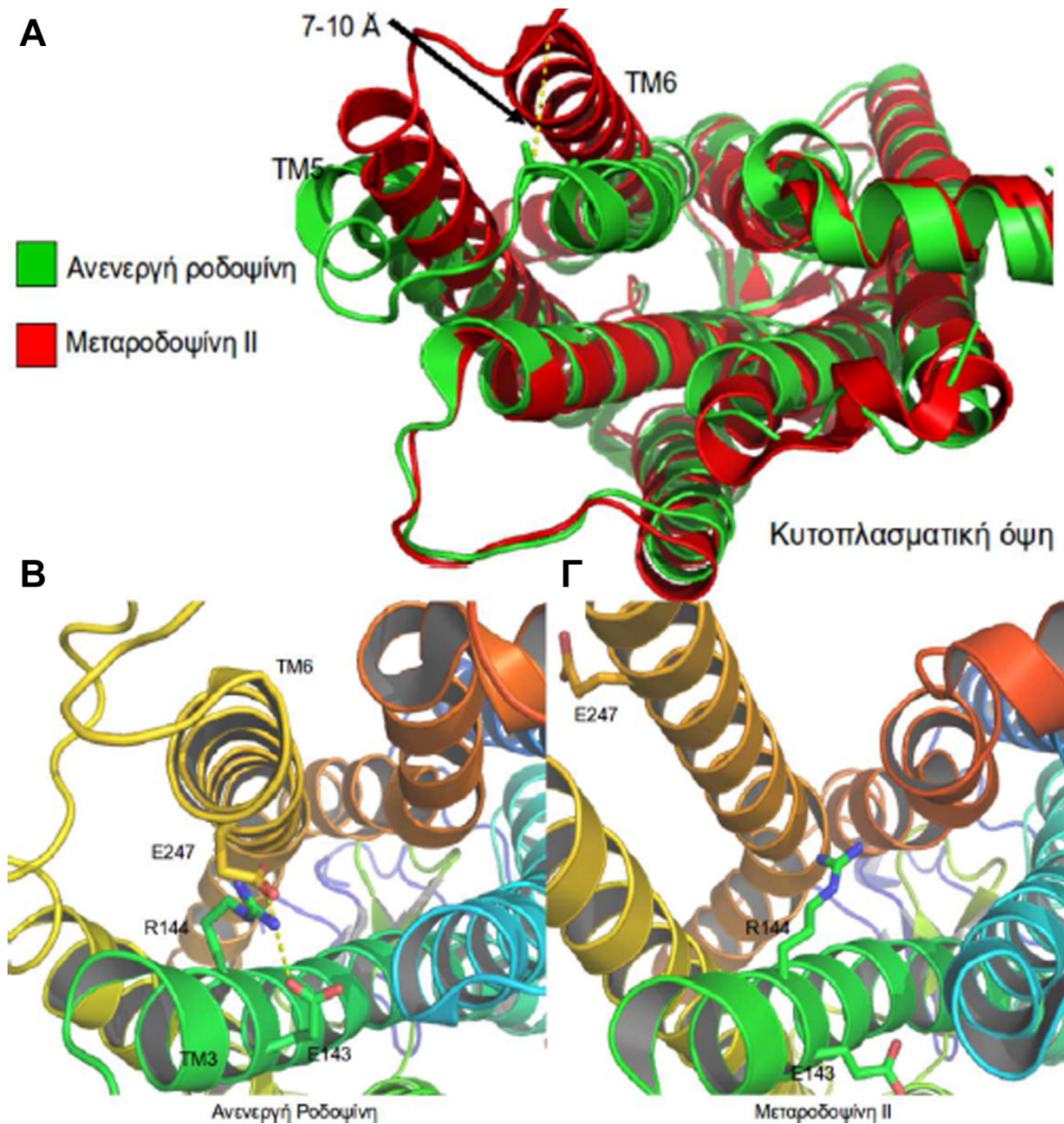
Οι GPCRs βρίσκονται σε μια ισορροπία μεταξύ δύο διακριτών στερεοδιατάξεων: μια ανενεργό κατάσταση και μια ενεργό κατάσταση, κατά την οποία ο υποδοχέας έχει την ικανότητα να επάγει την ανταλλαγή νουκλεοτιδίων στις G-πρωτεΐνες. Η σύνδεση του προσδέτη με τον υποδοχέα προκαλεί την έναρξη της μεταγωγής σήματος, μετατοπίζοντας την ισορροπία προς την ενεργό στερεοδιάταξη. Οι κρυσταλλογραφικά προσδιορισμένες δομές ενεργοποιημένων υποδοχέων για μια σειρά από GPCRs όπως η Ροδοψίνη, ο αδρενεργικός υποδοχέας $\beta 2$ και ο αδενοσινικός υποδοχέας A2A δείχνουν πως οι GPCRs της Κλάσης A ακολουθούν ένα κοινό πρότυπο ενεργοποίησης (Park et al., 2008; Rasmussen et al., 2011). Στην ανενεργή τους κατάσταση οι υποδοχείς σταθεροποιούνται μέσω μιας ιοντικής γέφυρας που σχηματίζεται ανάμεσα στα κυτοπλασματικά μέρη της 3ης (TM3) και της 6ης (TM6) διαμεμβρανικής έλικας. Συγκεκριμένα, στην κυτοπλασματική όψη της έλικας TM3 παρατηρείται η ύπαρξη ενός μοτίβου D(E)RY, τα φορτισμένα κατάλοιπα του οποίου συμμετέχουν στο σχηματισμό γεφυρών άλατος με αντίστοιχα φορτισμένα κατάλοιπα της έλικας TM6. Ο σχηματισμός αυτών των ηλεκτροστατικών αλληλεπιδράσεων σταθεροποιεί τη θέση των διαμεμβρανικών τμημάτων. Κατά την ενεργοποίηση του υποδοχέα, ωστόσο, αυτή η ιοντική γέφυρα σπάει, οδηγώντας στη μετατόπιση της έλικας TM6 κατά 6-15 Å, με μικρότερες

Κεφάλαιο I - Εισαγωγή

μετακινήσεις να παρατηρούνται επίσης για την 5η διαμεμβρανική έλικα (TM5). Οι μετατοπίσεις οδηγούν στο σχηματισμό μιας κοιλότητας στην κυτοπλασματική όψη του υποδοχέα, στην οποία εισέρχονται οι G-πρωτεΐνες. Ένα δεύτερο μοτίβο που φαίνεται να σχετίζεται άμεσα με τη διαδικασία της ενεργοποίησης είναι το NPxxYx(5,6)F, το οποίο εντοπίζεται στο κυτοπλασματικό άκρο της 7ης διαμεμβρανικής έλικας TM7. Σε αντίθεση με το D(E)RY, το συγκεκριμένο μοτίβο φαίνεται να συμβάλλει στη σταθεροποίηση των ενεργοποιημένων υποδοχέων, χρησιμοποιώντας τα κατάλοιπα τυροσίνης και φαινυλαλανίνης για να σταθεροποιήσει τη θέση της έλικας TM6 μέσω αλληλεπιδράσεων π-ηλεκτρονίων με αντίστοιχης φύσης κατάλοιπα στην TM6 (Park et al., 2008; Rasmussen et al., 2011).

Τα μοτίβα D(E)RY και NPxxYx(5,6)F παρατηρήθηκαν αρχικά στη δομή της Ροδοψίνης και η συντηρημένη παρουσία τους επιβεβαιώθηκε σε όλες τις υπόλοιπες δομές υποδοχέων της Κλάσης A, γεγονός που, σε συνδυασμό με τη σημαντική ομοιότητα των δομών των ενεργοποιημένων GPCRs, αποδεικνύει πως η διαδικασία ενεργοποίησης είναι κοινή για τα μέλη της Κλάσης A. Επιπρόσθετα, παρ'ότι μέχρι στιγμής δεν υπάρχει διαθέσιμη δομή για ενεργοποιημένους GPCRs άλλων οικογενειών, οι διαθέσιμες δομές των υποδοχέων των Κλάσεων B, C και F παρουσιάζουν είτε ιοντικές γέφυρες παρόμοιου χαρακτήρα με αυτές του μοτίβου D(E)RY ανάμεσα στην 3η και την 6η έλικα είτε μοτίβα παρόμοια με το NPxxYx(5,6)F στην έλικα TM7, γεγονός που υποδεικνύει ότι και οι υπόλοιπες κλάσεις των GPCRs πιθανώς ακολουθούν παρόμοιο πρότυπο ενεργοποίησης (Dore et al., 2014; Siu et al., 2013; Wang et al., 2013a).

Κεφάλαιο I - Εισαγωγή



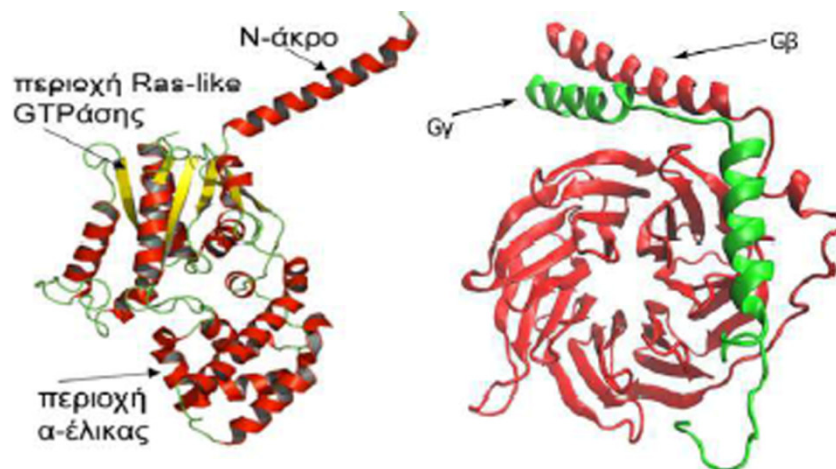
Εικόνα 16 Στερεοδιατάξεις αλλαγές κατά την ενεργοποίηση των GPCRs. (Α) Σύγκριση της Ροδοψίνης στην ανενεργή κατάσταση με την ενεργοποιημένη στερεοδιάταξή της, τη Μεταροδοψίνη II. Παρατηρείται σημαντική μετακίνηση του κυτοπλασματικού άκρου της διαμεμβρανικής έλικας TM6. (Β) Η γέφυρα άλατος ανάμεσα στην έλικα TM6 και το μοτίβο D(E)RY που βρίσκεται στην έλικα TM3 στην ανενεργή Ροδοψίνη. (Γ) Κατά την ενεργοποίηση των υποδοχέων η γέφυρα άλατος σπάει, και η έλικα TM6 απομακρύνεται από την έλικα TM3.

Κεφάλαιο I - Εισαγωγή

1.5.1.3 Οι ετεροτριμερείς G-πρωτεΐνες

Ο συνδετικός κρίκος ανάμεσα στους υποδοχείς GPCRs της μεμβράνης και στα μόρια εκτελεστές είναι οι ετεροτριμερείς G-πρωτεΐνες (Εικόνα 17). Λειτουργούν σαν μοριακοί «διακόπτες» που ενεργοποιούν ενδοκυτταρικούς σηματοδοτικούς «καταρράκτες» ως απόκριση στα εξωκυττάρια σήματα που ενεργοποιούν τους GPCRs και κατά συνέπεια, διαδραματίζουν σημαντικό ρόλο στον καθορισμό της ειδικότητας και των προσωρινών χαρακτηριστικών της κυτταρικής απόκρισης (Woehler & Ronimaskin, 2009). Οι G-πρωτεΐνες είναι μια ομάδα GTPασών με κοινή δομή. Είναι ετεροτριμερείς, αποτελούμενες από τις G α , G β και G γ υπομονάδες με μοριακά βάρη 39-45, 35-39 και 6-8 kDa αντίστοιχα. Η G α υπομονάδα αποτελείται από τρεις αυτοτελείς δομικές περιοχές (domains): μια αμινοτελική περιοχή με δομή α -έλικας, μια περιοχή με ενεργότητα GTPάσης, και μια περιοχή α -έλικας που αλληλεπιδρά με την περιοχή της GTPάσης. Οι δυο τελευταίες περιοχές δομούν μια βαθιά σχισμή, στην οποία συνδέεται ένα νουκλεοτίδιο GTP ή GDP. Η G β υπομονάδα ακολουθεί ένα πρότυπο διάταξης γνωστό σαν «έλικα επτά λεπίδων» (seven bladed propeller), το οποίο αποτελείται από επτά επαναλαμβανόμενες αλληλουχίες WD40. Η επανάληψη WD40 είναι ένα μικρό δομικό μοτίβο περίπου 40 καταλοίπων, που συνήθως τερματίζεται με ένα διπεπτίδιο τρυπτοφάνης-ασπαρτικού. Η G γ , τέλος, είναι μια μικρή υπομονάδα, αποτελούμενη από μια αμινοτελική και μια καρβοξυτελική α -έλικα, που ενώνονται μέσω ενός εκτεταμένου βρόχου (Wall et al., 1995). Σε αντίθεση με τη μεγάλη ποικιλομορφία των GPCRs, τα γονίδια που κωδικοποιούν τις G-πρωτεΐνες είναι σχετικά λίγα. Στα θηλαστικά υπάρχουν 21 G α υπομονάδες που κωδικοποιούνται από 16 γονίδια, 6 G β υπομονάδες που κωδικοποιούνται από 5 γονίδια και 12 G γ υπομονάδες. Οι G-πρωτεΐνες ταξινομούνται κυρίως με βάση τις G α υπομονάδες, οι οποίες κατατάσσονται σε τέσσερις μεγάλες οικογένειες, τις G α_s , G $\alpha_{i/o}$, G $\alpha_{q/11}$ και G $\alpha_{12/13}$ (Cabrera-Vera et al., 2003).

Κεφάλαιο I - Εισαγωγή



Εικόνα 17 Η δομή των ετεροτριμερών G-πρωτεϊνών. Η Gα υπομονάδα αποτελείται από μια αμινοτελική α-έλικα, μια α-ελικοειδή περιοχή και μια περιοχή με ενεργότητα GTPάσης. Η Gβ υιοθετεί το δίπλωμα της έλικας 7 λεπίδων και αποτελείται από 7 επαναλήψεις της περιοχής WD40, ενώ η Gγ αποτελείται από 2 α-έλικες που ενώνονται με έναν βρόχο. Η Gβ και η Gγ εμφανίζονται σχεδόν πάντα σαν ετεροδιμερές.

Όλες οι ετεροτριμερείς G-πρωτεΐνες ακολουθούν τον ίδιο κύκλο ενεργοποίησης–απενεργοποίησης. Όταν το GDP συνδέεται με την Gα υπομονάδα επιτρέπει τη σύνδεση της με το ετεροδιμερές Gβγ και την απενεργοποίηση του ετεροτριμερούς. Η πρόσδεση του ετεροδιμερούς Gβγ στην Gα υπομονάδα ενισχύει αυξάνει τη συγγένειά της για τον ενεργοποιημένο υποδοχέα. Με τη σύνδεση του αγωνιστή, ο υποδοχέας ενεργοποιείται, αυξάνοντας τη συγγένεια του για το ετεροτριμερές Gα-Gβγ. Η πρόσδεση του ετεροτριμερούς στον υποδοχέα οδηγεί στην απομάκρυνση του GDP από την Gα υπομονάδα και την αντικατάστασή του από GTP. Η αντικατάσταση αυτή ελαττώνει τη συγγένεια της Gα υπομονάδας για το Gβγ και την αποδιάταξη του ετεροτριμερούς. Η Gα υπομονάδα και το Gβγ στη συνέχεια ενεργοποιούν αντίστοιχα μόρια εκτελεστές. Η δράση της Gα υπομονάδας ως GTPάσης έχει ως αποτέλεσμα τη διάσπαση του GTP σε GDP και την επανασύνδεση και απενεργοποίηση του ετεροτριμερούς (Oldham & Hamm, 2008).

1.5.1.4 Αλληλεπιδράσεις Ολιγομερισμού στους GPCRs

Το κλασικό μοντέλο του σηματοδοτικού μηχανισμού των GPCRs παρουσιάζει τους υποδοχείς να λειτουργούν κυρίως ως μονομερή. Σύμφωνα με αυτήν την οπτική, ένας προσδέτης επάγει στην ενεργοποίηση του υποδοχέα, οδηγώντας στην ενεργοποίηση επιλεγμένων G-πρωτεϊνών και τη ρύθμιση συγκεκριμένων κυτταρικών αποκρίσεων. Ωστόσο, συνεχώς αυξανόμενος όγκος πειραματικών στοιχείων προτείνουν ότι οι GPCRs μπορούν επίσης να οργανωθούν σε διμερή ή ανώτερης τάξης ολιγομερή και να δράσουν συνεργατικά (Milligan, 2013). Αυτές οι τεταρτοταγείς δομές, σχηματιζόμενες είτε από πολλαπλά αντίγραφα του ίδιου υποδοχέα (ομομερή) είτε από διαφορετικούς GPCRs (ετερομερή) συχνά έχουν σαν αποτέλεσμα την περίπλοκη και ετερογενή αλλοστερική ρύθμιση σηματοδοτικών μηχανισμών. Επιπρόσθετα, μια σειρά από διμερή GPCRs συνδέονται με συνεχώς αυξανόμενο αριθμό σοβαρών παθολογικών καταστάσεων (Πίνακας 3), ανάμεσα στις οποίες συμπεριλαμβάνονται η ασθένεια του Parkinson, η σχιζοφρένεια, η ανοσολογική απόκριση σε διάφορα φάρμακα και ο εθισμός στα ναρκωτικά και τη νικοτίνη (Kamal & Jockers, 2011; Niswender & Conn, 2010; Pasternak & Pan, 2011). Κατά συνέπεια, το ενδιαφέρον για την περαιτέρω μελέτη των χαρακτηριστικών και της σημασίας του ολιγομερισμού των GPCRs, καθώς και τον πιθανό σχεδιασμό φαρμάκων με στόχους τα ολιγομερή υποδοχέων αυξάνεται συνεχώς (Dalrymple et al., 2008; Liu et al., 2009).

Ανάμεσα στα μέλη της μεγάλης υπερικογενείας των GPCRs, οι υποδοχείς της Κλάσης C είναι αυτοί που έχουν μελετηθεί περισσότερο ως προς τα χαρακτηριστικά του ολιγομερισμού. Αυτό συμβαίνει γιατί, όπως έχει αποδειχθεί, ο σχηματισμός διμερών και ολιγομερών είναι απαραίτητος για τη φυσιολογική λειτουργία αυτών των υποδοχέων (Kniazeff et al., 2011). Οι υποδοχείς GABA_{B1} και GABA_{B2} αλληλεπιδρούν για να σχηματίσουν έναν ετεροδιμερικό GABA_B υποδοχέα, υπεύθυνο για την αναγνώριση του γ-αμινοβουτυρικού οξέος. Αντίστοιχα, η αναγνώριση των διαφορετικών γεύσεων από τους υποδοχείς γεύσης επιτυγχάνεται μέσω του λειτουργικού ετεροδιμερισμού του υποδοχέα T1R3 με τους υποδοχείς T1R1 και T1R2. Από την άλλη, οι μεταβοτροπικοί υποδοχείς του γλουταμικού (mGluR), οι υποδοχείς ασβεστίου (CaS) και οι υποδοχείς φορτισμένων αμινοξέων (GPCRC6A) σχηματίζουν ομοδιμερή, στα οποία τα δυο πρωτομερή αλληλεπιδρούν μεταξύ τους όχι μόνο με ασθενείς αλληλεπιδράσεις αλλά και

Κεφάλαιο I - Εισαγωγή

ομοιοπολικά, συνδεόμενα με δισουλφιδικούς δεσμούς ανάμεσα στις εξωκυτταρικές περιοχές τους (Niswender & Conn, 2010). Τέλος, διμερή υποδοχέων του γλουταμικού έχουν βρεθεί να αλληλεπιδρούν μεταξύ τους σχηματίζοντας ομο- και ετεροτετραμερή. Τα παραπάνω έχουν οδηγήσει στο συμπέρασμα ότι ο διμερισμός και, πιθανώς ο ολιγομερισμός, είναι απαραίτητο στοιχείο στη φυσιολογική λειτουργία των υποδοχέων της Κλάσης C (Kniazeff et al., 2011).

Πίνακας 3 Παραδείγματα ολιγομερών GPCRs τα οποία έχουν συσχετιστεί με ασθένειες.

Ολιγομερές	Παθολογική σημασία
A1 – A2A	Ανοχή σε φάρμακα
A1 – D1	Ασθένεια Parkinson's, εθισμός στην κοκαΐνη
A1 – mGluR1	Σχιζοφρένεια
A1 – 5-HT2A	Σχιζοφρένεια
A2α – D2	Στόχος φαρμάκων για την ασθένεια Parkinson's, στόχος αντιψυχωτικών φαρμάκων, ιδιότητες αντίδρασης στην κοκαΐνη
A2αR – mGluR5	Ασθένεια Parkinson's, Σχιζοφρένεια, αποτοξίνωση από κοκαΐνη
A2αR – D2 – mGluR5	Ασθένεια Parkinson's, Σχιζοφρένεια, αποτοξίνωση από κοκαΐνη
A2αR – CB1	Σχιζοφρένεια
A2αR – CB1 – D2	Σχιζοφρένεια
A2αR – μOR	Εθισμός σε οπιοειδή
A3αR – A3αR	Ασθένεια Parkinson's, Σχιζοφρένεια
CB1 – D2	Ασθένεια Parkinson's, Σχιζοφρένεια
CRH1 – V1B	Κλινική κατάθλιψη
α1A – α1B	Έμφραγμα μυοκαρδίου
D1 – D2	Ασθένεια Parkinson's, Σχιζοφρένεια
D1 – D3	Εθισμός στην κοκαΐνη
D2 – D3	Ασθένεια Parkinson's, Σχιζοφρένεια
D4 – μOR	Εθισμός στα ναρκωτικά
mGluR2 – 5-HT2A	Στόχος αντιψυχωτικών φαρμάκων

Αντίθετα από τους υποδοχείς της Κλάσης C, η ιδέα του ολιγομερισμού παραμένει αμφιλεγόμενη για τις υπόλοιπες οικογένειες των GPCRs (Milligan, 2013). Την πιο περίπλοκη περίπτωση αποτελεί η Κλάση A, η οποία περιλαμβάνει την πλειονότητα των GPCRs. Συνολικά, πάνω από 100 υποδοχείς της Κλάσης A έχουν αποδειχθεί πειραματικά ότι συμμετέχουν στο σχηματισμό διμερών ή ολιγομερών, τόσο σαν ομομερή όσο και σαν ετερομερή (Borrotto-Escuela et al., 2014). Ωστόσο, οι ίδιοι υποδοχείς που συμμετέχουν σε ολιγομερή έχουν βρεθεί να είναι πλήρως λειτουργικοί και ως μονομερή (Bayburt et al., 2011). Επιπρόσθετα, μια σειρά

Κεφάλαιο I - Εισαγωγή

από μελέτες έχουν αμφισβητήσει τη σταθερότητα των αλληλεπιδράσεων ολιγομερισμού των υποδοχέων της Κλάσης A, προτείνοντας ότι αυτά τα διμερή δεν είναι μόνιμα σύμπλοκα αλλά παροδικά (Johnston et al., 2012; Lambert, 2010). Αντίθετα, τα διαθέσιμα δεδομένα προτείνουν ότι ο ολιγομερισμός των GPCRs της Κλάσης A λειτουργεί συμπληρωματικά ως προς τη ρύθμιση της λειτουργικότητάς τους, σαν μηχανισμός μέσω του οποίου οι υποδοχείς οδηγούνται σε μεγαλύτερη ποικιλομορφία στη ρύθμιση των σηματοδοτικών μονοπατιών. Για παράδειγμα, ο σχηματισμός διμερών υποδοχέων μπορεί να έχει σαν αποτέλεσμα τη μεταβολή στη συγγένεια των πρωτομερών για κάποιον προσδέτη. Σε άλλες περιπτώσεις, ο ολιγομερισμός δυο υποδοχέων μπορεί να οδηγήσει στη ρύθμιση διαφορετικών μονοπατιών μέσω των αλληλεπιδράσεων με διαφορετικές G-πρωτεΐνες (Franco et al., 2007). Τέλος, ο σχηματισμός διμερών GPCRs κατά το στάδιο της πρωτεϊνοσύνθεσής τους στο Ενδοπλασματικό Δίκτυο έχει βρεθεί να επάγει τη μεταφορά τους σε υποκυτταρικές θέσεις του κυττάρου διαφορετικές από την κυτταρική μεμβράνη, όπως οι μεμβράνες των οργανιδίων ή ο πυρηνικός φάκελος (Franco et al., 2007; Milligan, 2013).

Διάφορα χαρακτηριστικά της δομής και της λειτουργίας των GPCRs, συμπεριλαμβανομένης της ικανότητάς τους να σχηματίζουν διμερή ή/και ολιγομερή, έχουν βρεθεί να επηρεάζονται από τις αλληλεπιδράσεις τους με τα λιπίδια της μεμβράνης. Συγκεκριμένα, τόσο βιοχημικά πειράματα όσο και βιοπληροφορικές μελέτες έχουν δείξει ότι η ικανότητα σχηματισμού διμερών για υποδοχείς όπως η Ροδοψίνη και ο αδρενεργικός υποδοχέας β2 εξαρτάται επηρεάζεται από τις αλληλεπιδράσεις τους με τις πολικές κεφαλές λιπιδίων όπως οι φωσφατιδυλ-αιθανολαμίνες (PEs) ή οι φωσφατιδυλινωσιτόλες (PIs), καθώς και από τις παραμορφώσεις της μεμβράνης γύρω από τα άκρα των διαμεμβρανικών τμημάτων των υποδοχέων (Mondal et al., 2011). Ένα ακόμη χαρακτηριστικό της μεμβράνης, το οποίο έχει προταθεί σαν πιθανός ρυθμιστής των αλληλεπιδράσεων διμερισμού των GPCRs είναι η παρουσία χοληστερόλης στη λιπιδική διπλοστιβάδα. Προηγούμενες κρυσταλλογραφικές και υπολογιστικές μελέτες πάνω στον αδρενεργικό υποδοχέα β2 (Cherezon et al., 2007) και, πιο πρόσφατα, τον μεταβοτροπικό υποδοχέα του γλουταμικού mGluR1 (Dore et al., 2014) πρότειναν πως μόρια χοληστερόλης μπορούν να δράσουν συνεργατικά κατά το σχηματισμό διμερών, παρεμβαλλόμενα ανάμεσα στα δύο πρωτομερή και ενισχύοντας τις αλληλεπιδράσεις

Κεφάλαιο I - Εισαγωγή

μεταξύ τους. Παράλληλα, μια σειρά από βιοχημικές μελέτες έδειξαν ότι η παρουσία χοληστερόλης μπορεί να επάγει το σχηματισμό διμερών στους οπιοειδείς υποδοχείς (Zheng et al., 2012).

Παρά τη μεγάλη βιολογική σημασία του ολιγομερισμού των GPCRs, τα δεδομένα για την δομική φύση των αλληλεπιδράσεών τους είναι αμφιλεγόμενα. Μια σειρά βιοφυσικών και βιοχημικών μελετών, συμπεριλαμβανομένων πειραμάτων μεταλλαξιγένεσης, βιοχημικών τροποποιήσεων cross-linking και παρατηρήσεων με Μικροσκοπία Ατομικής Δύναμης (Atomic Force Microscopy, AFM), έχουν προτείνει την πιθανή συμμετοχή των διαμεμβρανικών ελίκων TM1, TM4 και TM5 στις αλληλεπιδράσεις ολιγομερισμού. Πιο πρόσφατα, προσδιορίστηκαν μια σειρά από κρυσταλλικές δομές GPCRs σε διάταξη διμερών ή ολιγομερών, οι οποίες παρουσιάζουν διαμοριακές αλληλεπιδράσεις που, εν μέρει, συμφωνούν με τα αποτελέσματα των βιοχημικών πειραμάτων. Ωστόσο, δεν είναι ξεκάθαρο το κατά πόσο αυτές οι επαφές αντιστοιχούν όντως στις αλληλεπιδράσεις ολιγομερισμού και κατά πόσο είναι αποτέλεσμα της διαδικασίας κρυστάλλωσης, ενώ άγνωστη επίσης παραμένει η πιθανή επιρροή των μεθόδων που χρησιμοποιούνται στην επίλυση των δομών των GPCRs, όπως η χρήση μεταλλαγών και χημικών μορίων και η κρυστάλλωση μέσω λιπιδικής μεσόφασης. Επιπρόσθετα, πολύ λίγα είναι γνωστά πάνω στον δομικό μηχανισμό με τον οποίο οι αλληλεπιδράσεις διμερισμού και ολιγομερισμού των GPCRs επηρεάζουν τη λειτουργία τους. Έτσι, η βιολογική σημασία των διαθέσιμων κρυσταλλογραφικών δεδομένων πάνω στον ολιγομερισμό των GPCRs παραμένει αντικείμενο έντονων διαφωνιών.

1.5.2 Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs)

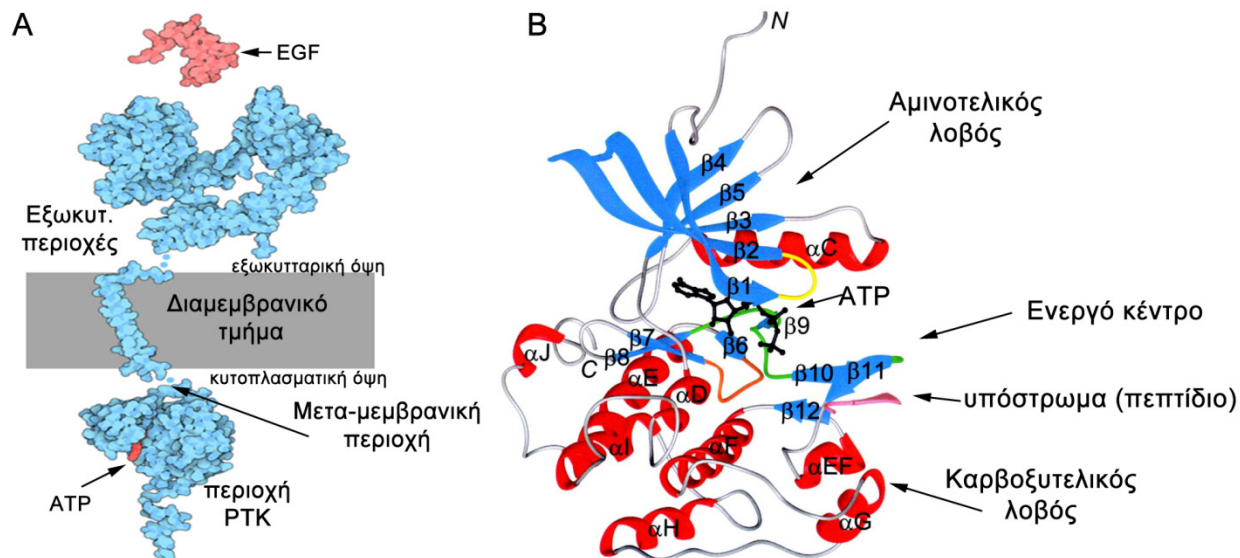
Μια κατηγορία υποδοχέων με ιδιαίτερα σημαντική λειτουργία στους ευκαρυωτικούς οργανισμούς είναι οι υποδοχείς με ενεργότητα τυροσινικής κινάσης (Receptor Tyrosine Kinases, RTKs). Εντοπίζονται κυρίως στην κυτταρική μεμβράνη των κυττάρων, με μερικά μέλη τους ωστόσο να έχουν βρεθεί και σε άλλα μέρη του κυττάρου, όπως οι μεμβράνες των ενδοσωμάτων και των μιτοχονδρίων και ο πυρηνικός φάκελος (Carpenter & Liao, 2013). Βασικό χαρακτηριστικό των RTKs είναι η ενζυμική ενεργότητα τυροσινικής κινάσης, μέσω της

Κεφάλαιο I - Εισαγωγή

οποίας ενεργοποιούνται και επιτελούν τους σηματοδοτικούς μηχανισμούς τους (Biarc et al., 2011). Οι RTKs αναγνωρίζουν μεγάλο εύρος πεπτιδικών προσδετών, ανάμεσα στις οποίες περιλαμβάνονται αυξητικοί παράγοντες όπως ο επιδερμικός αυξητικός παράγοντας EGF, ορμόνες όπως η ινσουλίνη, ανοσολογικοί παράγοντες όπως οι πρωτεΐνες του συμπληρώματος κλπ (Lemmon & Schlessinger, 2010). Μέσω των σηματοδοτικών μονοπατιών τους οι RTKs ρυθμίζουν σημαντικό αριθμό λειτουργιών, ανάμεσα στις οποίες συμπεριλαμβάνονται η έκφραση συγκεκριμένων γονιδίων, η κυτταρική αύξηση και ο κυτταρικός θάνατος, ο μεταβολισμός της γλυκόζης, η ανοσολογική απόκριση των λεμφοκυττάρων και η μετάδοση του νευρικού παλμού (Dengjel et al., 2009). Έτσι, οι RTKs έχουν εμπλακεί σε πολλές ασθένειες, οι οποίες περιλαμβάνουν πολλούς τύπους καρκίνου, μεταβολικά σύνδρομα όπως ο διαβήτης αλλά και διάφορες νευροεκφυλιστικές ασθένειες (Sie et al., 2015).

Παρά τη σημασία τους, οι RTKs είναι πολύ λιγότερο μελετημένοι σε πειραματικό επίπεδο συγκριτικά με άλλες κατηγορίες υποδοχέων όπως οι GPCRs. Μέχρι στιγμής δεν υπάρχει διαθέσιμη πειραματικά προσδιορισμένη δομή για κάποιον RTK, παρά μόνο δομές θραυσμάτων των υποδοχέων. Επιπρόσθετα, τα σηματοδοτικά μονοπάτια και η φαρμακολογία των RTKs έχουν μελετηθεί για συγκεκριμένες κατηγορίες όπως ο υποδοχέας ινσουλίνης και οι υποδοχείς των αυξητικών παραγόντων, ενώ πολύ λίγα είναι γνωστά για άλλους RTKs (Alexander et al., 2017). Τέλος, η πλειονότητα της πειραματικής έρευνας πάνω στους RTKs επικεντρώνεται στους υποδοχείς του ανθρώπου, του ποντικού και του αρουραίου (Lemmon & Schlessinger, 2010). Αντίθετα, πολύ λίγα είναι γνωστά για την ύπαρξη και τη λειτουργία των RTKs σε άλλα θηλαστικά, καθώς και σε άλλες κατηγορίες οργανισμών όπως τα φυτά, τα έντομα και τα πρωτόζωα.

Κεφάλαιο I - Εισαγωγή



Εικόνα 18 Η τρισδιάστατη δομή των RTKs. (Α) Θεωρητικό μοντέλο της πλήρους δομής του υποδοχέα EGFR. Το μοντέλο έχει κατασκευαστεί από το συνδυασμό θραυσμάτων του υποδοχέα με διαθέσιμη δομή. Προσαρμογή από Goodsell, 2010. (Β) Τρισδιάστατη δομή της περιοχής τυροσινικής κινάσης των RTKs (PTK). Η περιοχή PTK αποτελείται από έναν αμινοτελικό λοβό με δομή β-πτυχωτής επιφάνειας και έναν καρβοξυτελικό λοβό με κυρίως α-ελικοειδή δομή. Στην κοιλότητα ανάμεσα στους δύο λοβούς εντοπίζεται το ενεργό κέντρο το οποίο περιλαμβάνει μία θέση για την πρόσδεση του ATP και μία θέση για την αναγνώριση του πρωτεϊνικού υποστρώματος (μωβ). Οι βρόχοι που καταλύουν την ενζυμική ενεργότητα της περιοχής PTK χρωματίζονται με κίτρινο για την θέση πρόσδεσης του ATP, πράσινο για τον βρόχο ενεργοποίησης και κόκκινο για τον καταλυτικό βρόχο φωσφορύλωσης. Προσαρμογή από Hubbard, 1999.

1.5.2.1 Δομή και ταξινόμηση των RTKs

Οι RTKs είναι διαμεμβρανικές πρωτεΐνες με ένα διαμεμβρανικό τμήμα (Εικόνα 18) και κατατάσσονται στις διαμεμβρανικές πρωτεΐνες Τύπου I. Διαθέτουν ένα σηματοδοτικό πεπτίδιο στην αρχή της αλληλουχίας τους, το οποίο αποκόπτεται κατά την είσοδό τους στο Ενδοπλασματικό Δίκτυο. Το αμινοτελικό άκρο τους βρίσκεται εξωκυτταρικά και, με ελάχιστες εξαιρέσεις, περιλαμβάνει μία ή περισσότερες αυτοτελείς δομικές περιοχές (π.χ. περιοχή L, περιοχή Ig-like κλπ), οι οποίες είναι υπεύθυνες για την αναγνώριση των προσδετών (Lemmon & Schlessinger, 2010). Το διαμεμβρανικό τμήμα τους έχει μήκος 15-25 καταλοίπων και αποτελείται από κυρίως υδρόφοβα αμινοξέα, με τη συχνή παρουσία, ωστόσο, φορτισμένων καταλοίπων λυσίνης ή αργινίνης στα άκρα του, τα οποία αλληλεπιδρούν με τα φωσφορικά

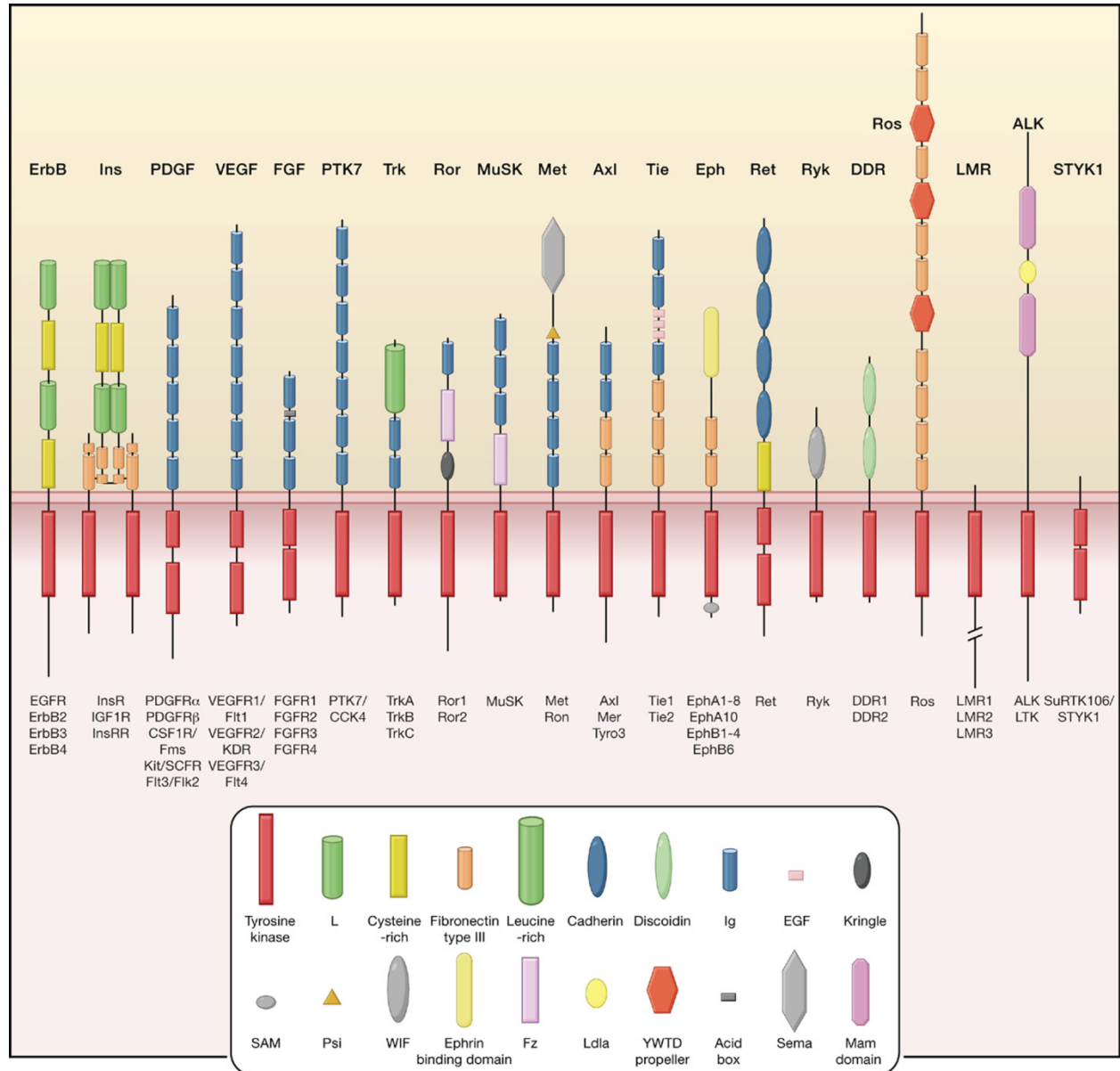
Κεφάλαιο I - Εισαγωγή

ιόντα των φωσφολιπιδίων. Το καρβοξυτελικό άκρο βρίσκεται κυτοπλασματικά και περιλαμβάνει μια αυτοτελή δομική περιοχή τυροσινικής κινάσης (Protein Tyrosine Kinase, PTK), η οποία επιτελεί την ενζυμική λειτουργία του υποδοχέα, καθώς και μια εκτεταμένη «ουρά» καταλοίπων (Hubbard, 1999). Τόσο η ίδια η περιοχή PTK όσο και η καρβοξυτελική ουρά περιέχουν θέσεις αυτο- και ετεροφωσφορυλίωσης, μέσω των οποίων ενεργοποιείται και ρυθμίζεται η δραστηριότητα των υποδοχέων (Lemmon & Schlessinger, 2010). Τέλος, η αλληλουχία ανάμεσα στο κυτοπλασματικό άκρο του διαμεμβρανικού τμήματος και την περιοχή τυροσινικής κινάσης χαρακτηρίζεται ως «μεταμεμβρανική περιοχή» (juxtamembrane region). Η μεταμεμβρανική περιοχή αποτελείται από 40-80 αμινοξικά κατάλοιπα και δρα συνεργατικά στη σηματοδότηση των RTKs, επηρεάζοντας τη στερεοδιάταξη του υποδοχέα κατά την ενεργοποίησή του μαζί με το διαμεμβρανικό τμήμα και ρυθμίζοντας την ενζυμική ενεργότητά του, μέσα από αλληλεπιδράσεις με την περιοχή PTK. Πρέπει να σημειωθεί πως, μέχρι στιγμής, δεν προσδιοριστεί πειραματικά η πλήρης δομή ενός RTK. Αντίθετα, υπάρχουν διαθέσιμες δομές στην PDB για θραύσματα των υποδοχέων όπως διάφορες εξωκυτταρικές περιοχές (Ferguson et al., 2003; Ogiso et al., 2002), η περιοχή PTK (Zhang et al., 2006) και το διαμεμβρανικό τμήμα (Bocharon et al., 2008), οι οποίες έχουν χρησιμοποιηθεί στην κατασκευή θεωρητικών μοντέλων για πλήρεις υποδοχείς (Goodsell, 2010).

Η ταξινόμηση των RTKs γίνεται, συνήθως, με βάση το είδος και τον αριθμό των εξωκυτταρικών περιοχών τους (Εικόνα 19), σε συνδυασμό με το είδος των προσδετών που αναγνωρίζουν και δεσμεύουν (Lemmon & Schlessinger, 2010). Το πιο κοινό σύστημα ταξινόμησης, το οποίο χρησιμοποιείται περισσότερο στη βιβλιογραφία είναι το σύστημα της Διεθνούς Ένωσης Βασικής και Κλινικής Φαρμακολογίας (IUPHAR), το οποίο κατατάσσει τους υποδοχείς σε 20 οικογένειες που παρουσιάζονται στον Πίνακα 4 (Alexander et al., 2017). Πρέπει να σημειωθεί ότι το παραπάνω σύστημα ταξινόμησης έχει σχεδιαστεί με βάση τους πειραματικά μελετημένους RTKs του ανθρώπου και του ποντικού, και θεωρείται ότι ισχύει γενικά για τους RTKs των σπονδυλοζώων. Ωστόσο, καθώς το σύστημα βασίζεται σε μεγάλο βαθμό στο είδος των εξωκυτταρικών περιοχών, είναι πολύ πιθανό να μην μπορεί να εφαρμοστεί στην ταξινόμηση RTKs άλλων οργανισμών (π.χ. έντομα, φυτά) οι οποίοι, λόγω διαφορών στη δομή του κυττάρου (π.χ. κυτταρικό τοίχωμα στα φυτά) και στα είδη των ιστών

Κεφάλαιο I - Εισαγωγή

τους αναμένεται να διαθέτουν διαφορετικές εξωκυτταρικές περιοχές και να αναγνωρίζουν διαφορετικά είδη προσδετών.



Εικόνα 19 Το σύστημα ταξινόμησης των RTKs. Διαγραμματική απεικόνιση των 20 υποοικογενειών των RTKs και της τοπολογίας τους. Για κάθε οικογένεια δίνονται διαγραμματικά το είδος και ο αριθμός των εξωκυτταρικών αυτοτελών δομικών περιοχών, η θέση της περιοχής PTK και τα μέλη της οικογένειας στο ανθρώπινο πρωτόμα. Προσαρμογή από Lemmon & Schlessinger, 2010.

Κεφάλαιο I - Εισαγωγή

Πίνακας 4 Ταξινόμηση των RTKs με βάση το σύστημα της IUPHAR.

Όνομα	Περιγραφή	Όνομα	Περιγραφή
Τύπος I EGFR/ErbB	Υποδοχείς ομοιάζοντες με τον υποδοχέα του επιδερμικού αυξητικού παράγοντα EGF	Τύπος XI AXL/TAM	Υποδοχείς που σχετίζονται με την πρόσδεση των πρωτεϊνών GAS6 και της πρωτεΐνης S
Τύπος II INS	Υποδοχείς ομοιάζοντες με τον υποδοχέα της ινσουλίνης	Τύπος XII Tie	Υποδοχείς της αγγειοποιητίνης
Τύπος III PDGFR	Υποδοχείς ομοιάζοντες με τον υποδοχέα του αυξητικού παράγοντα των αιμοπεταλίων PDGF	Τύπος XIII Ephrin	Υποδοχείς των εφρινών
Τύπος IV VEGFR	Υποδοχείς ομοιάζοντες με τον υποδοχέα του επιθηλιακού αυξητικού παράγοντα VEGF	Τύπος XIV RET	Υποοικογένεια των πρωτο-ογκογονιδίων ret
Τύπος V FGFR	Υποδοχείς ομοιάζοντες με τον υποδοχέα του αυξητικού παράγοντα ινοβλαστών FGF	Τύπος XV RYK	Υποοικογένεια υποδοχέων που σχετίζονται με την ενεργοποίηση περιοχών πρόσδεσης ATP
Τύπος VI PTK7/CCK4	Υποδοχείς που σχετίζονται με την πόλωση των επιθηλιακών κυττάρων και την ανάπτυξη του νευρικού ιστού	Τύπος XVI DDR	Υποδοχείς κολλαγόνου και δισκοειδίνης
Τύπος VII NGFR/TRK	Υποδοχείς της νευροτροφίνης και των τροπομοσινών	Τύπος XVII ROS	Υποοικογένεια των πρωτο-ογκογονιδίων c-ros
Τύπος VIII ROR	Μη χαρακτηρισμένη (Orphan) υποοικογένεια υποδοχέων	Τύπος XVIII LMR	Υποοικογένεια χωρίς γνωστή λειτουργία, η οποία δεν διαθέτει εξωκυτταρικές περιοχές
Τύπος IX MuSK	Κινάσες του μυϊκού ιστού	Τύπος XIX ALK/LTK	Υποδοχείς των λευκοκυττάρων
Τύπος X HGFR/MET	Υποδοχείς ομοιάζοντες με τον υποδοχέα του αυξητικού παράγοντα των ηπατικών κυττάρων HGF	Τύπος XX STYK1	Υποοικογένεια χωρίς γνωστή λειτουργία, η οποία δεν διαθέτει εξωκυτταρικές περιοχές

1.5.2.2 Ο μηχανισμός λειτουργίας των RTKs

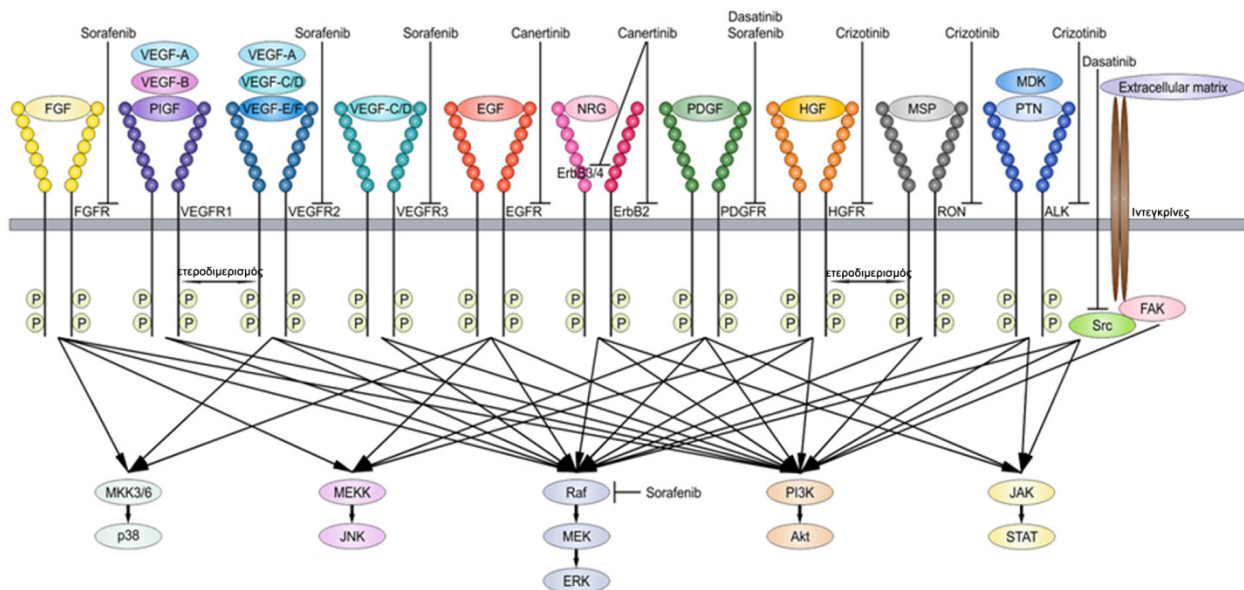
Το βασικό χαρακτηριστικό του μηχανισμού λειτουργίας των RTKs είναι ο διμερισμός των υποδοχέων (Lemmon & Schlessinger, 2010). Το πρώτο βήμα αυτού του μηχανισμού είναι η αναγνώριση και δέσμευση του κατάλληλου προσδέτη, η οποία οδηγεί στην αύξηση της συγγένειας ανάμεσα σε δύο υποδοχείς RTK και το σχηματισμό του διμερούς τους. Οι προσδέτες των RTKs, ανάλογα με τον τρόπο πρόσδεσής τους στους υποδοχείς, μπορούν να είναι μονομερείς (monovalent) ή διμερείς (bivalent). Οι μονομερείς προσδέτες (π.χ. EGF) δεσμεύονται κάθε φορά από έναν RTK, με αποτέλεσμα για τον σχηματισμού ενός διμερούς

Κεφάλαιο I - Εισαγωγή

RTKs να απαιτούνται δύο ξεχωριστοί προσδέτες, οι οποίοι αλληλεπιδρούν με δύο υποδοχείς. Από την άλλη, οι διμερείς προσδέτες (π.χ. Εφρίνες) διαθέτουν δύο θέσεις αναγνώρισης από RTKs και μπορούν να προσδένονται ταυτόχρονα σε δύο υποδοχείς. Σε κάθε περίπτωση, η δέσμευση του προσδέτη ωθεί τους υποδοχείς να έρθουν κοντά στο χώρο και να σχηματίσουν διαμοριακές αλληλεπιδράσεις, αρχικά μέσω των εξωκυτταρικών περιοχών τους και στη συνέχεια μέσω των διαμεμβρανικών τμημάτων τους (Lemmon & Schlessinger, 2010). Αυτά τα διμερή μπορούν να αποτελούνται από αντίγραφα του ίδιου υποδοχέα (ομοδιμερή) ή και από διαφορετικούς υποδοχείς (ετεροδιμερή), οδηγώντας αντίστοιχα στη ρύθμιση διαφορετικών κυτταρικών διεργασιών (Sie et al., 2015).

Η μετακίνηση των διαμεμβρανικών τμημάτων κατά το διμερισμό τους έχει σαν αποτέλεσμα την προσέγγιση των κυτοπλασματικών άκρων τους και τη μερική ενεργοποίηση της ενζυμικής ενεργότητας των περιοχών τυροσινικής κινάσης (PTK). Η μερικώς ενεργοποιημένη περιοχή PTK του κάθε υποδοχέα, στη συνέχεια, φωσφορυλιώνει κατάλοιπα τυροσίνης, είτε στον ίδιο υποδοχέα (αυτοφωσφορυλίωση) είτε στο δεύτερο υποδοχέα του συμπλόκου (ετεροφωσφορυλίωση). Η φωσφορυλίωση αυτών των καταλοίπων προκαλεί περαιτέρω στερεοδιαταξικές αλλαγές στην περιοχή PTK οδηγώντας σε πλήρη ενεργοποίησή του διμερούς, με αύξηση της ενεργότητας κινάσης κατά 50-200 φορές (Dengjel et al., 2009) και τη φωσφορυλίωση περισσότερων καταλοίπων τυροσίνης, κυρίως στη μεταμεμβρανική περιοχή. Η παραπάνω διαδικασία μπορεί, σε διάφορους RTKs, να ρυθμιστεί μέσα από μια σειρά από παράγοντες, όπως η μετακίνηση της μεταμεμβρανικής περιοχής, η οποία μπορεί να επάγει ή να αναστείλει τη λειτουργία της περιοχής PTK σε κάποιους υποδοχείς (π.χ. Υποδοχείς KIT των βλαστοκυττάρων) ή η καρβοξυτελική ουρά, η οποία αυτοφωσφορυλιώνεται και αυξάνει την ενεργότητα κινάσης στους υποδοχείς των αυξητικών παραγόντων των ινοβλαστών (Lemmon & Schlessinger, 2010).

Κεφάλαιο I - Εισαγωγή



Εικόνα 20 Σηματοδοτικά μονοπάτια των RTKs. Παρουσιάζονται τα σηματοδοτικά μονοπάτια μιας σειράς από RTKs στον άνθρωπο. Οι υποδοχείς αναγνωρίζουν μια σειρά από ορμόνες και σχηματίζουν ομοδιμερή για την ενεργοποίησή τους, ενώ σε κάποιες περιπτώσεις μπορούν να σχηματίσουν και ετεροδιμερή (π.χ. VEGFR1 και VEGFR2). Τα μονοπάτια πολλών RTKs αλληλοεπικαλύπτονται, ρυθμίζοντας μια σειρά από μεταβολικές οδούς μέσω ενδοκυτταρικών κινάσεων όπως οι κινάσες JAK, PI3K κλπ, οι οποίες ελέγχουν τη γονιδιακή έκφραση και την κυτταρική αύξηση. Επιπρόσθετα, τα μονοπάτια των RTKs αλληλοεπικαλύπτονται με τα σηματοδοτικά μονοπάτια άλλων υποδοχέων (π.χ. Ιντεγκρίνες), οδηγώντας σε αυξημένη πολυπλοκότητα. Η λειτουργία πολλών RTKs μπορεί να ανασταλεί μέσα από την αλληλεπίδρασή τους με διάφορα φάρμακα, παραδείγματα των οποίων παρουσιάζονται στην εικόνα. Προσαρμογή από Sie et al., 2015.

Οι πλήρως ενεργοποιημένοι RTKs, τέλος, αλληλεπιδρούν μέσω των φωσφορυλιωμένων καταλοίπων τους με μια σειρά από πρωτεΐνες που διαθέτουν ειδικές αυτοτελείς δομικές περιοχές αναγνώρισης, όπως οι ομόλογες με Src περιοχές 2 και 3 (SH2, SH3) και η περιοχή πρόσδεσης φωσφορυλιωμένων τυροσινών PTB. Η πλειονότητα αυτών των πρωτεϊνών διαθέτουν επίσης ενεργότητα τυροσινικής κινάσης και, με την προσδέσή τους στους RTKs, ενεργοποιούνται και φωσφορυλιώνουν με τη σειρά τους άλλες πρωτεϊνικές κινάσες (Εικόνα 20), οδηγώντας σε σηματοδοτικούς καταρράκτες φωσφορυλιώσεων, οι οποίοι καταλήγουν να ρυθμίζουν μια σειρά από κυτταρικές διεργασίες, με κυριότερη τη ρύθμιση της γονιδιακής έκφρασης (Dengjel et al., 2009).

Η παραπάνω διαδικασία, όπως περιγράφηκε, αποτελεί το γενικό μοντέλο του μηχανισμού λειτουργίας των RTKs. Ωστόσο, κάποιες ομάδες υποδοχέων παρουσιάζουν

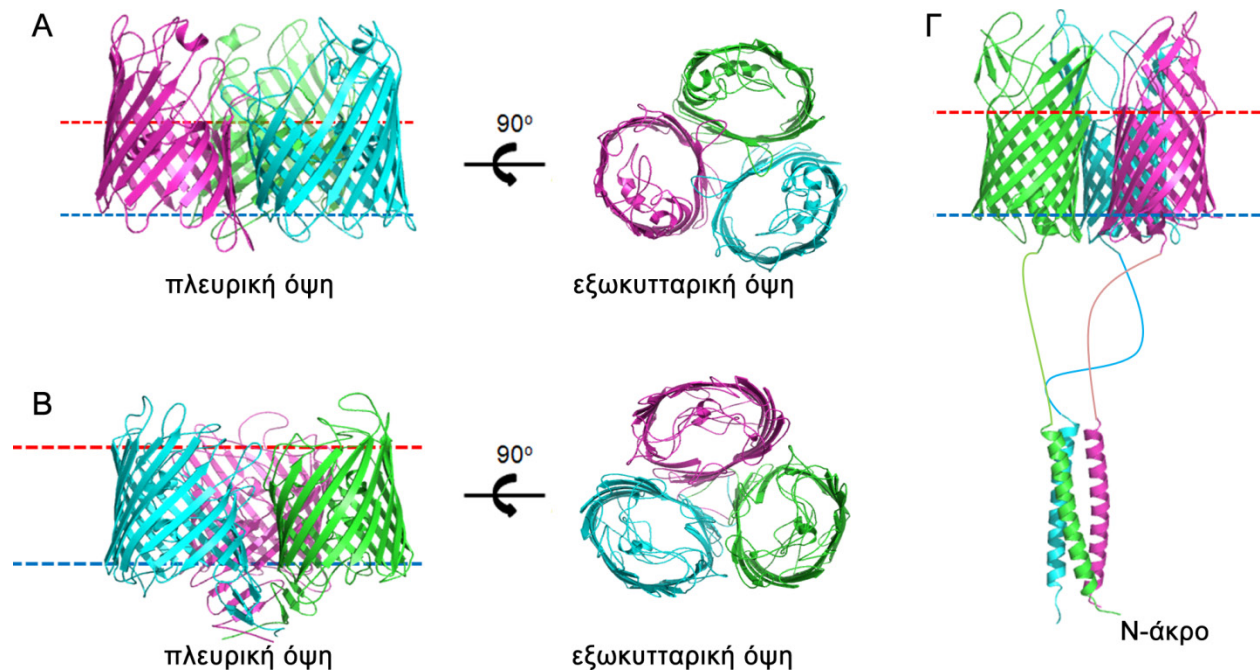
Κεφάλαιο I - Εισαγωγή

παρεκκλίσεις, σε μικρότερο ή μεγαλύτερο βαθμό, από αυτό το μηχανισμό. Το πιο χαρακτηριστικό παράδειγμα είναι η οικογένεια των υποδοχέων ινσουλίνης, οι οποίοι είναι υποχρεωτικά ετεροτετραμερή, αποτελούμενα από υπομονάδες α και β ενωμένες μεταξύ τους με δισουλφιδικούς δεσμούς (Lemmon & Schlessinger, 2010). Σε κάποια άλλα παραδείγματα υποδοχέων, πάλι, θεωρείται ότι η δομή του διμερούς είναι μερικώς σχηματισμένη πριν την δέσμευση του προσδέτη, η οποία αλλάζει τη στερεοδιάταξη του διμερούς, είτε στο επίπεδο των διαμεμβρανικών τμημάτων είτε στο επίπεδο της μεταμεμβρανικής περιοχής και των περιοχών PTK. Σε κάθε περίπτωση, η αναγνώριση και δέσμευση του προσδέτη είναι απαραίτητη για την πλήρη ενεργοποίηση των RTKs και την επαγωγή της ενζυμικής ενεργότητας των περιοχών PTK (Biaric et al., 2011).

1.5.3 Αλληλεπιδράσεις ολιγομερισμού στα διαμεμβρανικά β-βαρέλια

Όπως αναφέρθηκε στην Ενότητα 1.3.1.2, οι διαμεμβρανικές πρωτεΐνες που εντοπίζονται στην Εξωτερική Μεμβράνη των αρνητικών κατά Gram βακτηρίων είναι διαμεμβρανικά β-βαρέλια. Τα τελευταία λειτουργούν συχνά σαν πορίνες, οι οποίες ρυθμίζουν τη μεταφορά ιόντων, θρεπτικών ουσιών και διάφορων μεταβολιτών κατά μήκος της Εξωτερικής Μεμβράνης. Κάποιες από αυτές τις πορίνες (π.χ. OmpF) δεν παρουσιάζουν κάποια επιλεκτικότητα ως προς τις ουσίες που μεταφέρουν πέρα από το μέγεθός τους, το οποίο δεν πρέπει να ξεπερνά τη διάμετρο του πόρου. Από την άλλη, ειδικές πορίνες όπως οι μεταφορείς OmpP και FapF αναγνωρίζουν και μεταφέρουν επιλεκτικά συγκεκριμένα υποστρώματα, μέσω ειδικών θέσεων πρόσδεσης για αυτά πάνω στη δομή τους (Niramiiranon et al., 2016).

Κεφάλαιο I - Εισαγωγή



Εικόνα 21 Παραδείγματα ολιγομερισμού στα διαμεμβρανικά β-βαρέλια. (Α) Η πορίνη OmpF του βακτηρίου *E. coli*. (Β) Ο μεταφορέας φωσφορικών ιόντων OprP του βακτηρίου *P. aeruginosa*. (Γ) Ο μεταφορέας πρωτεϊνών των βιοϋμενίων FarF του βακτηρίου *P. aeruginosa*. Ο μεταφορέας σχηματίζει αλληλεπιδράσεις ολιγομερισμού τόσο μέσω των διαμεμβρανικών περιοχών του όσο και μέσω του αμινοτελικού άκρου του.

Ωστόσο, η ικανότητα πολλών πορινών να αναγνωρίζουν το υπόστρωμά τους έχει βρεθεί να σχετίζεται και με την ικανότητά τους να σχηματίζουν αλληλεπιδράσεις ολιγομερισμού. Συγκεκριμένα, διάφορες βακτηριακές πορίνες έχουν βρεθεί να σχηματίζουν ομοτρίμερη (Εικόνα 21), με τις αλληλεπιδράσεις ολιγομερισμού ανάμεσά τους να διαδραματίζει σημαντικό ρόλο τόσο στη δομή όσο και στη λειτουργία τους (Nirmitranon et al., 2016). Ο σχηματισμός ολιγομερών, και συγκεκριμένα τριμερών, έχει παρατηρηθεί δομικά κυρίως σε μια οικογένεια β-βαρελίων με 16 διαμεμβρανικούς β-κλώνους και αριθμό κάμψης 20 ($n=16$, $S=20$), γνωστή και με την ονομασία «τριμερείς πορίνες». Το πιο μελετημένο μέλος αυτής της οικογένειας είναι η πρωτεΐνη OmpF του βακτηρίου *E. coli*, μια γενική πορίνη η οποία μεταφέρει μη ειδικά ιόντα και αποτελεί στόχο αντιβιοτικών όπως η αμπικιλίνη και η τετρακυκλίνη (Ziervogel & Roux, 2013). Η OmpF έχει βρεθεί τόσο με βιοφυσικά πειράματα, μέσα από την επίλυση της τρισδιάστατης δομής της (Efremov & Sazanov, 2012), όσο και με βιοχημικές μελέτες (Surrey et al., 1996), να σχηματίζει λειτουργικά ομοτρίμερη στην Εξωτερική

Κεφάλαιο I - Εισαγωγή

Μεμβράνη της *E. coli*. Επιπρόσθετα, συμπληρωματικά δεδομένα από βιοφυσικές και υπολογιστικές μελέτες έχουν δείξει ότι η δομή της OmpF σε μονομερή κατάσταση είναι πιο ασταθής από ότι στην κατάσταση τριμερούς και παρουσιάζει μειωμένη δραστηριότητα (Niramiiranon et al., 2016).

Ωστόσο, αντίστοιχη οργάνωση τριμερών με βιολογικό ρόλο έχει παρατηρηθεί και σε β-βαρέλια που δεν διαθέτουν την αρχιτεκτονική « $n=16$, $S=20$ », υποδηλώνοντας ότι ο ολιγομερισμός δεν περιορίζεται αποκλειστικά σε πορίνες όπως η OmpF. Χαρακτηριστικό παράδειγμα αποτελεί ο μεταφορέας FarF, ένα λειτουργικό ομοτρίμερες β-βαρελίων το οποίο ρυθμίζει τη σύνθεση βιοϋμενίων (biofilms) στο βακτήριο *Pseudomonas aeruginosa* (Rouse et al., 2018). Η δομή του μεταφορέα FarF περιλαμβάνει μια αμινοτελική περιοχή με δομή α-έλικας η οποία βρίσκεται στον περιπλαστικό χώρο, ακολουθούμενη από μια δομή διαμεμβρανικού β-βαρελίου με αρχιτεκτονική « $n=12$, $S=14$ ». Τόσο τα ίδια τα διαμεμβρανικά β-βαρέλια (Rouse et al., 2017) όσο και οι αμινοτελικές α-έλικες του μεταφορέα έχουν δειχθεί κρυσταλλογραφικά ότι σχηματίζουν ομοτρίμερή με βιολογική σημασία (Εικόνα 21). Η αλληλεπίδραση των α-ελίκων συμμετέχει στην αγκυροβόληση των πρωτεϊνών-μονομερών των βιοϋμενίων (FarB, FarC και FarE) και την μεταφορά τους στην περιπλαστική όψη του FarF, ενώ τα β-βαρέλια λειτουργούν ως δίαυλοι για τη διέλευση των πρωτεϊνών, με τις αλληλεπιδράσεις ολιγομερισμού τους να ελέγχουν το άνοιγμα και το κλείσιμο του πόρου σε κάθε βαρέλι (Rouse et al., 2018).

Παρά τη βιολογική σημασία του ολιγομερισμού στα β-βαρέλια, πολύ λίγα είναι γνωστά πάνω στους δομικούς παράγοντες που επηρεάζουν το σχηματισμό των τριμερών τους. Υπολογιστικές μελέτες πάνω σε διαμεμβρανικά β-βαρέλια με γνωστή δομή έχουν υποδείξει ότι πολλά β-βαρέλια διαθέτουν περιοχές μειωμένης σταθερότητας («weakly-stable regions»), στις οποίες το δίκτυο υδρογονικών δεσμών που σταθεροποιεί τη δομή των β-κλώνων είναι σημαντικά πιο χαλαρό (Naveed et al., 2009). Η ανάλυση της αλληλουχίας β-βαρελίων με γνωστές αλληλεπιδράσεις ολιγομερισμού, όπως η OmpF, έχει δείξει ότι η επιφάνεια αλληλεπίδρασης των βαρελίων αποτελείται σε μεγάλο βαθμό από περιοχές μειωμένης σταθερότητας. Έτσι, έχει προταθεί ότι ο ολιγομερισμός χρησιμοποιείται από τα β-βαρέλια ως μέσο σταθεροποίησης αυτών των περιοχών (Naveed et al., 2009). Ένας δεύτερος παράγοντας

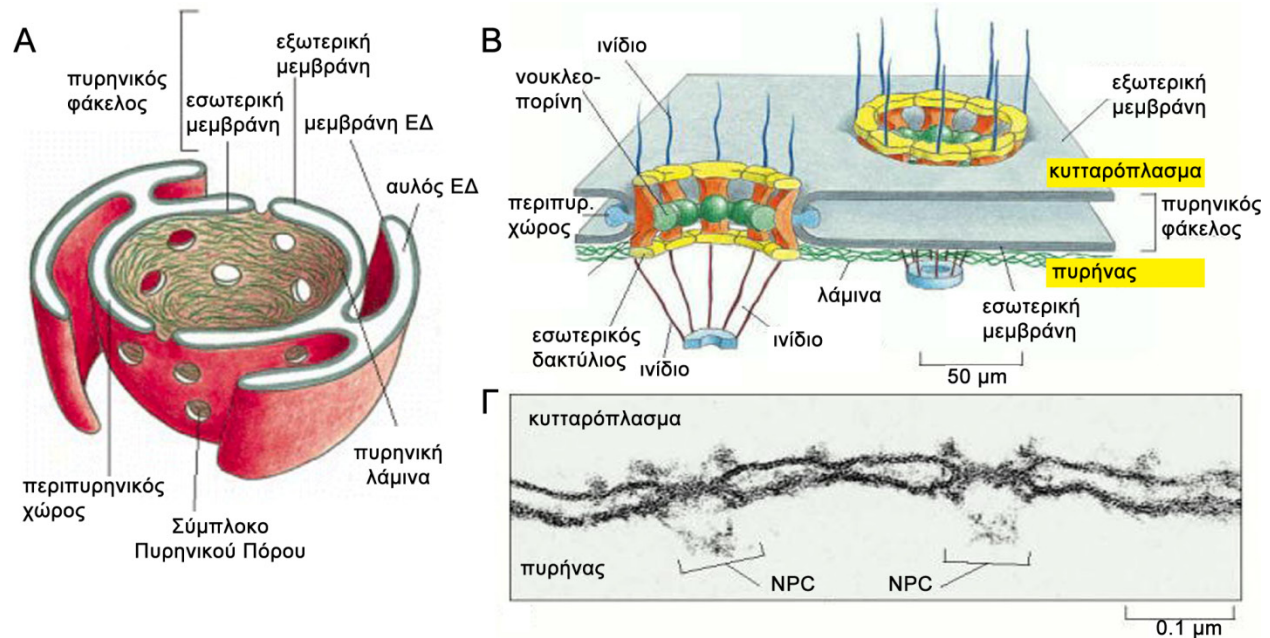
Κεφάλαιο I - Εισαγωγή

που έχει προταθεί να συμμετέχει στη ρύθμιση του ολιγομερισμού είναι οι αλληλεπιδράσεις των β-βαρελίων με μόρια λιποπολυσακχαριτών (βλ. Ενότητα 1.2.4) στην εξωκυτταρική στιβάδα της Εξωτερικής Μembrάνης. Συγκεκριμένα, βιοφυσικές και υπολογιστικές μελέτες πάνω στην OmpF έχουν δείξει ότι κατά τη βιοσύνθεση και μεταφορά της πρωτεΐνης στην Εξωτερική Μembrάνη, οι αλληλεπιδράσεις με μόρια λιποπολυσακχαριτών επηρεάζουν την πλευρική διάχυση των β-βαρελίων στη διπλοστιβάδα και κατευθύνουν το σχηματισμό τριμερών (Efremov & Sazanov, 2012; Ma et al., 2018; Patel et al., 2016). Ωστόσο, ο συγκεκριμένος μηχανισμός δεν έχει μελετηθεί, μέχρι στιγμής, σε άλλα παραδείγματα πορινών, και έτσι δεν είναι γνωστό αν ισχύει γενικά για τα β-βαρέλια ή ειδικά για την OmpF.

1.5.4 Οι πρωτεΐνες του πυρηνικού φακέλου

Ο πυρήνας των ευκαρυωτικών κυττάρων αποτελεί ένα από τα κυριότερα οργανίδια καθώς οργανώνει, προστατεύει και ρυθμίζει το γονιδίωμα, συμβάλλει στη διαδικασία της αντιγραφής του DNA, στη ρύθμιση της μεταγραφής των γονιδίων σε mRNA αλλά και στην πρωτεϊνοσύνθεση, καθώς επίσης και στην οργάνωση του κυτταροσκελετού στη μεσόφαση αλλά και στις διεργασίες της κυτταρικής διαίρεσης, είτε στη μίτωση είτε στη μείωση (Wilson & Berk, 2010). Ο πυρήνας περιβάλλεται από ένα ιδιαίτερο σύστημα διπλής μεμβράνης, γνωστό σαν «πυρηνικός φάκελος» (Εικόνα 22). Ο τελευταίος διαχωρίζει το κυτταρόπλασμα και το εσωτερικό του πυρήνα, και ελέγχει όλες τις διαδικασίες ανταλλαγής μακρομορίων μεταξύ των δύο. Επιπρόσθετα, συνεισφέρει στον καθορισμό της χωρικής οργάνωσης του πυρήνα. Τα κύρια δομικά στοιχεία του πυρηνικού φακέλου είναι η εσωτερική και η εξωτερική πυρηνική μεμβράνη, τα Σύμπλοκα Πυρηνικού Πόρου (NPCs) και η πυρηνική λάμινα (Bridger et al., 2007), η οποία επενδύει την εσωτερική μεμβράνη από την πλευρά του πυρηνοπλάσματος (Nigg, 1989).

Κεφάλαιο I - Εισαγωγή



Εικόνα 22 Η δομή του πυρηνικού φακέλου. (Α) Ο πυρηνικός φάκελος των ευκαρυωτικών κυττάρων. (Β) Η οργάνωση του πυρηνικού φακέλου και το Σύνπλοκο Πυρηνικού Πόρου (NPC). (Γ) Παρατήρηση του πυρηνικού φακέλου από Ηλεκτρονικό Μικροσκόπιο Διέλευσης. Διακρίνονται δύο Σύνπλοκα Πυρηνικού Πόρου. Προσαρμογή από Alberts et al., 2002.

Η εξωτερική μεμβράνη αποτελεί φυσική συνέχεια της μεμβράνης του ΕΔ (Εικόνα 22). Ωστόσο, δεν διαθέτει τις πρωτεΐνες που προσδίδουν στο ΕΔ το χαρακτηριστικό σωληνοειδές σχήμα του. Αντίθετα, περιέχει πλήθος άλλων πρωτεϊνών, σημαντικός αριθμός των οποίων αλληλεπιδρούν με τον κυτταροσκελετό (Bonne, 2014). Η εσωτερική μεμβράνη συναντά την εξωτερική στα σύνπλοκα του πυρηνικού πόρου, ωστόσο είναι διαφορετική τόσο σε λιπιδική σύσταση όσο και στις πρωτεΐνες που περιέχει, παρά την στενή σύνδεσή τους. Η εσωτερική πυρηνική μεμβράνη φέρει μια σειρά από μεμβρανικές πρωτεΐνες που εντοπίζονται αποκλειστικά στον πυρήνα και αλληλεπιδρούν με την πυρηνική λάμινα. Ο περιμεμβρανικός χώρος ανάμεσα στις δύο μεμβράνες είναι γνωστός σαν περιπυρηνικός χώρος και αποτελεί ουσιαστικά συνέχεια του αυλού του ΕΔ. Και οι δύο μεμβράνες είναι διαπερατές από μικρά υδρόφοβα μόρια, ενώ πολικές και φορτισμένες ουσίες και μεγαλύτερα σε μέγεθος μόρια, συμπεριλαμβανομένων των πρωτεϊνών και των μορίων RNA, περνούν μόνο από τα NPCs. Η επιφάνεια της εσωτερικής πυρηνικής μεμβράνης καλύπτεται ένα στρώμα ενδιάμεσων ινιδίων που διασυνδέονται με πληθώρα διαμεμβρανικών πρωτεϊνών και με παράγοντες που

Κεφάλαιο I - Εισαγωγή

σχετίζονται με την χρωματίνη, γνωστό σαν πυρηνικό έλασμα ή πυρηνική λάμινα (Bridger et al., 2007). Η πυρηνική λάμινα προσφέρει μία εξαιρετική σταθερότητα στον πυρηνικό φάκελο, και κατ' επέκταση στον πυρήνα, καθώς είναι εν μέρη υπεύθυνη για την διατήρηση του σχήματός του (Alberts et al., 2002).

Το Σύμπλοκο Πυρηνικού Πόρου (Nuclear Pore Complex, NPC) είναι ένα μεγάλο υπερμοριακό σύμπλοκο, αποτελούμενο από πολλαπλές υπομονάδες, το οποίο εντοπίζεται αποκλειστικά στον πυρηνικό φάκελο και δρα σαν υδατικό κανάλι και σαν μεταφορέας τόσο μικρών μορίων όσο και μακρομορίων (πρωτεΐνες, RNA), ελέγχοντας τη διέλευσή τους από και προς το εσωτερικό του πυρήνα (Alberts et al., 2002). Οι πρωτεΐνες που σχηματίζουν τα NPCs ομαδοποιούνται με το γενικό όρο «Νουκλεοπορίνες» (Nucleoporins, Nups). Πρόκειται για μια πολύ ετερογενή ομάδα πρωτεϊνών, η οποία περιλαμβάνει σφαιρικές υδατοδιαλυτές, δομικές ινώδεις και μεμβρανικές πρωτεΐνες. Τα NPCs των σπονδυλοζώων δομούνται από 30 διαφορετικές Nups, οι οποίες απαντώνται σε μεγάλο αριθμό αντιγράφων. Έτσι, ένα τυπικό NPC δομείται από περισσότερες από 500 διαφορετικές υπομονάδες, οι οποίες οργανώνονται σε ένα μεγάλο υπερμοριακό σύμπλεγμα μοριακού βάρους 12.5 MDa και σφαιρική συμμετρία όγδοης τάξης, με μία εξωτερική διάμετρο περίπου 100-120 nm και ένα κεντρικό κανάλι μεταφοράς διαμέτρου 40 nm (Strambio-De-Castillia et al., 2010). Σε κάθε πυρήνα υπάρχουν 3000-4000 NPCs, με αποτέλεσμα ο πυρηνικός φάκελος να είναι ουσιαστικά ένα διάτρητο στρώμα, από το οποίο διέρχονται συνεχώς διάφορα υποστρώματα (Alberts et al., 2002). Από το δίαυλο κάθε NPC μεταφέρονται επιλεκτικά πολικές και φορτισμένες ουσίες, μόρια mRNA και tRNA, πρωτεΐνες ή και ολόκληρα σύμπλοκα πρωτεϊνών-πρωτεϊνών και πρωτεϊνών-RNA. Η μεταφορά γίνεται τόσο από το εσωτερικό του πυρήνα στο κυτταρόπλασμα όσο και αντίστροφα (Strambio-De-Castillia et al., 2010). Τα μικρά μόρια μεταφέρονται παθητικά, ακολουθώντας τη ροή του διαλύτη μέσα από τον ανοικτό δίαυλο. Από την άλλη, οι πρωτεΐνες και τα μόρια RNA μεταφέρονται μέσω ενεργούς μεταφοράς, κατά την οποία γίνεται η αναγνώριση κατάλληλων πρωτεϊνών και RNA καθώς και η επιλεκτική τους μετακίνηση προς συγκεκριμένη κατεύθυνση (Alberts et al., 2002).

Κεφάλαιο I - Εισαγωγή

1.5.4.1 Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο

Η πυρηνική λάμινα και τα NPCs δεν είναι οι μοναδικές πρωτεΐνες του πυρηνικού φακέλου. Αντίθετα, πλήθος άλλων μεμβρανικών πρωτεϊνών, είτε διαμεμβρανικών είτε περιφερειακών, έχουν εντοπιστεί τόσο στην εξωτερική όσο και την εσωτερική πυρηνική μεμβράνη (Bonne, 2014). Κάποιες από αυτές τις πρωτεΐνες συμμετέχουν άμεσα στη φυσική σύνδεση της εξωτερικής πυρηνικής μεμβράνης με την εσωτερική, συμβάλλοντας με αυτόν στην επικοινωνία του κυτταροπλάσματος με το εσωτερικό του πυρήνα. Χαρακτηριστικό παράδειγμα αποτελούν οι πρωτεΐνες KASH και SUN, οι οποίες εντοπίζονται στην εξωτερική και την εσωτερική πυρηνική μεμβράνη, αντίστοιχα (Crisp et al., 2006). Εκτός από τις παραπάνω, μια μεγάλη σειρά από διαμεμβρανικές πρωτεΐνες έχουν βρεθεί να εκφράζονται αποκλειστικά στις μεμβράνες του πυρηνικού φακέλου, γνωστές με το γενικό όρο «Διαμεμβρανικές Πρωτεΐνες του Πυρηνικού Φακέλου» (Nuclear Envelope Transmembrane Proteins, NETs) (Wong et al., 2014). Ωστόσο, ο λειτουργικός ρόλος των περισσότερων NETs δεν έχει αποσαφηνιστεί πλήρως. Πολλές πρωτεΐνες του πυρηνοπλάσματος αλληλεπιδρούν επίσης με την εσωτερική μεμβράνη του πυρήνα ως περιφερειακές, ανάμεσα στις οποίες συγκαταλέγονται πολλοί μεταγραφικοί παράγοντες και διάφορα ένζυμα.

Τέλος, πρέπει να σημειωθεί ότι πολλές μεμβρανικές πρωτεΐνες, οι οποίες τυπικά εντοπίζονται σε διαφορετικές υποκυτταρικές θέσεις, έχουν βρεθεί να εκφράζονται και στον πυρηνικό φάκελο. Για παράδειγμα, η εξωτερική πυρηνική μεμβράνη έχει βρεθεί να περιέχει διαμεμβρανικές πρωτεΐνες της κυτταρικής μεμβράνης όπως διάφορους RTKs (Wang et al., 2015), GPCRs (Tadevosyan et al., 2012), αντλίες πρωτονίων και ιοντικά κανάλια (Bkaily et al., 2012), καθώς και περιφερειακές ή αγκυροβολημένες πρωτεΐνες όπως G-πρωτεΐνες και διάφορες φωσφολιπάσες (Bkaily et al., 2009). Η παρουσία αυτών των πρωτεϊνών σε κάποιες περιπτώσεις είναι κομμάτι του κύκλου λειτουργίας τους. Χαρακτηριστικό παράδειγμα αποτελούν RTKs όπως ο υποδοχέας EGFR (Carpenter & Liao, 2013), ο οποίος μετά την ενεργοποίησή του ενδοκυτώνεται και μεταφέρεται μέσω κυστιδίων στον πυρηνικό φάκελο (Wang et al., 2015). Σε περιπτώσεις όπως οι GPCRs και τα ιοντικά κανάλια, η μεταφορά τους στον πυρηνικό φάκελο αντί για την κυτταρική μεμβράνη πραγματοποιείται μετά την είσοδό τους στο ΕΔ, είτε σαν απόκριση σε κάποιο σηματοδοτικό μηχανισμό είτε σαν αποτέλεσμα

Κεφάλαιο I - Εισαγωγή

κάποιας ασθένειας (Bkaily et al., 2012; Tadevosyan et al., 2012). Τέλος, πολλές περιφερειακές και αγκυροβολημένες μεμβρανικές πρωτεΐνες μπορούν να βρεθούν σε διαφορετικές θέσεις στο κύτταρο, συμπεριλαμβανομένου του πυρήνα και των οργανιδίων, ανάλογα με το στάδιο του κυτταρικού κύκλου (Bonne, 2014).

Όλες οι προαναφερθείσες πρωτεΐνες συμμετέχουν σε πλήθος αλληλεπιδράσεων, τόσο με σταθερά στοιχεία του φακέλου όπως τα NPCs και η πυρηνική λάμινα όσο και μεταξύ τους αλλά και με μεγάλο αριθμό άλλων πρωτεϊνών του κυττάρου. Για την ακρίβεια, υπολογίζεται πλέον ότι το σύνολο των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών του πυρηνικού φακέλου προσεγγίζει σε ετερογένεια και πολυπλοκότητα αυτό της κυτταρικής μεμβράνης (Stewart, 2007). Μέσω αυτών των αλληλεπιδράσεων ο πυρηνικός φάκελος συμμετέχει σε πλήθος κυτταρικών λειτουργιών, από τη μεταγωγή σήματος, την κυτταρική αύξηση, τη διαίρεση και την πρωτεϊνοσύνθεση ως τη ρύθμιση της γονιδιακής έκφρασης και τον κυτταρικό θάνατο. Ωστόσο, ο βιολογικός ρόλος μεγάλου μέρους των πρωτεϊνών του φακέλου και των αλληλεπιδράσεών τους παραμένει, σε μεγάλο βαθμό, άγνωστος (Wong et al., 2014).

1.5.5 Οι Διαμεμβρανικοί Μεταφορείς SLC11

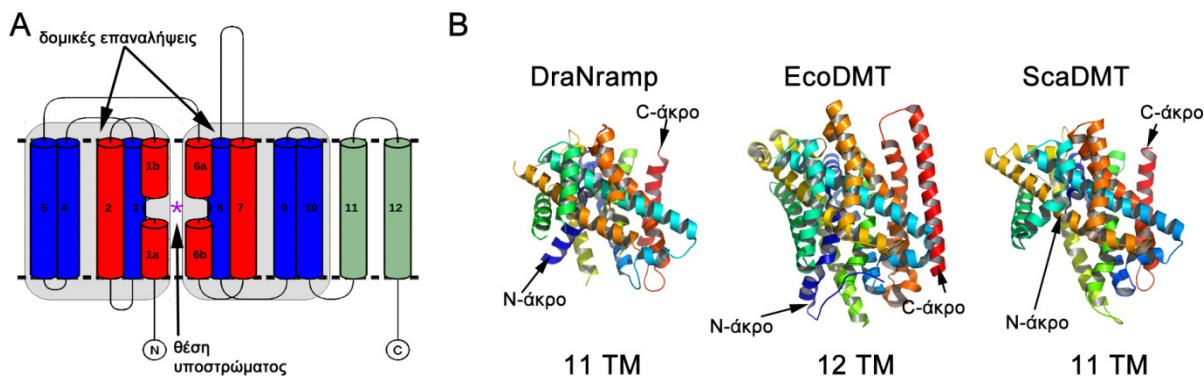
Η οικογένεια μεταφορέων 11 («Solute Carrier Family 11», SLC11), γνωστή και σαν «οικογένεια πρωτεϊνών των μακροφάγων που σχετίζονται με τη φυσική αντίσταση» («Natural Resistance Associate Macrophage Proteins», NRAMPs) περιλαμβάνει μια σειρά από μεταφορείς μεταλλικών ιόντων συζευγμένων με τη μεταφορά πρωτονίων (Cellier, 2012). Τα πρώτα μέλη της οικογένειας (SLC11A1 και SLC11A2) ανακαλύφθηκαν σε μακροφάγα σπονδυλοζώων και αρχικά θεωρήθηκε ότι συμμετέχουν στους μηχανισμούς άμυνας, λόγω της συσχέτισης του SLC11A1 με τη μόλυνση του οργανισμού από διάφορα παράσιτα (Cellier et al., 1994). Ωστόσο, πλέον είναι γνωστό ότι ομόλογοι μεταφορείς των SLC11A1 και SLC11A2, γνωστοί με τη γενικότερη ονομασία NRAMPs, υπάρχουν σε έντομα, νηματώδεις σκώληκες, φυτά, μύκητες και βακτήρια. Σε όλες τις παραπάνω περιπτώσεις, οι μεταφορείς SLC11 είναι υπεύθυνοι κυρίως για τη μεταφορά δισθενών μεταλλικών ιόντων (Fe^{2+} , Mg^{2+} , Mn^{2+} κλπ) κατά

Κεφάλαιο I - Εισαγωγή

μήκος της κυτταρικής μεμβράνης, ενώ η λειτουργία τους ρυθμίζεται από την παράλληλη μεταφορά πρωτονίων (Cellier, 2012).

Όλα τα μέλη της οικογένειας SLC11 χαρακτηρίζονται από την ύπαρξη μιας συντηρημένης αυτοτελούς δομικής περιοχής με κοινή μεμβρανική τοπολογία, γνωστής σαν περιοχή NRAMP (Εικόνα 23). Η περιοχή NRAMP αποτελείται από 10 διαμεμβρανικά τμήματα με δομή α-έλικας, με το N-άκρο να βρίσκεται στον κυτοπλασματικό χώρο (Cellier, 2012). Μελέτες φυλογενετικής ανάλυσης έχουν δείξει ότι η περιοχή παρουσιάζει απομακρυσμένη συγγένεια με τους βακτηριακούς μεταφορείς λευκίνης (LeuT) οι οποίοι, παρότι κατατάσσονται στην οικογένεια μεταφορέων 6 (SLC6) θεωρούνται αποτελούν απομακρυσμένα δομικά ομόλογα των SLC11 (Krishnamurthy & Gouaux, 2012). Εκτός από την περιοχή NRAMP, οι περισσότεροι μεταφορείς SLC11 περιλαμβάνουν επιπρόσθετα διαμεμβρανικά τμήματα με τον αριθμό τους να ποικίλει ανάμεσα στους οργανισμούς. Συγκεκριμένα, οι μεταφορείς SLC11 των σπονδυλοζών διαθέτουν 12 διαμεμβρανικά τμήματα και κυτοπλασματικό N- και C-άκρο, με ελάχιστες εξαιρέσεις όπως ο μεταφορέας NRAMP του οστεϊχθύος *Oncorhynchus mykiss*, ο οποίος διαθέτει 13 διαμεμβρανικά τμήματα. Από την άλλη, οι μεταφορείς SLC11 των ασπόνδυλων, των φυτών, των μυκήτων και των βακτηρίων μπορεί να διαθέτουν από 10 ως 13 διαμεμβρανικά τμήματα, με το C-άκρο τους να βρίσκεται ενδοκυτταρικά ή εξωκυτταρικά (Cellier, 2012). Η παραπάνω τοπολογία έχει μελετηθεί κυρίως με υπολογιστικές μεθόδους ανάλυσης ακολουθιών για τους ευκαρυωτικούς οργανισμούς, ενώ πιο πρόσφατα προσδιορίστηκαν μια σειρά από κρυσταλλικές δομές βακτηριακών μεταφορέων SLC11 (Εικόνα 23), οι οποίες επιβεβαίωσαν την ομοιότητα της περιοχής NRAMP με το δίπλωμα των μεταφορέων LeuT αλλά και τη συνολική τοπολογία των μεταφορέων SLC11 (Bozzi et al., 2016a; Ehrnstorfer et al., 2014; Ehrnstorfer et al., 2017).

Κεφάλαιο I - Εισαγωγή



Εικόνα 23 Οι διαμεμβρανικοί μεταφορείς SLC11. (Α) Διάγραμμα της τοπολογίας των SLC11 στη μεμβράνη. Η περιοχή NRAMP αποτελείται από 10 διαμεμβρανικά τμήματα τα οποία οργανώνονται σε δύο ανεστραμμένες δομικές επαναλήψεις (γκρίζα πλαίσια). Ανάμεσα στις δύο επαναλήψεις εντοπίζεται η θέση πρόσδεσης του υποστρώματος (μεταλλικά ιόντα), η οποία θεωρείται ότι σχηματίζεται από τα διαμεμβρανικά τμήματα 1, 2, 6 και 7 (κόκκινο χρώμα). Πολλοί μεταφορείς SLC11 διαθέτουν, εκτός από την περιοχή NRAMP, και επιπρόσθετα διαμεμβρανικά τμήματα, με τον αριθμό τους να ποικίλλει ανάμεσα σε διαφορετικά είδη. (Β) Οι τρισδιάστατες δομές των βακτηριακών ομολόγων των SLC11, DraNramp (*Deinococcus radiodurans*), EcoDMT (*Eremococcus coleocola*) και ScaDMT (*Staphylococcus capitis*). Ο μεταφορέας EcoDMT διαθέτει 12 διαμεμβρανικά τμήματα και κοινή τοπολογία για το N- και το C-άκρο, ενώ οι ScaDMT και DraNramp διαθέτουν 11 διαμεμβρανικά τμήματα και αντίστροφη τοπολογία για τα άκρα. Στην περίπτωση του ScaDMT το πρώτο μισό τμήμα του διαμεμβρανικού τμήματος TM1 δεν έχει προσδιοριστεί.

Στα θηλαστικά η οικογένεια SLC11 περιλαμβάνει δύο μέλη, τους μεταφορείς SLC11A1 και SLC11A2, γνωστούς και ως NRAMP1 και NRAMP2 (Cellier et al., 1994). Ο μεταφορέας SLC11A2, γνωστός και με το όνομα μεταφορέας δισθενών μεταλλικών ιόντων 1 (Divalent Metal Transporter 1, DMT1), εκφράζεται στο σύνολο των κυττάρων του οργανισμού και επιτελεί τη μεταφορά μεγάλου εύρους δισθενών ιόντων (Fe^{2+} , Mn^{2+} , Ca^{2+} , Zn^{2+} κλπ) από το εξωτερικό του κυττάρου προς το κυτταρόπλασμα, με παράλληλη συμμεταφορά πρωτονίων (Forbes & Gros, 2003). Από την άλλη, ο μεταφορέας SLC11A1 εντοπίζεται κυρίως στα μακροφάγα και, σε μικρότερο βαθμό, στα κύτταρα του ήπατος και μεταφέρει ιόντα Fe^{2+} και Mn^{2+} (Vidal et al., 1996). Η ορθή λειτουργία του SLC11A1 είναι ιδιαίτερα σημαντική για τη φυσιολογία του οργανισμού. Μεταλλαγές που οδηγούν σε απώλεια της ικανότητας του SLC11A1 στη μεταφορά σιδήρου έχουν βρεθεί να οδηγούν στη λανθασμένη φαγοκυττάρωση ερυθρών αιμοσφαιρίων, προκαλώντας αιμολυτική αναιμία (Soe-Lin et al., 2010), ενώ η δυσλειτουργία του στα ηπατικά κύτταρα έχει συσχετιστεί με την παθολογία ασθενών με διαβήτη Τύπου I (Paccagnini et al., 2009). Παράλληλα, μεταλλαγές και πολυμορφισμοί του SLC11A1 έχουν

Κεφάλαιο I - Εισαγωγή

βρεθεί να σχετίζονται με την αυξημένη ευαισθησία του οργανισμού στη μόλυνση από ενδοκυτταρικά παράσιτα. Συγκεκριμένα, ο ανθρώπινος SLC11A1 έχει συσχετιστεί με τη μόλυνση από *Salmonella enterica* (σαλμονέλωση), *Leishmania donovani* (λεϊσμανίαση) και διάφορα είδη του γένους *Mycobacterium* (φυματίωση, ασθένεια του Crohn, λέπρα, ρευματοειδής αρθρίτιδα, έλκη Bulceri κλπ), ενώ σε άλλα θηλαστικά όπως η αίγα (*Capra hircus*) και τα βοοειδή (*Bos taurus*), ο SLC11A1 έχει συσχετιστεί με τη μόλυνση από το βακτήριο *Mycobacterium avium* και την ασθένεια του John (παραφυματίωση) (Bellamy et al., 1998; Paccagnini et al., 2009; Roupie et al., 2008; Ruiz-Larrañaga, 2009; Sechi & Dow, 2015; Wessling-Resnick, 2015).

Παρά τη βιολογική και κλινική σημασία του SLC11A1, ωστόσο, ελάχιστα είναι γνωστά τόσο για τη δομή του όσο και για τον τρόπο λειτουργίας του. Η ανάλυση της αλληλουχίας του μεταφορέα παρουσιάζει υψηλή ομολογία (ταύτιση ~90%) με τον περισσότερο μελετημένο μεταφορέα SLC11A2/DMT1, χωρίς ωστόσο να υπάρχει πειραματικά προσδιορισμένη δομική πληροφορία από βιοφυσικά ή βιοχημικά πειράματα. Αντίστοιχα με τον SLC11A2/DMT1, η μεταφορά των ιόντων $\text{Fe}^{2+}/\text{Mn}^{2+}$ είναι συζευγμένη με την παράλληλη μεταφορά πρωτονίων. Ωστόσο, σε αντίθεση με τον SLC11A2/DMT1, ο οποίος είναι γνωστό ότι δρα σαν συμμεταφορέας ιόντων-πρωτονίων, η ακριβής λειτουργία του SLC11A1 δεν είναι πλήρως γνωστή, με κάποιες πειραματικές μελέτες να υποδεικνύουν ότι δρα σαν αντιμεταφορέας ιόντων-πρωτονίων, ενώ διαφορετικά δεδομένα υποστηρίζουν ότι η μεταφορά πρωτονίων μπορεί να συμβεί αμφίδρομα, ανάλογα με το είδος του οργανισμού (Techau et al., 2007). Επιπρόσθετα, ενώ ο SLC11A2/DMT1 αναγνωρίζει μεγάλο εύρος ιόντων με υψηλή συγγένεια προς το Fe^{2+} , ο SLC11A1 μεταφέρει σχεδόν αποκλειστικά ιόντα $\text{Fe}^{2+}/\text{Mn}^{2+}$ και παρουσιάζει υψηλή συγγένεια για το Mn^{2+} έναντι του Fe^{2+} (Cellier, 2012).

1.6 Οι επιρροές της μεμβράνης στις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών

Αντίστοιχα με τις αλληλεπιδράσεις των σφαιρικών υδατοδιαλυτών πρωτεϊνών, η φύση και η ισχύς των οποίων καθορίζεται όχι μόνο από τα αλληλεπιδρώντα κατάλοιπα αλλά και από την επίδραση του διαλύτη (Chothia, 1976; Chothia & Janin, 1975; Kastiris & Bonvin, 2013), οι αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών επηρεάζονται σε μεγάλο βαθμό από τη λιπιδική διπλοστιβάδα (Mondal et al., 2014; van Meer et al., 2008). Η παραπάνω παρατήρηση έχει επιβεβαιωθεί επανειλημμένα για τις αλληλεπιδράσεις διάφορων κατηγοριών διαμεμβρανικών πρωτεϊνών, από τους GPCRs (Periole, 2017; Sengupta & Chattopadhyay, 2015) και τα διαμεμβρανικά β-βαρέλια (Ma et al., 2018; Patel et al., 2016) ως τους RTKs (Hedger et al., 2015) και διάφορους διαμεμβρανικούς μεταφορείς (Pyle et al., 2018). Επιπρόσθετα, οι ιδιότητες της λιπιδικής διπλοστιβάδας έχουν βρεθεί να επηρεάζουν τις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών με μη-διαμεμβρανικές πρωτεΐνες (π.χ. αλληλεπιδράσεις GPCRs-G-πρωτεϊνών), καθώς και τις αλληλεπιδράσεις πρωτεϊνών που προσκολλούνται στην επιφάνεια της μεμβράνης, όπως είναι οι περιφερειακές και αγκυροβολημένες πρωτεΐνες αλλά και διάφορα πεπτίδια (Giocondi et al., 2010). Τέλος, οι επιρροές της μεμβράνης παρουσιάζουν, σε αρκετές περιπτώσεις, και φαρμακολογικό ενδιαφέρον, καθώς μπορούν να καθορίσουν την ικανότητα αλληλεπίδρασης σε πρωτεΐνες και πεπτίδια με θεραπευτικό ρόλο. Χαρακτηριστικό παράδειγμα αποτελούν τα αντιμικροβιακά πεπτίδια, η ικανότητα συσσωμάτωσης των οποίων συχνά εξαρτάται από τη σύσταση των βακτηριακών μεμβρανών στις οποίες προσδένονται (Bello et al., 2015).

Οι επιρροές της μεμβράνης πάνω στις πρωτεϊνικές αλληλεπιδράσεις μπορούν, γενικά, να χωριστούν σε δύο κατηγορίες, στις άμεσες και έμμεσες επιρροές (Sengupta & Chattopadhyay, 2015). Οι άμεσες επιρροές της μεμβράνης περιλαμβάνουν τις άμεσες επαφές ανάμεσα στις πρωτεΐνες και σε λιπίδια της μεμβράνης, και το αποτέλεσμα αυτών των επαφών πάνω στις αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών. Χαρακτηριστικό παράδειγμα αποτελούν οι αλληλεπιδράσεις των GPCRs με μόρια χοληστερόλης, τα οποία έχουν βρεθεί να επηρεάζουν την ικανότητα των υποδοχέων στο σχηματισμό διμερών και ολιγομερών (Sengupta &

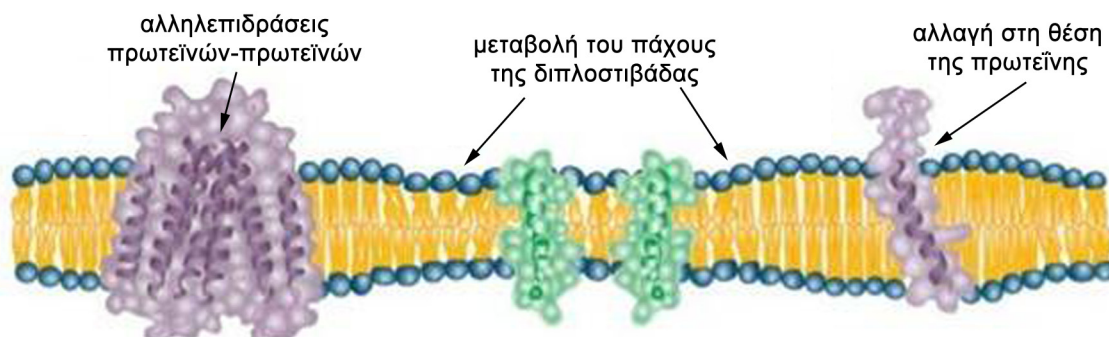
Κεφάλαιο I - Εισαγωγή

Chattopadhyay, 2015; Zheng et al., 2012). Αντίστοιχα, τόσο η βιοσύνθεση όσο και η ικανότητα των διαμεμβρανικών β-βαρελίων στο σχηματισμό ολιγομερών έχει συνδεθεί με την παρουσία λιποπολυσακχαριτών στην Εξωτερική Μεμβράνη (Arunmanee et al., 2016; Ma et al., 2018). Η παρουσία λιπιδίων με συγκεκριμένες πολικές κεφαλές έχει βρεθεί επίσης να επηρεάζει το διμερισμό των RTKs και τη δομή της μετα-μεμβρανικής περιοχής τους, οδηγώντας σε διαφορές στην ενεργοποίηση της ενεργότητας κινάσης των υποδοχέων (Hedger et al., 2015). Ο τρόπος με τον οποίο οι αλληλεπιδράσεις των πρωτεϊνών με τα λιπίδια επηρεάζει τις πρωτεϊνικές αλληλεπιδράσεις δεν έχει αποσαφηνιστεί πλήρως. Σε κάποιες περιπτώσεις τα λιπίδια φαίνονται να καθοδηγούν τις πρωτεΐνες κατά την πλευρική διάχυσή τους στη μεμβράνη, ωθώντας τις να έλθουν κοντά στο χώρο και να σχηματίσουν διαμοριακές επαφές. Από την άλλη, κάποια λιπίδια φαίνονται να δρουν αλλοστερικά πάνω στις πρωτεΐνες, επηρεάζοντας τη δομή, τη λειτουργία και τις αλληλεπιδράσεις τους με άλλες πρωτεΐνες (Hedger et al., 2015; Pyle et al., 2018; Sengupta & Chattopadhyay, 2015). Σε αντίθεση με τις άμεσες επιρροές που περιγράφηκαν παραπάνω, οι έμμεσες επιρροές της μεμβράνης προκύπτουν από την επίδραση των διάφορων μηχανικών ιδιοτήτων της διπλοστιβάδας πάνω στην πρωτεϊνική δομή (Sengupta & Chattopadhyay, 2015). Έτσι, χαρακτηριστικά όπως το πάχος της διπλοστιβάδας, η πλευρική διάχυση και ο βαθμός οργάνωσης των λιπιδίων, τα οποία σχετίζονται άμεσα με τη λιπιδική φάση της μεμβράνης (βλ. Ενότητα 1.2.5) μπορούν να επηρεάσουν τόσο τη θέση των διαμεμβρανικών πρωτεϊνών στο εσωτερικό της μεμβράνης όσο και την κινητικότητά τους, ρυθμίζοντας την ικανότητά τους να έρχονται κοντά στο χώρο και να σχηματίζουν σύμπλοκα (Giocondi et al., 2010; Periole, 2017).

Ένα από τα πιο χαρακτηριστικά παραδείγματα επίδρασης της διπλοστιβάδας πάνω στις αλληλεπιδράσεις πρωτεϊνών-πρωτεϊνών είναι η «υδροφοβική ασυμφωνία» (hydrophobic mismatch). Ως υδροφοβική ασυμφωνία χαρακτηρίζεται το φαινόμενο κατά το οποίο το πάχος της υδρόφοβης διπλοστιβάδας είναι διαφορετικό από το μήκος των διαμεμβρανικών τμημάτων μιας πρωτεΐνης (Mondal et al., 2011). Αυτή η ασυμφωνία μπορεί να συμβεί λόγω στερεοδιαταξικών αλλαγών στη δομή μιας πρωτεΐνης (π.χ. κατά την ενεργοποίηση ενός υποδοχέα, ή κατά το άνοιγμα ή κλείσιμο ενός μεταφορέα), ή κατά τη μεταφορά της πρωτεΐνης σε διαφορετικό μεμβρανικό περιβάλλον (π.χ. μεταφορά από το ΕΔ στην κυτταρική μεμβράνη).

Κεφάλαιο I - Εισαγωγή

Σε κάθε περίπτωση, η ασυμφωνία έχει σαν αποτέλεσμα την ατελή τοποθέτηση της πρωτεΐνης στο επίπεδο της διπλοστιβάδας, οδηγώντας σε μη ευνοϊκές αλληλεπιδράσεις ανάμεσα στα κατάλοιπα που βρίσκονται στα όρια των διαμεμβρανικών τμημάτων και το περιβάλλον τους. Έτσι, αν το πάχος της διπλοστιβάδας είναι μικρότερο από το μήκος του διαμεμβρανικού τμήματος (θετική ασυμφωνία), τα υδρόφοβα κατάλοιπα που δεν καλύπτονται από τη μεμβράνη εκτίθενται στον πολικό διαλύτη. Αν, αντίθετα, το πάχος της διπλοστιβάδας είναι μεγαλύτερο από το μήκος του διαμεμβρανικού τμήματος (αρνητική ασυμφωνία), τότε πολικά και φορτισμένα κατάλοιπα στους βρόχους της πρωτεΐνης ενδέχεται να θαφτούν στο υδρόφοβο εσωτερικό της μεμβράνης (Mondal et al., 2014).



Εικόνα 24 Προσαρμογές στο φαινόμενο της υδροφοβικής ασυμφωνίας. Η εξομάλυνση της υδροφοβικής ασυμφωνίας μπορεί να πραγματοποιηθεί μέσα από τοπικές αλλαγές στο πάχος της διπλοστιβάδας, με μεταβολές στη θέση των πρωτεϊνών στο επίπεδο της μεμβράνης ή μέσα από το σχηματισμό αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών.

Καθώς η υδροφοβική ασυμφωνία οδηγεί σε θερμοδυναμικά μη ευνοϊκές καταστάσεις, έχει προταθεί ότι τα βιολογικά συστήματα εφαρμόζουν μια σειρά από προσαρμογές για την εξομάλυνσή της, τόσο από την πλευρά της μεμβράνης όσο και από την πλευρά των πρωτεϊνών (Εικόνα 24). Συγκεκριμένα, έχει βρεθεί ότι σε περιπτώσεις υδροφοβικής ασυμφωνίας παρατηρούνται τοπικές παραμορφώσεις της λιπιδικής διπλοστιβάδας γύρω από τις πρωτεΐνες, με στόχο τη μείωση των μη ευνοϊκών επαφών. Έτσι, σε περιοχές με θετική ασυμφωνία παρατηρείται τοπική αύξηση του πάχους της διπλοστιβάδας, ώστε να καλυφθούν τα εκτεθειμένα υδρόφοβα κατάλοιπα της πρωτεΐνης, ενώ σε περιοχές με αρνητική ασυμφωνία παρατηρείται μείωση του πάχους, ώστε τα θαμμένα υδρόφιλα κατάλοιπα να εκτεθούν στο

Κεφάλαιο I - Εισαγωγή

διαλύτη. Ωστόσο, οι ελαστικές παραμορφώσεις της μεμβράνης χαρακτηρίζονται από υψηλό ενεργειακό κόστος, με αποτέλεσμα να μην είναι ευνοϊκές (Mondal et al., 2014). Επιπρόσθετα, πολύ συχνά η αλλαγή του πάχους της διπλοστιβάδας δεν οδηγεί σε πλήρη εξομάλυνση της ασυμφωνίας, καθώς πολλά κατάλοιπα των πρωτεϊνών παραμένουν εκτεθειμένα σε μη ευνοϊκό περιβάλλον. Από την πλευρά των πρωτεϊνών, μια προσαρμογή που συναντάται συχνά είναι η αλλαγή της θέσης τους σε σχέση με το επίπεδο της διπλοστιβάδας. Συγκεκριμένα, πολλές πρωτεΐνες έχουν βρεθεί να μεταβάλλουν τη γωνία του κατακόρυφου αξονά τους με το επίπεδο της μεμβράνης, προκειμένου να μεγιστοποιήσουν την ευνοϊκή κάλυψη των διαμεμβρανικών τμημάτων τους από τη διπλοστιβάδα. Η παραπάνω προσαρμογή έχει βρεθεί να είναι αποτελεσματική στις περιπτώσεις πρωτεϊνών με ένα διαμεμβρανικό τμήμα, στις οποίες η αλλαγή της θέσης για τη διαμεμβρανική α-έλικα είναι θερμοδυναμικά ευνοϊκή. Ωστόσο, στην πλειονότητα των διαμεμβρανικών πρωτεϊνών, οι οποίες αποτελούνται από διαμεμβρανικά τμήματα ποικίλου μήκους η μεταβολή της θέσης δεν είναι πάντα αρκετή, καθώς αδυνατεί να εξομαλύνει την ασυμφωνία στο σύνολο των διαμεμβρανικών τμημάτων της πρωτεΐνης. Έτσι, έχει προταθεί ότι ο σχηματισμός αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών μπορεί να λειτουργήσει ευεργετικά για την εξάλειψη της υδροφοβικής ασυμφωνίας, καθώς ο σχηματισμός διαμοριακών αλληλεπιδράσεων οδηγεί στην κάλυψη καταλοίπων τα οποία, στη μονομερή κατάσταση των πρωτεϊνών, θα ήταν εκτεθειμένα σε μη ευνοϊκό περιβάλλον. Η παραπάνω υπόθεση έχει ελεγχθεί για μια σειρά από διαμεμβρανικές πρωτεΐνες, ενώ η ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας έχει προταθεί σαν πιθανή μέθοδος αναγνώρισης αλληλεπιδρώντων καταλοίπων στον ολιγομερισμό διαμεμβρανικών πρωτεϊνών όπως οι GPCRs (Mondal et al., 2011; Mondal et al., 2013).

1.7 Στόχοι της διδακτορικής διατριβής

Η μελέτη της δομής, της λειτουργίας και των διαμοριακών αλληλεπιδράσεων των διαμεμβρανικών πρωτεϊνών αποτελεί ένα **πεδίο έντονου ερευνητικού ενδιαφέροντος**, τόσο λόγω του εξέχοντα ρόλου που διαδραματίζουν αυτές οι πρωτεΐνες στο κύτταρο όσο και λόγω της συσχέτισής τους με ασθένειες και του ρόλου τους σαν στόχοι για το σχεδιασμό φαρμάκων. Ειδικότερα, η μελέτη των **αλληλεπιδράσεων πρωτεϊνών–πρωτεϊνών** στο επίπεδο της μεμβράνης συμβάλλει, σε επίπεδο δομής, στην αποσαφήνιση του μηχανισμού λειτουργίας των διαμεμβρανικών πρωτεϊνών ενώ παράλληλα, σε επίπεδο συστήματος, προσφέρει τη δυνατότητα για ανάλυση των πρωτεϊνικών αλληλεπιδράσεων σε μεγάλη κλίμακα και την αποσαφήνιση του ρόλου των διαμεμβρανικών πρωτεϊνών σε σχέση με το σύνολο του πρωτεώματος στον οργανισμό. Ταυτόχρονα, **οι ιδιότητες των λιπιδίων και οι αλληλεπιδράσεις τους με τις πρωτεΐνες** αποτελούν κρίσιμο κομμάτι του δομικού και του λειτουργικού ρόλου της μεμβράνης στις πρωτεϊνικές αλληλεπιδράσεις το οποίο, παρά τη σημασία του, έχει μελετηθεί ελάχιστα. Με βάση τις παραπάνω παρατηρήσεις, αντικείμενο της παρούσας διδακτορικής διατριβής ήταν η **μελέτη των αλληλεπιδράσεων πρωτεϊνών–πρωτεϊνών σε βιολογικές μεμβράνες και διαμεμβρανικές πρωτεΐνες με τη χρήση υπολογιστικών μεθόδων**.

Η δομή και η δυναμική των διαμεμβρανικών πρωτεϊνών αποτελούν καθοριστικά χαρακτηριστικά των αλληλεπιδράσεών τους στο επίπεδο της μεμβράνης. Ωστόσο, η μελέτη τους με πειραματικές μεθόδους είναι δύσκολη, ενώ πολύ λίγα είναι γνωστά για τις επιρροές της μεμβράνης πάνω στη δομή και τις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών. Κατά συνέπεια, **πρωταρχικός στόχος** της διδακτορικής διατριβής ήταν η **μελέτη των δομικών και δυναμικών χαρακτηριστικών των πρωτεϊνικών αλληλεπιδράσεων** για μια σειρά από διαμεμβρανικές πρωτεΐνες **στο επίπεδο της μεμβράνης**. Συγκεκριμένα, η διδακτορική διατριβή επικεντρώθηκε στη μελέτη των **αλληλεπιδράσεων ολιγομερισμού των GPCRs**, τα διμερή των οποίων έχουν βρεθεί να αυξάνουν την πολυπλοκότητα των σηματοδοτικών μηχανισμών των υποδοχέων και έχουν συσχετιστεί με μεγάλο αριθμό ασθενειών. Παράλληλα, επιχειρήθηκε η δομική μελέτη της **Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων**, καθώς και των **αλληλεπιδράσεων ολιγομερισμού των διαμεμβρανικών β-**

Κεφάλαιο I - Εισαγωγή

βαρελιών, τα οποία διαδραματίζουν σημαντικό ρόλο στη Βιολογία πολλών βακτηρίων. Η μελέτη των παραπάνω αλληλεπιδράσεων πραγματοποιήθηκε μέσα από εκτεταμένη χρήση **προσομοιώσεων Μοριακής Δυναμικής**. Επιπρόσθετα, στα πλαίσια της μελέτης των Εξωτερικών Μεμβρανών κατασκευάστηκε μια μέθοδος για την **αυτοματοποιημένη προτυποποίηση της δομής Εξωτερικών Μεμβρανών των αρνητικών κατά Gram βακτηρίων**. Τέλος, προτυποποιήθηκαν η **τρισδιάστατη δομή και οι αλληλεπιδράσεις του διαμεμβρανικού μεταφορέα δισθενών μεταλλικών ιόντων SLC11A1**, ο οποίος διαδραματίζει σημαντικό ρόλο στη μεταφορά των ιόντων σιδήρου και έχει συσχετιστεί με μεγάλο αριθμό ασθενειών.

Παράλληλα με τη μελέτη των πρωτεϊνικών αλληλεπιδράσεων σε επίπεδο δομής, ένας **επιπρόσθετος στόχος** της διδακτορικής διατριβής ήταν η **μελέτη των υποδοχέων με ενεργότητα τυροσινικής κινάσης (RTKs)**, μιας ομάδας διαμεμβρανικών πρωτεϊνών οι οποίες, παρά τη σημασία τους, έχουν μελετηθεί ελάχιστα σε άλλους οργανισμούς πέρα από τον άνθρωπο και τον ποντικό. Συγκεκριμένα, στα πλαίσια της διδακτορικής διατριβής κατασκευάστηκε **μία μέθοδος για την πρόγνωση και το σχολιασμό πιθανών RTKs** από την αλληλουχία. Σημειώνεται πως, μέχρι στιγμής, **δεν υπάρχει** διαθέσιμη στη βιβλιογραφία κάποια αντίστοιχη μέθοδος για την πρόγνωση RTKs. Με δεδομένη την αύξηση του αριθμού των διαθέσιμων αλληλουχιών πρωτεϊνών, μέσα από τον προσδιορισμό ολόένα και περισσότερων γονιδιωμάτων, η παραπάνω μέθοδος μπορεί να φανεί ιδιαίτερα χρήσιμη στο **σχολιασμό ελάχιστα μελετημένων πρωτεωμάτων**.

Τέλος, πραγματοποιήθηκε η **μελέτη των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου** ο οποίος, παρά τη σημασία του για το ευκαρυωτικό κύτταρο, παραμένει ελάχιστα μελετημένος. Η μελέτη των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου πραγματοποιήθηκε σε επίπεδο συστήματος, με την **κατασκευή και ανάλυση του δικτύου αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών για τις πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου**. Επιπρόσθετα, κατασκευάστηκε μια **βάση δεδομένων για το σύνολο των πρωτεϊνών του πυρηνικού φακέλου στους ευκαρυωτικούς οργανισμούς**, με στόχο την οργάνωση της διαθέσιμης πληροφορίας πάνω στον πυρηνικό φάκελο και την πιο εύχρηστη και αποτελεσματική πρόσβαση σε αυτήν.

ΚΕΦΑΛΑΙΟ II

ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

Σε αυτήν την ενότητα παρουσιάζονται αναλυτικά οι μέθοδοι που χρησιμοποιήθηκαν στην παρούσα διδακτορική εργασία. Οι τρεις πρώτες ενότητες του κεφαλαίου παρουσιάζουν τις αρχές των προσομοιώσεων Μοριακής Δυναμικής (Ενότητα 2.1), των δικτύων βιολογικών αλληλεπιδράσεων (Ενότητα 2.2) και των Κρυφών Μοντέλων Markov (Ενότητα 2.3). Καθώς μεγάλο μέρος της έρευνας που πραγματοποιήθηκε στα πλαίσια της διδακτορικής διατριβής περιελάμβανε τη διεξαγωγή προσομοιώσεων Μοριακής Δυναμικής, δίνεται ιδιαίτερη έμφαση στην παρουσίαση της συγκεκριμένης μεθοδολογίας. Οι υπόλοιπες ενότητες του κεφαλαίου (2.4-2.7) παρουσιάζουν αναλυτικά την εφαρμογή των παραπάνω μεθόδων στη μελέτη των αλληλεπιδράσεων των διαμεμβρανικών πρωτεϊνών, όπως αυτή καθορίστηκε στους στόχους της διδακτορικής διατριβής (βλ. Ενότητα 1.7).

2.1 Προσομοιώσεις Μοριακής Δυναμικής

Οι προσομοιώσεις Μοριακής Δυναμικής (Molecular Dynamics) επιχειρούν την όσο το δυνατόν πιο ρεαλιστική αναπαράσταση του περιβάλλοντος των βιολογικών μακρομορίων σε μικροσκοπικό επίπεδο. Αναφέρονται στη μελέτη της δομής και δυναμικής των πρωτεϊνών, νουκλεϊκών οξέων και άλλων βιολογικών μορίων χρησιμοποιώντας θεωρητικές τεχνικές. Αντικείμενο της Μοριακής Δυναμικής είναι ουσιαστικά η προσομοίωση της κίνησης ενός συστήματος πολλών σωματιδίων για τη μελέτη της δυναμικής συμπεριφοράς του στη διάρκεια ενός ορισμένου χρονικού διαστήματος. Αυτό περιλαμβάνει τον υπολογισμό και την εφαρμογή

Κεφάλαιο II – Υλικά & Μέθοδοι

κατάλληλων δυνάμεων και τη μελέτη του συστήματος στη μονάδα του χρόνου (Karplus & McCammon, 2002).

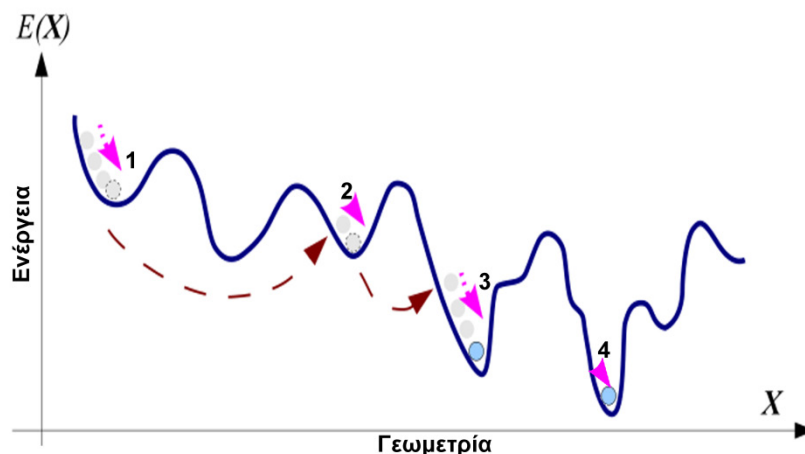
Οι προσομοιώσεις Μοριακής Δυναμικής βασίζονται στους θεμελιώδεις νόμους της στατιστικής και της κλασικής μηχανικής. Με τις εξισώσεις κίνησης των Νόμων του Νεύτωνα, οι οποίες εφαρμόζονται στην κλασική μηχανική επιτυγχάνεται η προσομοίωση μιας πορείας κινήσεων που περιγράφει τις θέσεις, τις ταχύτητες και τις επιταχύνσεις τμημάτων του βιολογικού συστήματος σε μικροσκοπικό επίπεδο. Από την άλλη, με την στατιστική μηχανική μελετώνται, μέσω μαθηματικών εκφράσεων, οι μακροσκοπικές ιδιότητες του συστήματος όπως η πίεση, η ενέργεια, η θερμότητα κλπ. Μια τυπική προσομοίωση Μοριακής Δυναμικής συνδυάζει τα χαρακτηριστικά των παραπάνω για την επίλυση του προβλήματος και περιλαμβάνει πάντα μια σειρά από συγκεκριμένα βήματα (Wilde & Singh, 1998). Αρχικά πραγματοποιείται ελαχιστοποίηση ενέργειας του συστήματος. Ακολούθως, για κάθε άτομο του προσομοιούμενου συστήματος ανατίθεται μια τιμή ταχύτητας. Από την κατανομή των ταχυτήτων μπορεί να υπολογιστεί η θερμοκρασία και η κινητική ενέργεια του συστήματος, ενώ η δυναμική ενέργεια υπολογίζεται από τις θέσεις των σωματιδίων. Οι ενεργειακές τιμές χρησιμοποιούνται για τον υπολογισμό των αντίστοιχων δυνάμεων, μέσω υπολογισμού των παραγώγων των ενεργειών. Με βάση την ενέργεια, την ταχύτητα και τη θέση εφαρμόζονται οι εξισώσεις του Νεύτωνα για υπολογισμό της κίνησης και πρόγνωση της νέας θέσης των ατόμων και, με τη μετάβασή τους στη νέα θέση, υπολογίζονται εκ νέου οι ενέργειες και η διαδικασία επαναλαμβάνεται για το σύνολο του χρόνου της προσομοίωσης (Leach, 2001).

2.1.1 Μέθοδοι Ελαχιστοποίησης Ενέργειας

Το πρώτο και βασικό βήμα σε κάθε προσομοίωση βιομορίων είναι η ελαχιστοποίηση της ενέργειας του συστήματος, δηλαδή η διαδικασία βελτιστοποίησης της γεωμετρίας και στερεοχημικής ποιότητας των μορίων ώστε αυτά να αποκτήσουν την βέλτιστη δυνατή στερεοδιάταξή τους, που θεωρείται πως είναι και αυτή με την ελάχιστη δυνατή ενέργεια. Η δυναμική ενέργεια ενός μακρομοριακού συστήματος είναι ένα ιδιαίτερα πολύπλοκο πεδίο, το οποίο υλοποιείται σε μεγάλο αριθμό διαστάσεων. Διαθέτει ένα ολικό ελάχιστο, το οποίο

Κεφάλαιο II – Υλικά & Μέθοδοι

αντιστοιχεί στη φυσική στερεοδιάταξη, καθώς επίσης και μεγάλο αριθμό τοπικών ελαχίστων, όπου όλες οι πρώτες παράγωγοι της δυναμικής ενέργειας είναι μηδενικές, ενώ οι δεύτεροι παράγωγοι είναι μη αρνητικές. Η γνώση όλων των πιθανών τοπικών ελαχίστων, όπως και του ολικού ελαχίστου, καθώς και όλων των δυνατών μονοπατιών ανάμεσά τους θα επέτρεπε την περιγραφή όλων των σχετικών δομών, των στερεοδιατάξεων και των ενεργειών τους, καθώς επίσης και της δυναμικής των δομικών μεταβάσεων ανάμεσά τους.



Εικόνα 25 Η πορεία της ελαχιστοποίησης ενέργειας. Δισδιάστατη διαγραμματική απεικόνιση της ενέργειας του συστήματος (E) σαν συνάρτηση της γεωμετρίας του (X). Αντικείμενο της ελαχιστοποίησης ενέργειας είναι η μετάβαση από μια αρχική κατάσταση υψηλής ενέργειας (Κατάσταση 1) στην κατάσταση χαμηλής ενέργειας η οποία αντιστοιχεί στο ολικό ενεργειακό ελάχιστο του συστήματος (Κατάσταση 4). Αυτή η μετάβαση πραγματοποιείται σταδιακά, μέσα από ενδιάμεσα βήματα (Καταστάσεις 2 και 3). Ωστόσο, το ολικό ελάχιστο του συστήματος είναι συχνά αδύνατο να προσδιοριστεί, λόγω της πολυπλοκότητας στην επιφάνεια της ελεύθερης ενέργειας του συστήματος. Έτσι, πολλές φορές προσδιορίζεται το καλύτερο δυνατό τοπικό ελάχιστο (Κατάσταση 3), το οποίο συνδυάζει τη βέλτιστη δυνατή στερεοχημική ποιότητα με την ελάχιστη δυνατή δυναμική ενέργεια.

Πρακτικά αυτή η γνώση είναι αδύνατη, λόγω της υψηλής πολυπλοκότητας του ενεργειακού πεδίου. Κατά συνέπεια, δεν υπάρχει καμία μέθοδος ελαχιστοποίησης ενέργειας που να εγγυάται το σίγουρο προσδιορισμό του ολικού ενεργειακού ελαχίστου σε οποιοδήποτε χρονικό διάστημα υπολογισμών. Ωστόσο, με δεδομένη μια αρχική κατάσταση, είναι δυνατό να βρεθεί το καλύτερο δυνατό ελάχιστο, δηλαδή η στερεοδιάταξη η οποία θα συνδυάζει την καλύτερη δυνατή στερεοχημική ποιότητα με την ελάχιστη δυνατή δυναμική ενέργεια. (Leach, 2001).

2.1.2 Ο βασικός αλγόριθμος των προσομοιώσεων Μοριακής Δυναμικής

Ο πιο απλός τρόπος εξήγησης του αλγορίθμου των προσομοιώσεων Μοριακής Δυναμικής είναι η επίλυσή του σε ένα ιδεατό, απομονωμένο σύστημα N σωματιδίων, κατανεμημένων σε ένα καθορισμένο τμήμα του χώρου με σταθερό όγκο V . Κάθε σωματίδιο i αυτού του συστήματος χαρακτηρίζεται από δύο μεγέθη, τη θέση του στο χώρο (r_i) και την ορμή (p_i). Καθώς η θέση και η ορμή κάθε σωματιδίου αλλάζουν στη διάρκεια του χρόνου της προσομοίωσης, το σύστημα στο σύνολό του μεταβάλλεται στο χώρο (r^N, p^N) και στη μονάδα του χρόνου. Ο συνολικός αριθμός των βαθμών ελευθερίας του συστήματος, δηλαδή των κινήσεων που πρέπει να ληφθούν υπόψη στη μελέτη του, είναι $6N-3$. Εφόσον αυτό το σύστημα είναι απομονωμένο, σύμφωνα με τον πρώτο νόμο της Θερμοδυναμικής, η ενέργειά του (E) παραμένει σταθερή. Αυτή η ενέργεια μπορεί να απεικονισθεί με τη χρήση μιας Χαμιλτονιανής συνάρτησης (H)

$$H(r^N, p^N) = E \quad (1)$$

όπου N είναι ο αριθμός των σωματιδίων, E είναι η συνολική ενέργεια και οι τιμές $r^N = (r_1, r_2, \dots, r_N)$ και $p^N = (p_1, p_2, \dots, p_N)$ αντιπροσωπεύουν το σύνολο των θέσεων και των ορμών των σωματιδίων. Η Χαμιλτονιανή συνάρτηση ουσιαστικά περιγράφει την ενέργεια σαν συνάρτηση των r^N και p^N και μπορεί να αναλυθεί στους όρους της κινητικής και της δυναμικής ενέργειας, όπως ορίζεται στην Εξίσωση 2

$$H(r^N, p^N) = K(p^N) + U(r^N) \quad (2)$$

όπου οι όροι K και U αντιπροσωπεύουν την κινητική και τη δυναμική ενέργεια, αντίστοιχα (Wilde & Singh, 1998).

Καθώς η ενέργεια παραμένει σταθερή, η παράγωγος της ως προς το χρόνο μηδενίζεται, όπως φαίνεται στην Εξίσωση 3

$$H(r^N, p^N) = E \Rightarrow \frac{dH(r^N, p^N)}{dt} = 0 \quad (3)$$

ενώ οι μερικές παράγωγοι των όρων ως προς τις μεταβλητές r^N και p^N ορίζονται σύμφωνα με την Εξίσωση 4

$$\frac{dH(r^N, p^N)}{dt} = 0 \Rightarrow \sum_{i=1}^N \partial \frac{K(p^N)}{\partial p_i} \cdot \dot{p}_i + \sum_{i=1}^N \frac{\partial U(r^N)}{\partial r_i} \cdot \dot{r}_i = 0 \quad (4)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου οι όροι $\dot{r}_i$ και $\dot{p}_i$ συμβολίζουν τις παραγώγους των μεγεθών r και p ως προς το χρόνο, με βάση τη σημειογραφία του Νεύτωνα, για κάθε σωματίδιο i . Σημειώνεται ότι η παράγωγος της θέσης προς το χρόνο, $\dot{r}_i$, είναι ουσιαστικά η ταχύτητα του σωματιδίου u_i ($u_i = dr_i/dt$).

Ο όρος της κινητικής ενέργειας μπορεί να αναπτυχθεί σαν συνάρτηση της ορμής (p) σύμφωνα με την Εξίσωση 5

$$K(p^N) = \sum_{i=1}^N \frac{p_i^2}{2m_i} \quad (5)$$

όπου m_i και p_i είναι η μάζα και η ορμή του σωματιδίου i , αντίστοιχα. Αντικαθιστώντας την κινητική ενέργεια στην Εξίσωση 2 με την Εξίσωση 5, η Χαμιλτονιανή συνάρτηση ορίζεται ως

$$H(r^N, p^N) = \sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r^N) \quad (6)$$

με την παράγωγό της να γίνεται

$$\sum_{i=1}^N \frac{p_i}{m_i} \cdot \dot{p}_i + \sum_{i=1}^N \frac{\partial U(r^N)}{\partial r_i} \cdot \dot{r}_i = 0 \quad (7)$$

Έτσι, καταλήγουμε στην ισότητα της Εξίσωσης 8:

$$\frac{-\partial U(r^N)}{\partial r_i} \cdot \dot{r}_i = \frac{p_i}{m_i} \cdot \dot{p}_i \quad (8)$$

Η ορμή κάθε σωματιδίου, p_i , ορίζεται ως το γινόμενο της μάζας του σωματιδίου επί την ταχύτητά του, σύμφωνα με την Εξίσωση 9

$$p_i = m_i \cdot u_i = m_i \cdot \dot{r}_i \quad (9)$$

Έτσι, εφαρμόζοντας την Εξίσωση 9 στην Εξίσωση 8, καταλήγουμε στην σχέση της Εξίσωσης 10:

$$\frac{-\partial U(r^N)}{\partial r_i} = \dot{p}_i, \forall i \in \{1, 2, 3, \dots, N\} \quad (10)$$

Η παραπάνω εξίσωση είναι γνωστή και ως Χαμιλτονιανή εξίσωση της κίνησης (Wilde & Singh, 1998). Σημειώνεται πως, με βάση την εξίσωση 9 για την ορμή, η παράγωγος της ορμής $\dot{p}_i$ μπορεί να οριστεί σύμφωνα με την Εξίσωση 12

$$\dot{p}_i = m_i \cdot \ddot{r}_i = m_i \cdot \frac{d^2 r_i}{dt^2} = m_i \cdot \alpha_i = F_i \quad (11)$$

όπου α_i είναι η επιτάχυνση και F_i είναι η εφαρμοζόμενη δύναμη. Έτσι, η Χαμιλτονιανή εξίσωση της κίνησης μπορεί να οριστεί με βάση τον δεύτερο Νόμο του Νεύτωνα, και οδηγούμαστε στην τελική της μορφή της εξίσωσης κίνησης, σύμφωνα με την Εξίσωση 12

$$F_i = \frac{-\partial U(r^N)}{\partial r_i} \quad (12)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

Η επίλυση του αριθμημένου ολοκληρώματος της παραπάνω εξίσωσης σε ένα δεδομένο στιγμιότυπο t θα δώσει σαν αποτέλεσμα τις δυνάμεις που ασκούνται σε κάθε σωματίδιο i του συστήματος, στη δεδομένη θέση r_i στην οποία βρίσκεται στο συγκεκριμένο στιγμιότυπο. Έπειτα υπολογίζεται το αριθμημένο ολοκλήρωμα της συνάρτησης ως προς το χρόνο με βάση ένα βήμα χρόνου dt και καθορίζονται οι νέες θέσεις και ταχύτητες των ατόμων στο χρόνο $t+dt$, μέσω των οποίων θα υπολογιστούν εκ νέου η κινητική ενέργεια και οι εφαρμοζόμενες δυνάμεις του συστήματος. Η επανάληψη της παραπάνω διαδικασίας για ένα οριζόμενο χρονικό διάστημα ουσιαστικά αποτελεί την προσομοίωση Μοριακής Δυναμικής του συστήματος (Wilde & Singh, 1998).

2.1.3 Υλοποίηση του αλγορίθμου των προσομοιώσεων Μοριακής Δυναμικής στη μονάδα του χρόνου

Όπως αναφέρθηκε ήδη, ο πυρήνας των προσομοιώσεων Μοριακής Δυναμικής είναι η επαναλαμβανόμενη επίλυση του ολοκληρώματος των εξισώσεων κίνησης στη μονάδα του χρόνου, για ένα ορισμένο χρονικό διάστημα προσομοίωσης. Αυτή η ακολουθία υπολογισμών υλοποιείται με την εφαρμογή μιας μεθόδου χρονικής ολοκλήρωσης (time integration). Οι μέθοδοι χρονικής ολοκλήρωσης λειτουργούν διαιρώντας το συνολικό χρόνο της προσομοίωσης σε διακριτά υποσύνολα που χωρίζονται μεταξύ τους με ένα βήμα χρόνου (dt), οδηγώντας έτσι στην υλοποίηση ενός πεπερασμένου πλέγματος που περιγράφει την προσομοίωση. Κάθε σημείο του πλέγματος αντιστοιχεί σε ένα συγκεκριμένο στιγμιότυπο t , για το οποίο οι θέσεις των ατόμων r , καθώς και οι παράγωγοί τους (ταχύτητα, επιτάχυνση κλπ), είναι άμεσα γνωστές ή μπορούν να υπολογιστούν αριθμητικά. Καθώς η προσομοίωση προχωρά στο επόμενο σημείο του πλέγματος, το οποίο απέχει από το προηγούμενο κατά dt (οπότε αντιστοιχεί στο στιγμιότυπο $t+dt$), οι προηγούμενες τιμές των θέσεων χρησιμοποιούνται για να υπολογίσουν τις νέες τιμές, χρησιμοποιώντας επεκτάσεις παρόμοιες με τις μαθηματικές σειρές Taylor:

Κεφάλαιο II – Υλικά & Μέθοδοι

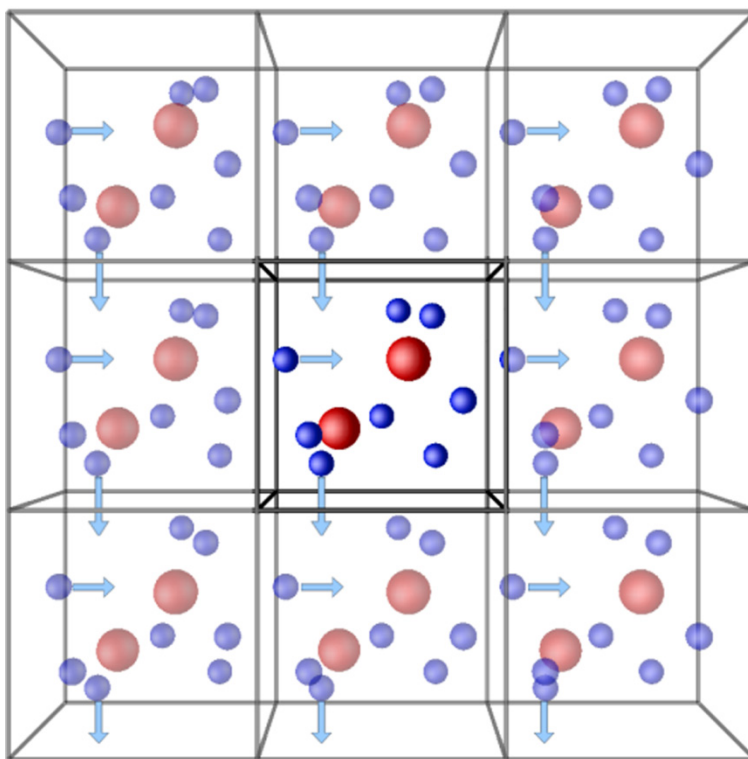
$$\begin{aligned}r(t + dt) &= r(t) + dtu(t) + \frac{1}{2}dt^2\alpha(t) + \frac{1}{6}dt^3b(t) + \frac{1}{24}dt^4c(t) + \dots \Rightarrow \\u(t + dt) &= u(t) + dt\alpha(t) + \frac{1}{2}dt^2b(t) + \frac{1}{6}dt^3c(t) + \dots \Rightarrow \\ \alpha(t + dt) &= \alpha(t) + dtb(t) + \frac{1}{2}dt^2c(t) + \dots\end{aligned}\quad (13)$$

όπου r είναι η θέση, u είναι η πρώτη παράγωγος της θέσης (ταχύτητα), α είναι η δεύτερη παράγωγος (επιτάχυνση), b είναι η τρίτη παράγωγος κλπ (Hockney et al., 1974).

Βασικό χαρακτηριστικό της υλοποίησης των προσομοιώσεων μέσω του παραπάνω αλγορίθμου είναι η τιμή του χρόνου βήματος (dt) ανάμεσα στα χρονικά στιγμιότυπα του πλέγματος, το οποίο ουσιαστικά συμβάλλει σε μεγάλο βαθμό στην ταχύτητα της προσομοίωσης. Καθώς ο αλγόριθμος των προσομοιώσεων λειτουργεί μέσω του υπολογισμού δυνάμεων για κάθε σωματίδιο σε διακριτά στιγμιότυπα, είναι απαραίτητο η τιμή του dt να είναι αρκετά μικρή, ώστε το ολοκλήρωμα της εξίσωσης κίνησης να επιλυθεί σωστά για όλα τα σωματίδια και τις αλληλεπιδράσεις τους. Χρήση μιας πολύ μεγάλης τιμής dt οδηγεί στον υπολογισμό μη ρεαλιστικών δυνάμεων, με αποτέλεσμα τη θερμοδυναμική και κινητική αστάθεια του συστήματος και την αποτυχία της προσομοίωσης. Ωστόσο, η χρήση μικρών τιμών dt έχει σαν μειονέκτημα τον περιορισμό της ταχύτητας των προσομοιώσεων και, κατά συνέπεια, του χρόνου που μπορεί να προσομοιωθεί.

2.1.4 Συνθήκες περιοδικών ορίων

Η πλειονότητα των εφαρμογών της Μοριακής Δυναμικής, ειδικά στην περίπτωση των βιομορίων, περιλαμβάνει τη μελέτη ενός μακρομοριακού συστήματος (π.χ. μιας πρωτεΐνης) στο φυσικό του περιβάλλον. Στην περίπτωση των σφαιρικών πρωτεϊνών αυτό το περιβάλλον είναι ο υδατικός διαλύτης, ενώ για τις μεμβρανικές πρωτεΐνες είναι το ετερογενές περιβάλλον της λιπιδικής διπλοστιβάδας με το διαλύτη εκατέρωθεν της. Αυτές οι συνθήκες προτυποποιούνται ουσιαστικά σαν ένα κελί ή κυψελίδα με ορισμένες διαστάσεις στο χώρο. Ωστόσο, αυτή η προσέγγιση οδηγεί σε μη ρεαλιστικές συνθήκες στα όρια του συστήματος προσομοίωσης. Επιπρόσθετα, η πιθανή μετακίνηση του μελετώμενου μακρομορίου από τα όρια του συστήματος κατά την προσομοίωση θα οδηγήσει σε αλληλεπιδράσεις με τα «τοιχώματα» του συστήματος.



Εικόνα 26 Συνθήκες περιοδικών ορίων. Κατά την εφαρμογή των συνθηκών περιοδικών ορίων θεωρείται ότι το σύστημα αποτελείται από αντίγραφα του πρωταρχικού κελιού προς όλες τις κατευθύνσεις.

Για την επίλυση αυτών των προβλημάτων εφαρμόζονται κατά την προσομοίωση μια σειρά από χωρικούς περιορισμούς, γνωστοί σαν «συνθήκες περιοδικών ορίων» (periodic boundary conditions). Σε αυτή τη μέθοδο, το υπό μελέτη μόριο τοποθετείται σε ένα συγκεκριμένο όγκο, ο οποίος αποτελεί το πρωταρχικό κελί. Για να προσομοιωθεί ο τεράστιος όγκος του περιβάλλοντος, θεωρείται ότι το σύστημα αποτελείται από αντίγραφα του πρωταρχικού κελιού προς όλες τις κατευθύνσεις. Ουσιαστικά, το σύστημα αποτελείται από περιοδικές επαναλήψεις του πρωταρχικού κελιού που ως στόχο έχουν τη δημιουργία ενός μακροσκοπικού δείγματος και την περιοδική έκφραση των θέσεων και των ορμών των ατόμων (Εικόνα 26). Τα κελιά χωρίζονται μεταξύ τους από νοητά σύνορα από τα οποία μπορούν ελεύθερα να περάσουν τα άτομα σε ένα διπλανό κελί. Όταν ένα άτομο ή ένα μέρος του μορίου περάσουν τα σύνορα από το πρωταρχικό κελί, άμεσα εισέρχεται το αντίστοιχο τμήμα από το αντιπαράλληλο κελί.

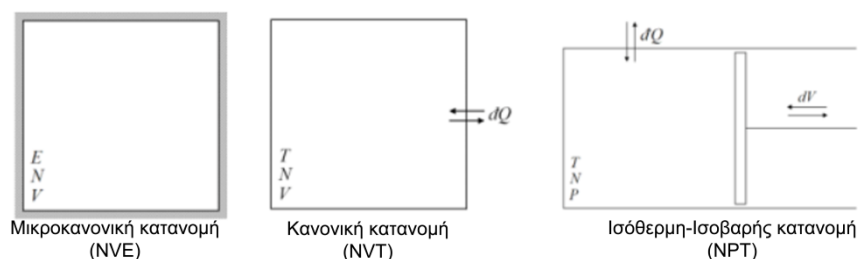
2.1.5 Προσομοίωση θερμοδυναμικών κατανομών

Στις προσομοιώσεις Μοριακής Δυναμικής ένα σύστημα προσομοίωσης καθορίζεται από συγκεκριμένες φυσικές και μηχανικές παραμέτρους. Η μηχανική κατάσταση ενός συστήματος N σωματιδίων χαρακτηρίζεται από τη θέση της και την ορμή (p) καθενός από τα σωματίδια που απαρτίζουν αυτό το σύστημα. Τα ζεύγη (r,p) μπορούν να θεωρηθούν ως οι συντεταγμένες ενός πολυδιάστατου χώρου φάσεων (phase space), στον οποίο κάθε μοναδικό σημείο αντιπροσωπεύει μια πιθανή μικροσκοπική κατάσταση (microstate). Οι φυσικές παράμετροι που χαρακτηρίζουν αυτές τις μικροσκοπικές καταστάσεις περιλαμβάνουν χαρακτηριστικά όπως ο όγκος (V), η πίεση (P) η θερμοκρασία (T) και η συνολική ενέργεια του συστήματος (E). Τα σημεία ή μικροκαταστάσεις του χώρου χαρακτηρίζονται από ένα κοινό συνδυασμό αυτών των συνθηκών και αντιπροσωπεύουν, συνολικά, μια **θερμοδυναμική κατανομή** ή **θερμοδυναμικό σύνολο** (thermodynamic ensemble), που ουσιαστικά περιγράφει μακροσκοπικά τις συνθήκες διεξαγωγής της προσομοίωσης (Εικόνα 27). Οι συνδυασμοί των παραπάνω παραμέτρων οδηγούν σε διαφορετικές συνθήκες προσομοιώσεων οι οποίες προσομοιάζουν τις αντίστοιχες θερμοδυναμικές μεταβολές (Wilde & Singh, 1998).

Ο βασικός αλγόριθμος των προσομοιώσεων Μοριακής Δυναμικής, όπως περιγράφηκε, προϋποθέτει την ύπαρξη ενός συστήματος στο οποίο ο αριθμός των σωματιδίων (N), ο όγκος (V) και η συνολική ενέργεια E δεν μεταβάλλονται. Η θερμοδυναμική κατάσταση αυτού του συστήματος είναι γνωστή σαν **μικροκανονική κατανομή** (Microcanonical Ensemble) ή **συνθήκες NVE**. Ένα σύστημα αυτού του τύπου προσομοιάζει τις συνθήκες μιας ισόχωρης-αδιαβατικής αντίδρασης και, κατά συνέπεια, μπορεί να απεικονίσει το φαινόμενο ανταλλαγής της κινητικής και δυναμικής ενέργειας, με τη συνολική ενέργεια του συστήματος να παραμένει σταθερή. Σε μια προσομοίωση NVE, χαρακτηριστικά όπως η πίεση (P) και η θερμοκρασία (T) μεταβάλλονται ενεργά, ώσπου να επιτευχθεί η σύγκλισή τους σε μία κατάσταση ισορροπίας, κατά την οποία θα αποκτήσουν τις τελικές τιμές τους. Ωστόσο, η εκ των προτέρων πρόβλεψη ή εκτίμηση αυτών των τιμών είναι πολύ δύσκολο να πραγματοποιηθεί. Επιπρόσθετα, η μελέτη των βιομοριακών αλληλεπιδράσεων συχνά προϋποθέτει την εφαρμογή συγκεκριμένων περιβαλλοντικών συνθηκών, οι οποίες δεν μπορούν να ελεγχθούν στην μικροκανονική κατανομή. Έτσι, οι προσομοιώσεις Μοριακής Δυναμικής συνήθως υιοθετούν διαφορετικές

Κεφάλαιο II – Υλικά & Μέθοδοι

θερμοδυναμικές κατανομές, οι οποίες περιλαμβάνουν τον έλεγχο της θερμοκρασίας, της πίεσης και του όγκου σε διαφορετικούς συνδυασμούς. Οι δυο πιο συχνά χρησιμοποιούμενες θερμοδυναμικές καταστάσεις είναι η ισόχωρη – ισόθερμη και η ισόθερμη-ισοβαρής κατανομή. Η **ισόθερμη-ισόχωρη κατανομή**, γνωστή και σαν **κανονική κατανομή** (Canonical Ensemble) ή **συνθήκες NVT**, περιλαμβάνει έλεγχο και σταθεροποίηση του αριθμού των σωματιδίων (N), του όγκου (V) και της θερμοκρασίας (T) του συστήματος, ενώ η πίεση αφήνεται ελεύθερη να μεταβληθεί. Η **ισόθερμη-ισοβαρής κατανομή (συνθήκες NPT)**, από την άλλη, διατηρεί σταθερές τιμές για τον αριθμό των σωματιδίων (N), την πίεση (P) και τη θερμοκρασία (T) του συστήματος, ενώ ο όγκος αφήνεται ελεύθερος να μεταβληθεί (Εικόνα 27).



Εικόνα 27 Διαγραμματική απεικόνιση των θερμοδυναμικών κατανομών. Η Μικροκανονική κατανομή (NVE) προτυποποιεί ένα κλειστό, απομονωμένο σύστημα χωρίς εξωτερικές επιρροές. Στην Κανονική Κατανομή (NVT) το σύστημα είναι συζευγμένο με μια εξωτερική πηγή θερμότητας για τον έλεγχο της θερμοκρασίας, ενώ στην Ισόθερμη-Ισοβαρή κατανομή (NPT) εφαρμόζεται επιπρόσθετα ένας βαροστάτης για τον έλεγχο της πίεσης.

2.1.5.1 Αλγόριθμοι Ελέγχου της Θερμοκρασίας

Βασικό στοιχείο τόσο στην Κανονική (NVT) όσο και στην ισόθερμη-ισοβαρή (NPT) θερμοδυναμική κατανομή είναι ο αποτελεσματικός έλεγχος της θερμοκρασίας (T) του συστήματος. Η θερμοκρασία συνδέεται άμεσα με την κινητική ενέργεια και την ταχύτητα, όπως φαίνεται από την Εξίσωση 14

$$K = \frac{1}{2} m u^2 = \frac{3}{2} N k_B T \quad (14)$$

όπου m είναι η μάζα, u η ταχύτητα, N ο αριθμός των σωματιδίων και k_B η σταθερά Boltzmann. Κατά συνέπεια, η ορθή ρύθμιση της θερμοκρασίας είναι απαραίτητη για τον υπολογισμό της

Κεφάλαιο II – Υλικά & Μέθοδοι

σωστής κατανομής της κινητικής ενέργειας του συστήματος και του σωστού υπολογισμού της συνολικής ενέργειας του συστήματος. Στην πραγματική κατανομή NVT (και, αντίστοιχα, στην κατανομή NPT) οι τιμές της ταχύτητας και, κατά συνέπεια, της κινητικής ενέργειας και της θερμοκρασίας ακολουθούν την κατανομή Maxwell-Boltzmann, σύμφωνα με την κινητική θεωρία των αερίων. Για το σωστό έλεγχο της θερμοκρασίας, λοιπόν, έχουν σχεδιαστεί μια σειρά από αλγορίθμους και μεθόδους που επιχειρούν να ελέγξουν σωστά τη θερμοκρασία κατά την προσομοίωση και να επιτύχουν ισόθερμες συνθήκες, είτε σε συνθήκες NVT είτε σε συνθήκες NPT.

Παρότι οι λεπτομέρειες στην υλοποίηση κάθε προσέγγισης μπορεί να διαφέρουν, όλες οι μέθοδοι ελέγχου της θερμοκρασίας λειτουργούν προσομοιώνοντας την ύπαρξη ενός θερμοστάτη, δηλαδή μιας συσκευής που πραγματοποιεί σύζευξη (coupling) του συστήματος προσομοίωσης σε μια εξωτερική πηγή θερμότητας, γνωστή σαν θερμικό ρεζερβουάρ (thermal reservoir) ή λουτήρας θερμότητας (heat bath), μιμούμενες με άλλα λόγια τις αντίστοιχες μεθόδους που εφαρμόζονται και πειραματικά για τον έλεγχο της θερμοκρασίας του συστήματος (Εικόνα 27). Η κάθε μέθοδος υλοποιεί αυτή τη διαδικασία διαφορετικά, εφαρμόζοντας συγκεκριμένες αρχές της θερμοδυναμικής και της κλασικής μηχανικής. Ωστόσο, ανάλογα με την προσέγγισή τους, οι περισσότερες μέθοδοι ελέγχου της θερμοκρασίας μπορούν να ταξινομηθούν σε δύο μεγάλες κατηγορίες:

- στις άμεσες μεθόδους ελέγχου: οι συγκεκριμένες μέθοδοι επιχειρούν την προσομοίωση ισόθερμης αντίδρασης παρεμβαίνοντας άμεσα στον ορισμό της Χαμιλτονιανής (H) του συστήματος με την εισαγωγή ενός ενεργειακού όρου. Χαρακτηρίζονται ως άμεσες, καθώς οι τροποποιήσεις τους επηρεάζουν απευθείας τον υπολογισμό της ενέργειας του συστήματος.
- στις έμμεσες μεθόδους ελέγχου: οι συγκεκριμένες μέθοδοι δεν επηρεάζουν τον ίδιο τον υπολογισμό της κινητικής ενέργειας. Αντίθετα, επιδρούν έμμεσα στο σύστημα, επηρεάζοντας τις τιμές των αποκλίσεων της θερμοκρασίας και της κινητικής ενέργειας με την εφαρμογή διορθωτικών παραγόντων.

Ανεξάρτητα από τον τρόπο δράσης τους, το αποτέλεσμα κάθε μεθόδου είναι μια τροποποιημένη μορφή της εξίσωσης κίνησης της Μοριακής Δυναμικής. Ανάλογα με τις

Κεφάλαιο II – Υλικά & Μέθοδοι

ανάγκες της κάθε προσομοίωσης (μέγεθος συστήματος, επίπεδο διακριτικότητας, χρόνος προσομοίωσης κλπ) κάποιες μέθοδοι μπορεί να είναι αποτελεσματικότερες από άλλες. Για το λόγο αυτό, πολύ συχνά στις προσομοιώσεις Μοριακής Δυναμικής δεν χρησιμοποιείται μόνο μία μέθοδος αλλά πολλές, ανάλογα με το στάδιο στο οποίο βρίσκεται κάθε φορά το σύστημα.

Στα πλαίσια της παρούσας διδακτορικής διατριβής έχουν χρησιμοποιηθεί διάφοροι αλγόριθμοι ελέγχου της θερμοκρασίας, τα βασικά χαρακτηριστικά των οποίων παρουσιάζονται παρακάτω. Συγκεκριμένα, θα παρουσιαστούν οι αλγόριθμοι του Berendsen (Berendsen et al., 1984) και της επανακλιμάκωσης των ταχυτήτων (Bussi et al., 2007), οι οποίοι ανήκουν στις έμμεσες μεθόδους ελέγχου, καθώς και ο αλγόριθμος Nosé-Hoover (Hoover, 1985; Nosé, 1984), ο οποίος ανήκει στις άμεσες μεθόδους. Επιπρόσθετα, θα παρουσιαστεί εν συντομία και η Δυναμική Langevin (Brünger, 1992), η οποία δεν εμπίπτει σε καμία από τις παραπάνω διαδικασίες, αλλά αποτελεί μια ξεχωριστή εναλλακτική προσέγγιση που συνδυάζει τη Μοριακή Δυναμική με τη Στοχαστική Φυσική. Τα βασικά χαρακτηριστικά των αλγορίθμων ισχύουν τόσο για την κανονική όσο και για την ισόθερμη-ισοβαρή κατανομή. Ωστόσο, για την πιο εύκολη κατανόησή τους, οι αλγόριθμοι παρουσιάζονται έχοντας ως γνώμονα την κανονική κατανομή.

Ο θερμοστάτης Berendsen: Ο θερμοστάτης του Berendsen (Berendsen thermostat) ελέγχει τη θερμοκρασία του συστήματος, προσομοιώνοντας την ασθενή σύζευξή του σε ένα υποθετικό εξωτερικό λουτήρα θερμότητας. Ο αλγόριθμος του Berendsen ελέγχει τη θερμοκρασία έμμεσα, μειώνοντας την απόκλισή της από τη θερμοκρασία αναφοράς T_0 εκθετικά, ώστε το σύστημα σταδιακά στη μονάδα του χρόνου να αποκτά και να διατηρεί θερμοκρασία κοντά στην τιμή T_0 . Αυτό ουσιαστικά οδηγεί σε καταστολή των διακυμάνσεων της κινητικής ενέργειας του συστήματος (Berendsen et al., 1984). Αλγοριθμικά η μέθοδος λειτουργεί μέσω της κλιμάκωσης της ταχύτητας καθενός από τα σωματίδια του συστήματος, εφαρμόζοντας έναν παράγοντα λ της μορφής

$$\lambda = \left[1 + \frac{\Delta T}{\tau_T} \cdot \left(\frac{T_0}{T} - 1 \right) \right]^{1/2} \quad (15)$$

όπου T είναι η θερμοκρασία του συστήματος, T_0 η θερμοκρασία αναφοράς και τ_T είναι μια χρονική σταθερά ανάλογη της περιόδου σύζευξης του συστήματος στο θερμοστάτη. Η

Κεφάλαιο II – Υλικά & Μέθοδοι

τροποποίηση της ταχύτητας μέσω του λ και η εφαρμογή της στο δεύτερο νόμο του Νεύτωνα οδηγεί σε μια τροποποιημένη εξίσωση κίνησης:

$$m_i \cdot \alpha_i = F_i - \frac{1}{2} \cdot \tau_T^{-1} \cdot \left(\frac{T_0}{T} - 1 \right) \cdot u_i \quad (16)$$

Καθώς ο θερμοστάτης του Berendsen λειτουργεί έμμεσα καταστέλλοντας τις μεταβολές της θερμοκρασίας, στην πράξη δεν επιτυγχάνονται πραγματικές συνθήκες NVT κατά την προσομοίωση, οδηγώντας έτσι σε σφάλμα στον υπολογισμό της κατανομής της κινητικής ενέργειας του συστήματος και της θερμοχωρητικότητας. Η τιμή του σφάλματος είναι αντιστρόφως ανάλογη του αριθμού των σωματιδίων ($1/N$), με αποτέλεσμα το φαινόμενο να είναι ιδιαίτερα έντονο για μικρά σε μέγεθος συστήματα. Όσο αυξάνεται ο αριθμός των σωματιδίων η τιμή του σφάλματος μειώνεται, όμως η παρουσία του μπορεί μακροπρόθεσμα να οδηγήσει σε λανθασμένα αποτελέσματα. Ωστόσο, ο αλγόριθμος έχει το πλεονέκτημα της εύκολης σύζευξης του συστήματος στο θερμοστάτη, μέσω της απλής ρύθμισης της τιμής τ_T . Αυτό έχει σαν αποτέλεσμα τη γρήγορη και υπολογιστικά ευέλικτη ρύθμιση της θερμοκρασίας του συστήματος. Έτσι, πρακτικά, στις προσομοιώσεις Μοριακής Δυναμικής συχνά ο θερμοστάτης του Berendsen χρησιμοποιείται κατά τα αρχικά στάδια εξισορρόπησης του συστήματος λόγω της υπολογιστικής ευελιξίας του, ενώ για στο στάδιο παραγωγής εφαρμόζονται μέθοδοι που μπορούν να επιτύχουν σωστή κατανομή για την κινητική ενέργεια.

Η μέθοδος επανακλιμάκωσης των ταχυτήτων: η μέθοδος επανακλιμάκωσης των ταχυτήτων (velocity rescaling, v-rescale) είναι μια τροποποιημένη εκδοχή του θερμοστάτη του Berendsen. Η μέθοδος λειτουργεί όπως και ο αρχικός αλγόριθμος, καταστέλλοντας εκθετικά τις διακυμάνσεις της θερμοκρασίας από την τιμή αναφοράς (Bussi et al., 2007). Ωστόσο, το σφάλμα της αρχικής μεθόδου διορθώνεται με την εφαρμογή ενός παράγοντα Wiener (W) κατά τον υπολογισμό της μεταβολής της κινητικής ενέργειας

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2 \cdot \left(\frac{K \cdot K_0}{N_f} \right)^{1/2} \cdot \frac{dW}{\sqrt{\tau_T}} \quad (17)$$

όπου K είναι η κινητική ενέργεια, K_0 η τιμή αναφοράς που προκύπτει από τη θερμοκρασία αναφοράς T_0 , N_f είναι ο αριθμός των βαθμών ελευθερίας του συστήματος, τ_T είναι η σταθερά της περιόδου σύζευξης. Η εφαρμογή αυτής της διόρθωσης επιτυγχάνει τη σωστή κατανομή της κινητικής ενέργειας, οδηγώντας έτσι σε μια κατάσταση που προσεγγίζει την κανονική κατανομή για την προσομοίωση (Bussi et al., 2007). Παράλληλα, καθώς η μέθοδος βασίζεται

Κεφάλαιο II – Υλικά & Μέθοδοι

στον αλγόριθμο του Berendsen, παραμένει υπολογιστικά ευέλικτη. Έτσι, η μέθοδος v-rescale χρησιμοποιείται πολύ συχνά, τόσο στην εξισορρόπηση του συστήματος όσο και στο στάδιο παραγωγής.

Η μέθοδος σύζευξης Nosé-Hoover: σε αντίθεση με τους αλγορίθμους του Berendsen και της επανακλιμάκωσης των ταχυτήτων, οι οποίοι ρυθμίζουν την κινητική ενέργεια του συστήματος έμμεσα, η μέθοδος σύζευξης Nosé-Hoover επιτυγχάνει την κανονική κατανομή άμεσα τροποποιώντας τη Χαμιλτονιανή του συστήματος, με την εισαγωγή μιας εξωτερικής πηγής θερμότητας που χαρακτηρίζεται από μια σταθερά μάζας Q και έναν παράγοντα τριβής ξ (Hoover, 1985; Nosé, 1984). Έτσι, η νέα Χαμιλτονιανή του συστήματος είναι της μορφής

$$H = K + U + \frac{p_\xi^2}{2 \cdot Q} + N_f \cdot k_B \cdot T \cdot \xi \quad (18)$$

όπου K και U είναι η κινητική και η δυναμική ενέργεια, αντίστοιχα, Q και ξ είναι οι παράμετροι του θερμοστάτη, p_ξ είναι μια τιμή ορμής που σχετίζεται με την τριβή του θερμοστάτη, k_B είναι η σταθερά Boltzmann και T η θερμοκρασία. Έτσι, με την εφαρμογή της μεθόδου Nosé-Hoover, η εξίσωση κίνησης του συστήματος τροποποιείται ως εξής:

$$\alpha_i = \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q} \cdot u_i = \frac{F_i}{m_i} - \frac{p_\xi}{Q} \frac{dr_i}{dt} \quad (19)$$

Σημειώνεται πως στην προσέγγιση των Nosé-Hoover, η παράγωγος της ορμής p_ξ μπορεί να εξισωθεί με τη διαφορά $T - T_0$. Επιπρόσθετα, η σταθερά Q μπορεί να αναλυθεί περαιτέρω σαν συνάρτηση της θερμοκρασίας T_0 και μιας χρονικής σταθεράς σύζευξης τ_T , αντίστοιχης της σταθεράς που χρησιμοποιείται στον αλγόριθμο του Berendsen:

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \quad (20)$$

Έτσι, ουσιαστικά η ρύθμιση του θερμοστάτη στη μέθοδο Nosé-Hoover μπορεί να πραγματοποιηθεί θέτοντας αντίστοιχες παραμέτρους (T_0 και τ_T) με το θερμοστάτη του Berendsen. Θεωρητικά, λοιπόν, η μέθοδος θα μπορούσε να χρησιμοποιηθεί στο σύνολο των προσομοιώσεων ενός συστήματος, τόσο για το στάδιο εξισορρόπησης όσο και για το στάδιο παραγωγής. Αλγοριθμικά, ωστόσο, η μέθοδος έχει βρεθεί να είναι 4-5 φορές πιο αργή σε σχέση με έμμεσες προσεγγίσεις. Έτσι, πρακτικά, η μέθοδος Nosé-Hoover χρησιμοποιείται μόνο στις προσομοιώσεις παραγωγής, ενώ στα στάδια εξισορρόπησης εφαρμόζονται ευέλικτες μέθοδοι όπως ο θερμοστάτης του Berendsen.

Κεφάλαιο II – Υλικά & Μέθοδοι

Η Δυναμική Langevin: Η Δυναμική Langevin προσφέρει μια εναλλακτική προσέγγιση στον έλεγχο της θερμοκρασίας. Η μέθοδος διαφοροποιείται από τους προηγούμενους αλγορίθμους που παρουσιάστηκαν, καθώς δεν επηρεάζει τη Χαμιλτονιανή του συστήματος ούτε τροποποιεί την απόκλιση των ταχυτήτων. Αντίθετα, η Δυναμική Langevin επηρεάζει άμεσα την κινητικότητα του συστήματος εισάγοντας ένα στοχαστικό παράγοντα, σε συμφωνία με τις βασικές αρχές της στοχαστικής δυναμικής που διέπουν τις κινήσεις Brown (Brünger, 1992). Η εξίσωση της κίνησης τροποποιείται και λαμβάνει τη μορφή

$$m_i \cdot \ddot{\alpha}_i = F_i - \gamma \cdot \dot{u}_i + R_i(t) \quad (21)$$

όπου γ είναι ένας παράγοντας τριβής, μετρούμενος σε αντίστροφες μονάδες χρόνου και $R_i(t)$ είναι μια τυχαία δύναμη, η μέση τιμή της οποίας ανάμεσα σε δύο σωματίδια i και j και χρονικά στιγμιότυπα t και t' προκύπτει από τη σχέση

$$\langle R_i(t), R_j(t') \rangle = \delta_{ij} \delta(t - t') \cdot 6 \cdot k_B \cdot m_{ij} \cdot \gamma \cdot T \quad (22)$$

όπου οι παράγοντες δ συμβολίζουν τις αντίστοιχες συναρτήσεις Dirac, k_B είναι η σταθερά του Boltzmann, T είναι η θερμοκρασία και m η μάζα, ενώ οι αγκύλες υποδεικνύουν τον υπολογισμό της μέσης τιμής για τα σωματίδια i και j και τα στιγμιότυπα t και t' . Σε κάθε στιγμιότυπο t , κάθε σωματίδιο i του συστήματος δέχεται αυτήν την δύναμη και η ταχύτητά του κλιμακώνεται με βάση τον εφαρμοζόμενο παράγοντα γ . Η συσχέτιση της δύναμης R με τον παράγοντα γ και τη θερμοκρασία, όπως παρουσιάζεται στην Εξίσωση 21 ουσιαστικά ικανοποιεί το θεώρημα διαταραχών-ανάλωσης του Nyquist, οδηγώντας έτσι σε πραγματική κανονική κατανομή για την προσομοίωση (Brünger, 1992).

2.1.5.2 Αλγόριθμοι Ελέγχου της Πίεσης

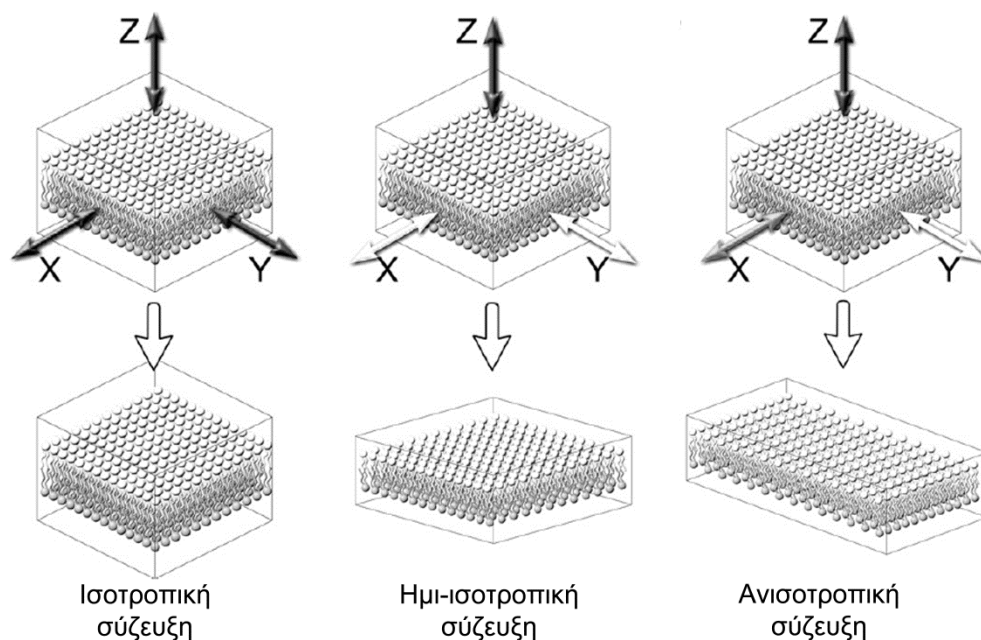
Η ισόθερμη-ισοβαρής κατανομή (NPT) χαρακτηρίζεται τόσο από σταθερή θερμοκρασία όσο και από σταθερή πίεση. Στις προσομοιώσεις Μοριακής Δυναμικής, η πίεση του συστήματος (P) υπολογίζεται μέσω του θεωρήματος Virial σαν συνάρτηση του όγκου, της θέσης κάθε σωματιδίου και της εφαρμοζόμενης σε αυτό δύναμης, σύμφωνα με την Εξίσωση

$$P = \frac{Nk_B T}{V} + \frac{1}{dV} \sum_{i=1}^N r_i F_i \quad (23)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου N είναι ο αριθμός των σωματιδίων, V ο όγκος, k_B η σταθερά Boltzmann, r_i η θέση κάθε σωματιδίου, d ο αριθμός των διαστάσεων του συστήματος ($d=3$ για τρισδιάστατα συστήματα) και F_i η συνολική δύναμη που ασκείται σε κάθε σωματίδιο (de Miguel & Jackson, 2006).

Αντίστοιχα με τη θερμοκρασία, ο έλεγχος της πίεσης πραγματοποιείται προσομοιώνοντας τη σύζευξη του συστήματος σε έναν εξωτερικό βαροστάτη, ο οποίος διατηρεί την πίεση κοντά στην τιμή αναφοράς. Ο έλεγχος της πίεσης μπορεί να πραγματοποιηθεί διαφορετικά στις τρεις διαστάσεις του χώρου, ανάλογα με τη φύση του συστήματος και τις ανάγκες της προσομοίωσης (Εικόνα 28). Συγκεκριμένα, η σύζευξη της πίεσης μπορεί να πραγματοποιηθεί με τρεις διαφορετικές προσεγγίσεις (Kandt et al., 2007):



Εικόνα 28 Οι τρεις διαφορετικοί τρόποι ελέγχου της πίεσης. Η ισοτροπική σύζευξη περιλαμβάνει τον κοινό έλεγχο της πίεσης και στις τρεις διαστάσεις. Οι διαστάσεις της κυψελίδας μεταβάλλονται ισόποσα και στους τρεις άξονες x , y και z . Η ημι-ισοτροπική σύζευξη παρέχει κοινό έλεγχο της πίεσης στους άξονες x και y και ανεξάρτητο έλεγχο στον άξονα z . Η επιφάνεια x - y της κυψελίδας παραμένει σταθερή. Η ανισοτροπική σύζευξη επιτρέπει ανεξάρτητο έλεγχο της πίεσης για κάθε άξονα του χώρου. Αυτός ο τρόπος ελέγχου μπορεί να οδηγήσει σε σημαντικές αλλαγές στις διαστάσεις της κυψελίδας σε κάθε άξονα. Προσαρμογή από Kandt et al., 2007.

Κεφάλαιο II – Υλικά & Μέθοδοι

- *Ισοτροπική σύζευξη*: η σύζευξη της πίεσης είναι κοινή και στις τρεις διαστάσεις x , y , z . Διακυμάνσεις του συστήματος ως προς τη μία κατεύθυνση, π.χ. τον άξονα x , συνοδεύονται από ισοδύναμες διακυμάνσεις στους άλλους δύο άξονες.
- *Ανισοτροπική σύζευξη*: ο έλεγχος της πίεσης πραγματοποιείται ανεξάρτητα για κάθε άξονα.
- *Ημι-ισοτροπική σύζευξη*: πραγματοποιείται κοινή σύζευξη για δύο από τους τρεις άξονες του χώρου, x και y , ενώ η πίεση στον άξονα z μεταβάλλεται ανεξάρτητα.

Στις προσομοιώσεις Μοριακής Δυναμικής που αφορούν βιομοριακά συστήματα, προσομοιώσεις που πραγματοποιούνται σε συνθήκες υδατικού διαλύτη (π.χ. μια σφαιρική υδατοδιαλυτή πρωτεΐνη στο περιβάλλον της) χρησιμοποιούν συνήθως *ισοτροπική σύζευξη* για τον έλεγχο της πίεσης, καθώς το περιβάλλον μοντελοποιείται σαν μια κυψελίδα διαστάσεων x , y , και z γεμάτη με υδατικό διαλύτη και θεωρείται ομοιογενές στις τρεις διαστάσεις του. Αντίθετα, οι προσομοιώσεις που περιλαμβάνουν βιολογικές μεμβράνες προτυποποιούν τη λιπιδική διπλοστιβάδα στο επίπεδο των αξόνων x και y , με τον άξονα z να είναι παράλληλος με τη μεσοκάθετο της μεμβράνης, και το διαλύτη εκατέρωθεν της να εκτείνεται στον άξονα z . Για την ορθή προσομοίωση των χαρακτηριστικών της μεμβράνης και την αναπαραγωγή σημαντικών χαρακτηριστικών (π.χ. επιφάνεια ανά λιπίδιο, πλευρική διάχυση, επιφανειακή τάση κλπ) χωρίς παραμορφώσεις, είναι απαραίτητο οι διαστάσεις της μεμβράνης να παραμένουν σταθερές, ενώ ο διαλύτης να μεταβάλλεται ελεύθερα. Έτσι, στις προσομοιώσεις NPT μεμβρανικών συστημάτων χρησιμοποιείται πάντα η *ημι-ισοτροπική σύζευξη* (Kandt et al., 2007).

Ανεξάρτητα από τον τρόπο σύζευξης της πίεσης, ο έλεγχος των τιμών της μπορεί να πραγματοποιηθεί με μια σειρά από αλγορίθμους. Στην παρούσα διδακτορική διατριβή έχουν χρησιμοποιηθεί τρεις διαφορετικές προσεγγίσεις, οι οποίες παρουσιάζονται παρακάτω:

Ο βαροστάτης Berendsen: ο βαροστάτης του Berendsen (Berendsen barostat) είναι μια μέθοδος ασθενούς σύζευξης της πίεσης. Κατά αντιστοιχία με τον ομώνυμο θερμοστάτη, ο βαροστάτης Berendsen λειτουργεί καταστέλλοντας τις διακυμάνσεις της πίεσης από την τιμή αναφοράς P_0 με έναν παράγοντα μ της μορφής

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_P} (P - P_0) \quad (24)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου P είναι η πίεση του συστήματος σε κάθε δεδομένη στιγμή, P_0 η πίεση αναφοράς, Δt είναι το βήμα χρόνου, β είναι η ισοθερμική συμπιεστότητα του συστήματος και τ_p είναι μια χρονική σταθερά μέσω της οποίας ρυθμίζεται η ισχύς της σύζευξης στο βαροστάτη (Berendsen et al., 1984). Ένα σημαντικό πλεονέκτημα του βαροστάτη Berendsen είναι η ικανότητα χρήσης τόσο ιστροπικής όσο και ανιστροπικής σύζευξης της πίεσης, η οποία υλοποιείται με τη χρήση διαφορετικών τιμών τ_p για κάθε βαθμό ελευθερίας. Ωστόσο, αντίστοιχα με το θερμοστάτη Berendsen ο βαροστάτης, παρ'ότι μπορεί να δώσει τιμές μέσης πίεσης στο σύστημα κοντά στην τιμή P_0 , αδυνατεί να προτυποποιήσει σωστά την κατανομή της πίεσης. Αυτό μπορεί να δημιουργήσει σφάλματα στις προσομοιώσεις, ειδικά σε περιπτώσεις όπου οι ταλαντώσεις της πίεσης του συστήματος παρουσιάζουν βιολογικό ενδιαφέρον (π.χ. στη μέτρηση της ροής ιόντων μέσα από ένα διαμεμβρανικό ιοντικό κανάλι, όπου οι ιδιότητες της μεμβράνης επηρεάζονται από την πίεση). Έτσι, συνηθίζεται ο βαροστάτης Berendsen να χρησιμοποιείται κατά το στάδιο της εξισορρόπησης, ενώ στο στάδιο της παραγωγής εφαρμόζονται πιο αυστηρές μέθοδοι ελέγχου.

Ο βαροστάτης Parrinello-Rahman: ο βαροστάτης Parrinello-Rahman υλοποιεί τις συνθήκες ισοβαρούς προσομοίωσης προτυποποιώντας τις διαστάσεις του συστήματος σαν ένα τρισδιάστατο πίνακα b , ο οποίος ακολουθεί τους κανόνες των εξισώσεων κίνησης σύμφωνα με την Εξίσωση 24

$$\frac{db^2}{dt^2} = V \cdot W^{-1} \cdot b'^{-1}(P - P_0) \quad (25)$$

όπου V είναι ο όγκος και W είναι μια παράμετρος του πίνακα σχετιζόμενη με την ισχύ της σύζευξης. Η παράμετρος W μπορεί περαιτέρω να αναλυθεί σαν συνάρτηση της συμπιεστότητας β και της σταθεράς σύζευξης τ_p , με βάση την Εξίσωση 25

$$W^{-1} = \frac{4\pi^2\beta}{3\tau_p L} \quad (26)$$

όπου L είναι η τιμή του μεγαλύτερου στοιχείου του πίνακα b (Parrinello & Rahman, 1981). Ο βαροστάτης Parrinello-Rahman μπορεί να δώσει σωστή κατανομή για την πίεση, ωστόσο, αδυνατεί να την προτυποποιήσει σωστά όταν η αρχική τιμή P του συστήματος διαφέρει σημαντικά από την τιμή αναφοράς P_0 . Έτσι, συνηθίζεται να χρησιμοποιείται μόνο στο στάδιο παραγωγής, ενώ για την εξισορρόπηση του συστήματος εφαρμόζονται μέθοδοι ασθενούς σύζευξης όπως ο βαροστάτης Berendsen.

Κεφάλαιο II – Υλικά & Μέθοδοι

Η μέθοδος πιστονιού *Langevin*: Η μέθοδος πιστονιού *Langevin* αποτελεί μια επέκταση της Δυναμικής *Langevin* για το συντονισμένο έλεγχο της θερμοκρασίας και της πίεσης. Η μέθοδος τροποποιεί την εξίσωση κίνησης, εισάγοντας περαιτέρω όρους για τον έλεγχο της πίεσης μέσω του όγκου του συστήματος:

$$m_i \ddot{\alpha}_i = F_i - \frac{1}{3} \frac{\dot{V}}{V} p_i \quad (27)$$

όπου $\dot{V}$ είναι η παράγωγος του όγκου και p_i η ορμή. Η τιμή $\dot{V}$ μπορεί περαιτέρω να συσχετιστεί με την πίεση μέσα από μια εξίσωση *Langevin*:

$$\dot{V} = \frac{1}{W} (P - P_0) - \gamma \dot{V} + R(t) \quad (28)$$

όπου γ είναι ο παράγοντας τριβής της δυναμικής *Langevin*, W είναι η παράμετρος μάζας του πιστονιού, σε αναλογία με την αντίστοιχη τιμή W άλλων βαροστατών όπως ο *Parrinello-Rahman*, και $R(t)$ είναι μια τυχαία εφαρμοζόμενη δύναμη, αντίστοιχη της δύναμης που εφαρμόζεται στον έλεγχο της θερμοκρασίας από τη Δυναμική *Langevin* (Feller et al., 1995). Πρέπει να σημειωθεί ότι η μέθοδος του πιστονιού *Langevin* εξαρτάται άμεσα από τη Δυναμική *Langevin* για την υλοποίησή της και μπορεί να συνδυαστεί μόνο με αυτήν για την επίτευξη NPT συνθηκών. Αντίθετα, οι βαροστάτες *Berendsen* και *Parrinello-Rahman* είναι πιο ευέλικτοι και μπορούν να συνδυαστούν με μεγαλύτερη ποικιλία θερμοστατών.

2.1.6 Συναρτήσεις επιμερισμού και θερμοδυναμικές ιδιότητες

Κάθε θερμοδυναμική κατανομή χαρακτηρίζεται από μία συνάρτηση επιμερισμού Q (partition function), η οποία περιλαμβάνει το σύνολο των διαθέσιμων καταστάσεων που μπορούν να παρατηρηθούν υπό τις συνθήκες της κατανομής. Έτσι, οι θερμοδυναμικές ιδιότητες κάθε συστήματος μπορούν να υπολογιστούν με βάση τη συνάρτηση επιμερισμού του, καθώς αυτή περιέχει το σύνολο της θερμοδυναμικής πληροφορίας για την προσομοίωση (Wilde & Singh, 1998). Στη μικροκανονική κατανομή (NVE) η συνάρτηση επιμερισμού (Q_{NVE}) ορίζεται σαν συνάρτηση της διαφοράς της εσωτερικής ενέργειας από τη συνολική ενέργεια του συστήματος

$$Q_{NVE} = \frac{1}{h^{3N} N!} \sum_{\Gamma} \delta_i(H_i - E) \quad (29)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου H_i είναι η τιμή της Χαμιλτονιανής σε κάθε κατάσταση i της κατανομής, E είναι η συνολική ενέργεια του συστήματος, h είναι η σταθερά Planck, N είναι ο αριθμός των σωματιδίων και $N!$ το παραγοντικό του N , ενώ δ_i είναι η συνάρτηση δ του Dirac. Στην κανονική κατανομή (NVT) η συνάρτηση επιμερισμού (Q_{NVT}) ορίζεται σαν συνάρτηση της εσωτερικής ενέργειας (E_i) και της θερμοκρασίας (T)

$$Q_{NVT} = \sum_0^\infty e^{-E_i/k_B T} \quad (30)$$

όπου k_B είναι η σταθερά Boltzmann. Ωστόσο, με το συνδυασμό όλων των καταστάσεων του συστήματος που έχουν κοινή ενέργεια (έστω E_j), η Q_{NVT} μπορεί να οριστεί και σαν συνάρτηση της Q_{NVE} :

$$Q_{NVT} = \sum_0^\infty Q_{NVE}(E_j, V, N) e^{-E_j/k_B T} \quad (31)$$

Τέλος, η ισόθερμη-ισοβαρής κατανομή (NPT) χαρακτηρίζεται από μια συνάρτηση επιμερισμού (Q_{NPT}) της μορφής

$$Q_{NPT} = \sum_{V_j=0}^\infty \sum_{i=0}^\infty e^{\frac{-(E_i + PV_j)}{k_B T}} \quad (32)$$

και μπορεί περαιτέρω να οριστεί και σαν συνάρτηση της Q_{NVT} :

$$Q_{NPT} = \frac{1}{V_0} \int Q_{NVT} e^{PV_j/k_B T} dV \quad (33)$$

όπου V είναι ο όγκος και P η πίεση.

Από τη συνάρτηση επιμερισμού κάθε κατανομής μπορούν να προκύψουν οι βασικές θερμοδυναμικές ιδιότητες του συστήματος. Σε κάθε περίπτωση, η υπολογιζόμενη θερμοδυναμική ιδιότητα ορίζεται σαν συνάρτηση του φυσικού λογαρίθμου της αντίστοιχης συνάρτησης επιμερισμού. Έτσι, στην κατανομή NVE η τιμή Q_{NVE} μπορεί να χρησιμοποιηθεί για τον υπολογισμό της εντροπίας του συστήματος (S):

$$S = k_B \ln(Q_{NVE}) \quad (34)$$

Στην κατανομή NVT, η συνάρτηση επιμερισμού μπορεί να δώσει την ελεύθερη ενέργεια Helmholtz (F):

$$F = -k_B T \ln(Q_{NVT}) \quad (35)$$

Τέλος, στην κατανομή NPT, η συνάρτηση επιμερισμού μπορεί να χρησιμοποιηθεί για τον υπολογισμό της ελεύθερης ενέργειας Gibbs:

$$G = -k_B T \ln(Q_{NPT}) \quad (36)$$

2.1.7 Δυναμικά Πεδία

Σε μια προσομοίωση Μοριακής Δυναμικής κάθε σωματίδιο του συστήματος δέχεται την επίδραση δυνάμεων, οι οποίες αντιπροσωπεύουν τις αλληλεπιδράσεις του σωματιδίου με τα υπόλοιπα στοιχεία του συστήματος. Η απαραίτητη πληροφορία για τον υπολογισμό αυτών των δυνάμεων δίνεται από τα δυναμικά πεδία (Force Fields). Τα δυναμικά πεδία αποτελούν μαθηματικά μοντέλα που χρησιμοποιούνται για τον υπολογισμό των δυνάμεων που λειτουργούν σε ατομικό επίπεδο, καθώς επίσης και την περιγραφή της δυναμικής ενέργειας (U) των υπό μελέτη συστημάτων ως συνάρτηση των αλληλεπιδράσεών τους και της θέσης των ατόμων στο χώρο. Η πληροφορία που περιέχεται σε ένα δυναμικό πεδίο περιλαμβάνει τόσο τις απαραίτητες παραμέτρους για την περιγραφή των μορίων όσο και τις κατάλληλες εξισώσεις για την προτυποποίηση των αλληλεπιδράσεων μεταξύ τους (Phillips et al., 2005). Οι παράμετροι που περιγράφουν τα άτομα περιλαμβάνουν χαρακτηριστικά όπως η ακτίνα van der Waals, και το φορτίο για κάθε άτομο, καθώς και πληροφορίες σχετικά με την τοπολογία των ατόμων κατά τη συμμετοχή τους στο σχηματισμό μορίων, δηλαδή πληροφορίες όπως το ποια άτομα ενώνονται ομοιοπολικά με δεσμούς και ποια όχι, το μήκος και η γωνία των δεσμών ανάμεσα σε άτομα, η ύπαρξη διέδρων γωνιών αλλά και πληροφορίες για τις μη ομοιοπολικές αλληλεπιδράσεις ανάμεσα σε διάφορες ομάδες (Χαμόδρακας, 1993).

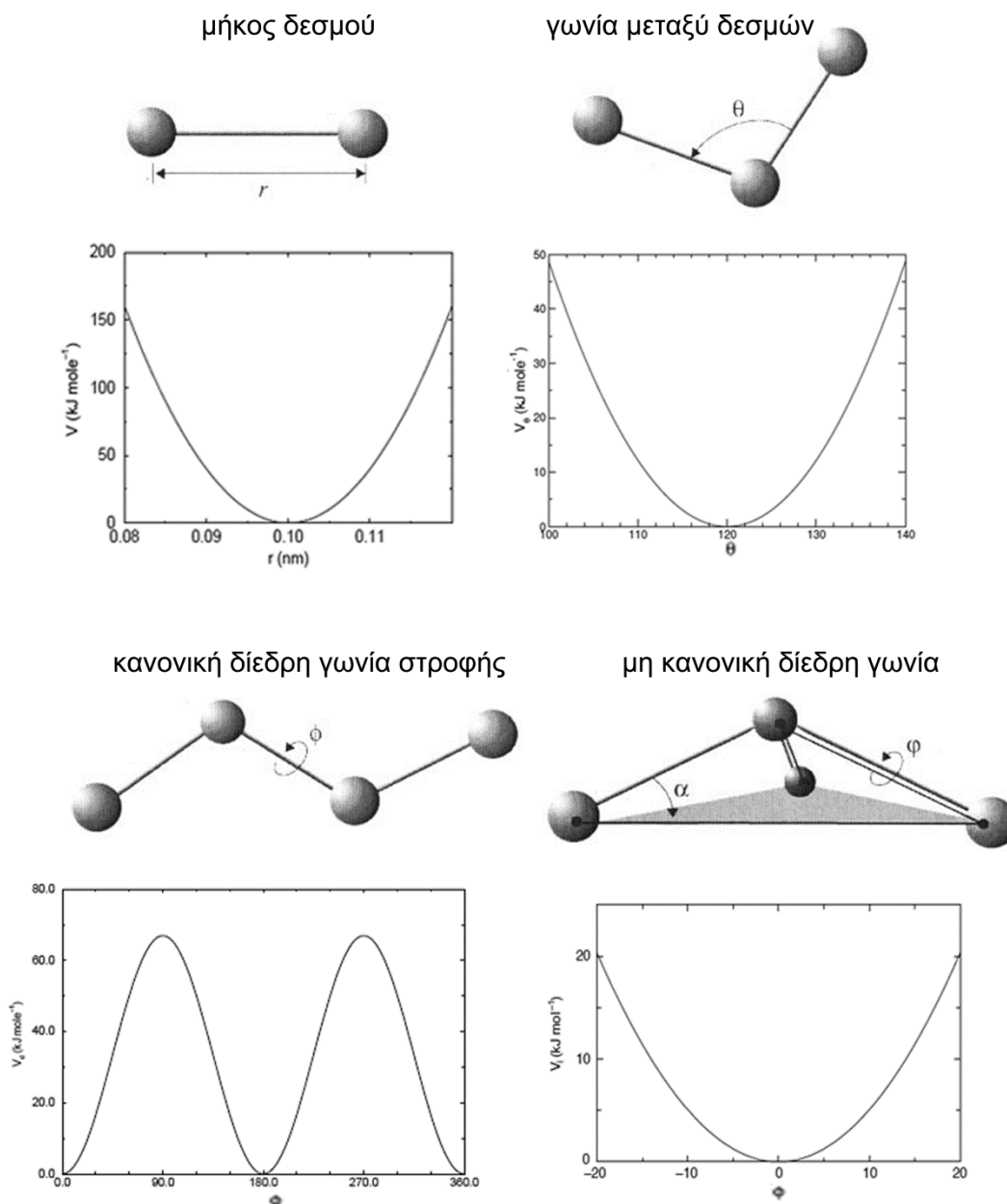
Οι εξισώσεις των δυναμικών πεδίων περιλαμβάνουν συναρτήσεις δυναμικών για την περιγραφή των αλληλεπιδράσεων ανάμεσα στα διάφορα άτομα και μόρια. Το άθροισμά τους συνιστά τη συνολική δυναμική ενέργεια του συστήματος. Τα περισσότερα δυναμικά πεδία περιγράφουν συνολικά τις αλληλεπιδράσεις χρησιμοποιώντας μια κοινή συνάρτηση, η οποία αποτελείται από τους παρακάτω συντελεστές:

$$U = U_{bond} + U_{angle} + U_{dihe} + U_{impro} + U_{vdW} + U_{elec} \quad (37)$$

Οι τέσσερις πρώτοι όροι της εξίσωσης συνολικά περιγράφουν το κομμάτι των ομοιοπολικών αλληλεπιδράσεων του συστήματος (Εικόνα 29). Οι συγκεκριμένοι όροι χαρακτηρίζονται και σαν δυναμικά παραμόρφωσης, αφού ουσιαστικά προτυποποιούν ενεργειακά τις αποκλίσεις των ομοιοπολικά συνδεδεμένων σωματιδίων από τις θέσεις ισορροπίας τους. Καθένα από αυτά τα δυναμικά εκφράζεται σαν ενεργειακή συνάρτηση του αντίστοιχου μεγέθους που χαρακτηρίζει τη θέση των σωματιδίων (μέτρο της απόσταση μεταξύ

Κεφάλαιο II – Υλικά & Μέθοδοι

ατόμων για το μήκος δεσμού, μέτρο της γωνίας για τα δυναμικά των γωνιών και των δίεδρων), με την επιπλέον εφαρμογή μιας σταθεράς ελατηρίου k . Σε όλες τις περιπτώσεις η αντίστοιχη εφαρμοζόμενη δύναμη προκύπτει παραγωγίζοντας τη συνάρτηση της ενέργειας ως προς τη θέση των σωματιδίων.



Εικόνα 29 Οι ομοιοπολικές αλληλεπιδράσεις στη συνάρτηση δυναμικής ενέργειας. Διαγραμματική απεικόνιση των ομοιοπολικών αλληλεπιδράσεων (πάνω) και του δυναμικού τους σαν συνάρτηση της θέσης (κάτω).

Κεφάλαιο II – Υλικά & Μέθοδοι

Οι δύο τελευταίοι όροι της συνάρτησης του δυναμικού πεδίου περιγράφουν τις μη ομοιοπολικές αλληλεπιδράσεις του συστήματος. Ο όρος U_{vdW} αντιστοιχεί στις αλληλεπιδράσεις van der Waals ανάμεσα σε δύο σωματίδια, οι οποίες περιγράφουν συλλογικά τις ελκτικές δυνάμεις διασποράς μεταξύ στιγμιαίου διπόλου – επαγόμενου διπόλου (δυνάμεις London) και τις απωστικές δυνάμεις ανταλλαγής από την αλληλοεπικάλυψη των ηλεκτρονιακών νεφών των ατόμων (Εικόνα 2.6). Ο όρος U_{elec} , τέλος, περιγράφει τις ηλεκτροστατικές αλληλεπιδράσεις ανάμεσα σε φορτισμένα σωματίδια, με βάση τις τιμές των φορτίων και την απόσταση μεταξύ τους. Αντίστοιχα με τις ομοιοπολικές αλληλεπιδράσεις, οι εφαρμοζόμενες δυνάμεις των μη ομοιοπολικών αλληλεπιδράσεων προκύπτουν από την παραγωγή των ενεργειακών συναρτήσεων ως προς τη θέση των σωματιδίων.

2.1.7.1 Ομοιοπολικές αλληλεπιδράσεις

Για κάθε ομοιοπολικό δεσμό i σχηματιζόμενο από δύο σωματίδια α και β , το δυναμικό του (U_{bond}) περιγράφει το μήκος r του δεσμού και προτυποποιείται με μια αρμονική εξίσωση της μορφής

$$U_{bond}(r) = \frac{1}{2}k_b(r_i - r_0)^2 \quad (38)$$

όπου r_i είναι η απόσταση των α και β , r_0 η τιμή αναφοράς και k_b η σταθερά ελατηρίου.

Η γωνία θ μεταξύ δύο συνεχόμενων ομοιοπολικών δεσμών, σχηματιζόμενων από τα σωματίδια α , β και γ , υλοποιείται επίσης σαν αρμονικό δυναμικό (U_{angle}) της μορφής

$$U_{angle}(\theta) = \frac{1}{2}k_\theta(\theta_i - \theta_0)^2 = \frac{1}{2}k_\theta^{cos}(\cos\theta_i - \cos\theta_0)^2 \quad (39)$$

όπου θ_i είναι η γωνία των σωματιδίων, θ_0 η τιμή αναφοράς και k_θ η σταθερά ελατηρίου. Σημειώνεται ότι σε κάποια δυναμικά πεδία ο παραπάνω όρος διαφοροποιείται ελαφρά, χρησιμοποιώντας όχι τις ίδιες τις γωνίες αλλά τα συνημίτονά τους. Σε αυτήν την περίπτωση εφαρμόζονται διαφορετικές τιμές για τη σταθερά ελατηρίου (k_θ^{cos}), προκειμένου το τελικό αποτέλεσμα να είναι το ίδιο.

Μια δίεδρη γωνία ϕ τυπικά σχηματίζεται από τέσσερα συνεχόμενα, άμεσα συνδεδεμένα μεταξύ τους άτομα α , β , γ και δ . Γωνίες αυτού του τύπου χαρακτηρίζονται και

Κεφάλαιο II – Υλικά & Μέθοδοι

σαν «κανονικές» (proper) δίδρες γωνίες. Το δυναμικό παραμόρφωσης των δίδρων γωνιών (U_{dih}) είναι μια συνάρτηση της μορφής

$$U_{dih}(\varphi) = k_{\varphi}(1 + \cos(n\varphi_i - \delta)) \quad (40)$$

όπου φ_i είναι η δίδρη γωνία, δ η φάση του δυναμικού και n η πολλαπλότητα της γωνίας. Ο τελευταίος όρος είναι ένας ακέραιος αριθμός που μπορεί να λάβει θετικές τιμές ($n=1, 2, 3, \dots$) και εφαρμόζεται στο δυναμικό καθώς μία δίδρη γωνία μπορεί να έχει περισσότερες από μία ενεργειακά βέλτιστες τιμές. Χαρακτηριστικό παράδειγμα αποτελούν οι δίδρες γωνίες φ και ψ των πρωτεϊνών, οι οποίες μπορούν να λάβουν περισσότερες από μία τιμές, με τους διαφορετικούς συνδυασμούς τους να χαρακτηρίζουν τα διαφορετικά στοιχεία δευτεροταγούς δομής. Έτσι, ανάλογα με το μέγεθος του n , μια δίδρη γωνία μπορεί να χαρακτηρίζεται με περισσότερα του ενός ζευγάρια τιμών k_{φ} και δ , ανάλογα με το συνολικό αριθμό των ενεργειακών ελαχίστων της.

Εκτός από τις κανονικές δίδρες γωνίες, τα περισσότερα δυναμικά πεδία περιλαμβάνουν έναν επιπλέον όρο, ο οποίος χαρακτηρίζει ένα σύνολο τεσσάρων σωματιδίων α, β, γ και δ , η θέση των οποίων μπορεί να περιγραφεί από τη δίδρη γωνία ταλάντωσης ενός επιπέδου. Οι γωνίες αυτού του τύπου χαρακτηρίζονται ως «μη κανονικές» (improper) δίδρες γωνίες και προτυποποιούνται από ένα αρμονικό δυναμικό (U_{impro}) της μορφής

$$U_{impro}(\varphi) = k_{imp}(\varphi_i - \varphi_0)^2 \quad (41)$$

όπου φ_i είναι η δίδρη γωνία, φ_0 η τιμή αναφοράς και k_{imp} η σταθερά ελατηρίου (Phillips et al., 2005).

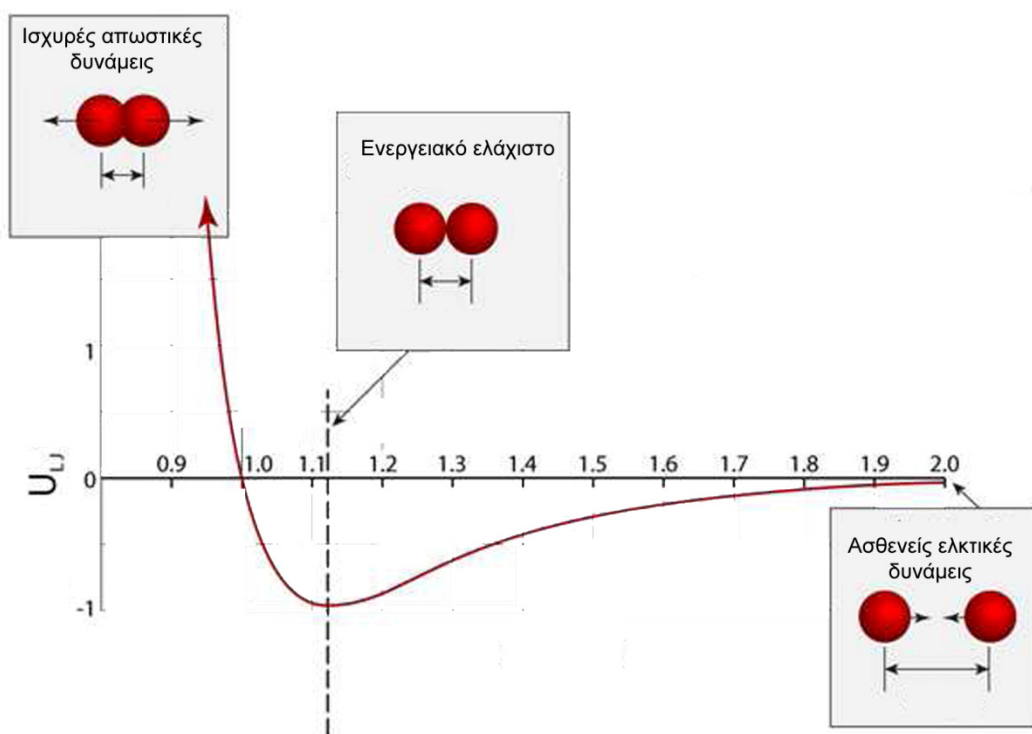
2.1.7.2 Μη ομοιοπολικές αλληλεπιδράσεις

Οι μη ομοιοπολικές αλληλεπιδράσεις περιλαμβάνουν τις αλληλεπιδράσεις van der Waals και τις ηλεκτροστατικές αλληλεπιδράσεις. Οι αλληλεπιδράσεις van der Waals (Εικόνα 30) ανάμεσα σε δύο σωματίδια i και j τυπικά υλοποιούνται με τη μορφή του δυναμικού Lennard-Jones

$$U_{vdw}(i, j) = \frac{B}{r_{ij}^{12}} - \frac{A}{r_{ij}^6} \quad (42)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου οι A και B είναι σταθερές εξαρτώμενες από τη φύση των αλληλεπιδρώντων σωματιδίων και r_{ij} είναι η απόσταση των σωματιδίων i και j (Χαμόδρακας, 1993). Ο πρώτος όρος της εξίσωσης (B/r_{ij}^{12}) περιγράφει το κομμάτι των απωστικών δυνάμεων ανταλλαγής, ενώ ο δεύτερος ($-A/r_{ij}^6$) το κομμάτι των ελκτικών δυνάμεων διασποράς (Lennard-Jones, 1931). Οι τιμές των σταθερών A και B διαφοροποιούνται για κάθε άτομο ανάλογα με τη φύση των χημικών ομάδων στις οποίες συμμετέχει, καθώς επίσης και για τα διαφορετικά ζεύγη ατόμων i και j. Έτσι, για παράδειγμα, οι τιμές των A και B ανάμεσα σε δύο άνθρακες που συμμετέχουν και οι δύο σε δυο μεθυλομάδες θα είναι διαφορετικές από τις αντίστοιχες τιμές A και B δυο ανθράκων εκ των οποίων ο πρώτος συμμετέχει σε μια μεθυλομάδα και ο δεύτερος σε ένα βενζολικό δακτύλιο.



Εικόνα 30 Το δυναμικό Lennard-Jones. Διαγραμματική απεικόνιση της ενέργειας Lennard-Jones σε συνάρτηση με την απόσταση ανάμεσα σε δύο άτομα. Σημαίνονται οι περιοχές των ελκτικών και απωστικών δυνάμεων. Το ενεργειακό ελάχιστο της αλληλεπίδρασης βρίσκεται στο χαμηλότερο σημείο της καμπύλης δυναμικού και αντιπροσωπεύεται από το φρέαρ δυναμικού (-ε). Όταν η απόσταση μεταξύ των ατόμων βρίσκεται αριστερά αυτού του σημείου τα άτομα απωθούνται, ενώ όταν βρίσκεται δεξιά έλκονται μεταξύ τους.

Κεφάλαιο II – Υλικά & Μέθοδοι

Η Εξίσωση 42, γνωστή και σαν *μορφή AB*, είναι η αρχική εκδοχή του δυναμικού, όπως περιγράφηκε από τον ίδιο τον Lennard-Jones το 1931 (Lennard-Jones, 1931). Η συγκεκριμένη εξίσωση χρησιμοποιείται ακόμα σε κάποια δυναμικά πεδία, όπως το GROMOS. Ωστόσο, στις περισσότερες περιπτώσεις πλέον χρησιμοποιούνται οι τροποποιημένες εκδοχές του δυναμικού, οι οποίες παρουσιάζονται στην Εξίσωση 43

$$U_{vdw}(i, j) = \varepsilon_{ij} \left[\left(\frac{r_{ij}^m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^m}{r_{ij}} \right)^6 \right] = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (43)$$

όπου ε_{ij} είναι το “φρέαρ δυναμικού”, δηλαδή το μέτρο του ενεργειακού ελάχιστου των αλληλεπιδράσεων van der Waals, r^m είναι η τιμή της απόστασης των σωματιδίων i και j στην οποία το δυναμικό van der Waals βρίσκεται στο ενεργειακό ελάχιστο ($-\varepsilon_{ij}$) και σ_{ij} είναι η τιμή της απόστασης στην οποία το δυναμικό Lennard-Jones μηδενίζεται. Η τιμή σ_{ij} ισούται με το άθροισμα των ακτινών van der Waals των δύο αλληλεπιδρώντων σωματιδίων i και j . Η τιμή r^m προκύπτει από την απόσταση σ μέσω της σχέσης $r^m = 2^{1/6}\sigma$. Οι τιμές ε και σ σχετίζονται επίσης με τις παραμέτρους A και B της αρχικής εξίσωσης, μέσω των σχέσεων $A = 4\varepsilon\sigma^6$ και $B = 4\varepsilon\sigma^{12}$.

Οι ηλεκτροστατικές αλληλεπιδράσεις μεταξύ δύο φορτισμένων σωματιδίων i και j , με μερικά φορτία q_i και q_j , περιγράφονται χρησιμοποιώντας το νόμο του Coulomb

$$U_{elec} = \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (44)$$

όπου q_i και q_j είναι οι τιμές των μερικών φορτίων, r_{ij} είναι η απόσταση των σωματιδίων και ϵ_0 είναι η διηλεκτρική σταθερά του μέσου, η οποία περιγράφει τις ιδιότητες του περιβάλλοντος. Σε προσομοιώσεις και υπολογισμούς στους οποίους το περιβάλλον (π.χ. υδατικός διαλύτης) προτυποποιείται μοριακά η ϵ_0 παίρνει την τιμή 1 (διηλεκτρική σταθερά του κενού), καθώς θεωρείται πως η επίδραση του περιβάλλοντος γύρω από τα σωματίδια i και j μπορεί να περιγραφεί από την επίλυση του νόμου για τα φορτία των μορίων του διαλύτη. Αντίθετα, σε περιπτώσεις όπου ο διαλύτης δεν προτυποποιείται (εννοούμενος διαλύτης), ο γύρω χώρος αντιμετωπίζεται σαν ένα ηλεκτροστατικό συνεχές με ιδιότητες διαλύτη και η ϵ_0 λαμβάνει τιμές 78.0-80.0, οι οποίες αντιστοιχούν στη διηλεκτρική σταθερά των υδατικών διαλυμάτων.

Ασθενείς αλληλεπιδράσεις, τυπικά, σχηματίζονται ανάμεσα σε όλα τα πιθανά ζεύγη μη ομοιοπολικά συνδεδεμένων ατόμων. Ωστόσο, λόγω του μεγάλου αριθμού των σωματιδίων σε μια προσομοίωση Μοριακής Δυναμικής, ο υπολογισμός όλων των πιθανών ζευγαριών

Κεφάλαιο II – Υλικά & Μέθοδοι

αλληλεπιδράσεων είναι υπολογιστικά πολύ δύσκολος. Επιπρόσθετα, στις περιπτώσεις των ατόμων που βρίσκονται σε πολύ μικρή ή μεγάλη απόσταση μεταξύ τους, η συμβολή των μη ομοιοπολικών αλληλεπιδράσεων στη συνολική δυναμική ενέργεια του συστήματος είναι στην πραγματικότητα αμελητέα. Συγκεκριμένα, για άτομα τα οποία συμμετέχουν σε κοινές ομοιοπολικές αλληλεπιδράσεις (δεσμοί, γωνίες δεσμών, δίεδρες γωνίες), η ισχύς των ομοιοπολικών αλληλεπιδράσεων είναι κατά πολύ μεγαλύτερη από αυτή των μη ομοιοπολικών αλληλεπιδράσεων, με αποτέλεσμα οι τελευταίες να συνεισφέρουν ελάχιστα στη συνολική δυναμική ενέργεια. Έτσι, για την υπολογιστική βελτιστοποίηση των προσομοιώσεων, συνηθίζεται να μη λαμβάνονται υπόψη οι ασθενείς αλληλεπιδράσεις ομοιοπολικά συνδεδεμένων μεταξύ τους ατόμων. Συμβατικά, αυτό πραγματοποιείται για άτομα τα οποία απέχουν από έναν έως τέσσερις συνεχόμενους ομοιοπολικούς δεσμούς.

Αντίστοιχα, για τα άτομα τα οποία βρίσκονται σε μεγάλη απόσταση οι αλληλεπιδράσεις van der Waals προσεγγίζουν το μηδέν, με αποτέλεσμα να είναι αμελητέες. Έτσι, κατά την προσομοίωση ορίζονται τιμές κατωφλίου, πέρα από το οποίο ο υπολογισμός των αλληλεπιδράσεων μετριάζεται ώστε να προσεγγίζει το μηδέν (van der Spoel & van Maaren, 2006). Παραδοσιακά, αυτός ο κανόνας ακολουθείται τόσο για τις αλληλεπιδράσεις van der Waals όσο και για τις αλληλεπιδράσεις Coulomb. Ωστόσο, αυτή η αντιμετώπιση τα τελευταία χρόνια έχει τεθεί υπό αμφισβήτηση στην περίπτωση των ηλεκτροστατικών αλληλεπιδράσεων, καθώς έχει βρεθεί πως η εμβέλεια των δυνάμεων Coulomb είναι κατά πολύ μεγαλύτερη από ότι πιστευόταν αρχικά. Έτσι, η απενεργοποίηση των ηλεκτροστατικών υπολογισμών σε αυξημένες αποστάσεις έχει σαν μειονέκτημα το μη υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλης εμβέλειας.

Για την επίλυση αυτού του προβλήματος, η χρήση κατωφλίου για τις μη ομοιοπολικές αλληλεπιδράσεις συνοδεύεται από την εφαρμογή επιπλέον μεθόδων για την ορθή προτυποποίηση των ηλεκτροστατικών αλληλεπιδράσεων. Η πιο συχνά εφαρμοζόμενη προσέγγιση είναι η μέθοδος Particle Mesh Ewald (PME). Η PME βασίζεται στην αθροιστική μέθοδο του Ewald, η οποία τροποποιεί το νόμο του Coulomb εισάγοντας ένα νέο όρο απόστασης n

$$U_{elec} = \frac{1}{2} \sum_n \sum_i \sum_{j \neq i} \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij} + n|} \quad (45)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

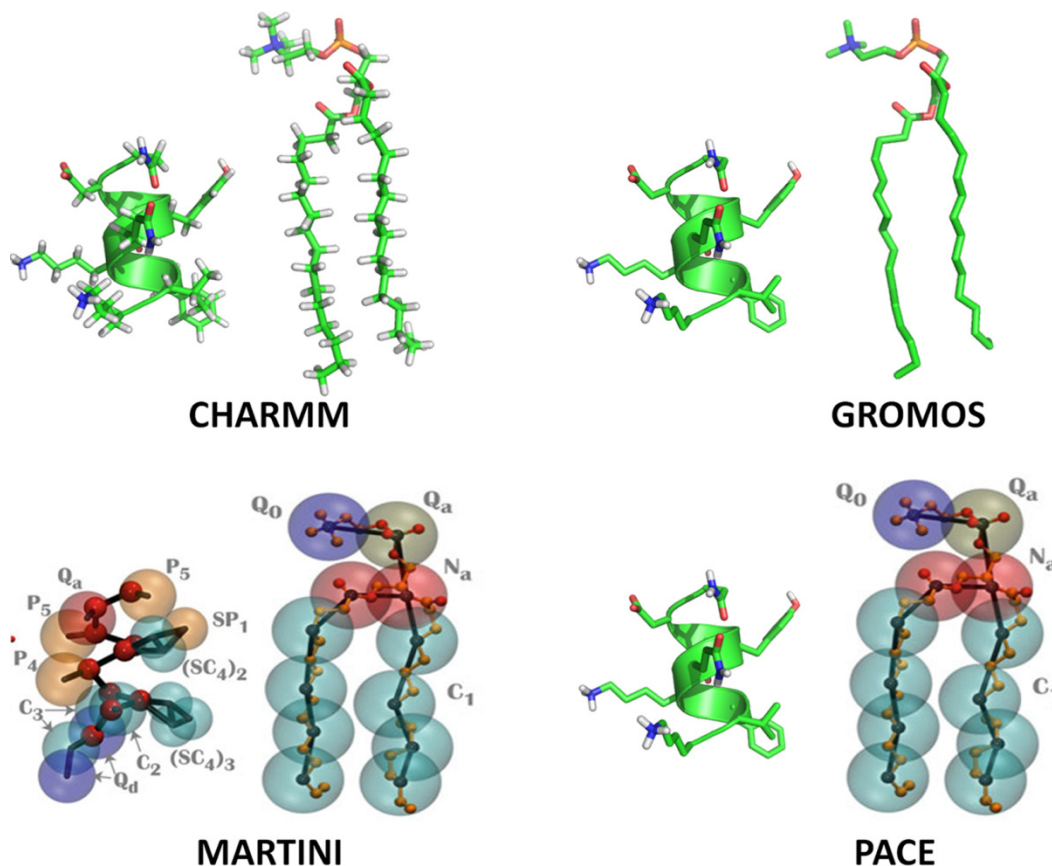
όπου το $n=(n_xL_x, n_yL_y, n_zL_z)$ με τις τιμές $L_{x,y,z}$ να αντιστοιχούν στις διαστάσεις του συστήματος και τις τιμές $n_{x,y,z}$ να περιγράφουν τη θέση κάθε σημείου στο πλέγμα της προσομοίωσης, όπως αυτό υλοποιείται κατά την εφαρμογή των συνθηκών περιοδικών ορίων. Η τροποποιημένη συνάρτηση της Εξίσωσης 45, λόγω της αυξημένης πολυπλοκότητάς της, συγκλίνει πολύ αργά στη μονάδα του χρόνου. Η επέκτασή της, μέσω της PME, λύνει αυτό το πρόβλημα χρησιμοποιώντας τον κλασικό νόμο του Coulomb για αλληλεπιδράσεις μικρής εμβέλειας, ενώ η διόρθωση του Ewald υπόκειται σε μετασχηματισμούς Fourier για τις αλληλεπιδράσεις μεγάλης εμβέλειας (Darden et al., 1993).

2.1.8 Δυναμικά πεδία και επίπεδα διακριτικότητας

Ανάλογα με το επίπεδο διακριτικότητας το οποίο χρησιμοποιούν για να περιγράψουν ένα σύστημα, τα δυναμικά πεδία μπορούν να διακριθούν σε διάφορες κατηγορίες, οι κυριότερες από τις οποίες είναι τα πεδία ενοποιημένων ατόμων, τα πεδία πλήρους ατομικής λεπτομέρειας και τα πεδία αδρής διακριτικότητας (Εικόνα 31). Τα πεδία ενοποιημένων ατόμων (united-atom), γνωστά και σαν πεδία πολικών υδρογόνων (polar hydrogen) περιγράφουν τα βιομόρια χρησιμοποιώντας πλήρεις συντεταγμένες για όλα τα βαριά άτομα, ωστόσο αγνοούν όλα τα άτομα υδρογόνου εκτός από αυτά των αμινομάδων της κεντρικής αλυσίδας των πρωτεϊνών και αυτά των πολικών ομάδων. Αυτή η απεικόνιση ακολουθείται από τις διάφορες παραλλαγές του δυναμικού πεδίου GROMOS (Huang et al., 2011). Τα πλήρως ατομικά (all-atom) δυναμικά πεδία όπως το CHARMM36 (Huang et al., 2017) και το AMBER (Salomon-Ferrer et al., 2013) περιγράφουν πλήρως τις συντεταγμένες των ατόμων, περιλαμβάνοντας και όλα τα άτομα υδρογόνου. Παρέχουν την καλύτερη δυνατή λεπτομέρεια, ωστόσο, αυτό έχει σαν τίμημα τις υψηλές απαιτήσεις σε υπολογιστική ισχύ και περιορισμούς τόσο στο χρόνο προσομοίωσης που δύναται να πραγματοποιηθεί όσο και στο μέγεθος του συστήματος που μπορεί να μελετηθεί. Τα πεδία αδρής διακριτικότητας (Coarse-Grained) επιτυγχάνουν προσομοιώσεις για πολύ μεγάλα χρονικά διαστήματα και ογκώδη συστήματα, απλοποιώντας την απεικόνιση των μορίων με τέτοιο τρόπο ώστε να επιταχύνεται η διαδικασία με τη μικρότερη δυνατή απώλεια πληροφορίας (Ingolfsson et al., 2016; Marrink et al., 2007). Τέλος,

Κεφάλαιο II – Υλικά & Μέθοδοι

μια νέα προσέγγιση στο χώρο των δυναμικών πεδίων αποτελούν τα μοντέλα υβριδικής διακριτικότητας (hybrid resolution), τα οποία συνδυάζουν διαφορετικά επίπεδα διακριτικότητας, χρησιμοποιώντας ατομική διακριτικότητα για τις πρωτεΐνες και αδρή διακριτικότητα για το διαλύτη (Han & Schulten, 2012).



Εικόνα 31 Κατηγορίες δυναμικών πεδίων. Το CHARMM είναι ένα πλήρως ατομικό (all-atom) δυναμικό πεδίο, στο οποίο προτυποποιείται το σύνολο των ατόμων. Το GROMOS είναι ένα δυναμικό πεδίο ενοποιημένων ατόμων (united-atom), στο οποίο προτυποποιούνται όλα τα βαριά άτομα, καθώς και τα άτομα υδρογόνου των πολικών ομάδων. Το MARTINI είναι ένα πεδίο αδρής διακριτικότητας (Coarse-Grained), στο οποίο τα άτομα ομαδοποιούνται σε σωματίδια αλληλεπίδρασης. Το PACE είναι ένα υβριδικό πεδίο, το οποίο συνδυάζει ένα πεδίο ενοποιημένων ατόμων για τις πρωτεΐνες με το MARTINI για τα λιπίδια και το διαλύτη.

Στα πλαίσια της παρούσας διδακτορικής διατριβής έχουν πραγματοποιηθεί προσομοιώσεις Μοριακής Δυναμικής χρησιμοποιώντας δυναμικά πεδία και από τις τέσσερις προαναφερόμενες κατηγορίες, τα πεδία CHARMM (Huang et al., 2017), GROMOS (Huang et al.,

Κεφάλαιο II – Υλικά & Μέθοδοι

2011), MARTINI (Marrink et al., 2007) και PACE (Han & Schulten, 2012). Τα βασικά χαρακτηριστικά τους παρουσιάζονται στις επόμενες υποενότητες, ενώ ιδιαίτερη έμφαση δίνεται στο πεδίο αδρής διακριτικότητας MARTINI το οποίο, τόσο λόγω της υπολογιστικής ευελιξίας που προσφέρει όσο και λόγω της ικανότητάς του να προσομοιώνει ρεαλιστικά βιολογικές μεμβράνες και διαμεμβρανικές πρωτεΐνες, χρησιμοποιήθηκε εκτεταμένα στις προσομοιώσεις της διδακτορικής διατριβής.

2.1.8.1 Το δυναμικό πεδίο πλήρους ατομικής διακριτικότητας CHARMM

Το CHARMM (Chemistry at Harvard Macromolecular Mechanics) είναι το πρώτο οργανωμένο δυναμικό πεδίο που αναπτύχθηκε για προσομοιώσεις Μοριακής Δυναμικής, καθώς και ένα από τα πιο γνωστά και ευρέως χρησιμοποιούμενα δυναμικά πεδία στις προσομοιώσεις βιομορίων όπως οι πρωτεΐνες και τα λιπίδια (Brooks et al., 2009). Το CHARMM αναπτύχθηκε αρχικά στο Πανεπιστήμιο του Harvard από την ομάδα του Martin Karplus, τόσο σαν δυναμικό πεδίο όσο και σαν ξεχωριστό πρόγραμμα προσομοιώσεων Μοριακής Δυναμικής (Brooks et al., 1983). Η πιο πρόσφατη έκδοση του πεδίου, η οποία χρησιμοποιείται στην παρούσα διδακτορική διατριβή είναι η έκδοση CHARMM36m (Huang et al., 2017), ένα πλήρως ατομικό πεδίο που περιλαμβάνει παραμέτρους για πρωτεΐνες, νουκλεϊκά οξέα, υδατάνθρακες, ανόργανα πολυμερή και ένα μεγάλο αριθμό λιπιδίων. Για την ακρίβεια το λιπιδικό πεδίο δυνάμεων του CHARMM είναι το πιο ποικιλόμορφο, παρέχοντας παραμέτρους για πάνω από 100 διαφορετικούς τύπους λιπιδίων. Η ευέλικτη φύση του πεδίου το καθιστά ιδιαίτερα εύκολο ως προς την υλοποίησή του, με αποτέλεσμα να είναι διαθέσιμο σε όλα σχεδόν τα υπάρχοντα προγράμματα Μοριακής Δυναμικής.

2.1.8.2 Το δυναμικό πεδίο ενοποιημένων ατόμων GROMOS

Το GROMOS (Groningen Molecular Simulations) είναι μια συλλογή από δυναμικά πεδία ενοποιημένων ατόμων (united-atom), καθώς και το όνομα μιας από τις πρώτες εφαρμογές προσομοιώσεων Μοριακής Δυναμικής που σχεδιάστηκαν. Αναπτύχθηκε αρχικά στο

Κεφάλαιο II – Υλικά & Μέθοδοι

Πανεπιστήμιο του Groningen από την ομάδα του Wilfred van Gunsteren (van Gunsteren & Berendsen, 1987). Τα δυναμικά πεδία του GROMOS έχουν σχεδιαστεί έτσι ώστε να αναπαράγουν επιτυχώς τις πειραματικά προσδιορισμένες ιδιότητες των αλκανίων και έχουν επεκταθεί ώστε να προτυποποιούν ορθά πρωτεΐνες, υδατάνθρακες, λιπίδια καθώς και πληθώρα φαρμάκων και λοιπών οργανικών ενώσεων. Η υλοποίηση του GROMOS υιοθετεί την προσέγγιση των πεδίων ενοποιημένων ατόμων, προτυποποιώντας ατομικά όλα τα βαριά άτομα των βιομορίων, καθώς και τα άτομα υδρογόνου των πολικών και φορτισμένων ομάδων, γεγονός που αυξάνει την υπολογιστική ευελιξία του. Παράλληλα, η παραμετροποίηση του πεδίου έχει γίνει με τέτοιο τρόπο ώστε οι ενεργειακές τιμές των αλληλεπιδράσεων, και κυρίως οι μη ομοιοπολικές αλληλεπιδράσεις, να υπολογίζονται σωστά παρά την έλλειψη πλήρους ατομικής διακριτικότητας (van Gunsteren et al., 1996). Έτσι, το GROMOS είναι ιδιαίτερα δημοφιλές στο σχεδιασμό και τον έλεγχο νέων φαρμάκων (Stroet et al., 2018). Στην παρούσα διδακτορική διατριβή χρησιμοποιείται η έκδοση GROMOS 54A7 του πεδίου (Huang et al., 2011), η οποία περιλαμβάνει παραμέτρους για πρωτεΐνες και μια σειρά από λιπίδια. Ωστόσο, σε σύγκριση με άλλα δυναμικά πεδία, το GROMOS διαθέτει πολύ λιγότερα λιπίδια και περιορίζεται κυρίως στις φωσφατιδυλοχολίνες, με τις άλλες κατηγορίες λιπιδίων να υποεκπροσωπούνται.

2.1.8.3 Το δυναμικό πεδίο αδρής διακριτικότητας MARTINI

Η αποτελεσματική προσομοίωση ενός μεμβρανικού συστήματος απαιτεί την ορθή προτυποποίηση των χαρακτηριστικών του. Αυτή περιλαμβάνει την κατάλληλη τοποθέτηση της πρωτεΐνης σε μια λιπιδική διπλοστιβάδα, αποτελούμενη από ένα ή περισσότερους τύπους λιπιδίων, η οποία είναι απαραίτητο να διαθέτει ικανό μέγεθος ώστε να καλύπτει αποτελεσματικά τον υποδοχέα, δίνοντας παράλληλα αρκετό περιθώριο ώστε να μην δημιουργούνται σφάλματα από αλληλεπιδράσεις λόγω περιοδικότητας. Επιπρόσθετα, απαιτείται η προσθήκη μορίων νερού και ιόντων εκατέρωθεν της μεμβράνης, τα οποία προσομοιώνουν το υδατικό περιβάλλον των εξωμεμβρανικών τμημάτων. Τα παραπάνω βήματα οδηγούν σε υπερβολικά ογκώδη συστήματα με πολύ μεγάλο αριθμό ατόμων που

Κεφάλαιο II – Υλικά & Μέθοδοι

συμμετέχουν σε μεγάλο αριθμό ομοιοπολικών και μη ομοιοπολικών αλληλεπιδράσεων. Έτσι, απαιτείται τόσο υψηλή υπολογιστική ισχύς όσο και μεγάλο χρονικό διάστημα για την προσομοίωση μερικών μόνο νανοδευτερολέπτων (ns) σε αυτά τα συστήματα (Marrink et al., 2007).

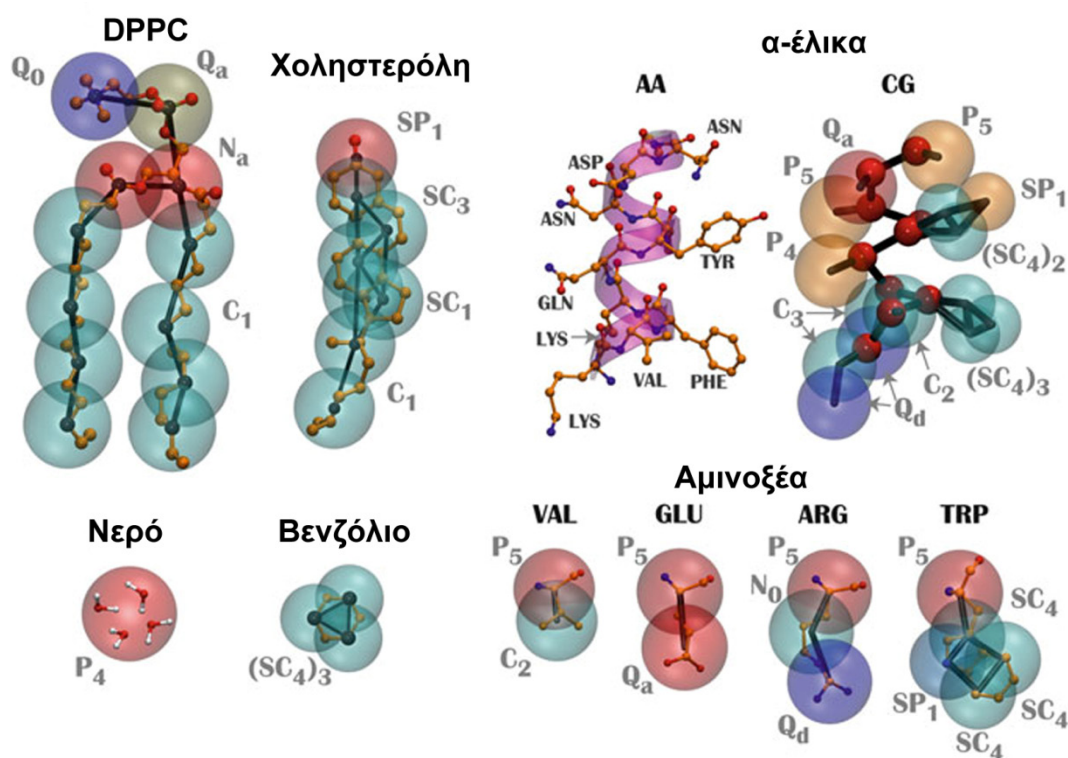
Μια λύση στο παραπάνω πρόβλημα αποτελούν οι προσομοιώσεις που χρησιμοποιούν δυναμικά πεδία αδρής διακριτικότητας (Coarse-Grained, CG). Σε αντίθεση με πεδία όπως το CHARMM και το GROMOS, τα πεδία CG δεν προσομοιώνουν τα βιομόρια με πλήρη λεπτομέρεια αλλά υιοθετούν πιο απλοποιημένες απεικονίσεις, στις οποίες τα άτομα που συμμετέχουν στο σχηματισμό μιας χημικής ομάδας με συγκεκριμένες ιδιότητες (π.χ. ο φώσφορος και τα οξυγόνα μιας φωσφορικής ομάδας) ομαδοποιούνται σε ένα σωματίδιο αλληλεπίδρασης το οποίο αντιπροσωπεύει αυτήν την ομάδα. Αυτή η μείωση του αριθμού των σωματιδίων συνοδεύεται και από αντίστοιχη μείωση του συνολικού αριθμού των βαθμών ελευθερίας στο σύστημα, καθιστώντας την προσομοίωσή του πιο ευέλικτη υπολογιστικά.

Ένα από τα πιο δημοφιλή και εξελιγμένα πεδία CG είναι το δυναμικό πεδίο MARTINI (de Jong et al., 2013; Marrink et al., 2007; Monticelli et al., 2008). Το MARTINI έχει σχεδιαστεί από στο πανεπιστήμιο του Groningen από την ομάδα του Siewert-Jan Marrink. Η αρχική μορφή του πεδίου σχεδιάστηκε για προσομοιώσεις λιπιδικών διπλοστιβάδων ενώ ακολούθησε η παραμετροποίηση και άλλων ειδών μακρομορίων, με το πεδίο δυνάμεων να προσφέρει πληροφόρία πλέον για λιπίδια, στερόλες, διάφορα είδη διαλύτη, πρωτεΐνες, υδατάνθρακες, νουκλεϊκά οξέα αλλά και ανόργανα πολυμερή. Όσον αφορά τις βιολογικές μεμβράνες, το MARTINI προσφέρει μεγάλη ποικιλία από παραμέτρους για λιπίδια με πάνω από 70 διαφορετικούς τύπους (Wassenaar et al., 2015) και υπολείπεται μόνο της λιπιδικής συλλογής του CHARMM.

Το μοντέλο του MARTINI χρησιμοποιεί ένα πρότυπο αντιστοιχίας 4:1 για την απεικόνιση των βιομορίων. Αυτό σημαίνει ότι κατά τη μετατροπή της ατομικής δομής σε δομή CG, τέσσερα βαριά άτομα (και τα αντίστοιχα άτομα υδρογόνου) αντιπροσωπεύονται από ένα σωματίδιο αλληλεπίδρασης (Marrink et al., 2007). Έτσι, για παράδειγμα, τα άτομα κάθε αμινοξικού καταλοίπου τα οποία συμμετέχουν στην κύρια ανθρακική αλυσίδα της πρωτεΐνης (N, Ca, C, O) αντιπροσωπεύονται από ένα κέντρο αλληλεπίδρασης ενώ οι πλευρικές ομάδες,

Κεφάλαιο II – Υλικά & Μέθοδοι

ανάλογα με το μέγεθός τους, μπορούν να αντιπροσωπεύονται από ένα ως τέσσερα σωματίδια. Αντίστοιχοι κανόνες ακολουθούνται και για άλλες κατηγορίες μορίων όπως τα φωσfolιπίδια. Για τη λεπτομερέστερη απεικόνιση πολλών μορίων, ωστόσο, ο κανόνας παρακάμπτεται. Έτσι, για μόρια με δακτυλίους, όπως οι στερόλες αλλά και οι πλευρικές ομάδες των αρωματικών αμινοξέων χρησιμοποιείται η αντιστοιχία 3:1 ή ακόμα και 2:1 για την περιγραφή των ατόμων. Αυτή η εξαίρεση οδηγεί σε πιο ογκώδεις αναπαραστάσεις για αυτές τις ομάδες, ωστόσο προσφέρει το πλεονέκτημα της λεπτομερέστερης απεικόνισης (Εικόνα 32).



Εικόνα 32 Το δυναμικό πεδίο MARTINI. Παρουσιάζεται η αντιστοίχιση ανάμεσα σε ατομική (AA) και αδρή διακριτικότητα (CG) για το φωσfolιπίδιο διπαλμιτοϋλ-φωσφατιδυλοχολίνη (DPPC), τη χοληστερόλη, ένα πεπτίδιο με δομή α-έλικας, το νερό, το βενζόλιο και τέσσερα αντιπροσωπευτικά αμινοξέα (βαλίνη, γλουταμικό, αργινίνη και τρυπτοφάνη). Τα σωματίδια CG παρουσιάζονται σαν ημιδιαφανείς σφαίρες van der Waals και σημαίνονται με βάση τις φυσικοχημικές ιδιότητές τους. Προσαρμογή από Periole & Marrink, 2013.

Τα σωματίδια αλληλεπίδρασης που απεικονίζουν ένα μακρομόριο συνδέονται μεταξύ τους ομοιοπολικά με βάση τη γεωμετρία του αρχικού, ατομικής διακριτικότητας μορίου και χαρακτηρίζονται από δυναμικά μήκους δεσμών, γωνιών δεσμών και, όπου είναι απαραίτητο,

Κεφάλαιο II – Υλικά & Μέθοδοι

δίδει γωνιών. Επιπρόσθετα, στην περίπτωση των πρωτεϊνών, οι ιδιότητες των σωματιδίων που απεικονίζουν κάθε αμινοξικό κατάλοιπο καθορίζονται με βάση τη δευτεροταγή δομή του στην αρχική ατομική δομή. Έτσι, κατάλοιπα που συμμετέχουν σε α-έλικες, β-κλώνους ή ακανόνιστες δομές απεικονίζονται διαφορετικά στο πεδίο δυνάμεων (Monticelli et al., 2008; Periole et al., 2009). Με αυτές τις παραδοχές, το MARTINI επιτρέπει μεγαλύτερη λεπτομέρεια στην προσομοίωση των ιδιοτήτων των βιολογικών συστημάτων από άλλα πεδία αδρής διακριτικότητας καθώς, παρά την απλοποίηση της απεικόνισής του, τα βιομόρια διατηρούν χαρακτηριστικά τα οποία αντιπροσωπεύουν τις χημικές ιδιότητές τους. Οι ομοιοπολικές αλληλεπιδράσεις προτυποποιούνται χρησιμοποιώντας τα κλασικά αρμονικά δυναμικά των δυναμικών πεδίων όπως αυτά περιγράφηκαν στις προηγούμενες ενότητες, με μόνη εξαίρεση την προτυποποίηση του δυναμικού της γωνίας δεσμών, για το οποίο εφαρμόζεται το συνημιτονοειδές δυναμικό.

Όσον αφορά τις μη ομοιοπολικές αλληλεπιδράσεις, το MARTINI καθορίζει διαφορετικά είδη κέντρων αλληλεπίδρασης για να περιγράψει τα σωματίδια των διαφορετικών ομάδων με βάση τις ιδιότητές τους. Ανάλογα με τη φύση τους, τα σωματίδια μπορούν να διακριθούν σε πολικά (P), μη πολικά (C), ενδιάμεσα (N), φορτισμένα (Q) και αρωματικά (SP, SN και SC) σωματίδια, με κάθε τύπο να διακρίνεται περαιτέρω σε υποτύπους, επιτρέποντας έτσι την ακριβή απεικόνιση της χημικής φύσης της αρχικής ατομικής δομής (για παράδειγμα, ο υποτύπος P4 θεωρείται ότι αντιπροσωπεύει πιο πολικές ομάδες από τον υποτύπο P1). Οι μη ομοιοπολικές αλληλεπιδράσεις προτυποποιούνται μέσω του δυναμικού Lennard – Jones και του νόμου του Coulomb για τις αλληλεπιδράσεις van der Waals και το ηλεκτροστατικό δυναμικό αντίστοιχα, εφαρμόζοντας επιπλέον συναρτήσεις μετασχηματισμού και ανάλογες τιμές κατωφλίου για τη χρήση τους με βάση τη φύση των σωματιδίων (Marrink et al., 2007).

Η ικανότητα του πεδίου MARTINI στην ορθή προσομοίωση βιολογικών συστημάτων έχειδειχθεί επανειλημμένα μέσα από μια σειρά εφαρμογών σε πρωτεΐνες, λιπιδικές μεμβράνες, νουκλεϊκά οξέα αλλά και ογκώδη υπερμοριακά συμπλέγματα όπως λιπιδικά κυστίδια, μικρά οργανίδια και ολόκληροι ιοί (de Jong et al., 2013; Ingolfsson et al., 2016; Ingolfsson et al., 2014; Wassenaar et al., 2015). Αξιοσημείωτη περίπτωση, μάλιστα, αποτελούν τα συστήματα τα οποία περιλαμβάνουν βιολογικές μεμβράνες, στα οποία το πεδίο δυνάμεων

Κεφάλαιο II – Υλικά & Μέθοδοι

αποδεικνύεται ιδιαίτερα ικανό, πιθανώς λόγω του ότι η αρχική μορφή του σχεδιάστηκε ειδικά για την προσομοίωση λιπιδικών διπλοστιβάδων. Η υπολογιστική ευελιξία του πεδίου, η οποία επιτρέπει τη χρήση βημάτων χρόνου της τάξης των 10-40 fs οδηγεί σε σημαντική επιτάχυνση των προσομοιώσεων σε σύγκριση με τις προσομοιώσεις ατομικής διακριτικότητας, στις οποίες εφαρμόζονται βήματα χρόνου της τάξης των 1-2 fs. Έτσι μπορούν να πραγματοποιηθούν προσομοιώσεις χρόνων της τάξης των δεκάδων ή εκατοντάδων μs ή ακόμη και ms, οι οποίες είναι σχεδόν ακατόρθωτες με δυναμικά πεδία ατομικής διακριτικότητας (Periole & Marrink, 2013). Τέλος, ένα ιδιαίτερο χαρακτηριστικό του πεδίου το οποίο προκύπτει από τον τρόπο ομαδοποίησης των ατόμων σε συνδυασμό με την υλοποίηση των μη ομοιοπολικών αλληλεπιδράσεων είναι η αυξημένη ταχύτητα της διάχυσης των μορίων κατά την προσομοίωση. Συγκεκριμένα, μετρήσεις του συντελεστή διάχυσης για το νερό καθώς και για φωσφολιπίδια έχουν δείξει ότι η κίνηση των μορίων σε μια προσομοίωση MARTINI γίνεται τέσσερις φορές πιο γρήγορα από ότι σε μια ατομική προσομοίωση, με αντίστοιχες τιμές συντελεστών διάχυσης να έχουν μετρηθεί και για την κίνηση μεμβρανικών πρωτεϊνών (Marrink et al., 2007; Monticelli et al., 2008). Έτσι, στη μελέτη φαινομένων στα οποία η διάχυση διαδραματίζει σημαντικό ρόλο, όπως οι πλευρικές κινήσεις των λιπιδίων, η ροή του νερού και των ιόντων μέσα από πόρους αλλά και ο αυθόρμητος σχηματισμός υπερμοριακών συμπλεγμάτων από πρωτεΐνες, οι προσομοιώσεις του MARTINI αναμένεται να προτυποποιούν τα παραπάνω φαινόμενα τέσσερις φορές πιο γρήγορα από ότι οι ατομικές προσομοιώσεις (π.χ. μια προσομοίωση αυτού του τύπου με το MARTINI αναμένεται να δώσει σε χρόνο 1 μs το ίδιο αποτέλεσμα που μια αντίστοιχη προσομοίωση με το CHARMM θα έδινε σε χρόνο 4 μs).

2.1.8.4 Το υβριδικό δυναμικό πεδίο PACE

Ένας από τους βασικότερους περιορισμούς των πεδίων αδράς διακριτικότητας όπως το MARTINI είναι η απώλεια λεπτομέρειας στην απεικόνιση της πρωτεϊνικής δομής. Συγκεκριμένα, η ομαδοποίηση των ατόμων της κεντρικής ανθρακικής αλυσίδας σε σωματίδια CG, και η αντίστοιχη μείωση των βαθμών ελευθερίας οδηγούν στην απώλεια σημαντικών χαρακτηριστικών της δευτεροταγούς και τριτοταγούς δομής των πρωτεϊνών, όπως οι δεσμοί

Κεφάλαιο II – Υλικά & Μέθοδοι

υδρογόνου και οι διέδρες γωνίες φ και ψ . Τα πεδία CG επιλύουν αυτό το πρόβλημα χρησιμοποιώντας πληροφορίες από την αρχική δομή ατομικής διακριτικότητας για να παραμετροποιήσουν το σύστημα αδρής διακριτικότητας. Για παράδειγμα, στο πεδίο MARTINI, οι ιδιότητες των σωματιδίων κάθε καταλοίπου καθορίζονται με βάση τη δευτεροταγή δομή του (Monticelli et al., 2008). Ωστόσο, αυτό οδηγεί στον περιορισμό του συστήματος στην αρχική στερεοδιάταξή του. Έτσι, ενώ τα πεδία CG είναι κατάλληλα για τη μελέτη συστημάτων μεγάλης κλίμακας και για την προτυποποίηση διαμοριακών αλληλεπιδράσεων ανάμεσα σε πλήρως διπλωμένες πρωτεΐνες, δεν μπορούν να χρησιμοποιηθούν στη μελέτη συστημάτων χωρίς καθορισμένη δομή, όπως οι εγγενώς μη δομημένες πρωτεΐνες, ή στη μελέτη φαινομένων στα οποία λαμβάνουν χώρα σημαντικές στερεοδιαταξικές αλλαγές, όπως η διαδικασία του διπλώματος των πρωτεϊνών.

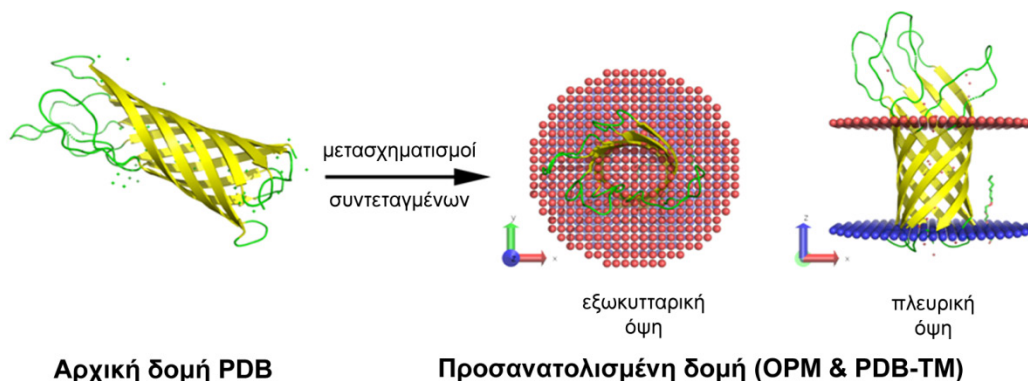
Ως λύση στα παραπάνω προβλήματα έχει προταθεί η χρήση δυναμικών πεδίων υβριδικής διακριτικότητας. Πεδία αυτού του τύπου λειτουργούν συνδυάζοντας την ατομική απεικόνιση για την προτυποποίηση των πρωτεϊνών με μοντέλα αδρής διακριτικότητας για την απεικόνιση του διαλύτη. Αυτός ο συνδυασμός επιτρέπει ουσιαστικά τη διεξαγωγή προσομοιώσεων οι οποίες, αφενός έχουν την ταχύτητα και υπολογιστική ευελιξία των προσομοιώσεων CG, αφετέρου επιτρέπουν τη λεπτομερή προτυποποίηση των πρωτεϊνών και την παρατήρηση στερεοδιαταξικών αλλαγών. Συνεπώς, τα υβριδικά δυναμικά πεδία μπορούν να βρουν εφαρμογή στη μελέτη β-δομών όπως τα β-βαρέλια, καθώς και στη μελέτη του διπλώματος των πρωτεϊνών. Στην παρούσα διδακτορική διατριβή έχει χρησιμοποιηθεί το πεδίο υβριδικής διακριτικότητας PACE (Protein Atoms in Coarse Environment), το οποίο αναπτύχθηκε αρχικά από την ομάδα του Yun-Dong Wu στο Πανεπιστήμιο του Πεκίνου, και βελτιστοποιήθηκε περαιτέρω από την ομάδα του Klaus Schulten στο Πανεπιστήμιο του Illinois (Han & Schulten, 2012; Wan et al., 2012). Το PACE χρησιμοποιεί την απεικόνιση ενοποιημένων ατόμων για την προτυποποίηση των πρωτεϊνών στα πρότυπα του GROMOS, ενώ ο διαλύτης και τα λιπίδια προτυποποιούνται χρησιμοποιώντας το πεδίο MARTINI. Η συνδυασμένη χρήση των δύο διαφορετικών επιπέδων διακριτικότητας επιτυγχάνεται με την χρήση πρόσθετων παραμέτρων Lennard-Jones για την αλληλεπίδραση των πρωτεϊνικών ατόμων με τα σωματίδια CG, οι οποίες έχουν οριστεί με βάση πειραματικές μετρήσεις του συντελεστή επιμερισμού μιας

Κεφάλαιο II – Υλικά & Μέθοδοι

ευρείας σειράς χημικών ομάδων σε πολικά και μη πολικά διαλύματα. Η ορθότητα του PACE έχει ελεγχθεί μέσα από επιτυχείς προσομοιώσεις πρωτεϊνικού διπλώματος για μια σειρά από πεπτίδια και μικρές πρωτεΐνες με γνωστή δομή (Han & Schulten, 2012). Επιπρόσθετα, το δυναμικό πεδίο έχει παραμετροποιηθεί και για τη μελέτη διαμεμβρανικών πρωτεϊνών σε περιβάλλον λιπιδικής διπλοστιβάδας (Qi et al., 2014; Wan et al., 2012), παρουσιάζοντας αυξημένη ταχύτητα σε συνδυασμό με ατομική λεπτομέρεια.

2.1.9 Προσομοιώσεις Μοριακής Δυναμικής για διαμεμβρανικές πρωτεΐνες

Ο σχεδιασμός μιας προσομοίωσης Μοριακής Δυναμικής για ένα σύστημα βιομορίων, όπως είναι μια διαμεμβρανική πρωτεΐνη στο φυσιολογικό περιβάλλον της, είναι μια διαδικασία που εξαρτάται σε μεγάλο βαθμό τόσο από τις ιδιότητες του εκάστοτε συστήματος όσο και από το είδος των ερωτημάτων που τίθενται από τον ερευνητή. Ωστόσο, όλες οι προσομοιώσεις διαμεμβρανικών πρωτεϊνών σε περιβάλλον λιπιδικής διπλοστιβάδας ακολουθούν, γενικά, μια κοινή πορεία εργασίας. Το πρώτο βήμα στην προετοιμασία ενός συστήματος προσομοίωσης είναι η κατάλληλη επεξεργασία της δομής της πρωτεΐνης. Αυτή η μπορεί να είναι κάποια πειραματικά προσδιορισμένη δομή από την PDB, λυμένη με κρυσταλλογραφία ακτίνων-Χ, κρυο-ηλεκτρονική μικροσκοπία ή NMR, να έχει προκύψει από υπολογιστικές μεθόδους όπως η ομόλογη προτυποποίηση ή να έχει κατασκευαστεί με κάποιο συνδυασμό των παραπάνω (π.χ. μια κρυσταλλική δομή, στην οποία τμήματα που απουσιάζουν προτυποποιούνται με ομόλογη προτυποποίηση). Επιπρόσθετα, στην περίπτωση των μεμβρανικών πρωτεϊνών, είναι απαραίτητο αυτή η δομή να διαθέτει τον κατάλληλο προσανατολισμό σε σχέση με το επίπεδο της μεμβράνης, προκειμένου να οριστεί σωστά η θέση της στη διπλοστιβάδα κατά την κατασκευή του συστήματος. Συμβατικά, στις προσομοιώσεις των βιολογικών μεμβρανών ορίζεται ότι το επίπεδο της διπλοστιβάδας είναι παράλληλο με το επίπεδο των αξόνων X και Y και κατακόρυφο ως προς τον άξονα Z και το γεωμετρικό κέντρο της μεμβράνης τοποθετείται στην αρχή των αξόνων (0,0,0). Έτσι, η δομή μιας διαμεμβρανικής πρωτεΐνης πρέπει να προσανατολιστεί κατάλληλα, ώστε να ακολουθεί τους παραπάνω κανόνες (Εικόνα 33).



Εικόνα 33 Προσανατολισμός μιας διαμεμβρανικής πρωτεΐνης στο επίπεδο της μεμβράνης. Παρουσιάζεται το παράδειγμα του διαμεμβρανικού β-βαρελίου OmpA. Οι συντεταγμένες της πρωτεΐνης υπόκεινται σε μετασχηματισμούς, έτσι ώστε ο κατακόρυφος κύριος άξονάς της να είναι παράλληλος με τον άξονα Z και κατακόρυφος στο επίπεδο της διπλοστιβάδας. Το γεωμετρικό κέντρο της μεμβράνης ορίζεται στην αρχή των αξόνων. Κατάλληλα προσανατολισμένες δομές υπάρχουν διαθέσιμες στις βάσεις δεδομένων OPM και PDB-TM.

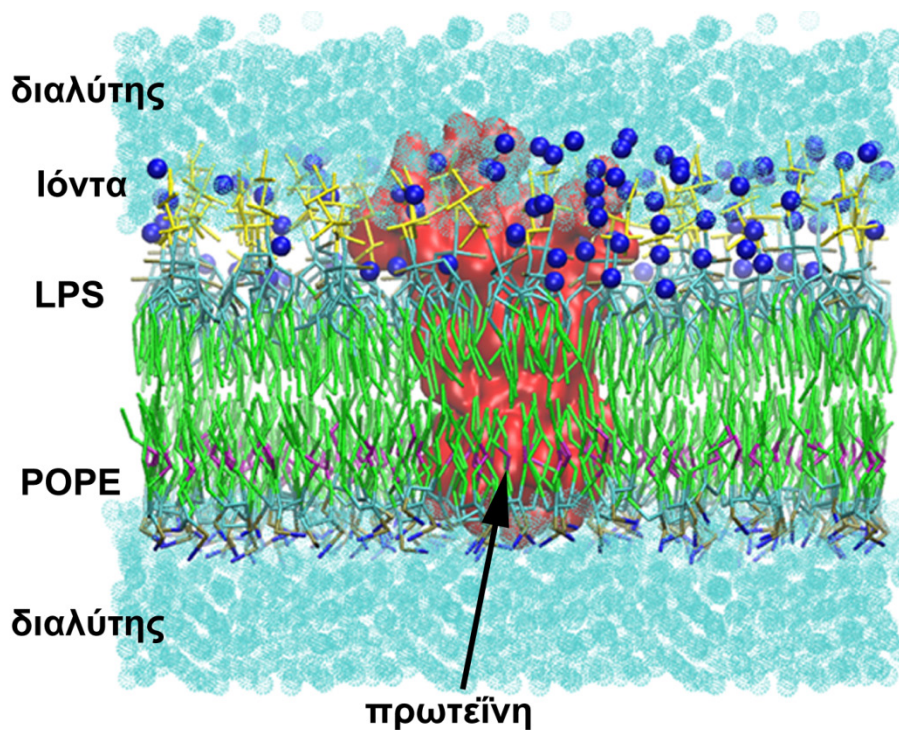
Κατάλληλα προσανατολισμένες δομές PDB σε σχέση με τη μεμβράνη μπορούν να βρεθούν στις βάσεις δεδομένων PDB-TM (Kozma et al., 2013; Tusnady et al., 2004) και OPM (Lomize et al., 2012), τις δύο μεγαλύτερες βάσεις δεδομένων για δομές μεμβρανικών πρωτεϊνών. Η βάση PDB-TM (<http://pdbtm.enzim.hu/>) διαθέτει προσανατολισμένες δομές PDB αποκλειστικά για διαμεμβρανικές πρωτεΐνες, ενώ η βάση OPM (<https://opm.phar.umich.edu/>) διαθέτει προσανατολισμένες δομές για διαμεμβρανικές, περιφερειακές και αγκυροβολημένες μεμβρανικές πρωτεΐνες. Επιπρόσθετα, η OPM προσφέρει στο χρήστη τη δυνατότητα να προσανατολίσει τις δομές της επιλογής του (π.χ. κάποιο θεωρητικό μοντέλο) σε σχέση με τη μεμβράνη, χρησιμοποιώντας τη μέθοδο PPM (Positioning of Proteins in the Membrane), η οποία χρησιμοποιείται για τον προσανατολισμό των δομών που περιέχονται στην ίδια τη βάση δεδομένων (Lomize et al., 2012).

Έχοντας αποκτήσει μια κατάλληλα προσανατολισμένη πρωτεϊνική δομή, το επόμενο βήμα είναι η επιλογή του κατάλληλου δυναμικού πεδίου και η περαιτέρω επεξεργασία της πρωτεΐνης με βάση αυτό. Αυτή η επεξεργασία περιλαμβάνει την προσθήκη ατόμων υδρογόνου, όπου χρειάζεται (όλα τα άτομα υδρογόνου για τα πλήρως ατομικά πεδία ή μόνο τα άτομα των πολικών/φορτισμένων ομάδων για τα πεδία ενοποιημένων ατόμων), τη διόρθωση τυχόν σφαλμάτων στη δομή και, τέλος, την δημιουργία της τοπολογίας της

Κεφάλαιο II – Υλικά & Μέθοδοι

πρωτεΐνης, δηλαδή τον καθορισμό του συνόλου των ομοιοπολικών και μη ομοιοπολικών αλληλεπιδράσεων με βάση το επιλεγμένο δυναμικό πεδίο. Επιπρόσθετα, στην περίπτωση που χρησιμοποιείται κάποιο δυναμικό πεδίο CG όπως το MARTINI, στην παραπάνω διαδικασία περιλαμβάνεται και η μετατροπή της δομής από μοντέλο ατομικής διακριτικότητας σε μοντέλο CG, με βάση τους κανόνες του δυναμικού πεδίου.

Το τρίτο βήμα στο σχεδιασμό της προσομοίωσης είναι η κατασκευή του περιβάλλοντος γύρω από την πρωτεΐνη. Αυτή περιλαμβάνει την προτυποποίηση της λιπιδικής διπλοστιβάδας και το υδατικού διαλύτη σε κάθε πλευρά της μεμβράνης (Εικόνα 34). Η λιπιδική διπλοστιβάδα μπορεί να είναι είτε ομοιογενής, αποτελούμενη από έναν τύπο λιπιδίου, είτε ετερογενής, αποτελούμενη από περισσότερους του ενός τύπους λιπιδίων σε διάφορες συγκεντρώσεις. Ο διαλύτης σε κάθε πλευρά της μεμβράνης είναι υδατικής φύσεως, αποτελούμενος από μόρια νερού και ιόντα (NaCl, CaCl₂ κλπ). Η κατασκευή του περιβάλλοντος γύρω από την πρωτεΐνη συνοδεύεται από την αντίστοιχη τροποποίηση της τοπολογίας του συστήματος, με την προσθήκη παραμέτρων για κάθε συστατικό του περιβάλλοντος. Το σύνολο της παραπάνω διαδικασίας μπορεί να πραγματοποιηθεί είτε χειροκίνητα είτε αυτοματοποιημένα, μέσω κάποιου εργαλείου για την προετοιμασία συστημάτων προσομοίωσης. Το πιο διαδεδομένο εργαλείο στην τελευταία κατηγορία είναι η διαδικτυακή υπηρεσία CHARMM-GUI (<http://www.charmm-gui.org>), η οποία προσφέρει τη δυνατότητα κατασκευής συστημάτων για προσομοιώσεις Μοριακής Δυναμικής με το δυναμικό πεδίο CHARMM (Jo et al., 2008; Lee et al., 2018). Ένα δεύτερο παράδειγμα είναι το πρόγραμμα INSANE («Insert Membrane»), το οποίο επιτρέπει την κατασκευή μεμβρανικών συστημάτων προσομοίωσης για το δυναμικό πεδίο MARTINI (Wassenaar et al., 2015). Και τα δύο εργαλεία έχουν χρησιμοποιηθεί στην προετοιμασία μεμβρανικών συστημάτων στην παρούσα διδακτορική διατριβή. Τέλος, σημειώνεται ότι στα πλαίσια της διατριβής έχει αναπτυχθεί το εργαλείο GNOMM, μια αυτοματοποιημένη μέθοδος για την κατασκευή μεμβρανικών συστημάτων προσομοίωσης για αρνητικά κατά Gram βακτήρια.



Εικόνα 34 Η δομή ενός συστήματος προσομοιώσεων Μοριακής Δυναμικής. Παρουσιάζεται το διαμεμβρανικό β-βαρέλι OmpA, τοποθετημένο σε μια λιπιδική διπλοστιβάδα με λιποπολυσακχαρίτες (LPS) και φωσfolιπίδια (POPE). Το σύστημα έχει προτυποποιηθεί με το δυναμικό πεδίο MARTINI.

Το τελικό σύστημα της πρωτεΐνης, της μεμβράνης και του διαλύτη, περιλαμβάνει τόσο τη δομή όσο και την τοπολογία του συστήματος και είναι έτοιμο να υποβληθεί σε προσομοιώσεις. Η διαδικασία της προσομοίωσης του συστήματος χωρίζεται, γενικά, σε τρία στάδια. Στο πρώτο στάδιο πραγματοποιείται εκτεταμένη ελαχιστοποίηση ενέργειας, προκειμένου να διορθωθούν τυχόν σφάλματα στη δομή (π.χ. στερεοχημικές παρεμποδίσεις) και το σύστημα να αποκτήσει τη βέλτιστη δυνατή ενεργειακή κατάσταση. Ακολουθεί ένα στάδιο προκαταρκτικών προσομοιώσεων, γνωστό με το γενικό όρο «εξισορρόπηση» (equilibration), κατά το οποίο πραγματοποιούνται προσομοιώσεις οι οποίες έχουν σαν στόχο να φέρουν σε κατάσταση ισορροπίας το περιβάλλον του συστήματος, τόσο ως προς τη θερμοκρασία και την πίεση όσο και ως προς τη φύση του διαλύτη και της μεμβράνης. Έτσι, το στάδιο της εξισορρόπησης περιλαμβάνει μια σειρά από προσομοιώσεις NVT και NPT, στις οποίες η θερμοκρασία και η πίεση ωθούνται σταδιακά στις τιμές ισορροπίας τους (τυπικά 300-310 K και 1 atm, αντίστοιχα), ενώ εφαρμόζονται περιορισμοί θέσης στα άτομα της πρωτεΐνης

Κεφάλαιο II – Υλικά & Μέθοδοι

και, συχνά, στα άτομα των πολικών κεφαλών των λιπιδίων. Οι παραπάνω περιορισμοί διατηρούν σταθερή τη θέση της πρωτεΐνης και των λιπιδικών κεφαλών στο σύστημα, επιτρέποντας την ελεύθερη κίνηση του διαλύτη και των λιπιδικών ουρών προκειμένου να αποκτήσουν τη ρευστή κατάσταση που το νερό και η μεμβράνη αναμένονται να έχουν *in vitro* και *in vivo*. Το τελικό στάδιο, γνωστό και σαν φάση παραγωγής της προσομοίωσης, ακολουθεί εφόσον το σύστημα έρθει σε κατάσταση θερμοδυναμικής ισορροπίας, και είναι ουσιαστικά η προσομοίωση Μοριακής Δυναμικής που θα χρησιμοποιηθεί για τη μελέτη του συστήματος. Κατά την φάση παραγωγής αφαιρούνται όλοι οι περιορισμοί που εφαρμόστηκαν στο στάδιο της εξισορρόπησης και το σύστημα αφήνεται πλήρως ελεύθερο να κινηθεί. Η πορεία της προσομοίωσης παραγωγής χρησιμοποιείται για την ανάλυση των αποτελεσμάτων, μέσα από τη μέτρηση μιας σειράς ιδιοτήτων, παραδείγματα των οποίων παρουσιάζονται στην επόμενη ενότητα.

Ένα πολύ σημαντικό κομμάτι στο σχεδιασμό και τη διεξαγωγή των προσομοιώσεων Μοριακής Δυναμικής είναι η χρήση των κατάλληλων αλγορίθμων και παραμέτρων για τον υπολογισμό των ιδιοτήτων του συστήματος (Πίνακας 5). Συγκεκριμένα, είναι απαραίτητο σε κάθε προσομοίωση να ορίζονται τα όρια στα οποία υπολογίζονται οι μη ομοιοπολικές αλληλεπιδράσεις ανάμεσα στα σωματίδια του συστήματος, σύμφωνα με τους κανόνες που παρουσιάζονται στην Ενότητα 2.1.7.2, καθώς επίσης και το αν θα χρησιμοποιηθεί κάποια μέθοδος για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλης εμβέλειας (π.χ. PME). Οι παραπάνω παράμετροι δεν είναι καθολικές, αλλά εξαρτώνται από το δυναμικό πεδίο που χρησιμοποιείται κάθε φορά στην προσομοίωση, με τα διαφορετικά δυναμικά πεδία να προτείνουν τη χρήση διαφορετικών τιμών για τα όρια του υπολογισμού των αλληλεπιδράσεων.

Κεφάλαιο II – Υλικά & Μέθοδοι

Πίνακας 5 Παράμετροι προσομοιώσεων Μοριακής Δυναμικής για τα δυναμικά πεδία CHARMM, GROMOS, MARTINI και PACE, οι οποίες χρησιμοποιήθηκαν στα πλαίσια της παρούσας διδακτορικής διατριβής.

Παράμετροι	CHARMM	GROMOS	MARTINI	PACE
Βήμα χρόνου	1-2 fs	1-2 fs	10-30 fs	5 fs
Θερμοστάτης	Berendsen	Berendsen	V-rescale	Berendsen
(προσ. εξισορρόπησης)	Duv. Langevin			Duv. Langevin
Θερμοστάτης	Nosé-Hoover	Nosé-Hoover	V-rescale	Nosé-Hoover
(προσ. παραγωγής)	Duv. Langevin			Duv. Langevin
Βαροστάτης	Berendsen	Berendsen	Berendsen	Berendsen
(προσ. εξισορρόπησης)	Langevin Piston			Langevin Piston
Βαροστάτης	Par.-Rahman	Par.-Rahman	Par.-Rahman	Par.-Rahman
(προσ. παραγωγής)	Langevin Piston			Langevin Piston
Όρια μη-ομοιοπολικών	LJ: 10-12	LJ: 10-14	LJ: 9-12	LJ: 9-12
αλ/σεων (Å)	Coulomb: 10-12	Coulomb: 10-14	Coulomb: 0-12	Coulomb: 0-12
Ηλεκτροστατικές Αλ/σεις	PME	PME	Όχι	Όχι
μεγάλης εμβέλειας				

Πέρα από τα όρια των αλληλεπιδράσεων, μία επιπρόσθετη παράμετρος η οποία πρέπει να καθοριστεί σε μια προσομοίωση είναι η τιμή του χρόνου βήματος (dt). Ο σωστός ορισμός του dt είναι ιδιαίτερα σημαντικός, καθώς καθορίζει τόσο τον υπολογισμό των τιμών της θέσης και της ταχύτητας και, κατά συνέπεια, των εφαρμοζόμενων δυνάμεων μέσω της διαδικασίας της χρονικής διακριτοποίησης, όσο και την ταχύτητα της προσομοίωσης. Παρά τη σημασία του, δυστυχώς δεν υπάρχει αποτελεσματικός τρόπος για τον ακριβή υπολογισμό της τιμής του dt που αρμόζει σε κάθε προσομοίωση. Αντίθετα, η τιμή ανατίθεται από το χρήστη εμπειρικά, με βάση μια σειρά από χαρακτηριστικά του συστήματος, το κυριότερο των οποίων είναι η ταχύτητα των θερμικών κινήσεων. Γενικά θεωρείται πως η τιμή του dt πρέπει να είναι 5 ως 10 φορές πιο μικρή από την περίοδο της πιο γρήγορης ταλάντωσης που παρατηρείται σε ένα σύστημα. Στην περίπτωση των χημικών και των βιομοριακών συστημάτων που προτυποποιούνται σε ατομικό επίπεδο διακριτικότητας, οι πιο γρήγορες ταλαντώσεις τυπικά αντιστοιχούν στις δονήσεις των ομοιοπολικών δεσμών ανάμεσα σε άτομα άνθρακα και άτομα υδρογόνου, η περίοδος των οποίων είναι της τάξης των ~ 10 fs, με αποτέλεσμα οι τυπικές τιμές του χρόνου βήματος στις προσομοιώσεις ατομικής λεπτομέρειας να είναι $dt=1.0-2.0$ fs. Έτσι, μια προσομοίωση 1 ns με χρόνο βήματος 1 fs θα χρειαζόταν 10^6 επαναλαμβανόμενες επιλύσεις της εξίσωσης κίνησης για να πραγματοποιηθεί ($1 \text{ ns} = 10^6 \text{ fs}$). Από την άλλη, σε

Κεφάλαιο II – Υλικά & Μέθοδοι

προσομοιώσεις που πραγματοποιούνται σε αδρό (Coarse-Grained) επίπεδο διακριτικότητας, η κοινή απεικόνιση πολλών ατόμων από μικρότερο αριθμό σωματιδίων αλληλεπίδρασης οδηγεί σε συστήματα με πιο αργές ταλαντώσεις, γεγονός που επιτρέπει την εφαρμογή μεγαλύτερου χρόνου βήματος. Έτσι, οι προσομοιώσεις αδρών συστημάτων μπορούν να έχουν χρόνο βήματος $dt=10-40$ fs ή ακόμα μεγαλύτερο, οδηγώντας σε αυξημένη ταχύτητα και κατά συνέπεια προσομοίωση μεγαλύτερου χρονικού διαστήματος. Γενικά, οι προσομοιώσεις ατομικής ή σχεδόν ατομικής διακριτικότητας με δυναμικά πεδία όπως το CHARMM και το GROMOS πραγματοποιούνται με τιμές $dt=1.0-2.0$ fs. Από την άλλη, τα πεδία αδρής διακριτικότητας όπως το MARTINI, λόγω της μείωσης του επιπέδου λεπτομέρειάς τους, επιτρέπουν τον ορισμό τιμών $dt=10-30$ fs, με την τιμή $dt=20$ fs να εμφανίζεται πιο συχνά στη βιβλιογραφία. Τέλος, το υβριδικό πεδίο PACE, λόγω του ότι περιγράφει τις πρωτεΐνες με σχεδόν ατομική διακριτικότητα και το διαλύτη/μεμβράνη με αδρή, τυπικά χρησιμοποιεί τιμές ενδιάμεσες των παραπάνω (π.χ. $dt=5$ fs).

Εκτός από τις παραπάνω ιδιότητες, κατά την διεξαγωγή των προσομοιώσεων Μοριακής Δυναμικής εφαρμόζονται επιπλέον αλγόριθμοι για τη ρύθμιση των συνθηκών του συστήματος. Ο ορισμός αυτών των αλγορίθμων εξαρτάται από το είδος της θερμοδυναμικής κατανομής που χρησιμοποιείται. Στη μικροκανονική κατανομή NVE δεν απαιτείται η χρήση κάποιου αλγορίθμου, αφού τόσο η θερμοκρασία όσο και η πίεση εξελίσσονται ελεύθερα κατά τη διάρκεια της προσομοίωσης. Στην κανονική κατανομή NVT, στην οποία η θερμοκρασία του συστήματος διατηρείται σταθερή, απαιτείται η χρήση ενός αλγορίθμου – θερμοστάτη για τον έλεγχο της θερμοκρασίας (βλ. Ενότητα 2.5.1.1), ενώ στην ισόθερμη-ισοβαρή κατανομή NPT απαιτείται επιπλέον και η χρήση ενός βαροστάτη (βλ. Ενότητα 2.5.1.2) για τον έλεγχο της πίεσης. Συνηθίζεται να χρησιμοποιούνται διαφορετικοί αλγόριθμοι για τα στάδια της εξισορρόπησης και παραγωγής, αντίστοιχα (Πίνακας 5). Στα στάδια εξισορρόπησης, αντικείμενο των οποίων είναι η επίτευξη ισορροπίας στις περιβαλλοντικές συνθήκες, χωρίς να είναι απαραίτητη η σωστή κατανομή της κινητικής ενέργειας, συνηθίζεται να χρησιμοποιούνται αλγόριθμοι όπως ο θερμοστάτης/βαροστάτης Berendsen, οι οποίοι επιτρέπουν τη γρήγορη ρύθμιση της θερμοκρασίας και της πίεσης, καθώς και ο θερμοστάτης V-rescale, ο οποίος αποτελεί παραλλαγή του αλγορίθμου Berendsen. Από την άλλη, στο

Κεφάλαιο II – Υλικά & Μέθοδοι

στάδιο παραγωγής, και ειδικά στην περίπτωση που αντικείμενο της προσομοίωσης είναι η μέτρηση ιδιοτήτων που εξαρτώνται από τη θερμοκρασία και την πίεση (π.χ. θερμοχωρητικότητα, επιφανειακή τάση, επιφάνεια ανά λιπίδιο κλπ) προτιμούνται αλγόριθμοι οι οποίοι προτυποποιούν πραγματικές συνθήκες NVT ή NPT, όπως οι θερμοστάτες V-rescale και Nosé-Hoover και ο βαροστάτης Parrinello-Rahman. Σημειώνεται ότι η χρήση θερμοστάτη/βαροστάτη εξαρτάται, συχνά, και από το δυναμικό πεδίο που χρησιμοποιείται, με κάποια πεδία να έχουν βρεθεί να αποδίδουν καλύτερα όταν χρησιμοποιούνται συγκεκριμένοι αλγόριθμοι θερμοκρασίας ή πίεσης. Τέλος, υπενθυμίζεται ότι στις προσομοιώσεις συστημάτων με βιολογικές μεμβράνες η πίεση ρυθμίζεται ημι-ισοτροπικά (βλ. Ενότητα 2.5.1.2), με τον έλεγχο της στον άξονα Z να πραγματοποιείται ανεξάρτητα από τους άξονες X και Y, προσομοιώνοντας έτσι τη γεωμετρία της λιπιδικής διπλοστιβάδας (Kandt et al., 2007).

2.1.10 Προσδιορισμός ιδιοτήτων σε αποτελέσματα προσομοιώσεων Μοριακής Δυναμικής

Τα αποτελέσματα των προσομοιώσεων Μοριακής Δυναμικής μπορούν να αναλυθούν τόσο μέσω της οπτικής παρατήρησης της πορείας (trajectory) όσο και μέσα από τον υπολογισμό μιας σειράς από ιδιότητες του συστήματος και των συστατικών του. Κάποιες από αυτές τις ιδιότητες είναι καθολικές για το σύστημα (π.χ. θερμοκρασία, πίεση, κινητική ενέργεια κλπ), ενώ άλλες υπολογίζονται ειδικά για συγκεκριμένα συστατικά του (π.χ. πρωτεΐνες, λιπίδια) μέσα από μαθηματικές εκφράσεις. Σε αυτήν την ενότητα παρουσιάζονται μερικές από τις πιο συχνά υπολογιζόμενες ιδιότητες των πρωτεϊνών και των λιπιδίων, οι οποίες χρησιμοποιούνται κατεξοχήν για την αξιολόγηση των προσομοιώσεων Μοριακής Δυναμικής.

Στην περίπτωση των πρωτεϊνών, η πιο κοινή ιδιότητα που προσδιορίζεται είναι η ρίζα της μέσης τετραγωνικής απόκλισης (Root Mean Square Deviation, RMSD). Το RMSD δύο στοιχισμένων δομών αποτελεί μέτρο της μέσης απόκλισης μεταξύ των ατόμων, συνήθως των κεντρικών ατόμων άνθρακα Cα και, για κάθε ζεύγος ατόμων i και j, υπολογίζεται σύμφωνα με την εξίσωση

Κεφάλαιο II – Υλικά & Μέθοδοι

$$RMSD = \sqrt{\frac{\sum (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}{N}} \quad (46)$$

όπου x_i, y_i, z_i και x_j, y_j, z_j είναι οι συντεταγμένες των ατόμων i και j , ενώ N είναι ο αριθμός των ατόμων στην πρωτεϊνική δομή. Στις προσομοιώσεις Μοριακής Δυναμικής η μέτρηση του RMSD πραγματοποιείται για κάθε στιγμιότυπο της πορείας σε σχέση με την αρχική δομή στη μονάδα του χρόνου, και χρησιμοποιείται για την αξιολόγηση της μεταβολής της πρωτεϊνικής δομής από την αρχική στερεοδιάταξή της. Μια δεύτερη ιδιότητα, σχετική με το RMSD, είναι η ρίζα της μέσης τετραγωνικής ταλάντωσης (Root Mean Square Fluctuation, RMSF) καθενός καταλοίπου στη δομή μιας πρωτεΐνης. Το RMSF εκφράζει την απόκλιση των ατόμων από μια θέση αναφοράς όπως και το RMSD, ωστόσο η απόκλιση μετράται για κάθε άτομο ξεχωριστά στο σύνολο του χρόνου σύμφωνα με την εξίσωση

$$RMSF = \sqrt{\frac{\sum (x_i(t) - x_i)^2 + (y_i(t) - y_i)^2 + (z_i(t) - z_i)^2}{T}} \quad (47)$$

όπου για κάθε στιγμιότυπο t του συνολικού χρόνου προσομοίωσης T μετρείται η απόκλιση των συντεταγμένων κάθε ατόμου i . Έτσι, ουσιαστικά το RMSF αποτελεί μέτρο της κινητικότητας κάθε ατόμου κατά τη διάρκεια της προσομοίωσης και ερμηνεύεται αντίστοιχα με τους θερμοκρασιακούς παράγοντες B των κρυσταλλικών δομών.

Εκτός από την κινητικότητα των πρωτεϊνών, ένα άλλο χαρακτηριστικό που αναλύεται κατά τη διάρκεια των προσομοιώσεών τους είναι η προσβάσιμη στο διαλύτη επιφάνειά τους. Ως προσβάσιμη στο διαλύτη επιφάνεια (Accessible Surface Area, ASA) ορίζεται η περιοχή που διαγράφεται από το κέντρο μιας διεισδύουσας σφαίρας που αντιπροσωπεύει ένα μόριο διαλύτη, καθώς αυτό κυλάει πάνω στην επιφάνεια του μελετώμενου μορίου. Ένα άτομο ή μια ομάδα ατόμων ενός μορίου χαρακτηρίζεται ως προσβάσιμο όταν ένα μόριο διαλύτη με συγκεκριμένο μέγεθος μπορεί να σχηματίσει επαφή van der Waals με αυτό (Kabsch & Sander, 1983). Η ASA είναι η περιοχή στην επιφάνεια μιας σφαίρας ακτίνας R , σε κάθε σημείο της οποίας το κέντρο ενός μορίου διαλύτη μπορεί να έρθει σε επαφή με αυτό το άτομο, χωρίς να συγκρούεται με άλλα άτομα του μορίου. Η ακτίνα R δίνεται από το άθροισμα της ακτίνας van der Waals του ατόμου και της ακτίνας του επιλεγμένου μορίου του διαλύτη. Ως διαλύτης συνήθως επιλέγεται το νερό, με ακτίνα $1.4 \text{ \AA}$ σε δυναμικά πεδία ατομικής διακρίτικότητας, ενώ στην περίπτωση αδρών δυναμικών πεδίων όπως το MARTINI επιλέγεται μεγαλύτερη σε

Κεφάλαιο II – Υλικά & Μέθοδοι

μέγεθος ακτίνα (π.χ. 5.2 Å), η οποία αντιστοιχεί στην ομαδοποίηση των ατόμων σε σωματίδια CG.

Αντίστοιχα με τις πρωτεΐνες, οι μεμβράνες διαθέτουν μια σειρά από δομικές ιδιότητες, η πορεία των οποίων καταγράφεται κατά τη διάρκεια μιας προσομοίωσης. Οι ιδιότητες αυτές έχουν να κάνουν κυρίως με την οργάνωση και τη γεωμετρία των λιπιδίων στη διπλοστιβάδα και περιλαμβάνουν μεγέθη όπως η επιφάνεια ανά λιπίδιο, το πάχος της λιπιδικής διπλοστιβάδας, η πλευρική διάχυση των μεμβρανικών συστατικών και η παράμετρος τάξης των λιπιδικών ουρών (Kandt et al., 2007).

Ως επιφάνεια ανά λιπίδιο (area per lipid) ορίζεται η επιφάνεια που καταλαμβάνει η πολική κεφαλή ενός λιπιδίου στο συνολικό εμβαδόν της επιφάνειας της κάθε στιβάδας. Σε απλές, μεμβράνες, αποτελούμενες μόνο από ένα είδος λιπιδίου, η επιφάνεια ανά λιπίδιο μπορεί να υπολογιστεί απλά με τη διαίρεση του συνολικού εμβαδού της στιβάδας διά τον αριθμό των λιπιδίων που τη συνθέτουν. Ωστόσο, η παραπάνω μέθοδος δεν μπορεί να εφαρμοστεί σε μικτές μεμβράνες, καθώς οι τελευταίες αποτελούνται από περισσότερα του ενός είδη λιπιδίων, τα οποία διαφέρουν στο μέγεθος και τη δομή τους και, κατά συνέπεια, διαθέτουν διαφορετικές τιμές επιφάνειας. Σε αυτές τις μεμβράνες, αντίθετα, η επιφάνεια ανά λιπίδιο υπολογίζεται μέσω κατακερματισμού (tessellation) της επιφάνειας της στιβάδας σε πολύγωνα Voronoi, με κάθε πολύγωνο να αντιστοιχεί στο χώρο που καταλαμβάνει το κάθε λιπίδιο. Η επιφάνεια κάθε λιπιδικού μορίου υπολογίζεται ως η επιφάνεια του πολυγώνου, ενώ η μέση επιφάνεια για κάθε τύπο λιπιδίων υπολογίζεται ως η μέση τιμή των επιφανειών των λιπιδικών μορίων αυτού του τύπου (Allen et al., 2009).

Το πάχος της λιπιδικής διπλοστιβάδας ορίζεται, τυπικά, ως η κατακόρυφη στο επίπεδο της διπλοστιβάδας απόσταση ανάμεσα στα επίπεδα των πολικών περιοχών των δύο στιβάδων, όπως αυτά καθορίζονται από τις φωσφορικές κεφαλές των λιπιδίων. Αλγοριθμικά, αυτή η απόσταση προσδιορίζεται μέσω του υπολογισμού της πυκνότητας μάζας των φωσφορικών ιόντων των λιπιδίων, ακολουθούμενη από τη μέτρηση της απόστασης ανάμεσα στις κορυφές της πυκνότητας. Το πάχος της διπλοστιβάδας εξαρτάται άμεσα από το μήκος αλλά και τον κορεσμό των υδρόφοβων ουρών των λιπιδίων. Σημειώνεται πως το πάχος δεν είναι απαραίτητα ίδιο σε όλο το πλάτος της διπλοστιβάδας. Αντίθετα, παρουσιάζονται

Κεφάλαιο II – Υλικά & Μέθοδοι

διακυμάνσεις, τόσο λόγω διαφορών στην οργάνωση των λιπιδίων σε διαφορετικά τμήματα της μεμβράνης όσο και λόγω της επίδρασης παραγόντων όπως η πίεση, η διαπερατότητα από διαλύτη και ιόντα αλλά και η επιφανειακή τάση των λιπιδίων. Ωστόσο, το μέσο πάχος της λιπιδικής διπλοστιβάδας θεωρείται πως παραμένει σταθερό γύρω από μια μέση τιμή και, μαζί με την επιφάνεια ανά λιπίδιο, χρησιμοποιείται τόσο σαν κριτήριο της σύγκλισης των προσομοιώσεων όσο και σαν κριτήριο αξιολόγησης της πιστότητάς της, μέσω σύγκρισης των αποτελεσμάτων της προσομοίωσης με αντίστοιχες, πειραματικά προσδιορισμένες τιμές για τα λιπίδια που συνθέτουν τη διπλοστιβάδα.

Μια ακόμη παράμετρος που μετράται για τα λιπιδικά μόρια είναι η παράμετρος τάξης (order parameter) των λιπιδικών ουρών. Η παράμετρος τάξης αποτελεί μέτρο της οργάνωσης των λιπιδίων στη διπλοστιβάδα, και μπορεί να χρησιμοποιηθεί για να καθορίσει τη φάση στην οποία βρίσκεται η μεμβράνη (βλ. Ενότητα 1.2.5). Η παράμετρος τάξης, γνωστή και με τον όρο παράμετρος τάξης δευτερίωσης (deuterium order parameter, S_{CD}) ή παράμετρος τάξης δεύτερης βαθμίδας (second rank order parameter, P_2) προσδιορίζεται πειραματικά μέσα από πειράματα NMR ή, πιο σπάνια, φασματοσκοπίας υπερύθρου. Στις προσομοιώσεις Μοριακής Δυναμικής η παράμετρος τάξης P_2 υπολογίζεται για κάθε ομοιοπολικό δεσμό C-C στις υδρόφοβες ουρές των λιπιδίων, σύμφωνα με τον τύπο

$$P_2 = \frac{1}{2}(3 \cdot \cos^2 \theta - 1) \quad (48)$$

όπου θ είναι η γωνία του δεσμού σε σχέση με την κάθετο ευθεία του επιπέδου της διπλοστιβάδας. Οι τιμές του P_2 στα λιπίδια κυμαίνονται από -0.5 ως 1 και υποδεικνύουν τη γεωμετρία του κάθε δεσμού. Έτσι, για τιμές κοντά στο $P_2=1$ ο δεσμός είναι παράλληλος με την κάθετο, για $P_2=-0.5$ είναι πλήρως αντιπαράλληλος, ενώ τιμές στην περιοχή $P_2=0$ υποδεικνύουν τυχαία διεύθυνση. Ο μέσος όρος των τιμών P_2 όλων των δεσμών σε ένα λιπιδικό μόριο αποτελεί ένδειξη της οργάνωσής του και, κατά συνέπεια, της οργάνωσης της λιπιδικής διπλοστιβάδας. Γενικότερα, λιπίδια με κορεσμένες υδρόφοβες ουρές (π.χ. DPPC) διαθέτουν υψηλές τιμές P_2 και είναι πιο οργανωμένα στη μεμβράνη, με αντίστοιχες επιπτώσεις στο πάχος του υδρόφοβου πυρήνα, την επιφάνεια ανά λιπίδιο και τη ρευστότητα της διπλοστιβάδας. Αντίθετα, λιπίδια με ακόρεστες ουρές (π.χ. POPC, DOPC) διαθέτουν χαμηλότερες τιμές P_2 και πιο χαλαρή οργάνωση. Σημειώνεται, τέλος, πως η τιμή του P_2 , όπως και άλλων παραμέτρων

Κεφάλαιο II – Υλικά & Μέθοδοι

των λιπιδίων, εξαρτάται και από τη θερμοκρασία του συστήματος. Έτσι, ένα λιπίδιο μπορεί να έχει υψηλότερες τιμές P_2 σε χαμηλές θερμοκρασίες και χαμηλότερες τιμές σε υψηλές θερμοκρασίες.

Τέλος, ένα πολύ σημαντικό μέγεθος που προσδιορίζεται κατά τη μελέτη των μεμβρανών και των συστατικών τους (λιπίδια αλλά και πρωτεΐνες) είναι ο συντελεστής διάχυσης (D), ένα μέτρο της ικανότητας των συστατικών της μεμβράνης να κινούνται ελεύθερα στο επίπεδο της διπλοστιβάδας. Ο συντελεστής διάχυσης για ένα σύστημα υπολογίζεται από τη συνάρτηση της μέσης τετραγωνικής μετατόπισης (Mean Square Displacement, MSD) στη μονάδα του χρόνου t , σύμφωνα με τη συνάρτηση Einstein

$$MSD = \langle \|r_i(t) - r_i(0)\|^2 \rangle = 2d \cdot D \cdot t \quad (49)$$

όπου r_i είναι η θέση (x,y,z) του κάθε μορίου i στο στιγμιότυπο t σε σχέση με την αρχή της προσομοίωσης, d είναι ο αριθμός των βαθμών ελευθερίας (διαστάσεις) του συστήματος και οι αγκύλες « $\langle \rangle$ » υποδηλώνουν το μέσο όρο των μετρήσεων. Στην περίπτωση της πλευρικής διάχυσης στις μεμβράνες οι βαθμοί ελευθερίας είναι $d=2$. Ο συντελεστής διάχυσης λαμβάνει τιμές σε μονάδες cm^2/s , με υψηλότερες τιμές να αντιστοιχούν σε πιο ευκίνητα μόρια, τα οποία διαχέονται πιο εύκολα στο επίπεδο της μεμβράνης (Wang et al., 2016).

2.1.11 Καθοδηγούμενη Μοριακή Δυναμική

Ο μηχανισμός λειτουργίας πολλών πρωτεϊνών συχνά περιλαμβάνει μια σειρά μεταβάσεων από μια κατάσταση ισορροπίας σε μια άλλη (π.χ. στερεοδιαταξικές αλλαγές κατά την ενεργοποίηση ενός υποδοχέα από έναν προσδέτη ή κατά το άνοιγμα και κλείσιμο ενός ιοντικού καναλιού). Ωστόσο, φαινόμενα αυτού του τύπου συχνά δεν μπορούν να προτυποποιηθούν με τυπικές προσομοιώσεις Μοριακής Δυναμικής, καθώς απαιτούνται για την παρατήρησή τους χρονικά διαστήματα της τάξης των ms ή ακόμα και δευτερολέπτων, κατά πολύ μεγαλύτερα από αυτά που μπορούν να επιτευχθούν μέσα από μια απλή προσομοίωση ισορροπίας. Μία λύση σε αυτό το πρόβλημα είναι η εφαρμογή εξωτερικών δυνάμεων κατά την προσομοίωση, οι οποίες χρησιμοποιούνται για να κατευθύνουν το σύστημα από μία κατάσταση σε μια άλλη (Phillips et al., 2005). Η μεθοδολογία βασίζεται στην αντίστοιχη πειραματική εφαρμογή δυνάμεων σε τεχνικές όπως η Μικροσκοπία Ατομικής Δύναμης (Atomic Force Microscopy, AFM) (Binnig et al., 1986) ή η χρήση οπτικών λαβίδων laser (laser optical tweezers) (Jagannathan & Marqusee, 2013), οι οποίες επίσης εφαρμόζουν εξωτερικές δυνάμεις σε ένα σύστημα για τη μελέτη των μηχανικών ιδιοτήτων σε κύτταρα, υποκυτταρικά οργανίδια και υπερμοριακά συμπλέγματα.

Η *in silico* εκδοχή αυτών των πειραματικών διαδικασιών, γνωστή με το όρο *Καθοδηγούμενη Μοριακή Δυναμική* (Steered Molecular Dynamics, SMD) περιλαμβάνει επίσης την εφαρμογή δυνάμεων σε βιομόρια κατά την προσομοίωση, τόσο για την αξιολόγηση των μηχανικών ιδιοτήτων τους όσο και για την επιτάχυνση φαινομένων τα οποία, υπό φυσιολογικές συνθήκες, απαιτούν πολύ μεγαλύτερο χρονικό διάστημα. Η προσεκτική εφαρμογή αυτών των δυνάμεων οδηγεί το σύστημα να υπερπηδήσει τυχόν ενεργειακά φράγματα ανάμεσα στις δύο καταστάσεις ισορροπίας, και επιτρέπει τη μελέτη του μονοπατιού της μετάβασης από τη μία κατάσταση στην άλλη. Στην πιο γενική μορφή της, η Καθοδηγούμενη Μοριακή Δυναμική λειτουργεί εφαρμόζοντας εξωτερικές επιρροές σε ένα υποσύνολο του συστήματος μελέτης (π.χ. ένα άτομο, μια ομάδα ατόμων, το κέντρο βάρους μιας πρωτεΐνης κλπ) ως προς μια κατεύθυνση. Αυτές οι επιρροές υλοποιούνται είτε εφαρμόζοντας μια συνεχή, σταθερή δύναμη F (προσομοιώσεις σταθερής δύναμης), είτε με τη

Κεφάλαιο II – Υλικά & Μέθοδοι

χρήση σταθερής ταχύτητας (προσομοιώσεις σταθερής ταχύτητας), χρησιμοποιώντας ένα αρμονικό δυναμικό της μορφής

$$U_{SMD}(\vec{r}_1, \vec{r}_2, \dots, t) = \frac{1}{2}k \left[u \cdot t - \left(\overrightarrow{R(t)} - \overrightarrow{R_0} \right) \cdot \vec{n} \right]^2 \quad (50)$$

όπου $(r_1, r_2, \dots)$ είναι οι θέσεις των ατόμων στα οποία εφαρμόζεται το δυναμικό, k είναι η σταθερά ελατηρίου, t είναι ο χρόνος, R_0 και $R(t)$ είναι η θέση του κέντρου μάζας των ατόμων στην αρχική κατάσταση και σε χρόνο t , αντίστοιχα, u είναι η ταχύτητα και n είναι η κατεύθυνση στην οποία ωθείται το σύστημα (Phillips et al., 2005). Ωστόσο, το δυναμικό της Εξίσωσης 50 μπορεί να τροποποιηθεί σημαντικά, ανάλογα με τις ανάγκες της προσομοίωσης.

2.1.12 Υπολογισμοί Ελεύθερης Ενέργειας

Οι υπολογισμοί των μεταβολών της ελεύθερης ενέργειας αποτελούν σημαντικό κομμάτι των προσομοιώσεων Μοριακής Δυναμικής στα βιομόρια. Μπορούν να χρησιμοποιηθούν για την περιγραφή του ενεργειακού τοπίου (γνωστού και σαν «Επιφάνεια Ελεύθερης Ενέργειας») των πρωτεϊνών κατά τη διαδικασία του διπλώματος, την αξιολόγηση της ισχύος των αλληλεπιδράσεων ανάμεσα σε δύο ή περισσότερα βιομόρια, την αξιολόγηση της επίδρασης μιας ή περισσότερων μεταλλαγών στην πρωτεϊνική δομή κλπ. Έτσι, οι ενεργειακοί υπολογισμοί βρίσκουν εφαρμογή σε πολλά πεδία, τόσο στη Βιοπληροφορική και την Υπολογιστική Βιοφυσική, με μεθόδους όπως η μοριακή προτυποποίηση και η αγκυροβόληση πρωτεϊνών-πρωτεϊνών και πρωτεϊνών – μικρών μορίων, όσο και στην πειραματική Βιολογία, την Ιατρική και τη Φαρμακολογία, σε πειράματα μεταλλαξιγένεσης αλλά και στο σχεδιασμό φαρμάκων (Knight & Brooks, 2009; Pohorille et al., 2010).

Στην πιο γενική μορφή της, η μεταβολή της ελεύθερης ενέργειας ενός συστήματος, ΔE , μπορεί να οριστεί ως η διαφορά της ενθαλπίας (ΔH) από την εντροπία του (ΔS), σύμφωνα με το γενικό τύπο:

$$\Delta E = \Delta H - T \cdot \Delta S \quad (51)$$

όπου ΔH είναι η ενθαλπία του συστήματος, ΔS η εντροπία και T η θερμοκρασία. Στην περίπτωση συστημάτων τα οποία ακολουθούν την κανονική θερμοδυναμική κατανομή NVT, ο

Κεφάλαιο II – Υλικά & Μέθοδοι

ενεργειακός όρος ΔE ταυτίζεται με τη μεταβολή στην ελεύθερη ενέργεια Helmholtz (ΔF), ενώ η ενθαλπία ΔH με την εσωτερική ενέργεια του συστήματος ΔE_i :

$$\Delta F = \Delta E_i - T \cdot \Delta S \quad (52)$$

όπου η εσωτερική ενέργεια, στις προσομοιώσεις Μοριακής Δυναμικής, αναλύεται στη Χαμιλτονιανή συνάρτηση $H(r,p)$ του συστήματος (Εξίσωση 2). Από την άλλη, στην ισόθερμη-ισοβαρή κατανομή NPT, στην οποία πραγματοποιούνται οι βιολογικές και βιοχημικές αντιδράσεις, συμπεριλαμβανομένων των προσομοιώσεών τους, υπολογίζεται η μεταβολή της ελεύθερης ενέργειας Gibbs (ΔG), σύμφωνα με τον τύπο

$$\Delta G = \Delta E_i + PV - T \cdot \Delta S \quad (53)$$

όπου ο ενθαλπικός όρος $\Delta H = \Delta E_i + PV$, εκτός από την εσωτερική ενέργεια ΔE_i περιλαμβάνει, τυπικά, και την ενεργειακή συνεισφορά του έργου πίεσης-όγκου PV (με μονάδες μέτρησης Pa και cm^3 για τα δύο μεγέθη και $1 \text{ Pa} \cdot \text{cm}^3 = 1 \text{ joule}$). Ωστόσο, στις τάξεις μεγεθών στις οποίες λαμβάνουν χώρα οι βιοχημικές αντιδράσεις ($\text{\AA}^3 \cdot \text{nm}^3$) η συνεισφορά του έργου PV είναι τόσο μικρή ($\sim 10^{-30} \text{ kJ/mol}$) ώστε να είναι, πρακτικά, αμελητέα. Έτσι, συνήθως δεν λαμβάνεται υπόψη και οι υπολογισμοί της ΔG χρησιμοποιούν μόνο τη συνεισφορά της εσωτερικής ενέργειας. Τέλος, στην περίπτωση των χημικών και βιοχημικών αντιδράσεων, η μεταβολή ΔG μπορεί να υπολογιστεί και μέσω του τύπου

$$\Delta G = -RT \ln K_a \text{ ή } \Delta G = RT \ln K_d \quad (54)$$

όπου R είναι η παγκόσμια σταθερά των αερίων, T είναι η θερμοκρασία και K_a και K_d είναι οι σταθερές ισορροπίας και διαστάσεως της αντίδρασης, αντίστοιχα, με $K_d = 1/K_a$. Σημειώνεται πως η τιμή της R προκύπτει από το γινόμενο $R = k_B \cdot N_A$, όπου k_B είναι η σταθερά Boltzmann και N_A είναι ο αριθμός Avogadro. Έτσι, η εξίσωση 50 αποτελεί ουσιαστικά μια τροποποιημένη εκδοχή της γενικής συνάρτησης υπολογισμού της ενέργειας Gibbs στην κατανομή NPT, όπως παρουσιάζεται στην Εξίσωση 36.

Στις υποενότητες που ακολουθούν παρουσιάζονται μια σειρά από μεθόδους υπολογισμού της ελεύθερης ενέργειας, οι οποίες εφαρμόζονται συχνά μέσα από προσομοιώσεις Μοριακής Δυναμικής για τη μελέτη των βιομοριακών αλληλεπιδράσεων. Δίνεται ιδιαίτερη έμφαση στον υπολογισμό του Δυναμικού Μέσης Δύναμης (PMF), καθώς και στη μεθοδολογία των διαταραχών ελεύθερης ενέργειας, δυο τεχνικών που χρησιμοποιούνται

Κεφάλαιο II – Υλικά & Μέθοδοι

κατά κόρον στον ακριβή υπολογισμό των μεταβολών ελεύθερης ενέργειας. Επιπρόσθετα, παρουσιάζονται δύο πιο εξειδικευμένες εφαρμογές, οι οποίες χρησιμοποιούν υπολογισμούς ελεύθερης ενέργειας στη μελέτη της επίδρασης των μεταλλαγών και στη μελέτη της επίδρασης της υδροφοβικής ασυμφωνίας.

2.1.12.1 Το Δυναμικό Μέσης Δύναμης

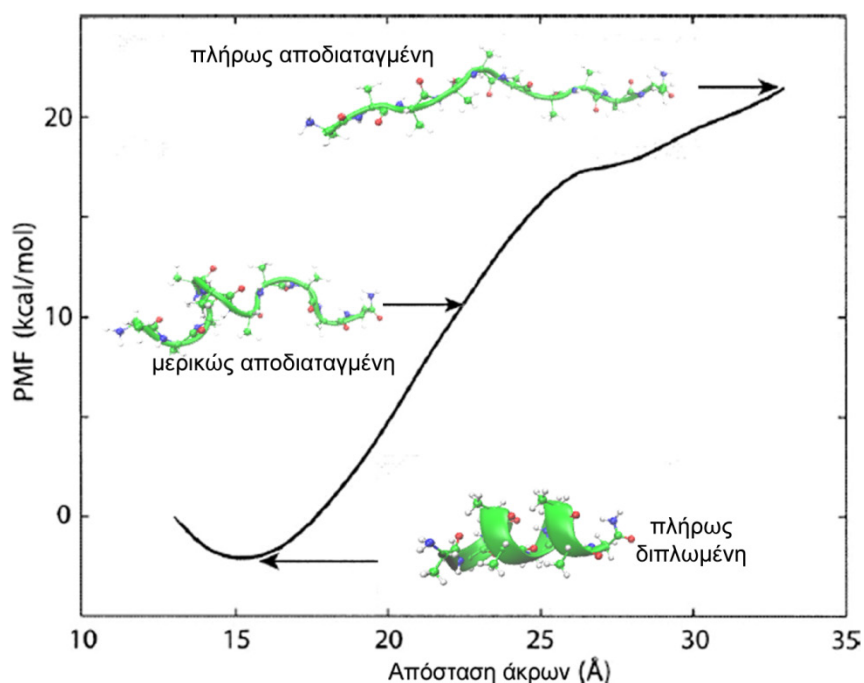
Το Δυναμικό Μέσης Δύναμης (Potential of Mean Force, PMF) είναι μια εκτίμηση της μεταβολής της ελεύθερης ενέργειας του συστήματος σαν συνάρτηση μιας συντεταγμένης αναφοράς (Kirkwood, 1935). Ο υπολογισμός του PMF για ένα σύστημα αποτελεί την πιο περιεκτική περιγραφή της θερμοδυναμικής συμπεριφοράς του. Οι καμπύλες PMF μπορούν να χρησιμοποιηθούν αποτελεσματικά για τη μελέτη της ελεύθερης ενέργειας του συστήματος και τη σχέση της με σύνθετα δομικά και θερμοδυναμικά δεδομένα, από στερεοδιαταξικές αλλαγές στη δομή μιας πρωτεΐνης (Εικόνα 35) έως την αλληλεπίδραση ανάμεσα σε δύο πρωτεΐνες ή σε μια πρωτεΐνη και ένα μικρό μόριο (Εικόνα 36).

Κατά τον υπολογισμό του Δυναμικού Μέσης Δύναμης οι μεταβολές του συστήματος μελετώνται με βάση μια οριζόμενη μεταβλητή, η οποία περιγράφει την μελετώμενη αντίδραση από την αρχική ως την τελική κατάσταση του συστήματος. Αυτό το μέγεθος είναι γνωστό σαν συντεταγμένη αντίδρασης (reaction coordinate) και τυπικά συμβολίζεται με το γράμμα «ξ». Το είδος της συντεταγμένης ξ μπορεί να περιγράφει π.χ. τη θέση ενός μορίου σε σχέση με ένα σημείο αναφοράς, τη γωνία ανάμεσα σε δύο αυτοτελείς δομικές περιοχές που συνδέονται με μια περιοχή-αρμό κλπ. Ο ορισμός και ο τρόπος μεταβολής της συντεταγμένης ουσιαστικά εξαρτάται από το είδος του φαινομένου που περιγράφεται. Για παράδειγμα, αν το σύστημα μελέτης είναι η αποδιάταξη της δομής ενός πεπτιδίου (Εικόνα 35), ως συντεταγμένη αντίδρασης μπορεί να οριστεί η απόσταση ανάμεσα στο αμινοτελικό και το καρβοξυτελικό άκρο του. Σε αυτήν την περίπτωση, το Δυναμικό Μέσης Δύναμης μπορεί να οριστεί, σαν μια παραλλαγή των εξισώσεων 35-36, ως

$$F(\xi) = -k_B T \ln P(\xi) \quad (55)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

όπου $F(\xi)$ είναι η τιμή του PMF σαν συνάρτηση της συντεταγμένης ξ , k_B είναι η σταθερά Boltzmann και $P(\xi)$ είναι η συνάρτηση κατανομής της πιθανότητας να βρεθεί το σύστημα σε μια συγκεκριμένη τιμή της συντεταγμένης ξ (Roux, 1995). Σημειώνεται πως, τυπικά, το PMF ορίζεται σαν έκφραση της ελεύθερης ενέργειας Helmholtz (F), αφού σύμφωνα με τον κλασικό ορισμό του υπολογίζεται μέσα από προσομοιώσεις NVT. Ωστόσο, καθώς η πλειονότητα των προσομοιώσεων των βιομορίων πραγματοποιούνται σε συνθήκες σταθερής θερμοκρασίας και πίεσης, ο υπολογισμός του PMF στην ισόθερμη-ισοβαρή κατανομή NPT μπορεί να θεωρηθεί και σαν έκφραση της ελεύθερης ενέργειας Gibbs.



Εικόνα 35 Το Δυναμικό Μέσης Δύναμης (PMF). Παρουσιάζεται ο υπολογισμός του PMF για την αντίδραση της αποδιάταξης της δομής ενός δεκαπεπτιδίου αλανίνης με δομή α -έλικας, ως συνάρτηση της απόστασης ανάμεσα στο N- και το C-άκρο του πεπτιδίου. Οι δομές της αρχικής και τελικής κατάστασης, καθώς και ένα ενδιάμεσο στιγμιότυπο της αντίδρασης αντιστοιχούνται στις ανάλογες περιοχές της καμπύλης του PMF. Οι υπολογισμοί πραγματοποιήθηκαν χρησιμοποιώντας την Ισότητα του Jarzynski.

Θεωρητικά, ο προσδιορισμός του PMF θα μπορούσε να πραγματοποιηθεί για οποιαδήποτε βιολογική αντίδραση μέσα από μια τυπική προσομοίωση ισορροπίας, εφόσον

Κεφάλαιο II – Υλικά & Μέθοδοι

επιλεγεί η κατάλληλη συντεταγμένη ξ και εφόσον η προσομοίωση δώσει επαρκή δειγματοληψία, ώστε να περιγράψει πλήρως το φαινόμενο. Στην πράξη, ωστόσο, αυτό πολλές φορές δεν είναι δυνατό, καθώς απαιτούνται χρονικά διαστήματα πολύ μεγαλύτερα από αυτά που μπορούν να επιτευχθούν μέσα από μια απλή προσομοίωση. Επιπρόσθετα, πολλά βιολογικά φαινόμενα περιλαμβάνουν μεταβάσεις ανάμεσα σε καταστάσεις που χωρίζονται από υψηλά ενεργειακά φράγματα, τα οποία είναι αδύνατο να προσπελαστούν μέσα από μια προσομοίωση ισορροπίας. Έτσι, οι υπολογισμοί του PMF πολύ συχνά πραγματοποιούνται εφαρμόζοντας προσομοιώσεις εκτός ισορροπίας (Lemkul, 2019).

Παρακάτω παρουσιάζονται δυο πολύ συχνά χρησιμοποιούμενες μέθοδοι για τον προσδιορισμό του PMF, η Δειγματοληψία Ομπρέλας και η Ισότητα του Jarzynski.

Δειγματοληψία Ομπρέλας: Μια πολύ συχνά χρησιμοποιούμενη μέθοδος για τον προσδιορισμό του PMF είναι αυτή της Δειγματοληψίας Ομπρέλας (Umbrella Sampling, US) (Roux, 1995). Η μέθοδος US βασίζεται στην τροποποίηση της συνολικής δυναμικής ενέργειας του συστήματος με την προσθήκη ενός αρμονικού δυναμικού $W(\xi)$ της μορφής

$$W(\xi) = k(\xi - \xi_0)^2 \quad (56)$$

όπου ξ είναι η συντεταγμένη αντίδρασης στην κατάσταση που επιθυμούμε να μελετήσουμε, ξ_0 είναι η αρχική κατάσταση ισορροπίας και k είναι η σταθερά ελατηρίου. Το δυναμικό $W(\xi)$, γνωστό και ως *δυναμικό μεροληψίας (biasing potential)*, μεταβάλλει την ενεργειακή επιφάνεια του συστήματος, ωθώντας το έτσι στο να υπερβεί τους ενεργειακούς φραγμούς που το εμποδίζουν και να μεταβεί από την κατάσταση ισορροπίας ξ_0 στην κατάσταση ξ (Hub et al., 2010).

Οι προσομοιώσεις US ουσιαστικά εφαρμόζουν την παραπάνω αρχή μελετώντας τη μετάβαση του συστήματος κατά μήκος του μονοπατιού που περιγράφεται από τη συντεταγμένη ξ σταδιακά (Εικόνα 252). Το μονοπάτι του συστήματος από ξ_0 ως ξ χωρίζεται σε N διαδοχικά στιγμιότυπα ή «παράθυρα», καθένα από τα οποία αντιστοιχεί και σε μια τιμή ξ_i της συντεταγμένης ξ , με $i=0,1,2,\dots,N$. Για κάθε μονοπάτι i πραγματοποιείται μια προσομοίωση US, κατά την οποία εφαρμόζεται στο σύστημα ένα δυναμικό μεροληψίας $W(\xi)$ και υπολογίζεται η τροποποιημένη συνάρτηση της κατανομής πυκνότητας $P'(\xi)$

Κεφάλαιο II – Υλικά & Μέθοδοι

$$P_i'(\xi) = \frac{Z_0}{Z_i} e^{-W_i(\xi)} P_0(\xi) \quad (57)$$

όπου $P_0(\xi)$ είναι η συνάρτηση κατανομής στην κατάσταση αναφοράς, ενώ Z_0 και Z_i είναι οι συναρτήσεις επιμερισμού της κατάστασης αναφοράς και του παραθύρου i αντίστοιχα. Η προσομοίωση όλων των παραθύρων οδηγεί στο συνολικό υπολογισμό της τροποποιημένης συνάρτησης $P'(\xi)$, μέσω της οποίας μπορεί να υπολογιστεί το PMF τροποποιώντας κατάλληλα την Εξίσωση 55:

$$F(\xi) = -k_B T \ln P'(\xi) - W(\xi) + K \quad (58)$$

όπου $W(\xi)$ είναι το δυναμικό μεροληψίας και K είναι μια ενεργειακή σταθερά (Roux, 1995). Η τιμή της K εξαρτάται από το δυναμικό $W(\xi)$ και μπορεί να υπολογιστεί μέσω μιας στατιστικής μεθόδου, γνωστής σαν Μέθοδος Ανάλυσης Σταθμισμένων Ιστογραμμάτων (Weighted Histogram Analysis Method, WHAM) (Kumar et al., 1992).

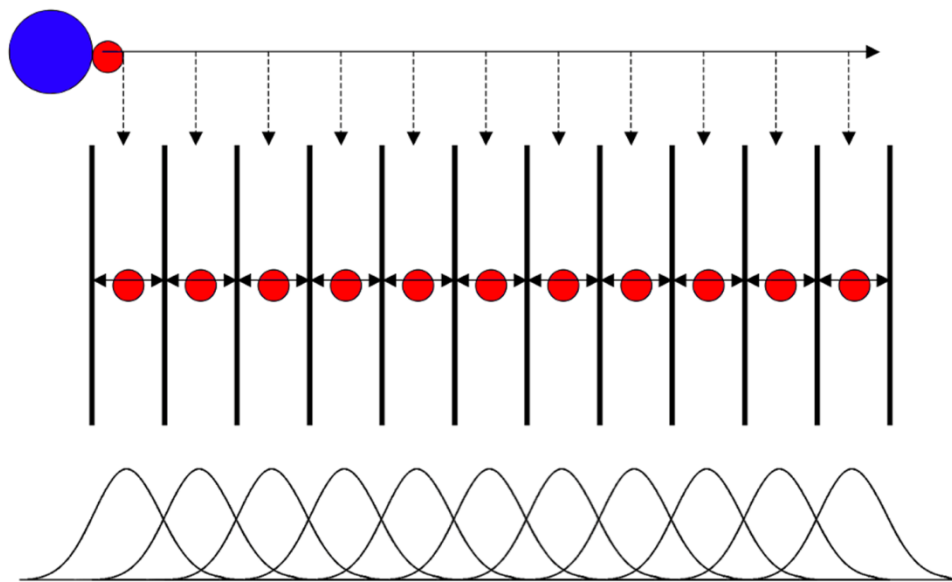
Η μέθοδος WHAM στηρίζεται στην ιδέα ότι καθένα από τα παράθυρα προσομοίωσης US μπορεί να απεικονιστεί ως ένα ιστόγραμμα (Εικόνα 36), το οποίο θα περιγράφει την κατανομή της πιθανότητας για την εμφάνιση της κατάστασης ενδιαφέροντος που περιγράφεται από το παράθυρο. Η συνάρτηση κατανομής στην κατάσταση αναφοράς $P_0(\xi)$ ισούται με το σταθμισμένο άθροισμα αυτών των ιστογραμμάτων, σύμφωνα με την Εξίσωση 59

$$P_0(\xi) = \frac{\sum_{i=1}^N n_i P_i'(\xi)}{\sum_{i=1}^N n_i e^{-(W_i(\xi) - K_i)/k_B T}} \quad (59)$$

όπου k_B και T είναι η σταθερά Boltzmann και η θερμοκρασία, αντίστοιχα (Hub et al., 2010). Από την παραπάνω εξίσωση προκύπτει η τιμή της σταθεράς K ως

$$K = -k_B T \ln \sum_{i=1}^N P_i'(\xi) e^{-W_i(\xi)/k_B T} \quad (60)$$

Η Δειγματοληψία Ομπρέλας αποτελεί μια από τις πιο ακριβείς μεθόδους για τον υπολογισμό του PMF, με αποτελέσματα τα οποία βρίσκονται συχνά σε συμφωνία με πειραματικές μετρήσεις, όπως προκύπτει από τη βιβλιογραφία (Hub et al., 2010). Ωστόσο, οι υπολογισμοί εξαρτώνται σε μεγάλο βαθμό από τον σωστό ορισμό τιμών για τη συντεταγμένη αναφοράς ξ , με αποτέλεσμα να παρουσιάζουν έντονο σφάλμα όταν αυτές οι τιμές δεν έχουν οριστεί σωστά. Επιπρόσθετα, η χρήση ξεχωριστών προσομοιώσεων για κάθε παράθυρο καθιστά τη μέθοδο υπολογιστικά απαιτητική, ιδιαίτερα για μεγάλα σε μέγεθος συστήματα όπως αυτά των διαμεμβρανικών πρωτεϊνών.



Εικόνα 36 Η μέθοδος της Δειγματοληψίας Ομπρέλας (Umbrella Sampling). Παρουσιάζεται διαγραμματικά το παράδειγμα του αποχωρισμού ενός συμπλόκου πρωτεϊνών-πρωτεϊνών, με τη σταδιακή απομάκρυνση της μιας πρωτεΐνης μέσα από μια προσομοίωση SMD. Ως συντεταγμένη της αντίδρασης ορίζεται η απόσταση ανάμεσα στις δύο πρωτεΐνες. Μια σειρά από ενδιάμεσες καταστάσεις (παράθυρα US) λαμβάνονται κατά την προσομοίωση SMD, για καθένα από τα οποία πραγματοποιείται μια προσομοίωση US με την εφαρμογή ενός κατάλληλου αρμονικού δυναμικού. Η κατανομή των πιθανοτήτων για κάθε παράθυρο προτυποποιείται μέσω αλληλεπικαλυπτόμενων ιστογραμμάτων, τα οποία χρησιμοποιούνται για τον προσδιορισμό του PMF μέσω της μεθόδου WHAM. Προσαρμογή από Lemkul, 2019.

Το θεώρημα της Ισότητας του Jarzynski: Το θεώρημα της Ισότητας του Jarzynski (Jarzynski Equality) είναι μια απλοποιημένη προσέγγιση για τον υπολογισμό του PMF. Διατυπώθηκε για πρώτη φορά από τον Christopher Jarzynski το 1997 στο Πανεπιστήμιο της Washington, σαν πιθανή λύση στο πρόβλημα του υπολογισμού του PMF από προσομοιώσεις εκτός ισορροπίας (Jarzynski, 1997a; Jarzynski, 1997b). Η ισότητα του Jarzynski μπορεί να εφαρμοστεί σε αποτελέσματα προσομοιώσεων Καθοδηγούμενης Μοριακής Δυναμικής, και έχει χρησιμοποιηθεί σε ενεργειακούς υπολογισμούς για μεγάλη ποικιλία βιομοριακών συστημάτων.

Η Ισότητα του Jarzynski επιχειρεί να συνδέσει τη διαφορά της ελεύθερης ενέργειας ανάμεσα σε δύο καταστάσεις, όπως αυτή παρατηρείται σε συνθήκες ισορροπίας, με το παραγόμενο έργο (W) που δίνεται από προσομοιώσεις εκτός ισορροπίας ως εξής: Ας

Κεφάλαιο II – Υλικά & Μέθοδοι

θεωρήσουμε ένα σύστημα μελέτης, το οποίο μεταβάλλεται από μια κατάσταση A σε μια κατάσταση B ακολουθώντας μια συντεταγμένη αντίδρασης ξ , σε χρόνο t . Σύμφωνα με το δεύτερο Θερμοδυναμικό Νόμο, η μεταβολή ελεύθερης ενέργειας αυτής της αντίδρασης, ΔF , σχετίζεται με το παραγόμενο έργο W μέσω της ανισότητας:

$$\Delta F_{AB} = F_B - F_A \leq W_{A \rightarrow B} \quad (61)$$

όπου W είναι το παραγόμενο έργο. Στην παραπάνω συνάρτηση, η ισότητα $\Delta F = W$ ικανοποιείται μόνο αν η αντίδραση είναι ψευδοστατική, δηλαδή συμβαίνει τόσο αργά ώστε να είναι πλήρως αντιστρέψιμη. Ωστόσο, η Ισότητα του Jarzynski υπολογίζει τη μεταβολή ΔF σαν συνάρτηση του παραγόμενου έργου W που προκύπτει από ένα σύνολο μη αντιστρέψιμων αντιδράσεων σύμφωνα με την παρακάτω συνάρτηση:

$$e^{-\Delta F/k_B T} = \langle e^{-W/k_B T} \rangle \quad (62)$$

όπου k_B είναι η σταθερά Boltzmann, T είναι η θερμοκρασία και οι αγκύλες « $\langle \rangle$ » υποδηλώνουν το μέσο όρο του συνόλου των αντιδράσεων (Jarzynski, 1997a; Jarzynski, 1997b). Το έργο W κάθε αντίδρασης για κάθε στιγμιότυπο t_i υπολογίζεται ως

$$W(t_i) = u \int_0^t f(t_i) dt \quad (63)$$

όπου $f(t)$ είναι η δύναμη που εφαρμόζεται στο σύστημα σε κάθε στιγμιότυπο και u είναι η ταχύτητα της προσομοίωσης καθοδηγούμενης Μοριακής Δυναμικής. Έτσι, το PMF τελικά ορίζεται με βάση τις Εξισώσεις 55, 62 και 63, ως

$$F(\xi) = -k_B T \ln \left(\left\langle e^{-u \int_0^t f(t_i) dt / k_B T} \right\rangle \right) \quad (64)$$

Πρακτικά, η μέθοδος υλοποιείται με τον ορισμό μιας συντεταγμένης αντίδρασης ξ και την πραγματοποίηση πολλαπλών προσομοιώσεων Καθοδηγούμενης Μοριακής Δυναμικής (SMD) κατά μήκος αυτής της συντεταγμένης (Park et al., 2003). Για κάθε προσομοίωση υπολογίζεται το παραγόμενο έργο, σύμφωνα με την Εξίσωση 59, και η τελική καμπύλη PMF προκύπτει ως η μέση τιμή όλων των υπολογισμένων τιμών του έργου. Σε αντίθεση με μεθόδους όπως η Δειγματοληψία Ομπρέλας, οι οποίες απαιτούν ξεχωριστές προσομοιώσεις για κάθε διακριτή τιμή της συντεταγμένης ξ , η Ισότητα του Jarzynski είναι πιο ευέλικτη υπολογιστικά, εφόσον χρησιμοποιεί απευθείας τις τιμές έργου των προσομοιώσεων SMD (Park et al., 2003). Ωστόσο, λόγω της απλότητάς της, η μέθοδος επηρεάζεται σε πολύ μεγάλο

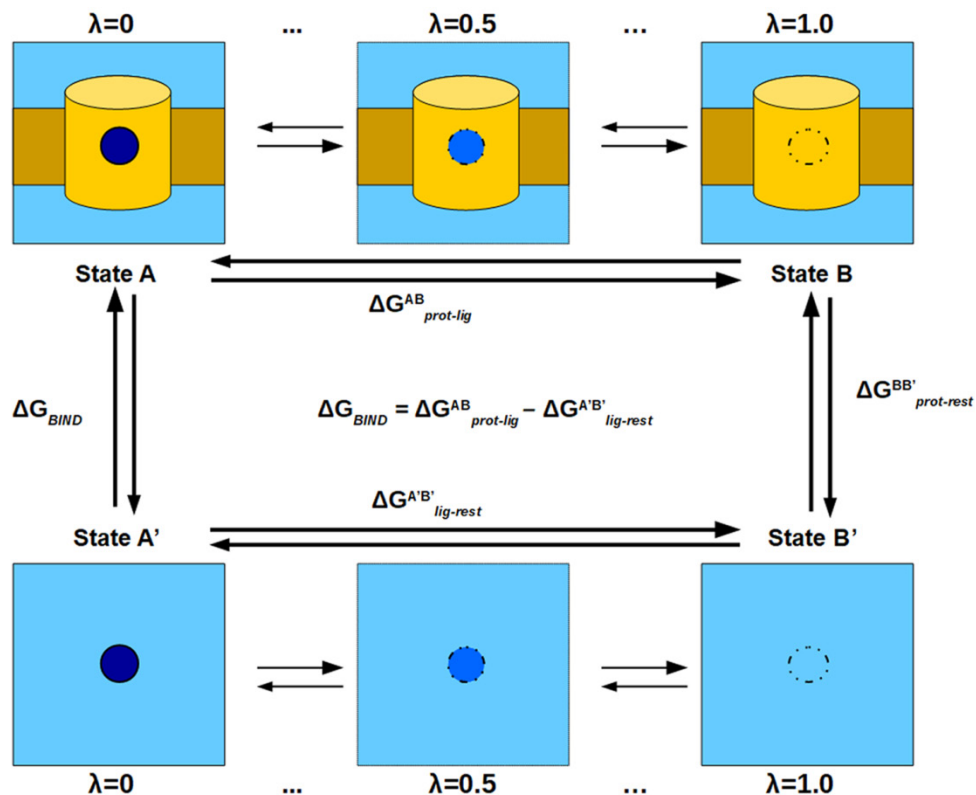
Κεφάλαιο II – Υλικά & Μέθοδοι

βαθμό από ιδιότητες των προσομοιώσεων SMD, όπως το μέγεθος της εφαρμοζόμενης δύναμης, η ταχύτητα της προσομοίωσης, αλλά και οι συνθήκες ελέγχου της θερμοκρασίας.

2.1.12.2 Διαταραχή Ελεύθερης Ενέργειας

Πέρα από τον υπολογισμό του PMF, μια άλλη ευρέως διαδεδομένη μέθοδος για την πραγματοποίηση υπολογισμών ελεύθερης ενέργειας είναι η Διαταραχή Ελεύθερης Ενέργειας (Free Energy Perturbation, FEP). Η μέθοδος FEP αποτελεί μια από τις παλαιότερες και πιο εμπειριστωμένες μεθόδους υπολογισμού της ενέργειας στο πεδίο των προσομοιώσεων. Οι προσομοιώσεις FEP μπορούν να χρησιμοποιηθούν σε υπολογισμούς όπως η μελέτη της συγγένειας πρόσδεσης μικρών μορίων σε πρωτεΐνες, ο υπολογισμός των ενεργειών υδάτωσης (hydration) ιόντων και βιομορίων, η μελέτη της πρωτεϊνικής σταθερότητας κατά την εφαρμογή μεταλλαγών και η διαφορά της συμπεριφοράς των μορίων ανάμεσα σε διαφορετικούς διαλύτες. Τα αποτελέσματα αυτών των μετρήσεων μπορούν να βρουν μεγάλο εύρος εφαρμογών, από το σχεδιασμό και τη βελτιστοποίηση νέων μοριακών παραμέτρων για δυναμικά πεδία έως το σχεδιασμό φαρμάκων και τη σύγκριση της δραστηρότητας ανάμεσα σε διαφορετικές ουσίες (Pohorille et al., 2010).

Αντίστοιχα με τον υπολογισμό του PMF, οι μετρήσεις FEP για μια αντίδραση πραγματοποιούνται ορίζοντας μια συντεταγμένη αντίδρασης, η οποία τυπικά συμβολίζεται με το γράμμα «λ» και περιγράφει την μετάβαση του συστήματος από μια αρχική κατάσταση A σε μια τελική κατάσταση B. Η διαφορά από τους υπολογισμούς PMF εντοπίζεται στο ότι στις μετρήσεις FEP δεν πραγματοποιείται στερεοδιαταξική μετάβαση του συστήματος από την κατάσταση A στη B, είτε με προσομοιώσεις Ισορροπίας είτε μέσω προσομοιώσεων SMD. Αντίθετα, η μέθοδος λειτουργεί «μεταμορφώνοντας» σταδιακά το σύστημα από την αρχική στην τελική κατάσταση, τροποποιώντας κατάλληλα τη δομή του και τις παραμέτρους του δυναμικού πεδίου που τη χαρακτηρίζουν (Knight & Brooks, 2009).



Εικόνα 37 Ο θερμοδυναμικός κύκλος του υπολογισμού της ενέργειας πρόσδεσης σε ένα σύμπλοκο υποδοχέα-προσδέτη με τη μέθοδο FEP. Ορίζονται δύο συστήματα, το σύστημα του συμπλόκου υποδοχέα-προσδέτη και το σύστημα του ελεύθερου προσδέτη σε υδατικό διάλυμα. Για κάθε σύστημα πραγματοποιούνται προσομοιώσεις FEP, στις οποίες γίνεται σταδιακή αποδέσμευση του προσδέτη μέσω του μονοπατιού που καθορίζεται από τη συντεταγμένη « λ ». Η τιμή ΔG_{BIND} ορίζεται ως η διαφορά ανάμεσα στις ενέργειες FEP των δύο συστημάτων.

Το είδος αυτής της «μεταμόρφωσης» εξαρτάται από το φαινόμενο που μελετάται. Για παράδειγμα, στον υπολογισμό της αλληλεπίδρασης ανάμεσα σε μια πρωτεΐνη και ένα μόριο - προσδέτη, ως αρχικές και τελικές καταστάσεις ορίζονται το σύμπλοκο πρωτεΐνη-προσδέτης και η απομονωμένη πρωτεΐνη αντίστοιχα. Σε αυτήν την περίπτωση, οι υπολογισμοί FEP πραγματοποιούνται προσομοιώνοντας την «εξαφάνιση» ή «αποδέσμευση» του μορίου από τη θέση πρόσδεσής του, μέσω της σταδιακής «απενεργοποίησης» των όρων δυναμικής ενέργειας στις παραμέτρους του μορίου. Από την άλλη, στη μελέτη της επίδρασης μιας μεταλλαγής στη δομή μιας πρωτεΐνης, ως αρχικές και τελικές καταστάσεις ορίζονται ο φυσικός τύπος και η μεταλλαγμένη πρωτεΐνη, αντίστοιχα. Σε αυτήν την περίπτωση, οι υπολογισμοί FEP πραγματοποιούνται προσομοιώνοντας τη σταδιακή «μετατροπή» της θέσης μεταλλαγής από

Κεφάλαιο II – Υλικά & Μέθοδοι

το αρχικό στο τελικό κατάλοιπο. Σε όλες τις περιπτώσεις, πραγματοποιούνται στερεοδιαταξικές και θερμοδυναμικές τροποποιήσεις οι οποίες δεν ανταποκρίνονται απαραίτητα σε πραγματικά φαινόμενα και, για αυτό το λόγο, οι προσομοιώσεις FEP συχνά αναφέρονται και ως «αλχημικές μεταμορφώσεις» (alchemical transformations). Ωστόσο, η ανάλυση των αποτελεσμάτων τους έχει βρεθεί ότι μπορεί να χρησιμοποιηθεί στον υπολογισμό της αλλαγής στην ελεύθερη ενέργεια από την αρχική στην τελική κατάσταση με σημαντική ακρίβεια (Knight & Brooks, 2009; Pohorille et al., 2010).

Ανεξάρτητα από τον τρόπο που πραγματοποιούνται οι μεταμορφώσεις κατά τις προσομοιώσεις FEP, αυτές συμβαίνουν πάντα σταδιακά, κατά μήκος της συντεταγμένης λ . Οι τιμές του λ κυμαίνονται από 0 ως 1, αντιστοιχώντας στην αρχική και την τελική κατάσταση, με τις ενδιάμεσες καταστάσεις να λαμβάνουν ενδιάμεσες δεκαδικές τιμές (π.χ. $\lambda=0,0.1,0.2,...,1.0$). Έτσι, η διαφορά στην ενέργεια κατά τη μετάβαση από την κατάσταση A στην κατάσταση B (ΔG_{AB}) μπορεί να οριστεί ως το άθροισμα

$$\Delta G_{AB} = \sum_{\lambda=0}^{1.0} (-k_B T \ln \langle e^{-\frac{H_{\lambda+\Delta\lambda}-H_{\lambda}}{k_B T}} \rangle) \quad (65)$$

όπου k_B είναι η σταθερά Boltzmann, T είναι η θερμοκρασία του συστήματος, λ και $\Delta\lambda$ είναι η συντεταγμένη της αντίδρασης και η διαφορά της από κατάσταση σε κατάσταση, αντίστοιχα, H_{λ} και $H_{\lambda+\Delta\lambda}$ είναι οι τροποποιημένες Χαμιλτονιανές συναρτήσεις των δύο καταστάσεων που συγκρίνονται κάθε φορά και οι αγκύλες « $\langle \rangle$ » υποδηλώνουν το μέσο όρο του θερμοδυναμικού συνόλου. Η τροποποίηση της Χαμιλτονιανής συνάρτησης του συστήματος σε κάθε βήμα της συντεταγμένης λ ορίζεται ως

$$H_{\lambda} = K + (1 - \lambda)U_A + \lambda U_B \quad (66)$$

όπου K είναι η κινητική ενέργεια του συστήματος, ενώ U_A και U_B είναι οι τιμές της δυναμικής ενέργειας στην αρχική και τελική κατάσταση, αντίστοιχα. Έτσι, για $\lambda=0$ και $\lambda=1.0$ η Χαμιλτονιανή συνάρτηση περιγράφει τις ακραίες καταστάσεις A και B, αντίστοιχα, ενώ οι ενδιάμεσες τιμές του λ οδηγούν σε Χαμιλτονιανές που περιγράφουν τις ενδιάμεσες, «υβριδικές» καταστάσεις. Η ενέργεια σε καθεμία από αυτές τις καταστάσεις υπολογίζεται μέσα από διαδοχικά παράθυρα προσομοιώσεων FEP, με τον αριθμό τους να αντιστοιχεί στον αριθμό των τιμών της συντεταγμένης λ . Το άθροισμα των παραθύρων αποτελεί την ενέργεια ΔG_{AB} .

Κεφάλαιο II – Υλικά & Μέθοδοι

Ο υπολογισμός της ελεύθερης ενέργειας, όπως περιγράφηκε παραπάνω, αντιπροσωπεύει το μονοπάτι $A \rightarrow B$, μέσα από τις «μεταμορφώσεις» FEP που πραγματοποιήθηκαν. Μονοπάτια αυτού του τύπου επαρκούν για σχετικά απλές προσομοιώσεις, όπως είναι ο υπολογισμός της ενέργειας υδάτωσης ενός μορίου (δηλαδή η μεταβολή ενέργειας κατά τη μεταφορά του μορίου από υδρόφοβο περιβάλλον σε υδατικό διαλύτη). Ωστόσο, για τις προσομοιώσεις περισσότερο περίπλοκων συστημάτων απαιτείται ο σχεδιασμός περισσότερων αλληλοσυμπληρούμενων μονοπατιών FEP, τα οποία οργανώνονται σε θερμοδυναμικούς κύκλους (Εικόνα 37). Χαρακτηριστικό παράδειγμα αποτελεί η μέτρηση της ενέργειας πρόσδεσης (ΔG_{bind}) ενός μορίου-προσδέτη σε μια πρωτεΐνη, μία αντίδραση που ουσιαστικά περιλαμβάνει τη μετάβαση του προσδέτη από το υδατικό διάλυμα στη θέση πρόσδεσης της πρωτεΐνης. Η ίδια η μετάβαση δεν μπορεί να προτυποποιηθεί άμεσα, ωστόσο, η ενεργειακή τιμή της μπορεί να προσδιοριστεί έμμεσα μέσα από δύο ξεχωριστές αντιδράσεις, την «αποδέσμευση» του προσδέτη από το σύμπλοκο πρωτεΐνης-προσδέτη και την «αποδέσμευση» του ίδιου του προσδέτη από το υδατικό διάλυμα. Ο κύκλος αντιδράσεων που σχηματίζεται, γνωστός και σαν «κύκλος διπλής αποδέσμευσης» (double annihilation cycle), μπορεί να δώσει την ενέργεια πρόσδεσης ΔG_{bind} σαν τη διαφορά ανάμεσα στις ενέργειες των δύο ενδιάμεσων αντιδράσεων (Pohorille et al., 2010):

$$\Delta G_{\text{bind}} = \Delta G_{\text{prot-lig}}^{AB} - \Delta G_{\text{lig-rest}}^{AB} \quad (67)$$

2.1.12.3 Υπολογιστική Σάρωση Αλανίνης

Η πειραματική μέθοδος της Σάρωσης Αλανίνης (Alanine Scanning) χρησιμοποιείται συχνά στη Μοριακή Βιολογία για τον προσδιορισμό της συνεισφοράς των καταλοίπων στη σταθερότητα ή τη λειτουργία μιας πρωτεΐνης. Η διαδικασία περιλαμβάνει τη σύνθεση μεταλλαγών της πρωτεΐνης, στις οποίες κάθε κατάλοιπο μεταλλάσσεται διαδοχικά σε αλανίνη. Η αλανίνη επιλέγεται για την αντικατάσταση καθώς έχει μικρή πλευρική αλυσίδα η οποία, παρ'όλα αυτά, μπορεί να μιμηθεί τις προτιμήσεις των άλλων καταλοίπων ως προς τη δευτεροταγή δομή. Ακολούθως, οι διαφορές ανάμεσα στην πρωτεΐνη φυσικού τύπου και κάθε μεταλλαγμένη, που προκύπτουν ως αποτέλεσμα της αντικατάστασης του κάθε καταλοίπου με

Κεφάλαιο II – Υλικά & Μέθοδοι

αλανίνη, μπορούν να μελετηθούν με διάφορες μεθόδους, όπως η Φασματοσκοπία Υπερερύθρου ή ο Πυρηνικός Μαγνητικός Συντονισμός (Morrison & Weiss, 2001). Η Υπολογιστική Σάρωση Αλανίνης (Computational Alanine Scanning) επιχειρεί να μιμηθεί την παραπάνω πειραματική διαδικασία προκειμένου να προβλέψει πιθανά κατάλοιπα-θέσεις ενδιαφέροντος πάνω στην επιφάνεια αλληλεπίδρασης ενός πρωτεϊνικού συμπλόκου. Αντίστοιχα με την πειραματική σάρωση αλανίνης, η διαδικασία περιλαμβάνει τη διαδοχική αντικατάσταση κάθε καταλοίπου της επιφάνειας αλληλεπίδρασης με αλανίνη μέσω ομόλογης προτυποποίησης, ακολουθούμενη από βελτιστοποίηση των νέων μοντέλων μέσα από ελαχιστοποίηση ενέργειας και σύντομες προσομοιώσεις. Η συνεισφορά κάθε καταλοίπου αξιολογείται μέσα από τον υπολογισμό της διαφοράς στην ενέργεια αλληλεπίδρασης, σαν αποτέλεσμα της μεταλλαγής ($\Delta\Delta G$):

$$\Delta\Delta G = \Delta G_{mut} - \Delta G_{WT} \quad (68)$$

όπου ΔG_{mut} είναι η ενέργεια αλληλεπίδρασης της μεταλλαγμένης πρωτεΐνης και ΔG_{WT} είναι η ενέργεια του φυσικού τύπου. Τα κατάλοιπα με τιμή $\Delta\Delta G > 1.5$ kcal/mol θεωρούνται σημεία ενδιαφέροντος («hot spots») για την επιφάνεια αλληλεπίδρασης του συμπλόκου (Guerois et al., 2002).

2.1.12.4 Ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας

Η υδροφοβική ασυμφωνία (hydrophobic mismatch) είναι το φαινόμενο κατά το οποίο το πάχος του υδρόφοβου πυρήνα της λιπιδικής διπλοστιβάδας διαφέρει από το μήκος των διαμεμβρανικών τμημάτων των πρωτεϊνών. Αυτή η ασυμφωνία μπορεί να οδηγήσει σε μη ευνοϊκές αλληλεπιδράσεις ανάμεσα στις διαμεμβρανικές πρωτεΐνες και το περιβάλλον τους, καθώς υδρόφοβα κατάλοιπα εκτίθενται στον υδατικό διαλύτη και, αντίστροφα, πολικά ή φορτισμένα κατάλοιπα θάβονται στην υδρόφοβη διπλοστιβάδα. Η ομάδα των Mondal et al το 2011 πρότεινε μία μέθοδο για τον υπολογισμό της ενεργειακής ποινής αυτών των μη ευνοϊκών αλληλεπιδράσεων μέσα από προσομοιώσεις Μοριακής Δυναμικής (Mondal et al., 2014; Mondal et al., 2011). Η μέθοδος, γνωστή με το όνομα Μοριακή Δυναμική Συνεχούς (Continuum-Molecular Dynamics, CTMD), διαχωρίζει την ενεργειακή ποινή της υδροφοβικής

Κεφάλαιο II – Υλικά & Μέθοδοι

ασυμφωνίας σε δύο όρους, την ενέργεια παραμόρφωσης της διπλοστιβάδας (ΔG_{def}) και την ενέργεια της υπολειπόμενης υδροφοβικής ασυμφωνίας (ΔG_{RHM}).

Ο όρος ΔG_{def} περιγράφει τις ελαστικές παραμορφώσεις της μεμβράνης γύρω από τα εκτεθειμένα τμήματα των πρωτεϊνών. Αυτές οι παραμορφώσεις περιλαμβάνουν τοπικές αλλαγές στο πάχος της λιπιδικής διπλοστιβάδας με σκοπό την προσαρμογή της στη φύση των καταλοίπων στις περιοχές αυτές. Έτσι, το τοπικό πάχος σε καθεμία από τις στιβάδες της μεμβράνης αυξάνεται γύρω από εκτεθειμένα υδρόφοβα κατάλοιπα, με σκοπό να τα αποκρύψει από το διαλύτη και, αντίστροφα, μειώνεται γύρω από πολικά και αφόρτιστα κατάλοιπα. Αυτές οι παραμορφώσεις της μεμβράνης εκφράζονται μέσω ενός δείκτη τοπικών παραμορφώσεων $u(x,y)$, σύμφωνα με τον τύπο

$$u(x,y) = \frac{1}{2}(d(x,y) - d_0) \quad (69)$$

Όπου x και y είναι οι συντεταγμένες των σημείων στα οποία συμβαίνουν οι τοπικές παραμορφώσεις, $d(x,y)$ είναι το πάχος της διπλοστιβάδας σε αυτά τα σημεία, ενώ d_0 είναι η μέση τιμή του πάχους της διπλοστιβάδας σε κατάσταση ισορροπίας. Ο δείκτης u στη συνέχεια χρησιμοποιείται για τον προσδιορισμό της ενεργειακής ποινής των παραμορφώσεων σε κάθε στιβάδα σύμφωνα με τον τύπο

$$\Delta G_{\text{def}} = \frac{1}{2} \int \left\{ K_a \frac{(2u)^2}{d_0^2} + K_c \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - C_0 \right)^2 + a \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 \right) \right\} d\Omega \quad (70)$$

όπου οι τιμές K_a και K_c είναι οι δείκτες διαστρέβλωσης (splay distortion) και συμπίεσης-έκτασης της μεμβράνης, αντίστοιχα, C_0 είναι η αυθόρμητη καμπυλότητα της μονοστιβάδας και η τιμή a είναι η επιφανειακή τάση. Οι τιμές K_a , K_c , C_0 και a ορίζονται με βάση των τύπο λιπιδίων που συνθέτουν τη λιπιδική διπλοστιβάδα, με πειραματικά προσδιορισμένες τιμές για διάφορα λιπίδια (POPC, DPPC, χοληστερόλη κλπ) να υπάρχουν διαθέσιμες στη βιβλιογραφία.

Οι παραμορφώσεις της μεμβράνης γύρω από τις πρωτεΐνες, όπως αναφέρθηκε και στην Ενότητα 1.5, συχνά δεν είναι αρκετές για την πλήρη εξάλειψη της υδροφοβικής ασυμφωνίας. Έτσι, ο δεύτερος ενεργειακός όρος της μεθόδου (ΔG_{RHM}) περιγράφει την ενεργειακή ποινή της υπολειπόμενης υδροφοβικής ασυμφωνίας (residual hydrophobic mismatch, RHM), δηλαδή της ασυμφωνίας η οποία παραμένει στο σύστημα ακόμη και μετά τις

Κεφάλαιο II – Υλικά & Μέθοδοι

παραμορφώσεις της μεμβράνης. Η τιμή του ΔG_{RHM} ορίζεται σαν συνάρτηση της εκτεθειμένης επιφάνειας των καταλοίπων στα όρια των διαμεμβρανικών τμημάτων, σύμφωνα με τον τύπο

$$\Delta G_{RHM} = \sum_{i=1}^{N_{TM}} \sigma_{res} SA_{res,i} \quad (71)$$

όπου N_{TM} είναι ο αριθμός των διαμεμβρανικών τμημάτων, SA_{res} είναι το τμήμα της προσβάσιμης επιφάνειας των καταλοίπων το οποίο έρχεται σε επαφή με μη ευνοϊκό περιβάλλον και $\sigma_{res}=0.028$ kcal/mol. Η τελευταία τιμή αποτελεί τη μέση μεταβολή της ελεύθερης ενέργειας κατά τη μετάβαση των αμινοξικών καταλοίπων από περιβάλλον υδατικού διαλύτη σε περιβάλλον ανάλογο της λιπιδικής διπλοστιβάδας (διάλυμα οκτανόλης), όπως έχει προσδιοριστεί πειραματικά σε oligοπεπτίδια με μελέτες θερμιδομετρίας. Η τιμή SA_{res} υπολογίζεται από τα αποτελέσματα των προσομοιώσεων, και ο ορισμός της εξαρτάται από τη φύση του κάθε καταλοίπου. Για υδρόφοβα κατάλοιπα η SA_{res} είναι το τμήμα της επιφάνειας το οποίο εκτίθεται στο διαλύτη, ενώ για πολικά και φορτισμένα κατάλοιπα είναι το τμήμα της επιφάνειας το οποίο θάβεται στον πυρήνα της διπλοστιβάδας (Mondal et al., 2011).

2.1.13 Προγράμματα προσομοιώσεων Μοριακής Δυναμικής

Τα πρώτα προγράμματα προσομοιώσεων Μοριακής Δυναμικής που εμφανίστηκαν στη βιβλιογραφία ήταν τα CHARMM, AMBER και GROMOS, τα οποία αναπτύχθηκαν και συνεχίζουν να αναπτύσσονται παράλληλα με τα ομώνυμα δυναμικά πεδία, σαν υπολογιστικές υλοποιήσεις των τελευταίων. Ωστόσο, οι εξελίξεις στο χώρο της Πληροφορικής, σε συνδυασμό με την αλματώδη πρόοδο στην τεχνολογία των υπολογιστών και την αύξηση της υπολογιστικής ισχύος έδωσαν τη δυνατότητα για το σχεδιασμό νεότερων, πιο ισχυρών προγραμμάτων προσομοιώσεων, τα οποία εκμεταλλεύονται τις νέες τεχνολογίες, όπως η παράλληλη επεξεργασία (parallel processing) και η χρήση συνεπεξεργαστών γραφικών (GPUs) στους υπολογισμούς, με σκοπό την ταχύτερη προσομοίωση μεγαλύτερων σε μέγεθος συστημάτων. Έτσι, πλέον υπάρχουν διαθέσιμα δεκάδες διαφορετικά προγράμματα για προσομοιώσεις Μοριακής Δυναμικής, η πλειονότητα των οποίων είναι δημόσια διαθέσιμα σαν λογισμικό ανοιχτού κώδικα (open source). Στα πλαίσια της παρούσας διδακτορικής διατριβής έχουν

Κεφάλαιο II – Υλικά & Μέθοδοι

χρησιμοποιηθεί δύο διαφορετικά προγράμματα προσομοιώσεων, τα GROMACS και NAMD, τα οποία παρουσιάζονται συνοπτικά παρακάτω.

GROMACS: Το GROMACS είναι ένα ελεύθερα διαθέσιμο πακέτο λογισμικού ανοιχτού κώδικα για προσομοιώσεις Μοριακής Δυναμικής υψηλής απόδοσης, το οποίο αναπτύχθηκε αρχικά στο Πανεπιστήμιο του Groningen στην Ολλανδία και πλέον συντηρείται από μεγάλο αριθμό ερευνητικών ομάδων παγκοσμίως. Χαρακτηριστικό προτέρημά του είναι ότι υποστηρίζει όλες τις διαφορετικές υλοποιήσεις της συνάρτησης δυναμικής ενέργειας και, κατά συνέπεια, μπορεί να πραγματοποιήσει προσομοιώσεις με σχεδόν όλα τα δυναμικά πεδία που έχουν εμφανιστεί στη βιβλιογραφία. Επιπρόσθετα, το GROMACS διαθέτει υλοποιήσεις για μεγάλο αριθμό αλγορίθμων, συμπεριλαμβανομένων μιας σειράς από αλγορίθμους για τον έλεγχο της θερμοκρασίας και της πίεσης, και αυτοματοποιημένων μεθόδων για τη διεξαγωγή και ανάλυση προσομοιώσεων καθοδηγούμενης Μοριακής Δυναμικής, υπολογισμών PMF και ενεργειακών διαταραχών FEP. Το GROMACS υποστηρίζει τόσο την παράλληλη επεξεργασία όσο και τη χρήση GPUs στη διεξαγωγή προσομοιώσεων και είναι εξαιρετικά γρήγορο στη χρήση για προσομοιώσεις πρωτεϊνών, λιπιδίων και μεμβρανικών συστημάτων σε χρόνους της τάξης των μικροδευτερολέπτων (Abraham et al., 2015).

NAMD: Το NAMD είναι ένα ελεύθερα διαθέσιμο πρόγραμμα προσομοιώσεων Μοριακής Δυναμικής, το οποίο αναπτύχθηκε και συντηρείται από το εργαστήριο Θεωρητικής και Υπολογιστικής Βιοφυσικής στο Πανεπιστήμιο του Illinois των ΗΠΑ. Το NAMD είναι σχεδιασμένο για την προσομοίωση πολύ μεγάλων σε μέγεθος ($>10^6$ άτομα) συστημάτων, όπως υπερμοριακά συμπλέγματα δεκάδων ή εκατοντάδων πρωτεϊνών καθώς και ολόκληροι ιοί. Το NAMD υποστηρίζει κυρίως τα δυναμικά πεδία CHARMM, AMBER και OPLS και περιλαμβάνει μια σειρά από αλγορίθμους για εξειδικευμένες προσομοιώσεις, όπως η καθοδηγούμενη Μοριακή Δυναμική και διάφοροι ενεργειακοί υπολογισμοί αλλά και η χρήση πειραματικών δεδομένων, όπως δεδομένα NMR ή χάρτες πυκνότητας ηλεκτρονικής μικροσκοπίας, για την καθοδήγηση των προσομοιώσεων. Επιπρόσθετα, υποστηρίζει τόσο την παράλληλη επεξεργασία όσο και τη χρήση GPUs στη διεξαγωγή προσομοιώσεων. Τέλος, μέσω της

Κεφάλαιο II – Υλικά & Μέθοδοι

γλώσσας προγραμματισμού Tcl, προσφέρεται η δυνατότητα στο χρήστη να δημιουργήσει τις δικές του μεθόδους προσομοίωσης και ανάλυσης, επιτρέποντας έτσι το σχεδιασμό πιο εξειδικευμένων συστημάτων (Phillips et al., 2005).

Εκτός από τα προγράμματα GROMACS και NAMD, στα πλαίσια της διδακτορικής διατριβής έχουν χρησιμοποιηθεί μια σειρά από προγράμματα μοριακής προτυποποίησης (Molecular Modeling) και αγκυροβόλησης (Docking), τα οποία εφαρμόζουν μεθόδους Μοριακής Δυναμικής για την εξειδικευμένη προτυποποίηση, βελτιστοποίηση και ανάλυση των χαρακτηριστικών της δομής των πρωτεϊνών:

MODELLER: Το MODELLER είναι ένα πακέτο λογισμικού σχεδιασμένο για την προτυποποίηση πρωτεϊνικών δομών, εφαρμόζοντας τη μέθοδο της προτυποποίησης μέσω ομολογίας (Homology Modeling). Το MODELLER πραγματοποιεί προτυποποιήσεις μέσω ομολογίας εφαρμόζοντας μια τεχνική εμπνευσμένη από τον Πυρηνικό Μαγνητικό Συντονισμό (NMR) γνωστή ως ικανοποίηση χωρικών περιορισμών (Satisfaction of Spatial Restraints), σύμφωνα με την οποία ένα σύνολο γεωμετρικών κριτηρίων χρησιμοποιούνται για τη δημιουργία μιας συνάρτησης πυκνότητας που θα καθορίσει τη θέση κάθε ατόμου στο μοντέλο με βάση τη δομή-πρότυπο. Επιπρόσθετα, το πρόγραμμα περιλαμβάνει μια σειρά από αλγορίθμους για ελαχιστοποίηση ενέργειας και προσομοιώσεις Μοριακής Δυναμικής, μέσω των οποίων πραγματοποιείται η βελτιστοποίηση των αποτελεσμάτων. Εκτός από το πρότυπο με βάση το οποίο θα κατασκευαστεί το μοντέλο, το MODELLER απαιτεί επιπρόσθετα μια στοίχιση αλληλουχιών ανάμεσα στην αμινοξική ακολουθία της δομής – προτύπου και την αλληλουχία – στόχο. Η χρήση του προγράμματος πραγματοποιείται μέσω της συγγραφής προγραμμάτων στη γλώσσα προγραμματισμού Python, καθώς το MODELLER είναι ουσιαστικά μια συλλογή κλάσεων και αντικειμένων που υλοποιούν τις μεθόδους προτυποποίησης σαν λειτουργίες της Python (Webb & Sali, 2017).

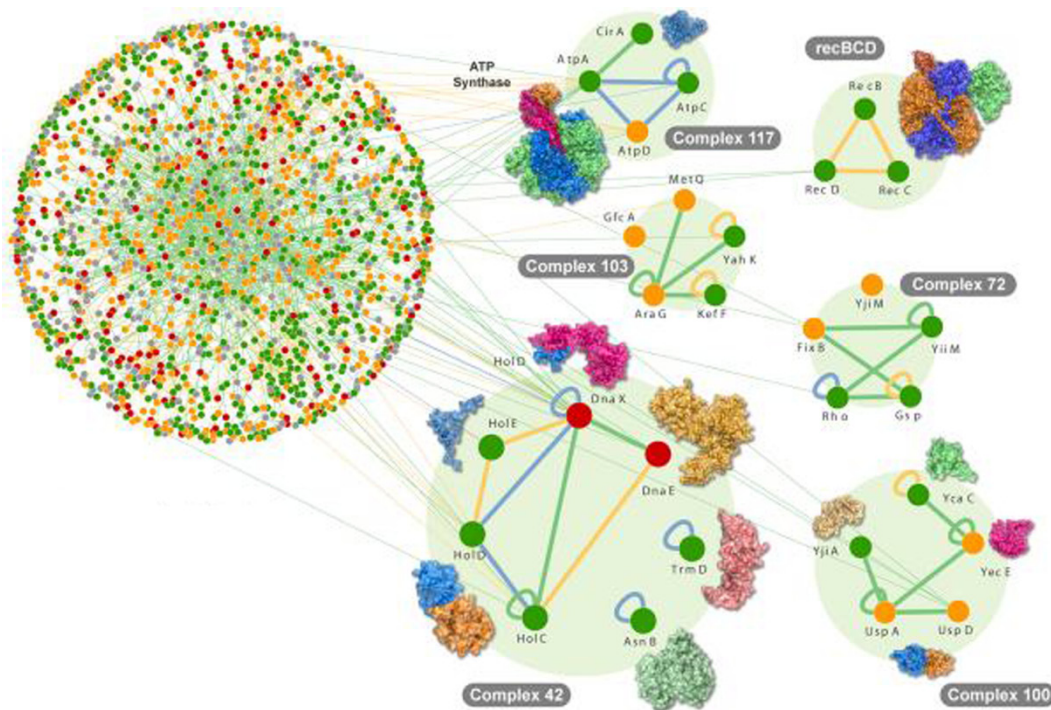
Κεφάλαιο II – Υλικά & Μέθοδοι

HADDOCK: Το HADDOCK (High Ambiguity Data-driven Docking) είναι μια κατευθυνόμενη από δεδομένα μέθοδος αγκυροβόλησης μεταξύ πρωτεϊνών, η οποία αξιοποιεί δομικές πληροφορίες (Ambiguous Interactions Restraints, AIRs) προερχόμενες από προηγούμενα πειραματικά ή υπολογιστικά δεδομένα για να κατευθύνει την πορεία της αγκυροβόλησης. Το πρωτόκολλο αγκυροβόλησης του HADDOCK αποτελείται από τρία στάδια: ελαχιστοποίηση ενέργειας των στατικών (rigid body) δομών με βάση τους υπολογισμένους περιορισμούς, ημι-ευέλικτη βελτιστοποίηση των διέδρων γωνιών σε συνθήκες κενού και μια τελική προσομοίωση Μοριακής Δυναμικής σε συνθήκες διαλύτη. Το τελικό αποτέλεσμα του HADDOCK περιλαμβάνει το σύνολο των προβλεφθέντων συμπλόκων, ομαδοποιημένα με βάση τη δομική ομοιότητά τους, ενώ κάθε σύμπλοκο βαθμολογείται με ένα ενεργειακό κριτήριο (HADDOCK Score), το οποίο προκύπτει σαν σταθμισμένο άθροισμα των τιμών της ενέργειας αλληλεπίδρασης για τις αλληλεπιδράσεις van der Waals (U_{vdW}), τις ηλεκτροστατικές αλληλεπιδράσεις (U_{elec}) και την επίδραση του διαλύτη (U_{desolv}), της θαμμένης επιφάνειας του συμπλόκου και των παραβιάσεων των αρχικών περιορισμών που τίθενται από το χρήστη (van Zundert et al., 2016).

Fold-X: Το Fold-X είναι ένα πρόγραμμα μοριακής προτυποποίησης, δομικής ανάλυσης και μοριακών προσομοιώσεων, σχεδιασμένο για τη μελέτη της σταθερότητας των πρωτεϊνικών δομών. Βασίζεται σε ενεργειακούς υπολογισμούς και προσομοιώσεις Μοριακής Δυναμικής για τη μελέτη της επίδρασης μεταλλαγών πάνω στην πρωτεϊνική δομή, την αξιολόγηση της σταθερότητας των ενδομοριακών και διαμοριακών αλληλεπιδράσεων ανάμεσα σε αμινοξικά κατάλοιπα και τη μελέτη της επίδρασης του διαλύτη, της θερμοκρασίας και της ιοντικής ισχύος πάνω στο πρωτεϊνικό δίπλωμα. Επιπρόσθετα, περιλαμβάνει μια σειρά από αλγορίθμους για τη διεξαγωγή προσομοιώσεων και ενεργειακών υπολογισμών όπως η Υπολογιστική Σάρωση Αλανίνης (Schymkowitz et al., 2005).

2.2 Δίκτυα Αλληλεπιδράσεων

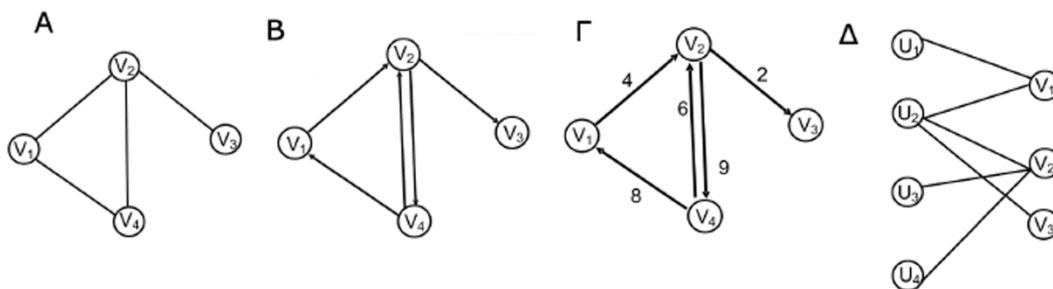
Η χρήση μοντέλων δικτύων για την απεικόνιση μεγάλου όγκου δεδομένων γίνεται όλο και πιο συχνή τα τελευταία χρόνια και αποτελεί αντικείμενο ενεργής έρευνας σε μεγάλο εύρος επιστημών, από την Πληροφορική και τα εφαρμοσμένα Μαθηματικά ως τις Κοινωνικές Επιστήμες αλλά και τη Στατιστική Φυσική και διάφορους κλάδους της Βιολογίας. Αναλύσεις δικτύων έχουν χρησιμοποιηθεί για τη μελέτη και περιγραφή ποικίλων φαινομένων, από τον παγκόσμιο ιστό και τα κοινωνικά δίκτυα ως την επιδημιολογία και την οργάνωση διάφορων οικοσυστημάτων. Ιδιαίτερη σημασία για τη Βιολογία παρουσιάζουν τα βιολογικά δίκτυα αλληλεπιδράσεων, τα οποία μπορούν να οπτικοποιήσουν και να αναλύσουν μεγάλο όγκο δεδομένων για την περιγραφή αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών (Εικόνα 38), τη ρύθμιση της γονιδιακής έκφρασης και τη συσχέτιση γονιδίων με τη λειτουργία τους αλλά και την ανάλυση πολύπλοκων μηχανισμών όπως οι μηχανισμοί μεταγωγής σήματος και τα μεταβολικά μονοπάτια (Pavlioroulos et al., 2011).



Εικόνα 38 Το δίκτυο αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών της *Escherichia coli*. Παρουσιάζεται το πλήρες δίκτυο, καθώς και μια σειρά από υπο-ομάδες του, οι οποίες αντιστοιχούν σε σύμπλοκα με γνωστή δομή ή καθορισμένο βιολογικό ρόλο.

2.2.1 Θεωρία Γράφων

Ο πυρήνας της κατασκευής και ανάλυσης δικτύων αλληλεπιδράσεων βασίζεται στη μαθηματική Θεωρία Γράφων (Graph Theory). Η Θεωρία Γράφων είναι ένα γνωστικό αντικείμενο της Πληροφορικής που περιλαμβάνει τον ορισμό και τη μελέτη των γράφων και των σχέσεών τους, οι οποίες χρησιμοποιούνται ευρύτατα στη μελέτη διαφόρων ειδών δικτύων (Huber et al., 2007). Σύμφωνα με τη θεωρία γράφων, ένας γράφος G μπορεί να οριστεί ένα σύνολο τιμών (V, E) , όπου με V συμβολίζονται οι κορυφές ή κόμβοι (nodes) του συνόλου, ενώ με E συμβολίζονται οι συνδέσεις ή ακμές (edges) που συνδέουν τους κόμβους που συσχετίζονται μεταξύ τους. Ένα δίκτυο αλληλεπιδράσεων αποτελεί στην ουσία ένα μεγάλο γράφο (V, E) , με τους κόμβους να αντιπροσωπεύουν τα στοιχεία του και τις ακμές να αντιπροσωπεύουν τις επαφές ανάμεσα σε αυτά τα στοιχεία. Στην περίπτωση των βιολογικών δικτύων οι κόμβοι αντιπροσωπεύουν τα βιομόρια (π.χ. πρωτεΐνες, γονίδια, μεταβολικά προϊόντα), ενώ οι ακμές ανάμεσά τους αντιπροσωπεύουν τις σχέσεις μεταξύ τους (π.χ. αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών, γονιδιακή έκφραση, μεταβολικές πορείες).



Εικόνα 39 Οι τέσσερις βασικές κατηγορίες των γράφων. (Α) Μη-κατευθυνόμενος γράφος. (Β) Κατευθυνόμενος γράφος, στον οποίο κάθε ακμή διαθέτει κατεύθυνση. (Γ) Σταθμισμένος γράφος, στον οποίο κάθε ακμή διαθέτει μια τιμή βάρους. (Δ) Διμερής γράφος, στον οποίο τα δεδομένα χωρίζονται σε δύο διακριτά σύνολα.

2.2.2 Κατηγορίες Γράφων

Ανάλογα με την οργάνωση των στοιχείων τους οι γράφοι (και, κατά συνέπεια, τα δίκτυα αλληλεπιδράσεων) μπορούν να διακριθούν σε τέσσερις κύριες κατηγορίες (Εικόνα 39), τους μη κατευθυνόμενους, τους κατευθυνόμενους, τους σταθμισμένους και τους διμερείς γράφους (Pavlopoulos et al., 2011). Οι μη κατευθυνόμενοι γράφοι παρουσιάζουν μέσω των ακμών τις αλληλεπιδράσεις ανάμεσα στους κόμβους, χωρίς να δίνουν περισσότερη πληροφορία για την ύπαρξη κάποιας κατευθυνόμενης πορείας. Ένας μη κατευθυνόμενος γράφος είναι απλά ένα σύνολο κόμβων και ακμών $G=(V, E)$, με την κάθε ακμή E να ορίζεται ως

$$E = \{(i, j) | i, j \in V\} \quad (72)$$

όπου i και j είναι δύο κόμβοι που συνδέονται άμεσα μεταξύ τους. Οι συνδεδεμένοι κόμβοι του γράφου χαρακτηρίζονται και ως «γείτονες», με τους άμεσα συνδεδεμένους μεταξύ τους κόμβους να χαρακτηρίζονται σαν «πρώτοι γείτονες», τους κόμβους οι οποίοι ενώνονται έμμεσα μέσω ενός ενδιάμεσου κόμβου να χαρακτηρίζονται σαν «δεύτεροι γείτονες» (π.χ. ένας κόμβος A συνδέεται με έναν κόμβο B , ο οποίος συνδέεται με έναν κόμβο Γ) κλπ. Η πλειονότητα των πρωτεϊνικών δικτύων αλληλεπιδράσεων ανήκουν σε αυτήν την κατηγορία (Pavlopoulos et al., 2011).

Αντίθετα, οι κατευθυνόμενοι γράφοι ορίζονται σαν σύνολα τριών τιμών $G=(V, E, f)$, όπου V και E είναι οι κόμβοι και οι ακμές, αντίστοιχα, ενώ με f ορίζεται μια συνάρτηση η οποία εκφράζει την κατεύθυνση της ακμής από έναν κόμβο i σε ένα κόμβο j . Οι ακμές των κατευθυνόμενων γράφων συμβολίζονται με βέλη, τα οποία υποδεικνύουν την κατεύθυνση των αλληλεπιδράσεων. Οι αλληλεπιδράσεις μπορεί να είναι μονοκατευθυνόμενες, οπότε συμβολίζονται με απλό βέλος ή δικατευθυνόμενες, οπότε συμβολίζονται με διπλά, αντιπαράλληλα βέλη. Στην κατηγορία των κατευθυνόμενων γράφων μπορούν να καταταχθούν τα δίκτυα των μεταβολικών μονοπατιών (KEGG, REACTOME κλπ), καθώς και τα δίκτυα μεταγωγής σήματος, στα οποία υπάρχει κατεύθυνση της πληροφορίας (Fabregat et al., 2018; Kanehisa & Goto, 2000).

Οι σταθμισμένοι γράφοι μπορούν να οριστούν σαν $G=(V, E, w)$, όπου V και E είναι οι κόμβοι και οι ακμές, αντίστοιχα, ενώ κάθε ακμή σταθμίζεται με μια τιμή βάρους w , η οποία αποτελεί δείκτη της ισχύος της στο δίκτυο. Τα βάρη των ακμών μπορούν να οριστούν με

Κεφάλαιο II – Υλικά & Μέθοδοι

διάφορους τρόπους, ανάλογα με το είδος της πληροφορίας που περιγράφουν (π.χ. απόσταση ανάμεσα στους κόμβους, αξιοπιστία της σύνδεσης κλπ). Ένας σταθμισμένος γράφος μπορεί να είναι, επιπρόσθετα, κατευθυνόμενος ή μη κατευθυνόμενος. Χαρακτηριστικό παράδειγμα σταθμισμένων γράφων αποτελούν τα δομικά δίκτυα αλληλεπιδράσεων (βλ. Ενότητα 2.2.4), μια ειδική εφαρμογή των δικτύων στην απεικόνιση της τρισδιάστατης δομής των βιομορίων όπου κάθε κόμβος είναι ένα αμινοξικό κατάλοιπο (πρωτεΐνες) ή μικρό μόριο (π.χ. ένας προσδέτης) και οι ακμές είναι οι αλληλεπιδράσεις ανάμεσα στα κατάλοιπα, ενώ τα βάρη των ακμών υποδεικνύουν την ισχύ ή τη σταθερότητα των αλληλεπιδράσεων (Eargle & Luthey-Schulten, 2012).

Τέλος, οι διμερείς γράφοι μπορούν να οριστούν σαν $G=(U,V,E)$, όπου οι κόμβοι κατατάσσονται σε δύο διακριτά μεταξύ τους σύνολα $U=(u_1, u_2,...,u_N)$ και $V=(v_1,v_2,...,v_K)$, ενώ οι ακμές αποτελούν τις συνδέσεις ανάμεσα στους κόμβους των δύο συνόλων. Οι διμερείς γράφοι αποτελούν βρίσκουν ευρεία εφαρμογή στα πεδία της Πληροφορικής και της Στατιστικής. Στην περίπτωση των βιολογικών δικτύων, χαρακτηριστικό παράδειγμα διμερούς γράφου αποτελούν τα δίκτυα ασθενειών (Goh et al., 2007), στα οποία η πρώτη κατηγορία κόμβων περιέχει τις ασθένειες, ενώ η δεύτερη περιέχει τα στοιχεία που σχετίζονται με αυτές (πρωτεΐνες, γονίδια, μεταβολικά προϊόντα κλπ).

2.2.3 Ιδιότητες δικτύων αλληλεπιδράσεων

Σημαντικό κομμάτι της Θεωρίας Γράφων είναι ο καθορισμός και υπολογισμός των ιδιοτήτων ενός δικτύου αλληλεπιδράσεων. Αυτές οι ιδιότητες καθορίζουν την εσωτερική οργάνωση του δικτύου, τη συνολική γεωμετρία του, την κατανομή των στοιχείων του και την ύπαρξη ή όχι σημαντικών στοιχείων που διατηρούν την ακεραιότητά του. Η ανάλυση αυτών των ιδιοτήτων, ιδιαίτερα στην περίπτωση των βιολογικών δικτύων αλληλεπιδράσεων, μπορεί να δώσει χρήσιμες πληροφορίες για την οργάνωση των βιομορίων στο κύτταρο και την πολυπλοκότητα των σχέσεων μεταξύ τους. Σε αυτήν την ενότητα παρουσιάζονται μια σειρά από ιδιότητες των δικτύων αλληλεπιδράσεων και περιγράφεται ο τρόπος υπολογισμού τους (Pavlopoulos et al., 2011). Καθώς η πλειονότητα των πρωτεϊνικών δικτύων αλληλεπιδράσεων

Κεφάλαιο II – Υλικά & Μέθοδοι

είναι μη κατευθυνόμενοι γράφοι, οι παρακάτω ιδιότητες παρουσιάζονται κυρίως με γνώμονα τα χαρακτηριστικά των τελευταίων.

Βαθμός κόμβου: ο βαθμός (k) ενός κόμβου είναι ο αριθμός των ακμών στις οποίες συμμετέχει ο κόμβος στο δίκτυο ή, εναλλακτικά, ο αριθμός των πρώτων γειτόνων του. Στην περίπτωση των μη κατευθυνόμενων γράφων, ο βαθμός του κόμβου είναι το σύνολο των ακμών στις οποίες συμμετέχει. Από την άλλη, στους κατευθυνόμενους γράφους μπορούν να οριστούν δύο βαθμοί για κάθε κόμβο, ένας για τις εξερχόμενες ακμές, οι οποίες ξεκινούν από το συγκεκριμένο κόμβο και κατευθύνονται σε άλλα σημεία του δικτύου και ένας για τις εισερχόμενες ακμές, οι οποίες έχουν σαν προορισμό το συγκεκριμένο κόμβο.

Συνδεσιμότητα και πυκνότητα δικτύου: Η συνδεσιμότητα και η πυκνότητα ενός δικτύου αποτελούν μέτρο της συνδεσιμότητας των στοιχείων του. Η ολική *συνδεσιμότητα* περιγράφει τον ελάχιστο αριθμό στοιχείων (κόμβων ή ακμών) που πρέπει να αφαιρεθούν ώστε τα υπόλοιπα περιεχόμενα του δικτύου να χωριστούν σε απομονωμένα υπο-δίκτυα. Η συνδεσιμότητα ορίζεται ως

$$c = \frac{E}{N(N-1)} \quad (73)$$

όπου E και N είναι ο συνολικός αριθμός των ακμών και των κόμβων, αντίστοιχα. Η *πυκνότητα* του δικτύου περιγράφει το πόσο αραιά ή πυκνά συνδεδεμένο είναι. Ορίζεται ως ο αριθμός των συνδέσεων για κάθε σύνολο κόμβων, σύμφωνα με τη συνάρτηση

$$D = \frac{2|E|}{|V|(|V|-1)} \quad (74)$$

όπου $|E|$ και $|V|$ είναι ο αριθμός των ακμών και των κόμβων. Οι τιμές της πυκνότητας κυμαίνονται από 0-1, με μεγαλύτερες τιμές να αντιστοιχούν σε πιο πυκνά δίκτυα. Η πυκνότητα του δικτύου εξαρτάται σε μεγάλο βαθμό από την τιμή του $|E|=O(|V|^k)$ με $1 < k < 2$. Οι αραιά συνδεδεμένοι γράφοι έχουν τιμές κοντά στο $|E| \sim |V|$, ενώ οι πυκνά συνδεδεμένοι γράφοι έχουν τιμές κοντά στο $|E| \sim |V|^2$. Σε γενικές γραμμές, τα βιολογικά δίκτυα αλληλεπιδράσεων θεωρούνται *αραιά συνδεδεμένα* ($D < 0.1$), ένα χαρακτηριστικό που τους προσδίδει εξελικτικό πλεονέκτημα για τη διατήρηση της ευρωστίας τους (Pavlopoulos et al., 2011).

Κεφάλαιο II – Υλικά & Μέθοδοι

Χαρακτηριστικό μήκος μονοπατιού και διάμετρος δικτύου: η διαδρομή από έναν κόμβο i σε έναν κόμβο j σε ένα δίκτυο, μέσω των ενδιάμεσων ακμών και κόμβων που παρεμβάλλονται μεταξύ τους, χαρακτηρίζεται ως *μονοπάτι*. Στην πλειονότητα των περιπτώσεων οι δύο κόμβοι μπορούν να συνδεόνται με πολλαπλά άλλα στοιχεία ανάμεσά τους, με αποτέλεσμα να υπάρχουν πολλοί συνδυασμοί διαδρομών ανάμεσά τους και, κατά συνέπεια, πολλαπλά μονοπάτια. Ωστόσο, για αξιολόγηση της *απόστασης* ανάμεσα στους κόμβους επιλέγεται το *συντομότερο μονοπάτι*, δηλαδή το μονοπάτι μέσω του οποίου η διαδρομή ανάμεσα στους κόμβους i και j περνά από το μικρότερο δυνατό αριθμό ενδιάμεσων κόμβων και ακμών. Για το συντομότερο μονοπάτι, η απόσταση ή το μήκος του, $\delta_{\min}(i,j)$ ορίζεται τυπικά ως ο αριθμός των ακμών που το συνθέτουν. Έτσι, για N αριθμό κόμβων, η τιμή θα είναι $\delta_{\min}=N-1$. Το *χαρακτηριστικό μήκος μονοπατιού* του δικτύου ορίζεται ως η διορθωμένη μέση τιμή όλων των συντομότερων μονοπατιών του, σύμφωνα με τη σχέση

$$\delta = \frac{1}{N(N-1)} \sum_{i,j=1, i \neq j}^N \delta_{\min}(i,j) \quad (75)$$

όπου N είναι ο αριθμός των κόμβων στο δίκτυο. Το χαρακτηριστικό μήκος μονοπατιού ουσιαστικά υποδηλώνει το μέσο αριθμό βημάτων που χρειάζονται για να επικοινωνήσουν δυο οποιοιδήποτε κόμβοι στο δίκτυο. Μια άλλη παράμετρος του δικτύου, η οποία υπολογίζεται επίσης μέσω των συντομότερων μονοπατιών, είναι η *διάμετρος* του δικτύου, η οποία ουσιαστικά είναι η μέγιστη τιμή του συντομότερου μονοπατιού.

Συντελεστής ομαδοποίησης: ο συντελεστής ομαδοποίησης (clustering) του δικτύου είναι ένα μέτρο της τάσης των στοιχείων του να οργανώνονται σε ομάδες. Μια ομάδα είναι ένα υποσύνολο των κόμβων του δικτύου που ενώνονται μεταξύ τους με πολλές ακμές, οι οποίες συνδέουν ισχυρά αυτούς τους κόμβους. Ο μέσος συντελεστής ομαδοποίησης του δικτύου ορίζεται ως

$$C = \frac{1}{N} \sum_{i=1}^N \frac{E_i}{k_i(k_i-1)} \quad (76)$$

όπου N είναι ο αριθμός των κόμβων στο δίκτυο, E_i είναι ο αριθμός των ακμών γύρω από έναν κόμβο i και k_i είναι ο βαθμός κάθε κόμβου. Τα βιολογικά δίκτυα έχουν σημαντικά υψηλότερο μέσο συντελεστή ομαδοποίησης σε σχέση με τα τυχαία δίκτυα, κάτι το οποίο αποδεικνύει τη

Κεφάλαιο II – Υλικά & Μέθοδοι

φύση τους να οργανώνονται σε ομάδες. Η παραπάνω παρατήρηση έχει και βιολογική σημασία, καθώς πολλές κυτταρικές λειτουργίες διενεργούνται από υποσύνολα βιολογικών μακρομορίων που αλληλεπιδρούν (Pavlopoulos et al., 2011).

Κεντρικότητες: Οι κεντρικότητες (centralities) αποτελούν ενδεικτικές μετρικές της οργάνωσης και της ιεράρχησης των δικτύων αλληλεπιδράσεων, καθώς παρέχουν πληροφορίες για τη σημασία της συνεισφοράς των κόμβων στο δίκτυο, τη συμμετοχή τους στην ακεραιότητα του δικτύου αλλά και την επίδραση της δυσλειτουργίας ή της αφαίρεσής τους στην οργάνωση και τη συνοχή του δικτύου (Pavlopoulos et al., 2011). Επιπρόσθετα, οι κεντρικότητες μπορούν να χρησιμοποιηθούν περαιτέρω σε πιο εξειδικευμένες αναλύσεις των συστατικών του δικτύου, όπως η εύρεση της τοπολογίας του αλλά και η αλγοριθμική ανάλυση ομαδοποίησης. Στη Θεωρία Γράφων υπάρχουν πολλές κατηγορίες κεντρικοτήτων, οι οποίες περιγράφουν διάφορα χαρακτηριστικά των κόμβων και των ακμών του δικτύου. Στη μελέτη των βιολογικών δικτύων και κυρίως των πρωτεϊνικών δικτύων αλληλεπιδράσεων χρησιμοποιούνται ευρέως η *κεντρικότητα βαθμού* (degree centrality), η *ενδιάμεση κεντρικότητα* (betweenness centrality) και η *κεντρικότητα εγγύτητας* (closeness centrality).

Η *κεντρικότητα βαθμού* (C_{deg}) υποδεικνύει κόμβους οι οποίοι εμπλέκονται σε μεγάλο αριθμό συνδέσεων. Η τιμή της για κάθε κόμβο i εξισώνεται ουσιαστικά με το βαθμό (k_i) αυτού του κόμβου. Κόμβοι με υψηλή κεντρικότητα βαθμού χαρακτηρίζονται σαν «κεντρικοί κόμβοι» (hubs), καθώς συνδέονται με μεγάλο αριθμό ακμών. Οι κεντρικοί κόμβοι διαδραματίζουν βασικό ρόλο καθώς, λόγω του μεγάλου αριθμού των ακμών στις οποίες συμμετέχουν, η πιθανή αφαίρεσή τους μπορεί να οδηγήσει στην μερική ή και πλήρη κατάρρευση του δικτύου. Στα βιολογικά δίκτυα η ύπαρξη κεντρικών κόμβων είναι ιδιαίτερα σημαντική, καθώς παρότι είναι ανθεκτικά απέναντι στις τυχαίες μεταβολές των κόμβων, η αφαίρεση κεντρικών κόμβων μπορεί να προκαλέσει αστοχία στο σύστημα (Levy & Siegal, 2008).

Η *ενδιάμεση κεντρικότητα* (C_b) υποδεικνύει σημαντικούς ενδιάμεσους κόμβους, οι οποίοι βρίσκονται ανάμεσα σε γειτονικούς κόμβους, κατατάσσονται υψηλότερα και έχουν σημαντική θέση μέσα στο δίκτυο. Χωρίς αυτούς τους κόμβους δεν θα υπήρχε τρόπος για δυο γείτονες να επικοινωνήσουν ο ένας με τον άλλο. Επομένως, η ενδιάμεση κεντρικότητα

Κεφάλαιο II – Υλικά & Μέθοδοι

αναπαριστά σημαντικούς κόμβους που βρίσκονται σε μεγάλο ποσοστό των μονοπατιών μεταξύ άλλων κόμβων του δικτύου. Η ενδιάμεση κεντρικότητα ενός κόμβου w , ανάμεσα σε δύο άλλους κόμβους i, j υπολογίζεται σύμφωνα με τη σχέση

$$C_b(w) = \sum_{i,j \in V} \frac{\sigma_{ij}(w)}{\sigma_{ij}} \quad (77)$$

όπου $\sigma_{ij}(w)$ είναι ο αριθμός των συντομότερων μονοπατιών ανάμεσα στους δύο κόμβους τα οποία διέρχονται μέσω του w , ενώ σ_{ij} είναι το σύνολο των συντομότερων μονοπατιών των δύο κόμβων γενικότερα. Οι κόμβοι με υψηλή ενδιάμεση κεντρικότητα χαρακτηρίζονται σαν «στενωποί» (bottlenecks) του δικτύου και λειτουργούν σαν σύνδεσμοι με απαραίτητες λειτουργικές και δυναμικές ιδιότητες (Barabasi et al., 2011). Σε βιολογικά δίκτυα οι στενωποί αναπαριστούν στοιχεία υπεύθυνα για τη σύνδεση πολλαπλών βιολογικών διεργασιών (π.χ. μεταβολίτες στους οποίους συναντώνται διαφορετικά μεταβολικά μονοπάτια).

Η *κεντρικότητα εγγύτητας* (C_{cl}), τέλος, υποδεικνύει σημαντικούς κόμβους οι οποίοι μπορούν να επικοινωνήσουν «γρήγορα» με άλλους κόμβους του δικτύου. Σε ένα μη κατευθυνόμενο γράφο, η τιμή της υπολογίζεται σύμφωνα με τη σχέση

$$C_{cl} = \frac{1}{\sum_{i,j} \delta_{min}(i,j)} \quad (78)$$

όπου το δ_{min} υποδηλώνει την απόσταση του συντομότερου μονοπατιού ανάμεσα στους κόμβους i και j . Κόμβοι με υψηλή κεντρικότητα εγγύτητας είναι γενικά, σε μικρότερη απόσταση με την πλειονότητα των άλλων κόμβων του δικτύου και μπορούν να επικοινωνήσουν μαζί τους χρησιμοποιώντας μονοπάτια με μικρό μήκος. Στα βιολογικά δίκτυα, η κεντρικότητα εγγύτητας μπορεί να χρησιμοποιηθεί για την εύρεση των πιο κεντρικών μεταβολιτών σε ένα κύτταρο. Έτσι, οι μετρήσεις εγγύτητας έχουν εφαρμοστεί στη σύγκριση μονοκύτταρων και πολυκύτταρων ευκαρυωτικών οργανισμών ως προς την εξέλιξη των μεταβολικών και βιοχημικών μονοπατιών τους (Mazurie et al., 2010).

Τοπολογία του γράφου: Η τοπολογία ενός γράφου αποτελεί ένδειξη της γεωμετρίας του, καθώς και του τρόπου κατανομής των στοιχείων του. Οι τρεις μεγάλες κατηγορίες τοπολογίας δικτύων είναι τα *τυχαία δίκτυα*, τα *δίκτυα μικρού κόσμου* και τα *ανεξάρτητα από κλίμακα δίκτυα*. Το πιο γνωστό μοντέλο *τυχαίων δικτύων* είναι το μοντέλο Erdős-Rényi. Σε

Κεφάλαιο II – Υλικά & Μέθοδοι

αυτό το μοντέλο, ένα τυχαίο δίκτυο αποτελείται από N αριθμό κόμβων που συνδέονται από $N(N-1)/2$ αριθμό πιθανών ακμών τυχαία. Τυχαία δίκτυα αυτού του τύπου είναι θεωρητικά, και χρησιμοποιούνται περισσότερο σαν μέτρο σύγκρισης άλλων δικτύων, προκειμένου να εξακριβωθεί κατά πόσο η οργάνωσή τους είναι τυχαία ή όχι (Barabasi & Albert, 1999). Τα δίκτυα μικρού κόσμου είναι δίκτυα στα οποία οι περισσότεροι κόμβοι δεν σχετίζονται άμεσα μεταξύ τους, ωστόσο οι γείτονες των μη συνδεδεμένων κόμβων έχουν μεγάλη πιθανότητα να συνδέονται άμεσα μεταξύ τους και οι περισσότεροι κόμβοι του δικτύου μπορούν να επικοινωνήσουν μεταξύ τους με μικρό αριθμό βημάτων. Η τοπολογία τους αντιστοιχείται κυρίως σε κάποιες κατηγορίες κοινωνικών δικτύων, καθώς και σε κάποια είδη βιολογικών δικτύων όπως τα δίκτυα μερικών μεταβολικών μονοπατιών (Watts & Strogatz, 1998). Τέλος, τα *ανεξάρτητα από κλίμακα δίκτυα* αντιστοιχούν ουσιαστικά στα δίκτυα πραγματικού κόσμου, και περιγράφουν τόσο την οργάνωση φυσικών δικτύων όπως ο Παγκόσμιος Ιστός όσο και την πλειονότητα των βιολογικών δικτύων αλληλεπιδράσεων (Barabasi et al., 2011; Rajula et al., 2018). Ιδιαίτερο χαρακτηριστικό των ανεξάρτητων από κλίμακα δικτύων είναι η κατανομή του βαθμού των κόμβων που τα αποτελούν, η οποία ακολουθεί την κατανομή νόμου δύναμης

$$P(k) \sim k^{-\gamma} \quad (79)$$

όπου k είναι ο βαθμός του κόμβου και γ είναι μια πραγματική σταθερά. Τιμές του γ μικρότερες του 2 ($\gamma < 2$) υποδεικνύουν ότι ο ρόλος των κεντρικών κόμβων του δικτύου είναι σημαντικός στη λειτουργία του, καθώς πιθανή καταστροφή ή δυσλειτουργία αυτών των κόμβων μπορεί να οδηγήσει σε κατάρρευση του δικτύου (Pavlopoulos et al., 2011).

2.2.4 Αλγόριθμοι Ομαδοποίησης

Τα δίκτυα αλληλεπιδράσεων και, ειδικότερα, τα βιολογικά δίκτυα αλληλεπιδράσεων, συνήθως αποτελούνται από στοιχεία τα οποία μπορούν να οργανωθούν περαιτέρω σε υποδίκτυα ή ομάδες (clusters). Χαρακτηριστικό στοιχείο αυτών των ομάδων είναι ότι οι κόμβοι που τα αποτελούν σχετίζονται περισσότερο μεταξύ τους από ότι με τα υπόλοιπα μέρη του δικτύου. Η διαδικασία της δημιουργίας ομάδων σε ένα δίκτυο είναι γνωστή με τον όρο ανάλυση ομαδοποίησης (clustering). Ένα μέτρο ένδειξης για την ικανότητα ομαδοποίησης των

Κεφάλαιο II – Υλικά & Μέθοδοι

στοιχείων ενός δικτύου είναι ο συντελεστής ομαδοποίησης, ο οποίος περιγράφηκε στην Ενότητα 2.2.1.2. Επιπρόσθετα, η ομαδοποίηση μπορεί συχνά να πραγματοποιηθεί ποιοτικά, με τη χρήση χαρακτηριστικών όπως π.χ. η κοινή έκφραση γονιδίων σε συγκεκριμένους ιστούς, ή ο εντοπισμός πρωτεϊνών σε συγκεκριμένες υποκυτταρικές θέσεις. Ωστόσο, στη θεωρία δικτύων συνήθως επιχειρείται η υπολογιστική ανάλυση της ομαδοποίησης των δικτύων, μέσα από τη χρήση αλγορίθμων. Έχουν προταθεί διάφοροι αλγόριθμοι ομαδοποίησης, με αρκετούς να χρησιμοποιούνται ευρέως και στη μελέτη των βιολογικών δικτύων αλληλεπιδράσεων. Δύο πολύ συχνά χρησιμοποιούμενοι αλγόριθμοι, οι οποίοι έχουν εφαρμοστεί και στην παρούσα διπλωματική εργασία, είναι οι αλγόριθμοι Markov Cluster (Enright et al., 2002; van Dongen, 2000) και Girvan-Newman (Girvan & Newman, 2002), οι οποίοι παρουσιάζονται συνοπτικά παρακάτω.

Ο αλγόριθμος Markov Cluster (MCL): Ο αλγόριθμος Markov Cluster (MCL) είναι μια μέθοδος μη επιβλεπόμενης ομαδοποίησης, η οποία προσομοιάζει στοχαστικά τη ροή πληροφοριών σε έναν γράφο μέσα από μια επαναληπτική διαδικασία. Ο MCL εντοπίζει ομάδες μέσα σε ένα δίκτυο $G=(V,E)$ υπολογίζοντας τις πιθανότητες τυχαίων διαδρομών στο γράφο και, χρησιμοποιώντας δύο εναλλασσόμενους τελεστές και στοχαστικούς πίνακες γειτνίασης (πίνακες Markov), παράγει αλληλομετατρεπόμενα σύνολα πιθανοτήτων. Οι δύο τελεστές που χρησιμοποιούνται καλούνται διαστολή (expansion) και διόγκωση (inflation). Η διαστολή συμπίπτει με την ύψωση ενός στοχαστικού πίνακα σε κάποια δύναμη (π.χ. στο τετράγωνο), ενώ η διόγκωση αντιστοιχεί στην ύψωση του προϊόντος της διαστολής σε δύναμη Hadamard, που ακολουθείται από έναν μετασχηματισμό του πίνακα σε στοχαστικό. Η διαστολή ουσιαστικά αντιστοιχεί στους τυχαίες διαδρομές μεγάλου μήκους, οι οποίες απαντώνται συχνότερα εντός μίας ομάδας από ότι μεταξύ διαφορετικών ομάδων, με αποτέλεσμα οι πιθανότητες των ζευγών κόμβων που βρίσκονται εντός μιας ομάδας να είναι υψηλότερες. Η διόγκωση ενισχύει τις πιθανότητες των διαδρομών εντός μιας ομάδας και ταυτόχρονα θα υποβιβάσει τις αντίστοιχες πιθανότητες των διαδρομών ανάμεσα σε διαφορετικές ομάδες. Ολόκληρη η διαδικασία που περιγράφηκε πραγματοποιείται χωρίς πρότερη γνώση του αριθμού των ομάδων ή της δομής αυτών. Το τελικό αποτέλεσμα είναι η

Κεφάλαιο II – Υλικά & Μέθοδοι

ανάδειξη της δομής των ομάδων που υπάρχουν στο δίκτυο, αποκόπτοντας ακμές μεταξύ κόμβων που δεν συνδέονται επαρκώς (Enright et al., 2002; van Dongen, 2000).

Ο αλγόριθμος ανάλυσης κοινοτήτων Girvan-Newman: η μέθοδος ομαδοποίησης των Girvan και Newman περιλαμβάνει την οργάνωση των στοιχείων ενός δικτύου σε κοινότητες, με βάση τη συνδεσιμότητά τους. Η ομαδοποίηση του αλγορίθμου χρησιμοποιεί σαν μέτρο την ενδιάμεση κεντρικότητα των κόμβων. Μέσω του υπολογισμού της ενδιάμεσης κεντρικότητας ο αλγόριθμος Girvan – Newman οργανώνει τα στοιχεία του δικτύου σε κοινότητες. Συγκεκριμένα, η μέθοδος πραγματοποιεί μια αναζήτηση «από πάνω προς τα κάτω» (top-down), μέσω της οποίας εντοπίζει σε κάθε κύκλο την ακμή με την υψηλότερη κεντρικότητα, την αφαιρεί από το μη ομαδοποιημένο δίκτυο και υπολογίζει εκ νέου τις τιμές κεντρικότητας για τις υπόλοιπες ακμές, επαναλαμβάνοντας τη διαδικασία ώσπου να μη μείνουν καθόλου μη ομαδοποιημένες ακμές στο δίκτυο. Το αποτέλεσμα είναι η οργάνωση των στοιχείων του δικτύου σε κοινότητες, οι οποίες κατατάσσονται περαιτέρω ιεραρχικά ως προς τη συνδεσιμότητα μεταξύ τους μέσω ιεραρχικής ομαδοποίησης (Girvan & Newman, 2002). Σε αντίθεση με τον MCL, ο αλγόριθμος Girvan-Newman δεν διακόπτει τις ακμές ανάμεσα σε διαφορετικές κοινότητες. Αντίθετα, οι κόμβοι αυτών των ακμών χαρακτηρίζονται σαν «κρίσιμοι κόμβοι» (critical nodes), διαθέτουν υψηλές τιμές ενδιάμεσης κεντρικότητας και λειτουργούν σαν στενωποί για την επικοινωνία ανάμεσα σε διαφορετικές κοινότητες.

2.2.5 Ανάλυση λειτουργικού εμπλουτισμού

Στη μελέτη των βιολογικών δικτύων αλληλεπιδράσεων, ο λειτουργικός χαρακτηρισμός των στοιχείων του δικτύου είναι εξίσου σημαντικός με τη μαθηματική ανάλυση των χαρακτηριστικών τους. Η ανάλυση των λειτουργικών χαρακτηριστικών των δικτύων περιλαμβάνει κυρίως την εύρεση κοινών λειτουργιών στους κόμβους με στατιστικά σημαντική συσχέτιση, μια διαδικασία γνωστή σαν «Λειτουργικός Εμπλουτισμός» (functional enrichment). Έτσι, αν σε μια ομάδα πρωτεϊνών βρεθεί στατιστικά σημαντική εκπροσώπηση ενός όρου π.χ. γονιδιακής οντολογίας, αυτός ο όρος μπορεί να θεωρηθεί ότι αντιπροσωπεύει το σύνολο των

Κεφάλαιο II – Υλικά & Μέθοδοι

στοιχείων αυτής της ομάδας και να τη χαρακτηρίσει λειτουργικά. Η ανάλυση λειτουργικού εμπλουτισμού, τυπικά, πραγματοποιείται χρησιμοποιώντας όρους γονιδιακής οντολογίας από τη βάση δεδομένων Gene Ontology (Ashburner et al., 2000). Ωστόσο, αντίστοιχη ανάλυση εμπλουτισμού μπορεί να πραγματοποιηθεί και για άλλες κατηγορίες δεδομένων, όπως τα μεταβολικά μονοπάτια και η συσχέτιση με ασθένειες και φάρμακα.

2.2.6 Ανάλυση δυναμικών δικτύων αλληλεπιδράσεων σε προσομοιώσεις Μοριακής Δυναμικής

Η ανάλυση δικτύων έχει χρησιμοποιηθεί κυρίως για την περιγραφή μεγάλου όγκου δεδομένων, στα οποία κάθε κόμβος αντιστοιχεί σε μια ξεχωριστή οντότητα, π.χ. ένα γονίδιο ή μια πρωτεΐνη. Ωστόσο, η απεικόνιση δικτύων μπορεί να χρησιμοποιηθεί και για την περιγραφή αλληλεπιδράσεων στο επίπεδο της δομής των βιομορίων, ενώ επιπλέον πληροφορία για τη σημαντικότητα των αλληλεπιδράσεων μπορεί να δοθεί από δεδομένα κινητικότητας ή από εξελικτική πληροφορία. Η ανάλυση δυναμικών δικτύων (Dynamical Network Analysis) εφαρμόζει στοιχεία της θεωρίας δικτύων στη μελέτη βιομοριακών δομών, οπτικοποιώντας μια δομή σαν δίκτυο κόμβων συνδεόμενων από ακμές, τα χαρακτηριστικά του οποίου σταθμίζονται με βάση τη δυναμική συμπεριφορά του βιομορίου (Eargle & Luthey-Schulten, 2012; Sethi et al., 2009). Σε ένα δομικό δίκτυο αλληλεπιδράσεων οι κόμβοι αντιπροσωπεύουν άτομα, μόρια, νουκλεοτίδια ή αμινοξικά κατάλοιπα, ενώ οι ακμές που συνδέουν τους κόμβους αντιπροσωπεύουν τις μη ομοιοπολικές αλληλεπιδράσεις. Επιπρόσθετη πληροφορία για την κατασκευή και την ανάλυση του δικτύου παρέχεται από δεδομένα προερχόμενα από μελέτες της δυναμικής συμπεριφοράς των βιομορίων, όπως οι προσομοιώσεις Μοριακής Δυναμικής ή οι δομές NMR, τα αποτελέσματα των οποίων μπορούν να χρησιμοποιηθούν για την στάθμιση και αξιολόγηση της σημαντικότητας των αλληλεπιδράσεων στο δίκτυο, αλλά και για την ομαδοποίηση των στοιχείων του δικτύου με βάση τα χαρακτηριστικά τους. Μέσω της ανάλυσης δικτύων για αποτελέσματα Μοριακής Δυναμικής γίνεται δυνατή η παρατήρηση βιολογικά σημαντικών συσχετίσεων στις κινήσεις των βιομορίων, η αναγνώριση ισχυρών αλληλεπιδράσεων ανάμεσα στα αλληλεπιδρώντα βιομόρια αλλά και η ανίχνευση πιθανών

Κεφάλαιο II – Υλικά & Μέθοδοι

αλλοστερικών μηχανισμών μέσω των οποίων μπορούν να ρυθμιστούν σημαντικές βιολογικές διεργασίες (Sethi et al., 2009).

Τα δυναμικά δομικά δίκτυα αλληλεπιδράσεων είναι σταθμισμένα δίκτυα. Οι ακμές ανάμεσα στους κόμβους σταθμίζονται με τιμές βαρών, οι οποίες χαρακτηρίζουν την ισχύ των αλληλεπιδράσεων. Οι τιμές των βαρών υπολογίστηκαν με βάση τα δεδομένα μοριακής δυναμικής. Το βάρος μιας ακμής ανάμεσα σε δυο κόμβους i και j (w_{ij}) ορίζεται σαν συνάρτηση της αντίστοιχη συσχέτισης της κινητικότητας των καταλοίπων i και j . Η κατά ζεύγη συσχέτιση υπολογίζεται ως η κανονικοποιημένη συνδιακύμανση της κίνησης των i και j (C_{ij}). Έχοντας υπολογίσει τη συσχέτιση μεταξύ των i και j , η τιμή του βάρους w_{ij} της ακμής ανάμεσα στους δύο κόμβους ορίζεται ως

$$w_{ij} = -\log(|C_{ij}|) \quad (80)$$

Έτσι, οι ακμές με τα μικρότερα βάρη αντιστοιχούν στις πιο συσχετισμένες και πιο ισχυρές αλληλεπιδράσεις του δικτύου. Στην ανάλυση δυναμικών δικτύων, τέλος, μπορεί να πραγματοποιηθεί ανάλυση ομαδοποίησης, χρησιμοποιώντας, τυπικά, τον αλγόριθμο οργάνωσης κοινοτήτων Girvan-Newman (Girvan & Newman, 2002). Στα πλαίσια ενός δικτύου Μοριακής Δυναμικής, οι κοινότητες του δικτύου μπορούν να αντιστοιχιστούν με δομικά στοιχεία που κινούνται συντονισμένα, και μπορούν να χρησιμοποιηθούν για την περιγραφή δομικών και κινητικών χαρακτηριστικών που σχετίζονται με τη λειτουργικότητα μιας πρωτεΐνης (Sethi et al., 2009).

2.2.7 Προγράμματα οπτικοποίησης και ανάλυσης δικτύων αλληλεπιδράσεων

Στην επιστημονική κοινότητα έχουν αναπτυχθεί μια σειρά από αλγορίθμους και προγράμματα για την κατασκευή, την οπτικοποίηση και την ανάλυση δικτύων αλληλεπιδράσεων. Κάποια από αυτά τα προγράμματα επικεντρώνονται σε συγκεκριμένες κατηγορίες δικτύων, ενώ άλλα διαθέτουν πιο γενικές λειτουργίες. Στην παρούσα διδακτορική διατριβή έχουν χρησιμοποιηθεί τρία προγράμματα ανάλυσης δικτύων, τα Cytoscape (Shannon et al., 2003), NetworkView (Eargle & Luthey-Schulten, 2012) και NetworkX (Hagberg et al., 2008), τα οποία παρουσιάζονται συνοπτικά παρακάτω:

Κεφάλαιο II – Υλικά & Μέθοδοι

Cytoscape: Το Cytoscape είναι ένα από τα πιο γνωστά και ευρέως χρησιμοποιούμενα προγράμματα οπτικοποίησης και ανάλυσης δικτύων. Είναι μια ανοιχτού κώδικα πλατφόρμα λογισμικού που χρησιμοποιείται για την οπτικοποίηση και το σχολιασμό δικτύων μοριακών αλληλεπιδράσεων και βιολογικών μονοπατιών, ικανή για τη μελέτη μεγάλων δικτύων αλληλεπιδράσεων, αποτελούμενων από χιλιάδες ή και εκατομμύρια κόμβους και ακμές. Ιδιαίτερα προτερήματα του Cytoscape αποτελούν η ικανότητά του να αναγνωρίζει μεγάλη ποικιλία τύπων αρχείων και μορφών δικτύων, καθώς και η πλούσια βιβλιοθήκη του από εφαρμογές για την ανάλυση δικτύων, η οποία περιλαμβάνει εργαλεία για ανάλυση με τη θεωρία γράφων, για ομαδοποίηση με πλήθος αλγορίθμων, αλλά και το λειτουργικό εμπλουτισμό των δικτύων. Επιπρόσθετα, η ιδιότητά του σαν λογισμικό ανοιχτού κώδικα επιτρέπει το σχεδιασμό και την προσθήκη νέων λειτουργιών από το χρήστη, με αποτέλεσμα να υπάρχουν πλέον εκατοντάδες διαθέσιμες εφαρμογές για αυτό, σχεδιασμένες από την επιστημονική κοινότητα (Shannon et al., 2003).

NetworkView: Το NetworkView είναι μια εφαρμογή για το σχεδιασμό, την οπτικοποίηση και την ανάλυση δομικών δικτύων αλληλεπιδράσεων από αποτελέσματα προσομοιώσεων Μοριακής Δυναμικής, η οποία βρίσκεται ενσωματωμένη στο πρόγραμμα μοριακών γραφικών Visual Molecular Dynamics (VMD) (Humphrey et al., 1996). Το NetworkView υποστηρίζει την κατασκευή δικτύων από προσομοιώσεις πρωτεϊνών και νουκλεϊκών οξέων (Sethi et al., 2009), τη χρήση δεδομένων από προσομοιώσεις για τον υπολογισμό της ισχύος των αλληλεπιδράσεων και τον καθορισμό των τιμών βάρους των ακμών, καθώς και την ανάλυση ομαδοποίησης αυτών των δικτύων με τον αλγόριθμο Girvan-Newman (Girvan & Newman, 2002). Επιπρόσθετα, επιτρέπει τη συγγραφή προγραμμάτων από το χρήστη για την περαιτέρω ανάλυση των δικτύων, μέσω της γλώσσας προγραμματισμού Tcl (Eargle & Luthey-Schulten, 2012).

NetworkX: Το NetworkX είναι μια βιβλιοθήκη εντολών για τη γλώσσα προγραμματισμού Python, μέσω της οποίας ο χρήστης μπορεί να κατασκευάσει και να

Κεφάλαιο II – Υλικά & Μέθοδοι

αναλύσει δίκτυα αλληλεπιδράσεων προγραμματιστικά. Διαθέτει υλοποιημένες λειτουργίες για τη δημιουργία δικτύων, τόσο τυχαίων όσο και από πραγματικά βιολογικά δεδομένα και υποστηρίζει την κατασκευή όλων των κατηγοριών δικτύων (κατευθυνόμενα, μη κατευθυνόμενα, σταθμισμένα κλπ) καθώς και όλων των κατηγοριών τοπολογίας (τυχαία, ανεξάρτητα από κλίμακα και δίκτυα μικρού κόσμου). Επιπρόσθετα, περιλαμβάνει λειτουργίες για την ανάλυση δικτύων μέσω της θεωρίας γράφων και την ανάλυση ομαδοποίησης μέσα από διάφορους αλγορίθμους.

2.2.8 Βάσεις δεδομένων πρωτεϊνικών αλληλεπιδράσεων

Η κατασκευή ενός δικτύου αλληλεπιδράσεων απαιτεί την ύπαρξη δεδομένων για αυτές τις αλληλεπιδράσεις, είτε πρόκειται για αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών είτε πρόκειται για δεδομένα μεταβολικών οδών, προτύπων γονιδιακής έκφρασης ή μονοπατιών μεταγωγής σήματος. Ιδανικά, αυτά τα δεδομένα μπορούν να ληφθούν άμεσα από πειραματικές μελέτες. Ωστόσο, η δυνατότητα πρόσβασης στις υποδομές για τη διεξαγωγή πειραμάτων δεν είναι πάντα δυνατή. Έτσι, πολύ συχνά οι μελέτες δικτύων αλληλεπιδράσεων είναι αποκλειστικά υπολογιστικές και βασίζονται σε υπάρχοντα δεδομένα, προερχόμενα είτε άμεσα από τη βιβλιογραφία είτε έμμεσα από βάσεις δεδομένων. Σε κάποιες περιπτώσεις τα δεδομένα λαμβάνονται άμεσα από τη βιβλιογραφία μέσα από προσεκτική αναζήτηση και εκτεταμένη μελέτη των ανάλογων δημοσιεύσεων ή αυτοματοποιημένα με μεθόδους εξόρυξης κειμένου (text mining). Ωστόσο, η παραπάνω διαδικασία συνήθως είναι χρονοβόρα και δύσκολη. Κατά συνέπεια, οι ερευνητές που ασχολούνται με τη μελέτη των δικτύων αλληλεπιδράσεων καταφεύγουν στις περισσότερες περιπτώσεις σε εξειδικευμένες βάσεις δεδομένων, οι οποίες περιέχουν κωδικοποιημένη την πληροφορία που απαιτείται τόσο για την κατασκευή των δικτύων αλληλεπιδράσεων όσο και για το λειτουργικό σχολιασμό τους. Στη βιβλιογραφία υπάρχουν διαθέσιμες πολλές βάσεις δεδομένων οι οποίες προσφέρουν πληροφορίες για βιομοριακές αλληλεπιδράσεις. Στην περίπτωση των πρωτεϊνικών αλληλεπιδράσεων, οι τρεις πιο διαδεδομένες βάσεις δεδομένων είναι η IntAct (Orchard et al., 2014), η BioGRID και η STRING.

Κεφάλαιο II – Υλικά & Μέθοδοι

Η βάση δεδομένων IntAct (<https://www.ebi.ac.uk/intact/>) είναι μία ελεύθερα διαθέσιμη βάση μοριακών αλληλεπιδράσεων, με καλά σχολιασμένες εγγραφές που προκύπτουν από πειράματα που έχουν καταγραφεί στη βιβλιογραφία. Κάθε πρωτεϊνική αλληλεπίδραση στην IntAct είναι πειραματικά επιβεβαιωμένη, ενώ η πιστότητα των δεδομένων ελέγχεται συνεχώς, μέσα από την εφαρμογή μιας σειράς κριτηρίων για την αξιολόγηση των πειραματικών διεργασιών που έχουν ακολουθηθεί (Orchard et al., 2014). Η βάση δεδομένων BioGRID (<https://thebiogrid.org/>) είναι βάση τόσο γενετικών όσο και πρωτεϊνικών αλληλεπιδράσεων διαφόρων οργανισμών-μοντέλων, μεταξύ των οποίων και του ανθρώπου. Περιλαμβάνει τόσο πειραματικά επιβεβαιωμένες αλληλεπιδράσεις, προερχόμενες από πειράματα μεγάλης κλίμακας, όσο και αποτελέσματα από εξόρυξη κειμένου αλλά και αναθέσεις αλληλεπιδράσεων με υπολογιστική ανάθεση όρων γονιδιακής οντολογίας, ή μέσω ομολογίας μεταξύ διαφορετικών ειδών (Chatr-Aryamontri et al., 2017). Τέλος, η βάση STRING (<https://string-db.org/>) συνδυάζει γνωστά δεδομένα αλληλεπιδράσεων από τις παραπάνω βάσεις, τα οποία εμπλουτίζει με πληροφορία από εξόρυξη κειμένου καθώς και από άλλες βάσεις σημαντικής βιολογικής γνώσης, όπως π.χ. είναι οι βάσεις μοριακής σηματοδότησης (Szkarczyk et al., 2017). Από τις παραπάνω βάσεις, η IntAct διαθέτει τη μικρότερη πληροφορία, ωστόσο, τα δεδομένα της, καθώς ελέγχονται και επιβεβαιώνονται χειροκίνητα, διαθέτουν σημαντικά μεγαλύτερη πιστότητα. Από την άλλη, οι βάσεις BioGRID και STRING διαθέτουν σημαντικά μεγαλύτερο όγκο αλληλεπιδράσεων. Ωστόσο, καθώς μεγάλο μέρος των αλληλεπιδράσεών τους προκύπτει από υπολογιστικές μεθόδους πρόγνωσης ή ανάλυσης, η αξιοπιστία των δεδομένων τους είναι μικρότερη.

2.3 Κρυφά Μοντέλα Markov

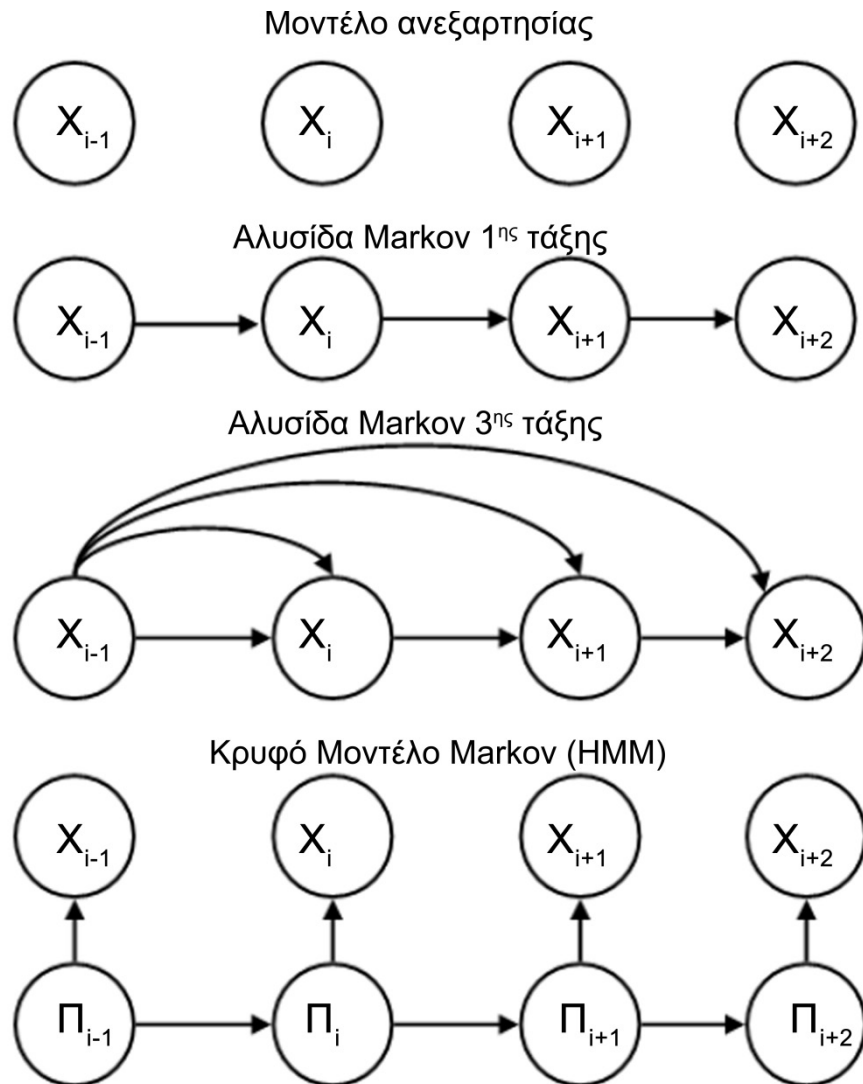
Οι βιολογικές ακολουθίες (πρωτεΐνες, νουκλεϊκά οξέα) μπορούν συνήθως να κατηγοριοποιηθούν σε οικογένειες συγγενικών αλληλουχιών και δομών (Henikoff et al., 1997). Διαφορετικά κατάλοιπα σε μια λειτουργική αλληλουχία υπόκεινται σε διαφορετικές εξελικτικές πιέσεις. Πολλαπλές στοιχίσεις μιας οικογένειας ακολουθιών αποκαλύπτουν αυτό το χαρακτηριστικό στο πρότυπο συντήρησής τους, καθώς κάποιες θέσεις στην αλληλουχία είναι πιο συντηρημένες σε σχέση με άλλες, ενώ κάποιες περιοχές μιας πολλαπλής στοίχισης φαίνεται να «ανέχονται» παρεμβολές ή εξαλείψεις περισσότερο από άλλες. Για την εξειδικευμένη και ακριβή μελέτη και αναζήτηση αυτών των χαρακτηριστικών απαιτείται συχνά η χρήση πληροφορίας ειδικής για κάθε θέση των στοιχίσεων, μια διαδικασία που συνήθως πραγματοποιείται με τη δημιουργία προφίλ (Gribskov et al., 1987). Μια από τις πιο συχνά χρησιμοποιούμενες μεθόδους στο συγκεκριμένο τομέα είναι τα μοντέλα Markov και οι παραλλαγές τους, τα κρυφά μοντέλα Markov (Hidden Markov Models, HMMs) και τα προφίλ HMMs (profile HMMs, pHMMs).

2.3.1 Αλυσίδες Markov και κρυφά μοντέλα Markov (HMMs)

Τα μοντέλα Markov (Markov models) ή αλυσίδες Markov (Markov chains) στηρίζονται στην ιδέα ότι κάθε ένα από τα ενδεχόμενα εξαρτάται από το αμέσως προηγούμενό του, ή αλλιώς το κάθε ενδεχόμενο καθορίζει με κάποια πιθανότητα το αμέσως επόμενο του. Η παραπάνω ιδέα έχει τις απαρχές της στην ανάλυση μοτίβων πάνω στις φυσικές γλώσσες που χρησιμοποιούνται στην ανθρώπινη επικοινωνία. Για παράδειγμα, στην αγγλική γλώσσα το γράμμα Q ακολουθείται σχεδόν πάντοτε από το U. Συνεπώς, η πιθανότητα να εμφανιστεί το U σε μια θέση δεν είναι πάντα ίδια αλλά εξαρτάται από το αν προηγήθηκε το Q. Αντίστοιχες αρχές φαίνονται να ακολουθούνται όχι μόνο από τις ανθρώπινες γλώσσες αλλά και από πολλά συστήματα κωδικοποιημένης πληροφορίας, συμπεριλαμβανομένων και των αλληλουχιών μακρομορίων όπως το DNA και οι πρωτεΐνες. Κατά συνέπεια, τα μοντέλα Markov και οι διάφορες παραλλαγές τους συγκαταλέγονται ανάμεσα στις πιο ικανές μεθόδους ανάλυσης

Κεφάλαιο II – Υλικά & Μέθοδοι

ακολουθιών και έχουν χρησιμοποιηθεί επανειλημμένα στην κατασκευή αλγορίθμων ανάλυσης και πρόγνωσης μεγάλου εύρους χαρακτηριστικών των βιολογικών μακρομορίων, από την πρόγνωση της δομής και της οργάνωσης των γονιδίων ως την πρόγνωση της τοπολογίας και της δομής των διαμεμβρανικών πρωτεϊνών (Μπάγκος, 2015).



Εικόνα 40 Σύγκριση του μοντέλου ανεξαρτησίας με τις αλυσίδες Markov και τα HMMs. Σε ένα ανεξάρτητο μοντέλο θεωρείται ότι κάθε ενδεχόμενο X είναι ανεξάρτητο από τα υπόλοιπα. Μια αλυσίδα Markov 1^{ης} Τάξης ορίζει ότι ένα ενδεχόμενο X_i εξαρτάται από το αμέσως προηγούμενό του και επηρεάζει το αμέσως επόμενο του. Αλυσίδες Markov ανώτερης τάξης (π.χ. 3^{ης} Τάξης) ορίζουν ότι ένα ενδεχόμενο μπορεί να επηρεάζεται από περισσότερα του ενός προηγούμενα ενδεχόμενα. Τέλος, σε ένα HMM οι ιδιότητες των αλυσίδων Markov δεν ισχύουν για τα ίδια τα ενδεχόμενα X , τα οποία παραμένουν κρυφά, αλλά για τις καταστάσεις Π .

Κεφάλαιο II – Υλικά & Μέθοδοι

Η προτυποποίηση μιας ακολουθίας ενδεχομένων με τη μορφή αλυσίδας Markov στηρίζεται στην ιδέα ότι το κάθε ενδεχόμενο καθορίζει, με κάποια πιθανότητα το αμέσως επόμενο του. Με βάση αυτήν την εξάρτηση, η ακολουθία ενδεχομένων ουσιαστικά αποτελεί μια αλυσίδα Markov 1^{ης} τάξης (Εικόνα 40). Αυτή η εξάρτηση μπορεί, ωστόσο, να επεκταθεί και σε περισσότερα του ενός (2,3,4,...,N) προηγούμενα ενδεχόμενα. Σε αυτήν την περίπτωση το μοντέλο που περιγράφει την ακολουθία επεκτείνεται, και πλέον αναφερόμαστε σε αλυσίδες Markov 2^{ης}, 3^{ης}, 4^{ης} τάξης κλπ. Στην περίπτωση των βιολογικών ακολουθιών, ως καταστάσεις ορίζονται τα σύμβολα της ακολουθίας τα οποία ανήκουν σε ένα πεπερασμένο αλφάβητο (τα τέσσερα νουκλεοτίδια στην περίπτωση του DNA ή τα 20 αμινοξέα στην περίπτωση των πρωτεϊνών).

Για παράδειγμα, έστω ότι ορίζουμε μια πρωτεϊνική ακολουθία X μήκους L της μορφής $X=(x_1, x_2, \dots, x_{L-1}, x_L)$. Αν θεωρήσουμε την κατανομή των καταλοίπων x_i σε κάθε θέση i κατά μήκος της αλληλουχίας σαν τυχαία μεταβλητή, τότε μπορούμε να ορίσουμε μια αλυσίδα Markov 1^{ης} τάξης, η οποία να περιγράφει αυτήν την αλληλουχία ως μια στοχαστική ανέλιξη διακριτών καταστάσεων σε διακριτό χρόνο. Αυτή η ανέλιξη ουσιαστικά αποτελείται από την ακολουθία των τυχαίων μεταβλητών x , η οποία παίρνει τιμές σε ένα «χώρο καταστάσεων» οριζόμενο από το συγκεκριμένο αλφάβητο που χρησιμοποιείται στην αλληλουχία. Η Μαρκοβιανή ιδιότητα της αλυσίδας ορίζει ότι για κάθε παρατήρηση x_i , η δεσμευμένη κατανομή των «μελλοντικών» παρατηρήσεων x_{i+1} , x_{i+2} , x_{i+3} , δεδομένων των προηγούμενων παρατηρήσεων $x_1, x_2, x_3, \dots, x_{i-1}, x_i$, εξαρτάται από το παρελθόν μόνο μέσω του x_i . Αυτή η Μαρκοβιανή ιδιότητα μπορεί να παρασταθεί με την παρακάτω σχέση πιθανοτήτων για κάθε θέση της ακολουθίας i :

$$P(x_i \vee x_{i-1}, x_{i-2}, \dots, x_1) = P(x_i \vee x_{i-1}) \quad (81)$$

Μια διακριτή αλυσίδα Markov χαρακτηρίζεται από έναν πίνακα πιθανοτήτων αυτού του τύπου, γνωστές σαν πιθανότητες μετάβασης (transition probabilities), οι οποίες περιγράφουν την πιθανότητα το σύστημα να μεταβεί από μία κατάσταση του μοντέλου i , στην επόμενη κατάσταση $i+1$ (Μπάγκος, 2015).

Η πιο συχνή εφαρμογή των μοντέλων Markov στη Βιοπληροφορική είναι τα κρυφά μοντέλα Markov ή Hidden Markov Models (HMMs). Ένα HMM είναι ουσιαστικά ένα στατιστικό μοντέλο στο οποίο το σύστημα, το οποίο μοντελοποιείται, θεωρείται μια αλυσίδα

Κεφάλαιο II – Υλικά & Μέθοδοι

Markov με μη παρατηρούμενες καταστάσεις (Εικόνα 40). Σε μια αλυσίδα Markov, κάθε κατάσταση του μοντέλου είναι απευθείας ορατή από τον παρατηρητή και γι' αυτό οι πιθανότητες μετάβασης μεταξύ καταστάσεων είναι οι μόνες παράμετροι. Αντίθετα, σε ένα HMM οι παρατηρήσεις αποδεδειγμένα από τις καταστάσεις. Το μοντέλο αποτελείται από ένα σύνολο κρυφών καταστάσεων, ένα σύνολο παρατηρούμενων συμβόλων, και δύο σύνολα πιθανοτήτων, τις πιθανότητες μετάβασης (transition) και τις πιθανότητες εκπομπής (emission). Κάθε κατάσταση έχει μια κατανομή πιθανοτήτων για τα πιθανά σύμβολα, ενώ η ακολουθία συμβόλων που παράγεται από ένα HMM δίνει πληροφορίες για την ακολουθία καταστάσεων. Έτσι, σε ένα HMM, δεν είναι προτυποποιείται η ίδια η ακολουθία των συμβόλων σαν αλυσίδα Markov, αλλά η αλληλουχία των κρυφών καταστάσεων που σχετίζονται με αυτά (Tsaousis et al., 2017a; Tsaousis et al., 2017b).

Για παράδειγμα, έστω ότι ορίζουμε μια πρωτεϊνική ακολουθία X μήκους L της μορφής $X=(x_1, x_2, \dots, x_{L-1}, x_L)$. Στο HMM, οι παρατηρήσεις x_i αποδεδειγμένα από τις καταστάσεις. Το σύνολο των καταστάσεων είναι πλέον «κρυφό» από τον παρατηρητή και μοντελοποιείται από μια πιθανότητα κατά την οποία η μία κατάσταση δίνει μετάβαση προς την επόμενη και ορίζεται ως

$$a_{kl} = P(\pi_i = l \vee \pi_{i-1} = k) \quad (82)$$

όπου k, l είναι δυο καταστάσεις που συνδέονται από μια πιθανότητα μετάβασης, η οποία συμβολίζει την πιθανότητα η κατάσταση k να δώσει μετάβαση στην κατάσταση l . Από αυτές τις καταστάσεις σχηματίζεται μία αλυσίδα Markov 1^{ης} τάξης. Η σύνδεση των συμβόλων της παρατήρησης με της κατάστασης, γίνεται μέσω των πιθανοτήτων εκπομπής

$$e_k(b) = P(x_i = b \vee \pi_i = k) \quad (83)$$

οι οποίες δηλώνουν την πιθανότητα εμφάνισης σε κάθε θέση i της ακολουθίας ενός συγκεκριμένου συμβόλου b (αμινοξικού καταλοίπου) δεδομένου ότι το σύστημα βρίσκεται στη συγκεκριμένη κατάσταση k . Έτσι, αν για παράδειγμα το HMM προτυποποιεί τις καταστάσεις k = διαμεμβρανικό τμήμα, l =μη-διαμεμβρανικό τμήμα, για κάθε θέση της αλληλουχίας που περιγράφεται από το μοντέλο οι καταστάσεις μετάβασης περιγράφουν την πιθανότητα ένα κατάλοιπο να μεταβεί από τη μία κατάσταση (πχ διαμεμβρανικό τμήμα) στην άλλη, ενώ οι

Κεφάλαιο II – Υλικά & Μέθοδοι

πιθανότητες εκπομπής περιγράφουν την πιθανότητα κάποιο σύμβολο του αλφαβήτου (π.χ. κάποιο αμινοξικό κατάλοιπο) να εμφανιστεί σε αυτήν την κατάσταση.

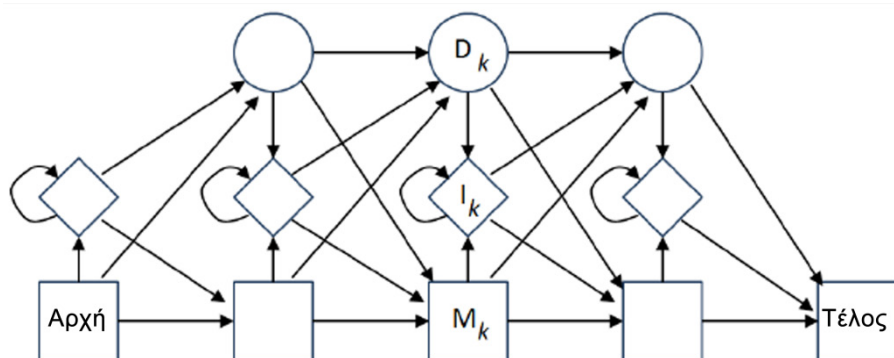
Όταν ένα HMM σχεδιαστεί, ανεξάρτητα από την πολυπλοκότητά του, οι ίδιοι βασικοί προγραμματιστικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν για να στοιχηθούν και να βαθμονομηθούν οι ακολουθίες με βάση το μοντέλο. Οι παράμετροι ενός HMM μπορούν να οριστούν με δυο τρόπους. Ένα HMM μπορεί να εκπαιδευτεί από αρχικά μη στοιχισμένες ακολουθίες ή εναλλακτικά, το μοντέλο να κατασκευαστεί από προ-στοιχισμένες ακολουθίες (όπου η ακολουθίες καταστάσεων θεωρούνται γνωστές). Στη δεύτερη περίπτωση, η εκτίμηση των παραμέτρων είναι απλά ζήτημα της μετατροπής των παρατηρηθέντων μετρήσεων των εκπομπών των συμβόλων και των μεταβάσεων μεταξύ καταστάσεων σε πιθανότητες. Οι αλγόριθμοι εκπαίδευσης είναι σημαντικοί, γιατί μπορεί να μην είναι γνωστή μια αποδεκτή στοίχιση για τις ακολουθίες ενδιαφέροντος. Όλοι οι αλγόριθμοι εκπαίδευσης βελτιστοποιούν τοπικά και αυτός είναι ο λόγος που είναι σημαντικό τα HMMs να κατασκευάζονται από προ-στοιχισμένα δεδομένα, όταν αυτό είναι δυνατόν. Σε αντίθεση με την εκτίμηση των παραμέτρων, ο σχεδιασμός μιας κατάλληλης αρχιτεκτονικής για ένα HMM είναι συνήθως ζήτημα σχεδίασης με το χέρι (Durbin et al., 1998).

2.3.2 Προφίλ Hidden Markov Model (pHMMs)

Μια ειδική κατηγορία των HMMs, προσανατολισμένη στη μελέτη βιολογικών ακολουθιών, είναι τα προφίλ HMM (profile Hidden Markov Models, pHMMs). Ένα pHMM είναι ουσιαστικά ένα HMM το οποίο περιγράφει μία πολλαπλή στοίχιση αλληλουχιών, είτε νουκλεϊκών οξέων (DNA, RNA) είτε πρωτεϊνών. Η βασική εξειδίκευση ενός pHMM, σε σχέση με ένα κλασικό HMM είναι το ότι σε ένα pHMM κάθε κατάσταση περιγράφει μια συγκεκριμένη θέση (στήλη) της πολλαπλής στοίχισης που έχει χρησιμοποιηθεί για την κατασκευή του (Eddy, 1998). Κατά συνέπεια, το μοντέλο του pHMM διαθέτει συγκεκριμένες παραμέτρους ανά θέση. Επιπρόσθετα, ακολουθώντας τη ροή της πληροφορίας στην πολλαπλή στοίχιση, η κατεύθυνση των μεταβάσεων των pHMMs είναι πάντα μονόδρομη, από τα αριστερά προς τα δεξιά (μοντέλα «left-to-right»). Τέλος, μια ιδιαιτερότητα των pHMMs είναι η ύπαρξη μιας επιπλέον

Κεφάλαιο II – Υλικά & Μέθοδοι

κατηγορίας καταστάσεων («σιωπηλές καταστάσεις»), οι οποίες δεν εκπέμπουν κάποιο σύμβολο (αμινοξύ ή νουκλεοτίδιο) αλλά κωδικοποιούν τη μετάβαση από μία κατάσταση/θέση σε κάποια άλλη αρκετά κατάλοιπα μακριά, χωρίς να υπάρχει ανάγκη για πολλαπλές μεταβάσεις από αυτήν την κατάσταση.



Εικόνα 41 Σχηματική αναπαράσταση ενός τυπικού rHMM. Οι καταστάσεις σύμπτωσης (M_k), εισαγωγής (I_k) και απαλοιφής (D_k) συμβολίζονται με τετράγωνα, ρόμβους και κύκλους, αντίστοιχα. Οι καταστάσεις συνδέονται με αντίστοιχες πιθανότητες μετάβασης οι οποίοι συμβολίζονται με βέλη, ενώ αντίστοιχα ορίζονται οι πιθανότητες γέννησης με τις οποίες κάθε κατάσταση εκπέμπει ένα σύμβολο.

Το παράδειγμα ενός rHMM που έχει προκύψει από μια πολλαπλή στοίχιση παρουσιάζεται στην Εικόνα 41. Το μοντέλο αποτελείται από μια σειρά από καταστάσεις, οι οποίες ανήκουν σε τρεις κατηγορίες, τις καταστάσεις σύμπτωσης (Match states), τις καταστάσεις εισαγωγής (Insertion states) και τις καταστάσεις απαλοιφής (Deletion states). Οι καταστάσεις σύμπτωσης και εισαγωγής είναι κανονικές καταστάσεις, οι οποίες διαθέτουν τις αντίστοιχες τιμές πιθανοτήτων μετάβασης και εκπομπής. Οι καταστάσεις σύμπτωσης αντιστοιχούν στις στήλες της πολλαπλής στοίχισης οι οποίες έχουν στοιχηθεί σωστά, ενώ οι καταστάσεις εισαγωγής αντιστοιχούν στις περιοχές των ακολουθιών όπου υπάρχει εισαγωγή χαρακτήρων (νουκλεοτίδια ή αμινοξέα) που δεν στοιχίζονται σωστά. Στις ακολουθίες όπου αυτές οι περιοχές δεν εντοπίζονται, η θέση τους αντικαθίσταται από κενά, τα οποία ελέγχονται από τις καταστάσεις απαλοιφής (Eddy, 1998; Tsaousis et al., 2017b).

Η χρήση των rHMMs, μέσω της εισαγωγής διαφορετικών καταστάσεων σύμπτωσης και εισαγωγής, οδηγεί σε σημαντική τομή στο πεδίο των πολλαπλών στοίχισεων σε σχέση με τις κλασικές μεθόδους οι οποίες, καθώς δεν προϋποθέτουν ένα μοντέλο, δεν διαχωρίζουν τις

Κεφάλαιο II – Υλικά & Μέθοδοι

πληροφοριακές θέσεις στη στοίχιση από τις απλές τυχαίες εισαγωγές. Επιπλέον, οι ποινές για την εισαγωγή κενών δεν τίθενται εκ των προτέρων αλλά εκτιμώνται από τα δεδομένα και αναπαρίστανται με έναν καθαρά πιθανοθεωρητικό τρόπο, αποκλείοντας την οποιαδήποτε υποκειμενική παρέμβαση. Έτσι, η χρήση των pHMMs επιτρέπει την πραγματοποίηση ιδιαίτερα ευαίσθητων αναζητήσεων και τον εντοπισμό απομακρυσμένων ομολόγων τα οποία οι παραδοσιακοί αλγόριθμοι στοίχισης δεν θα μπορούσαν να εντοπίσουν (Μπάγκος, 2015).

2.3.3 Κατασκευή και χρήση των pHMMs στην Υπολογιστική Βιολογία

Η χρησιμότητα των pHMMs ξεκινάει από τη δημιουργία πολλαπλών στοιχίσεων και φτάνει στη δημιουργία μοντέλων μέσω των οποίων μπορεί να γίνει ευαίσθητη αναζήτηση απομακρυσμένων ομολογιών αλλά και χαρακτηρισμός της δομής και της λειτουργίας αλληλουχιών με άγνωστο ρόλο (Durbin et al., 1998; Eddy, 1998). Ιδιαίτερα με την τελευταία μέθοδο, και τη δημιουργία όλο και περισσότερων μοντέλων για την ταξινόμηση πρωτεϊνικών οικογενειών, έχουν κατασκευαστεί ειδικές βάσεις δεδομένων όπως η Pfam (El-Gebali et al., 2019), η INTERPRO (Mitchell et al., 2019) και η TIGRFAMs (Haft et al., 2001), οι οποίες περιέχουν pHMMs που αντιπροσωπεύουν συντηρημένες περιοχές πρωτεϊνών και χρησιμοποιούνται στο χαρακτηρισμό πρωτεϊνών με άγνωστη λειτουργία. Στα πλαίσια της παρούσας διδακτορικής διατριβής έχουν χρησιμοποιηθεί μια σειρά από pHMMs, είτε κατασκευασμένα για τις ανάγκες της διατριβής με το εργαλείο HMMER (Eddy, 2011) είτε προερχόμενα από τη βάση δεδομένων Pfam. Τόσο το εργαλείο HMMER όσο και η Pfam παρουσιάζονται συνοπτικά παρακάτω.

2.3.3.1 Το πακέτο λογισμικού HMMER

Το HMMER (<http://hmmer.org/>) είναι το πιο γνωστό πακέτο λογισμικού για την κατασκευή και τη χρήση pHMMs. Είναι μια ελεύθερα διαθέσιμη συλλογή προγραμμάτων ανοιχτού κώδικα για τη διαχείριση βιολογικών αλληλουχιών (πρωτεΐνες και νουκλεϊκά οξέα), την κατασκευή pHMMs και τη χρήση τους στην αναζήτηση ομολογίας αλλά και την κατασκευή πολλαπλών στοιχίσεων (Eddy, 2011). Το HMMER είναι η πιο ευρέως χρησιμοποιούμενη μέθοδος pHMMs στην Υπολογιστική Βιολογία, με το σύστημα αρχείων του να αποτελεί το επίσημο σύστημα μεγάλων βάσεων δεδομένων που χρησιμοποιούν pHMMs, όπως η βάση Pfam (βλ. Ενότητα 2.3.3.2). Οι κυριότερες λειτουργίες του HMMER, οι οποίες χρησιμοποιούνται εκτεταμένα από την επιστημονική κοινότητα, περιλαμβάνουν τα παρακάτω:

hmmbuild: Πρόγραμμα για την κατασκευή ενός pHMM από μια πολλαπλή στοίχιση ακολουθιών. Με τη λειτουργία αυτή ο χρήστης μπορεί να κατασκευάσει δικά του μοντέλα pHMMs, τα οποία στη συνέχεια μπορεί να χρησιμοποιήσει στην ανάλυση ακολουθιών.

hmmcompress: Κατασκευή μιας βιβλιοθήκης μοντέλων από ένα ή περισσότερα pHMMs σε δυαδική (binary) μορφή. Η βιβλιοθήκη μπορεί στη συνέχεια να χρησιμοποιηθεί για την αναζήτηση ομολογίας μέσω λειτουργιών όπως η *hmmsearch*.

hmmalign: Πρόγραμμα για την πολλαπλή στοίχιση ακολουθιών με πρότυπο ένα pHMM. Καθεμία από τις αλληλουχίες στοιχίζεται διαδοχικά με το μοντέλο του pHMM και τα αποτελέσματά των στοιχίσεων αυτών συνδυάζονται για το σχηματισμό της πολλαπλής στοίχισης.

hmmsearch: Πρόγραμμα με το οποίο πραγματοποιούνται αναζητήσεις μίας ή περισσότερων ακολουθιών ενάντια σε μια βιβλιοθήκη pHMMs, η οποία έχει κατασκευαστεί μέσω του *hmmcompress*. Μέσω του *hmmsearch* μπορούν να βρεθούν τμήματα των ακολουθιών που παρουσιάζουν ομοιότητα με τα pHMMs της βιβλιοθήκης.

hmmscan: Πρόγραμμα με το οποίο πραγματοποιούνται αναζητήσεις ενός pHMM ενάντια σε μια βάση δεδομένων. Μέσω του *hmmscan* μπορεί να βρεθεί αν το μοντέλο του pHMM αντιπροσωπεύεται ή όχι στις ακολουθίες, καθώς και οι θέσεις των τμημάτων των ακολουθιών που μοιάζουν με το μοντέλο.

Κεφάλαιο II – Υλικά & Μέθοδοι

phmmer: Πρόγραμμα με το οποίο πραγματοποιείται αναζήτηση μιας ή περισσότερων ακολουθιών ενάντια σε μια βάση δεδομένων ακολουθιών με κατά ζεύγη στοιχίσεις, στα πρότυπα του BLAST.

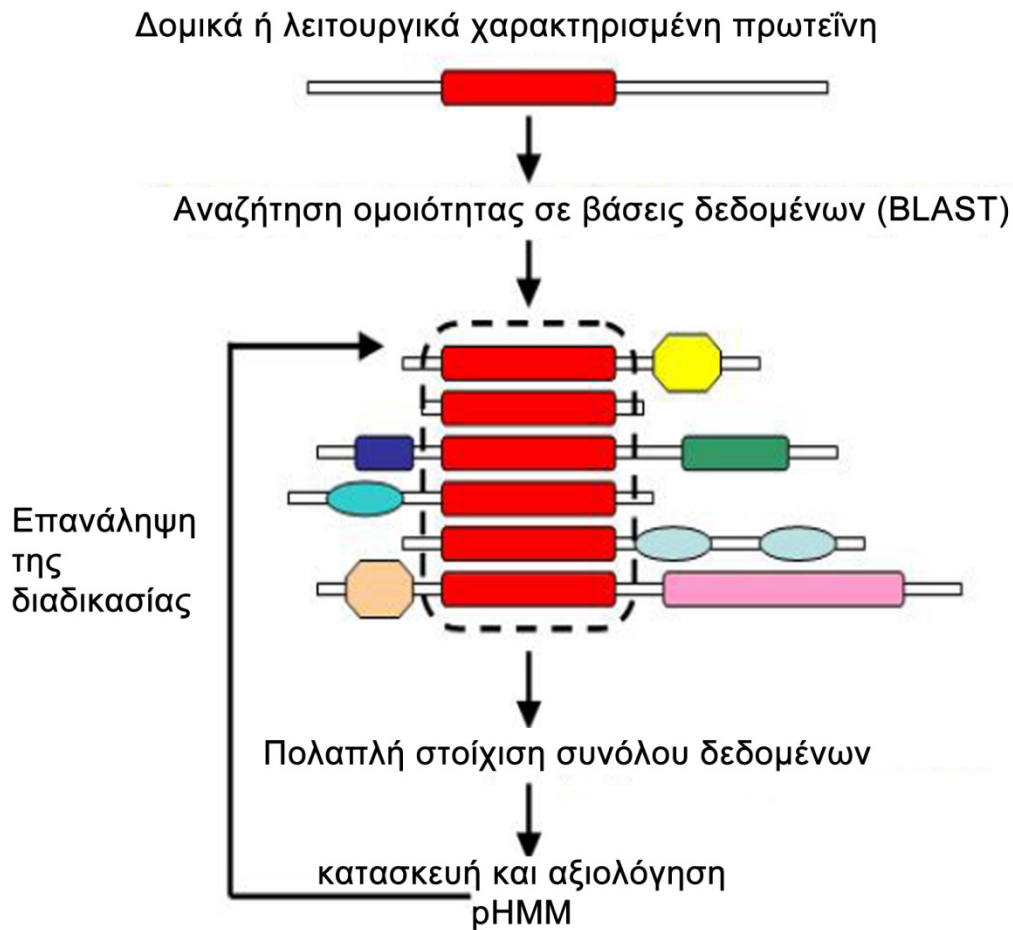
jackhmmer: Πρόγραμμα με το οποίο πραγματοποιούνται επαναλαμβανόμενες αναζητήσεις μιας ή περισσότερων ακολουθιών ενάντια σε μια βάση δεδομένων ακολουθιών με κατά ζεύγη στοιχίσεις και επέκταση μέσω ομολογίας (Eddy, 2011).

2.3.3.2 Η βάση δεδομένων Pfam

Η Pfam (<https://pfam.xfam.org/>) είναι μια δημόσια διαθέσιμη βάση δεδομένων πρωτεϊνικών οικογενειών, η οποία ταξινομεί τις πρωτεϊνικές αλληλουχίες και τις αυτοτελείς δομικές περιοχές σε οικογένειες με βάση την ομοιότητα της αμινοξικής ακολουθίας τους (El-Gebali et al., 2019). Ως μέλη μιας «οικογένειας» στην Pfam χαρακτηρίζονται είτε ολόκληρες πρωτεΐνες είτε συγκεκριμένες περιοχές της αλληλουχίας, οι οποίες διαθέτουν συντηρημένα χαρακτηριστικά. Πρωτεΐνες ή περιοχές που ανήκουν στην ίδια οικογένεια έχουν πιθανότατα παρόμοια λειτουργία και προέρχονται από ένα κοινό πρόγονο. Η ανάλυση αμινοξικών αλληλουχιών πρωτεϊνών που ανήκουν στην ίδια οικογένεια, μέσω μιας πολλαπλής στοίχισης, είναι πιθανό να οδηγήσει σε ένα «αποτύπωμα» χαρακτηριστικό για κάθε οικογένεια, ικανό να τη διαχωρίζει από τις πρωτεϊνικές αλληλουχίες που δεν ανήκουν σε αυτήν την οικογένεια. Η εύρεση και η ταξινόμηση των οικογενειών στην Pfam γίνεται χρησιμοποιώντας pHMMs. Η τρέχουσα έκδοση της βάσης (Έκδοση 32.0) διαθέτει δεδομένα για 17929 οικογένειες, με κάθε οικογένεια να αντιπροσωπεύεται από ένα ειδικό για αυτήν pHMM (El-Gebali et al., 2019). Υπολογίζεται ότι πάνω από 75% των συνολικών πρωτεϊνικών εγγραφών της UniProt αντιπροσωπεύεται από αυτές τις οικογένειες. Η Pfam χρησιμοποιεί το πρότυπο αρχείων του HMMER για τα pHMMs και τις πολλαπλές στοιχίσεις και, για κάθε εγγραφή της, διαθέτει ελεύθερα στο χρήστη τόσο το pHMM και τη στοίχιση που χρησιμοποιήθηκε στην εκπαίδευσή του, όσο και τα αποτελέσματα αναζητήσεων που έχουν πραγματοποιηθεί με αυτό το pHMM στα πρωτεώματα αναφοράς της UniProt.

2.3.3.3 Η διαδικασία εκπαίδευσης ενός pHMM

Η τυπική διαδικασία της κατασκευής και «εκπαίδευσης» ενός pHMM, το οποίο χαρακτηρίζει μια πρωτεϊνική οικογένεια, παρουσιάζεται στην πιο γενική μορφή της στην Εικόνα 42. Ξεκινώντας από μια καλά σχολιασμένη αλληλουχία αυτής της οικογένειας (ολόκληρη πρωτεΐνη ή αυτοτελής δομική περιοχή) πραγματοποιείται αρχικά αναζήτηση ομολογίας σε βάσεις δεδομένων γνωστών αλληλουχιών όπως η UniProt/Swiss-Prot, συνήθως με τη χρήση κλασικών εργαλείων όπως το BLAST (Altschul et al., 1990) ή, αν τα αποτελέσματα δεν είναι ικανοποιητικά, με τη χρήση πιο εντατικών μεθόδων όπως το PSI-BLAST (Altschul et al., 1997) ή η μέθοδος *phmmer* του πακέτου λογισμικού HMMER (Eddy, 2011). Από την αναζήτηση αυτή συγκεντρώνεται ένα καλά σχολιασμένο, αντιπροσωπευτικό σύνολο αλληλουχιών της οικογένειας (σύνολο εκπαίδευσης), το οποίο χρησιμοποιείται για την κατασκευή μιας αρχικής πολλαπλής στοίχισης, είτε χειροκίνητα είτε μέσω κάποιου προγράμματος πολλαπλών στοίχισεων. Τα αποτελέσματα της στοίχισης, ανάλογα με την περίπτωση, διορθώνονται ή εμπλουτίζονται από επιπλέον δεδομένα (π.χ. πρόγνωση δευτεροταγούς δομής, πρόγνωση διαμεμβρανικών τμημάτων, πειραματικά δεδομένα κλπ). Η τελική στοίχιση χρησιμοποιείται για την κατασκευή ενός μοντέλου pHMM (π.χ. μέσω της λειτουργίας *hmmbuild* του HMMER) που αξιολογείται ως προς την ποιότητά του μέσα από αναζητήσεις σε βάσεις δεδομένων για τη σωστή εύρεση αλληλουχιών, οι οποίες ανήκουν στην πρωτεϊνική οικογένεια του pHMM αλλά δεν χρησιμοποιήθηκαν στο σύνολο εκπαίδευσης. Η παραπάνω διαδικασία επαναλαμβάνεται, ώσπου να επιτευχθεί η καλύτερη δυνατή απόδοση του pHMM, χωρίς ωστόσο να παρατηρείται υπερπροσαρμογή («overfitting») στα δεδομένα εκπαίδευσης, η οποία θα μπορούσε να μειώσει την ικανότητα του pHMM στον εντοπισμό απομακρυσμένων ομόλογων αλληλουχιών.



Εικόνα 42 Η τυπική διαδικασία κατασκευής ενός rHMM. Προσαρμογή από Μπάγκος, 2015.

Η δημιουργία rHMMs για την περιγραφή πρωτεϊνικών οικογενειών παρουσιάζει συχνά δυσκολίες, ειδικά σε περιπτώσεις στις οποίες πρωτεΐνες από διαφορετικές οικογένειες παρουσιάζουν υψηλή ομολογία μεταξύ τους. Χαρακτηριστικό παράδειγμα αποτελούν οι πρωτεϊνικές κινάσες, οι οποίες καταλύουν τη φωσφορυλίωση αμινοξικών καταλοίπων σε πρωτεϊνικές αλληλουχίες. Παρ'ότι διαφορετικές κατηγορίες κινασών όπως τυροσινικές κινάσες και οι κινάσες Ser/Thr ανήκουν σε διαφορετικές οικογένειες, παρουσιάζουν υψηλή ομολογία στην αλληλουχία τους, με αποτέλεσμα να είναι δύσκολη η κατασκευή ειδικών rHMMs για το διαχωρισμό τους. Για την επίλυση αυτού του προβλήματος έχουν προταθεί διάφορες μέθοδοι, οι οποίες επιχειρούν την κατασκευή πιο εξειδικευμένων rHMMs. Μια

Κεφάλαιο II – Υλικά & Μέθοδοι

τέτοια μέθοδος, η οποία έχει χρησιμοποιηθεί στην παρούσα διδακτορική διατριβή, είναι η μέθοδος HMM-ModE (Sinha & Lynn, 2014; Srivastava et al., 2007).

Η μέθοδος HMM-ModE επιχειρεί τη βελτιστοποίηση της εκπαίδευσης των pHMMs με τη χρήση τόσο θετικού όσο και αρνητικού συνόλου εκπαίδευσης. Για την εκπαίδευση του pHMM μέσω της μεθόδου απαιτείται αρχικά η συλλογή δύο συνόλων εκπαίδευσης, ενός θετικού συνόλου, αποτελούμενου από μέλη της οικογένειας για την οποία θα κατασκευαστεί το μοντέλο, καθώς και ενός αρνητικού συνόλου, αποτελούμενου από πρωτεΐνες που δεν ανήκουν στην οικογένεια. Οι αλληλουχίες των συνόλων ομαδοποιούνται με βάση την ομοιότητά τους, χρησιμοποιώντας το BLAST (Altschul et al., 1990) για τη δημιουργία στοιχίσεων κατά ζεύγη σε συνδυασμό με μια παραλλαγή του αλγορίθμου ομαδοποίησης Markov Cluster (MCL), σχεδιασμένη για την ανάλυση και ομαδοποίηση αλληλουχιών (van Dongen, 2000). Για τις πρωτεΐνες του θετικού συνόλου εκπαίδευσης κατασκευάζεται ένα πρώιμο pHMM, το οποίο ελέγχεται ενάντια στις αλληλουχίες του αρνητικού συνόλου εκπαίδευσης. Τα ψευδώς θετικά αποτελέσματα (False Positives, FP) από την εφαρμογή του pHMM στο αρνητικό σύνολο χρησιμοποιούνται για την κατασκευή ενός pHMM των FP. Τα δύο προφίλ υποβάλλονται σε στοίχιση προφίλ-προφίλ, μέσω της οποίας αναγνωρίζονται οι θέσεις του μοντέλου pHMM της οικογένειας για τις οποίες απαιτείται τροποποίηση των πιθανοτήτων εκπομπής, μέσω του υπολογισμού της σχετικής εντροπίας

$$RE_i = \sum p_{i,x} \frac{p_{i,x}}{q_{i,x}} \quad (84)$$

όπου $p_{i,x}$ και $q_{i,x}$ είναι οι πιθανότητες εμφάνισης του καταλοίπου x στη θέση i του pHMM στο θετικό και το αρνητικό σύνολο εκπαίδευσης, αντίστοιχα. Υπολογίζονται επίσης οι τιμές της σχετικής εντροπίας για ένα μηδενικό (null) μοντέλο, το οποίο ουσιαστικά αντιστοιχεί στο pHMM των ψευδώς θετικών αποτελεσμάτων. Οι θέσεις του προφίλ οι οποίες αναγνωρίζονται από τη σύγκριση των παραπάνω κριτηρίων υπόκεινται σε τροποποίηση των πιθανοτήτων εκπομπής, μέσω ενός ευριστικού μοντέλου συμβατού με την αρχιτεκτονική του πακέτου λογισμικού HMMER. Η παραπάνω πορεία πραγματοποιείται μέσα από μια διαδικασία διασταυρωτής επικύρωσης επί 10 φορές (10-fold cross-validation), κατά την οποία το σύνολο των αλληλουχιών (θετικό+αρνητικό) χωρίζεται διαδοχικά σε 10 υποσύνολα, από τα οποία κάθε φορά τα εννιά χρησιμοποιούνται στην εκπαίδευση του pHMM και το ένα λειτουργεί σαν

σύνολο ελέγχου για την αξιολόγηση του rHMM. Το τελικό rHMM διαθέτει τροποποιημένες πιθανότητες εκπομπής, ικανές για την αναγνώριση των περιοχών διαχωρισμού ανάμεσα στα μέλη και τα μη-μέλη της οικογένειας, καθώς και βέλτιστες τιμές κατωφλίων για τη χρήση του προφίλ στην αναζήτηση ομολογίας (Srivastava et al., 2007).

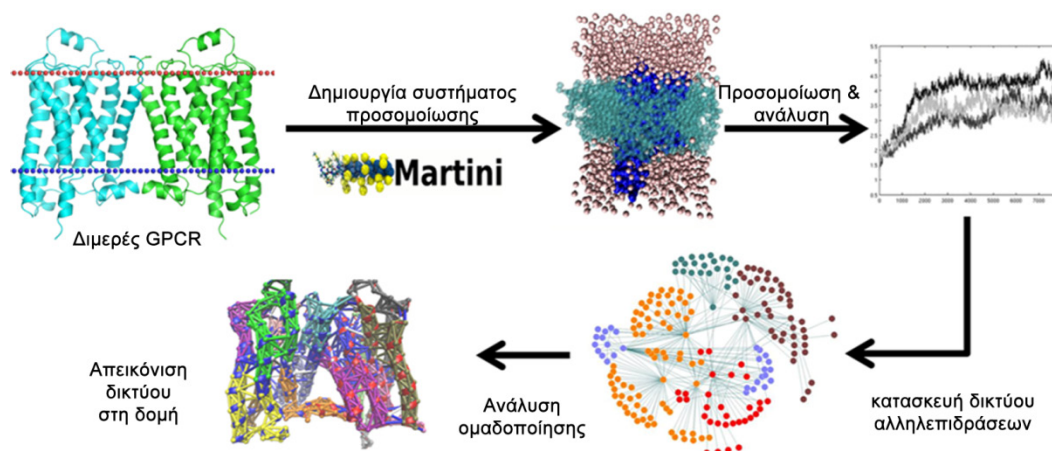
2.4 Αλληλεπιδράσεις ολιγομερισμού σε Συζευγμένους με G-πρωτεΐνες Υποδοχείς

2.4.1 Μελέτη των αλληλεπιδράσεων ολιγομερισμού των GPCRs με προσομοιώσεις Μοριακής Δυναμικής

Στόχος της παρούσας ενότητας ήταν η μελέτη των δομικών και των δυναμικών χαρακτηριστικών που διέπουν τον ολιγομερισμό των GPCRs με τη χρήση προσομοιώσεων Μοριακής Δυναμικής. Για την επίτευξή του αρχικά πραγματοποιήθηκε εκτεταμένη αναζήτηση στη βάση δεδομένων PDB (Berman et al., 2000) για τη συλλογή των διαθέσιμων κρυσταλλικών δομών των GPCRs που παρουσιάζουν τους υποδοχείς σε διάταξη διμερούς. Ως δομές του συνόλου επιλέχθηκαν οι δομές των υποδοχέων που, σύμφωνα με τις βιβλιογραφικές αναφορές τους, χαρακτηρίζονταν ως πιθανά διμερή υποδοχέων. Παράλληλα με τη συλλογή των κρυσταλλικών δομών, πραγματοποιήθηκε εκτεταμένη βιβλιογραφική αναζήτηση, με στόχο τη συλλογή πειραματικών δεδομένων από βιοφυσικά και βιοχημικά πειράματα για την ύπαρξη αλληλεπιδράσεων ολιγομερισμού ανάμεσα σε GPCRs, τα οποία χρησιμοποιήθηκαν για την κατασκευή τρισδιάστατων μοντέλων διμερών GPCRs μέσα από δοκιμασίες αγκυροβόλησης πρωτεϊνών-πρωτεϊνών. Το τελικό σύνολο δεδομένων, αποτελούμενο τόσο από κρυσταλλικές δομές όσο και από θεωρητικά μοντέλα, περιελάμβανε 21 διαφορετικά διμερή GPCRs. Καθένα από αυτά τα σύμπλοκα υποβλήθηκε σε προσομοιώσεις Μοριακής Δυναμικής σε συνθήκες μεμβρανικού περιβάλλοντος (Εικόνα 43), με χρόνους προσομοίωσης της τάξης των μικροδευτερολέπτων (μs). Λόγω του μεγάλου αριθμού των συστημάτων, αλλά και του χρόνου

Κεφάλαιο II – Υλικά & Μέθοδοι

προσομοίωσης, η προτυποποίηση και οι προσομοιώσεις των διμερών πραγματοποιήθηκαν με το πεδίο αδρής διακριτικότητας MARTINI (Marrink et al., 2007; Monticelli et al., 2008), χρησιμοποιώντας το πακέτο προσομοιώσεων GROMACS (Abraham et al., 2015).



Εικόνα 43 Η πορεία εργασίας στις προσομοιώσεις Μοριακής Δυναμικής των διμερών GPCRs. Κάθε διμερές GPCR μετατράπηκε σε μοντέλο CG με το δυναμικό πεδίο MARTINI και ενσωματώθηκε σε μια μεμβράνη λιπιδίων. Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν για την κατασκευή και ανάλυση ενός δυναμικού δικτύου αλληλεπιδράσεων, το οποίο οπτικοποιήθηκε πάνω στη δομή του διμερούς.

Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν για την ανάλυση των διαμοριακών αλληλεπιδράσεων ανάμεσα στα πρωτομερή κάθε διμερούς. Οι αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών κάθε καταλοίπου υπολογίστηκαν αρχικά μέσα από τον προσδιορισμό της επιφάνειας αλληλεπίδρασης, ή Θαμμένης Επιφάνειας (Buried Surface Area, BSA), ως τη διαφορά της προσβάσιμης στο διαλύτη επιφάνειας (Accessible Surface Area, ASA) σύμφωνα με τον τύπο:

$$BSA = ASA_A + ASA_B - ASA_{AB} \quad (85)$$

όπου BSA είναι η θαμμένη επιφάνεια, ενώ ASA_A , ASA_B και ASA_{AB} είναι οι τιμές προσβάσιμης επιφάνειας για τα πρωτομερή A, B και το διμερές AB. Η σημασία της συνεισφοράς κάθε καταλοίπου στην επιφάνεια αλληλεπίδρασης μελετήθηκε περαιτέρω, χρησιμοποιώντας τρεις διαφορετικές προσεγγίσεις:

Κεφάλαιο II – Υλικά & Μέθοδοι

1. Κατηγοριοποίηση καταλοίπων στον πυρήνα ή την περιφέρεια της θαμμένης επιφάνειας: Τα αλληλεπιδρώντα κατάλοιπα κατηγοριοποιήθηκαν χρησιμοποιώντας τη μέθοδο του Levy (Levy, 2010), η οποία χωρίζει τη θαμμένη επιφάνεια σε δύο διακριτά μέρη, τον «πυρήνα» (core) και την «περιφέρεια» (rim). Η κατηγοριοποίηση των καταλοίπων πραγματοποιήθηκε υπολογίζοντας τη σχετική προσβάσιμη επιφάνεια (Relative Surface Area, RSA) κάθε καταλοίπου i στο πρωτεϊνικό σύμπλοκο, καθώς και στα απομονωμένα πρωτομερή του, σύμφωνα με τον τύπο

$$RSA_i = \frac{ASA_i}{Max.ASA_i} \times 100 \quad (86)$$

όπου RSA_i είναι η σχετική προσβάσιμη επιφάνεια του καταλοίπου i , ASA_i είναι η απόλυτη τιμή της επιφάνειάς του καταλοίπου και $Max.ASA_i$ είναι η μέγιστη δυνατή επιφάνεια που μπορεί να έχει το συγκεκριμένο κατάλοιπο, όπως αυτή έχει υπολογιστεί για εκτεταμένα τριπεπτίδια Gly-X_i-Gly (Chothia, 1976; Miller et al., 1987). Τα αμινοξικά κατάλοιπα που έχουν τιμή $RSA > 25\%$ στα ελεύθερα πρωτομερή και $RSA < 25\%$ στο πρωτεϊνικό σύμπλοκο αποτελούν τον πυρήνα της θαμμένης επιφάνειας. Αντίθετα, κατάλοιπα με τιμές $RSA > 25\%$ στο πρωτεϊνικό σύμπλοκο κατατάσσονται στην περιφέρεια της θαμμένης επιφάνειας. Σύμφωνα με τη μέθοδο του Levy, τα κατάλοιπα που σχηματίζουν τον πυρήνα της θαμμένης επιφάνειας θεωρούνται ότι συνεισφέρουν περισσότερο στην ισχύ των αλληλεπιδράσεων (Levy, 2010).

2. Υπολογιστική Σάρωση Αλανίνης: Κάθε κατάλοιπο αξιολογήθηκε μέσα από προσομοιώσεις Υπολογιστικής Σάρωσης Αλανίνης (βλ. Ενότητα 2.1.12.3). Ως σημεία ενδιαφέροντος (hot spots) επιλέχθηκαν τα κατάλοιπα με τιμές $\Delta\Delta G > 1.5$ kcal/mol (Guerois et al., 2002).

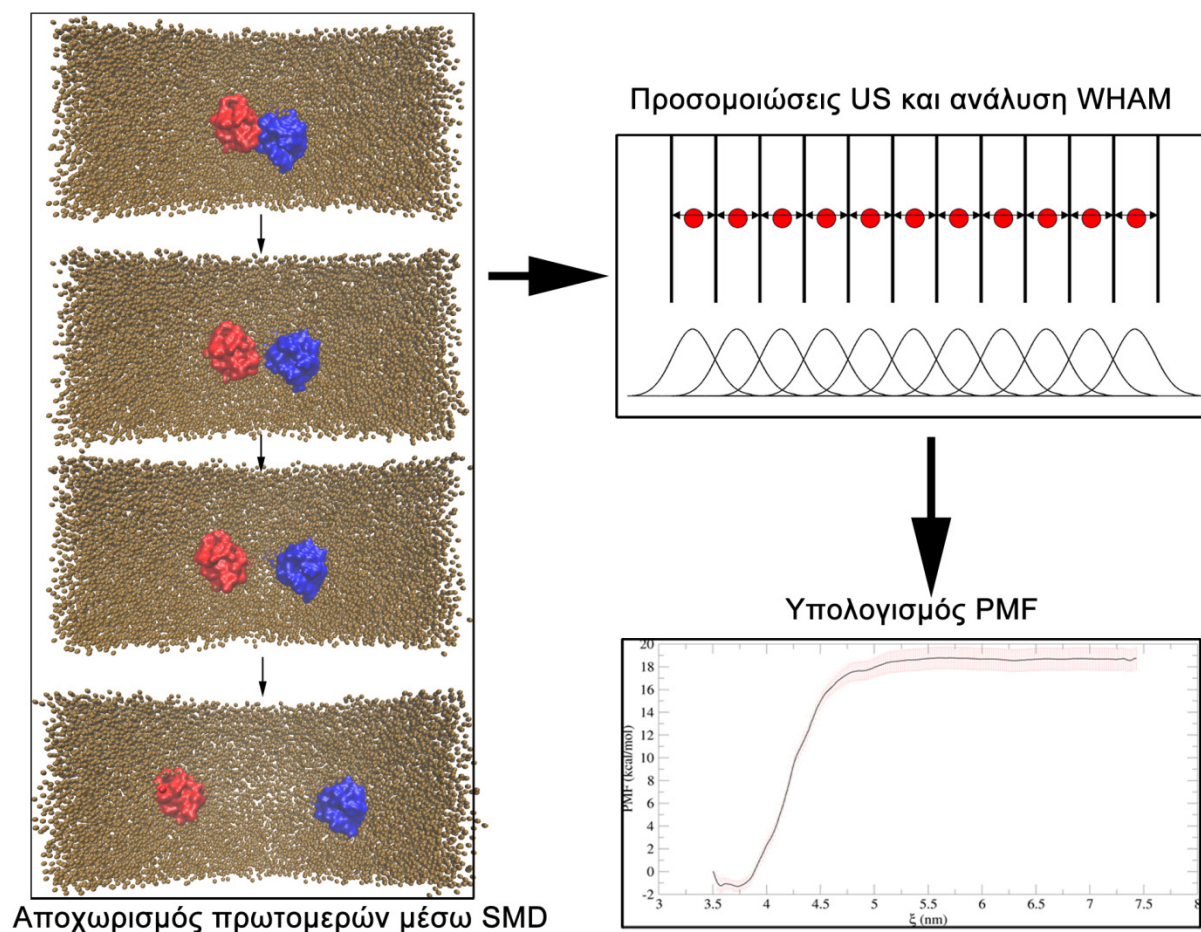
3. Αλληλεπιδράσεις π-ηλεκτρονίων: τα αποτελέσματα των προηγούμενων δύο μεθόδων έδειξαν ότι μεγάλος αριθμός των καταλοίπων που εντοπίζονται στον πυρήνα των αλληλεπιδράσεων ή/και χαρακτηρίζονται ως σημεία ενδιαφέροντος περιλαμβάνουν κατάλοιπα με επίπεδους δακτυλίους (Phe, Tyr, Trp, His), τα οποία μπορούν να σχηματίζουν αλληλεπιδράσεις π-ηλεκτρονίων (π-stacking). Έτσι, πραγματοποιήθηκε επιπρόσθετη αναζήτηση αλληλεπιδράσεων π-ηλεκτρονίων ανάμεσα σε αρωματικούς δακτυλίους (π-π

stacking), σε αρωματικών δακτυλίου και δακτυλίου προλίνης (π-Pro stacking) και σε αρωματικούς δακτυλίου – θετικά φορτισμένες ομάδες (π-cation stacking).

Τέλος, κάθε προσομοίωση χρησιμοποιήθηκε για την κατασκευή ενός αντίστοιχου δυναμικού δικτύου αλληλεπιδράσεων (Sethi et al., 2009), το οποίο χρησιμοποιήθηκε για την αξιολόγηση της κινητικότητας των GPCRs κατά την προσομοίωση των διμερών τους και για την ανίχνευση πιθανών συσχετίσεων ανάμεσα στις αλληλεπιδράσεις των διμερών και σε περιοχές των GPCRs με λειτουργικό ρόλο (Εικόνα 259). Τα δίκτυα προετοιμάστηκαν και οπτικοποιήθηκαν με το πρόγραμμα NetworkView (Eargle & Luthey-Schulten, 2012), ενώ πραγματοποιήθηκε ανάλυση ομαδοποίησης χρησιμοποιώντας τον αλγόριθμο Girvan-Newman για την ανίχνευση οργανωμένων κοινοτήτων (Girvan & Newman, 2002).

2.4.2 Θερμοδυναμική ανάλυση των αλληλεπιδράσεων ολιγομερισμού

Παράλληλα με τις προσομοιώσεις ισορροπίας που περιγράφηκαν, πραγματοποιήθηκε θερμοδυναμική ανάλυση στις αλληλεπιδράσεις ολιγομερισμού των GPCRs, μέσα από ενεργειακή μέτρηση της υδροφοβικής ασυμφωνίας και υπολογισμούς του Δυναμικού Μέσης Δύναμης (PMF) πάνω σε διμερή υποδοχέων. Συγκεκριμένα, πραγματοποιήθηκαν επιπρόσθετες προσομοιώσεις Μοριακής Δυναμικής για επιλεγμένα διμερή, τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για την ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας, χρησιμοποιώντας τη μέθοδο της Υπολειπόμενης Υδροφοβικής Ασυμφωνίας (RHM) των *Mondal et al.* Σχεδιάστηκαν και πραγματοποιήθηκαν προσομοιώσεις ισορροπίας για τα διμερή TM1-TM2-H8, TM4-TM5 και TM5-TM6 τριών υποδοχέων (Ροδοψίνη, β1AR και μOR) καθώς και για τα απομονωμένα πρωτομερή τους. Οι τρεις υποδοχείς επιλέχθηκαν τόσο λόγω της πειραματικά επιβεβαιωμένης δυνατότητάς τους να σχηματίζουν διμερή, όσο και επειδή έχουν διαθέσιμες δομές διμερών στην PDB. Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν για τον υπολογισμό της ενεργειακής ποινής του RHM (ΔG_{RHM}) καθενός από τα 7 διαμεμβρανικά τμήματα του κάθε υποδοχέα, στην κατάσταση μονομερούς, καθώς και όταν συμμετέχει σε κάποιο διμερές.



Εικόνα 44 Υπολογισμοί του Δυναμικού Μέσης Δύναμης σε διμερή GPCRs. Πραγματοποιείται αρχικά ο αποχωρισμός των δύο υποδοχέων στο διμερές μέσα από μια προσομοίωση SMD. Η πορεία SMD χρησιμοποιείται στον καθορισμό παραθύρων για τις προσομοιώσεις Δειγματοληψίας Ομπρέλας (US). Τα αποτελέσματα των προσομοιώσεων US αναλύονται με τη μέθοδο WHAM για τον υπολογισμό της καμπύλης του Δυναμικού Μέσης Δύναμης (PMF).

Εκτός από την υδροφοβική ασυμφωνία μελετήθηκε η ισχύς των αλληλεπιδράσεων διμερισμού των GPCRs σε διαφορετικές κατηγορίες διμερών, μέσω υπολογισμού του PMF κατά την διάσπαση του διμερούς (Εικόνα 44). Οι υπολογισμοί πραγματοποιήθηκαν χρησιμοποιώντας τη μέθοδο της Δειγματοληψίας Ομπρέλας (US) (Roux, 1995). Ως συντεταγμένη αντίδρασης ξ επιλέχθηκε η απόσταση ανάμεσα στα γεωμετρικά κέντρα των δύο πρωτομερών του διμερούς, καθώς αυτά απομακρύνονται το ένα από το άλλο (r). Για την πραγματοποίηση των προσομοιώσεων, κάθε διμερές τοποθετήθηκε σε μια λιπιδική διπλοστιβάδα POPC, επιμηκυμένη κατά μήκος του άξονα X . Ο αποχωρισμός των πρωτομερών

Κεφάλαιο II – Υλικά & Μέθοδοι

στο διμερές πραγματοποιήθηκε μέσω προσομοίωσης Καθοδηγούμενης Μοριακής Δυναμικής (SMD) σταθερής ταχύτητας, εφαρμόζοντας μια δύναμη έλξης στο ένα από τα δύο πρωτομερή του διμερούς και ελέγχοντας τη διαδικασία του αποχωρισμού με ένα αρμονικό δυναμικό, σύμφωνα με την Εξίσωση 50. Τα αποτελέσματα της προσομοίωσης SMD χρησιμοποιήθηκαν για τον καθορισμό των αρχικών στιγμιότυπων (παράθυρα US). Καθένα από τα παράθυρα υποβλήθηκε σε προσομοίωση US για 800 ns. Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν για τον υπολογισμό του PMF, εφαρμόζοντας τη μέθοδο WHAM (Kumar et al., 1992), και οπτικοποιήθηκαν με τη μορφή καμπυλών ελεύθερης ενέργειας (Hub et al., 2010).

2.4.3 Μελέτη του αυθόρμητου διμερισμού των υποδοχέων A2A και mGluR5

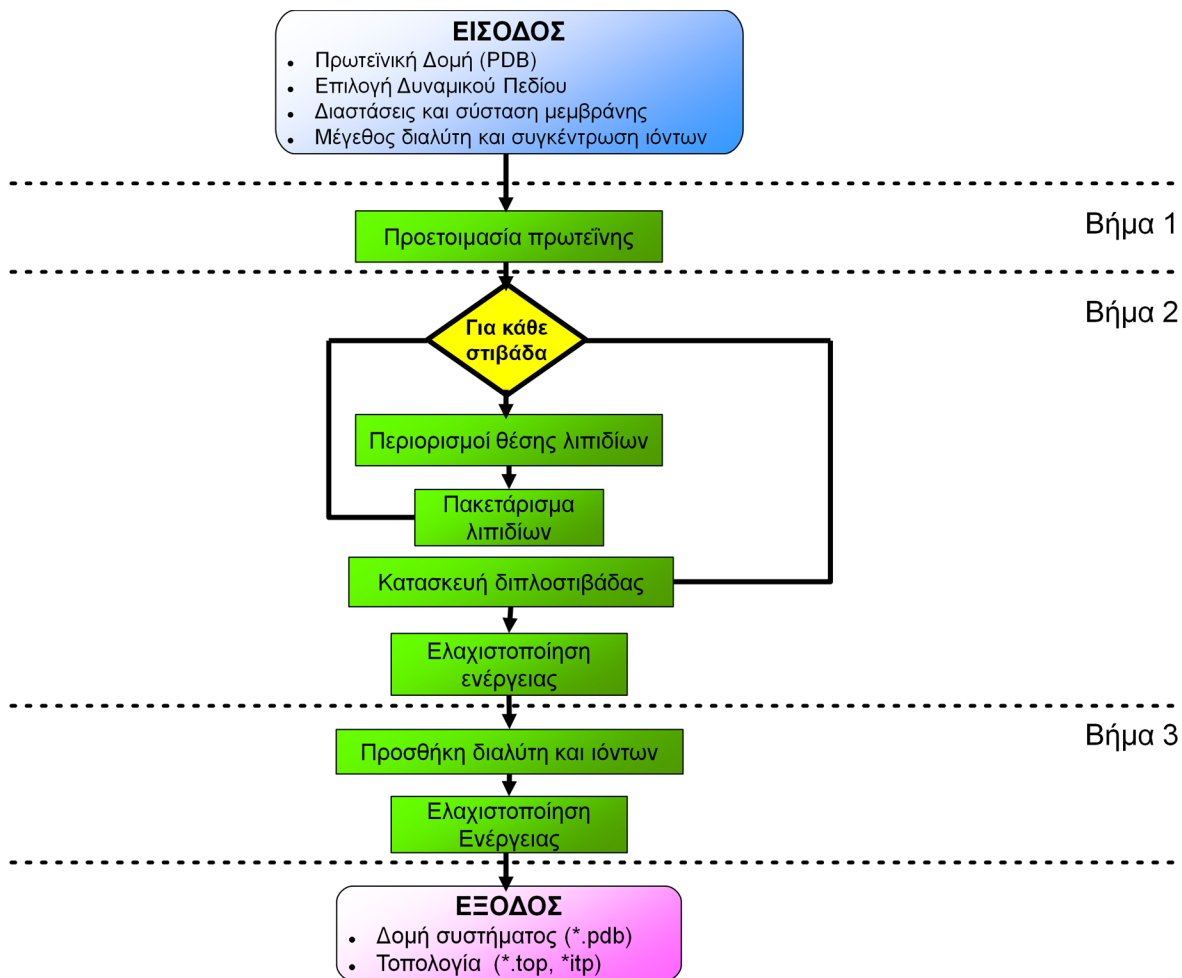
Στην παρούσα υποενότητα διερευνήθηκε η αυθόρμητη ικανότητα σχηματισμού διμερών GPCRs από απομακρυσμένα μεταξύ τους πρωτομερή, αλλά και η επίδραση στοιχείων του μεμβρανικού περιβάλλοντος, μέσα από προσομοιώσεις Μοριακής Δυναμικής μεγάλης χρονικής κλίμακας. Συγκεκριμένα, μελετήθηκε η ικανότητα διμερισμού δύο υποδοχέων με φαρμακολογικό ενδιαφέρον, του υποδοχέα της αδενοσίνης A2A και του μεταβοτροπικού υποδοχέα του γλουταμικού 5 (mGluR5), τα ομο- και ετεροδιμερή των οποίων έχουν προταθεί σαν στόχοι φαρμάκων για την αντιμετώπιση της ασθένειας του Parkinson (Cabello et al., 2009). Παράλληλα, διερευνήθηκε η επίδραση της χοληστερόλης, ενός σημαντικού συστατικού των βιολογικών μεμβρανών, πάνω στα χαρακτηριστικά των υποδοχέων. Αντίγραφα του κάθε υποδοχέα τοποθετήθηκαν σε μια μεμβράνη, με τυχαίο προσανατολισμό και απόσταση $\sim 3\text{\AA}$ μεταξύ τους, και αφέθηκαν να κινηθούν ελεύθερα κατά τη διάρκεια των προσομοιώσεων, με απώτερο στόχο το σχηματισμό σταθερών διμερών. Οι προσομοιώσεις πραγματοποιήθηκαν τόσο σε απλές φωσφολιπιδικές μεμβράνες POPC όσο και σε μεμβράνες με POPC και χοληστερόλη, σε περιεκτικότητα 10% και 30%.

2.5 Εξωτερικές Μembrάνες των αρνητικών κατά Gram βακτηρίων

2.5.1 Κατασκευή μεθόδου για την προτυποποίηση Εξωτερικών Μembrανών: η μέθοδος προτυποποίησης GNOMM

Όπως έχει αναφερθεί ήδη, οι λιποπολυσακχαρίτες (LPSs) αποτελούν σημαντικό χαρακτηριστικό της εξωτερικής μεμβράνης των αρνητικών κατά Gram βακτηρίων, διαδραματίζοντας σημαντικό ρόλο στην ορθή λειτουργία του βακτηριακού κυττάρου και στην προστασία του από αντιβιοτικά και άλλες παθογόνες προς αυτό ουσίες. Επιπρόσθετα, οι αλληλεπιδράσεις των LPSs με τα διαμεμβρανικά β-βαρέλια έχουν βρεθεί να επηρεάζουν σημαντικά στοιχεία της βιογένεσης, της δομής και της λειτουργίας τους. Τέλος, πολλοί LPSs έχουν βρεθεί να προκαλούν ανοσολογικές αποκρίσεις στον άνθρωπο και έχουν συσχετιστεί με διάφορες ασθένειες.

Παρά τη σημασία τους, η πειραματική μελέτη της εξωτερικής μεμβράνης και των στοιχείων της είναι δύσκολη. Έτσι, συχνά χρησιμοποιούνται προσομοιώσεις Μοριακής Δυναμικής για την προτυποποίηση των διαμεμβρανικών β-βαρελίων στο φυσικό λιπιδικό περιβάλλον τους (Ma et al., 2018; Niramitranon et al., 2016; Patel et al., 2016). Ωστόσο, η διαδικασία κατασκευής μιας εξωτερικής μεμβράνης, η οποία να περιέχει LPSs, δεν είναι εύκολη, κυρίως λόγω της αυξημένης πολυπλοκότητας των τελευταίων σε σχέση με τα απλά φωσφολιπίδια. Για την επίλυση αυτού του προβλήματος αναπτύχθηκε η μέθοδος GNOMM (Gram-Negative Outer Membrane Modeler), ένα εργαλείο για την αυτοματοποιημένη κατασκευή μοντέλων εξωτερικών μεμβρανών και τη χρήση τους σε προσομοιώσεις Μοριακής Δυναμικής. Το GNOMM επιτρέπει την κατασκευή μικτών, μη-συμμετρικών λιπιδικών διπλοστιβάδων αποτελούμενων από LPSs και φωσφολιπίδια, καθώς και την ενσωμάτωση πρωτεϊνών σε αυτές. Η μέθοδος υποστηρίζει τη χρήση τεσσάρων διαφορετικών δυναμικών πεδίων και παράγει συστήματα προσομοιώσεων συμβατά με την ελεύθερα διαθέσιμη μηχανή προσομοιώσεων GROMACS.



Εικόνα 45 Η πορεία εργασίας του GNOMM. Παρουσιάζεται η διαδικασία για την κατασκευή μιας Εξωτερικής Μεμβράνης γύρω από μια διαμεμβρανική πρωτεΐνη. Στην περίπτωση που ο χρήστης επιθυμεί την κατασκευή μιας διπλοστιβάδας χωρίς πρωτεΐνες, το πρώτο βήμα παραλείπεται.

Η πορεία εργασίας του GNOMM παρουσιάζεται στο διάγραμμα ροής της Εικόνας 45. Αρχικά, ο χρήστης καλείται να επιλέξει τον τύπο του συστήματος που επιθυμεί να κατασκευάσει και το δυναμικό πεδίο που θα χρησιμοποιήσει. Το GNOMM υποστηρίζει τρεις επιλογές για την κατασκευή συστημάτων:

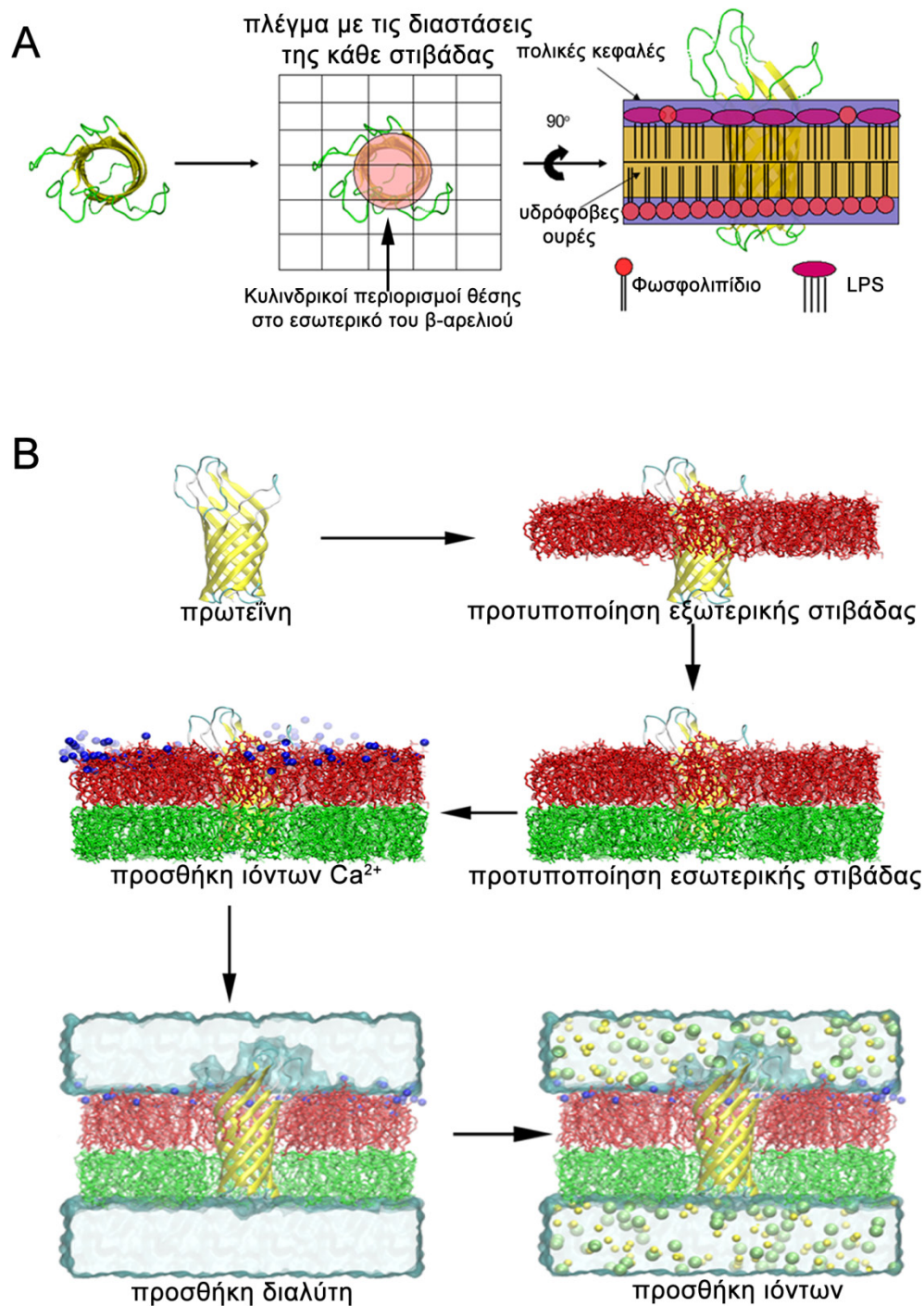
1. Σύστημα πρωτεΐνης/μεμβράνης: Προετοιμασία μιας πρωτεΐνης με βάση ένα δυναμικό πεδίο και προτυποποίηση μιας εξωτερικής μεμβράνης γύρω από αυτήν. Ο χρήστης καλείται να δώσει σαν είσοδο τις συντεταγμένες μιας πρωτεΐνης σε μορφή PDB. Το GNOMM θα προετοιμάσει τη δομή της πρωτεΐνης σύμφωνα με τους κανόνες του επιλεγμένου δυναμικού πεδίου και θα κατασκευάσει μια εξωτερική μεμβράνη γύρω της.

Κεφάλαιο II – Υλικά & Μέθοδοι

2. Σύστημα μεμβράνης: προετοιμασία μιας εξωτερικής μεμβράνης χωρίς πρωτεΐνες. Το GNOMM απλά θα κατασκευάσει μια λιπιδική διπλοστιβάδα.
3. Σύστημα προ-επεξεργασμένης δομής και τοπολογίας: Προετοιμασία μιας εξωτερικής μεμβράνης γύρω από μια προ-επεξεργασμένη από το χρήστη δομή. Η συγκεκριμένη επιλογή είναι χρήσιμη σε περιπτώσεις που η δομή ενδιαφέροντος περιέχει μη τυπικές συντεταγμένες (π.χ. χημικές τροποποιήσεις στη δομή, σύμπλοκο της πρωτεΐνης με κάποιο φάρμακο κλπ), οι οποίες δεν περιγράφονται από τις παραμέτρους των δυναμικών πεδίων του GNOMM. Ο χρήστης καλείται να δώσει σαν είσοδο την κατάλληλα προετοιμασμένη δομή σε μορφή PDB, καθώς και όσα επιπλέον αρχεία δυναμικών πεδίων (τοπολογία του συστήματος, παραμέτρους φαρμάκων κλπ) απαιτούνται. Το GNOMM επεκτείνει την δομή και τοπολογία που θα δώσει ο χρήστης, κατασκευάζοντας την εξωτερική μεμβράνη γύρω του.

Επιπρόσθετα, το GNOMM υποστηρίζει τέσσερα δυναμικά πεδία τα οποία αντιστοιχούν σε τέσσερα διαφορετικά επίπεδα διακριτικότητας, τα CHARMM36 (πλήρως ατομικό πεδίο), GROMOS 54A7 (πεδίο ενοποιημένων ατόμων), MARTINI (πεδίο αδρής διακριτικότητας) και PACE (υβριδικό πεδίο).

Στην περίπτωση που ο χρήστης επιλέξει την κατασκευή ενός συστήματος πρωτεΐνης/μεμβράνης, είτε μέσω της πρώτης είτε μέσω της τρίτης επιλογής, είναι απαραίτητο η δομή αυτής της πρωτεΐνης να διαθέτει τον κατάλληλο προσανατολισμό σε σχέση με το επίπεδο της μεμβράνης (βλ. Ενότητα 2.1.11.3). Προσανατολισμένες δομές μπορούν να βρεθούν στις βάσεις δεδομένων OPM και PDB-TM, ενώ αν ο χρήστης επιθυμεί να εισάγει κάποιο θεωρητικό μοντέλο, μπορεί να του προσδώσει τον κατάλληλο προσανατολισμό μέσω του εργαλείου PPM Server που προσφέρεται από τη βάση OPM. Έχοντας εισάγει τη δομή και επιλέξει το δυναμικό πεδίο που τον ενδιαφέρει, ο χρήστης καλείται στη συνέχεια να επιλέξει χαρακτηριστικά όπως η κατάσταση των φορτίων στο N- και το C-άκρο της πρωτεΐνης του, καθώς και το αν θα πραγματοποιηθεί διόρθωση τυχόν σφαλμάτων της δομής, μέσω του προγράμματος MODELLER.



Εικόνα 46 Προτυποποίηση της μεμβράνης γύρω από ένα β-βαρέλι με το GNOMM. (A) Σχηματική αναπαράσταση των περιορισμών θέσης που ορίζονται κατά το πακετάρισμα των λιπιδίων. Η κάθε στιβάδα της μεμβράνης ορίζεται γύρω από το β-βαρέλι σαν πλέγμα. Κυλινδρικοί περιορισμοί θέσης εφαρμόζονται στο χώρο που καταλαμβάνεται από το β-βαρέλι, για την αποφυγή τοποθέτησης λιπιδίων στο εσωτερικό του πόρου του βαρελίου. Επιπρόσθετοι περιορισμοί θέσεις εφαρμόζονται κατά την τοποθέτηση των λιπιδίων, ωθώντας τις πολικές κεφαλές στο εξωτερικό της μεμβράνης και τις υδρόφοβες ουρές στο εσωτερικό της. (B) Δομική αναπαράσταση της κατασκευής μιας Εξωτερικής Μεμβράνης γύρω από το β-βαρέλι OmpA.

Κεφάλαιο II – Υλικά & Μέθοδοι

Το επόμενο βήμα είναι ο καθορισμός των διαστάσεων και των περιεχομένων της μεμβράνης. Σημειώνεται ότι αν ο χρήστης είχε επιλέξει την κατασκευή Συστήματος Μεμβράνης, το συγκεκριμένο βήμα είναι το πρώτο βήμα στη διαδικασία του συστήματός του, ενώ το προαναφερόμενο βήμα προετοιμασίας της πρωτεΐνης παραλείπεται. Σε κάθε περίπτωση, σε αυτό το βήμα ο χρήστης καλείται να επιλέξει τις διαστάσεις X και Y της μεμβράνης, καθώς και τη λιπιδική σύσταση της κάθε στιβάδας. Το GNOMM μπορεί να κατασκευάσει τόσο ομοιογενείς στιβάδες, αποτελούμενες από ένα μόνο είδος λιπιδίου όσο και ετερογενείς, αποτελούμενες από περισσότερα του ενός λιπίδια σε διάφορες αναλογίες. Στη βιβλιοθήκη λιπιδίων του GNOMM περιλαμβάνονται παραλλαγές του Λιπιδίου A από τέσσερα διαφορετικά βακτήρια (*E. coli*, *P. aeruginosa*, *H. pylori* και *N. meningitidis*), διάφοροι λιποπολυσακχαρίτες, καθώς και μια σειρά από φωσφολιπίδια με γνωστή παρουσία στις εξωτερικές μεμβράνες των αρνητικών κατά Gram βακτηρίων (Πίνακας 6).

Η προτυποποίηση της μεμβράνης πραγματοποιείται θέτοντας αρχικά το σημείο της αρχής των αξόνων ($X, Y, Z=0,0,0$) ως το γεωμετρικό κέντρο της διπλοστιβάδας και ορίζοντας δύο ορθογώνια παραλληλεπίπεδα πλέγματα, τα οποία αντιπροσωπεύουν τις δύο στιβάδες της μεμβράνης. Οι διαστάσεις X και Y (μήκος και πλάτος) κάθε πλέγματος είναι οι τιμές που έχουν οριστεί από το χρήστη, ενώ η διάσταση Z (πάχος) καθορίζεται από το μήκος των λιπιδικών ουρών των λιπιδίων που έχει επιλέξει ο χρήστης για την κάθε στιβάδα. Στην περίπτωση που ο χρήστης έχει επιλέξει την κατασκευή Συστήματος Πρωτεΐνη/Μεμβράνη, υπολογίζεται ο χώρος που καταλαμβάνεται από την πρωτεΐνη σε κάθε στιβάδα και ορίζονται περιορισμοί θέσης (position restraints), οι οποίοι θα αποτρέψουν την εισαγωγή λιπιδίων στη συγκεκριμένη περιοχή (Εικόνα 46). Οι παραπάνω περιορισμοί εφαρμόζονται τόσο για το σωστό υπολογισμό του αριθμού των λιπιδίων στη στιβάδα, όσο και για την αποφυγή της λανθασμένης τοποθέτησης λιπιδίων στο εσωτερικό της πρωτεΐνης. Αν, αντίθετα, ο χρήστης δεν έχει συμπεριλάβει κάποια πρωτεΐνη στο σύστημά του, τότε το σύνολο του χώρου χρησιμοποιείται για την τοποθέτηση των λιπιδίων. Τέλος, αν ο χρήστης έχει εισάγει μια πρωτεΐνη η οποία καταλαμβάνει χώρο μόνο στη μία πλευρά της μεμβράνης (π.χ. μια περιφερειακή πρωτεΐνη), υπολογίζονται στερεοδιαταξοί περιορισμοί μόνο για την στιβάδα στην οποία ενσωματώνεται η πρωτεΐνη.

Κεφάλαιο II – Υλικά & Μέθοδοι

Πίνακας 6 Τα διαθέσιμα λιπίδια του GNOMM

Λιποπολυσακχαρίτες (LPSs)	
<i>Escherichia coli</i>	Λιπίδιο A (ECLI), Re LPS (ReMP), Ra LPS (RaMP)
<i>Pseudomonas aeruginosa</i>	5-ακυλικό Λιπίδιο A (PAL5), 6-ακυλικό Λιπίδιο A (PAL6), PAO1 LPS (PAO1)
<i>Helicobacter pylori</i>	Λιπίδιο A (HPLI)
<i>Neisseria meningitidis</i>	Λιπίδιο A (NMEN)
Φωσφολιπίδια	
Φωσφατιδυλ-αιθανολαμίνες (PEs)	POPE, DOPE, DPPE, DSPE, DMPE, DLPE
Φωσφατιδυλ-χολίνες (PCs)	POPC, DOPC, DPPC, DSPC, DMPC, DLPC
Φωσφατιδυλ-γλυκερόλες (PGs)	POPG, DOPG, DPPG, DSPG, DMPG, DLPG
Φωσφατιδυλ-σερίνες (PSs)	POPS, DOPS, DPPS, DSPS, DMPS, DLPS
Καρδιολιπίνες (CDLs)	Μονοανιονική Καρδιολιπίνη (CDL1), Διανιονική Καρδιολιπίνη (CDL2)

Το υπόλοιπο του χώρου κάθε στιβάδας, το οποίο δεν καταλαμβάνεται από πρωτεΐνες, χρησιμοποιείται για την τοποθέτηση των λιπιδικών μορίων. Ο αριθμός των μορίων υπολογίζεται με βάση τη σύνθεση της λιπιδικής διπλοστιβάδας και τις διαστάσεις X και Y του πλέγματος. Έτσι, αν ο χώρος στον οποίο θα τοποθετηθούν τα λιπίδια έχει εμβαδό $100 \times 100 = 10000 \text{ \AA}^2$, και η στιβάδα αποτελείται από Ra LPS (επιφάνεια ανά λιπίδιο = $\sim 170 \text{ \AA}^2$) και POPE (επιφάνεια ανά λιπίδιο = $\sim 60 \text{ \AA}^2$), σε αναλογία μορίων 70 % : 30 %, η σύστασή της θα είναι 41 μόρια Ra LPS και 19 μόρια POPE. Η τοποθέτηση των λιπιδίων γίνεται, γενικά, τυχαία. Ωστόσο, προκειμένου να εξασφαλιστεί η σωστή τοποθέτηση των λιπιδίων στην κάθε στιβάδα, εφαρμόζονται μια σειρά από επιπρόσθετους περιορισμούς θέσεων, οι οποίοι κατευθύνουν τις πολικές κεφαλές στην επιφάνεια της μεμβράνης και τις άκρες των λιπιδικών ουρών στο εσωτερικό του υδρόφοβου πυρήνα (Εικόνα 46). Μετά την κατασκευή της κάθε στιβάδας πραγματοποιείται η οργάνωση του πλήρους συστήματος, το οποίο υποβάλλεται σε εκτεταμένη ελαχιστοποίηση ενέργειας για τη διόρθωση μη επιτρεπτών διαμοριακών επαφών και τη βελτιστοποίηση της γεωμετρίας του. Στην περίπτωση που στη μεμβράνη υπάρχουν μόρια LPS, τοποθετούνται δισθενή ιόντα Ca^{2+} γύρω από τις φωσφορικές ομάδες του Λιπιδίου A. Η παρουσία αυτών των ιόντων, όπως αναφέρθηκε και στην εισαγωγή, είναι απαραίτητη για

Κεφάλαιο II – Υλικά & Μέθοδοι

την εξουδετέρωση του υψηλού αρνητικού φορτίου του Λιπιδίου A και, κατά συνέπεια, τη σωστή οργάνωση των LPSs στη μεμβράνη.

Το τελευταίο βήμα στην κατασκευή του συστήματος είναι η προτυποποίηση του διαλύτη σε κάθε πλευρά της μεμβράνης. Ο χρήστης καλείται να επιλέξει το μέγεθος του στρώματος του διαλύτη, το είδος των ιόντων (NaCl, KCl, CaCl₂) και τη συγκέντρωσή τους (Εικόνα 46). Μετά την προτυποποίηση του διαλύτη, το σύστημα υποβάλλεται σε μια τελική ελαχιστοποίηση ενέργειας. Το τελικό αποτέλεσμα του GNOMM είναι ένα πακέτο αρχείων, το οποίο περιλαμβάνει τη δομή του τελικού συστήματος σε μορφή PDB, καθώς και όλα τα διαθέσιμα αρχεία τοπολογίας και παραμέτρων του δυναμικού πεδίου, τα οποία απαιτούνται για τη διεξαγωγή προσομοιώσεων Μοριακής Δυναμικής.

Όλη η διαδικασία που περιγράφηκε παραπάνω πραγματοποιείται αυτοματοποιημένα. Οι γεωμετρικοί και στερεοδιαταξικοί υπολογισμοί, συμπεριλαμβανομένων των υπολογισμών του διαθέσιμου χώρου σε κάθε στιβάδα, του αριθμού των λιπιδικών μορίων και των περιορισμών θέσης για τις πρωτεΐνες και τα λιπίδια, πραγματοποιούνται μέσω ενός προγράμματος γραμμένου στη γλώσσα Python το οποίο χρησιμοποιεί επιπρόσθετα τη βιβλιοθήκη BioPython, ένα πακέτο εντολών για την ανάλυση βιολογικών δεδομένων (Cock et al., 2009). Από την άλλη, η ίδια η τοποθέτηση των λιπιδίων και η βελτιστοποίηση της θέσης τους πραγματοποιείται μέσω του αλγορίθμου PACKMOL, μιας μεθόδου σχεδιασμένης για τη δημιουργία πολυμερών δομών, αποτελούμενων από μονομερή μόρια (Martinez et al., 2009). Τέλος, όλοι οι ενεργειακοί υπολογισμοί, συμπεριλαμβανομένων των υπολογισμών του ηλεκτροστατικού φορτίου κατά τον καθορισμό της συγκέντρωσης των ιόντων και των βημάτων ελαχιστοποίησης ενέργειας, πραγματοποιούνται μέσω του GROMACS (Abraham et al., 2015). Η μέθοδος GNOMM έχει υλοποιηθεί σαν διαδικτυακή διεπαφή, χρησιμοποιώντας τις γλώσσες προγραμματισμού PHP, JavaScript και Python, καθώς και τις γλώσσες μορφοποίησης HTML5 και CSS3.

2.5.2 Εφαρμογή του GNOMM σε προσομοιώσεις Εξωτερικών Μembrανών

Για την αξιολόγηση της σταθερότητας και πιστότητας, από θερμοδυναμική άποψη, των συστημάτων προσομοίωσης που παράγονται από το GNOMM, πραγματοποιήθηκαν μια σειρά από προσομοιώσεις Μοριακής Δυναμικής, τόσο για απλές διπλοστιβάδες όσο και για μικτές μεμβράνες με σύσταση παρόμοια με αυτή των πραγματικών Εξωτερικών Μembrανών, καθώς επίσης και προσομοιώσεις διαμεμβρανικών β-βαρελίων σε μεμβράνες με LPSs. Όλα τα συστήματα προσομοίωσης κατασκευάστηκαν μέσω της διαδικτυακής διεπαφής της μεθόδου GNOMM. Οι προσομοιώσεις Μοριακής Δυναμικής πραγματοποιήθηκαν χρησιμοποιώντας το GROMACS (Abraham et al., 2015). Αρχικά, πραγματοποιήθηκαν προσομοιώσεις σε απλές λιπιδικές διπλοστιβάδες Λιπιδίου A για κάθε τύπο του λιπιδίου σε όλα τα διαθέσιμα δυναμικά πεδία, τα αποτελέσματα των οποίων χρησιμοποιήθηκαν σε μετρήσεις της επιφάνειας ανά λιπίδιο και του πάχους της διπλοστιβάδας. Ο στόχος των παραπάνω προσομοιώσεων ήταν ο έλεγχος της ποιότητας των παραμέτρων των δυναμικών πεδίων για το Λιπίδιο A, καθώς και η επιβεβαίωση της ύπαρξης σύγκλισης ανάμεσα στα διαφορετικά δυναμικά πεδία. Επιπρόσθετα, πραγματοποιήθηκαν και μια σειρά από προσομοιώσεις μικτών μεμβρανών, αποτελούμενων τόσο από LPSs όσο και από φωσφολιπίδια. Συγκεκριμένα, προτυποποιήθηκαν μεμβράνες με λιπιδική σύσταση η οποία προσομοιάζει τη σύσταση της εξωτερικής μεμβράνης του βακτηρίου *E. coli*, όπως αυτή είναι γνωστή από τη βιβλιογραφία (Lugtenberg & Peters, 1976). Κατασκευάστηκαν δύο συστήματα μεμβρανών. Στο πρώτο σύστημα η εξωτερική στιβάδα αποτελούταν αποκλειστικά από το λιποπολυσακχαρίτη Ra LPS (RaMP) και η εσωτερική στιβάδα αποτελούταν από τα φωσφολιπίδια POPE, POPG και CDL2 σε αναλογία 2:2:1. Από την άλλη, στο δεύτερο σύστημα η εξωτερική στιβάδα ήταν μικτή, αποτελούμενη από RaMP με μικρές παρεμβολές φωσφολιπιδίων POPE, προσομοιάζοντας τις ατέλειες που έχουν βρεθεί να προκαλούνται στην εξωτερική μεμβράνη της *E. coli* από την παρουσία φωσφολιπιδίων (Hsu et al., 2016; Nikaïdo, 2003), τα οποία οδηγούν σε αστάθεια της εξωτερικής στιβάδας και αυξάνουν τη διαπερατότητά της σε βλαβερές για το βακτήριο ουσίες.

Πέρα από τις προσομοιώσεις των μεμβρανών, πραγματοποιήθηκαν μια σειρά από προσομοιώσεις συστημάτων πρωτεΐνης/μεμβράνης, στα οποία διαμεμβρανικά β-βαρέλια ενσωματώθηκαν σε διπλοστιβάδες με LPS στην εξωτερική στιβάδα και φωσφολιπίδια POPE

Κεφάλαιο II – Υλικά & Μέθοδοι

στην εσωτερική. Συγκεκριμένα, πραγματοποιήθηκαν προσομοιώσεις για τα διαμεμβρανικά β-βαρέλια OmpA (PDB: 1BXW) (Pautsch & Schulz, 1998) και OmpF (PDB: 3POX) (Efremov & Sazanov, 2012) της *E. coli* σε διαφορετικά δυναμικά πεδία, τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για την αξιολόγηση της σταθερότητας των πρωτεϊνικών δομών και των αλληλεπιδράσεών τους με μόρια LPS. Τέλος, πραγματοποιήθηκαν προσομοιώσεις του βαρελιού OmpF σε σύμπλοκο με το αντιβιοτικό Αμπικιλλίνη (PDB: 4GCP) (Ziervogel & Roux, 2013), προκειμένου να διερευνηθεί η ικανότητα του GNOMM στην κατασκευή συστημάτων προσομοίωσης με μικρά μόρια (π.χ. φάρμακα).

2.5.3 Μελέτη των αλληλεπιδράσεων ολιγομερισμού σε διαμεμβρανικά β-βαρέλια

Στόχος της παρούσας υποενότητας ήταν η μελέτη των αλληλεπιδράσεων ολιγομερισμού των διαμεμβρανικών β-βαρελίων. Για την επίτευξη αυτού του στόχου προετοιμάστηκαν και πραγματοποιήθηκαν προσομοιώσεις Μοριακής Δυναμικής για μια σειρά από ολιγομερή β-βαρελίων με γνωστή δομή. Συγκεκριμένα, πραγματοποιήθηκαν προσομοιώσεις για τα ομοτμερή των διαμεμβρανικών β-βαρελίων OmpF και OmpP. Η πορίνη OmpF της *E. coli* (PDB: 3POX) και ο μεταφορέας OmpP του *P. aeruginosa* (PDB: 2O4V) είναι ενδεικτικά μέλη της υπερικογένειας των β-βαρελίων με αρχιτεκτονική «n=16,S=20», τα οποία έχουν βρεθεί να σχηματίζουν λειτουργικά ομοτμερή και χαρακτηρίζονται και ως υπερικογένεια τριμερών πορινών (Efremov & Sazanov, 2012; Moraes et al., 2007). Για καθένα από τα δύο ομοτμερή β-βαρελίων κατασκευάστηκαν συστήματα προσομοίωσης σε συνθήκες μεμβρανικού περιβάλλοντος της Εξωτερικής Μembrάνης. Το τριμερές της OmpF τοποθετήθηκε σε μια μεμβράνη αποτελούμενη από το Λιπίδιο A της *E. coli* (ECL1) στην εξωτερική στιβάδα και φωσφολιπίδια POPE στην εσωτερική στιβάδα, ενώ το τριμερές του OmpP τοποθετήθηκε σε μεμβράνη με το 6-ακυλικό Λιπίδιο A του *P. aeruginosa* (PAL6) στην εξωτερική στιβάδα και POPE στην εσωτερική στιβάδα. Όλες οι προσομοιώσεις προετοιμάστηκαν χρησιμοποιώντας το εργαλείο GNOMM, που αναπτύχθηκε στα πλαίσια της διατριβής, και πραγματοποιήθηκαν χρησιμοποιώντας το GROMACS (Abraham et al., 2015) και το δυναμικό πεδίο MARTINI (Periole

Κεφάλαιο II – Υλικά & Μέθοδοι

et al., 2009), σε χρόνους της τάξης του 1 μ s (1000 ns). Τα αποτελέσματα των προσομοιώσεων αναλύθηκαν ως προς τη σταθερότητα των πρωτεϊνικών συμπλόκων, ενώ πραγματοποιήθηκε επιπρόσθετα ανάλυση των αλληλεπιδράσεων ολιγομερισμού ανάμεσα στα μονομερή.

Τέλος, διερευνήθηκε η επίδραση των μορίων LPSs στην ισχύ των αλληλεπιδράσεων ολιγομερισμού, μέσω υπολογισμού του Δυναμικού Μέσης Δύναμης (PMF) κατά την διάσπαση του τριμερούς της OmpF. Για το σκοπό αυτό πραγματοποιήθηκαν προσομοιώσεις σε τέσσερα διαφορετικά είδη μεμβρανών: απλές, φωσφολιπιδικές μεμβράνες POPE και στις δύο στιβάδες, μεμβράνες με Λιπίδιο A στην εξωτερική στιβάδα τους, μεμβράνες με τον «αδρό» λιποπολυσακχαρίτη Re LPS (ReMP), ο οποίος περιλαμβάνει τμήμα του ολιγοσακχαριτικού πυρήνα, και μεμβράνες με τον «λείο» λιποπολυσακχαρίτη Ra LPS (RaMP), ο οποίος διαθέτει πλήρη ολιγοσακχαριτικό πυρήνα. Ο αποχωρισμός του ενός μονομερούς OmpF από το τριμερές κατά μήκος της διπλοστιβάδας προτυποποιήθηκε μέσα από πολλαπλές προσομοιώσεις Καθοδηγούμενης Μοριακής Δυναμικής (SMD) σταθερής ταχύτητας, εφαρμόζοντας μια δύναμη έλξης στο ένα από τα δύο πρωτομερή του διμερούς και ελέγχοντας τη διαδικασία του αποχωρισμού με ένα αρμονικό δυναμικό. Για κάθε σύστημα πραγματοποιήθηκαν 20 διαφορετικές προσομοιώσεις SMD (συνολικά 80 προσομοιώσεις για τα τέσσερα συστήματα), τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για τον υπολογισμό του PMF, εφαρμόζοντας τη μέθοδο της Ισότητας του Jarzynski (Jarzynski, 1997a; Jarzynski, 1997b; Park et al., 2003).

2.6 Ο διαμεμβρανικός μεταφορέας SLC11A1

Στόχος της παρούσας ενότητας ήταν η μελέτη της δομής του μεταφορέα SLC11A1 και των αλληλεπιδράσεών του με το φυσικό υπόστρωμά του, τα δισθενή ιόντα Fe^{2+} και Mn^{2+} . Για την επίτευξη αυτού του στόχου πραγματοποιήθηκε αρχικά η κατασκευή μοντέλων της δομής του μεταφορέα σε διαφορετικές στερεοδιατάξεις. Συγκεκριμένα, κατασκευάστηκαν μοντέλα του SLC11A1 στην ανοιχτή προς τον εξωκυττάριο χώρο (outward-facing) στερεοδιάταξη και την ανοιχτή προς το κυτταρόπλασμα (inward-facing) στερεοδιάταξη, καθώς και σύμπλοκα του μεταφορέα με Fe^{2+} ή Mn^{2+} στη θέση πρόσδεσης των ιόντων. Επιπρόσθετα, κατασκευάστηκαν μοντέλα του μεταφορέα με μεταλλαγές σε τρεις θέσεις με γνωστή συμμετοχή στην πρόσδεση ιόντων (Ala-68-Gly, Asp-71-Ala και Met-250-Ala). Ως πρότυπα για την κατασκευή των μοντέλων χρησιμοποιήθηκαν οι δομές τριών βακτηριακών ομολόγων του SLC11A1 (Bozzi et al., 2016a; Ehrnstorfer et al., 2014; Ehrnstorfer et al., 2017), οι οποίες είναι διαθέσιμες στην PDB (PDB: 4GWV, 5M95, 5KTE). Η σταθερότητα και δυναμική συμπεριφορά κάθε μοντέλου ελέγχθηκε μέσα από προσομοιώσεις Μοριακής Δυναμικής σε περιβάλλον μεμβράνης με το GROMACS και το δυναμικό πεδίο CHARMM36m (Huang et al., 2017). Τέλος, οι αλληλεπιδράσεις του SLC11A1 με δισθενή ιόντα Fe^{2+} και Mn^{2+} μελετήθηκαν με ενεργειακούς υπολογισμούς, χρησιμοποιώντας τη μέθοδο των Διαταραχών Ελεύθερης Ενέργειας (FEP).

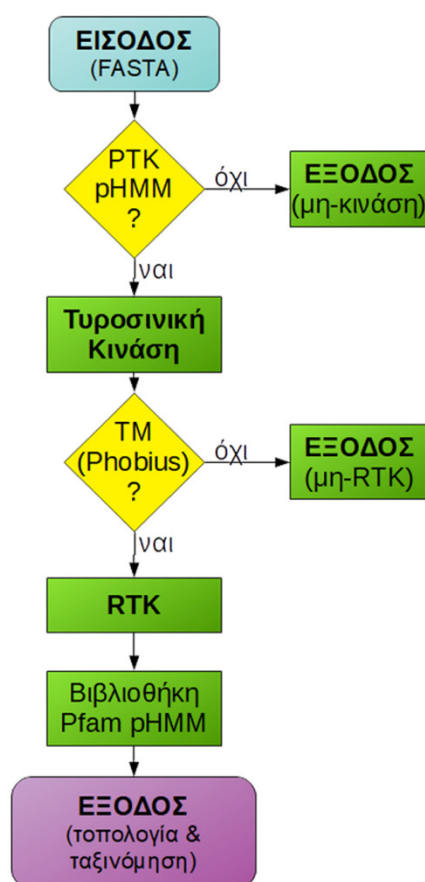
2.7 Υποδοχείς με Ενεργότητα Τυροσινικής Κινάσης (RTKs)

2.7.1 Κατασκευή μεθόδου πρόγνωσης και σχολιασμού των RTKs: ο αλγόριθμος RTK-PRED

Οι υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs) επιτελούν μια σειρά από σημαντικές λειτουργίες στους ευκαρυωτικούς οργανισμούς, όπως η ρύθμιση της γονιδιακής έκφρασης, η κυτταρική αύξηση και διαφοροποίηση και ο κυτταρικός θάνατος, και έχουν συσχετιστεί με μια σειρά από ασθένειες στον άνθρωπο όπως διάφοροι τύποι καρκίνου, μεταβολικά σύνδρομα και νευροεκφυλιστικές ασθένειες. Παρά τη σημασία τους, ωστόσο,

Κεφάλαιο II – Υλικά & Μέθοδοι

έχουν μελετηθεί ελάχιστα σε οργανισμούς εκτός του ανθρώπου, του ποντικού και του αρουραίου. Επιπρόσθετα, δεν υπάρχει κάποια μέθοδος ανίχνευσης των RTKs από την αλληλουχία τους, η οποία θα μπορούσε να χρησιμοποιηθεί στον εντοπισμό πιθανών RTKs για το σχολιασμό πρωτεωμάτων. Με βάση τα παραπάνω, αναπτύχθηκε στα πλαίσια της παρούσας διδακτορικής διατριβής η μέθοδος RTK-PRED, η οποία πραγματοποιεί εντοπισμό και σχολιασμό RTKs από την αλληλουχία τους με τη χρήση pHMMs. Στόχος της μεθόδου είναι ο εντοπισμός νέων RTKs σε ελάχιστα μελετημένα πρωτεώματα, ο σχολιασμός της τοπολογίας τους και η κατάταξή τους σε υποοικογένειες.



Εικόνα 47 Η πορεία εργασίας του RTK-PRED. Διάγραμμα ροής του αλγορίθμου πρόγνωσης RTK-PRED για την αναγνώριση και το σχολιασμό RTKs από την αλληλουχία.

Η διαδικασία λειτουργίας της μεθόδου RTK-PRED παρουσιάζεται στο διάγραμμα ροής της Εικόνας 47. Η μέθοδος δέχεται σαν είσοδο μία ή περισσότερες πρωτεϊνικές αλληλουχίες σε μορφή FASTA και πραγματοποιεί την πρόγνωση των RTKs σε τρία βήματα. Στο πρώτο βήμα

Κεφάλαιο II – Υλικά & Μέθοδοι

γίνεται η ανίχνευση των πρωτεϊνών με ενεργότητα τυροσινικής κινάσης (PTKs) από το σύνολο των αλληλουχιών που εξετάζονται, μέσω ενός pHMM ειδικού για την αναγνώριση της περιοχής τυροσινικής κινάσης. Στο δεύτερο βήμα, οι PTKs που αναγνωρίζονται ελέγχονται ως προς την τοπολογία τους σε σχέση με τη μεμβράνη, χρησιμοποιώντας τον αλγόριθμο Phobius (Kall et al., 2004). Οι PTKs που διαθέτουν τη σωστή τοπολογία χαρακτηρίζονται ως RTKs και, στο τρίτο βήμα, χρησιμοποιούνται για την ανίχνευση των εξωκυτταρικών αυτοτελών δομικών περιοχών τους, μέσω μιας βιβλιοθήκης pHMMs ειδικών για τις αυτοτελείς δομικές περιοχές των RTKs από τη βάση δεδομένων Pfam (El-Gebali et al., 2019), ενώ παράλληλα γίνεται και ταξινόμηση των RTKs σε υποοικογένειες, χρησιμοποιώντας ως κριτήριο το είδος των εξωκυτταρικών περιοχών τους.

Το πρώτο βήμα στη μέθοδο RTK-PRED, όπως αναφέρθηκε, είναι η αναγνώριση των πρωτεϊνών με ενεργότητα τυροσινικής κινάσης. Για το σκοπό αυτό χρησιμοποιείται ένα pHMM ειδικό για την αυτοτελή δομική περιοχή τυροσινικής κινάσης (PTK). Σημειώνεται πως στη βάση δεδομένων Pfam υπάρχει διαθέσιμο ένα προφίλ για την περιοχή τυροσινικής κινάσης (Pfam: PF07714). Ωστόσο, το συγκεκριμένο προφίλ δεν είναι αρκετά ειδικό, με αποτέλεσμα να αναγνωρίζει λανθασμένα και άλλες κατηγορίες κινασών όπως οι κινάσες Ser/Thr, οδηγώντας σε μεγάλο αριθμό ψευδώς θετικών αποτελεσμάτων. Έτσι, για τις ανάγκες της μεθόδου κατασκευάστηκε ένα νέο pHMM ειδικό προς την περιοχή PTK. Για την κατασκευή και την εκπαίδευση του προφίλ χρησιμοποιήθηκε η μέθοδος εκπαίδευσης HMM-ModE για το HMMER v. 3 (Sinha & Lynn, 2014), η οποία επιτρέπει τη δημιουργία pHMMs με τη χρήση τόσο θετικού όσο και αρνητικού συνόλου εκπαίδευσης, με σκοπό τη δημιουργία πιο εξειδικευμένων μοντέλων.

Για την κατασκευή των συνόλων εκπαίδευσης πραγματοποιήθηκαν αναζητήσεις στη UniProtKB/Swiss-Prot (The UniProt Consortium, 2019) για τη συλλογή καλά σχολιασμένων πρωτεϊνικών αλληλουχιών. Για το θετικό σύνολο εκπαίδευσης συγκεντρώθηκαν αρχικά 194 καλά σχολιασμένες αλληλουχίες τυροσινικών κινασών, οι οποίες περιελάμβαναν τόσο RTKs όσο και σφαιρικές υδατοδιαλυτές τυροσινικές κινάσες. Τα τμήματα των αλληλουχιών που περιελάμβαναν την περιοχή PTK απομονώθηκαν με βάση το σχολιασμό των πρωτεϊνών στη Swiss-Prot, και οι αλληλουχίες τους χρησιμοποιήθηκαν για την κατασκευή του συνόλου

Κεφάλαιο II – Υλικά & Μέθοδοι

εκπαίδευσης. Αντίστοιχα, για την κατασκευή του αρνητικού συνόλου εκπαίδευσης συγκεντρώθηκαν 1986 αλληλουχίες, που περιελάμβαναν πρωτεΐνες από τις υπόλοιπες κατηγορίες κινασών (κινάσες σερίνης/θρεονίνης, ιστιδίνης, γλουταμικού και μη πρωτεϊνικές κινάσες), από τις οποίες απομονώθηκαν οι αλληλουχίες των περιοχών κινάσης του κάθε τύπου, καθώς επίσης και ένα ετερογενές σύνολο άλλων πρωτεϊνών χωρίς ενεργότητα κινάσης. Τόσο το θετικό όσο και το αρνητικό σύνολο εκπαίδευσης είναι διαθέσιμα στη διεύθυνση <http://hector.biol.uoa.gr/RTK-PRED/about.html>.

Το δεύτερο βήμα στη μέθοδο RTK-PRED είναι ο διαχωρισμός των PTKs που ανιχνεύθηκαν στο πρώτο βήμα σε RTKs και σφαιρικές υδατοδιαλυτές PTKs. Για το σκοπό αυτό χρησιμοποιήθηκε ο αλγόριθμος Phobius (<http://phobius.sbc.su.se/>), ο οποίος πραγματοποιεί συνδυαστική πρόγνωση του σηματοδοτικού πεπτιδίου και των διαμεμβρανικών τμημάτων για α-ελικοειδείς πρωτεΐνες με τη χρήση HMMs. Ο αλγόριθμος είναι ικανός να αναγνωρίσει αν μια πρωτεΐνη είναι διαμεμβρανική ή όχι και, στην περίπτωση των διαμεμβρανικών πρωτεϊνών, να προβλέψει την τοπολογία τους, συμπεριλαμβανομένης της παρουσίας ή όχι σηματοδοτικού πεπτιδίου, της θέσης του N-άκρου και τον αριθμό και τη θέση των διαμεμβρανικών τμημάτων (Kall et al., 2004). Στη μέθοδο RTK-PRED ο αλγόριθμος χρησιμοποιείται για να αναγνωρίσει τις αλληλουχίες των PTKs που διαθέτουν τη σωστή τοπολογία (σηματοδοτικό πεπτίδιο, 1 διαμεμβρανικό τμήμα, εξωκυτταρικό N-άκρο και ενδοκυτταρικό C-άκρο), οι οποίες χαρακτηρίζονται ως RTKs, ενώ οι υπόλοιπες αλληλουχίες χαρακτηρίζονται ως μη-RTKs (nRTKs), οδηγώντας στη λήξη της διαδικασίας.

Το τρίτο βήμα της μεθόδου RTK-PRED είναι ο σχολιασμός της τοπολογίας των αλληλουχιών που προβλέπονται ως πιθανοί RTKs. Αυτός ο σχολιασμός περιλαμβάνει τις θέσεις του σηματοδοτικού πεπτιδίου, του διαμεμβρανικού τμήματος και της περιοχής PTK, καθώς και το είδος και τις θέσεις των εξωκυτταρικών περιοχών του κάθε υποδοχέα. Η θέση της περιοχής PTK εντοπίζεται μέσα από τη στοίχιση της κάθε αλληλουχίας με το pHMM της περιοχής PTK που κατασκευάστηκε για τη μέθοδο, ενώ οι θέσεις του σηματοδοτικού πεπτιδίου και του διαμεμβρανικού τμήματος λαμβάνονται από τα αποτελέσματα του Phobius. Τέλος, για την αναγνώριση των εξωκυτταρικών περιοχών κατασκευάστηκε μια βιβλιοθήκη pHMMs χρησιμοποιώντας τη λειτουργία *hmmpress* του πακέτου HMMER, η οποία περιέχει τα προφίλ

Κεφάλαιο II – Υλικά & Μέθοδοι

των αυτοτελών δομικών περιοχών με γνωστή παρουσία στο N-άκρο των RTKs από τη βάση δεδομένων Pfam (Πίνακας 7).

Πίνακας 7 Η βιβλιοθήκη των pHMMs της Pfam τα οποία χρησιμοποιούνται στη μέθοδο RTK-PRED.

Όνομα	Κωδικός Pfam	Όνομα	Κωδικός Pfam
WIF	PF02019	Kringle	PF00051
TIG	PF01833	Ig domain	PF00047
Sema	PF01403	Fz domain	PF01392
L-domain	PF01030	Furin-like Cysteine-rich	PF00757
MAM	PF00629	Frizzled	PF01534
Leucine-Rich	PF13855	Fn Type 3	PF00041
LDL Class A	PF00057	Discoidin	PF00754
Eph binding	PF01404	Cadherin	PF00028

Το τελευταίο μέρος της μεθόδου είναι η κατηγοριοποίηση των RTKs. Συγκεκριμένα, χρησιμοποιείται το σύστημα ταξινόμησης της IUPHAR (Alexander et al., 2017), το οποίο κατατάσσει τους RTKs σε υποοικογένειες με κριτήριο το είδος των εξωκυτταρικών περιοχών που διαθέτουν. Έτσι, η ταξινόμηση των RTKs πραγματοποιείται με βάση τα αποτελέσματα από το σχολιασμό της τοπολογίας τους ως προς τις εξωκυτταρικές περιοχές τους, όπως αυτός πραγματοποιήθηκε στο προηγούμενο βήμα. Στην περίπτωση των RTKs που δεν περιλαμβάνουν κάποια από τις γνωστές περιοχές της βιβλιοθήκης pHMMs, αυτοί ταξινομούνται στην κατηγορία «Other», η οποία περιλαμβάνει τους υποδοχείς που δεν αναγνωρίζονται από το σύστημα της IUPHAR.

Η μέθοδος RTK-PRED έχει υλοποιηθεί με τη μορφή ενός προγράμματος γραμμένου στη γλώσσα Perl, το οποίο πραγματοποιεί αυτοματοποιημένα τη διαδικασία που περιγράφηκε. Όλες οι διαδικασίες που περιλαμβάνουν τη χρήση pHMMs πραγματοποιούνται χρησιμοποιώντας τις λειτουργίες *hmmsearch* και *hmmalign* του πακέτου HMMER v.3 (Eddy, 2011), ενώ η πρόγνωση της τοπολογίας γίνεται καλώντας τον αλγόριθμο Phobius. Η μέθοδος RTK-PRED είναι ελεύθερα διαθέσιμη σαν διαδικτυακή διεπαφή, κατασκευασμένη χρησιμοποιώντας τις γλώσσες προγραμματισμού PHP, JavaScript και Perl, καθώς και τις γλώσσες μορφοποίησης HTML5 και CSS3.

2.7.2 Αξιολόγηση της μεθόδου RTK-PRED

Η αξιολόγηση της μεθόδου RTK-PRED πραγματοποιήθηκε χρησιμοποιώντας μια σειρά από δείκτες/μέτρα. Συγκεκριμένα, προσδιορίστηκαν η Ευαισθησία (Sensitivity, Sn), η Ειδικότητα (Specificity, Sp) και η Ακρίβεια (Accuracy, Acc) της μεθόδου, καθώς επίσης και ο Συντελεστής Συσχέτισης Matthews (Matthews Correlation Coefficient, MCC). Στον προσδιορισμό όλων των παραπάνω τιμών χρησιμοποιούνται τέσσερις όροι για την περιγραφή των αποτελεσμάτων της πρόγνωσης:

1. Αληθώς θετικά (True Positives, TP): τα αποτελέσματα που προβλέφθηκαν σωστά ως θετικά.
2. Αληθώς αρνητικά (True Negatives, TN): τα αποτελέσματα που προβλέφθηκαν σωστά ως αρνητικά.
3. Ψευδώς θετικά (False Positives, FP): τα αποτελέσματα που προβλέφθηκαν λανθασμένα ως θετικά ενώ είναι αρνητικά.
4. Ψευδώς αρνητικά (False Negatives): τα αποτελέσματα που προβλέφθηκαν λανθασμένα ως αρνητικά ενώ είναι θετικά.

Ως «θετικό» ή «αρνητικό» αποτέλεσμα ορίζεται η ιδιότητα της κάθε μέτρησης σε σχέση με το αντικείμενο της μεθόδου πρόγνωσης. Έτσι, για το pHMM της περιοχής PTK ως θετικό αποτέλεσμα ορίζεται το να είναι μια πρωτεΐνη τυροσινική κινάση, ενώ για τη μέθοδο RTK-PRED, ως θετικό αποτέλεσμα ορίζεται να είναι μια πρωτεΐνη RTK.

Η Ευαισθησία εκφράζει το ποσοστό των αληθώς θετικών (TP) προβλέψεων στο σύνολο των θετικών τιμών, σύμφωνα με τον τύπο:

$$Sn = \frac{TP}{TP+FN} \cdot 100 \quad (87)$$

Η Ειδικότητα εκφράζει το ποσοστό των αληθώς αρνητικών (TN) προβλέψεων στο σύνολο των αρνητικών τιμών, σύμφωνα με τον τύπο:

$$Sp = \frac{TN}{TN+FP} \cdot 100 \quad (88)$$

Η Ακρίβεια εκφράζει το ποσοστό των επιτυχών προβλέψεων (TP+TN) στο σύνολο των προβλέψεων, σύμφωνα με τον τύπο:

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \cdot 100 \quad (89)$$

Κεφάλαιο II – Υλικά & Μέθοδοι

Τέλος, ο συντελεστής συσχέτισης Matthews εκφράζει την ποιότητα της πρόγνωσης σε συστήματα «δυαδικού» τύπου (θετικό/αρνητικό). Λειτουργεί ως μέτρο της συσχέτισης ανάμεσα στα αποτελέσματα της πρόβλεψης και τα πειραματικά δεδομένα και λαμβάνει τιμές από -1 ως +1, με τιμές κοντά στο +1 να υποδηλώνουν θετική συσχέτιση, τιμές κοντά στο -1 να υποδηλώνουν αρνητική συσχέτιση, ενώ τιμές κοντά στο μηδέν υποδηλώνουν απουσία συσχέτισεων. Ο συντελεστής υπολογίζεται σύμφωνα με τον παρακάτω τύπο:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (90)$$

Πραγματοποιήθηκε αρχικά η αξιολόγηση της απόδοσης του ειδικού pHMM της περιοχής PTK στον εντοπισμό πρωτεϊνών με ενεργότητα τυροσινικής κινάσης, καθώς και η σύγκρισή του με το αντίστοιχο προφίλ της Pfam. Για το σκοπό αυτό κατασκευάστηκαν δύο μη ομόλογα σύνολα ελέγχου, μέσα από αναζητήσεις στη βάση δεδομένων UniProtKB/Swiss-Prot: ένα θετικό σύνολο, αποτελούμενο από 76 πρωτεΐνες με ενεργότητα τυροσινικής κινάσης, καθώς και ένα αρνητικό σύνολο, αποτελούμενο από 640 πρωτεΐνες με ενεργότητα κινασών άλλων κατηγοριών. Και τα δύο σύνολα αποτελούνται από αλληλουχίες οι οποίες δεν περιλαμβάνονταν στο σύνολο εκπαίδευσης των δύο pHMMs. Στη συνέχεια, πραγματοποιήθηκε η αξιολόγηση της μεθόδου RTK-PRED στο σύνολό της, ως προς την ικανότητα αναγνώρισης RTKs. Για το σκοπό αυτό κατασκευάστηκαν δύο σύνολα ελέγχου: ένα θετικό σύνολο, αποτελούμενο από 44 RTKs και ένα αρνητικό σύνολο, αποτελούμενο από 32 PTKs οι οποίες δεν είναι RTKs (hRTKs), καθώς και 640 άλλες πρωτεΐνες, ανάμεσα στις οποίες συμπεριλαμβάνονται κινάσες με διαφορετικό υπόστρωμα, διαμεμβρανικές πρωτεΐνες οι οποίες δεν είναι RTKs, καθώς και διάφορες σφαιρικές πρωτεΐνες. Σημειώνεται πως τόσο το θετικό σύνολο των 44 RTKs όσο και οι 32 PTKs του αρνητικού συνόλου δεν περιλαμβάνονταν στα σύνολα εκπαίδευσης της μεθόδου. Τέλος, αξιολογήθηκε η ικανότητα της μεθόδου στη σωστή ταξινόμηση των RTKs σε υποοικογένειες. Για το σκοπό αυτό κατασκευάστηκε ένα σύνολο 269 RTKs από όλες τις υποοικογένειες των υποδοχέων, και ελέγχθηκε η ικανότητα της μεθόδου στη σωστή κατηγοριοποίησή τους.

2.8 Μεμβρανικές πρωτεΐνες του Πυρηνικού Φακέλου

2.8.1 Κατασκευή δικτύου αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου στον άνθρωπο

Στόχος της παρούσας ενότητας ήταν η κατασκευή του δικτύου αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου στον άνθρωπο. Για την επίτευξη αυτού του στόχου συγκεντρώθηκαν αρχικά όλες οι ανθρώπινες πρωτεΐνες με γνωστή παρουσία στον πυρηνικό φάκελο από τη βάση δεδομένων UniProt/SwissProt (The UniProt Consortium, 2019). Ως κριτήριο για την επιλογή των πρωτεϊνών χρησιμοποιήθηκε το πεδίο της Υποκυτταρικής Θέσης («Subcellular Location»), της UniProt, στο οποίο παρατίθενται όλες οι διακριτές υποκυτταρικές θέσεις, η περιγραφή και τα συνώνυμά τους, ενώ σε κάθε θέση ανατίθεται ένας μοναδικός κωδικός (π.χ. SL-0179). Στην περίπτωση του πυρηνικού φακέλου, η UniProt ορίζει μια σειρά από διακριτές υποκυτταρικές θέσεις, τόσο για τον ίδιο πυρηνικό φάκελο όσο και για τα υποσύνολά του (The UniProt Consortium, 2019). Μια λίστα με αυτές τις θέσεις και την περιγραφή τους παρατίθεται στον Πίνακα 8. Επιλέχθηκαν όλες οι πρωτεΐνες του ανθρώπου (*Homo sapiens*) με ανάθεση σε τουλάχιστον μία από αυτές τις υποκυτταρικές θέσεις και, στη συνέχεια, κατηγοριοποιήθηκαν ως προς τη θέση τους στον πυρηνικό φάκελο και ως προς την τοπολογία τους.

Το αρχικό σύνολο πρωτεϊνών, το οποίο προέκυψε από την παραπάνω διαδικασία, ήταν 432 ανθρώπινες πρωτεΐνες. Αυτές οι πρωτεΐνες χρησιμοποιήθηκαν για την αναζήτηση αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών, μέσω της βάσης δεδομένων βιομοριακών αλληλεπιδράσεων IntAct (Orchard et al., 2014). Η IntAct αξιολογεί κάθε αλληλεπίδραση για την αξιοπιστία της με κριτήρια τη μέθοδο προσδιορισμού της, τον αριθμό των πειραμάτων επιβεβαίωσης και το είδος της αλληλεπίδρασης. Οι αλληλεπιδράσεις βαθμολογούνται χρησιμοποιώντας μια κανονικοποιημένη κλίμακα βαθμολόγησης (Mi-Score) η οποία μπορεί να λάβει τιμές από 0.0 ως 1.0, με υψηλότερες τιμές να αντιστοιχούν σε πιο αξιόπιστα δεδομένα. Στην αναζήτηση των αλληλεπιδράσεων ορίστηκε ένα κατώφλι αξιοπιστίας, χρησιμοποιώντας την τιμή Mi-Score της βάσης δεδομένων και ορίζοντας σαν κατώτατο όριο αξιοπιστίας την τιμή

Κεφάλαιο II – Υλικά & Μέθοδοι

0.5. Η παραπάνω διαδικασία πραγματοποιήθηκε για την υπολογιστική βελτιστοποίηση του δικτύου αλληλεπιδράσεων, καθώς και για την αποφυγή χρήσης πληροφορίας χαμηλής ποιότητας, η οποία θα μπορούσε να μειώσει την αξιοπιστία του δικτύου. Επιπρόσθετα, προηγούμενες υπολογιστικές μελέτες πάνω στα δίκτυα πρωτεϊνικών αλληλεπιδράσεων έχουν αποδείξει πως η εφαρμογή κατωφλίου αξιοπιστίας μπορεί να παράξει δίκτυα αλληλεπιδράσεων με χαρακτηριστικά πραγματικών βιολογικών δικτύων, τα οποία περιέχουν πληροφορία υψηλότερης ποιότητας και μπορούν να περιγράφουν επαρκώς τις ιδιότητες του αρχικού συνόλου δεδομένων (Mora & Donaldson, 2012; Nastou et al., 2014).

Πίνακας 8 Οι υποκυτταρικές θέσεις που ορίζονται από τη UniProt για τον πυρηνικό φάκελο.

Όνομα	Κωδικός	Περιγραφή (σύμφωνα με τη UniProt)
Πυρηνικός φάκελος	SL-0178	Το συνολικό σύστημα μεμβρανών που περιβάλλει τον πυρήνα των ευκαρυωτικών κυττάρων. Αποτελείται από την πυρηνική λάμινα, τα συμπλέγματα πυρηνικού πόρου και δύο πυρηνικές μεμβράνες.
Πυρηνική μεμβράνη	SL-0182	Η μεμβράνη που περιβάλλει τον πυρήνα. Ο συγκεκριμένος όρος χρησιμοποιείται όταν δεν είναι γνωστό αν η πρωτεΐνη βρίσκεται στην εσωτερική ή την εξωτερική πυρηνική μεμβράνη.
Εξωτερική πυρηνική μεμβράνη	SL-0183	Η εξωτερική μεμβράνη του φακέλου βρίσκεται στην πλευρά του κυτταροπλάσματος. Στα θηλαστικά αποτελεί συνέχεια του Ενδοπλασματικού Δικτύου και περιέχει πολλά ριβοσώματα.
Εσωτερική πυρηνική μεμβράνη	SL-0179	Η εσωτερική μεμβράνη του φακέλου βρίσκεται στην πλευρά του πυρηνοπλάσματος. Στα θηλαστικά βρίσκεται σε επαφή με την πυρηνική λάμινα και τη χρωματίνη.
Περιπυρηνικός ενδομεμβρανικός χώρος	SL-0184	Ο χώρος ανάμεσα στην εξωτερική και την εσωτερική μεμβράνη του φακέλου.
Σύμπλεγμα πυρηνικού πόρου	SL-0185	Ο διάυλος μεταφοράς ουσιών από και προς το πυρηνόπλασμα.
Πυρηνική Λάμινα	SL-0180	Ένα στρώμα ενδιάμεσων ινιδίων (λαμίνες), οι οποίες προσκολλώνται στην εσωτερική πυρηνική μεμβράνη.
Περι-πυρηνική περιοχή	SL-0198	Η περιοχή του κυτταροπλάσματος που βρίσκεται σε άμεση επαφή με τον πυρηνικό φάκελο.

Κεφάλαιο II – Υλικά & Μέθοδοι

Τα αποτελέσματα της παραπάνω αναζήτησης έδωσαν αλληλεπιδράσεις για 117 από τις 432 πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου, οι οποίες αλληλεπιδρούν με 649 πρώτους γείτονες. Σημειώνεται ότι σε αρκετές περιπτώσεις, οι πρώτοι γείτονες μιας πρωτεΐνης του πυρηνικού φακέλου ήταν άλλες πρωτεΐνες του πυρηνικού φακέλου, ωστόσο, η πλειονότητα των πρώτων γειτόνων περιελάμβανε πρωτεΐνες από άλλες υποκυτταρικές θέσεις. Για τον περαιτέρω εμπλουτισμό του συνόλου δεδομένων και την παραγωγή ενός πιο αληθοφανούς δικτύου αλληλεπιδράσεων, πραγματοποιήθηκε επιπρόσθετη αναζήτηση στην IntAct για την εύρεση αλληλεπιδράσεων ανάμεσα στους πρώτους γείτονες, χρησιμοποιώντας τα ίδια κριτήρια αναζήτησης, όπως αυτά περιγράφηκαν.

Το τελικό σύνολο δεδομένων που προέκυψε και το οποίο χρησιμοποιήθηκε για την κατασκευή του δικτύου αλληλεπιδράσεων περιελάμβανε 766 πρωτεΐνες (κόμβοι) που συνδέονται από 2271 αλληλεπιδράσεις (ακμές). Από αυτές, οι 839 ακμές αφορούν τις αλληλεπιδράσεις των πρωτεϊνών του πυρηνικού φακέλου (117), ενώ οι υπόλοιπες ακμές αντιστοιχούν στις αλληλεπιδράσεις μεταξύ των γειτόνων. Το δίκτυο κατασκευάστηκε και οπτικοποιήθηκε χρησιμοποιώντας το λογισμικό οπτικοποίησης δικτύων Cytoscape v. 3.5 (Shannon et al., 2003). Πραγματοποιήθηκε ανάλυση του δικτύου αλληλεπιδράσεων χρησιμοποιώντας τη θεωρία γράφων, μέσω της εφαρμογής Network Analyzer (Assenov et al., 2008), η οποία υπάρχει διαθέσιμη στο Cytoscape, καθώς επίσης και μέσα από προγράμματα γραμμένα στη γλώσσα προγραμματισμού Python, χρησιμοποιώντας τη βιβλιοθήκη NetworkX (Hagberg et al., 2008). Στα πλαίσια της ανάλυσης του δικτύου πραγματοποιήθηκε περαιτέρω μελέτη του μέσω ανάλυσης ομαδοποίησης με δύο προσεγγίσεις: α) με βάση την υποκυτταρική θέση και β) υπολογιστικά, χρησιμοποιώντας τον αλγόριθμο Markov Clustering (MCL) (Enright et al., 2002; van Dongen, 2000). Τέλος, πραγματοποιήθηκε λειτουργικός εμπλουτισμός του δικτύου με δεδομένα από τη βάση Gene Ontology, χρησιμοποιώντας την εφαρμογή BiNGO του Cytoscape (Maere et al., 2005) και το πρόγραμμα WebGestalt (Wang et al., 2013b; Wang et al., 2017; Zhang et al., 2005).

2.8.2 Κατασκευή βάσης δεδομένων για τις πρωτεΐνες του Πυρηνικού Φακέλου: η βάση δεδομένων NucEnvDB

Έχοντας ολοκληρώσει την ανάλυση του δικτύου αλληλεπιδράσεων για τον ανθρώπινο πυρηνικό φάκελο, επόμενος στόχος ήταν η κατασκευή μιας βάσης δεδομένων για όλες τις καλά σχολιασμένες πρωτεΐνες του πυρηνικού φακέλου. Για το σκοπό αυτό αναπτύχθηκε η NucEnvDB, μια σχεσιακή βάση δεδομένων των πρωτεϊνών του πυρηνικού φακέλου και των αλληλεπιδράσεών τους. Η συλλογή των δεδομένων για την κατασκευή της βάσης πραγματοποιήθηκε εφαρμόζοντας τη διαδικασία αναζήτησης που εφαρμόστηκε στο δίκτυο αλληλεπιδράσεων του ανθρώπινου πυρηνικού φακέλου και επεκτείνοντάς την στο σύνολο των ευκαρυωτικών οργανισμών. Τα δεδομένα των πρωτεϊνών του πυρηνικού φακέλου, συμπεριλαμβανομένων των ετεροαναφορών τους σε μεγάλες βάσεις βιολογικών δεδομένων και του λειτουργικού σχολιασμού τους, ανασύρθηκαν από τη UniProt/SwissProt ενώ οι αλληλεπιδράσεις τους με άλλες πρωτεΐνες ανασύρθηκαν από την IntAct. Από την παραπάνω αναζήτηση προέκυψαν δεδομένα για 2863 πρωτεΐνες από 394 οργανισμούς, οι οποίες έχουν συσχετιστεί με 4680 όρους γονιδιακής οντολογίας και συμμετέχουν σε 20378 αλληλεπιδράσεις με πρώτους γείτονες.

Τα δεδομένα του συνόλου χρησιμοποιήθηκαν για την κατασκευή μιας βάσης δεδομένων SQL. Η διαχείριση των δεδομένων πραγματοποιήθηκε μέσω των γλωσσών προγραμματισμού Perl και Python, ενώ η βάση δεδομένων κατασκευάστηκε χρησιμοποιώντας το σύστημα διαχείρισης MySQL. Για την οπτικοποίηση των δεδομένων και την πρόσβαση του χρήστη στη βάση σχεδιάστηκε μια διαδικτυακή διεπαφή, χρησιμοποιώντας τις γλώσσες προγραμματισμού PHP, JavaScript, Python και R, καθώς και τις γλώσσες μορφοποίησης HTML5 και CSS3. Στη διεπαφή ενσωματώθηκαν τα εργαλεία BLAST (Altschul et al., 1990) και HMMER (Eddy, 2011) για την πραγματοποίηση αναζητήσεων αλληλουχιών και pHMMs, καθώς και το πρόγραμμα μοριακών γραφικών LiteMol (Mir et al., 2018; Sehnal et al., 2017) για την οπτικοποίηση πρωτεϊνών με γνωστή δομή. Επιπρόσθετα, σχεδιάστηκε μια εφαρμογή για την κατασκευή, ανάλυση και οπτικοποίηση δικτύων αλληλεπιδράσεων με τα περιεχόμενα της βάσης από το χρήστη, χρησιμοποιώντας το πρόγραμμα Cytoscape.js (Franz et al., 2016), μια

Κεφάλαιο II – Υλικά & Μέθοδοι

εκδοχή του Cytoscape γραμμένη σε JavaScript για χρήση σε ιστοσελίδες, και τη βιβλιοθήκη NetworkX για τη γλώσσα Python (Hagberg et al., 2008). Τέλος, στη βάση ενσωματώθηκε μια εφαρμογή για το λειτουργικό εμπλουτισμό των περιεχομένων της αλλά και των δικτύων αλληλεπιδράσεων του χρήστη, χρησιμοποιώντας το πρόγραμμα WebGestaltR, μια εκδοχή της μεθόδου WebGestalt (Wang et al., 2013b; Wang et al., 2017; Zhang et al., 2005) στη γλώσσα R, η οποία έχει σχεδιαστεί για τον αυτοματοποιημένο εμπλουτισμό μεγάλου όγκου δεδομένων.

ΚΕΦΑΛΑΙΟ III

ΑΠΟΤΕΛΕΣΜΑΤΑ – ΣΥΖΗΤΗΣΗ

Τα αποτελέσματα της διδακτορικής διατριβής, τα οποία παρουσιάζονται στο παρόν κεφάλαιο, χωρίζονται σε πέντε διακριτές ενότητες. Στην πρώτη ενότητα παρουσιάζονται τα αποτελέσματα πάνω στην μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs όπως αυτές μελετήθηκαν με προσομοιώσεις Μοριακής Δυναμικής. Η δεύτερη ενότητα περιλαμβάνει τα αποτελέσματα πάνω στη μελέτη των εξωτερικών μεμβρανών των Gram-αρνητικών βακτηρίων και περιλαμβάνει τα αποτελέσματα της μεθόδου GNOMM για την προτυποποίηση μεμβρανών, καθώς και της εφαρμογής της μεθόδου στη μελέτη αλληλεπιδράσεων ολιγομερισμού σε διαμεμβρανικά β-βαρέλια, μέσα από προσομοιώσεις Μοριακής Δυναμικής. Η τρίτη ενότητα περιγράφει τα αποτελέσματα της δομικής μελέτης πάνω στον μεταφορέα SLC11A1. Η τέταρτη ενότητα παρουσιάζει τα αποτελέσματα του αλγορίθμου RTK-PRED για τη μελέτη των RTKs ενώ, τέλος, η πέμπτη ενότητα παρουσιάζει τα αποτελέσματα από τη συστηματική ανάλυση των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο, καθώς και τη βάση δεδομένων NucEnvDB.

3.1 Μελέτες των αλληλεπιδράσεων ολιγομερισμού στους GPCRs

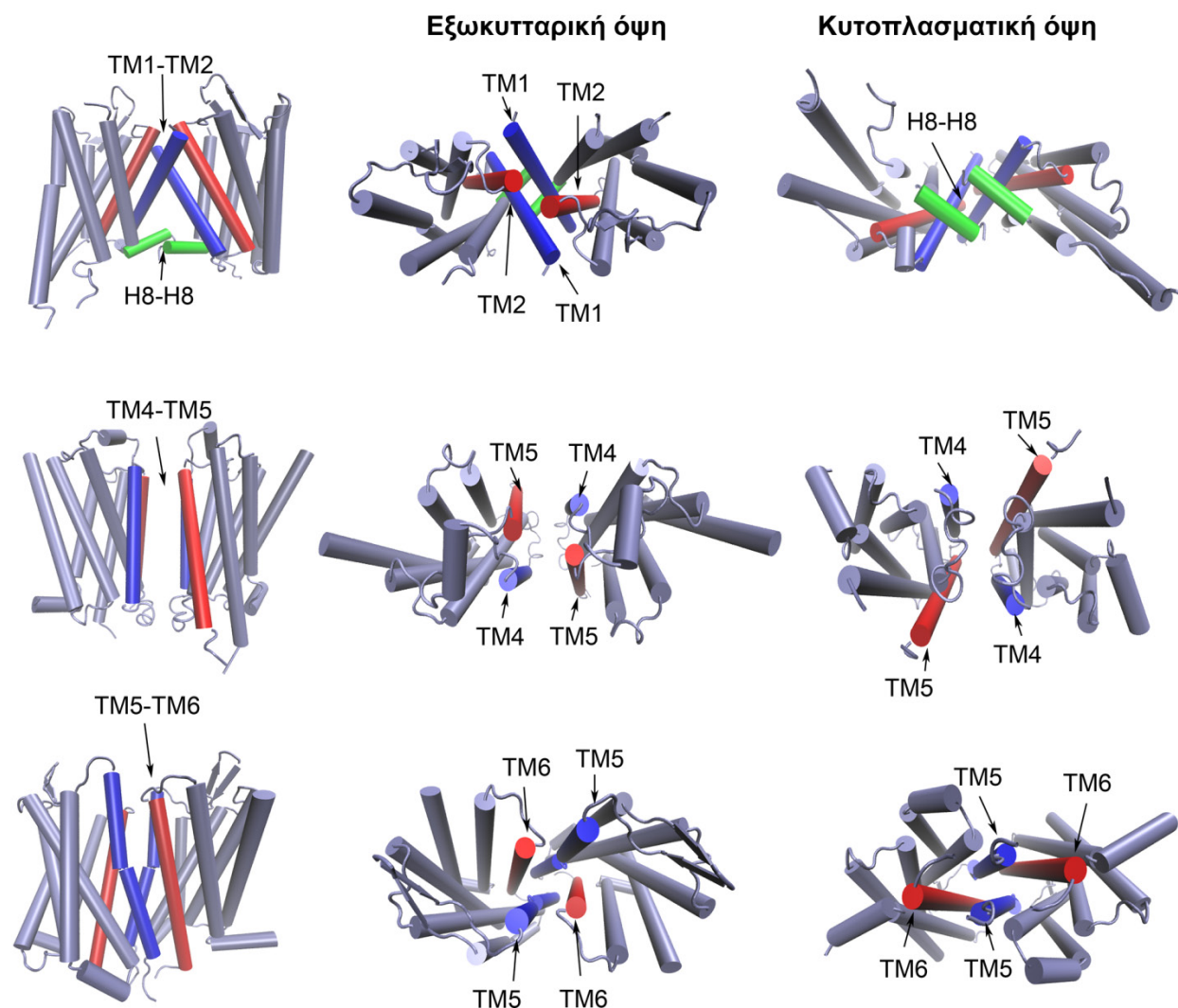
3.1.1 Δομική και δυναμική ανάλυση των αλληλεπιδράσεων ολιγομερισμού των GPCRs

Για τη μελέτη των αλληλεπιδράσεων ολιγομερισμού των GPCRs πραγματοποιήθηκε αρχικά συλλογή όλων των διαθέσιμων δομών ομοδιμερών υποδοχέων από τη βάση δεδομένων PDB. Αυτές περιελάμβαναν σύμπλοκα πρότυπων GPCRs όπως η Ροδοψίνη και ο $\beta 2$ αδρενεργικός υποδοχέας, καθώς επίσης και υποδοχέων με έντονο φαρμακολογικό ενδιαφέρον, όπως οι οπιοειδείς υποδοχείς κOR και μ OR, ο αδενοσινικός υποδοχέας A_{2A} , ο χημειοκινικός υποδοχέας CXCR4, ο υποδοχέας Smoothened και ο μεταβοτροπικός υποδοχέας του γλουταμικού mGluR1. Παράλληλα, κατασκευάστηκαν μια σειρά από θεωρητικά μοντέλα διμερών υποδοχέων, βασισμένα σε προηγούμενα πειραματικά δεδομένα πάνω στο διμερισμό ή/και ολιγομερισμό μιας σειράς από υποδοχείς, συμπεριλαμβανομένων των οπιοειδών υποδοχέων δ OR και μ OR και του υποδοχέα του γλουταμικού mGluR1. Το τελικό σύνολο των διμερών GPCRs, συμπεριλαμβανομένων τόσο των κρυσταλλικών δομών όσο και των θεωρητικών μοντέλων, περιελάμβανε συνολικά 21 διαφορετικά σύμπλοκα (Πίνακας 9). Κάθε διμερές μελετήθηκε με προσομοιώσεις Μοριακής Δυναμικής, σε χρόνους της τάξης των 1000 ns (1 μ s).

Τα αποτελέσματα από τις προσομοιώσεις Μοριακής Δυναμικής των ομοδιμερών έδειξαν ότι οι GPCRs στο σύνολό τους φαίνονται να προτιμούν το σχηματισμό τριών διακριτών τύπων διμερών, με βάση τον τρόπο αλληλεπίδρασης των πρωτομερών στο σύμπλοκο. Συγκεκριμένα, τα διμερή μπορούν να σχηματιστούν από τα διαμεμβρανικά τμήματα TM1, TM2 και την έλικα H8 (διμερή TM1-TM2-H8), τα διαμεμβρανικά τμήματα TM4 και TM5 (διμερή TM4-TM5) ή τα διαμεμβρανικά τμήματα TM5 και TM6 (διμερή TM5-TM6). Χαρακτηριστικά παραδείγματα των τριών τύπων παρουσιάζονται στην Εικόνα 48. Όπως φάνηκε και από τις προσομοιώσεις, διμερή υποδοχέων τα οποία ανήκαν σε μια από τις τρεις κατηγορίες διατηρούν τον γενικό προσανατολισμό των πρωτομερών και τα κύρια δομικά χαρακτηριστικά τους καθ'όλη τη διάρκεια των προσομοιώσεων, υποδεικνύοντας τη σταθερότητα των

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

αλληλεπιδράσεών τους. Αντίθετα, τα διμερή με διαφορετικά στοιχεία αλληλεπίδρασης υπέστησαν πιο εκτεταμένες αλλαγές κατά τη διάρκεια των προσομοιώσεων, ώσπου τελικά υιοθέτησαν τη στοιχειομετρία διμερών TM1-TM2-H8, TM4-TM5 ή TM5-TM6 (Εικόνα 49).



Εικόνα 48 Χαρακτηριστικά παραδείγματα διμερών GPCRs. Απεικονίζονται οι δομές των διμερών TM1-TM2-H8 του υποδοχέα κOR (επάνω), TM4-TM5 του υποδοχέα β1AR (μέση) και TM5-TM6 του υποδοχέα μOR (κάτω). Απεικονίζεται η πλευρική, η εξωκυτταρική και η κυτοπλασματική όψη των υποδοχέων. Οι α-έλικες των διαμεμβρανικών τμημάτων απεικονίζονται σαν κύλινδροι. Τα τμήματα που συμμετέχουν σε αλληλεπιδράσεις διμερισμού σημαίνονται με βέλη και χρωματίζονται μπλε, κόκκινα ή πράσινα, ενώ το υπόλοιπο τμήμα των δομών χρωματίζεται γκριζο.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 9 Σύνοψη των αποτελεσμάτων των προσομοιώσεων Μοριακής Δυναμικής (MD) για τα διμερή GPCRs.

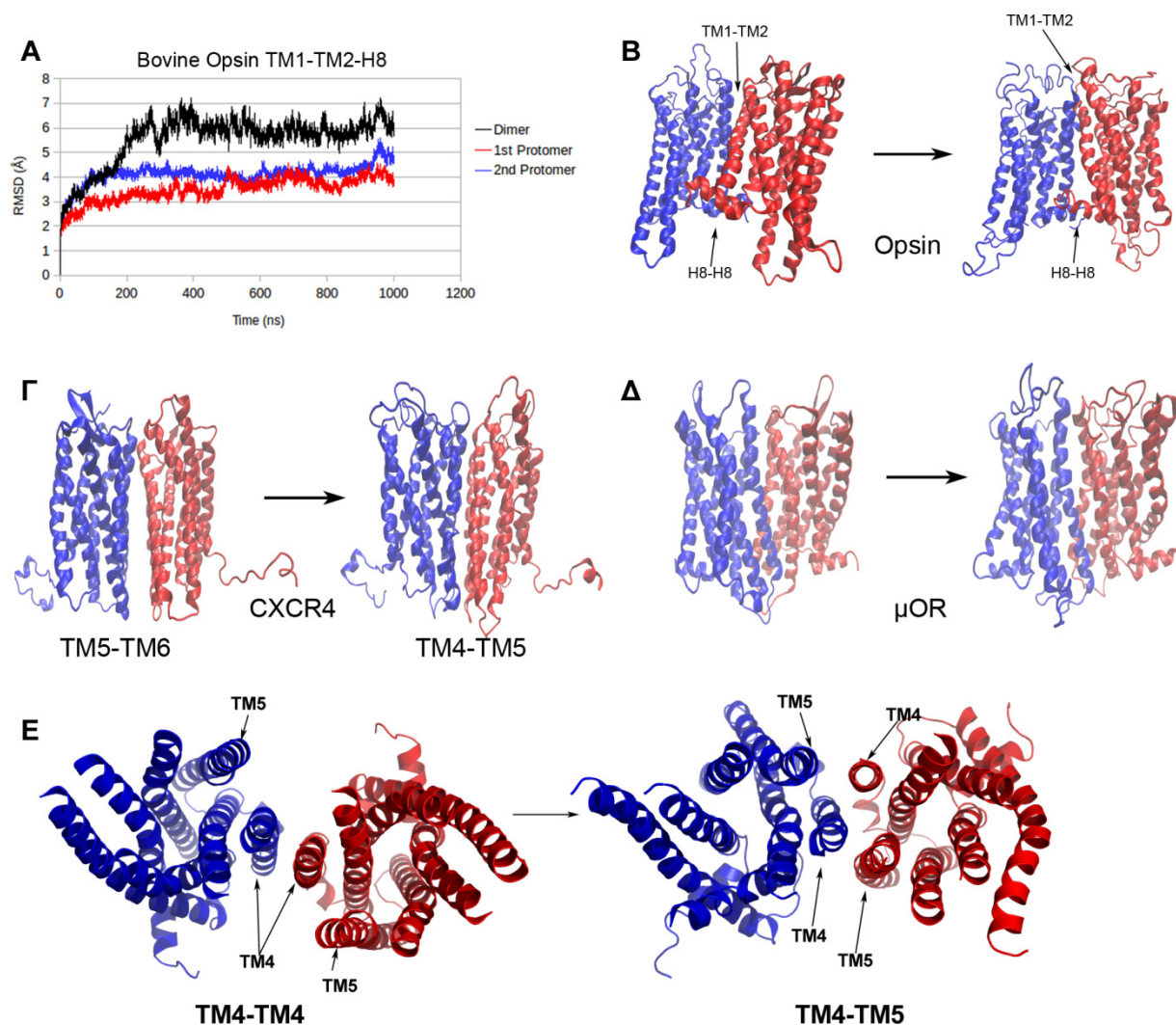
Υποδοχέας	PDB	Αρχικό Διμερές	Προσομοιώσεις MD	Τελικό Διμερές
Ανενεργή Ροδοψίνη (<i>Bos taurus</i>)	2I35	TM1-TM2-H8	3 x 1 μ s	TM1-TM2-H8
Ενεργή Ροδοψίνη (Οψίνη) (<i>Bos taurus</i>)	3CAP	TM1-TM2-H8	3 x 1 μ s	TM1-TM2-H8
Ανενεργή Ροδοψίνη (<i>Bos taurus</i>)	Θεωρητικό Μοντέλο	TM1-TM1	3 x 1 μ s	TM1-TM2-H8
Ανενεργή Ροδοψίνη (<i>Bos taurus</i>)	Θεωρητικό Μοντέλο	TM4-TM5	3 x 1 μ s	TM4-TM5
Οπιοειδής Υποδοχέας κOR (<i>Homo sapiens</i>)	4DJH	TM1-TM2-H8	1 x 10 μ s 3 x 1 μ s	TM1-TM2-H8
Οπιοειδής υποδοχέας μOR (<i>Mus musculus</i>)	4DKL	TM1-TM2-H8	3 x 1 μ s	TM1-TM2-H8
Οπιοειδής υποδοχέας μOR (<i>Mus musculus</i>)	4DKL	TM5-TM6	1 x 10 μ s 3 x 1 μ s	TM5-TM6
Οπιοειδής υποδοχέας μOR (<i>Mus musculus</i>)	Θεωρητικό Μοντέλο	TM4-TM3	1 x 10 μ s 3 x 1 μ s	TM4-TM5
Οπιοειδής υποδοχέας δOR (<i>Homo sapiens</i>)	Θεωρητικό Μοντέλο	TM4-TM3	1 x 10 μ s 3 x 1 μ s	TM4-TM5
Αδρενεργικός υποδοχέας β1AR (<i>Gallus gallus</i>)	4GPO	TM1-TM2-H8	3 x 1 μ s	TM1-TM2-H8
Αδρενεργικός υποδοχέας β1AR (<i>Gallus gallus</i>)	4GPO	TM4-TM5	3 x 1 μ s	TM4-TM5
Ροδοψίνη καλαμαριού (<i>Todarodes pacificus</i>)	2Z73	TM4-TM5	3 x 1 μ s	TM4-TM5
Ροδοψίνη καλαμαριού (<i>Todarodes pacificus</i>)	2Z73	TM5-TM5	3 x 1 μ s	TM5-TM6
Αδενোসινικός υποδοχέας A2A (<i>Homo sapiens</i>)	4EIY	TM4-TM5	3 x 1 μ s	TM4-TM5
Υποδοχέας Smoothened (<i>Homo sapiens</i>)	4JKV	TM4-TM5	3 x 1 μ s	TM4-TM5
Χημειοκινικός υποδοχέας CXCR4 (<i>Homo sapiens</i>)	3ODU	TM5-TM6	1 x 10 μ s 3 x 1 μ s	TM4-TM5
Χημειοκινικός υποδοχέας CXCR4 (<i>Homo sapiens</i>)	3OEO	TM4-TM5-TM6	3 x 1 μ s	TM4-TM5

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Μεταβοτροφικός υποδοχέας του γλουταμικού mGluR1 (<i>Homo sapiens</i>)	4OR2	TM1-TM2	3 x 10 μs με κρυσταλ. μόρια χοληστερόλης 3 x 10 μs χωρίς στερόλες 3 x 10 μs σε μεμβράνη 9:1 POPC: χοληστερόλη	TM1-TM2-H8
Μεταβοτροφικός υποδοχέας του γλουταμικού mGluR1 (<i>Homo sapiens</i>)	Θεωρητικό Μοντέλο	TM4-TM4	3 x 1μs	TM4-TM5
Αδρενεργικός υποδοχέας β2AR (<i>Homo sapiens</i>)	2RH1	TM1-TM1	3 x 1μs	TM1-TM2-H8
Αδρενεργικός υποδοχέας β2AR (<i>Homo sapiens</i>)	2RH1	TM1-H8	3 x 10 μs με κρυσταλ. μόρια χοληστερόλης 3 x 10 μs χωρίς στερόλες 3 x 10 μs σε μεμβράνη 9:1 POPC: χοληστερόλη	TM1-TM2-H8

Χαρακτηριστικά παραδείγματα περιλαμβάνουν τα κρυσταλλογραφικά διμερή TM4-TM5-TM6 του χημειοκινικού υποδοχέα CXCR4 και το κρυσταλλογραφικό διμερές TM5-TM5 της Ροδοψίνης του καλαμαριού, καθώς και τα θεωρητικά διμερή TM3-TM4 των οπιοειδών υποδοχέων δOR και μOR. Στην περίπτωση του υποδοχέα CXCR4, οι δομές των διμερών υπέστησαν αλλαγές στον προσανατολισμό των πρωτομερών, με αποτέλεσμα να χαθούν οι αλληλεπιδράσεις των καταλοίπων στο διαμεμβρανικό τμήμα TM6 και να σχηματιστούν νέες, πιο εκτεταμένες αλληλεπιδράσεων των διαμεμβρανικών τμημάτων TM4 και TM5. Έτσι, τα διμερή του υποδοχέα κατέληξαν να υιοθετήσουν τα χαρακτηριστικά των διμερών TM4-TM5. Αντίστοιχα, στο διμερές TM5-TM5 της Ροδοψίνης του καλαμαριού παρατηρήθηκε περιστροφή του ενός πρωτομερούς σε σχέση με το άλλο στο σύμπλοκο, με αποτέλεσμα το σχηματισμό νέων αλληλεπιδράσεων στα διαμεμβρανικά τμήματα TM6 (διμερές TM5-TM6). Ακόμη πιο εκτεταμένες αλλαγές πραγματοποιήθηκαν στα θεωρητικά μοντέλα των διμερών TM3-TM4, τα οποία υιοθέτησαν χαρακτηριστικά των διμερών TM4-TM5.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 49 Αποτελέσματα των προσομοιώσεων Μοριακής Δυναμικής για τα διμερή των GPCRs. (A) Μετρήσεις RMSD για το διμερές της ενεργής Ροδοψίνης (Οψίνη) του *Bos taurus*. Μετρήσεις RMSD δίνονται για ολόκληρο το διμερές (μαύρο), καθώς και για κάθε πρωτομέρες του (μπλε, κόκκινο). (B) Αρχική (αριστερά) και τελική (δεξιά) στερεοδιάταξη του διμερούς TM1-TM2-H8 της Οψίνης. Οι δομές και ο συνολικός προσανατολισμός των πρωτομερών παραμένουν ανέπαφα, ενώ μοναδική αλλαγή είναι η ελαφρά περιστροφή του ενός πρωτομερούς σε σχέση με το άλλο, οδηγώντας σε πιο ευνοϊκό πακετάρισμα. (Γ) Αρχική και τελική στερεοδιάταξη του διμερούς του υποδοχέα CXCR4. Το αρχικό διμερές TM4-TM5-TM6 υπόκειται σε αλλαγές και μεταβαίνει σε στερεοδιάταξη διμερούς TM4-TM5. (Δ) Αρχική και τελική στερεοδιάταξη του διμερούς TM5-TM6 του υποδοχέα μOR. Η δομή του διμερούς παραμένει ανέπαφη κατά τη διάρκεια των προσομοιώσεων. Οι δομές των B-Δ παρουσιάζονται με την πλευρική όψη τους. (E) Αρχική και τελική στερεοδιάταξη του θεωρητικού διμερούς TM3-TM4 του υποδοχέα δOR. Παρουσιάζεται η εξωκυτταρική όψη των υποδοχέων. Το αρχικό διμερές υπόκειται σε αλλαγές του προσανατολισμού των πρωτομερών του και γίνεται διμερές TM4-TM5.

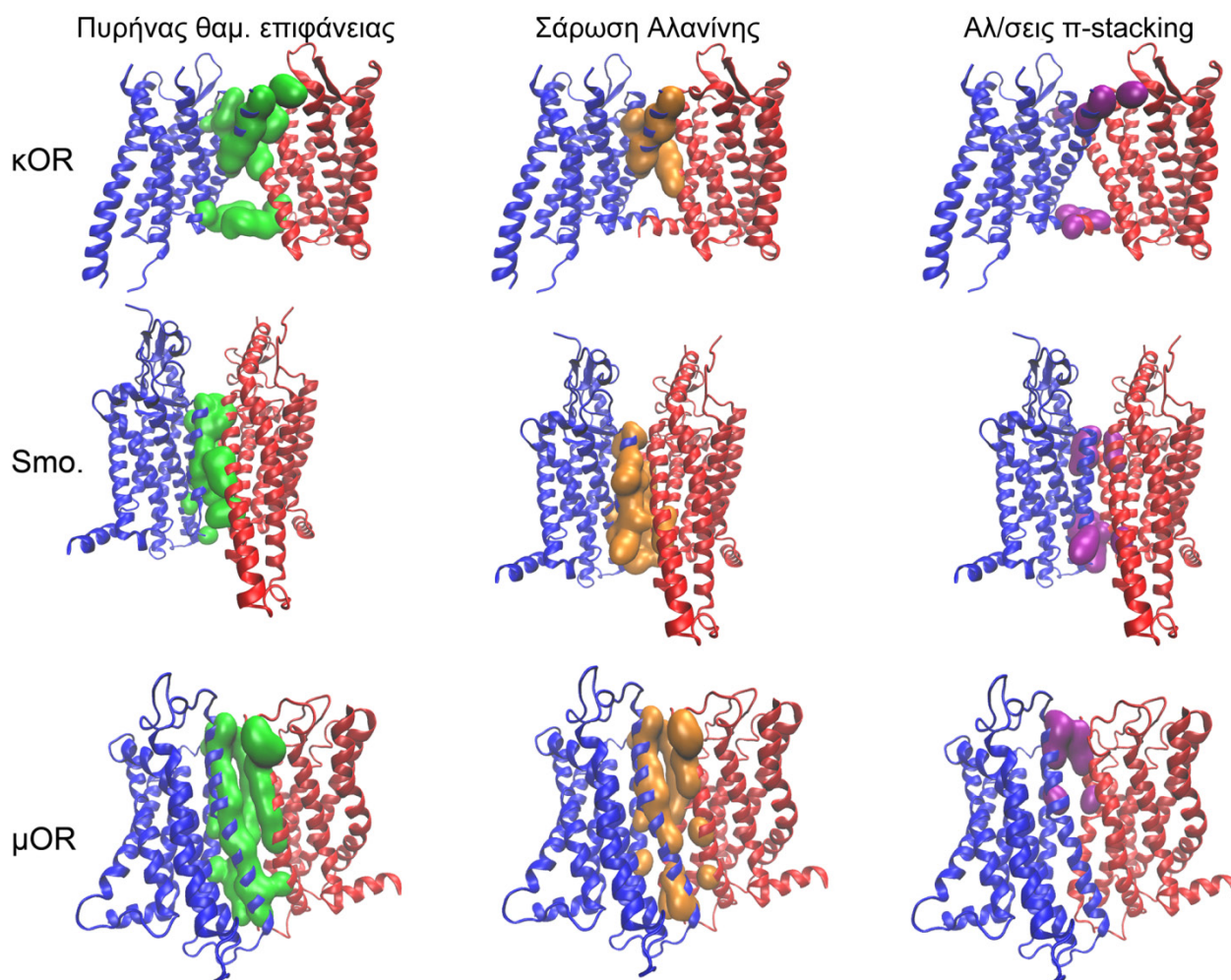
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πρέπει να σημειωθεί πως σε καμία από τις προσομοιώσεις που πραγματοποιήθηκαν δεν παρατηρήθηκαν στερεοδιαταξικές αλλαγές στο δίπλωμα των πρωτομερών, οι οποίες θα μπορούσαν να κατευθύνουν τις αλλαγές στον προσανατολισμό και να επηρεάσουν το αποτέλεσμα. Αντίθετα, όλοι οι υποδοχείς διατήρησαν ανέπαφο το δίπλωμα των GPCRs, με κινήσεις αποκλειστικά στα μη δομημένα τμήματά τους (ενδο- και εξωκυτταρικοί βρόχοι). Έτσι, όλες οι μεταβολές οι οποίες οδήγησαν στο σχηματισμό νέων αλληλεπιδράσεων οφείλονταν αποκλειστικά στην αλλαγή του προσανατολισμού του ενός πρωτομερούς σε σχέση με το άλλο. Τέλος, πρέπει να σημειωθεί ότι σε όλες τις περιπτώσεις διμερών στα οποία συνέβησαν οι μεταβολές που περιγράφηκαν, το σύμπλοκο απέκτησε την τελική στερεοδιάταξή του στα πρώτα 100-200 ns των προσομοιώσεων και παρέμεινε σταθερό ως το τέλος του χρόνου προσομοίωσης (1000 ns).

Οι διαμοριακές αλληλεπιδράσεις κάθε διμερούς μελετήθηκαν περαιτέρω μέσα από μια σειρά μεθόδων για την αξιολόγηση της σημαντικότητας των αμινοξικών καταλοίπων (Εικόνα 50). Χρησιμοποιήθηκε ένα κριτήριο σχετικής προσβάσιμης επιφάνειας (RSA) για την κατηγοριοποίηση των καταλοίπων στον πυρήνα ή την περιφέρεια της θαμμένης επιφάνειας. Παράλληλα, εφαρμόστηκαν μετρήσεις υπολογιστικής σάρωσης αλανίνης, μιας διαδικασίας που περιλαμβάνει αντικατάσταση κάθε καταλοίπου από αλανίνη και εκτίμηση στη μεταβολή στην ενέργεια αλληλεπίδρασης ως αποτέλεσμα αυτής της μεταλλαγής. Τα αποτελέσματα των μεθόδων έδειξαν ότι τα περισσότερα κατάλοιπα του πυρήνα των αλληλεπιδράσεων εντοπίζονται στις έλικες TM1, TM4 και TM5 των διμερών TM1-TM2-H8, TM4-TM5 και TM5-TM6, αντίστοιχα. Σε μικρότερο βαθμό, κατάλοιπα του πυρήνα μπορούν να εντοπιστούν επίσης και στην έλικα TM5, όπως και στην κυτοπλασματική έλικα H8. Αντίστοιχα αποτελέσματα έδωσαν και οι μετρήσεις σάρωσης αλανίνης, με τα περισσότερα σημεία ενδιαφέροντος να εντοπίζονται στις έλικες TM1 και TM2 των διμερών TM1-TM2-H8 και στην έλικα TM5 των διμερών TM4-TM5 ή TM5-TM6. Τα αμινοξικά κατάλοιπα που κατηγοριοποιήθηκαν ως μέλη του πυρήνα και, παράλληλα, ως σημεία ενδιαφέροντος στη σάρωση αλανίνης, περιλαμβάνουν κυρίως αρωματικά κατάλοιπα που συμμετέχουν σε αλληλεπιδράσεις π-ηλεκτρονίων και, σε μικρότερο βαθμό, πολικά ή φορτισμένα κατάλοιπα που σχηματίζουν δεσμούς υδρογόνου και ηλεκτροστατικές αλληλεπιδράσεις, είτε μεταξύ τους είτε με τις πολικές κεφαλές των

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

παρακείμενων φωσφολιπιδίων. Η πλειονότητα αυτών των καταλοίπων εντοπίζονται κυρίως κοντά στα εξωκυτταρικά και κυτοπλασματικά όρια των διαμεμβρανικών τμημάτων TM1, TM4 και TM5 οι οποίες βρίσκονται κοντά στα όρια της μεμβράνης, καθώς και στην έλικα H8.



Εικόνα 50 Απεικόνιση των σημαντικών καταλοίπων αλληλεπίδρασης στα διμερή των GPCRs. Παρουσιάζονται οι δομές των διμερών TM1-TM2-H8 του κOR (επάνω), TM4-TM5 του Smoothened (Smo.) (μέση) και TM5-TM6 του μOR (κάτω), στην πλευρική όψη τους. Τα κατάλοιπα που συγκροτούν τον πυρήνα σύμφωνα με τη μέθοδο Levy, χαρακτηρίζονται σαν σημεία ενδιαφέροντος από τη σάρωση αλανίνης ή σχηματίζουν επαφές π-ηλεκτρονίων χρωματίζονται πράσινα, πορτοκαλί ή μωβ, αντίστοιχα.

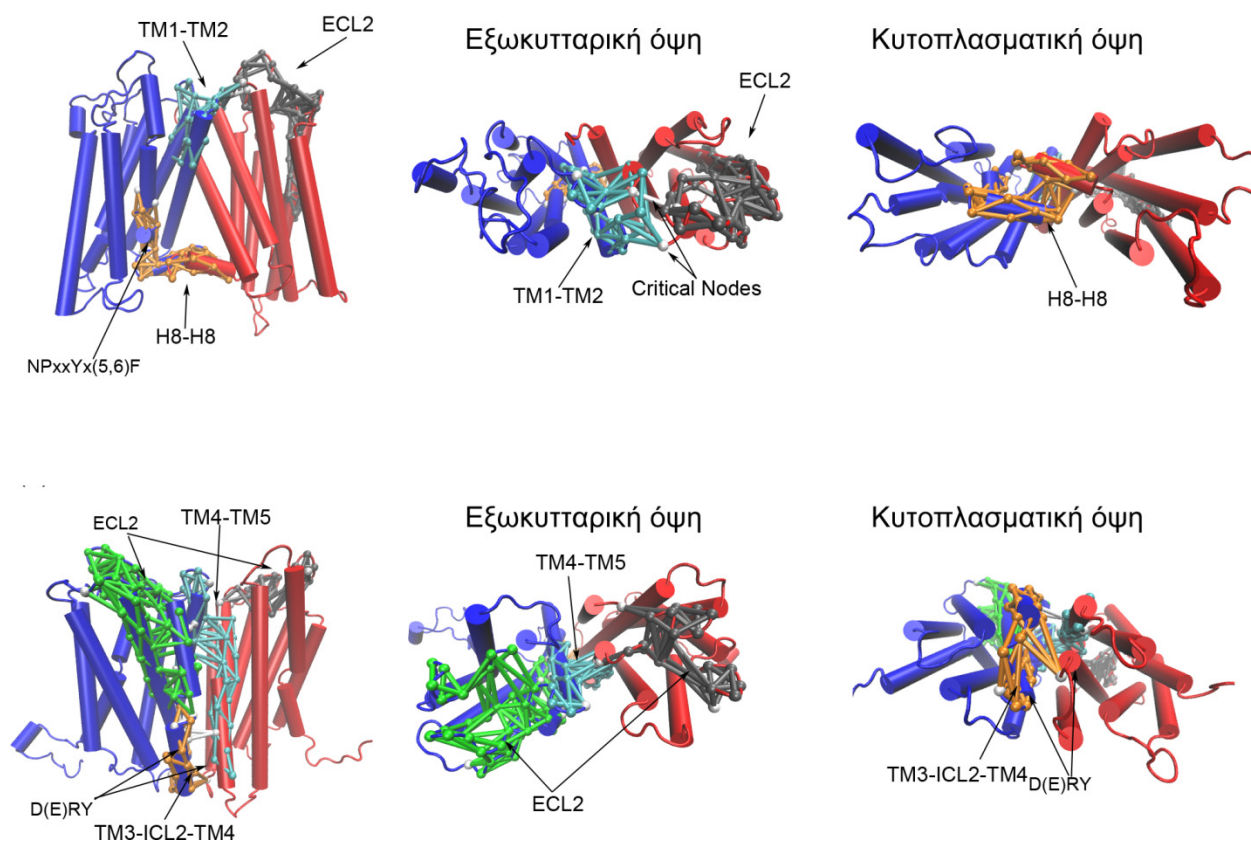
Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν περαιτέρω για την κατασκευή δομικών δικτύων αλληλεπιδράσεων, χρησιμοποιώντας την μεθοδολογία της Ανάλυσης Δυναμικών Δικτύων. Στα δίκτυα που δημιουργήθηκαν κάθε κόμβος

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

αντιπροσωπεύει ένα αμινοξικό κατάλοιπο. Δυο μη γειτονικοί κόμβοι συνδέονται με μια ακμή αν τα αμινοξικά κατάλοιπά τους αλληλεπιδρούν για το μεγαλύτερο μέρος (>75%) της προσομοίωσης που αναλύεται. Όπως φάνηκε από τα αποτελέσματα της ανάλυσης δικτύων, κάθε δομή διμερούς GPCR μπορεί να απεικονιστεί σαν ένα ιδιαίτερα πυκνό δίκτυο αλληλεπιδράσεων. Η πλειονότητα αυτών των αλληλεπιδράσεων, όπως είναι αναμενόμενο, περιλαμβάνει τις ενδομοριακές αλληλεπιδράσεις του κάθε πρωτομερούς. Ωστόσο, σημαντικός αριθμός ακμών σχηματίζονται και ανάμεσα στους κόμβους των διαφορετικών πρωτομερών, οι οποίες αντιπροσωπεύουν τις διαμοριακές αλληλεπιδράσεις που αποτελούν την επιφάνεια αλληλεπίδρασης των διμερών.

Σε κάθε δίκτυο αλληλεπιδράσεων πραγματοποιήθηκε ανάλυση ομαδοποίησης χρησιμοποιώντας τον αλγόριθμο Girvan-Newman για την εύρεση κοινοτήτων. Στην ανάλυση ομαδοποίησης, οι κόμβοι που ανήκουν στην ίδια κοινότητα είναι πιο πυκνά συνδεδεμένοι μεταξύ τους σε σχέση με κόμβους εκτός της κοινότητας και μπορούν να επικοινωνήσουν μεταξύ τους σχετικά εύκολα μέσω πολλαπλών οδών. Καθώς όλα τα χαρακτηριστικά του κάθε δικτύου υπολογίζονται με βάση τα δομικά και δυναμικά χαρακτηριστικά του αντίστοιχου διμερούς, οι κοινότητες των δυναμικών δικτύων των προσομοιώσεων ουσιαστικά αντιπροσωπεύουν τμήματα των υποδοχέων, που κινούνται σε συμφωνία μεταξύ τους. Όπως φάνηκε από την ανάλυση ομαδοποίησης, τα στοιχεία των δομικών δικτύων των διμερών οργανώνονται σε μια σειρά από κοινότητες. Αν και δεν είναι απαραίτητο οι παρατηρούμενες κοινότητες να αντιπροσωπεύουν όντως δομικά χαρακτηριστικά, σε όλες τις περιπτώσεις παρατηρούνται κάποιες βιολογικά σημαντικές συσχετίσεις κινήσεων. Εξίσου σημαντικό επίσης είναι το γεγονός ότι παρατηρείται συντηρημένη εμφάνιση συγκεκριμένων κοινοτήτων σε όλα τα προσομοιωμένα συστήματα, αποκαλύπτοντας πιθανά κοινά χαρακτηριστικά των GPCRs ως προς τη δομή και τη δυναμική συμπεριφορά τους, αλλά και ως προς τις αλληλεπιδράσεις ολιγομερισμού (Εικόνα 51).

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 51 Σημαντικές κοινότητες στα δομικά δίκτυα αλληλεπιδράσεων των διμερών GPCRs. Παρουσιάζονται παραδείγματα από τα διμερή TM1-TM2-H8 (επάνω) και TM4-TM5 (κάτω). Απεικονίζεται η πλευρική, η εξωκυτταρική και η κυτοπλασματική όψη κάθε διμερούς. Στις δομές απεικονίζονται με τη μορφή δικτύων τα τμήματα των υποδοχέων που συμμετέχουν σε οργανωμένες κοινότητες με βιολογικό ρόλο και συσχέτιση με τις αλληλεπιδράσεις διμερισμού. Οι κοινότητες των δικτύων προτυποποιούνται χρησιμοποιώντας σφαίρες για τους κόμβους και σωλήνες για τις ακμές ανάμεσά τους. Οι κρίσιμες συνδέσεις ανάμεσα σε κοινότητες απεικονίζονται με λευκό χρώμα. Οι θέσεις των αλληλεπιδράσεων διμερισμού που αντιστοιχούν σε κοινότητες των δικτύων και τα μοτίβα D(E)RY και NPxxYx(5,6)F σημαίνονται με βέλη.

Συγκεκριμένα, τα δίκτυα από τις προσομοιώσεις των διμερών TM1-TM2-H8 παρουσιάζουν τη δημιουργία δύο κοινοτήτων που περιλαμβάνουν στοιχεία της επιφάνειας αλληλεπίδρασης (Εικόνα 51). Η πρώτη κοινότητα περιλαμβάνει τις διαμοριακές αλληλεπιδράσεις ανάμεσα στα διαμεμβρανικά τμήματα TM1 και TM2 του κάθε πρωτομερούς, καθώς και τμήματα του 2ου εξωκυτταρικού βρόχου (ECL2) των υποδοχέων, ο οποίος έχει βρεθεί σε αρκετούς υποδοχείς να συμμετέχει στην επιλεκτική αναγνώριση προσδετών. Η συνύπαρξη του βρόχου ECL2 με την επιφάνεια αλληλεπίδρασης στην ίδια κοινότητα αποτελεί ένδειξη ότι η κινητικότητα του επηρεάζεται από τις αλληλεπιδράσεις ολιγομερισμού. Η

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

δεύτερη κοινότητα περιλαμβάνει τις διαμοριακές αλληλεπιδράσεις των καρβοξυτελικών ελίκων H8, καθώς και το συντηρημένο μοτίβο NPx(2)Yx(5,6)F από κάθε υποδοχέα, το οποίο έχει συσχετιστεί με τη διαδικασία ενεργοποίησης των GPCRs και, συγκεκριμένα, τη σταθεροποίηση της ενεργής στερεοδιάταξης των υποδοχέων (Park et al., 2008). Αντίστοιχα με τα διμερή TM1-TM2-H8, τα δίκτυα από τις προσομοιώσεις των διμερών TM4-TM5 παρουσιάζουν το σχηματισμό κοινοτήτων οι οποίες περιλαμβάνουν τις διαμοριακές αλληλεπιδράσεις των τμημάτων TM4 και TM5. Επιπρόσθετα, η πλειονότητα των διμερών TM4-TM5 παρουσιάζει το σχηματισμό μίας δεύτερης διαμοριακής κοινότητας, η οποία περιλαμβάνει τις αλληλεπιδράσεις των κυτοπλασματικών όψεων των τμημάτων TM4 και TM5, τον 2ο ενδοκυτταρικό βρόχο (ICL2) και το συντηρημένο μοτίβο D(E)RY, το οποίο έχει επίσης συσχετιστεί με τη διαδικασία ενεργοποίησης των GPCRs και, συγκεκριμένα, συμμετέχει στη σταθεροποίηση της ανενεργής στερεοδιάταξης των υποδοχέων (Park et al., 2008).

Η συντηρημένη παρουσία αλλά και η υψηλή ομοιότητα αυτών των χαρακτηριστικών στην οργάνωση δικτύου των διμερών GPCRs αποκαλύπτουν την πιθανή ύπαρξη συντηρημένων χαρακτηριστικών στη δομική και δυναμική φύση των GPCRs. Μάλιστα, η παρουσία των παραπάνω κοινοτήτων εντοπίζεται όχι μόνο στα διμερή των υποδοχέων Κλάσης A αλλά και στα διμερή υποδοχέων Κλάσης C (mGluR1) και F (Smoothed). Επιπρόσθετα, ο σχηματισμός κοινοτήτων σε περιοχές διαμοριακών επαφών μεταξύ των πρωτομερών δείχνει την πυκνή συνδεσιμότητα αυτών των περιοχών, υποδεικνύοντας τη σημαντική ισχύ των διαμοριακών αλληλεπιδράσεων στις συγκεκριμένες περιοχές. Τέλος, η ταυτόχρονη παρουσία τμημάτων της επιφάνειας αλληλεπίδρασης και τμημάτων των υποδοχέων με λειτουργική σημασία αποτελεί ένδειξη της δομικής συσχέτισής τους, σε συμφωνία με ότι είναι γνωστό πάνω στη λειτουργική σημασία του ολιγομερισμού των GPCRs. Συγκεκριμένα, η δομική και δυναμική συσχέτιση της κίνησης στοιχείων όπως ο βρόχος ECL2 με την κίνηση της επιφάνειας αλληλεπίδρασης στα διμερή TM1-TM2-H8 αποτελεί ένδειξη της αλλοστερικής σύνδεσής τους, γεγονός που είναι σύμφωνο με πειραματικά δεδομένα, τα οποία υποστηρίζουν ότι οι αλληλεπιδράσεις ολιγομερισμού επηρεάζουν την επιλεκτικότητα των υποδοχέων απέναντι στους προσδέτες τους. Αντίστοιχα, η κοινή παρουσία του μοτίβου D(E)RY και των διαμοριακών αλληλεπιδράσεων στην ίδια κοινότητα, στα διμερή TM4-TM5, υποδεικνύει την κοινή

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

κινητικότητά τους. Καθώς έχει προταθεί ότι ο ολιγομερισμός των GPCRs μπορεί να επάγει ή να αναστέλλει την ενεργοποίηση των υποδοχέων, είναι πιθανό οι αλληλεπιδράσεις μεταξύ των πρωτομερών του διμερούς TM4-TM5 στο συγκεκριμένο σημείο να συμμετέχουν στη ρύθμιση της ενεργοποίησης, επηρεάζοντας την κίνηση του μοτίβου D(E)RY. Τέλος, η συσχέτιση της κινητικότητας των διαμοριακών αλληλεπιδράσεων στα τμήματα H8-H8 και το μοτίβο NPxxYx(5,6)F υποδεικνύει ότι η αλληλεπίδραση των πρωτομερών στο συγκεκριμένο σημείο να μπορεί να ρυθμίσει την κινητικότητα του μοτίβου και, μέσω αυτού, την σταθεροποίηση της ενεργής στερεοδιάταξης των υποδοχέων.

Συνολικά, όλα τα αποτελέσματα από τις προσομοιώσεις MD των διμερών GPCRs και την ανάλυσή τους, υποδεικνύουν ότι τα διμερή TM1-TM2-H8, TM4-TM5 και TM5-TM6 αποτελούν τις πιο ευνοϊκές δομές για το σχηματισμό αλληλεπιδράσεων διμερισμού και, ίσως ολιγομερισμού στους GPCRs. Επιπρόσθετα, η παρουσία κοινών γνωρισμάτων στο σύνολο των διμερών υποδοχέων, τόσο για υποδοχείς Κλάσης A όσο και για υποδοχείς Κλάσης C και F υποδεικνύει ότι τα δομικά και δυναμικά χαρακτηριστικά του ολιγομερισμού δεν περιορίζονται αποκλειστικά στους υποδοχείς με ομοιότητα ως προς τη Ροδοψίνη αλλά, πιθανότατα, αποτελούν κοινό χαρακτηριστικό όλων των GPCRs. Τέλος, η ανίχνευση συσχέτισης της δυναμικής συμπεριφοράς ανάμεσα στην επιφάνεια αλληλεπίδρασης των διμερών και σε δομικά στοιχεία των GPCRs με γνωστό λειτουργικό ρόλο αποτελεί ένδειξη του δομικού μηχανισμού με τον οποίο ο διμερισμός και, ίσως, ο ολιγομερισμός των GPCRs, επηρεάζει στοιχεία της λειτουργίας τους (Εικόνα 51).

3.1.2 Ενεργειακή ανάλυση και θερμοδυναμική σταθερότητα των αλληλεπιδράσεων διμερισμού των GPCRs

Τα αποτελέσματα από τις προσομοιώσεις Μοριακής Δυναμικής για τα ομοδιμερή των GPCRs (βλ. Ενότητα 3.1.1) έδειξαν ότι συγκεκριμένα μόνο τμήματα των GPCRs (TM1, TM2, TM4, TM5, TM6 και H8) μπορούν να συμμετάσχουν σε αλληλεπιδράσεις ολιγομερισμού. Παράλληλα, η ανάλυση και αξιολόγηση των διαμοριακών αλληλεπιδράσεων υπέδειξε ότι τα πιο σημαντικά αλληλεπιδρώντα κατάλοιπα εντοπίζονται σε περιοχές κοντά στα άκρα των διαμεμβρανικών

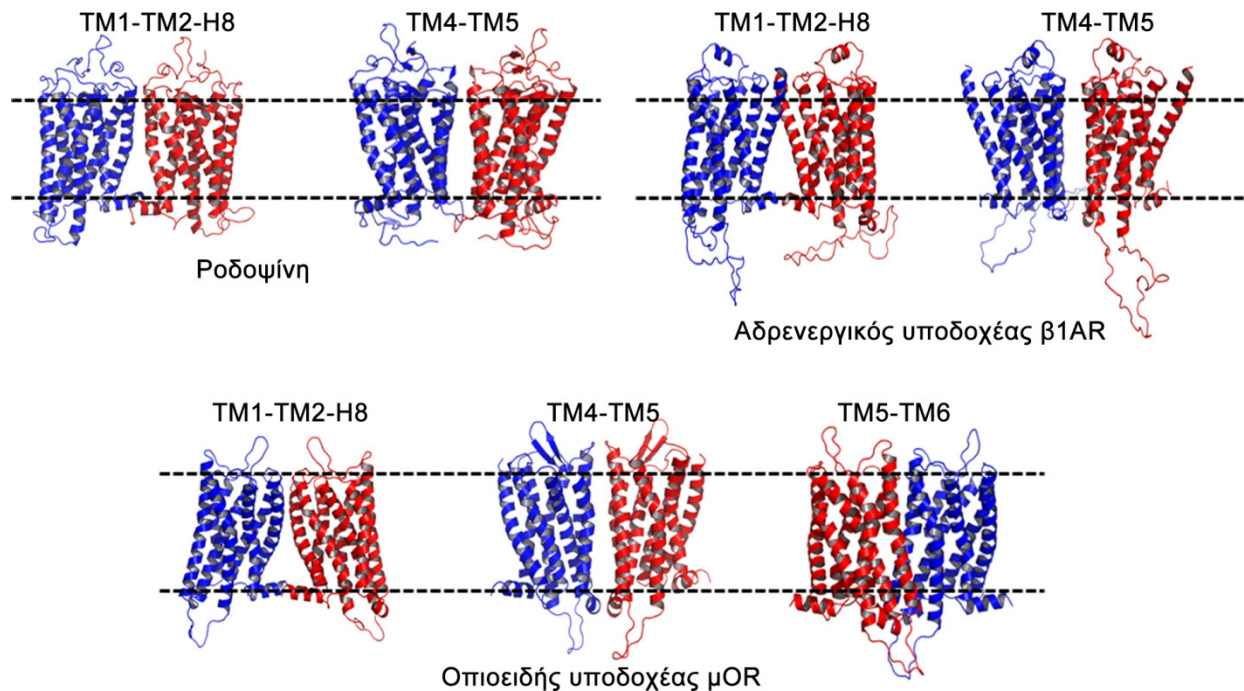
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

τμημάτων (Εικόνα 50), οι οποίες βρίσκονται στα όρια του υδρόφοβου πυρήνα της λιπιδικής διπλοστιβάδας. Η θέση τους κοντά στα όρια της μεμβράνης υποδεικνύει ότι ο σχηματισμός των επαφών διμερισμού θα μπορούσε να αποτελεί τρόπο προσαρμογής των υποδοχέων στο φαινόμενο της υδροφοβικής ασυμφωνίας (hydrophobic mismatch), προστατεύοντας υδρόφοβες ομάδες όπως τα αρωματικά κατάλοιπα από το να βρεθούν σε υδατικό περιβάλλον, αλλά και πολικές ομάδες από το να βρεθούν στο εσωτερικό της υδρόφοβης διπλοστιβάδας. Για τη διερεύνηση αυτής της υπόθεσης πραγματοποιήθηκαν επιπρόσθετες προσομοιώσεις Μοριακής Δυναμικής για επιλεγμένα διμερή (Εικόνα 52), τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για την ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας, χρησιμοποιώντας τη μέθοδο της Υπολειπόμενης Υδροφοβικής Ασυμφωνίας (RHM) (βλ. Ενότητα 2.1.12.4). Συγκεκριμένα, πραγματοποιήθηκαν προσομοιώσεις ισορροπίας για τα διμερή TM1-TM2-H8, TM4-TM5 και TM5-TM6 τριών υποδοχέων (Ροδοψίνη, $\beta 1AR$ και μOR) καθώς και για τα απομονωμένα πρωτομερή τους. Τα αποτελέσματα των προσομοιώσεων χρησιμοποιήθηκαν για τον υπολογισμό της ενεργειακής ποινής του RHM (ΔG_{RHM}) καθενός από τα 7 διαμεμβρανικά τμήματα του κάθε υποδοχέα, στην κατάσταση μονομερούς, καθώς και όταν συμμετέχει σε κάποιο διμερές.

Τα αποτελέσματα των μετρήσεων παρουσιάζονται στην Εικόνα 53α. Οι μετρήσεις του ΔG_{RHM} στα μονομερή έδειξαν οι πιο σημαντικές ενεργειακές ποινές του RHM εντοπίζονται στα διαμεμβρανικά τμήματα TM1, TM4 και TM5 όλων των υποδοχέων, καθώς επίσης και στην έλικα TM2, στην περίπτωση της Ροδοψίνης. Οι υψηλές τιμές του ΔG_{RHM} υποδεικνύουν ότι σε συνθήκες μεμβρανικού περιβάλλοντος, τα κατάλοιπα στα άκρα αυτών των διαμεμβρανικών τμημάτων εντοπίζονται σε μη ευνοϊκό για αυτά περιβάλλον (π.χ. υδρόφοβα κατάλοιπα εκτίθενται στον υδατικό διαλύτη ή πολικά κατάλοιπα θάβονται στην υδρόφοβη διπλοστιβάδα), με αποτέλεσμα οι αλληλεπιδράσεις τους με το περιβάλλον να είναι θερμοδυναμικά μη-ευνοϊκές. Αντίθετα, τα υπόλοιπα διαμεμβρανικά τμήματα παρουσιάζουν μικρότερες τιμές ΔG_{RHM} , οι οποίες υποδεικνύουν ότι η συμμετοχή τους στη συνολική ποινή του RHM είναι λιγότερο σημαντική. Ο υπολογισμός του ΔG_{RHM} για τα διμερή των υποδοχέων και η σύγκρισή τους με τα μονομερή αποκάλυψε ότι, κατά το σχηματισμό διμερών, οι ποινές RHM των τμημάτων που συμμετέχουν στις αλληλεπιδράσεις διμερισμού μειώνονται σημαντικά. Για

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

παράδειγμα, στο διμερές TM1-TM2-H8 της Ροδοψίνης το ΔG_{RHM} μειώνεται από τα 13.9 kT στα 4.4 kT για το TM1, ενώ μηδενίζεται για το TM2. Αντίστοιχα, στο διμερές TM4-TM5 του υποδοχέα παρατηρείται σημαντική μείωση του ΔG_{RHM} για το TM4, ενώ παρόμοια συμπεριφορά παρατηρείται και στους υποδοχείς $\beta 1AR$ και μOR .



Εικόνα 52 Δομές διμερών των GPCRs που μελετήθηκαν με ενεργειακούς υπολογισμούς. Παρουσιάζονται τα διμερή της Ροδοψίνης, του αδρενεργικού υποδοχέα $\beta 1AR$ και του οπιοειδή υποδοχέα μOR μετά από 1 μs προσομοιώσεων Μοριακής Δυναμικής. Αυτά τα διμερή χρησιμοποιήθηκαν για την κατασκευή συστημάτων προσομοίωσης στη μέτρηση της υδροφοβικής ασυμφωνίας, καθώς και στις προσομοιώσεις για τον υπολογισμό του Δυναμικού Μέσης Δύναμης. Τα πρωτομερή του κάθε διμερούς χρωματίζονται μπλε και κόκκινο, ενώ τα όρια της λιπιδικής διπλοστιβάδας απεικονίζονται με στικτές γραμμές.

Από τις μετρήσεις του ΔG_{RHM} φάνηκε ότι οι συνεισφορές των τμημάτων TM1, TM2, TM4 και TM5 στο RHM δεν είναι πάντα ισόποσες. Για παράδειγμα, στη Ροδοψίνη, οι τιμές ΔG_{RHM} των TM1 και TM2 είναι μεγαλύτερες από αυτές των TM4 και TM5. Αντίστοιχα με τα μονομερή, η μείωση του RHM στα διμερή δεν είναι ισόποση, με το διμερές TM1-TM2-H8 να παρουσιάζει σημαντικότερες πτώσεις του ΔG_{RHM} σε σχέση με το διμερές TM4-TM5. Παρόμοια αποτελέσματα καταγράφονται και για τον $\beta 1AR$ στα τα τμήματα TM1, TM4 και TM5, ενώ η

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

αντίστροφη συμπεριφορά παρατηρείται στον μOR , στον οποίο παρατηρούνται υψηλότερες τιμές για τα τμήματα TM4 και TM5 παρά για τα τμήματα TM1 και TM2. Αυτή η διαφοροποίηση πιθανώς αποτελεί ένδειξη ότι τα διαφορετικά διμερή των υποδοχέων ίσως διαφοροποιούνται και ως προς τη θερμοδυναμική σταθερότητά τους. Προκειμένου να ελεγχθεί η παραπάνω υπόθεση, πραγματοποιήθηκαν υπολογισμοί του Δυναμικού Μέσης Δύναμης (PMF) για κάθε διμερές, χρησιμοποιώντας τη μέθοδο της Δειγματοληψίας Ομπρέλας (US). Οι μετρήσεις των τιμών PMF έγιναν θέτοντας σαν συντεταγμένη αναφοράς την απόσταση ανάμεσα στα κέντρα μάζας των πρωτομερών (r) κατά τον σταδιακό αποχωρισμό τους, ενώ οι αρχικές συντεταγμένες για τις προσομοιώσεις US δημιουργήθηκαν μέσω προσομοιώσεων Καθοδηγούμενης Μοριακής Δυναμικής (SMD).

Πίνακας 10 Αποτελέσματα προσομοιώσεων US για τα διμερή των $\beta 1\text{AR}$, μOR και Ροδοψίνης.

Υποδοχέας	Διμερές	Εύρος ξ ($\text{\AA}$)	$\xi_{\min}$ ($\text{\AA}$) ¹	$\Delta F_{\min}$ (kcal/mol) ²
Ροδοψίνη	TM1-TM2-H8	35.06 – 74.32	37.42	-19.98 $\pm$ 0.47
	TM4-TM5	32.52 – 72.62	35.49	-13.95 $\pm$ 0.46
Αδρενεργικός υποδοχέας $\beta 1\text{AR}$	TM1-TM2-H8	42.93 – 73.99	43.56	-13.82 $\pm$ 0.46
	TM4-TM5	31.48 – 73.94	32.25	-9.43 $\pm$ 0.49
Οπιοειδής υποδοχέας μOR	TM1-TM2-H8	36.11 – 74.30	38.22	-9.55 $\pm$ 0.15
	TM4-TM5	29.03 – 73.53	30.37	-24.88 $\pm$ 0.18
	TM5-TM6	28.84 – 74.01	30.88	-17.47 $\pm$ 0.10

¹ Η τιμή της συντεταγμένης ξ στην οποία εντοπίζεται το ενεργειακό ελάχιστο $\Delta F_{\min}$.

² Η τιμή του ενεργειακού ελάχιστου ($\Delta F_{\min}$) του PMF (μέση τιμή $\pm$ σφάλμα).

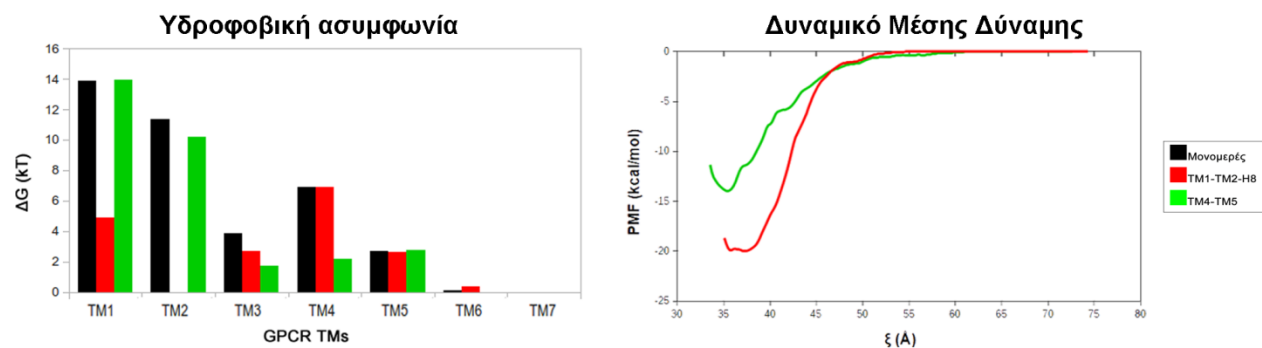
Τα αποτελέσματα από τον υπολογισμό των τιμών PMF παρουσιάζονται στην Εικόνα 53β και στον Πίνακα 10. Οι μετρήσεις του PMF επιβεβαίωσαν ότι τα διαφορετικά διμερή κάθε υποδοχέα παρουσιάζουν διαφορετική θερμοδυναμική συμπεριφορά. Κάθε καμπύλη PMF παρουσιάζει ένα ολικό ενεργειακό ελάχιστο ($\Delta F_{\min}$), το οποίο αντιστοιχεί στην απόσταση ανάμεσα στα πρωτομερή στην οποία το διμερές έχει τη θερμοδυναμικά βέλτιστη δομή του ($\xi_{\min}$). Οι τιμές των $\Delta F_{\min}$ μπορούν, ουσιαστικά, να ταυτιστούν με την ελεύθερη ενέργεια αλληλεπίδρασης του κάθε διμερούς. Στο σύνολό τους, οι μετρήσεις του PMF επιβεβαιώνουν την υπόθεση ότι η θερμοδυναμική συμπεριφορά των διμερών διαφέρει ανάλογα με τον τύπο

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

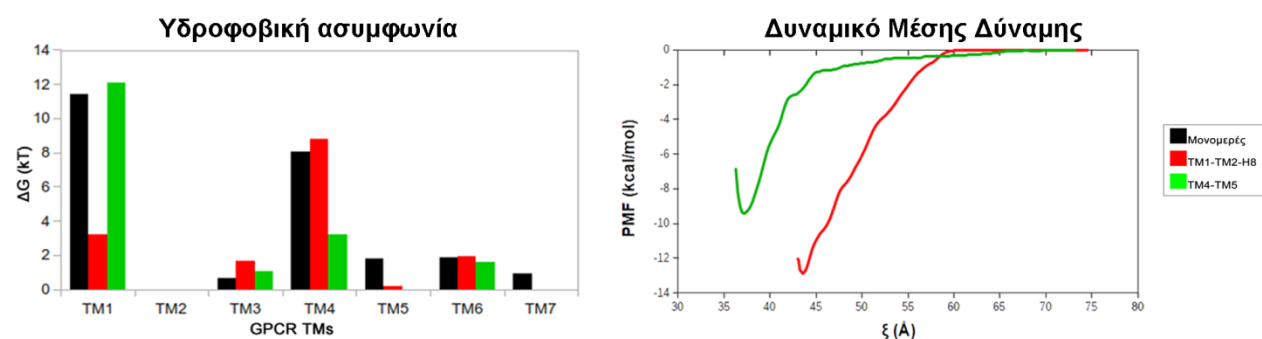
τους σε κάθε υποδοχέα. Έτσι, στη Ροδοψίνη το διμερές TM1-TM2-H8 φαίνεται να είναι θερμοδυναμικά πιο ευνοϊκό (-19.98 ± 0.47 kcal/mol) από το διμερές TM4-TM5 (-13.95 ± 0.46 kcal/mol). Αντίστοιχα αποτελέσματα παρατηρούνται και για τον υποδοχέα β1AR. Αντίθετα, στον υποδοχέα μOR το πιο ευνοϊκό θερμοδυναμικά σύμπλοκο φαίνεται να είναι το διμερές TM4-TM5 (-24.88 ± 0.18 kcal/mol), ακολουθούμενο από το διμερές TM5-TM6 (-17.47 ± 0.10 kcal/mol), ενώ το διμερές TM1-TM2-H8 παρουσιάζει σημαντικά πιο ασθενή ενέργεια αλληλεπίδρασης (-9.55 ± 0.15 kcal/mol). Η πιο σημαντική παρατήρηση, ωστόσο είναι το ότι τα παραπάνω αποτελέσματα συμφωνούν με τις παρατηρήσεις από τις μετρήσεις RHM των υποδοχέων, με τα διαμεμβρανικά τμήματα τα οποία παρουσίαζαν τις υψηλότερες τιμές ΔG_{RHM} (και, αντίστοιχα, τη μεγαλύτερη μείωσή τους κατά το διμερισμό) να συμμετέχουν στα θερμοδυναμικά πιο ευνοϊκά διμερή.

Συνολικά, τα αποτελέσματα από τις μετρήσεις του RHM και τους υπολογισμούς του PMF είναι σύμφωνα με την υπόθεση της υδροφοβικής ασυμφωνίας και υποδεικνύουν ότι ο σχηματισμός των αλληλεπιδράσεων διμερισμού των GPCRs με τα τμήματα TM1, TM2 (για τη Ροδοψίνη), TM4 και TM5 είναι ευνοϊκός για τη βελτιστοποίηση των αλληλεπιδράσεών τους με το μεμβρανικό περιβάλλον. Η ταύτιση των μετρήσεων του RHM με την ανάλυση αλληλεπιδράσεων (Ενότητα 3.1.1) αποδεικνύει περαιτέρω ότι τα συγκεκριμένα τμήματα των GPCRs αποτελούν τις προτιμώμενες θέσεις αλληλεπίδρασης. Τέλος, οι υπολογισμοί του PMF δείχνουν ότι ο βαθμός της μείωσης της υδροφοβικής ασυμφωνίας κατά το σχηματισμό των διμερών αποτελεί ένδειξη της θερμοδυναμικής σταθερότητά τους. Συνολικά, όλα τα παραπάνω αποτελέσματα δείχνουν ότι η σταθερότητα των αλληλεπιδράσεων στα διμερή των GPCRs σχετίζεται σε σημαντικό βαθμό με τις αλληλεπιδράσεις τους με τη λιπιδική διπλοστιβάδα, και αποτελούν περαιτέρω ένδειξη ότι οι ιδιότητες της μεμβράνης μπορούν να επηρεάσουν τις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών.

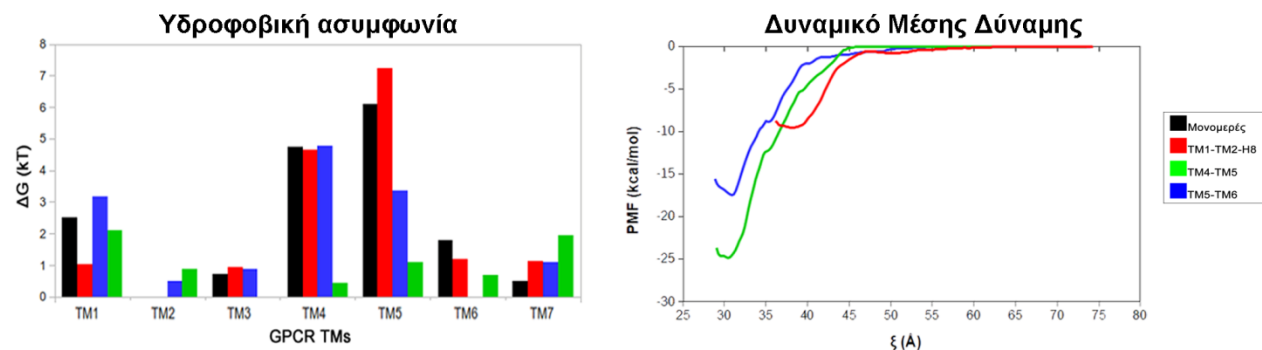
Ροδοψίνη



β1AR



μOR



Εικόνα 53 Θερμοδυναμική Ανάλυση των αλληλεπιδράσεων διμερισμού των GPCRs. (Α) Μετρήσεις των τιμών ΔG_{RHM} για τα διαμεμβρανικά τμήματα της Ροδοψίνης (αριστερά), του β1AR (μέση) και του μOR (δεξιά). Ο οριζόντιος άξονας κάθε διαγράμματος απεικονίζει τα 7 διαμεμβρανικά τμήματα των GPCRs, ενώ ο κατακόρυφος τις τιμές του ΔG (σε μονάδες kT). Οι μετρήσεις των μονομερών απεικονίζονται με μαύρες μπάρες, ενώ οι μετρήσεις των διμερών TM1-TM2-H8, TM4-TM5 και TM5-TM6 με κόκκινο, πράσινο και μπλε χρώμα, αντίστοιχα. (Β). Καμπύλες του Δυναμικού Μέσης Δύναμης (PMF) για τα διμερή των τριών υποδοχέων. Ο οριζόντιος άξονας κάθε διαγράμματος απεικονίζει την συντεταγμένη της αντίδρασης ξ (απόσταση ανάμεσα στα κέντρα μάζας των πρωτομερών) σε Å, ενώ ο κατακόρυφος άξονας απεικονίζει τις τιμές του PMF σε kcal/mol. Οι καμπύλες PMF των διμερών χρησιμοποιούν τα χρώματα των διαγραμμάτων στο (Α).

3.1.3 Η επίδραση της χοληστερόλης στον αυθόρμητο σχηματισμό διμερών GPCRs

Στα πλαίσια της παρούσας διδακτορικής διατριβής διερευνήθηκε η επίδραση της παρουσίας χοληστερόλης στη λιπιδική διπλοστιβάδα σε διαφορετικές συγκεντρώσεις, πάνω στην ικανότητα δύο GPCRs στο σχηματισμό ομο- και ετεροδιμερών. Οι υποδοχείς που επιλέχθηκαν για αυτή τη μελέτη ήταν ο αδενোসινικός υποδοχέας A2A και ο μεταβοτροπικός υποδοχέας του γλουταμικού mGluR5. Και οι δύο υποδοχείς έχουν βρεθεί πειραματικά να σχηματίζουν ομο- και ετεροδιμερή, καθώς και ανώτερης τάξης ολιγομερή, με πιθανό ρόλο σε ασθένειες όπως η Parkinson's και το σύνδρομο του εύθραυστου χρωμοσώματος X (Fragile X Syndrome). Επιπρόσθετα, η ορθή λειτουργία και των δύο υποδοχέων έχει βρεθεί να σχετίζεται με τις αλληλεπιδράσεις τους με τη χοληστερόλη. Ωστόσο, ο μηχανισμός με τον οποίο οι υποδοχείς σχηματίζουν τα διμερή τους παραμένει άγνωστος.

Για την μελέτη αυτού του μηχανισμού, σχεδιάστηκαν και πραγματοποιήθηκαν προσομοιώσεις Μοριακής Δυναμικής σε μεγάλη χρονική κλίμακα, στις οποίες δύο απομακρυσμένα πρωτομερή τοποθετήθηκαν σε τυχαίες θέσεις σε μια λιπιδική διπλοστιβάδα και αφέθηκαν να κινηθούν ελεύθερα στο χώρο και να σχηματίσουν αυθόρμητα αλληλεπιδράσεις μεταξύ τους. Πραγματοποιήθηκαν προσομοιώσεις τόσο για τον ομοδιμερισμό των υποδοχέων όσο και για τον ετεροδιμερισμό τους. Για τη διερεύνηση του ρόλου της χοληστερόλης, οι προσομοιώσεις πραγματοποιήθηκαν σε λιπιδικές διπλοστιβάδες POPC, με διαφορετική περιεκτικότητα χοληστερόλης (0%, 10% και 30%). Όλες οι προσομοιώσεις πραγματοποιήθηκαν σε χρόνους της τάξης των 20-35 μ s σε πολλαπλά αντίγραφα, με το συνολικό χρόνο προσομοιώσεων να ξεπερνά τα 500 μ s.

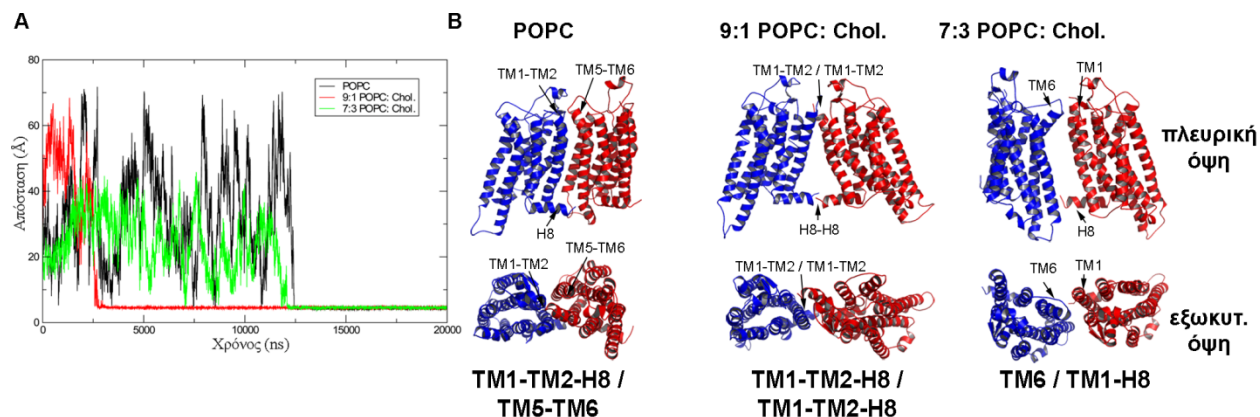
Τα αποτελέσματα από την ανάλυση των προσομοιώσεων έδειξαν ότι τα δύο πρωτομερή του κάθε υποδοχέα (είτε του A2A είτε του mGluR5) διαχέονται ελεύθερα στη μεμβράνη και σχηματίζουν διαμοριακές αλληλεπιδράσεις μεταξύ τους. Πρέπει να σημειωθεί πως τα πρώτα διμερή που παρατηρήθηκαν να σχηματίζονται δεν ήταν τα τελικά σύμπλοκα που κάθε συστήματος. Αντίθετα, σε κάθε προσομοίωση παρατηρήθηκε επανειλημμένα ο αυθόρμητος σχηματισμός παροδικών συμπλόκων ακολουθούμενος από τη διάσπασή τους,

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

πριν τον σχηματισμό ενός σταθερού συμπλόκου που παρέμεινε ανέπαφο ως το τέλος της προσομοίωσης, και το οποίο θεωρήθηκε ως η τελική δομή του συστήματος. Η παραπάνω συμπεριφορά παρατηρήθηκε στο σύνολο σχεδόν των συστημάτων προσομοίωσης, με μοναδική εξαίρεση το σύστημα ομοδιμερισμού του mGluR5 σε μεμβράνες με περιεκτικότητα χοληστερόλης 30%, στις οποίες σχηματίστηκαν μόνο παροδικά σύμπλοκα και δεν προέκυψε κάποια σταθερή τελική δομή.

Ομοδιμερή A2A-A2A: Όλες οι προσομοιώσεις ομοδιμερισμού του αδενοσινικού υποδοχέα A2A έδωσαν σαν τελικό αποτέλεσμα σταθερά σύμπλοκα ομοδιμερών του υποδοχέα, τόσο στις πλήρως φωσφολιπιδικές μεμβράνες POPC, όσο και στις μεμβράνες με παρουσία χοληστερόλης (Εικόνα 54). Στις μεμβράνες POPC, το τελικό σύμπλοκο που προέκυψε από τα συστήματα προσομοίωσης ήταν ένα πυκνά συνδεδεμένο (μεγάλος αριθμός αλληλεπιδράσεων) μη συμμετρικό διμερές, αποτελούμενο από τα τμήματα TM1, TM2 και H8 του ενός πρωτομερούς και τα τμήματα TM5 και TM6 του δεύτερου (διμερές TM1-TM2-H8/TM5-TM6). Ο σχηματισμός του τελικού διμερούς πραγματοποιήθηκε περίπου στα 12-13 μς των προσομοιώσεων, ενώ παρατηρήθηκε μεγάλος αριθμός παροδικών συμπλόκων, υποδεικνύοντας ότι οι υποδοχείς χρειάστηκε να δοκιμάσουν πολλαπλές παραλλαγές του συμπλόκου πριν υιοθετήσουν τον τελικό προσανατολισμό τους. Πρέπει να σημειωθεί πως παρόμοια, μη συμμετρικά διμερή, έχουν παρατηρηθεί με πειράματα κρυο-ηλεκτρονικής μικροσκοπίας για τη Ροδοψίνη, καθώς και σε προηγούμενες μελέτες του ολιγομερισμού των GPCRs με προσομοιώσεις Μοριακής Δυναμικής. Ωστόσο, η ενεργειακή σταθερότητα αυτών των διμερών έχει βρεθεί να είναι κατά πολύ ασθενέστερη από αυτή των συμμετρικών διμερών TM1-TM2-H8, TM4-TM5 και TM5-TM6, με αποτέλεσμα τα μη συμμετρικά διμερή να μην θεωρούνται βιολογικά σημαντικά.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

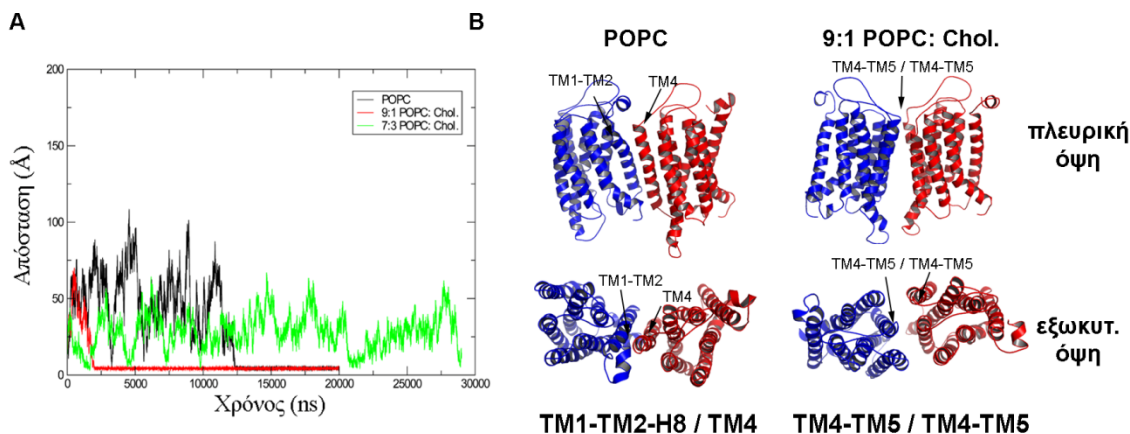


Εικόνα 54 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ομοδιμερισμό του υποδοχέα A2A. (Α) Διάγραμμα της απόστασης ανάμεσα στα πρωτομερή σε κάθε σύστημα μεμβράνης. Ο οριζόντιος άξονας απεικονίζει το χρόνο (σε ns) ενώ ο κατακόρυφος άξονας απεικονίζει την απόσταση (σε Å). Οι μετρήσεις των τριών μεμβρανικών συστημάτων απεικονίζονται με μαύρο, κόκκινο και πράσινο χρώμα, αντίστοιχα. (Β) Χαρακτηριστικές δομές των τελικών διμερών που προέκυψαν από τις προσομοιώσεις.

Σε αντίθεση με τις προσομοιώσεις των μεμβρανών POPC, η εισαγωγή της χοληστερόλης στη μεμβράνη οδήγησε στο σχηματισμό διαφορετικών διμερών για τον A2A. Συγκεκριμένα, οι προσομοιώσεις που πραγματοποιήθηκαν σε μεμβράνες με περιεκτικότητα χοληστερόλης 10% οδήγησαν στο σχηματισμό ενός συμμετρικού διμερούς TM1-TM2-H8 για τους υποδοχείς. Το διμερές παρουσιάζει μεγάλη ομοιότητα με τα αντίστοιχα διμερή TM1-TM2-H8 που αναλύθηκαν στην Ενότητα 3.1.1 και φαίνεται να μοιράζεται τα δομικά και δυναμικά χαρακτηριστικά τους. Αξίζει να σημειωθεί ότι, αντίθετα από τις προσομοιώσεις σε μεμβράνες POPC, οι προσομοιώσεις παρουσία χοληστερόλης οδήγησαν στο σχηματισμό του τελικού διμερούς σε πολύ μικρότερο χρόνο (~2.5 μ s), ενώ παρατηρήθηκε και πολύ μικρότερος αριθμός παροδικών συμπλόκων. Αντίθετα, φαίνεται ότι η παρουσία της χοληστερόλης μπορεί ουσιαστικά να επιταχύνει την πλευρική διάχυση των πρωτομερών και να «κατευθύνει» τον προσανατολισμό τους, ωθώντας τα να σχηματίσουν το τελικό σύμπλοκο. Σε αντίθεση με τα παραπάνω αποτελέσματα, οι προσομοιώσεις που πραγματοποιήθηκαν σε μεμβράνες με περιεκτικότητα χοληστερόλης 30% οδήγησαν στο σχηματισμό ενός μη συμμετρικού διμερούς, αποτελούμενο από τα τμήματα TM1 και H8 του ενός πρωτομερούς και το τμήμα TM6 του δεύτερου. Ωστόσο, σε σχέση με το διμερές από τη μεμβράνη POPC, το συγκεκριμένο

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

σύμπλοκο είναι χαλαρά συνδεδεμένο, παρουσιάζοντας πολύ λιγότερες διαμοριακές αλληλεπιδράσεις.



Εικόνα 55 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ομοδιμερισμό του υποδοχέα mGluR5. (Α) Διάγραμμα της απόστασης ανάμεσα στα πρωτομερή σε κάθε σύστημα μεμβράνης. Ο οριζόντιος άξονας απεικονίζει το χρόνο (σε ns) ενώ ο κατακόρυφος άξονας απεικονίζει την απόσταση (σε Å). Οι μετρήσεις των τριών μεμβρανικών συστημάτων απεικονίζονται με μαύρο, κόκκινο και πράσινο χρώμα, αντίστοιχα. (Β) Χαρακτηριστικές δομές των τελικών διμερών που προέκυψαν από τις προσομοιώσεις.

Ομοδιμερή mGluR5-mGluR5: Οι προσομοιώσεις ομοδιμερισμού του υποδοχέα του γλουταμικού mGluR5 οδήγησαν στο σχηματισμό σταθερών διμερών σε μεμβράνες POPC, καθώς και σε μεμβράνες με 10% περιεκτικότητα χοληστερόλης (Εικόνα 55). Συγκεκριμένα, στις μεμβράνες POPC παρατηρήθηκε ο σχηματισμός ενός σταθερού, μη συμμετρικού διμερούς TM1-TM2-H8/TM4, παρόμοια με το μη συμμετρικό διμερές του A2A στην ίδια μεμβράνη. Ο σχηματισμός του τελικού συμπλόκου πραγματοποιήθηκε περίπου στα 13 μ s, με μεγάλο αριθμό παροδικών συμπλόκων να προηγούνται. Αντίθετα, οι μεμβράνες με περιεκτικότητα χοληστερόλης 10% οδήγησαν στο σχηματισμό ενός συμμετρικού διμερούς TM4-TM5, πολύ παρόμοιου με τα διμερή που έχουν παρατηρηθεί στις κρυσταλλικές δομές και σε πειραματικές μελέτες διάφορων GPCRs. Όπως συνέβη και στην περίπτωση του A2A, ο σχηματισμός του διμερούς πραγματοποιήθηκε σε πολύ μικρό χρόνο (~ 2 μ s) ενώ παρατηρήθηκαν πολύ λίγα παροδικά σύμπλοκα.

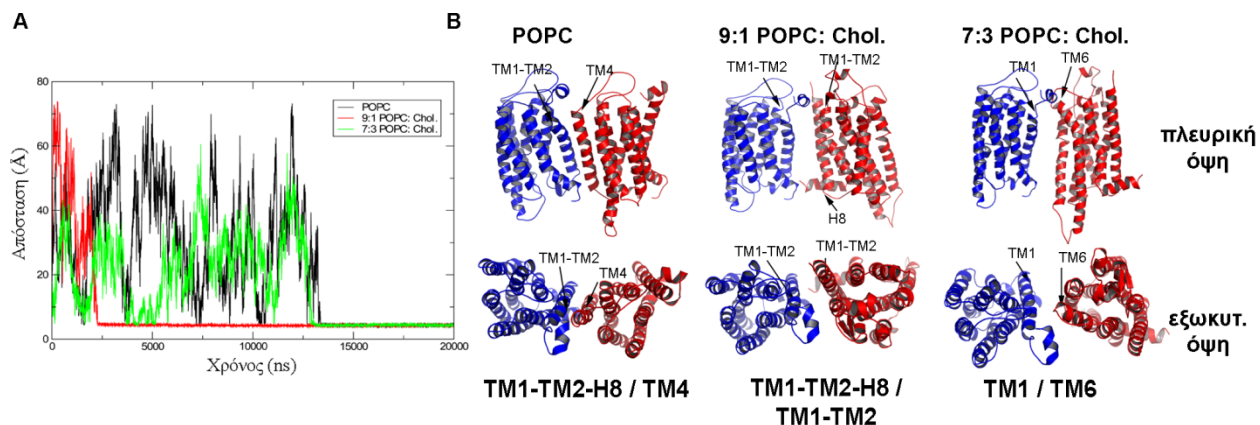
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Σε αντίθεση με τα παραπάνω αποτελέσματα, οι προσομοιώσεις του mGluR5 σε μεμβράνες με περιεκτικότητα χοληστερόλης 30% δεν έδωσαν κάποιο τελικό σύμπλοκο. Αντίθετα, παρατηρήθηκε πολύ πιο αργή κίνηση των πρωτομερών στη λιπιδική διπλοστιβάδα, καθώς και ο σχηματισμός μικρού αριθμού παροδικών συμπλόκων, τα οποία αποχωρίστηκαν σε πολύ μικρό χρόνο (<50 ns). Η ανάλυση των διαμοριακών επαφών σε αυτά τα σύμπλοκα έδειξε αλληλεπιδράσεις ανάμεσα στο τμήμα TM1 του ενός πρωτομερούς και είτε το τμήμα TM4 του δεύτερου είτε τα τμήματα TM5 και TM6 του δεύτερου, με πολύ μικρό αριθμό αλληλεπιδρώντων καταλοίπων. Προκειμένου να διερευνηθεί ο πιθανός σχηματισμός τελικού συμπλόκου σε μεγαλύτερο χρονικό διάστημα, ο χρόνος των προσομοιώσεων για τον υποδοχέα επεκτάθηκε από τα 20 μ s των υπόλοιπων συστημάτων στα 35 μ s, ωστόσο δεν προέκυψε κάποιο τελικό σύμπλοκο. Τα παραπάνω αποτελέσματα υποδεικνύουν ότι η παρουσία χοληστερόλης σε περιεκτικότητα 30% δρα ανασταλτικά στην ικανότητα του mGluR5 να σχηματίζει ομοδιμερή.

Ετεροδιμερή A2A-mGluR5: Παράλληλα με τα ομοδιμερή, διερευνήθηκε και η ικανότητα των υποδοχέων A2A και mGluR5 στο να σχηματίζουν ετεροδιμερή μέσα από την προσομοίωσή τους σε μεμβράνες POPC παρουσία ή απουσία χοληστερόλης. Αντίστοιχα με τις προσομοιώσεις ομοδιμερισμού, και οι προσομοιώσεις ετεροδιμερισμού οδήγησαν στο σχηματισμό διμερών, με διαφορετικά αποτελέσματα για κάθε τύπο μεμβράνης (Εικόνα 56). Συγκεκριμένα, οι προσομοιώσεις σε μεμβράνη POPC οδήγησαν στο σχηματισμό ενός μη συμμετρικού διμερούς, αποτελούμενου από την όψη TM1-TM2-H8 του υποδοχέα A2A και από το τμήμα TM4 του mGluR5. Οι προσομοιώσεις σε μεμβράνη με περιεκτικότητα χοληστερόλης 10% οδήγησαν στο σχηματισμό ενός ψευδο-συμμετρικού διμερούς TM1-TM2-H8 (A2A) / TM1-TM2 (mGluR5). Αντίστοιχα με τον ομοδιμερισμό, οι προσομοιώσεις του ετεροδιμερισμού οδήγησαν στο τελικό σύμπλοκο σε σχετικά μικρό χρόνο (~2.4 μ s). Τέλος, οι προσομοιώσεις σε μεμβράνη με περιεκτικότητα χοληστερόλης 30% οδήγησαν στο σχηματισμό ενός χαλαρά συνδεδεμένου TM6 (A2A) / TM1 (mGluR5). Συνολικά, τα αποτελέσματα των προσομοιώσεων ετεροδιμερισμού φαίνονται να ακολουθούν το μοτίβο των προσομοιώσεων ομοδιμερισμού

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

του A2A παρά του mGluR5, υποδηλώνοντας ότι ο αδενοσινικός υποδοχέας, πιθανότατα, κατευθύνει με κάποιο τρόπο τη διαδικασία σχηματισμού των συμπλόκων.



Εικόνα 56 Αποτελέσματα των προσομοιώσεων για τον αυθόρμητο ετεροδιμερισμό των υποδοχέων A2A και mGluR5. (A) Διάγραμμα της απόστασης ανάμεσα στα πρωτομερή σε κάθε σύστημα μεμβράνης. Ο οριζόντιος άξονας απεικονίζει το χρόνο (σε ns) ενώ ο κατακόρυφος άξονας απεικονίζει την απόσταση (σε Å). Οι μετρήσεις των τριών μεμβρανικών συστημάτων απεικονίζονται με μαύρο, κόκκινο και πράσινο χρώμα, αντίστοιχα. (B) Χαρακτηριστικές δομές των τελικών διμερών που προέκυψαν από τις προσομοιώσεις.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 11 Μετρήσεις του συντελεστή πλευρικής διάχυσης (D_{lat}) στις προσομοιώσεις διμερισμού των A2A και mGluR5.

Μεμβράνη	D_{lat} (Πρωτομερές A) [$10^{-7} \text{ cm}^2/\text{s}$]	D_{lat} (Πρωτομερές B) [$10^{-7} \text{ cm}^2/\text{s}$]
<i>Ομοδιμερισμός A2A-A2A</i>		
POPC	0.08 ± 0.01	0.015 ± 0.01
POPC με 10% χοληστερόλη	0.20 ± 0.01	0.102 ± 0.01
POPC με 30% χοληστερόλη	0.027 ± 0.02	0.022 ± 0.01
<i>Ομοδιμερισμός A2A-mGluR5</i>		
POPC	0.03 ± 0.01	0.03 ± 0.01
POPC με 10% χοληστερόλη	0.18 ± 0.02	0.19 ± 0.02
POPC με 30% χοληστερόλη	$(6.38 \pm 1.19) \cdot 10^{-3}$	$(4.52 \pm 1.32) \cdot 10^{-3}$
<i>Ετεροδιμερισμός A2A-mGluR5¹</i>		
POPC	0.092 ± 0.01	0.03 ± 0.02
POPC με 10% χοληστερόλη	0.221 ± 0.01	0.21 ± 0.03
POPC με 30% χοληστερόλη	0.030 ± 0.02	$(3.98 \pm 1.59) \cdot 10^{-3}$

¹ Στα συστήματα ετεροδιμερισμού ο A2A είναι το πρωτομερές A και ο mGluR5 είναι το πρωτομερές B.

Τα αποτελέσματα όλων των προσομοιώσεων, τόσο στον ομοδιμερισμό όσο και στον ετεροδιμερισμό έδειξαν ότι σε σύγκριση με τις αμιγώς φωσφολιπιδικές μεμβράνες, η παρουσία χοληστερόλης φαίνεται να επιδρά άμεσα τόσο στην ικανότητα διμερισμού των υποδοχέων όσο και στον τύπο του διμερούς που σχηματίζεται. Με βάση τα αποτελέσματα των προσομοιώσεων, η βέλτιστη περιεκτικότητα χοληστερόλης στη μεμβράνη για το σχηματισμό ομο- και ετεροδιμερών φαίνεται να είναι 10%. Αυτή η τιμή είναι σύμφωνη με την πειραματικά γνωστή περιεκτικότητα χοληστερόλης (6-15%) των μεμβρανών σε κύτταρα του ανθρώπινου νευρικού συστήματος, στα οποία οι A2A και mGluR5 εκφράζονται. Όπως φάνηκε από τα αποτελέσματα των προσομοιώσεων σε περιεκτικότητα χοληστερόλης 10%, η παρουσία των στερολών επιταχύνει το σχηματισμό σταθερών διμερών για τους υποδοχείς, ενώ φαίνεται να τους ωθεί στο να σχηματίσουν συμμετρικά διμερή (Εικόνες 54-56). Αυτή η συμπεριφορά είναι σύμφωνη με προηγούμενα πειραματικά δεδομένα από άλλους GPCRs (π.χ. οπιοειδείς υποδοχείς), στους οποίους έχει παρατηρηθεί θετική επίδραση της χοληστερόλης στο σχηματισμό διμερών. Αντίθετα, η περαιτέρω αύξηση της περιεκτικότητας της

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

χοληστερόλης στα 30% φαίνεται να έχει το αντίθετο αποτέλεσμα, δυσχεραίνοντας το σχηματισμό διμερών και, στην περίπτωση του mGluR5, εμποδίζοντάς τον πλήρως.

Η επίδραση της χοληστερόλης πάνω στην κινητικότητα των υποδοχέων μπορεί να φανεί μέσα από τον υπολογισμό του συντελεστή πλευρικής διάχυσης (D_{lat}), ενός μέτρου της κινητικότητας στο επίπεδο της λιπιδικής διπλοστιβάδας (Πίνακας 3.3). Όπως φαίνεται από τις μετρήσεις Πίνακα 11, οι υποδοχείς σε μεμβράνες POPC έχουν τιμές D_{lat} στην περιοχή των $0.03-0.09 \cdot 10^{-7} \text{ cm}^2/\text{s}$. Η προσθήκη της χοληστερόλης σε περιεκτικότητα 10% αυξάνει το συντελεστή διάχυσης κατά μία τάξη μεγέθους, σε τιμές $0.1-0.2 \cdot 10^{-7} \text{ cm}^2/\text{s}$, υποδεικνύοντας ότι οι υποδοχείς διαχέονται πιο γρήγορα στη μεμβράνη. Περαιτέρω αύξηση της χοληστερόλης οδηγεί σε μείωση της διάχυσης για τον A2A και κυρίως στον mGluR5, του οποίου οι τιμές D_{lat} είναι τρεις τάξεις μεγέθους μικρότερες από αυτές του A2A, υποδεικνύοντας ότι οι κινήσεις του υποδοχέα δυσχεραίνονται σε πολύ μεγαλύτερο βαθμό. Συνολικά, οι παραπάνω παρατηρήσεις είναι σύμφωνες με προηγούμενες μελέτες πάνω στην επίδραση της χοληστερόλης σε φωσφολιπιδικές μεμβράνες. Δεδομένου ότι ο σχηματισμός πρωτεϊνικών αλληλεπιδράσεων, συμπεριλαμβανομένων των διαμεμβρανικών πρωτεϊνών, θεωρείται ότι επηρεάζεται από την ικανότητα διάχυσης των πρωτεϊνών, οι σημαντικές διαφορές στην πλευρική διάχυση των πρωτομερών ίσως αποτελούν μέρος του μηχανισμού με τον οποίο η χοληστερόλη επηρεάζει το διμερισμό τους. Έτσι, η αυξημένη διάχυση των υποδοχέων σε περιεκτικότητα χοληστερόλης 10% διευκολύνει το σχηματισμό των συμπλόκων τους, ενώ περαιτέρω αύξηση της περιεκτικότητας οδηγεί σε μειωμένη διάχυση, δυσχεραίνοντας το διμερισμό τους.

Συνολικά, τα αποτελέσματα των προσομοιώσεων ομο- και ετεροδιμερισμού των A2A και mGluR5 δείχνουν ότι και οι δύο υποδοχείς έχουν την ικανότητα να σχηματίζουν αυθόρμητα υπερμοριακά συμπλέγματα, τόσο ομοδιμερή όσο και ετεροδιμερή. Τα διμερή φαίνονται να σχηματίζονται από συγκεκριμένα τμήματα υποδοχέων με γνωστή εμπλοκή στις αλληλεπιδράσεις ολιγομερισμού (TM1, TM2, TM4, TM5, TM6, H8). Αξίζει να σημειωθεί ότι τα παραπάνω τμήματα συμμετέχουν στην επιφάνεια αλληλεπίδρασης τόσο στα συμμετρικά όσο και στα μη συμμετρικά διμερή, υποδεικνύοντας ότι αποτελούν, γενικά, τις προτιμώμενες θέσεις διμερισμού των δύο υποδοχέων. Η παραπάνω παρατήρηση είναι σε συμφωνία με τα

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

αποτελέσματα των Ενοτήτων 3.1.1 και 3.1.2, και επιβεβαιώνει περαιτέρω ότι οι GPCRs χρησιμοποιούν συγκεκριμένα μόνο τμήματα στις αλληλεπιδράσεις ολιγομερισμού. Πρέπει να σημειωθεί ότι η επίδραση της μεταβολής των επιπέδων χοληστερόλης στο διμερισμό, όπως φάνηκε από τα αποτελέσματα των προσομοιώσεων, ίσως έχει βιολογική σημασία. Κύτταρα διαφορετικών ιστών διαθέτουν κυτταρική μεμβράνη με διαφορετική σύσταση. Έτσι, οι διαφορές στα επίπεδα της χοληστερόλης μπορεί να επιτρέπουν το σχηματισμό διμερών των υποδοχέων σε έναν ιστό (π.χ. εγκέφαλος) και να τον εμποδίζουν σε κάποιον άλλο (π.χ. λευκοκύτταρα), ή να οδηγούν στο σχηματισμό διαφορετικών διμερών. Επιπρόσθετα, είναι πειραματικά γνωστό ότι τα επίπεδα χοληστερόλης στην κυτταρική μεμβράνη μεταβάλλονται σημαντικά με την αύξηση της ηλικίας. Κάποιες από τις ασθένειες με τις οποίες οι A2A και mGluR5 έχουν συσχετιστεί (π.χ. ασθένεια Parkinson's) έχουν βρεθεί σχετίζονται με την ηλικία του ασθενή. Έτσι, μπορεί να υποτεθεί ότι η μεταβολή των επιπέδων χοληστερόλης και η επίδρασή της στη δομή και τις αλληλεπιδράσεις υποδοχέων όπως ο A2A και ο mGluR5 είναι ένας από τους παράγοντες που συμμετέχουν στην παθολογία των ασθενών.

3.2 Εξωτερικές Μembrάνες των αρνητικών κατά Gram βακτηρίων

3.2.1 Η μέθοδος προτυποποίησης Εξωτερικών Μembrανών GNOMM

Στα πλαίσια της παρούσας διδακτορικής διατριβής αναπτύχθηκε το εργαλείο GNOMM (Gram-Negative Outer Membrane Modeler), μια αυτοματοποιημένη μέθοδος για την προτυποποίηση μεμβρανών με λιποπολυσακχαρίτες, για τη διεξαγωγή προσομοιώσεων Μοριακής Δυναμικής σε συστήματα Εξωτερικών Μembrανών. Η μέθοδος είναι διαθέσιμη στη διεύθυνση <http://bioinformatics.biol.uoa.gr/GNOMM> και υποστηρίζει την κατασκευή συστημάτων προσομοιώσεων σε τέσσερα δυναμικά πεδία (CHARMM36, GROMOS 54A7, MARTINI και PACE) για το ελεύθερα διαθέσιμο πρόγραμμα προσομοιώσεων GROMACS. Από τη σελίδα της μεθόδου ο χρήστης μπορεί να επιλέξει το είδος του συστήματος που επιθυμεί να κατασκευάσει, καθώς και το δυναμικό πεδίο με το οποίο επιθυμεί να προτυποποιήσει το σύστημά του (Εικόνα 57).

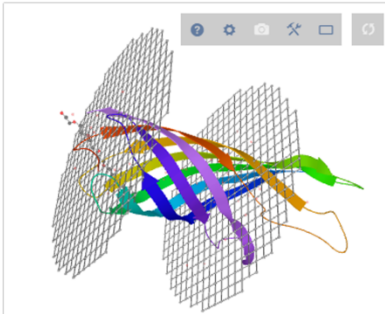
The screenshot displays the GNOMM web interface. At the top, the title 'GRAM-NEGATIVE OUTER MEMBRANE MODELER' is shown, along with a subtitle 'Automated building of Lipopolysaccharide-rich Bacterial Outer Membranes'. A navigation bar includes links for Home, Submit, Retrieve Results, Help, Downloads, Links, and About. The 'Submit' link is highlighted with a red box and a red arrow labeled '1'. Below the navigation bar, the main form is titled '1. Choose the type of system you want to build:'. It contains two radio buttons: 'Protein/Membrane' (selected) and 'Membrane-only'. The 'Protein/Membrane' section includes a text box for 'Enter PDB ID:', a 'Source:' dropdown menu, and a 'Choose File' button. A note indicates that the PDB database should be selected as the source. The 'Membrane-only' section includes a radio button for 'Upload pre-processed PDB and Topology (for advanced users)'. Below this, the form is titled '2. Choose a Force Field:', with a 'Select Force Field' dropdown menu highlighted by a red box and a red arrow labeled '3'. A 'Submit' button is located at the bottom right of the form.

Εικόνα 57 Η ιστοσελίδα της μεθόδου GNOMM. Παρουσιάζεται η φόρμα εισόδου της μεθόδου. Ο χρήστης, μέσω της σελίδας «Υποβολή» (1) μπορεί να επιλέξει το είδος του συστήματος που θα κατασκευαστεί (2), καθώς και το δυναμικό πεδίο (3).

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

A Step 1: Protein Preparation

1.1 Structure Viewer



1.2 Protein Manipulation

Select N- and C-Termini types:

N-ter:
☒ NH₃⁺
☐ NH₂
☐ None

C-ter:
☒ COO⁻
☐ COOH
☐ None

Your PDB may have missing coordinates or non-standard atoms, which can cause problems with your chosen force field. GNOMM can attempt to fix these problems using MODELLER.

Do you want to repair your structure? ☒ Yes ☐ No

B Step 2: Membrane Definition

2.1 System Size

Please enter X and Y values for the dimensions of your membrane (in Å)

Length of X:

Length of Y:

Your submitted PDB's dimensions are 40.86 Å x 33.43 Å x 61.55 Å. Based on those, we would recommend a bilayer of at least 66 Å x 58 Å.

Note: For efficiency, membrane area size cannot exceed 90000 Å² (300 Å x 300 Å). If you want to build a larger bilayer, please contact us.

2.2 Lipid Composition

Fill the following table with the percentage ratios of the lipids you want in your bilayer. For example, if you want a 50% ratio of POPE in the lower leaflet, set the appropriate value at 50. Change only the values of the lipids you want to use and leave the rest at zero (0). Make sure that the total sum of each leaflet is 100% (100).

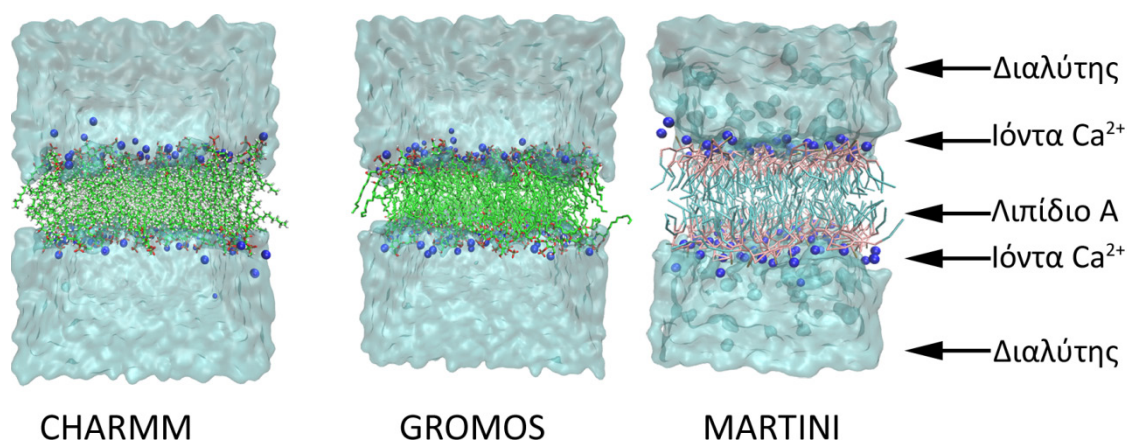
Lipopolysaccharides (LPSs)				
Lipid	Abbreviation	No. of Tails	Upper Leaflet	Lower Leaflet
<i>Escherichia coli</i>				
Lipid A	ECLI	6	0	0
Re LPS	REMP	6	0	0
Ra LPS	RAMP	6	100	0
<i>Pseudomonas aeruginosa</i>				
Lipid A Type 1	PAL6	6	0	0
Lipid A Type 2	PAL5	5	0	0
PAO1 LPS	PAO1	5	0	0
<i>Helicobacter pylori</i>				
Lipid A	HPLI	4	0	0
<i>Neisseria meningitidis</i>				
Lipid A	NMEN	6	0	0

Phospholipids				
Lipid	Abbreviation	Lipid Tails	Upper	Lower
Phosphatidyl-cholines (PCs)				
1-palmitoyl-2-oleyl-phosphatidyl-choline	POPC	16:0/18:1	0	0
di-oleyl-phosphatidyl-choline	DOPC	18:1/18:1	0	0
di-stearoyl-phosphatidyl-choline	DSPC	18:0/18:0	0	0
di-palmitoyl-phosphatidyl-choline	DPPC	16:0/16:0	0	0
di-myristoyl-phosphatidyl-choline	DMPC	14:0/14:0	0	0
di-lauroyl-phosphatidyl-choline	DLPC	12:0/12:0	0	0
Phosphatidyl-ethanolamines (PEs)				

Εικόνα 58 Η σελίδα προτυποποίησης του συστήματος προσομοίωσης. Σε αυτή τη σελίδα ο χρήστης καλείται να καθορίσει τα χαρακτηριστικά του συστήματός του (ιδιότητες πρωτεΐνης, μεμβράνης και διαλύτη). Παρατίθενται τμήματα της φόρμας για τις πρωτεΐνες (A) και τη λιπιδική διπλοστιβάδα (B). Σημαντικά σημεία της φόρμας εισόδου σημαίνονται με κόκκινα πλαίσια.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Έχοντας καθορίσει το είδος του συστήματος και το δυναμικό πεδίο της επιλογής του, ο χρήστης στη συνέχεια καλείται να καθορίσει τα χαρακτηριστικά του συστήματος, τα οποία περιλαμβάνουν μια σειρά από επιλογές για την τοπολογία της πρωτεΐνης (π.χ. φορτίο N- και C- άκρου), τις διαστάσεις και τη σύνθεση της μεμβράνης, καθώς επίσης και τα χαρακτηριστικά του υδατικού διαλύτη (Εικόνα 58). Τα αποτελέσματα της μεθόδου GNOMM περιλαμβάνουν τη δομή του πλήρους συστήματος προσομοίωσης στο σύστημα αρχείων PDB. Επιπρόσθετα, δίνονται στο χρήστη όλα τα απαραίτητα στοιχεία της τοπολογίας του συστήματος στο σύστημα αρχείων του προγράμματος προσομοίωσης GROMACS. Το σύνολο των παραπάνω αρχείων μπορούν να χρησιμοποιηθούν απευθείας για τη διεξαγωγή προσομοιώσεων Μοριακής Δυναμικής με το GROMACS, χωρίς να απαιτείται κάποια περαιτέρω τροποποίηση της δομής ή της τοπολογίας από το χρήστη.



Εικόνα 59 Προσομοιώσεις Μοριακής Δυναμικής σε μεμβράνες Λιπιδίου A. Παρουσιάζονται τα συστήματα προσομοίωσης για το Λιπίδιο A της *E. coli* (ECL1) σε τρία δυναμικά πεδία.

Η αξιολόγηση της πιστότητας και σταθερότητας των συστημάτων προσομοίωσης που παράγονται από τη μέθοδο GNOMM πραγματοποιήθηκε αρχικά μέσα από την προσομοίωση πρότυπων διπλοστιβάδων Λιπιδίου A σε τρία δυναμικά πεδία, τα CHARMM36, GROMOS 54A7 και MARTINI (Εικόνα 59). Τα αποτελέσματα των προσομοιώσεων αναλύθηκαν χρησιμοποιώντας ως κριτήρια δύο ιδιότητες των λιπιδίων, την επιφάνεια ανά λιπίδιο και το πάχος της υδρόφοβης διπλοστιβάδας. Τα αποτελέσματα των μετρήσεων παρουσιάζονται στον

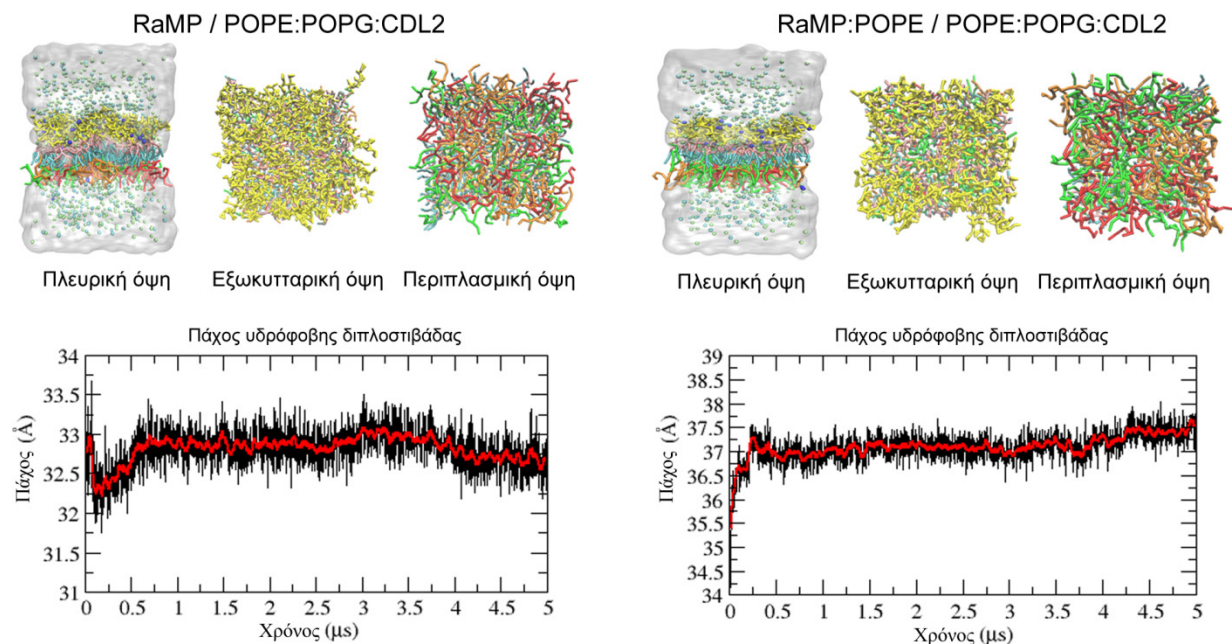
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακα 12. Όπως φαίνεται και από τις μετρήσεις, τα αποτελέσματα των προσομοιώσεων για κάθε τύπο Λιπιδίου A παρουσιάζουν συμφωνία ανάμεσα στα διαφορετικά δυναμικά πεδία, τόσο για την επιφάνεια όσο και για το πάχος της διπλοστιβάδας. Επιπρόσθετα, στην περίπτωση του Λιπιδίου A της *E. coli* (ECLI) και του πεντα-ακυλιωμένου Λιπιδίου A του *P. aeruginosa* (PAL5), οι μετρήσεις της επιφάνειας ανά λιπίδιο συμφωνούν με προηγούμενα πειραματικά δεδομένα. Συγκεκριμένα, οι τιμές επιφάνειας του ECLI σε όλα τα δυναμικά πεδία ($148.15\text{-}154.57 \text{ \AA}^2$) βρίσκονται μέσα στα όρια των πειραματικά προσδιορισμένων τιμών του ECLI σε διάφορες θερμοκρασίες ($130\text{-}156 \text{ \AA}^2$). Αντίστοιχα, οι τιμές επιφάνειας του PAL5 σε όλα τα πεδία ($139.17\text{-}143.55 \text{ \AA}^2$) προσεγγίζουν τις μετρήσεις Λιπιδίων A του *P. aeruginosa* και άλλων συγγενικών ειδών στην υγροκρυσταλλική φάση ($125\text{-}142 \text{ \AA}^2$). Στο σύνολό τους, τα παραπάνω αποτελέσματα δείχνουν ότι η τοπολογία και οι παράμετροι των διάφορων τύπων Λιπιδίου A του GNOMM είναι θερμοδυναμικά σταθερές, καθώς και ότι τα αποτελέσματα των διαφορετικών δυναμικών πεδίων συμφωνούν μεταξύ τους, καθώς επίσης και με τα διαθέσιμα πειραματικά δεδομένα, ανεξάρτητα από το επίπεδο διακριτικότητας του μοντέλου.

Πίνακας 12 Μετρήσεις της επιφάνειας ανά λιπίδιο και του πάχους της λιπιδικής διπλοστιβάδας από τα αποτελέσματα των προσομοιώσεων για τις ομοιογενείς μεμβράνες Λιπιδίου A.

Λιπίδιο	CHARMM36	GROMOS 54A7	MARTINI
<i>Επιφάνεια ανά λιπίδιο ($\text{\AA}^2$)</i>			
ECLI	154.57 ± 1.75	148.15 ± 1.84	152.35 ± 1.56
PAL5	143.55 ± 1.89	141.48 ± 1.54	139.17 ± 1.70
PAL6	146.87 ± 1.90	142.54 ± 1.48	139.10 ± 1.78
HPLI	110.12 ± 2.92	107.50 ± 2.78	103.11 ± 2.85
NMEN	179.70 ± 1.84	179.72 ± 1.69	175.80 ± 1.48
<i>Πάχος διπλοστιβάδας ($\text{\AA}$)</i>			
ECLI	35.44 ± 0.80	36.01 ± 0.70	36.89 ± 0.78
PAL5	31.70 ± 0.97	31.54 ± 0.76	30.10 ± 0.46
PAL6	32.01 ± 0.82	31.85 ± 0.55	30.87 ± 0.59
HPLI	46.20 ± 0.65	45.90 ± 0.58	46.30 ± 0.75
NMEN	43.10 ± 0.75	42.56 ± 0.77	41.27 ± 0.84

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 60 Αποτελέσματα προσομοιώσεων των μικτών μεμβρανών. Παρουσιάζονται οι δομές των μικτών μεμβρανών (πάνω), καθώς και μετρήσεις του πάχους της διπλοστιβάδας στη μονάδα του χρόνου (κάτω). Οι δομές παρουσιάζονται χρησιμοποιώντας κίτρινο χρώμα για τον πυρήνα του λιποπολυσακχαρίτη RaMP, ροζ για τις πολικές κεφαλές του Λιπιδίου A και κυανό για τις υδρόφοβες ουρές του, πράσινο για τα λιπίδια POPE, κόκκινο για τα λιπίδια POPG και πορτοκαλί για τις καρδιολιπίνες CDL2. Στις καμπύλες του πάχους της διπλοστιβάδας, η κόκκινη καμπύλη αντιπροσωπεύει το μέσο όρο των μετρήσεων και η μαύρη καμπύλη τα όρια των τιμών.

Εκτός από τις προσομοιώσεις των διπλοστιβάδων του Λιπιδίου A, πραγματοποιήθηκαν μια σειρά από προσομοιώσεις μικτών μεμβρανών, αποτελούμενων από λιποπολυσακχαρίτες στην εξωτερική στιβάδα και φωσφολιπίδια στην εσωτερική, προκειμένου να διερευνηθεί η ικανότητα προτυποποίησης της μη συμμετρικής φύσης της Εξωτερικής Μεμβράνης (Εικόνα 60). Συγκεκριμένα, πραγματοποιήθηκαν προσομοιώσεις μεμβρανών με μίξη φωσφολιπιδίων POPE, POPG και CDL2 στην εσωτερική στιβάδα και είτε πλήρη λιποπολυσακχαρική εξωτερική στιβάδα, αποτελούμενη αποκλειστικά από Ra LPS (RaMP), είτε μια μικτή RaMP : POPE στιβάδα, η οποία προσομοιάζει τις ατέλειες που έχουν βρεθεί να προκαλούνται στην εξωτερική μεμβράνη της *E. coli* από την παρουσία φωσφολιπιδίων. Τα αποτελέσματα των προσομοιώσεων των δύο μεμβρανών έδειξαν ότι και οι δύο διπλοστιβάδες ήταν σχετικά άκαμπτες στην εξωτερική όψη τους, γεγονός που οφείλεται κυρίως στην ογκώδη δομή του

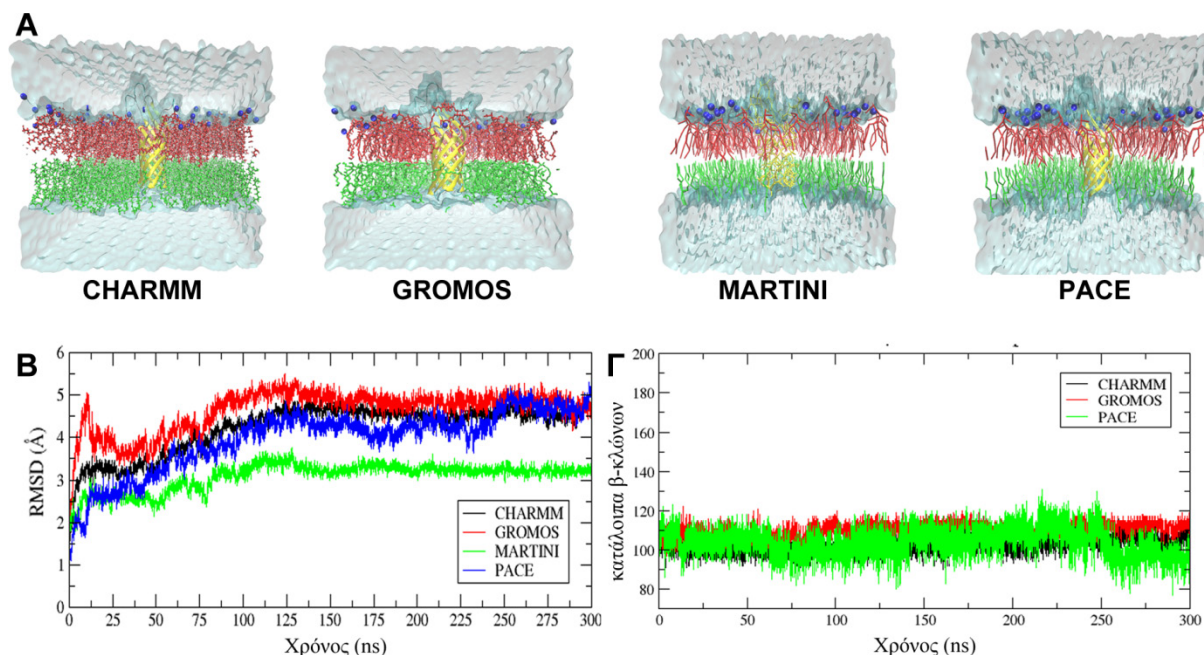
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

RaMP. Αντίθετα, τα φωσφολιπίδια της εσωτερικής στιβάδας παρουσίασαν φυσιολογική πλευρική διάχυση, ενώ οι μετρήσεις της επιφάνειάς τους έδωσαν τιμές πολύ κοντά στις πειραματικά προσδιορισμένες τιμές αναφοράς για κάθε λιπιδικό τύπο (Πίνακας 13). Από την άλλη, οι μετρήσεις για την εξωτερική στιβάδα έδειξαν ότι η παρουσία του POPE ωθεί σε μείωση της επιφάνειας ανά λιπίδιο του RaMP. Μια δεύτερη, πιο έντονη συμπεριφορά παρατηρείται στο πάχος της υδρόφοβης διπλοστιβάδας, το οποίο είναι σημαντικά μεγαλύτερο στην «ατελή» διπλοστιβάδα με RaMP:POPE από ότι στην διπλοστιβάδα με RaMP (Εικόνα 60). Για την ακρίβεια, παρατηρείται ότι το πάχος της διπλοστιβάδας αυξάνεται κατά 4.5 Å στη «ατελή» μεμβράνη με ένθετα μόρια POPE στην εξωτερική στιβάδα. Αυτές οι μετρήσεις καταγράφονται σε σχετικά μικρό χρόνο (~1 μs) και παραμένουν σταθερές για το υπόλοιπο του χρόνου προσομοίωσης και στα δύο συστήματα. Πρέπει να σημειωθεί πως τα παραπάνω αποτελέσματα είναι σύμφωνα με προηγούμενες μετρήσεις πάνω στο λιποπολυσακχαρίτη RaMP, καθώς και με όσα είναι γνωστά για την επίδραση των φωσφολιπιδικών παρεμβολών στο στρώματα των λιποπολυσακχαριτών. Συγκεκριμένα, τόσο η αύξηση του πάχους της διπλοστιβάδας όσο και η μείωση της επιφάνειας ανά λιπίδιο για το λιποπολυσακχαρίτη RaMP υποδεικνύουν χαλάρωση στη δομή της μεμβράνης, γεγονός που την καθιστά πιο ασταθή. Αυτή η αυξημένη αστάθεια είναι σύμφωνη με τα υπάρχοντα πειραματικά δεδομένα, σύμφωνα με τα οποία ατέλειες στην εξωτερική στιβάδα των λιποπολυσακχαριτών αυξάνουν την αστάθεια της Εξωτερικής Μεμβράνης και διευκολύνουν την διέλευση βλαβερών ουσιών (π.χ. αντιβιοτικά) στο εσωτερικό του βακτηριακού κυττάρου.

Πίνακας 13 Τιμές επιφάνειας ανά λιπίδιο σε Å² για τα λιπίδια των μικτών μεμβρανών.

Μόριο	Μεμβράνη RaMP / POPE:POPG:CDL2	Μεμβράνη RaMP:POPE / POPE:POPG:CDL2
<i>Εξωτερική στιβάδα</i>		
RaMP	175.35 ± 2.60	164.01 ± 2.76
POPE	–	66.06 ± 3.96
<i>Περιπλαστική Στιβάδα</i>		
POPE	59.21 ± 2.43	57.53 ± 3.85
POPG	60.05 ± 2.90	62.90 ± 3.34
CDL2	106.02 ± 5.11	104.07 ± 7.30

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

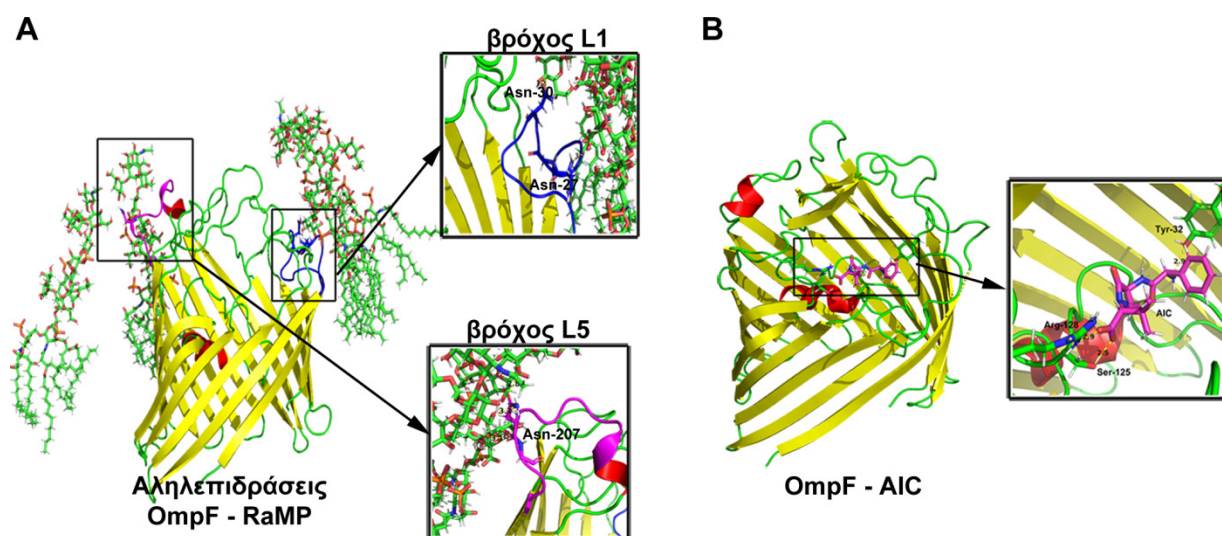


Εικόνα 61 Προσομοιώσεις Μοριακής Δυναμικής για το β-βαρέλι OmpA. (Α) Οι δομές των συστημάτων προσομοίωσης της OmpA σε τέσσερα δυναμικά πεδία. Το β-βαρέλι παρουσιάζεται χρησιμοποιώντας κίτρινα βέλη για την απεικόνιση των β-κλώνων. Η εξωτερική στιβάδα, αποτελούμενη από Λιπίδιο A, απεικονίζεται με κόκκινο χρώμα ενώ η εσωτερική στιβάδα, αποτελούμενη από φωσφολιπίδια POPE, απεικονίζεται με πράσινο. (Β) Μετρήσεις RMSD για τις προσομοιώσεις της OmpA στη μονάδα του χρόνου. (Γ) Μετρήσεις του αριθμού των καταλοίπων που συμμετέχουν σε διαμεμβρανικούς β-κλώνους στη μονάδα του χρόνου.

Η ικανότητα του GNOMM στην προτυποποίηση συστημάτων πρωτεϊνών/μεμβράνης ελέγχθηκε περαιτέρω μέσα από μια σειρά προσομοιώσεων των διαμεμβρανικών β-βαρελίων OmpA και OmpF του βακτηρίου *E. coli*, σε λιπιδικές διπλοστιβάδες με λιποπολυσακχαρίτες και φωσφολιπίδια. Για την OmpA, πραγματοποιήθηκαν προσομοιώσεις του β-βαρελίου σε λιπιδικές διπλοστιβάδες Λιπιδίου A και POPE (ECLIP/POPE) και στα τέσσερα δυναμικά πεδία του GNOMM (CHARMM, GROMOS, MARTINI και PACE) για 300 ns (Εικόνα 61). Τα αποτελέσματα των προσομοιώσεων ελέγχθηκαν μέσα από μετρήσεις RMSD στη μονάδα του χρόνου για την αξιολόγηση της σταθερότητας του βαρελίου. Όπως προκύπτει από τις παραπάνω μετρήσεις, το β-βαρέλι σε όλα τα δυναμικά πεδία φαίνεται να φτάνει σε κατάσταση ισορροπίας σχετικά νωρίς στην προσομοίωση (~100-125 ns), με τιμές RMSD στην περιοχή 3.0-4.5 Å. Για την αξιολόγηση της σταθερότητας της δομής των διαμεμβρανικών τμημάτων του β-βαρελίου, πραγματοποιήθηκαν επίσης και μετρήσεις των στοιχείων δευτεροταγούς δομής για τις

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

προσομοιώσεις των πεδίων CHARMM, GROMOS και PACE, στα οποία η πρωτεΐνη προτυποποιείται σε πλήρως ή σχεδόν ατομική διακριτικότητα. Συγκεκριμένα, μετρήθηκε ο αριθμός των καταλοίπων που συμμετέχουν στους διαμεμβρανικούς β-κλώνους του βαρελιού (Εικόνα 61). Τα αποτελέσματα των μετρήσεων δείχνουν ότι και στα τρία δυναμικά πεδία η β-πτυχωτή επιφάνεια του βαρελιού παραμένει ανέπαφη, καθώς ο αριθμός των καταλοίπων που συμμετέχουν στη β-πτυχωτή επιφάνεια διατηρείται σχεδόν σταθερός, υποδεικνύοντας ότι τα συστήματα προσομοίωσης της πρωτεΐνης, όπως έχουν κατασκευαστεί από το GNOMM, προτυποποιούν ορθά τη σταθερότητα της δομής κατά την προσομοίωση.



Εικόνα 62 Προσομοιώσεις Μοριακής Δυναμικής για το β-βαρέλι OmpF. (Α) Αλληλεπιδράσεις της OmpF με παρακείμενα μόρια RaMP κατά τη διάρκεια των προσομοιώσεων με το δυναμικό πεδίο CHARMM. Απεικονίζονται οι αλληλεπιδράσεις μορίων RaMP με τους βρόχους L1 και L5 του β-βαρελιού, οι οποίοι αποτελούν επίτοπους για μονοκλωνικά αντισώματα. (Β) Τελική δομή του συμπλόκου της OmpF με το αντιβιοτικό Αμικικιλίνη (AIC).

Στην περίπτωση της OmpF, πραγματοποιήθηκαν αρχικά προσομοιώσεις του β-βαρελιού σε μια διπλοστιβάδα με λιποπολυσακχαρίτες RaMP στην εξωτερική όψη και φωσφολιπίδια POPE στην περιπλασμική όψη, χρησιμοποιώντας τα δυναμικά πεδία CHARMM και MARTINI. Και σε αυτές τις προσομοιώσεις το β-βαρέλι διατήρησε σταθερή τη δομή του, αντίστοιχα με τα αποτελέσματα των προσομοιώσεων της OmpA. Επιπρόσθετα, η ανάλυση των αλληλεπιδράσεων ανάμεσα στην OmpF και τα παρακείμενα μόρια RaMP αποκάλυψε μια

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

σειρά από αλληλεπιδράσεις ανάμεσα στο βαρέλι και στους λιποπολυσακχαρίτες (Εικόνα 62α). Συγκεκριμένα, οι διαμεμβρανικοί β-κλώνοι αλληλεπιδρούν κυρίως με τις λιπιδικές ουρές του Λιπιδίου A, καθώς και με τα μόρια γλυκοζαμίνης και τις φωσφορικές ομάδες του λιπιδίου, μέσω της ζώνης αρωματικών καταλοίπων στην εξωκυτταρική όψη του βαρελίου. Από την άλλη, οι εξωκυτταρικοί βρόχοι είτε κινούνται ελεύθεροι στο διαλύτη είτε αλληλεπιδρούν με σάκχαρα του πυρήνα του λιποπολυσακχαρίτη. Αξίζει να σημειωθεί ότι κατά τη διάρκεια των προσομοιώσεων, παρατηρήθηκε ειδικά η πρόσδεση των εξωκυτταρικών βρόχων L1 και L5 του βαρελίου στους ολιγοσακχαρικούς πυρήνες παρακείμενων μορίων RaMP (Εικόνα 62α). Οι συγκεκριμένοι βρόχοι έχουν βρεθεί ότι είναι επίτοποι για την πρόσδεση μονοκλωνικών αντισωμάτων που στοχεύουν ειδικά την OmpF. Ωστόσο, πειραματικά δεδομένα έχουν δείξει ότι παρουσία λιποπολυσακχαριτών, η πρόσδεση των αντισωμάτων στη OmpF αναστέλλεται, και έχει υποτεθεί ότι οι αλληλεπιδράσεις του β-βαρελίου με τους λιποπολυσακχαρίτες επηρεάζει την προσβασιμότητα των βρόχων L1 και L5 στο διαλύτη. Τα αποτελέσματα των προσομοιώσεων του βαρελίου με RaMP στην παρούσα εργασία συμφωνούν με αυτά τα πειραματικά δεδομένα, καθώς δείχνουν ότι η αλληλεπίδραση των μορίων RaMP με την OmpF ουσιαστικά «καλύπτει» τους βρόχους L1 και L5, μειώνοντας την έκθεσή τους στο διαλύτη και, πιθανώς, εμποδίζοντας την αναγνώρισή τους από αντισώματα. Επιπρόσθετα, η ανάλυση των αλληλεπιδράσεων ανάμεσα στην OmpF και τα μόρια RaMP παρουσιάζει το σχηματισμό δεσμών υδρογόνου ανάμεσα στα κατάλοιπα Asn-27, Asn-30 (βρόχος L1) και Asn-207 (βρόχος L5) της OmpF και τα σάκχαρα του ολιγοσακχαρικού πυρήνα των RaMP, σε συμφωνία με προηγούμενα πειραματικά και υπολογιστικά δεδομένα από τη βιβλιογραφία (Εικόνα 62α).

Τέλος, μελετήθηκαν οι αλληλεπιδράσεις της OmpF με ένα μικρό μόριο, το αντιβιοτικό Αμπικιλίνη (Ampicillin, AIC), μέσα από προσομοιώσεις του συμπλόκου OmpF-AIC σε μεμβράνες RaMP/POPE. Προηγούμενα πειραματικά αποτελέσματα έχουν δείξει ότι το συγκεκριμένο αντιβιοτικό προσδένεται πάνω στο «βρόχο-κάλυμμα» (constriction loop) του πόρου του βαρελίου και επηρεάζει την κινητικότητά του, εμποδίζοντας τη ροή του διαλύτη και άλλων ουσιών μέσω του πόρου. Επιπρόσθετα, μια σειρά από κατάλοιπα της OmpF (Tyr-32, Ser-125 και Arg-128) έχουν βρεθεί να σχηματίζουν υδρογονικούς δεσμούς με το AIC, σταθεροποιώντας την πρόσδεσή του. Τα αποτελέσματα των προσομοιώσεων (Εικόνα 62β)

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

έδειξαν ότι το αντιβιοτικό παραμένει προσδεδεμένο στο εσωτερικό του πόρου το βαρελίου. Συγκεκριμένα, κατά τη διάρκεια των προσομοιώσεων το μόριο παρουσιάζει διάφορες ταλαντώσεις, ωστόσο δεν εξέρχεται από τη θέση πρόσδεσής του αλλά συνεχίζει να κλείνει, μαζί με το βρόχο-κάλυμμα, τον πόρο της OmpF. Επιπρόσθετα, οι υδρογονικοί δεσμοί των καταλοίπων Tyr-32, Ser-125 και Arg-128 διατηρούνται ανέπαφοι. Συνολικά, τα παραπάνω αποτελέσματα δείχνουν ότι το σύμπλοκο OmpF-AIC παραμένει θερμοδυναμικά σταθερό και παρουσιάζει βιολογικά σημαντικές αλληλεπιδράσεις.

Στο σύνολό τους, τα αποτελέσματα όλων των παραπάνω προσομοιώσεων δείχνουν ότι το εργαλείο GNOMM είναι ικανό να προτυποποιήσει θερμοδυναμικά σταθερά συστήματα προσομοίωσης και για τα τέσσερα υποστηριζόμενα δυναμικά πεδία, τόσο στην περίπτωση των λιπιδίων όσο και στην περίπτωση των πρωτεϊνών. Επιπρόσθετα, η σύγκριση των προσομοιώσεων του ίδιου συστήματος σε διαφορετικά πεδία δείχνει συμφωνία ανάμεσα στα αποτελέσματα, υποδεικνύοντας ότι οι ιδιότητες του κάθε συστήματος μπορούν να προτυποποιηθούν ανεξάρτητα από το επίπεδο διακριτικότητας που χαρακτηρίζει το κάθε δυναμικό πεδίο. Τέλος, η συμφωνία των μετρήσεων από τα αποτελέσματα των προσομοιώσεων με τα διαθέσιμα πειραματικά δεδομένα, τόσο στην περίπτωση των μεμβρανών όσο και στην περίπτωση των πρωτεϊνών, υποδεικνύει ότι τα συστήματα προσομοίωσης που παράγονται από το GNOMM είναι κατάλληλα για τη μελέτη της δομής και της δυναμικής των Εξωτερικών Μεμβρανών με ακρίβεια.

3.2.2 Μελέτη των αλληλεπιδράσεων ολιγομερισμού σε διαμεμβρανικά β-βαρέλια

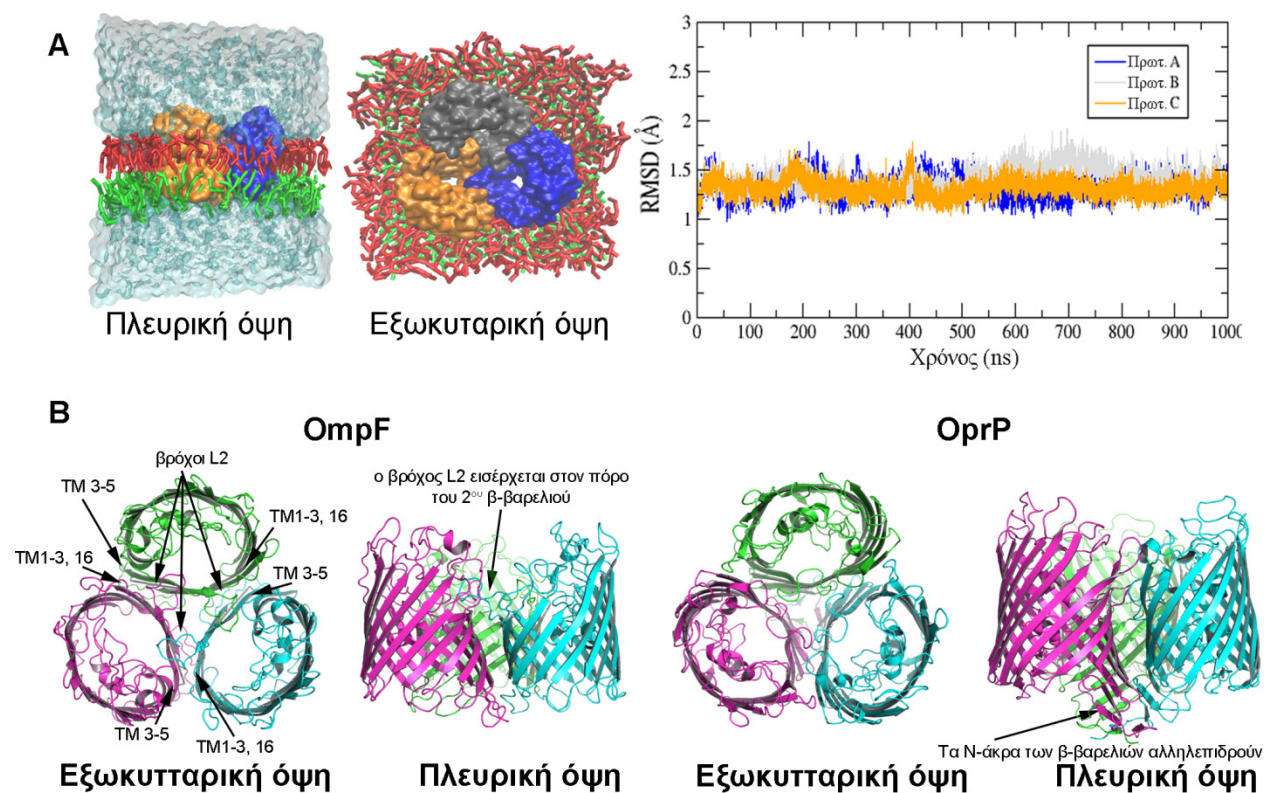
Αντικείμενο της παρούσας υποενότητας ήταν η μελέτη των αλληλεπιδράσεων ολιγομερισμού στα διαμεμβρανικά β-βαρέλια των αρνητικών κατά Gram βακτηρίων. Για την επίτευξη αυτού του στόχου σχεδιάστηκαν και πραγματοποιήθηκαν προσομοιώσεις Μοριακής Δυναμικής για ομοτρίμερη β-βαρελίων, χρησιμοποιώντας το εργαλείο GNOMM για την κατασκευή των συστημάτων προσομοίωσης. Στη μελέτη των αλληλεπιδράσεων

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

ολιγομερισμού χρησιμοποιήθηκαν οι δομές ομοτριμερών για δύο β-βαρέλια, την πορίνη OmpF της *E. coli* και τον μεταφορέα OmpP του *P. aeruginosa*. Οι παραπάνω πρωτεΐνες ανήκουν στην υπερικογένεια των «τριμερών πορινών», μια κατηγορία β-βαρελίων με αρχιτεκτονική « $n=16, S=20$ », τα οποία είναι γνωστά για το σχηματισμό ομοτριμερών. Προκειμένου να διερευνηθεί κατά πόσο η κοινή αρχιτεκτονική των τριμερών OmpF και OmpP εντοπίζεται σε άλλα β-βαρέλια, πραγματοποιήθηκε αρχικά αναζήτηση στη βάση δεδομένων PDB για τον εντοπισμό αντίστοιχων δομών ομοτριμερών, τα οποία αναλύθηκαν ως προς τις αλληλεπιδράσεις ολιγομερισμού τους. Στο σύνολο των β-βαρελίων που εξετάστηκαν βρέθηκε ότι η δομή του τριμερούς ακολουθεί το ίδιο μοτίβο με αυτό των OmpF και OmpP υποδηλώνοντας ότι η οργάνωση των τριμερών αποτελεί κοινό χαρακτηριστικό τους. Έτσι, πραγματοποιήθηκαν προσομοιώσεις για τα βαρέλια των OmpF και OmpP, τα αποτελέσματα των οποίων θεωρείται ότι αντιπροσωπεύουν το σύνολο των τριμερών πορινών της αρχιτεκτονικής « $n=16, S=20$ ».

Τα αποτελέσματα των προσομοιώσεων για τα τριμερή των OmpF και OmpP δείχνουν ότι τα σύμπλοκα παραμένουν ανέπαφα κατά τη διάρκεια της προσομοίωσης, ενώ οι δομές των μονομερών τους παρουσιάζουν αξιοσημείωτη σταθερότητα, με τιμές RMSD στην περιοχή 1.5-2.5 Å (Εικόνα 63). Η ανάλυση των διαμοριακών επαφών δείχνει ότι οι αλληλεπιδράσεις ολιγομερισμού φαίνονται να σχηματίζονται από τους διαμεμβρανικούς β-κλώνους 1-5 και 16 κάθε βαρελιού. Τα β-βαρέλια αλληλεπιδρούν ανά δύο, με τους β-κλώνους 1-3 και 16 του ενός βαρελιού να έρχονται σε επαφή με τους β-κλώνους 3-5 του επόμενου.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 63 Προσομοιώσεις Μοριακής Δυναμικής των ομοτριμερών β-βαρελίων. (Α) Η τελική δομή του ομοτριμερούς της OmpF (αριστερά) και μετρήσεις του RMSD στη μονάδα του χρόνου (δεξιά). Τα μόρια Λιπιδίου A και POPE χρωματίζονται κόκκινα και πράσινα, αντίστοιχα. (Β) Αλληλεπιδράσεις ολιγομερισμού στα τριμερή των β-βαρελίων OmpF (αριστερά) και OmpR (δεξιά). Δίνονται η εξωκυτταρική και η πλευρική όψη. Οι διαμεμβρανικοί (TM) β-κλώνοι που σχηματίζουν τις επιφάνειες ολιγομερισμού σημαίνονται στο τριμερές της OmpF και αντιστοιχούνται και στο τριμερές του OmpR, το οποίο δείχνεται με τον ίδιο προσανατολισμό. Στο τριμερές της OmpF σημαίνονται οι βρόχοι L2, οι οποίοι αποτελούν μέρος της επιφάνειας αλληλεπίδρασης, με το βρόχο του κάθε β-βαρελίου να εισέρχεται στον πόρο του επόμενου. Στο τριμερές του OmpR σημαίνονται τα N-άκρα των βαρελίων, τα οποία σχηματίζουν αλληλεπιδράσεις στην περιπλαστική όψη.

Η παραπάνω αρχιτεκτονική του τριμερούς είναι κοινή τόσο για την OmpF όσο και για τον OmpR (Εικόνα 63). Ωστόσο, οι δύο πορίνες παρουσιάζουν μια σειρά από επιπρόσθετες αλληλεπιδράσεις, οι οποίες φαίνονται να είναι ειδικές για την κάθε πρωτεΐνη. Συγκεκριμένα, η πορίνη OmpF, εκτός από τις αλληλεπιδράσεις των διαμεμβρανικών β-κλώνων, παρουσιάζει επιπρόσθετες αλληλεπιδράσεις στην εξωκυτταρική όψη της, στην οποία ο δεύτερος εξωκυτταρικός βρόχος (βρόχος L2) του κάθε μονομερούς β-βαρελίου φαίνεται να αγκυροβολεί στο εσωτερικό του πόρου του παρακείμενου β-βαρελίου και να αλληλεπιδρά με τμήμα του βρόχου-καλύμματος (constriction loop), ο οποίος ελέγχει το άνοιγμα και το κλείσιμο του

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

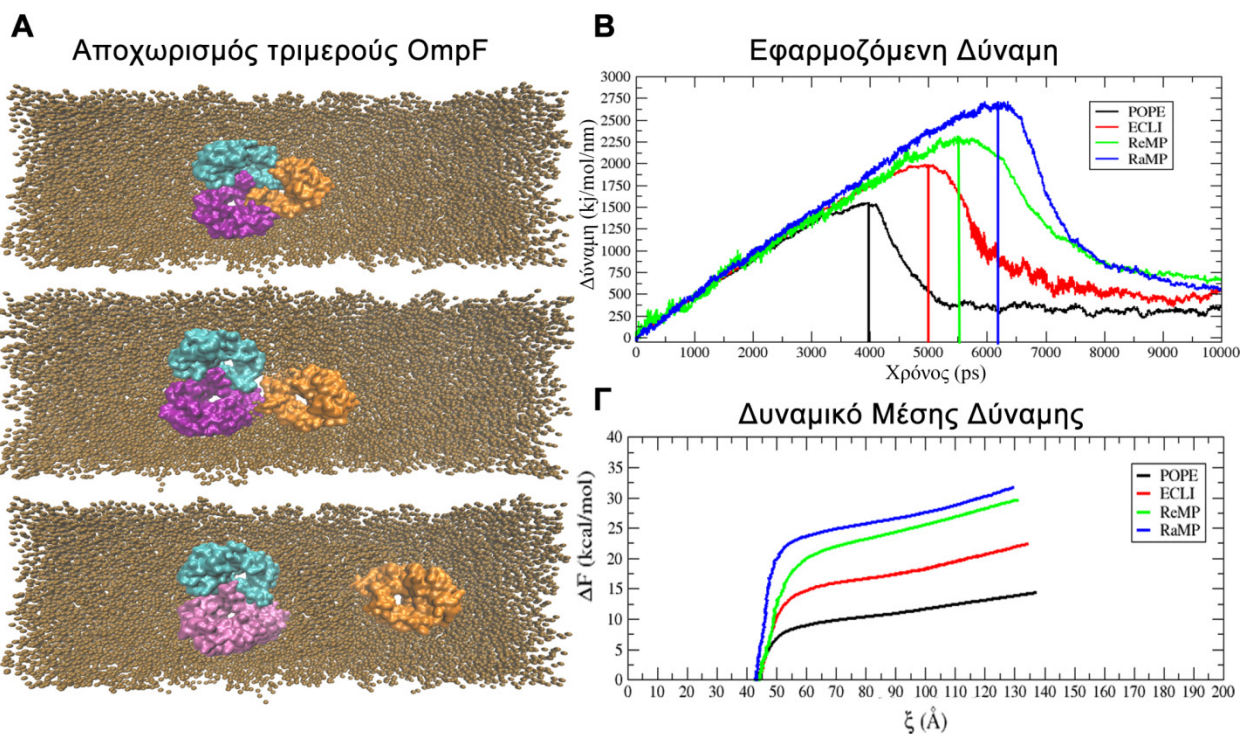
πόρου. Η παραπάνω αλληλεπίδραση παρατηρείται τόσο στην αρχική δομή όσο και στα αποτελέσματα των προσομοιώσεων και συμβαίνει και στα τρία βαρέλια του συμπλόκου. Η θέση του συγκεκριμένου βρόχου στην αλληλεπίδραση ανάμεσα στα β-βαρέλια της OmpF ίσως έχει λειτουργικό ρόλο, καθώς φαίνεται να συνδέει τις αλληλεπιδράσεις ολιγομερισμού με το άνοιγμα/κλείσιμο του πόρου των βαρελιών. Σε αντίθεση με την OmpF, δεν παρατηρείται κάποια ανάλογη αλληλεπίδραση στον μεταφορέα OtrP. Ωστόσο, τα μονομερή του τελευταίου φαίνονται να σχηματίζουν επιπρόσθετες αλληλεπιδράσεις χρησιμοποιώντας τα αμινοτελικά άκρα τους, τα οποία βρίσκονται στον περιπλαστικό χώρο. Συγκεκριμένα, το N-άκρο καθενός από τα τρία μονομερή φαίνεται να σχηματίζει μια μικρή β-πτυχωτή επιφάνεια τριών αντιπαράλληλων β-κλώνων και να αλληλεπιδρά με τα αντίστοιχα N-άκρα των άλλων δύο μονομερών, σχηματίζοντας έναν τριγωνικό σχηματισμό. Η συγκεκριμένη αλληλεπίδραση πιθανότατα συμβάλλει στη σταθεροποίηση της δομής του τριμερούς, καθώς προηγούμενα πειραματικά δεδομένα υποστηρίζουν ότι το N-άκρο του OtrP είναι απαραίτητο για το σχηματισμό του ομοτριμερούς του.

Τέλος, διερευνήθηκε η επίδραση των αλληλεπιδράσεων της OmpF με λιποπολυσακχαρίτες πάνω στη σταθερότητα του ομοτριμερούς της, μέσα από προσομοιώσεις Καθοδηγούμενης Μοριακής Δυναμικής (SMD), συνοδευόμενες από υπολογισμούς του Δυναμικού Μέσης Δύναμης (PMF) μέσω της Ισότητας του Jarzynski. Συγκεκριμένα, πραγματοποιήθηκαν προσομοιώσεις SMD για τον αποχωρισμό ενός μονομερούς από το τριμερές της OmpF σε μεμβράνες με διαφορετική σύσταση στην εξωτερική στιβάδα, αποτελούμενη από φωσφολιπίδια POPE, Λιπίδιο A (ECL1), τον αδρό λιποπολυσακχαρίτη ReMP και τον λείο λιποπολυσακχαρίτη RaMP. Τα αποτελέσματα των προσομοιώσεων παρουσιάζονται στην Εικόνα 64 και δείχνουν ότι σε όλες τις προσομοιώσεις SMD πραγματοποιείται ο αποχωρισμός και απομάκρυνση του μονομερούς από το ομοτριμερές. Ωστόσο, τόσο ο χρόνος αποχωρισμού όσο και η εφαρμοζόμενη δύναμη φαίνονται να διαφέρουν ανάμεσα στις προσομοιώσεις που πραγματοποιήθηκαν σε διαφορετικό μεμβρανικό περιβάλλον. Σε συνθήκες φωσφολιπιδικής εξωτερικής στιβάδας (POPE), ο διαχωρισμός του τριμερούς φαίνεται να συμβαίνει σχετικά νωρίς (~4 ns) στην προσομοίωση, με μια εφαρμοζόμενη δύναμη της τάξης των 1500 kJ/mol/nm. Αντίθετα, η αντικατάσταση του

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

POPE από Λιπίδιο A στην εξωτερική στιβάδα φαίνεται να δρα ανασταλτικά, αυξάνοντας το χρόνο που απαιτείται για το διαχωρισμό του συμπλόκου και, κυρίως, το μέγεθος της δύναμης που απαιτείται να εφαρμοστεί (~ 2000 kJ/mol/nm). Περαιτέρω προσθήκη στοιχείων του ολιγοσακχαριτικού πυρήνα, με τη μορφή των λιποπολυσακχαριτών ReMP (ατελής πυρήνας) και RaMP (πλήρης πυρήνας) αυξάνει περισσότερο την τιμή της απαιτούμενης δύναμης (~ 2200 kJ/mol/nm και ~ 2600 kJ/mol/nm αντίστοιχα), καθώς και το χρόνο επίτευξης του διαχωρισμού. Τα παραπάνω αποτελέσματα υποδεικνύουν ότι, σε σχέση με τα απλά φωσφολιπίδια, η παρουσία λιποπολυσακχαριτών δυσχεραίνει την αποδιάταξη του τριμερούς.

Η επίδραση των λιποπολυσακχαριτών στη σταθερότητα του συμπλόκου καταδεικνύεται ακόμη περισσότερο με τον υπολογισμό του Δυναμικού Μέσης Δύναμης (PMF) για κάθε μεμβράνη (Εικόνα 3.18γ). Όπως προκύπτει από τα αποτελέσματα των μετρήσεων, η ενέργεια που απαιτείται για τον αποχωρισμό του μονομερούς από το τριμερές σε μεμβράνες με POPE είναι $\Delta F = 14.37 \pm 2.45$ kcal/mol. Με αντικατάσταση της εξωτερικής στιβάδας της μεμβράνης από μια στιβάδα Λιπιδίου A η ενέργεια αυτή αυξάνεται στα $\Delta F = 22.35 \pm 2.70$ kcal/mol, δείχνοντας ότι η παρουσία του Λιπιδίου A δρα ανασταλτικά στην απομάκρυνση του μονομερούς από το υπόλοιπο σύμπλοκο. Τέλος, στις μεμβράνες με λιποπολυσακχαρίτες ReMP και RaMP οι ενεργειακές τιμές είναι ακόμη μεγαλύτερες ($\Delta F = 29.65 \pm 2.30$ kcal/mol και $\Delta F = 31.66 \pm 2.48$ kcal/mol, αντίστοιχα).



Εικόνα 64 Υπολογισμοί του Δυναμικού Μέσης Δύναμης για την OmpF σε διαφορετικές μεμβράνες. (Α) Στιγμιότυπα από τις προσομοιώσεις SMD για τον αποχωρισμό ενός μονομερούς της OmpF από το τριμερές. Εμφανίζονται η αρχική (πάνω) και η τελική (κάτω) στερεοδιάταξη του συστήματος, καθώς και το στιγμιότυπο έναρξης του αποχωρισμού (μέση). Παρουσιάζεται η εξωκυτταρική όψη του συστήματος. (Β) Μετρήσεις της εφαρμοζόμενης δύναμης για τον αποχωρισμό του συμπλόκου στη μονάδα του χρόνου κατά τις προσομοιώσεις SMD σε μεμβράνες με διαφορετική σύσταση στην εξωτερική στιβάδα. Ο οριζόντιος άξονας απεικονίζει το χρόνο της προσομοίωσης (σε ps) ενώ ο κατακόρυφος άξονας την εφαρμοζόμενη δύναμη (σε kJ/mol/nm). Σε κάθε καμπύλη, η κατακόρυφη γραμμή απεικονίζει το χρονικό στιγμιότυπο έναρξης του αποχωρισμού του τριμερούς. (Γ) Καμπύλες του Δυναμικού Μέσης Δύναμης (PMF) για το τριμερές OmpF για μεμβράνες με διαφορετική σύσταση στην εξωτερική στιβάδα. Ο οριζόντιος άξονας του διαγράμματος απεικονίζει την συντεταγμένη της αντίδρασης ξ (απόσταση κέντρου μάζας του πρωτομερούς από το κέντρο μάζας του υπόλοιπου συμπλόκου) σε Å, ενώ ο κατακόρυφος άξονας απεικονίζει τις τιμές του PMF σε kcal/mol.

Τα παραπάνω αποτελέσματα δείχνουν ότι η παρουσία λιποπολυσακχαριτών στην εξωτερική στιβάδα της μεμβράνης φαίνεται να επηρεάζει τη σταθερότητα του ομοτριμερούς της OmpF. Η διαφορά του δυναμικού PMF ανάμεσα στις μεμβράνες με απλά φωσφολιπίδια και στις μεμβράνες με Λιπίδιο A δείχνουν ότι η παρουσία του τελευταίου συμβάλλει στην αυξημένη σταθερότητα του συμπλόκου, γεγονός το οποίο μπορεί να οφείλεται τόσο στη μειωμένη πλευρική διάχυση των στοιχείων της Εξωτερικής Μεμβράνης όσο και στις αλληλεπιδράσεις ανάμεσα στα β-βαρέλια και το Λιπίδιο A. Επιπρόσθετα, η διαφορά του PMF

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

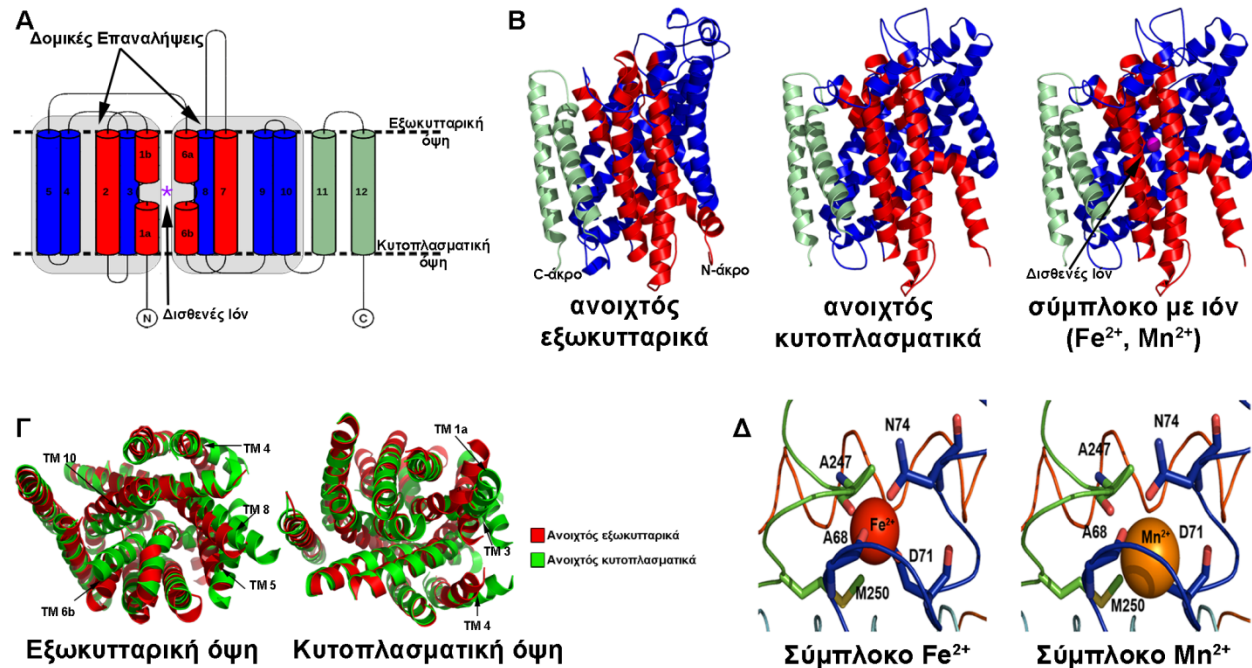
ανάμεσα στο Λιπίδιο A και τους λιποπολυσακχαρίτες RaMP και ReMP υποδεικνύει περαιτέρω ότι η προσθήκη του ολιγοσακχαριτικού πυρήνα είναι σημαντική για τη σταθερότητα του συμπλόκου, πιθανώς λόγω των αλληλεπιδράσεων ανάμεσα στα σάκχαρα των λιποπολυσακχαριτών με τους βρόχους της OmpF, όπως παρουσιάστηκε και στην Ενότητα 3.2.1. Ωστόσο, η σχετικά μικρή διαφορά ανάμεσα στις τιμές του PMF για τους ReMP και RaMP δείχνει ότι οι διαφορές των δύο μορίων στο μέγεθος του ολιγοσακχαριτικού πυρήνα συμβάλλουν σε μικρότερο βαθμό στη σταθερότητα του συμπλόκου.

Συνολικά, τα παραπάνω αποτελέσματα δείχνουν ότι τα διαμεμβρανικά β-βαρέλια που σχηματίζουν τριμερή ακολουθούν μια συγκεκριμένη, κοινή αρχιτεκτονική στις αλληλεπιδράσεις ολιγομερισμού. Η επιφάνεια επαφής φαίνεται να αποτελείται από συγκεκριμένα διαμεμβρανικά τμήματα των βαρελιών (1-5 και 15-16), ενώ τυχόν επιπρόσθετες αλληλεπιδράσεις πιθανότατα σχετίζονται με την λειτουργική εξειδίκευση κάθε β-βαρελιού. Τέλος, μέσω των υπολογισμών του PMF φάνηκε ότι η παρουσία λιποπολυσακχαριτών στην Εξωτερική Μembrάνη είναι σημαντική για τη σταθερότητα των αλληλεπιδράσεων ολιγομερισμού στην OmpF. Δεδομένου ότι η αρχιτεκτονική του ομοτριμερούς της OmpF φαίνεται να είναι κοινή για την υπεροικογένεια των τριμερών πορινών, είναι πιθανό η επίδραση των λιποπολυσακχαριτών στη σταθερότητα των αλληλεπιδράσεων ολιγομερισμού να μην περιορίζεται για το παρόν σύστημα μελέτης αλλά να ισχύει για το σύνολό των τριμερών β-βαρελιών.

3.3 Ο μεταφορέας δισθενών μεταλλικών ιόντων SLC11A1

Αντικείμενο της παρούσας ενότητας ήταν η μελέτη της δομής και της λειτουργίας του SLC11A1, ενός διαμεμβρανικού μεταφορέα δισθενών μεταλλικών ιόντων ο οποίος ανήκει στην οικογένεια των μεταφορέων SLC11. Για το σκοπό αυτό κατασκευάστηκαν αρχικά μοντέλα της τρισδιάστατης δομής του μεταφορέα, χρησιμοποιώντας ως πρότυπο τις κρυσταλλικές δομές μιας σειράς από βακτηριακά ομόλογα της οικογένειας των μεταφορέων (Bozzi et al., 2016a; Ehrnstorfer et al., 2014; Ehrnstorfer et al., 2017). Τα μοντέλα που κατασκευάστηκαν προτυποποιούν τρεις διακριτές στερεοδιατάξεις, τον ανοιχτό εξωκυτταρικά μεταφορέα (outward-facing), τον ανοιχτό κυτοπλασματικά (inward-facing) και τα σύμπλοκα του μεταφορέα με προσδεδεμένο ιόν Fe^{2+} ή Mn^{2+} (ion-bound). Η δομή του SLC11A1, όπως προκύπτει από τα μοντέλα που κατασκευάστηκαν, διαθέτει 12 διαμεμβρανικά τμήματα με δομή α -έλικας, με N- και C-άκρο στον κυτοπλασματικό χώρο (Εικόνα 65). Τα πρώτα 10 από τα 12 διαμεμβρανικά τμήματα υιοθετούν τη συντηρημένη δομή της αυτοτελούς δομικής περιοχής των NRAMPs, η οποία παρουσιάζει ομοιότητες με το δίπλωμα των βακτηριακών μεταφορέων LeuT (Krishnamurthy & Gouaux, 2012). Συγκεκριμένα, η δομική περιοχή περιλαμβάνει δύο δομικές επαναλήψεις αποτελούμενες από τα διαμεμβρανικά τμήματα TM 1-5 και TM 6-10, αντίστοιχα, με τη δεύτερη επανάληψη να είναι ανεστραμμένη σε σχέση με την πρώτη. Τα τελευταία δύο διαμεμβρανικά τμήματα, TM11 και TM12, δεν ανήκουν στην περιοχή NRAMP και εμφανίζονται κυρίως στα ευκαρυωτικά μέλη της οικογένειας των NRAMPs, ενώ απουσιάζουν από κάποια βακτηριακά μέλη της. Στην περιοχή NRAMP, τα διαμεμβρανικά τμήματα TM1 και TM6 διαθέτουν δυο περιοχές με ακανόνιστη δομή στο μέσο του υδρόφοβου πυρήνα, οι οποίες χωρίζουν το κάθε τμήμα σε δύο μικρότερες α -έλικες (1a και 1b για το TM1, 6a και 6b για το TM6). Αυτές οι περιοχές είναι πλούσιες σε πολικά και φορτισμένα κατάλοιπα (Asp-71, Asn-74, His-252 κλπ), αποτελούν την περιοχή πρόσδεσης των μεταλλικών ιόντων και συμβάλλουν στην ταυτόχρονη μεταφορά μεταλλικών ιόντων και πρωτονίων (Εικόνα 65).

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 65 Η προτεινόμενη δομή του μεταφορέα SLC11A1. (Α) Διάγραμμα της τοπολογίας του μεταφορέα στη μεμβράνη. Οι δύο δομικές επαναλήψεις της περιοχής NRAMP σημαίνονται με δύο πλαίσια. Με κόκκινο χρώμα σημαίνονται τα διαμεμβρανικά τμήματα TM1a, TM1b, TM2, TM6a, TM6b και TM7, τα οποία σχηματίζουν το «δεμάτιο» (bundle) και συμμετέχουν στην πρόσδεση και τη μεταφορά ιόντων, ενώ με μπλε χρώμα τα διαμεμβρανικά τμήματα TM3, TM4, TM5, TM8, TM9 και TM10, τα οποία σχηματίζουν το «ικρίωμα» (scaffold). Τα επιπρόσθετα διαμεμβρανικά τμήματα TM11 και TM12, τα οποία δεν συμμετέχουν στην περιοχή NRAMP, χρωματίζονται με γκριζό χρώμα. Η θέση πρόσδεσης του ιόντος σημαίνεται με ένα μωβ αστερίσκο. (Β) Η τρισδιάστατη δομή του μεταφορέα σε τρεις διαφορετικές στερεοδιατάξεις. Τα στοιχεία της δομής χρωματίζονται με βάση το διάγραμμα της τοπολογίας στο (Α). (Γ) Σύγκριση των δύο στερεοδιατάξεων (ανοιχτός εξωκυτταρικά και κυτοπλασματικά) του SLC11A1 μέσω δομικής υπέρθεσης. (Δ) Απεικόνιση της περιοχής πρόσδεσης ιόντων στα σύνπλοκα του μεταφορέα με ιόν Fe^{2+} ή Mn^{2+} μετά από 150 ns προσομοιώσεων Μοριακής Δυναμικής.

Παρότι και οι τρεις στερεοδιατάξεις του μεταφορέα ακολουθούν την ίδια τοπολογία και τρισδιάστατη οργάνωση των διαμεμβρανικών τμημάτων τους, παρουσιάζουν σημαντικές διαφορές μεταξύ τους. Οι πιο εκτεταμένες διαφορές παρατηρούνται ανάμεσα στις δομές του ανοιχτού εξωκυτταρικά και του ανοιχτού κυτοπλασματικά μεταφορέα, ενώ το σύνπλοκο SLC11A1-ιόν παρουσιάζει σημαντικές ομοιότητες με τον ανοιχτό κυτοπλασματικά μεταφορέα, με τις διαφορές τους να εντοπίζονται κυρίως στην περιοχή πρόσδεσης του ιόντος (βρόχοι TM1a-TM1b και TM6a-TM6b). Η σύγκριση ανάμεσα στον ανοιχτό εξωκυτταρικά και τον ανοιχτό κυτοπλασματικά μεταφορέα αποκαλύπτει την ύπαρξη μιας σειράς από

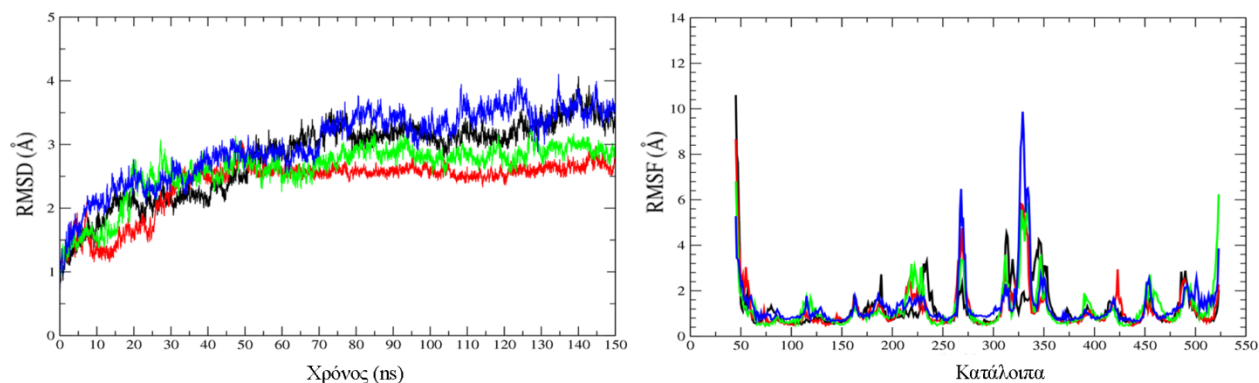
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

στερεοδιαταξικές αλλαγές (Εικόνα 65γ). Συγκεκριμένα, οι πιο εκτεταμένες διαφορές παρουσιάζονται ανάμεσα στα διαμεμβρανικά τμήματα TM1a, TM1b, TM2, TM6a, TM6b και TM7, τα οποία υιοθετούν διαφορετικές θέσεις στα δύο μοντέλα. Αυτές οι διαφορές φαίνονται να σχετίζονται άμεσα με τη λειτουργία του SLC11A1 στις δύο στερεοδιατάξεις. Έτσι, στον ανοιχτό εξωκυτταρικά μεταφορέα παρατηρείται άνοιγμα της δομής στον εξωκυτταρικό χώρο, με τις κινήσεις των τμημάτων TM1b, TM6a, TM7 και του εξωκυτταρικού βρόχου ανάμεσα στα τμήματα TM7 και TM8 (βρόχος 7-8), οι οποίες επιτρέπουν την είσοδο του ιόντος από το εξωτερικό του κυττάρου στη θέση πρόσδεσης του μεταφορέα. Από την άλλη, στον ανοιχτό κυτοπλασματικά μεταφορέα παρατηρείται άνοιγμα της δομής στον κυτοπλασματικό χώρο με τις κινήσεις των τμημάτων TM1a, TM2, TM6b και του βρόχου ανάμεσα στο TM6b και το TM7 (βρόχος 6-7), οι οποίες επιτρέπουν την έξοδο του ιόντος από τη θέση πρόσδεσης προς το κυτταρόπλασμα. Εκτός από τις παραπάνω αλλαγές, διαφορές μικρότερης κλίμακας παρατηρούνται για τα διαμεμβρανικά τμήματα TM4, TM5, TM8 και TM10. Ωστόσο, αυτές οι κινήσεις φαίνονται να είναι συμπληρωματικές ως προς τις κινήσεις των περιοχών που σχηματίζουν τη θέση πρόσδεσης των ιόντων.

Η σταθερότητα της δομής για καθεμία από τις διαφορετικές στερεοδιατάξεις του SLC11A1 ελέγχθηκε μέσα από εκτεταμένες προσομοιώσεις Μοριακής Δυναμικής σε μεμβράνες POPC. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι και οι τρεις στερεοδιατάξεις του μεταφορέα είναι σταθερές οντότητες, διατηρώντας τόσο τη γενική οργάνωση των διαμεμβρανικών τμημάτων τους όσο και τα ξεχωριστά χαρακτηριστικά που τις διακρίνουν μεταξύ τους (Εικόνα 66). Συγκεκριμένα, όπως αποκαλύπτεται από τις μετρήσεις RMSD των προσομοιώσεων, τα μοντέλα και των τριών στερεοδιατάξεων φτάνουν σε κατάσταση ισορροπίας σε σχετικά μικρό χρόνο (~100 ns) και διατηρούνται σε αυτήν την κατάσταση για το υπόλοιπο των προσομοιώσεων. Οι μετρήσεις των τιμών RMSD στην κατάσταση ισορροπίας κυμαίνονται στην περιοχή των 2.0-4.0 Å, υποδεικνύοντας σχετικά περιορισμένες στερεοδιαταξικές αλλαγές σε κάθε μοντέλο. Η ανά κατάλοιπο μελέτη της κινητικότητας, μέσα από μετρήσεις RMSF, δείχνει ότι οι πιο ευκίνητες περιοχές του μεταφορέα είναι περιοχές χωρίς εγγενή δομή, όπως τα N- και C-άκρα και οι εξω- και ενδοκυτταρικοί βρόχοι, ενώ τα διαμεμβρανικά τμήματα έχουν τιμές $RMSF < 2.0$ Å, οι οποίες καταδεικνύουν τη σταθερότητά

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

τους. Η κυριότερη παρατήρηση είναι πως σε καμία από τις προσομοιώσεις δεν παρατηρήθηκε μετάβαση του μεταφορέα από τη μία στερεοδιάταξη στην άλλη (π.χ. από outward-facing σε inward-facing ή το ανάποδο). Αυτή η παρατήρηση δείχνει ότι τα μοντέλα των ξεχωριστών στερεοδιατάξεων είναι θερμοδυναμικά σταθερές και προσδίδει επιπλέον εγκυρότητα στα συμπεράσματα που προκύπτουν από τη δομική ανάλυσή τους.



Εικόνα 66 Προσομοιώσεις Μοριακής Δυναμικής του SLC11A1. Μετρήσεις RMSD (αριστερά) και RMSF (δεξιά) για τα μοντέλα του μεταφορέα. Οι μετρήσεις για τις διαφορετικές στερεοδιατάξεις του μεταφορέα (ανοιχτός εξωκυτταρικά, ανοιχτός κυτοπλασματικά, σύμπλοκο Fe^{2+} , σύμπλοκο Mn^{2+}) παρουσιάζονται με μαύρο, κόκκινο, μπλε και πράσινο χρώμα, αντίστοιχα.

Πίνακας 14 Αποτελέσματα ενεργειακών μετρήσεων (FEP) για τον μεταφορέα SLC11A1 με ιόντα Fe^{2+} και Mn^{2+} .

Σύστημα μεταφορέα SLC11A1	Σύμπλοκο με Fe^{2+}	Σύμπλοκο με Mn^{2+}
Φυσικός Τύπος	-13.22 ± 2.40	-22.21 ± 2.08
Ala-68-Gly	-7.17 ± 2.31	-20.63 ± 1.78
Met-250-Ala	-19.85 ± 1.74	-19.87 ± 1.98

Τα δομικά και θερμοδυναμικά χαρακτηριστικά του SLC11A1 μελετήθηκαν επιπρόσθετα μέσα από ενεργειακούς υπολογισμούς για την ισχύ των αλληλεπιδράσεών του με τα δισθενή ιόντα Fe^{2+} και Mn^{2+} (Πίνακας 14). Οι ενεργειακοί υπολογισμοί πραγματοποιήθηκαν με τη μεθοδολογία της Διαταραχής Ελεύθερης Ενέργειας (FEP) στα σύμπλοκα του SLC11A1 με Fe^{2+} και Mn^{2+} , όπως αυτά προέκυψαν από τις προσομοιώσεις Μοριακής Δυναμικής. Σε αυτά τα σύμπλοκα, το ιόν (Fe^{2+} ή Mn^{2+}) βρίσκεται στη θέση πρόσδεσης που σχηματίζεται από τις μη

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

δομημένες περιοχές TM1a-TM1b και TM6a-TM6b και αλληλεπιδρά με τα κατάλοιπα Ala-68, Asp-71, Asn-74, Ala-247 και Met-250. Το τελευταίο κατάλοιπο είναι ιδιαίτερα σημαντικό, καθώς αντίστοιχα, συντηρημένα κατάλοιπα μεθειονίνης έχουν βρεθεί να σχηματίζουν επαφές με δισθενή ιόντα τόσο σε βακτηριακά ομόλογα του SLC11A1 (Bozzi et al., 2016b) όσο και στον συγγενικό μεταφορέα SLC11A2/DMT1, ενώ μεταλλαγές του καταλοίπου από μεθειονίνη σε αλανίνη οδηγούν σε απώλεια της επιλεκτικότητας των μεταφορέων προς το υπόστρωμά τους (Bozzi et al., 2016a; Bozzi et al., 2016b). Τα αποτελέσματα των μετρήσεων FEP έδειξαν ότι τα σύμπλοκα SLC11A1-Mn²⁺ στο φυσικό τύπο του μεταφορέα παρουσιάζουν πιο ευνοϊκές τιμές ΔG_{BIND} σε σχέση με τα σύμπλοκα SLC11A1-Fe²⁺, οι οποίες υποδεικνύουν ότι η πρόσδεση ιόντων Mn²⁺ στον μεταφορέα είναι θερμοδυναμικά πιο ευνοϊκή (Πίνακας 14). Αυτά τα αποτελέσματα είναι σύμφωνα με προηγούμενα πειραματικά δεδομένα, σύμφωνα με τα οποία ο SLC11A1 παρουσιάζει μεγαλύτερη συγγένεια για τα ιόντα Mn²⁺ σε σχέση με τα ιόντα Fe²⁺.

Προκειμένου να διερευνηθεί περαιτέρω η δομική φύση των αλληλεπιδράσεων, κατασκευάστηκαν μοντέλα του μεταφορέα με μεταλλαγές σε κατάλοιπα που συμμετέχουν στην πρόσδεση των ιόντων. Συγκεκριμένα, κατασκευάστηκαν μοντέλα με τις μεταλλαγές Ala-68-Gly, Asp-71-Ala και Met-250-Ala. Η μεταλλαγή της Ala-68 σε γλυκίνη πραγματοποιήθηκε καθώς το συγκεκριμένο κατάλοιπο εμφανίζεται στην αντίστοιχη θέση σε βακτηριακά ομόλογα του SLC11A1 (Ehrnstorfer et al., 2014; Ehrnstorfer et al., 2017), ενώ οι μεταλλαγές των Asp-71 και Met-250 σε αλανίνη πραγματοποιήθηκαν με γνώμονα προηγούμενα πειραματικά αποτελέσματα από μελέτες μεταλλαξιγένεσης στις αντίστοιχες θέσεις του μεταφορέα SLC11A2/DMT1 (Bozzi et al., 2016a). Τα μοντέλα των μεταλλαγών χρησιμοποιήθηκαν για την κατασκευή συμπλόκων με ιόντα Fe²⁺ και Mn²⁺, τα οποία υποβλήθηκαν σε προσομοιώσεις Μοριακής Δυναμικής για τον έλεγχο της σταθερότητάς τους και σε ενεργειακούς υπολογισμούς FEP για την εκτίμηση της ισχύος των αλληλεπιδράσεών τους με ιόντα. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι τα σύμπλοκα των μεταλλαγών Ala-68-Gly και Met-250-Ala διατήρησαν το ιόν Fe²⁺/Mn²⁺ στη θέση πρόσδεσής του, και τα σύμπλοκα αυτά χρησιμοποιήθηκαν στη συνέχεια σε ενεργειακούς υπολογισμούς FEP. Αντίθετα, στα σύμπλοκα της μεταλλαγής Asp-71 παρατηρήθηκε ο αποχωρισμός και η έξοδος του ιόντος από τη θέση πρόσδεσης του μεταφορέα προς στο διαλύτη. Πιο ξεκάθαρη εικόνα για την επίδραση των

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

μεταλλαγών έδωσαν τα αποτελέσματα των υπολογισμών FEP για τις μεταλλαγές Ala-68-Gly και Met-250-Ala (Πίνακας 14). Συγκεκριμένα, η μεταλλαγή της Ala-68 φαίνεται να έχει μικρή επίδραση στην ικανότητα του μεταφορέα να προσδένει ιόντα, ενώ δεν φαίνεται να επηρεάζει τη διαφορά συγγένειας των Mn^{2+} και Fe^{2+} . Αντίθετα, η μεταλλαγή Met-250-Ala έδωσε σχεδόν πανομοιότυπες τιμές ΔG_{BIND} για τα δύο ιόντα, υποδεικνύοντας ότι η αντικατάσταση της μεθειονίνης από αλανίνη επηρεάζει σημαντικά την ικανότητα του SLC11A1 να ξεχωρίζει τα δύο ιόντα. Στο σύνολό τους, τα παραπάνω αποτελέσματα υποδεικνύουν ότι τα πιο σημαντικά κατάλοιπα για την αναγνώριση και πρόσδεση των ιόντων στη θέση πρόσδεσης του SLC11A1 είναι τα κατάλοιπα Asp-71 και Met-250. Η παρουσία του ασπαρτικού στη θέση 71 φαίνεται να είναι απαραίτητη για το σχηματισμό και τη σταθερότητα του συμπλόκου SLC11A1-ιόντος, καθώς η αρνητικά φορτισμένη πλευρική ομάδα του συμμετέχει στην αγκυροβόληση των θετικά φορτισμένων κατιόντων στη θέση πρόσδεσης. Ανάλογο ρόλο θα μπορούσε να επιτελεί και το πολικό κατάλοιπο Asn-74, χωρίς ωστόσο να έχει επιβεβαιωθεί μέχρι σήμερα η σημασία του. Από την άλλη, η παρουσία της μεθειονίνης στη θέση πρόσδεσης των ιόντων, ενώ δεν είναι απαραίτητη για την πρόσδεση του ιόντος, φαίνεται να καθορίζει την επιλεκτικότητα του μεταφορέα απέναντι στο υπόστρωμά του.

Στο σύνολό τους, οι παραπάνω παρατηρήσεις είναι σύμφωνες τόσο με όσα είναι γνωστά για τη λειτουργία των βακτηριακών ομολόγων των NRAMPs, όσο και με όσα είναι γνωστά για τους αμινοξικούς μεταφορείς LeuT. Στους τελευταίους, η δομή των μεταφορέων οργανώνεται σε δύο περιοχές, ένα «δεμάτιο» (bundle), το οποίο περιλαμβάνει τη θέση πρόσδεσης του υποστρώματος και παρουσιάζει εκτεταμένες κινήσεις κατά τη μετάβαση του μεταφορέα από τη μία στερεοδιάταξη στην άλλη, και ένα «ικρίωμα» (scaffold), το οποίο δομείται από τα υπόλοιπα διαμεμβρανικά τμήματα γύρω από την περιοχή-δεμάτιο και ελέγχει την κινητικότητα της τελευταίας (Krishnamurthy & Gouaux, 2012). Οι δομικές διαφορές που παρατηρούνται στις outward-facing και inward-facing στερεοδιατάξεις του SLC11A1, σε συνδυασμό με τη δομική ομολογία των NRAMPs με τους μεταφορείς LeuT, υποδεικνύουν ότι και ο SLC11A1 ίσως ακολουθεί παρόμοια οργάνωση. Έτσι, με βάση τα αποτελέσματα από τη σύγκριση των μοντέλων και τις προσομοιώσεις τους, προτείνεται ότι τα διαμεμβρανικά

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

τμήματα TM1a, TM1b, TM2, TM6a, TM6b και TM7 σχηματίζουν την περιοχή-δεμάτιο του μεταφορέα SLC11A1 και εμπλέκονται άμεσα στη μεταφορά του υποστρώματος (ιόντα Fe^{2+} ή Mn^{2+}) από τον εξωκυτταρικό χώρο στο κυτταρόπλασμα, ενώ τα υπόλοιπα διαμεμβρανικά τμήματα της περιοχής NRAMP σχηματίζουν την περιοχή-ικρίωμα, η οποία ρυθμίζει την κινητικότητα της περιοχής-δεματίου. Τέλος, με βάση τα αποτελέσματα από τους ενεργειακούς υπολογισμούς FEP, φαίνεται ότι τα κατάλοιπα Asp-71 και Met-250 διαδραματίζουν ιδιαίτερα σημαντικό ρόλο στην περιοχή πρόσδεσης του υποστρώματος του μεταφορέα.

3.4 Υποδοχείς με ενεργότητα τυροσινικής κινάσης (RTKs)

3.4.1 Ο αλγόριθμος πρόγνωσης RTK-PRED

Για τη μελέτη των υποδοχέων με ενεργότητα τυροσινικής κινάσης (RTKs) αναπτύχθηκε, στα πλαίσια της παρούσας διδακτορικής διατριβής, η μέθοδος RTK-PRED, ένα εργαλείο για τον αυτοματοποιημένο εντοπισμό RTKs από την αλληλουχία τους, το σχολιασμό της τοπολογίας τους και την ταξινόμησή τους σε υποοικογένειες. Η απόδοση της μεθόδου αξιολογήθηκε σε τρία στάδια. Στο πρώτο στάδιο πραγματοποιήθηκε ο έλεγχος της απόδοσης του pHMM για τον ειδικό εντοπισμό των PTKs και η σύγκρισή της με την αντίστοιχη απόδοση του γενικού pHMM για την περιοχή τυροσινικής κινάσης, το οποίο υπάρχει διαθέσιμο στη βάση δεδομένων Pfam. Τα αποτελέσματα της αξιολόγησης (Πίνακας 15) δείχνουν ότι το pHMM της μεθόδου RTK-PRED παρουσιάζει σημαντικά καλύτερες τιμές ειδικότητας, ακρίβειας και συντελεστή MCC (98%, 98% και 0.92) σε σύγκριση με το pHMM της Pfam (83%, 85% και 0.58), γεγονός που οφείλεται κυρίως στο μικρότερο αριθμό ψευδώς θετικών αποτελεσμάτων (12 FPs, έναντι 110 FPs του pHMM της Pfam). Στο σύνολό τους, αυτά τα αποτελέσματα υποδεικνύουν ότι το pHMM της μεθόδου είναι πιο ικανό στο σωστό εντοπισμό PTKs από ότι το προφίλ της Pfam.

Πίνακας 15 Αξιολόγηση της μεθόδου RTK-PRED ως προς τον εντοπισμό RTKs σε ένα σύνολο ελέγχου 716 πρωτεϊνών. Παρουσιάζονται η αξιολόγηση του ειδικού για τυροσινικές κινάσες pHMM (PTK) και η σύγκριση με το αντίστοιχο προφίλ της Pfam (PF07714) ως προς τον εντοπισμό της περιοχής PTK, καθώς και η αξιολόγηση του συνολικού αλγορίθμου RTK-PRED για τον εντοπισμό RTKs. Η αξιολόγηση πραγματοποιήθηκε σε ένα σύνολο ελέγχου 716 αλληλουχιών (44 RTKs, 32 μη διαμεμβρανικές PTKs, 640 άλλες πρωτεΐνες).

	TP ¹	TN	FP	FN	Sn.	Sp.	Acc.	MCC
<i>Αξιολόγηση των pHMMs ως προς τον εντοπισμό PTKs</i>								
Pfam (PF07714)	76	530	110	0	100.00%	82.81%	84.64%	0.58
PTK pHMM	76	628	12	0	100.00%	98.12%	98.32%	0.92
<i>Αξιολόγηση της μεθόδου RTK-PRED ως προς τον εντοπισμό RTKs</i>								
RTK-PRED	44	670	2	0	100.00%	99.70%	99.72%	0.97

¹ TP: αληθώς θετικά, TN: αληθώς αρνητικά, FP: ψευδώς θετικά, FN: ψευδώς αρνητικά, Sn.: ευαισθησία, Sp.: ειδικότητα, Acc.: ακρίβεια, MCC: Συντελεστής συσχέτισης Matthews.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Στο δεύτερο στάδιο αξιολόγησης πραγματοποιήθηκε ο έλεγχος της απόδοσης της μεθόδου στον σωστό εντοπισμό των RTKs. Τα αποτελέσματα της μεθόδου (Πίνακας 15) δείχνουν ότι η μέθοδος RTK-PRED παρουσιάζει ευαισθησία 100%, με αντίστοιχα υψηλή ειδικότητα (99%), ακρίβεια (99.7%) και συντελεστή συσχέτισης Mathews (0.92). Αυτή η υψηλή απόδοση της μεθόδου οφείλεται τόσο στην υψηλή απόδοση του pHMM για την εύρεση των PTKs όσο και στην υψηλή απόδοση του αλγόριθμου Phobius. Στο τελικό στάδιο αξιολόγησης πραγματοποιήθηκε ο έλεγχος της ικανότητας του RTK-PRED στην ταξινόμηση των RTKs σε υποοικογένειες. Για την αξιολόγηση κατασκευάστηκε ένα σύνολο ελέγχου αποτελούμενο από καλά σχολιασμένους 269 RTKs, προερχόμενους από όλες τις υποοικογένειες των υποδοχέων. Όπως προκύπτει από τα αποτελέσματα, η μέθοδος προβλέπει σωστά την οικογένεια για 266 από τους 269 RTKs, με τιμές ευαισθησίας, ειδικότητας και ακρίβειας 100% για την πλειονότητα των οικογενειών, με ελάχιστες εξαιρέσεις (Πίνακας 16). Συνολικά, τα αποτελέσματα από την αξιολόγηση δείχνουν ότι η μέθοδος RTK-PRED είναι ικανή για τη σωστή αναγνώριση, την ταξινόμηση και το σχολιασμό της τοπολογίας των RTKs από το σύνολο ελέγχου.

Πίνακας 16 Αξιολόγηση της ικανότητας του RTK-PRED στην κατηγοριοποίηση των RTKs στο σύστημα της IUPHAR, με βάση ένα σύνολο ελέγχου 269 καλά σχολιασμένων RTKs.

IUPHAR	Sn.	Sp.	Acc.	IUPHAR	Sn.	Sp.	Acc.
<i>EGFR</i>	100%	100%	100%	<i>MET</i>	100%	100%	100%
<i>INS</i>	100%	100%	100%	<i>TAM</i>	100%	100%	100%
<i>PDGFR</i>	100%	100%	100%	<i>TIE</i>	80.00%	100%	99.60%
<i>VEGFR</i>	100%	100%	100%	<i>EPH</i>	100%	100%	100%
<i>FGFR</i>	100%	100%	100%	<i>RET</i>	100%	100%	100%
<i>CCK4</i>	100%	100%	100%	<i>RYK</i>	100%	100%	100%
<i>NGFR</i>	100%	100%	100%	<i>DDR</i>	100%	100%	100%
<i>ROR</i>	100%	99.60%	99.60%	<i>ROS</i>	100%	99.60%	99.60%
<i>MusK</i>	75.10%	100%	99.60%	<i>ALK</i>	100%	100%	100%

¹Sn.: ευαισθησία, Sp.: ειδικότητα, ACC: ακρίβεια

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

A

RTK-PRED

HOME ABOUT HELP

RTK-PRED is a web tool that uses profile Hidden Markov Models (pHMMs) for Receptor Tyrosine Kinase (RTK) prediction and classification.

```
>query_sequence1
PQNLLINTEGAIKLADFLARAFGVFVRTYTHEVVTLMYRAPEILLGCKYYSTAVDIWS
LGCI FAENVTRRALFPDSEIDQLFRIFRTLGTDFEVWFGVTSMPDYKPSFKWARQD
FSKVVFPLEDDGRLLSQMLHYDPNKRISAKAALAHPPQDVTKPVPHLAL

>query_sequence2
MGSNKSFKDASQRRRSLEPAENVHGAGGGAFFASQTFSPASADGHRGSAFAFAAAA
EFKLFQGFNSSDTVTSFQRAGFLAGGVTFVALYDYESRTETDLSFKKERLQIVNNT
GDWWLAHSLSTGQTGYIPSNYVAPSDSIQAEWYFGKITRRESERLLNNAENPRGTFV
RESEETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDGGGFVITSRTQFNSLQQLVAYYKH
ADGLCHRLTTCVPTSKFQTQGLAKDAWEIPRESLRLEVKLGQGFVWMTWNGTRV
AIKTLKFGTMSFEAFLQEAQVMKKLRHEKLVQLYAVVSEEPYIVYTEYMSKGSLLDFLK
GETGRYLRLPOLVDMQAQIASGMAYVERMNVHRDLRAANILVGENLVCKVADFGLARL
IEDNEYTARQGAFFPKWTAPEALYGRFTIKSDVWSFGILLTELTTGRVVPYGMVNR
```

or upload a fasta file: No file chosen

B

query_sequence1			
Not Tyrosine Kinase			
query_sequence2			
non Receptor Tyrosine Kinase (nRTK)			
Tyrosine Kinase Domain		259	525
query_sequence3			
Receptor Tyrosine Kinase (RTK)			
Signal Peptide		1	24
Extracellular Domain	Recep_L domain(PF01030.23)	57	167
Extracellular Domain	Recep_L domain(PF01030.23)	361	480
Extracellular Domain	Furin-like(PF00757.19)	185	358
Extracellular Domain	Furin-like(PF00757.19)	474	557
Extracellular Domain	Furin-like(PF00757.19)	554	598
Transmembrane		646	667
Tyrosine Kinase Domain		712	978
Classification	EGF receptor subfamily		

Γ

```
>> Query: query_sequence1
Classification: Not Tyrosine Kinase
//
>> Query: query_sequence2
Classification: non Receptor Tyrosine Kinase (nRTK)
Kinase Domain: From:259 To:525 Score:337.0
//
>> Query: query_sequence3
Classification: Receptor Tyrosine Kinase (RTK)
SIGNALP: From:1 To:24
EC Domains(PF01030.23): From:57 To:167 Recep_L_domain Score: 104.6
EC Domains(PF01030.23): From:361 To:480 Recep_L_domain Score: 97.4
EC Domains(PF00757.19): From:185 To:358 Furin-like Score: 101.3
EC Domains(PF00757.19): From:474 To:557 Furin-like Score: 23.8
EC Domains(PF00757.19): From:554 To:598 Furin-like Score: 9.5
TRANSMEM: From:646 To:667
Kinase Domain: From:712 To:978 Score:288.7
RTK subfamily: Type 1 (EGF receptor subfamily)
//
```

Εικόνα 67 Η ιστοσελίδα της μεθόδου RTK-PRED. Παρουσιάζονται η φόρμα εισόδου της μεθόδου (Α), στην οποία τα πλαίσια εισαγωγής σημαίνονται με κόκκινο χρώμα, καθώς και τα αποτελέσματα σε μορφή γραφικής απεικόνισης (Β) και αρχείου κειμένου (Γ).

Η μέθοδος RTK-PRED είναι ελεύθερα διαθέσιμη σαν διαδικτυακή διεπαφή στη διεύθυνση <http://bioinformatics.biol.uoa.gr/RTK-PRED>. Από την κεντρική σελίδα της μεθόδου (Εικόνα 67) ο χρήστης μπορεί να δώσει σαν είσοδο μία ή περισσότερες πρωτεϊνικές αλληλουχίες σε μορφή FASTA. Σημειώνεται πως η μέθοδος μπορεί να εφαρμοστεί και στο σχολιασμό ολόκληρων πρωτεωμάτων, αποτελούμενων από χιλιάδες ή εκατοντάδες χιλιάδες αλληλουχίες. Τα αποτελέσματα της μεθόδου για κάθε αλληλουχία περιγράφουν αν η πρωτεΐνη είναι τυροσινική κινάση (PTK) ή όχι, καθώς και αν είναι υποδοχέας RTK ή όχι. Στην περίπτωση των RTKs, παρουσιάζονται επίσης η πρόγνωση της τοπολογίας του, συμπεριλαμβανομένης της θέσης του σηματοδοτικού πεπτιδίου, των εξωκυτταρικών περιοχών, του διαμεμβρανικού τμήματος και της περιοχής τυροσινικής κινάσης, ενώ τέλος δίνεται και η ταξινόμηση του υποδοχέα σε μια υποοικογένεια. Τα αποτελέσματα μπορούν να

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

δοθούν είτε σε μορφή γραφικής απεικόνισης είτε σε μορφή αρχείου κειμένου κατάλληλου για ανάγνωση από προγράμματα που χρησιμοποιούν κανονικές εκφράσεις και αναζητήσεις μοτίβων, γραμμένα σε γλώσσες προγραμματισμού όπως η Perl και η Python.

3.4.2 Εντοπισμός RTKs σε ευκαρυωτικά πρωτεώματα

Η μέθοδος RTK-PRED χρησιμοποιήθηκε περαιτέρω στον εντοπισμό και το σχολιασμό πιθανών RTKs σε μια σειρά από ευκαρυωτικά πρωτεώματα (Πίνακες 17 και 18). Συγκεκριμένα, πραγματοποιήθηκε αναζήτηση πιθανών RTKs σε 37 ευκαρυωτικά πρωτεώματα αναφοράς, τα οποία ανασύρθηκαν από τη συλλογή πρωτεωμάτων της UniProt και αντιπροσωπεύουν όλες τις κατηγορίες ευκαρυωτικών οργανισμών (Μετάζωα, Πρωτόζωα, Φυτά και Μύκητες). Η μόνη ομάδα οργανισμών στην οποία δεν βρέθηκε κανένας RTK ήταν οι μύκητες. Αντίθετα, όλες οι υπόλοιπες ομάδες οργανισμών (Μετάζωα, Πρωτόζωα και Φυτά) έδωσαν αποτελέσματα, με τους περισσότερους RTKs να εντοπίζονται, όπως ήταν αναμενόμενο, στα Μετάζωα. Τέλος, σημειώνεται ότι στην περίπτωση των καλά σχολιασμένων ζωικών πρωτεωμάτων *H. sapiens*, *M. musculus* και *R. norvegicus*, προβλέφθηκαν ορθά όλοι οι καλά σχολιασμένοι RTKs που εντάσσονται στη SwissProt, καθώς επίσης και όλες οι εγγραφές TrEMBL των πιθανών RTKs των εν λόγω πρωτεωμάτων.

Η κατάταξη των RTKs σύμφωνα με το σύστημα ταξινόμησης της IUPHAR έδειξε ότι η πλειονότητα των RTKs των Μεταζώων κατατάσσονται στις οικογένειες του συστήματος, με σχετικά μικρό αριθμό (116 από 785) να παραμένουν αταξινομήτες. Στην περίπτωση των τελευταίων, η πλειονότητά τους ήταν είτε θραύσματα υποδοχέων από την TrEMBL είτε RTKs με πολύ μικρό εξωκυτταρικό N-άκρο, οι οποίοι δεν διέθεταν εξωκυτταρικές περιοχές και, κατά συνέπεια, δεν μπορούσαν να ταξινομηθούν.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 17 Αποτελέσματα από την εφαρμογή της μεθόδου RTK-PRED σε 37 πρωτεύματα αναφοράς.

Πρωτέωμα	NCBI Tax. ID	Αριθμός αλληλουχιών	Πιθανοί RTKs
Μετάζωα			
<i>Homo sapiens</i>	9606	73928	112 (0.15%)
<i>Mus musculus</i>	10090	54185	118 (0.22%)
<i>Rattus norvegicus</i>	10116	29951	99 (0.31%)
<i>Bos taurus</i>	9913	23965	60 (0.25%)
<i>Gallus gallus</i>	9031	29474	97 (0.33%)
<i>Drosophila melanogaster</i>	7227	21923	38 (0.17%)
<i>Anopheles gambiae</i>	7165	13515	17 (0.13%)
<i>Caenorhabditis elegans</i>	6239	26898	55 (0.20%)
<i>Danio rerio</i>	7955	46926	150 (0.32%)
<i>Branchiostoma floridae</i>	7739	28544	45 (0.16%)
Πρωτόζωα			
<i>Dictyostelium discoideum</i>	44689	12746	6 (0.05%)
<i>Toxoplasma gondii</i>	432359	8404	0 (0.00%)
<i>Tetrahymena thermophila</i>	312017	26976	0 (0.00%)
<i>Capsaspora owczarzaki</i>	595528	9794	105 (1.07%)
<i>Ectocarpus siliculosus</i>	2880	16334	1 (0.01%)
<i>Naegleria gruberi</i>	5762	15636	5 (0.03%)
<i>Plasmodium falciparum</i>	36329	5449	0 (0.00%)
Φυτά			
<i>Arabidopsis thaliana</i>	3702	39380	81 (0.21%)
<i>Zea mays</i>	4577	99374	43 (0.04%)
<i>Glycine max</i>	3847	75672	217 (0.29%)
<i>Solanum lycopersicum</i>	4081	33952	47 (0.13%)
<i>Volvox carteri</i>	3068	14335	0 (0.00%)
<i>Vitis vinifera</i>	29760	29907	72 (0.24%)
<i>Oryza sativa subsp. japonica</i>	39947	48914	41 (0.08%)
<i>Nicotiana tabacum</i>	4097	73605	107 (0.14%)
<i>Triticum aestivum</i>	4565	130673	249 (0.19%)
<i>Rosa chinensis</i>	74649	45026	113 (0.25%)
Μύκητες			
<i>Saccharomyces cerevisiae</i>	559292	6049	0 (0.00%)
<i>Emericella nidulans (Aspergillus nidulans)</i> ¹	227321	10557	0 (0.00%)
<i>Neosartorya fumigata (Aspergillus fumigatus)</i> ¹	330879	9647	0 (0.00%)
<i>Aspergillus flavus</i>	332952	13500	0 (0.00%)
<i>Candida albicans</i>	237561	6035	0 (0.00%)
<i>Neurospora crassa</i>	367110	10258	0 (0.00%)
<i>Beauveria bassiana</i>	655819	10363	0 (0.00%)
<i>Penicillium digitatum</i>	1170229	9101	0 (0.00%)
<i>Trichoderma asperellum</i>	1042311	12547	0 (0.00%)
<i>Schizosaccharomyces pombe</i>	284812	5142	0 (0.00%)

¹ Σε παρένθεση δίνεται το παλαιότερο όνομα του είδους, το οποίο εμφανίζεται συχνά στη βιβλιογραφία.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 18 Λειτουργική κατηγοριοποίηση των πιθανών RTKs από τα πρωτεώματα αναφοράς.

IUPHAR	Μετάζωα	Πρωτόζωα	Φυτά	IUPHAR	Μετάζωα	Πρωτόζωα	Φυτά
<i>EGFR</i>	51	0	0	<i>MET</i>	18	0	0
<i>INS</i>	38	0	0	<i>TAM</i>	28	0	0
<i>PDGFR</i>	49	0	0	<i>TIE</i>	18	0	0
<i>VEGFR</i>	33	0	0	<i>EPH</i>	141	0	0
<i>FGFR</i>	96	0	0	<i>RET</i>	12	0	0
<i>CCK4</i>	10	0	0	<i>RYK</i>	17	0	0
<i>NGFR/TRK</i>	35	48	164	<i>DDR</i>	30	0	0
<i>ROR</i>	31	0	0	<i>ROS</i>	37	5	0
<i>MusK</i>	12	0	0	<i>ALK</i>	13	1	0
				Άλλοι RTKs ¹	116	64	806

¹ Στην κατηγορία «άλλοι RTKs» εντάσσονται όσοι υποδοχείς δεν έχουν σχολιαστεί λειτουργικά.

Αντίθετα από τα Μετάζωα, οι RTKs των Φυτών και των Πρωτοζώων, στην πλειονότητά τους, δεν μπόρεσαν να ταξινομηθούν σε κάποια οικογένεια με ελάχιστες εξαιρέσεις. Συγκεκριμένα, στην περίπτωση των Φυτών, η μόνη οικογένεια του υπάρχοντος συστήματος που εντοπίστηκε στα πρωτεώματα ήταν η οικογένεια NGFR/TRK, η οποία περιλαμβάνει υποδοχείς νευροτροφινών και τροπομοσινών. Δεδομένου ότι ομόλογες πρωτεΐνες των νευροτροφινών έχουν βρεθεί σε φυτικά πρωτεώματα, είναι πιθανό η ύπαρξη RTKs της συγκεκριμένης οικογένειας να σχετίζεται με τη λειτουργία αυτών των ομόλογων πρωτεϊνών. Αντίστοιχα, μεγάλος αριθμός υποδοχέων NGFR/TRK βρέθηκαν και στα Πρωτόζωα, μαζί με πολύ μικρό αριθμό αλληλουχιών από τις οικογένειες ROS και ALK.

Τα παραπάνω αποτελέσματα σχετίζονται σε μεγάλο βαθμό με τον τρόπο ταξινόμησης των RTKs, ο οποίος βασίζεται κυρίως στον τύπο των εξωκυτταρικών περιοχών και στο είδος του προσδέτη που αναγνωρίζουν οι υποδοχείς. Η πλειονότητα των RTKs, με βάση αυτό το σύστημα, λειτουργούν ως υποδοχείς ορμονών, νευροδιαβιβαστών και αυξητικών παραγόντων, μόρια τα οποία εκφράζονται από διαφορετικούς κυτταρικούς τύπους και μεταφέρονται στο σώμα μέσω του κυκλοφορικού συστήματος, ή συμμετέχουν σε μηχανισμούς όπως η μετάδοση του νευρικού παλμού και η ανοσολογική απόκριση. Έτσι, είναι αναμενόμενο να εντοπίζονται μέλη αυτών των οικογενειών στα Μετάζωα, τα οποία διαθέτουν τέτοιους μηχανισμούς, καθώς και σε κάποια Πρωτόζωα, τα οποία συχνά ζουν συμβιωτικά ή παρασιτικά μέσα σε μετάζωα. Από την άλλη, τα Φυτά και η πλειονότητα των Πρωτοζώων δεν υιοθετούν τους παραπάνω

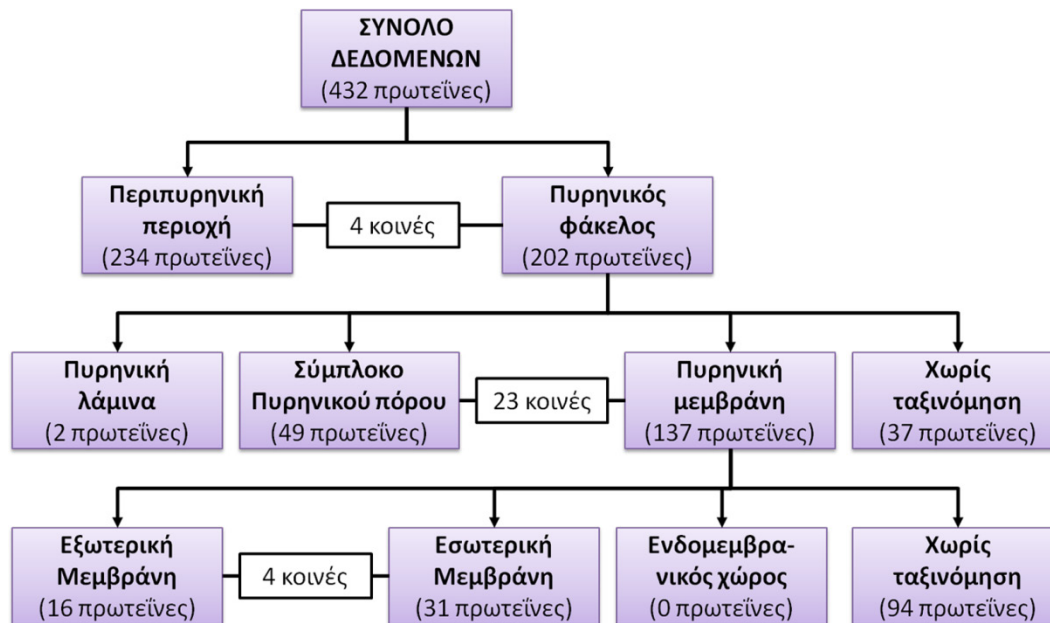
Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

μηχανισμούς λειτουργίας, ενώ συχνά διαθέτουν ειδικά χαρακτηριστικά (π.χ. κυτταρικό τοίχωμα) τα οποία δεν υπάρχουν στα μετάζωα. Αυτό έχει σαν αποτέλεσμα οι υποδοχείς των συγκεκριμένων οργανισμών να αναγνωρίζουν διαφορετικές κατηγορίες μορίων και, κατά συνέπεια, να διαθέτουν διαφορετικές εξωκυτταρικές περιοχές, οι οποίες δεν περιλαμβάνονται στο υπάρχον σύστημα ταξινόμησης. Σε κάθε περίπτωση, τα αποτελέσματα επιβεβαιώνουν την αντίληψη ότι το υπάρχον σύστημα ταξινόμησης των RTKs μπορεί να εφαρμοστεί επιτυχώς μόνο στα Μετάζωα, ενώ για την περίπτωση των άλλων οργανισμών χρειάζεται, πιθανώς, ο σχεδιασμός ενός νέου συστήματος, το οποίο να συμπεριλαμβάνει τις αυτοτελείς δομικές περιοχές αυτών των οργανισμών.

3.5 Πρωτεΐνες του πυρηνικού φακέλου

3.5.1 Κατασκευή δικτύου αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου

Συγκεντρώθηκαν συνολικά 432 καλά σχολιασμένες πρωτεΐνες του ανθρώπου (*Homo sapiens*) με γνωστή υποκυτταρική θέση τους τον πυρηνικό φάκελο. Η ταξινόμησή τους στα διάφορα διαμερίσματα του πυρηνικού φακέλου, σύμφωνα με την οργάνωση των υποκυτταρικών θέσεων όπως αυτή δίνεται από τη UniProt/SwissProt, παρουσιάζεται στην Εικόνα 68. Η ταξινόμηση των υποκυτταρικών θέσεων γίνεται ιεραρχικά, ξεκινώντας από τις γενικές υποκυτταρικές θέσεις «Πυρηνικός Φάκελος» και «Περιπυρηνική Περιοχή». Η θέση «Πυρηνικός Φάκελος» διαιρείται περαιτέρω στις θέσεις «Πυρηνική Μembrάνη», «Σύμπλοκο Πυρηνικού Πόρου», και «Πυρηνική Λάμινα», με τη θέση «Πυρηνική Μembrάνη» να διαιρείται επιπρόσθετα στην «Εξωτερική» και «Εσωτερική Πυρηνική Μembrάνη» και τον «Περιπυρηνικό ενδομεμβρανικό χώρο».



Εικόνα 68 Οι πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου. Ιεραρχική ταξινόμηση του συνόλου των πρωτεϊνών με βάση την υποκυτταρική τοποθεσία τους στον πυρηνικό φάκελο.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Όπως φαίνεται από τα αποτελέσματα (Εικόνα 68), 202 από τις 432 πρωτεΐνες του συνόλου κατατάσσονται στη γενική υποκυτταρική θέση «Πυρηνικός Φάκελος», ενώ 234 πρωτεΐνες κατατάσσονται στη θέση «Περιπυρηνική Περιοχή». Για 4 πρωτεΐνες δίνονται ως υποκυτταρικές θέσεις τόσο η «Περιπυρηνική Περιοχή» όσο και ο «Πυρηνικός Φάκελος». Για 37 από τις 202 πρωτεΐνες του Πυρηνικού Φακέλου δεν είναι γνωστό σε ποιο τμήμα του φακέλου εντοπίζονται, ενώ οι υπόλοιπες ταξινομούνται περαιτέρω στα υποτμήματά του. Αντίστοιχα, στην περίπτωση της θέσης «Πυρηνική Μembrάνη» 94 πρωτεΐνες δεν διαθέτουν περαιτέρω εξειδίκευση, καθώς δεν είναι γνωστό σε ποια από τις δύο μεμβράνες εντοπίζονται, ενώ οι υπόλοιπες 43 διακρίνονται στην Εξωτερική και την Εσωτερική πυρηνική μεμβράνη. Κάποιες πρωτεΐνες φαίνονται να εντοπίζονται σε περισσότερες από μία διακριτές θέσεις, π.χ. ταυτόχρονα στον πυρηνικό φάκελο και στην περιπυρηνική περιοχή, ή ταυτόχρονα στην πυρηνική μεμβράνη και το Σύμπλοκο Πυρηνικού Πόρου. Αυτές χαρακτηρίζονται σαν κοινές πρωτεΐνες των δύο θέσεων, και σημαίνονται αντίστοιχα στην ταξινόμηση της Εικόνας 3.24. Τέλος, πρέπει να σημειωθεί πως για την υποκυτταρική θέση «Περιπυρηνικός Χώρος» δεν βρέθηκε καμία πρωτεΐνη. Αυτό δεν σημαίνει ότι στο συγκεκριμένο διαμέρισμα του πυρηνικού φακέλου δεν εντοπίζονται πρωτεΐνες, αλλά ότι δεν υπάρχουν διαθέσιμα στοιχεία για τις πρωτεΐνες του στο πρωτόμα του ανθρώπου. Οι πρωτεΐνες του συνόλου δεδομένων ταξινομήθηκαν, με βάση την τοπολογία τους στη μεμβράνη, σε διαμεμβρανικές, περιφερειακές και αγκυροβολημένες μεμβρανικές πρωτεΐνες. Από την ταξινόμηση προέκυψε ότι η πλειονότητα των πρωτεϊνών του συνόλου είναι περιφερειακές (285), ακολουθούμενες από τις διαμεμβρανικές (124) και μικρό αριθμό αγκυροβολημένων πρωτεϊνών (23). Στην περίπτωση των διαμεμβρανικών πρωτεϊνών περίπου οι μισές (63) διαθέτουν ένα μόνο διαμεμβρανικό τμήμα, ενώ οι υπόλοιπες 61 μπορεί να διαθέτουν από 2 ως 12 διαμεμβρανικά τμήματα.

Οι πρωτεΐνες του ανθρώπινου πυρηνικού φακέλου χρησιμοποιήθηκαν, στη συνέχεια, για την εύρεση των αλληλεπιδράσεων στις οποίες συμμετέχουν από τη βάση δεδομένων πρωτεϊνικών αλληλεπιδράσεων IntAct. Συγκεντρώθηκαν αρχικά οι αλληλεπιδράσεις ανάμεσα στις πρωτεΐνες του πυρηνικού φακέλου και τους πρώτους γείτονές τους και, στη συνέχεια, οι αλληλεπιδράσεις μεταξύ των ίδιων των πρώτων γειτόνων. Κατά τη συλλογή των

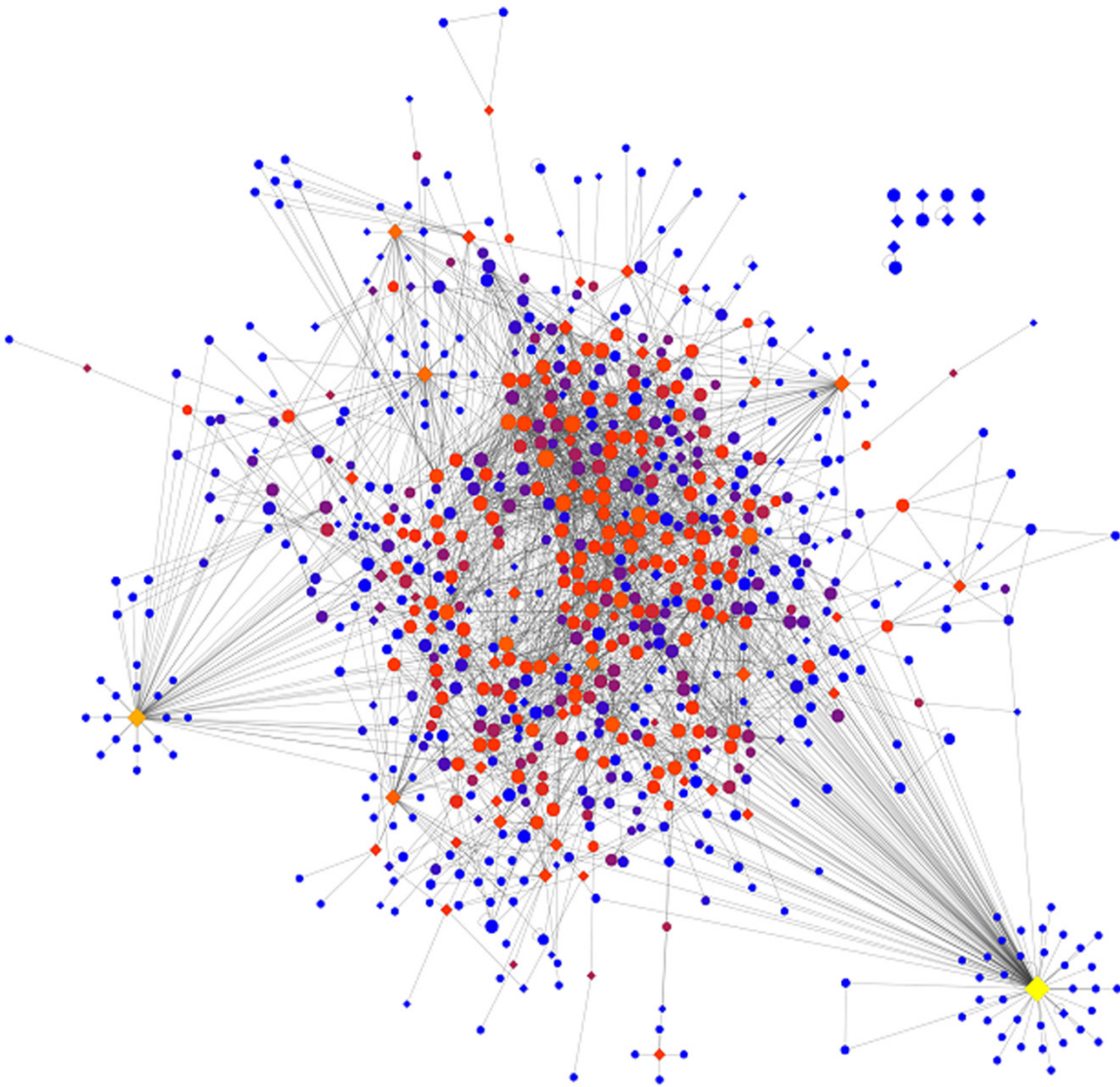
Κεφάλαιο III – Αποτελέσματα-Συζήτηση

αλληλεπιδράσεων εφαρμόστηκε ένα κατώφλι αξιοπιστίας, χρησιμοποιώντας το σύστημα αξιολόγησης των αλληλεπιδράσεων της IntAct. Το τελικό σύνολο δεδομένων που προέκυψε και χρησιμοποιήθηκε για την κατασκευή του δικτύου αλληλεπιδράσεων περιελάμβανε 724 πρωτεΐνες (128 πρωτεΐνες του πυρηνικού φακέλου και 598 πρώτοι γείτονες), οι οποίες συμμετείχαν σε 1986 αλληλεπιδράσεις. Από αυτές, οι 839 αλληλεπιδράσεις ήταν οι επαφές των πρωτεϊνών του πυρηνικού φακέλου είτε μεταξύ τους είτε με πρώτους γείτονες, ενώ οι υπόλοιπες ήταν οι αλληλεπιδράσεις μεταξύ των γειτόνων.

Τα παραπάνω δεδομένα χρησιμοποιήθηκαν στην κατασκευή του δικτύου αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου, το οποίο παρουσιάζεται στην Εικόνα 69. Το δίκτυο αποτελεί έναν μη κατευθυνόμενο γράφο, καθώς τα δεδομένα των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών δεν διαθέτουν κατεύθυνση. Αρχικό βήμα στη μελέτη των χαρακτηριστικών του δικτύου ήταν η ανάλυση των ιδιοτήτων του με βάση τη θεωρία γράφων. Αυτές οι ιδιότητες περιελάμβαναν την πυκνότητα και τη διάμετρο του δικτύου, το μέσο συντελεστή ομαδοποίησης των κόμβων και το χαρακτηριστικό μήκος μονοπατιού. Επιπρόσθετα, προσδιορίστηκε το μοντέλο της κατανομής του βαθμού των κόμβων, σαν μέτρο αξιολόγησης της τοπολογίας του. Όπως προκύπτει από τα αποτελέσματα της ανάλυσης (Πίνακας 19), το δίκτυο αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου υιοθετεί τα τυπικά χαρακτηριστικά των βιολογικών δικτύων αλληλεπιδράσεων. Συγκεκριμένα, το δίκτυο διαθέτει συντελεστή ομαδοποίησης 0.149, η τιμή του οποίου είναι σημαντικά μεγαλύτερη από την τιμή ενός τυχαίου δικτύου με τον ίδιο αριθμό κόμβων ($1/N = 1/724 = 0.001$). Επιπρόσθετα, η πυκνότητα του δικτύου βρέθηκε να είναι αρκετά μικρή (0.007), υποδηλώνοντας ότι το δίκτυο είναι αραιά συνδεδεμένο. Η εξίσωση της κατανομής του δικτύου βρέθηκε να είναι $P(k) = 354.36 * k^{-1.503}$, υποδεικνύοντας ότι το δίκτυο ακολουθεί κατανομή νόμου δύναμης και διαθέτει τοπολογία ανεξάρτητου από κλίμακα (scale free) δικτύου. Η τιμή $\gamma=1.503$ είναι μικρότερη του 2, υποδεικνύοντας ότι ο ρόλος των κεντρικών κόμβων (hubs) είναι σημαντικός στο δίκτυο, καθώς πιθανή αφαίρεσή τους (η οποία, βιολογικά, θα αντιστοιχούσε είτε στη μη έκφραση της πρωτεΐνης ή σε μεταλλαγή που οδηγεί στην απώλεια της λειτουργίας της) μπορεί να οδηγήσει σε κατάρρευση του δικτύου. Στο σύνολό τους, τα παραπάνω αποτελέσματα είναι σύμφωνα με τα χαρακτηριστικά των πραγματικών

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

βιολογικών δικτύων, τα οποία είναι αραιά συνδεδεμένα, διαθέτουν σημαντικά μεγαλύτερες τιμές συντελεστή ομαδοποίησης από αυτές των τυχαίων δικτύων και είναι ανεξάρτητα από κλίμακα, ενώ σημαντικό ρόλο διαδραματίζουν στη συγκρότηση της ενότητάς τους οι κεντρικοί κόμβοι.



Εικόνα 69 Το δίκτυο αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου. Με ρόμβους συμβολίζονται οι πρωτεΐνες που ανήκουν στον πυρηνικό φάκελο, ενώ με κυκλικούς κόμβους συμβολίζονται οι πρωτεΐνες που αλληλεπιδρούν με αυτές του πυρηνικού φακέλου. Το μέγεθος των κόμβων είναι ανάλογο της τιμής του βαθμού (degree) τους, με κόμβους μεγαλύτερου μεγέθους να διαθέτουν μεγαλύτερο βαθμό. Οι κόμβοι χρωματίζονται με μια παλέτα χρωμάτων από το σκούρο μπλε ως το ανοιχτό κίτρινο, με τα ψυχρά και θερμά χρώματα να αντιστοιχούν σε μικρότερες και μεγαλύτερες τιμές ενδιάμεσης κεντρικότητας (betweenness centrality), αντίστοιχα.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 19 Ανάλυση των ιδιοτήτων του δικτύου με βάση τη Θεωρία Γράφων.

Ιδιότητα Δικτύου	Τιμή
Συνδεδεμένα στοιχεία	10
Μέσος συντελεστής ομαδοποίησης	0.149
Χαρακτηριστικό μήκος μονοπατιού	3.984
Διάμετρος δικτύου	9
Πυκνότητα δικτύου	0.007
Μέσος όρος γειτόνων	5.381
Συνάρτηση κατανομής βαθμού $P(k) \sim k^{-\gamma}$	$P(k) = 354.36 * k^{-1.503}$

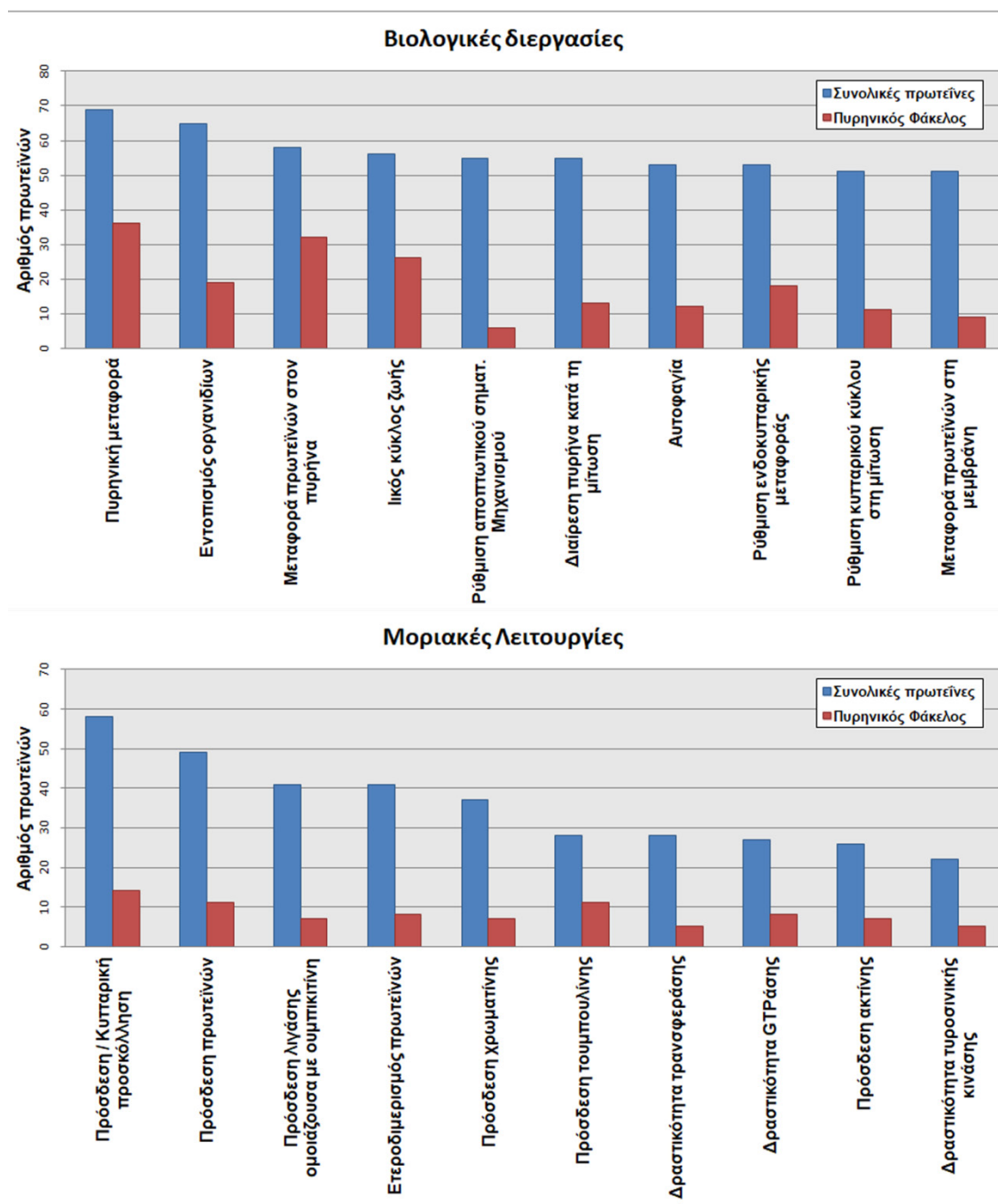
Δύο πολύ σημαντικά χαρακτηριστικά του δικτύου, σύμφωνα με τη θεωρία γράφων, είναι οι κεντρικοί κόμβοι (hubs) και οι στενωποί (bottlenecks). Οι κεντρικοί κόμβοι διαθέτουν το μεγαλύτερο βαθμό κεντρικότητας στο δίκτυο και, όπως αναφέρθηκε παραπάνω, ρυθμίζουν ενεργά τη δομή του δικτύου. Από την άλλη, οι στενωποί είναι κόμβοι με υψηλή ενδιάμεση κεντρικότητα και αποτελούν κρίσιμους συνδέσμους ανάμεσα σε διακριτές υποομάδες των δικτύων, καθώς από αυτούς διέρχεται μεγάλος αριθμός μονοπατιών. Οι κυριότεροι κεντρικοί κόμβοι και στενωποί του δικτύου παρουσιάζονται στον Πίνακα 20. Όπως φαίνεται από τα αποτελέσματα, η πλειονότητά τους περιλαμβάνει πρωτεΐνες του πυρηνικού φακέλου. Επιπρόσθετα, μεγάλος αριθμός των παραπάνω κόμβων είναι ταυτόχρονα κεντρικοί κόμβοι και στενωποί, υποδεικνύοντας ότι διαδραματίζουν εξαιρετικά σημαντικό ρόλο στο δίκτυο. Συγκεκριμένα, οι παραπάνω κόμβοι περιλαμβάνουν πρωτεΐνες όπως ο υποδοχέας EGFR, ένας RTK που ρυθμίζει πολύ μεγάλο αριθμό σηματοδοτικών μονοπατιών και ο οποίος, σε ενεργοποιημένη κατάσταση, μεταφέρεται από την κυτταρική μεμβράνη στον πυρηνικό φάκελο. Δυο άλλες πρωτεΐνες που είναι τόσο κεντρικοί κόμβοι όσο και στενωποί είναι η πρωτεΐνη KASH5 και Νεσπρίνη-4 της εξωτερικής πυρηνικής μεμβράνης, οι οποίες ρυθμίζουν τις επαφές του πυρήνα με τον κυτταροσκελετό. Τέλος, στους κεντρικούς κόμβους και στενωπούς περιλαμβάνεται και η νουκλεοπορίνη Nur62, η οποία συμμετέχει στο NPC. Στο σύνολό τους οι παραπάνω παρατηρήσεις δείχνουν ότι οι πρωτεΐνες του πυρηνικού φακέλου διαδραματίζουν σημαντικό ρόλο στη συνοχή του δικτύου.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 20 Λίστα με τους 10 κυριότερους κεντρικούς κόμβους (hubs) και τους 10 κυριότερους κόμβους-στενωπούς (bottlenecks) του δικτύου. Οι πρωτεΐνες του πυρηνικού φακέλου σημαίνονται με έντονο μαύρο χρώμα.

Κεντρικοί κόμβοι				
AC	Όνομα γονιδίου	Όνομα πρωτεΐνης	Υποκυτταρική θέση	Βαθμός κόμβου
P00533	EGFR	Υποδοχέας αυξητικού παράγοντα EGF	Κυτταρική μεμβράνη, πυρηνική μεμβράνη	124
Q8N6L0	CCDC155	Πρωτεΐνη KASH5	Εξωτερική πυρηνική μεμβράνη	73
P37198	NUP62	Νουκλεοπορίνη p62	Σύμπλοκο Πυρηνικού Πόρου	48
Q8N205	SYNE4	Νεσπρίνη-4	Εξωτερική πυρηνική μεμβράνη	47
P63165	SUMO1	Πρωτεΐνη SUMO1	Πυρηνική μεμβράνη	42
Q6A162	KRT40	Κυτοκερατίνη 40	Κυτταρόπλασμα	42
P14373	TRIM27	Πρωτεΐνη δακτύλων Zn RFP	Πυρήνας και κυτταρόπλασμα	38
Q5JR59	MTUS2	Συζευγμένη με μικροσωληνίσκους πρωτεΐνη MTUS2/TIP150	Κυτταρόπλασμα, κυτταροσκελετός	36
P50402	EMD	Εμερίνη	Εξωτερική και εσωτερική πυρηνική μεμβράνη	31
Q9BRK4	LZTS2	Πρωτεΐνη LZTS2/LAPSER1	Κυτταρόπλασμα	31
Στενωποί				
AC	Όνομα γονιδίου	Όνομα πρωτεΐνης	Υποκυτταρική θέση	Ενδιάμεση κεντρικότητα
P00533	EGFR	Υποδοχέας αυξητικού παράγοντα EGF	Κυτταρική μεμβράνη, πυρηνική μεμβράνη	0.235
Q8N6L0	CCDC155	Πρωτεΐνη KASH5	Εξωτερική πυρηνική μεμβράνη	0.146
Q8N205	SYNE4	Νεσπρίνη-4	Εξωτερική πυρηνική μεμβράνη	0.068
P50402	EMD	Εμερίνη	Εξωτερική και εσωτερική πυρηνική μεμβράνη	0.059
P63165	SUMO1	Πρωτεΐνη SUMO1	Πυρηνική μεμβράνη	0.054
Q12933	TRAF2	Σχετιζόμενος με TNF παράγοντας 2	Κυτταρόπλασμα	0.053
P37198	NUP62	Νουκλεοπορίνη p62	Σύμπλοκο Πυρηνικού Πόρου	0.050
P62993	GRB2	Πρωτεΐνη προσαρμογής GRB2	πυρήνας	0.050
P19320	VCAM1	Πρωτεΐνη προσκόλλησης αγγειώδους επιθηλίου 1	Κυτταρική μεμβράνη	0.048
P14373	TRIM27	Πρωτεΐνη δακτύλων Zn RFP	Πυρήνας και κυτταρόπλασμα	0.036

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

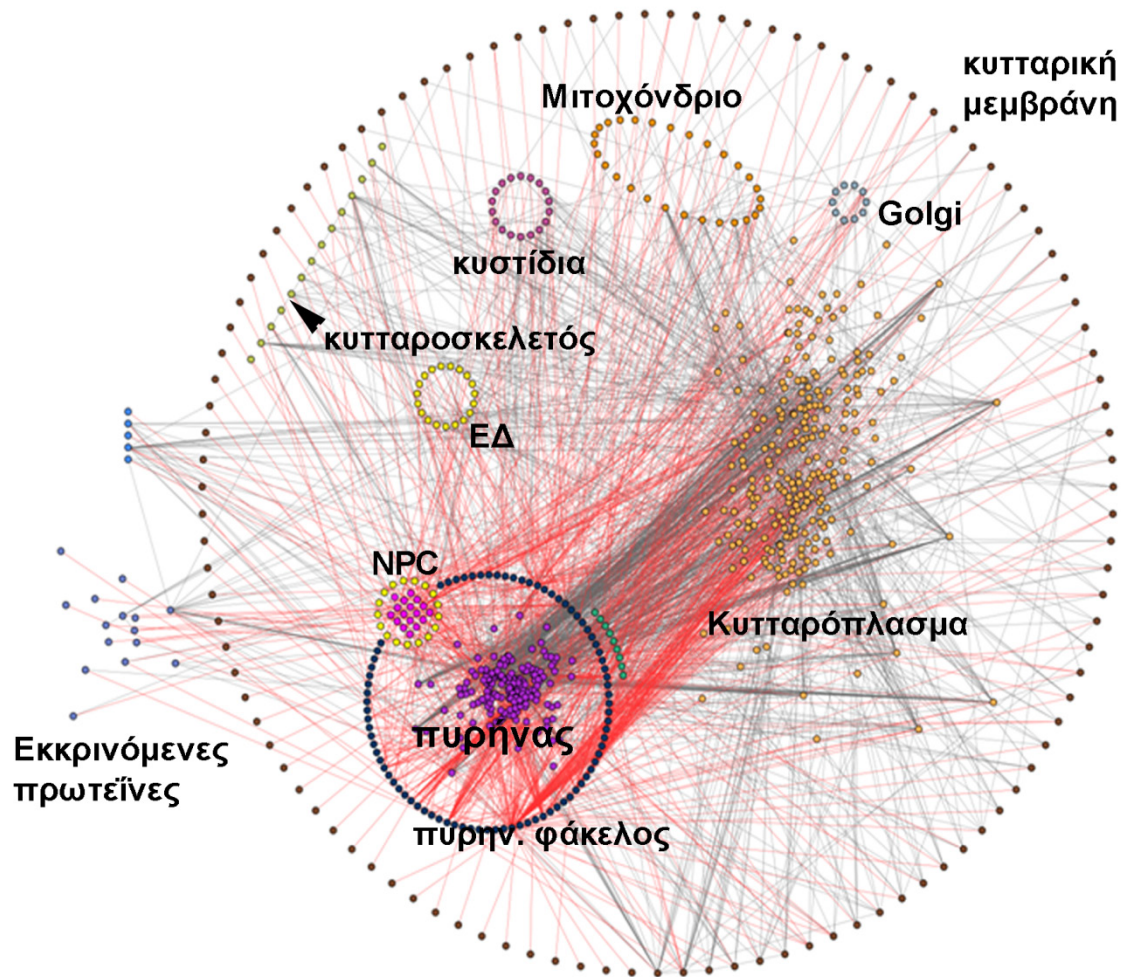


Εικόνα 70 Λειτουργικός σχολιασμός του δικτύου αλληλεπιδράσεων. Παρουσιάζονται οι πιο υπερ-εκπροσωπούμενοι όροι Γονιδιακής Οντολογίας στο δίκτυο, για τις κατηγορίες «Βιολογική Διεργασία» και «Μοριακή Λειτουργία», ενώ δίνεται ο αριθμός των πρωτεϊνών που αντιστοιχούν σε αυτούς τους όρους, τόσο για το συνολικό δίκτυο γενικά όσο και για τον Πυρηνικό Φάκελο ειδικά.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Έχοντας πραγματοποιήσει την υπολογιστική ανάλυση του δικτύου μέσω της θεωρίας γράφων, το επόμενο βήμα ήταν η λειτουργική ανάλυση και ερμηνεία του. Αυτή πραγματοποιήθηκε μέσα από τη μέθοδο του λειτουργικού εμπλουτισμού, χρησιμοποιώντας όρους γονιδιακής οντολογίας από τη βάση δεδομένων Gene Ontology (Ashburner et al., 2000). Στην Gene Ontology οι όροι γονιδιακής οντολογίας χωρίζονται σε τρεις μεγάλες κατηγορίες, τις Βιολογικές Διεργασίες, τις Μοριακές Λειτουργίες και τα Κυτταρικά Συστατικά. Καθώς ο τελευταίος όρος ουσιαστικά είναι ταυτόσημος της υποκυτταρικής τοποθεσίας, η ανάλυση εμπλουτισμού επικεντρώθηκε στις δύο πρώτες κατηγορίες. Τα αποτελέσματα της ανάλυσης (Εικόνα 70) έδειξαν ότι οι πρωτεΐνες του δικτύου εμπλέκονται σε μεγάλο βαθμό σε βιολογικές διεργασίες που σχετίζονται άμεσα με τον πυρήνα όπως η πυρηνική μεταφορά ουσιών αλλά και πρωτεϊνών, η διαίρεση του πυρήνα και η ρύθμιση του κυτταρικού κύκλου, αλλά και σε άλλες διεργασίες όπως η μεταφορά και εντοπισμός των οργανιδίων, η αυτοφαγία, ο έλεγχος της απόπτωσης, η μεταφορά πρωτεϊνών στη μεμβράνη και ο κύκλος ζωής διάφορων ιών. Ως προς τις μοριακές λειτουργίες, οι πιο υπερ-εκπροσωπημένοι όροι περιλαμβάνουν τόσο όρους ενζυμικής ενεργότητας (κινάσες, GTPάσες κλπ) όσο και όρους που αφορούν την πρόσδεση στοιχείων του κυτταροσκελετού (τουμπουλίνη, ακτίνη) αλλά και όρους που σχετίζονται με τις αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών γενικότερα.

Το τελικό κομμάτι της ανάλυσης του δικτύου αλληλεπιδράσεων ήταν η προσπάθεια ομαδοποίησης των στοιχείων του. Αυτή η ομαδοποίηση πραγματοποιήθηκε μέσα από δύο προσεγγίσεις, μια βιολογική προσέγγιση με βάση την υποκυτταρική θέση των πρωτεϊνών και μια υπολογιστική προσέγγιση, χρησιμοποιώντας τον αλγόριθμο MCL. Η ομαδοποίηση του δικτύου με βάση την υποκυτταρική θέση παρουσιάζεται στην Εικόνα 71.



Εικόνα 71 Ομαδοποίηση των στοιχείων του δικτύου με βάση την υποκυτταρική θέση των πρωτεϊνών. Σημαίνονται οι κύριες υποκυτταρικές θέσεις του ευκαρυωτικού κυττάρου: πυρήνας, πυρηνικός φάκελος (ομαδοποίηση όλων των θέσεων του φακέλου εκτός από το Σύμπλοκο Πυρηνικού Πόρου), Σύμπλοκο Πυρηνικού Πόρου (NPC), Ενδοπλασματικό Δίκτυο (ΕΔ), κυτταροσκελετός, σύστημα Golgi, κυτταρικά κυστίδια (λυσosώματα, ενδοσώματα και εξωσώματα κλπ), κυτταρόπλασμα και κυτταρική μεμβράνη. Στην κατηγορία «εκκρινόμενες πρωτεΐνες» εντάσσονται οι πρωτεΐνες που εκκρίνονται από το κύτταρο στο περιβάλλον του.

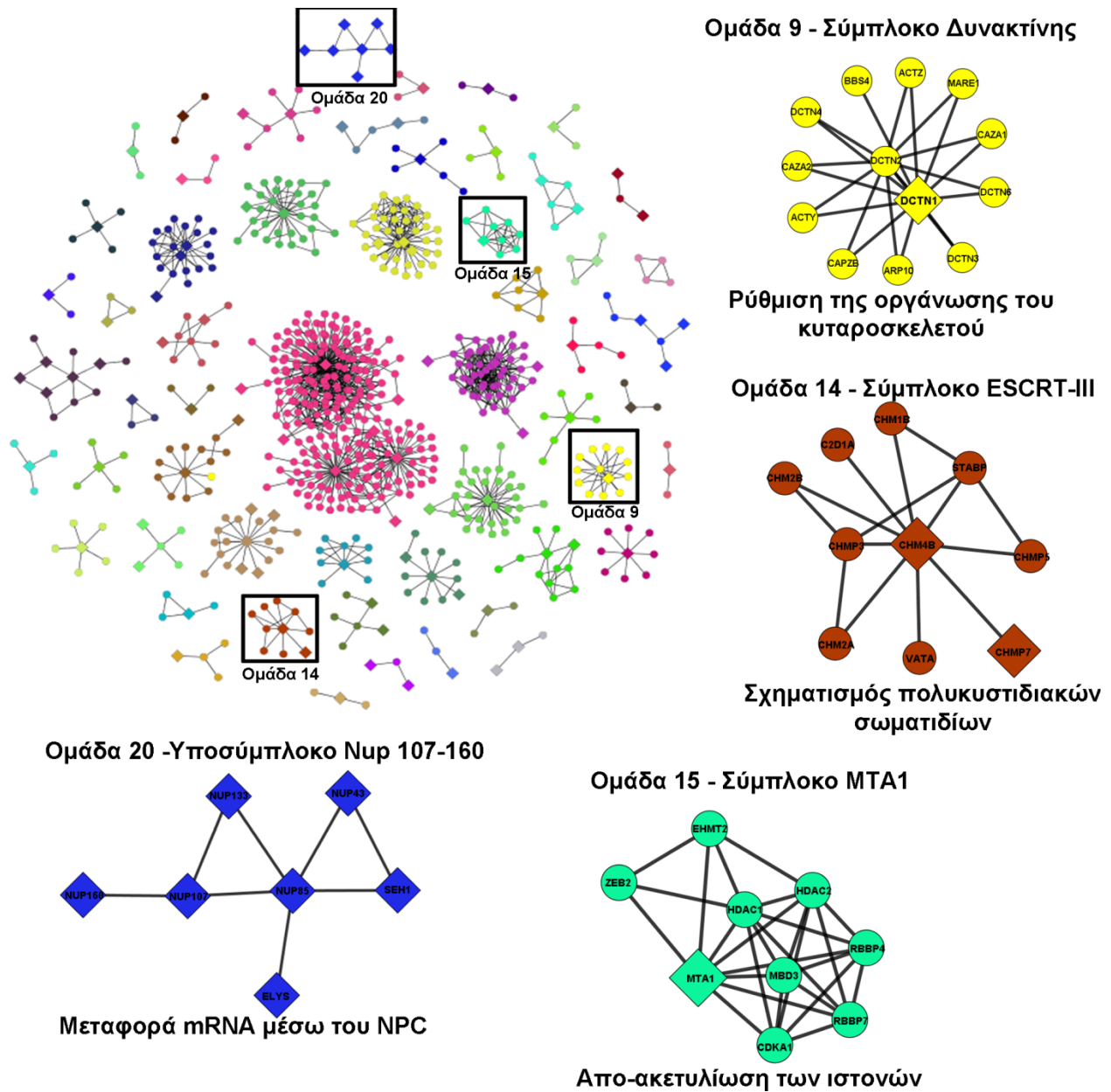
Όπως προκύπτει, οι πρωτεΐνες του πυρηνικού φακέλου αλληλεπιδρούν τόσο μεταξύ τους όσο και με μεγάλο αριθμό πρωτεϊνών σε άλλα διαμερίσματα του κυττάρου. Η πλειονότητα αυτών των πρωτεϊνών εντοπίζονται στο κυτταρόπλασμα, ενώ μεγάλος αριθμός γειτόνων υπάρχουν στο εσωτερικό του πυρήνα και στον κυτταροσκελετό. Και οι τρεις θέσεις έρχονται σε άμεση επαφή με τον πυρηνικό φάκελο, γεγονός που κάνει την παρουσία των πρωτεϊνών σε αυτές αναμενόμενη. Ωστόσο, η ομαδοποίηση αποκάλυψε ότι οι πρωτεΐνες του

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

πυρηνικού φακέλου αλληλεπιδρούν με αρκετούς πρώτους γείτονες σε υποκυτταρικές θέσεις όπως η κυτταρική μεμβράνη και τα μιτοχόνδρια, καθώς και με κάποιες εκκρινόμενες πρωτεΐνες. Ως ένα βαθμό, αυτές οι αλληλεπιδράσεις οφείλονται στις επαφές των πρώτων γειτόνων με στοιχεία του πυρηνικού φακέλου όπως το NPC, το οποίο ρυθμίζει τη μεταφορά πρωτεϊνών από και προς το εσωτερικό του πυρήνα. Ωστόσο, μεγάλος αριθμός αυτών των αλληλεπιδράσεων οφείλεται και σε πρωτεΐνες με πολλαπλές υποκυτταρικές θέσεις, οι οποίες είτε εκφράζονται ταυτόχρονα σε περισσότερες από μία υποκυτταρικές θέσεις (π.χ. GPCRs στα καρδιακά κύτταρα), είτε μεταφέρονται από τον πυρηνικό φάκελο σε άλλα σημεία του κυττάρου και το αντίστροφο, ανάλογα με το στάδιο του κυτταρικού κύκλου (π.χ. RTKs όπως ο EGFR). Σε κάθε περίπτωση, η ομαδοποίηση δείχνει ότι μέσω των αλληλεπιδράσεών τους, οι πρωτεΐνες του πυρηνικού φακέλου μπορούν να επικοινωνήσουν με πολλαπλά υποκυτταρικά διαμερίσματα, γεγονός που αναδεικνύει περαιτέρω τη σημασία του πυρηνικού φακέλου στο συνολικό δίκτυο αλληλεπιδράσεων του κυττάρου και, κατά συνέπεια, στη φυσιολογία και τη λειτουργία του.

Μια δεύτερη προσπάθεια ομαδοποίησης πραγματοποιήθηκε υπολογιστικά, χρησιμοποιώντας τον αλγόριθμο ομαδοποίησης MCL. Η εφαρμογή του MCL οδήγησε στον υπολογισμό 55 ομάδων (clusters) στο δίκτυο (Εικόνα 72). Για κάθε ομάδα του δικτύου πραγματοποιήθηκε ανάλυση λειτουργικού εμπλουτισμού με όρους γονιδιακής οντολογίας, με σκοπό την ανίχνευση στατιστικά σημαντικών λειτουργικών ομοιοτήτων ανάμεσα στις πρωτεΐνες της ομάδας. Τα αποτελέσματα της ανάλυσης οδήγησαν σε λειτουργικό χαρακτηρισμό για 46 από τις 55 ομάδες του δικτύου. Κάποια χαρακτηριστικά παραδείγματα παρουσιάζονται οπτικά στην Εικόνα 72, ενώ τα αποτελέσματα για το σύνολο των ομάδων παρατίθενται στον Πίνακα 21.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση



Εικόνα 72 Ομαδοποίηση των στοιχείων του δικτύου με τον αλγόριθμο MCL. Οι διαφορετικές ομάδες (clusters) του δικτύου εμφανίζονται με διαφορετικό χρώμα. Παραδείγματα ομάδων με λειτουργικό ρόλο σημαίνονται με πλαίσια και παρατίθενται λεπτομερώς. Οι πρωτεΐνες του πυρηνικού φακέλου παρουσιάζονται ως ρόμβοι, ενώ οι υπόλοιπες πρωτεΐνες ως κύκλοι.

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Πίνακας 21 Λειτουργικός σχολιασμός των ομάδων του δικτύου με όρους Γονιδιακής Οντολογίας. Για κάθε ομάδα δίνονται οι πιο σημαντικοί όροι Βιολογικής Διεργασίας και Μοριακής Λειτουργίας, όπως προέκυψαν από την ανάλυση υπερεκπροσώπησης.

Ομάδα	Βιολογικές Διεργασίες	Μοριακές Λειτουργίες
1	1) Σηματοδοτικό μονοπάτι που πυροδοτείται από διαμεμβρανικούς υποδοχείς με δράση τυροσινικής κινάσης (GO-ID:7169) 2) Τροποποιήσεις πρωτεϊνών, με κύρια την φωσφορυλίωση (GO-ID:16310)	1) Ενεργότητα ρύθμισης κινασών (GO-ID: 19887) 2) Ενεργότητα μεταφοράς (GO-ID: 16740)
2	1) Μεταγραφή DNA (GO-ID: 6351) 2) Ρύθμιση της μεταγραφής (GO-ID:6355)	1) Πρόσδεση DNA (GO-ID: 3677) 2) Πρόσδεση πρωτεϊνών του κυτταροσκελετού (GO-ID: 8092)
3	1) Εξαγωγή πρωτεϊνών από τον πυρήνα (GO-ID: 6611) 2) Αποδιοργάνωση του πυρηνικού φακέλου (GO-ID: 51081)	Δομικό συστατικό του πυρηνικού πόρου (GO-ID: 17056)
4	Αρνητική ρύθμιση της μεταγραφής (GO-ID:45892)	Πρόσδεση σε ρυθμιστικές περιοχές της μεταγραφής του DNA (GO-ID: 44212)
5	1) Αρνητική ρύθμιση στον μεταβολισμό των νουκλεϊκών οξέων (GO-ID:45934) 2) Σουμοϋλίωση πρωτεϊνών (GO-ID:16925)	Πρόσδεση στο DNA (GO-ID: 3677) και μεταγραφική ρύθμιση (GO-ID: 140110)
6	1) Ρύθμιση της απόπτωσης (GO-ID:6915) 2) Προγραμματισμένος κυτταρικός θάνατος (GO-ID:12501) 3) Απελευθέρωση κυτοχρώματος c (GO-ID:1836)	1) Πρόσδεση σε περιοχές ομόλογες με Bcl-2 (GO-ID: 51400) 2) Ενεργότητα ετεροδιμερισμού (GO-ID: 46982) και ομοδιμερισμού (GO-ID: 46983)
7	Οργάνωση του πυρήνα (GO-ID: 6997)	Πρόσδεση λαμινών (GO-ID: 5221)
8	1) Κυτταρική προσκόλληση (GO-ID: 7155) 2) Οργάνωση του κυτταροσκελετού ακτίνης (GO-ID: 30036)	1) Πρόσδεση ακτίνης (GO-ID: 3779) 2) Ενεργότητα υποδοχέα ICAM-3 (GO-ID: 30369)
9	1) Αντιγονοπαρουσίαση αντιγόνου MHC II (GO-ID: 2495) 2) Ρύθμιση της οργάνωσης του κυτταροσκελετού (GO-ID: 51493)	Πρόσδεση πρωτεϊνών του κυτταροσκελετού (GO-ID: 8092)
10	1) Οργάνωση κυστιδίων (GO-ID: 16050) 2) Μεταφορά μέσω κυστιδίων (GO-ID: 16192)	Πρόσδεση ασβέστιο-εξαρτώμενων πρωτεϊνών (GO-ID: 48306)
11	Πυρηνοκυτταροπλασματική μεταφορά (GO-IDs: 6913, 51169, 51170, 51168, 17038, 50658)	1) Πρόσδεση RAN/RAS GTPασών (GO-IDs: 8536 & 17016) 2) Ενεργότητα μεταφοράς πρωτεϊνών (GO-ID: 8565)
12	1) Αγκυροβόληση του κυτταροσκελετού στην πυρηνική μεμβράνη (GO-ID: 90286) 2) Οργάνωση της πυρηνικής μήτρας (GO-ID: 43578)	Πρόσδεση λαμινών (GO-ID: 5221)
13	1) Κίνηση επί των μικροσωληνίσκων (GO-ID: 7018) 2) Συγκρότηση της μιτωτικής ατράκτου (GO-ID: 51293)	Κινητήρια ενεργότητα (GO-ID: 3774)

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

14	3) Οργάνωση του κυτταροσκελετού (GO-ID: 7010) 1) Αποσυναρμολόγηση συστήματος ESCRT (GO-ID: 1904896) 2) Κυτταροκίνηση (GO-ID: 920)	Πρόσδεση καντερινών (GO-ID: 45296)
15	1) Αποακετυλίωση των ιστονών (GO-ID: 16575) 2) Οργάνωση της χρωματίνης (GO-ID: 6325)	Αποακετυλάση ιστονών (GO-ID: 4407)
16	1) Βιοσύνθεση χολικών οξέων (GO-ID: 6699) 2) Μεταφορά λιπιδίων (GO-ID: 6869)	Πρόσδεση στερολών (GO-ID: 32934)
17	1) Ρύθμιση του κυτταρικού θανάτου (GO-ID: 10941) 2) Μονοπάτι ERAD (GO-ID: 36503)	Δεν βρέθηκε κάποιος όρος
18	1) Εισαγωγή πρωτεϊνών που φέρουν NLS στον πυρήνα (GO-ID: 6607) 2) Ρύθμιση του γενετικού ανασυνδυασμού (GO-ID: 18)	1) Πρόσδεση σε αλληλουχία πυρηνικού εντοπισμού (GO-ID: 8139) 2) Μεταφορά πρωτεϊνών (GO-ID: 8565)
19	1) Ενεργοποίηση της έμφυτης ανοσίας (GO-ID: 2218) 2) Επιδιόρθωση του DNA (GO-ID: 6281)	Ενεργότητα μεταφοράς ουβικιτίνης (GO-ID: 4842)
20	1) Οργάνωση του συμπλόκου του πυρηνικού πόρου (GO-ID: 6999) 2) Εξαγωγή του mRNA (GO-ID: 6406) 3) Μεταφορά πρωτεϊνών (GO-ID: 15031) 4) Σύνδεση αδελφών χρωματίδων (GO-ID: 7062)	Δομικό συστατικό του πυρηνικού πόρου (GO-ID: 17056)
21	Αυτοφαγία (GO-ID: 6914)	1) Πρόσδεση σε λιγάση ουβικιτίνης (GO-ID: 31625) 2) Πρόσδεση σε GABA υποδοχέα (GO-ID: 50811) και πρόσδεση σε β-τουμπουλίνη (GO-ID: 48487)
22	Σηματοδοτικό μονοπάτι νευροτροφινών (GO-ID: 38179)	Ενεργότητα υποδοχέα νευροτροφίνης (GO-ID: 5030)
23	Θετική ρύθμιση στο στάδιο επιμήκυνσης της μετάφρασης (GO-ID: 45901)	Ενεργότητα παράγοντα επιμήκυνσης της μετάφρασης (GO-ID: 3746)
24	Σηματοδοτικό μονοπάτι οιστρογόνων (GO-ID: 30520)	Ενεργότητα υποδοχέα οιστρογόνων (GO-ID: 30284)
25	Απόκριση του ενδοπλασματικού δικτύου σε μη διπλωμένες πρωτεΐνες (GO-ID: 30968)	Πρόσδεση καντερινών (GO-ID: 45296)
26	1) Ρύθμιση της διατήρησης των τελομερών (GO-ID: 32204) 2) Ρύθμιση της χρωμοσωμικής οργάνωσης (GO-ID: 33044)	Πρόσδεση στα τελομερή (GO-ID: 42162)
27	Δεν βρέθηκε κάποιος όρος	Πρόσδεση καρβοξυτελικών άκρων (GO-ID: 8022)
28	Τροποποίηση ασπαραγίνης (GO-ID: 18196)	Ενεργότητα υποδοχέα ασιαλογλυκοπρωτεϊνών (GO-ID: 4873)
29	Δεν βρέθηκε κάποιος όρος	Δεν βρέθηκε κάποιος όρος
30	Κυστιδιακή μεταφορά από το αδρό ΕΔ στο cis-Golgi (GO-ID: 48207)	Δεν βρέθηκε κάποιος όρος
31	Δεν βρέθηκε κάποιος όρος	Δεν βρέθηκε κάποιος όρος

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

32	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
33	1) Εκπόλωση της μεμβράνης (GO-ID: 51899) 2) Ρύθμιση αυτοφαγίας (GO-ID: 10506) 3) Είσοδος ιόντων ασβεστίου στο κυτταρόπλασμα (GO-ID: 97533)	Ενεργότητα καναλιού ασβεστίου (GO-ID: 5262)
34		Πρόσδεση RNA (GO-ID: 3723)
35	Απόκριση στην ιντερλευκίνη-15 (GO-ID: 70672)	Πρόσδεση κυτοκινών (GO-ID: 19955)
36	Εισαγωγή πρωτεϊνών που φέρουν NLS στον πυρήνα (GO-ID:6607)	Πρόσδεση σε αλληλουχίες πυρηνικού εντοπισμού (GO-ID: 8139)
37	Εμπλοκή στην μορφογένεση (GO-ID: 48646)	Πρόσδεση στο DNA (GO-ID: 3677)
38	1) Ρύθμιση της διακοπής του κυτταρικού κύκλου (GO-ID: 71156) ως απόκριση στις βλάβες DNA (GO-ID: 6977) 2) Απόκριση στην υποξία (GO-ID: 1666)	1) Πρόσδεση σε λιγάνση ουβικιτίνης (GO-ID: 31625) 2) Πρόσδεση ιόντων ψευδαργύρου (GO-ID: 8270)
39	Αρνητική ρύθμιση στον πολλαπλασιασμό των οστεοβλαστών (GO-ID: 33689)	<i>Δεν βρέθηκε κάποιος όρος</i>
40	1) Ρύθμιση της καρδιακής σύσπασης (GO-ID: 1903779) 2) Ρύθμιση της ενεργότητας υδρολασών (GO-ID: 51336)	Ενεργότητα ρύθμισης καναλιού ασβεστίου (GO-ID: 5246)
41	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
42	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
43	Εκτύλιξη δευτεροταγών δομών RNA (GO-ID: 10501)	Ενεργότητα ελικάσης που εξαρτάται από πουρινικά NTPs (GO-ID: 70035)
44	Βιοσύνθεση φωσφατιδυλοϊνοσιτόλης (GO-ID: 6661)	Ενεργότητα φωσφατάσης (GO-ID: 16791)
45	Μεταμεταγραφική γονιδιακή σίγαση μέσω RNA (GO-IDs: 16441 & 31047)	1) Πρόσδεση siRNA (GO-ID: 35197) 2) Ενεργότητα νουκλεάσης (GO-ID: 4518)
46	Διαφοροποίηση των νευρώνων (GO-ID:30182)	Πρόσδεση των πρωτεϊνών S100 (GO-ID: 44548)
47	1) Ρύθμιση της πυρηνοκυτταροπλασματικής μεταφοράς (GO-ID: 46822) 2) Τροποποίηση αμινοξικών καταλοίπων (GO-ID: 18193)	<i>Δεν βρέθηκε κάποιος όρος</i>
48	Ρύθμιση της απόκρισης στο στρες (GO-ID: 80134)	Ενεργότητα οξειδοαναγωγής (GO-ID:16701 & 16702)
49	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
50	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
51	<i>Δεν βρέθηκε κάποιος όρος</i>	<i>Δεν βρέθηκε κάποιος όρος</i>
52	Αποσυναρμολόγηση πρωτεϊνικών συμπλόκων (GO-ID: 32984)	1) Ενεργότητα πεπτιδάσης (GO-ID: 8233) 2) Πρόσδεση MIT αυτοτελούς δομικής περιοχής (GO-ID: 90541)
53	Μεταφορά mRNA (GO-ID:51028) εκτός του πυρήνα (GO-ID: 6611)	Πρόσδεση σε αλληλουχία πυρηνικού εντοπισμού (GO-ID: 8139)
54	Διαφοροποίηση των λιποκυττάρων (GO-ID: 45444)	Ενεργότητα αναγωγής του γλυοξυλικού(NADP) (GO-ID: 30267)
55	Ανασυνδυασμός του DNA (GO-ID: 6310)	<i>Δεν βρέθηκε κάποιος όρος</i>

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Στο σύνολό τους, η πλειονότητα των ομάδων που προέκυψαν μέσω της ομαδοποίησης του αλγορίθμου MCL σχετίζονται με όρους Βιολογικών Διεργασιών άμεσα συνυφασμένων με τη λειτουργικότητα του πυρήνα, όπως η οργάνωση του πυρηνικού φακέλου, η μεταφορά μορίων RNA και πρωτεϊνών η λειτουργία των NPCs, η τροποποίηση της χρωματίνης κλπ, αλλά και με λειτουργίες σχετικές με τη ρύθμιση του κυτταρικού κύκλου και πολλαπλασιασμού, όπως η οργάνωση του κυτταροσκελετού και η συγκρότηση και αποσυγκρότηση των μικροσωληνίσκων. Επιπρόσθετα, κάποιες ομάδες φαίνονται να σχετίζονται με μηχανισμούς όπως η κυτταρική διαφοροποίηση σε διάφορους ιστούς, η ρύθμιση της αυτοφαγίας και της απόπτωσης. Όσον αφορά τους όρους Μοριακής Λειτουργίας, πολλές ομάδες αντιστοιχούνται σε υπερμοριακά σύμπλοκα που σχετίζονται με ενζυμική ενεργότητα, όπως σύμπλοκα GTPασών, κινασών, φωσφατασών κλπ. Ωστόσο, μεγάλος αριθμός ομάδων βρέθηκαν να σχετίζονται κυρίως με τη ρύθμιση των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών, είτε συμμετέχοντας σε μηχανισμούς όπως τα NPCs είτε ρυθμίζοντας μηχανισμούς μεταγωγής σήματος, το σχηματισμό ομο- και ετεροδιμερών και την πρόσδεση υπομονάδων ουμπικιτίνης σε πρωτεΐνες που σημαίνονται για αποικοδόμηση.

Συνολικά, τα αποτελέσματα από την κατασκευή και την ανάλυση του δικτύου αλληλεπιδράσεων έδειξαν ότι οι πρωτεΐνες του πυρηνικού φακέλου επιτελούν μια σειρά από ιδιαίτερα σημαντικούς ρόλους στο κύτταρο. Το ίδιο το δίκτυο βρέθηκε να διαθέτει τα χαρακτηριστικά πραγματικών βιολογικών δικτύων, ενώ σημαντικό στοιχείο στη δομή και τη συνοχή του αποτελούν οι πρωτεΐνες του πυρηνικού φακέλου, παρά το σχετικά μικρό αριθμό τους. Τέλος, μέσω της ανάλυσης ομαδοποίησης και του λειτουργικού σχολιασμού του δικτύου, φάνηκε πως οι πρωτεΐνες του πυρηνικού φακέλου διαδραματίζουν ιδιαίτερα σημαντικό ρόλο στις διεργασίες του κυττάρου, τόσο μέσα από τη ρύθμιση σημαντικών κυτταρικών μηχανισμών, όσο και μέσα από τις αλληλεπιδράσεις τους με πρωτεΐνες από διάφορα υποκυτταρικά διαμερίσματα.

3.5.2 Η βάση δεδομένων NucEnvDB

Η βάση δεδομένων NucEnvDB είναι μια δημόσια διαθέσιμη βάση δεδομένων για τις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις τους. Περιέχει συνολικά 2863 πρωτεΐνες του πυρηνικού φακέλου, προερχόμενες από 394 οργανισμούς. Οι παραπάνω πρωτεΐνες συμμετέχουν σε 20378 αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών, ενώ σχετίζονται με 4680 όρους γονιδιακής οντολογίας. Το 30.77% των πρωτεϊνών της βάσης (881 πρωτεΐνες) είναι διαμεμβρανικές. Ακολουθούν οι περιφερειακές πρωτεΐνες (558, 19.49%) και οι αγκυροβολημένες μεμβρανικές πρωτεΐνες (169, 5.90%), ενώ οι υπόλοιπες πρωτεΐνες (1255, 43.84%) είτε είναι σφαιρικές υδατοδιαλυτές, είτε δεν διαθέτουν γνωστή τοπολογία σε σχέση με τη μεμβράνη. Διαθέσιμες αλληλεπιδράσεις υπάρχουν για 19 από τους 394 οργανισμούς, με τη συντριπτική πλειονότητά τους να προέρχονται από τον άνθρωπο (12082) και τον ποντικό (1308).

Η βάση δεδομένων NucEnvDB είναι διαθέσιμη διαδικτυακά μέσα από τη διεύθυνση <http://bioinformatics.biol.uoa.gr/nucenvdb>. Ο χρήστης έχει τη δυνατότητα να περιηγηθεί στα περιεχόμενα της βάσης μέσω των σελίδων Πλοήγησης («Browse») για τις πρωτεΐνες («Browse by Proteins»), τις υποκυτταρικές θέσεις του πυρηνικού φακέλου («Browse by Envelope Locations»), το είδος του οργανισμού («Browse by Organism») ή τους όρους γονιδιακής οντολογίας («Browse by Ontology Terms»). Η αναζήτηση των περιεχομένων της βάσης μπορεί να πραγματοποιηθεί είτε μέσω απλής αναζήτησης είτε μέσω της σελίδας Εξειδικευμένης Αναζήτησης («Advanced Search»). Μέσω της τελευταίας, ο χρήστης μπορεί να πραγματοποιήσει αναζητήσεις με όρους όπως το όνομα της πρωτεΐνης, το όνομα του γονιδίου, το όνομα του οργανισμού ή να εισάγει έναν ή περισσότερους κωδικούς AC της UniProt. Επιπρόσθετα, μπορεί να περιορίσει τις αναζητήσεις του σε εγγραφές με γνωστή δομή ή με γνωστές αλληλεπιδράσεις και να πραγματοποιήσει αναζητήσεις σε όλη τη βάση ή μόνο σε πρωτεώματα αναφοράς. Τέλος, μπορεί να επιλέξει την αναζήτηση σε συγκεκριμένες υποκυτταρικές θέσεις, να αναζητήσει πρωτεΐνες με συγκεκριμένη τοπολογία (π.χ. διαμεμβρανικές) και να περιορίσει την αναζήτηση των τελευταίων με βάση την πηγή προέλευσης της τοπολογίας (Εικόνα 73).

Advanced Search

Protein Name:
e.g. Nucleoporin

Gene Name:
e.g. NUP43

Accession Code(s): Separate multiple accession codes by spaces
e.g. Q8NFH3 Q7Z3B4 Q9BTX1

Organism:
Enter species name OR NCBI Taxonomy ID
e.g. Homo sapiens, 9606

OR
select your organism from the drop-down list:
--Organism--

Proteome Status: ☒ Any ☐ Reference ☐ Not Reference

Limit search to:
☐ Entries with known 3D structure ☐ Entries with protein-protein interactions

Subcellular Location

Nuclear Envelope Locations:

- ☒ All
- ☒ Nuclear Envelope
- ☒ Nuclear Inner Membrane
- ☒ Nuclear Lamina
- ☒ Nuclear Membrane
- ☒ Nuclear Outer Membrane
- ☒ Nuclear Intermembrane Space
- ☒ Nuclear Pore Complex
- ☒ Perinuclear Region
- ☒ Host Perinuclear Region
- ☒ Host Nuclear Envelope
- ☒ Host Nuclear Lamina
- ☒ Host Nuclear Membrane
- ☒ Host Nuclear Inner Membrane
- ☒ Host Nuclear Outer Membrane

Membrane Topology:

- ☒ All
- ☒ Transmembrane
- ☒ Peripheral
- ☒ Lipid-Anchored
- ☒ Other/Unknown

Topology Annotation Source:

- ☒ All
- ☒ Experimental Evidence
- ☒ By Similarity
- ☒ Curator Inference
- ☒ Sequence Analysis
- ☒ Other/Unknown

Search **Clear**

Εικόνα 73 Αναζήτηση πληροφοριών στη βάση δεδομένων NucEnDB. Παρουσιάζεται η φόρμα της εξειδικευμένης αναζήτησης.

Η σελίδα της κάθε εγγραφής στη βάση περιλαμβάνει μια λεπτομερή περιγραφή των χαρακτηριστικών της (Εικόνες 74-75). Συγκεκριμένα, παρατίθενται τα βασικά στοιχεία κάθε εγγραφής (όνομα πρωτεΐνης, όνομα γονιδίου, κωδικός AC, οργανισμός προέλευσης και τύπος πρωτεώματος), καθώς και η αμινοξική αλληλουχία της. Στην περίπτωση πρωτεϊνών με διαθέσιμη τρισδιάστατη δομή δίνεται η δυνατότητα οπτικοποίησής της, με την ενσωμάτωση του προγράμματος μοριακών γραφικών LiteMol. Για κάθε εγγραφή περιλαμβάνονται στοιχεία για την υποκυτταρική θέση και τη λειτουργία της, συμπεριλαμβανομένων όλων των σχετικών όρων γονιδιακής οντολογίας (Εικόνα 74). Σε περίπτωση που η πρωτεΐνη της εγγραφής συμμετέχει σε αλληλεπιδράσεις με άλλες πρωτεΐνες, αυτές παρατίθενται σε ξεχωριστό πεδίο και διακρίνονται σε δύο κατηγορίες: τις αλληλεπιδράσεις με άλλες πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις με άλλες πρωτεΐνες (Εικόνα 75). Τέλος, παρατίθενται οι

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

ετεροαναφορές της εγγραφής σε μεγάλες βάσεις βιολογικών δεδομένων (UniProt, PDB, PROSITE, Pfam, DisGeNET και OMIM).

NucEnvDB

Browse ▾

Advanced Search

Tools ▾

Downloads

Manual

Contact

Quick Search

Q

Nucleoporin p54 (*Homo sapiens*)

FASTA

TEXT

XML

Subcellular Location

Biology & Function

Interactions

Cross-References

General Information

Protein Information

Protein Name

Nucleoporin p54

Accession Code

Q7Z3B4

Gene

NUP54

Organism

Homo sapiens (Taxonomy: 9606)

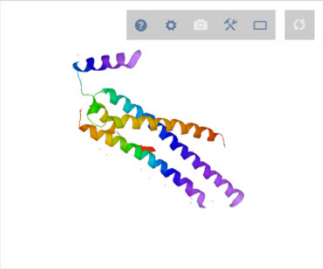
Part of Reference Proteome?

Yes

Sequence (Length: 507)

Show

Structure Viewer (PDB: 4JO9)



Select PDB ID ▾ Go

Subcellular Location

Description

Show

Position in the Nuclear Envelope

Location	Location ID	Description
Nuclear Envelope	SL-0178	The nuclear envelope is a membrane system which surrounds the nucleoplasm of eukaryotic cells. It is composed of the nuclear lamina, nuclear pore complexes and two nuclear membranes. The space between the two membranes is called the nuclear intermembrane space.
Nuclear Membrane	SL-0182	The membrane surrounding the nucleus. This term is used when it is not known if the protein is found in or associated with the inner or outer nuclear membrane.
Nuclear Pore Complex	SL-0185	The nuclear pore complex (NPC) constitutes the exclusive means of nucleocytoplasmic transport. NPCs allow the passive diffusion of ions and small molecules and the active bidirectional transport of macromolecules such as proteins, RNAs etc across the double-membrane nuclear envelope. The NPC is composed of at least 30 distinct subunits known as Nucleoporins (NUPs).

Membrane Topology

Topology	Source	Annotation Type
Peripheral	UniProt	By Similarity (ECO:0000250 UniProtKB:P70582)

Assigned Ontology terms

Show

Εικόνα 74 Η βάση δεδομένων NucEnvDB. Παρουσιάζεται το πρώτο τμήμα της σελίδας εγγραφής για την Νουκλεοπορίνη p54 του ανθρώπου (UniProt: Q7Z3B4). Παρατίθενται οι γενικές πληροφορίες πάνω στην πρωτεΐνη, η τρισδιάστατη δομή της, μέσω της εφαρμογής LiteMol και το πεδίο της Υποκυτταρικής Τοποθεσίας, στο οποίο φαίνονται οι θέσεις εντοπισμού της πρωτεΐνης στον πυρηνικό φάκελο. Τμήματα της εγγραφής μπορούν να εμφανιστούν ή κρυφτούν πατώντας τα κουμπιά «Show/Hide».

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

Protein-Protein Interactions

Interactions with Nuclear Envelope proteins			
Hide			
Partner (NucEnvDB)		IntAct	Confidence score
Q7Z3B4	Self	EBI-741048,EBI-741048	0.37
Q9BVL2	Nucleoporin p58/p45	EBI-2811583,EBI-741048	0.82
P57740	Nuclear pore complex protein Nup107	EBI-295687,EBI-741048	0.35
P37198	Nuclear pore glycoprotein p62	EBI-741048,EBI-347978	0.86
P61970	Nuclear transport factor 2	EBI-741048,EBI-591778	0.49
Q9NV70	Exocyst complex component 1	EBI-1045313,EBI-741048	0.49
Q8IYI6	Exocyst complex component 8	EBI-741048,EBI-742102	0.49
Q8WUM0	Nuclear pore complex protein Nup133	EBI-741048,EBI-295695	0.37
Q96HA1	Nuclear envelope pore membrane protein POM 121	EBI-741048,EBI-739990	0.37
Interactions with other proteins			
Show			

[Back to top](#)

Cross-References

Database	Links
UNIPROT	Q7Z3B4
PDB	4JNU 4JNV 4JO7 4JO9 5IJN 5IJO
Pfam	PF13874
OMIM	607607

[Back to top](#)

National and Kapodistrian University of Athens

Department of Biology

Biophysics & Bioinformatics Laboratory

Εικόνα 75 Η βάση δεδομένων NucEnvDB (2). Παρουσιάζεται το δεύτερο τμήμα της σελίδας εγγραφής για την Νουκλεοπορίνη p54 του ανθρώπου (UniProt: Q7Z3B4). Παρατίθενται οι αλληλεπιδράσεις στις οποίες συμμετέχει η πρωτεΐνη, τόσο με άλλες πρωτεΐνες του πυρηνικού φακέλου όσο και με πρωτεΐνες από άλλες υποκυτταρικές θέσεις. Επιπρόσθετα, παρατίθεται το πεδίο των ετεροαναφορών της πρωτεΐνης σε μεγάλες βάσεις δεδομένων. Τμήματα της εγγραφής μπορούν να εμφανιστούν ή κρυφτούν πατώντας τα κουμπιά «Show/Hide».

Η NucEnvDB διαθέτει μια σειρά από εργαλεία, μέσω των οποίων ο χρήστης μπορεί να πραγματοποιήσει αναλύσεις στα περιεχόμενα της βάσης. Αρχικά, στη βάση έχει ενσωματωθεί ο αλγόριθμος BLAST, μέσω του οποίου ο χρήστης μπορεί να πραγματοποιήσει αναζητήσεις ομολογίας μιας ή περισσότερων άγνωστων αλληλουχιών με τα περιεχόμενα της βάσης, μέσω στοιχίσεων κατά ζεύγη (Εικόνα 76α και γ). Εκτός από το BLAST, στη βάση υπάρχει

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

ενσωματωμένο και το πακέτο λογισμικού HMMER, επιτρέποντας την πιο εξειδικευμένη αναζήτηση ομολογίας μέσω rHMMs (Εικόνα 76 β και δ). Συγκεκριμένα, ο χρήστης μπορεί να πραγματοποιήσει αναζήτηση για μία ή περισσότερες αλληλουχίες έναντι των αλληλουχιών της βάσης, μέσω της μεθόδου *phmmer*, αντίστοιχα με το BLAST. Εναλλακτικά, μπορεί να πραγματοποιήσει αναζήτηση έναντι των προφίλ rHMMs των πρωτεϊνών της βάσης, μέσω της μεθόδου *hmmsearch*, και να εντοπίσει αν η αλληλουχία του διαθέτει κάποια αυτοτελή δομική περιοχή με γνωστή παρουσία στον πυρηνικό φάκελο. Τέλος, μπορεί να εισάγει είτε ένα rHMM της επιλογής του είτε μια πολλαπλή στοίχιση (η οποία θα χρησιμοποιηθεί για την αυτοματοποιημένη κατασκευή ενός rHMM), και να αναζητήσει πρωτεΐνες οι οποίες διαθέτουν το μοτίβο (ή την αυτοτελή δομική περιοχή) που περιγράφεται από το rHMM, μέσω της μεθόδου *hmmsearch*.

Πέρα από το BLAST και το HMMER, η NucEnvDB προσφέρει ένα αυτοματοποιημένο εργαλείο για την κατασκευή, οπτικοποίηση και ανάλυση δικτύων αλληλεπιδράσεων από το χρήστη. Στη σελίδα «Ανάλυση Δικτύων» («Network Analysis») ο χρήστης μπορεί να εισάγει μια λίστα κωδικών AC από πρωτεΐνες της βάσης, για τις οποίες θα πραγματοποιηθεί η αναζήτηση αλληλεπιδράσεων και η κατασκευή του δικτύου. Αυτή η λίστα μπορεί να εισαχθεί είτε χειροκίνητα είτε αυτοματοποιημένα, από τα αποτελέσματα μιας αναζήτησης στη βάση. Εναλλακτικά, ο χρήστης μπορεί να επιλέξει την κατασκευή ενός δικτύου για το σύνολο των πρωτεϊνών ενός οργανισμού από τη σχετική λίστα. Σε κάθε περίπτωση, ο χρήστης καλείται να επιλέξει τις αλληλεπιδράσεις που θα συμπεριληφθούν στο δίκτυο, καθώς και το αν θα πραγματοποιηθεί ανάλυση ομαδοποίησης με τον αλγόριθμο MCL (Εικόνα 77).

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

A

BLAST Search

Step 1. Paste your sequence(s) in the form below:

>UNKNOWN_SEQUENCE
AVVSVNITSLGLRLSEISDGVIIISGNKLCYANTINNKKEIKLVLS
GAFQTVYKGLMIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAVDNPHVCRLLG
CLTSTVQLITQGLNFFGCLLDVREHKDNIQSQYLLNMCVQIAKGIWYLEDORRL
RNVLVKTPQHVKITDFGLAKLLGAEKEYHAEGGKVPKIMPALESILHRIYTHQSDVNS
GVTVWEIMTFGSEFDTGTPASEISSLEKGERLPQPPICTIDVYMIHVKCMIH
FRELIIFFSGHARDPQRYLVIQGDERHMLPSPDTSNFRALMDEEDMDVDAEYLIPQ
QOFFSPSTRTFLSSLSATSNISTVACIDRNLGSCPIKEDSFLGRYSDDPTGALTED
SIDOTFLPVFEIYNGVFRKAGSVQVNVNKHQFLNFAKSDPHYQDPHSTAVONPEYLN
TVQPTCVNSTFDSFAHMAKQSGHQISLQNPFDYQDFFFEAKPHGIFKOSTAEHAYLAV
APQSEFTIG

Note: Sequences must be in FASTA format.

Step 2. Define search options

E-value: 10.0

Submit Clear

B

HMMSCAN
protein sequence vs profile HMM database

Step 1. Paste your sequence(s) in the form below:

>UNKNOWN_SEQUENCE
AVVSVNITSLGLRLSEISDGVIIISGNKLCYANTINNKKEIKLVLS
GAFQTVYKGLMIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAVDNPHVCRLLG
CLTSTVQLITQGLNFFGCLLDVREHKDNIQSQYLLNMCVQIAKGIWYLEDORRL
RNVLVKTPQHVKITDFGLAKLLGAEKEYHAEGGKVPKIMPALESILHRIYTHQSDVNS
GVTVWEIMTFGSEFDTGTPASEISSLEKGERLPQPPICTIDVYMIHVKCMIH
FRELIIFFSGHARDPQRYLVIQGDERHMLPSPDTSNFRALMDEEDMDVDAEYLIPQ
QOFFSPSTRTFLSSLSATSNISTVACIDRNLGSCPIKEDSFLGRYSDDPTGALTED
SIDOTFLPVFEIYNGVFRKAGSVQVNVNKHQFLNFAKSDPHYQDPHSTAVONPEYLN
TVQPTCVNSTFDSFAHMAKQSGHQISLQNPFDYQDFFFEAKPHGIFKOSTAEHAYLAV
APQSEFTIG

Note: Sequences must be in FASTA format.

Search options

Cut-off type: Gathering (GA) E-value: 0.01 Bit-score: 25

Submit Clear

Γ

Receptor tyrosine-protein kinase erbB-2 (*Canis lupus familiaris*) Accession: O18735

No.	AC	Length	Bit-score	E-value	Query	Hit	Identities	Positives	Gaps	Alignment
13	O18735	1259	553.132	0	43-527	721-1253	306/550 (55.64 %)	363/550 (66.00 %)	82/550 (14.91 %)	Hide

Alignment:

Query: 43 KIKVLGSGAGFTVYKGLMIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAVDNPHVCRLLGICLTSTVQLITQVIFPGCLLDVREHKDNIQSQYLLNMCVQIAKGIWYLEDORRL 162
+KXVLGSGAGFTVYKGLMIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMA V +P+V RLLGICLTSTVQL+TQVIFPGCLLD+VREH+ +GSO LLNMCVQIAKGIWYLEDORRL
Hit: 721 KXVLGSGAGFTVYKGLMIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMAVDNPHVCRLLGICLTSTVQLITQVIFPGCLLDVREHGRRLGSDQLNMCVQIAKGIWYLEDORRL 840

Query: 163 VHRDLAARNVLVKTPQHVKITDFGLAKLLGAEKEYHAEGGKVPKIMPALESILHRIYTHQSDVNSVGVTVWEIMTFGSKPYDGIPIASEISSLEKGERLPQPPICTIDVYMIHVKCMIH 282
VHRDLAARNVLVK+P HVKITDFGLA+LL +E EYHA+GGKVPKIMPALES I R +THQSDVNSVGVTVWEIMTFG+KPYDGIPA EI +LEKGERLPQPPICTIDVYMIHVKCMIH
Hit: 841 VHRDLAARNVLVKSPHMKITDFGLARLLDIDETEHADGGKVPKIMPALESIPPRFTTHQSDVNSVGVTVWEIMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIHVKCMIH 960

Query: 283 DADSRPKFRELIIFFSGHARDPQRYLVIQGDERHMLPSPDTSNFRALMDEEDMDVDAEYLIPQDGFSSPST-----SRTPLSSL5 368
D++ RP+FRE+ EF5+HARDPQR++VIQ +E + SP DS FYR+L+++DM D+VDA+EYL+PQDGF T ++PL S
Hit: 961 DSECRPRFRELVAEFSHARDPQRFFVIQ-NEDLGASPLDSTFYRSLLEDDMDGLVDAEYLVQDGFCEPTPGAGGTAAHRRSSSTRNGGGELTGLPESEEPKSPPLAPSE 1079

Query: 369 ATSN--NSTVACIDRNLGSCPIKEDSFLGRYSDDPTGALTEDSIDO---TFLPVPEYINQSV-----PKRPGAGSVQNPVYHQNLPAPSROPHYQD--PHSTAVGNPE 466
A S+ + + GLQS P + S LQRYSDPT L ++ T P PEY+NQ P RPAG+ P +P ++ +D +AV NPE
Hit: 1080 AGSDVFDGDLGGAAGLQSLPSQDPSPLQRYSEDTVPLPPTDGGKVPKIMPALESIPPRFTTHQSDVNSVGVTVWEIMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYMIHVKCMIH 1199

Query: 467 YL-----NTVQPTCVNSTFDSFAHMAKQSGHQISLQNPFDYQDFFFEAKPHGIFKGS-TAENAYL 527
YL P + FD+ +W Q S + S P F+G+ TAEN EYL
Hit: 1200 YLAPRGAAPQPPPPAFSFAFDONLYMDQDPSERGS-----PPSTFEGTPTAENPEYL 1253

Δ

Protein tyrosine kinase (Pfam: PF07714)

No.	Accession	Bit-score	E-value	Query	Hit	Alignment	Assoc. Entries
1	PF07714	295.3	1.2e-89	44-294	4-257	Show	Show

Protein kinase domain (Pfam: PF00069)

No.	Accession	Bit-score	E-value	Query	Hit	Alignment	Assoc. Entries
2	PF00069	135.6	6.6e-41	44-293	4-258	Show	Show

Receptor L domain (Pfam: PF01030)

No.	Accession	Bit-score	E-value	Query	Hit	Alignment	Assoc. Entries
3	PF01030	30.2	1.3e-08	6-42	74-111	Hide	Show

Alignment:

T-SBEE-TTT-ESSSSEEEEESSSTTSSTTTSHHHHT CS
PF01030.23 74 nleelglplskltsvsvvskNpkLyseteidwkel 111
n+++Lgl+slkei+g v+is N++Lcy +t i+w+k+l
UNKNOWN_SEQUENCE 6 NITSLGLRLSEISDGVIIISGNKLCYANTINNKKEIKLVLS 42
799*****986 PP

Εικόνα 76 Αναζήτηση ομολογίας στη βάση NucEnnDB. (Α) Η φόρμα αναζήτησης του εργαλείου BLAST. (Β) Η φόρμα αναζήτησης του εργαλείου HMMER. (Γ) Τμήμα των αποτελεσμάτων του BLAST. (Δ) Τμήμα των αποτελεσμάτων του HMMER.

~ 271 ~

Protein-Protein Interaction Network Analysis

Step 1. Prepare dataset

Enter a list of NucEnvDB Accession Codes in the box below (separate ACs by spaces):

A8CG34 O00159 O15504 O60318 O75694 O95271 P12270 P20592 P35658 P37198 P49790
P49792 P52948 P55735 P57740 P61970 P63241 Q12769 Q14764 Q53GS7 Q5SRE5 Q723B4
Q8N1F7 Q8NFB3 Q8NFB4 Q8NFB5 Q8TD16 Q8TEM1 Q8WUM0 Q8WYF5 Q92621 Q96EE3 Q96HA1
Q99567 Q9BTX1 Q9BVL2 Q9BN27 Q9GZV4 Q9HC62 Q9NFA8 Q9NR21 Q9NRG9 Q9NUV7 Q9UIA9
Q9UKX7 Q9UMR2 Q9UND3

OR

Retrieve all NucEnvDB entries for species:

Εγγραφές NucEnvDB

Step 2. Select protein-protein interactions

2.1 Define which interactions to include:

☐ Use only interactions between the Nuclear Envelope proteins themselves

☒ Add interactions between Nuclear Envelope proteins and their (non Envelope) first neighbors

☐ Also add interactions between the first neighbors (Slower)

2.2 Filter interaction data:

Remove self-interactions? ☒ No ☐ Yes

Use interaction cut-off? ☒ No ☐ Yes

Ορισμός αλληλεπιδράσεων

Step 3. Perform Network Analysis

WARNING! Network analysis can be slow for large networks.

Perform graph theory analysis with NetworkX? ☒ Yes ☐ No

Cluster network with Markov Clustering (MCL)? ☒ Yes ☐ No

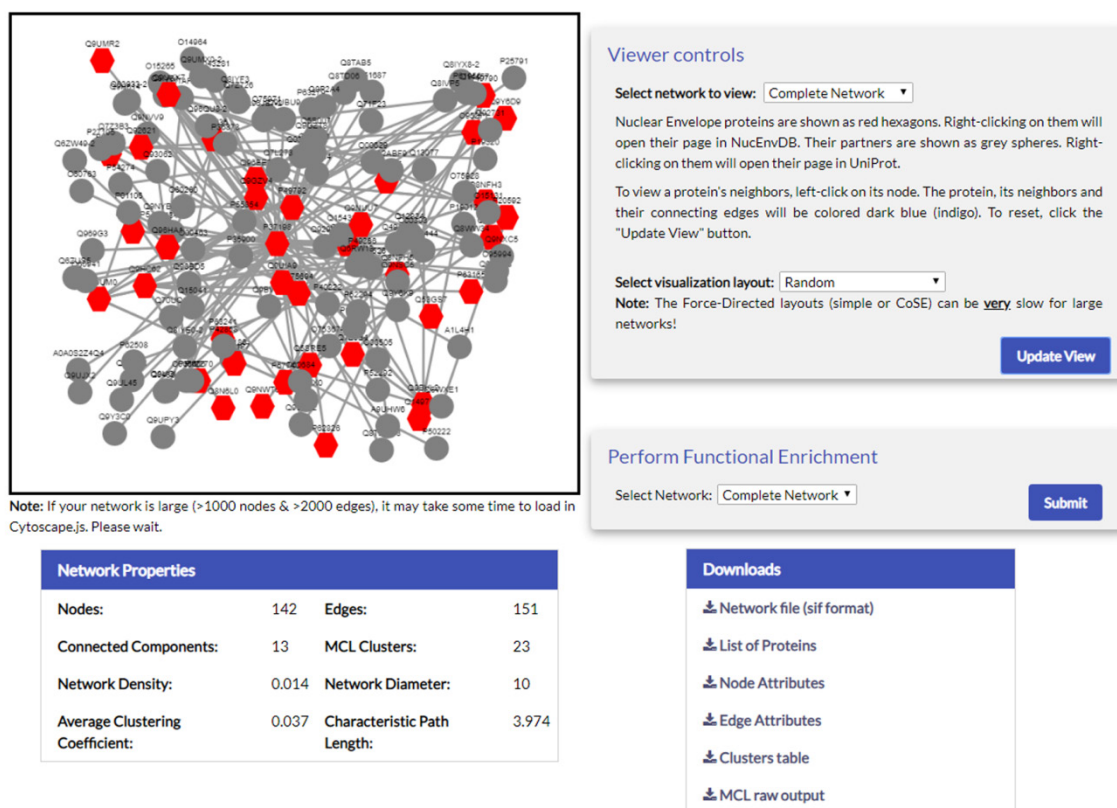
Inflation:

☒ Remove Loops

**Επιλογές
ανάλυσης δικτύου**

Εικόνα 77 Κατασκευή δικτύου αλληλεπιδράσεων μέσω της NucEnvDB. Παρατίθεται η φόρμα εισαγωγής για την κατασκευή ενός δικτύου αλληλεπιδράσεων. Ο χρήστης αρχικά δίνει ως είσοδο ένα ή περισσότερους κωδικούς AC της βάσης, είτε χειροκίνητα είτε αυτοματοποιημένα, μέσα από τα αποτελέσματα αναζητήσεων. Εναλλακτικά, ο χρήστης μπορεί να επιλέξει ως είσοδο το σύνολο των πρωτεϊνών της βάσης NucEnvDB για ένα συγκεκριμένο πρωτόωμα (π.χ. άνθρωπος). Στο δεύτερο βήμα, ο χρήστης καλείται να επιλέξει το είδος των αλληλεπιδράσεων που θα συμπεριληφθούν, καθώς και το αν θα εφαρμόσει κάποιο κατώφλι αξιοπιστίας. Στο τελικό βήμα, ο χρήστης καλείται να επιλέξει αν θα πραγματοποιηθεί ανάλυση δικτύου με βάση τη Θεωρία Γράφων και ανάλυση ομαδοποίησης με τον αλγόριθμο MCL.

Network Analysis Results



Εικόνα 78 Αποτελέσματα της ανάλυσης δικτύου. Παρουσιάζονται τα αποτελέσματα από την κατασκευή και ανάλυση ενός δικτύου αλληλεπιδράσεων για το ανθρώπινο Σύμπλοκο Πυρηνικού Πόρου. Η σελίδα των αποτελεσμάτων περιλαμβάνει μια εφαρμογή για την οπτικοποίηση του δικτύου, τα αποτελέσματα από την ανάλυση μέσω της Θεωρίας Γράφων, καθώς και μια σειρά αρχείων με το σύνολο των δεδομένων, τα οποία μπορούν να αναλυθούν περαιτέρω μέσω του Cytoscape.

Τα αποτελέσματα της ανάλυσης δικτύου οπτικοποιούνται χρησιμοποιώντας το Cytoscape.js, μια διαδικτυακή εκδοχή του Cytoscape σχεδιασμένη για ενσωμάτωση και χρήση σε ιστοσελίδες. Εκτός από την οπτικοποίηση του δικτύου παρατίθενται και τα αποτελέσματα από την ανάλυσή του με βάση τη θεωρία γράφων. Επιπρόσθετα, δίνεται η δυνατότητα ανάλυσης λειτουργικού εμπλουτισμού, τόσο για το συνολικό δίκτυο όσο και για τις υπο-ομάδες του (βλ. παρακάτω). Τόσο το ίδιο το δίκτυο όσο και τα αποτελέσματα από την ανάλυσή του είναι διαθέσιμα για λήψη, σε αρχεία συμβατά με το πρόγραμμα Cytoscape, για περαιτέρω ανάλυση (Εικόνα 78). Τέλος, η NucEnvDB παρέχει στο χρήστη μια τη δυνατότητα να πραγματοποιήσει ανάλυση λειτουργικού εμπλουτισμού («Functional Enrichment») με τη μέθοδο WebGestalt (Εικόνα 79). Η ανάλυση μπορεί να πραγματοποιηθεί τόσο για τα

Κεφάλαιο III – Αποτελέσματα-Συζήτηση

αποτελέσματα αναζητήσεων στη βάση όσο και για αποτελέσματα από την ανάλυση δικτύων, όπως περιγράφηκε παραπάνω. Ο λειτουργικός εμπλουτισμός μπορεί να πραγματοποιηθεί χρησιμοποιώντας όρους γονιδιακής οντολογίας από την Gene Ontology (Ashburner et al., 2000) ή πληροφορίες μεταβολικών μονοπατιών από τις βάσεις δεδομένων KEGG (Kanehisa & Goto, 2000) και REACTOME (Fabregat et al., 2018). Επιπρόσθετα, μπορεί να πραγματοποιηθεί η αναζήτηση συσχετίσεων με ασθένειες, αντλώντας πληροφορία από τις βάσεις DisGeNET (Pintero et al., 2017) και OMIM (Amberger & Hamosh, 2017) ή με φάρμακα, με πληροφορίες από τη βάση δεδομένων DrugBank (Wishart et al., 2018). Τα αποτελέσματα της ανάλυσης περιλαμβάνουν κάθε φορά τους λειτουργικούς όρους οι οποίοι βρέθηκαν να είναι στατιστικά σημαντικοί για το σύνολο των πρωτεϊνών που αναλύθηκαν.

Functional Enrichment Analysis

The screenshot shows the 'Functional Enrichment Analysis' web interface. It is divided into four steps:

- Step 1. Prepare Dataset:** A text box contains a list of Accession Codes (ACs) separated by spaces. A red box highlights this text box, with the red text label 'Σύνολο δεδομένων' (Dataset) to its right.
- Step 2. Select organism functional database for reference background:** A dropdown menu shows '--Organism--'. A red box highlights this dropdown, with the red text label 'Είδος οργανισμού' (Organism type) to its right.
- Step 3. Select database for analysis:** Radio buttons are used to select a database. 'Gene Ontology' is selected. A red box highlights this section, with the red text label 'κατηγορία πληροφορίας για εμπλουτισμό' (Information category for enrichment) to its right.
- Step 4. Define Sampling Parameters:** This section includes a 'Multiple Test Adjustment Method' dropdown (set to 'BH'), a 'Significance Level' section with 'FDR: 0.05' selected, a 'Number of categories visualized in the report' input (set to 20), and 'Annotated genes for each category' with 'Minimum: 10' and 'Maximum: 500'. A red box highlights this entire section, with the red text label 'Επιλογές στατιστικής ανάλυσης' (Statistical analysis options) to its right.

At the bottom, there are 'Submit' and 'Clear' buttons.

Εικόνα 79 Ανάλυση λειτουργικού εμπλουτισμού μέσω της NucEnvDB. Παρατίθεται η φόρμα εισαγωγής για τη διεξαγωγή ανάλυσης λειτουργικού εμπλουτισμού. Ο χρήστης αρχικά δίνει ως είσοδο μια σειρά από κωδικούς AC, οι οποίοι μπορεί να προέρχονται από τα αποτελέσματα αναζητήσεων στη βάση ή από τα αποτελέσματα της ανάλυσης δικτύου. Στη συνέχεια, ο χρήστης επιλέγει το είδος του οργανισμού που θα χρησιμοποιηθεί σαν υπόβαθρο στην ανάλυση, καθώς και το είδος της πληροφορίας που θα χρησιμοποιηθεί για εμπλουτισμό (όροι γονιδιακής οντολογίας, μεταβολικά μονοπάτια, συσχέτιση με ασθένειες ή συσχέτιση με φάρμακα). Τέλος, ο χρήστης μπορεί να ορίσει μια σειρά από επιλογές πάνω στη στατιστική ανάλυση που θα πραγματοποιηθεί.

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

Η βάση δεδομένων NucEnvDB αποτελεί, μέχρι στιγμής, τη μοναδική δημόσια διαθέσιμη πηγή εξειδικευμένης πληροφορίας πάνω στις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις στις οποίες συμμετέχουν. Σημειώνεται ότι το περιεχόμενο κάθε εγγραφής διατίθεται για λήψη σε τρεις διαφορετικούς τύπους αρχείου (TXT, FASTA και XML). Επιπρόσθετα, τόσο το σύνολο των πρωτεϊνών της βάσης όσο και μια σειρά από υποσύνολά του, ομαδοποιημένα με βάση την ταυτότητα των αμινοξικών αλληλουχιών τους με τον αλγόριθμο CD-HIT (Fu et al., 2012), είναι διαθέσιμα για λήψη στους τρεις παραπάνω τύπους αρχείων, μέσα από τη σελίδα «Μεταφορτώσεις» («Downloads»). Με αυτόν τον τρόπο, δίνεται στο χρήστη η δυνατότητα να διαχειριστεί προγραμματιστικά τα περιεχόμενα της βάσης, διευκολύνοντας έτσι την αυτοματοποιημένη ανάλυση των πρωτεϊνών του πυρηνικού φακέλου σε μεγάλη κλίμακα. Τέλος, μέσω των εργαλείων που διατίθενται στη βάση, καθίσταται δυνατή η ανάλυση των πρωτεϊνών του πυρηνικού φακέλου ως προς την ομολογία, τις αλληλεπιδράσεις και τη λειτουργικότητά τους, τόσο για τον άνθρωπο όσο και για άλλους, λιγότερο μελετημένους οργανισμούς.

Κεφάλαιο ΙΙΙ – Αποτελέσματα-Συζήτηση

ΚΕΦΑΛΑΙΟ IV

ΣΥΜΠΕΡΑΣΜΑΤΑ

Συνοπτικά, η παρούσα διδακτορική διατριβή επιχείρησε να μελετήσει το φαινόμενο των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών σε διαμεμβρανικές πρωτεΐνες, με την εφαρμογή μεθόδων από τα πεδία της Βιοπληροφορικής, της Υπολογιστικής Βιολογίας και της Θεωρητικής Βιοφυσικής. Η προσέγγιση αυτού του στόχου πραγματοποιήθηκε σε διάφορα επίπεδα, τόσο μικροσκοπικά, με τη μελέτη των βιομοριακών αλληλεπιδράσεων για μια σειρά από πρωτεΐνες στο επίπεδο της δομής όσο και μακροσκοπικά, με τη μελέτη του συνόλου των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου, ενός υποκυτταρικού διαμερίσματος που έχει δεχτεί ελάχιστη προσοχή από την επιστημονική κοινότητα. Παράλληλα, διερευνήθηκαν οι επιρροές της λιπιδικής διπλοστιβάδας, ένα χαρακτηριστικό που παρά τη σημασία του για τη δομή και τη λειτουργία των διαμεμβρανικών πρωτεϊνών, παραμένει ελάχιστα μελετημένο. Κλείνοντας, λοιπόν, την παρούσα διδακτορική διατριβή, θα επιχειρηθεί μία σύνοψη των σημαντικότερων συμπερασμάτων που προέκυψαν, σε όλα τα στάδια της μελέτης. Θα γίνει μια προσπάθεια ανάδειξης της σημασίας των μεθόδων που αναπτύχθηκαν και των αναλύσεων που διεξήχθησαν και, τέλος, θα τεθούν οι μελλοντικοί στόχοι και οι προεκτάσεις που θα μπορούσαν να έχουν όλες οι πτυχές της παρούσας μελέτης.

4.1 Μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs

Στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε η συστηματική μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs στο επίπεδο της δομής. Αρχικά, πραγματοποιήθηκε η μελέτη των διαμοριακών αλληλεπιδράσεων ολιγομερισμού των GPCRs, μέσα από εκτεταμένες προσομοιώσεις Μοριακής Δυναμικής σε ένα μεγάλο σύνολο διμερών υποδοχέων. Η μελέτη και ανάλυση τόσο των διαθέσιμων δομικών δεδομένων όσο και των

Κεφάλαιο IV - Συμπεράσματα

αποτελεσμάτων των προσομοιώσεων Μοριακής Δυναμικής αποκάλυψε μια σειρά από συντηρημένα χαρακτηριστικά στα διμερή των GPCRs. Συγκεκριμένα, οι αλληλεπιδράσεις ολιγομερισμού βρέθηκαν να επικεντρώνονται κυρίως σε περιοχές όπως οι διαμεμβρανικές έλικες TM1, TM4 και TM5 και η κυτοπλασματική έλικα H8 του καρβοξυτελικού άκρου των υποδοχέων. Αρωματικά κατάλοιπα, τα οποία συμμετέχουν σε επαφές π-ηλεκτρονίων, βρέθηκαν να σχηματίζουν τις πιο ισχυρές αλληλεπιδράσεις μεταξύ των πρωτομερών, με την πλειονότητά αυτών των αλληλεπιδράσεων να διατηρούνται σταθερές κατά τις προσομοιώσεις Μοριακής Δυναμικής. Επιπρόσθετα, η παρουσία αυτών των καταλοίπων σε συγκεκριμένες θέσεις, κοντά στα όρια των διαμεμβρανικών τμημάτων, υποδεικνύει τη συμμετοχή τους σε φαινόμενα που σχετίζονται με την υδροφοβική ασυμφωνία ανάμεσα στα υδρόφοβα τμήματα των διαμεμβρανικών πρωτεϊνών και το πάχος των βιολογικών μεμβρανών. Τέλος, τα αποτελέσματα των προσομοιώσεων Μοριακής Δυναμικής συνδυάστηκαν με μεθόδους ανάλυσης προερχόμενες από τα δίκτυα αλληλεπιδράσεων. Μέσω της ανάλυσης δικτύων των δυναμικών κινήσεων αποκαλύφθηκε η ύπαρξη συντηρημένων χαρακτηριστικών οργάνωσης στα διμερή των GPCRs. Επιπρόσθετα, παρατηρήθηκε θετική συσχέτιση της κινητικότητας ανάμεσα σε περιοχές που συμμετέχουν σε αλληλεπιδράσεις ολιγομερισμού και σε δομικά στοιχεία που σχετίζονται με τη λειτουργικότητα των GPCRs, όπως τα μοτίβα που συμμετέχουν στη διαδικασία ενεργοποίησης των υποδοχέων. Η παρουσία αυτών των δυναμικών συσχετίσεων προτείνει την ύπαρξη μηχανισμών ρύθμισης μέσω των οποίων οι αλληλεπιδράσεις ολιγομερισμού μπορούν να επηρεάσουν τη λειτουργία των υποδοχέων.

Η συντηρημένη παρουσία ισχυρών αλληλεπιδράσεων ολιγομερισμού στα όρια των διαμεμβρανικών τμημάτων, η οποία φάνηκε μέσα από την ανάλυση των παραπάνω προσομοιώσεων, οδήγησε στο ερώτημα του κατά πόσο το φαινόμενο της υδροφοβικής ασυμφωνίας μπορεί να επηρεάσει τη δομή και, ίσως, τη σταθερότητα των διαμοριακών αλληλεπιδράσεων στους GPCRs. Για τη διερεύνησή του πραγματοποιήθηκαν επιπρόσθετες προσομοιώσεις πάνω σε διμερή GPCRs, τα αποτελέσματα των οποίων χρησιμοποιήθηκαν για την ενεργειακή ανάλυση της υδροφοβικής ασυμφωνίας, ενώ παράλληλα διερευνήθηκε η θερμοδυναμική σταθερότητα των διμερών, μέσα από εξειδικευμένες προσομοιώσεις για τον υπολογισμό του Δυναμικού Μέσης Δύναμης. Τα αποτελέσματα αυτής της μελέτης έδειξαν ότι

Κεφάλαιο IV - Συμπεράσματα

οι περιοχές που συμμετέχουν σε αλληλεπιδράσεις ολιγομερισμού (TM1, TM4, TM5) αποτελούν, σε μονομερείς GPCRs, θέσεις με υψηλή υδροφοβική ασυμφωνία, στις οποίες παρατηρούνται μη ευνοϊκές αλληλεπιδράσεις ανάμεσα στα κατάλοιπα των διαμεμβρανικών τμημάτων και το περιβάλλον τους (μεμβράνη και διαλύτης). Αντίθετα, ο σχηματισμός αλληλεπιδράσεων ολιγομερισμού φάνηκε να μειώνει τις ενεργειακές ποινές αυτής της ασυμφωνίας, υποδεικνύοντας ότι η οργάνωση των GPCRs σε υπερμοριακά συμπλέγματα δρα ευεργετικά στη μείωση της υδροφοβικής ασυμφωνίας και, κατά συνέπεια, στην σταθεροποίηση της δομής των υποδοχέων στη μεμβράνη. Επιπρόσθετα, παρατηρήθηκε θετική συσχέτιση ανάμεσα στο βαθμό μείωσης της υδροφοβικής ασυμφωνίας και τη θερμοδυναμική σταθερότητα των διμερών GPCRs, όπως αυτή προσδιορίστηκε από το Δυναμικό Μέσης Δύναμης. Στο σύνολό τους, τα παραπάνω αποτελέσματα καταδεικνύουν τη σημασία του μεμβρανικού περιβάλλοντος, και των αλληλεπιδράσεών του με τους GPCRs, στην ικανότητα οργάνωσης και σταθεροποίησης των τελευταίων σε υπερμοριακά συμπλέγματα.

Τέλος, διερευνήθηκε η ικανότητα των GPCRs ως προς τον αυθόρμητο σχηματισμό ομο- και ετεροδιμερών, καθώς και η επίδραση της χοληστερόλης, ενός ιδιαίτερα σημαντικού συστατικού των μεμβρανών, μέσα από εκτεταμένες προσομοιώσεις των υποδοχέων A2A και mGluR5, δύο GPCRs με κλινική σημασία σε ασθένειες όπως η σχιζοφρένεια και η ασθένεια του Parkinson. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι η παρουσία της χοληστερόλης στη μεμβράνη, σε διαφορετικές συγκεντρώσεις, μπορεί να επηρεάσει τόσο την ικανότητα των υποδοχέων να σχηματίσουν διμερή όσο και το είδος των διαμοριακών αλληλεπιδράσεων που σχηματίζεται. Δεδομένου ότι η χοληστερόλη αποτελεί κρίσιμο συστατικό των μεμβρανών στα ευκαρυωτικά κύτταρα, το οποίο επηρεάζει τη δομή και τη λειτουργία μεγάλου αριθμού πρωτεϊνών συμπεριλαμβανομένων και των GPCRs, τα αποτελέσματα της διδακτορικής διατριβής μπορούν να φανούν ιδιαίτερα χρήσιμα στην περαιτέρω μελέτη του μηχανισμού με τον οποίο οι GPCRs και, ίσως, οι α-ελικοειδείς διαμεμβρανικές πρωτεΐνες γενικότερα, επηρεάζονται από τις μεμβρανικές στερόλες. Παράλληλα, τα μοντέλα των διμερών A2A και mGluR5, τα οποία προέκυψαν από την παραπάνω μελέτη, μπορούν να χρησιμοποιηθούν στη μελέτη των ασθενειών που σχετίζονται με αυτούς τους υποδοχείς και, πιθανώς, στο σχεδιασμό νέων φαρμάκων για την αντιμετώπισή τους.

Κεφάλαιο IV - Συμπεράσματα

Συνολικά, τα αποτελέσματα από τη μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs αναδεικνύουν τόσο τα δομικά γνωρίσματα αυτών των αλληλεπιδράσεων όσο και τη λειτουργική σημασία τους, ενώ παράλληλα ρίχνουν φως σε στοιχεία του μηχανισμού με τον οποίο το μεμβρανικό περιβάλλον επηρεάζει αυτές τις αλληλεπιδράσεις. Αξίζει να σημειωθεί ότι η παρούσα μελέτη αποτελεί την πρώτη προσπάθεια για διερεύνηση της δυναμικής συμπεριφοράς του ολιγομερισμού των GPCRs σε συλλογικό επίπεδο. Δεδομένου του συνεχώς αυξανόμενου ενδιαφέροντος για τη μελέτη και αποσαφήνιση της δομικής φύσης και βιολογικής σημασίας του ολιγομερισμού και των επιδράσεων του στη φυσιολογική και μη λειτουργία των υποδοχέων, τα αποτελέσματα της παρούσας εργασίας μπορούν να βρουν εφαρμογή στο σχεδιασμό πειραματικών μελετών πάνω σε διμερή ή ολιγομερή GPCRs, καθώς και στη μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στις διαμεμβρανικές πρωτεΐνες γενικότερα.

4.2 Μελέτη της Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων

Στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε εκτεταμένη μελέτη των δομικών και δυναμικών στοιχείων της Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων, τόσο ως προς τα λιπίδια που την αποτελούν, όσο και ως προς τις πρωτεΐνες της, τα διαμεμβρανικά β-βαρέλια. Συγκεκριμένα, πραγματοποιήθηκε η μελέτη της δομής και της δυναμικής των λιποπολυσακχαριτών, ενός τύπου λιπιδίων που εμφανίζεται αποκλειστικά στα αρνητικά κατά Gram βακτήρια, ενώ παράλληλα κατασκευάστηκε το εργαλείο GNOMM, μια αυτοματοποιημένη μέθοδος για την προτυποποίηση μοντέλων της Εξωτερικής Μεμβράνης. Επιπρόσθετα, πραγματοποιήθηκε μελέτη των αλληλεπιδράσεων ανάμεσα στα λιπίδια της Εξωτερικής Μεμβράνης και σε διαμεμβρανικά β-βαρέλια, ενώ, τέλος, διερευνήθηκαν οι αλληλεπιδράσεις ολιγομερισμού των β-βαρελίων, αλλά και η πιθανή επίδραση της παρουσίας λιποπολυσακχαριτών σε αυτές τις αλληλεπιδράσεις.

Το πρώτο μέρος της παρούσας ενότητας επικεντρώθηκε στη μελέτη της Εξωτερικής Μεμβράνης και, συγκεκριμένα, στην κατασκευή μιας μεθόδου για την προτυποποίηση της, με

Κεφάλαιο IV - Συμπεράσματα

γνώμονα τις προσομοιώσεις Μοριακής Δυναμικής. Οι τελευταίες έχουν καθιερωθεί πλέον ως μια από τις πιο αποτελεσματικές μεθόδους διερεύνησης της δομής, της δυναμικής και της λειτουργίας των βιολογικών μεμβρανών και των διαμεμβρανικών πρωτεϊνών. Η παραπάνω παρατήρηση καταδεικνύεται από το συνεχώς αυξανόμενο αριθμό προσομοιώσεων μεμβρανών και μεμβρανικών πρωτεϊνών στη βιβλιογραφία, συμπεριλαμβανομένων των προσομοιώσεων που επικεντρώνονται στα διαμεμβρανικά β-βαρέλια. Παρά την εξέλιξη του πεδίου, ωστόσο, η πρόοδος όσον αφορά τη σύσταση της λιπιδικής διπλοστιβάδας είναι μικρή, καθώς οι προσομοιώσεις των μεμβρανών περιορίζονται σε μεγάλο βαθμό στην προτυποποίηση απλών ομοιογενών διπλοστιβάδων. Αντίθετα, η προτυποποίηση πιο «ρεαλιστικών», σύνθετων μεμβρανών, συμπεριλαμβανομένων των Εξωτερικών Μεμβρανών των αρνητικών κατά Gram βακτηρίων, αποτελεί μια πολύ πρόσφατη εξέλιξη στο χώρο των προσομοιώσεων. Παρ'ότι κάποιες ιδιότητες των Εξωτερικών Μεμβρανών, όπως η σταθερότητα των διαμεμβρανικών β-βαρελίων, μπορούν να απεικονιστούν, ως ένα βαθμό, μέσα από προσομοιώσεις σε απλές, φωσφολιπιδικές διπλοστιβάδες, η μελέτη πιο εξειδικευμένων χαρακτηριστικών απαιτεί τη ρεαλιστική απεικόνιση του μεμβρανικού περιβάλλοντος των αρνητικών κατά Gram βακτηρίων, με χαρακτηριστικά όπως οι λιποπολυσακχαρίτες και η μη συμμετρική δομή της Εξωτερικής Μεμβράνης να είναι κρίσιμα.

Έτσι, είναι αναγκαία τόσο η δημιουργία νέων παραμέτρων στα δυναμικά πεδία για μόρια όπως οι λιποπολυσακχαρίτες, όσο και η δημιουργία μεθόδων για την κατασκευή κατάλληλων συστημάτων προσομοίωσης για τις Εξωτερικές Μεμβράνες. Επιπρόσθετα, είναι αναγκαίο αυτά τα συστήματα να είναι υπολογιστικά και θερμοδυναμικά σταθερά και, ταυτόχρονα, να μπορούν να προτυποποιήσουν ορθά τόσο τις ιδιότητες της Εξωτερικής Μεμβράνης όσο και τις αλληλεπιδράσεις πρωτεϊνών – λιπιδίων στα διαμεμβρανικά β-βαρέλια. Η μέθοδος GNOMM (<http://bioinformatics.biol.uoa.gr/GNOMM>), η οποία σχεδιάστηκε με στόχο την προτυποποίηση Εξωτερικών Μεμβρανών για προσομοιώσεις Μοριακής Δυναμικής, εκπληρώνει τους παραπάνω στόχους. Όπως φάνηκε και από τα αποτελέσματα των προσομοιώσεων της παρούσας διατριβής, τα συστήματα Εξωτερικών Μεμβρανών που προτυποποιούνται από το εργαλείο GNOMM παρουσιάζουν συμφωνία με τα υπάρχοντα πειραματικά δεδομένα, τόσο ως προς τις ιδιότητες των λιποπολυσακχαριτών, όσο και ως προς

Κεφάλαιο IV - Συμπεράσματα

τις ιδιότητες της μεμβράνης γενικότερα. Επιπρόσθετα, οι προσομοιώσεις διαμεμβρανικών β-βαρελίων σε μεμβράνες κατασκευασμένες με το GNOMM έδειξαν σημαντική σταθερότητα για τις πρωτεΐνες αλλά και βιολογικά σημαντικές αλληλεπιδράσεις ανάμεσα στις πρωτεΐνες και τους λιποπολυσακχαρίτες, καταδεικνύοντας την ικανότητα του εργαλείου στην προτυποποίηση θερμοδυναμικά ορθών συστημάτων. Τα παραπάνω αποτελέσματα παρατηρήθηκαν τόσο στα πεδία ατομικής διακριτικότητας, όπως το CHARMM και το GROMOS, όσο και σε πεδία αδρής διακριτικότητας όπως το MARTINI και το PACE, καθώς παρατηρήθηκε συμφωνία στα αποτελέσματα των δυναμικών πεδίων για τους λιποπολυσακχαρίτες αλλά και για τις πρωτεΐνες. Η παραπάνω παρατήρηση είναι ιδιαίτερα σημαντική, καθώς δείχνει ότι τα χαρακτηριστικά των Εξωτερικών Μεμβρανών και των β-βαρελίων μπορούν να προτυποποιηθούν τόσο από λεπτομερή όσο και από απλοποιημένα μοντέλα με σχεδόν παραπλήσιο βαθμό ακρίβειας. Συνολικά, όλα τα αποτελέσματα της διδακτορικής διατριβής δείχνουν ότι η μέθοδος GNOMM είναι κατάλληλη για τη σωστή προτυποποίηση και προσομοίωση των Εξωτερικών Μεμβρανών και των περιεχομένων τους, παρέχοντας στο χρήστη μια αυτοματοποιημένη διαδικασία για την εύκολη μελέτη της δομής των λιποπολυσακχαριτών και των διαμεμβρανικών β-βαρελίων.

Το δεύτερο μέρος της παρούσας ενότητας επικεντρώθηκε στις αλληλεπιδράσεις ολιγομερισμού των διαμεμβρανικών β-βαρελίων, μέσα από προσομοιώσεις Μοριακής Δυναμικής σε αντιπροσωπευτικά β-βαρέλια με γνωστή ικανότητα ολιγομερισμού. Τα αποτελέσματα των προσομοιώσεων έδειξαν ότι τα διαμεμβρανικά β-βαρέλια υιοθετούν συντηρημένα χαρακτηριστικά στις αλληλεπιδράσεις ολιγομερισμού τους, μέσω του σχηματισμού ομοτριμερών στο επίπεδο της μεμβράνης. Επιπρόσθετα, διερευνήθηκε η πιθανή επίδραση της παρουσίας λιποπολυσακχαριτών στη σταθερότητα αυτών των αλληλεπιδράσεων, μέσα από εξειδικευμένες προσομοιώσεις του τριμερούς της OmpF και από υπολογισμούς του Δυναμικού Μέσης Δύναμης. Τα αποτελέσματα έδειξαν ότι η σταθερότητα των υπερμοριακών συμπλόκων του βαρελίου εξαρτώνται σε μεγάλο βαθμό από τη σύνθεση της εξωτερικής στιβάδας της μεμβράνης, με την παρουσία λιποπολυσακχαριτών να αποτελεί σημαντικό στοιχείο στη σταθεροποίηση των αλληλεπιδράσεων ολιγομερισμού. Δεδομένου ότι η αρχιτεκτονική του ομοτριμερούς της OmpF φαίνεται να είναι κοινή για την υπεροικογένεια

Κεφάλαιο IV - Συμπεράσματα

των τριμερών πορινών, είναι πιθανό η επίδραση των λιποπολυσακχαριτών στη σταθερότητα των αλληλεπιδράσεων ολιγομερισμού να μην περιορίζεται για το παρόν σύστημα μελέτης αλλά να ισχύει για το σύνολό των τριμερών β-βαρελίων.

4.3 Μελέτη της δομής του διαμεμβρανικού μεταφορέα SLC11A1

Στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε η μελέτη των δομικών και δυναμικών χαρακτηριστικών του μεταφορέα δισθενών μεταλλικών ιόντων SLC11A1, μέσα από την προτυποποίηση και προσομοίωση της τρισδιάστατης δομής του σε διαφορετικές στερεοδιατάξεις, καθώς και μέσα από τη μελέτη των αλληλεπιδράσεων του μεταφορέα με το υπόστρωμά του, τα ιόντα Fe^{2+} και Mn^{2+} , μέσα από ενεργειακούς υπολογισμούς για τον φυσικό τύπο αλλά και για μεταλλαγές της πρωτεΐνης. Τα αποτελέσματα της παραπάνω μελέτης προτείνουν μια τρισδιάστατη δομή για τον SLC11A1 τόσο στην ανοιχτή εξωκυτταρικά όσο και στην ανοιχτή ενδοκυτταρικά στερεοδιάταξή του και προτείνουν μια σειρά από στερεοδιαταξικές αλλαγές του οποίου ο μεταφορέας μεταβαίνει ανάμεσα στις δύο καταστάσεις και επιτελεί τη λειτουργία του. Παράλληλα, οι ενεργειακοί υπολογισμοί των αλληλεπιδράσεων του SLC11A1 με το υπόστρωμά του αποκαλύπτουν τη δομή και την ενεργειακή συμπεριφορά της θέσης πρόσδεσης των ιόντων, ενώ συμπληρωματικοί υπολογισμοί πάνω σε μεταλλαγές του μεταφορέα στη θέση αυτή προτείνουν μια σειρά από αμινοξικά κατάλοιπα που διαδραματίζουν σημαντικό ρόλο στη λειτουργία του SLC11A1, τόσο ως προς την αναγνώριση και πρόσδεση των ιόντων όσο και ως προς την επιλεκτικότητα του μεταφορέα απέναντι στο υπόστρωμά του. Στο σύνολό τους, τα παραπάνω αποτελέσματα αποκαλύπτουν στοιχεία του μηχανισμού λειτουργίας του SLC11A1 και, πιθανώς, του συνόλου των μεταφορέων SLC11 γενικότερα, σε δομικό επίπεδο, κάτι το οποίο δεν έχει πραγματοποιηθεί για άλλους ευκαρυωτικούς μεταφορείς SLC11 μέχρι σήμερα. Δεδομένης της κλινικής σημασίας του SLC11A1 σε διάφορες ασθένειες, τόσο στον άνθρωπο όσο και σε άλλους ζωικούς οργανισμούς, τα παραπάνω αποτελέσματα μπορούν να φανούν ιδιαίτερα χρήσιμα στην περαιτέρω μελέτη της δομής του μεταφορέα, της εμπλοκής του σε παθολογικές καταστάσεις και, πιθανώς, στο σχεδιασμό φαρμάκων γι αυτόν.

4.4 Κατασκευή μεθόδου πρόγνωσης και σχολιασμού των RTKs

Για τη μελέτη των RTKs αναπτύχθηκε, στα πλαίσια της διδακτορικής διατριβής, ο αλγόριθμος RTK-PRED (<http://bioinformatics.biol.uoa.gr/RTK-PRED>), μια μέθοδος πρόγνωσης και σχολιασμού των RTKs από την αμινοξική αλληλουχία. Η μέθοδος είναι ικανή τόσο για τον εντοπισμό πιθανών RTKs σε ολόκληρα πρωτεώματα, όσο και για την εύρεση της τοπολογίας τους αλλά και το λειτουργικό σχολιασμό τους. Τα αποτελέσματα από την αξιολόγηση της μεθόδου έδειξαν σημαντικά υψηλές τιμές ευαισθησίας, ειδικότητας και αξιοπιστίας για το RTK-PRED. Παράλληλα, η εφαρμογή του αλγορίθμου σε μια σειρά από αντιπροσωπευτικά πρωτεώματα αναφοράς αποκάλυψε τόσο την ύπαρξη πιθανών RTKs σε οργανισμούς προερχόμενους από τα Μετάζωα, τα Πρωτόζωα και τα Φυτά, όσο και την λειτουργική κατανομή τους στις παραπάνω κατηγορίες πρωτεωμάτων. Πρέπει να σημειωθεί ότι το RTK-PRED αποτελεί, μέχρι στιγμής τη μοναδική διαθέσιμη μέθοδο για την πρόγνωση και το σχολιασμό RTKs από την αλληλουχία. Δεδομένης της σημασίας των RTKs στη φυσιολογία πολλών ευκαρυωτικών οργανισμών, η μέθοδος RTK-PRED μπορεί να φανεί ιδιαίτερα χρήσιμη στον αυτοματοποιημένο σχολιασμό ελάχιστα μελετημένων πρωτεωμάτων.

4.5 Μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο

Στα πλαίσια της παρούσας διδακτορικής διατριβής πραγματοποιήθηκε εκτεταμένη ανάλυση των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στον πυρηνικό φάκελο, ένα υποκυτταρικό διαμέρισμα το οποίο, παρά τη μεγάλη σημασία του για το κύτταρο, έχει μελετηθεί ελάχιστα μέχρι σήμερα. Η μελέτη των αλληλεπιδράσεων πραγματοποιήθηκε σε επίπεδο συστήματος, αρχικά μέσα από την κατασκευή και ανάλυση του δικτύου αλληλεπιδράσεων των πρωτεϊνών του ανθρώπινου πυρηνικού φακέλου. Επιπρόσθετα, στα πλαίσια της μελέτης αυτών των αλληλεπιδράσεων, σχεδιάστηκε η βάση δεδομένων NucEnvDB, μια διαδικτυακή πηγή πληροφοριών για τις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις τους στο σύνολο των διαθέσιμων ευκαρυωτικών πρωτεωμάτων.

Κεφάλαιο IV - Συμπεράσματα

Τα αποτελέσματα από τη συστηματική μελέτη των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου έδειξαν ότι οι τελευταίες αποτελούν κρίσιμο κομμάτι του συνόλου των αλληλεπιδράσεων στο ανθρώπινο πρωτόμα. Το δίκτυο αλληλεπιδράσεων του πυρηνικού φακέλου διαθέτει τα χαρακτηριστικά πραγματικών βιολογικών δικτύων, υιοθετώντας μια ανεξάρτητη από κλίμακα τοπολογία. Οι πρωτεΐνες του πυρηνικού φακέλου βρέθηκαν να αποτελούν ιδιαίτερα σημαντικούς κόμβους, αλληλεπιδρώντας τόσο με στοιχεία στο εσωτερικό του πυρήνα και το κυτταρόπλασμα όσο και με πρωτεΐνες από το σύνολο σχεδόν των υποκυτταρικών διαμερισμάτων, συμπεριλαμβανομένης της κυτταρικής μεμβράνης, των οργανιδίων και των μεμβρανικών κυστιδίων. Παράλληλα, η λειτουργική ανάλυση του δικτύου έδειξε ότι οι πρωτεΐνες του πυρηνικού φακέλου συμμετέχουν σε μια σειρά από βιολογικές διεργασίες και μοριακές λειτουργίες κρίσιμες για τη λειτουργία του κυττάρου. Σημειώνεται ότι η παρούσα μελέτη αποτελεί την πρώτη συστηματική μελέτη των πρωτεϊνών του πυρηνικού φακέλου σε επίπεδο συστήματος.

Η βάση δεδομένων NucEnvDB (<http://bioinformatics.biol.uoa.gr/nucenvdb>) είναι μια δημόσια διαθέσιμη πηγή πληροφοριών για τις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις στις οποίες συμμετέχουν. Η βάση περιέχει το σύνολο των καλά σχολιασμένων πρωτεϊνών με γνωστή παρουσία στον πυρηνικό φάκελο, οργανωμένο σε μια εύκολη στην πρόσβαση από το χρήστη εφαρμογή. Τα δεδομένα της βάσης μπορούν να αξιοποιηθούν για τη συστηματική μελέτη των πρωτεϊνών του πυρηνικού φακέλου όχι μόνο στον άνθρωπο αλλά και σε μεγάλη ποικιλία ευκαρυωτικών οργανισμών. Παράλληλα, η βάση προσφέρει στο χρήστη μια σειρά από εργαλεία για τη μελέτη των πρωτεϊνών του πυρηνικού φακέλου, τόσο για την κατασκευή και ανάλυση δικτύων αλληλεπιδράσεων όσο και για το λειτουργικό σχολιασμό τους. Σημειώνεται ότι η NucEnvDB αποτελεί τη μοναδική, μέχρι στιγμής, διαθέσιμη βάση δεδομένων πάνω στα περιεχόμενα του πυρηνικού φακέλου.

4.6 Μελλοντικές κατευθύνσεις

Όλες οι κυτταρικές διεργασίες σε όλους ανεξαιρέτως τους οργανισμούς, από τα βακτήρια ως τα ανώτερα θηλαστικά και τον άνθρωπο, επιτελούνται στη συντριπτική

Κεφάλαιο IV - Συμπεράσματα

πλειονότητά τους από τις πρωτεΐνες του κυττάρου. Οι τελευταίες δεν δρουν μεμονωμένα αλλά συνεργατικά, σχηματίζοντας υπερμοριακά σύμπλοκα μέσα από αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών, οδηγώντας σε ένα πυκνό δίκτυο αλληλεπιδράσεων, το οποίο είναι υπεύθυνο για την εύρυθμη λειτουργία του κυττάρου. Έτσι, η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών, τόσο σε επίπεδο δομής όσο και συνολικά, σε επίπεδο συστήματος, είναι απαραίτητη για την αποσαφήνιση του λειτουργικού ρόλου των πρωτεϊνών. Στην περίπτωση των διαμεμβρανικών πρωτεϊνών, ωστόσο, αυτή η μελέτη παρουσιάζει δυσκολίες, τόσο στον προσδιορισμό της δομής των διαμεμβρανικών πρωτεϊνών και των υπερμοριακών συμπλόκων τους όσο και στη μελέτη των επιδράσεων της λιπιδικής διπλοστιβάδας. Με βάση τα παραπάνω, τα αποτελέσματα της παρούσας διδακτορικής διατριβής μπορούν να φανούν ιδιαίτερα χρήσιμα στην περαιτέρω μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών στο επίπεδο της μεμβράνης.

Τα αποτελέσματα από τη μελέτη των αλληλεπιδράσεων ολιγομερισμού στους GPCRs και στα διαμεμβρανικά β-βαρέλια αποκάλυψαν μια σειρά από συντηρημένα χαρακτηριστικά στα υπερμοριακά συμπλέγματά τους, καθώς και στοιχεία των μηχανισμών με τους οποίους αυτές οι αλληλεπιδράσεις επηρεάζουν τη λειτουργία των πρωτεϊνών. Σημειώνεται ότι τόσο για τους GPCRs όσο και για τα β-βαρέλια, η φύση και η σταθερότητα των αλληλεπιδράσεων βρέθηκε να εξαρτάται όχι μόνο από τα αμινοξικά κατάλοιπα των ίδιων των πρωτεϊνών αλλά και από τις ιδιότητες της λιπιδικής διπλοστιβάδας και τις αλληλεπιδράσεις ανάμεσα στις πρωτεΐνες και τα λιπίδια γύρω τους. Μια μελλοντική κατεύθυνση, λοιπόν, είναι η περαιτέρω μελέτη των επιδράσεων της μεμβράνης πάνω στη δομή και τις αλληλεπιδράσεις τόσο των υποδοχέων όσο και των β-βαρελίων, μέσα από πιο εξειδικευμένη μελέτη της δομής και της δυναμικής τους σε μεμβράνες με μεγαλύτερη ποικιλία λιπιδικών μορίων, προκειμένου να αποσαφηνιστούν οι επιρροές που δέχονται οι διαμεμβρανικές πρωτεΐνες από το μεμβρανικό περιβάλλον γύρω τους. Απώτερος στόχος είναι η πιθανή εφαρμογή αυτών των αποτελεσμάτων στο σχεδιασμό μιας μεθόδου πρόγνωσης για τις αλληλεπιδράσεις των διαμεμβρανικών πρωτεϊνών, η οποία θα μπορεί να λαμβάνει υπόψη και τις επιρροές του μεμβρανικού περιβάλλοντος.

Κεφάλαιο IV - Συμπεράσματα

Η μέθοδος GNOMM αποτελεί ένα ιδιαίτερα χρήσιμο εργαλείο για τη μελέτη της Εξωτερικής Μεμβράνης των αρνητικών κατά Gram βακτηρίων τόσο ως προς τη δομή όσο και ως προς τη δυναμική της. Η μέθοδος αποτελεί μέχρι στιγμής το μοναδικό διαθέσιμο εργαλείο που επιτρέπει την κατασκευή Εξωτερικών Μεμβρανών σε τέσσερα διαφορετικά δυναμικά πεδία, προσφέροντας μια σειρά από μόρια λιποπολυσακχαριτών. Μελλοντικός στόχος είναι η περαιτέρω βελτιστοποίηση και εξέλιξη της μεθόδου, με τον εμπλουτισμό της με περισσότερα μόρια λιποπολυσακχαριτών και άλλων λιπιδίων, καθώς και με την υποστήριξη περισσότερων δυναμικών πεδίων, αυξάνοντας έτσι τις επιλογές του χρήστη στην κατασκευή συστημάτων προσομοίωσης.

Η μελέτη της δομής του μεταφορέα SLC11A1 προσφέρει μια σειρά από χρήσιμα δεδομένα για την αποσαφήνιση της λειτουργίας του, καθώς και της εμπλοκής του σε ασθένειες. Δεδομένης της συντηρημένης τοπολογίας των μεταφορέων SLC11, κάποιες από τις παρατηρήσεις πάνω στη δομή και τη λειτουργία του SLC11A1 μπορούν να εφαρμοστούν, ως ένα βαθμό, στη μελέτη και των υπόλοιπων μελών της οικογένειας. Έτσι, ένας μελλοντικός στόχος είναι η μελέτη περισσότερων μεταφορέων SLC11, τόσο σε ευκαρυωτικούς όσο και προκαρυωτικούς οργανισμούς, με σκοπό την αποσαφήνιση του γενικού μηχανισμού λειτουργίας των SLC11.

Η μέθοδος RTK-PRED αποτελεί ένα χρήσιμο εργαλείο για τον εντοπισμό RTKs σε μεγάλο αριθμό αλληλουχιών και, έτσι, μπορεί να χρησιμοποιηθεί στον αυτοματοποιημένο σχολιασμό πρωτεωμάτων. Ήδη στα πλαίσια της διδακτορικής διατριβής πραγματοποιήθηκε η μελέτη μιας σειράς από πρωτεώματα για την εύρεση πιθανών RTKs. Απώτερος στόχος είναι η μελέτη του συνόλου των διαθέσιμων πρωτεωμάτων και γονιδιωμάτων, και η πιθανή ενσωμάτωση των αποτελεσμάτων σε μια δημόσια διαθέσιμη βάση δεδομένων, εξειδικευμένη για τους RTKs.

Τέλος, η μελέτη των αλληλεπιδράσεων των πρωτεϊνών του πυρηνικού φακέλου αποκαλύπτει μια σειρά από λειτουργικά γνωρίσματα για το συγκεκριμένο υποκυτταρικό διαμέρισμα. Παράλληλα, η βάση δεδομένων NucEnvDB αποτελεί ένα χρήσιμο αποθετήριο πληροφοριών πάνω στις πρωτεΐνες του πυρηνικού φακέλου. Ωστόσο, σε μεγάλο βαθμό τα χαρακτηριστικά και η λειτουργική σημασία του πυρηνικού φακέλου παραμένουν ανεξερεύνητα, ενώ άλλοι οργανισμοί πέρα από τον άνθρωπο παραμένουν ελάχιστα

Κεφάλαιο IV - Συμπεράσματα

μελετημένοι ως προς το συγκεκριμένο χαρακτηριστικό. Έτσι, μελλοντικός στόχος είναι η επέκταση της μελέτης των αλληλεπιδράσεων του πυρηνικού φακέλου σε άλλους οργανισμούς, καθώς και η περαιτέρω ενημέρωση και εξέλιξη της βάσης NucEnvDB.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19-25.
- Abu-Farha, M., Elisma, F., Figeys, D., 2008. Identification of protein-protein interactions by mass spectrometry coupled techniques. *Advances in biochemical engineering/biotechnology* 110, 67-80.
- Akira, S., Uematsu, S., Takeuchi, O., 2006. Pathogen recognition and innate immunity. *Cell* 124, 783-801.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. *Molecular Biology of the Cell*. 4th ed. Garland Science, New York.
- Alexander, S.P., Fabbro, D., Kelly, E., Marrion, N.V., Peters, J.A., Faccenda, E., Harding, S.D., Pawson, A.J., Sharman, J.L., Southan, C., Davies, J.A., 2017. THE CONCISE GUIDE TO PHARMACOLOGY 2017/18: Catalytic receptors. *British journal of pharmacology* 174 Suppl 1, S225-S271.
- Allen, W.J., Lemkul, J.A., Bevan, D.R., 2009. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *Journal of computational chemistry* 30, 1952-1958.
- Almen, M.S., Nordstrom, K.J., Fredriksson, R., Schioth, H.B., 2009. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC biology* 7, 50.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *Journal of molecular biology* 215, 403-410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 25, 3389-3402.
- Amberger, J.S., Hamosh, A., 2017. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. *Current protocols in bioinformatics* 58, 1 2 1-1 2 12.
- Arunmanee, W., Pathania, M., Solovyova, A.S., Le Brun, A.P., Ridley, H., Basle, A., van den Berg, B., Lakey, J.H., 2016. Gram-negative trimeric porins have specific LPS binding sites that are essential for porin biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 113, E5034-5043.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25, 25-29.

Βιβλιογραφία

- Assenov, Y., Ramirez, F., Schelhorn, S.E., Lengauer, T., Albrecht, M., 2008. Computing topological parameters of biological networks. *Bioinformatics* 24, 282-284.
- Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., Hamodrakas, S.J., 2004. PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic acids research* 32, W400-404.
- Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509-512.
- Barabasi, A.L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews. Genetics* 12, 56-68.
- Bartke, N., Hannun, Y.A., 2009. Bioactive sphingolipids: metabolism and function. *Journal of lipid research* 50 Suppl, S91-96.
- Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P., Whittle, H.C., Hill, A.V., 1998. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N Engl J Med* 338, 640-644.
- Bello, G., Eriksson, J., Terry, A., Edwards, K., Lawrence, M.J., Barlow, D., Harvey, R.D., 2015. Characterization of the aggregates formed by various bacterial lipopolysaccharides in solution and upon interaction with antimicrobial peptides. *Langmuir : the ACS journal of surfaces and colloids* 31, 741-751.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., 1984. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics* 81, 3684-3690.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. *Nucleic acids research* 28, 235-242.
- Biarç, J., Chalkley, R.J., Burlingame, A.L., Bradshaw, R.A., 2011. Receptor tyrosine kinase signaling--a proteomic perspective. *Advances in enzyme regulation* 51, 293-305.
- Binnig, G., Quate, C.F., Gerber, C., 1986. Atomic force microscope. *Physical review letters* 56, 930-933.
- Birgbauer, E., Chun, J., 2006. New developments in the biological functions of lysophospholipids. *Cellular and molecular life sciences : CMLS* 63, 2695-2701.
- Bkaily, G., Avedanian, L., Jacques, D., 2009. Nuclear membrane receptors and channels as targets for drug development in cardiovascular diseases. *Canadian journal of physiology and pharmacology* 87, 108-119.
- Bkaily, G., Avedanian, L., Al-Khoury, J., Ahmarani, L., Perreault, C., Jacques, D., 2012. Receptors and ionic transporters in nuclear membranes: new targets for therapeutical pharmacological interventions. *Canadian journal of physiology and pharmacology* 90, 953-965.
- Bocharov, E.V., Mineev, K.S., Volynsky, P.E., Ermolyuk, Y.S., Tkach, E.N., Sobol, A.G., Chupin, V.V., Kirpichnikov, M.P., Efremov, R.G., Arseniev, A.S., 2008. Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *The Journal of biological chemistry* 283, 6950-6956.
- Bonne, G., 2014. Nuclear envelope proteins in health and diseases. *Seminars in cell & developmental biology* 29, 93-94.
- Borroto-Escuela, D.O., Brito, I., Romero-Fernandez, W., Di Palma, M., Oflijan, J., Skieterska, K., Duchou, J., Van Craenenbroeck, K., Suarez-Boomgaard, D., Rivera, A., Guidolin, D., Agnati, L.F., Fuxe, K., 2014. The G protein-coupled receptor heterodimer network

Βιβλιογραφία

- (GPCR-HetNet) and its hub components. *International journal of molecular sciences* 15, 8570-8590.
- Bos, M.P., Robert, V., Tommassen, J., 2007. Biogenesis of the gram-negative bacterial outer membrane. *Annual review of microbiology* 61, 191-214.
- Bozzi, A.T., Bane, L.B., Weihofen, W.A., Singharoy, A., Guillen, E.R., Ploegh, H.L., Schulten, K., Gaudet, R., 2016a. Crystal Structure and Conformational Change Mechanism of a Bacterial Nramp-Family Divalent Metal Transporter. *Structure* 24, 2102-2114.
- Bozzi, A.T., Bane, L.B., Weihofen, W.A., McCabe, A.L., Singharoy, A., Chipot, C.J., Schulten, K., Gaudet, R., 2016b. Conserved methionine dictates substrate preference in Nramp-family divalent metal transporters. *Proceedings of the National Academy of Sciences of the United States of America* 113, 10310-10315.
- Bridger, J.M., Foeger, N., Kill, I.R., Herrmann, H., 2007. The nuclear lamina. Both a structural framework and a platform for genome organization. *The FEBS journal* 274, 1354-1361.
- Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry* 4, 187-217.
- Brooks, B.R., Brooks, C.L., 3rd, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M., Karplus, M., 2009. CHARMM: the biomolecular simulation program. *Journal of computational chemistry* 30, 1545-1614.
- Brünger, A.T. 1992. X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry.
- Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. *The Journal of chemical physics* 126, 014101.
- Cabello, N., Gandia, J., Bertarelli, D.C., Watanabe, M., Lluís, C., Franco, R., Ferre, S., Lujan, R., Ciruela, F., 2009. Metabotropic glutamate type 5, dopamine D2 and adenosine A2a receptors form higher-order oligomers in living cells. *Journal of neurochemistry* 109, 1497-1507.
- Cabrera-Vera, T.M., Vanhauwe, J., Thomas, T.O., Medkova, M., Preininger, A., Mazzoni, M.R., Hamm, H.E., 2003. Insights into G protein structure, function, and regulation. *Endocrine reviews* 24, 765-781.
- Caffrey, M., 2003. Membrane protein crystallization. *Journal of structural biology* 142, 108-132.
- Carpenter, G., Liao, H.J., 2013. Receptor tyrosine kinases in the nucleus. *Cold Spring Harbor perspectives in biology* 5, a008979.
- Ceccarelli, M., Danelon, C., Laio, A., Parrinello, M., 2004. Microscopic Mechanism of Antibiotics Translocation through a Porin. *Biophysical journal* 87, 58-64.
- Cellier, M., Govoni, G., Vidal, S., Kwan, T., Groulx, N., Liu, J., Sanchez, F., Skamene, E., Schurr, E., Gros, P., 1994. Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. *J Exp Med* 180, 1741-1752.
- Cellier, M.F., 2012. Nramp: from sequence to structure and mechanism of divalent metal import. *Curr Top Membr* 69, 249-293.

Βιβλιογραφία

- Chapman, M.J., Margulis, L., 1998. Morphogenesis by symbiogenesis. *International microbiology : the official journal of the Spanish Society for Microbiology* 1, 319-326.
- Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., Stark, C., Breitkreutz, B.J., Dolinski, K., Tyers, M., 2017. The BioGRID interaction database: 2017 update. *Nucleic acids research* 45, D369-D379.
- Cherezov, V., Abola, E., Stevens, R.C., 2010. Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods Mol Biol* 654, 141-168.
- Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K., Stevens, R.C., 2007. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318, 1258-1265.
- Chothia, C., 1976. The nature of the accessible and buried surfaces in proteins. *Journal of molecular biology* 105, 1-12.
- Chothia, C., Janin, J., 1975. Principles of protein-protein recognition. *Nature* 256, 705-708.
- Chothia, C., Hubbard, T., Brenner, S., Barns, H., Murzin, A., 1997. Protein folds in the all-beta and all-alpha classes. *Annual review of biophysics and biomolecular structure* 26, 597-627.
- Chou, K.C., Shen, H.B., 2007. MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochemical and biophysical research communications* 360, 339-345.
- Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422-1423.
- Crisp, M., Liu, Q., Roux, K., Rattner, J.B., Shanahan, C., Burke, B., Stahl, P.D., Hodzic, D., 2006. Coupling of the nucleus and cytoplasm: role of the LINC complex. *The Journal of cell biology* 172, 41-53.
- Dalrymple, M.B., Pflieger, K.D., Eidne, K.A., 2008. G protein-coupled receptor dimers: functional consequences, disease states and drug targets. *Pharmacology & therapeutics* 118, 359-371.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of chemical physics* 98, 10089-10092.
- de Jong, D.H., Singh, G., Bennett, W.F., Arnarez, C., Wassenaar, T.A., Schafer, L.V., Periole, X., Tieleman, D.P., Marrink, S.J., 2013. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of chemical theory and computation* 9, 687-697.
- De Las Rivas, J., Fontanillo, C., 2010. Protein-protein interactions essentials: key concepts to building and analyzing interactome networks. *PLoS computational biology* 6, e1000807.
- de Miguel, E., Jackson, G., 2006. The nature of the calculation of the pressure in molecular simulations of continuous models from volume perturbations. *The Journal of chemical physics* 125, 164109.
- de Planque, M.R., Boots, J.W., Rijkers, D.T., Liskamp, R.M., Greathouse, D.V., Killian, J.A., 2002. The effects of hydrophobic mismatch between phosphatidylcholine bilayers and transmembrane alpha-helical peptides depend on the nature of interfacially exposed aromatic and charged residues. *Biochemistry* 41, 8396-8404.

Βιβλιογραφία

- Dengjel, J., Kratchmarova, I., Blagoev, B., 2009. Receptor tyrosine kinase signaling: a view from quantitative proteomics. *Molecular bioSystems* 5, 1112-1121.
- Dobson, L., Remenyi, I., Tusnady, G.E., 2015. The human transmembrane proteome. *Biology direct* 10, 31.
- Dore, A.S., Okrasa, K., Patel, J.C., Serrano-Vega, M., Bennett, K., Cooke, R.M., Errey, J.C., Jazayeri, A., Khan, S., Tehan, B., Weir, M., Wiggin, G.R., Marshall, F.H., 2014. Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* 511, 557-562.
- Durbin, R., Eddy, S., Krogh, A., Mitchison, G., 1998. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge University Press, Cambridge, UK.
- Eargle, J., Luthey-Schulten, Z., 2012. NetworkView: 3D display and analysis of protein.RNA interaction networks. *Bioinformatics* 28, 3000-3001.
- Eddy, S.R., 1998. Profile hidden Markov models. *Bioinformatics* 14, 755-763.
- Eddy, S.R., 2011. Accelerated Profile HMM Searches. *PLoS computational biology* 7, e1002195.
- Efremov, R.G., Sazanov, L.A., 2012. Structure of Escherichia coli OmpF porin from lipidic mesophase. *Journal of structural biology* 178, 311-318.
- Ehrnstorfer, I.A., Geertsma, E.R., Pardon, E., Steyaert, J., Dutzler, R., 2014. Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. *Nature structural & molecular biology* 21, 990-996.
- Ehrnstorfer, I.A., Manatschal, C., Arnold, F.M., Laederach, J., Dutzler, R., 2017. Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family. *Nature communications* 8, 14033.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. *Nucleic acids research* 47, D427-D432.
- Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic acids research* 30, 1575-1584.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., K€orninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2018. The Reactome Pathway Knowledgebase. *Nucleic acids research* 46, D649-D655.
- Feller, S.E., Zhang, Y., Pastor, R.W., Brooks, B.R., 1995. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of chemical physics* 103, 4613-4621.
- Ferguson, A.D., Welte, W., Hofmann, E., Lindner, B., Holst, O., Coulton, J.W., Diederichs, K., 2000. A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins. *Structure* 8, 585-592.
- Ferguson, K.M., Berger, M.B., Mendrola, J.M., Cho, H.S., Leahy, D.J., Lemmon, M.A., 2003. EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Molecular cell* 11, 507-517.
- Forbes, J.R., Gros, P., 2003. Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane. *Blood* 102, 1884-1892.

Βιβλιογραφία

- Franco, R., Casado, V., Cortes, A., Ferrada, C., Mallol, J., Woods, A., Lluís, C., Canela, E.I., Ferre, S., 2007. Basic concepts in G-protein-coupled receptor homo- and heterodimerization. *TheScientificWorldJournal* 7, 48-57.
- Franz, M., Lopes, C.T., Huck, G., Dong, Y., Sumer, O., Bader, G.D., 2016. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* 32, 309-311.
- Fredriksson, R., Schioth, H.B., 2005. The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Molecular pharmacology* 67, 1414-1425.
- Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150-3152.
- Giocondi, M.C., Yamamoto, D., Lesniewska, E., Milhiet, P.E., Ando, T., Le Grimmeléc, C., 2010. Surface topography of membrane domains. *Biochimica et biophysica acta* 1798, 703-718.
- Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7821-7826.
- Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabasi, A.L., 2007. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* 104, 8685-8690.
- Goodsell, D. 2010. PDB-101: Molecule of the Month: Epidermal Growth Factor. RCSB PDB.
- Gribnikov, M., McLachlan, A.D., Eisenberg, D., 1987. Profile analysis: detection of distantly related proteins. *Proceedings of the National Academy of Sciences of the United States of America* 84, 4355-4358.
- Griffiss, J.M., Schneider, H., Mandrell, R.E., Yamasaki, R., Jarvis, G.A., Kim, J.J., Gibson, B.W., Hamadeh, R., Apicella, M.A., 1988. Lipooligosaccharides: the principal glycolipids of the neisserial outer membrane. *Reviews of infectious diseases* 10 Suppl 2, S287-295.
- Guerois, R., Nielsen, J.E., Serrano, L., 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology* 320, 369-387.
- Haft, D.H., Loftus, B.J., Richardson, D.L., Yang, F., Eisen, J.A., Paulsen, I.T., White, O., 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic acids research* 29, 41-43.
- Hagberg, A.A., Schult, D.A., Swart, P.J. 2008. Exploring network structure, dynamics, and function using NetworkX., pp. 11-15 *Proceedings of the 7th Python in Science Conference (SciPy2008)*. Gael Varoquaux, Travis Vaught, and Jarrod Millman (Eds), Pasadena, CA USA.
- Han, W., Schulten, K., 2012. Further optimization of a hybrid united-atom and coarse-grained force field for folding simulations: Improved backbone hydration and interactions between charged side chains. *Journal of chemical theory and computation* 8, 4413-4424.
- Hayat, S., Elofsson, A., 2012. BOCTOPUS: improved topology prediction of transmembrane beta barrel proteins. *Bioinformatics* 28, 516-522.
- Hedger, G., Sansom, M.S., Koldso, H., 2015. The juxtamembrane regions of human receptor tyrosine kinases exhibit conserved interaction sites with anionic lipids. *Scientific reports* 5, 9198.
- Henikoff, S., Greene, E.A., Pietrokovski, S., Bork, P., Attwood, T.K., Hood, L., 1997. Gene families: the taxonomy of protein paralogs and chimeras. *Science* 278, 609-614.

Βιβλιογραφία

- Hockney, R.W., Goel, S.P., Eastwood, J.W., 1974. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics* 14, 148-158.
- Hoover, W.G., 1985. Canonical dynamics: Equilibrium phase-space distributions. *Physical review. A, General physics* 31, 1695-1697.
- Hsu, P.C., Jefferies, D., Khalid, S., 2016. Molecular Dynamics Simulations Predict the Pathways via Which Pristine Fullerenes Penetrate Bacterial Membranes. *The journal of physical chemistry. B* 120, 11170-11179.
- Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmuller, H., MacKerell, A.D., Jr., 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* 14, 71-73.
- Huang, W., Lin, Z., van Gunsteren, W.F., 2011. Validation of the GROMOS 54A7 Force Field with Respect to beta-Peptide Folding. *Journal of chemical theory and computation* 7, 1237-1243.
- Hub, J.S., de Groot, B.L., van der Spoel, D., 2010. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *Journal of chemical theory and computation* 6, 3713-3720.
- Hubbard, S.R., 1999. Structural analysis of receptor tyrosine kinases. *Progress in biophysics and molecular biology* 71, 343-358.
- Huber, W., Carey, V.J., Long, L., Falcon, S., Gentleman, R., 2007. Graphs in molecular biology. *BMC bioinformatics* 8 Suppl 6, S8.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics* 14, 33-38, 27-38.
- Ingolfsson, H.I., Arnarez, C., Periole, X., Marrink, S.J., 2016. Computational 'microscopy' of cellular membranes. *Journal of cell science* 129, 257-268.
- Ingolfsson, H.I., Melo, M.N., van Eerden, F.J., Arnarez, C., Lopez, C.A., Wassenaar, T.A., Periole, X., de Vries, A.H., Tieleman, D.P., Marrink, S.J., 2014. Lipid organization of the plasma membrane. *Journal of the American Chemical Society* 136, 14554-14559.
- Jagannathan, B., Marqusee, S., 2013. Protein folding and unfolding under force. *Biopolymers* 99, 860-869.
- James, C.E., Mahendran, K.R., Molitor, A., Bolla, J.M., Bessonov, A.N., Winterhalter, M., Pages, J.M., 2009. How beta-lactam antibiotics enter bacteria: a dialogue with the porins. *PloS one* 4, e5453.
- Janin, J., Chothia, C., 1990. The structure of protein-protein recognition sites. *The Journal of biological chemistry* 265, 16027-16030.
- Jarzynski, C., 1997a. Nonequilibrium Equality for Free Energy Differences. *Physical review letters* 78, 2690-2693.
- Jarzynski, C., 1997b. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E* 56, 5018-5035.
- Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 29, 1859-1865.
- Johnston, J.M., Wang, H., Provasi, D., Filizola, M., 2012. Assessing the relative stability of dimer interfaces in g protein-coupled receptors. *PLoS computational biology* 8, e1002649.
- Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.

Βιβλιογραφία

- Kall, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. *Journal of molecular biology* 338, 1027-1036.
- Kamal, M., Jockers, R., 2011. Biological Significance of GPCR Heteromerization in the Neuro-Endocrine System. *Frontiers in endocrinology* 2, 2.
- Kandt, C., Ash, W.L., Tieleman, D.P., 2007. Setting up and running molecular dynamics simulations of membrane proteins. *Methods* 41, 475-488.
- Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28, 27-30.
- Karplus, M., McCammon, J.A., 2002. Molecular dynamics simulations of biomolecules. *Nature structural biology* 9, 646-652.
- Kastritis, P.L., Bonvin, A.M., 2013. Molecular origins of binding affinity: seeking the Archimedean point. *Current opinion in structural biology* 23, 868-877.
- Kim, S., Patel, D.S., Park, S., Slusky, J., Klauda, J.B., Widmalm, G., Im, W., 2016. Bilayer Properties of Lipid A from Various Gram-Negative Bacteria. *Biophysical journal* 111, 1750-1760.
- Kirkwood, J.G., 1935. Statistical Mechanics of Fluid Mixtures. *The Journal of chemical physics* 3, 300-313.
- Klebba, P.E., Benson, S.A., Bala, S., Abdullah, T., Reid, J., Singh, S.P., Nikaido, H., 1990. Determinants of OmpF porin antigenicity and structure. *The Journal of biological chemistry* 265, 6800-6810.
- Kniazeff, J., Prezeau, L., Rondard, P., Pin, J.P., Goudet, C., 2011. Dimers and beyond: The functional puzzles of class C GPCRs. *Pharmacology & therapeutics* 130, 9-25.
- Knight, J.L., Brooks, C.L., 3rd, 2009. Lambda-dynamics free energy simulation methods. *Journal of computational chemistry* 30, 1692-1700.
- Koebnik, R., Locher, K.P., Van Gelder, P., 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular microbiology* 37, 239-253.
- Kolakowski, L.F., Jr., 1994. GCRDb: a G-protein-coupled receptor database. *Receptors & channels* 2, 1-7.
- Kozma, D., Simon, I., Tusnady, G.E., 2013. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic acids research* 41, D524-529.
- Krishnamurthy, H., Gouaux, E., 2012. X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* 481, 469.
- Kristiansen, K., 2004. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacology & therapeutics* 103, 21-80.
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 305, 567-580.
- Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., Kollman, P.A., 1992. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of computational chemistry* 13, 1011-1021.
- Lambert, N.A., 2010. GPCR dimers fall apart. *Science signaling* 3, pe12.
- Leach, A.R., 2001. *Molecular Modelling, Principles and Applications*. 2nd ed. Pearson Education Limited.
- Lee, J., Patel, D.S., Stahle, J., Park, S.J., Kern, N.R., Kim, S.H., Cheng, X., Valvano, M.A., Holst, O., Knirel, Y.A., Qi, Y., Jo, S., Klauda, J.B., Widmalm, G., Im, W., 2018.

Βιβλιογραφία

- CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of chemical theory and computation*.
- Lemkul, J.A., 2019. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. *Living J. Comp. Mol. Sci.* 1 (1), 5068.
- Lemmon, M.A., Schlessinger, J., 2010. Cell signaling by receptor tyrosine kinases. *Cell* 141, 1117-1134.
- Lennard-Jones, J.E., 1931. Cohesion. *Proceedings of the Physical Society* 43, 461-482.
- Levy, E.D., 2010. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *Journal of molecular biology* 403, 660-670.
- Levy, S.F., Siegal, M.L., 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS biology* 6, e264.
- Liu, W.M., 1998. Shear numbers of protein beta-barrels: definition refinements and statistics. *Journal of molecular biology* 275, 541-545.
- Liu, Z., Zhang, J., Zhang, A., 2009. Design of multivalent ligand targeting G-protein-coupled receptors. *Current pharmaceutical design* 15, 682-718.
- Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research* 40, D370-376.
- Lugtenberg, E.J., Peters, R., 1976. Distribution of lipids in cytoplasmic and outer membranes of *Escherichia coli* K12. *Biochimica et biophysica acta* 441, 38-47.
- Ma, H., Khan, A., Nangia, S., 2018. Dynamics of OmpF Trimer Formation in the Bacterial Outer Membrane of *Escherichia coli*. *Langmuir : the ACS journal of surfaces and colloids* 34, 5623-5634.
- Maere, S., Heymans, K., Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448-3449.
- Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., de Vries, A.H., 2007. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry. B* 111, 7812-7824.
- Martinez, L., Andrade, R., Birgin, E.G., Martinez, J.M., 2009. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of computational chemistry* 30, 2157-2164.
- Mazurie, A., Bonchev, D., Schwikowski, B., Buck, G.A., 2010. Evolution of metabolic network organization. *BMC systems biology* 4, 59.
- Melo, M.N., Ingolfsson, H.I., Marrink, S.J., 2015. Parameters for Martini sterols and hopanoids based on a virtual-site description. *The Journal of chemical physics* 143, 243152.
- Melton, L., 2004. Protein arrays: proteomics in multiplex. *Nature* 429, 101-107.
- Miller, S., Janin, J., Lesk, A.M., Chothia, C., 1987. Interior and surface of monomeric proteins. *Journal of molecular biology* 196, 641-656.
- Milligan, G., 2013. The prevalence, maintenance, and relevance of G protein-coupled receptor oligomerization. *Molecular pharmacology* 84, 158-169.
- Mir, S., Alhroub, Y., Anyango, S., Armstrong, D.R., Berrisford, J.M., Clark, A.R., Conroy, M.J., Dana, J.M., Deshpande, M., Gupta, D., Gutmanas, A., Haslam, P., Mak, L., Mukhopadhyay, A., Nadzirin, N., Paysan-Lafosse, T., Sehnal, D., Sen, S., Smart, O.S.,

Βιβλιογραφία

- Varadi, M., Kleywegt, G.J., Velankar, S., 2018. PDBe: towards reusable data delivery infrastructure at protein data bank in Europe. *Nucleic acids research* 46, D486-D492.
- Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., Gough, J., Haft, D.R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A.P., Paysan-Lafosse, T., Pesseat, S., Potter, S.C., Qureshi, M.A., Rawlings, N.D., Redaschi, N., Richardson, L.J., Rivoire, C., Salazar, G.A., Sangrador-Vegas, A., Sigrist, C.J.A., Sillitoe, I., Sutton, G.G., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Yong, S.Y., Finn, R.D., 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research* 47, D351-D360.
- Mondal, S., Khelashvili, G., Weinstein, H., 2014. Not just an oil slick: how the energetics of protein-membrane interactions impacts the function and organization of transmembrane proteins. *Biophysical journal* 106, 2305-2316.
- Mondal, S., Khelashvili, G., Shan, J., Andersen, O.S., Weinstein, H., 2011. Quantitative modeling of membrane deformations by multihelical membrane proteins: application to G-protein coupled receptors. *Biophysical journal* 101, 2092-2101.
- Mondal, S., Johnston, J.M., Wang, H., Khelashvili, G., Filizola, M., Weinstein, H., 2013. Membrane driven spatial organization of GPCRs. *Scientific reports* 3, 2909.
- Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.J., 2008. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of chemical theory and computation* 4, 819-834.
- Mora, A., Donaldson, I.M., 2012. Effects of protein interaction data integration, representation and reliability on the use of network properties for drug target prediction. *BMC bioinformatics* 13, 294.
- Moraes, T.F., Bains, M., Hancock, R.E., Strynadka, N.C., 2007. An arginine ladder in OprP mediates phosphate-specific transfer across the outer membrane. *Nature structural & molecular biology* 14, 85-87.
- Morrison, K.L., Weiss, G.A., 2001. Combinatorial alanine-scanning. *Current opinion in chemical biology* 5, 302-307.
- Munk, C., Isberg, V., Mordalski, S., Harpsoe, K., Rataj, K., Hauser, A.S., Kolb, P., Bojarski, A.J., Vriend, G., Gloriam, D.E., 2016. GPCRdb: the G protein-coupled receptor database - an introduction. *British journal of pharmacology* 173, 2195-2207.
- Munk, C., Mutt, E., Isberg, V., Nikolajsen, L.F., Bibbe, J.M., Flock, T., Hanson, M.A., Stevens, R.C., Deupi, X., Gloriam, D.E., 2019. An online resource for GPCR structure determination and analysis. *Nature methods* 16, 151-162.
- Nastou, K.C., Tsaousis, G.N., Kremizas, K.E., Litou, Z.I., Hamodrakas, S.J., 2014. The human plasma membrane peripherome: visualization and analysis of interactions. *BioMed research international* 2014, 397145.
- Naveed, H., Jackups, R., Jr., Liang, J., 2009. Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America* 106, 12735-12740.
- Nigg, E.A., 1989. The nuclear envelope. *Current opinion in cell biology* 1, 435-440.
- Nikaido, H., 2003. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and molecular biology reviews* : MMBR 67, 593-656.

Βιβλιογραφία

- Nina, M., Berneche, S., Roux, B., 2000. Anchoring of a monotopic membrane protein: the binding of prostaglandin H2 synthase-1 to the surface of a phospholipid bilayer. *European biophysics journal* : EBJ 29, 439-454.
- Niramitrannon, J., Sansom, M.S., Pongprayoon, P., 2016. Why do the outer membrane proteins OmpF from *E. coli* and OprP from *P. aeruginosa* prefer trimers? Simulation studies. *Journal of molecular graphics & modelling* 65, 1-7.
- Niswender, C.M., Conn, P.J., 2010. Metabotropic glutamate receptors: physiology, pharmacology, and disease. *Annual review of pharmacology and toxicology* 50, 295-322.
- Nosé, S., 1984. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* 52, 255-268.
- Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M., Yokoyama, S., 2002. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110, 775-787.
- Ohvo-Rekila, H., Ramstedt, B., Leppimäki, P., Slotte, J.P., 2002. Cholesterol interactions with phospholipids in membranes. *Progress in lipid research* 41, 66-97.
- Oldham, W.M., Hamm, H.E., 2008. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature reviews. Molecular cell biology* 9, 60-71.
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., Hermjakob, H., 2014. The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic acids research* 42, D358-363.
- Paccagnini, D., Sieswerda, L., Rosu, V., Masala, S., Pacifico, A., Gazouli, M., Ikonopoulou, J., Ahmed, N., Zanetti, S., Sechi, L.A., 2009. Linking chronic infection and autoimmune diseases: *Mycobacterium avium* subspecies paratuberculosis, SLC11A1 polymorphisms and type-1 diabetes mellitus. *PloS one* 4, e7109.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M., 2000. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739-745.
- Park, J.H., Scheerer, P., Hofmann, K.P., Choe, H.W., Ernst, O.P., 2008. Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454, 183-187.
- Park, S., Khalili-Araghi, F., Tajkhorshid, E., Schulten, K., 2003. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of chemical physics* 119, 3559-3566.
- Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52, 7182-7190.
- Pasquier, C., Promponas, V.J., Hamodrakas, S.J., 2001. PRED-CLASS: cascading neural networks for generalized protein classification and genome-wide applications. *Proteins* 44, 361-369.
- Pasternak, G.W., Pan, Y.X., 2011. Mix and match: heterodimers and opioid tolerance. *Neuron* 69, 6-8.

Βιβλιογραφία

- Patel, D.S., Re, S., Wu, E.L., Qi, Y., Klebba, P.E., Widmalm, G., Yeom, M.S., Sugita, Y., Im, W., 2016. Dynamics and Interactions of OmpF and LPS: Influence on Pore Accessibility and Ion Permeability. *Biophysical journal* 110, 930-938.
- Pautsch, A., Schulz, G.E., 1998. Structure of the outer membrane protein A transmembrane domain. *Nature structural biology* 5, 1013-1017.
- Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., Bagos, P.G., 2011. Using graph theory to analyze biological networks. *BioData mining* 4, 10.
- Periole, X., 2017. Interplay of G Protein-Coupled Receptors with the Membrane: Insights from Supra-Atomic Coarse Grain Molecular Dynamics Simulations. *Chemical reviews* 117, 156-185.
- Periole, X., Marrink, S.J., 2013. The Martini coarse-grained force field. *Methods Mol Biol* 924, 533-565.
- Periole, X., Cavalli, M., Marrink, S.J., Ceruso, M.A., 2009. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *Journal of chemical theory and computation* 5, 2531-2543.
- Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 8, 785-786.
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* 26, 1781-1802.
- Pinero, J., Bravo, A., Queralt-Rosinach, N., Gutierrez-Sacristan, A., Deu-Pons, J., Centeno, E., Garcia-Garcia, J., Sanz, F., Furlong, L.I., 2017. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research* 45, D833-D839.
- Pogozheva, I.D., Lomize, A.L., 2018. Evolution and adaptation of single-pass transmembrane proteins. *Biochimica et biophysica acta. Biomembranes* 1860, 364-377.
- Pohorille, A., Jarzynski, C., Chipot, C., 2010. Good practices in free-energy calculations. *The journal of physical chemistry. B* 114, 10235-10253.
- Pontes, F.J., Rusu, V.H., Soares, T.A., Lins, R.D., 2012. The Effect of Temperature, Cations, and Number of Acyl Chains on the Lamellar to Non-Lamellar Transition in Lipid-A Membranes: A Microscopic View. *Journal of chemical theory and computation* 8, 3830-3838.
- Pyle, E., Kalli, A.C., Amillis, S., Hall, Z., Lau, A.M., Hanyaloglu, A.C., Diallinas, G., Byrne, B., Politis, A., 2018. Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell chemical biology* 25, 840-848 e844.
- Qi, Y., Bar-Joseph, Z., Klein-Seetharaman, J., 2006. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. *Proteins* 63, 490-500.
- Qi, Y., Cheng, X., Han, W., Jo, S., Schulten, K., Im, W., 2014. CHARMM-GUI PACE CG Builder for solution, micelle, and bilayer coarse-grained simulations. *Journal of chemical information and modeling* 54, 1003-1009.
- Raetz, C.R., Whitfield, C., 2002. Lipopolysaccharide endotoxins. *Annual review of biochemistry* 71, 635-700.
- Rajagopala, S.V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S.B., Phanse, S., Ceol, A., Hauser, R., Siszler, G., Wuchty, S., Emili, A., Babu,

Βιβλιογραφία

- M., Aloy, P., Pieper, R., Uetz, P., 2014. The binary protein-protein interaction landscape of *Escherichia coli*. *Nature biotechnology* 32, 285-290.
- Rajula, H.S.R., Mauri, M., Fanos, V., 2018. Scale-free networks in metabolomics. *Bioinformatics* 14, 140-144.
- Rasmussen, S.G., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., Mathiesen, J.M., Shah, S.T., Lyons, J.A., Caffrey, M., Gellman, S.H., Steyaert, J., Skiniotis, G., Weis, W.I., Sunahara, R.K., Kobilka, B.K., 2011. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477, 549-555.
- Rosenbaum, D.M., Rasmussen, S.G., Kobilka, B.K., 2009. The structure and function of G-protein-coupled receptors. *Nature* 459, 356-363.
- Rosenberg, I.M., 2005. *Protein Analysis and Purification* Birkhauser.
- Roupie, V., Rosseels, V., Piersoel, V., Zinniel, D.K., Barletta, R.G., Huygen, K., 2008. Genetic resistance of mice to *Mycobacterium paratuberculosis* is influenced by *Slc11a1* at the early but not at the late stage of infection. *Infection and immunity* 76, 2099-2105.
- Rouse, S.L., Stylianou, F., Wu, H.Y.G., Berry, J.L., Sewell, L., Morgan, R.M.L., Sauerwein, A.C., Matthews, S., 2018. The FapF Amyloid Secretion Transporter Possesses an Atypical Asymmetric Coiled Coil. *Journal of molecular biology* 430, 3863-3871.
- Rouse, S.L., Hawthorne, W.J., Berry, J.L., Chorev, D.S., Ionescu, S.A., Lambert, S., Stylianou, F., Ewert, W., Mackie, U., Morgan, R.M.L., Otzen, D., Herbst, F.A., Nielsen, P.H., Dueholm, M., Bayley, H., Robinson, C.V., Hare, S., Matthews, S., 2017. A new class of hybrid secretion system is employed in *Pseudomonas amyloid* biogenesis. *Nature communications* 8, 263.
- Roux, B., 1995. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* 91, 275-282.
- Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., Vidal, M., 2005. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173-1178.
- Ruiz-Larriaga, O., 2009. Identification of single nucleotide polymorphisms in the bovine solute carrier family 11 member 1 (SLC11A1) gene and their association with infection by *Mycobacterium avium* subspecies *paratuberculosis*. *Journal of Dairy Science* 93, 1713-1721.
- Salomon-Ferrer, R., Case, D.A., Walker, R.C., 2013. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 3, 198-210.
- Schlame, M., Rua, D., Greenberg, M.L., 2000. The biosynthesis and functional role of cardiolipin. *Progress in lipid research* 39, 257-288.
- Schulz, G.E., 2003. Transmembrane beta-barrel proteins. *Advances in protein chemistry* 63, 47-70.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic acids research* 33, W382-388.

Βιβλιογραφία

- Sechi, L.A., Dow, C.T., 2015. Mycobacterium avium ss. paratuberculosis Zoonosis - The Hundred Year War - Beyond Crohn's Disease. *Frontiers in immunology* 6, 96.
- Sehna, D., Deshpande, M., Varekova, R.S., Mir, S., Berka, K., Midlik, A., Pravda, L., Velankar, S., Koca, J., 2017. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nature methods* 14, 1121-1122.
- Sela-Culang, I., Kunik, V., Ofran, Y., 2013. The structural basis of antibody-antigen recognition. *Frontiers in immunology* 4, 302.
- Sengupta, D., Chattopadhyay, A., 2015. Molecular dynamics simulations of GPCR-cholesterol interaction: An emerging paradigm. *Biochimica et biophysica acta* 1848, 1775-1782.
- Sethi, A., Eargle, J., Black, A.A., Luthey-Schulten, Z., 2009. Dynamical networks in tRNA:protein complexes. *Proceedings of the National Academy of Sciences of the United States of America* 106, 6620-6625.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13, 2498-2504.
- Sie, M., den Dunnen, W.F., Lourens, H.J., Meeuwsen-de Boer, T.G., Scherpen, F.J., Zomerman, W.W., Kampen, K.R., Hoving, E.W., de Bont, E.S., 2015. Growth-factor-driven rescue to receptor tyrosine kinase (RTK) inhibitors through Akt and Erk phosphorylation in pediatric low grade astrocytoma and ependymoma. *PloS one* 10, e0122555.
- Singer, S.J., Nicolson, G.L., 1972. The fluid mosaic model of the structure of cell membranes. *Science* 175, 720-731.
- Sinha, S., Lynn, A.M., 2014. HMM-ModE: implementation, benchmarking and validation with HMMER3. *BMC research notes* 7, 483.
- Siu, F.Y., He, M., de Graaf, C., Han, G.W., Yang, D., Zhang, Z., Zhou, C., Xu, Q., Wacker, D., Joseph, J.S., Liu, W., Lau, J., Cherezov, V., Katritch, V., Wang, M.W., Stevens, R.C., 2013. Structure of the human glucagon class B G-protein-coupled receptor. *Nature* 499, 444-449.
- Snyder, S., Kim, D., McIntosh, T.J., 1999. Lipopolysaccharide bilayer structure: effect of chemotype, core mutations, divalent cations, and temperature. *Biochemistry* 38, 10758-10767.
- Soe-Lin, S., Apte, S.S., Mikhael, M.R., Kayembe, L.K., Nie, G., Ponka, P., 2010. Both Nramp1 and DMT1 are necessary for efficient macrophage iron recycling. *Exp Hematol* 38, 609-617.
- Sohlenkamp, C., Geiger, O., 2016. Bacterial membrane lipids: diversity in structures and pathways. *FEMS microbiology reviews* 40, 133-159.
- Srivastava, P.K., Desai, D.K., Nandi, S., Lynn, A.M., 2007. HMM-ModE--improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences. *BMC bioinformatics* 8, 104.
- Stevens, R.C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., Wuthrich, K., 2013. The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. *Nature reviews. Drug discovery* 12, 25-34.
- Stewart, M., 2007. Molecular mechanism of the nuclear protein import cycle. *Nature reviews. Molecular cell biology* 8, 195-208.

Βιβλιογραφία

- Strambio-De-Castillia, C., Niepel, M., Rout, M.P., 2010. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nature reviews. Molecular cell biology* 11, 490-501.
- Stroet, M., Caron, B., Visscher, K.M., Geerke, D.P., Malde, A.K., Mark, A.E., 2018. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *Journal of chemical theory and computation* 14, 5834-5845.
- Surrey, T., Schmid, A., Jahnig, F., 1996. Folding and membrane insertion of the trimeric beta-barrel protein OmpF. *Biochemistry* 35, 2283-2288.
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* 45, D362-D368.
- Tadevosyan, A., Vaniotis, G., Allen, B.G., Hebert, T.E., Nattel, S., 2012. G protein-coupled receptor signalling in the cardiac nuclear membrane: evidence and possible roles in physiological and pathophysiological function. *The Journal of physiology* 590, 1313-1330.
- Taylor, I.W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., Bull, S., Pawson, T., Morris, Q., Wrana, J.L., 2009. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology* 27, 199-204.
- Techau, M.E., Valdez-Taubas, J., Popoff, J.F., Francis, R., Seaman, M., Blackwell, J.M., 2007. Evolution of differences in transport function in Slc11a family members. *The Journal of biological chemistry* 282, 35646-35656.
- Teichmann, S.A., Murzin, A.G., Chothia, C., 2001. Determination of protein function, evolution and interactions by structural genomics. *Current opinion in structural biology* 11, 354-363.
- The UniProt Consortium, 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* 47, D506-D515.
- Titz, B., Schlesner, M., Uetz, P., 2004. What do we learn from high-throughput protein interaction data? *Expert review of proteomics* 1, 111-121.
- Tsaousis, G.N., Bagos, P.G., Hamodrakas, S.J., 2014. HMMpTM: improving transmembrane protein topology prediction using phosphorylation and glycosylation site prediction. *Biochimica et biophysica acta* 1844, 316-322.
- Tsaousis, G.N., Hamodrakas, S.J., Bagos, P.G., 2017a. Predicting Beta Barrel Transmembrane Proteins Using HMMs. *Methods Mol Biol* 1552, 43-61.
- Tsaousis, G.N., Theodoropoulou, M.C., Hamodrakas, S.J., Bagos, P.G., 2017b. Predicting Alpha Helical Transmembrane Proteins Using HMMs. *Methods Mol Biol* 1552, 63-82.
- Tsirigos, K.D., Peters, C., Shu, N., Kall, L., Elofsson, A., 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic acids research* 43, W401-407.
- Tusnady, G.E., Dosztanyi, Z., Simon, I., 2004. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* 20, 2964-2972.
- van der Spoel, D., van Maaren, P.J., 2006. The Origin of Layer Structure Artifacts in Simulations of Liquid Water. *Journal of chemical theory and computation* 2, 1-11.
- van Dongen, S., 2000. Graph Clustering by Flow Simulation. PhD Thesis, University of Utrecht, Utrecht, the Netherlands.

Βιβλιογραφία

- van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P., Tironi, I.G., 1996. Biomolecular Simulation: The GROMOS96 manual and user guide, Zürich, Switzerland: Hochschulverlag AG an der ETH Zürich.
- van Gunsteren, W.F., Berendsen, H.J.C., 1987. The Gromos-87 Manual, Biomos BV Nijenborgh 4, 9747 AG Groningen, The Netherlands.
- van Meer, G., Voelker, D.R., Feigenson, G.W., 2008. Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology* 9, 112-124.
- van Zundert, G.C.P., Rodrigues, J., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., Bonvin, A., 2016. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of molecular biology* 428, 720-725.
- Vidal, S.M., Pinner, E., Lepage, P., Gauthier, S., Gros, P., 1996. Natural resistance to intracellular infections: Nramp1 encodes a membrane phosphoglycoprotein absent in macrophages from susceptible (Nramp1 D169) mouse strains. *J Immunol* 157, 3559-3568.
- Viklund, H., Elofsson, A., 2008. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* 24, 1662-1668.
- Viklund, H., Bernsel, A., Skwark, M., Elofsson, A., 2008. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* 24, 2928-2929.
- von Heijne, G., 1991. Proline kinks in transmembrane alpha-helices. *Journal of molecular biology* 218, 499-503.
- von Heijne, G., 2007. The membrane protein universe: what's out there and why bother? *Journal of internal medicine* 261, 543-557.
- Walhout, A.J., Vidal, M., 2001. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* 24, 297-306.
- Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., Sprang, S.R., 1995. The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* 83, 1047-1058.
- Wallin, E., von Heijne, G., 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein science : a publication of the Protein Society* 7, 1029-1038.
- Wan, C.K., Han, W., Wu, Y.D., 2012. Parameterization of PACE Force Field for Membrane Environment and Simulation of Helical Peptides and Helix-Helix Association. *Journal of chemical theory and computation* 8, 300-313.
- Wang, C., Wu, H., Katritch, V., Han, G.W., Huang, X.P., Liu, W., Siu, F.Y., Roth, B.L., Cherezov, V., Stevens, R.C., 2013a. Structure of the human smoothed receptor bound to an antitumour agent. *Nature* 497, 338-343.
- Wang, J., Duncan, D., Shi, Z., Zhang, B., 2013b. WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic acids research* 41, W77-83.
- Wang, J., Vasaikar, S., Shi, Z., Greer, M., Zhang, B., 2017. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic acids research* 45, W130-W137.
- Wang, Y., Gkeka, P., Fuchs, J.E., Liedl, K.R., Cournia, Z., 2016. DPPC-cholesterol phase diagram using coarse-grained Molecular Dynamics simulations. *Biochimica et biophysica acta* 1858, 2846-2857.

Βιβλιογραφία

- Wang, Y.N., Huo, L., Hsu, J.L., Hung, M.C., 2015. Biochemical fractionation of membrane receptors in the nucleus. *Methods Mol Biol* 1234, 99-112.
- Wassenaar, T.A., Ingolfsson, H.I., Bockmann, R.A., Tieleman, D.P., Marrink, S.J., 2015. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of chemical theory and computation* 11, 2144-2155.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.
- Webb, B., Sali, A., 2017. Protein Structure Modeling with MODELLER. *Methods Mol Biol* 1654, 39-54.
- Wessling-Resnick, M., 2015. Nramp1 and Other Transporters Involved in Metal Withholding during Infection. *The Journal of biological chemistry* 290, 18984-18990.
- Wilde, R.E., Singh, S., 1998. *Statistical Mechanics, Fundamentals and Modern Applications* John Wiley & Sons, Inc, New York.
- Wilson, K.L., Berk, J.M., 2010. The nuclear envelope at a glance. *Journal of cell science* 123, 1973-1978.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., Wilson, M., 2018. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic acids research* 46, D1074-D1082.
- Woehler, A., Ponimaskin, E.G., 2009. G protein--mediated signaling: same receptor, multiple effectors. *Current molecular pharmacology* 2, 237-248.
- Wong, X., Luperchio, T.R., Reddy, K.L., 2014. NET gains and losses: the role of changing nuclear envelope proteomes in genome regulation. *Current opinion in cell biology* 28, 105-120.
- Wu, E.L., Engstrom, O., Jo, S., Stuhlsatz, D., Yeom, M.S., Klauda, J.B., Widmalm, G., Im, W., 2013. Molecular dynamics and NMR spectroscopy studies of E. coli lipopolysaccharide structure and dynamics. *Biophysical journal* 105, 1444-1455.
- Zhang, B., Kirov, S., Snoddy, J., 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research* 33, W741-748.
- Zhang, X., Gureasko, J., Shen, K., Cole, P.A., Kuriyan, J., 2006. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* 125, 1137-1149.
- Zheng, H., Pearsall, E.A., Hurst, D.P., Zhang, Y., Chu, J., Zhou, Y., Reggio, P.H., Loh, H.H., Law, P.Y., 2012. Palmitoylation and membrane cholesterol stabilize mu-opioid receptor homodimerization and G protein coupling. *BMC cell biology* 13, 6.
- Ziervogel, B.K., Roux, B., 2013. The binding of antibiotics in OmpF porin. *Structure* 21, 76-87.
- Λέκκα, Μ.Ε., Γαλανοπούλου, Ν., Λεονταρίτης, Γ., Κητσιούλη, Ε., 2015. Βιολογικές Μεμβράνες. Από τη δομή στις λειτουργίες. Θεωρία και πειραματικές προσεγγίσεις. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα.
- Μπάγκος, Π.Γ., 2015. Βιοπληροφορική. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα.
- Χαμόδρακας, Σ.Ι., 1993. Θέματα Μοριακής Βιοφυσικής Εκδόσεις Συμμετρία, Αθήνα.

ΠΑΡΑΡΤΗΜΑΤΑ

Παράρτημα Ι – Σύντομο βιογραφικό

Ο Φώτιος Μπαλτούμας είναι απόφοιτος του *Τμήματος Βιολογίας του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών*, με *Μεταπτυχιακό Δίπλωμα Ειδίκευσης στη Βιοπληροφορική*. Τα ερευνητικά ενδιαφέροντά του επικεντρώνονται στα πεδία της Δομικής Βιολογίας, της Βιοπληροφορικής και της Υπολογιστικής Βιοφυσικής και περιλαμβάνουν τη δομική και υπολογιστική *μελέτη της δομής, της δυναμικής και του λειτουργικού ρόλου των βιολογικών μεμβρανών, των μεμβρανικών πρωτεϊνών και των αλληλεπιδράσεών τους*. Παράλληλα, στα ερευνητικά ενδιαφέροντά του συμπεριλαμβάνονται *η μελέτη των μηχανισμών του διπλώματος των πρωτεϊνών και η μελέτη του μηχανισμού αυτοσυγκρότησης και συσσωμάτωσης πεπτιδίων και πρωτεϊνών κατά το σχηματισμό των αμυλοειδών ινιδίων*. Για τη μελέτη των παραπάνω εφαρμόζει μια σειρά από μεθόδους και τεχνικές από τα πεδία της Δομικής Βιοπληροφορικής και της Υπολογιστικής Βιοφυσικής, όπως προσομοιώσεις Μοριακής Δυναμικής, χρήση της Θεωρίας Γράφων για την ανάλυση βιολογικών δικτύων αλληλεπιδράσεων και μοντέλα μηχανικής μάθησης όπως τα Κρυφά Μοντέλα Markov (HMMs).

Εκπαίδευση

- 2014-2019** *Υποψ. Διδάκτορας στη Βιοπληροφορική, Τμήμα Βιολογίας, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών*
Τίτλος Διατριβής: «Υπολογιστικές Μελέτες Αλληλεπιδράσεων Πρωτεϊνών – Πρωτεϊνών σε Διαμεμβρανικές Πρωτεΐνες» (Επιβλέπων: Ομοτ. Καθ. Σ.Ι. Χαμόδρακας)
- 2014-2015** *Μεταπτυχιακό Δίπλωμα Ειδίκευσης στη Βιοπληροφορική, Π.Μ.Σ. «Βιοπληροφορική», Τμήμα Βιολογίας, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών*
Βαθμός: Άριστα (8.89/10)
Διπλωματική Εργασία: «Υπολογιστικές Μελέτες Αλληλεπιδράσεων Ολιγομερισμού σε Συζευγμένους με G-πρωτεΐνες Υποδοχείς» (Επιβλέπουσα: Επικ. Καθ. Β. Α. Οικονομίδου)
- 2006-2013** *Πτυχίο στη Βιολογία, Τμήμα Βιολογίας, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών*
Βαθμός: Λίαν Καλώς (6.53/10)
Διπλωματική Εργασία: «Υπολογιστικές Μελέτες Αλληλεπιδράσεων GPCRs, G-πρωτεϊνών και μορίων-εκτελεστών» (Επιβλέπων: Καθ. Σ.Ι. Χαμόδρακας)

Παράρτημα II – Λίστα υπολογιστικών εργαλείων και δημοσιεύσεων

Υπολογιστικά εργαλεία τα οποία αναπτύχθηκαν στα πλαίσια της διδακτορικής διατριβής:

1. **GNOMM** (<http://bioinformatics.biol.uoa.gr/GNOMM>): μία αυτοματοποιημένη μέθοδος προτυποποίησης εξωτερικών μεμβρανών των αρνητικών κατά Gram βακτηρίων.
2. **RTK-PRED** (<http://bioinformatics.biol.uoa.gr/RTK-PRED>): ένας αλγόριθμος για την πρόγνωση και το σχολιασμό υποδοχέων με ενεργότητα τυροσινικής κινάσης (RTKs).
3. **NucEnvDB** (<http://bioinformatics.biol.uoa.gr/nucenvdb>): μία βάση δεδομένων για τις πρωτεΐνες του πυρηνικού φακέλου και τις αλληλεπιδράσεις τους.

Δημοσιεύσεις σε περιοδικά με σύστημα κριτών, οι οποίες προέκυψαν από τα αποτελέσματα της διδακτορικής διατριβής

1. **Baltoumas, F.A.**, Theodoropoulou, M.C. & Hamodrakas, S.J. (2016) **Molecular dynamics simulations and structure-based network analysis reveal structural and functional aspects of G-protein coupled receptor dimer interactions.** *J. Comput. Aid. Mol. Des.* 30 (6):489-512
2. Triantaphyllopoulos, K.A., **Baltoumas, F.A.** & Hamodrakas, S.J. (2019) **Structural Characterization and Molecular Dynamics simulations of the caprine and bovine Solute Carrier Family 11 A1 (SLC11A1).** *J. Comput. Aid. Mol. Des.* 33 (2): 265-285
3. **Baltoumas, F.A.**, Hamodrakas, S.J. & Iconomidou, V.A. (2019) **The Gram-Negative Outer Membrane Modeler: automated building of lipopolysaccharide-rich bacterial outer membranes in four force fields.** *J. Comput. Chem.* 9999:1-9, doi: 10.1002/jcc.25823
4. **Baltoumas, F.A.***, Sofras, D.T.*, Hamodrakas, S.J. & Iconomidou, V.A. (2019) **NucEnvDB: a database of nuclear envelope proteins and their interactions.** *In preparation*

**Equally contributing authors*

5. **Baltoumas, F.A.**, Theodoropoulou, M.C. & Hamodrakas, S.J. (2013) **Interactions of the α -subunits of heterotrimeric G-proteins with GPCRs, effectors and RGS proteins: a critical review and analysis of interacting surfaces, conformational shifts, structural diversity and electrostatic potentials.** *J. Struct. Biol.* 182 (3):209-18
6. Louros, N.N., **Baltoumas, F.A.**, Hamodrakas, S.J. & Iconomidou, V.A. (2016) **A β -solenoid model of the Pmel17 repeat domain: insights to the formation of functional amyloid fibrils.** *J. Comput. Aid. Mol. Des.* 30 (2):153-164
7. Louros, N.N., Tsiolaki, P.L., **Baltoumas, F.A.**, Chryssikos, G.D., Gionis, V., Hamodrakas, S.J. & Iconomidou, V.A. (2017) **Tracking the amyloidogenic core of IAPP amyloid fibrils: insights from micro-Raman spectroscopy.** *J. Struct. Biol.* 199 (2):140-152
8. Tsiolaki, P.L., Nasi, G.I., **Baltoumas, F.A.**, Louros, N.N., Magafa, V., Hamodrakas, S.J. & Iconomidou, V.A. (2018) **α CGRP, another amyloidogenic member of the CGRP family.** *J. Struct. Biol.* 203 (1):27-36
9. Tsiolaki, P.L., Katsafana, A.D., **Baltoumas, F.A.**, Louros, N.N. & Iconomidou, V.A. (2019) **Hidden Aggregation Hot-Spots on Human Apolipoprotein E: A Structural Study.** *Int. J. Mol. Sci.* 20(9): E2274
10. Tsiolaki, P.L., Nasi, G.I., **Baltoumas, F.A.**, Fishman, S., Tu, H.C. & Iconomidou, V.A. (2019) **Delving into the amyloidogenic core of human Leukocyte Chemotactic Factor 2.** *Under Review*

Παράρτημα ΙΙΙ – Δημοσιεύσεις των αποτελεσμάτων της διατριβής σε επιστημονικά περιοδικά

Molecular dynamics simulations and structure-based network analysis reveal structural and functional aspects of G-protein coupled receptor dimer interactions

Fotis A. Baltoumas¹ · Margarita C. Theodoropoulou^{1,2} · Stavros J. Hamodrakas¹

Received: 27 January 2016 / Accepted: 22 June 2016 / Published online: 27 June 2016
© Springer International Publishing Switzerland 2016

Abstract A significant amount of experimental evidence suggests that G-protein coupled receptors (GPCRs) do not act exclusively as monomers but also form biologically relevant dimers and oligomers. However, the structural determinants, stoichiometry and functional importance of GPCR oligomerization remain topics of intense speculation. In this study we attempted to evaluate the nature and dynamics of GPCR oligomeric interactions. A representative set of GPCR homodimers were studied through Coarse-Grained Molecular Dynamics simulations, combined with interface analysis and concepts from network theory for the construction and analysis of dynamic structural networks. Our results highlight important structural determinants that seem to govern receptor dimer interactions. A conserved dynamic behavior was observed among different GPCRs, including receptors belonging in different GPCR classes. Specific GPCR regions were highlighted as the core of the interfaces. Finally, correlations of motion were observed between parts of the dimer interface and GPCR segments participating in ligand binding and receptor activation, suggesting the existence of mechanisms through which dimer formation may affect GPCR

function. The results of this study can be used to drive experiments aimed at exploring GPCR oligomerization, as well as in the study of transmembrane protein–protein interactions in general.

Keywords G-protein coupled receptors · Oligomerization · Molecular dynamics · Network analysis

Introduction

G-protein coupled receptors (GPCRs) are one of the largest and most diverse superfamilies of membrane receptors in eukaryotic cells. They regulate the majority of cell responses to stimuli and have been implicated in a wide range of diseases, including neurological syndromes, cardiac diseases, HIV infection and various types of cancer [1]. As a result, today GPCRs are targets for more than 40 % of pharmaceuticals on the market [2]. Most GPCR functions are conducted through heterotrimeric G-proteins, composed by G α subunits and G $\beta\gamma$ heterodimers which, in turn, regulate the function for a wide variety of effectors. However, a number of alternative signaling pathways, either complementary to the G-protein pathway or completely independent, have also been identified [1, 3].

GPCRs are usually grouped into six classes (A–F), four of which (namely A, B, C and F) are present in Metazoa [4]. GPCRs from all Classes share a common topology, comprised by an extracellular N-terminus, seven transmembrane (TM) α -helices connected by three intracellular (ICL) and three extracellular (ECL) loops and followed by an amphiphilic 8th helix (H8) and a cytoplasmic C-terminus [2]. This topology has been confirmed by the growing number of GPCR crystal structures, including important Class A GPCRs such as the photoreceptor Rhodopsin [5], β 1AR and

Electronic supplementary material The online version of this article (doi:10.1007/s10822-016-9919-y) contains supplementary material, which is available to authorized users.

✉ Stavros J. Hamodrakas
shamodr@biol.uoa.gr

¹ Department of Cell Biology and Biophysics, Faculty of Biology, National and Kapodistrian University of Athens, Panepistimiopolis, 15701 Athens, Greece

² Present Address: Department of Computer Science and Biomedical Informatics, University of Central Greece, 35131 Lamia, Greece

β 2AR adrenergic receptors [6, 7], the A2A adenosine receptor [8] and the entire opioid receptor subfamily [9–11], the Smoothed receptor from Class F [12], the CRF1R corticotropin release factor 1 and glucagon receptors from Class B [13, 14] and metabotropic glutamate receptors mGluR1 and mGluR5 from Class C [15, 16].

An emerging paradigm in GPCR research is the notion that receptors do not act exclusively as monomers, but also form functionally relevant dimers and oligomers [17]. Numerous biophysical and biochemical trials have demonstrated the spatial organization of GPCRs in higher order assemblies, both in vitro and in vivo [18, 19]. Dimer formation has been repeatedly shown to be obligatory for canonical receptor function in class C GPCRs [20], while receptors from other classes have also been shown to be functional both as monomers and as oligomers [18]. Furthermore, biological fingerprint experiments have shown that formation of homo- and heterodimers or oligomers may influence important aspects in GPCR signaling pathways, such as ligand binding affinity, receptor activation or internalization and interactions with heterotrimeric G-proteins [18]. Finally, a significant number of GPCR oligomers have been implicated in various pathological conditions such as schizophrenia, Parkinson's disease, drug addiction, heart failure and asthma [21] and, in many cases, receptor homo- and heterodimers are believed to be more suitable targets for the design of novel drugs than monomers [17, 22–24].

Despite the accumulating experimental evidence and the functional significance of oligomerization, the structural determinants of GPCR dimers and oligomers remain controversial. Cross-linking and mutagenesis studies for several GPCRs [25–30], as well as Atomic Force and Cryo-Electron Microscopy studies for Rhodopsin [31–33] have identified transmembrane segments TM1, TM4, TM5 and TM6 and the cytoplasmic H8 helix as potential participants in the dimer interface, while recently solved crystal structures display symmetric parallel homodimers with interfaces resembling the ones proposed by experiments and involving interactions between TM1, TM2 and H8 (TM1–TM2–H8 dimers), TM4 and TM5 (TM4–TM5 dimers) or TM5 and TM6 (TM5–TM6 dimers) [17] (Fig. 1). A number of bioinformatics and computational biology studies, including all-atom and Coarse-Grained Molecular Dynamics simulations, sequence-based interface predictions and co-evolution analyses have also investigated the formation and stability of the afore mentioned interfaces [34–40]. However, the functional relevance of these crystallographic dimers remains a subject of intense debate.

The goal of this study was to investigate and evaluate the structural nature and dynamics of GPCR oligomeric interactions, based on the available structural data. The dynamic nature of GPCR dimers and the dimer interface

stoichiometry were explored through Coarse-Grained Molecular Dynamics (CG-MD) simulations in model membranes, followed by reconstruction in atomistic detail. Original crystallographic data and simulation results were analyzed for the identification of important interface hot spots. Simulation results were combined with concepts from network theory for the analysis of dynamic networks and the study of correlated motions between GPCR structural aspects. The results of this study can be used to drive theoretical and experimental trials in the study of GPCR oligomers, as well as in the study of transmembrane protein–protein interactions in general.

Materials and methods

Structural data collection and manipulation

A set of representative crystal structures, containing GPCRs in parallel homodimers, was compiled through extensive search in the literature and the Protein Data Bank (PDB) [41]. These structures were used as the input in Molecular Dynamics simulations, followed by interface analysis and classification. In cases where different PDB entries represented the same crystallographic dimers, the structure with the best possible quality was selected. In some cases, multiple structures were included for a specific receptor; these involved occasions where the structures contained either different receptor conformations or different dimer interfaces for the same receptor. In cases when the crystal interfaces were not part of the asymmetric unit but inferred biological assemblies, the dimers were generated using the Protein Interfaces, Structures and Assemblies (PISA) service through the PDBePISA server [42]. Where necessary, T4 Lysozyme or soluble cytochrome b562 coordinates and non-essential heteroatoms were removed and missing loop segments were reconstructed.

In addition to the crystallographic data, a number of theoretical models for GPCR dimers, based on existing structural evidence, were subjected to simulation and analysis. Specifically, a model of oligomerized Rhodopsin by *Liang and co-workers* [31, 32], based on restraints derived from Atomic-Force Microscopy (AFM) and Cryo-Electron Microscopy (Cryo-EM), was retrieved from the RCSB PDB Theoretical Model archive and included in the dataset. This model features two different interfaces for the Rhodopsin dimer; namely, a TM4–TM5 and a TM1–TM1 dimer. Finally, a number of putative GPCR dimers were constructed, featuring alternative dimer interfaces not appearing in crystal structures. Specifically, TM4–TM4 dimers were constructed for δ OR, μ OR and mGluR1 receptors. These models were designed to accommodate restraints from biochemical evidence [28, 43], but feature

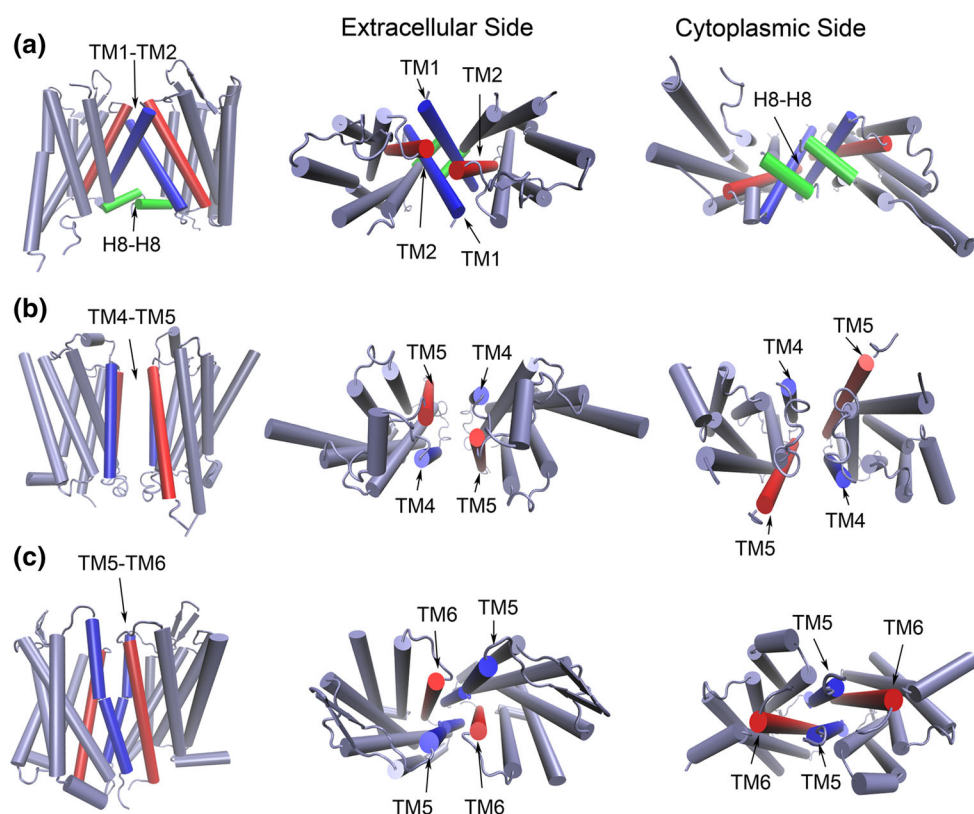


Fig. 1 Examples of commonly found GPCR crystallographic dimer interfaces. **a** The κ opioid receptor (κ OR) TM1–TM2–H8 dimer. **b** The β 1 adrenergic receptor (β 1AR) TM4–TM5 dimer. **c** The μ opioid receptor (μ OR) TM5–TM6 dimer. Helical segments are shown

as cylinders, with interacting segments colored blue, red and green, while non-interacting segments are colored grey. The third intracellular loop (ICL3) loop segments are not shown for clarity

an alternative dimer interface that has been classified as energetically unfavorable, compared to the crystallographic dimers [36, 40]. Monomers from the mouse μ OR, human δ OR and mGluR1 crystal structures were used to construct the initial conformations. Where necessary, T4 coordinates and heteroatoms were removed and missing loop segments were reconstructed. The resulting models were used to build starting conformations, by manually positioning the protomers in configurations compatible with symmetric contacts of the TM4 or TM6 helices.

Manipulation, rendering and visualization of the structures were performed with PyMOL [44] and Visual Molecular Dynamics (VMD) [45]. Molecular modeling was performed using MODELLER v. 9.14 [46]. All relevant amino acid sequences were retrieved from UniProt [47] and alignments for modeling were performed using Clustal X v. 2.1 [48]. The final dataset of simulated GPCR dimers is presented in Tables 1 and 2 and Supplementary Table S1.

Coarse-grained molecular dynamics simulations

Coarse-Grained Molecular Dynamics simulations (CG-MD) were performed using the MARTINI Coarse-Grained

force field and its extension to proteins [49, 50]. MARTINI utilizes a four to one mapping rule, meaning that, on average, four heavy atoms are represented by one Coarse-Grained particle. The accuracy and robustness of the force field has been validated for a number of biomolecules and processes, including the accurate reproduction of lipid bilayer properties such as the area per lipid, hydrophobic thickness and membrane curvature, the successful reproduction of dimerization free energy values obtained from all-atom simulations or thermodynamic data [49, 50] and simulations for several transmembrane proteins, including the study of Rhodopsin [34, 36] and adrenergic receptor homo-oligomerization [38, 51] and the hetero-oligomerization of opioid receptors [28, 39].

Each dimer was converted to Coarse-Grained (CG) representations using the MARTINI mapping scheme [52] and embedded in a lipid bilayer composed by 1-palmitoyl-2-oleyl-phosphatidyl-choline (POPC) phospholipids (Supplementary Figure S1). A special case was made for the cholesterol-containing β 2AR and mGluR1 dimers, which were simulated with and without the original cholesterol molecules in a pure POPC membrane, as well as in a bilayer with a 9:1 POPC: cholesterol ratio [53]. A simple

Table 1 Overview of simulated GPCR crystallographic homodimers, including a brief description of the initial dimer interfaces and the changes observed during simulation

Receptor	PDB	Original dimer interface	Notes on the initial structure	Number of simulations ^d	Changes after simulation	Final dimer interface
<i>Closely packed dimers</i>						
Bovine Rhodopsin ^{a,b} (Class A) ^c	2I35	TM1–TM2–H8	The protomers form a ~20° angle between their principal axes	3 × 1 μs CG ^d	The angle between the protomers' principal axes is reduced; the dimer resembles κOR and β1AR	TM1–TM2–H8
Bovine Ligand-free Opsin ^a (Class A)	3CAP	TM1–TM2–H8	The protomers form a ~20° angle between their principal axes; different interface residues than 2I35	3 × 1 μs CG	The angle between the protomers' principal axes is reduced; the dimer resembles κOR and β1AR	TM1–TM2–H8
Human κ opioid receptor (κOR) (Class A)	4DJH	TM1–TM2–H8	Almost parallel protomers; extensive H8–H8 contacts	1 × 10 μs CG 3 × 1 μs CG	Very few changes are observed	TM1–TM2–H8
Mouse μ opioid receptor (μOR) ^{a,b} (Class A)	4DKL	TM1–TM2–H8	The protomers form a ~20° angle between their principal axes	3 × 1 μs CG	The angle between the protomers' principal axes is reduced; the dimer resembles κOR and β1AR	TM1–TM2–H8
Turkey β1 adrenergic receptor (β1AR) (Class A)	4GPO	TM1–TM2–H8	Almost parallel protomers	3 × 1 μs CG	Very few changes are observed	TM1–TM2–H8
Squid Rhodopsin ^{a,b} (Class A)	2Z73	TM4–TM5	Extensive TM4–TM5 contacts	3 × 1 μs CG	The inter-protomer distance is reduced	TM4–TM5
Human A2A adenosine receptor ^b (Class A)	4E1Y	TM4–TM5	More loose interface compared to other TM4–TM5 dimers	3 × 1 μs CG	The inter-protomer distance is reduced	TM4–TM5
Turkey β1 adrenergic receptor (β1AR) ^b (Class A)	4GPO	TM4–TM5	More loose interface compared to other TM4–TM5 dimers	3 × 1 μs CG	The inter-protomer distance is reduced	TM4–TM5
Human Smoothed receptor (Class F)	4JKV	TM4–TM5	Extensive TM4–TM5 contacts	3 × 1 μs CG	Very few changes are observed	TM4–TM5
Human CXCR4 (Ligand: It1) ^a (Class A)	3ODU	TM5–TM6	Loose TM5–TM6 contacts; T4 Lysozymes form part of the interface	1 × 10 μs CG 3 × 1 μs CG	The dimer adopts a TM4–TM5 orientation	TM4–TM5
Human CXCR4 (Ligand: CVX15) ^{a,b} (Class A)	3OE0	TM4–TM5–TM6	Loose TM5–TM6 contacts; T4 Lysozymes form part of the interface	3 × 1 μs CG	The dimer adopts a TM4–TM5 orientation	TM4–TM5
Mouse μ opioid receptor (μOR) ^b (Class A)	4DKL	TM5–TM6	Extensive TM5–TM6 contacts; T4 Lysozymes form part of the interface	1 × 10 μs CG 3 × 1 μs CG	Very few changes are observed	TM5–TM6

Table 1 continued

Receptor	PDB	Original dimer interface	Notes on the initial structure	Number of simulations ^d	Changes after simulation	Final dimer interface
<i>Loosely packed dimers</i>						
Squid Rhodopsin ^{a,b} (Class A)	2Z73	TM5–TM5	A loose interface in the cytoplasmic side of TM5	3 × 1 μs CG	The dimer adopts a TM5–TM6 orientation	TM5–TM6
Human β2 adrenergic receptor (β2AR) ^b (Class A)	2RH1	TM1–TM1	A loose interface in the extracellular side of TM1	3 × 1 μs CG	The dimer adopts a TM1–TM2–H8 orientation	TM1–TM2–H8
<i>Dimers with intermediate cholesterol molecules</i>						
Human mGluR1 Metabotropic glutamate receptor (Class C) ^e	4OR2	TM1–TM2	cholesterol molecules are stacked between the protomers in the upper bilayer leaflet	3 × 10 μs CG w. original cholesterol 3 × 10 μs CG w/o cholesterol 3 × 10 μs CG in 9:1 POPC; cholesterol	Increased contacts for the TM1–TM2 interface; partial H8–H8 contacts; cholesterol molecules leave the interface	TM1–TM2–H8
Human β2 adrenergic receptor (β2AR) ^{b,e} (Class A)	2RH1	TM1–H8	Loose TM1–H8 contacts; cholesterol molecules are stacked between the protomers in the lower bilayer leaflet	3 × 10 μs CG w. original cholesterol 3 × 10 μs CG w/o cholesterol 3 × 10 μs CG in 9:1 POPC; cholesterol	The dimer adopts a TM1–TM2–H8 orientation different to the others; original cholesterol molecules leave the interface;	Alt. TM1–TM2–H8

^a Dimers are also represented by PDB entries 2I36 and 2I37 for Rhodopsin, 3DQB, 4A4 M and 3PXO for activated Rhodopsin/Metarhodopsin II/Opisin, 5C1 M for μOR TM1–TM2–H8, 3NYA for Squid Rhodopsin, 3OE6, 3OE8 & 3OE9 for CXCR4

^b Biological assemblies, as generated with PDBePISA

^c Receptor classification based on the Kolakowski A–F system

^d Number and length of different simulations. CG, Coarse-Grained

^e The simulations were performed in the presence of original cholesterol molecules, in the absence of any cholesterol and in a 9:1 POPC: cholesterol membrane

Table 2 Overview of simulated theoretical models of GPCR homodimers, including a brief description of the initial dimer interfaces and the changes observed during simulation

Receptor	Original dimer interface	Notes on the initial structure	Number of simulations ^a	Changes after simulation	Final dimer interface
AFM/Cryo-EM Bovine Rhodopsin (Class A)	TM4–TM5	Inferred from AFM/Cryo-EM restraints; More loose interface compared to other TM4–TM5 dimers	$3 \times 1 \mu\text{s}$ CG	The inter-protomer distance is reduced	TM4–TM5
AFM/Cryo-EM Bovine Rhodopsin (Class A)	TM1–TM1	Inferred from AFM/Cryo-EM restraints; A loose interface in the extracellular side of TM1	$3 \times 1 \mu\text{s}$ CG	The dimer adopts a TM1–TM2–H8 orientation	TM1–TM2–H8
Human mGluR1 Metabotropic glutamate receptor (Class C)	TM4–TM3	TM4–TM4 contacts; This dimer has been classified as energetically weak	$3 \times 1 \mu\text{s}$ CG	The dimer adopts a TM4–TM5 interface similar to crystallographic dimers; Cross-linking restraints are satisfied	TM4–TM5
Human δ opioid receptor (δ OR) (Class A)	TM4–TM3	TM4–TM4 contacts; This dimer has been classified as energetically weak	$3 \times 1 \mu\text{s}$ CG	The dimer adopts a TM4–TM5 interface similar to crystallographic dimers; Cross-linking restraints are satisfied	TM4–TM5
Mouse μ opioid receptor (μ OR) (Class A)	TM4–TM3	TM4–TM4 contacts; This dimer has been classified as energetically weak	$3 \times 1 \mu\text{s}$ CG	The dimer adopts a TM4–TM5 interface similar to crystallographic dimers; Previous modeling and experimental restraints are satisfied	TM4–TM5

CG Coarse-Grained

^a Number and length of different simulations

elastic network was applied to protein backbone particles, and secondary structure elements were defined using restraints derived from secondary structure calculations with DSSP [54]. Details on the properties of the network constraints are given in the Supplementary Material, available online. The system was solvated using the standard MARTINI water and ion model, with a salt concentration of 0.15 M NaCl. Each system was subjected to energy minimization, followed by equilibration over 8 ns with gradually decreasing restraints on the protein particles to let the membrane and solvent relax around the proteins. Subsequently, each system was simulated without any restraints, using a 20 fs time step. The properties of each simulation system are presented in Tables 1 and 2 and Supplementary Table S1. All CG-MD simulations were performed using GROMACS v. 4.5.7 [55].

Reverse coarse-graining and all-atom molecular dynamics simulations

The lowest energy frame from the last nanosecond in each CG-MD simulation was used for the reconstruction of atomistic protein models. The group of atoms in each initial all-atom system corresponding to a protein CG particle was translated to the simulation results, such that the center of mass of the group was on the CG particle's final position

[56, 57]. The Backward method was applied to reconstruct all-atom representations, using a geometric projection algorithm for reverse Coarse-Graining, followed by a position restrained, force field-based relaxation refinement [58]. This treatment relaxed the unphysical bonds while keeping the overall structure and interactions of the simulated systems intact. Backmapping simulations were performed using GROMACS, the Backward pipeline and the CHARMM36 force field for proteins and lipids [59–61].

A number of subsequent all-atom simulations were performed, to further refine the all-atom models produced by Backward. In cases when it was required, the peptide bond geometry was optimized and localized, biased MD simulations [62] were used to strengthen the integrity of the helical segments, by steering the backbone atoms towards their original secondary structure for short (0.5–2 ns) times. Afterwards, the dimers were re-inserted into an explicit POPC bilayer and, after a number of initial equilibration stages with gradually decreasing restraints, were subjected to a simulated annealing refinement with no position restraints. All-atom simulations were performed using NAMD v. 2.10 [63] and the CHARMM36 force field [59–61]. Detailed information on the simulation protocols can be found in the Supplementary Material, available online. All simulation results and models were analyzed using various GROMACS utilities and analysis scripts

[64], VMD and the Carma suite for structural dynamics analysis [65]. Visualization and plots were built with VMD and the Matplotlib module [66].

Interface analysis and classification

Protein–protein and protein–lipid interactions for CG-MD results were defined using a distance cut-off of 7.5 Å between particles, based on the properties of the MARTINI force field [49, 50]. An additional, time dependent analysis was performed for the change in the Buried Surface Area (BSA) during simulation, calculated as the difference between the Accessible Surface Area (ASA) values of the complex and the isolated protomers. All Coarse-Grained surface calculations were performed using the GROMACS implementation of the Double Cube Lattice algorithm [67] and a 5.2 Å probe radius, based on the properties of the MARTINI water model.

For the initial crystal structures and the final all-atom representations, surface calculations were performed using a 1.4 Å probe radius and DSSP [54] for the calculations. Interface residues for the starting structures as well as the all-atom representations of the final dimer models were further classified as parts of the core or the rim of the BSA, using the Levy classification system [68]. The Relative Accessible Surface Area (RSA) of each residue was calculated as its absolute ASA value, normalized by the maximum ASA value for this residue, as measured in extended Gly-x-Gly tripeptides [69, 70]. Interface residues with RSA < 25 % in the complex and RSA > 25 % in the unbound state were recognized as interface core residues, while residues with RSA > 25 % both in the complex and in the unbound chains were assigned to the rim [68]. An additional interface classification was performed through computational alanine scanning (Ala-scan) mutagenesis. The process involves substitution of each interface residue to alanine, followed by estimation of the change in interaction energy as a result of the mutation ($\Delta\Delta G$). Residues with a $\Delta\Delta G$ value above a specified cut-off are considered as potential Ala-scan hot spots. Alanine scanning was performed with FoldX v. 3.0b6 [71, 72] for protein–protein interactions, using a 1.5 kcal/mol cut-off for the identification of hot spots, while protein–heteroatom interactions were analyzed using the ABS-Scan web server for protein–ligand alanine scanning [73], with a cut-off value of 0.5 kcal/mol.

Dynamical network analysis

The results of the CG-MD simulations were used for the construction and analysis of dynamic networks. Dynamical network analysis applies concepts from network theory to the study of biomolecular structures, by visualizing a

structure as a network defined by nodes connected by edges where the nodes represent atoms or molecules, while the edges are the non-bonded interactions. Additional information is provided by the results of Molecular Dynamics simulations, used to weigh the importance of interactions in the network as well as the clustering of network elements through community analysis.

In this work, the procedure originally developed by *Sethi and co-workers* [74] was adapted, slightly modified to accommodate the nature of Coarse-Grained simulations. The last 100 ns of each CG-MD simulation were used as input for network preparation and analysis. Each CG amino acid was represented by a single node, centered at its backbone particle. Protein–protein contacts were defined using the appropriate for MARTINI CG representations distance cut-off. Edges were defined to connect pairs of nodes if the corresponding elements were in contact for more than 75 % of the analyzed trajectory [74]. Covalently bonded residues and nearest neighbors were not considered to be in contact, as they lead to a number of trivial sub-optimal paths in the dynamical network; these also included cysteine pairs forming disulfide bonds, as well as all elastic network restraints used in CG-MD simulations. The weight (w_{ij}) of an edge between nodes i and j was derived from the appropriate pairwise correlation value (C_{ij}) calculated as the normalized covariance for the MD trajectory. Covariance calculations were performed through Cartesian Principal Component Analysis (PCA), and the results were subsequently used to derive the edge weights, with the weight value defined as $w_{ij} = -\log(C_{ij})$.

The physical network of nodes and edges contains substructures of nodes or communities that are more densely interconnected to each other than to other network elements. In the concept of Molecular Dynamics, these network communities define structural elements which move in concert with each other. The optimum community structure is found by maximizing the modularity value (Q), which is a measure of difference in probability of intra- and inter- community edges. Q can have a maximum value of 1; large values of Q indicate better community structure. The shortest paths between pairs of nodes belonging to 2 different communities are calculated and analyzed for communication across communities in the network. Of these intercommunity links, all edges connecting any two of these communities are identified. Edges with the greatest betweenness are pinpointed, and the nodes connected by these edges are established as critical for communication between the structural elements represented by these communities [75]. Trajectory analysis, network preparations and visualization were performed with a slightly modified version of the NetworkView plugin in VMD [53], designed for Dynamical Network Analysis. Correlation (normalized covariance) calculations were performed

through PCA using Carma [65]. Community analysis was performed with the Girvan-Newmann algorithm [75], using the Luthy-Schulten Network Tools [74].

Additional methods

Additional information on Molecular Dynamics simulations and dynamical network analysis can be found in the Supplementary Material, available on line.

Results and discussion

Structural features and diversity in crystallographic GPCR homodimers

A number of structures involving parallel homodimers are available for various GPCRs, featuring symmetric interfaces belonging to the TM1–TM2–H8, TM4–TM5 and TM5–TM6 dimerization types (Fig. 1). Although the overall orientation between different receptor dimers of the same type is generally similar, several features show significant diversity between different receptors and even between dimers of the same receptor. Based on characteristics like the extent of the Buried Surface Area (BSA) and the presence or absence of intermediate hereroatoms, the available crystallographic evidence on GPCR dimer interfaces may be divided into three main categories. Receptor dimers may appear as closely packed interfaces, featuring extensive contacts between the protomers, as loosely packed interfaces, featuring a similar orientation as the closely packed dimers but with a significantly smaller interface and as dimers which, in addition to protein–protein contacts, involve intermediate cholesterol molecules between the protomers (Table 1).

Closely packed homodimers involving the TM1–TM2–H8 interface have been crystallized for bovine Rhodopsin, both in its ground—state form [76] and the ligand—free Opsin intermediate [77], the κ opioid receptor (κ OR) [11] and the first of the two different interfaces found in the crystals of μ opioid (μ OR) [10] and β 1 adrenergic (β 1AR) receptors [78]. The structures of κ OR and μ OR dimers involve fusions with T4 Lysozyme in the protomers' cytoplasmic sides; however, no contacts between the T4L subunits are observed in the TM1–TM2–H8 dimers. Despite the similar orientation of the dimers, significant differences can be observed. An important factor of discrimination between the TM1–TM2–H8 dimers is the relative orientation of each protomer with respect to the other. While in the β 1AR and κ OR dimers the two protomers are almost parallel to each other and would be expected to sit perpendicular to a membrane plane, in other cases the protomers' principal axes are offset by an angle up to 20°.

Characteristic examples include the Rhodopsin and Opsin TM1–TM2–H8 dimers; the different angle of the protomers in these two structures leads to formation of different protein–protein contacts in the two structures. It should be noted that electron crystallography and Cryo-EM experiments for Rhodopsin have shown the protomers adopting a parallel orientation [33, 36]. As such, it can be hypothesized that this angle between the protomers possibly occurred due to the lack of a suitable membrane analogue during crystallization. Another feature of variability observed is the orientation of the TM1 helix, which protrudes away from the helix bundle in the κ OR and β 1AR but is closer to TM2 in other structures. These differences in orientation essentially lead to a significant diversity in the observed TM1–TM2–H8 dimers.

TM4–TM5 dimers are observed in the structures of squid Rhodopsin [79], the A2A adenosine receptor [80] and the ligand—free β 1AR, as well as the class F Smoothened receptor. Cytochrome b562 subunits replace the N-terminal domains but do not form interactions with each other in the Smoothened structure. Similarly, the T4L subunits in the A2A structure do not come into any contact. Therefore, these dimers are unlikely to have been influenced by the presence of these fusions. A model of oligomerized Rhodopsin, based on experimental restraints derived from AFM and Cryo-EM data [31, 32] also displays a TM4–TM5 interface, bearing significant resemblance to the β 1AR dimer. Although the extent of the interface differs among structures, all TM4–TM5 dimers display a similar orientation, involving contacts in the transmembrane regions as well as the intracellular ICL2 loop, with some dimers also displaying contacts with the cytoplasmic side of TM3. However, the inter-protomer distance is larger in the cases of β 1AR, A2A and the AFM Rhodopsin model, while the Smoothened and Squid Rhodopsin structures form more closely packed dimers.

Available TM5–TM6 dimer structures include the second, more prominent dimer interface in the μ OR structure, as well as two different interfaces for the CXCR4 chemokine receptor [81]. The μ OR dimer consists of extensive contacts between the TM5 and TM6 helices from each protomer that form a transmembrane four helical bundle. Another TM5–TM6 interface is observed for the It1t-bound CXCR4 receptor, however, this dimer involves almost exclusively residues in the extracellular side of the receptors, with very few contacts in the cytoplasmic end of TM5. Furthermore, a variant dimer is formed in the CVX15-bound CXCR4 crystal structure, involving the same interface on the extracellular side but with extensive contacts involving TM4 and the ICL2 loop on the cytoplasmic side. Thus, this interface seems to appear as an intermediate between a TM4–TM5 and a TM5–TM6

dimer. It is important to note that in all three aforementioned structures T4 Lysozyme constructs have been fused in the place of the ICL3 loop connecting the TM5 and TM6 helices and that in the μ OR and It1t-bound CXCR4 dimers the two T4L subunits form contacts with each other, although the degree in which T4L participates in the interface varies. As such, it is unclear whether the TM5–TM6 dimer orientation has been influenced by the presence of the T4L fusions.

Apart from the closely packed interfaces referenced above, a number of loosely packed interfaces exist for some GPCRs. The high resolution crystal structure of the β 2 adrenergic receptor (β 2AR) [6] features a parallel homodimer involving a few contacts between the TM1 helix extracellular ends. The AFM-based model of oligomerized Rhodopsin also displays a dimer with almost identical orientation to β 2AR, as part of a Rhodopsin oligomer that features multiple dimer interfaces (Supplementary Fig. S1). Interestingly, loosely packed TM1–TM1 interfaces have been observed in self-assembly Molecular Dynamics simulations of β 1AR and β 2AR oligomers [82]. It is possible that this interface could be the precursor to the formation of a more tightly packed TM1–TM2–H8 dimer. A different loose interface is observed in the Squid Rhodopsin structure, involving small contacts in the cytoplasmic side of TM5. However, whether this interface could be a precursor to a TM4–TM5 or a TM5–TM6 dimer is unknown.

Finally, two GPCR structures present parallel homodimers which involve cholesterol molecules. These include a second interface for β 2AR, as well as the recently solved mGluR1 metabotropic glutamate receptor transmembrane structure [16]. In both cases the receptors adopt orientations resembling a TM1–TM2–H8 dimer. It should be noted that both GPCRs have been solved as chimeric constructs with T4 Lysozyme (β 2AR) and cytochrome b562 (mGluR1). While β 2AR shows no contacts between the Lysozymes, a small portion of the BSA in mGluR1 is formed by cytochrome subunits, located at the extracellular side, substituting the large Venus Fly-trap domains (VFD) of Class C GPCRs. A number of protein–protein contacts are observed in both structures, however, a major part of the interface is formed by cholesterol molecules that stack against each other between the protomers, forming a sterol “bridge” that seems to mediate the dimer through protein–cholesterol and cholesterol–cholesterol interactions. The extent of observed protein–protein, protein–cholesterol and cholesterol–cholesterol intermolecular contacts differs between the two structures; the β 2AR dimer involves less protein–protein interactions and seems to rely more on adjacent cholesterol molecules than mGluR1, which presents more extensive protein–protein interactions.

Stability and dynamics of GPCR homodimers

Examination of the available structural data reveals that, even though GPCRs seem to form similar dimer interfaces, a significant diversity is observed in the configuration of the protein–protein interfaces. To some extent, this diversity can be attributed to each receptor’s distinct structural characteristics. However, several aspects of the interfaces seem to have been influenced by the structure determination process. In order to examine these potential influences and estimate the overall stability and dynamics of GPCR homodimers, each interface was subjected to Molecular Dynamics simulations.

A total of 21 different CG systems were considered for simulation. These dimers were inserted in lipid bilayers and simulated with the MARTINI force field (Supplementary Figure S1). Details for each simulation system are reported in Tables 1 and 2 and Supplementary Table S1. To establish a common time scale for performing simulations, a series of 10 μ s long simulations were performed for three selected case studies, namely, the κ OR TM1–TM2–H8 dimer, the CXCR4 dimer and the μ OR TM5–TM6 dimer. In all cases, structural rearrangements occur very early in the simulations, with the Root Mean Square Deviation (RMSD) of each protein system reaching a plateau within the first 500–1000 ns and remaining relatively stable for the remainder of the simulations (Supplementary Figure S2). Shorter, 1 μ s long simulations for the same systems showed a similar dynamic behavior, with all structural changes occurring in the same time scale as the 10 μ s runs. Convergence for each system was evaluated by measuring RMSD from the starting conformation as well as the change in the dimers’ Buried Surface Area, as measured with a probe adopting the properties of the MARTINI solvent; the analysis showed that the 1 μ s time scale is adequate for observing the stabilization of the different dimer interfaces to a stable orientation (Supplementary Figures S3–S7). As a result, all CG-MD systems were simulated for 1 μ s, with each simulation replicated in triplicate for validation; an exception was made for all simulations of systems involving cholesterol, which were extended to 10 μ s to obtain better sampling for cholesterol movements and interactions (Supplementary Figure S6). The total simulation time accumulates to approximately 270 μ s for the Coarse-Grained systems. The lowest energy representations of the final CG-MD conformations were subsequently used for the reconstruction of atomistic representations, using a reverse Coarse-Graining procedure followed by refinement simulations and these models were further used in interface analysis.

It is important to note that none of the simulations that were performed resulted in the dissociation of a dimer interface into its protomers; this included both the 1 μ s

long simulations conducted for all dimers and the 10 μ s runs conducted for the case studies and cholesterol-associated dimers. However, this observation is in good agreement with what is already known regarding the stability of GPCR dimers and oligomers during simulation. A wide range of self-assembly simulations for several receptors [34, 36, 38, 51, 82] have shown that, even though GPCRs may undergo several early binding and unbinding events before stable assemblies are formed, eventually the simulation systems result in dimers and oligomers that mostly resemble crystallographic dimers such as TM1–TM2–H8 and TM4–TM5, and that remain stable for the remainder of the simulation time, even in simulations conducted in the sub-millisecond scale. In fact, biased Molecular Dynamics and Umbrella Sampling simulations for various dimer interfaces of Rhodopsin [36], β 2AR [40], μ OR and κ OR [83] have shown that known crystallographic TM1–TM2–H8, TM4–TM5 and TM5–TM6 interfaces are characterized by deep energy wells that would be expected to preclude dissociation in the time scales achieved by unbiased Molecular Dynamics simulations. In agreement with these simulations, kinetic analysis derived from Umbrella Sampling simulations for different dimers in β 1AR and β 2AR have shown that strong dimers, such as the TM1–TM2–H8 interface have an estimated lifetime in the scale of seconds or even minutes, while even weaker interfaces such as TM4–TM4 have an estimated lifetime of milliseconds to seconds [40]. Taken together, all these observations suggest that a complete and accurate description of spontaneous association and dissociation events in GPCR oligomerization may not be achieved by MD simulations, even in cases where simplified models are utilized.

In all simulations the overall GPCR 7TM fold remains stable, with most major fluctuations observed mainly for the loop regions. This is reflected in the RMSD measurements, which range between the values of 5–7 Å when the entirety of the protein particles are included in the calculations, but are in the ranges of 3–4 Å when only the helical segments are considered. These differences can be attributed to the nature of the elastic network used for the Coarse-Grained systems, which applies harmonic restraints to the transmembrane helical fold but not the loop region. Although such a network structure may lead to increased RMSD values, it minimizes the bias to the stoichiometry of the dimer, which has been suggested to be significant when strict elastic restraints are applied to the loops [83].

With a few notable exceptions, explored in greater detail in the following sections, the overall stoichiometry of the crystallographic dimer interfaces remained the same, indicating that the known crystallographic dimers are stable entities. Furthermore, simulations of alternative interfaces for a number of putative dimers for the μ OR,

δ OR and mGluR1 receptors resulted in rearrangements for the assemblies, which adopted features very similar to the crystallographic interfaces, as will be described. These observations are in good agreement with spontaneous self-assembly and Umbrella Sampling simulations for different dimer interfaces in Rhodopsin [36], which have indicated that symmetric dimers resembling those produced by crystal structures are among the most stable oligomerization interfaces, while other, alternative interfaces, even in cases when they meet experimental evidence, are significantly weaker.

However, the observed RMSD values for each dimer were higher than the RMSDs observed for its individual protomers in almost all cases (Fig. 2a, see also Supplementary Table S1). These observations suggest that the main contributions to movements observed during simulation come from rearrangements in the orientation of the interfaces rather than alteration in the structure of the protomers themselves. Even in cases where the overall dimer stoichiometry does not change, these rearrangements lead to the formation of novel protein–protein interactions missing from the original structures; interestingly, several of these novel putative contacts have been identified as parts of GPCR dimer interfaces by several experimental evidence, such as cross-linking and synthetic peptide studies, as it will be outlined. Overall, these results suggest that the orientation of the protomers in the original crystal structures may not be optimal in all cases but that at least some aspects of these crystallographic interfaces could be considered artifacts produced by the applied structure determination processes.

Structural and dynamic behavior of closely packed dimers

The closely packed TM1–TM2–H8 dimers do not display significant deviations, retaining most contacts observed in the original interfaces. Among the different dimers of the dataset, the Rhodopsin, Opsin and μ OR dimers displayed the most profound movements regarding the orientation of the protomers, where an angle between the receptors' principal axes is observed. In all three cases this angle is reduced, enabling the formation of more extensive contacts involving residues in the TM1 helix (Fig. 2b). New contacts also involve residues in the first extracellular loop (ECL1) as well as the extracellular end of TM2. The μ OR dimer also displays the formation of more extensive contacts in the H8 helix, closely resembling the ones in κ OR, while the H8–H8 interface in the Rhodopsin and Opsin dimers resembles more the one in the β 1AR dimer. On the other hand, the κ OR and β 1AR TM1–TM2–H8 dimers display very few changes, retaining their initial configuration. It is important to note that the κ OR interface has

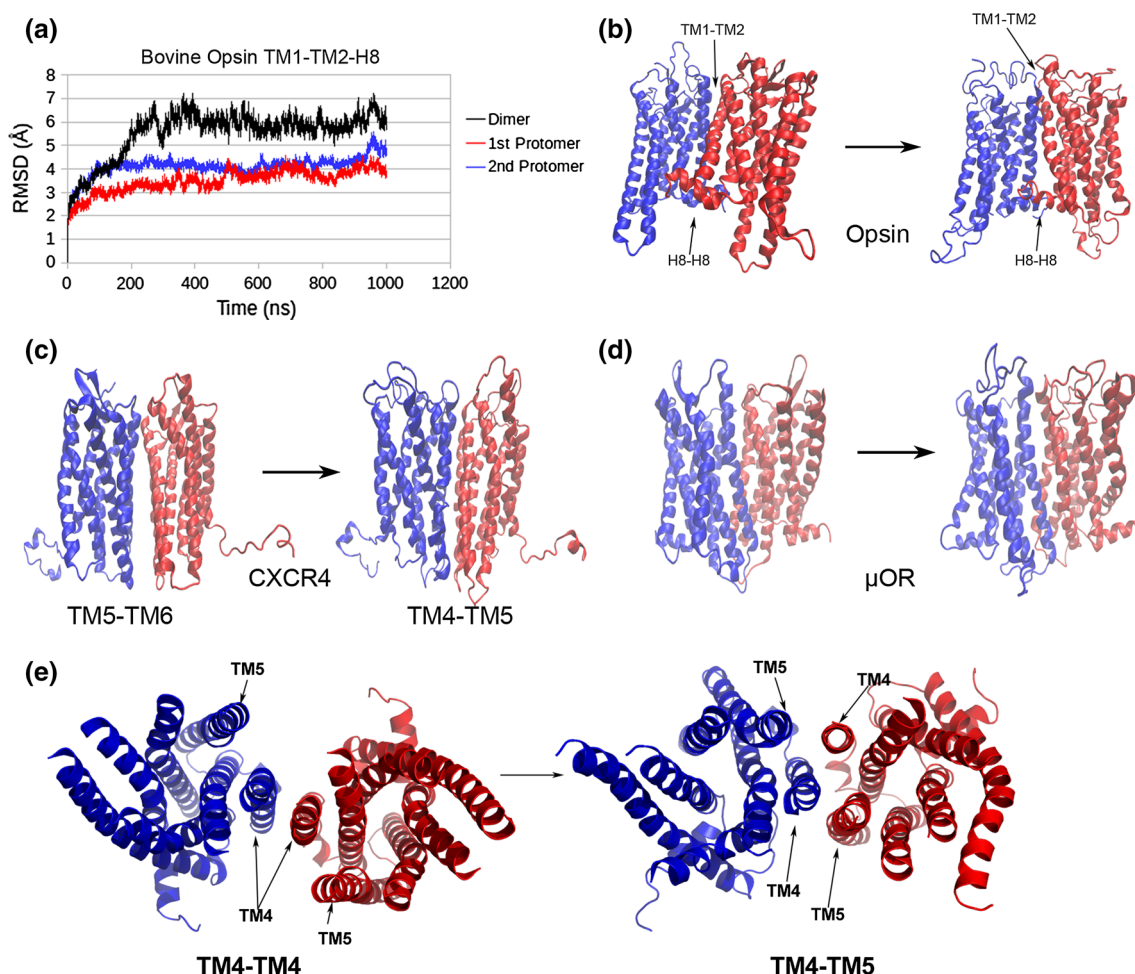


Fig. 2 Examples of results from GPCR homodimer simulations. **a** Time—dependent Root Mean Square Deviation (RMSD) measurement for the Opsin TM1–TM2–H8 dimer CG-MD simulation. RMSD has been measured with respect to the starting conformation, using the coordinates of the backbone particles. The *horizontal axis* displays time (in ns), while the *vertical axis* displays RMSD (in Å). All protein segments are considered, including both the transmembrane segments and the interhelical loops. The RMSD of the entire dimer is colored black, while the RMSDs of the isolated protomers are colored blue and red, respectively. **b, d** Initial (*left*) and final conformations of the

Opsin (**b**), CXCR4 (**c**) and μOR TM5–TM6 (**d**) interfaces. **e** Initial (*left*) and final (*right*) conformations of the alternative dimer interface for δOR. The dimer starts as a TM4–TM4 interface and, during simulation, shifts to a TM4–TM5 interface with features similar to those of crystallographic TM4–TM5 dimers. All structures are shown in cartoon representation, with different protomers colored blue and red. Initial conformations are represented by the starting crystal structures, while final conformations are represented by all-atom reconstructions of the CG-MD results

been found to fit well to available Cryo-EM volume data for the Metarhodopsin I dimer [33, 36], displaying better correlation in comparison to the Rhodopsin and Opsin original crystallographic interfaces. Overall, these observations suggest that the receptor orientation found in the original κOR structure and proposed by the results of the CG-MD simulations may be more favorable than the original TM1–TM2–H8 structures.

The TM4–TM5 dimers also preserve their overall orientation, with the β1AR and A2A adenosine receptor dimers showing the most profound changes, while the Smoothened dimer remains largely unchanged, with most significant movements observed, as expected, in the

extracellular and cytoplasmic regions. All TM4–TM5 dimer simulations result in an almost identical configuration, resembling the one observed in the Smoothened crystal structure. These involve more extensive contacts in TM4 and TM5 and formation of new contacts, formed by residues in the ICL2 loop and the cytoplasmic end of TM3. It is important to note that the latter involves residues in the area of the D(E)RY motif and the TM3–TM6 ionic lock, an important feature of Class A GPCRs that is involved in the process of GPCR activation. The motif seems to form part of the dimer interface in all Class A TM4–TM5 dimer simulations. Interestingly, similar contacts are observed both in the original structure and the simulation result of

Smoothened, although the latter lacks the D(E)RY motif. However, a closer examination of the available Smoothened receptor structures reveals a tight network of contacts between TM3 and TM6 in each protomer, involving residues that are conserved among Class F GPCRs that may serve the same purpose as the TM3–TM6 ionic lock of the D(E)RY motif of Class A receptors [12].

In contrast to the above, significant deviations are observed for some of the TM5–TM6 dimers, in which the original structures contained T4 Lysozyme fusions in close proximity. This is particularly true for the two simulated CXCR4 dimers in the absence of the T4L constructs, which display significant movements. Simulations for both CXCR4 dimers give very similar results. The protomers come closer to each other, forming contacts using residues in transmembrane helices TM4 and TM5 and the second intracellular loop (ICL2), while most TM6 contacts from the original interface dissociate, with the exception of a few interactions involving residues in the second extracellular loop (Fig. 2c). Additional contacts involving the cytoplasmic end of TM3 are also observed. The final models bear a striking resemblance to the other TM4–TM5 dimers. Geometry analysis reveals that most rearrangements occur during the first 30–50 ns of the simulations; subsequently, both dimers remain relatively stable for the rest of the simulations. It is important to note that these Coarse-Grained results are in good agreement with previously published results from much shorter all-atom simulations of CXCR4 [37, 84]. Furthermore, the formation of a TM4–TM5 dimer for CXCR4 is in agreement with experimental studies involving synthetic peptides, which have shown that CXCR4 forms homodimers using residues in TM4 [85] in living malignant cells. Overall, these observations clearly display the important influence the T4 Lysozyme's presence may have had during the structure determination process.

Contrary to CXCR4, the simulated μ OR TM5–TM6 interface displays less extensive changes. This crystallographic dimer also includes T4 Lysozymes forming contacts with each other as well as the receptors, but their contributions to the interface are much smaller, and the majority of the buried surface area is formed by extensive TM5–TM6 contacts from each protomer, that form a transmembrane four helical bundle. As such, the interface remains largely unchanged, with significant movement observed mostly for the cytoplasmic ends of the TM5 helices (Fig. 2d). Considering that the T4 Lysozyme was fused in place of the ICL3 loop connecting TM5 and TM6, the original orientation of these transmembrane segments was probably influenced by the presence of T4L during the structure determination process and, therefore, such movements are to be expected. The overall orientation of the protomers remains unchanged, showing the significant

stability of the μ OR TM5–TM6 interface. However, it is still difficult to assess the potential influence that the T4 Lysozyme fusions may have had in the packing and orientation of the receptors during crystallization.

Simulation of loosely packed dimers leads to closely packed interfaces

Significant structural movements are observed in all simulations performed for the loosely packed Squid Rhodopsin TM5–TM5 and β 2AR TM1–TM1 crystallographic homodimers (Supplementary Fig. S8). The squid Rhodopsin TM5–TM5 dimer changes its configuration through the slight rotation of each receptor with respect to their vertical axes, leading to the formation of contacts between the TM5 and TM6 helices. These rearrangements are observed within the first 40–60 ns of the simulation, with the dimer being relatively stable afterwards. This new TM5–TM6 interface bears some resemblance to the μ OR TM5–TM6 dimer, but the interacting transmembrane helices do not seem to form a helical bundle. This, however, can be attributed to the distinct structural characteristics of squid Rhodopsin, which displays TM5 and TM6 helices with significantly longer intracellular segments adopting very different conformations compared to the ones observed in vertebrate GPCRs.

More extensive rearrangements are observed in the case of the β 2AR TM1–TM1 dimer. The distance between the protomers was decreased, eventually leading to the formation of a TM1–TM2–H8 interface. The new dimer is very similar to other closely packed TM1–TM2–H8 interfaces. Combined with observations from self-assembly β 2AR simulations [82], this result could suggest a potential mechanism for the formation of TM1–TM2–H8 GPCR dimers, in which two receptors first come in contact using the extracellular ends of their TM1 helices.

Rearrangements of the interfaces in cholesterol-bound mGluR1 and β 2AR dimers

A special case needs to be made for the mGluR1 and β 2AR simulations, since a considerable part of both dimers' original interface areas is not formed by the receptors themselves but rather by adjacent cholesterol molecules that seem to mediate the formation and stabilization of the assemblies. An abundance of experimental and computational evidence has suggested that the presence and concentration of cholesterol seems to be connected with the ability of many GPCRs to oligomerize, contributing to the formation and stabilization of quaternary structures or the preference of receptors towards specific types of interfaces [38, 86]. In order to evaluate the features of these cholesterol mediated configurations in the framework of the

present study, the mGluR1 and β 2AR dimers were simulated in the presence of cholesterol molecules appearing in the original crystal structures. Although major protein movements occurred in the 1 μ s time frame used for other GPCR dimers, these simulations were extended to 10 μ s, to achieve better sampling for cholesterol.

CG-MD simulations for both systems displayed a similar behavior for the receptors. During the first steps of both simulations the protomers were observed to decrease the distance between them. Interestingly, cholesterol molecules seemed to remain in their original positions between the receptors, despite the fluidity of the membrane (Supplementary Figs. S9 and S10). Cholesterol–receptor contacts were retained, while contacts between the sterols themselves became more extensive. These movements were observed in both dimer simulations, with the β 2AR dimer displaying more profound movements compared to mGluR1. The above observations seem to support the idea that cholesterol enhances the movement of the protomers, essentially driving them towards decreasing their distance. Cholesterol molecules remained in place until the two protomers reached a certain distance from each other. This occurs in the first 100 ns for the mGluR1 dimer, with more time required for the β 2AR system (120–150 ns), presumably due to the increased inter-protomer separation. Subsequently, both simulations displayed a gradual exchange between cholesterol molecules and adjacent POPC lipids, eventually leading to removal of almost all cholesterol molecules from their original positions between the receptors, allowing the protomers to further decrease the distance between them.

With regards to the configuration of the final dimer interface, both mGluR1 and β 2AR simulations resulted in an increase of TM1–TM2 contacts, as well as contacts involving the ECL1 loop. Furthermore, mGluR1 also displays the formation of new contacts in the cytoplasmic ends of TM7, suggesting a potential H8–H8 interface. However, the final configuration of the β 2AR dimer differs from the other TM1–TM2–H8 interfaces, with regards to the orientation of the H8 helices. While other crystallographic TM1–TM2–H8 dimers display direct H8–H8 contacts involving the “outward” side of H8, the β 2AR dimer shows TM1–H8/TM1–H8 contacts in the “inward” side, that also faces the cytoplasmic end of TM1 (Supplementary Fig. S11). Interestingly, similar configurations of H8–H8 contacts have been observed in self-assembly simulations of both β 2AR [51] and Rhodopsin [36], suggesting it is possible that multiple variations may exist for the TM1–TM2–H8 interface.

In order to further assess the possible contribution of cholesterol towards strengthening the interface, additional simulations were performed for these receptors without the original, crystallographic cholesterol molecules; these

included a set of 10 μ s long simulations in a purely phospholipid bilayer as well as a bilayer with a 9:1 POPC: cholesterol ratio (Supplementary Figures S6, S12 and S13), mimicking the ratio of monoolein: cholesterol that is usually applied during GPCR crystallization [6, 16]. These simulations resulted in dimer orientations similar to the ones produced by the original simulations. However, the two protomers required significantly longer time to decrease the distance between them when no cholesterol was used, as shown by time—dependent measurement of the distance between the two protomers’ centers of mass for β 2AR (Supplementary Fig. S13). It is important to note that self-assembly simulations for Rhodopsin [36, 49] and β 2AR [51] in bilayers composed exclusively by phospholipids have shown that, even in the absence of cholesterol, simulation setups will eventually produce all known dimerization interfaces. Furthermore, cholesterol has been proposed to enhance dimerization for some GPCRs but has no apparent influence in others. These observations would suggest that, although important, cholesterol is not a prerequisite for the formation of known dimer interfaces. On the other hand, simulations of β 2AR featuring different levels of cholesterol in the membrane have shown that cholesterol may influence the stabilization of specific dimer interfaces by interacting with specific binding sites on the receptors [38], suggesting that even though dimer formation may occur in the absence of cholesterol, the presence of the latter in the membrane may help guide the process. Interestingly, these simulations have also shown that cholesterol may enhance the speed of dimer association in some cases. Our results are in agreement with these observations, suggesting that the presence of cholesterol may accelerate the formation of GPCR dimers.

It should be noted that, in both systems, not all cholesterol molecules leave the dimer interface. Instead, cholesterol molecules appear at the rim of the dimer interface, acting as intermediates between the protomers. Furthermore, some of the original cholesterol molecules relocated to different binding sites upon the receptors, including the cleft formed by TM2 and TM4 in both the mGluR1 and β 2AR simulations and the lower leaflet side of TM4 in β 2AR (Supplementary Fig. S12a). To further investigate the behavior of these putative cholesterol binding sites, we performed evaluations for the distribution of cholesterol in the 9:1 POPC: cholesterol simulations by measuring its partial mass density and distribution around the proteins. Impressively, cholesterol displayed a significant preference for these regions, validating the observations from the simulations featuring only the original cholesterol molecules of the crystal structures (Supplementary Figure S12b). Interestingly, these potential cholesterol binding sites have also been proposed by all-atom and Coarse-Grained simulations for a number of Class A GPCRs

[38, 86] and the observations from the mGluR1 simulation suggest that similar binding sites may exist in receptors from other Classes as well. In fact, some of these sites have been implicated in regulating dimer formation. For example, a highly occupied cholesterol binding site is located in the TM4 helix of β 2AR. This site has been proposed to regulate the participation of TM4 in dimer formation during β 2AR simulations [38]. Furthermore, experimental studies on protein-cholesterol interactions have also shown that the presence of cholesterol in that region may influence receptor oligomerization in μ OR [87]. Although different GPCRs are expected to interact with cholesterol in a different manner, it could be speculated that the presence of cholesterol in that site could have an impact in the formation of TM4–TM5 inter-protomer contacts in general. However, it should be noted that the simulation setups for these systems may not be adequate to capture the dynamics of protein-cholesterol and lipid-cholesterol interactions accurately. The main limitation is the lack of adequate sampling of binding/unbinding events for all observed interaction sites, which would require significantly larger simulation times that are outside the scope of the present study. In the absence of sampling, at least some of the measured properties are expected to be biased. As such, although the results produced by our simulations on GPCR-cholesterol interactions are interesting, they should be interpreted with great care.

Simulations of theoretical dimer models reproduce known crystallographic interfaces

In addition to the crystallographic dimers, CG-MD simulations were performed for a number of theoretical dimer models. These included additional models for Rhodopsin, based on restraints derived from biophysical evidence as well as putative dimers for mGluR1, δ OR and μ OR, featuring an alternative dimer interface not appearing in crystal structures. As with the crystal structures of the dataset, these theoretical models were also subjected to Coarse-Grained Molecular Dynamics simulations, reverse Coarse-Graining and refinement. The additional Rhodopsin systems involved primarily two dimer interfaces obtained from the model of oligomerized Rhodopsin, based on restraints derived from Atomic Force Microscopy (AFM) and Cryo-EM experiments [31, 32]. These dimers feature a closely packed TM4–TM5 interface, bearing significant resemblance to β 1AR and A2A dimers and a loosely packed TM1–TM1 interface, very similar to the loose TM1–TM1 dimer from the β 2AR crystal structure. As expected, the CG-MD simulations of these models produced very similar results to the ones reported by CG-MD simulations of the crystallographic dimers. Specifically, the TM4–TM5 model of AFM Rhodopsin displayed the same

behavior as the β 1AR crystallographic TM4–TM5 dimer, by forming more extensive interactions in TM4 and TM5 and new contacts involving the ICL2 loop and the cytoplasmic end of TM3. Interestingly, residues in the D(E)RY motif also form part of the interface, as it is observed for the crystallographic dimers. Similarly, the loose TM1–TM1 model of AFM Rhodopsin proceeded in forming a closely packed TM1–TM2–H8 interface with features akin to crystallographic Rhodopsin TM1–TM2–H8 dimers, in the same manner as the β 2AR TM1–TM1 simulation (Figure S8a).

Special attention must be given to the alternative dimer interfaces that were investigated for mGluR1, δ OR and μ OR. Experimental evidence, involving cross-linking [28, 43] and mutagenesis studies [87] and computational studies including molecular modeling and correlated mutation analysis [88] have proposed that these receptors may use residues in TM4 to form their homo- and heterodimer interfaces. Although restraints defined by such studies can be satisfied by TM4–TM5 dimers produced by crystal structures, other, alternative assemblies can be proposed, including a putative dimer interface featuring symmetric contacts involving mainly the TM4 helix and a part of TM3, henceforth mentioned as TM4–TM4, instead of a TM4–TM5 interface. However, its stability has been found to be questionable. Specifically, Umbrella Sampling simulations have shown that, compared to crystallographic dimers such as TM1–TM2–H8 or TM4–TM5, a TM4–TM4 dimer would be significantly weaker and have classified the interface as transient [36, 40]. In agreement with these simulations, a scoring-based docking assay has recently shown that a dimer involving TM4 and TM3 is favorable only when TM5 is also part of the interface [89]. Finally, additional experimental evidence has proposed potential interface contacts involving the TM5 helix along with TM4 [43]. Taken together, these studies would propose that the stoichiometry of a dimer satisfying such experimental restraints would be a TM4–TM5 interface rather than TM4–TM4.

To further test this hypothesis in the framework of the current study, we modeled putative TM4–TM4 dimers for the δ OR, μ OR and mGluR1 receptors. The models were constructed in a manner that restricted all protein–protein contacts to the TM4 and TM3 helices, excluding TM5, while at the same time satisfying experimental restraints about interface residues in TM4. The models were subsequently subjected to CG-MD simulations for 1 μ s, during which significant rearrangements of the orientation of the protomers were observed. Specifically, in all three cases the protomers rotated with respect to one another, leading to the formation of contacts between residues in TM4 and TM5 and essentially resulting in TM4–TM5 interfaces very similar to the β 1AR crystallographic dimer (Fig. 2e).

Impressively, these new assemblies satisfied all original experimental restraints for TM4, as well as restraints for interface residues in TM5, providing further mitigation towards their validity. As such, our results from these simulations show that the TM4–TM5 assemblies proposed by crystallographic evidence, compared to other putative dimers, are preferable entities.

Apart from shifts in the interface, the new TM4–TM5 dimers present novel features that were also observed during the simulation of crystallographic TM4–TM5 dimers, including the formation of contacts with the cytoplasmic end of TM3. In both δ OR and μ OR receptors, belonging to Class A these contacts included regions around the D(E)RY motif participating in the ionic lock between TM3 and TM6, as it was in the case of the crystallographic interfaces. Interestingly, similar contacts were observed for the mGluR1 TM4–TM5 dimer, belonging to Class C. Although Class C GPCRs lack the D(E)RY motif, the recent crystal structures of mGluR1 and mGluR5 have shown the existence of a similar ionic lock between charged residues in TM3 (R3.53 in the B&W numbering) and TM6 (E6.35 in the B&W numbering) that are conserved among Class C GPCRs and have been proposed as the equivalent of the Class A feature [15, 16].

Contribution of polar and aromatic contacts near the membrane boundaries

Interface analysis was conducted on both the original crystal structures and simulation results, using a surface area definition accompanied by classification of interacting residues based on surface burial. A relative ASA (RSA) term was used, taking into account the differences in size and properties for each amino acid side chain to identify residues as parts of the Buried Surface Area's core or rim. Interface analysis reveals that most core residues are located in the TM1 helix of TM1–TM2–H8 dimers and TM5 of the various TM5 dimers. To a lesser extent, some core residues also appear in the TM4 and H8 areas. Not surprisingly, hydrophobic interactions form the majority of the interface. However, a significant presence of polar and charged residues is also observed. In most cases, these polar residues seem to interact with other similar groups, with a few hydrogen bonds also observed, formed with other polar or charged groups, elements of the backbone or, in some cases, the –OH group of a tyrosine side chain.

With regards to hydrophobic interactions, approximately 9–30 % of each dimer's BSA includes aromatic residues. A more detailed interatomic contacts analysis reveals that a significant number of these aromatic residues form π -stacking interactions, while aromatic-Proline and aromatic-amino group contacts are also observed. Further classification of interacting residues was attempted using

computational alanine scanning mutagenesis, a process that involves substitution of each interface residue to alanine, followed by estimation of the change in binding energy upon mutation. Core residues and Ala-scan hot spots seem to be in good agreement and the overall surface of the Ala-scan hot spots overlaps with the interface core (Fig. 3). More importantly, the most prominent hot spots in almost all cases include intermolecular hydrogen bonds and aromatic residues participating in π -stacking interactions. All identified interface hot spots are included in Supplementary Table S2.

The extent to which these hot spots appear in the interface differs among dimers. Several stacking interactions and polar contacts are observed in the κ OR and CXCR4 dimers and a complex π -stacking network is observed in the class F Smoothed TM4–TM5 dimer (Supplementary Fig. S14). Stacking pairs and hydrogen bonds are also formed in the ground-state Rhodopsin and Opsin TM1–TM2–H8 dimers. The TM4–TM5 dimers in the Rhodopsin AFM model and the A2A and Squid Rhodopsin structures also have some aromatic stacking pairs. On the other hand, the μ OR TM1–TM2–H8 dimer shows no such hot spots in its original crystallographic structure, possibly due to its differences in orientation with κ OR and other TM1–TM2–H8 dimers, as already described. These differences also affect the Rhodopsin and Opsin dimers, which display hot spots in different regions. The β 1AR TM4–TM5 and μ OR TM5–TM6 also show very few aromatic interactions but, instead, have an increased number of polar contacts.

Impressively, some of the most profound hot spots that were identified through this analysis involved residues that have been experimentally found to contribute to GPCR dimer interfaces (see underlined residues in Supplementary Table S2). Examples include stacking interactions involving W175 and Y206, which are located in transmembrane segments TM4 and TM5 of Rhodopsin, respectively. Impressively, those same residues have been identified as parts of a dimer interface in Rhodopsin through cross-linking experiments [90]. Similarly, C316 was also identified as a hot spot in all Rhodopsin TM1–TM2–H8 dimers after simulation, again, in agreement with cross-linking. However, it should be noted that no such contacts were observed in the starting conformation of the dimers, suggesting that rearrangements are required for the original interfaces to be in agreement with experimental evidence. Such hot spots were also observed for the simulated TM4–TM5 dimer models of μ OR, δ OR and mGluR1, again, in agreement with cross-linking evidence for these receptors.

It is important to note that for all simulated dimers, stacking interactions and hydrogen bonds that were classified as core residues or Ala-scan hot spots in the initial structures were retained throughout the simulations.

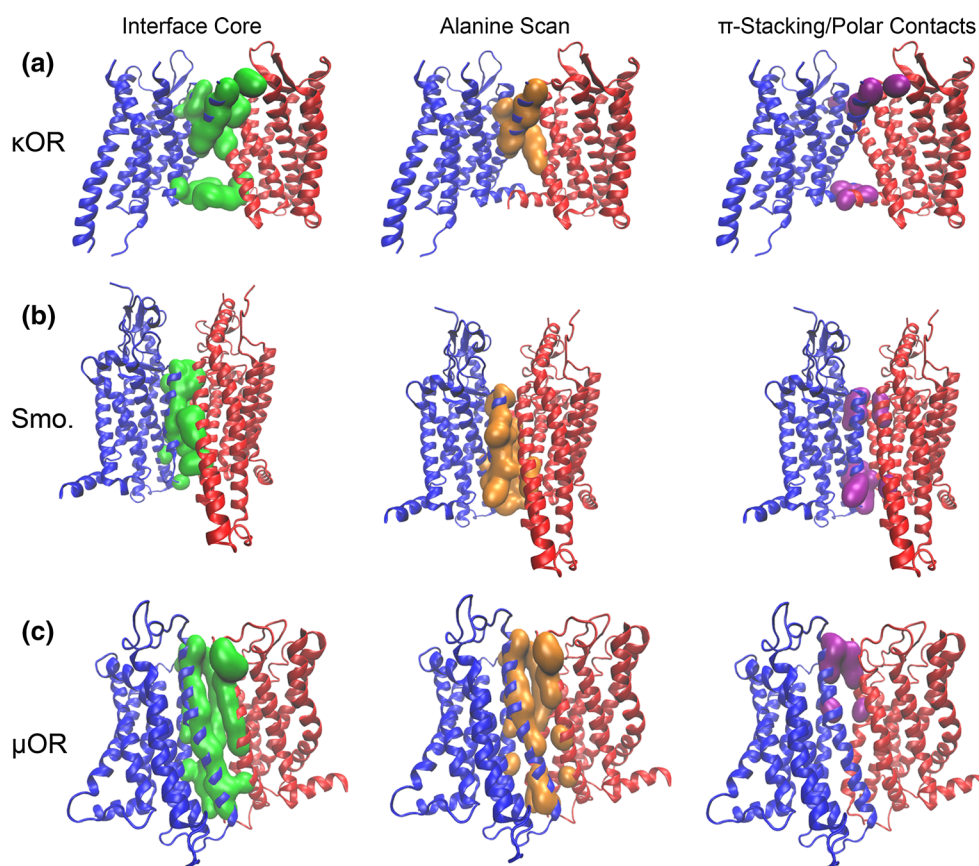


Fig. 3 Examples of interface classification, Ala-scan hot spots and aromatic stacking interactions for the κ OR TM1–TM2–H8 (a), Smo. TM4–TM5 (b) and μ OR TM5–TM6 (c) dimers. Receptors are shown in cartoon orientation, with different protomers

colored blue and red. Interface core residues, hot spots and stacking pairs are shown in isosurface representation and colored green, orange and purple, respectively

Furthermore, in dimers where extensive structural movements were reported, new contacts also included a significant amount of π -stacking interactions, which were formed during the early stages and were retained for the remainder of the simulations (Supplementary Table S2). Interesting cases include the μ OR TM1–TM2–H8 dimer which, after assuming its new configuration, displays the formation of multiple stacking interactions resembling the ones observed in κ OR, as well as the Rhodopsin and Opsin dimers, which resulted in dimers with seemingly identical stacking pairs. Impressively, the TM1–TM2–H8 interface formed by the loosely packed TM1–TM1 dimer also showed interactions in the same regions. Subsequent interface analysis and Ala-scan calculations showed that these new stacking interactions were also classified as core residues and Ala-scan hot spots.

An interesting case of hot spots appears in the mGluR1 and β 2AR dimers with adjacent cholesterol molecules. Protein-cholesterol contact residues in mGluR1 include, among others, W588 and F646 from each protomer, which form stacking—like interactions with the ring segments of

adjacent cholesterol molecules. Similar interactions are observed between cholesterol molecules and F49 from each protomer in the β 2AR—cholesterol dimer. A protein—ligand Ala-scan trial on the initial structures shows that these aromatic residues contribute the most to cholesterol binding, displaying the highest $\Delta\Delta G$ values. As already described, protein—cholesterol contacts are retained for a large part of the mGluR1 and β 2AR CG-MD simulations. Interestingly, after the departure of cholesterol from the mGluR1 dimer interface, these protein—cholesterol hot spots participate in stacking interactions that are classified as protein—protein interface hot spots, very similar to those observed in other GPCR dimers (Supplementary Table S2).

With a few notable exceptions, such as the hydrogen bond formed between Ser residues in the μ OR and κ OR TM1–TM2–H8 dimers (a residue that appears to be conserved in the opioid receptor subfamily), most stacking and polar interactions are located primarily near the cytoplasmic and extracellular ends of the transmembrane helices TM1, TM4 and TM5, as well as the C-terminal H8 helix

(Fig. 3), meaning that these hot spots are mainly located near the membrane boundaries. Impressively, cross-linking experiments aimed at studying the dimer interfaces of various GPCRs, including mammalian Class A GPCRs such as Rhodopsin [90], D2 dopamine [26] and δ OR [28], the Class D Ste2 pheromone receptor in yeast [27] and, more recently, the Class C metabotropic glutamate receptor subfamily [43], have implicated residues in these positions as potential protein–protein interface hot spots. Furthermore, recent experiments involving synthetic peptides and mass spectrometry have shown that oligomerization in Rhodopsin can be disrupted by blocking regions near the membrane boundaries of TM1, TM2 and TM4 [91]. Results from our analysis are in good agreement with these studies, indicating that the significant contributions of residues near the membrane boundaries can be a common feature in several GPCR dimers.

In monomeric GPCRs, residues in such positions could lead to unalleviated hydrophobic mismatch due to difference between the length of the transmembrane segments and the membrane's hydrophobic thickness, resulting to high energy costs due to unfavorable exposure of nonpolar residues to the solvent or polar and charged residues to the lipid bilayer's hydrophobic core. It has been suggested that the energetics of residual hydrophobic mismatch can be an important factor in the aggregation of transmembrane proteins, leading to the formation of protein–protein interactions as a means to alleviate the high energy penalty of the mismatch [92]. Furthermore, Coarse-Grained Molecular Dynamics simulations have proposed that phospholipids with different lipid tail lengths may display different behavior patterns with regards to hydrophobic mismatch, eventually affecting the aggregation rate in the oligomerization of Rhodopsin in model membranes [34]. The close proximity of the strongest hot spots to the membrane boundaries in GPCR dimers, as proposed by the current study, could be a result of this phenomenon and a further indication towards a role for the membrane as a structural determinant in the spatial organization of GPCRs.

Network properties and community organization of GPCR dimers

The dynamics of GPCR dimers, as described by the performed simulation, were further evaluated through the construction and analysis of dynamical networks. The last 100 ns of each CG-MD simulation were used to prepare dynamical networks (Supplementary Table S3) for all studied GPCR dimers. In the networks constructed, each residue of the simulated dimers was represented as a node of the network. Any two non neighboring nodes are connected by an edge if they are in contact for the majority

(>75 %) of the analyzed simulation, i.e. if the distance between them is less than a defined cutoff. Individual edges may have associated weights or lengths based on properties such as correlated motions, energies or physical distance. Edges between nodes are weighted by correlation values obtained from PCA calculations for the simulation, so that the distance between the two nodes reduces as the correlation (or energy of interaction) between the monomers increases. In this sense, edges with the lowest weight values represent the most correlated and, therefore, strongest residue–residue contacts in the network.

A path between two nodes is simply a set of nodes and edges connecting one node to the other, and the path length is the sum of weights for edges in the path. Multiple paths may exist between two nodes. If there are multiple communication paths nearly equal in length, then not all residues along these paths need be considered as important. Instead, only residues or interactions that occur in the highest number of suboptimal pathways need to be conserved to guarantee an effective pathway for communication in the complex. Instead, for two nodes within a connected network, there exists at least one optimal, shortest path between them, and slightly longer paths are referred to as suboptimal.

These paths are considered to be crucial for network communication, representing interatomic contacts that play an important role in the dynamics of the studied structure. A time-averaged, dynamic study of the network's connectivity may utilize these characteristics to identify the substructure of communities into the network. Network communities are defined as a sum of nodes belonging to the same subnetwork, that are more densely interconnected to one another compared to nodes outside this community and can communicate with each other relatively easy through multiple routes. These communities, which can contain both amino acids and nucleotides, are thought to be similar to domains, but are defined by the dynamics of the biomolecules. As such, from a structural viewpoint, these communities can represent parts of the proteins that move in concert with each other. Two different communities can often be in contact or overlapping, with nodes belonging to both clusters. Communication across these groups is estimated by identifying the shortest paths between these intercommunity nodes. Such edges are defined as critical, representing contacts that may prove to be important for communication between two structural segments.

Interestingly, all networks presented a modularity score (Q) in the area of 0.7–0.8 (Supplementary Table S3), displaying an optimal community structure and nearing the values often found in real-world networks [75]. Although these communities may not necessarily correspond to structural features, a number of motion correlations with potential structural significance are reported. More

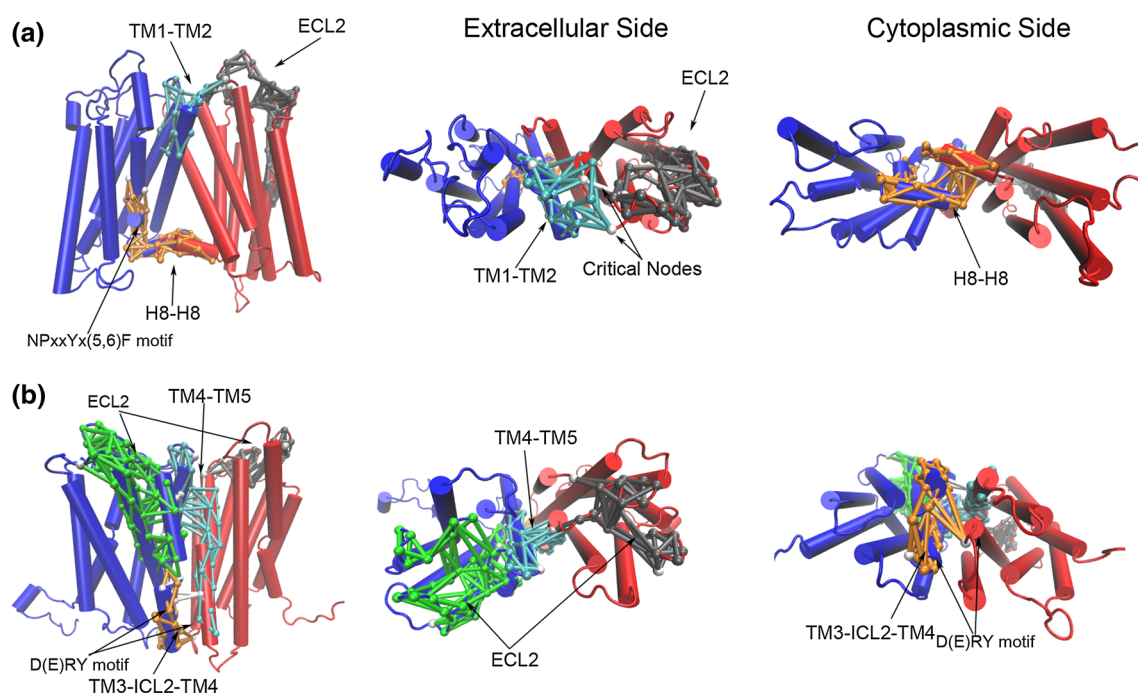


Fig. 4 Inter-protomer and Extracellular Loop 2 (ECL2) network communities for the κ OR TM1–TM2–H8 (a) and CXCR4 TM4–TM5 (b) dynamic networks. Side, extracellular and cytoplasmic views are shown. Receptors are colored blue and red, with helical segments shown as cylinders. Network elements are represented with spheres for nodes and sticks for edges and different communities are depicted

using distinct colors. Critical connections (nodes and edges) between communities are colored white. Dimer interfaces and other important GPCR features are labeled accordingly. Full network representations, with all communities, are shown for these structures in Supplementary Figures S8 and S9

importantly, a conserved appearance of specific communities is observed, revealing possible common features in GPCR structure and dynamics, as well as dimer interactions.

Specifically, simulated TM1–TM2–H8 dimers display the formation of a community containing the dimer interface of the TM1 and TM2 helices (Fig. 4a, see also Supplementary Fig. S15). Such a community is reported for all Class A receptor dimers, including the new AFM Rhodopsin and β 2AR assemblies formed during their respective simulations. A second dimerization community appearing in some of the TM1–TM2–H8 simulations is formed by the cytoplasmic ends of TM7 and the H8 helices, containing not only the H8–H8 interface but also the conserved NPxxYx(5,6)F motif (Fig. 4a), present in both Class A and Class C GPCRs. Specifically, such communities are reported for the Rhodopsin, κ OR and μ OR dimers and, surprisingly, a similar community is observed in the case of mGluR1, despite the lack of complete coordinates for the H8 helices. Finally, although not directly involved with dimerization, community structures containing the ECL2 loop are observed for both protomers in all simulations, in which the ECL2 loops are either part of the TM1–TM2/TM1–TM2 community themselves or are

connected with it through critical nodes and edges (Supplementary Fig. S15).

Although not as extensive as the one reported for the TM1–TM2–H8 dimers, a similar conservation of network features is observed for TM4–TM5 dimers (Fig. 4b, see also Supplementary Fig. S16). In each case two inter-protomer communities are formed, one containing part of the dimer interface area between TM4 and TM5 and the other containing the interface between the ICL2 loops and the cytoplasmic ends of TM3; the latter contains the D(E)RY areas, which also participate in inter-protomer contacts (Fig. 4b). The only exception to these rules is the β 1AR TM4–TM5 dimer, in which no TM4–TM5 community is observed. This can be partially attributed to the overall interface of the β 1AR dimer, which is considerably loosely packed compared to other TM4–TM5 dimers, even after the simulations. Interestingly, this particular dimer contains the least amount of stacking/polar interactions among TM4–TM5 dimers. Similar to the TM1–TM2–H8 dimers, communities containing the ECL2 loop and, depending on its presence, residues from the N-terminus are formed, connected to inter-protomer communities through critical nodes and edges. In fact, in some instances, these communities also contain residues of TM4 and TM5,

participating in dimerization interactions. Such communities are observed not only for the Class A receptors but also for the Class F Smoothened and Class C mGluR1 TM4–TM5 dimers. Furthermore, they are also observed for the two dimers of CXCR4 that resulted during simulation, providing further support to the notion that CXCR4 adopts features of a TM4–TM5 dimer rather than TM5–TM6 (Supplementary Fig. S16).

In contrast to the above cases, the μ OR TM5–TM6 dimer stands out, in the sense that none of the previously mentioned features are observed. Communities are formed containing the TM1 and TM2 helices as well as the TM6 and TM7 helices, suggesting a co-dependent movement of these transmembrane segments. Small communities are also observed involving the D(E)RY motif and residues in TM5 and TM6, but without any inter-protomer contacts involved. Instead, almost all significant inter-protomer contacts are modeled as critical nodes connecting the otherwise independent communities of each protomer. Interestingly, part of the TM5–TM6 dimer interface forms a community with the ECL2 loop in each receptor protomer. While these results are interesting, the lack of other available crystallographic TM5–TM6 dimers hinders any attempt to detect any conserved features, since the CXCR4 dimers shifted to a TM4–TM5 orientation during simulation. However, it should be noted that the simulated TM5–TM6 dimer for squid Rhodopsin shows an almost identical community network organization (Supplementary Fig. S17).

The presence and similarity of these network elements in the simulated dimers suggest a conserved nature in GPCR structure and dynamic behavior. Furthermore, the observation of inter-protomer communities displays the dense interconnectivity of these regions, suggesting the considerable strength of inter-protomer contacts in these regions. A comparison of the dimers' network structure with results obtained through Alanine scanning and interface analysis reveals that the location of these communities overlaps with regions rich in interface core residues and Ala-scan hot spots (Figs. 3, 4). Finally, it should be noted that all stable stacking interactions, observed through interface analysis, appear to be part of these inter-protomer communities.

Conserved features in GPCR homodimer interfaces

Overall, the proposed Molecular Dynamics, interface analysis and community network results suggest the presence of conserved structural features in TM1–TM2–H8, TM4–TM5 and TM5–TM6 GPCR dimers, despite the often low sequence similarity between the receptors. It is important to note that the final orientation of all dimers appears to be relatively stable for more than half the simulation time in each system, suggesting favorable

thermodynamic properties for these configurations. Interestingly, the Class C mGluR1 and Class F Smoothened dimers display significant structural and dynamic similarities with Class A GPCR dimers. It should be noted that all Smoothened and mGluR1 simulations were both performed using only the transmembrane segments of the receptors, making it impossible to ascertain the possible influence of their large extracellular domains. These influences can be of great importance, especially in the case of metabotropic glutamate receptors, for which a significant amount of evidence shows that the extracellular Venus Fly-trap domains form part of the dimer interface [93]. However, the proposed results, combined with the significant structural similarity of the transmembrane α -helical bundle between these receptors and other GPCRs suggests that the observed oligomerization features may be common for all GPCR classes.

Structural and dynamic insights on the functional impact of GPCR dimers

A pivotal aspect in GPCR oligomerization research is studying the possible implications of oligomeric interactions upon receptor function. In numerous occasions, GPCR dimers and oligomers have been reported to influence or be influenced by ligand binding, regulate receptor activation and G-protein binding or initiate internalization and the β -arrestin signaling path [18]. However, despite the abundance of experimental evidence, little is known concerning the structural nature of these influences.

Regarding GPCR activation, a number of crystal structures for activated receptors in various stages have become available, including various Rhodopsin intermediates [76, 77] and the β 2AR-G-protein complex [94]. GPCRs in all of these structures display the same structural patterns for activation, which include breaking the ionic lock formed between TM6 and the D(E)RY motif in TM3, followed by movements of the TM5 and TM6 cytoplasmic segments and the ICL3 loop, with some minor rearrangements in the TM3 helix also observed. This motif is a common feature in all Class A receptors. GPCRs from other classes, including the Class C mGluR1 and Class F Smoothened receptors, lack this feature. However, a similar ionic lock has been observed in the recently solved mGluR1 and mGluR5 structures [15, 16] and a tight network of interactions between TM3 and TM6 exists in the Smoothened structure. A second motif, NPxxYx(5,6)F that in the cytoplasmic end of TM7 is also implicated in the process, seemingly stabilizing the active state by forming contacts with TM6. The significant similarity of these features in all activated GPCR structures, as well as the high conservation of the above mentioned motifs among

the members of class A suggests that the pattern of these movements is highly conserved in Rhodopsin—like GPCRs. Furthermore, the observation of a potential ionic lock between TM3 and TM6 in recently solved Class B and C structures, as well as the existence of variations for the NPxxYx(5,6)F motif in non Class A receptors [13–16] could be indications that other GPCRs may follow a similar activation process.

As expected, these structural rearrangements appear in the structure of the Opsin dimer. Furthermore, structural alignment of an activated receptor to any TM1–TM2–H8 dimer shows no potential clashes between the receptor's active state and the dimer interface (Supplementary Figure S18). Considering that no regions participating in activation are actually part of the TM1–TM2–H8 interface, the above observation is not surprising. In any case, it is clear that a TM1–TM2–H8 dimer can allow receptor activation and, consequently, canonical GPCR signal transduction. Similarly, Opsin and the adenosine bound active A2A adenosine receptor can be aligned to their respective TM4–TM5 dimers without introducing any bumps. It can be surmised, therefore, that the TM4–TM5 dimer may also allow GPCR activation, although participation of residues in the TM3 and TM5 helices to the dimer interface could affect the activation process. On the other hand, alignment of an active GPCR to the μ opioid TM5–TM6 interface displays serious clashes, due to the orientation of the TM5 and TM6 helices in the dimer. These clashes are observed both for the original μ OR structure and for the result of the CG-MD simulation. The above observations indicate the inability of the receptors in such a dimer towards proper activation (Supplementary Figure S18). Considering that there is experimental evidence implicating opioid receptor heterodimerization with inhibitory mechanisms, the hypothesis that a TM5–TM6 dimer could be a part of an inhibitory process could be viable.

An evaluation through structural alignment such as the one described provides hints towards each dimerization type's ability to shift to the active GPCR state, but offers little towards unveiling more detailed information regarding how the TM1–TM2–H8 and TM4–TM5 dimers may regulate the process. However, the study of simulation results through network analysis and community clustering shows correlations between elements of the dimer interface and regions implicated in activation that may be part of these regulatory mechanisms. Specifically, all the Class A GPCR TM4–TM5 dimer networks display the formation of inter—protomer communities including the region of the D(E)RY motif, the ICL2 loop and TM4 from each protomer, suggesting that these segments move in concert during simulation (Fig. 4b, see also Supplementary Figure S16). Surprisingly, a similar community is observed for the Smoothened and mGluR1 TM4–TM5 dimers even

though they lack the motif; however, the similar packing of the TM3 and TM6 helices in Smoothened and the existence of an ionic lock in mGluR1 would suggest that these receptors may be activated in a similar manner. A second correlation between elements of the activation process and dimerization is also observed for the μ OR, κ OR, Rhodopsin, β 2AR and mGluR1 TM1–TM2–H8 interfaces, which display the formation of communities involving the H8 helices and parts of the TM7 helix, including the NPxxYx(5,6)F motif, with the dense connectivity of these community elements hinting towards regulation of the motif by the H8–H8 interface (Fig. 4a). These correlations of motions between these regions and the dimer interfaces suggests that the dynamic behavior of these motifs and, possibly, their participation in the activation process by stabilizing the inactive or active state, could be subject to influences from the presence of a second receptor protomer.

The collective study of ligand binding mechanisms in GPCRs is challenging, both due to the presence of multiple binding sites on the same receptor and due to the significant diversity of GPCR ligands, which range from small molecules to large peptides and steroid hormones. Despite these limitations, a number of GPCR structures with various antagonists, synthetic or even native agonists have become available, including some receptors appearing in GPCR dimers. Furthermore, experimental evidence has implicated the contribution of the second extracellular loop (ECL2) to ligand recognition and selectivity for a number of different GPCRs [95].

Combining the information regarding the binding sites of these ligands with structural information, results from Molecular Dynamics and the dimers' network properties can propose possible mechanisms for the relationship between oligomerization and ligand recognition. Regarding the relations between GPCR oligomerization and the various ligand binding sites, it should be noted that many of the available dimer interfaces contain regions that also form part of ligand binding sites. The CXCR4 chemokine receptor is a characteristic example, with its TM4–TM5 interface overlapping with the It1t, CVX15 and vMIP-II binding pockets. Similarly, the adenosine site in A2A overlaps with its TM4–TM5 dimer, while the carvedilol and carazolol interacting residues are in close proximity to the dimer interfaces of β 1AR and β 2AR, respectively (Supplementary Figure S19). Finally, network analysis displays correlation between the dimer interfaces and the protomers' ECL2 loops, with the latter being either a direct part of an inter—protomer community or connected to one through critical nodes and edges (Fig. 4, see also Supplementary Figures S8–S10). It should be noted that a different network approach, combining structural features and co-evolution information, has suggested an evolutionary

correlation between ECL2 and a TM4–TM5 interface obtained through all-atom MD simulations for CXCR4 [84]. Our results, based solely on structural and dynamic GPCR features examined through Coarse-Grained Molecular Dynamics are in good agreement with these observations, since a similar correlation is observed. In fact, correlations between ECL2 and the dimer interface are reported in all examined dimerization interfaces, suggesting the possible connection between the interfaces and a GPCR region implicated in ligand recognition and selectivity, as well as hinting towards the existence of potential mechanisms through which TM1–TM2–H8, TM4–TM5 and TM5–TM6 dimers may regulate, or be regulated by ligand recognition.

Limitations of the model and methodology

It is important to state some of the limitations underlying the proposed models and the methodology used. With regards to the applied methodology, an important limitation may be the choice of using a Coarse-Grained force field to model biomolecular interactions. Our simulations with the MARTINI force field have shown good agreement with atomistic simulations for specific case studies and, more importantly, experimental evidence on the nature of the dimer interfaces. Still, the chosen model results in a loss of detail that limits the accurate representation of intermolecular contacts. A second limitation may be the poor sampling of membrane properties, particularly in the cases of simulations involving cholesterol. While the chosen simulation setups may describe the dynamic behavior of protein–protein interactions adequately, they fail to capture the membrane's slow dynamics, such as the phospholipid and cholesterol diffusion or measuring the stability of protein–cholesterol interactions. Adequate sampling for these phenomena would require significantly larger simulation times that are outside the scope of the present study. As such, any observations regarding these properties should be interpreted with care.

Regarding the methods used for intermolecular contacts analysis, it should be noted that both surface-based classification of the interface and computational alanine scanning have been developed with soluble proteins in mind and have not been extensively validated against transmembrane proteins. However, the good agreement between simulation results and interface analysis suggests that both methods can be used in transmembrane protein–protein interactions.

Finally, it should be noted that the present study focuses solely on symmetric dimer interfaces featuring TM1, TM4 and TM5, mainly due to their increased observation in crystal structures, biochemical evidence and Molecular Dynamics simulations. However, a few GPCR structures

[32, 81, 96] and a number of self-assembly Molecular Dynamics simulations have also proposed the potential formation of dimers involving different regions or even non-symmetric dimers, either as a result of specific environmental conditions such as membrane composition [38] or as a means through which more than two receptors may interact to form higher order oligomers [36, 39, 82]. While the strength of these alternative dimer interfaces has been challenged by free-energy calculations [36], the potential structural and physiological relevance or irrelevance of these assemblies warrants further investigation. Furthermore, the current study offers very little towards unveiling the nature of protein–protein interactions in GPCR hetero-oligomerization, although it should be noted that both the available evidence and the conserved structural nature of the GPCR transmembrane bundle would suggest that GPCR heteromers may display similar features as GPCR homomers. Despite these limitations, the results proposed by this study reveal important aspects of the structural nature in GPCR oligomerization.

Conclusions

In this study we explored aspects of the structural nature and dynamic behavior of GPCR dimer interactions. Molecular Dynamics simulations showed structural movements that can occur in GPCR dimers, suggesting possible rearrangements of the observed crystallographic interfaces. Additionally, the contribution of inter–protomer aromatic interactions and polar contacts near the membrane boundaries was explored, hinting at the existence of potential interface hot spots in GPCR oligomeric interactions as well as the potential role of the membrane, which may act as a structural determinant in driving the formation of dimer interactions in GPCRs. Finally, potential aspects of the influence dimer formation may have upon GPCR function were highlighted and possible regions of interest for the study of regulatory mechanisms were proposed. To our knowledge, this is one of the first studies to collectively examine the dynamic behavior of the available GPCR oligomerization structural evidence, as well as examine the structural nature of GPCR oligomerization for receptors outside of Class A. Given the rising interest in unveiling the implications of GPCR oligomerization and the significant structural conservation among GPCRs, the results of our study could be applicable in the design of experimental studies involving GPCR dimers and oligomers, as well as the study of transmembrane protein–protein interactions in general.

Acknowledgments We would like to thank the scientific and administrative staff of the “Bioinformatics” Master’s Program at the

Faculty of Biology of the University of Athens, for its generous support. M.C.T. was financially supported as a postdoctoral fellow by Greek State Scholarships Foundation, through the Siemens Program: “IKY Fellowships of Excellence for Postgraduate Studies in Greece – Siemens Program (2014–2015)”. This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility—ARIS under project ID “PR001025-M.D.S.B.M.S.”. Finally, we would like to sincerely thank the anonymous reviewers for their very valuable and constructive criticism, which helped us to considerably improve the manuscript, as well as the Editor-in-Chief for properly handling it. The authors declare no conflicts of interest.

References

- Rosenbaum DM, Rasmussen SG, Kobilka BK (2009) The structure and function of G-protein-coupled receptors. *Nature* 459(7245):356–363. doi:[10.1038/nature08144](https://doi.org/10.1038/nature08144)
- Cherezov V, Abola E, Stevens RC (2010) Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods Mol Biol* 654:141–168. doi:[10.1007/978-1-60761-762-4_8](https://doi.org/10.1007/978-1-60761-762-4_8)
- Oldham WM, Hamm HE (2008) Heterotrimeric G protein activation by G-protein-coupled receptors. *Nat Rev Mol Cell Biol* 9(1):60–71. doi:[10.1038/nrm2299](https://doi.org/10.1038/nrm2299)
- Kolakowski LF Jr (1994) GCRDB: a G-protein-coupled receptor database. *Recept Channels* 2(1):1–7
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289(5480):739–745
- Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi HJ, Kuhn P, Weis WI, Kobilka BK, Stevens RC (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318(5854):1258–1265. doi:[10.1126/science.1150577](https://doi.org/10.1126/science.1150577)
- Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454(7203):486–491. doi:[10.1038/nature07101](https://doi.org/10.1038/nature07101)
- Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao ZG, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A2A adenosine receptor. *Science* 332(6027):322–327. doi:[10.1126/science.1202793](https://doi.org/10.1126/science.1202793)
- Granier S, Manglik A, Kruse AC, Kobilka TS, Thian FS, Weis WI, Kobilka BK (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485(7398):400–404. doi:[10.1038/nature11111](https://doi.org/10.1038/nature11111)
- Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S (2012) Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* 485(7398):321–326. doi:[10.1038/nature10954](https://doi.org/10.1038/nature10954)
- Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC (2012) Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 485(7398):327–332. doi:[10.1038/nature10939](https://doi.org/10.1038/nature10939)
- Wang C, Wu H, Katritch V, Han GW, Huang XP, Liu W, Siu FY, Roth BL, Cherezov V, Stevens RC (2013) Structure of the human smoothed receptor bound to an antitumour agent. *Nature* 497(7449):338–343. doi:[10.1038/nature12167](https://doi.org/10.1038/nature12167)
- Hollenstein K, Kean J, Bortolato A, Cheng RK, Dore AS, Jazayeri A, Cooke RM, Weir M, Marshall FH (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* 499(7459):438–443. doi:[10.1038/nature12357](https://doi.org/10.1038/nature12357)
- Siu FY, He M, de Graaf C, Han GW, Yang D, Zhang Z, Zhou C, Xu Q, Wacker D, Joseph JS, Liu W, Lau J, Cherezov V, Katritch V, Wang MW, Stevens RC (2013) Structure of the human glucagon class B G-protein-coupled receptor. *Nature* 499(7459):444–449. doi:[10.1038/nature12393](https://doi.org/10.1038/nature12393)
- Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B, Weir M, Wiggan GR, Marshall FH (2014) Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* 511(7511):557–562. doi:[10.1038/nature13396](https://doi.org/10.1038/nature13396)
- Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* 344(6179):58–64. doi:[10.1126/science.1249489](https://doi.org/10.1126/science.1249489)
- Ferre S, Casado V, Devi LA, Filizola M, Jockers R, Lohse MJ, Milligan G, Pin JP, Guitart X (2014) G protein-coupled receptor oligomerization revisited: functional and pharmacological perspectives. *Pharmacol Rev* 66(2):413–434. doi:[10.1124/pr.113.008052](https://doi.org/10.1124/pr.113.008052)
- Franco R, Casado V, Cortes A, Ferrada C, Mallol J, Woods A, Lluís C, Canela EI, Ferre S (2007) Basic concepts in G-protein-coupled receptor homo- and heterodimerization. *Sci World J* 7:48–57. doi:[10.1100/tsw.2007.197](https://doi.org/10.1100/tsw.2007.197)
- Boroto-Escuela DO, Brito I, Romero-Fernandez W, Di Palma M, Oflijan J, Skietarska K, Duchou J, Van Craenenbroeck K, Suarez-Boomgaard D, Rivera A, Guidolin D, Agnati LF, Fuxe K (2014) The G protein-coupled receptor heterodimer network (GPCR-HetNet) and its hub components. *Int J Mol Sci* 15(5):8570–8590. doi:[10.3390/ijms15058570](https://doi.org/10.3390/ijms15058570)
- Kniazeff J, Prezeau L, Rondard P, Pin JP, Goudet C (2011) Dimers and beyond: the functional puzzles of class C GPCRs. *Pharmacol Ther* 130(1):9–25. doi:[10.1016/j.pharmthera.2011.01.006](https://doi.org/10.1016/j.pharmthera.2011.01.006)
- Dalrymple MB, Pflieger KD, Eidne KA (2008) G protein-coupled receptor dimers: functional consequences, disease states and drug targets. *Pharmacol Ther* 118(3):359–371. doi:[10.1016/j.pharmthera.2008.03.004](https://doi.org/10.1016/j.pharmthera.2008.03.004)
- Niswender CM, Conn PJ (2010) Metabotropic glutamate receptors: physiology, pharmacology, and disease. *Annu Rev Pharmacol Toxicol* 50:295–322. doi:[10.1146/annurev.pharmtox.011008.145533](https://doi.org/10.1146/annurev.pharmtox.011008.145533)
- Lee CW, Ho IK (2013) Pharmacological profiles of oligomerized mu-opioid receptors. *Cells* 2(4):689–714. doi:[10.3390/cells2040689](https://doi.org/10.3390/cells2040689)
- Moreno JL, Holloway T, Gonzalez-Maeso J (2013) G protein-coupled receptor heterocomplexes in neuropsychiatric disorders. *Prog Mol Biol Transl Sci* 117:187–205. doi:[10.1016/B978-0-12-386931-9.00008-8](https://doi.org/10.1016/B978-0-12-386931-9.00008-8)
- Wade SM, Dalman HM, Yang SZ, Neubig RR (1994) Multisite interactions of receptors and G proteins: enhanced potency of dimeric receptor peptides in modifying G protein function. *Mol Pharmacol* 45(6):1191–1197
- Guo W, Shi L, Javitch JA (2003) The fourth transmembrane segment forms the interface of the dopamine D2 receptor homodimer. *J Biol Chem* 278(7):4385–4388. doi:[10.1074/jbc.C200679200](https://doi.org/10.1074/jbc.C200679200)
- Wang HX, Konopka JB (2009) Identification of amino acids at two dimer interface regions of the alpha-factor receptor (Ste2). *Biochemistry* 48(30):7132–7139. doi:[10.1021/bi900424h](https://doi.org/10.1021/bi900424h)
- Johnston JM, Aburi M, Provati D, Bortolato A, Urizar E, Lambert NA, Javitch JA, Filizola M (2011) Making structural sense of dimerization interfaces of delta opioid receptor homodimers. *Biochemistry* 50(10):1682–1690. doi:[10.1021/bi101474v](https://doi.org/10.1021/bi101474v)

29. Kaczor AA, Selent J (2011) Oligomerization of G protein-coupled receptors: biochemical and biophysical methods. *Curr Med Chem* 18(30):4606–4634
30. Hu J, Thor D, Zhou Y, Liu T, Wang Y, McMillin SM, Mistry R, Challiss RA, Costanzi S, Wess J (2012) Structural aspects of M(3) muscarinic acetylcholine receptor dimer formation and activation. *FASEB J* 26(2):604–616. doi:[10.1096/fj.11-191510](https://doi.org/10.1096/fj.11-191510)
31. Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K (2003) Atomic-force microscopy: rhodopsin dimers in native disc membranes. *Nature* 421(6919):127–128. doi:[10.1038/421127a](https://doi.org/10.1038/421127a)
32. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem* 278(24):21655–21662. doi:[10.1074/jbc.M302536200](https://doi.org/10.1074/jbc.M302536200)
33. Ruprecht JJ, Mielke T, Vogel R, Villa C, Schertler GF (2004) Electron crystallography reveals the structure of metarhodopsin I. *EMBO J* 23(18):3609–3620. doi:[10.1038/sj.emboj.7600374](https://doi.org/10.1038/sj.emboj.7600374)
34. Periole X, Huber T, Marrink SJ, Sakmar TP (2007) G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J Am Chem Soc* 129(33):10126–10132. doi:[10.1021/ja0706246](https://doi.org/10.1021/ja0706246)
35. Simpson LM, Taddese B, Wall ID, Reynolds CA (2010) Bioinformatics and molecular modelling approaches to GPCR oligomerization. *Curr Opin Pharmacol* 10(1):30–37. doi:[10.1016/j.coph.2009.11.001](https://doi.org/10.1016/j.coph.2009.11.001)
36. Periole X, Knepp AM, Sakmar TP, Marrink SJ, Huber T (2012) Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *J Am Chem Soc* 134(26):10959–10965. doi:[10.1021/ja303286e](https://doi.org/10.1021/ja303286e)
37. Rodriguez D, Gutierrez-de-Teran H (2012) Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor. *Proteins* 80(8):1919–1928. doi:[10.1002/prot.24099](https://doi.org/10.1002/prot.24099)
38. Prasanna X, Chattopadhyay A, Sengupta D (2014) Cholesterol modulates the dimer interface of the beta(2)-adrenergic receptor via cholesterol occupancy sites. *Biophys J* 106(6):1290–1300. doi:[10.1016/j.bpj.2014.02.002](https://doi.org/10.1016/j.bpj.2014.02.002)
39. Provasi D, Boz MB, Johnston JM, Filizola M (2015) Preferred supramolecular organization and dimer interfaces of opioid receptors from simulated self-association. *PLoS Comput Biol* 11(3):e1004148. doi:[10.1371/journal.pcbi.1004148](https://doi.org/10.1371/journal.pcbi.1004148)
40. Johnston JM, Wang H, Provasi D, Filizola M (2012) Assessing the relative stability of dimer interfaces in g protein-coupled receptors. *PLoS Comput Biol* 8(8):e1002649. doi:[10.1371/journal.pcbi.1002649](https://doi.org/10.1371/journal.pcbi.1002649)
41. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
42. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372(3):774–797. doi:[10.1016/j.jmb.2007.05.022](https://doi.org/10.1016/j.jmb.2007.05.022)
43. Xue L, Rovira X, Scholler P, Zhao H, Liu J, Pin JP, Rondard P (2015) Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nat Chem Biol* 11(2):134–140. doi:[10.1038/nchembio.1711](https://doi.org/10.1038/nchembio.1711)
44. Schrödinger, LLC (2010) The PyMOL molecular graphics system, Version 1.7
45. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14(1):33–38
46. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-y, Pieper U, Sali A (2001) Comparative protein structure modeling using MODELLER. In: Coligan JE (ed) *Current protocols in protein science*, Wiley, Hoboken. doi:[10.1002/0471140864.ps0209s50](https://doi.org/10.1002/0471140864.ps0209s50)
47. Consortium TU (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43(D1):D204–D212. doi:[10.1093/nar/gku989](https://doi.org/10.1093/nar/gku989)
48. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21):2947–2948. doi:[10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404)
49. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 111(27):7812–7824. doi:[10.1021/jp071097f](https://doi.org/10.1021/jp071097f)
50. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J (2008) The MARTINI coarse-grained force field: extension to proteins. *J Chem Theory Comput* 4(5):819–834. doi:[10.1021/ct700324x](https://doi.org/10.1021/ct700324x)
51. Ghosh A, Sonavane U, Joshi R (2014) Multiscale modelling to understand the self-assembly mechanism of human beta2-adrenergic receptor in lipid bilayer. *Comput Biol Chem* 48:29–39. doi:[10.1016/j.compbiolchem.2013.11.002](https://doi.org/10.1016/j.compbiolchem.2013.11.002)
52. Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ (2015) Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *J Chem Theory Comput* 11(5):2144–2155. doi:[10.1021/acs.jctc.5b00209](https://doi.org/10.1021/acs.jctc.5b00209)
53. Eargle J, Luthey-Schulten Z (2012) NetworkView: 3D display and analysis of protein-RNA interaction networks. *Bioinformatics* 28(22):3000–3001. doi:[10.1093/bioinformatics/bts546](https://doi.org/10.1093/bioinformatics/bts546)
54. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637. doi:[10.1002/bip.360221211](https://doi.org/10.1002/bip.360221211)
55. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29(7):845–854. doi:[10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055)
56. Shih AY, Freddolino PL, Arkhipov A, Schulten K (2007) Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *J Struct Biol* 157(3):579–592. doi:[10.1016/j.jsb.2006.08.006](https://doi.org/10.1016/j.jsb.2006.08.006)
57. Shih AY, Freddolino PL, Sligar SG, Schulten K (2007) Disassembly of nanodiscs with cholate. *Nano Lett* 7(6):1692–1696. doi:[10.1021/nl0706906](https://doi.org/10.1021/nl0706906)
58. Wassenaar TA, Pluhackova K, Bockmann RA, Marrink SJ, Tieleman DP (2014) Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J Chem Theory Comput* 10(2):676–690. doi:[10.1021/ct400617g](https://doi.org/10.1021/ct400617g)
59. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614. doi:[10.1002/jcc.21287](https://doi.org/10.1002/jcc.21287)
60. Lim JB, Rogaski B, Klauda JB (2012) Update of the cholesterol force field parameters in CHARMM. *J Phys Chem B* 116(1):203–210. doi:[10.1021/jp207925m](https://doi.org/10.1021/jp207925m)
61. Huang J, MacKerell AD Jr (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* 34(25):2135–2145. doi:[10.1002/jcc.23354](https://doi.org/10.1002/jcc.23354)
62. Schlitter J, Engels M, Kruger P (1994) Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* 12(2):84–89

63. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802. doi:[10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289)
64. Castillo N, Monticelli L, Barnoud J, Tieleman DP (2013) Free energy of WALP23 dimer association in DMPC, DPPC, and DOPC bilayers. *Chem Phys Lipids* 169:95–105. doi:[10.1016/j.chemphyslip.2013.02.001](https://doi.org/10.1016/j.chemphyslip.2013.02.001)
65. Glykos NM (2006) Software news and updates. Carma: a molecular dynamics analysis program. *J Comput Chem* 27(14):1765–1768. doi:[10.1002/jcc.20482](https://doi.org/10.1002/jcc.20482)
66. Hunter JD (2007) Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9(3):90–95
67. Eisenhaber F, Lijnzaad P, Argos P, Sander C, Scharf M (1995) The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J Comput Chem* 16(3):273–284. doi:[10.1002/jcc.540160303](https://doi.org/10.1002/jcc.540160303)
68. Levy ED (2010) A simple definition of structural regions in proteins and its use in analyzing interface evolution. *J Mol Biol* 403(4):660–670. doi:[10.1016/j.jmb.2010.09.028](https://doi.org/10.1016/j.jmb.2010.09.028)
69. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. *J Mol Biol* 105(1):1–12
70. Miller S, Janin J, Lesk AM, Chothia C (1987) Interior and surface of monomeric proteins. *J Mol Biol* 196(3):641–656
71. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320(2):369–387. doi:[10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4)
72. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33(Web Server issue):W382–W388. doi:[10.1093/nar/gki387](https://doi.org/10.1093/nar/gki387)
73. Anand P, Nagarajan D, Mukherjee S, Chandra N (2014) ABS-Scan: in silico alanine scanning mutagenesis for binding site residues in protein-ligand complex. *F1000Research* 3:214. doi:[10.12688/f1000research.5165.2](https://doi.org/10.12688/f1000research.5165.2)
74. Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci USA* 106(16):6620–6625. doi:[10.1073/pnas.0810961106](https://doi.org/10.1073/pnas.0810961106)
75. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826. doi:[10.1073/pnas.122653799](https://doi.org/10.1073/pnas.122653799)
76. Salom D, Lodowski DT, Stenkamp RE, Le Trong I, Golczak M, Jastrzebska B, Harris T, Ballesteros JA, Palczewski K (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc Natl Acad Sci USA* 103(44):16123–16128. doi:[10.1073/pnas.0608022103](https://doi.org/10.1073/pnas.0608022103)
77. Park JH, Scheerer P, Hofmann KP, Choe HW, Ernst OP (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454(7201):183–187. doi:[10.1038/nature07063](https://doi.org/10.1038/nature07063)
78. Huang J, Chen S, Zhang JJ, Huang XY (2013) Crystal structure of oligomeric beta1-adrenergic G protein-coupled receptors in ligand-free basal state. *Nat Struct Mol Biol* 20(4):419–425. doi:[10.1038/nsmb.2504](https://doi.org/10.1038/nsmb.2504)
79. Murakami M, Kouyama T (2008) Crystal structure of squid rhodopsin. *Nature* 453(7193):363–367. doi:[10.1038/nature06925](https://doi.org/10.1038/nature06925)
80. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, IJzerman AP, Cherezov V, Stevens RC (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* 337(6091):232–236. doi:[10.1126/science.1219218](https://doi.org/10.1126/science.1219218)
81. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330(6007):1066–1071. doi:[10.1126/science.1194396](https://doi.org/10.1126/science.1194396)
82. Mondal S, Johnston JM, Wang H, Khelashvili G, Filizola M, Weinstein H (2013) Membrane driven spatial organization of GPCRs. *Sci Rep* 3:2909. doi:[10.1038/srep02909](https://doi.org/10.1038/srep02909)
83. Johnston JM, Filizola M (2014) Differential stability of the crystallographic interfaces of mu- and kappa-opioid receptors. *PLoS ONE* 9(2):e90694. doi:[10.1371/journal.pone.0090694](https://doi.org/10.1371/journal.pone.0090694)
84. Nichols SE, Hernandez CX, Wang Y, McCammon JA (2013) Structure-based network analysis of an evolved G protein-coupled receptor homodimer interface. *Protein Sci Publ Protein Soc* 22(6):745–754. doi:[10.1002/pro.2258](https://doi.org/10.1002/pro.2258)
85. Wang J, He L, Combs CA, Roderiquez G, Norcross MA (2006) Dimerization of CXCR4 in living malignant cells: control of cell migration by a synthetic peptide that reduces homologous CXCR4 interactions. *Mol Cancer Ther* 5(10):2474–2483. doi:[10.1158/1535-7163.MCT-05-0261](https://doi.org/10.1158/1535-7163.MCT-05-0261)
86. Cang X, Du Y, Mao Y, Wang Y, Yang H, Jiang H (2013) Mapping the functional binding sites of cholesterol in beta2-adrenergic receptor by long-time molecular dynamics simulations. *J Phys Chem B* 117(4):1085–1094. doi:[10.1021/jp3118192](https://doi.org/10.1021/jp3118192)
87. Zheng H, Pearsall EA, Hurst DP, Zhang Y, Chu J, Zhou Y, Reggio PH, Loh HH, Law PY (2012) Palmitoylation and membrane cholesterol stabilize mu-opioid receptor homodimerization and G protein coupling. *BMC cell biology* 13:6. doi:[10.1186/1471-2121-13-6](https://doi.org/10.1186/1471-2121-13-6)
88. Filizola M, Olmea O, Weinstein H (2002) Prediction of heterodimerization interfaces of G-protein coupled receptors with a new subtractive correlated mutation method. *Protein Eng* 15(11):881–885
89. Kaczor AA, Guixà-González R, Carrió P, Poso A, Dove S, Pastor M, Selent J (2015) Multi-component protein–protein docking based protocol with external scoring for modeling dimers of G protein-coupled receptors. *Mol Inf* 34(4):246–255. doi:[10.1002/minf.201400088](https://doi.org/10.1002/minf.201400088)
90. Kota P, Reeves PJ, Rajbhandary UL, Khorana HG (2006) Opsin is present as dimers in COS1 cells: identification of amino acids at the dimeric interface. *Proc Natl Acad Sci USA* 103(9):3054–3059. doi:[10.1073/pnas.0510982103](https://doi.org/10.1073/pnas.0510982103)
91. Jastrzebska B, Chen Y, Orban T, Jin H, Hofmann L, Palczewski K (2015) Disruption of rhodopsin dimerization with synthetic peptides targeting an interaction interface. *J Biol Chem* 290(42):25728–25744. doi:[10.1074/jbc.M115.662684](https://doi.org/10.1074/jbc.M115.662684)
92. Mondal S, Khelashvili G, Weinstein H (2014) Not just an oil slick: how the energetics of protein-membrane interactions impacts the function and organization of transmembrane proteins. *Biophys J* 106(11):2305–2316. doi:[10.1016/j.bpj.2014.04.032](https://doi.org/10.1016/j.bpj.2014.04.032)
93. Kunishima N, Shimada Y, Tsuji Y, Sato T, Yamamoto M, Kumasaka T, Nakanishi S, Jingami H, Morikawa K (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* 407(6807):971–977. doi:[10.1038/35039564](https://doi.org/10.1038/35039564)
94. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477(7366):549–555. doi:[10.1038/nature10361](https://doi.org/10.1038/nature10361)
95. Seibt BF, Schiedel AC, Thimm D, Hinz S, Sherbiny FF, Muller CE (2013) The second extracellular loop of GPCRs determines subtype-selectivity and controls efficacy as evidenced by loop exchange study at A2 adenosine receptors. *Biochem Pharmacol* 85(9):1317–1329. doi:[10.1016/j.bcp.2013.03.005](https://doi.org/10.1016/j.bcp.2013.03.005)
96. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341(6152):1387–1390. doi:[10.1126/science.1241475](https://doi.org/10.1126/science.1241475)



Structural characterization and molecular dynamics simulations of the caprine and bovine solute carrier family 11 A1 (SLC11A1)

Kostas A. Triantaphyllopoulos¹ · Fotis A. Baltoumas² · Stavros J. Hamodrakas²

Received: 12 July 2018 / Accepted: 3 December 2018 / Published online: 12 December 2018
© Springer Nature Switzerland AG 2018

Abstract

Natural Resistance-Associated Macrophage Proteins are a family of transmembrane divalent metal ion transporters, with important implications in life of both bacteria and mammals. Among them, the Solute Carrier family 11 member A1 (SLC11A1) has been implicated with susceptibility to infection by *Mycobacterium avium* subspecies *paratuberculosis* (MAP), potentially causing Crohn's disease in humans and paratuberculosis (PTB) in ruminants. Our previous research had focused on sequencing the mRNA of the caprine *slc11a1* gene and pinpointed polymorphisms that contribute to caprine SLC11A1's susceptibility to infection by MAP in PTB. Despite its importance, little is known on the structural/dynamic features of mammalian SLC11A1 that may influence its function under normal or pathological conditions at the protein level. In this work we studied the structural architecture of SLC11A1 in *Capra hircus* and *Bos taurus* through molecular modeling, molecular dynamics simulations in different, functionally relevant configurations, free energy calculations of protein-metal interactions and sequence conservation analysis. The results of this study propose a three dimensional structure for SLC11A1 with conserved sequence and structural features and provide hints for a potential mechanism through which divalent metal ion transport is conducted. Given the importance of SLC11A1 in susceptibility to PTB, this study provides a framework for further studies on the structure and dynamics of SLC11A1 in other organisms, to gain 3D structural insight into the macromolecular arrangements of SLC11A1 but also suggesting a potential mechanism which divalent metal ion transport is conducted.

Keywords SLC11A1 · NRAMP · Paratuberculosis · Ion transport · Molecular dynamics · Energy calculations

Abbreviations

Slc11a1 Solute carrier family 11 member 1 gene
SLC11A1 Solute carrier family 11 member A1
NRAMPs Natural resistance-associated macrophage proteins

DMT1 Divalent metal ion transporter 1
MAP *Mycobacterium avium* subspecies *paratuberculosis*
ScaDMT *Staphylococcus capitis* divalent metal ion transporter
EcoDMT *Eremococcus coleocola* divalent metal ion transporter
DraNramp *Deinococcus radiodurans* NRAMP homolog
TM Transmembrane
MD Molecular dynamics
FEP Free energy perturbation
EDA Essential dynamics analysis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10822-018-0179-x>) contains supplementary material, which is available to authorized users.

✉ Kostas A. Triantaphyllopoulos
ktrianta@aia.gr

¹ Department of Animal Breeding and Husbandry, Faculty of Animal Science and Aquaculture, School of Agricultural Production, Infrastructure and Environment, Agricultural University of Athens, 75 Iera Odos St., 11855 Athens, Greece

² Section of Cell Biology and Biophysics, Department of Biology, School of Sciences, National and Kapodistrian University of Athens, Panepistimiopolis, 15701 Athens, Greece

Introduction

There is growing evidence that limiting the spread of Johne's disease (JD) or paratuberculosis (PTB), caused by *Mycobacterium avium* subsp. *paratuberculosis* (MAP), has a strong potential through selective breeding of animals based on

their genetic susceptibility to the disease [1]. Biomarkers have been proposed for establishing the diagnosis, prognosis, screening and monitoring of slow progressive infectious diseases such as, the infection caused by MAP that could potentially be the cause of zoonotic diseases [2–4]. Disease targets that have been scrutinized for susceptibility in humans to *Mycobacterium* infection have so far involved a number of biomarkers, including the solute carrier family 11 member A1 (SLC11A1), formerly known as Natural Resistance-Associated Macrophage Protein 1 (NRAMP1) [5–9]. Likewise, veterinary and livestock research on ruminant SLC11A1 has shown genetic correlation between single nucleotide polymorphisms (SNPs) in the bovine and caprine *slc11a1* gene, with susceptibility to infection by MAP [7, 10–13].

The product of the *slc11a1* gene belongs to Natural Resistance-Associated Macrophage Protein (NRAMP) family of transporters, also known as Solute Carrier Family 11 (SLC11) [14]. SLC11A1 is a multi-spanning, α -helical transmembrane protein that primarily functions as a transporter of divalent cations such as Fe^{2+} , Mg^{2+} , Mn^{2+} and other divalent metals across membranes. It is primarily expressed in the phagosomes of macrophages, where it has an essential role in the host defense mechanism against pathogens; however, the protein has also been found to be expressed in the lung, liver and spleen [14–17]. Like most NRAMPs, SLC11A1 transports unbound Fe^{2+} and other divalent ions into the cytoplasm; this transport seems to be coupled to the concurrent transport of protons alongside divalent ions. A similar function is also served by another NRAMP, Divalent Metal Transporter 1 (DMT1), also known as SLC11A2 or NRAMP2 [14, 18]. However, while DMT1 is known to act as a symporter, co-catalyzing the transport of metal ions alongside the proton gradient, the exact functionality of SLC11A1 is not yet clear. Instead, while some studies suggest that SLC11A1 is a symporter like DMT1, others have proposed that it is, in fact, an antiporter, transporting iron against the proton gradient [19].

Eukaryotic NRAMPs in general and SLC11A1 and DMT1 in particular are required for Fe^{2+} and Mn^{2+} homeostasis by metal trafficking and intercellular signaling. Several pathologies may result from defects in SLC11A1-dependent Fe^{2+} or Mn^{2+} transport and most importantly play a role in host innate immunity. Thus, mutations that impair SLC11A1 or DMT1 for proper function or deficiency in mice lacking SLC11A1 or DMT1 affect iron recycling and lead to iron accumulation within the liver and spleen, contributing to erythrophagocytosis and hemolytic anemia [20, 21]. Another important aspect of SLC11A1 is its relationship with the organism's susceptibility to various pathogens, among them *Mycobacterium avium* subsp. *paratuberculosis* (MAP). *Slc11a1* gene orthologs from various species including *Capra hircus* (goat), *Bos taurus* (bovine), *Mus*

musculus (mouse), *Homo sapiens* (human) and others, have been implicated with susceptibility to MAP, causing PTB in ruminants and Crohn's disease, leprosy, Buruli ulcers etc in humans and mice [22–24]. At the same time, SLC11A1 may play a pivotal role in innate immunity and has been implicated in autoimmune diseases such as Type I diabetes in humans [6]. Thus, broadly speaking, there have been several reports in the literature for the genetic susceptibility of the host to MAP infection and the relationships between polymorphisms in *Slc11a1* and diseases such as, tuberculosis, Crohn's disease and others [5, 7–9], where in all cases the role of SLC11A1 is most central. In the survival “battle” between the parasite and the host cell, the latter competes against the pathogen for divalent ions by starving the invader of an essential element, in this case, iron, a phenomenon often referred to as “nutritional immunity”. Hence, knowledge of protein-metal interactions in the SLC11A1 ion-binding site can enhance molecular design strategies to intervene and, potentially, inhibit the survival of *Mycobacterium* and other similar pathogens inside the host [25].

Despite its importance in mammalian physiology, information regarding the structure of SLC11A1 is sparse. Current information on its topology comes primarily from sequence analysis and transmembrane topology predictions through various computational methods. Members of the NRAMP family, including SLC11A1, were initially predicted to resemble the fold of bacterial LeuT, a group of amino acid transporters with 12 transmembrane segments that are also capable of ion transport [18]. One key differentiation of NRAMPs from LeuT, however, involves the total number of transmembrane segments that could discern sequences among species. In general, most mammalian NRAMPs feature 12 transmembrane α -helices with their N- and C-termini in the cytoplasmic side, while the NRAMP bacterial homologs can have either 11 or 12 segments, with their N-terminus in the cytoplasmic side and their C-terminus in either the extracellular or cytoplasmic side, depending on the number of their transmembrane segments [18]. More recently, a number of crystal structures of bacterial SLC11 homologs were solved, namely, the *Staphylococcus capitis* divalent metal transporter (ScaDMT), the *Deinococcus radiodurans* NRAMP homolog (DraNramp) and the *Eremococcus coleocola* manganese transporter (EcoDMT) [26–28]. All structures displayed a common architecture for the transporters' NRAMP domain, confirming its similarity with the LeuT fold. Interestingly, the ScaDMT and DraNramp structures revealed a transmembrane topology featuring 11 transmembrane segments, shared by many bacterial NRAMPs, while the EcoDMT structure presented an organization of 12 transmembrane segments, resembling mammalian NRAMPs. Most importantly, these structures revealed two distinct NRAMP configurations, namely, an outward-facing state [28] that was found in ScaDMT and DraNramp

structures open to the extracellular space and capable of receiving divalent ions, and an inward-facing state found in EcoDMT structure [26, 27] open to the cytoplasm, to release the substrate into the cell interior.

So far, no structural information is available for mammalian NRAMPs, which could enable studying the structural features that regulate their function and may influence their susceptibility to pathogens. More importantly, the available SLC11A1 structural, functional and clinical annotation records in protein databases such as UniProt [29] are only limited to human, mouse and rat sequences, while equivalent records in ruminants and other animals are either absent or computationally assigned by their similarity to the more studied species. Considering that genetic aberrations and/or mutations in functionally significant regions accompanied by higher structural arrangements may lead to molecular structure perturbations that may affect infection susceptibility to pathogens [30], this lack of annotation hinders research on SLC11A1 and its connection to MAP infection for ruminants such as *Bos taurus*, *Capra hircus* or other potentially affected mammals.

Given the physiological and clinical significance of MAP infection with the associated *Slc11a1*, and its implication in disease susceptibility in both humans and animals, our goal in the current study was the structural and dynamic characterization of SLC11A1 in two ruminant species, namely, *Capra hircus* and *Bos taurus*, in which the transporter has been implicated to the host's infection with MAP [11, 12, 25, 31]. Previous research in this area has primarily focused on sequencing the *slc11a1* gene and its mRNA in different species, and studying the connections of various polymorphisms with MAP infection and paratuberculosis [7, 10, 12, 22]. In this work, by performing molecular modeling and molecular dynamics simulations, coupled by sequence conservation analysis, we arrive at a potential structure for the caprine and bovine SLC11A1 transmembrane domain. At the same time, we attempt to model a number of different, functionally relevant structural states of SLC11A1 and, by using molecular dynamics simulations coupled with essential dynamics analysis and free energy perturbation calculations, investigate and compare their features. Our results propose a potential structural framework for SLC11A1, possibly including other mammalian NRAMPs, and provide hints for the mechanism through which SLC11A1 function is regulated.

Methods

SLC11A1 sequence annotation and analysis

The SLC11A1 amino acid sequences of *Capra hircus* (UniProt: B8XJG7), *Bos taurus* (UniProt: Q27981), *Homo*

sapiens (UniProt: P49279), *Mus musculus* (UniProt: P41251), *Ovis aries* (UniProt: P49280) and the sequences of four bacterial SLC11 homologs, namely, ScaDMT from *Staphylococcus capitis* (UniProt: A0A0S4MEX1), DraN-ramp from *Deinococcus radiodurans* (UniProt: Q9RTP8), EcoDMT from *Eremococcus coleocola* (UniProt: E4KPW4) and BsuMntH from *Bacillus subtilis* (UniProt: P96593) were retrieved from the UniProtKB database [29]. The potential presence of signal peptides was checked using SignalP v. 4.1 [32] and none of the sequences were found to have a signal peptide, therefore, no sequence segments were excluded from further analysis. The basic physicochemical properties of all sequences, including their molecular weight, isoelectric point, instability index, aliphatic index and the GRand Δ Verage of hYdropathy (GRAVY) index, were estimated using ProtParam [33]. The transmembrane topology of SLC11A1 was predicted by considering the results of five distinct prediction algorithms, namely, TMHMM2.0 [34], Phobius [35], OCTOPUS [36], MEMSAT3 [37] and HMMpTM [38]. The latter, which combines the prediction of both transmembrane topology and eukaryotic post-translational modifications (PTMs) using Hidden Markov Models [38], was also used to predict potential phosphorylation and glycosylation sites in mammalian SLC11A1 alongside the NetNglyc server, which also predicts N-Glycosylation sites using artificial neural networks [39, 40]. The predicted PTMs were further validated and enhanced by the ScanProsite tool detecting all the signature matches, as previously reported [41]. Secondary structure predictions for the caprine and bovine SLC11A1 amino acid sequences were performed using PSIPRED [42]. The position of the NRAMP domain in each sequence (Pfam: PF01566) was identified by scanning the latter against the NRAMP Hidden Markov Model profile (pHMM) available in Pfam [43], using the *hmmsearch* function in HMMER v. 3.1 [44]. Snake plot representations, displaying significant annotation features for *C. hircus* and *B. taurus* SLC11A1 and their bacterial homologs were prepared using PROTTTER [45] (Fig. 1A, Supplementary Fig. S1 & S2). A multiple sequence alignment (MSA) of the sequences was prepared using PSI/TM-Coffee, a variation of T-Coffee designed for the alignment of α -helical transmembrane proteins through homology extension [46]. Further alignment editing and visualization was performed with Clustal Omega [47], MEGA v. 4.0 [48] and JalView 2 [49] (Fig. 2). Phylogenetic tree analysis was performed with Phylogeny.fr/MrBayes [50, 51], using the implemented Bayesian approach. The caprine, ovine, human, murine and bovine SLC11A1 amino acid sequences were used at the out-group. All alignment gaps were included in the analysis and the Monte Carlo parameters algorithm was applied to the aligned sequences. A summarized annotation of the analyzed sequences is presented in Supplementary Figure S5.

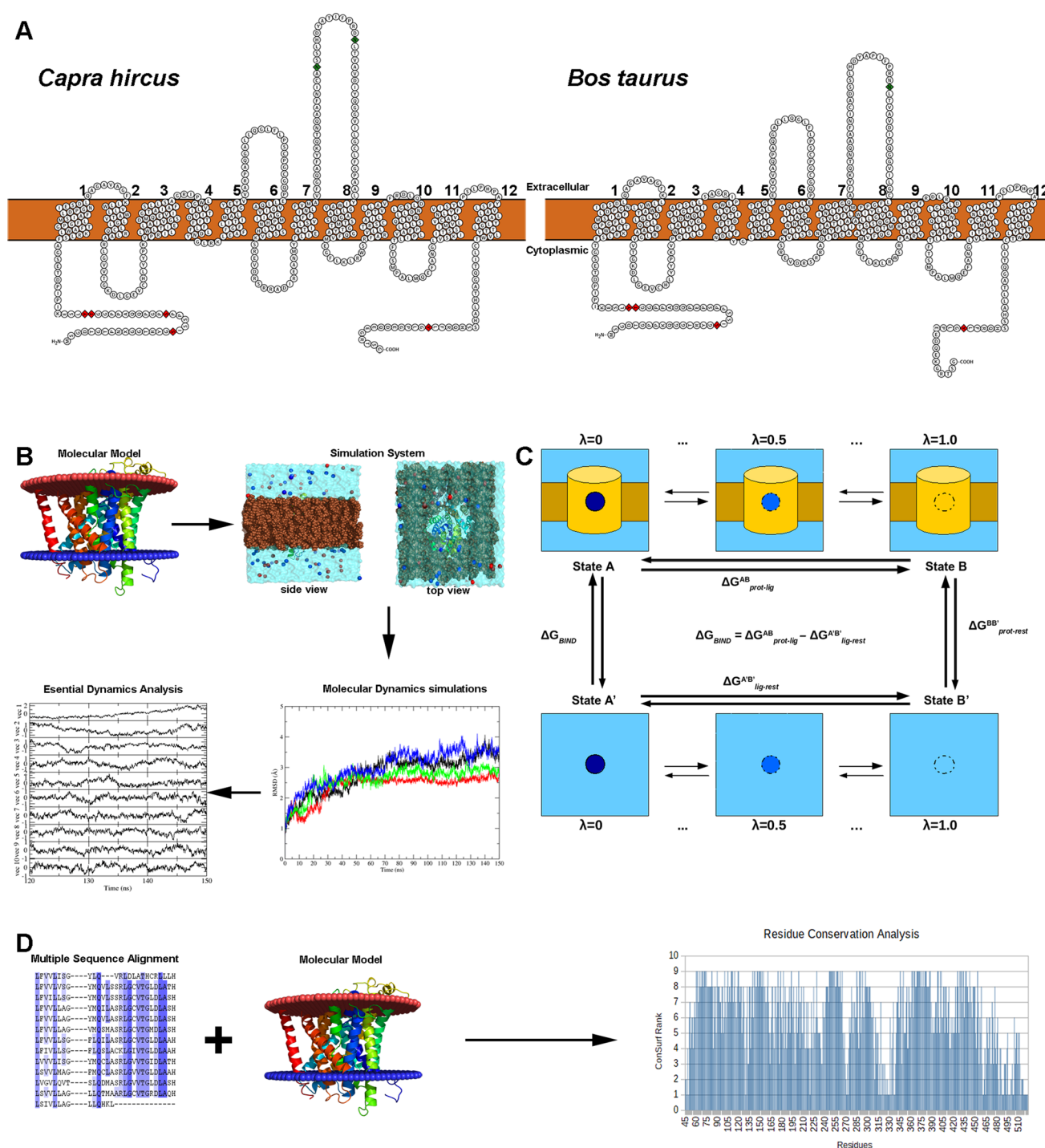


Fig. 1 Overview of the computational methods used in this study. **a** Sequence annotation of SLC11A1 in *Capra hircus* (left) and *Bos taurus* (right). Both sequences are presented as snake plots showing the topology of the transporters, created using Protter. Potential phosphorylation and glycosylation sites are shown as rhombi and colored red and green, respectively. **b** Workflow of the procedure followed in molecular dynamics (MD) simulations and associated analysis. SLC11A1 is inserted in a box composed by POPC lipids and aqueous solvent and simulated. MD simulation results are then used in essential dynamics analysis (EDA) to detect collective motions. **c** Graphic representation of the thermodynamic cycle used in Free

Energy Perturbation calculations. Two simulation systems are used, a system comprising of the ion-bound transporter in membrane/solvent (prot-lig) and a system comprising of a free ion in a solvent box (lig-rest). In both cases, non-bonded interactions are gradually turned off using a pathway described by a lambda coordinate. The final binding energy value is then taken to be the difference between the free energy values for the prot-lig and lig-rest systems. **d** Overview of the sequence conservation analysis performed using ConSurf. A multiple sequence analysis is combined with a three-dimensional structure to calculate, through Bayesian analysis, the per-residue conservation rank of each amino acid position

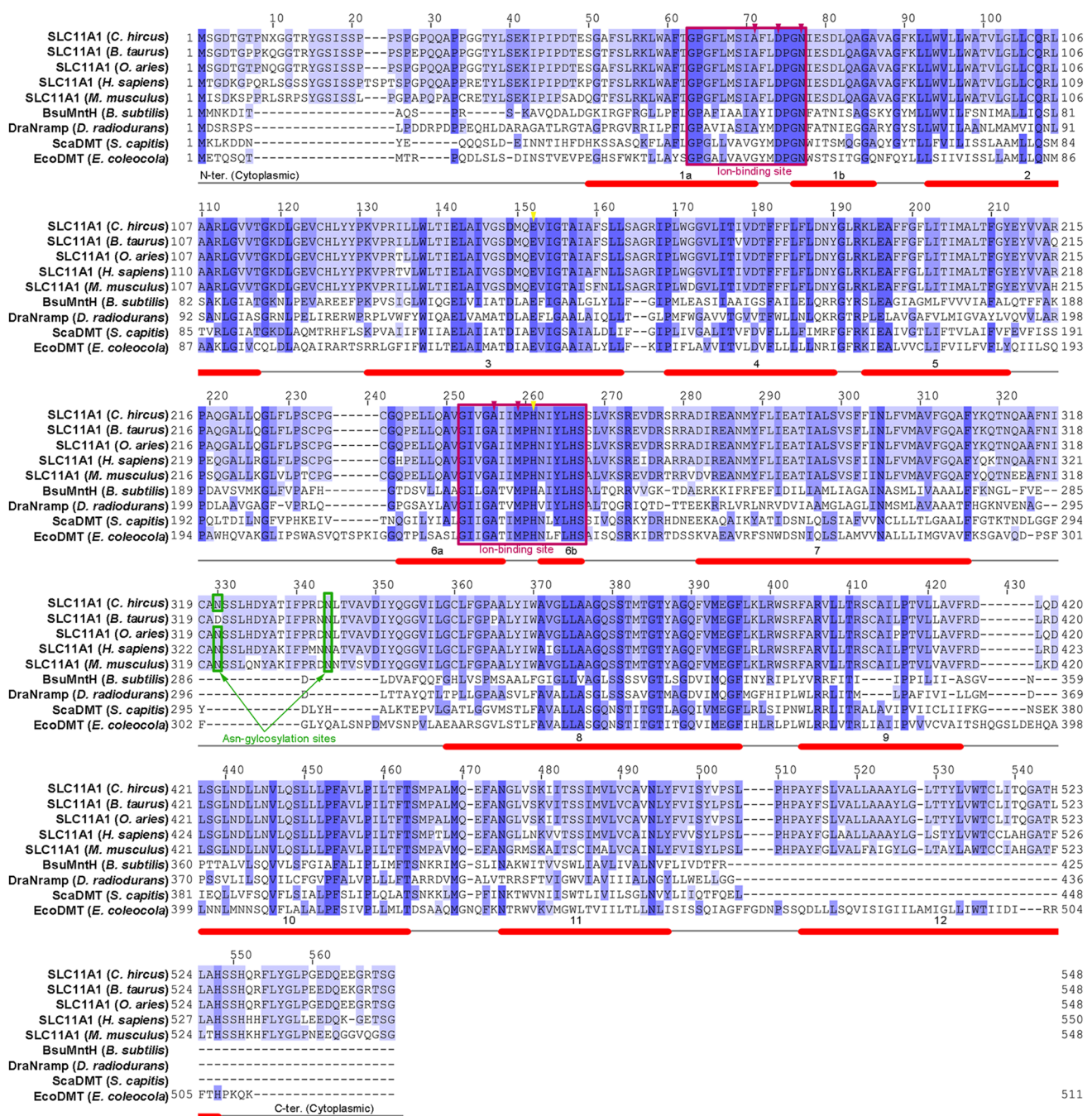


Fig. 2 Comparison of SLC11A1 and its homologs across different species. Multiple sequence alignment of SLC11A1 sequences from *C. hircus*, *B. taurus*, *O. aries*, *H. sapiens* and *M. musculus* with bacterial homologs ScaDMT, DraNrap, EcoDMT and BsuMnH. Alignment positions are indicated by the numerical scale on top of each alignment row. The transmembrane topology of SLC11A1, as shown in the models, was generated through homology modeling below each alignment row using red cylinders to represent α -helices. Residues

are colored using shades of blue to reflect their percentage identity in the alignment, light blue, blue and deep blue correspond to 40%, 60% and 80% sequence identity, respectively. The 1a–1b and 6a–6b regions, comprising the ion-binding site, are labeled using magenta frames. Residues potentially participating in protein-metal interactions or in proton co-transport are labeled using magenta and yellow arrowheads, respectively. Potential Asn-glycosylation sites in mammalian SLC11A1 are labeled using green frames

Molecular modeling and structural analysis

The currently available representative structures of the NRAMP architecture in the Protein Data Bank (PDB) [52]

include the structures of three prokaryotic SLC11 homologs, namely, the *S. capitis* ScaDMT (PDB: 4WGV, 5M94, 5M95, UniProt: A0A0S4MEX1), the *D. radiodurans* DraNrap (PDB: 5KTE, UniProt: Q9RTP8) and the *E. coleocola*

EcoDMT (PDB: 5M87, UniProt: E4KPW4) [26–28]. The sequences and structures of all three homologs were retrieved from UniProt and PDB, respectively, and tested against caprine and bovine SLC11A1 through pairwise and multiple sequence alignments. The ScaDMT and DraNramp homologs represent the inward-facing state of the NRAMP architecture; furthermore, ScaDMT has been solved in complex with a divalent ion, providing information on the binding site. However, both structures feature 11 transmembrane segments [26, 27]. On the other hand, EcoDMT has 12 transmembrane segments, more relevant to mammalian NRAMPs, and represents the ion-free, outward-facing NRAMP architecture; however, no information is given regarding the binding of ions in this particular state [28].

Molecular models for the caprine and bovine SLC11A1 transmembrane segments were constructed and evaluated through homology modeling. In fact, both the outward-facing and the inward-facing conformations of the transporters were modeled. The outward-facing conformation models were constructed using the structure of monomeric EcoDMT (PDB: 5M87), as the best matched template. For the inward-facing conformation the structure of ScaDMT (PDB: 5M95) was chosen over DraNramp (PDB: 5KTE), first due to its higher sequence identity (35% over 31%) with SLC11A1, with an overall sequence coverage of 79% and second due to the higher resolution in the experimental data (3.4 Å over 3.9 Å). Furthermore, the ScaDMT structure contains almost the entire NRAMP domain with the exception of the N-terminal 1a α -helical segment, while in the DraNramp structure a number of segments are missing, including a significant portion of the 8th transmembrane segment and several inter-helical loops. Finally, the ScaDMT structure provides information regarding the NRAMP ion-binding site at atomic resolution. The ScaDMT structure has been solved as a crystallographic heterotetramer, containing two copies of the transporter in complex with antibody fragments that were used during crystallization, however, the functional unit of the protein is considered to be the monomer. As a result, chain A of the ScaDMT structure (PDB: 5M95), corresponding to a monomer of the transporter was chosen to model the majority of the transmembrane region (see section for the proposed structure), while segments missing from the ScaDMT structure were modeled using the other SLC11 homolog structures. Specifically, the missing coordinates of the 1a helical segment, absent from ScaDMT, were modeled using the equivalent 1a segment in DraNramp (PDB: 5KTE), combined with secondary structure predictions for this particular segment obtained from PSIPRED. Finally, the 12th transmembrane segment, absent from both ScaDMT and DraNramp, was modeled using the equivalent 12th transmembrane segment in EcoDMT (PDB: 5M87).

All molecular modeling was performed using MODELLER v. 9.19 [53]. Sequence alignments for modeling

were prepared with PSI/TM-coffee and further edited with JalView 2. The wild type models of Caprine SLC11A1 in its inward state were subsequently used to generate mutations in three residue positions with potential relevance to metal ion binding. Specifically, Alanine at position 68 was substituted for Glycine, its equivalent in ScaDMT, while additional mutations were performed for Asp-71 and Met-250, which were both substituted for Alanine. The mutant models (Ala-68-Gly, Asp-71-Ala and Met-250-Ala) were generated using the Mutation Wizard in PyMOL. All derived models, both wild type and mutant, were subjected to a short energy minimization to remove steric clashes and optimize geometry, performed with UCSF Chimera [54]. The final models were evaluated using PROCHECK [55] to generate Ramachandran plots and QMEANbrn for per-residue energetic quality analysis, which is a variation of QMEAN designed for the validation of transmembrane protein models [56]. Graphs of the models' topology were prepared with TopDraw [57]. Molecule visualization and rendering were performed with UCSF Chimera, visual molecular dynamics (VMD) [58] and PyMOL [59].

Molecular dynamics simulations

The structure and dynamics of the proposed SLC11A1 models were evaluated through equilibrium molecular dynamics (MD) simulations (Fig. 1b), performed using GROMACS v. 5.1.4 [60]. For each of the two SLC11A1 orthologs, four different simulation systems were prepared, namely, the outward-facing state, the inward-facing state, the inward-facing state with a bound Fe^{2+} ion (Fe-bound) and the inward-facing state with a bound Mn^{2+} ion (Mn-bound). The position of the ions in the latter two systems was based on the equivalent position of Mn^{2+} in the ScaDMT structure (PDB: 5M95), since soaking of ScaDMT crystals with different ions, among them Fe^{2+} showed that the same binding site can interact with several divalent metal ions [27]. In addition to the wild type transporters, metal-bound simulations were performed for the models of the three SLC11A1 mutants, Ala-68-Gly, Asp-71-Ala and Met-250-Ala. Each structure was inserted into a model lipid bilayer, composed by 1-palmitoyl-2-oleyl-phosphatidyl-choline (POPC) lipids, built with the CHARMM-GUI Membrane Builder [61, 62], and solvated using the TIP3P water model with neutralizing NaCl counter-ions. The proteins, membrane and solvent were modeled using the CHARMM36m [63, 64] and CHARMM36 lipid force fields [65], while additional parameters for divalent Fe^{2+} and Mn^{2+} ions, compatible with the CHARMM TIP3P water model, were adapted from the work of Li and co-workers [66, 67]. Each system was subjected to thorough energy minimization, followed by 100 ps of heating and 5 ns of equilibration in the isothermal-isobaric

(NPT) ensemble, with position restraints of gradually decreasing strength applied on the protein coordinates. Subsequently, another 20 ns of NPT equilibration followed without any restraints. Finally, each system was simulated for 150 ns, with each production simulation repeated twice for validation. A summary of the simulation systems is presented in Supplementary Table S1.

All simulations were performed with a 2 fs time step using periodic boundary conditions. A switching function at 1.0 nm and a non-bonded cut-off at 1.2 nm were applied to screen van der Waals and short-range Coulomb interactions, while long-range electrostatic forces were calculated using the Particle Mesh Ewald method. Throughout the simulations, temperature and pressure were maintained at 310 K and 1.013 bar, respectively, with separate couplings for the protein, membrane and solvent. The Berendsen weak coupling ensemble was used during equilibration [68], while production simulations used the Nosé-Hoover thermostat [69] with a coupling constant of 1.0 ps for temperature and the Parrinello–Rahman barostat with a coupling constant of 5.0 ps and a compressibility modulus of $4.5 \times 10^{-5} \text{ bar}^{-1}$ for pressure [70]. Simulation analyses including rendering, visualization, root mean square deviation (RMSD) and root mean square fluctuation (RMSF) calculations were performed using various GROMACS utilities, VMD, PyMOL and in-house scripts.

Essential dynamics analysis

The collective motions sampled by SLC11A1 and their conformational differences between simulations were evaluated through essential dynamics analysis (EDA) [71]. In EDA, the conformational space of a simulation separated in two distinct subspaces, namely, a physically “constrained” subspace describing fast, local fluctuations and an “essential” subspace with few degrees of freedom, which describe the slower overall protein movements; the latter is hypothesized to describe structural transitions that are likely to be relevant for the protein’s function [71]. The method is based on performing principal component analysis (PCA) for the simulation data, i.e. the diagonalization of the covariance matrix of atomic fluctuations, as defined by Eq. 1

$$C_{ij} = \langle (a_i - \langle a_i \rangle) \cdot (a_j - \langle a_j \rangle) \rangle \quad (1)$$

where $a_i (x_i, y_i, z_i)$ and $a_j (x_j, y_j, z_j)$ are the C α carbon coordinates of residues i and j , respectively, and the brackets (“ $\langle \rangle$ ”) symbolize averages over an ensemble of configurations, i.e. the simulation trajectory. The application of PCA yields eigenvectors that describe the directions of correlated motions in the studied protein, while the corresponding eigenvalues describe the root mean square fluctuations

along these directions. The EDA hypothesis states that only the eigenvectors with large eigenvalues describe the overall protein’s collective motions; by convention, these include the first few eigenvectors of the diagonalized matrix [71]. EDA calculations in our study were performed for all simulated SLC11A1 states. Each of the different simulation systems was subjected to standard EDA calculations, using the last 30 ns of each simulation trajectory, where, in all cases, the coordinates of C α carbons were selected in the analysis. EDA calculations were performed using the GROMACS package and the results were analyzed with in-house scripts and PyMOL.

Free energy calculations

The energetics of protein-ion interactions in the Fe²⁺- and Mn²⁺-bound SLC11A1 models were evaluated by performing Free Energy Perturbation (FEP) calculations (Fig. 1c) [72–74]. Similar estimations of protein-ion energetics have been previously performed for the LeuT transporter in complex with various ions [75]. In FEP calculations, the change in free energy between two end states, A and B, (ΔG^{AB}) is estimated by defining a thermodynamic transition pathway of intermediate states between A and B, along a reaction coordinate. The coordinate is represented by a lambda variable (λ) that receives values ranging from 0 to 1 corresponding to the end-states of A and B, respectively. ΔG^{AB} is then defined as a function of lambda coordinate using the Eq. 2

$$\Delta G^{AB} = \sum_{\lambda=0}^1 \left(-k_B T \ln \left\langle \exp \left(- \frac{H_{\lambda+\Delta\lambda} - H_{\lambda}}{k_B T} \right) \right\rangle_{\lambda} \right) \quad (2)$$

where k_B is the Boltzmann constant, T is the temperature, the brackets (“ $\langle \rangle$ ”) denote averaging over the different intermediate λ states and H is the Hamiltonian describing each λ state, as given in Eq. 3

$$H_{\lambda} = E_k + (1 - \lambda)V_A + \lambda V_B \quad (3)$$

where E_k is the kinetic energy, V_A and V_B are the potential energy contributions associated with the two endpoints and the λ parameter indicates the distance along the transformational pathway.

For the purposes of this study, FEP calculations were performed using the thermodynamic cycle presented in Fig. 1c. For each calculation ensemble two systems were considered, namely, the SLC11A1-ion complex (Fe²⁺ or Mn²⁺), as obtained from the results of equilibrium MD simulations and an ion–solvent system, i.e. a Fe²⁺ or Mn²⁺ ion embedded in solvent. Both the wild type and the mutant systems were used in the calculations. In each case, FEP calculations involved the “annihilation” of the ion by gradually decoupling its Lennard-Jones and Coulomb interactions with the

rest of the system; the latter was performed both for the SLC11A1-ion complex and for the ion-solvent system, thus resulting in a “double-annihilation” cycle as described by *Gilson and co-workers* [76]. The standard binding free energy (ΔG_{BIND}) was then taken to be the difference between the ΔG^{AB} values of the protein-ion and ion-solvent systems, as shown in Fig. 1c.

All FEP simulations and subsequent energy calculations were performed using the GROMACS package. The two end-states of the systems, i.e. the systems featuring a fully coupled or fully decoupled ion were represented by λ values of 0.0 and 1.0, respectively, with a $\Delta\lambda$ step of 0.1. To avoid any instability issues or unphysical interactions, the decoupling of Lennard-Jones and Coulomb terms was performed using separate, sequential transformation paths, i.e. Lennard-Jones interactions were decoupled first followed by electrostatic interactions, resulting in a total of 21 λ windows for each FEP calculation. Each window was equilibrated for 1 ns and simulated for 3 ns, with the first nanosecond also discarded as equilibration and the last 2 ns subsequently used for FEP calculations. All FEP simulations were performed using the leap-frog stochastic dynamics algorithm implemented in GROMACS; the rest of the simulation ensemble was modeled using the same configuration as in equilibrium MD simulations. A summary of the FEP calculations is given in Supplementary Table S1 alongside MD simulations.

Residue conservation analysis

Conserved, functionally relevant residues on the sequence and structure of SLC11A1 were identified and analyzed using the approach implemented in ConSurf [77]. ConSurf utilizes structural information, provided by either an experimentally determined structure or a theoretical model, with sequence conservation information, provided by a carefully calibrated multiple sequence alignment. The degree to which an amino acid position is evolutionarily conserved is strongly dependent on its structural and functional importance; rapidly evolving positions are variable while slowly evolving positions are conserved. Thus, conservation analysis of positions among members from the same family can often reveal the importance of each position for the protein (or nucleic acid) structure or function.

An initial dataset of protein sequences containing the NRAMP domain was compiled by scanning UniProt's Reference Proteomes, using the caprine SLC11A1 sequence, the Pfam pHMM profile for the NRAMP domain (Pfam: PF01566) [43] and HMMER v. 3.1 [44]. The search resulted in a set of 5394 sequences containing the NRAMP domain, which were retrieved from UniProt [29]. A multiple sequence alignment (MSA) was performed using the NRAMP pHMM and HMMER's *hmmalign* function and

was further edited using Clustal Omega [47] and JalView [49], to improve the alignment quality of any misaligned regions and/or segments, not contained in the NRAMP profile.

The produced MSA was subsequently refined to reduce “background noise”, guided by the targets sequences of *Capra hircus* and *Bos taurus* SLC11A1. The refinement process was based on a procedure previously described by *Liu and co-workers* [78] and consisted of the following three steps: (i) iterative implementation of the Smith-Waterman algorithm [79] for pairwise alignment using SLC11A1 as the guide and elimination of those sequences below a 62% sequence identity, (ii) deletion of MSA columns that correspond to insertions with respect to the consensus sequence, and (iii) removal of the sequences containing more than 20% gaps in the alignment. These three steps resulted in a final MSA of 537 sequences with N=548 columns, corresponding to the sequence length of the caprine and bovine SLC11A1. MSA refinement was performed using the Evol module in the ProDy-Evol Python package [80]. Sequence logos of alignment regions with particular interest were created using WebLogo [81]. The final MSA was used in combination with the *C. hircus* SLC11A1 theoretical model produced by homology modeling as input in the ConSurf server. The Bayesian method was chosen to calculate rate of evolution values, as it has been found to outperform other methods such as Maximum Likelihood, particularly when the number of sequences in the MSA is small [77].

Results and discussion

Sequence analysis and annotation of SLC11A1

A detailed computational sequence analysis was performed for the caprine and bovine SLC11A1, and the results were then compared against other, well-studied mammalian ortholog sequences which are presented in Table 1. Both caprine and bovine SLC11A1 have 548 residues and feature 12 transmembrane (TM) segments connected by six extracellular and five cytoplasmic loops, with their N- and C-termini located at the cytoplasmic side of the membrane (Fig. 1a). This topology is successfully predicted by four out of five tested transmembrane topology prediction algorithms (Phobius, Octopus, MEMSAT3 and HMMpTM), with the exception of one algorithm (TMHMM) predicting 11 TM segments and the N-terminal domain as extracellular, while missing the first helical transmembrane segment (Supplementary Table S2). Secondary structure predictions performed using PSIPRED (Supplementary Figs. S3 & S4) showed that the two sequences are largely α -helical (~47.8%), and only a small (~3–5%) portion of the predicted β -strands scored by low prediction confidence, while the

Table 1 Sequence annotation of mammalian SLC11A1 and comparison with bacterial SLC11 homologs of known structure

Protein (species)	UniProt AC (Seq. status)	Length	TM segments	NRAMP position	Phosph. sites	Glycosyl. sites ^a
SLC11A1 (<i>C. hircus</i>)	B8XJG7 (TrEMBL)	548	12	75–460	17,23,34,35,534	321,335
SLC11A1 (<i>B. taurus</i>)	Q27981 (SwissProt)	548	12	75–460	17,34,25,534	335
SLC11A1 (<i>O. aries</i>)	P49280 (SwissProt)	548	12	75–460	17,23,34,35,534	321,335
SLC11A1 (<i>H. sapiens</i>)	P49279 (SwissProt)	550	12	78–463	13,15,26,37,38	324, 338
SLC11A1 (<i>M. musculus</i>)	P41251 (SwissProt)	548	12	75–460	15,17,34,35	321, 335
ScaDMT (<i>S. capitis</i>)	A0A0S4MEX1 (TrEMBL)	448	11	53–420	N/A ^b	N/A ^b
EcoDMT (<i>E. coleocola</i>)	E4KPW4 (SwissProt)	511	12	55–439	N/A ^b	N/A ^b
DraNramp (<i>D. radiodurans</i>)	Q9RTP8 (SwissProt)	436	11	60–409	N/A ^b	N/A ^b
BsuMntH (<i>B. subtilis</i>)	P96593 (SwissProt)	425	11	50–399	N/A ^b	N/A ^b

^aFor *C. hircus* and *B. taurus* SLC11A1, glycosylation sites were predicted using HMMpTM. For *O. aries*, *H. sapiens* and *M. musculus* SLC11A1, information on glycosylation sites was obtained from their respective UniProt records. (See also Supplementary Table S5)

^bN/A Not available. Glycosylation and phosphorylation sites were not predicted for bacterial SLC11 homologs

rest of the sequence was predicted as random coil. In addition, the predicted α -helical segments largely correspond to the detected TM topology, providing further validation. Scanning the caprine and bovine SLC11A1 sequences with the Pfam NRAMP pHMM profile shows that the NRAMP domain most likely occupies the first 10 TM helices with a large part of TM1 excluded (residues 75–460), while the last 2 segments seem to form a separate entity (Table 1).

Further comparative analysis of the caprine and bovine SLC11A1 sequences was performed against other, closely related protein sequences, namely, the SLC11A1 orthologs from *Ovis aries*, *Homo sapiens* and *Mus musculus*, as well as three bacterial NRAMP homologs with known structures (ScaDMT, EcoDMT, DraNramp). As shown by their multiple sequence alignment (Fig. 2), SLC11A1 is highly conserved among all five investigated mammals, while differences with its bacterial homologs are larger; still, regions potentially implicated in ion binding and transportation are conserved. The most significant sequence variations observed mainly involve the last, 12th TM segment, which is absent from ScaDMT and DraNramp but present in EcoDMT as well as the N- and C-termini, which differentiate among all aligned sequences (Fig. 2). Apart from the above, differences are also observed for the extracellular loop connecting TM 7 and 8 (7–8 loop). This loop is generally smaller in bacterial NRAMPs and its sequence is divergent among the aligned species. A phylogenetic evaluation of the sequences (Supplementary Fig. S5), performed using Bayesian analysis of the phylogenetic tree derived from their multiple sequence alignment, shows that the bovine, caprine and ovine SLC11A1 sequences are more closely related to one another compared to mouse and human SLC11A1, as well as the bacterial homologs, and vice versa. Interestingly, comparison among the bacterial homologs shows that DraNramp, alongside its close relative, BsuMntH, stands apart from ScaDMT and EcoDMT, possibly due to sequence

differences (including the missing last transmembrane segment in DraNramp) that may reflect *D. radiodurans*'s extremophile nature.

Regarding the transmembrane organization of SLC11A1, its comparative analysis among different mammalian SLC11A1 and bacterial homologs shows a consistent agreement for most of the methods, with only a few cases diverging from the general trend. Specifically, the mammalian orthologs seem to follow a common transmembrane organization, with ovine SLC11A1 showing virtually identical prediction results to those of caprine and bovine SLC11A1, while mouse and human orthologs follow similar trends with a few differences, most likely due to their sequence differences in the non-transmembrane (i.e. N- and C-termini and inter-helical loops) regions. In all five cases of mammalian SLC11A1, only one of the algorithms (TMHMM) failed to predict the correct NRAMP topology. Similarly, the transmembrane topology of the three bacterial NRAMPs was also tested with the same algorithms and their results were evaluated against their known transmembrane positions, as defined in the crystal structures. Prediction results were largely successful with only one exception (EcoDMT), in which one of the algorithms (Phobius) severely underperformed, failing to predict TM segments 1 and 6, as well as the 7–8 loop. Apart from the aforementioned examples, the results of the different methods are generally in agreement both with one another and, in the case of the three bacterial homologs, with experimental evidence from the crystal structures. The largest deviations mainly occur in the correct prediction of transmembrane segments TM1, TM6 and TM8; however these likely occur due to the increased presence of charged residues in TMs 1 and 6 and the variable length and composition of the 7–8 loop, which is very short in bacterial NRAMPs and significantly more extended in mammalian SLC11A1, while the rest of the topology is largely predicted the same. Overall, the predicted topology,

secondary structure composition and domain annotation of *C. hircus* and *B. taurus* SLC11A1 are in agreement with the known topology of other mammalian NRAMPs, as well as the topology of the *E. coleocola* (EcoDMT) crystal structure that most closely resembles the mammalian SLC11A1 secondary structure (Table 1 and Supplementary Table S2).

The physicochemical properties and the amino acid distribution in the protein consist of an important aspect in the characterization of the sequence information, especially across the protein interface and on its surface, which are a principal key in the molecular interactions with the protein. To further understand the biological function of members in SLC11 family, the ruminant's SLC11A1 physicochemical properties were compared with the non-ruminant's and bacterial's properties, as predicted using ProtParam (see “Methods”), and the results are shown in Table 2. In general, all tested sequences showed features indicative of transmembrane proteins, i.e. a high aliphatic index, corresponding to the increased presence of aliphatic residues [82] and positive GRAVY values in the 0.5–0.7 range, corresponding to increased hydrophobic stretches indicative of transmembrane α -helices. Charged residues are mainly located in interhelical loops and at the ends of predicted transmembrane segments, while positively-charged residues are located primarily at the cytoplasmic side, in agreement with the “positive inside” rule that characterizes α -helical transmembrane proteins, although depending on the pH (Supplementary Figures S10–S12 and Table S8) [83]. A few charged residues also appear in the middle of transmembrane segments (e.g. Asp-71, His-252 in segments 1 and 6 of caprine SLC11A1), however, these are located in regions that are likely to participate in metal transportation (e.g. the NRAMP metal-ion site) and are, therefore, expected. Interestingly, the isoelectric point (pI) values of caprine, bovine and ovine SLC11A1 are 6.25, 6.39 and 6.39, respectively, classifying

the proteins as slightly negatively charged, in contrast to murine and human SLC11A1, which have pI values closer to neutral conditions. This divergence may be indicative of the ruminant species' more pronounced susceptibility to infection by MAP, especially considering the fact that infection by various *Mycobacterium* species and subspecies has been connected with pH conditions in phagosomes. As far as the bacterial homologs are concerned, EcoDMT and DraNrap have significantly larger pI values (9.13 and 9.55, respectively). The predicted instability index of the sequences has values in the 41–47 range and classifies all mammalian SLC11A1 orthologs, as well as DraNrap, as unstable, an observation that is in agreement both with the nature of transmembrane proteins and with prior knowledge on the stability of mammalian NRAMPs in general. Interestingly, ScaDMT and EcoDM display significantly lower values for the instability index (34.21 and 32.52, respectively) and are classified as stable; this is in agreement with experimental evidence on the latter two transporters, which indicates that, compared to mammalian NRAMPs, these two homologs are more stable and therefore, easier to purify and crystallize [27, 28]; thus, judging from its instability index and similar type of evidence we hypothesize that purification of mammalian NRAMPs could be laborious.

Potential post-translational modifications in SLC11A1

Potential post-translational modification (PTM) positions in the SLC11A1 sequence were detected using HMMpTM, which is an algorithm that combines transmembrane topology prediction with the detection of PTMs, to increase prediction confidence as well as the ScanProsite and NetNGlyc tools (see “Methods”). HMMpTM and ScanProsite

Table 2 Physicochemical properties of mammalian SLC11A1 sequences and their bacterial homologs

Species	UniProt AC	Length	Mol. weight (Da)	Isoel. point (pI)	Aliph. index (AI)	Instability index (U/S) ^a	GRAVY index
<i>C. hircus</i>	B8XJG7	548	59403.70	6.25	112.86	44.24 (U)	0.565
<i>B. taurus</i>	Q27981	548	59567.86	6.39	113.92	47.40 (U)	0.545
<i>O. aries</i>	P49280	548	59411.46	6.39	111.44	43.38 (U)	0.537
<i>H. sapiens</i>	P49279	550	59872.25	7.58	113.00	40.29 (U)	0.542
<i>M. musculus</i>	P41251	548	59741.19	7.52	113.21	42.17 (U)	0.550
<i>S. capitis</i> ^b	A0A0S4MEX1	448	49261.38	6.71	132.14	34.21 (S)	0.719
<i>E. coleocola</i> ^b	E4KPW4	511	55701.95	9.13	130.72	32.52 (S)	0.692
<i>D. radiodurans</i>	Q9RTP8	436	46652.33	9.55	124.20	41.24 (U)	0.693
<i>B. subtilis</i>	P96593	425	45685.61	9.54	131.11	30.32 (S)	0.901

^aU Unstable, S stable

^bScaDMAT, EcoDMT, BsuMnH (S): This value classifies the proteins as stable compared to the other proteins that were classified as unstable, as predicted by the ProtParam online program (see “Methods”)

predictions detected the presence of several phosphorylation sites at the N-terminus and one site at the C-terminus of both caprine and bovine SLC11A1, located at the cytoplasmic side of the membrane (Table 1); the same pattern was observed for the N-terminus of human and mouse SLC11A1 orthologs, despite the sequence differences in that particular region. Detailed list of the motifs of the relevant PTM sites include also N-myristoylation and two types of phosphorylation sites, as shown in Supplementary Table S5. Phosphorylation, occurring primarily at the cytoplasmic side, can induce signaling pathways in receptors or regulate enzymatic activity at membrane-bound enzymes. Most importantly, phosphorylation is often used to mark proteins for endocytosis and subsequent recycling or proteolysis [84]. As such, the phosphorylation sites predicted for SLC11A1 in this study through sequence analysis, appearing primarily at the N-terminus with another position at the C-terminus, could be potential sites for further studying the aforementioned features. Although the mechanisms of SLC11A1 phosphorylation and its effects on SLC11A1 activity remain largely unknown, specific types of phosphorylation such as Src family kinases are required for phosphorylation and activation of SLC11A1 in macrophages [85].

Asn-glycosylation, occurring primarily at the extracellular side, can influence proper folding and translocation of transmembrane proteins and, in some cases, is required for their proper function. This also applies to SLC11A1, as experimental evidence has shown that incomplete N-glycosylation of the transporter can lead to defective intracellular targeting and, therefore, loss of SLC11A1 action [24]. Thus, potential Asn-glycosylation sites were also searched and detected to be located at sequence positions 321 and 335 in *C. hircus* and *O. aries* SLC11A1 and in sequence position 335 in *B. taurus* SLC11A1. The results for the caprine and bovine SLC11A1 were then compared with the known PTM sites in the human and mouse SLC11A1 sequences (Table 1, Supplementary Table S5, Fig. 2 and Supplementary Fig. S1 and S2), as retrieved from their annotation records in UniProt. As shown (Table 1, Supplementary Table S5 and Fig. 2), these positions are highly conserved among the tested mammalian SLC11A1 sequences, which display similar glycosylation patterns with the exception of bovine SLC11A1, in which Asn-321 is substituted for aspartate, thus reducing the number of glycosylation sites for *Bos taurus*. Notably, both glycosylation positions in all species reside at the 7–8 loop towards the extracellular compartment. Interestingly, this particular loop region was recently proposed to participate in the regulation of ion transport for the DrnRamp SLC11 homolog [26]. Considering that experimental trials have also linked the glycosylation state of SLC11A1 with the regulation of its canonical function [24], it is very likely that the presence of glycosylation sites at the 7–8 loop could be indicative of a mechanism by which PTMs affect SLC11A1

function in *C. hircus*, *B. taurus* and, potentially, other species as well.

Ultimately, post-translational modifications such as N-glycosylation in this region as well as other regions have functional consequences affecting membrane protein trafficking. Thus, the amino acid positions 314–340, are less conserved with peaks at positions 320–325, and 330–339 where also N-glycosylation takes place, are shown in Supplementary Table S5, as well as the consensus tripeptide Asn-Xaa-Ser/Thr (sequons) including also full information on all potential Asparagines with high score that may not be glycosylated [39, 40] (Supplementary Table S7). A recent example of the importance of glycosylation is the case of the Lysosomal Protein CLN5. Research with mutants has shown that lack of N-glycosylation on certain sites due to mutation, caused retention in the endoplasmic reticulum of the Lysosomal Protein CLN5, which is not only important for protein folding but also essential for CLN5 trafficking to the lysosome [86].

A proposed three-dimensional structure for SLC11A1

Models for the three-dimensional structure of the *C. hircus* and *B. taurus* SLC11A1 transmembrane region were constructed, using the experimentally determined structures of bacterial homologs EcoDMT, ScaDMT and DrnRamp as templates (Fig. 3). For both species, models were constructed for the outward and the inward-facing state, the former using the EcoDMT structure as template and the latter being based mostly on ScaDMT, with additional segments modeled, where necessary, after the equivalent regions of DrnRamp and EcoDMT (see “Methods” for details). The ion-bound state of the transporters with either Fe^{2+} or Mn^{2+} was also modeled, using the inward-facing state and experimental information on ion binding from the ScaDMT homolog. Structural analysis of the models, using PROCHECK to construct Ramachandran plots (Supplementary Fig. S6) and QMEANbrane (Supplementary Fig. S7) to evaluate their energetic and structural integrity, showed that the proposed SLC11A1 structure is well-defined, with the majority (~92%) of the residues adopting phi and psi torsion values in the most favorable Ramachandran regions, while QMEANbrane energetic values were in ranges indicative of transmembrane proteins. Thus, only a few outliers were observed, mostly located in unstructured regions such as the N- and C-termini, or the interhelical loops connecting transmembrane segments.

Both the outward and the inward-facing state of SLC11A1 adopt the same general architecture, featuring 12 α -helical TM segments. The first 10 segments form a LeuT-like fold of two pseudosymmetry-related structural repeats, comprised

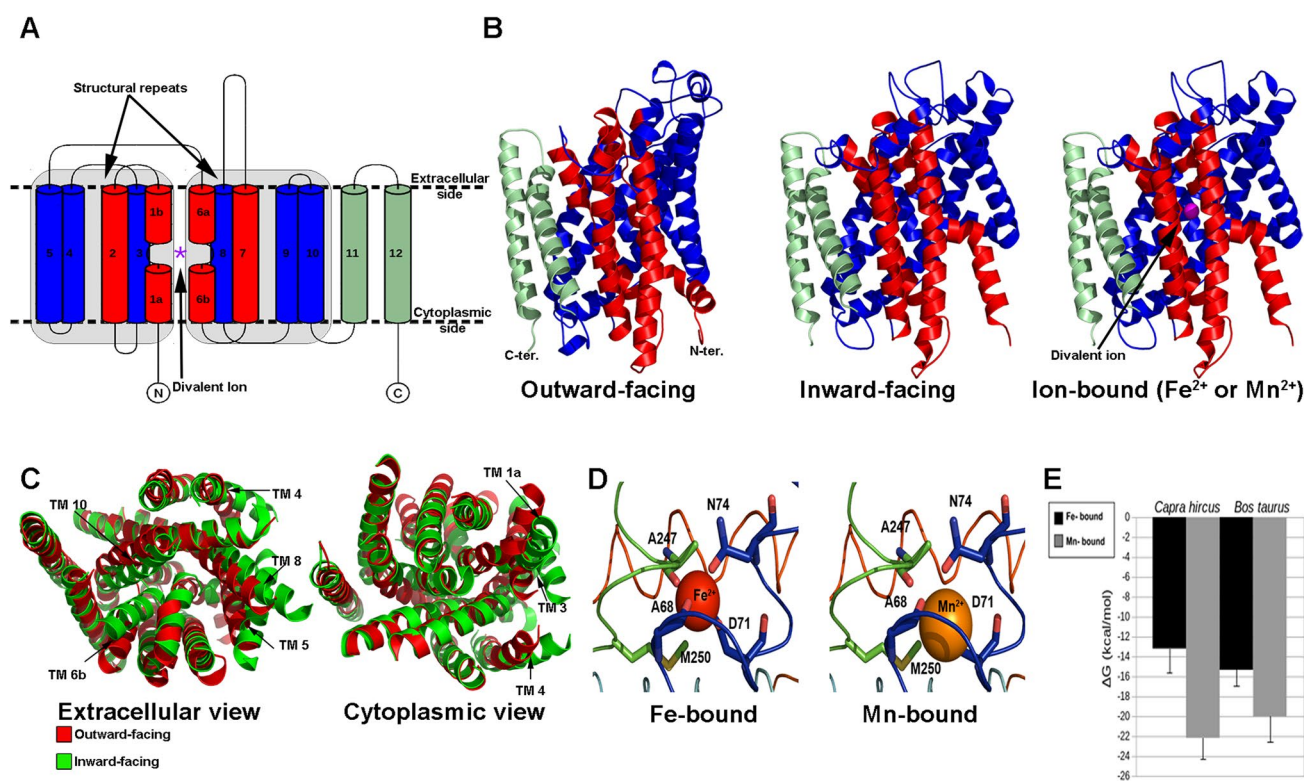


Fig. 3 Structural features of SLC11A1. **a** Topology diagram of the proposed SLC11A1 structure, as derived through homology modeling for *C. hircus* and *B. taurus* (see “Methods”). Transmembrane α -helices are shown as cylinders and labeled accordingly. Helices participating in the “bundle” that handles ion support are colored in red (1a, 1b, 2, 6a, 6b and 7), while helices forming the “scaffold” are colored in blue (3, 4, 5, 8, 9 and 10). The final two transmembrane helices, which do not form part of the NRAMP domain, are colored in light green (11 and 12). The regions forming the two pseudo-symmetric structural repeats of the NRAMP domain are indicated by grey boxes in the background. The ion-binding site is labeled using a purple asterisk pointed by an arrow. **b** Proposed three dimensional structures of the outward-facing, inward-facing and ion-bound states of SLC11A1. The models of *C. hircus* SLC11A1 are shown, with the bovine transporter adopting a very similar architecture (data not shown). The structures are shown in cartoon representation and colored using the same scheme as in the topology diagram of (a).

The bound ion in the ion-bound structure is shown as a purple sphere (arrow) and labeled accordingly. Similar structural features were obtained for the bovine SLC11A1 (data not shown). **c** Comparison of the outward-facing (red) and inward-facing (green) states through structure superimposition. The extracellular and cytoplasmic views of caprine SLC11A1 are shown. Loops between transmembrane helices are not rendered for clarity. Segments showing significant differences are labeled accordingly. **d** Protein-ion interactions of the Fe-bound (left) and Mn-bound SLC11A1 (right) after 150 ns of Molecular Dynamics simulations. The two ions are shown as van der Waals spheres, while interacting residues are shown as sticks and labeled accordingly. **e** Bar chart showing the results of Free Energy Perturbation calculations for *C. hircus* and *B. taurus* SLC11A1. Binding energy values (in kcal/mol) are given in the vertical axis. Measurements are colored black for the Fe-bound and grey for the Mn-bound systems

by TM segments 1–5 and 6–10, respectively; these form the NRAMP domain that exists in all known SLC11 homologs (Fig. 3a). Two additional TM α -helices follow that are seemingly unrelated to the fold; these are common to mammalian NRAMPs and also appear in the EcoDMT structure, but are missing from other bacterial homologs such as ScaDMT and DraNrap. TM segments 1 and 6 become unwound and bent near the center of the membrane’s hydrophobic core, thus leading to two separate helices per segment (i.e. 1a–1b for TM 1 and 6a–6b for TM 6), connected by small hinges. These unwound hinges contain a number of polar or charged residues (e.g. Asp-71, Asn-74, His-252 etc) and most likely form the main ion-binding site of the

transporters, as evidenced by previous experimental information on the ScaDMT homolog. Generally speaking, the topology of our model is in good agreement with the results of the transmembrane topology predictions. Still, differences are observed for the first two TM segments, which also deviated in the predictions used in the sequence analysis; the differences, however, can be attributed to the unique nature of TM 1, which is split into segments 1a and 1b and features several polar residues in the hinge region between them, and therefore, influencing the results of transmembrane topology prediction algorithms. Overall, the organization of the SLC11A1 transmembrane region seems to adopt similar characteristics with bacterial SLC11 homologs and LeuT

transporters: TM segments 1 (1a–1b), 2, 6 (6a–6b) and 7 form a “bundle” that seemingly handles ion transportation, TM segments 3, 4, 5, 8, 9 and 10 form a “scaffold” that supports the “bundle”, while the last two segments, 11 and 12 are not part of the NRAMP domain and seemingly remain distinct from the rest of the structure (Fig. 3a, b).

Despite their similar architecture, a number of differences are observed between the inward and outward-facing states of SLC11A1 (Fig. 3c). Specifically, significant structural variations are observed for TM segments 1a, 1b, 2, 6a, 6b and 7 forming the “bundle”, while some differences also occur for TM segments 4, 5 and 8 and 10, participating in the “scaffold”. These structural variations essentially lead to the formation of the two distinct states, as the movement of segments 1a, 6b and the cytoplasmic end of 2 contribute to opening of the ion-binding site towards the cytoplasm for the inward-facing state, while equivalent variations of 1b, 6a and the extracellular end of 7 lead to an opening towards the extracellular space for the outward-facing state (Fig. 3c). We hypothesize that, changes in the “scaffold” around these segments, as indicated by the movement of transmembrane segments 4, 5, 8 and 10, may contribute to the stabilization of each distinct state, as equivalent, functionally relevant transitions have been previously reported for other transporter

families with a similar “bundle”-“scaffold” organization [87]. Overall, these structural variations can provide some hints as to what structural transitions may occur for SLC11A1 during metal ion transportation. Interestingly, significant differences were also observed in the extracellular 7–8 loop. Although these differences can be attributed to the random nature of loop construction during homology modeling, it is important to note that the structural diversity of this particular segment may have functional implications for the ion transport function in NRAMPs.

Stability, dynamics and collective motions of SLC11A1 in different states

The SLC11A1 models were further validated by performing equilibrium Molecular Dynamics (MD) simulations (Fig. 4, see also Supplementary Fig. S8). The outward-facing, inward-facing and ion-bound (Fe^{2+} or Mn^{2+}) states were embedded in model lipid bilayers and simulated using the recently developed CHARMM36m force field [64]. In all cases the organization and the architecture of the NRAMP domain remain stable, with no dramatic structural changes, providing an indication towards all proposed models' validity. A geometric analysis of the simulation results is given in

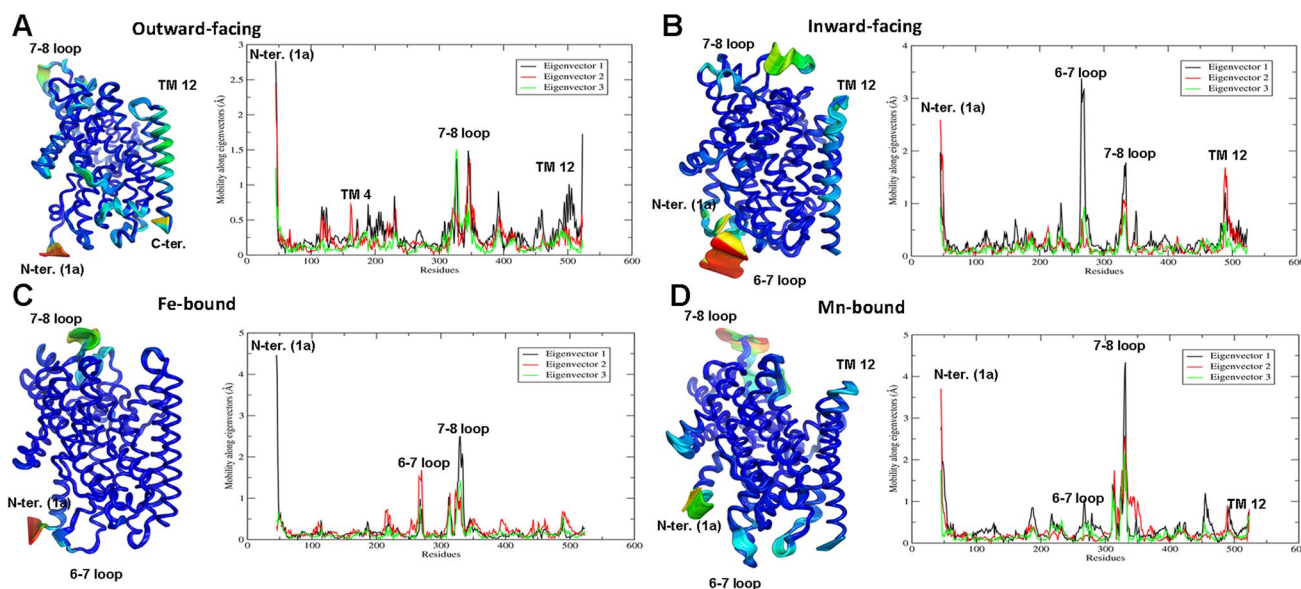


Fig. 4 Characteristic dynamics of SLC11A1. Results are shown for the caprine SLC11A1 in its outward-facing (a) and inward-facing (b) state, Fe-bound (c) and Mn-bound (d) inward-facing states. For each state, a structural projection of motion along the first eigenvector is shown (left), accompanied by per-residue mobility analysis of the projections from the first three eigenvectors (right). The structural projections are shown in ribbon orientation, colored using a rainbow scale ranging from deep blue to dark red, with the size and color of each region corresponding to its mobility. Regions with increasing movement are shown in bigger size and warm color

shades, with residues of the most extensive collective motions colored in dark red and having the largest size. The per-residue mobility plots follow the representation pattern of RMSF calculations. The horizontal axis depicts the residue index and the vertical axis depicts mobility (in Å). The first three eigenvectors are shown, colored black, red and green. Structural regions with potential interest are labeled accordingly both in structural depictions and in mobility plots. Similar dynamics results were obtained for the bovine SLC11A1 (data not shown)

Supplementary Fig. S8, including Root Mean Square Deviation (RMSD) calculations from the starting configuration and per-residue Root Mean Square Fluctuation (RMSF) calculations from the average structure. In both *Capra hircus* and *Bos taurus*, all simulated SLC11A1 states seem to equilibrate well within the simulation time scale, with their RMSD curves reaching a plateau at approximately 80–100 ns and remaining relatively stable onwards. The RMSD values of the last 30 ns for the different SLC11A1 states are within the range of 2.5–4.0 Å, indicating moderate rather than extensive structural transitions (Supplementary Table S3). The per-residue mobility analysis, performed through RMSF calculations, shows that the aforementioned value range can be attributed to the movement of the models' unstructured regions that are expected to be highly mobile, namely, the N- and C-termini and the loops connecting adjacent transmembrane segments, while the RMSF values of residues belonging in the transmembrane segments themselves are almost exclusively below 2.0 Å (Supplementary Fig. S8), indicating structural stability. More importantly, the comparison of the initial and final configurations for each SLC11A1 state, as indicated by their RMSD values, shows that throughout the simulations all systems retained their unique structural characteristics, and no transition was observed from one state to another (for example, a shift of inward-facing SLC11A1 to its outward-facing conformation, or vice versa). Overall, the simulation results indicate that the tested models are stable and structurally valid entities.

The dynamics of SLC11A1 were further analyzed by identifying and evaluating the collective motions of the proteins during simulation using EDA. EDA attempts to identify the potentially relevant structural motions of a protein by separating them from the local fluctuations of thermal motion through principal component analysis, so that the slow, “essential” structural transitions are represented by a few principal components, typically corresponding to the first 1–3 eigenvectors of the analysis [71]. Such analysis was performed for the SLC11A1 model in its different states, and presented in Fig. 4 for the *Capra hircus* models, with similar results for *Bos taurus* (data not shown). For each model, the collective motions were identified and studied by projecting the mobility of the proteins along the first three eigenvectors in an RMSF-like fashion, as well as 3d projections of the mobility along the first eigenvector upon the SLC11A1 structure (Fig. 4).

As revealed by EDA calculations, the most characteristic motions of SLC11A1 are focused on specific segments of the protein, namely, the N-terminal region, transmembrane segment TM 4, a part of the cytoplasmic loop connecting segments 6b and 7 (6–7 loop), a part of the extracellular loop connecting TM 7 and 8 (7–8 loop) and the extracellular region of TM 12 (loop 11–12) (Figs. 3a, 4). Some of these segments, namely, the N- and C-termini and the 6–7 and

7–8 loops are among the models' least structured regions and are likely to fluctuate, while other segments, namely, TM 4 and TM 12, have a more rigid, helical structure. These fluctuations are also observed in the standard RMSF analysis (Supplementary Fig. S8), however, the application of EDA, enabling the removal of fast fluctuations, reveals their magnitude more clearly. At the same time, the co-existence of these motions in the same eigenvectors indicates that the corresponding segments could move in concert with one another and could be used as indicators of potential dynamic connections between these parts of SLC11A1.

It is important to note that, with the exception of transmembrane segment 12, the other segments can be linked to the regulation of NRAMP function. Specifically, both the N-terminus near TM 1a and the 6–7 loop are close to the ion-binding site of the NRAMP domain, formed by the segments 1a–1b and 6a–6b. At the same time, the N-terminus and segment 1a and the 7–8 loop have been proposed to co-regulate ion transport in the DraNramp bacterial SLC11 homolog [26]. Specifically, the unencumbered movement of the N-terminus and segment 1a is believed to be important for the conformational changes of DraNramp, while the movement of the 7–8 loop has been proposed to participate in revealing and/or occluding the transporter's metal binding site [26]. TM 4 has been shown to differentiate in the structures of the outward and inward-facing states, both in the ScaDMT and EcoDMT homologs and in the models presented in the current study. Movements of the 6b region and its adjacent 6–7 cytoplasmic loop have also been implicated in ion transport in both SLC11 and LeuT transporters [26, 88]. Furthermore, mutagenesis experiments for a conserved histidine in close proximity to the 6b region, performed for the EcoDMT bacterial homolog, have implicated that particular region with the proton co-transport function of NRAMPs [28]. The observation of collective motions involving these segments, particularly the 1a segment and the 7–8 loop, observed in the simulations of the current study indicates a potential functional correlation for these segments that is in agreement with the aforementioned experimental observations for the bacterial SLC11 homologs, and could be an indication that mammalian SLC11A1 and, perhaps, NRAMPs in general, may follow similar structural and functional patterns.

Interestingly, the collective motions' profile displays deviations between the different SLC11A1 states. Specifically, the outward-facing state's collective motions involve mainly the 7–8 loop and the N-terminal region near 1a with additional movements of TM 4, TM 12 and the C-terminus. On the other hand, the inward-facing state also involves movements in the 6–7 loop that are absent in the outward-facing state. This observation hints at the differences in the dynamics of the two states, as the inward-facing state would be expected to have a more mobile cytoplasmic side, while the opposite would be true for the outward-facing

state. The fluctuation of the 7–8 loop is apparent in both states; however, considering the functional implications of this particular site, as already outlined, its mobility is not surprising. Interestingly, the dynamics of the Fe- and Mn-bound states also differ from the empty inward-facing state, specifically, in the contributions of the 6–7 loop and TM 12 (compare Fig. 4a with Fig. 4b–d). While both the outward-facing and the inward-facing states have a relatively mobile TM 12, the same does not apply to the ion-bound models. This could be partially attributed to the modeling procedure followed to construct TM 12 for the inward-facing and ion-bound models; as stated in the "Methods" section, the TM 12 helix of the outward-facing template (EcoDMT) was used to model the segment in the inward-facing models, since their templates (ScaDMT and DraNrp) lacked this particular segment. It is possible the followed procedure could have influenced the dynamics of TM 12 during the simulations of the inward-facing and ion-bound states. Still, it is worth noting that the mobility of TM 12 is reduced only in ion-bound SLC11A1 and not in the inward-facing state, with the latter displaying collective motions for the segment that resemble the outward-facing state. This could indicate a potential influence of the metal ion in the binding site on TM 12. However, a putative structural perturbation emanated from the ion binding site (TMs 1a–1b, 6a–6b) having an effect on a remote loop or distant TM, such as TM 12, is currently unknown and remains to be experimentally validated. Finally, fluctuations are also observed in the 6–7 loop between the ion-bound and empty states of SLC11A1. In the empty state, inward-facing SLC11A1 (Fig. 4b–d) seems to have a highly mobile cytoplasmic side with large fluctuations of the 6–7 loop. More importantly, the presence of an ion in the binding site of inward-facing state seems to cause a distinct mobility alteration in the 6–7 loop between the metal-bound states (Fig. 4c, d).

SLC11A1 preference for Mn^{2+} over Fe^{2+} ion binding

To further illustrate the dynamics of divalent metal ion binding in SLC11A1, MD simulations and Free Energy Perturbation (FEP) calculations were performed in GROMACS (see "Methods"), featuring the binding of a Fe^{2+} or Mn^{2+} ion into the SLC11A1 metal ion binding site. In all cases, the simulations were performed using the inward-facing state of the transporters. Initial protein-ion interactions were modeled using the equivalent protein-ion interactions observed in the ScaDMT structure [27].

In the caprine and bovine SLC11A1 orthologs these residues included Asp-71 and Asn-74 from the first transmembrane segment and Ala-247 and Met-250 from the sixth transmembrane segment; notably, all the aforementioned residues were identical among the different sequences of SLC11A1 under study (Figs. 2, 3a, d). These interactions

were all retained throughout all MD simulations for both Fe^{2+} - and Mn^{2+} -bound SLC11A1, while additional interactions were also formed between the ions and an adjacent alanine (Ala-68) (Fig. 3d). The stability of these interatomic contacts throughout the simulations suggests the thermodynamic validity of the proposed interactions in the binding site. Strikingly, this applies not only to the residues corresponding to the ion binding site of ScaDMT, from which the ion-bound complexes of the current study were initially modeled, but also to Ala-68, which formed contacts with Fe^{2+} and Mn^{2+} during the simulations. It should be noted that Ala-68 is substituted for glycine in the ScaDMT and EcoDMT structures; this change could be part of the reason why no interaction of the positions with the ions was observed in ScaDMT, since the more flexible glycine could result in slightly different conformations for the binding site. To evaluate the overall strength of interactions between the divalent ions and the binding site, binding energy (ΔG_{BIND}) estimations were performed using FEP calculations (Fig. 3E, see also Supplementary Table S6). The ΔG_{BIND} results showed that Mn^{2+} -bound simulations display more favorable protein-ion interactions compared to the Fe^{2+} -bound simulations. Strictly speaking with thermodynamic values, ΔG_{BIND} for the Mn^{2+} -bound complexes were -20.04 ± 2.59 kcal/mol for *B. taurus* and -22.21 ± 2.08 kcal/mol for *C. hircus*, while the equivalent values for the Fe^{2+} -bound complexes were -15.37 ± 1.57 kcal/mol (*B. taurus*) and -13.22 ± 2.40 kcal/mol (*C. hircus*), respectively (Fig. 3e). Remarkably, these observations are in good agreement with previous experimental evidence that propose a preference of SLC11A1 for Mn^{2+} over Fe^{2+} [14].

Interestingly, previous experimental and crystallographic evidence for the ScaDMT bacterial homolog have implicated the presence of a methionine (Met-230) located in the hinge between segments 6a and 6b (Met-230 corresponds to Met-250 in caprine SLC11A1 and the conserved residue position 259 on the top of the alignment, see Fig. 2), with protein-ion interactions of the binding site and the co-transportation of protons alongside divalent metals [27, 89]. Mutations of this residue in the DraNrp homolog have also been shown to alter DraNrp's selectivity towards different ions [26]. As shown by the ion-bound simulations of the caprine and bovine SLC11A1 and FEP calculations in the current study, the equivalent methionine (Met-250) in SLC11A1 and bacterial transporters also form part of the protein-ion interface. Along with the aforementioned work from the bacterial DMTs, supportive evidence of the ubiquitous presence of Met-230 among the SLC11 family members attracts our attention in the position of Met-230, that has been proven to be the conserved ligand for divalent metal-ion binding [89]. Thus, both system's work invigorates the agreement of the current study's prediction models and simulation results

with the experimental evidence for the SLC11 family, especially given the highly conserved features of the ion-binding site, supports the validity of the proposed mammalian SLC11A1 structure.

To further test the metal-binding properties of the proposed SLC11A1 structure, mutant models of the caprine transporter were simulated in the presence of a metal ion (Fe^{2+} and Mn^{2+}) at the binding site. Specifically, three residue positions, namely Ala-68, Asp-71 and Met-250 were considered. All three residues had been implicated in protein-metal interactions during the simulations of the wild type model. Ala-68 was mutated back to glycine, its equivalent residue in the structures of the ScaDMT and EcoDMT homologs, while Asp-71 and Met-250 were both mutated to alanine, in line with previous experimental evidence for the DraNrapm homolog [26].

As shown by the simulation results (Supplementary Figure S9), the Ala-68-Gly and Met-250-Ala simulation systems retained their ion-binding capabilities, as the metal ion remained bound in the 1a-1b/6a-6b binding site throughout their simulations. To further investigate the role of the two residues, additional FEP calculations (Supplementary Table S6) were performed for the two mutants. The calculations showed that the Ala-68-Gly substitution had a mild effect upon the binding energy compared to the wild type but did not impair SLC11A1's selectivity towards different metals (ΔG values were -7.17 ± 2.31 for the Fe-bound and -20.63 ± 1.78 kcal/mol for the Mn-bound model). This could mean that the role of the newly-formed Ala-68 contact with the ions could be auxiliary in stabilizing the interface, rather than participating in ion selectivity. On the other hand, FEP calculations for the Met-250-Ala mutant showed nearly identical values for both Fe^{2+} and Mn^{2+} (-19.85 and -19.87 kcal/mol, respectively), indicating that the substitution of methionine may lead to a loss for SLC11A1 metal ion selectivity. In contrast to the Ala-68-Gly and Met-250-Ala, simulations of the Asp-71-Ala mutant with either Fe^{2+} or Mn^{2+} bound displayed very limited ion-binding capabilities, as the metal ion was shown to exit the site.

Taken together, the simulation results for the mutant constructs suggest that the most important determinants for metal binding would be Asp-71 and Met-250. Asp-71, having a negative charge, would seem to be the main anchor for holding the positively charged Fe^{2+} or Mn^{2+} cations in the binding site. On the other hand, Met-250 could be part of the mechanism through which SLC11A1 discriminates against different ions. Impressively, this observation is in agreement with experimental evidence for DraNrapm, in which methionine to alanine mutations resulted in non-selective transporter constructs that could bind ions without selectivity [26].

Conservation analysis of the SLC11A1 sequence, structure and potential functional sites

The modeling and MD simulation results of this study were complemented by residue conservation analysis of SLC11A1 (Fig. 5). A multiple sequence alignment (MSA), representing proteins of the NRAMP family, was used alongside the proposed SLC11A1 structure to calculate the conservation status of each residue with ConSurf [77]. The results were mapped upon the structure of caprine SLC11A1 and used to evaluate the segments of the protein that are more conserved (Fig. 5a). The conservation was also reflected by a per-residue analysis, using the ConSurf conservation score and its accompanying rank order to evaluate each residue position (Fig. 5b). The ConSurf scoring function uses negative values for conserved positions and positive values for variable ones. This is also translated in a rank ordering scheme, in which each residue is assigned a rank ranging from 1 to 9, corresponding to the least and most conserved positions, respectively. Finally, the sequence profiles of specific SLC11A1 segments with special interest were evaluated by creating sequence logos (Fig. 5c).

As it can be seen, the least conserved regions, and those displaying fluctuations in the simulations (Fig. 4b–d), comprise of the 11th and 12th TM helices, the C-terminus and N-terminal residues near TM 1a, a portion of the cytoplasmic 6–7 loop and a portion of the extracellular 7–8 loop. Their low conservation was expected, since these regions differentiate significantly among NRAMPs, as some members of the family can have fewer TM segments or shorter loops. As far as the rest of SLC11A1 sequence is concerned, the most conserved regions include the buried part of the first 10 TM segments, comprising the NRAMP domain. However, a number of variable spots also appear within the TM “bundle”, corresponding to inter-helical loops and positions with high variability among different sequences. Among the functionally relevant regions of SLC11A1, the most prominent ones include the 1a-1b and 6a-6b regions comprising the ion-binding site; this can also be seen in the sequence logos of the pertinent positions (Fig. 5c). These regions share high sequence similarity among all NRAMPs sampled in the MSA, since they perform the most basic function in the family, which is divalent metal ion binding and transportation. However, not all regions with potential functional implications are highly conserved. An example of this involves TM 4, which shows structural transitions between the outward and inward-facing states of SLC11A1. As shown by conservation analysis (Fig. 5b, c), a number of positions in this particular region display variability.

A more detailed analysis on conservation propensity and ranking was also performed (Supplementary Table S4), focusing on residues with potentially significant roles in the function of SLC11A1. These included residues that may

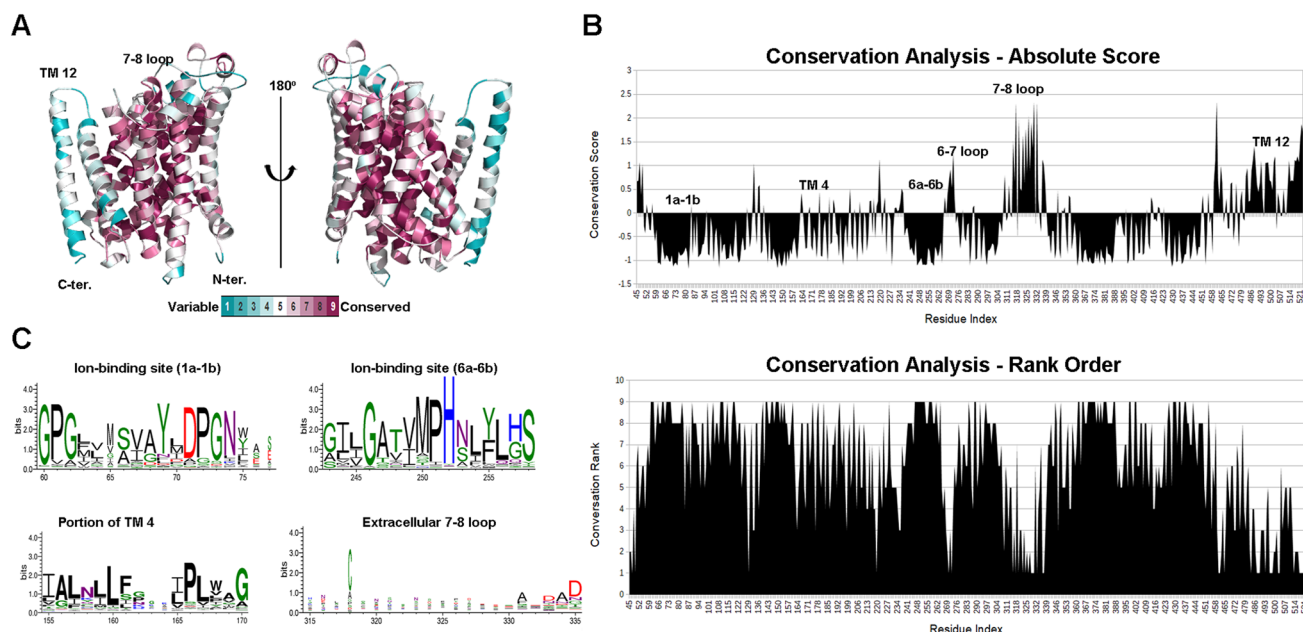


Fig. 5 Sequence and Structure Conservation Analysis for SLC11A1. **a** The residue conservation profile of SLC11A1 as calculated by ConSurf is mapped upon the modeled structure of *C. hircus* SLC11A1. The protein is shown in cartoon representation and each segment is colored by its conservation status with colors ranging from blue to purple, corresponding to highly variable and highly conserved sites, respectively. **b** Per-residue conservation analysis plots, showing absolute scores (top) and conservation ranking (bottom) for each residue in the modeled SLC11A1 structure. The horizontal axis in both graphs depicts residue indices. In the top panel more negative score values indicate more conserved residues while more positive score values indicate more variable positions. In the bottom panel the rank

order ranges from 1 to 9, corresponding to the most variable and most conserved positions, respectively. **c** Sequence logos of four segments from the NRAMP multiple sequence alignment, corresponding to the regions with special interest. The horizontal axes display residue indices, while the vertical axes display bit score values as calculated from the alignment. Residues in the logos are colored by their physico-chemical properties (black for A, L, V, I, P, M, C, F and W, green for G, S, T and Y, purple for N and Q, blue for H, R and K, red for D and E) and sized based on their relative frequency in the alignment, with larger letters corresponding to residues occurring more frequently. Similar conservation analysis and results were obtained for the bovine SLC11A1 (data not shown)

participate in the basic functionality of NRAMPs, like the protein-metal ion interactions or confer on the transporter proton co-transport capabilities. As already stated, a number of residues were found to participate in strong protein-metal interactions during the SLC11A1 simulations, namely, Ala-68, Asp-71, Asn-74, Ala-247 and Met-250. Most of these residues have also been experimentally implicated in metal transportation in bacterial NRAMP homologs [26, 27, 89]. As shown by the conservation analysis (Fig. 5), all these residues are in the highest rank of conservation for SLC11A1 (rank 9). Interestingly, the variability of Ala-68 between eukaryotic and bacterial NRAMP homologs is reflected in the conservation score of the residue position. Although all ion-binding residues are assigned to the top conservation rank, Ala-68 has the least negative score among them and is more prone to variability. In any case, conservation analysis confirms the significance of the presented residue positions and validates them as part of the protein-ion interface.

Previous work in the EcoDMT bacterial homolog has identified two residues with potential relevance for the proton co-transport mechanism accompanying iron

transportation in NRAMPs [28]. The first residue (Glu-129 in EcoDMT) is the equivalent of Glu-149 in SLC11A1, which is located in transmembrane segment 3 (TM3). Mutation of this residue to Gln or Ala was found to have a mild negative effect upon the proton transportation capabilities of EcoDMT [28]. As shown in Supplementary Table S4, this position is not only identical in SLC11A1 but also highly conserved in the NRAMP family, ranked 9 in the ConSurf conservation scale. The second residue implicated in proton co-transport is His-236 in EcoDMT (or His-228 and His-232 for ScaDMT and DraDMT, respectively-Fig. 2), corresponding to His-252 in the ruminant and mouse SLC11A1 sequence and to His-255 in the human sequence (Fig. 2). Mutations of this residue to alanine in EcoDMT showed a significant impact in proton co-transport, although they did not impair the transporter's metal binding capabilities [28]. This residue is located in the 6a-6b region of the binding site, among the most conserved residues, by being histidine the most common amino acid for that particular position (ConSurf rank 9), as shown in the sequence logo of the segment (Fig. 5c). Overall, both of the aforementioned residues

participating in proton co-transport in EcoDMT are not only identical in SLC11A1 but also highly conserved; this would indicate that they could be part of a similar transportation mechanism.

A potential structural mechanism for divalent metal transport in caprine and bovine SLC11A1

The structural mechanism through which SLC11A1 and other mammalian NRAMPs conduct divalent metal ion transport is currently unknown. Previous experimental work involving cysteine substitutions, performed in combination with determining the DraNrap crystal structure, has proposed a potential mechanism through which DraNrap may perform ion transport [26]. This metal permeation pathway primarily involves the coordinated movement of two structurally distinct regions, namely, the N-terminus, TM 1a and intracellular 6–7 loop, which control the opening and closing of the cytoplasmic side, and the extracellular 7–8 loop, which opens and closes the extracellular side. Additional mutagenesis experiments on human DMT1/NRAMP2, performed in the same study showed comparable results to the bacterial homolog, indicating that DMT1 may follow a similar structural pattern [26]. However, it is not clear whether such a behavior is global to all NRAMPs or restricted to specific members of the family. The results presented in the current study propose that SLC11A1 may follow a similar structural and dynamic pattern, as the same segments differentiate significantly between the proposed outward and inward-facing states, with their structural differences retained throughout the performed MD simulations. Most importantly, the collective motions of all tested SLC11A1 states sampled by EDA (Fig. 4) showed a characteristic movement of the extracellular 7–8 loop in agreement with the mechanism proposed for DraNrap, that signifies the importance of this region for SLC11A1's NRAMP activity, ion binding and transport.

The study of collective motions also revealed the movement of the intracellular 6–7 loop connecting 6b with transmembrane segment 7 (TM7). Given the fact that increased participation of the 6–7 loop was only observed for the inward-facing and ion-bound models—albeit with variable participation in each state - it could be speculated that movements of this particular segment may contribute to stabilizing the inward-facing state and, potentially, ion transportation. Parallelism between NRAMP's TM7 and the equivalent region of the LeuT transporter, sharing a similar fold with NRAMPs, has been also found to destabilize the inward-facing state when mutated [88]. It remains to be seen whether this segment is indeed important for SLC11A1 function.

Finally, the ion-binding capabilities of the proposed models were tested through MD simulations with divalent ions in the proposed ion-binding site, coupled with FEP

calculations on the strength of the interactions when either Fe^{2+} or Mn^{2+} is present. The MD simulations showed that specific residues contribute to protein-metal interactions in SLC11A1, while the residue conservation analysis showed that they are highly invariable. Among these residues, Asp-71 and Met-250 seem to be most central in the formation of protein-metal interactions, with the negatively charged aspartate stabilizing the interface, while methionine serves as the binding site's selectivity determinant towards different ions. At the same time, FEP calculations displayed an increased preference of the models for interacting with Mn^{2+} over Fe^{2+} that is in line with the experimental evidence [14]; this would be a further validation of the proposed SLC11A1 models, particularly for the nature of the ion-binding sites.

Conclusions

In this study we attempted to determine the structural characteristics of SLC11A1 in *Capra hircus* and *Bos taurus* through computational means, expanding our previous molecular experimental work in the caprine *slc11a1* gene to the protein structure level. We proposed a potential model for the SLC11A1 transmembrane domain in both the outward-facing and the inward-facing states and further investigated the structural and dynamic differences between these states. This work has revealed a number of significant collective motions corresponding to regions with experimentally verified interest for NRAMP function which, combined with sequence conservation analysis, gives insight to the metal binding and an explanation of the two-way cation binding mechanism at the binding site. Although the link(s) between the collective motions and the SLC11A1 ion transport mechanism are not entirely clear, there is an agreement between our results and experimental evidence that supports the prediction of a potential mechanism through which the various features of SLC11A1 ion transport may be regulated; and that is our next investigation line ahead in order to establish these connections.

To the best of our knowledge, this is the first study to attempt investigating the structural and dynamic features of SLC11A1 in ruminant species, as well as one of the very few studies to investigate the structure of metazoan NRAMPs in general. While additional data are necessary to gain detailed understanding of the mechanism of SLC11 Me^{2+} transport, our results could act as a referral work for structural and functional features of SLC11A1 and pave the way to foster further research on elucidating similar structures among diverse species and characterize the function of similar members in the SLC11 family. Overall, our findings provide a previously unavailable structural framework for caprine and bovine SLC11A1 that could help to conduct further research on these sequences at the protein level. Given the

importance of SLC11A1 in mammalian physiology in general, characterized as top affiliated gene (malacards) for the susceptibility of ruminant paratuberculosis as well as human diseases such as, Crohn's, *Mycobacterium tuberculosis*, etc., the results of this study could be used to initiate multifaceted studies aimed to an, in depth, analysis of the structural and functional features of SLC11A1 and the means through which is implicated in disease.

Acknowledgements We thank the authors from both the Agricultural University of Athens and the National and Kapodistrian University of Athens for the conception, enthusiasm and continuous support of the work. We would also like to thank the anonymous reviewers for their comments and the associate editor for their proper handling of the manuscript. This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility—ARIS under project IDs “PR002041-S.C.S.M.P.” and “PR004006-BioMemPro-MD”.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

- Dekkers JC, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet* 3(1):22–32. <https://doi.org/10.1038/nrg701>
- Sechi LA, Dow CT (2015) *Mycobacterium avium* ss. paratuberculosis zoonosis—the hundred year war - Beyond Crohn's Disease. *Front Immunol* 6:96. <https://doi.org/10.3389/fimmu.2015.00096>
- Shariat N, Dudley EG (2008) CRISPRs: molecular signatures used for pathogen subtyping. *Appl Environ Microbiol* 80(2):430–439. <https://doi.org/10.1128/AEM.02790-13>
- Walzl G, Ronacher K, Hanekom W, Scriba TJ, Zumla A (2011) Immunological biomarkers of tuberculosis. *Nat Rev Immunol* 11(5):343–354. <https://doi.org/10.1038/nri2960>
- Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV (1998) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N Engl J Med* 338(10):640–644. <https://doi.org/10.1056/NEJM199803053381002>
- Paccagnini D, Sieswerda L, Rosu V, Masala S, Pacifico A, Gazouli M, Ikononopoulos J, Ahmed N, Zanetti S, Sechi LA (2009) Linking chronic infection and autoimmune diseases: *Mycobacterium avium* subspecies paratuberculosis, SLC11A1 polymorphisms and type-1 diabetes mellitus. *PLoS ONE* 4(9):e7109. <https://doi.org/10.1371/journal.pone.0007109>
- Ruiz-Larrañaga O (2009) Identification of single nucleotide polymorphisms in the bovine solute carrier family 11 member 1 (SLC11A1) gene and their association with infection by *Mycobacterium avium* subspecies paratuberculosis. *J Dairy Sci* 93:1713–1721
- Sophie M, Hameed A, Muneer A, Samdani AJ, Saleem S, Azhar A (2017) SLC11A1 polymorphisms and host susceptibility to cutaneous leishmaniasis in Pakistan. *Parasit Vectors* 10(1):12. <https://doi.org/10.1186/s13071-016-1934-2>
- Stewart LC, Day AS, Pearson J, Barclay ML, Gearry RB, Roberts RL, Bentley RW (2010) SLC11A1 polymorphisms in inflammatory bowel disease and *Mycobacterium avium* subspecies paratuberculosis status. *World J Gastroenterol* 16(45):5727–5731
- Korou LM, Liandris E, Gazouli M, Ikononopoulos J (2010) Investigation of the association of the SLC11A1 gene with resistance/sensitivity of goats (*Capra hircus*) to paratuberculosis. *Vet Microbiol* 144(3–4):353–358. <https://doi.org/10.1016/j.vetmic.2010.01.009>
- Taka S, Gazouli M, Sotirakoglou K, Liandris E, Andreadou M, Triantaphyllopoulos K, Ikononopoulos J (2015) Functional analysis of 3'UTR polymorphisms in the caprine SLC11A1 gene and its association with the *Mycobacterium avium* subsp. paratuberculosis infection. *Vet Immunol Immunopathol* 167(1–2):75–79. <https://doi.org/10.1016/j.vetimm.2015.06.004>
- Taka S, Liandris E, Gazouli M, Sotirakoglou K, Theodoropoulos G, Bountouri M, Andreadou M, Ikononopoulos J (2013) In vitro expression of the SLC11A1 gene in goat monocyte-derived macrophages challenged with *Mycobacterium avium* subsp paratuberculosis. *Infect Genet Evol* 17:8–15. <https://doi.org/10.1016/j.meegid.2013.03.033>
- Vacca GM, Pazzola M, Pisano C, Carcangiu V, Diaz ML, Nieddu M, Robledo R, Mezzanotte R, Dettori ML (2011) Chromosomal localisation and genetic variation of the SLC11A1 gene in goats (*Capra hircus*). *Vet J* 190(1):60–65. <https://doi.org/10.1016/j.tvjl.2010.09.028>
- Forbes JR, Gros P (2003) Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane. *Blood* 102(5):1884–1892. <https://doi.org/10.1182/blood-2003-02-0425>
- Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, Sanchez F, Skamene E, Schurr E, Gros P (1994) Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. *J Exp Med* 180(5):1741–1752
- Forbes JR, Gros P (2001) Divalent-metal transport by NRAMP proteins at the interface of host-pathogen interactions. *Trends Microbiol* 9(8):397–403
- Vidal SM, Pinner E, Lepage P, Gauthier S, Gros P (1996) Natural resistance to intracellular infections: Nramp1 encodes a membrane phosphoglycoprotein absent in macrophages from susceptible (Nramp1 D169) mouse strains. *J Immunol* 157(8):3559–3568
- Cellier MF (2012) Nramp: from sequence to structure and mechanism of divalent metal import. *Curr Top Membr* 69:249–293. <https://doi.org/10.1016/B978-0-12-394390-3.00010-0>
- Techau ME, Valdez-Taubas J, Popoff JF, Francis R, Seaman M, Blackwell JM (2007) Evolution of differences in transport function in Slc11a family members. *J Biol Chem* 282(49):35646–35656. <https://doi.org/10.1074/jbc.M707057200>
- Howitt J, Putz U, Lackovic J, Doan A, Dorstyn L, Cheng H, Yang B, Chan-Ling T, Silke J, Kumar S, Tan SS (2009) Divalent metal transporter 1 (DMT1) regulation by Ndfip1 prevents metal toxicity in human neurons. *Proc Natl Acad Sci USA* 106(36):15489–15494. <https://doi.org/10.1073/pnas.0904880106>
- Soe-Lin S, Apte SS, Mikhael MR, Kayembe LK, Nie G, Ponka P (2010) Both Nramp1 and DMT1 are necessary for efficient macrophage iron recycling. *Exp Hematol* 38(8):609–617. <https://doi.org/10.1016/j.exphem.2010.04.003>
- Roupie V, Rosseels V, Piersoel V, Zinniel DK, Barletta RG, Huygen K (2008) Genetic resistance of mice to *Mycobacterium paratuberculosis* is influenced by Slc11a1 at the early but not at the late stage of infection. *Infection immunity* 76(5):2099–2105. <https://doi.org/10.1128/IAI.01137-07>
- Stienstra Y, van der Werf TS, Oosterom E, Nolte IM, van der Graaf WT, Etuafu S, Raghunathan PL, Whitney EA, Ampadu EO, Asamo K, Klutse EY, te Meerman GJ, Tappero JW, Ashford DA, van der Steege G (2006) Susceptibility to Buruli ulcer is associated with the SLC11A1 (NRAMP1) D543N polymorphism. *Genes Immun* 7(3):185–189. <https://doi.org/10.1038/sj.gene.6364281>

24. White JK, Stewart A, Popoff JF, Wilson S, Blackwell JM (2004) Incomplete glycosylation and defective intracellular targeting of mutant solute carrier family 11 member 1 (SLC11A1). *Biochem J* 382(Pt 3):811–819. <https://doi.org/10.1042/BJ20040808>
25. Wessling-Resnick M (2015) Nramp1 and other transporters involved in metal withholding during infection. *J Biol Chem* 290(31):18984–18990. <https://doi.org/10.1074/jbc.R115.643973>
26. Bozzi AT, Bane LB, Weihofen WA, Singharoy A, Guillen ER, Ploegh HL, Schulten K, Gaudet R (2016) Crystal structure and conformational change mechanism of a bacterial Nramp-family divalent metal transporter. *Structure* 24(12):2102–2114. <https://doi.org/10.1016/j.str.2016.09.017>
27. Ehrnstorfer IA, Geertsma ER, Pardon E, Steyaert J, Dutzler R (2014) Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport. *Nat Struct Mol Biol* 21(11):990–996. <https://doi.org/10.1038/nsmb.2904>
28. Ehrnstorfer IA, Manatschal C, Arnold FM, Laederach J, Dutzler R (2017) Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family. *Nat Commun* 8:14033. <https://doi.org/10.1038/ncomms14033>
29. UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45(D1):D158–D169. <https://doi.org/10.1093/nar/gkw1099>
30. Montalbetti N, Simonin A, Kovacs G, Hediger MA (2013) Mammalian iron transporters: families SLC11 and SLC40. *Mol Aspects Med* 34(2–3):270–287. <https://doi.org/10.1016/j.mam.2013.01.002>
31. Martinez R, Dunner S, Barrera G, Canon J (2008) Novel variants within the coding regions of the SLC11A1 gene identified in *Bos taurus* and *Bos indicus* breeds. *J Anim Breed Genet* 125(1):57–62. <https://doi.org/10.1111/j.1439-0388.2007.00690.x>
32. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8(10):785–786. <https://doi.org/10.1038/nmeth.1701>
33. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552
34. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182
35. Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35(Web Server issue):W429–W432. <https://doi.org/10.1093/nar/gkm256>
36. Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* 24(15):1662–1668. <https://doi.org/10.1093/bioinformatics/btn221>
37. Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* 23(5):538–544. <https://doi.org/10.1093/bioinformatics/btl677>
38. Tsaousis GN, Bagos PG, Hamodrakas SJ (2014) HMMpTM: improving transmembrane protein topology prediction using phosphorylation and glycosylation site prediction. *Biochim Biophys Acta* 1844(2):316–322. <https://doi.org/10.1016/j.bbapa.2013.11.001>
39. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4(6):1633–1649. <https://doi.org/10.1002/pmic.200300771>
40. Gupta R, Brunak S (2002) Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 2002:310–322
41. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34(Web Server issue):W362–W365. <https://doi.org/10.1093/nar/gkl124>
42. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res* 41(Web Server issue):W349–W357. <https://doi.org/10.1093/nar/gkt381>
43. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279–D285. <https://doi.org/10.1093/nar/gkv1344>
44. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7(10):e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
45. Omasits U, Ahrens CH, Muller S, Wollscheid B (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* 30(6):884–886. <https://doi.org/10.1093/bioinformatics/btt607>
46. Floden EW, Tommaso PD, Chatzou M, Magis C, Notredame C, Chang JM (2016) PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases. *Nucleic Acids Res* 44(W1):W339–W343. <https://doi.org/10.1093/nar/gkw300>
47. Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079:105–116. https://doi.org/10.1007/978-1-62703-646-7_6
48. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24(8):1596–1599. <https://doi.org/10.1093/molbev/msm092>
49. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
50. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36(Web Server issue):W465–W469. <https://doi.org/10.1093/nar/gkn180>
51. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754–755
52. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242
53. Webb B, Sali A (2017) Protein structure modeling with MODELLER. *Methods Mol Biol* 1654:39–54. https://doi.org/10.1007/978-1-4939-7231-9_4
54. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612. <https://doi.org/10.1002/jcc.20084>
55. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12(4):345–364. <https://doi.org/10.1002/prot.340120407>
56. Studer G, Biasini M, Schwede T (2014) Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* 30(17):i505–i511. <https://doi.org/10.1093/bioinformatics/btu457>
57. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19(2):311–312

58. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Gr* 14(1):33–38, 27–38
59. Schrodinger LLC (2010) The PyMOL molecular graphics system, Version 1.7
60. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
61. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 29(11):1859–1865. <https://doi.org/10.1002/jcc.20945>
62. Wu EL, Cheng X, Jo S, Rui H, Song KC, Davila-Contreras EM, Qi Y, Lee J, Monje-Galvan V, Venable RM, Klauda JB, Im W (2014) CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J Comput Chem* 35(27):1997–2004. <https://doi.org/10.1002/jcc.23702>
63. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* 8(9):3257–3273. <https://doi.org/10.1021/ct300400x>
64. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* 14(1):71–73. <https://doi.org/10.1038/nmeth.4067>
65. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, Vorobyov I, MacKerell AD Jr, Pastor RW (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* 114(23):7830–7843. <https://doi.org/10.1021/jp101759q>
66. Li P, Roberts BP, Chakravorty DK, Merz KM Jr (2013) Rational design of particle mesh Ewald compatible Lennard-Jones parameters for + 2 metal cations in explicit solvent. *J Chem Theory Comput* 9(6):2733–2748. <https://doi.org/10.1021/ct400146w>
67. Li P, Song LF, Merz KM Jr (2015) Parameterization of highly charged metal ions using the 12–6–4 LJ-type nonbonded model in explicit water. *J Phys Chem B* 119(3):883–895. <https://doi.org/10.1021/jp505875v>
68. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684–3690. <https://doi.org/10.1063/1.448118>
69. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol Phys* 52(2):255–268. <https://doi.org/10.1080/00268978400101201>
70. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52(12):7182–7190. <https://doi.org/10.1063/1.328693> doi
71. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. *Proteins* 17(4):412–425. <https://doi.org/10.1002/prot.340170408>
72. Klimovich PV, Shirts MR, Mobley DL (2015) Guidelines for the analysis of free energy calculations. *J Comput Aided Mol Des* 29(5):397–411. <https://doi.org/10.1007/s10822-015-9840-9>
73. Knight JL, Brooks CL (2009) Lambda-dynamics free energy simulation methods. *J Comput Chem* 30(11):1692–1700. <https://doi.org/10.1002/jcc.21295>
74. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *J Phys Chem B* 114(32):10235–10253. <https://doi.org/10.1021/jp102971x>
75. Caplan DA, Subbotina JO, Noskov SY (2008) Molecular mechanism of ion-ion and ion-substrate coupling in the Na⁺-dependent leucine transporter LeuT. *Biophys J* 95(10):4613–4621. <https://doi.org/10.1529/biophysj.108.139741>
76. Gilson MK, Given JA, Bush BL, McCammon JA (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys J* 72(3):1047–1069. [https://doi.org/10.1016/S0006-3495\(97\)78756-3](https://doi.org/10.1016/S0006-3495(97)78756-3)
77. Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44(W1):W344–W350. <https://doi.org/10.1093/nar/gkw408>
78. Liu Y, Gierasch LM, Bahar I (2010) Role of Hsp70 ATPase domain intrinsic dynamics and sequence evolution in enabling its functional interactions with NEFs. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1000931>
79. Smith TF, Waterman MS (1981) Identification of common molecular subsequences. *J Mol Biol* 147(1):195–197
80. Bakan A, Dutta A, Mao W, Liu Y, Chennubhotla C, Lezon TR, Bahar I (2014) Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* 30(18):2681–2683. <https://doi.org/10.1093/bioinformatics/btu336>
81. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190. <https://doi.org/10.1101/gr.849004>
82. Ikai A (1980) Thermostability and aliphatic index of globular proteins. *J Biochem* 88(6):1895–1898
83. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157(1):105–132
84. Sirota FL, Maurer-Stroh S, Eisenhaber B, Eisenhaber F (2015) Single-residue posttranslational modification sites at the N-terminus, C-terminus or in-between: To be or not to be exposed for enzyme access. *Proteomics* 15(14):2525–2546. <https://doi.org/10.1002/psmic.201400633>
85. Xu YZ, Thuraisingam T, Kanagaratham C, Tao S, Radzioch D (2018) c-Src kinase is involved in the tyrosine phosphorylation and activity of SLC11A1 in differentiating macrophages. *PLoS ONE* 13(5):e0196230. <https://doi.org/10.1371/journal.pone.0196230>
86. Moharir A, Peck SH, Budden T, Lee SY (2013) The role of N-glycosylation in folding, trafficking, and functionality of lysosomal protein CLN5. *PLoS ONE* 8(9):e74299. <https://doi.org/10.1371/journal.pone.0074299>
87. Krishnamurthy H, Gouaux E (2012) X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. *Nature* 481:469. <https://doi.org/10.1038/nature10737>
88. Kazmier K, Sharma S, Quick M, Islam SM, Roux B, Weinstein H, Javitch JA, McHaourab HS (2014) Conformational dynamics of ligand-dependent alternating access in LeuT. *Nat Struct Mol Biol* 21(5):472–479. <https://doi.org/10.1038/nsmb.2816>
89. Bozzi AT, Bane LB, Weihofen WA, McCabe AL, Singharoy A, Chipot CJ, Schulten K, Gaudet R (2016) Conserved methionine dictates substrate preference in Nrap-family divalent metal transporters. *Proc Natl Acad Sci USA* 113(37):10310–10315. <https://doi.org/10.1073/pnas.1607734113>

The Gram-Negative Outer Membrane Modeler: Automated Building of Lipopolysaccharide-rich Bacterial Outer Membranes in Four Force Fields

Fotis A. Baltoumas , Stavros J. Hamodrakas, and Vassiliki A. Iconomidou *

Outer membranes are a crucial component of Gram-negative bacteria, containing standard lipids in their inner leaflet, lipopolysaccharides (LPSs) in their outer leaflet, and transmembrane β -barrels known as outer membrane proteins (OMPs). OMPs regulate functions such as substrate transport and cell movement, while LPSs act as a protective barrier for bacteria and can cause toxic reactions in humans. However, the experimental study of outer membranes is challenging. Molecular dynamics simulations are often used for the computational study of membrane systems, but the preparation of complex, LPS-rich outer membranes is not straightforward. The Gram-Negative Outer Membrane

Modeler (GNOMM) is an automated pipeline for preparing simulation systems of OMPs embedded in LPS-containing membranes in four different force fields. Given the physiological and clinical importance of outer membranes and their components, GNOMM can be a useful tool in the study of their structure, function, and implications in diseases. GNOMM is available at <http://bioinformatics.biol.uoa.gr/GNOMM>. © 2019 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.25823

Introduction

The cell envelope of Gram-negative bacteria is a complex structure, composed by an inner and an outer membrane. The inner membrane is a mostly symmetric lipid bilayer that consists of standard lipids and contains α -helical transmembrane proteins. Conversely, the outer membrane (OM) is a highly asymmetric lipid bilayer, containing standard phospholipids in its inner leaflet and lipopolysaccharides (LPSs) in its outer leaflet.^[1] Proteins of the OM are mostly transmembrane β -barrels that often act as porins or ion channels, commonly known as outer membrane proteins (OMPs).^[2]

LPSs form the majority of the OM's outer leaflet.^[1] They are composed of three main components, namely, a lipid moiety (Lipid A), a short, oligosaccharide chain (core), and a highly variable polysaccharide of repeating units (O-antigen)^[3] (Fig. 1a). Lipid A is composed of a phosphorylated glycosamine dimer, bound to acyl chains that serve as lipid tails. The number of lipid tails and phosphates can be different among LPSs (Fig. S1).^[4] The oligosaccharide core is covalently linked to Lipid A and is composed of a few copies of rare sugars such as 3-deoxy-D-manno-2-octulosonic acid (kdo) or L-glycero-D-manno heptose (Hep), as well as standard hexoses such as D-glucose or D-galactose. Similar to Lipid A, the core is also phosphorylated. Finally, the O-antigen is formed by multiple glycan repeats. Its composition is highly variable and often depends on the serological environment.^[5] Because both Lipid A and the core are phosphorylated, LPSs have a high negative charge and require divalent cations such as Ca^{2+} or Mg^{2+} to form stable membrane structures. This ionic interaction is a major determinant in maintaining the OM's integrity.^[4]

The OM and its components are crucial for Gram-negative cells. OMPs can regulate several functions such as substrate

transport, bacterial conjugation, host infection, and cell movement.^[2] As such, OMPs are actively studied as potential drug targets.^[6–8] Interestingly, the proper function of some OMPs has been found to depend on their interactions with neighboring LPSs.^[9,10] The LPS layer of the OM itself is also functionally relevant; it serves as a protective barrier, preventing permeability by potentially toxic substances such as antibiotics, antimicrobial peptides, and bile salts.^[11] Finally, the lipid A component of LPSs, also known as Endotoxin, contributes to the toxicity of Gram-negative bacteria causing fever and septic shock.^[12] As a result, LPSs from several species have been implicated in a number of pathological conditions, ranging from simple cases of food poisoning and urinary tract infections to severe diseases such as meningitis and the septicemic plague.^[13–15] Overall, the study of LPSs, including their organization in the OM, their interactions with OMPs and their implication in immune responses is vital for the elucidation of bacterial cell mechanisms and, ultimately, the design of novel, more potent antimicrobial drugs.

Despite their importance, the experimental study of LPSs and OMPs is challenging. Molecular Dynamics (MD) simulations are a useful tool for modeling the structure and dynamics of biological macromolecules and have been routinely used in the study of biological membranes and membrane proteins, including OMPs.^[16,17] However, while the construction of membrane models with phospholipids has become a rudimentary process, designing MD simulation systems for OMs is not as straightforward,

[a] F. A. Baltoumas, S. J. Hamodrakas, V. A. Iconomidou
Section of Cell Biology and Biophysics, Department of Biology, School of Sciences, National and Kapodistrian University of Athens, Panepistimiopolis, 15701, Athens, Greece
E-mail: veconom@biol.uoa.gr

© 2019 Wiley Periodicals, Inc.

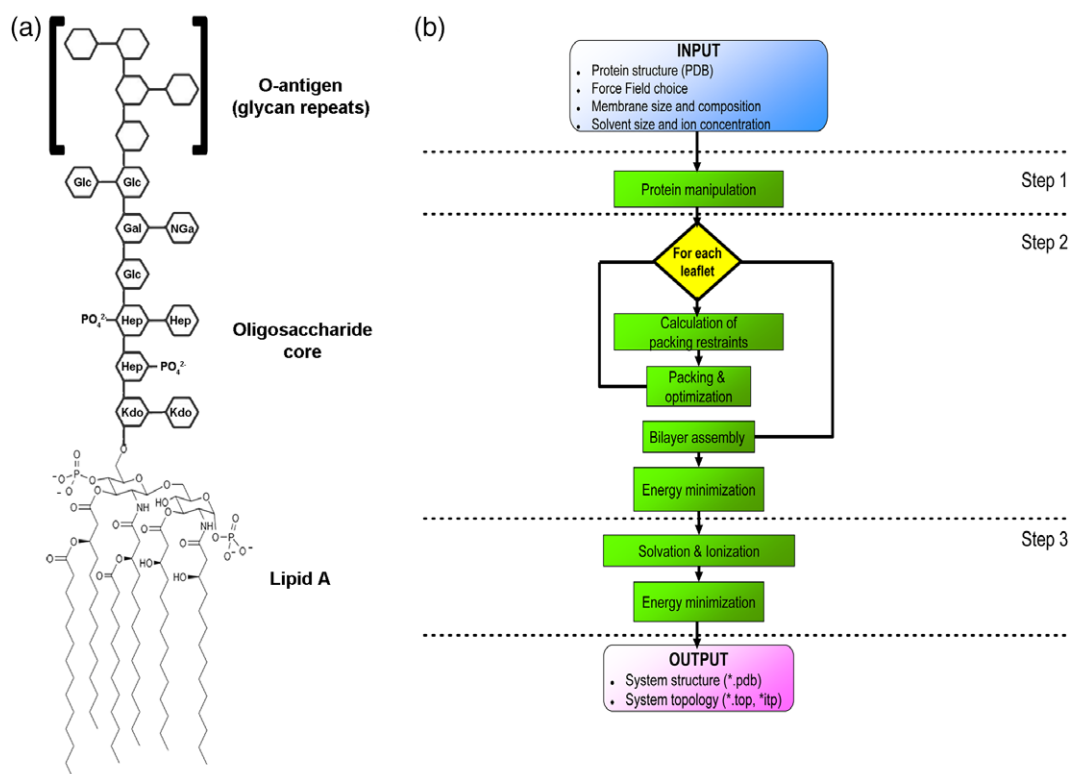


Figure 1. a) Example structure of a lipopolysaccharide (LPS). The *Escherichia coli* lipid A and R3 oligosaccharide core are depicted. Lipid A is presented as a chemical structure, while for the core and O-antigen a simplified representation is used, due to their complexity. The carbohydrates commonly found in the core are labeled: 3-deoxy-D-manno-2-octulosonic acid (Kdo), L-glycero-D-manno heptose (Hep), D-glucose (Glc), D-galactose (Gal), and N-acetyl-D-glucosamine (NGa). Glycan repeats forming the O-antigen are indicated by brackets; their composition and number may vary, depending on the bacterial species and environmental conditions. b) Flowchart of the GNOMM modeling procedure. A detailed description is given in the Methods section. [Color figure can be viewed at wileyonlinelibrary.com]

mostly due to the increased complexity of LPSs compared to standard lipids. Currently, the only fully automated solutions to this problem are offered by the CHARMM-GUI web service.^[18–20] The recently published CHARMM-GUI LPS modeler supports the creation of LPSs from various species and their insertion in lipid bilayers through the service's *Membrane Builder*.^[20] However, the LPS modeler is limited to the CHARMM all-atom force field.^[21] Another option is the CHARMM-GUI Martini Maker,^[19] which currently offers two *Escherichia coli* LPS molecules (Re and Ra LPS) for the MARTINI Coarse-Grained force field but does not, at present, support any other species. Some glycolipid and glycan modeling tools can also be used to model LPS molecules, such as doGlycans^[22] for the AMBER/GLYCAM^[23] and OPLS^[24] force fields; however, these methods are not able to construct LPS-containing bilayers, either homogeneous or mixed. As such, apart from CHARMM-GUI, there are very limited options to automatically create LPS-containing membranes with force fields or molecules outside the aforementioned systems (e.g., GROMOS, or MARTINI with LPSs from species other than *E. coli*), even though valid Lipid A and LPS parameters exist.^[4,25–28] Instead, the literature shows that LPS-containing simulation systems in such cases are, for the most part, built by hand,^[4,25–29] a process that can be time consuming and difficult, particularly for inexperienced users.

In this work we present the Gram-Negative Outer Membrane Modeler (GNOMM), a publicly available, automated pipeline for

the creation of complex, LPS-rich simulation systems. GNOMM supports the creation of LPS-containing membranes in four different force fields representing different levels of resolution, namely, CHARMM^[30] (All-Atom, AA), GROMOS^[31] (United-Atom, UA), MARTINI^[32] (Coarse-Grained, CG), and PACE^[33] (hybrid UA/CG). Users can insert structures of membrane proteins in lipid bilayers containing LPSs from different bacterial species, as well as other lipids commonly found in OMs, and produce simulation systems for the freely available GROMACS^[34,35] biomolecular simulation program.

Methods

The GNOMM modeling procedure

The automated modeling procedure followed by GNOMM involves three main steps (Fig. 1). In the first step, users select the type of system they want to model and choose a force field representation. To model a protein-membrane system, a pre-oriented Protein Data Bank (PDB) structure with respect to the membrane plane is required, that is, a structure with its z-axis parallel to the membrane normal. Such structures can be retrieved from databases such as OPM^[36] or PDB-TM^[37] or generated through tools like the PPM server.^[36] Both single protein chains and multichain protein complexes can be used as input. Upon submission, protein coordinates are used to generate the

relevant topology and parameters, in accordance with the chosen force field. If the users elect to build a membrane-only system, the aforementioned step is skipped, and only the force field of the system is chosen.

Step 2 involves modeling the membrane around the proteins. The user defines the X and Y dimensions of the membrane and selects its composition, namely, the number and ratios of different lipid types in each leaflet. It is important to note that GNOMM allows the definition of different X and Y values, enabling the creation of nonequidimensional membranes. Using the information provided by the user, GNOMM builds the lipid bilayer by setting its center at $(x, y, z) = (0, 0, 0)$ and defining two rectangular boxes representing its two membrane leaflets. In each box the space occupied by the protein(s) is determined and position restraints are calculated to preclude the positioning of any lipids in that space. The rest of the box is then filled with lipids, distributed randomly over the available space. If no protein coordinates exist in the leaflet (such as in the case of membrane-only systems, as well as peripheral proteins that bind to only one side of the membrane) the entire leaflet space is used to position the lipids.

To preclude the generation of any abnormal orientations, position restraints are used to constrain each lipid's head group in the area corresponding to the leaflet's head-group region and its hydrocarbon tail ends to the hydrophobic interior of the membrane. Multiple cycles of molecule repositioning and rigid-body optimization are performed to ensure the best possible packing for the lipids. After the two leaflets are built, the system is assembled and subjected to a short energy minimization to remove any steric clashes. If any of the leaflets contains LPSs, Ca^{2+} ions are packed around the phosphates of Lipid A and the oligosaccharide core to neutralize their charge; this is performed because divalent cations have been found to contribute to the stabilization of the OM, by interacting with the LPS phosphates.^[4] In the final step, the user chooses whether the system will be solvated and ionized. If yes, the type of water model,^[38–40] the size of the solvent layer on each membrane size, the type of ions (NaCl, KCl or CaCl_2) and their concentration are determined. A short energy minimization of the system follows to alleviate steric clashes and any unphysical interactions.

The output of GNOMM consists of the simulation system structure in PDB format, accompanied by the system's force field topology and parameters in the GROMACS (TOP/ITP) format. Force field files and example configuration parameters are also available for download through the GNOMM website.

Implementation

Python and the BioPython module^[41] are utilized for the majority of structural calculations, including the manipulation of protein coordinates, calculations of system size and stoichiometry, and the generation of packing restraints. Modeling of missing atom coordinates, where necessary, is performed using MODEL-ER.^[42] The LiteMol PDB viewer is used to inspect structures.^[43] Force field operations are performed with GROMACS.^[34,35] Lipid positioning, packing, and optimization during the membrane

building stage are performed using the GENCAN local optimization algorithm implemented in PackMol.^[44]

GNOMM has been implemented as a web server and is publicly available through <http://bioinformatics.biol.uoa.gr/GNOMM>. Users can create various types of simulation systems depending on their needs, by choosing the appropriate option from the service's *Submit* page. With the *Protein-Membrane* option, users can upload a protein structure in PDB format and build a membrane around it, while the *Membrane-Only* option allows them to simply construct a lipid bilayer. Finally, users can select the *Pre-Processed PDB and Topology* option, to upload a processed structure and force field topology of their own; this last option can be useful in cases where the simulation system contains elements not available in standard force field parameters, for example, modifications to the protein structure, rationally designed compounds or even inorganic polymer structures like carbon nanotubes or fullerenes.^[26,45]

Lipids and force field parameters

GNOMM offers a selection of 34 different lipids (Table 1, see also Supporting Information Table S1). These include variations of Lipid A from four different species (*E. coli*, *Pseudomonas aeruginosa*, *Helicobacter pylori* and *Neisseria meningitidis*) (Supporting Information Fig. S1), as well as three experimentally studied LPSs, containing Lipid A and the oligosaccharide core (Re and Ra LPS for *E. coli* and PAO1 LPS for *P. aeruginosa*).^[4,5,21,25–27] In addition, a number of standard lipids are also offered, representing the major categories most commonly found in bacteria.^[1,46] Simulation systems can be built for four force fields, namely, CHARMM36^[30], GROMOS 54A7^[31], MARTINI,^[32,47] and PACE.^[33] Information regarding the availability of LPSs and standard lipids for each force field is given in Supporting Information Table S1. Details on the implementation of each force field are given in the Supporting Information.

Table 1. LPS and phospholipid molecules currently available in GNOMM.

Lipopolysaccharides (LPSs)	
<i>Escherichia coli</i>	Lipid A (ECLI), Re LPS (ReMP), Ra LPS (RaMP)
<i>Pseudomonas aeruginosa</i>	penta-acyl Lipid A (PAL5), hexa-acyl Lipid A (PAL6), PAO1 LPS (PAO1)
<i>Helicobacter pylori</i>	Lipid A (HPLI)
<i>Neisseria meningitidis</i>	Lipid A (NMEN)
Phospholipids	
Phosphatidyl-ethanolamines (PEs)	POPE, DOPE, DPPE, DSPE, DMPE, DLPE
Phosphatidyl-cholines (PCs)	POPC, DOPC, DPPC, DSPC, DMPC, DLPC
Phosphatidyl-glycerols (PGs)	POPG, DOPG, DPPG, DSPG, DMPG, DLPG
Phosphatidyl-serines (PSs)	POPS, DOPS, DPPS, DSPS, DMPS, DLPS
Cardiolipins (CDLs)	Mono-anionic cardiolipin (CDL1), Di-anionic cardiolipin (CDL2)

Molecular dynamics simulations

A number of MD simulation systems were prepared with GNOMM and simulated with GROMACS.^[35] A summary of all systems is given in Supporting Information Table S2. Details are given in the Supporting Information.

Results and Discussion

To verify the robustness of the membrane systems produced by GNOMM, a number of MD simulations were prepared and performed. These included model LPS bilayers, mixed membranes containing LPSs and standard lipids, various protein-membrane systems for two well-studied OMPs and a complex of an OMP with a bound ligand. The total simulation time, including AA, UA, CG, and hybrid UA/CG simulations, reaches 35 μ s (Supporting Information Table S2). The results for each type of simulation system are presented in the following sections.

Simulation results and physical properties of LPS-containing lipid bilayers

The validity of the topology and parameters offered by GNOMM was initially tested through MD simulations of uniform Lipid A bilayers (Supporting Information Fig. S2). The simulations were evaluated by measuring two important lipid parameters for the Lipid A component, namely, the average area per lipid (ApL) and hydrophobic thickness (H_T) of the bilayers. As shown by the results (Supporting Information Table S3), ApL and H_T values are consistent among force fields for all tested lipid A variants. Interestingly, the results for *E. coli* Lipid A (ECL1) and *P. aeruginosa* penta-acyl Lipid A (PAL5) ApL values are in agreement with experimental evidence. Specifically, the ECL1 ApL values in all force fields (148.15–154.57 $\text{\AA}^2$) fall within the experimentally determined ranges previously reported in the literature, measured at various temperatures (130–156 $\text{\AA}^2$).^[5,48] Similarly, PAL5 values (139.17–143.55 $\text{\AA}^2$) are in agreement with measurements of Lipid A at the liquid crystalline state (125–142 $\text{\AA}^2$) for *P. aeruginosa* and other related species.^[49] Finally, it should be noted that the simulation results of all Lipid A variants are in agreement with previous simulation evidence for the CHARMM36^[21], MARTINI^[25–27] and GROMOS^[4,28] force fields. Overall, these results show that the Lipid A topologies and simulation systems offered by GNOMM are, in terms of lipid bilayer properties, valid and consistent among different levels of resolution.

Apart from lipid properties, such as ApL and H_T , another criterion for the evaluation of structure and dynamics is the proper conformation of the lipids during simulation. For LPSs this includes two features, namely, the conservation of chirality for the Lipid A hydrophobic tails and the dynamics of the sugars forming the oligosaccharide core. The chirality of all Lipid A variants during simulation was investigated by computing the angle distribution of the improper dihedrals defined around each chiral atom of the molecules (Supporting Information Figs. S1 and S3). As shown by the distributions (Supporting Information Fig. S3), all chiral centers retained their proper

configuration throughout the simulations, with no abnormal chirality changes recorded. As far as the carbohydrate core of LPSs is concerned, it is important that its sugars maintain their chair conformation throughout the simulations. To investigate this, all-atom MD simulations were performed for Ra LPS (RaMP) with the CHARMM36 force field, the results of which were used to analyze the conformation of the core sugars, by applying the Cremer-Pople^[50] definitions for carbohydrate ring puckering (see Supporting Information Methods section for details). As shown by the measurements (Supporting Information Table S4), all sugars forming the R3 core of RaMP retain their original structural features, as evidenced by the puckering amplitude ($Q \sim 0.60$) and θ angle values ($\theta = 0\text{--}12^\circ$), which are in ranges commonly associated with the carbohydrate chair conformation.^[51] Overall, the investigation of all aforementioned properties shows that LPSs in the membranes generated by GNOMM can maintain their pertinent configuration.

In addition to the model LPS membranes, systems of mixed LPS-phospholipid bilayers were prepared and simulated in the microsecond scale, to approximate the nonsymmetric nature of actual OMs and evaluate the dynamics of LPSs in them. Due to the timescale of the simulations and the complexity of the membranes, all systems were modeled and simulated with the MARTINI CG force field. The simulation systems featured either a pure LPS outer leaflet, composed by Ra LPS (RaMP), or a mixed RaMP:POPE bilayer in a 3:2 ratio, representing the small defects caused by the presence of POPE, often observed in bacterial OMs^[11]; both systems also had a mixed inner leaflet, composed by POPE, POPG, and CDL2 in a 2:2:1 ratio (Fig. 2).

As shown by the simulations, the systems were rather rigid in the outer leaflet for both the pure RaMP and the RaMP:POPE outer leaflets. Contrary to the outer leaflet, all lipid types diffused freely in the inner leaflet, with each lipid species displaying the expected ApL values (Supporting Information Table S5). The behavior of the outer leaflet can be attributed to the extremely slow dynamics of RaMP, due to the presence of the oligosaccharide core, as it has previously been reported in the literature.^[5,19] It should be noted that no phase separation was reported for the lipids of either system; however, this is not surprising, considering both the complexity of RaMP and the timescales in which phase separation is expected to occur,^[52,53] which are much longer than the simulation time of the systems (5 μ s). Overall, this behavior is comparable to previous simulation evidence for mixed, OM-like bilayers conducted with the MARTINI force field.^[19]

The comparison of the pure RaMP and RaMP:POPE “defect” leaflets reveals two differences with potential significance. The ApL value of RaMP (Supporting Information Table S3) is slightly lower in the RaMP:POPE leaflet ($\sim 164.01 \text{ \AA}^2$), compared to pure RaMP ($\sim 175.35 \text{ \AA}^2$). The second observation is a rather significant difference in the average thickness of the membranes ($\sim 4.50 \text{ \AA}$), with the “defect” system having a larger thickness (~ 37.30 over $\sim 32.80 \text{ \AA}$). In both systems, these measurements were observed relatively early in the simulation ($\sim 1.0 \mu$ s), with the membranes retaining the same dynamic behavior onward. These results are consistent with previous simulation studies by Hsu and co-workers,^[26] performed using the same force field

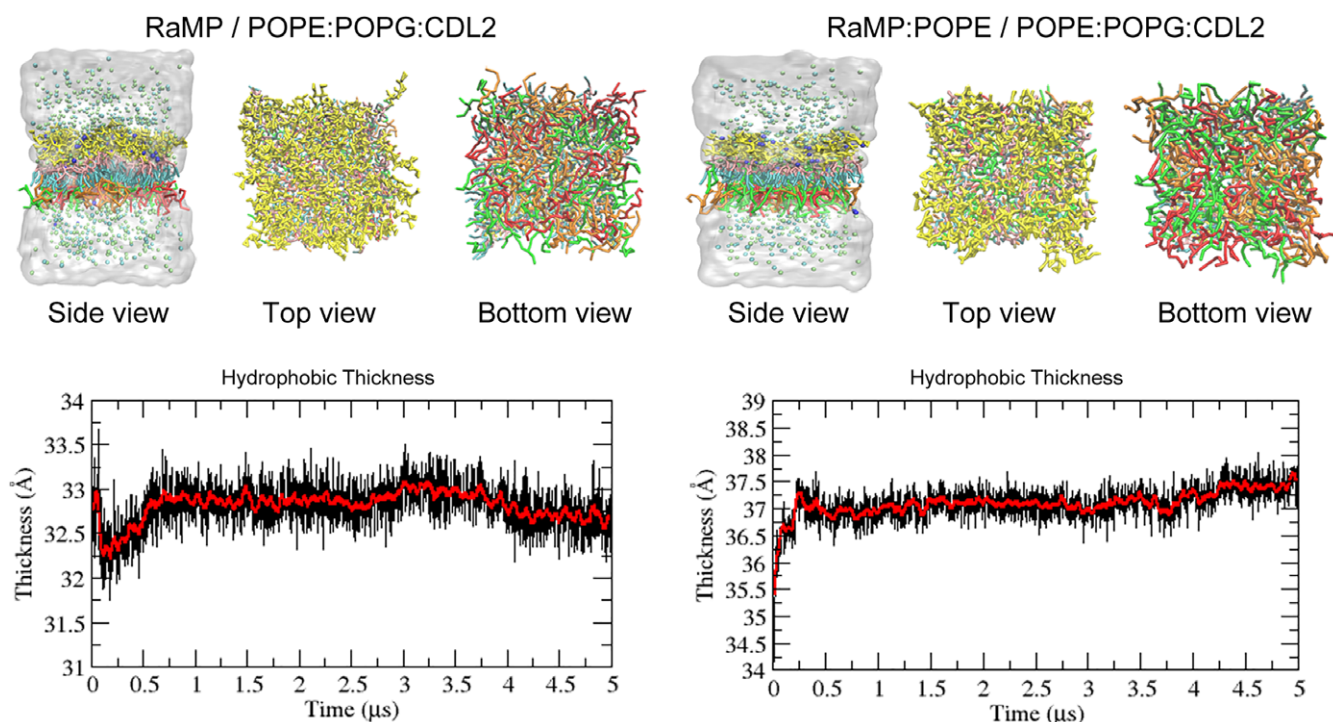


Figure 2. Membrane structures (top) and hydrophobic thickness profiles (bottom) for the two mixed bilayer systems after 5 μs of simulation. The side, top, and bottom views are shown for each system. All lipids are shown in stick representation. Ra LPS (RaMP) is depicted using cyan for the lipid tails, pink for the glucosamine dimer (head-group region), and yellow for the oligosaccharide core. POPE, POPG, and CDL2 are uniformly colored using green, red, and orange, respectively. The solvent is shown as a transparent surface, with ions shown as spheres and colored blue for Ca^{2+} , lime for Na^{+} , and light blue for Cl^{-} . In the top and bottom views the solvent and ions are not rendered for more clarity. In the graphs, the hydrophobic thickness is shown as a black curve, while the running average is shown as a red curve. [Color figure can be viewed at wileyonlinelibrary.com]

model for RaMP, in which a decrease of ApL and increase of thickness were observed for RaMP:POPE “defect” bilayers, compared to pure RaMP bilayers. Furthermore, they are indicative of the disruptions introduced to the LPS layer of the OM, caused by the insertion of phospholipids.

Simulations of OMPs in LPS-containing membranes

To test the integrity of the β -barrel structure in the membranes modeled by GNOMM, MD simulations were initially performed for two well-studied OMPs, namely, OmpA^[54] and OmpF^[55] from *E. coli* (Supporting Information Fig. S4), embedded in model Lipid A/POPE membranes. For OmpA, a monomer was embedded in a Lipid A/POPE bilayer and simulated in all four supported force fields, with each simulation performed for 300 ns. The robustness of the protein was monitored by measuring the root mean square deviation (RMSD) from the initial crystal structure. The β -sheet content of the atomic and near-atomic resolution simulations was also monitored, to evaluate the integrity of the transmembrane β -strands of OmpA. As shown by RMSD measurements, the protein reaches structural equilibrium early (~ 100 – 125 ns) in all force fields and remains relatively stable for the remainder of the simulations (Supporting Information Fig. S4). The values of the final conformations are in acceptable values (~ 3.0 – 4.8 Å) for all force fields. More importantly, the measurement of OmpA's β -sheet content during the simulation indicates stable β -strands, suggesting that its transmembrane β -barrel structure remains intact. In the case

of OmpF, a 1 μs (1000 ns) long simulation of the OmpF homotrimer in a Lipid A/POPE bilayer, modeled with the MARTINI force field was prepared, in order to investigate the performance of the simulation system in longer timescales, as well as the integrity of the OmpF trimer in an LPS-containing bilayer. As shown by the simulation results, the OmpF homotrimer retains a remarkable structural stability throughout the 1 μs run, with RMSD values in the area of ~ 1.5 – 1.8 Å for the proto-mers (Supporting Information Fig. S4).

Having established the stability of OMPs in membranes containing Lipid A, the next step was to investigate the interactions of β -barrels with RaMP, containing both Lipid A and the oligosaccharide core. Toward this end, simulations were performed for OmpF embedded in a RaMP/POPE membrane with the CHARMM force field (Fig. 3a). OmpF was chosen because its interactions with LPSs have previously been investigated. Specifically, both experimental assays^[56] and MD simulations^[57] have revealed that OmpF can form interactions with LPSs at various sites. Two of these sites, located at extracellular loops L1 and L5, correspond to epitopes recognized by monoclonal antibodies. The binding of LPSs at these sites blocks them from being recognized by the antibodies.^[56,57]

As shown by the simulation results, OmpF remained stable throughout the simulation, similar to the systems embedded in membranes with lipid A. Inspection of the system during the simulation revealed that OmpF interacts with both Lipid A and the oligosaccharide core of RaMP. Specifically, the transmembrane segments of the protein and some of the short loops

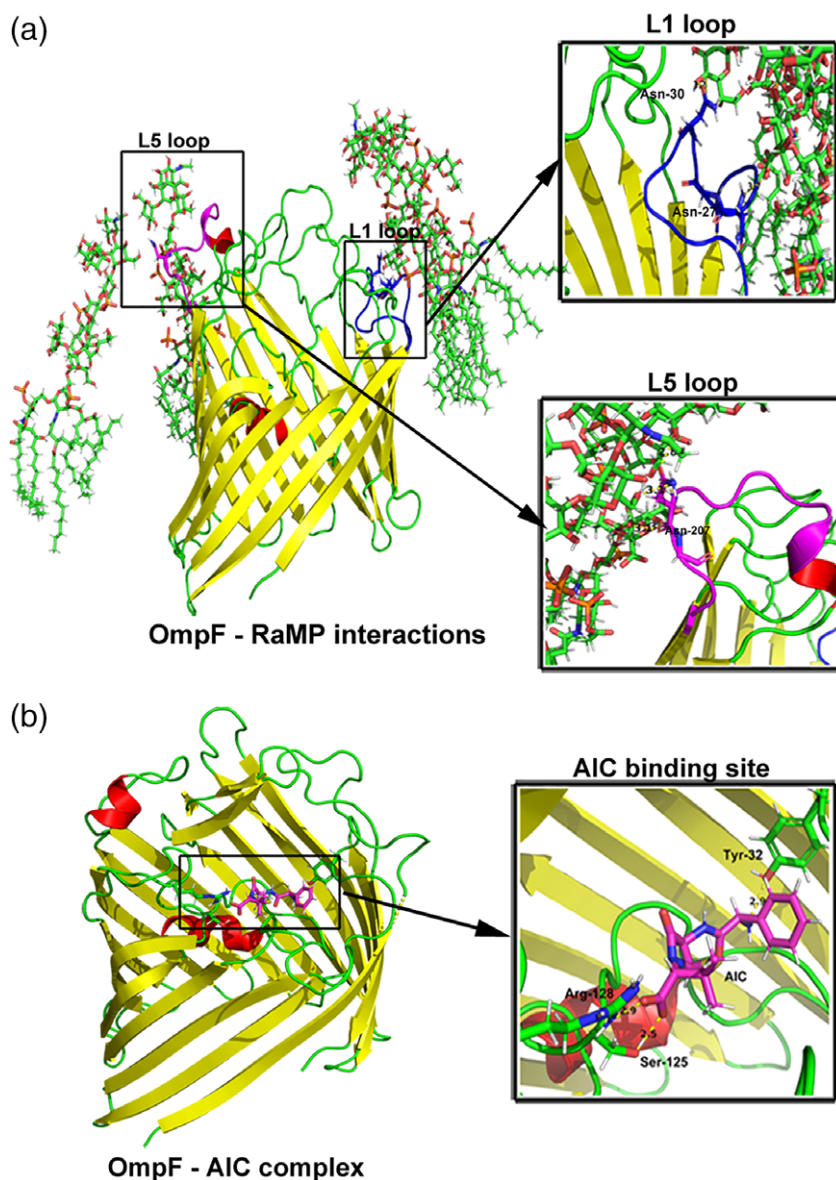


Figure 3. a) Interactions of OmpF with Ra LPS (RaMP). The L1 and L5 loops are colored blue and magenta, respectively. Adjacent RaMP molecules are shown as sticks; the rest of the simulation system is not rendered for more clarity. Close-up views of the L1 and L5 loops are also shown, in which residues Asn-27, Asn-30 (L1), and Asn-207 (L5) that interact with the RaMP core are shown as sticks. b) Snapshot of OmpF – ampicillin (AIC) interactions during simulation. AIC and its adjacent OmpF residues are shown as sticks. A close-up of the AIC binding site is also shown, with AIC-interacting residues Tyr-32, Ser-125, and Arg-128 indicated. [Color figure can be viewed at wileyonlinelibrary.com]

connecting them form contacts with the tails and phosphorylated glucosamine dimer of Lipid A. Conversely, the longer extracellular loops either fluctuate freely in the solvent or form contacts with the sugars of the RaMP core. Impressively, tight OmpF–sugar contacts were observed for the L1 and L5 loops. The inspection of binding at these regions (Fig. 3a) showed that adjacent RaMP molecules “cap” the two sites almost completely, effectively hiding them from the solvent and, therefore, could prevent their recognition by antibodies. Furthermore, the observed OmpF–LPS interactions at these sites involve polar residues (Asn-27, Asn-30, and Asn-207) contacting the hydroxyl groups of core sugars, the same as previously recognized in simulations of OmpF.^[57]

Simulations of OmpF in complex with an antibiotic

In addition to the protein-membrane systems described in the previous section, MD simulations were also performed for OmpF in complex with the antibiotic ampicillin (AIC), a β -lactam

compound. Experimental evidence has shown that AIC acts by binding to the pore cap segment of OmpF, constraining its movement and effectively blocking the flow of substrates through the barrel's pore and has pinpointed a number of OmpF residues that directly influence its binding.^[8] To reproduce these results in the context of a membrane system resembling actual OMs, a simulation system was prepared for the OmpF–AIC complex embedded in a RaMP/POPE membrane, built with the CHARMM force field. The system was built using GROMACS's *Pre-processed PDB and Topology* option for the Builder, applying user-generated ligand topologies for AIC obtained from the CHARMM General Force Field (GenFF).^[58]

As shown by the results (Fig. 3b, Supporting Information Fig. S5), AIC remains tightly bound in its binding site upon OmpF. Throughout the simulation the molecule displayed various fluctuations in the binding site (Supporting Information Fig. S4); however, its position remains, ultimately, unchanged, limiting the flow of water through the pore. Furthermore, OmpF residues Tyr-32, Ser-125, and Arg-128, retained their polar

interactions with AIC, in agreement with previous experimental and computational evidence on the compound's activity.^[8] Overall, the simulation results showed that the OmpF-AIC system remains stable during simulation and displays structurally and functionally relevant interactions.

GNOMM produces stable simulation systems capable of modeling the features of outer membranes

Computational studies in general and MD simulations in particular have become a firmly established, widely used method for investigating the properties of biological membranes and membrane proteins. Still, the progress of membrane simulations has, so far, been mostly limited to the modeling of "standard" phospholipid systems, while more complex membrane structures, including the OM of Gram-negative bacteria, have only recently emerged in the field. While some of the properties of OM components, such as the stability of OMPs, can be partially investigated by using simple lipid bilayers, the accurate sampling of OM and OMP structure and dynamics requires simulation systems that can model the structural complexity of LPSs and their interactions. This calls both for the development of force field topology and parameters for LPSs and for modeling approaches that can readily build such complex systems according to the user's needs. At the same time, the produced simulation systems must be numerically stable and accurately reproduce the dynamics of LPSs and their membranes, both in terms of experimentally measured properties and in terms of their interactions with transmembrane β -barrels.

As shown by the simulations performed in this study, GNOMM can achieve this goal. The lipid properties of the membranes modeled by GNOMM, display an agreement both among different force fields and with experimental evidence.^[5,48,49] At the same time, mixed LPS/phospholipid bilayers effectively model the dynamics of such complex systems as previously reported in the literature.^[26,52,53] With regards to protein-membrane simulations, the systems of widely studied OmpA and OmpF β -barrels showed adequate stability for the proteins and comparable behavior among AA, UA, CG, and hybrid systems. The latter observation is of particular importance, as it shows that simplified models, such as the MARTINI CG and PACE hybrid UA/CG force fields, can reproduce the dynamics observed in more detailed models such as CHARMM. Finally, the simulations of OmpF in RaMP membranes and in complex with ampicillin were able to reproduce previously reported intermolecular contacts, both for OmpF-LPS interactions and for the binding of the antibiotic. While, of course, this consistency can be mainly attributed to the accurate parameterization of the proteins and LPSs in the force fields, it is also dependent upon the simulation system having a valid configuration. Overall, all simulation results indicate that the membrane systems produced by GNOMM are numerically stable, consistent among different levels of resolution and with experimental evidence^[5,43,44] and, therefore, suitable for the simulation of OMs and transmembrane β -barrels in OM-like conditions.

The main strength of GNOMM, compared to other similar services, is the availability of multiple different force field models to choose from. This feature offers the users flexibility in creating

simulation systems in terms of force field choice and level of resolution. In addition, GNOMM is, to our knowledge, the only membrane modeling server for the GROMOS force field that also allows the insertion of proteins in lipid bilayers with Lipid A. As GROMOS is a continuously evolving model,^[4,28] with good performance in modeling the dynamics of β -sheet structures^[31] including β -barrels, we believe that the availability of such a tool for this force field is important. At the same time, the modeling procedure followed by GNOMM offers users additional flexibility in building their simulation systems, including not only the choice of different solvent models for each force field, but also the ability to build nonequidimensional membranes. It is important to note that, with a few notable exceptions (e.g., the MARTINI *Insane*^[59] tool which, however, does not support LPSs), most membrane modeling services, such as CHARMM-GUI, support only the generation of square membrane patches, that is, with identical X and Y values. Conversely, GNOMM allows the generation of nonequidimensional lipid bilayers, enabling users to construct membranes extended along either axis. Although this feature may initially seem trivial, its availability can be important in easily constructing membranes for simulations that require such directionality, such as Steered Molecular Dynamics and Potential of Mean Force calculations for transmembrane protein-protein complexes.^[29,60] Finally, it is important to note that the membrane modeling process of GNOMM is force field-independent; the same procedure can be applied to any pertinent molecular system, regardless of its resolution or complexity, with minimal changes. Therefore, as force field models progress and more LPS parameters become available, the GNOMM pipeline can be easily extended to include them, both in terms of providing more molecules for the four already available force fields and in terms of extending the support to other force field models (e.g., GLYCAM^[61]).

Conclusions

We have presented GNOMM, a novel, automated pipeline for building LPS-containing simulation systems that can represent the outer membranes of Gram-negative bacteria. GNOMM is currently the only available web service capable of building LPS-containing simulation systems in four different force fields, thus enabling a selection of simulation ensembles at various levels of resolution. Given the physiological and clinical significance of Gram-negative bacteria and the rising interest in accurately representing the environment of transmembrane β -barrels during simulation, we expect GNOMM to help ease the process of designing complex LPS-containing simulation systems, by replacing the tedious and often confusing manual membrane modeling processes with a user-friendly, automated procedure that requires minimal manual intervention.

Acknowledgments

We thank the National & Kapodistrian University of Athens for support. This work was supported by computational time granted from the Greek Research & Technology Network (GRNET) in the National HPC facility-ARIS, under project ID "PR004006-BioMemPro-MD." We would like to thank the anonymous

reviewers for their valuable comments, which helped improve the manuscript, as well as the editor for properly handling it.

Keywords: outer membrane · gram-negative bacteria · lipopolysaccharide · β -Barrel · molecular dynamics

How to cite this article: F. A. Baltoumas, S. J. Hamodrakas, V. A. Iconomidou. *J. Comput. Chem.* **2019**, 9999, 1–8. DOI: 10.1002/jcc.25823

 Additional Supporting Information may be found in the online version of this article.

- [1] M. P. Bos, V. Robert, J. Tommassen, *Annu. Rev. Microbiol.* **2007**, 61, 191.
- [2] R. Koebnik, K. P. Locher, P. Van Gelder, *Mol. Microbiol.* **2000**, 37, 239.
- [3] C. R. Raetz, C. Whitfield, *Annu. Rev. Biochem.* **2002**, 71, 635.
- [4] F. J. Pontes, V. H. Rusu, T. A. Soares, R. D. Lins, *Journal of chemical theory and computation* **2012**, 8, 3830.
- [5] E. L. Wu, O. Engstrom, S. Jo, D. Stuhlsatz, M. S. Yeom, J. B. Klauda, G. Widmalm, W. Im, *Biophys. J.* **2013**, 105, 1444.
- [6] M. Ceccarelli, C. Danelon, A. Laio, M. Parrinello, *Biophys. J.* **2004**, 87, 58.
- [7] C. E. James, K. R. Mahendran, A. Molitor, J. M. Bolla, A. N. Bessonov, M. Winterhalter, J. Pages, *M. PLoS One* **2009**, 4, e5453.
- [8] B. K. Ziervogel, B. Roux, *Structure* **2013**, 21, 76.
- [9] W. Arunmanee, M. Pathania, A. S. Solovyova, A. P. Le Brun, H. Ridley, A. Basle, B. van den Berg, J. H. Lakey, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113, E5034.
- [10] A. D. Ferguson, W. Welte, E. Hofmann, B. Lindner, O. Holst, J. W. Coulton, K. Diederichs, *Structure* **2000**, 8, 585.
- [11] H. Nikaido, *Microbiol. Mol. Biol. Rev.* **2003**, 67, 593.
- [12] S. Akira, S. Uematsu, O. Takeuchi, *Cell* **2006**, 124, 783.
- [13] J. M. Griffiss, H. Schneider, R. E. Mandrell, R. Yamasaki, G. A. Jarvis, J. J. Kim, B. W. Gibson, R. Hamadeh, M. A. Apicella, *Rev. Infect. Dis.* **1988**, 10, S287.
- [14] Y. A. Knirel, A. P. Anisimov, *Acta Naturae* **2012**, 4, 46.
- [15] H. Xu, A. E. Murdaugh, W. Chen, K. E. Aidala, M. A. Ferguson, E. M. Spain, M. E. Nunez, *Langmuir* **2013**, 29, 3000.
- [16] A. Boags, P. C. Hsu, F. Samsudin, P. J. Bond, S. Khalid, *J Phys Chem Lett* **2017**, 8, 2513.
- [17] J. Parkin, M. Chavent, S. Khalid, *Biophys. J.* **2015**, 109, 461.
- [18] S. Jo, T. Kim, V. G. Iyer, W. Im, *J. Comput. Chem.* **2008**, 29, 1859.
- [19] P. C. Hsu, B. M. H. Bruininks, D. Jefferies, P. Cesar Telles de Souza, J. Lee, D. S. Patel, S. J. Marrink, Y. Qi, S. Khalid, W. Im, *J. Comput. Chem.* **2017**, 38, 2354.
- [20] J. Lee, D. S. Patel, J. Stahle, S. J. Park, N. R. Kern, S. H. Kim, X. Cheng, M. A. Valvano, O. Holst, Y. A. Knirel, Y. Qi, S. Jo, J. B. Klauda, G. Widmalm, W. Im, *Journal of chemical theory and computation* **2019**, 15, 775.
- [21] S. Kim, D. S. Patel, S. Park, J. Slusky, J. B. Klauda, G. Widmalm, W. Im, *Biophys. J.* **2016**, 111, 1750.
- [22] R. Danne, C. Poojari, H. Martinez-Seara, S. Rissanen, F. Lolicato, T. Rog, I. Vattulainen, *J. Chem. Inf. Model.* **2017**, 57, 2401.
- [23] A. Singh, M. B. Tessier, K. Pederson, X. Wang, A. P. Venot, G. J. Boons, J. H. Prestegard, R. J. Woods, *Can. J. Chem.* **2016**, 94, 927.
- [24] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyan, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, *Journal of chemical theory and computation* **2016**, 12, 281.
- [25] H. Ma, F. J. Irudayanathan, W. Jiang, S. Nangia, *The journal of physical chemistry. B* **2015**, 119, 14668.
- [26] P. C. Hsu, D. Jefferies, S. Khalid, *B* **2016**, 120, 11170.
- [27] H. Ma, D. D. Cummins, N. B. Edelstein, J. Gomez, A. Khan, M. D. Llewellyn, T. Picudella, S. R. Willsey, S. Nangia, *Journal of chemical theory and computation* **2017**, 13, 811.
- [28] D. E. S. Santos, L. Pol-Fachin, R. D. Lins, T. A. Soares, *J. Chem. Inf. Model.* **2017**, 57, 2181.
- [29] H. Ma, A. Khan, S. Nangia, *Langmuir* **2018**, 34, 5623.
- [30] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmuller, A. D. MacKerell, Jr., *Nat. Methods* **2017**, 14, 71.
- [31] W. Huang, Z. Lin, W. F. van Gunsteren, *Journal of chemical theory and computation* **2011**, 7, 1237.
- [32] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *The journal of physical chemistry. B* **2007**, 111, 7812.
- [33] W. Han, K. Schulten, *Journal of chemical theory and computation* **2012**, 8, 4413.
- [34] C. Kutzner, S. Pall, M. Fechner, A. Esztermann, B. L. de Groot, H. Grubmuller, *J. Comput. Chem.* **2015**, 36, 1990.
- [35] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, 26, 1701.
- [36] M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, *Nucleic Acids Res.* **2012**, 40, D370.
- [37] D. Kozma, I. Simon, G. E. Tusnady, *Nucleic Acids Res.* **2013**, 41, D524.
- [38] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926.
- [39] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, 91, 6269.
- [40] S. O. Yesylevsky, L. V. Schafer, D. Sengupta, S. J. Marrink, *PLoS computational biology* **2010**, 6, e1000810.
- [41] P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M. J. de Hoon, *Bioinformatics* **2009**, 25, 1422.
- [42] B. Webb, A. Sali, *Methods Mol. Biol.* **2017**, 1654, 39.
- [43] D. Sehnal, M. Deshpande, R. S. Varkova, S. Mir, K. Berka, A. Midlik, L. Pravda, S. Velankar, J. Koca, *Nat. Methods* **2017**, 14, 1121.
- [44] L. Martinez, R. Andrade, E. G. Birgin, J. M. Martinez, *J. Comput. Chem.* **2009**, 30, 2157.
- [45] S. Baoukina, L. Monticelli, D. P. Tieleman, *The journal of physical chemistry. B* **2013**, 117, 12113.
- [46] C. Sohlenkamp, O. Geiger, *FEMS Microbiol. Rev.* **2016**, 40, 133.
- [47] D. H. de Jong, G. Singh, W. F. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, S. Marrink, *J. Journal of chemical theory and computation* **2013**, 9, 687.
- [48] S. Snyder, D. Kim, T. J. McIntosh, *Biochemistry* **1999**, 38, 10758.
- [49] K. Brandenburg, S. S. Funari, M. H. Koch, U. Seydel, *J. Struct. Biol.* **1999**, 128, 175.
- [50] D. Cremer, J. A. J. Pople, *Am. Chem. Soc.* **1975**, 97, 1354.
- [51] A. D. Hill, P. Reilly, *J. Journal of chemical information and modeling* **2007**, 47, 1031.
- [52] H. I. Ingolfsson, M. N. Melo, F. J. van Eerden, C. Arnarez, C. A. Lopez, T. A. Wassenaar, X. Periole, A. H. de Vries, D. P. Tieleman, S. J. Marrink, *J. Am. Chem. Soc.* **2014**, 136, 14554.
- [53] H. J. Risselada, S. J. Marrink, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 17367.
- [54] A. Pautsch, G. E. Schulz, *Nat. Struct. Biol.* **1998**, 5, 1013.
- [55] R. G. Efremov, L. A. Sazanov, *J. Struct. Biol.* **2012**, 178, 311.
- [56] P. E. Klebba, S. A. Benson, S. Bala, T. Abdullah, J. Reid, S. P. Singh, H. Nikaido, *J. Biol. Chem.* **1990**, 265, 6800.
- [57] D. S. Patel, S. Re, E. L. Wu, Y. Qi, P. E. Klebba, G. Widmalm, M. S. Yeom, Y. Sugita, W. Im, *Biophys. J.* **2016**, 110, 930.
- [58] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, Jr., *J. Comput. Chem.* **2010**, 31, 671.
- [59] T. A. Wassenaar, H. I. Ingolfsson, R. A. Bockmann, D. P. Tieleman, S. Marrink, *J. Journal of chemical theory and computation* **2015**, 11, 2144.
- [60] X. Periole, A. M. Knepp, T. P. Sakmar, S. J. Marrink, T. Huber, *J. Am. Chem. Soc.* **2012**, 134, 10959.
- [61] K. N. Kirschner, R. D. Lins, A. Maass, T. A. Soares, *Journal of chemical theory and computation* **2012**, 8, 4719.

Received: 21 January 2019

Revised: 18 February 2019

Accepted: 3 March 2019