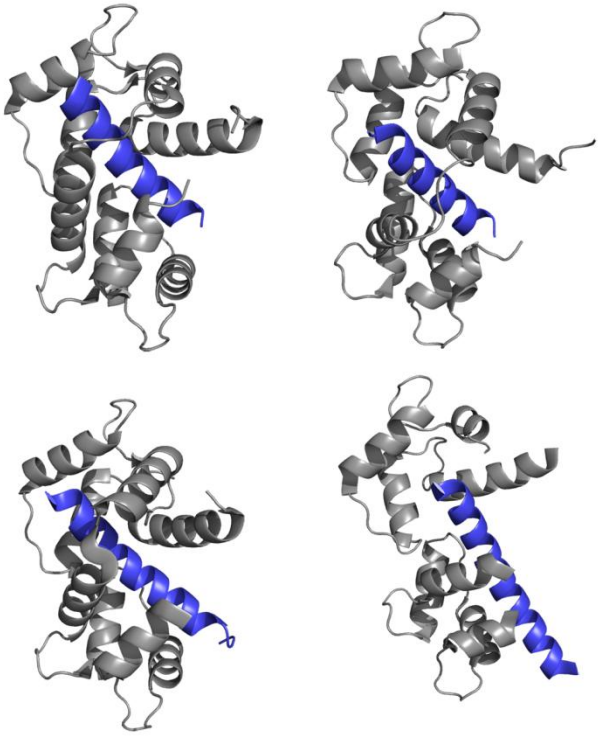
ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η  Ε Ρ Γ Α Σ Ι Α

# «Υπολογιστική ανάλυση οικογένειας κινασών»



**Βάσια Μεραντζή**

Πτυχιούχος Βιολογίας, ΕΚΠΑ

**ΑΘΗΝΑ 2018**

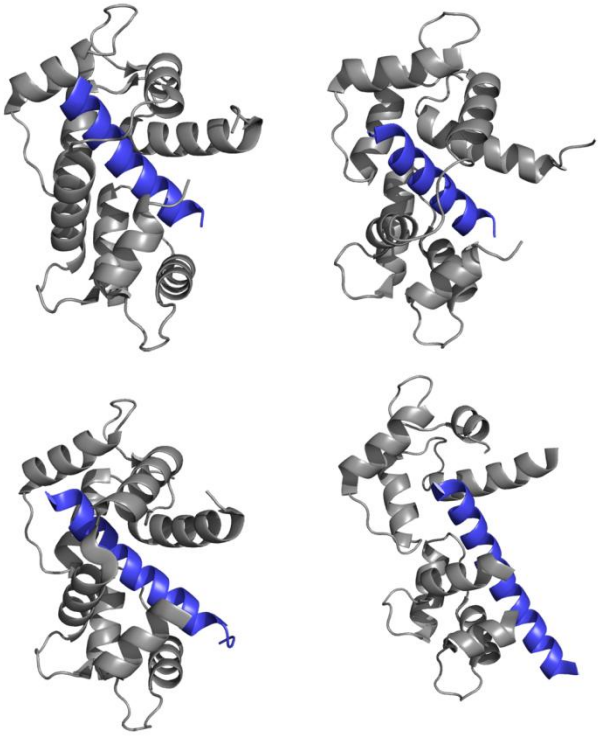
HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
FACULTY OF BIOLOGY

MASTER IN "BIOINFORMATICS"

**Master Diploma Thesis**

# «Computational analysis of a kinase superfamily»



**Vasia Merantzi**

BSc Biology, NKUA

**ATHENS 2018**

**ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ**
**Εθνικό και Καποδιστριακό**
**Πανεπιστήμιο Αθηνών**

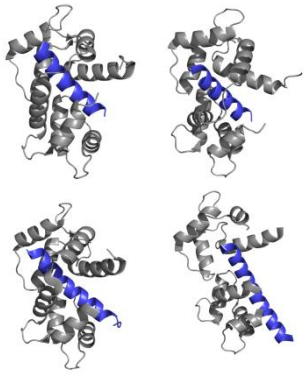
**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**
**ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ**

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ**
**«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»**

---

**Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η  Ε Ρ Γ Α Σ Ι Α**

# «Υπολογιστική ανάλυση οικογένειας κινασών»



**Τριμελής εξεταστική επιτροπή**

**Καθηγητής Κωνσταντίνος Βοργιάς (Επιβλέπων)**
*Τομέας Βιοχημείας & Μοριακής Βιολογίας*
*Τμήμα Βιολογίας*
*Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών*

**Αναπληρωτής Καθηγητής Παντελής Μπάγκος**
*Εργαστήριο Μοριακής και Υπολογιστικής Βιολογίας και Γενετικής*
*Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική*
*Πανεπιστήμιο Θεσσαλίας*

**Επίκουρη Καθηγήτρια Βασιλική Οικονομίδου**
*Τομέας Βιολογίας Κυττάρου & Βιοφυσικής*
*Τμήμα Βιολογίας*
*Εθνικό & Καποδιστριακό Πανεπιστήμιο Αθηνών*

# Περίληψη

Το ασβέστιο (Ca2+) αποτελεί έναν από τους πιο ευρέως διαδεδομένους και σημαντικούς δεύτερους αγγελιαφόρους. Λειτουργεί ως δεύτερο μήνυμα σε μια ποικιλία κυτταρικών αποκρίσεων ρυθμίζοντας καταρροϊκά μονοπάτια μεταγωγής σήματος, συμμετέχοντας σε πληθώρα σημαντικών βιολογικών διαδικασιών. Η καλμοντουλίνη (CaM), που δρα ως ο πρωταρχικός ενδοκυτταρικός υποδοχέας ασβεστίου σχεδόν σε όλα τα ευκαρυωτικά κύτταρα, δημιουργεί σύμπλοκο με το ασβέστιο (Ca2+-CaM) και με τον τρόπο αυτό ελέγχει την ενεργότητα περισσότερων από 120 διαφορετικών ενζύμων και πρωτεϊνών. Μια από τις πιο σημαντικές πρωτεΐνες-στόχους της καλμοντουλίνης είναι η οικογένεια των εξαρτώμενων από την καλμοντουλίνη κινασών (Calcium-Calmodulin Dependent Kinases – CAMKs). Η αλληλεπίδραση κινασών-καλμοντουλίνης οδηγεί σε κρίσιμες στερεοδιατακτικές αλλαγές στο μόριο της κινάσης, οι οποίες είναι αναγκαίες για την τέλεση των φυσιολογικών λειτουργιών της. Τα σύμπλοκα Ca2+-CaM-κινάσης ρυθμίζουν πολύ διαφορετικές κυτταρικές λειτουργίες, μεταξύ των οποίων η ρύθμιση της έκφρασης διαφορετικών γονιδίων, ο κυτταρικός θάνατος (απόπτωση), η αναδιοργάνωση του κυτταροσκελετού, η μάθηση, η μνήμη και πολλές άλλες.

Η μελέτη του τρόπου αλληλεπίδρασης των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης παίζει καθοριστικό ρόλο τόσο στην αποσαφήνιση των μηχανισμών απόκρισης των κυττάρων στις αλλαγές της συγκέντρωσης του ασβεστίου, όσο και στη ρύθμιση των λειτουργιών ενός οργανισμού γενικότερα. Ωστόσο, τα διαθέσιμα δομικά δεδομένα που αφορούν τους μηχανισμούς που διέπουν τις αλληλεπιδράσεις μεταξύ των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης είναι ελλιπή, καθώς μόνο σε δύο περιπτώσεις πειραματικά προσδιορισμένων συμπλόκων εντοπίζονται ολόκληρα τα μόρια της εξαρτώμενης από καλμοντουλίνη κινάσης με την αντίστοιχη καλμοντουλίνη. Σήμερα η δομική πληροφορία έχει αποκτηθεί κυρίως από σύμπλοκα της καλμοντουλίνης με μικρά πεπτίδια των κινασών. Επιπροσθέτως, ενώ η αλληλεπίδραση εμφανίζει υψηλή εξειδίκευση, δεν έχει καταγραφεί κάποια συντηρημένη ακολουθία αναγνώρισης των πρωτεϊνών-στόχων από την καλμοντουλίνη.

Με βάση τα παραπάνω, στόχος της παρούσας διπλωματικής εργασίας ήταν η υπολογιστική μελέτη των αλληλεπιδράσεων των εξαρτώμενων από την καλμοντουλίνη κινασών με την καλμοντουλίνη με υπολογιστικές μεθόδους που εκμεταλλεύονται τη συνεξελικτική πληροφορία.

Το υπολογιστικό πρωτόκολλο συνεξέλιξης που εφαρμόστηκε κατά τη μελέτη των αλληλεπιδράσεων μεταξύ των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης συνέβαλε τόσο στην εύρεση πιθανών θέσεων συνεξέλιξης μεταξύ των δύο πρωτεϊνικών ομάδων, όσο και στο λειτουργικό σχολιασμό αλληλεπιδράσεων κομβικών αμινοξικών καταλοίπων. Συγκεκριμένα, η πληροφορία που προέκυψε από τις μεθόδους συνεξέλιξης κατά τη μελέτη των παραπάνω συμπλόκων επιβεβαιώθηκε τόσο από τις πειραματικά προσδιορισμένες δομές των συμπλόκων, όσο και με τη βοήθεια μεθόδων κατευθυνόμενης αγκυροβόλησης. Τέλος, ανέδειξε τα τμήματα εκείνα των πρωτεϊνών που συμμετέχουν ενεργά στις στερεοδιατακτικές τροποποιήσεις κατά την αλληλεπίδραση των μορίων.

Τα αποτελέσματα της παρούσας διπλωματικής εργασίας αφενός ενισχύουν την άποψη πως η χρήση της συνεξελικτικής πληροφορίας αποτελεί ένα σημαντικό εργαλείο για τη διερεύνηση και το σχολιασμό των λειτουργιών οικογενειών πρωτεϊνικών μορίων. Αφετέρου, τα δεδομένα συνεξέλιξης που αφορούν την καλμοντουλίνη και τις εξαρτώμενες από την καλμοντουλίνη κινάσες εμπλουτίζουν τις προσπάθειες μελέτης του μηχανισμού και της φύσης των μεταξύ τους αλληλεπιδράσεων , ένα ερευνητικό πεδίο που αποτελεί στόχο για το σχεδιασμό και την ανάπτυξη φαρμάκων για πληθώρα παθολογικών καταστάσεων, όπως διάφορες παθήσεις του νευρικού συστήματος ή διαφορετικές μορφές καρκίνου.

# Αντί προλόγου

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Βιοπληροφορική υπό την επίβλεψη του Καθηγητή κ. Κωνσταντίνου Βοργιά.

Αρχικά θα ήθελα να ευχαριστήσω τον Ομότιμο Καθηγητή κ. Σταύρο Χαμόδρακα, τον Καθηγητή κ. Κωνσταντίνο Βοργιά, τον Καθηγητή κ Αναστάσιο Λεγάκι και την Επίκουρη Καθηγήτρια κ. Βασιλική Οικονομίδου για την ευκαιρία που μου έδωσαν, επιλέγοντάς με να συμμετάσχω στο συγκεκριμένο Μεταπτυχιακό Δίπλωμα Ειδίκευσης, να γνωρίσω μια «άλλη όψη» της Βιολογίας που καθόρισε σε μεγάλο βαθμό τον τρόπο που προσεγγίζω διάφορα επιστημονικά ζητήματα, αλλά και τις επιστήμες στην ολότητά τους, να καθοδηγηθώ σε ανεξερεύνητα για μένα μονοπάτια και να αποκτήσω γνώσεις αλλά κυρίως τα εργαλεία να τις διαχειριστώ.

Στον Καθηγητή και επιβλέποντα της παρούσας εργασίας, κ. Κωνσταντίνο Βοργιά, οφείλω ένα μεγάλο ευχαριστώ για την καθοδήγηση και τη βοήθεια που μου παρείχε καθ΄ όλη τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας. Τον ευχαριστώ επίσης θερμά για την εμπιστοσύνη, την ηθική στήριξη, την επιμονή και την υπομονή που μου έδειξε.

Θα ήθελα να ευχαριστήσω, επίσης, τα μέλη της Τριμελούς Εξεταστικής Επιτροπής, την Επίκουρη Καθηγήτρια κ. Βασιλική Οικονομίδου, τον Αναπληρωτή Καθηγητή κ. Παντελή Μπάγκο και τον Καθηγητή κ. Κωνσταντίνο Βοργιά, τόσο για την τιμή που μου έκαναν να συμμετάσχουν σε αυτήν όσο και για τις πολύτιμες και εποικοδομητικές υποδείξεις τους. Τους ευχαριστώ, επιπλέον, για τα πολύ σημαντικά και ενδιαφέροντα μαθήματά τους στα πλαίσια του Μεταπτυχιακού Προγράμματος Σπουδών, τα οποία επηρέασαν σε μεγάλο βαθμό την πορεία της παρούσας διπλωματικής εργασίας.

Εξίσου θερμές ευχαριστίες οφείλω στο σύνολο των διδασκόντων του Μεταπτυχιακού Προγράμματος Σπουδών για το αμείωτο ενδιαφέρον τους να μοιραστούν και να συζητήσουν γνώσεις και πληροφορίες. Ιδιαίτερες ευχαριστίες οφείλω στη Δρ. Μαργαρίτα Θεοδωροπούλου και στον Δρ. Νικόλαο Παπανδρέου που αφιέρωσαν χρόνο στον καθένα μας προσωπικά κάνοντας ιδιαίτερα απαιτητικούς τομείς να μοιάζουν σαν ένα ενδιαφέρον εγκεφαλικό παιχνίδι.

Θα ήθελα να ευχαριστήσω από καρδιάς την Δρ. Εβίτα Τσιολάκη και τις συμφοιτήτριές μου Χριστίνα Κουτσοθανάση και Μαρίνα Γαβριηλίδου που ήταν δίπλα μου σε κάθε βήμα, αφιερώνοντας σημαντικό κομμάτι από τον προσωπικό τους χρόνο, και με πολύτιμες συμβουλές και συζητήσεις συνέβαλαν καθοριστικά στην πορεία και την ολοκλήρωση της παρούσας διπλωματικής εργασίας.

Τέλος θα ήθελα να ευχαριστήσω όλους τους μοναδικά ενδιαφέροντες ανθρώπους που γνώρισα, διδάσκοντες και συμφοιτητές, για την ανθρωπιά και το ανιδιοτελές ενδιαφέρον τους.

Η παρούσα διπλωματική είναι αφιερωμένη στον αδερφό μου, Δημήτρη, και στο ειλικρινές απέραντο χαμόγελό του που είναι η δύναμη για να συνεχίζω να προχωράω και στους ανθρώπους που προχωράμε μαζί.

# Περιεχόμενα

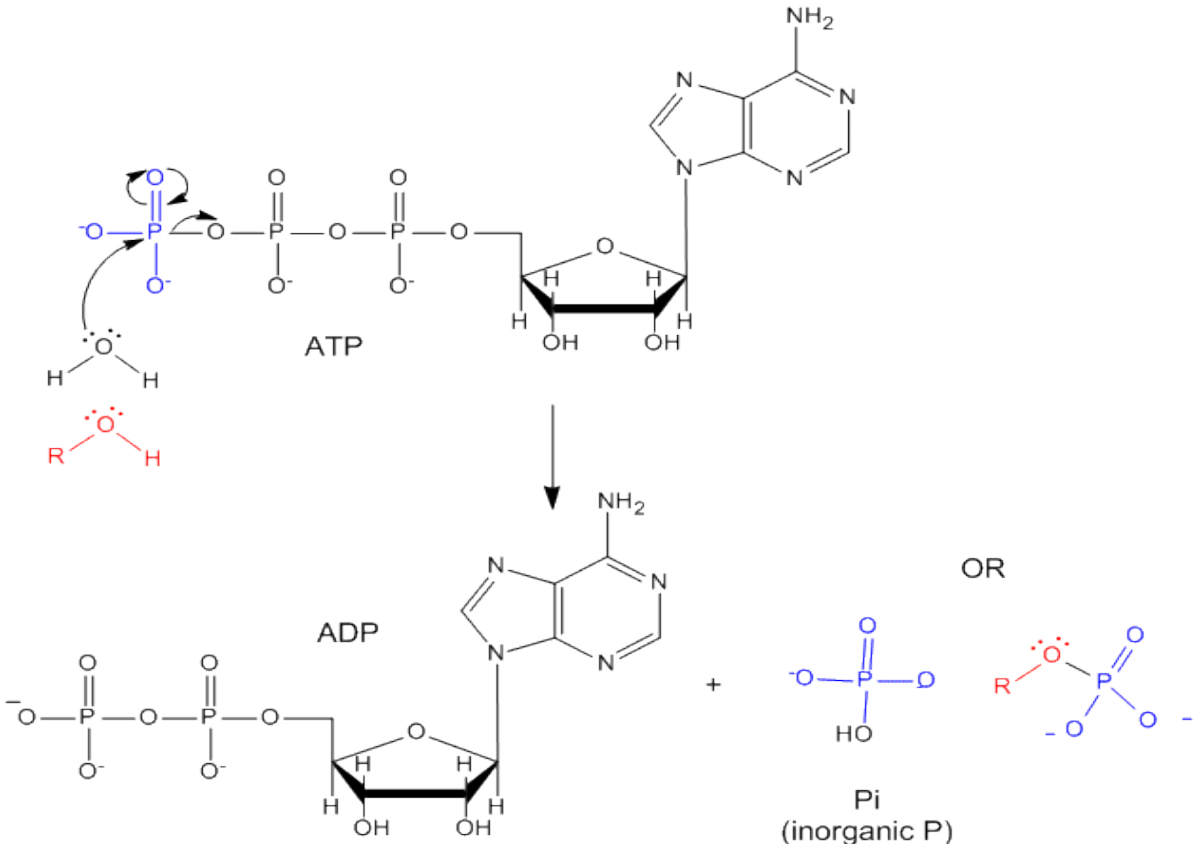
Περίληψη ..... 7
Αντί προλόγου ..... 9
1. Εισαγωγή ..... 1
1.1. Εισαγωγή στις πρωτεϊνικές κινάσες ..... 1
1.1.1. Βακτηριακές κινάσες ..... 5
1.1.2. Ευκαρυωτικές κινάσες ..... 7
1.1.3. Αλληλεπιδράσεις κινασών ..... 15
1.2 Εισαγωγή στις υπό μελέτη ομάδες πρωτεϊνών ..... 19
1.2.1. $Ca^{2+}$ ένας σημαντικός δεύτερος αγγελιαφόρος ..... 19
1.2.2. CaM πρωτεΐνες ..... 21
1.2.3. Εξαρτώμενες από $Ca^{2+}$/CaM κινάσες ..... 26
1.3. Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών ..... 66
1.3.1. Εισαγωγή ..... 66
1.3.2. Πειραματικές μέθοδοι προσδιορισμού αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών ..... 67
1.3.3. Υπολογιστικές μέθοδοι για την μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών ..... 69
1.3.4. Υπολογιστικές μέθοδοι για την μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών βασισμένες σε δεδομένα συνεξέλιξης ..... 79
1.3.5. Στατιστικά μοντέλα για την πρόβλεψη αλληλεπιδρώντων καταλοίπων μέσω της συνεξελικτικής πληροφορία ..... 82
1.4. Σκοπός ..... 109
2. Υλικά και μέθοδοι ..... 110
2.1. Κατασκευή του συνόλου δεδομένων ..... 110
2.2. Πολλαπλές στοιχίσεις ακολουθιών ..... 113
2.3. Επιλογή server συνεξέλιξης ..... 114
2.3.1. Εισαγωγή ..... 114
2.3.2. Ο server i-coms ..... 114
2.4. Επιλογή στατιστικής μεθόδου - Οπτικοποίηση των δομών ..... 116
2.5. Κατασκευή δομικών μοντέλων των συμπλόκων με τη βοήθεια μεθόδων αγκυροβόλησης ..... 117
2.6. Δομική στοίχιση ..... 119
3. Αποτελέσματα και συζήτηση ..... 125
3.1. Κατασκευή του συνόλου δεδομένων ..... 125
3.2. Πολλαπλές στοιχίσεις ακολουθιών ..... 129
3.3. Επιλογή στατιστικής μεθόδου πρόβλεψης συνεξελισσόμενων καταλοίπων ..... 132

3.4. Αξιολόγηση αποτελεσμάτων του server i-coms ..... 133
3.4.1. Μελέτη συμπλόκου CaMK Iα – $Ca^{2+}$-CaM και συμπεράσματα ..... 135
3.4.2. Μελέτη συμπλόκου CaMK IIδ – $Ca^{2+}$-CaM2 και συμπεράσματα ..... 141
3.4.3. Μελέτη του συμπλόκου MLCK I – $Ca^{2+}$-CaM και συμπεράσματα ..... 146
3.4.4. Μελέτη συμπλόκου DAPK1 – $Ca^{2+}$-CaM και συμπεράσματα ..... 151
3.4.5. Μελέτη του συμπλόκου DAPK2 – $Ca^{2+}$-CaM και συμπεράσματα ..... 157
3.4.6. Συγκριτική μελέτη των $Ca^{2+}$-CaM πρωτεϊνών και συμπεράσματα ..... 162
3.5. Γενικά συμπεράσματα ..... 167
4. Βιβλιογραφία..... 170

# 1.Εισαγωγή

## 1.1. Εισαγωγή στις πρωτεϊνικές κινάσες

Οι πρωτεϊνικές κινάσες ορίζονται ως τα ένζυμα που καταλύουν τις αντιδράσεις φωσφορυλίωσης. Η αντίδραση αυτή αποτελεί την επικρατούσα αντιστρέψιμη ομοιοπολική τροποποίηση, εντοπίζεται σχεδόν σε κάθε μεταβολική πορεία και ρυθμίζεται μέσω αυτής η δραστικότητα πολλών ενζύμων, διαύλων μεμβράνης καθώς και άλλων πρωτεϊνών στόχων. Ο μηχανισμός της αντίδρασης αυτής αφορά τη μεταφορά της τελικής (γ) φωσφορικής ομάδας του ATP(δότης) σε ένα καθορισμένο σημείο συγκεκριμένου υποστρώματος (δέκτης).




**Εικόνα 1.1** Βασική αντίδραση φωσφορυλίωσης

Οι κινάσες έχουν την δυνατότητα να φωσφορυλιώνουν πρωτεΐνες, λιπίδια καθώς και υδατάνθρακες. Θεωρούνται μόρια ζωτικής σημασίας αφού συμμετέχουν στην ρύθμιση του μεταβολισμού, της κυτταρικής σηματοδότησης και μεταφοράς, της πρωτεϊνικής σύνθεσης, των διάφορων εκκριτικών διαδικασιών και άλλων κυτταρικών μονοπατιών.

Συγκεκριμένα, τα βασικά χαρακτηριστικά των κινασών είναι τα εξής:

Είναι υπεύθυνες για την σταθερότητα της αντίδρασης μεταφοράς της φωσφορικής ομάδας κατά την υδρόλυση του δεσμού υψηλής ενέργειας.

Έχουν την δυνατότητα να αυξήσουν τον ρυθμό της αντίδρασης, προσανατολίζοντας κατάλληλα το υπόστρωμα και την φωσφορική ομάδα στα ενεργά τους κέντρα.

Συχνά χρησιμοποιούν φορτισμένα αμινοξικά κατάλοιπα, που σταθεροποιούν ηλεκτροστατικά την μεταβατική κατάσταση, αλληλεπιδρώντας με την αρνητικά φορτισμένη φωσφορική ομάδα. Εναλλακτικά, κάποιες κινάσες χρησιμοποιούν δεσμευμένους συμπαράγοντες μετάλλων στο ενεργό τους κέντρο για να συντονίσουν τις φωσφορικές ομάδες.

Το προστιθέμενο φορτίο και η δομή της φωσφορικής ομάδας μπορεί να τροποποιήσει την δραστικότητα πολλών ενζύμων (καθώς και των περισσότερων κινασών), να αλλάξει τις αλληλεπιδράσεις μεταξύ πρωτεϊνών, την θέση ή την αλληλεπίδραση η των πρωτεϊνών με τα υποστρώματά τους.

Πολλές πρωτεΐνες φέρουν πολλαπλές θέσεις φωσφορυλίωσης, οι οποίες μπορεί να φέρουν διαφορετικές ή ακόμα και αντίθετες επιπτώσεις στην πρωτεΐνη – υπόστρωμα.

Η φωσφορυλίωση αποτελεί ένα ιδιαίτερα αποτελεσματικό μέσο ελέγχου δραστικότητας των πρωτεϊνών για δομικούς, θερμοδυναμικούς, κινητικούς και ρυθμιστικούς λόγους. Συγκεκριμένα:

Μία φωσφορική ομάδα προσθέτει δύο αρνητικά φορτία σε μία τροποποιημένη πρωτεΐνη. Οι ηλεκτροστατικές αλληλεπιδράσεις στη μη τροποποιημένη πρωτεΐνη είναι δυνατόν να διακοπούν και να σχηματιστούν νέες ηλεκτροστατικές δυνάμεις. Τέτοιες δομικές αλλαγές μπορούν να μεταβάλουν σημαντικά την πρόσδεση του υποστρώματος και την καταλυτική δραστικότητα.

Μία φωσφορική ομάδα μπορεί να σχηματίσει τρεις ή τέσσερις δεσμούς υδρογόνου. Η τετραεδρική γεωμετρία της φωσφορικής ομάδας σε μεγάλο βαθμό οδηγεί αναγκαστικά αυτούς τους δεσμούς υδρογόνου προς την υποδεικνυόμενη κατεύθυνση, επιτρέποντας ειδικές αλληλεπιδράσεις με δότες δεσμών υδρογόνου.

Η ελεύθερη ενέργεια της φωσφορυλίωσης είναι μεγάλη. Από τα -12kcal/mol (-50kJ/mol) που παρέχονται από το ATP, περίπου το μισό καταναλώνεται στο να γίνει η φωσφορυλίωση μη αντιστρεπτή, το άλλο μισό διατηρείται στη φωσφορυλιωμένη πρωτεΐνη. Είναι γνωστό ότι η αλλαγή της ελεύθερης ενέργειας των 1,36 kcal/mol (5,69 kJ/mol) αντιστοιχεί σε μεταβολή μιας σταθεράς ισορροπίας κατά έναν παράγοντα του 10. Επομένως, η φωσφορυλίωση μπορεί να αλλάξει την στερεοδιατακτική ισορροπία μεταξύ διαφορετικών λειτουργικά καταστάσεων μέχρι και $10^4$ φορές.

Η φωσφορυλίωση και η αποφωσφορυλίωση μπορούν να λάβουν χώρα σε λιγότερο από ένα δευτερόλεπτο ή σε διάστημα ωρών. Η κινητική της αντίδρασης έχει την ικανότητα να προσαρμόζεται έτσι ώστε να αντιμετωπίσει τις χρονικές απαιτήσεις μιας φυσιολογικής πορείας της αντίδρασης.

Η φωσφορυλίωση προκαλεί συχνά αποτελέσματα υψηλής ενίσχυσης. Μία μοναδική ενεργοποιημένη κινάση μπορεί να φωσφορυλιώσει σε σύντομο διάστημα εκατοντάδες πρωτεΐνες στόχους. Επιπλέον υψηλή ενίσχυση μπορεί να λάβει χώρα στην περίπτωση που οι πρωτεΐνες στόχοι είναι ένζυμα, που το κάθε ένα πρέπει να μετασχηματίσει έναν μεγάλο αριθμό μορίων υποστρώματος.

Το ATP είναι το νόμισμα ενέργειας του κυττάρου. Η χρήση της ένωσης αυτής ως δότη φωσφορικής ομάδας συνδέει την ενεργειακή κατάσταση του κυττάρου με τη ρύθμιση του μεταβολισμού.

Ακόμα, οι κινάσες ανάλογα με τα αμινοξικά κατάλοιπα τα οποία φωσφορυλιώνουν χωρίζονται στις σερίνης/θρεονίνης, τυροσίνης και στις διπλής ειδικότητας, οι οποίες έχουν την δυνατότητα να φωσφορυλιώνουν και τα τρία αμινοξικά κατάλοιπα (Dhanasekaran & Reddy, 1998). Ακόμα, υπάρχουν και κινάσες που έχουν την δυνατότητα να φωσφορυλιώνουν και άλλα αμινοξέα π.χ. την ιστιδίνη, όπως θα αναλυθεί παρακάτω. Ενώ υπάρχουν ενδείξεις φωσφορυλίωσης και άλλων αμινοξικών καταλοίπων, επειδή όμως αυτά τα φωσφορυλιωμένα κατάλοιπα είναι ιδιαίτερα ασταθή, η μελέτη τους είναι ιδιαίτερα δύσκολη και δεν έχουν προσδιοριστεί ακόμα.

Αξίζει να αναφερθεί η μεγάλη ποικιλία βάσεων δεδομένων που υπάρχει διαθέσιμη για τις πρωτεϊνικές κινάσες, κάποιες από τις οποίες αναγράφονται στον πίνακα 1.1.

**Πίνακας 1.1** Βάσεις δεδομένων πρωτεϊνικών κινασών

| Βάση Δεδομένων | Γενικά χαρακτηριστικά | Διεύθυνση |
|---|---|---|
| **Kinase.com** | Περιέχει πληροφορίες και εργαλεία για τα kinomes 11 οργανισμών | http://www.kinase.com/web/current/ |
| **KinG** | Περιέχει συλλογή των κινασών Ser/Thr και Tyr καθώς και πληροφορίες για λειτουργικά δομικά αυτοτελείς περιοχές από όλους τους οργανισμούς με αλληλουχημένα γονιδιώματα | http://king.mbu.iisc.ernet.in/ |
| **RKD** | Περιέχει πληροφορίες για όλες τις κινάσες που εντοπίζονται στο ρύζι | http://ricephylogenomics.ucdavis.edu/kinase/genInfo.shtml |
| **Kinase SARfari** | Περιέχει πληροφορία σχετικά με τις ακολουθίες, τις δομές, τα σύμπλοκα που σχηματίζουν οι κινάσες με άλλες πρωτεΐνες ή άλλα μόρια, π.χ.: φάρμακα | https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari/ |
| **MoKCa** | Περιέχει πληροφορία σχετικά με τις μεταλλάξεις που παρατηρούνται στις κινάσες | http://strubiol.icr.ac.uk/extra/mokca/ |

| | | |
|---|---|---|
| | και σχετίζονται με την εμφάνιση καρκίνου | |
| **PhosphoSitePlus** | Περιέχει πληροφορίες αλλά και διαθέσιμα εργαλεία σχετικά με διάφορες Μέτα-μεταφραστικές τροποποιήσεις | https://www.phosphosite.org/homeAction.action |
| **PhosphoNetworks** | Περιέχει πληροφορίες σχετικά με τις αλληλεπιδράσεις κινασών-υποστρωμάτων και τις θέσεις φωσφορυλίωσης | http://www.phosphonetworks.org/ |
| **PhosphoELM** | Περιέχει πληροφορίες σχετικά με τις αλληλεπιδράσεις κινασών-υποστρωμάτων και τις θέσεις φωσφορυλίωσης | http://phospho.elm.eu.org/ |
| **KIDFamMap** | Περιέχει πληροφορίες για αναστολείς κινασών αλλά και τις ασθένειες με τις οποίες σχετίζονται οι κινάσες | http://gemdock.life.nctu.edu.tw/KIDFamMap/ |
| **KID** | Περιέχει πληροφορίες που αφορούν τις κινάσες και αλληλεπιδράσεις αυτών με άλλες πρωτεΐνες στη ζύμη | http://www.moseslab.csb.utoronto.ca/KID/ |
| **KANPHOS** | Περιέχει πληροφορίες για τις περιοχές φωσφορυλίωσης πρωτεϊνών – στόχων | https://kanphos.neuroinf.jp/ |
| **BYKdb** | Περιέχει πληροφορίες για βακτηριακές κινάσες τυροσίνης | https://bykdb.ibcp.fr/BYKdb/ |
| **KLIFS** | Περιέχει πληροφορίες για τις αλληλεπιδράσεις αναστολέων με το καταλυτικό κέντρο των κινασών στον άνθρωπο και το ποντίκι | http://klifs.vu-compmedchem.nl/ |
| **EKPD** | Περιέχει πληροφορίες για τις ευκαρυωτικές κινάσες και φωσφατάσες | http://ekpd.biocuckoo.org/ |
| **Kinomer** | Περιέχει πληροφορία σχετικά με την εξέλιξη και την λειτουργία διαφόρων κινασών με βάση τη σύγκριση των ακολουθιών τις | http://www.compbio.dundee.ac.uk/kinomer/ |
| **KinomeFEATURE** | Περιέχει δομές από 180 ανθρώπινες κινάσες και πληροφορίες σχετικές με τις διαφορετικές περιοχές αλληλεπίδρασης αυτών με πρωτεΐνες ή άλλα μόρια | https://simtk.org/projects/kdb# |
| **ProKinO** | Περιέχει πληροφορίες για το σύνολο των κινασών του ανθρώπου | http://vulcan.cs.uga.edu/prokino/about/browser |
| **P2CS** | Περιέχει πληροφορίες σχετικά με τις κινάσες ιστιδίνης και το σύστημα μετάδοσης δύο συστατικών | http://www.p2cs.org/ |

### 1.1.1. Βακτηριακές κινάσες

Η ύπαρξη βακτηριακών κινασών έγινε αποδεκτή στα τέλη της δεκαετίας του 70, καθώς επικρατούσε η άποψη πως η φωσφορυλίωση πρωτεϊνών περιοριζόταν σε πολύπλοκους - πολυκύτταρους οργανισμούς (Kennelly & Potts, 1996) (Johnson & Lewis, 2001). Μία δεκαετία αργότερα, με την ανακάλυψη του συστήματος δύο συστατικών αποδείχθηκε η αναμφισβήτητη παρουσία των κινασών στους προκαρυωτικούς οργανισμούς. Βέβαια, οι κινάσες που είχαν ως στόχο την υδροξυλομάδα καταλοίπων όπως η σερίνη, η θρεονίνη και η τυροσίνη κατατάσσονταν ως αποκλειστικά ευκαρυωτικές, ενώ στις προκαρυωτικές κινάσες κατατάσσονταν εκείνες με στόχο την ιστιδίνη (Kennelly & Potts, 1996) (Johnson & Lewis, 2001). Με την ανάπτυξη, όμως, μεθόδων στον τομέα της γονιδιωματικής και της πρωτεωμικής παρατηρήθηκε η φωσφορυλίωση της τυροσίνης, της σερίνης και της θρεονίνης και στα βακτήρια.

#### 1.1.1.1. Κινάσες Ιστιδίνης

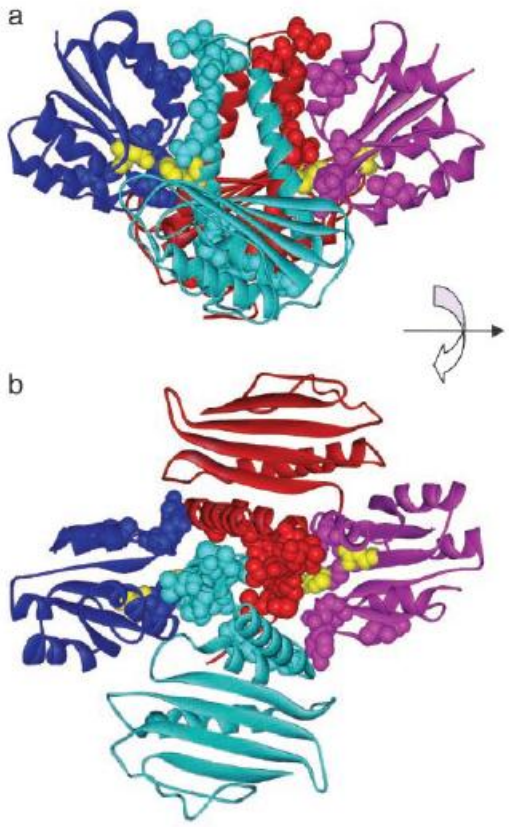
Οι κινάσες ιστιδίνης φωσφορυλιώνουν κατάλοιπα ιστιδίνης και ασπαρτικού οξέος κατά το μηχανισμό που ονομάζεται σύστημα μετάδοσης δύο συστατικών και αποτελεί τον βασικό μηχανισμό σύζευξης ερεθίσματος απόκρισης σε μεταβαλλόμενες περιβαλλοντικές συνθήκες. Πιο συγκεκριμένα καταλύουν την αντίδραση :

ATP + protein L-histidine $\rightleftharpoons$ ADP + protein N-phospho-L-histidine

Είναι πολυλειτουργικές, συνήθως διαμεμβρανικές πρωτεΐνες και με την μεταφορά μιας φωσφορικής ομάδας συμμετέχουν στη μεταγωγή σήματος κατά το μήκος της κυτταρικής μεμβράνης (Wolanin, Thomason, & Stock, 2002). Η συντριπτική πλειοψηφία των κινασών ιστιδίνης είναι ομοδιμερή που εμφανίζουν δραστικότητα κινάσης και φωσφατάσης. Ακόμα, δρουν και ως κυτταρικοί υποδοχείς για σηματοδοτικά μόρια με ανάλογο τρόπο με εκείνο των υποδοχέων κινασών τυροσίνης.

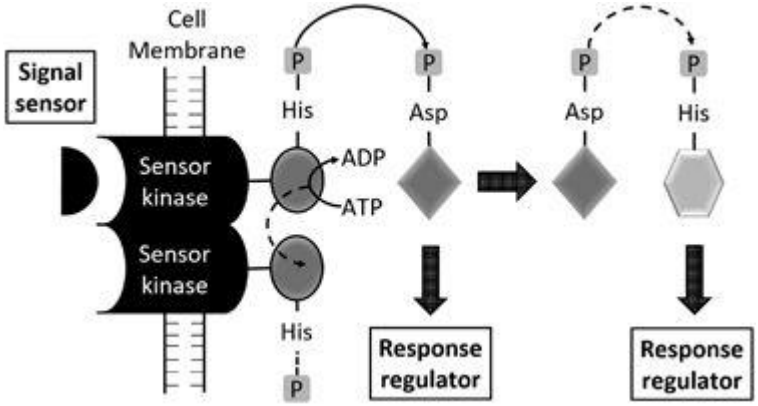
Το σύστημα μετάδοσης 2 συστατικών αποτελείται από την κινάση ιστιδίνης και την συγγενή αυτής πρωτεΐνη υπεύθυνη για την απόκριση – RR(Response Regulator). Οι κινάσες ιστιδίνης αποτελούνται από διακριτά τμήματα: Μία δομικά αυτοτελή περιοχή που δρα ως υποδοχέας του μηνύματος και την περιοχή με δραστικότητα πρωτεϊνικής κινάσης. Η δεύτερη αποτελείται από 2 δομικά αυτοτελείς περιοχές: το καταλυτικό κέντρο και την περιοχή υπεύθυνη για τον διμερισμό (DD). Το καταλυτικό κέντρο υδρολύει το ATP και μεταφέρει την φωσφορική ομάδα σε ένα συντηρημένο κατάλοιπο ιστιδίνης στην DD περιοχή. (Lewyn Li, Shakhnovich, & Mirny, 2003).

Η RR πρωτεΐνη με τη σειρά της αποτελείται και αυτή από 2 τμήματα: την αυτοτελή δομική περιοχή αποδέκτη (receiver domain) η οποία διαθέτει μια συντηρημένη α-β Rossman στερεοδιάταξη και την αυτοτελή δομική περιοχή τελεστή (effector domain) η οποία εμφανίζει διαφορετικές δομές μεταξύ διαφορετικών συστημάτων.




**Εικόνα 1.2** Απεικονίζεται το σύστημα μετάδοσης 2 συστατικών. Με κόκκινο και γαλάζιο η διμερισμένη κινάση ιστιδίνης (οι έλικες στο κέντρο αποτελούν τις DD περιοχές) ενώ με μπλε και ματζέντα η περιοχή αποδέκτης της RR. Με κίτρινο απεικονίζονται τα κατάλοιπα ιστιδίνης30 και ασπαρτικού54 (Lewyn Li et al., 2003).

Ο μηχανισμός έχει ως εξής : Η φωσφορική ομάδα μεταφέρεται από το κατάλοιπο της ιστιδίνης της DD περιοχής σε ένα κατάλοιπο ασπαρτικού στην δομικά αυτοτελή περιοχή αποδέκτη της RR. Αυτό ενεργοποιεί την περιοχή-τελεστή, η οποία προσδένεται στο DNA και με τη σειρά της ενεργοποιεί ή καταστέλλει την έκφραση καθορισμένων γονιδίων. Η μόνη γνωστή εξαίρεση είναι το χημειοτακτικό σύστημα 2 συστατικών που αρχικά δεν φωσφορυλιώνεται κατάλοιπο ιστιδίνης στην DD περιοχή αλλά σε μια δομικά αυτοτελή περιοχή υπεύθυνη για την μεταφορά της φωσφορικής ομάδας.




**Εικόνα 1.3** Βασική αντίδραση κινάσης ιστιδίνης

Όπως παρατηρούμε από την αντίδραση, τα δύο υποστρώματα του ενζύμου είναι το ATP και η πρωτεΐνη που φέρει το κατάλοιπο ιστιδίνης και τα προϊόντα της αντίδρασης είναι το ADP και η πρωτεΐνη που φέρει την φωσφορυλιωμένη ιστιδίνη. Οι κινάσες ιστιδίνης συμμετέχουν κυρίως σε σηματοδοτικά μονοπάτια διαφόρων κυτταρικών διεργασιών που ρυθμίζουν τον μεταβολισμό, την τοξικότητα και την ομοιόσταση. Ο μηχανισμός της αντίδρασης που καταλύουν οι κινάσες ιστιδίνης δεν έχει διευκρινιστεί πλήρως, αλλά έχει προταθεί ότι η καταλυτική περιοχή μίας κινάσης ιστιδίνης που σχηματίζει ομοδιμερές με μία δεύτερη κινάση μπορεί να περιστραφεί κατά τέτοιο τρόπο έτσι ώστε το η θέση δέσμευσης του ATP της κινάσης αυτής να έρθει σε επαφή με την συγκεκριμένη ιστιδίνη της δεύτερης κινάσης που συμμετέχει στο σχηματισμό του διμερούς, έτσι ώστε να πραγματοποιηθεί πυρηνόφιλη προσθήκη της φωσφορικής ομάδας στο κατάλοιπο αυτό της ιστιδίνης (Marina, Waldburger, & Hendrickson, 2005).

## 1.1.2. Ευκαρυωτικές κινάσες

Ανεξάρτητα από το αν μοιάζουν περισσότερο σε ένα πρότυπο, οι κινάσες στο σύνολό τους μοιράζονται στην αμινοξική τους αλληλουχία το ίδιο συντηρημένο κοινό καταλυτικό κέντρο και άρα, κοινό μηχανισμό κατάλυσης.

Συγκεκριμένα, η αντιπροσωπευτική δομή των κινασών αποτελείται από 2 διακριτούς, δομικά και λειτουργικά, λοβούς που συνεισφέρουν με μοναδικό τρόπο στην κατάλυση της φωσφορυλίωσης και χωρίζονται σε δώδεκα διακριτά αυτοτελείς δομικές υπό-περιοχές, οι οποίες αριθμούνται από το I έως το XII (Barker et al, 1982, Hanks et al, 1988).

Δομικές περιοχές κινασών :

**Ν-λοβός** : Σχηματίζεται από το αμινοτελικό άκρο (N-lobe) της αμινοξικής αλυσίδας και είναι μικρότερος από τον C-λοβό. Αποτελείται από μια β-πτυχωτή επιφάνεια 5 β-κλώνων και μια α-έλικα, την αC-έλικα.

Αναλυτικότερα, εντοπίζονται δύο υψηλά συντηρημένα μοτίβα αμινοξικής ακολουθίας στην περιοχή των τριών πρώτων β-κλώνων:

**Η πλούσια σε Γλυκίνη θηλιά (Glycine-rich loop - GxGxxG):** βρίσκεται ανάμεσα στους β1 και β2, διπλώνει πάνω από το νουκλεοτίδιο και προσανατολίζει την γ-φωσφορική ομάδα του ATP για την κατάλυση. Αποτελεί το πιο εύκαμπτο τμήμα του αμινοτελικού λοβού. Η κορυφή της θηλιάς κλείνει τελείως όταν η βάση της είναι αγκυροβολημένη στην γ-φωσφορική ομάδα του ATP. Οι δύο β-κλώνοι περικλείουν τον δακτύλιο αδενίνης κατά την αλληλεπίδραση. Επίσης, περιέχει μια άκρως συντηρημένη βαλίνη που αλληλεπιδρά με υδρόφοβες αλληλεπιδράσεις με τη βάση του ATP.

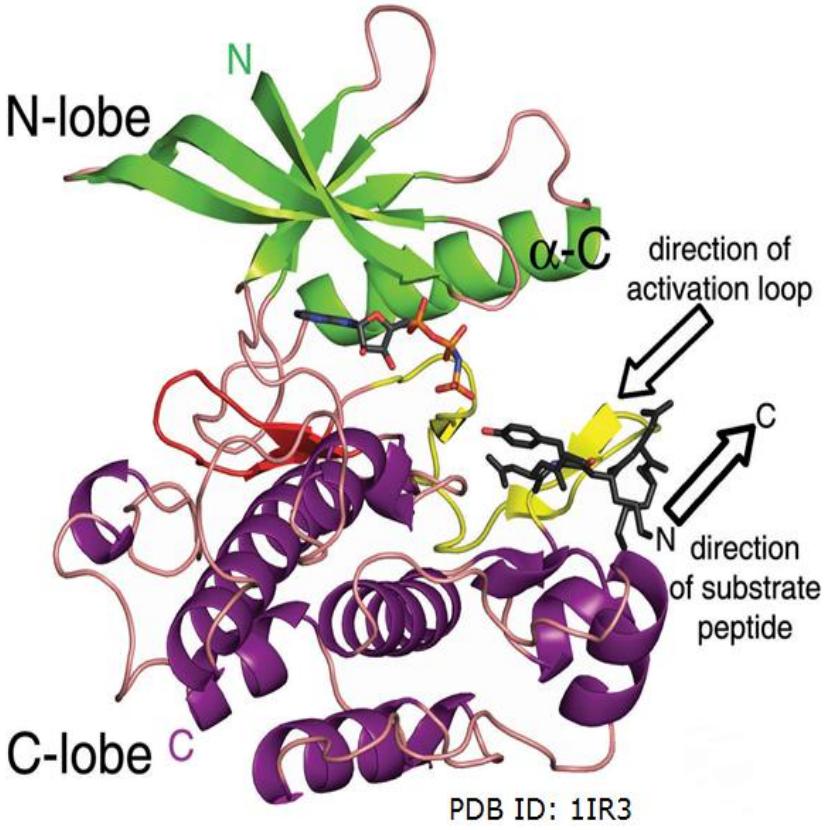
**Η P-θηλιά ή Walker-A μοτίβο (GxxxxGKT/S)**: Συμμετέχει στην αγκυροβόληση των φωσφορικών ομάδων χωρίς να έρθει σε επαφή με το πουρινικό τμήμα του ATP και περιέχει κατάλοιπα σερίνης ή θρεονίνης που αλληλεπιδρούν με την γ-φωσφορική ομάδα του ATP, σε αντίθεση με την πλούσια σε Γλυκίνη θηλιά.

**(AxK) μοτίβο:** Βρίσκεται στον 3 β-κλώνο. Περιέχει μια λυσίνη που συνδέει τις φωσφορικές ομάδες του ATP με την αC-έλικα.

Η αC-έλικα αποτελεί ένα μοναδικό και πολύ δυναμικό ρυθμιστικό στοιχείο για την πρωτεϊνική κινάση. Από πλευράς ακολουθίας ανήκει στον N-λοβό, αν και καταλαμβάνει μια στρατηγικά σημαντική θέση ανάμεσα στους 2 λοβούς. Συνδέεται με αρκετά διαφορετικά τμήματα του μορίου και έτσι λειτουργεί σαν 'μοτίβο ενσωμάτωσης του σήματος' (signal integration motif). Το καρβοξυτελικό της άκρο είναι αγκυροβολημένο στον στατικό κορμό (rigid body) του C-λοβού μέσω της θηλιάς αC–β4, ενώ το αμινοτελικό του άκρο αλληλεπιδρά με τη θηλιά ενεργοποίησης.

**C-λοβός**: Σχηματίζεται από το καρβοξυτελικό άκρο (C-lobe) της αμινοξικής αλυσίδας και είναι μεγαλύτερος από τον πρώτο και ιδιαίτερα σταθερός, γεγονός που τον καθιστά ικανό για την πρόσδεση του υποστρώματος. Κυριαρχούν στοιχεία α-έλικας καθώς και μία β-πτυχωτή επιφάνεια. Όσον αφορά τις έλικες αξίζει να σημειωθεί ότι τα αμίδια του σκελετού των ελίκων (D, E, F και H) είναι καλά προστατευμένα από τον διαλύτη, όπως αποδεικνύεται από πειράματα ανταλλαγής υδρογόνου/δευτερίου (H/D exchange). Εξαίρεση αποτελεί η G-έλικα, η οποία είναι εκτεθειμένη στον διαλύτη. Όσον αφορά τη β-πτυχωτή επιφάνεια αξίζει να αναφερθεί ότι φέρει 4 μικρούς σε μήκος β-κλώνους (6-9). Αποτελεί σημαντικό τμήμα του καταλυτικού μηχανισμού που σχετίζεται με την μεταφορά της φωσφορικής ομάδας από το ATP στο πρωτεϊνικό υπόστρωμα και είναι

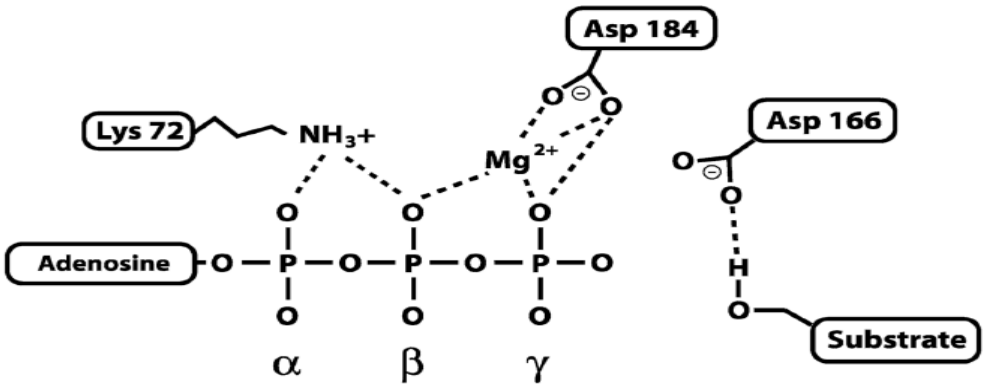
αγκυροβολημένη με τη βοήθεια υδροφοβικών καταλοίπων στον ελικοειδή πυρήνα. Η καταλυτική θηλιά που σχηματίζεται ανάμεσα στους β6 και β7 κλώνους εμφανίζει το μεγαλύτερο μέρος της καταλυτικής ενεργότητας, ενώ οι β8 και β9 κλώνοι πλαισιώνουν το DFG μοτίβο, όπου το ασπαρτικό είναι απαραίτητο για την αναγνώριση ενός εκ των προσδεμένων στο ATP ιόντων μαγνησίου.



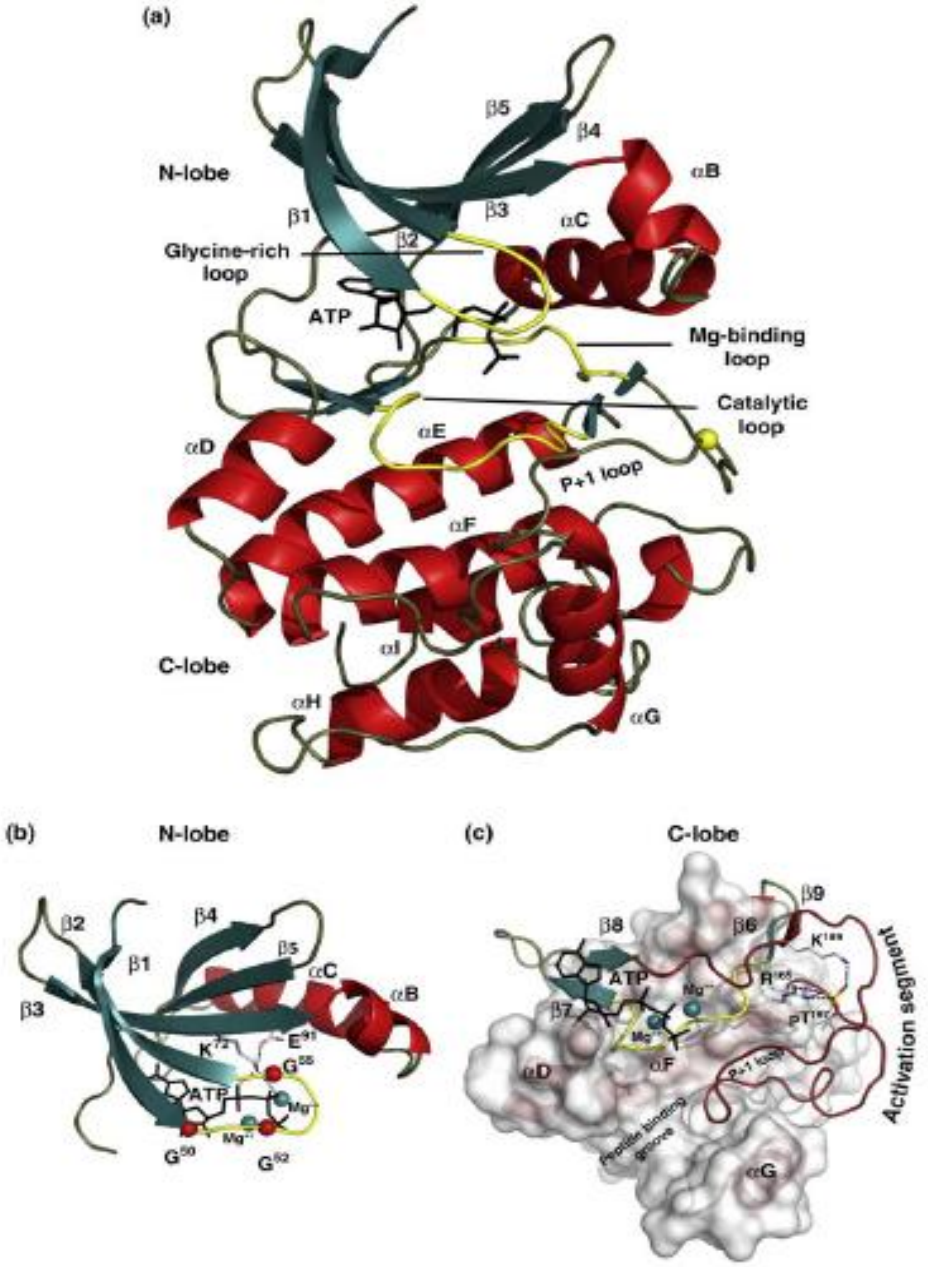

**Εικόνα 1.4** Δομή καταλυτικού κέντρου της κινάσης IRK (PDB 1IR3). Με πράσινο απεικονίζεται ο N-λοβός, με μοβ ο C-λοβός, με κίτρινο η θηλιά ενεργοποίησης, με μαύρο το πεπτίδιο-υπόστρωμα και ένα ανάλογο ATP. Σημειώνονται επίσης η έλικα-C που περιέχει κατάλοιπα σημαντικά για την κατάλυση καθώς και η έλικα G που συμμετέχει στον ελλιμενισμό του υποστρώματος. (Pellicena & Kuriyan, 2006)

Οι δύο λοβοί ενώνονται με ένα τμήμα ακανόνιστης δομής, μια σχετικά μικρή και μέτριας κινητικότητας θηλιά που ονομάζεται περιοχή αρμού (hinge region). Γύρω από την περιοχή αυτή δομείται το ενεργό κέντρο του ενζύμου. Σε άμεση γειτνίαση με το ενεργό κέντρο βρίσκονται τμήματα της πολυπεπτιδικής αλυσίδας που σχηματίζουν στοιχεία ακανόνιστης δευτεροταγούς δομής. Η διαμόρφωση των θηλιών είναι ιδιαίτερης σημασίας για την ρύθμιση της καταλυτικής λειτουργίας του ενζύμου.

Πιο συγκεκριμένα, το ενεργό κέντρο σχηματίζεται ως μια ευρύχωρη εσοχή με υδρόφοβα χαρακτηριστικά ανάμεσα στον αμινοτελικό και τον καρβοξυτελικό λοβό της κινάσης και οριοθετείτε από τις πλευρικές αλυσίδες και τον πρωτεϊνικό σκελετό αμινοξέων που ανήκουν και στους δύο λοβούς. Το μόριο του ATP προσδένεται στο ενεργό κέντρο μέσω δύο δεσμών υδρογόνου, οι οποίοι σχηματίζονται ανάμεσα στο δακτύλιο της πουρίνης του νουκλεοτιδίου και στον πρωτεϊνικό σκελετό της κινάσης.




**Εικόνα 1.5** Διαγραμματική απεικόνιση των καταλοίπων που συμμετέχουν στη μεταφορά της φωσφορικής ομάδας. Το Asp184 αλληλεπιδρά με το Mg2+, το οποίο είναι υπεύθυνο για τον σωστό προσανατολισμό της γ-φωσφορικής ομάδας και την κάλυψη του φορτίου της ώστε να μειώνονται οι ηλεκτροστατικές απωστικές δυνάμεις που θα δεχτεί το υδροξύλιο του υποστρώματος. Η Lys72 διευκολύνει τη μεταφορά και σταθεροποιεί τις αλληλεπιδράσεις με τις α και β φωσφορικές ομάδες του ATP. Το Asp166 βοηθάει στην σωστή τοποθέτηση του υδροξυλίου της πρωτεΐνης-υποστρώματος, μέσω της δημιουργίας δεσμών υδρογόνου. (Swulius & Waxham, 2008)




**Εικόνα 1.6** (α) Η χαρακτηριστική στερεοδιάταξη 2 λοβών. Ένα μόριο ATP εντοπίζεται προσδεμένο σε μια βαθιά εσοχή μεταξύ των 2 λοβών. Με κίτρινο απεικονίζονται σημαντικές για την κατάλυση θηλιές. (β) Δομή του Ν-λοβού: περιέχει 5 β-πτυχωτές επιφάνειες και μια υψηλά συντηρημένη aC-έλικα. Η πλούσια σε Gly θηλιά είναι υπεύθυνη για τον σωστό προσανατολισμό των φωσφορικών ομάδων του ATP. Με κόκκινες σφαίρες απεικονίζονται 3 συντηρημένες Gly. Η Lys72 από την 3$^{η}$ β-πτυχωτή επιφάνεια βοηθάει στο «ζευγάρωμα» των φωσφορικών ομάδων με την C-έλικα. (γ) Δομή του C-λοβού. Το τμήμα ενεργοποίησης περιέχει μια περιοχή φωσφορυλίωσης στην β9-πτυχωτη επιφάνεια (K189) και την R165 από το HRD μοτίβο. (Taylor SS & Kornev PA 2011)

Λειτουργικές περιοχές κινασών :

Στις περισσότερες περιπτώσεις μεταξύ της ενεργοποιημένης και της ανενεργής μορφής του ενζύμου υφίστανται σημαντικές δομικές διαφορές, όπως ο προσανατολισμός της C-έλικας στον χώρο και η στερεοδιάταξη των βρόχων οι οποίοι γειτνιάζουν με το ενεργό κέντρο (Huse & Kuriyan, 2002). Ο προσανατολισμός των άμινο- και καρβόξυ- τελικών άκρων κατά την ενεργοποίηση της κινάσης αποτελεί ένα από τα πιο σημαντικά βήματα για αποτελεσματική κατάλυση. Η απόσταση μεταξύ του αμινοτελικού άκρου της C-έλικας και του βρόχου ενεργοποίησης καθορίζει τις ανοιχτές και τις κλειστές στερεοδιατάξεις οι οποίες είναι απαραίτητες για κατάλυση. Αντίθετα, σε ανενεργές κινάσες το αμινοτελικό άκρο της C-έλικας συναντάται αρκετές φορές εγγενώς μη δομημένο ή απλά περιεστραμμένο σε έναν προσανατολισμό ακατάλληλο για κατάλυση.

Η C-έλικα περιέχει ακόμα ένα συντηρημένο κατάλοιπο γλουταμικού που δημιουργεί γέφυρα άλατος με την λυσίνη του β3-κλώνου. Η γέφυρα αυτή είναι υψηλά συντηρημένη σε ολόκληρη την οικογένεια των πρωτεϊνικών κινασών και συχνά θεωρείται σαν χαρακτηριστικό της ενεργής κατάστασης. Στην ενεργή κατάσταση, όπου η C-έλικα είναι δεσμευμένη στον πυρήνα της β-πτυχωτής επιφάνειας, ο N-λοβός μετακινείται σαν συμπαγές σώμα που ανοίγει και κλείνει ανάλογα με τις απαιτήσεις του κύκλου της κατάλυσης. Ο πλούσιος σε γλυκίνη βρόχος, αν και εύκαμπτος, μετακινείται μαζί με τον N-λοβό ανεξάρτητα από το αν η κινάση είναι ανοιχτή ή κλειστή. Μοναδική εξαίρεση του N-λοβού αποτελεί ο βρόχος αC–β4, ο οποίος παραμένει στενά αγκυροβολημένος στον C-λοβό.

**Ρυθμιστική Ραχοκοκαλιά** (R-spine): Υπάρχει ένα υψηλά συντηρημένο μοτίβο στο χώρο που εμφανίζεται σε κάθε ενεργή κινάση αλλά απουσιάζει από τις ανενεργές. Το μοτίβο αυτό αποτελείται από 4 μη-διαδοχικά υδρόφοβα κατάλοιπα, 2 του N-λοβού (Leu από τον β4 και Leu91 από την C-έλικα), και 2 από τον C-λοβό (Phe από το βρόχο ενεργοποίησης και Tyr από τον καταλυτικό βρόχο). Μπορεί να θεωρηθεί και ως η υδροφοβική ραχοκοκαλιά που συνδέει τους 2 λοβούς. Δεν είναι δυνατή η ταυτοποίησή της αποκλειστικά από την ακολουθία, και δεν έχει μελετηθεί επαρκώς ο ρόλος αυτών των 4 συντηρημένων καταλοίπων. Καθώς το ενδιάμεσο τμήμα αυτού του μοτίβου, δηλαδή η C-έλικα και ο βρόχος ενεργοποίησης, μπορεί να είναι αρκετά ευκίνητο η υδροφοβική ραχοκοκαλιά μπορεί δυναμικά να συγκροτηθεί και να ανασυγκροτηθεί, ρυθμίζοντας με αυτόν τον τρόπο την ενεργότητα της πρωτεϊνικής κινάσης.

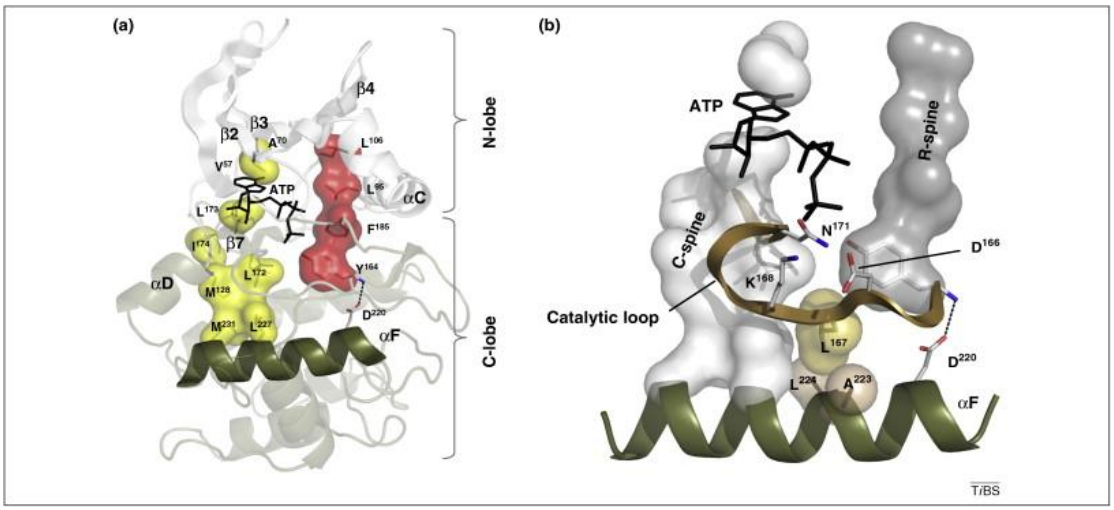
**Καταλυτική Ραχοκοκαλιά** (C-spine): Περιλαμβάνει κατάλοιπα και από τους δύο λοβούς. Διαφοροποιείται από την ρυθμιστική ραχοκοκαλιά στο γεγονός ότι ολοκληρώνεται με την προσθήκη του δακτυλίου αδενίνης του ATP. Δύο κατάλοιπα της καταλυτικής ραχοκοκαλιάς που εντοπίζονται στον N-λοβό, μία βαλίνη του β2-κλώνου και μία αλανίνη του AxK μοτίβου του β3-κλώνου, είναι συνδεδεμένα απευθείας στον δακτύλιο αδενίνης του ATP. Ακόμα, ένα κατάλοιπο λευκίνης του C-

λοβού είναι συνδεδεμένο απευθείας στον δακτύλιο αδενίνης, εντοπίζεται στο μέσο του β7-κλώνου και πλαισιώνεται από 2 υδρόφοβα κατάλοιπα.

**Υδρόφοβη F-έλικα**: Τα υψηλά ποσοστά υδρόφοβων καταλοίπων που περιέχει οδηγούν τους αλγορίθμους πρόβλεψης στην πρόβλεψή της ως διαμεμβρανική έλικα. Αποτελεί το στοιχείο οργάνωσης ολόκληρου του πυρήνα της κινάσης. Και οι δύο ραχοκοκαλιές που αναφέρθηκαν παραπάνω είναι αγκυροβολημένες στην F-έλικα, συγκεκριμένα, η καταλυτική ραχοκοκαλιά απευθείας στο υδρόφοβο καρβοξυτελικό της άκρο και η ρυθμιστική στον αμινοτελικό. Αρκετά μοτίβα του C-λοβού, είναι στενά αγκυροβολημένα μέσω υδρόφοβων αλληλεπιδράσεων στην F-έλικα.

**Καταλυτική θηλιά** : βρίσκεται αγκυροβολημένη στην F-έλικα. Δύο κατάλοιπα της θηλιάς (μία λευκίνη και μία προλίνη) συνδέονται απευθείας με την F-έλικα και αυτό έχει σαν αποτέλεσμα η κύρια αλυσίδα της θηλιάς να είναι συμπαγής τόσο στην αποφωσφορυλιωμένη όσο και στην ανενεργή κατάσταση.

**Θηλιά ενεργοποίησης** (activation loop): Εντοπίζεται κοντά στην καταλυτική θηλιά και είναι ιδιαίτερα εύκαμπτη. Οι περισσότερες κινάσες ενεργοποιούνται έπειτα από φωσφορυλίωση συγκεκριμένων καταλοίπων της θηλιάς ενεργοποίησης, τα οποία με τη σειρά τους δημιουργούν γέφυρες άλατος με τη θετικά φορτισμένη αργινίνη του μοτίβου HRD της καταλυτικής θηλιάς (αυτό έχει ως αποτέλεσμα την σταθεροποίηση της DFG-in στερεοδιάταξης που θα αναφερθεί παρακάτω). Οι περισσότερες κινάσες έχουν ένα κατάλοιπο θρεονίνης στο βρόχο της οποίας η φωσφορυλίωση απαιτείται για μέγιστη ενεργότητα. Λόγω του καταλοίπου αυτού η θηλιά ονομάζεται και T-θηλιά.

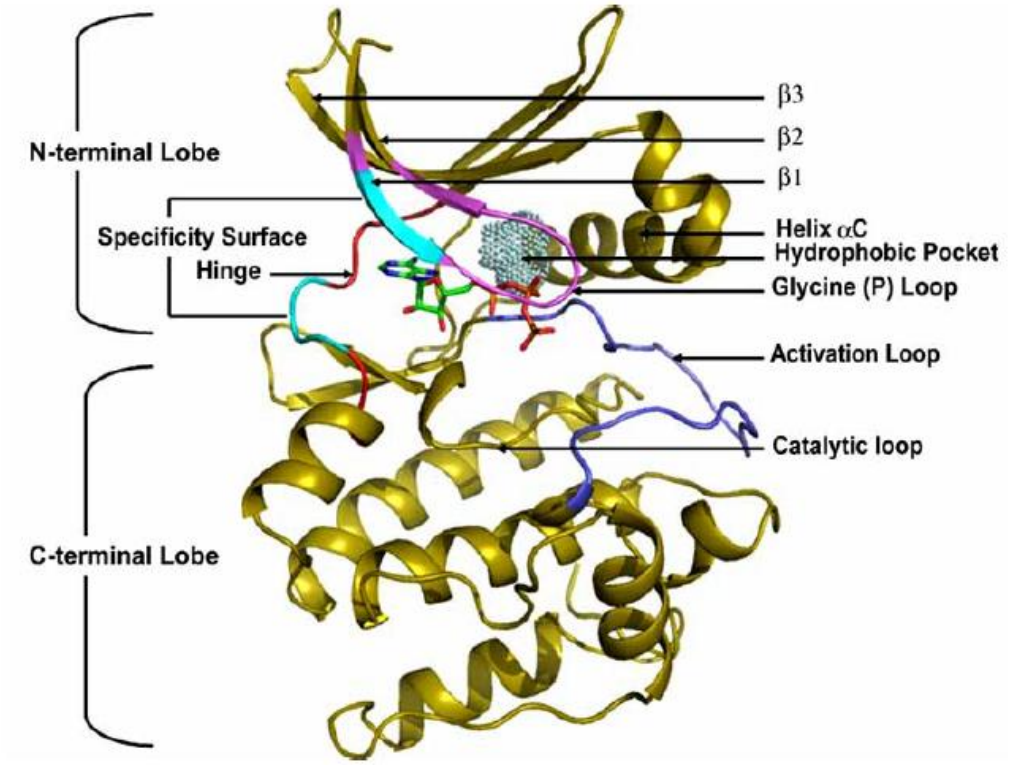



**Εικόνα 1.7** Η F έλικα και οι 2 υδρόφοβες ραχοκοκαλιές καθορίζουν τη βασική αρχιτεκτονική των πρωτεϊνικών κινασών. (a) Η ρυθμιστική ραχοκοκαλιά (κόκκινη επιφάνεια) και η καταλυτική ραχοκοκαλιά (κίτρινη επιφάνεια) είναι αγκυροβολημένες στην F-έλικα (πράσινο) στη μέση του στατικού κορμού του C-λοβού. Η καταλυτική ραχοκοκαλιά ολοκληρώνεται με το ATP και μαζί με την ρυθμιστική ραχοκοκαλιά εξασφαλίζουν ισχυρή αλλά εύκαμπτη σύνδεση μεταξύ των δύο λοβών. (b) Κύρια στοιχεία του ενεργού κέντρου

επίσης είναι αγκυροβολημένα στην F-έλικα είτε απ' ευθείας, είτε μέσω των δύο ραχοκοκαλιών. (Taylor SS & Kornev PA 2011).

**P+1 θηλιά** : εντοπίζεται αμέσως μετά τη θηλιά ενεργοποίησης. Εκεί συνήθως προσδένεται το κατάλοιπο του υποστρώματος που έπεται του φωσφορυλιωμένου (P+1 κατάλοιπο).

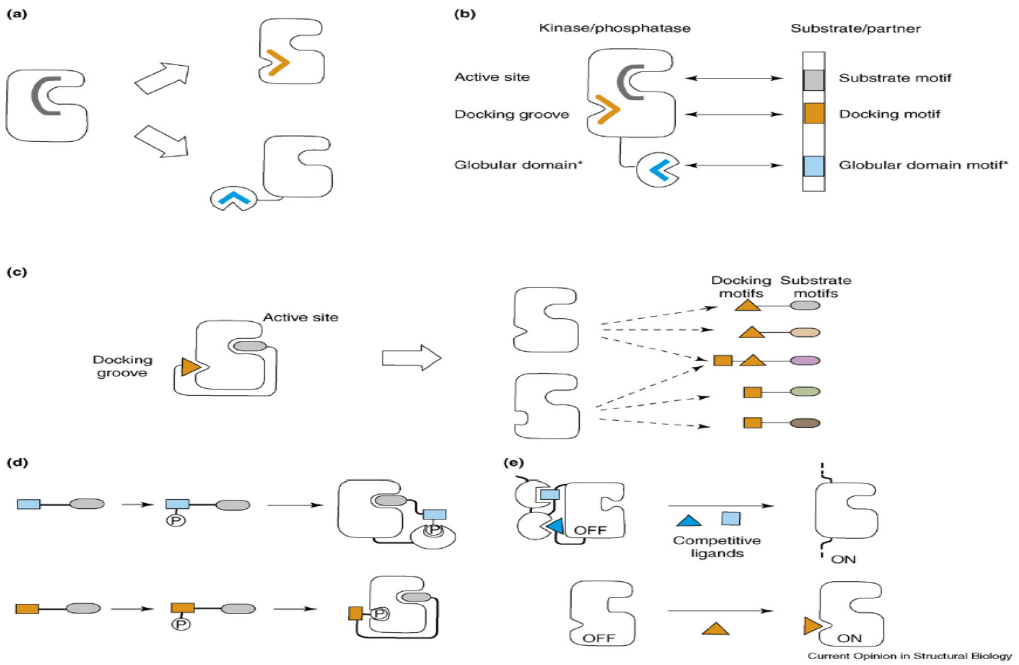
Η θηλιά ενεργοποίησης μαζί με την P+1 θηλιά συγκροτούν το **τμήμα ενεργοποίησης** (activation segment). Στο τμήμα αυτό εντοπίζεται το μοτίβο DFG που προαναφέρθηκε. Το κατάλοιπο ασπαρτικού συμμετέχει στον προσανατολισμό του $Mg^{2+}$ με τις φωσφορικές ομάδες του ATP. Στις περισσότερες κινάσες το τμήμα αποκτά δύο στερεοδιατάξεις: 1) την DFG-in στερεοδιάταξη η οποία παρατηρείται στην ενεργή κατάσταση των κινασών και το τμήμα εμφανίζεται ανεστραμμένο προς το εσωτερικό, και 2) την DFG-out στερεοδιάταξη η οποία παρατηρείται στην ανενεργή κατάσταση. Κατά την DFG-out στερεοδιάταξη η θηλιά ενεργοποίησης παρεμποδίζει τη διευθέτηση των καταλυτικών καταλοίπων, παρεμποδίζοντας την είσοδο του ATP (η πλευρική ομάδα της φαινυλαλανίνης εκτίθεται και εισχωρεί στη σχισμή πρόσδεσης του ATP) και του υποστρώματος. Αυτό έχει ως αποτέλεσμα το άνοιγμα μιας εκτενούς υδρόφοβης εσοχής κατάλληλης για πρόσδεση των τύπου II αναστολέων των κινασών (στοχεύουν τόσο την περιοχή πρόσδεσης του ATP όσο και της εσοχής που περιγράφηκε).




**Εικόνα 1.8** Δομή της PKA σε σύμπλοκο με το ATP (PDB: 1ATP). (Thaimattam, Banerjee, Miglani, & Iqbal, 2007).

## 1.1.3. Αλληλεπιδράσεις κινασών

Οι κινάσες αποτελούν μία από τις μεγαλύτερες γνωστές οικογένειες πρωτεϊνών. Ειδικότερα για τον άνθρωπο περίπου 1.7% των γονιδίων κωδικοποιούν 518 διαφορετικούς τύπους κινασών, οι οποίες κατηγοριοποιούνται ιεραρχικά σε 10 ομάδες, 134 οικογένειες και 201 υποοικογένειες κυρίως με βάση την ομολογία του καταλυτικού τους κέντρου.(Manning, Whyte, Martinez, & Hunter, 2002). Αυτή η ποικιλία των κινασών επιτρέπει τον ακριβή συντονισμό της ρύθμισης με βάση τη δράση τους σε συγκεκριμένο ιστό, σε συγκεκριμένο χρόνο ή υπόστρωμα.




**Εικόνα 1.9** Σύγκριση των αλληλεπιδράσεων που αναπτύσσονται μεταξύ της κινάσης και του υποστρώματος : (a)Το καταλυτικό κέντρο, εκτός από τις προτιμήσεις του ενεργού κέντρου ως προς το υπόστρωμα, μπορεί να απαιτεί επιπλέον αλληλεπιδράσεις είτε με DG είτε μέσω GD.(b) Η πρωτεΐνη που αλληλεπιδρά με την κινάση θα πρέπει να διαθέτει μικρά μοτίβα τα οποία αναγνωρίζονται και αλληλεπιδρούν με τα GD και DG αντίστοιχα. (c)η διαδικασία αυτή αυξάνει την εξειδίκευση των κινασών ως προς την αναγνώριση του υποστρώματός τους, καθώς επιλέγεται αυτό που διαθέτει το απαιτούμενο μοτίβο για ελλιμενισμό αυτού με το DG της επιφάνειας της κινάσης. (d)Πολλές από τις αλληλεπιδράσεις αυτές εξαρτώνται από την φωσφορυλίωση συγκεκριμένων καταλοίπων. (e) Αλληλεπιδράσεις ελλιμενισμού μπορούν επίσης να συμμετέχουν στην αλλοστερική ρύθμιση του καταλυτικού κέντρου των κινασών, διαμορφώνοντας κατάλληλα τη στερεοδιάταξή του. Χαρακτηριστικό παράδειγμα είναι οι περιπτώσεις αυτοαναστολής. . (Remenyi A., Good M. and Lim W.,2006)

Οι κινάσες ποικίλλουν και στον βαθμό της εξειδίκευσης. Εξειδικευμένες κινάσες φωσφορυλιώνουν μία κατηγορία υποστρωμάτων ενώ πολυλειτουργικές κινάσες τροποποιούν πολλούς διαφορετικούς στόχους και μπορούν να συντονίζουν διάφορες πορείες. Συγκρίσεις αμινοξικών ακολουθιών πολλαπλών θέσεων

φωσφορυλίωσης δείχνουν ότι μία πολυλειτουργική κινάση αναγνωρίζει συγγενικές αλληλουχίες. Ένα τέτοιο παράδειγμα αποτελεί η πρωτεϊνική κινάση A η οποία αναγνωρίζει την αλληλουχία R-R-X-S-Z ή R-R-X-T-Z, όπου X είναι ένα μικρό κατάλοιπο, Z είναι ένα μεγάλο υδρόφοβο κατάλοιπο και S ή T είναι η θέση της φωσφορυλίωσης. Πρέπει να σημειωθεί ότι αυτή η αλληλουχία δεν είναι απολύτως απαραίτητη, αφού η αντικατάσταση ενός από τα κατάλοιπα αργινίνης με λυσίνη επιφέρει μερική απώλεια της δραστικότητας. Μικρά συνθετικά πεπτίδια που περιέχουν ένα παρόμοιο μοτίβο φωσφορυλιώνονται σχεδόν πάντοτε από τις πρωτεϊνικές κινάσες σερίνης-θρεονίνης. Άρα, η εξειδίκευση καθορίζεται κατά κύριο λόγο από την αλληλουχία των αμινοξέων που περιστοιχίζουν τη θέση φωσφορυλίωσης. Όμως έχει παρατηρηθεί ότι και απομακρυσμένα κατάλοιπα μπορούν να συνεισφέρουν στην εξειδίκευση. Για παράδειγμα, αλλαγές στη στερεοδιάταξη της πρωτεΐνης μπορούν να μεταβάλουν την ευκολία προσέγγισης μιας πιθανής θέσης φωσφορυλίωσης.

Γενικά, το ενεργό κέντρο των κινασών διαθέτει περιορισμένες ικανότητες αναγνώρισης της πρωτεΐνης-υποστρώματος. Αν και μια τεράστια ποικιλία υποστρωμάτων πληρούν τις στερεοχημικές απαιτήσεις για κατάλυση, κάθε κινάση μπορεί και αναγνωρίζει συγκεκριμένη πρωτεΐνη-στόχο. Οι μηχανισμοί με τους οποίους οι κινάσες αλληλεπιδρούν εξειδικευμένα με άλλες πρωτεΐνες περιλαμβάνουν:

τον υποκυτταρικό συνεντοπισμό τους,

αλληλεπιδράσεις με τη βοήθεια μιας τρίτης πρωτεΐνης ικριώματος (βοηθάει στον συνεντοπισμό της κινάσης και του υποστρώματος και σε κάποιες περιπτώσεις το διμερισμό της κινάσης),

αλληλεπιδράσεις μέσω σφαιρικών αυτοτελών δομικών περιοχών (Globular Domains - GD) που εντοπίζονται εκτός του καταλυτικού κέντρου και που αναγνωρίζουν μικρά γραμμικά πεπτιδικά μοτίβα (3-10 αμινοξικά κατάλοιπα). Η διαδικασία αυτή κυριαρχεί στις κινάσες τυροσίνης. Έχει παρατηρηθεί επίσης η περίπτωση η κινάση να διαθέτει το μοτίβο αναγνώρισης από το GD και η πρωτεΐνη με την οποία αλληλεπιδρά το GD και

εξειδικευμένες συνδέσεις μέσω ενός αυλακιού που διαθέτουν στην επιφάνειά τους (Docking Groove - DG) που αναγνωρίζει μικρά γραμμικά πεπτιδικά μοτίβα (3-10 αμινοξικά κατάλοιπα) - διαδικασία που κυριαρχεί στις κινάσες σερίνης-θρεονίνης.

Οι διαθέσιμες πειραματικές μέθοδοι για τον εντοπισμό των περιοχών αλληλεπίδρασης κινασών με άλλες πρωτεΐνες είναι και επίπονες και υψηλού κόστους. Αυτό έχει οδηγήσει στην ανάπτυξη μιας σειράς υπολογιστικών μεθόδων που αφορούν την πρόβλεψη περιοχών φωσφορυλίωσης και βασίζονται στην αμινοξική ακολουθία. Κάποιες μέθοδοι προβλέπουν την περιοχή φωσφορυλίωσης

χωρίς να παρέχουν πληροφορία για την κινάση που πιθανά αλληλεπιδρά με την προς αναζήτηση πρωτεΐνη. Παρουσιάζονται αναφορικά κάποιες στον πίνακα 1.2.

**Πίνακας 1.2** Υπολογιστικές μέθοδοι πρόβλεψης περιοχών φωσφορυλίωσης ανεξάρτητες από την υπεύθυνη κινάση

| Υπολογιστική μέθοδος | Αναφορά |
|---|---|
| **DISPHOS** | (Iakoucheva et al., 2004) |
| **CKSAAP_PhSite** | (Zhao, Zhang, Xu, Ma, & Yin, 2012) |
| **PPRED** | (Biswas, Noman, & Sikder, 2010) |
| **Netphos** | (Blom, Gammeltoft, & Brunak, 1999) (MacDonald, Mackey, Pearson, & Haystead, 2002) (MacDonald et al., 2002) |
| **PHOSIDA** | (Gnad et al., 2007) |
| **AMS** | (D. Plewczynski, Tkacz, Wyrwicz, & Rychlewski, 2005) |

Άλλες μέθοδοι, εκτός από το να προβλέπουν αν εντοπίζεται στην προς αναζήτηση πρωτεΐνη περιοχή φωσφορυλίωσης παρέχουν και πληροφορία για την κινάση που πιθανά την φωσφορυλιώνει. Στον πίνακα 1.3 παρουσιάζονται αναφορικά κάποιες από τις μεθόδους αυτές.

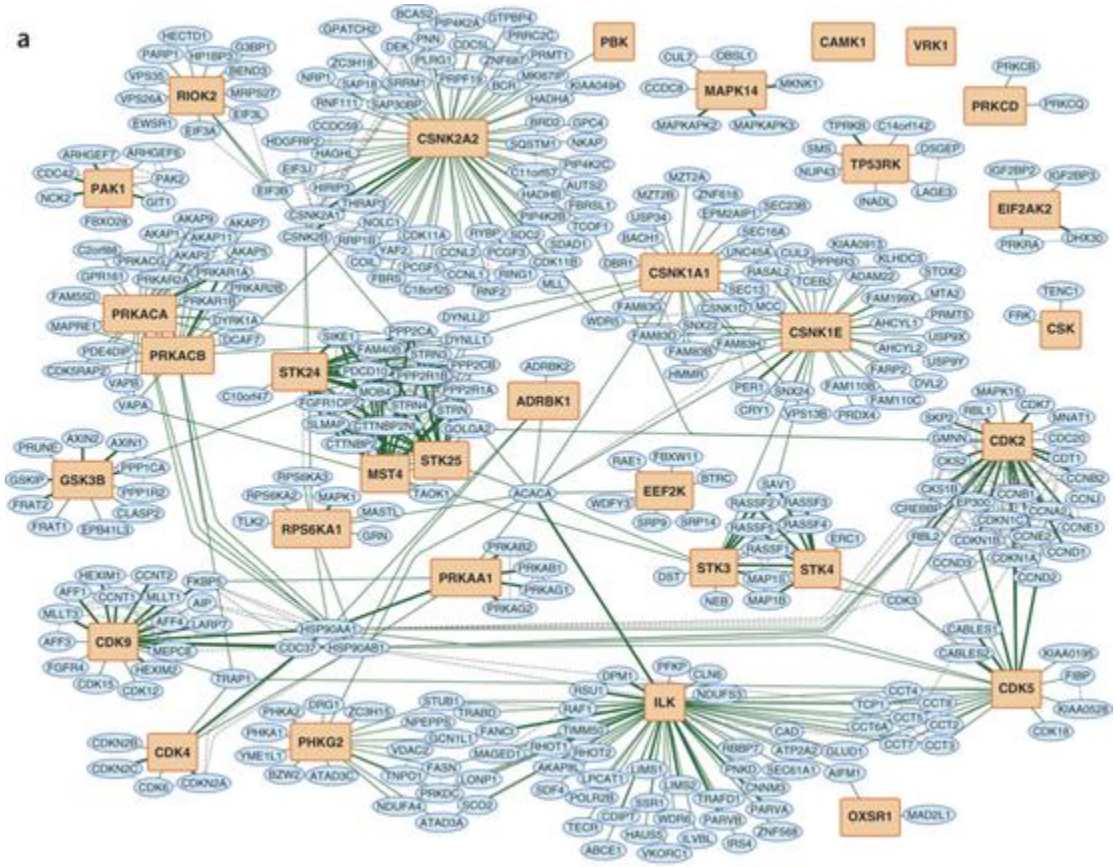
**Πίνακας 1.3** Υπολογιστικές μέθοδοι πρόβλεψης περιοχών φωσφορυλίωσης βασισμένοι στην εξειδίκευση κινασών.

| Υπολογιστική μέθοδος | Αναφορά |
|---|---|
| **Scansite** | (Obenauer, Cantley, & Yaffe, 2003) |
| **KinasePhos** | (Huang, Lee, Tzeng, & Horng, 2005) |
| **NetphosK** | (Blom, Sicheritz-Pontén, Gupta, Gammeltoft, & Brunak, 2004) |
| **PPSP** | (Yu Xue, Li, Wang, Feng, & Yao, 2006) |
| **GPS** | (Yu Xue et al., 2008) (Y. Xue et al., 2011) |
| **Postmod** | (Jung, Matsuyama, Yoshida, & Kim, 2010) |
| **BAE** | (Yu, Deng, Wong, & Tan, 2010) |
| **AMS 4.0 Server** | (Dariusz Plewczynski, Basu, & Saha, 2012) |
| **Metapred** | (Wan et al., 2008) |

Επίσης, πληθώρα συνδυαστικών υπολογιστικών μεθόδων έχει αναπτυχθεί τόσο για την πρόβλεψη του υποκυτταρικού εντοπισμού των πρωτεϊνών (Qian Xu, Pan, Xue, & Qiang Yang, 2011)(Liqi Li et al., 2012), γεγονός μέγιστης σημασίας για τη μελέτη αλληλεπίδρασης κινασών-πρωτεϊνών όπως αναφέρθηκε, όσο και για την μελέτη αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών γενικά. Οι τελευταίες θα αναφερθούν εκτενώς στην ενότητα 1.3.

Τέλος, οι αλληλεπιδράσεις κινασών – πρωτεϊνών έχουν μελετηθεί και με τη βοήθεια δικτύων αλληλεπίδρασης, με χρήση βιοχημικών δεδομένων ή δεδομένων που προκύπτουν με τη βοήθεια της κβαντικής χημείας, της μοριακής δυναμικής καθώς και της μεταγωγής σήματος (Herce, Deng, Helma, Leonhardt, & Cardoso, 2013). Τα δίκτυα αυτά ενισχύουν την περαιτέρω γνώση για τα βιοχημικά μονοπάτια που συμμετέχουν οι κινάσες καθώς και για τις ασθένειες με τις οποίες μπορεί να

σχετίζονται. Ένα τέτοιο δίκτυο παρατηρείται και στην παρακάτω εικόνα, όπου φαίνονται οι πειραματικά προσδιορισμένες αλληλεπιδράσεις κινασών (συμπαγής πράσινη γραμμή). Οι 32 κινάσες αναπαρίστανται με πορτοκαλί παραλληλόγραμμους κόμβους και οι πρωτεΐνες με τις οποίες αλληλεπιδρούν με μπλε κυκλικούς κόμβους. Οι παχιές γραμμές αναφέρονται σε αλληλεπιδράσεις που βρέθηκαν από βάσεις δεδομένων.




**Εικόνα 1.10** Δίκτυο προσδιορισμένων αλληλεπιδράσεων κινασών με πρωτεΐνες-στόχους.

## 1.2 Εισαγωγή στις υπό μελέτη ομάδες πρωτεϊνών

### 1.2.1. $Ca^{2+}$ ένας σημαντικός δεύτερος αγγελιαφόρος

Το ασβέστιο αποτελεί έναν από τους πιο ευρέως διαδεδομένους και σημαντικούς δεύτερους αγγελιαφόρους. Λειτουργεί ως δεύτερο μήνυμα σε μια ποικιλία κυτταρικών αποκρίσεων, τόσο στην περίπτωση που το πρώτο μήνυμα είναι χημικής, όσο και στην περίπτωση που είναι ηλεκτρικής φύσης. Ρυθμίζει καταρροϊκά μονοπάτια μεταγωγής σήματος, συμμετέχοντας σε πληθώρα βιολογικών διαδικασιών, όπως, η σηματοδότηση μέσω ορμονών ή παραγόντων ανάπτυξης, η ρύθμιση του κυτταρικού κύκλου, η γονιδιακή έκφραση και η απόπτωση. Τα σύμπλοκα του ασβεστίου με φωσφορυλιωμένες ή καρβοξυλιωμένες ενώσεις είναι συνήθως αδιάλυτα, αλλά και θεμελιώδη για πολλές βιοχημικές διεργασίες στο κύτταρο. Κατά συνέπεια, τα ενδοκυτταρικά επίπεδα ασβεστίου πρέπει να διατηρούνται χαμηλά έτσι ώστε να εμποδίζεται η καθίζηση αυτών των ενώσεων. Τα χαμηλά επίπεδα ασβεστίου συντηρούνται με δύο τρόπους :

Από τις ΑΤΡάσες-Ca που αποτελούν διαμεμβρανικές πρωτεΐνες (αντλίες) πρωτογενούς ενεργού μεταφοράς ασβεστίου. Εντοπίζονται στην κυτταρική μεμβράνη καθώς και στις μεμβράνες ορισμένων οργανιδίων (όπως σε αυτή του ενδοπλασματικού δικτύου) και είναι υπεύθυνες για την ενεργό απομάκρυνση του ασβεστίου από το κυτταρόπλασμα, είτε προς τον εξωκυττάριο χώρο, είτε προς τον αυλό κάποιου οργανιδίου και

Από αντιμεταφορείς Na-Ca που χρησιμοποιούν την κίνηση του νατρίου σύμφωνα με την διαβάθμιση της συγκέντρωσής του προκειμένου να αντλήσουν έξω από το κύτταρο τα ιόντα ασβεστίου, μέσω δευτερογενούς μεταφοράς (μηχανισμός που κυριαρχεί στα νευρικά κύτταρα και στα ραβδία).

Με τον τρόπο αυτό τα επίπεδα ασβεστίου στο κυτταρόπλασμα ανερέθιστων κυττάρων είναι τυπικά 100nM, συγκέντρωση περίπου 20.000 φορές μικρότερη από την εξωκυτταρική (η συγκέντρωσή του στο αίμα φτάνει στα 2.5 mM).

Η κυτταροπλασματική συγκέντρωση ασβεστίου μπορεί να αυξηθεί αιφνίδια ή σε κύματα (μπορεί να φτάσει στα 1-2 μM) μέσω :

Ενεργοποίησης ειδικών υποδοχέων, που οδηγεί στην διάνοιξη διαύλων ασβεστίου της κυτταρικής μεμβράνης. Ο υποδοχέας του μηνύματος, είτε συμπεριλαμβάνει το δίαυλο ασβεστίου στη δομή του, είτε ενεργοποιεί μια G πρωτεΐνη, η οποία με τη σειρά της οδηγεί, άμεσα ή έμμεσα, σε άνοιγμα του διαύλου.

Ενεργοποίησης ειδικών υποδοχέων (μέσω δεύτερων μηνυμάτων) που οδηγεί σε απελευθέρωση ασβεστίου στο κυτταρόπλασμα από το ενδοπλασματικό

δίκτυο (ενδοκυτταρική αποθήκη ασβεστίου) ως απόκριση σε δεύτερα μηνύματα, τα οποία μπορεί να είναι είτε το ίδιο το ασβέστιο (επαγόμενης από ασβέστιο απελευθέρωση ασβεστίου) είτε η $IP_3$.

Ενεργοποίησης ειδικών υποδοχέων (μέσω δεύτερων μηνυμάτων), που οδηγούν στην αναστολή της ενεργούς απομάκρυνσης ασβεστίου.

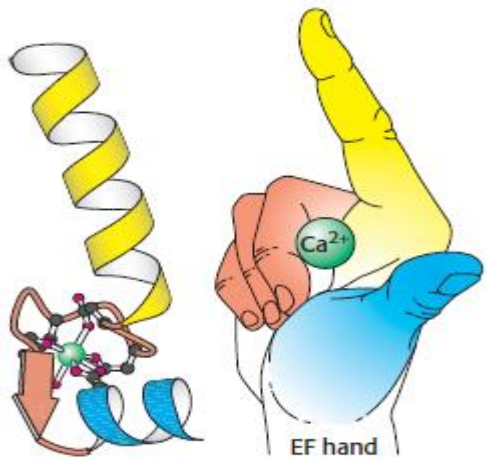
Διάνοιξης τασεο-ελεγχόμενων διαύλων ασβεστίου (εξαρτώμενων από το δυναμικό της μεμβράνης) ως απόκριση σε ηλεκτρικό ερέθισμα.

Όσον αφορά τη ρύθμιση της συγκέντρωσης ασβεστίου στο εσωτερικό του πυρήνα οι απόψεις διίστανται. Κάποιοι υποστηρίζουν πως ρυθμίζεται ανεξάρτητα από τη συγκέντρωση ασβεστίου στο κυτταρόπλασμα με τη μεσολάβηση αντλιών ασβεστίου του πυρηνικού φακέλου, ενώ άλλοι υποστηρίζουν πως η συγκέντρωση ασβεστίου αυξάνεται στο εσωτερικό του πυρήνα λόγω απότομης διάχυσης από το κυτταρόπλασμα. Το ασβέστιο προσδένεται, ισχυρά και επιλεκτικά, στις πρωτεΐνες στόχους και έτσι ρυθμίζεται η δράση τους. Η ικανότητα του ασβεστίου να συνδέεται ομοιοπολικά με 6-8 άτομα οξυγόνου το καθιστά ικανό να επάγει σημαντικές στερεοδιατακτικές αλλαγές στην πρωτεΐνη στόχο.

## 1.2.2. CaM πρωτεΐνες

Η καλμοντουλίνη αποτελεί τον πρωταρχικό ενδοκυτταρικό υποδοχέα ασβεστίου σχεδόν σε όλα τα ευκαρυωτικά κύτταρα. Είναι μια υψηλά συντηρημένη πρωτεΐνη 148 αμινοξέων (16,680 daltons). Στον άνθρωπο κωδικοποιείται από 3 διακριτά γονίδια (CALM1, CALM2, CALM3), κάθε ένα από τα οποία υπόκεινται σε διαφορετική ρύθμιση της έκφρασής του, ιστοειδικότητα και εναλλακτική συρραφή. Είναι εντυπωσιακό το γεγονός πως παράγουν ολόιδιες πρωτεΐνες (Berchtold, Egli, Rhyner, Hameister, & Strehler, 1993).

Η CaM αποτελείται από 4 μοτίβα έλικας-θηλιάς-έλικας που ονομάζονται EF-hand μοτίβα, 2 συγκροτούν την δομική αυτοτελή περιοχή του αμινοτελικού άκρου και τα άλλα 2 την αντίστοιχη του καρβοξυτελικού. Κάθε EF-hand είναι ικανό να δεσμεύει ιόντα $Ca^{2+}$ και αποτελείται από μια έλικα, την E, μια θηλιά και μια δεύτερη έλικα, την F. Απουσία του $Ca^{2+}$ οι CaM αποκτούν ποικιλία διαφορετικών στερεοδιατάξεων με τους α άνθρακες του σκελετού να εμφανίζουν μεταξύ τους RMSD της τάξης των 2 Å.

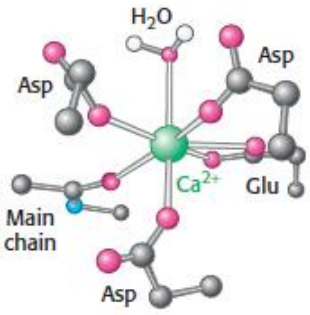



**Εικόνα 1.11** Απεικόνιση cartoon του EF hand μοτίβου (Με κίτρινο χρώμα απεικονίζεται η E έλικα, με μπλε η F και ως πράσινη σφαίρα το ασβέστιο) καθώς και απεικόνιση ανθρώπινου δεξιού χεριού, που αποτελεί και τον λόγο που ονομάστηκε έτσι το συγκεκριμένο μοτίβο.

Η CaM ενεργοποιείται από τη δέσμευση $Ca^{2+}$ όταν τα κυτοσολικά επίπεδα ασβεστίου υπερβαίνουν τα 500 nM. Πολλά από τα αμινοξικά κατάλοιπα που συμμετέχουν στη δέσμευση ασβεστίου βρίσκονται, κατάλληλα προσανατολισμένα, στην επιφάνεια των αυτοτελών δομικών περιοχών του αμινοτελικού και του καρβοξυτελικού άκρου. Με τη δέσμευση των ιόντων ασβεστίου σε κάθε EF-hand μοτίβο οι αυτοτελείς δομικές περιοχές αλλάζουν στερεοδιάταξη: οι θέσεις δέσμευσης του ασβεστίου στρέφονται προς το εσωτερικό για να δεσμεύουν $Ca^{2+}$ και τα υδρόφοβα κατάλοιπα μετακινούνται από το εσωτερικό στο εξωτερικό των περιοχών αυτών. Έτσι δημιουργούνται υδρόφοβα τμήματα στην επιφάνεια κάθε αυτοτελούς δομικής περιοχής, κατάλληλα για αλληλεπίδραση με άλλες πρωτεΐνες. Αναλυτικά η δέσμευση του ασβεστίου έχει σαν αποτέλεσμα οι E και F έλικες να πλησιάζουν μεταξύ τους. Αυτή η μικρή κίνηση μεταδίδεται και σε μια μικρή β-

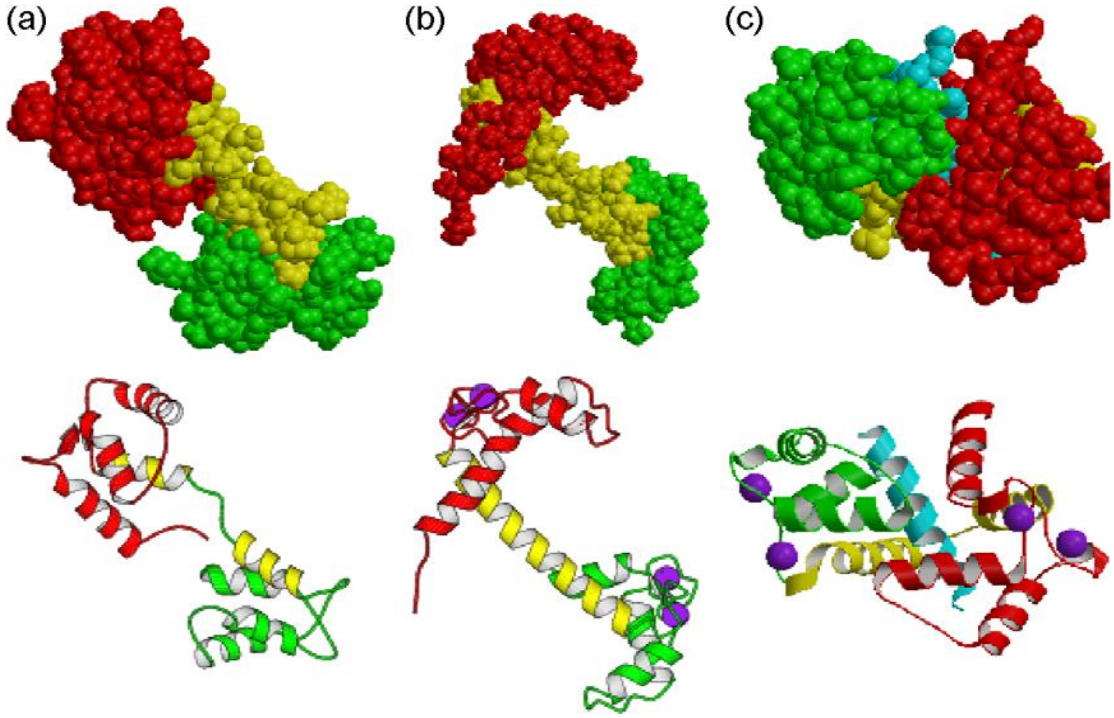
πτυχωτή επιφάνεια που ενώνει στο χώρο το κάθε EF hand με την δεύτερη περιοχή πρόσδεσης ασβεστίου της ίδιας αυτοτελούς δομικής περιοχής, αυξάνοντας έτσι τη συγγένεια πρόσδεσής της με το ασβέστιο. Πρόσδεση και του δεύτερου ασβεστίου έχει σαν αποτέλεσμα την απομάκρυνση των δυο EF hands μεταξύ τους, τόσο ώστε να αποκαλυφθεί μια υδρόφοβη θήκη στο εσωτερικό της κάθε δομικά αυτοτελούς περιοχής. Αναφέρεται πως, παρά το γεγονός ότι έχουν παρόμοιες δομές, τα 2 EF hands που συγκροτούν την C-τελική αυτοτελή δομική περιοχή διαθέτουν υψηλότερη συγγένεια πρόσδεσης για το ασβέστιο από ότι τα αντίστοιχα της N-τελικής. Η CaM έχει χαμηλή συγγένεια πρόσδεσης για το ασβέστιο (Kd=10 μm). Η συγγένεια πρόσδεσης όμως αυξάνεται κατακόρυφα με την αλληλεπίδραση της CaM με την πρωτεΐνη-στόχο. Η έκθεση των υδρόφοβων θηκών καθιστά ικανή την CaM να αναγνωρίζει τις περιοχές πρόσδεσης στην πρωτεΐνη-στόχο.

Επτά άτομα οξυγόνου συμμετέχουν στο σύμπλοκο με κάθε $Ca^{2+}$, τα 6 ανήκουν σε αμινοξικά κατάλοιπα της πρωτεΐνης και ένα σε ένα δεσμευμένο μόριο νερού. Αρνητικά φορτισμένα άτομα οξυγόνου (από τις πλευρικές αλυσίδες του γλουταμινικού και του ασπαραγινικού) και μη φορτισμένα άτομα οξυγόνου (από τις καρβονυλικές ομάδες των κύριων αλυσίδων και τις πλευρικές ομάδες της γλουταμίνης και της ασπαραγίνης) δεσμεύουν ισχυρά το $Ca^{2+}$.




**Εικόνα 1.12** Πρόσδεση του $Ca^{2+}$ στην CaM (δεσμοί που συγκρατούν το ασβέστιο σε κάθε EF hand)

Η δέσμευση του $Ca^{2+}$ από την CaM δεν προκαλεί μόνο τοπικές αλλαγές στην στερεοδιάταξη του κάθε EF hand, αλλά και συνολικά στην πρωτεΐνη, αυξάνοντας σημαντικά το ποσοστό των αμινοξέων που συμμετέχουν σε α-έλικες. Ποιο σημαντικός είναι ο σχηματισμός μιας έλικας-συνδέτη 8 στροφών, που ονομάζεται έλικα IV και συνδέει την N-τελική και C-τελική αυτοτελή δομική περιοχή, με αποτέλεσμα τον σχηματισμό μιας στερεοδιάταξης σχήματος αλτήρα. Οι τρισδιάστατες δομές της $Ca^{2+}$-CaM σε σύμπλοκο με πεπτίδια-στόχους υποδεικνύουν ότι, προκειμένου να είναι εφικτή η αλληλεπίδραση των υδρόφοβων περιοχών της N- και C-τελικής περιοχής με το πεπτίδιο-στόχο, η κεντρική έλικα υποβάλλεται σε $120^o$ χαλάρωση και $100^o$ κάμψη. Η έλικα αυτή επιτρέπει στην CaM να αλληλεπιδρά με πολύ διαφορετικές εξαρτώμενες από CaM πρωτεΐνες.




**Εικόνα 1.13** Σχηματική απεικόνιση, τόσο με cartoon (κάτω) όσο και χωροπληρωτικά μοντέλα (επάνω), 3 δομών της καλμοντουλίνης. Με κόκκινο απεικονίζεται η δομικά αυτοτελής περιοχή του Ν-τελικού άκρου, με πράσινο αυτή του C-τελικού άκρου, και κίτρινο η έλικα που συνδέει τις 2 περιοχές, με μοβ τα ιόντα ασβεστίου και με γαλάζιο το πεπτίδιο της MLCK α) apo-CaM (PDB: 1CFD) β) Ca-CaM (PDB: 1CLL) και γ) σύμπλοκο CaM με πεπτίδιο από την MLCK (PDB: 1CDL) (Yang, C. et al. 2004)

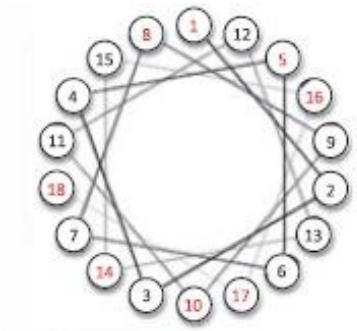
Το σύμπλοκο $Ca^{2+}$-CaM ελέγχει την ενεργότητα περισσότερων από 120 ενζύμων και άλλων πρωτεϊνών, μεταξύ των οποίων και άλλα μόρια όπως μεταγραφικοί παράγοντες, ιοντικοί δίαυλοι και αντλίες, φωσφατάσες και κινάσες. Μια από τις πιο σημαντικές πρωτεΐνες-στόχους είναι η οικογένεια των εξαρτώμενων από CaM κινασών.

Τα σύμπλοκα $Ca^{2+}$-CaM-κινάσες ρυθμίζουν πολύ διαφορετικές κυτταρικές λειτουργίες, όπως τη μεταγραφή των γονιδίων, τον κυτταρικό θάνατο (απόπτωση), την αναδιοργάνωση του κυτταροσκελετού, τη μάθηση, τη μνήμη και πολλές ακόμα.

Οι δομικές αυτοτελείς περιοχές πρόσδεσης στην CaM έχουν μήκος 9-26 κατάλοιπα και συνήθως είναι αμφιπαθείς α-έλικες που διασχίζουν το υδρόφοβο «τούνελ» που έχει δημιουργηθεί από την αποκάλυψη των υδρόφοβων θηκών σε κάθε EF hand ως αποτέλεσμα της πρόσδεσης του ασβεστίου, όπως προαναφέρθηκε. Κάθε μία από τις υδρόφοβες περιοχές της CaM περιέχει τέσσερα συντηρημένα κατάλοιπα μεθειονίνης, μερικά από τα οποία συμμετέχουν σε κρίσιμες επαφές με τις πρωτεΐνες-στόχους.

Στις περισσότερες περιπτώσεις η πρόσδεση της καλμοντουλίνης γίνεται με αντιπαράλληλο τρόπο. Η αυτοτελής δομική περιοχή του καρβοξυτελικού άκρου της καλμοντουλίνης αλληλεπιδρά με ένα υδρόφοβο κατάλοιπο αγκυροβόλησης του αμινοτελικού άκρου της ακολουθίας-στόχου και η αυτοτελής δομική περιοχή του αμινοτελικού άκρου με ένα υδρόφοβο κατάλοιπο αγκυροβόλησης του καρβοξυτελικού άκρου.

Παρά το γεγονός ότι δεν έχει καταγραφεί κάποια συντηρημένη ακολουθία αναγνώρισης των πρωτεϊνών στόχων, τα σύμπλοκα που δημιουργούνται εμφανίζουν υψηλή εξειδίκευση και συγγένεια αλληλεπίδρασης. Περισσότερες από 300 αλληλουχίες αναγνώρισης από την CaM είναι διαθέσιμες στην βάση δεδομένων Calmodulin Target Database (http://calcium.uhnres.utoronto.ca/ctdb/no_flash.htm). Η μεγάλη στερεοδιατακτική διαφορά της $Ca^{2+}$-CaM σε σχέση με την apo-CaM αυξάνει την πολυπλοκότητα του τρόπου με τον οποίο η καλμοντουλίνη αναγνωρίζει την πρωτεΐνη-στόχο, καθώς για τη δημιουργία του συμπλόκου απαιτούνται σημαντικές στερεοδιατακτικές αλλαγές τόσο στην CaM όσο και στην πρωτεΐνη-στόχο ώστε να επιτευχθεί ενεργοποίηση της τελευταίας. Επιπλέον, επειδή οι περισσότερες από τις πρωτεΐνες-στόχους είναι μεγάλες και πολυμερή είναι δύσκολη η μελέτη των συμπλόκων σε μοριακό επίπεδο. Κατά συνέπεια, η διαθέσιμη δομική πληροφορία έχει αποκτηθεί κυρίως από σύμπλοκα της CaM με μικρά πεπτίδια των πρωτεϊνών-στόχων.




**Εικόνα 1.14** Σχηματική αναπαράσταση σε μορφή Helical wheel στην οποία εμφανίζονται οι θέσεις των καταλοίπων που συμμετέχουν στην αγκυροβόληση (κόκκινο) σε μια ιδανική έλικα. Σε μοντέλα αλληλεπίδρασης με περισσότερα από ένα κατάλοιπα (π.χ. 1-5-10), η πρώτη και η τελευταία θέση συνήθως αποτελούν κατάλοιπα αγκυροβόλησης.

**Πίνακας 1.4** Δομές συμπλόκων CaM – πρωτεΐνη στόχος

| Κωδικός PDB | Πρωτεΐνη - στόχος | Οργανισμός | Θέσεις σημαντικών για την αλληλεπίδραση καταλοίπων | Μέθοδος |
|---|---|---|---|---|
| **1CDM** | CaMK II | B. taurus | 1-5-10 | X-ray |
| **1CM1/1CM4** | CaMK IIα | B. taurus | 1-5-10 | X-ray |
| **3GP2** | CaMK II δ | H. sapiens | 1-5-10 | X-ray |
| **2WEL** | CaMK II δ | H. sapiens | 1-5-10 | X-ray |
| **1CKK** | CaMKK | R. norvegicus | 1-10-16 | NMR |
| **1IQ5** | CaMKK | C. elegans | 1-10-16 | X-ray |
| **1MXE** | CaMK I | R. norvegicus | 1-(10)-14 | X-ray |
| **2L7L** | CaMK I α | R. norvegicus | 1-10-14 | NMR |
| **1CDL** | MLCK | G. gallus | 1-14 | X-ray |
| **1QS7** | MLCK | G. gallus | 1-14 | X-ray |
| **1QTX** | MLCK | G. gallus | 1-14 | X-ray |
| **2BBM** | MLCK | D.melanogaster | 1-14 | NMR |
| **2K0F** | MLCK | H. sapiens | 1-14 | NMR |
| **2O5G** | MLCK | G. gallus | 1-14 | NMR |
| **1WRZ/1YR5/1ZUZ** | DAPK-1 | H. sapiens | 1-(10)-14 | X-ray |
| **2X0G** | DAPK-1 | H. sapiens | 1-14 | X-ray |
| **2Y4V** | DAPK-1 | H. sapiens | 1-(10)-14 | X-ray |

## 1.2.3. Εξαρτώμενες από $Ca^{2+}$/CaM κινάσες

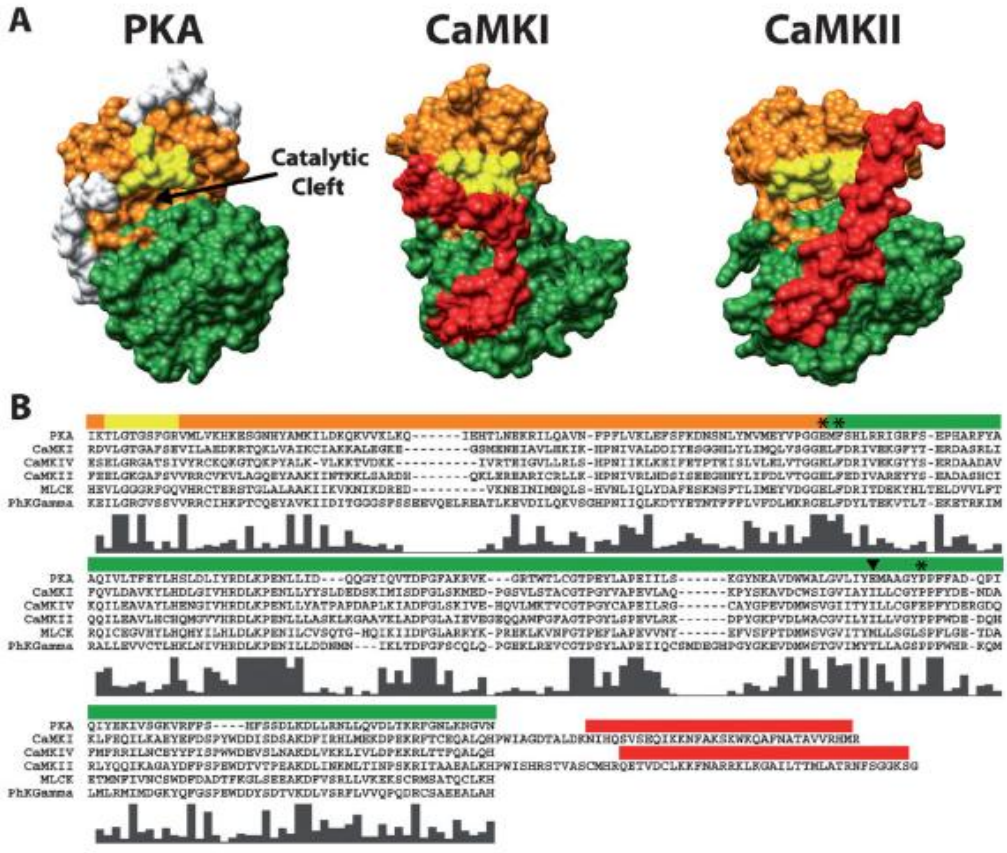
Μια από τις λειτουργίες της CaM είναι να ενεργοποιεί καταρροϊκά μέλη μιας οικογένειας πρωτεϊνικών κινασών Ser/Thr που ονομάζονται εξαρτώμενες από $Ca^{2+}$/CaM κινάσες. Κατηγοριοποιούνται σε 2 κατηγορίες με βάση το αν φωσφορυλιώνουν ένα και μοναδικό υπόστρωμα (όπως οι CaMK III, οι MLCK και οι Phosphorylase κινάσες) ή αν φωσφορυλιώνουν ποικιλία διαφορετικών υποστρωμάτων – πολυλειτουργικές κινάσες (όπως οι CaMK I, CaMK II, CaMK IV και DAPK)(Hook, S. & Means, A.,2001).

Η γενική αρχιτεκτονική της δομής τους περιλαμβάνει :

Μια αυτοτελή δομική περιοχή που αποτελεί το καταλυτικό κέντρο με ενεργότητα πρωτεϊνικής κινάσης στο αμινοτελικό άκρο.

Μια ρυθμιστική περιοχή, που περιέχει μια αυτοτελή δομική περιοχή υπεύθυνη για την αυτοαναστολή και

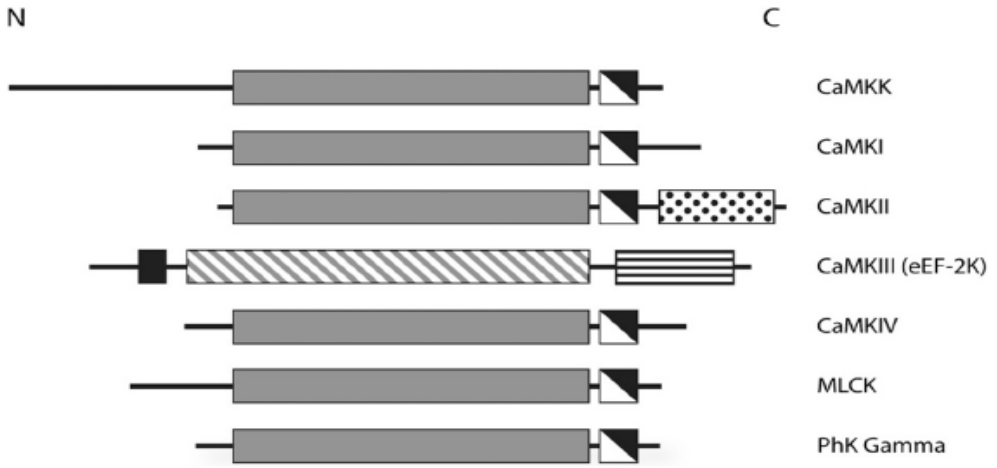
μια αυτοτελή δομική περιοχή αλληλεπίδρασης με την CaM μεταξύ των οποίων παρατηρείται επικάλυψη.




**Εικόνα 1.15** (α) Απεικονίζονται οι συντηρημένες δομικές και λειτουργικές περιοχές των περιοχών με δραστικότητα πρωτεϊνικής κινάσης των πρωτεϊνών: PKA (PDB 1J3H), CaMK I (PDB 1A06) και CaMK II (PDB 1J3H). Με πορτοκαλί απεικονίζεται ο λοβός πρόσδεσης ATP, με πράσινο ο λοβός πρόσδεσης υποστρώματος και με κίτρινο η πλούσια σε γλυκίνη θηλιά. Με κόκκινο απεικονίζονται οι περιοχές αυτοαναστολής των CaMK I και CaMK II, ενώ με

λευκό τα κατάλοιπα του άμινο- και καρβόξυ- τελικού άκρου της PKA που δεν εμφανίζονται συντηρημένα στις CaMK. Παρατηρείται η συντηρημένη δομή των 2 λοβών με την καταλυτική θήκη μεταξύ τους. (β) Πολλαπλή στοίχιση αμινοξικών ακολουθιών. Τα χρώματα που εμφανίζονται πάνω από τις ακολουθίες ακολουθούν τον χρωματικό κώδικα των δομών. Τα ιστογράμματα απεικονίζουν τον υψηλό βαθμό συντήρησης. Τα κατάλοιπα που σημειώνονται είναι συντηρημένα στις ειδικές για το υπόστρωμα CaMK και συμμετέχουν στην πρόσδεση υποστρώματος.(Swulius & Waxham, 2008)

Στις κινάσες που συγκροτούν πολυμερή (όπως οι phosphorylase κινάσες και οι CaMKII) αμέσως μετά, σε επίπεδο ακολουθίας, εντοπίζεται η περιοχή που είναι υπεύθυνη για τον πολυμερισμό τους.




**Εικόνα 1.16** Απεικονίζονται οι συντηρημένες δομικά αυτοτελείς περιοχές, με γκρι η καταλυτική περιοχή ενεργότητας κινάσης, με λευκό η περιοχή αυτοαναστολής και με μαύρο η περιοχή αλληλεπίδρασης με την CaM. Η καταλυτική περιοχή των CaMK III διαφέρει σημαντικά σε σχέση με τις υπόλοιπες, ενώ στο καρβοξυτελικό της άκρο απεικονίζεται η περιοχή αλληλεπίδρασης με τον eEF-2. Στο καρβοξυτελικό άκρο της CaMK II απεικονίζεται η περιοχή που συμμετέχει στον πολυμερισμό τους.(Swulius & Waxham, 2008)

Σε χαμηλά επίπεδα ασβεστίου παραμένουν στην ανενεργή κατάσταση, με την περιοχή υπεύθυνη για αυτοαναστολή να αλληλεπιδρά με τέτοια τρόπο ώστε να μπλοκάρει την πρόσδεση του υποστρώματος ή να αλλάζει την στερεοδιάταξη του ενεργού κέντρου με τρόπο που αυτό καθίσταται μη λειτουργικό. Πολλές φορές παρατηρείται συνδυασμός αυτών των 2 τεχνικών.

Αύξηση των επιπέδων ασβεστίου έχει ως αποτέλεσμα στερεοδιατακτικές αλλαγές στην CaM, όπως αναφέρθηκε, οι οποίες τις επιτρέπουν να αλληλεπιδρά με την εκάστοτε κινάση. Η Ca2+/CaM ενεργοποιεί τα ένζυμα μέσω της πρόσδεσής της στην περιοχή αναγνώρισής της και κατ’επέκταση απελευθέρωση της περιοχής υπεύθυνης για αυτοαναστολή από το καταλυτικό κέντρο της κινάσης. Αυτό οδηγεί στην ενεργοποίηση της κινάσης και στην επακόλουθη πρόσδεση του ATP και του υποστρώματος με στόχο να διεκπεραιωθεί επιτυχώς η αντίδραση φωσφορυλίωσης

Κάποιες από τις κινάσες είναι ικανές να γίνονται πλήρως ανεξάρτητες ως προς την $Ca^{2+}$/CaM έπειτα από την ενεργοποίησή τους ενώ άλλες απαιτούν περαιτέρω τροποποιήσεις, όπως φωσφορυλίωση, για να ενεργοποιηθούν πλήρως. Αυτές οι διαφορές όσον αφορά την ρύθμισή τους επιτρέπουν σε μια τόσο μικρή ομάδα πρωτεϊνών τον τόσο εκλεπτυσμένο έλεγχο ποικιλίας κυτταρικών λειτουργιών.

Οι εξαρτώμενες από CaM κινάσες κατηγοριοποιούνται ως εξής: στις εξειδικευμένες ως προς το υπόστρωμα κινάσες (όπως οι CaMKIII, phosphorylase κινάσες και MLCK) και στις πολυλειτουργικές (όπως οι CaMKK, CaMKI, CaMKII και CaMKIV). Ενεργοποίηση των πολυλειτουργικών κινασών μπορεί να επηρεάσει πληθώρα μονοπατιών καταρροϊκά ενώ οι εξειδικευμένες κινάσες συνήθως υπηρετούν μια καθορισμένη λειτουργία στο κύτταρο ή στον ιστό στον οποίο εκφράζονται.

Δεν είναι γνωστός ο ακριβής μηχανισμός μέσω του οποίου οι εξειδικευμένες εξαρτώμενες από CaM κινάσες εμφανίζουν τόσο υψηλή εξειδίκευση και ξεκάθαρη διάκριση υποστρωμάτων σε σχέση με τις πολυλειτουργικές, καθώς είναι ελλιπής η δομική πληροφορία σχετικά με τις κινάσες σε σύμπλοκο με τα υποστρώματα τους. Εντοπίζεται η συντηρημένη ακολουθία υποστρώματος Arg-X-X-Ser/Thr που χρησιμοποιείται και για πρόβλεψη πιθανών θέσεων φωσφορυλίωσης στο εσωτερικό πρωτεϊνών, όμως εντοπίζονται και εξαιρέσεις (Kennelly & Krebs, 1991)(Pearson & Kemp, 1991). Εκτός από την αμινοξική ακολουθία των περιοχών αναγνώρισης υποστρώματος στην αναγνώριση παίζουν σημαντικό ρόλο και οι διαφορετικές στερεοδιατάξεις των υποστρωμάτων.

**Πίνακας 1.5** Ιδιότητες πολυλειτουργικών εξαρτώμενων από $Ca^{2+}$/CaM κινασών

| | CaMK I | CaMK IV | CaMK II |
|---|---|---|---|
| **Ιστοειδικότητα** | Δεν εμφανίζει | Εμφανίζει αυξημένη | Δεν εμφανίζει, αλλά η κατανομή των ισομορφών εμφανίζει ιστοειδικότητα |
| **Υποκυτταρικός εντοπισμός** | Κυτταρόπλασμα | Πυρήνας και κυτταρόπλασμα | Κυρίως κυτταρόπλασμα, έκτος από τις ισομορφές $\alpha_B$, $\gamma_A$ και $\delta_B$ που εντοπίζονται και στον πυρήνα |
| **Σύσταση των υπομονάδων** | Μονομερές | Μονομερές | Όμο- ή έτερο-πολυμερές |
| **Γνωστές απαιτήσεις για πλήρη ενεργοποίηση** | Σύνδεση με $Ca^{2+}$/CaM<br>Θηλιά ενεργοποίησης Φωσφορυλίωση | Σύνδεση με $Ca^{2+}$/CaM<br>Θηλιά ενεργοποίησης Αυτοφωσφορυλίωση | Σύνδεση με $Ca^{2+}$/CaM |
| **Λειτουργεί ανεξάρτητα από $Ca^{2+}$/CaM** | Όχι | Ναι (μέχρι και 20% από την εξαρτώμενη από $Ca^{2+}$/CaM λειτουργικότητα) | Ναι (μετά από αυτοφωσφορυλίωση μέχρι και 80% από την εξαρτώμενη από $Ca^{2+}$/CaM λειτουργικότητα |

| | | | |
|---|---|---|---|
| **$K_{CaM}$** | 14 nM (αποφωσφορυλιωμένη) 4 nM (με φωσφορυλιωμένη τη θηλιά ενεργοποίησης) | 158 nM(recombinant) 32 nM (tissue purified) | 20–100 nM μειώνεται στα 60 pM με αυτοφωσφορυλίωση |
| **Συναινετική ακολουθία υποστρώματος** | Hyd-X-R-X-X-S/T-X-X-X-Hyd | Hyd-X-R-X-X-S/T | Hyd-X-R-NB-X-S/T |

### 1.2.3.1. CaMK I κινάσες

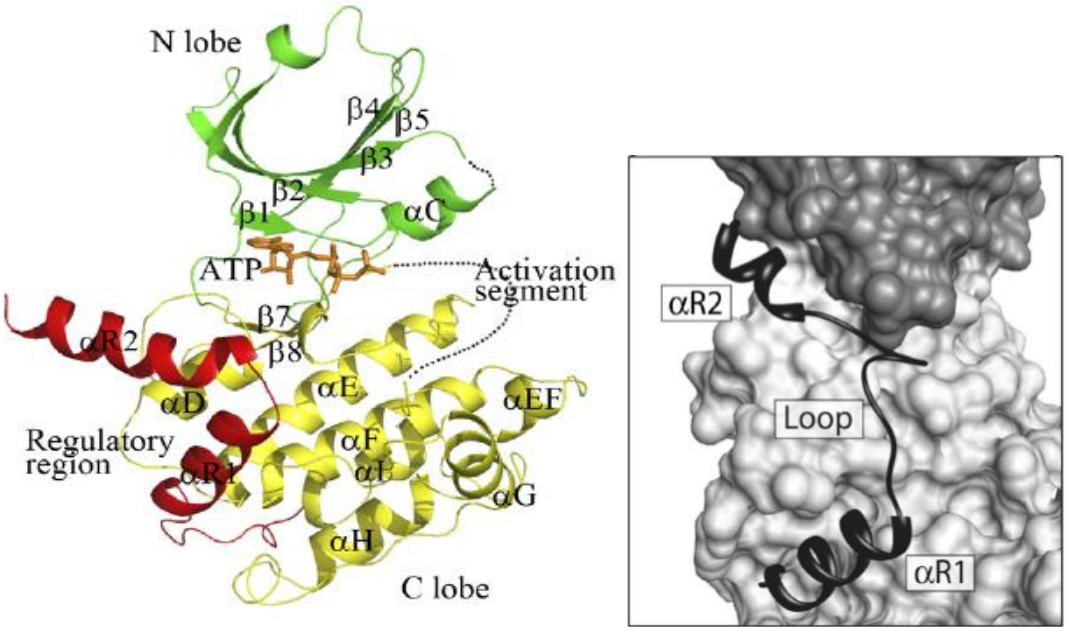
Η ομάδα των CaMKI κινασών στα θηλαστικά αποτελείται από 4 μέλη, που κωδικοποιούνται από 4 διαφορετικά γονίδια : CAMK1 ή CAMK1α, PNCK ή CAMK1β, CAMK1γ και CAMK1δ. Εντοπίζεται ως μονομερές των 42 kDa και εκφράζεται στα κύτταρα σχεδόν όλων των ιστών. Αποτελείται από 374 αμινοξικά κατάλοιπα. Η CaMKI αποτελεί έναν από τους 2 καταρροϊκούς στόχους του καταρράκτη των CaMK.

Τα κατάλοιπα 10-275 συμμετέχουν στο σχηματισμό του καταλυτικού κέντρου ενώ τα κατάλοιπα 276-316 συμμετέχουν στο σχηματισμό της ρυθμιστικής περιοχής, η οποία αποτελείται από την δομική αυτοτελή περιοχή πρόσδεσης της CaM και τη δομική αυτοτελή περιοχή αυτοαναστολής. Το καταλυτικό κέντρο της CaMK I εμφανίζει κάποιες διαφορές από το αντίστοιχο της PKA, που αναλύθηκε στην προηγούμενη ενότητα. Εμφανίζεται εκτεταμένο, με τους N- και C-λοβούς απομακρυσμένους σε γωνία $18^{o}$ μεταξύ τους, γεγονός που εκθέτει το ενεργό κέντρο στο διαλύτη. Επιπλέον, η θηλιά ενεργοποίησης της CaMK I εμφανίζεται εγγενώς μη δομημένη. Η δομή της CaMK I φανερώνει πως για να ενεργοποιηθεί απαιτείται τόσο η αποδέσμευση της δομικά αυτοτελούς περιοχής από το ενεργό κέντρο, όσο και μια περιστροφή μεταξύ των 2 λοβών και πιθανά φωσφορυλίωση της θηλιάς ενεργοποίησης. Αξίζει να σημειωθεί πως δομικά ανάλογα των σημαντικών για την πρόσδεση του υποστρώματος καταλοίπων Glu127, Phe129, Glu230 και Pro236 της PKA είναι τα Glu102, Phe104, Ile210 και Pro216 για την CaMKI. Το Glu102 αλληλεπιδρά με βασικά κατάλοιπα στην περιοχή αναγνώρισης υποστρώματος ενώ τα Phe104, Ile210 και Pro216 σχηματίζουν μια υδρόφοβη θήκη που αναγνωρίζει υδρόφοβα κατάλοιπα στις πρωτεΐνες-υποστρώματα.

Για την ενεργοποίηση του μονομερούς απαιτείται τόσο η πρόσδεση της $Ca^{2+}$/CaM για την απελευθέρωση του ενεργού κέντρου από την περιοχή υπεύθυνη για την αυτοαναστολή, όσο και η φωσφορυλίωση της Thr-l77 στην θηλιά ενεργοποίησης από τις CaMKK. Η πρόσδεση της $Ca^{2+}$/CaM είναι απαραίτητη και για την έκθεση της Thr177 που εντοπίζεται στη θηλιά ενεργοποίησης και αποτελεί υπόστρωμα για φωσφορυλίωση από την CaMKK. Όταν φωσφορυλιώνεται η CaMKI στην Thr177, η δραστικότητά της αυξάνεται 10 με 20 φορές, αλλά, σε αντίθεση με την CaMKK, η CaMKI εξακολουθεί να εξαρτάται από την $Ca^{2+}$/CaM. Όταν η

συγκέντρωση ασβεστίου επανέρχεται σε βασικά επίπεδα, η ενεργότητα της CaMKI θα μειωθεί ανεξάρτητα από τα αν η Thr177 βρίσκεται φωσφορυλιωμένη ή αποφωσφορυλιωμένη. Σε αυτό το σημείο αξίζει να αναφερθεί πως η ισομορφή CaMKIδ εμφανίζει, λόγω της μοναδικής δομής του αμινοτελικού της άκρου, ασυνήθιστη αντίσταση στην αποφωσφορυλίωσή της. Αυτό σημαίνει πως παραμένει φωσφορυλιωμένη, και άρα ενεργή, ακόμα και σε χαμηλές συγκεντρώσεις ασβεστίου, εξασφαλίζοντας έτσι την διαρκή ρύθμιση της μεταγωγής σήματος στους νευρώνες (Senga, Ishida, Shigeri, Kameshita, & Sueyoshi, 2015).

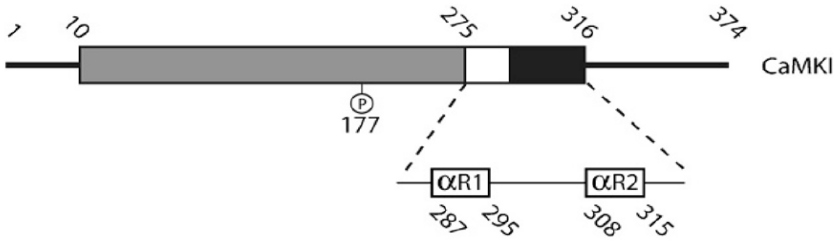
Το 1996, η CaMKI ήταν η πρώτη κινάση από την οικογένεια των CaMK που προσδιορίστηκε με περίθλαση ακτίνων-Χ (PDB : 1A06) (Goldberg, Nairn, & Kuriyan, 1996). Το γεγονός ότι η κινάση ήταν στην ανενεργή κατάσταση επέτρεψε την κατανόηση του μηχανισμού καταστολής της ενεργότητας της κινάσης μέσω αλληλεπίδρασης της περιοχής αυτοαναστολής με το ενεργό κέντρο. Αρχικά πίστευαν πως η αυτοαναστολή επιτυγχάνεται επειδή η υπεύθυνη δομικά αυτοτελής περιοχή της κινάσης δρα σαν ψευδοϋπόστρωμα, μιμούμενη την φυσιολογική πρωτεΐνη-υπόστρωμα της κινάσης, μπλοκάροντας το ενεργό κέντρο λόγω στερεοχημικής παρεμπόδισης.



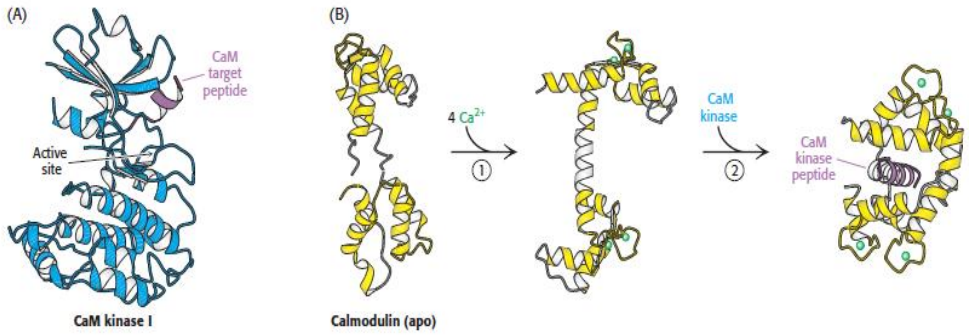

**Εικόνα 1.17** Δομή της ανενεργής CaMK I (Zha et al., 2012) και σε μεγέθυνση η περιοχή αλληλεπίδρασης της δομικά αυτοτελούς περιοχής υπεύθυνης για αυτοαναστολή της CaMK I με το καταλυτικό της κέντρο.

Η αυτοτελής δομική περιοχή υπεύθυνη για την αυτοαναστολή αποτελείται από 2 α-έλικες, τις αR1 και αR2, που συνδέονται με μια θηλιά. Η αR1 αλληλεπιδρά με αμινοξικά κατάλοιπα της περιοχής πρόσδεσης του υποστρώματος ενώ η αR2 με κατάλοιπα της περιοχής πρόσδεσης του ATP.

Με αυτόν τον τρόπο η αR1 εμφανίζεται να δρα ως ψευδοϋπόστρωμα ενώ η αR2 τροποποιεί την στερεοδιάταξη της περιοχής πρόσδεσης του ATP παρεμβαίνοντας έτσι στην ικανότητά της να προσδένει νουκλεοτίδια. Έχει υποτεθεί πως η $Ca^{2+}$/CaM αλληλεπιδρά αρχικά με την αR2, παρεμποδίζοντας την επαφή της με το πάνω τμήμα του C-λοβού. Αυτό δίνει τη δυνατότητα στο λοβό να αποκτά διαφορετική στερεοδιάταξη, ικανή για πρόσδεση του ATP. Αμέσως μετά, συνεπακόλουθη αλλαγή της στερεοδιάταξης της αR1 ή, πιθανά, αλληλεπίδραση αυτής με την $Ca^{2+}$/CaM, οδηγεί στην απομάκρυνση του ψευδοϋποστρώματος στο κατώτερο τμήμα του C-λοβού. Έτσι το καταλυτικό κέντρο της CaMKI είναι ελεύθερο για πρόσδεση του ATP και της πρωτεΐνης-υποστρώματος και έναρξη της αντίδρασης φωσφορυλίωσης. Αξίζει να αναφερθεί πως στην κρυσταλλογραφικά προσδιορισμένη δομή της CaMKI η θηλιά ενεργοποίησης, που περιέχει και την Thr177 που προαναφέρθηκε, δεν έχει λυθεί. Το γεγονός αυτό υποδηλώνει πως στην ανενεργή κατάσταση τα κατάλοιπα που απαρτίζουν την θηλιά είναι κινητά και χωρίς καθορισμένη στερεοδιάταξη. Αυτό είναι πιθανό να συνεισφέρει στην μειωμένη ενεργότητα της κινάσης πριν την φωσφορυλίωση στην Thr177. Σταθεροποίηση της θηλιάς ενεργοποίησης και των αλληλεπιδράσεων με κατάλοιπα γειτονικά του ενεργού κέντρου, μέσω της φωσφορυλίωσης από την CaMKK, αυξάνει την ενεργότητα της κινάσης.

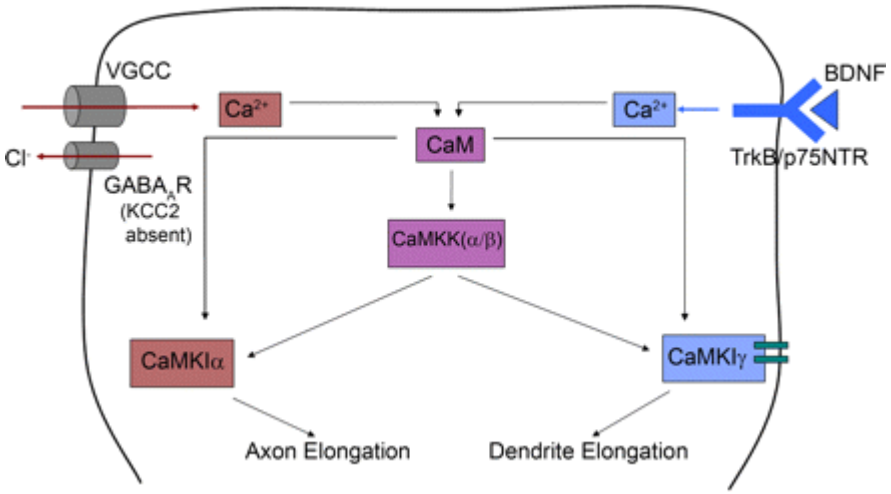



**Εικόνα 1.18** Λεπτομερής διαγραμματική απεικόνιση των αυτοτελών δομικών περιοχών της CaMK I. Με γκρι απεικονίζεται το καταλυτικό κέντρο, με λευκό η περιοχή υπεύθυνη για την αυτοαναστολή (παρουσιάζονται και τα κατάλοιπα που συμμετέχουν στο σχηματισμό των 2 α-ελίκων) και με μαύρο η περιοχή πρόσδεσης της CaM.




**Εικόνα 1.19** Μηχανισμός αλληλεπίδρασης CaMK I – $Ca^{2+}$-CaM

Στις πρωτεΐνες-υποστρώματα αναγνωρίζει την συναινετική ακολουθία [MVLIF]-x-R-$x_2$-[ST]-$x_3$-[MVLIF]. Εντοπίζεται και δρα κυτταροπλασματικά. Ρυθμίζει διάφορες λειτουργίες, μέσω φωσφορυλίωσης των υποστρωμάτων της, όπως την ενεργότητα μεταγραφικών παραγόντων, τον κυτταρικό κύκλο (ιδιαίτερα την G1 φάση της μεσόφασης), την παραγωγή ορμονών (όπως αλδοστερόνη), την κυτταρική διαφοροποίηση (ιδιαίτερα των μυϊκών και νευρικών κυττάρων), την οργάνωση των ινιδίων ακτίνης και την ανάπτυξη των νευρώνων (ιδιαίτερα την επέκταση του νευράξονα, την ανάπτυξη δενδριτών και τη δημιουργία συνάψεων στον ιππόκαμπο και την παρεγκεφαλίδα), την λειτουργία των συνάψεων (ιδιαίτερα την κινητικότητα των κυστιδίων που περιέχουν νευροδιαβιβαστές στο προσυναπτικό άκρο των νευρώνων), τη μνήμη και την μάθηση (συμμετέχοντας στις συνάψεις υπεύθυνες για την μακρόχρονη ενδυνάμωση -LTP- διαδικασία κατά την οποία οι συνάψεις υπόκεινται σε βελτίωση της αποτελεσματικότητάς τους όταν χρησιμοποιούνται έντονα), τη μετανάστευση κυττάρων μυελοβλαστώματος,




**Εικόνα 1.20** Παρουσιάζονται δύο ισομορφές της CaMKI. Η κυτταροπλασματική CaMKIα, που προωθεί την ανάπτυξη του νευράξονα, και η αγκυροβολημένη στη μεμβράνη CaMKIγ (ο συγκεκριμένος κυτταρικός εντοπισμός της οφείλεται στην ύπαρξη καταλοίπων κυστεΐνης στο καρβοξυτελικό της άκρο), που διευκολύνει την ανάπτυξη των δενδριτών. (Takemoto-Kimura et al. 2010; Horigane et al. 2016; Ageta-Ishihara et el. 2009).

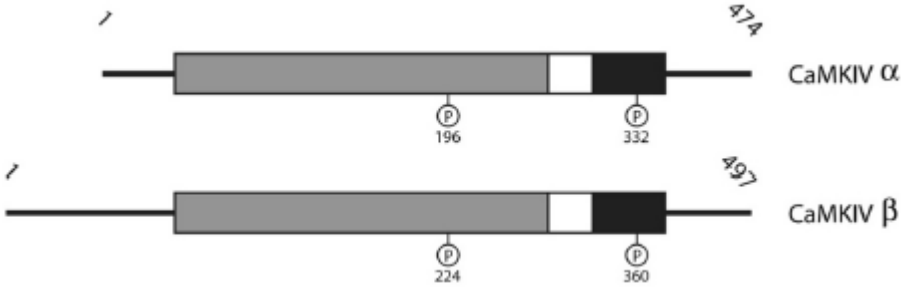
### 1.2.3.2. CaMK IV κινάσες

Η CaMKIV συμμετέχει και αυτή στον καταρράκτη των CaMK. Εντοπίζονται 2 ισομορφές, η α και η β, που κωδικοποιούνται από το ίδιο γονίδιο. Λειτουργούν ως μονομερή των 65 kDa. Η CaMKIVα κυριαρχεί στα κύτταρα του νευρικού ιστού, στα Τ-λεμφοκύτταρα και στους όρχις. Η β ισομορφή προκύπτει με εναλλακτική συρραφή, περιέχει επιπλέον 28 αμινοξικά κατάλοιπα στο αμινοτελικό άκρο και εκφράζεται διαφορικά σε όλα τα στάδια της ανάπτυξης στα κύτταρα της κοκκιώδους στιβάδας της παρεγκεφαλίδας.

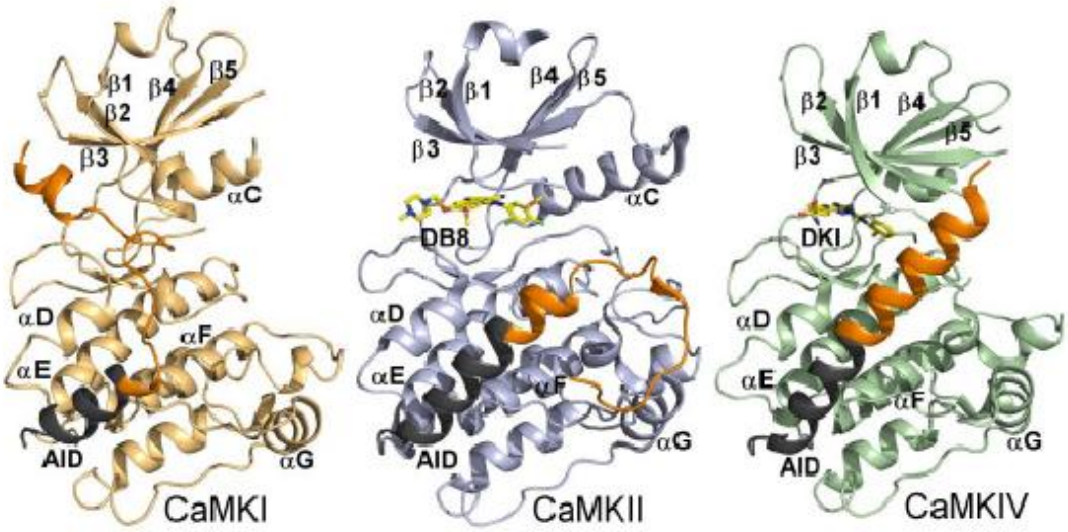
Για την ενεργοποίηση της CaMKIV απαιτείται πρόσδεση στην $Ca^{2+}$/CaM και φωσφορυλίωση από την CaMKK στην Thr196 που εντοπίζεται στην θηλιά ενεργοποίησης. Σε αντίθεση με την CaMKI, μετά την φωσφορυλίωση από την CaMKK, μπορεί να ακολουθήσει μια αυτοφωσφορυλίωση σε μια περιοχή πλούσια σε κατάλοιπα σερίνης και θρεονίνης που εντοπίζεται στο αμινοτελικό της άκρο. Αυτό έχει ως αποτέλεσμα τόσο την μεγιστοποίηση της ενεργότητας της CaMKIV όσο και την απόκτηση ανεξαρτησίας από την $Ca^{2+}$/CaM. Με αυτόν τον τρόπο η κινάση παραμένει λειτουργική και μετά το πέρας της παροδικής αύξησης της συγκέντρωσης του ασβεστίου που πυροδότησε την ενεργοποίησή της.

Διαθέτει ένα επιπλέον επίπεδο αυτορρύθμισης μέσω αυτοφωσφορυλίωσης της Ser332 που εντοπίζεται στο εσωτερικό της δομικά αυτοτελούς περιοχής πρόσδεσης της $Ca^{2+}$/CaM. Αυτό παρεμποδίζει την πρόσδεση της $Ca^{2+}$/CaM μέχρι το κατάλοιπο σερίνης να αποφωσφορυλιωθεί από κάποια διαθέσιμη φωσφατάση. Η φωσφατάση PP2a σχηματίζει με την CaMKIV ένα σύμπλοκο σηματοδότησης που επιτρέπει τον ακριβή έλεγχο των γεγονότων φωσφορυλίωσης/αποφωσφορυλίωσης της θηλιάς ενεργοποίησης της κινάσης. Η αλληλεπίδραση αυτή έχει σαν αποτέλεσμα η κινάση να μην αποκτά αυτονομία και έτσι να μην μπορεί να ενεργοποιήσει τη μεταγραφή γονιδίων (όπως θα αναφερθεί και παρακάτω).

Η διάταξη των δομικών αυτοτελών περιοχών της CaMKIV είναι παρόμοια με τις υπόλοιπες CaMK, με το καταλυτικό κέντρο στο αμινοτελικό άκρο, ακολουθεί η δομική αυτοτελής περιοχή υπεύθυνη για αυτοαναστολή και αμέσως μετά η δομική αυτοτελής περιοχή που προσδένεται στην $Ca^{2+}$/CaM.




**Εικόνα 1.21** Διάταξη των αυτοτελών δομικών περιοχών των CaMKIV α και β καθώς και τα κατάλοιπα που φωσφορυλιώνει η CaMKK στο καταλυτικό κέντρο και τα κατάλοιπα που αυτοφωσφορυλιώνονται και έτσι παρεμποδίζεται η πρόσδεση της $Ca^{2+}$/CaM στην περιοχή πρόσδεσης της $Ca^{2+}$/CaM. Απεικονίζονται επίσης τα 28 κατάλοιπα του αμινοτελικού άκρου στη β ισομορφή.




**Εικόνα 1.22** Αναπαράσταση των καταλυτικών κέντρων καθώς , των δομικά αυτοτελών περιοχών υπεύθυνων για αυτοαναστολή (σκούρο γκρι) και των περιοχών πρόσδεσης με την $Ca^{2+}$/CaM (πορτοκαλί) από τις CaMKI (PDB : 1A06), CaMKII (PDB : 3SOA), και τη μοναδική κρυσταλλογραφικά προσδιορισμένη δομή της CaMKIV (PDB : 2W4O) (Kylarova, Psenakova, Herman, & Obsil, 2018).

Η CaMKIV εντοπίζεται τόσο στο κυτταρόπλασμα, όσο και στον πυρήνα.

Στον πυρήνα είναι υπεύθυνη για την εξαρτώμενη από $Ca^{2+}$ ρύθμιση της μεταγραφής γονιδίων μέσω φωσφορυλίωσης διαφόρων μεταγραφικών παραγόντων, όπως οι CREB (cAMP response element binding protein), CBP (πρωτεΐνη που προσδένεται στον CREB και διευκολύνει τη σταθεροποίηση των αλληλεπιδράσεων στη μεταγραφική μηχανή)(Matthews et al. 1994;Bito et al. 1996; Silva et al. 1998), SRF (serum response factor), MEF2, HDAC4 (histone deacetylase 4), JUN, RORA κ.α. Έτσι συμμετέχει στη ρύθμιση διαφόρων βιολογικών διαδικασιών, όπως η ανοσολογική απόκριση (ιδιαίτερα ρύθμιση της οντογένεσης των T-λεμφοκυττάρων και της λειτουργίας των CD4 T-λεμφοκυττάρων μνήμης), η

φλεγμονή, η κυτταρική διαφοροποίηση (ιδιαίτερα στη διαφοροποίηση των οστεοκλαστών και των δενδριτικών κυττάρων), η μνήμη και η μάθηση (όπως στην μακρόχρονη ενδυνάμωση στον ιππόκαμπο, στη μακρόχρονη καταστολή –LTD- η οποία μειώνει την αποτελεσματικότητα των νευρικών συναπτικών επαφών στην παρεγκεφαλίδα, στη ρύθμιση του συναπτικού δυναμικού και στη μάθηση και στον συντονισμό κινήσεων). Για να συμβούν αυτά απαιτείται η CaMKIV να βρίσκεται στην αυτόνομη κατάσταση (ανεξάρτητη από Ca2+ ).

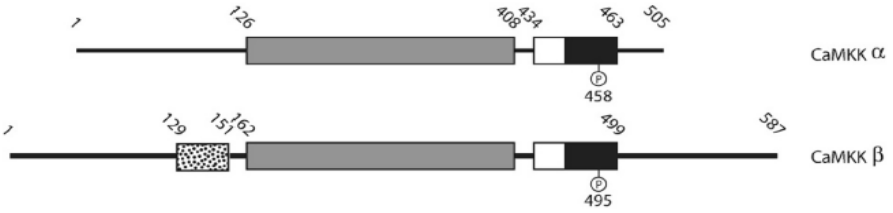
Στο κυτταρόπλασμα ο ρόλος της CaMKIV δεν είναι πλήρως κατανοητός. Είναι γνωστό πως φωσφορυλιώνει την ογκοπρωτεΐνη18 στην Ser16, αποτρέποντας έτσι την αλληλεπίδρασή της με την τουμπουλίνη.

### 1.2.3.3. CaMKK κινάσες

Όπως προαναφέρθηκε, κάποια μέλη της οικογένειες των CaMK αποτελούν υποστρώματα για άλλες κινάσες της ίδιας οικογένειας. Ο καταρράκτης αυτός ανακαλύφθηκε το 1994 όταν οι Lee & Edelmen και Tokumitsu et al. ανεξάρτητα απομόνωσαν μια εξαρτώμενη από $Ca^{2+}$/CaM πρωτεΐνη υπεύθυνη για την ενεργοποίηση των CaMKI και CaMKIV αντίστοιχα. Λίγο αργότερα ανακαλύφθηκε ότι πρόκειται για το ίδιο μόριο, που σήμερα είναι γνωστό ως εξαρτώμενη από $Ca^{2+}$/CaM πρωτεϊνική κινάση (CaMKK), και αποτελεί την πρώτη πρωτεΐνη στον καταρράκτη των CaMK που περιλαμβάνει και τις 3 κινάσες που αναφέρθηκαν. Τα ακριβή πλεονεκτήματα από την ύπαρξη ενός καταρράκτη που όλοι οι συμμετέχοντες αποκρίνονται στο ίδιο σήμα (πρόσδεση σε $Ca^{2+}$/CaM) δεν είναι γνωστά, αλλά ένα πιθανό πλεονέκτημα μπορεί να αποτελεί το γεγονός πως τα ένζυμα δεν θα είναι πλήρως ενεργοποιημένα αν δεν είναι επαρκές το σήμα από την αύξηση της συγκέντρωσης του ασβεστίου, ως προς τη συγκέντρωση και τη διάρκεια, ώστε να προσδεθούν όλα στην $Ca^{2+}$/CaM αλλά και να αλληλεπιδράσουν μεταξύ τους. Η άποψη αυτή ενισχύεται και από το γεγονός πως τα ένζυμα του CaMK καταρράκτη ρυθμίζουν γεγονότα που αφορούν την έκφραση γονιδίων, τα προϊόντα των οποίων οδηγούν σε μακράς διάρκειας αλλαγές για το κύτταρο, και έτσι η πυροδότησή τους θα πρέπει ελέγχεται από μηχανισμό που δεν ενέχει σφάλματα.

Υπάρχουν δύο ισομορφές των CaMKK, η α ισομορφή μήκους 505 αμινοξικών καταλοίπων και η β ισομορφή μήκους 587 αμινοξικών καταλοίπων. Κάθε ισομορφή κωδικοποιείται από διαφορετικά γονίδια. Και οι δυο ισομορφές λειτουργούν στα κύτταρα ως μονομερή. Η CaMKK συντίθεται με αυξημένο ρυθμό στα κύτταρα του εγκεφάλου, αλλά η ισομορφή α εντοπίζεται και στον θύμο και στο σπλήνα. Ενδοκυτταρικά εντοπίζεται τόσο στο κυτταρόπλασμα, όσο και στον πυρήνα και συνεπώς είναι υπεύθυνη για την απόκριση στις αλλαγές συγκέντρωσης του ασβεστίου και στις δυο περιοχές.

Η διάταξη των δομικά αυτοτελών περιοχών των ισομορφών των CaMKK είναι παρόμοια με αυτή των υπόλοιπων CaMK. Έτσι, από το άμινο- προς το καρβόξυ- τελικό άκρο, εντοπίζεται το καταλυτικό κέντρο με ενεργότητα πρωτεϊνικής κινάσης, μια περιοχή υπεύθυνη για αυτοαναστολή και μια περιοχή πρόσδεσης της $Ca^{2+}$/CaM.




**Εικόνα 1.23** Λεπτομερής διαγραμματική απεικόνιση των δομικά αυτοτελών περιοχών των 2 ισομορφών της CaMKK, α και β. Με γκρι απεικονίζεται το καταλυτικό κέντρο, με λευκό η περιοχή υπεύθυνη για αυτοαναστολή και με μαύρο η περιοχή πρόσδεσης της CaM. Στην β ισομορφή απεικονίζεται στο Ν-τελικό άκρο του καταλυτικού κέντρου και η περιοχή (μήκους 23 καταλοίπων) που συμμετέχει στη ρύθμιση της αυτόνομης δραστικότητας της CaMKK. Τέλος σημειώνεται το κατάλοιπο σερίνης που φωσφορυλιώνεται από την PKA.(Swulius & Waxham, 2008).

Το καταλυτικό κέντρο της CaMKK παρουσιάζει διαφορές από τα αντίστοιχα των υπόλοιπων CaMK. Σύγκριση των αμινοξικών ακολουθιών αποκάλυψε την παρουσία μιας επιπλέον μοναδικής ακολουθίας πλούσιας σε προλίνη και αργινίνη (Arg-Pro-rich insert – RP-insert) στο εσωτερικό του Ν-λοβού. Η ακολουθία αυτή καθιστά δυνατή την αναγνώριση και την επακόλουθη ενεργοποίηση των CaMKI και CaMKIV.

Παρά το γεγονός πως και οι δυο ισομορφές εμφανίζουν υψηλή ομολογία των αμινοξικών ακολουθιών τους, διαφέρουν μεταξύ τους ως προς τη ρύθμιση της ενεργότητάς τους. Η πρόσδεση στην $Ca^{2+}$/CaM είναι απαραίτητη για την ενεργοποίηση της CaMKKα, ενώ η CaMKKβ παρατηρείται ενεργή και απουσία $Ca^{2+}$/CaM.

Η κρυσταλλογραφικά λυμένη δομή του καταλυτικού κέντρου της CaMKKβ (PDB: 2ZV2) αποκάλυψε πως στο C-λοβό δεν εντοπίζεται η αD-έλικα που συμμετείχε σε υδρόφοβες αλληλεπιδράσεις με την περιοχή υπεύθυνη για αυτοαναστολή στις CaMKI και CaMKII (Yoshikawa et al., 2011). Η περιοχή αυτοαναστολής αλληλεπιδρά με μια επιφάνεια που σχηματίζεται από τους κλώνους β4 και β5 καθώς και από την αC-έλικα. Η επιφάνεια αυτή εντοπίζεται αντιδιαμετρικά σε σχέση με τις περιοχές που συμμετέχουν στην αυτοαναστολή για τις CaMKI και CaMKII.Η CaMKKβ, μετά την ενεργοποίησή της από την $Ca^{2+}$/CaM, μπορεί να λειτουργήσει ανεξάρτητα από την $Ca^{2+}$/CaM. Πειραματικές διαδικασίες βασισμένες σε μεταλλάξεις αποκάλυψαν την ύπαρξη μιας περιοχής μήκους 23 αμινοξικών καταλοίπων (κατάλοιπα 129–151)

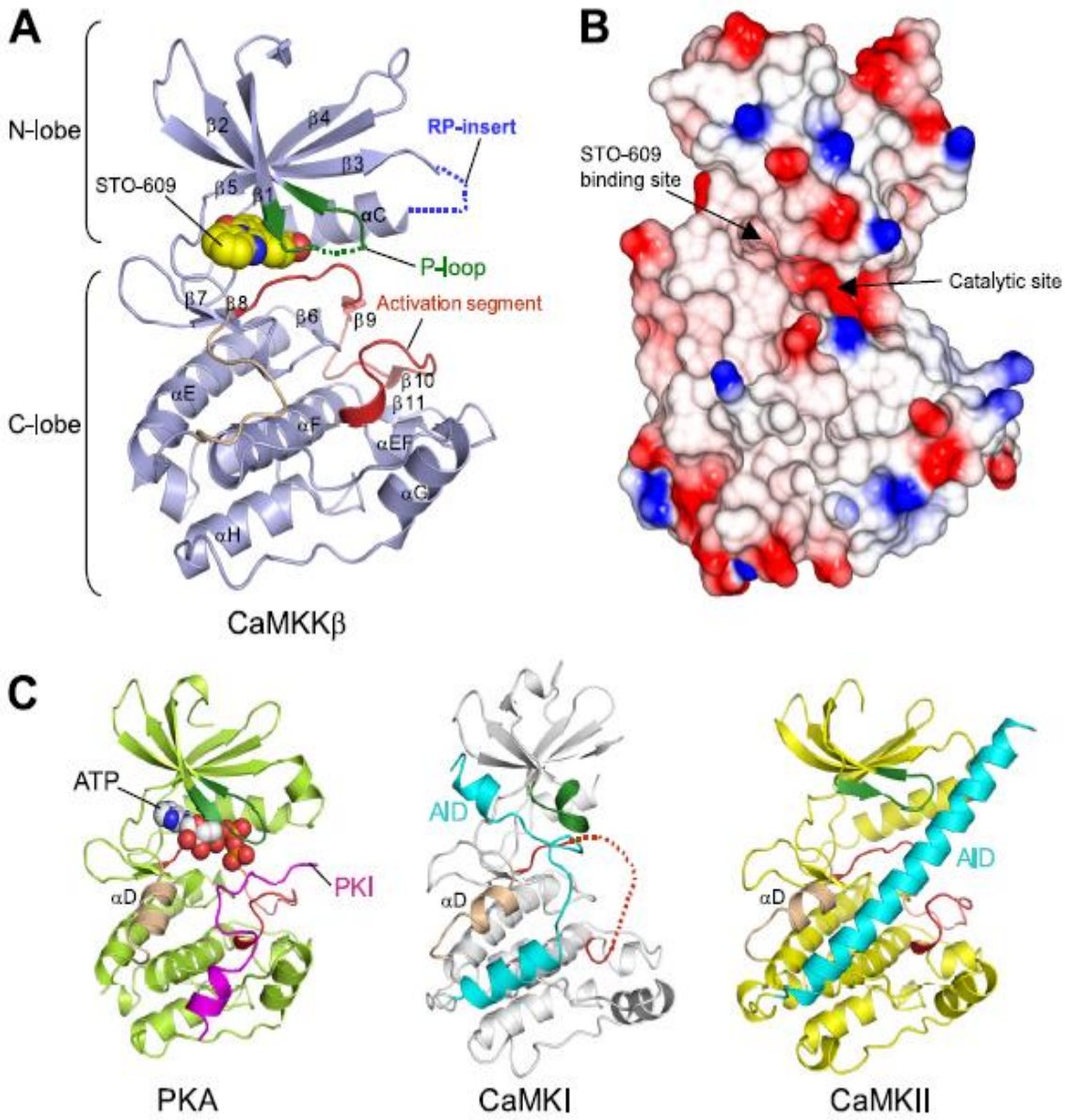
που συμμετέχει στην απελευθέρωση της περιοχής υπεύθυνης για την αυτοαναστολή από το ενεργό κέντρο, έχοντας ως αποτέλεσμα την αυτόνομη λειτουργία της κινάσης. Αυτό σημαίνει πως στα κύτταρα εκείνα, ή στις περιοχές των κυττάρων, στα οποία εντοπίζεται η β ισομορφή ο καταρράκτης θα παραμείνει ενεργός ανεξάρτητα από τη συγκέντρωση του ασβεστίου που πυροδότησε την ενεργοποίησή του.

Σημαντικά κατάλοιπα όπως αποδεικνύεται από σημειακές μεταλλάξεις είναι:

Φωσφορυλίωση των Ser129, Ser133 και Ser137 έχει ως αποτέλεσμα την αναστολή της αυτονομίας της κινάσης

Φωσφορυλίωση της Ser137 από την κινάση καθοδηγούμενη από προλίνη (Pro-K) πυροδοτεί την φωσφορυλίωση δύο ακόμα καταλοίπων σερίνης από την κινάση της συνθετάσης του γλυκογόνου-3 (GSK3). Αντιθέτως η αυτονομία της CaMKKβ αυξάνεται μέσω αυτοφωσφορυλίωσης των καταλοίπων Thr85 και Thr482, πιθανά γιατί έχει ως αποτέλεσμα την παρεμπόδιση του μηχανισμού αυτοαναστολής.

Εκτός από τις δομικές διαφορές που αναφέρθηκαν, η ακολουθία και το μήκος των περιοχών αυτοαναστολής διαφέρουν από τις αντίστοιχες των CaMKI και CaMKII. Αυτό υποδηλώνει πως θα διαφέρουν και οι αλληλεπιδράσεις που αναπτύσσονται μεταξύ της περιοχής αυτοαναστολής και του καταλυτικού κέντρου της κινάσης, αν και ο ακριβής μηχανισμός δεν είναι ακόμα γνωστός.

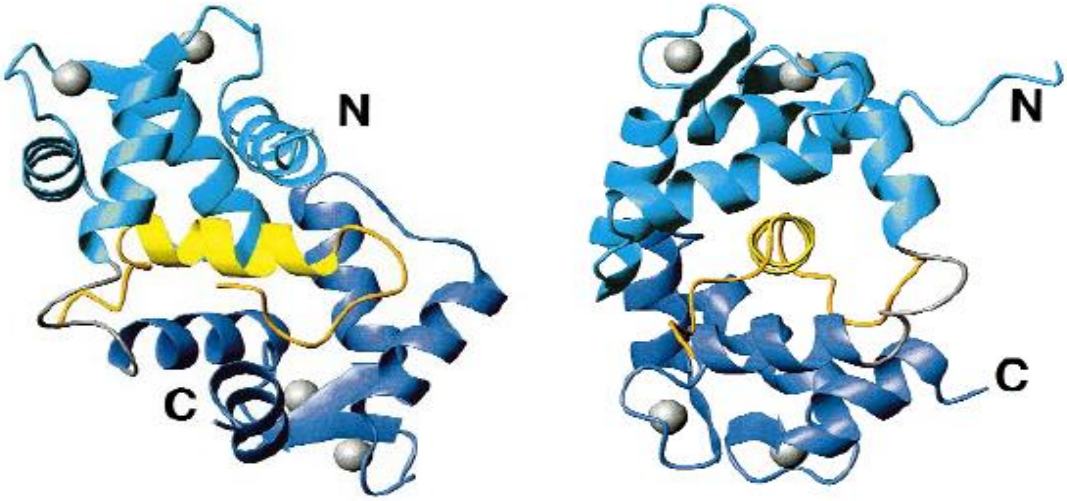



**Εικόνα 1.24** Απεικόνιση του καταλυτικού κέντρου της CaMKK2 (PDB : 2ZV2). Με πράσινο απεικονίζεται η P-θηλιά, η περιοχή ενεργοποίησης με κόκκινο και με μπεζ η θηλιά β5-αE. Ο αναστολέας STO-609 αναπαρίσταται με σφαίρες. Στο κάτω μέρος της εικόνας γίνεται σύγκριση μεταξύ της δομής της CaMKK2 και της PKA σε σύμπλοκο με τον αναστολέα PKI, του καταλυτικού κέντρου και της περιοχής αυτοαναστολής από την CaMK I και του καταλυτικού κέντρου και της περιοχής αυτοαναστολής από την CaMK II.(Yoshikawa et al., 2011).

Η λύση της δομής της $Ca^{2+}$/CaM σε σύμπλοκο με την περιοχή πρόσδεσής της από την CaMKK1 αποκάλυψε κάτι πολύ ενδιαφέρον. Ο προσανατολισμός κατά τον οποίο προσδένεται το πεπτίδιο της CaMKK1 στην $Ca^{2+}$/CaM είναι αντίθετος από τον προσανατολισμό πρόσδεσης των πεπτιδίων της MLCK και της CaMKII.




**Εικόνα 1.25** Στοίχιση των ακολουθιών των περιοχών πρόσδεσης στην $Ca^{2+}$/CaM διαφόρων πρωτεϊνών με βάση τα κατάλοιπα-κλειδιά (γκρι) του Ν-τελικού άκρου. Στα δεξιά απεικονίζεται ο προσανατολισμός κατά την πρόσδεση με την $Ca^{2+}$/CaM (Osawa et al., 1999).




**Εικόνα 1.26** Δομή της $Ca^{2+}$/CaM σε σύμπλοκο με την περιοχή πρόσδεσής της από την CaMKK1 από 2 οπτικές γωνίες(Osawa et al., 1999).

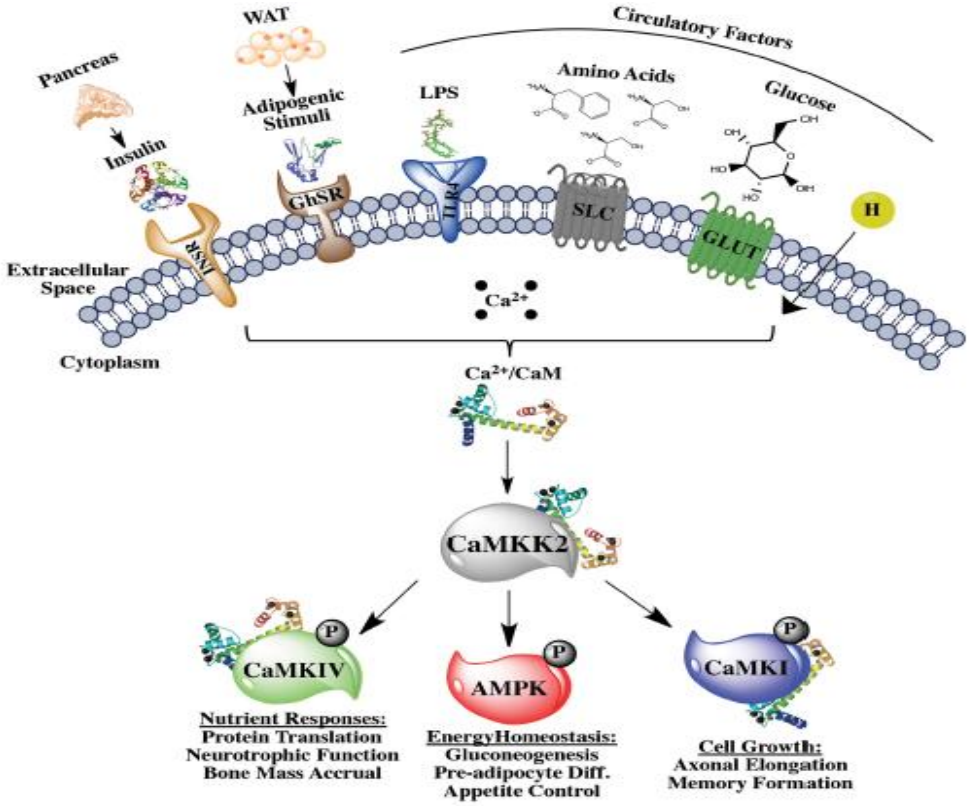
Η CaMKK αποτελεί υπόστρωμα για την ενεργή εξαρτώμενη από cAMP πρωτεϊνική κινάση (PKA). Η PKA φωσφορυλιώνει στην CaMKKβ στην Ser100 και στην Ser511, μειώνοντας έτσι τη δυνατότητα πρόσδεσης στην πρωτεΐνη-ικρίωμα 14-3-3. Επίσης φωσφορυλιώνει την CaMKKβ στην Thr145 και την CaMKKα στην Thr108, ρυθμίζοντας έτσι την ενεργότητα της κινάσης ως προς την κινάση-υπόστρωμά της. Τέλος, φωσφορυλιώνει την CaMKKα στην Ser458 και την CaMKKβ στην Ser495, κατάλοιπα που εντοπίζονται στο εσωτερικό της περιοχής πρόσδεσης της $Ca^{2+}$/CaM. Αυτό έχει ως αποτέλεσμα να παρεμποδίζεται η πρόσδεση της $Ca^{2+}$/CaM και μέσω αυτού η ικανότητα της CaMKK να αποκρίνεται στην αύξηση της συγκέντρωσης του ασβεστίου. Το γεγονός αυτό είναι αυξημένης σημασίας για τη σωστή απόκριση στο εκάστοτε σήμα. Για παράδειγμα, αν αυξηθούν τα επίπεδα του cAMP λόγω σηματοδότησης μέσω της forskolin (μιας πρωτεΐνης υπεύθυνης για την ενεργοποίηση της αδενυλικής κυκλάσης, που με τη σειρά της συνθέτει το cAMP και ενεργοποιεί την PKA) πριν από την αύξηση των επιπέδων ασβεστίου, η PKA θα

φωσφορυλιώσει την CaMKK αναστέλλοντας έτσι την ενεργοποίησή της από την $Ca^{2+}$/CaM και κατ'επέκταση παρεμποδίζοντας την ολοκληρωμένη ενεργοποίηση των CaMKI και CaMKIV. Αν, όμως, αυξηθούν τα επίπεδα του ασβεστίου πριν από το σήμα μέσω της forskolin, η $Ca^{2+}$/CaM θα προσδεθεί στην CaMKK, παρεμποδίζοντας την φωσφορυλίωση του καταλοίπου σερίνης από την PKA, και επιτρέποντας έτσι την περαιτέρω ενεργοποίηση των καταρροϊκών στόχων του καταρράκτη.

Η ενεργή CaMKKβ επίσης φωσφορυλιώνει την α υπομονάδα της ενεργοποιούμενης από AMP πρωτεϊνικής κινάσης. Αυτό έχει ως αποτέλεσμα τη δημιουργία ενός συμπλόκου που αποτελείται από την $Ca^{2+}$/CaM, την CaMKKβ και τις α και β υπομονάδες της AMPK. Οι CaMKKβ και AMPK αλληλεπιδρούν μέσω των καταλυτικών τους κέντρων και η CaMKKβ πρέπει να βρίσκεται στην ενεργή στερεοδιάταξή της (γεγονός που δεν απαιτείται για την ενεργοποίηση των CaMKI και CaMKIV). Η γ υπομονάδα της AMPK, που ανταποκρίνεται στις αλλαγές στη διακύμανση των ATP και AMP, δεν απαιτείται για τη δημιουργία του συμπλόκου, και έτσι η AMPK με αυτόν τον τρόπο ενεργοποιείται ως απόκριση στην αύξηση της συγκέντρωσης του ασβεστίου.

Η CaMKK αποτελεί επίσης υπόστρωμα και για την σχετιζόμενη με τον θάνατο πρωτεϊνική κινάση (death associated protein kinase - DAPK), επεκτείνοντας τις κυτταρικές λειτουργίες στις οποίες συμμετέχει. Η CaMKKβ στον εγκέφαλο συμμετέχει στον έλεγχο της όρεξης, στην ομοιόσταση της ενέργειας, στην ανάπτυξη κοκκιωδών κυττάρων στην παρεγκεφαλίδα, στην ρύθμιση του συμπαθητικού τόνου, στη μακρόχρονη ενδυνάμωση της μνήμης, , στη συσσώρευση οστικής μάζας, στην προστασία των νευρώνων από τραύματα που προέρχονται από ισχαιμικά επεισόδια, στην διατήρηση του αιματοεγκεφαλικού φραγμού, ρύθμιση του λιπώδη ιστού, στην ομοιόσταση της γλυκόζης, στην αιμοποίηση.

Έχει συνδεθεί με αρκετές παθολογικές καταστάσεις, όπως, οι αντιδράσεις φλεγμονής, ο καρκίνος (ρυθμίζει τη σύνθεση πρωτεϊνών και έτσι ελέγχει την ανάπτυξη του όγκου στην περίπτωση καρκίνου του ήπατος), το άγχος, η διπολική διαταραχή, η αθηροσκλήρωση και θεωρείται πιθανός στόχος για θεραπευτικές επεμβάσεις (Racioppi & Means, 2012).

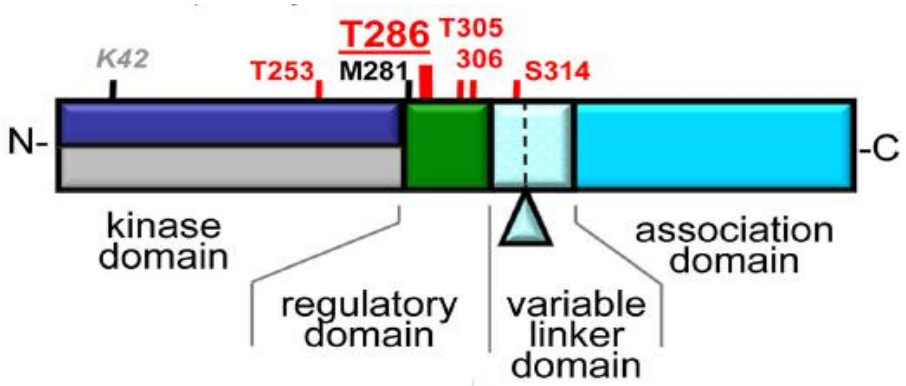



**Εικόνα 1.27** Ο καταρράκτης των εξαρτώμενων από $Ca^{2+}$/CaM κινασών συμμετέχει σε ποικίλες αποκρίσεις σε φυσιολογικές ή παθολογικές καταστάσεις. Αναρροϊκά εξωκυττάρια σήματα, όπως ινσουλίνη από το πάγκρεας, σήμα από τον λευκό λιπώδη ιστό και λιποπολυσακχαρίτες, αμινοξέα, ορμόνες και γλυκόζη από την κυκλοφορία που προσδένονται στο εκάστοτε υποδοχέα, πυροδοτούν την αύξηση της εκάστοτε ενδοκυτταρικής συγκέντρωσης ασβεστίου και την επακόλουθη ενεργοποίηση στόχων της $Ca^{2+}$/CaM,όπως η CaMKKβ. Αυτό έχει σαν αποτέλεσμα την ενεργοποίηση των CaMK IV, CaMK I και AMPK. Ενεργοποίηση της CaMK I έχει ως αποτέλεσμα την ρύθμιση της ανάπτυξης του κυττάρου, όπως παρατηρείται κατά την επιμήκυνση του άξονα και τη δημιουργία διακλαδώσεων στους νευρώνες ή κατά τη διάρκεια του κυτταρικού κύκλου. Ενεργοποίηση της AMPK έχει ως αποτέλεσμα την ρύθμιση του ισοζυγίου ενέργειας, κυρίως στα κύτταρα του εγκεφάλου, του ήπατος και του λιπώδη ιστού. Ρύθμιση την ενεργότητας της CaMKIV, η οποία έπειτα από την φωσφορυλίωση λειτουργεί ανεξάρτητα από την $Ca^{2+}$/CaM, οδηγεί στον έλεγχο της έκφρασης γονιδίων, ως απόκριση σε θρεπτικά και ορμόνες (Kathrina L. Marcelo, 2017).

### 1.2.3.4. CaMK II κινάσες

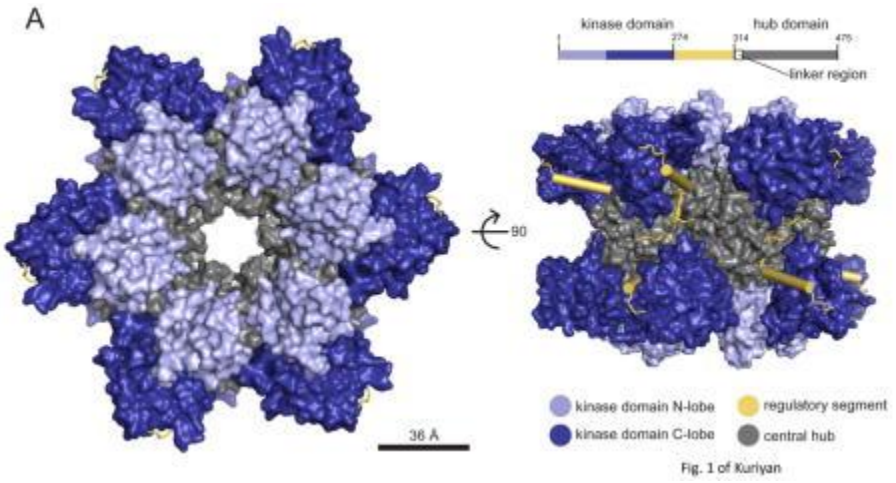
Υπάρχουν 4 ισομορφές των CaMKII που κωδικοποιούνται από 4 γονίδια: η CaMKIIα, η CaMKIIβ, η CaMKIIγ και η CaMKIIδ. Έχουν καταγραφεί και περισσότερες από 40 ισομορφές που προκύπτουν από εναλλακτική συρραφή, και έτσι το μέγεθός τους ποικίλει από 50 έως 68 kDa.

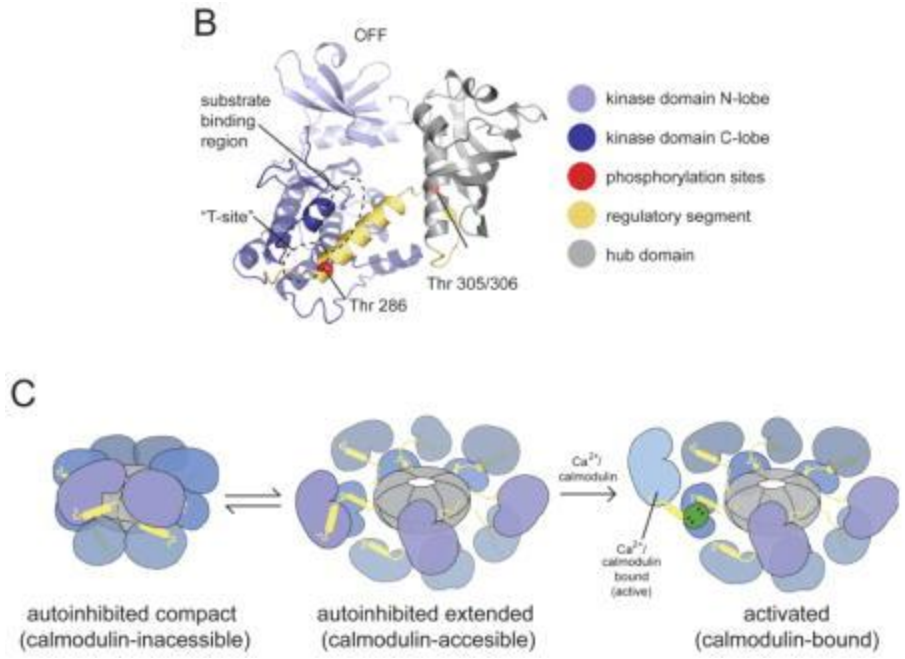
Η κύρια δομή τους αποτελείται από το καταλυτικό κέντρο με ενεργότητα πρωτεϊνικής κινάσης, μια δομικά αυτοτελή περιοχή υπεύθυνη για αυτοαναστολή (auto-inhibitory domain – AID), μια δομικά αυτοτελή περιοχή πρόσδεσης στην $Ca^{2+}$/CaM ($Ca^{2+}$/CaM -binding domain – CBD), μια περιοχή που ποικίλει μεταξύ των διαφορετικών ισομορφών και μια δομικά αυτοτελή περιοχή υπεύθυνη για τον σχηματισμό των πολυμερών (self-association domain - SAD). Το ολοένζυμο είναι ένα δωδεκαμερές, που συγκροτείται μέσω αλληλεπίδρασης των SAD της κάθε υπομονάδας (Chao et al., 2011; Stratton et al. 2013). Στην ανενεργή κατάσταση η AID δρα σαν ψευδοϋπόστρωμα και προσδένεται στην εσοχή πρόσδεσης του υποστρώματος. Όταν αυξηθεί η συγκέντρωση του ασβεστίου ενδοκυτταρικά, το σύμπλοκο $Ca^{2+}$/CaM θα προσδεθεί σε μια από τις υπομονάδες του ολοενζύμου και αυτό θα πυροδοτήσει την απελευθέρωση του AID από το ενεργό κέντρο επιτρέποντας έτσι την ενεργοποίηση του CaMKII ολοενζύμου. Όταν 2 γειτονικές υπομονάδες ενεργοποιηθούν, το κατάλοιπο Thr286 της μίας θα αποτελέσει υπόστρωμα για την γειτονική υπομονάδα. Αυτός ο καταρράκτης συνεχόμενων αντιδράσεων φωσφορυλίωσης, που συμβαίνει σε *trans* κατάσταση στο σύμπλοκο του ολοενζύμου αναφέρεται ως βήμα αυτοφωσφορυλίωσης. Το βήμα αυτοφωσφορυλίωσης είναι απαραίτητο για να αποκτήσει αυτονομία, ή ανεξάρτητη από $Ca^{2+}$/CaM δραστικότητα, το CaMKII ολοένζυμο (Coultrap & Bayer, 2012; Lisman, Yasuda, & Raghavachari, 2012; Soderling & Stull, 2001).




**Εικόνα 1.28** Απεικόνιση της διάταξης των δομικά αυτοτελών περιοχών στην αμινοξική ακολουθία της CaMK IIα (στην περιοχή-συνδέτη σημειώνεται το σημείο που έχει παρατηρηθεί επιπλέον αμινοξική ακολουθία λόγω εναλλακτικής συρραφής)(Coultrap & Bayer, 2012).

Αναλυτικά με βάση την εικόνα 1.28, το ολοένζυμο αποτελείται από 12 υπομονάδες, κάθε μια από τις οποίες στο καρβοξυτελικό της άκρο διαθέτει μια SAD (γκρι) που αλληλεπιδρούν μεταξύ τους και σχηματίζουν έναν κεντρικό οργανωτικό άξονα (ο οποίος εμφανίζει έναν κεντρικό πόρο με άγνωστη μέχρι τώρα λειτουργία), πάνω στον οποίο τα καταλυτικά κέντρα της κάθε υπομονάδας (μπλε) διπλώνουν σχηματίζοντας 2 δαχτυλίδια των 6 καταλυτικών κέντρων. Τα καταλυτικά κέντρα έχουν την συντηρημένη αρχιτεκτονική των 2 λοβών με την περιοχή πρόσδεσης του ATP μεταξύ των λοβών. Μια ρυθμιστική περιοχή (κίτρινο) ρυθμίζει την ενεργότητα της κινάσης ως προς το υπόστρωμα. Μια περιοχή-συνδέτης ενώνει την ρυθμιστική περιοχή με τον άξονα. Στην ανενεργή κατάσταση, λόγω αυτοαναστολής, η AID περιοχή σχηματίζει μια α-έλικα που λειτουργεί ως ψευδοϋπόστρωμα και παρεμποδίζει την περιοχή πρόσδεσης του υποστρώματος (B). Η κατάσταση αυτή σταθεροποιείται περισσότερο λόγω της απομόνωσης της θέσης φωσφορυλίωσης στην Thr286 στο εσωτερικό μιας υδρόφοβης αύλακας (ονομάζεται περιοχή-T). Η κρυσταλλογραφικά προσδιορισμένη δομή αποκαλύπτει μια συμπαγή κατάσταση κατά την οποία η περιοχή πρόσδεσης στην $Ca^{2+}$/CaM βρίσκεται στο εσωτερικό του άξονα, κάνοντας έτσι αδύνατη την αλληλεπίδραση μαζί της.





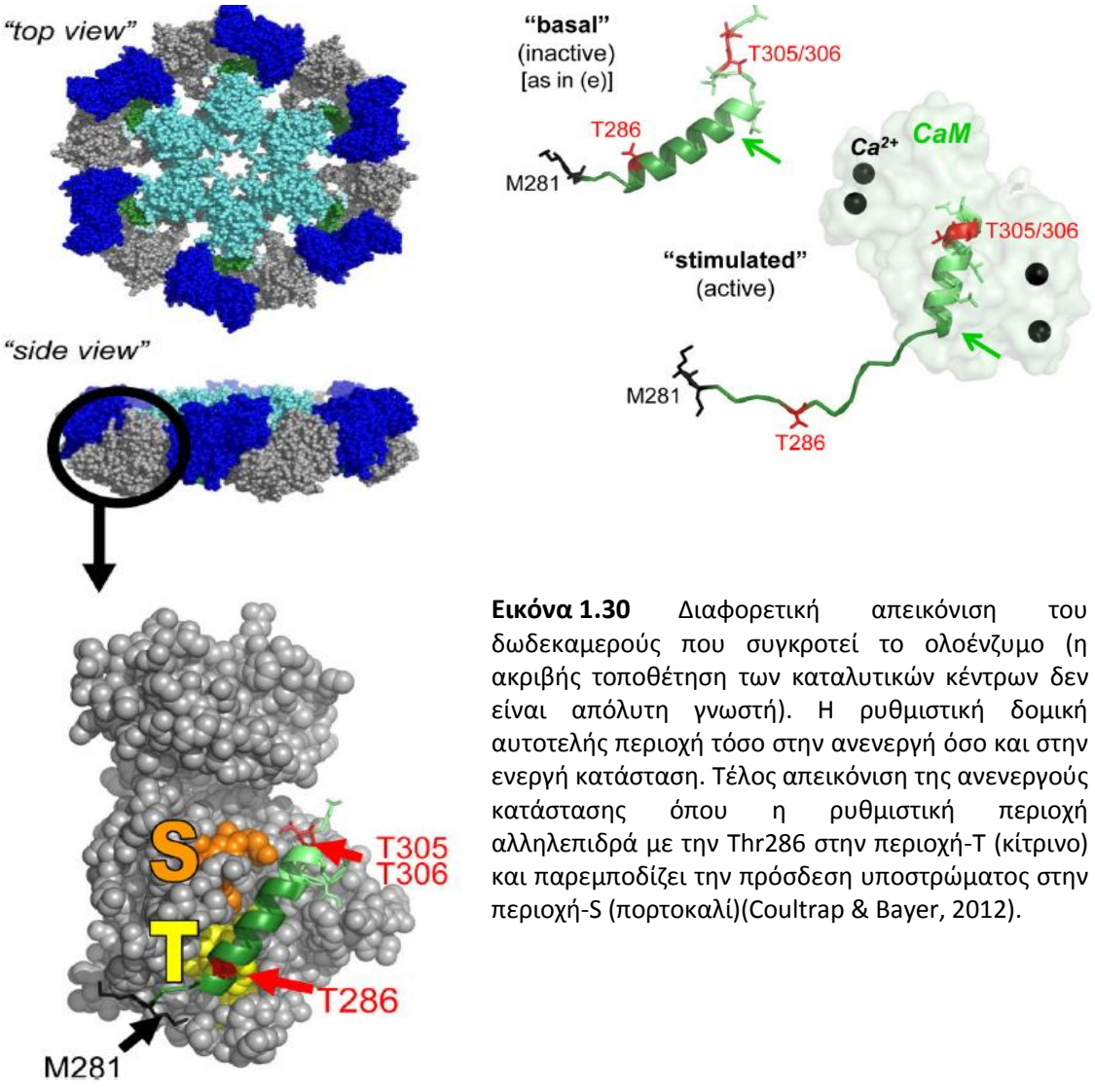

**Εικόνα 1.29** Απεικονίζονται (Α) η δομή του ολοενζύμου, (Β) η δομή της κάθε υπομονάδας και (Γ) οι 3 καταστάσεις του ολοενζύμου, όπως περιγράφονται αναλυτικά στο κείμενο(Lisman et al., 2012).

Η ενεργοποίηση είναι εφικτή επειδή η συμπαγής κατάσταση βρίσκεται σε δυναμική ισορροπία με την εκτεταμένη, στην οποία είναι εφικτή η αλληλεπίδραση με την $Ca^{2+}$/CaM.

Η περιοχή-συνδέτης, που διαφέρει μεταξύ των ισομορφών της CaMK II, καθορίζει το σημείο αναφοράς της ισορροπίας καθώς και τα επίπεδα ασβεστίου που απαιτούνται για ενεργοποίηση. Μετά την πρόσδεση της καλμοντουλίνης η AID απομακρύνεται από την περιοχή πρόσδεσης του υποστρώματος, επιτρέποντας έτσι την πρόσδεση και φωσφορυλίωση του υποστρώματος. Ταυτόχρονα ξεκινάει και η, σημαντικά πιο αργή, διαδικασία της αυτοφωσφορυλίωσης των καταλοίπων Thr286. Σε περίπτωση που η φωσφορυλίωση της Thr286 δεν έχει συμβεί και τα επίπεδα ασβεστίου μειωθούν, άμεσα το ολοένζυμο καθίσταται ανενεργό. Αν συμβεί η φωσφορυλίωση, το ένζυμο βρίσκεται στην αυτόνομη κατάσταση και η περιοχή-Τ μπορεί να αλληλεπιδράσει (για παράδειγμα με την NR2B υπομονάδα του NMDA υποδοχέα γλουταμικού). Τέλος η φωσφορυλίωση της Thr286 ισχυροποιεί την πρόσδεση της $Ca^{2+}$/CaM, η διαδικασία αναφέρεται και ως παγίδευση. Αξίζει να αναφερθεί πως η ικανότητα της CaMKII, στην αυτόνομη κατάσταση, να φωσφορυλιώνει ποικίλα υποστρώματα είναι έως και 5 φορές μεγαλύτερη με προσδεμένη την καλμοντουλίνη από ότι χωρίς. Τέλος, η κατάσταση αυτή χαρακτηρίζεται ως υπερευαίσθητη και αυτό οφείλεται σε μια σειρά από συνεργατικά γεγονότα. Απαιτούνται 4 ιόντα ασβεστίου για την πλήρη ενεργοποίηση της καλμοντουλίνης και αυτό υποβοηθάτε με πρόσδεση στο ολοένζυμο, όπως έχει αναφερθεί. Επιπλέον η $Ca^{2+}$/CaM μπορεί να ενεργοποιήσει μια υπομονάδα, αλλά η

κάθε υπομονάδα μπορεί να φωσφορυλιώσει τη γειτονική της μόνο εάν αυτή είναι προσδεμένη στην $Ca^{2+}$/CaM (ώστε η Thr286 να είναι εκτεθειμένη).

Εντοπίζεται και μια δεύτερη θέση αυτοφωσφορυλίωσης στα κατάλοιπα Thr305 και/ή Thr306 η οποία αποτρέπει την ενεργοποίηση λόγω πρόσδεσης της $Ca^{2+}$/CaM, συνεισφέροντας έτσι στην αναστολή της ενζυμικής δραστικότητας της CaMKII. Αν και παρεμποδίζει την πρόσδεση της $Ca^{2+}$/CaM δεν παρεμποδίζει την μετάβασή του ολοενζύμου στην αυτόνομη κατάσταση, εφόσον έχει προηγηθεί φωσφορυλίωση της Thr286 πριν την φωσφορυλίωση των Thr305/306, γεγονός που καθιστά τη σειρά με την οποία θα συμβούν κομβικής σημασίας για την ρύθμιση της ενεργότητας του ολοενζύμου. Αυτοφωσφορυλίωση των Thr305/306 όμως αποτρέπει την περαιτέρω αυτοφωσφορυλίωση εφόσον η $Ca^{2+}$/CaM πρέπει να είναι προσδεμένη, έτσι ώστε το καταλυτικό κέντρο να είναι ικανό να φωσφορυλιώσει άλλα υποστρώματα ή να αποτελέσει το ίδιο υπόστρωμα για φωσφορυλίωση.




**Εικόνα 1.30** Διαφορετική απεικόνιση του δωδεκαμερούς που συγκροτεί το ολοένζυμο (η ακριβής τοποθέτηση των καταλυτικών κέντρων δεν είναι απόλυτη γνωστή). Η ρυθμιστική δομική αυτοτελής περιοχή τόσο στην ανενεργή όσο και στην ενεργή κατάσταση. Τέλος απεικόνιση της ανενεργούς κατάστασης όπου η ρυθμιστική περιοχή αλληλεπιδρά με την Thr286 στην περιοχή-T (κίτρινο) και παρεμποδίζει την πρόσδεση υποστρώματος στην περιοχή-S (πορτοκαλί)(Coultrap & Bayer, 2012).

Αξίζει να σημειωθεί πως η κρυσταλλογραφικά λυμένη δομή της CaMKII αποκάλυψε ένα διμερές, με τις δύο αυτοτελείς δομικές περιοχές αυτοαναστολής να σχηματίζουν ένα ακανόνιστη/τυχαία δομή που παρεμποδίζει την είσοδο τόσο του ATP όσο και του υποστρώματος. Αυτή η αλληλεπίδραση πιθανά αποτρέπει την τυχαία αυτοφωσφορυλίωση σε ένα περιβάλλον (ολοένζυμο CaMKII) στο οποίο η συγκέντρωση των καταλυτικών κέντρων είναι υψηλή.

Οι α,β ισομορφές εντοπίζονται αποκλειστικά στον εγκέφαλο και μαζί αποτελούν το 1% των πρωτεϊνών του και το 2% των πρωτεϊνών του ιππόκαμπου (μια δομή του εγκεφάλου που συμμετέχει στη μάθηση και στη μνήμη). Οι ισομορφές γ και δ παράγονται σε πολύ μικρότερη συγκέντρωση σε όλο το υπόλοιπο σώμα.

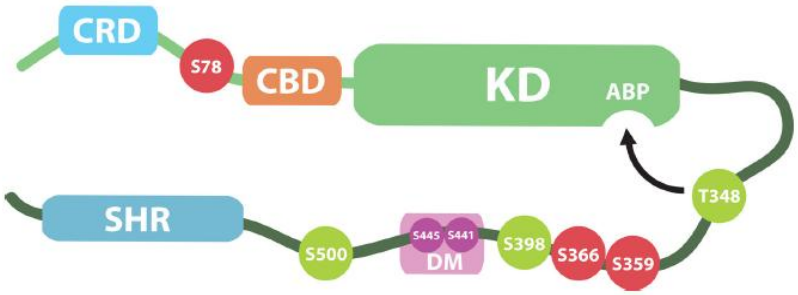
Στο εσωτερικό των κυττάρων συχνά οι υπομονάδες των διαφορετικών ισομορφών της CaMKII αλληλεπιδρούν μεταξύ τους συγκροτώντας ένα ετερομερές ολοένζυμο που αποτελείται από ποικίλες αναλογίες των διαφορετικών ισομορφών. Το γεγονός αυτό αυξάνει κατά πολύ την ποικιλία των βιολογικών λειτουργιών που συμμετέχει. Για παράδειγμα, η β ισομορφή περιέχει μια περιοχή πρόσδεσης στην ακτίνη και η συγγένεια της ως προς την ακτίνη μειώνεται όσο αυξάνεται η συγκέντρωση του ασβεστίου. Έτσι ένα ολοένζυμο που θα αποτελείται τόσο από α όσο και από β ισομορφές θα προσδένεται στον κυτταροσκελετό ακτίνης σε χαμηλές συγκεντρώσεις ασβεστίου και έτσι πιθανά θα περιορίζεται η ικανότητά του να ενεργοποιηθεί και να φωσφορυλιώσει πιθανά υποστρώματα. Σε περίπτωση αύξησης της συγκέντρωσης του ασβεστίου οι β υπομονάδες του ολοενζύμου θα απομακρυνθούν από τον κυτταροσκελετό και θα κατευθυνθούν προς το κυτταρόπλασμα όπου θα μπορούν να φωσφορυλιώσουν τις πρωτεΐνες-στόχους.

Συμμετέχει σε πληθώρα διαδικασιών που αφορούν το νευρικό σύστημα, όπως: η ρύθμιση της πλαστικότητας των συνάψεων, η ρύθμιση ενεργότητας υποδοχέων στο προσυναπτικό άκρο, η οργάνωση των πρωτεϊνών των συνάψεων (δρα ως πρωτεΐνη-ικρίωμα στο μετασυναπτικό άκρο), η μνήμη και η μάθηση (συμμετέχοντας στις συνάψεις υπεύθυνες για την μακρόχρονη ενδυνάμωση –LTP) αλλά και την καρδιά, όπως η ρύθμιση της σχέσης μεταξύ της δύναμης της κάθε σύσπασης αλλά και της συχνότητας των καρδιακών παλμών κατά την άσκηση.

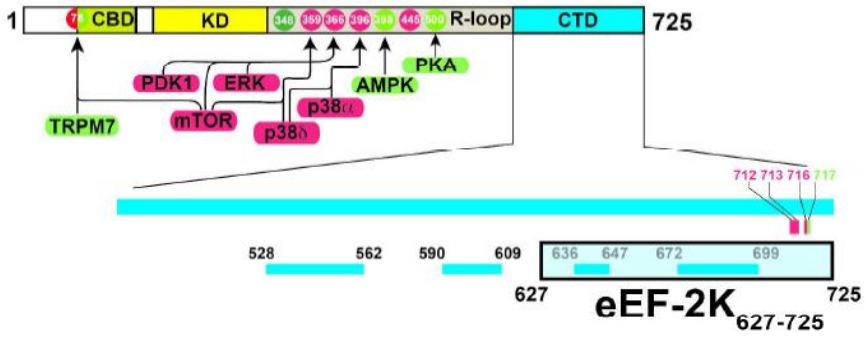
### 1.2.3.5. CaMK III κινάσες ή Elongation Factor 2 κινάσες

Η κινάση του ευκαρυωτικού παράγοντα επιμήκυνσης 2 (Eukaryotic elongation factor 2 kinase - eEF-2K), αποτελεί μέλος της άλφα οικογένειας κινασών (atypical kinase family). Εντοπίζεται στο κυτταρόπλασμα και δρα ως μονομερές (100 kDa). Πρόκειται για εξειδικευμένη κινάση, καθώς η μοναδική γνωστή πρωτεΐνη που αναγνωρίζει ως υπόστρωμα είναι ο παράγοντας επιμήκυνσης 2 (eukaryotic elongation factor 2 - eEF-2), του οποίου την ενεργότητα αναστέλλει μέσω φωσφορυλίωσής του στην Thr56. Ο ανενεργός eEF-2 δεν μπορεί να προσδεθεί πλέον στο ριβόσωμα, και αυτό έχει ως αποτέλεσμα την μείωση του ρυθμού πρωτεϊνοσύνθεσης.

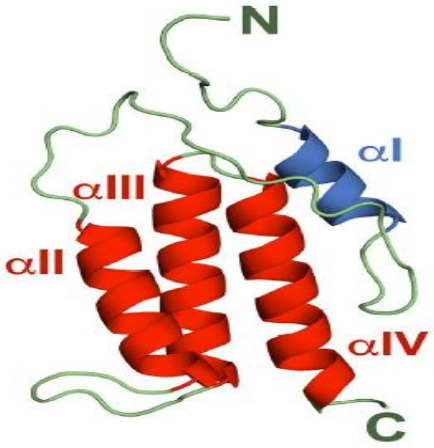
Οι Tavares et al.(2014) έδειξαν πως η eEF-2K ενεργοποιείται μέσω μιας μοναδικής διαδικασίας 2 βημάτων. Η διαδικασία αυτή δεν ακολουθεί το πρότυπο «απελευθέρωση της δομικά αυτοτελούς περιοχής αυτοαναστολής» που έχει περιγραφεί για τις υπόλοιπες CaMK, και αυτό σε συνδυασμό με το διαφορετικό, από άποψη ακολουθίας και δομής, καταλυτικό της κέντρο την διαφοροποιεί σημαντικά από αυτές.




**Εικόνα 1.31** Απεικονίζονται οι δομικά αυτοτελείς περιοχές καθώς και θέσεις φωσφορυλίωσης για την eEF-2K (κατάλοιπα 1–725). Πιθανή *CRD*: ρυθμιστική δομικά αυτοτελής περιοχή πρόσδεσης ασβεστίου, λόγω της ύπαρξης του μοτίβου D*X*D*X*DG, *CBD :* δομικά αυτοτελής περιοχή πρόσδεσης CaM, *KD* : το μη-τυπικό καταλυτικό κέντρο με μια περιοχή αλλοστερικής ρύθμισης (*ABP*), *DM* : μοτίβο degron – DSG*XX*S υπεύθυνο μεσολαβούμενη από ουβικιτίνη αποικοδόμηση και *SHR :* δομικά αυτοτελής περιοχή υπεύθυνη για την πρόσδεση του eEF-2 που περιέχει 3 πιθανές SEL1-like έλικες. Θέσεις φωσφορυλίωσης : στο κατάλοιπο Ser78 από μια άγνωστη ρυθμιζόμενη από TOR κινάση, στο κατάλοιπο Thr348 που αποτελεί θέση αυτοφωσφορυλίωσης, στο κατάλοιπο Ser359 από την p38 και το σύμπλοκο Cdc2-cyclin B, στο κατάλοιπο Ser366 από την κινάση p70 S6 και την p90, στο κατάλοιπο Ser398 από την AMPK, στο κατάλοιπο Ser441 από άγνωστη κινάση, στο κατάλοιπο Ser445 που αποτελεί θέση αυτοφωσφορυλίωσης και στο κατάλοιπο Ser500 που αποτελεί θέση αυτοφωσφορυλίωσης αλλά και φωσφορυλίωσης από την PKA. Πρόσδεση της φωσφορυλιωμένης Thr348 στο ABP διευκολύνει την πλήρη ενεργοποίηση της eEF-2K. Με πράσινο απεικονίζονται οι περιοχές που ενισχύουν την ενεργότητα της κινάσης ενώ με κόκκινο αυτές που την αναστέλλουν.(J Tavares et al., 2014)




**Εικόνα 1.32** Πιο αναλυτική απεικόνιση της διάταξης των δομικά αυτοτελών περιοχών και των θέσεων φωσφορυλίωσης. (Will et al., 2016) Η δομικά αυτοτελής περιοχή του C-τελικού άκρου (CTD) που έχει προβλεφθεί πως περιέχει αρκετά SLRs (SEL-1 like helical repeats), δομικό μοτίβο υπεύθυνο για αλληλεπιδράσεις μεταξύ πρωτεϊνών. Η περιοχή αυτή ρυθμίζει την ενεργότητα της κινάσης μέσω άμεσων αλληλεπιδράσεων με το καταλυτικό κέντρο. Επίσης μεταξύ των καταλοίπων 336-725 περιέχεται η περιοχή πρόσδεσης του eEF-2, ενώ μεταξύ των καταλοίπων 711-725 η περιοχή υπεύθυνη για την αναγνώριση του eEF-2. (Will et al., 2016)



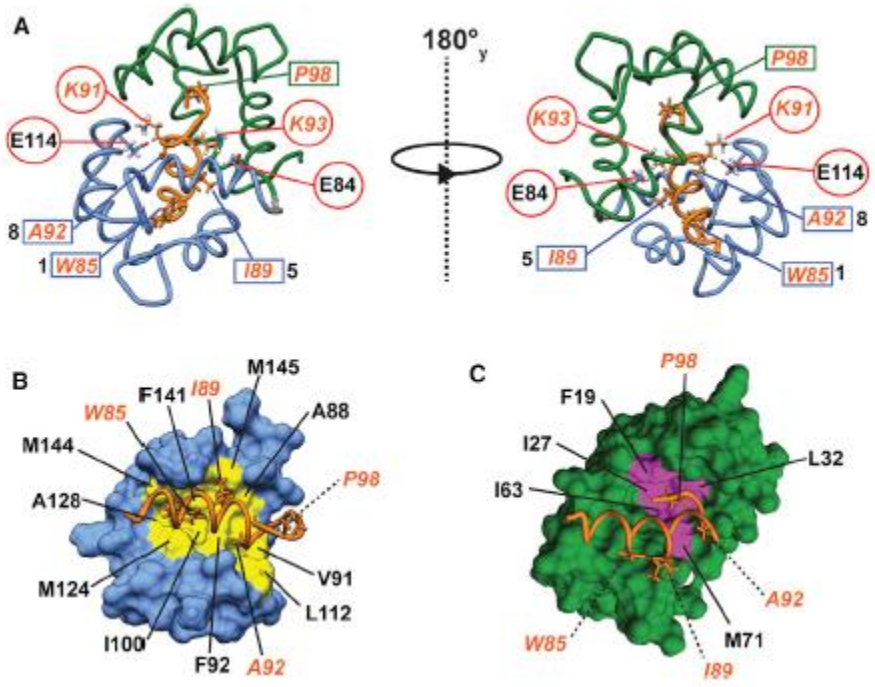

**Εικόνα 1.33** Δομή του CTD από τους Will et al., 2016. (Δεν υπάρχει ακόμα διαθέσιμη λυμένη δομή ολόκληρης της κινάσης)

Τα βήματα ενεργοποίησης της κινάσης έχουν ως εξής :

1. Αρχικά η $Ca^{2+}$-CaM αλληλεπιδρά κυρίως μέσω του C-λοβού με μια περιοχή στο αμινοτελικό άκρο της κινάσης. Αξίζει να σημειωθεί η απουσία ιόντων ασβεστίου στον C-λοβό, ακόμα και σε υψηλές συγκεντρώσεις. Η πρόσδεση της $Ca^{2+}$-CaM στην κινάση δεν είναι εφικτή χωρίς την πρόσδεση ιόντων ασβεστίου στον N-λοβό της CaM.

2. Η πρόσδεση της $Ca^{2+}$-CaM έχει ως αποτέλεσμα στερεοδιατακτικές αλλαγές στο καταλυτικό κέντρο της κινάσης. Αυτό έχει ως αποτέλεσμα την αύξηση, πάνω από 103 φορές, της ικανότητάς της να αυτοφωσφορυλιώνει την Thr348 που

εντοπίζεται στο εσωτερικό της ρυθμιστικής θηλιάς (R-loop) αμέσως μετά το καταλυτικό κέντρο.




**Εικόνα 1.34** (A) Το σύμπλοκο CaM/eEF-2KCBD σταθεροποιείται μέσω : υδρόφοβων αλληλεπιδράσεων των καταλοίπων 1-5-8 της περιοχής πρόσδεσης της καλμοντουλίνης της eEF-2K στον καρβοξυτελικό λοβό της CaM, μέσω αλληλεπιδράσεων μεταξύ της Pro98 της έλικας της κινάσης και του αμινοτελικού λοβού της CaM και μέσω ηλεκτροστατικών αλληλεπιδράσεων βασικών καταλοίπων της eEF-2K (απεικονίζονται με κόκκινους κύκλους) με όξινα κατάλοιπα της CaM. (B) Απεικονίζεται ο C-λοβός της CaM και (Γ) ο N-λοβός (K. Lee et al., 2016).

3. Έπειτα ακολουθεί αλλοστερική πρόσδεση του φωσφορυλιωμένου καταλοίπου θρεονίνης σε μια θήκη αλλοστερικής πρόσδεσης στο εσωτερικό του καταλυτικού κέντρου. Αυτό διευκολύνει μια δεύτερη στερεοδιατακτική αλλαγή της θηλιάς R που έχει ως αποτέλεσμα την μέγιστη ικανότητα φωσφορυλίωσης του υποστρώματος.

Η ρύθμιση της πρωτεϊνοσύνθεσης από τον eEF-2K είναι κομβικής σημασίας για πληθώρα κυτταρικών διαδικασιών, όπως ο έλεγχος του κυτταρικού κύκλου και η ρύθμιση της σύνθεσης των πρωτεϊνών που είναι απαραίτητες για τη διαμόρφωση των συνάψεων. Συνεπώς απαιτείται η φυσιολογική λειτουργία της eEF-2K για τη σωστή λειτουργία των νευρώνων (σε διαδικασίες σχετικές με την μάθηση και την μνήμη), ενώ η μη φυσιολογική λειτουργία της έχει συνδεθεί με αρκετές ασθένειες, όπως κάποιες μορφές καρκίνου (ορμονοεξαρτώμενος καρκίνος του στήθους, γλοιοβλάστωμα κ.α.), η νόσος Alzheimer's, η κατάθλιψη και διάφορες παθήσεις στην καρδιά.

### 1.2.3.6. Κινάσες της ελαφριάς αλυσίδας της μυοσίνης (Myosin Light Chain κινάσες - MLCK)

Εντοπίζονται 3 ισομορφές της κινάσης της ελαφριάς αλυσίδας της μυοσίνης (Myosin Light Chain Kinase – MLCK), οι οποίες κωδικοποιούνται από 3 γονίδια.

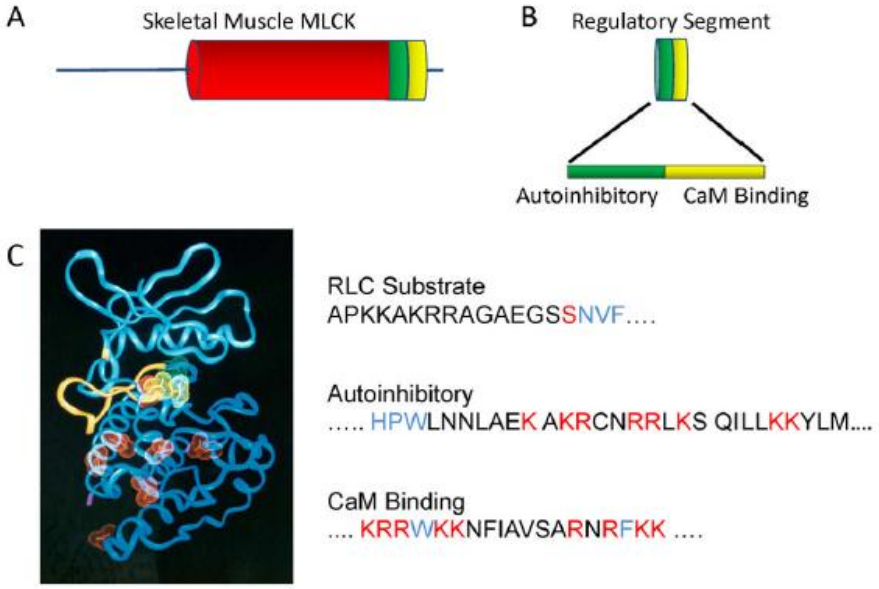
**skMLCK (skeletal muscle MLCK)**: Κωδικοποιείται από το γονίδιο MYLK2 και τα μεγέθη της ποικίλουν από 69 έως 150 kDa. Βασική μονάδα του κάθε μυοϊνιδίου αποτελεί το σαρκομέριο, που αποτελείται από τις βαριές αλυσίδες των μορίων μυοσίνης και τα λεπτά νημάτια που αποτελούνται από μόρια ακτίνης, τροπονίνης και τροπομυοσίνης. Ο χώρος μεταξύ γειτονικών παχιών και λεπτών νηματίων γεφυρώνεται από τις σφαιρικές κεφαλές των μορίων μυοσίνης, που ονομάζονται εγκάρσιες γέφυρες και προεξέχουν προς τα λεπτά νημάτια. Εκεί εντοπίζονται και οι ελαφριές αλυσίδες της μυοσίνης. Κατά τη μυϊκή συστολή οι γέφυρες αυτές εφάπτονται των λεπτών νηματίων και τους εξασκούν δύναμη. Η εγκάρσια γέφυρα της μυοσίνης περιέχει την επιφάνεια πρόσδεσης στην ακτίνη και μια θέση πρόσδεσης ATP και συνδέεται με τον πυρήνα του παχέος νήματος μέσω μιας α-έλικας. Γύρω από κάθε α-έλικα εντοπίζονται δύο μικρές υπομονάδες, η ελαφριά αλυσίδα και η ρυθμιστική ελαφριά αλυσίδα (Regulatory Light Chain - RLC) και προσφέρουν μηχανική στήριξη. Απελευθέρωση ασβεστίου στα σαρκομερή ενεργοποιεί και την skMLCK που με τη σειρά της φωσφορυλιώνει την RLC. Η φωσφορυλίωση της RLC δεν έχει σημαντικό αντίκτυπο στην ενεργότητα ATP-άσης της μυοσίνης, λόγω ενεργοποίησης από την ακτίνη, αλλά προωθεί την κίνηση της κεφαλής της μυοσίνης πάνω στα νημάτια ακτίνης, γεγονός που συμβαίνει μόνο όταν η τροπομυοσίνη έχει απομακρυνθεί από τις θέσεις πρόσδεσης της μυοσίνης στην ακτίνη ως αποτέλεσμα της πρόσδεσης ιόντων ασβεστίου στην τροπονίνη.

Η SkMLCK είναι μονομερές. Περιέχει μια περιοχή στο αμινοτελικό άκρο άγνωστης λειτουργίας, έπειτα ακολουθεί το καταλυτικό κέντρο με ενεργότητα κινάσης, η δομικά αυτοτελής περιοχή υπεύθυνη για αυτοαναστολή και η δομικά αυτοτελής περιοχή πρόσδεσης στην Ca2+-CaM. Το εύρος των μεγεθών της οφείλεται στις διαφορετικού μήκους αλληλουχίες της περιοχής άγνωστης λειτουργίας μεταξύ οργανισμών διαφορετικών ειδών. Παρόλο που δεν υπάρχει διαθέσιμη δομή του καταλυτικού κέντρου της skMLCK, λόγω της ομοιότητας της αμινοξικής της ακολουθίας με άλλες κινάσες, υποθέτουμε πως εμφανίζει τη χαρακτηριστική δομή των 2 λοβών.

Απουσία ασβεστίου η κινάση βρίσκεται σε κατάσταση αυτοαναστολής και η CaM δεν είναι προσδεμένη. Η περιοχή υπεύθυνη για αυτοαναστολή συμμετέχει σε ένα εκτεταμένο δίκτυο αλληλεπιδράσεων με κατάλοιπα που εκτείνονται από την επιφάνεια του C-λοβού μέχρι και το ενεργό κέντρο. Η περιοχή πρόσδεσης της καλμοντουλίνης παρεμποδίζει την πρόσδεση της RLC, αλλά όχι του ATP.

Σε αυξημένες συγκεντρώσεις ασβεστίου η skMLCK ενεργοποιείται μέσω της πρόσδεσης της Ca2+-CaM στην αντίστοιχη περιοχή και την επακόλουθη απελευθέρωση του ενεργού κέντρου. Έτσι προσδένεται η RLC, και φωσφορυλιώνεται στην Ser15.

Η Ca2+-CaM εμφανίζει υψηλή συγγένεια πρόσδεσης με την skMLCK (KCaM = 1 nM), γεγονός σημαντικό για τη φυσιολογία των σκελετικών μυών. Πρόσδεση της Ca2+-CaM στην α-έλικα της skMLCK (που περιέχει σημαντικά για την αλληλεπίδραση βασικά και υδρόφοβα κατάλοιπα) επάγει στερεοδιατακτικές αλλαγές στην Ca2+-CaM. Η έλικα-συνδέτης της Ca2+-CaM λυγίζει και στρίβει με αποτέλεσμα να πλησιάσουν οι δύο δομικά αυτοτελείς περιοχές, του αμινοτελικού και του καρβοξυτελικού άκρου, μεταξύ τους σχηματίζοντας ένα τούνελ το οποίο διαπερνά η α-έλικα της κινάσης με αντιπαράλληλο τρόπο.

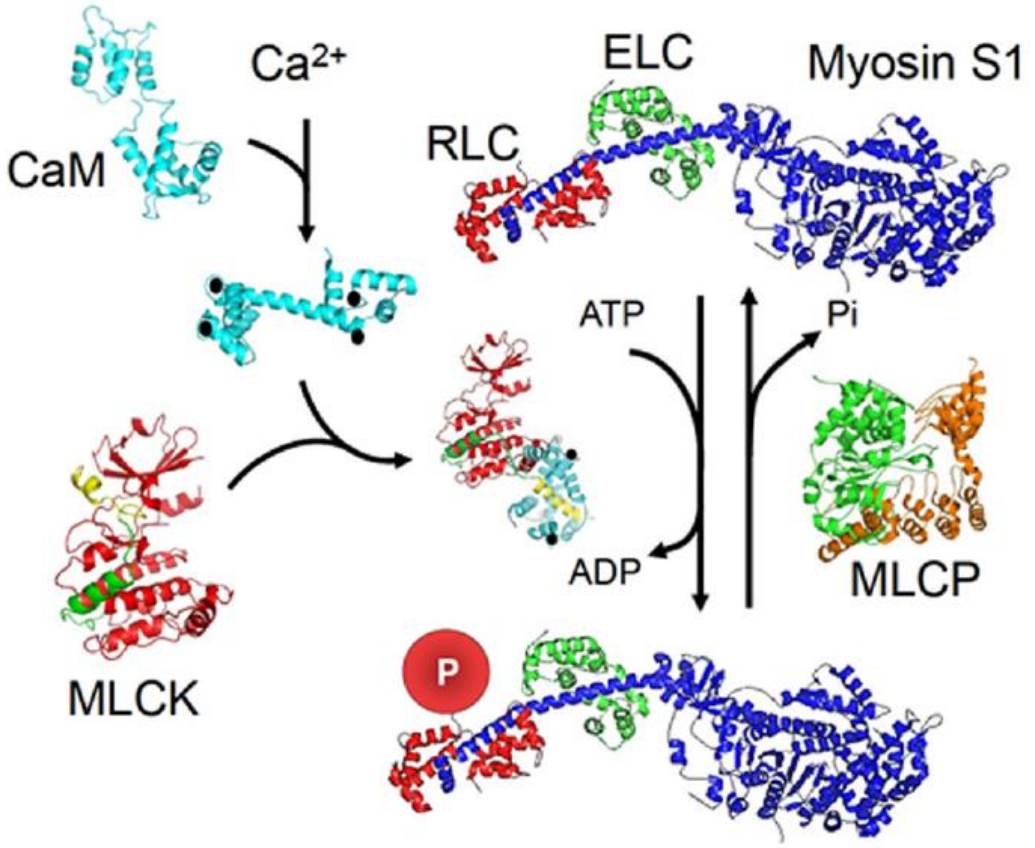



**Εικόνα 1.35** Σημαντικά δομικά χαρακτηριστικά για την skMLCK. (A), (B) Διάταξη των δομικά αυτοτελών περιοχών της SkMLCK, με κόκκινο απεικονίζεται το καταλυτικό κέντρο, με πράσινο η περιοχή υπεύθυνη για αυτοαναστολή και με κίτρινο η δομικά αυτοτελής περιοχή πρόσδεσης της καλμοντουλίνης. (C) Αριστερά, στη δομή της κινάσης απεικονίζονται με κόκκινο κατάλοιπα που πιθανά αλληλεπιδρούν με κατάλοιπα της περιοχής αυτοαναστολής και με κίτρινο κατάλοιπα σημαντικά για τον σωστό προσανατολισμό του υποστρώματος. Δεξιά, αρχικά απεικονίζονται τα σημαντικά κατάλοιπα του υποστρώματος (μπλε) για την φωσφορυλίωση της σερίνης (κόκκινο), στη μέση απεικονίζονται τα κατάλοιπα της περιοχής αυτοαναστολής, όπου συμμετέχουν κατάλοιπα του καταλυτικού κέντρου (μπλε), ενώ τα βασικά κατάλοιπα (κόκκινο) προσδένονται σε όξινα κατάλοιπα της επιφάνειας του καταλυτικού κέντρου, τέλος απεικονίζονται τα κατάλοιπα της περιοχής πρόσδεσης της καλμοντουλίνης, όπου σημειώνονται 2 υδρόφοβα κατάλοιπα (μπλε) και βασικά κατάλοιπα (κόκκινα) σημαντικά για την αλληλεπίδραση (James T. Stull, Kristine E. Kamm, 2012).

Όσον αφορά το σύνολο αλληλεπιδράσεων που αναπτύσσονται μεταξύ της δομικά αυτοτελούς περιοχής πρόσδεσης της καλμοντουλίνης και της καλμοντουλίνης αναφέρονται τα εξής: 1ον) στην περιοχή πρόσδεσης ένα κατάλοιπο τρυπτοφάνης και ένα φαινυλαλαλίνης σημειώνονται ως εκείνα που συμμετέχουν στην αγκυροβόληση (όπως απεικονίζονται στην εικόνα με μπλε χρώμα). 2ον) Αναπτύσσονται υδρόφοβες αλληλεπιδράσεις μεταξύ μιας υδρόφοβης επιφάνειας της μιας πλευράς του τούνελ και μιας υδρόφοβης επιφάνειας της α-έλικας. 3ον) Σχηματίζονται γέφυρες άλατος μεταξύ των βασικών καταλοίπων της κινάσης και καταλοίπων της καλμοντουλίνης. 4ον) Τα όρια των αλληλεπιδράσεων στις άκρες του πεπτιδίου πλαισιώνονται από ηλεκτροστατικές αλληλεπιδράσεις και δεσμούς υδρογόνου.

Όπως αναφέρθηκε και εισαγωγικά, αρχικά προσδένεται η δομική αυτοτελής περιοχή του καρβοξυτελικού άκρου της καλμοντουλίνης και ακολουθεί η πρόσδεση της αμινοτελικής. Η κεντρική έλικα της καλμοντουλίνης λειτουργεί ως ένας ευέλικτος συνδέτης και διευκολύνει την πρόσδεση της αμινοτελικής περιοχής της καλμοντουλίνης στο C-τελικό τμήμα της δομική αυτοτελούς περιοχής πρόσδεσης της καλμοντουλίνης και την επακόλουθη ενεργοποίηση της κινάσης. Το κατάλοιπο αγκυροβόλησης Trp στο N-τελικό άκρο βρίσκεται με την πλευρική του αλυσίδα εκτεθειμένη στο διαλύτη, με τέτοιο τρόπο ώστε να είναι προσβάσιμο από το C-τελικό άκρο της Ca2+-CaM. Σε αντίθεση, το C-τελικό άκρο της περιοχής πρόσδεσης της Ca2+-CaM εντοπίζεται θαμμένο κοντά στην εσοχή μεταξύ των δύο λοβών την κινάσης και πιθανά δεν δύναται αρχικά να αλληλεπιδράσει με το N-τελικό άκρο της Ca2+-CaM.

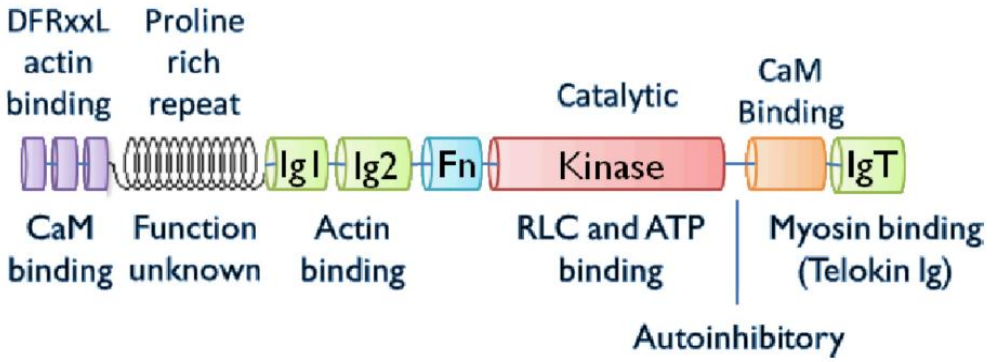
Πρόσδεση της Ca2+-CaM έχει σαν αποτέλεσμα την απελευθέρωση του ενεργού κέντρου από την δομική αυτοτελή περιοχή αυτοαναστολής και την επακόλουθη πρόσδεση του N-τελικού άκρου της RLC. Η εσοχή μεταξύ των λοβών έτσι κλείνει και επιτελείται η μεταφορά της φωσφορικής ομάδας από το ATP στην RLC.



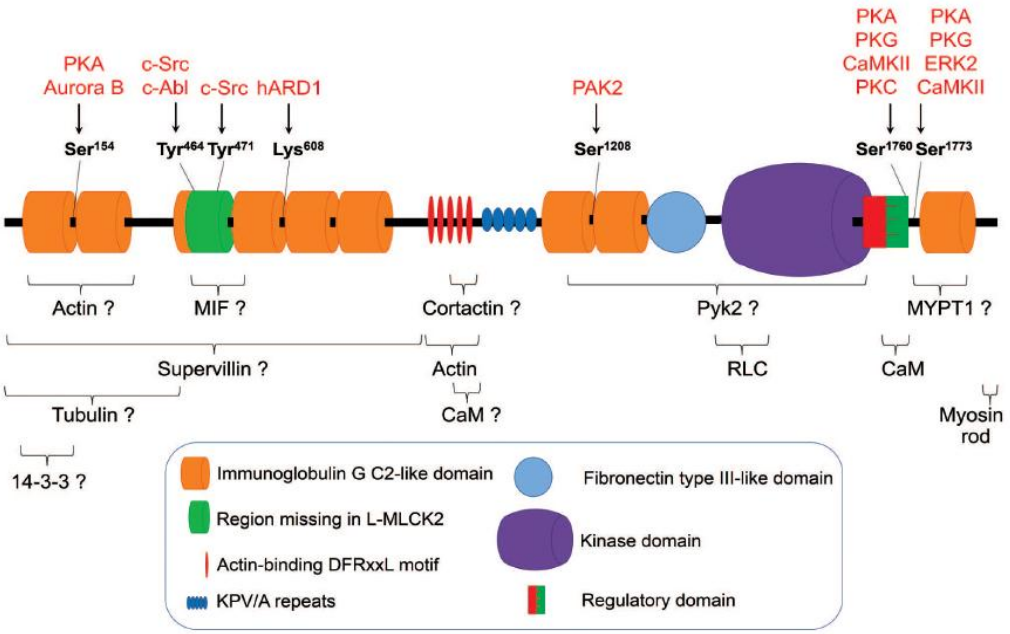

**Εικόνα 1.36** Σχηματική αναπαράσταση της φωσφορυλίωσης της RLC της μυοσίνης στους ραβδωτούς μύες βασισμένη στις δομές των πρωτεϊνών που συμμετέχουν. Η δομική αυτοτελής περιοχή υπεύθυνη για αυτοαναστολή απεικονίζεται με πράσινο ενώ η δομική αυτοτελής περιοχή πρόσδεσης της $Ca^{2+}$-CaM με κίτρινο. Το καταλυτικό κέντρο απεικονίζεται με κόκκινο. Η $Ca^{2+}$-CaM απεικονίζεται με γαλάζιο. Με μπλε απεικονίζεται ένα τμήμα της κεφαλής της μυοσίνης (βαριά αλυσίδα) καθώς και η έλικα που αλληλεπιδρά τόσο με την ελαφριά αλυσίδα (Essential Light Chain – ECL, πράσινο), όσο και με την RLC (κόκκινο). Η εγγενώς μη δομημένη περιοχή της RLC εντοπίζεται στο Ν-τελικό της άκρο (δεν απεικονίζεται) και βρίσκεται εκτεθειμένη για φωσφορυλίωση από την ενεργοποιημένη skMLCK. Τέλος απεικονίζεται και η φωσφατάση της ελαφριάς αλυσίδας της μυοσίνης (Myosin light chain phosphatase - MLCP) που είναι υπεύθυνη για την αποφωσφορυλίωση της RLC, με πράσινο η καταλυτική υπομονάδα και με πορτοκαλί η ρυθμιστική υπομονάδα. (James T. Stull, Kristine E. Kamm, 2012).

**smMLCK (smooth muscle MCLK)**: Κωδικοποιείται από το γονίδιο MYLK1 και τα μεγέθη της ποικίλουν από 130 έως 150 kDa. Εντοπίζονται δύο ισομορφές της κινάσης, μια στις λείες μυϊκές ίνες των λείων μυών και στη βιβλιογραφία εντοπίζεται ως smooth muscle MLCK ή light MLCK ή Short MLCK (στο εξής SsmMLCK) και η δεύτερη σε πολλούς διαφορετικούς ιστούς, πλην των μυϊκών και στη βιβλιογραφία εντοπίζεται ως non-muscle MLCK ή heavy MLCK ή Long MLCK (στο εξής LsmMLCK) . Στο αμινοτελικό άκρο της LsmMLCK εντοπίζεται μια επιπλέον περιοχή, σε σύγκριση με την SsmMLCK, μήκους περίπου 900 αμινοξικών καταλοίπων. Εντοπίζονται επίσης αρκετές ισομορφές της LsmMLCK στον άνθρωπο, που προκύπτουν από εναλλακτική συρραφή.

Στο αμινοτελικό άκρο της SsmMLCK εντοπίζεται μια δομική αυτοτελής περιοχή πρόσδεσης ακτίνης. Η περιοχή αυτή αποτελείται από τρία μοτίβα DFRxxL. Στα μοτίβα αυτά προσδένεται η ακτίνη, αλλά και η Ca2+-CaM, γεγονός που εξασθενεί την πρόσδεση της ακτίνης. Η ακτίνη προσδένεται επίσης στις περιοχές Ig1 και/ή Ig2. Το καταλυτικό κέντρο με ενεργότητα κινάσης προσδένει το ATP καθώς και την αμινοτελική περιοχή της ρυθμιστικής περιοχής της ελαφριάς αλυσίδας της μυοσίνης (myosin Regulatory Light Chain - RLC), φωσφορυλιώνοντάς την στην Ser19. Η περιοχή IgT προσδένεται στη μυοσίνη των λείων μυών (Smooth Muscle Myosin - SMM). Συγκεκριμένα, προσδένεται τόσο σε ινίδια SMM, όσο και σε μονομερή SMM, στην περιοχή μεταξύ των δύο κεφαλών της SMM και της ουράς. Οι λειτουργίες της δομικής αυτοτελούς περιοχής της φιμπρονεκτίνης καθώς και της περιοχής πλούσιας σε προλίνη δεν είναι γνωστές.



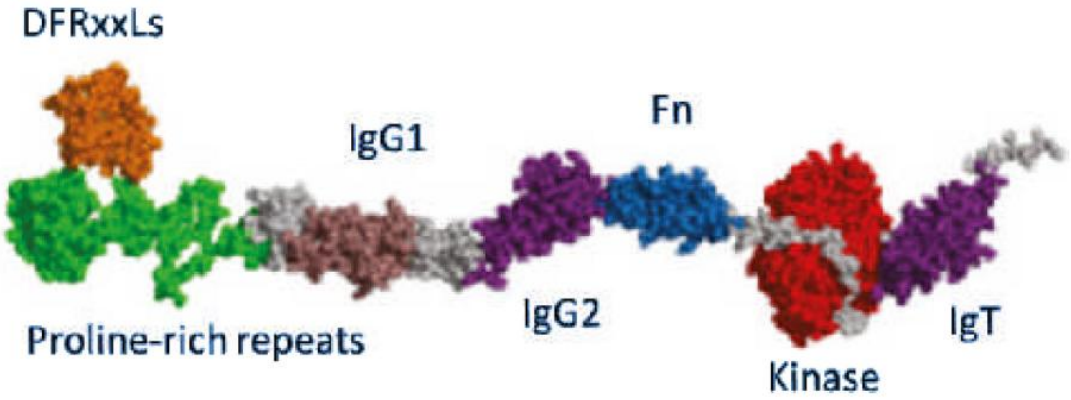

**Εικόνα 1.37** Διάταξη των δομικά αυτοτελών περιοχών της SsmMLCK, όπως εντοπίζονται συντηρημένες στα θηλαστικά. Αξίζει να αναφερθεί πως στα πτηνά δεν εντοπίζεται η περιοχή πλούσια σε προλίνη. (Hong, Haldeman, Jackson, Carter, & Baker, 2012)




**Εικόνα 1.38** Διάταξη των δομικών αυτοτελών περιοχών, των αλληλεπιδράσεων πρωτεΐνης-πρωτεΐνης καθώς και των μετα-μεταφραστικών τροποποιήσεων της LsmMLCK. Η

διαφορά με την SsmMLCK εντοπίζεται στην ύπαρξη έξι επιπλέον IG περιοχών στο αμινοτελικό άκρο της πρωτεΐνης. (PKG : protein kinase G, PKC : protein kinase C, CaMKII : Ca-CaM-dependent kinase of type II (Khapchaev & Shirinsky, 2016).

Γενικά η MLCK είναι ένα επίμηκες και πιθανά ευέλικτο μόριο. Δεν υπάρχει διαθέσιμη δομή για ολόκληρη την κινάση. Συγκρίσεις με διαθέσιμες δομές πρωτεϊνών, με τις οποίες εμφανίζει υψηλή ομολογία, έχουν οδηγήσει στο συμπέρασμα πως μπορεί να αποκτήσει επιμήκεις και συμπαγείς στερεοδιατάξεις, γεγονός που την καθιστά αρκετά ευέλικτο μόριο. Το τελικό μήκος που δύναται να αποκτήσει η επιμήκης στερεοδιάταξη της MLCK την καθιστά ικανή να εκτείνεται μεταξύ των παχιών και των λεπτών νηματίων του λείου μυ (η μεταξύ τους απόσταση σε χαλαρό μυ είναι περίπου 26 nm).




**Εικόνα 1.39** Μοριακό μοντέλο πρόβλεψης δομής της MLCK(Hong et al., 2012).

Σε χαμηλές συγκεντρώσεις ασβεστίου η κινάση είναι ανενεργή. Η δομική αυτοτελής περιοχή αυτοαναστολής δεν προσδένεται ως ψευδοϋπόστρωμα στο ενεργό κέντρο (σημειώνονται ως σημαντικά για την αυτοαναστολή κατάλοιπα τα υδρόφοβα Tyr, Met και Ala, ενώ δεν εντοπίζονται αντίστοιχα σημαντικά για την πρόσδεση της RLC στο ενεργό κέντρο). Η αναστολή επιτυγχάνεται με ενδοστερική ρύθμιση, η δομική αυτοτελής περιοχή αυτοαναστολής αλληλεπιδρά με κατάλοιπα του καταλυτικού κέντρου, χωρίς όμως να αλληλεπιδρά με κατάλοιπα υπεύθυνα για την πρόσδεση του υποστρώματος. Η $Ca^{2+}$-CaM αποτελεί τον κύριο ρυθμιστή της ενεργότητας της MLCK. Ακόμα και σε χαμηλά επίπεδα ασβεστίου, η καρβοξυτελική περιοχή της CaM βρίσκεται προσδεμένη στην MLCK. Αύξηση της ενδοκυτταρικής συγκέντρωσης ασβεστίου και πρόσδεσή του στην αμινοτελική περιοχή της CaM επιφέρει στερεοδιατακτικές αλλαγές, τόσο στην $Ca^{2+}$-CaM, όσο και στην κινάση. Αναπτύσσονται περαιτέρω αλληλεπιδράσεις μεταξύ της $Ca^{2+}$-CaM και της δομικής αυτοτελούς περιοχής πρόσδεσης της $Ca^{2+}$-CaM (η οποία εντοπίζεται στο καρβοξυτελικό άκρο της περιοχής αυτοαναστολής με σημαντική επικάλυψη μεταξύ τους και πριν το αμινοτελικό άκρο της IgT δομικής αυτοτελούς περιοχής) και έτσι

απομακρύνεται η δομική αυτοτελής περιοχή αυτοαναστολής από την επιφάνεια του καταλυτικού κέντρου, επιτρέποντας έτσι την πρόσδεση του υποστρώματος (RLC).

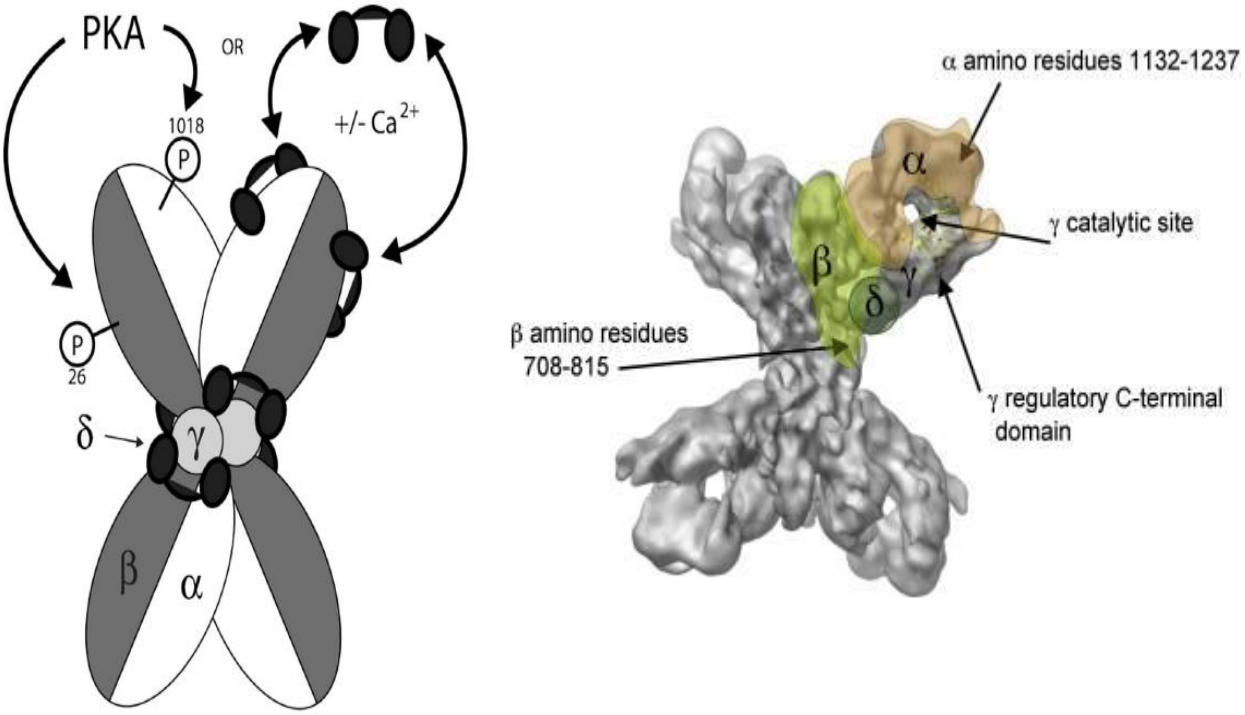
Αξίζει σε αυτό το σημείο να αναφερθεί πως η smMLCK (και συγκεκριμένα το καρβοξυτελικό άκρο της δομικής αυτοτελούς περιοχής πρόσδεσης της $Ca^{2+}$-CaM) αποτελεί υπόστρωμα και για την CaMKII. Η φωσφορυλίωση αυτή οδηγεί σε μείωση της φωσφορυλίωσης της RLC και στην επακόλουθη χαλάρωση του μυ.

Το καταλυτικό κέντρο των smMLCK εμφανίζει υψηλή ομολογία με το καταλυτικό κέντρο της PKA που περιγράφηκε εισαγωγικά. Εντοπίζεται η συναινετική ακολουθία Gly-X-Gly-X-X-Gly-(X)14-Lys στην περιοχή πρόσδεσης του $Mg^{2+}$ και του ATP. Μια περιοχή κοντά στο αμινοτελικό άκρο στην οποία εντοπίζονται αρκετά όξινα κατάλοιπα, διαδραματίζει σημαντικό ρόλο στην πρόσδεση του υποστρώματος (RLC). Η RLC προσδένεται στην κινάση όταν βρίσκονται ήδη προσδεμένο το $Mg^{2+}$ και το ATP. Όσον αφορά το υπόστρωμα, στην RLC το κατάλοιπο Arg16 είναι υπεύθυνο για την πρόσδεση του υποστρώματος και τον προσανατολισμό του καταλοίπου Ser19 της RLC στο ενεργό κέντρο της κινάσης και την επακόλουθη φωσφορυλίωσή του. Έχει παρατηρηθεί μια δεύτερη θέση φωσφορυλίωσης, στο κατάλοιπο Thr18 που έχει ως αποτέλεσμα την ενίσχυση της ενεργότητας της RLC, στην περίπτωση που έχει προηγηθεί φωσφορυλίωση στο Ser19, ενώ αν δεν έχει προηγηθεί φωσφορυλίωση συνδέεται με την χαλάρωση των λείων μυών.

**cMLCK (cardiac muscle MCLK)**: Κωδικοποιείται από το γονίδιο MYLK3. Είναι απαραίτητη για την φωσφορυλίωση της MLC2 στο μυοκάρδιο (συγκεκριμένα στην οργάνωση των σαρκομερών) (Wazen et al., 2014). Στον καρδιακό μυ ο κυρίαρχος ρυθμιστικός μηχανισμός βασίζεται στη δράση της τροπονίνης (μια πρωτεΐνη που προσδένει ιόντα ασβεστίου και εντοπίζεται στα λεπτά νημάτια). Πρόσδεση ασβεστίου στην τροπονίνη οδηγεί τον μυ σε σύσπαση. Εντοπίζεται μια κλίση στην φωσφορυλίωση της MLC2, από τη βάση στην κορυφή και από το ενδοκάρδιο στο επικάρδιο. Αυτό οδηγεί στο συμπέρασμα ότι συμμετέχει στην διατήρηση της κανονικής συστολής της καρδιάς. Επιπλέον η φωσφορυλίωση της MLC2 μπορεί να επηρεάζει τις αλληλεπιδράσεις που αναπτύσσονται μεταξύ των νηματίων ακτίνης και μυοσίνης, ανεξάρτητα από τις προσδεμένες στην ακτίνη πρωτεΐνες (τροπονίνη και τροπομυοσίνη)(Taniguchi et al., 2015). Ο ακριβής μηχανισμός ρύθμισης της MLCK3 από την καλμοντουλίνη δεν είναι γνωστός.

### 1.2.3.7. Κινάσες της φωσφορυλάσης

Είναι από τις μεγαλύτερες και πιο πολύπλοκες, δομικά, μεταξύ των γνωστών CaMK. Αποτελείται από 4 υπομονάδες, τις α,β,γ και δ, που αλληλεπιδρούν σχηματίζοντας ένα ετεροτετραμερές. Τέσσερα ετεροτετραμερή αλληλεπιδρούν σχηματίζοντας το ολοένζυμο της κινάσης της φωσφορυλάσης (1,300 kDa), όπως απεικονίζεται στις παρακάτω εικόνες.(Swulius & Waxham, 2008)(Vénien-Bryan et al., 2009)




**Εικόνα 1.40** Απεικόνιση του ολοενζύμου της κινάσης της φωσφορυλάσης, που αποτελείται από τα ετεροτετραμερή των υπομονάδων α,β,γ,δ.

Αναλόγως τον κυτταρικό τύπο παρατηρείται ποικιλία ολοενζύμων, γεγονός που οφείλεται στις διαφορετικές ισομορφές τις κάθε υπομονάδας που προκύπτουν λόγω εναλλακτικής συρραφής και διαφορετικής ρύθμισης της έκφρασης των αντίστοιχων γονιδίων.

Η υπομονάδα γ αποτελεί την υπομονάδα με ενεργότητα κινάσης (μέγεθος 45kDa) και εμφανίζει ομοιότητες με τις υπόλοιπες CaMK. Έτσι εντοπίζεται στο αμινοτελικό άκρο το καταλυτικό κέντρο, έπειτα η δομικά αυτοτελής περιοχή υπεύθυνη για αυτοαναστολή και τέλος η δομικά αυτοτελής περιοχή υπεύθυνη για πρόσδεση της $Ca^{2+}$-CaM. Το καταλυτικό κέντρο της γ υπομονάδας μοιάζει σε μεγάλο βαθμό με αυτό της PKA που περιγράφηκε εισαγωγικά.

Οι α,β,δ υπομονάδες είναι ρυθμιστικές. Αξίζει να αναφερθεί πως η δ υπομονάδα είναι καλμοντουλίνη και αποτελεί σταθερό μέλος του ολοενζύμου,

ακόμα και αν η συγκέντρωση ενδοκυττάριου ασβεστίου βρίσκεται σε επίπεδα nM (χωρίς να είναι γνωστός ξεκάθαρα ο μηχανισμός). Έχει υποτεθεί πως, λόγω υψηλής ομοιότητας της περιοχής πρόσδεσης στην CaM της γ υπομονάδας με την περιοχή πρόσδεσης στην τροπονίνη-C, μια πρωτεΐνη ομόλογη της καλμοντουλίνης, της τροπονίνης I (σύμπλοκο που παραμένει προσδεμένο ανεξάρτητα από τη συγκέντρωση ασβεστίου), η προσδεμένη στην γ υπομονάδα CaM εντοπίζεται στην εκτεταμένη της μορφή (όπως έχει αναλυθεί εισαγωγικά) επιτρέποντας έτσι την ανάπτυξη περισσότερων διαπρωτεϊνικών επαφών που σταθεροποιούν την αλληλεπίδραση. Παρατηρούνται επιπλέον επαφές με την προσθήκη και των υπόλοιπων ρυθμιστικών υπομονάδων.

Παρουσία χαμηλών συγκεντρώσεων ασβεστίου η κινάση της φωσφορυλάσης παραμένει σε αναστολή, με την περιοχή υπεύθυνη για αυτοαναστολή προσδεμένη στο ενεργό κέντρο, παρεμποδίζοντας έτσι την πρόσδεση της φωσφορυλάσης του γλυκογόνου (πρωτεΐνη υπόστρωμα της κινάσης που συμμετέχει και ρυθμίζει το μεταβολισμό του γλυκογόνου στους μύες και στο ήπαρ).

Ενεργοποιείται με αύξηση της συγκέντρωσης του ασβεστίου και πρόσδεση των ιόντων στη δ υπομονάδα, επάγονται στερεοδιατακτικές αλλαγές που έχουν ως αποτέλεσμα την απελευθέρωση της περιοχής υπεύθυνης για αυτοαναστολή και την επακόλουθη πρόσδεση και φωσφορυλίωση του υποστρώματος.

Οι υπομονάδες α και β είναι σχετικά μεγάλες πρωτεΐνες ,αποτελούν το 80% του μεγέθους του ολοενζύμου της κινάσης και συμμετέχουν στην αναστολή της δραστικότητας κινάσης του ολοενζύμου. Εντοπίζονται στο εσωτερικό τους περιοχές φωσφορυλίωσης, μεταξύ άλλων στην Ser1018 στην α υπομονάδα και στην Ser26 στη β υπομονάδα, τις οποίες φωσφορυλιώνει η PKA κατά τον καταρράκτη αποικοδόμησης του γλυκογόνου ως απόκριση στην πρόσδεση ορμονών (όπως η επινεφρίνη) στους υποδοχείς που αλληλεπιδρούν με G πρωτεΐνες. Η φωσφορυλίωση της β υπομονάδας έχει ως αποτέλεσμα την άμεση αύξηση της ενεργότητας της γ υπομονάδας, ενώ φωσφορυλίωση της α υπομονάδας ενισχύει περαιτέρω την ενεργότητά της (δεν αρκεί όμως για την ενεργοποίηση της γ υπομονάδας). Αξίζει τέλος να σημειωθεί πως δεν υπάρχουν πολλές δομικές πληροφορίες σχετικά με τις α,β υπομονάδες, αλλά τα κατάλοιπα 918–1093 της β υπομονάδας εμφανίζουν ομοιότητες με τις calcineurin B-like πρωτεΐνες, που αποτελούνται από 2 ζευγάρια μοτίβων EF-hand και ίσως έτσι συμμετέχουν στην ρύθμιση της κινάσης από το ασβέστιο.

Ένας επιπλέον μηχανισμός ρύθμισης εξαρτώμενος από τα επίπεδα ασβεστίου αφορά μια υπομονάδα που ονομάζεται δ’ υπομονάδα ή εξωγενής $Ca^{2+}$-CaM. Η δ’ υπομονάδα δεν παραμένει προσδεμένη στο ετεροτετραμερές, αλλά σε υψηλά επίπεδα ασβεστίου προσδένεται τόσο στην α όσο και στη β υπομονάδα. Η πρόσδεσή της δεν επιφέρει περαιτέρω ενεργοποίηση της κινάσης, αν έχει

προηγηθεί η ενεργοποίησή της, όμως αποτελεί έναν εναλλακτικό μηχανισμό ρύθμισης μέσω του οποίου αυξάνεται η ενεργότητα της κινάσης χωρίς να έχει προηγηθεί φωσφορυλίωση των α και β υπομονάδων.

Εκτός από την εξαρτώμενη από ασβέστιο ρύθμισή της, καθώς και την εξαρτώμενη από cAMP, μέσω της φωσφορυλίωσης από την PKA, ρυθμίζεται περαιτέρω μέσω της πρόσδεσης ADP και μέσω αλλαγών στο pH. Έτσι είναι ικανή να ρυθμίζει τον μεταβολισμό του γλυκογόνου σε ποικιλία ενδοκυτταρικών συνθηκών.

Η κινάση της φωσφορυλάσης εντοπίζεται σε όλους τους ιστούς, αλλά κυριαρχεί στους σκελετικούς μύες (όπου ρυθμίζει την σύσπαση καθώς το έναυσμα της σύσπασης είναι η απελευθέρωση ασβεστίου από το σαρκοπλασματικό δίκτυο), μάλιστα αποτελώντας το 0.5 με 1% των σφαιρικών πρωτεϊνών του μυ, και στο ήπαρ (όπου συμμετέχει στον ομοιοστατικό μηχανισμό ρύθμισης της συγκέντρωσης της γλυκόζης στο αίμα).

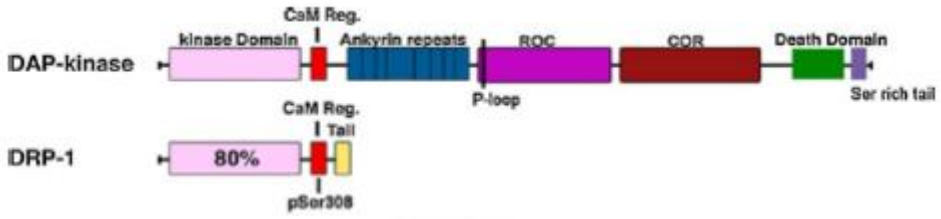
### 1.2.3.8. Death Associated Protein κινάσες (DAPK)

Στις σχετιζόμενες με τον κυτταρικό θάνατο πρωτεϊνικές κινάσες (Death Associated Protein Kinases – DAPKs) εντοπίζονται 2 μέλη των οποίων η ενεργότητα εξαρτάται από την αλληλεπίδρασή τους με την καλμοντουλίνη. Οι DAPK1 και οι DRP-1 (DAPK related protein 1 ή DAPK2). Εμφανίζουν ομοιότητα των καταλυτικών τους κέντρων περίπου στο 80% και 44% ομοιότητα των καταλυτικών τους κέντρων με τις MLCKs που αναλύθηκαν στην προηγούμενη ενότητα.

Η DAPK1 σχετίζεται με τις λειτουργίες του κυτταροσκελετού, έχει μέγεθος 160kDa και μήκος 1430 αμινοξικά κατάλοιπα. Αποτελεί την μεγαλύτερη πρωτεΐνη των εξαρτώμενων από καλμοντουλίνη κινασών. Συμμετέχει στην καταπολέμηση των όγκων και στη ρύθμιση της απόπτωσης και της αυτοφαγίας. Μέχρι σήμερα υπάρχουν κατατεθειμένες στην PDB 51 κρυσταλλογραφικά προσδιορισμένες δομές της DAPK1, καμία όμως δεν περιλαμβάνει την κινάση σε όλο της το μήκος. Πέντε δομές περιλαμβάνουν τόσο το καταλυτικό κέντρο, όσο και τις δομικές αυτοτελείς περιοχές αυτορρύθμισης. Μεταξύ αυτών είναι και η κινάση με PDB:2X0G που θα αναλυθεί εκτενώς στα πλαίσια της παρούσας διπλωματικής.

Η DAPK2 είναι κυτταροπλασματική πρωτεΐνη, φωσφορυλιώνει την ελαφριά αλυσίδα της μυοσίνης (MLC), συμμετέχει στη μετάδοση του σήματος που οδηγεί σε απόπτωση, έχει μήκος 370 αμινοξικά κατάλοιπα και μέγεθος 43kDa. Υπάρχουν 8 κρυσταλλογραφικά προσδιορισμένες δομές (οι δύο περιέχουν την περιοχή αυτορρύθμισης και οι 6 το καταλυτικό κέντρο και την περιοχή αυτορρύθμισης).

Σε όλα τα μέλη της οικογένειας εντοπίζεται το συντηρημένο καταλυτικό κέντρο, καθώς και δομικές αυτοτελείς περιοχές υπεύθυνες για την καταλυτική ενεργότητά τους, την σταθερότητά τους, τον ενδοκυτταρικό τους εντοπισμό και/ή την αλληλεπίδρασή τους με τα υποστρώματά τους.



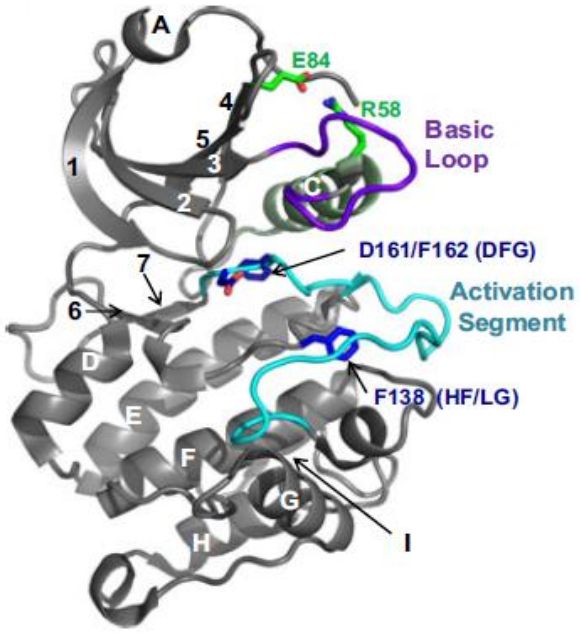

**Εικόνα 1.41** Σχηματική αναπαράσταση της διάταξης των δομικών αυτοτελών περιοχών των DAPK. Με ροζ απεικονίζεται το καταλυτικό κέντρο (το ποσοστό αφορά την ομοιότητα που εμφανίζουν σε σχέση με το καταλυτικό κέντρο της DAPK1). Στην DAPK1 εντοπίζονται με τη σειρά στο καρβοξυτελικό άκρο του καταλυτικού κέντρου οι εξής περιοχές: ρυθμιστική περιοχή πρόσδεσης της $Ca^{2+}$/CaM, δομικά αυτοτελείς περιοχές που σχετίζονται με τον υποκυτταρικό εντοπισμό (Ankyrin επαναλήψεις –υπεύθυνες για την στοίχισή της στα νημάτια ακτίνης- και περιοχή αλληλεπίδρασης με τον κυτταροσκελετό), δομική αυτοτελής περιοχή ROC και δομική αυτοτελής περιοχή του καρβοξυτελικού άκρου της Roc (COR) – υπεύθυνες για την πρόσδεση GTP μέσω μιας θηλιάς πλούσιας σε κατάλοιπα προλίνης στην περιοχή ROC και την υδρόλυσή του GTP (η πρόσδεση του GTP σταθεροποίει το διμερές και προωθεί τη φωσφορυλίωση της Ser308 ελέγχοντας έτσι την ενεργότητα της κινάσης), δομικά αυτοτελής περιοχή κυτταρικού θανάτου (DD) και μια ουρά πλούσια σε κατάλοιπα σερίνης. Στην DAPK2 εντοπίζονται: η ρυθμιστική δομικά αυτοτελής περιοχή πρόσδεσης της $Ca^{2+}$/CaM και μια μοναδική ουρά 40 καταλοίπων στο καρβοξυτελικό άκρο που μεσολαβεί στον ομοδιμερισμό (Shiloh, Bialik, & Kimchi, 2014).

**<u>Καταλυτικό κέντρο</u>** : Η δομή των N- και C-λοβών ακολουθεί το πρότυπο που περιγράφηκε εισαγωγικά. Η περιοχή-αρμός που ενώνει τους δύο λοβούς αποτελείται από 3 αμινοξικά κατάλοιπα (θέσεις 94–96 στην DAPK1) και επιτρέπει την ευέλικτη κίνηση των λοβών. Στην περιοχή πρόσδεσης του ATP εντοπίζεται, στην DAPK1, ένα συντηρημένο κατάλοιπο λυσίνης (Lys42) που συμμετέχει στη μεταφορά της φωσφορικής ομάδας. Για τις DAPK1 και 2 δεν υπάρχουν δεδομένα που να υποστηρίζουν πως απαιτείται φωσφορυλίωσή τους για να ενεργοποιηθούν.

Σε όλα τα μέλη της οικογένειας των DAPK απουσιάζει το μοτίβο HRD που απαιτεί φωσφορυλίωση της θηλιάς ενεργοποίησης για ενεργοποίηση της κινάσης. Στην θέση της αργινίνης στις DAPK εντοπίζεται ένα συντηρημένο κατάλοιπο φαινυλαλανίνης (θέση 138 στην DAPK1) το οποίο προσδένεται στον υδρόφοβο πυρήνα (σχηματίζεται από τα κατάλοιπα Ala136, Ile168, Ile177, Ile188, Leu194 και Ala198 στην DAPK1). Έτσι παραμένουν μόνιμα στη DFG-in στερεοδιάταξη (ενεργή κινάση).

Ένα μοναδικό χαρακτηριστικό των DAPK κινασών αποτελεί η ύπαρξη μιας «βασικής θηλιάς». Αποτελείται από περίπου 12 κατάλοιπα (μοτίβο ακολουθίας : KRXXXXSRRGV, κατάλοιπα 46–56 στην DAPK1), στην πλειοψηφία τους θετικά φορτισμένα. Τα κατάλοιπα αυτά εντοπίζονται εκτεθειμένα και προσανατολισμένα προς το ενεργό κέντρο. Διαδραματίζει σημαντικό ρόλο στον ομοδιμερισμό των DAPK1 και DAPK2 και την επακόλουθη αυτοφωσφορυλίωσή τους, αλλά και στην ρύθμιση της ενεργότητάς τους, σε συνδυασμό με την ρυθμιστική περιοχή. Δε συμμετέχει άμεσα στην πρόσδεση του υποστρώματος.

Το καταλυτικό κέντρο των DAPKs αλληλεπιδρά με την Hsp90 (heat shock protein 90 – μια τσαπερόνη που αλληλεπιδρά με πληθώρα πρωτεϊνών στο εσωτερικό του κυττάρου). Στην αλληλεπίδραση συμμετέχουν συγκεκριμένα ουδέτερα και θετικά φορτισμένα κατάλοιπα της θηλιάς μεταξύ της αC-έλικας και του β4-κλώνου. Η αλληλεπίδραση αυτή είναι κομβικής σημασίας για την σταθερότητα και την ενεργότητα των κινασών. Ανικανότητα αλληλεπίδρασης με την Hsp90 επάγει την αποικοδόμηση των κινασών.




**Εικόνα 1.42** Αναπαράσταση του καταλυτικού κέντρου των DAPK κινασών (Temmerman, Simon, & Wilmanns, 2013).

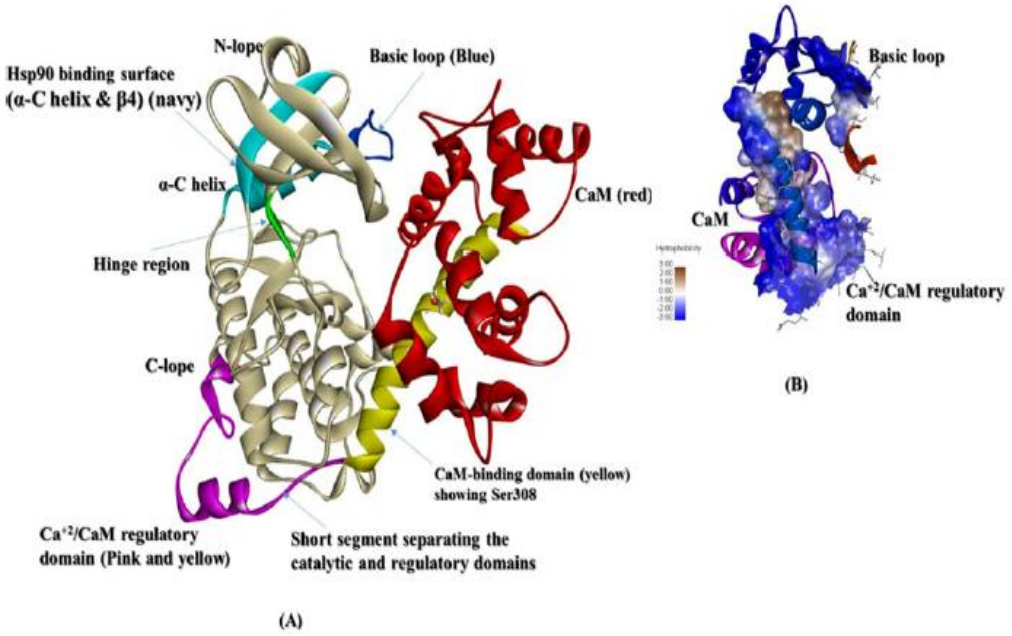
**<u>Ρυθμιστική περιοχή</u>** : Αποτελείται από 42 αμινοξικά κατάλοιπα και εμφανίζεται υψηλά συντηρημένη μεταξύ των DAPK1 και DAPK2. Μεταξύ αυτής και του καταλυτικού κέντρου μεσολαβεί ένα μικρό τμήμα (κατάλοιπα 292–298 στη DAPK1) που επιτρέπει την κίνηση των δύο περιοχών. Πρόσδεση της δομικής αυτοτελούς περιοχής αυτοαναστολής στο καταλυτικό κέντρο της κινάσης έχει ως αποτέλεσμα την παρεμπόδιση της φωσφορυλίωσης του υποστρώματος. Με την

πρόσδεση της καλμοντουλίνης στην αντίστοιχη περιοχή αποσταθεροποιούνται οι αλληλεπιδράσεις μεταξύ του καταλυτικού κέντρου και της περιοχής αυτοαναστολής. Η δεύτερη απομακρύνεται από την καταλυτική εσοχή καθιστώντας έτσι δυνατή την πρόσδεση και επακόλουθη φωσφορυλίωση του υποστρώματος.

Η κρυσταλλογραφικά προσδιορισμένη δομή της DAPK1 που αναφέρθηκε και εισαγωγικά αποκαλύπτει πως η δομική αυτοτελής περιοχή πρόσδεσης της καλμοντουλίνης σχηματίζει μια μακριά α-έλικα επτά στροφών (κατάλοιπα 293–320). Η έλικα αυτή ενώνεται με μια μικρή θηλιά (κατάλοιπα 289–292) η οποία πιθανά εξυπηρετεί στην κίνηση της έλικας. Το κατάλοιπο Ser289 στο εσωτερικό αυτής της θηλιάς αποτελεί στόχο για φωσφορυλίωση από την RSK.

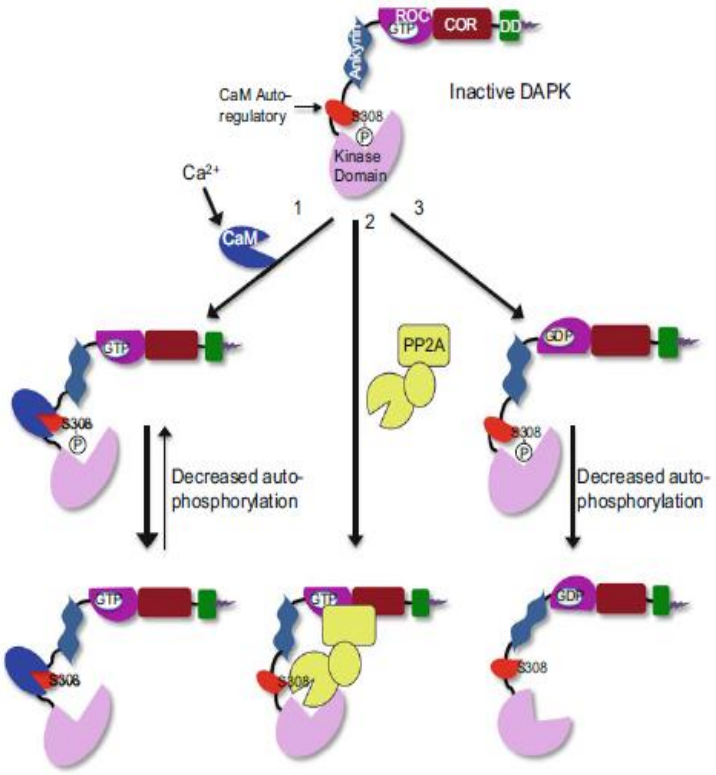
Φωσφορυλίωση του καταλοίπου Ser308 στην DAPK1 και Ser318 στην DAPK2 που εντοπίζεται στο εσωτερικό της δομικής αυτοτελούς περιοχής πρόσδεσης της $Ca^{2+}$/CaM έχει ως αποτέλεσμα την αναστολή της ενεργότητας της κινάσης. Ο προσανατολισμός της φωσφορυλιωμένης σερίνης είναι υπεύθυνος για τη δημιουργία περισσότερων περιοχών αλληλεπίδρασης μεταξύ της περιοχής πρόσδεσης της $Ca^{2+}$/CaM και του καταλυτικού κέντρου, σταθεροποιώντας έτσι την αγκυροβόληση της περιοχής αυτοαναστολής στην περιοχή πρόσδεσης του υποστρώματος. Η φωσφορυλιωμένη DAPK εμφανίζει χαμηλότερη συγγένεια πρόσδεσης με την $Ca^{2+}$/CaM ενισχύοντας περαιτέρω την ανατολή της ενεργότητά της. Η αντίδραση φωσφορυλίωσης συμβαίνει σε *cis* κατάσταση, και όχι *trans*-φωσφορυλίωση μεταξύ των πρωτεϊνών του ομοδιμερούς.

Αξίζει να αναφερθεί πως ο ομοδιμερισμός που μεσολαβείται από το καταλυτικό κέντρο έχει ως αποτέλεσμα τη δημιουργία μιας μικρής επιφάνειας πρόσδεσης στην $Ca^{2+}$/CaM (με χαμηλή συγγένεια πρόσδεσης). Στην αλληλεπίδραση αυτή συμμετέχει η βασική θηλιά και είναι εφικτή μόνο μετά τον ομοδιμερισμό της κινάσης.



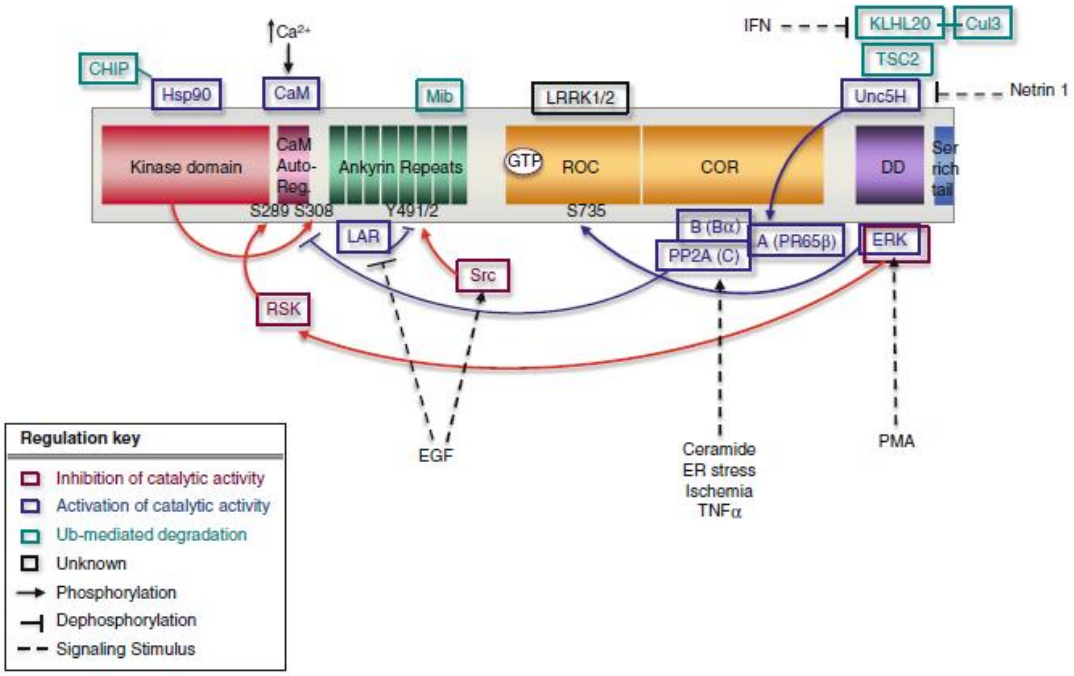

**Εικόνα 1.43** Διαγραμματική αναπαράσταση του καταλυτικού κέντρου και των δομικών αυτοτελών περιοχών της DAPK1 σε σύμπλοκο με την καλμοντουλίνη (PDB : 2X0G).Η δομή παρουσιάζει μεγάλο ενδιαφέρον καθώς βοήθησε στην κατανόηση των μοριακών μηχανισμών που διέπουν την αλληλεπίδραση.(A) Η βασική θηλιά απεικονίζεται με μπλε, η περιοχή αρμός με πράσινο, η ρυθμιστική περιοχή αποτελείται από την δομική αυτοτελή περιοχή αυτοαναστολής (ροζ) και την δομική αυτοτελή περιοχή πρόσδεσης της CaM (κίτρινο), η επιφάνεια αλληλεπίδρασης με την Hsp90 (που σχηματίζεται από την αC-β4 θηλιά) απεικονίζεται με γαλάζιο. Μεταξύ του καταλυτικού κέντρου (κατάλοιπα : 13–267) και της περιοχής αυτορρύθμισης μεσολαβεί ένα μικρό τμήμα (κατάλοιπα : 267–275) που επιτρέπει την κίνηση των 2 περιοχών. (B) Η μοριακή επιφάνεια μεταξύ της περιοχής αυτορρύθμισης και της καλμοντουλίνης. Απεικονίζεται η συμμετοχή της βασικής θηλιάς στον σχηματισμό του συμπλόκου (Farag & Roh, 2018).

Η φωσφατάση PP2A ενεργοποιεί την DAPK1 μέσω αποφωσφορυλίωσης της Ser308. Αξίζει να σημειωθεί πως η φωσφορυλιωμένη έλικα αλληλεπιδρά με την καλμοντουλίνη, αν βρίσκονται προσδεμένα τουλάχιστον τρία ιόντα ασβεστίου, ενώ η αποφωσφορυλιωμένη έλικα αλληλεπιδρά με την καλμοντουλίνη ακόμα και με ένα ιόν ασβεστίου προσδεμένο σε αυτήν. Αυτό αποκαλύπτει μια ενδιαφέρουσα σχέση μεταξύ της μεσολαβούμενης από ασβέστιο σηματοδότησης και της αποφωσφορυλίωσης των καταλοίπων σερίνης, όσον αφορά την ρύθμιση της ενεργότητας των κινασών. Ενεργοποίηση μιας φωσφατάσης και επακόλουθη αποφωσφορυλίωση του καταλοίπου σερίνης επάγει υψηλότερη ευαισθησία των κινασών ως προς το ασβέστιο και έτσι οι κινάσες ενεργοποιούνται σε χαμηλές συγκεντρώσεις ασβεστίου. Αντιθέτως, ελλείψει ενεργότητας φωσφατάσης οδηγεί στην ενεργοποίηση των κινασών μόνο σε υψηλά τοπικά ενδοκυτταρικά επίπεδα ασβεστίου. Με αυτόν τον τρόπο είναι πιθανή η ενεργοποίηση των κινασών από διαφορετικά μονοπάτια μεταγωγής σήματος. Τέλος, πρόσδεση της καλμοντουλίνης μειώνει τη φωσφορυλίωση του καταλοίπου σερίνης, όπως αποκαλύπτεται από τη δομή της DAPK1 με την CaM, καθώς η σερίνη βρίσκεται ενωμένη και με τους δύο λοβούς της καλμοντουλίνης.

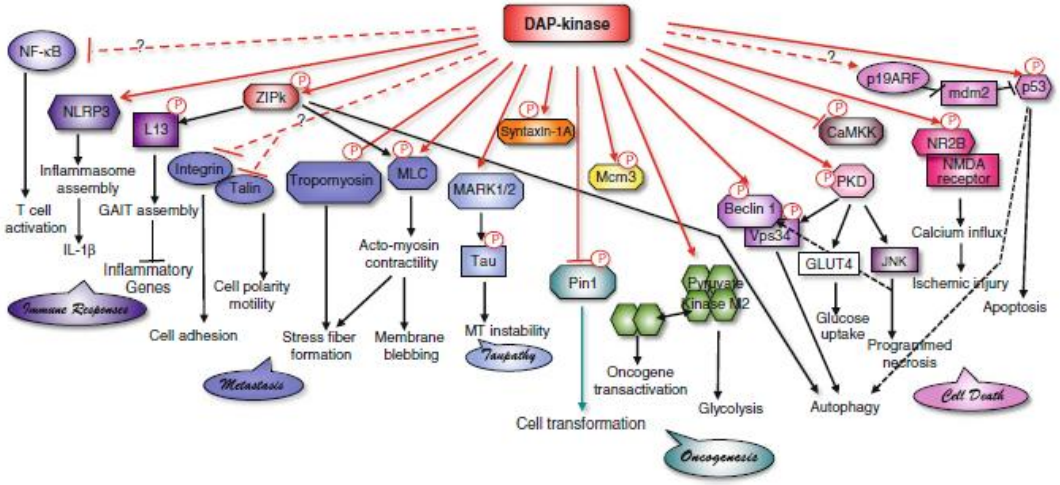



**Εικόνα 1.44** Μηχανισμοί ρύθμισής της ενεργότητας της DAPK1. Όσον αφορά την ρύθμιση μέσω αλληλεπίδρασης με την καλμοντουλίνη αναλύεται εκτενώς στο κείμενο. Πρόσδεση του GTP στην ROC δομική αυτοτελή περιοχή ενισχύει την αυτό-φωσφορυλίωση της Ser308. Υδρόλυση του GTP σε GDP από την ROC δομική αυτοτελή περιοχή έχει ως αποτέλεσμα στερεοδιατακτικές αλλαγές στο αμινοτελικό άκρο που, τελικά, μειώνουν την αυτοφωσφορυλίωση της Ser308. Η ακολουθία των γεγονότων δεν είναι γνωστή. (Bialik & Kimchi, 2014).

Συμμετέχουν σε πληθώρα βιολογικών διαδικασιών, όπως, η απόπτωση (υπερέκφρασή τους στα κύτταρα έχει ως αποτέλεσμα την εκτεταμένη δημιουργία προεξοχών της μεμβράνης), η αυτοφαγία, η πλαστικότητα των συνάψεων και η αντιϊική ανοσολογική απόκριση. Έχουν συνδεθεί με διάφορες ασθένειες, όπως, καρκίνος (καταστέλλουν την ανάπτυξη των όγκων και ρυθμίζουν το οξειδωτικό στρες), νόσος Alzheimer's και αιμορραγικά και ισχαιμικά εγκεφαλικά επεισόδια.




**Εικόνα 1.45** Παρουσιάζονται οι αλληλεπιδράσεις της DAPK1 αναρροϊκά.




**Εικόνα 1.46** Παρουσιάζονται οι αλληλεπιδράσεις της DAPK1 καταρροϊκά.

## 1.3. Αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών

### 1.3.1. Εισαγωγή

Με τον όρο αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών (protein – protein interactions , PPIs) αναφερόμαστε στις αλληλεπιδράσεις δύο ή περισσότερων πρωτεϊνικών μορίων , όμοιων ή διαφορετικών, προκειμένου να επιτελεστεί μια συγκεκριμένη βιολογική λειτουργία ή να σχηματιστεί κάποιος δομικός σχηματισμός με συγκεκριμένα χαρακτηριστικά. Ο πιο περιεκτικός και σύντομος ορισμός για να κατανοήσει κανείς την φύση της ισχυρής αλληλεπίδρασης δυο πρωτεϊνών περιέχεται στη φράση «φυσικές επαφές μεταξύ πρωτεϊνικών μορίων». Η επαφή συνίσταται στο ότι τα μόρια αυτά θα έρθουν τόσο κοντά ώστε να αναπτυχθούν αλληλεπιδράσεις ανάμεσα στα άτομα των μορίων (δεσμοί υδρογόνου, ηλεκτροστατικές αλληλεπιδράσεις, αλληλεπιδράσεις van der Waals, υδρόφοβες αλληλεπιδράσεις - λαμβάνοντας πάντα υπόψιν και τις απωστικές ηλεκτροστατικές δυνάμεις ή τις αλληλεπιδράσεις van der Waals που καθορίζουν την ελάχιστη απόσταση στην οποία μπορούν να βρεθούν τα αλληλεπιδρώντα μόρια). Πέραν της «απόστασης», ένας άλλος παράγοντας που καθορίζει μια αλληλεπίδρασης μεταξύ πρωτεϊνικών μορίων είναι ο χρόνος αλληλεπίδρασης. Συχνά υπάρχει η παρανόηση ότι αλληλεπίδραση σημαίνει «σύνδεση» δύο ή περισσότερων πρωτεϊνικών μορίων για μεγάλο χρονικό διάστημα, μη λαμβάνοντας υπόψιν πρωτεϊνικές αλληλεπιδράσεις στις οποίες τα επιμέρους μόρια αλληλεπιδρούν παροδικά, όπως συμβαίνει σε αρκετές περιπτώσεις αλληλεπιδράσεων μεταξύ κινασών και πρωτεϊνών-υποστρωμάτων ή άλλων πρωτεϊνών –όπως στην περίπτωση της καλμοντουλίνης.

Η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών αποτελεί μια από τις μεγαλύτερες προκλήσεις της μετα-γονιδιωματικής εποχής, καθώς διαδραματίζουν σημαντικό ρόλο σε πληθώρα σημαντικών βιολογικών διαδικασιών, όπως είναι η μεταγωγή σήματος, η κυτταρική επικοινωνία και διαφοροποίηση, οι καταρράκτες μεταβολικών αντιδράσεων κ.α. (Titz, Schlesner, & Uetz, 2004). Η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών μπορεί να συμβάλλει στη διερεύνηση της λειτουργίας των πρωτεϊνών, στην ανίχνευση νέων – μη καταγεγραμμένων – λειτουργιών πρωτεϊνών αλλά και στον προσδιορισμό της λειτουργίας «άγνωστων» πρωτεϊνών (Letovsky & Kasif, 2003)(Letovsky & Kasif, 2003), ενώ ένας άλλος τομέας στον οποίο πρωταγωνιστεί είναι η εξερεύνηση βιολογικών μονοπατιών που συνδέονται με διάφορες ασθένειες(Taylor et al., 2009).

Η μελέτη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών είναι ένας τομέας όπου η βιοπληροφορική και η μοριακή βιολογία «συνεργάζονται» στη προσπάθεια προσδιορισμού των φυσικών αλληλεπιδράσεων μεταξύ πρωτεϊνικών μορίων.

## 1.3.2. Πειραματικές μέθοδοι προσδιορισμού αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών

Οι πειραματικές μέθοδοι για τον προσδιορισμό των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών μπορούν να χωριστούν σε δύο βασικές κατηγορίες : α) μέθοδοι υψηλής απόδοσης *(high throughput)* και χαμηλής απόδοσης *(low throughput).* Για τη μελέτη του συνόλου των πρωτεϊνικών αλληλεπιδράσεων σε έναν οργανισμό, εφαρμόζονται «μεγάλης κλίμακας» πειραματικές μεθοδολογίες *(large-scale experiments)* (Rual et al., 2005; Walhout & Vidal, 2001). Οι χιλιάδες αλληλεπιδράσεις μεταξύ πρωτεϊνικών μορίων που προκύπτουν επιτρέπουν τη δημιουργία ενός «σκελετού» του δικτύου των πρωτεϊνικών αλληλεπιδράσεων που λαμβάνουν χώρα στο υπό μελέτη σύστημα. Είναι ευνόητο πως τα δεδομένα από μεγάλης κλίμακας πειράματα εμπεριέχουν μεγάλο αριθμό ψευδώς θετικών αλληλεπιδράσεων (false positives), δηλαδή αλληλεπιδράσεων που είτε είναι αναληθείς, είτε δεν συμβαίνουν μέσα στο κύτταρο (Bork et al., 2004). Για τη μελέτη μιας πρωτεΐνης ή ενός περιορισμένου αριθμού πρωτεϊνών συγχρόνως και των αλληλεπιδράσεων τους εφαρμόζονται μέθοδοι χαμηλής απόδοσης οι οποίες παρουσιάζουν πολύ χαμηλότερα ποσοστά σφάλματος σε σχέση με της μεθόδους υψηλής απόδοσης, αλλά έχουν μεγάλο κόστος και είναι αρκετά χρονοβόρες.

### 1.3.2.1. Μικρής κλίμακας

Οι πειραματικές μέθοδοι διακρίνονται στις *in vivo* που πλεονεκτούν καθώς εφαρμόζονται στο φυσικό κυτταρικό περιβάλλον και στις *in vitro* που παρουσιάζονται περιληπτικά στον πίνακα 1.6. Τέλος υπάρχουν και οι τεχνικές του Πυρηνικού Μαγνητικού Συντονισμού (NMR) και της κρυσταλλογραφίας ακτίνων Χ που ειδικεύονται στην ανάλυση των αλληλεπιδράσεων μεταξύ πρωτεϊνικών μορίων σε ατομικό επίπεδο.

**Πίνακας 1.6** In vitro μέθοδοι για τον προσδιορισμό αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών

| In vitro μέθοδος | Σύντομη περιγραφή |
|---|---|
| **Surface Plasmon Resonance (SPR)** | Ανίχνευση αλληλεπιδράσεων μεταξύ διαλυτών υποκαταστατών και ακινητοποιημένων υποδοχέων. Η πρόσδεση του υποκαταστάστη στον υποδοχέα προκαλεί αλλαγές στην ανάκλαση του φωτός της επιφάνειας όπου είναι ακινητοποιημένος ο υποδοχέας. |
| **Συν-ανοσοκατακρήμνιση (Co- immunoprecipitation)** | Προσδιορισμός άγνωστων πρωτεϊνικών αλληλεπιδράσεων για ενδογενείς και όχι υπερεκφρασμένες ή σημασμένες πρωτεΐνες. Απομόνωση πρωτεϊνικών συμπλόκων και περαιτέρω ταυτοποίηση αλληλεπιδρώντων πρωτεϊνών με Western blotting. |
| **Fluorescence Resonance Energy Transfer (FRET)** | Φασματοσκοπική τεχνική για ανίχνευση της αλληλεπίδρασης σε πραγματικό χρόνο που βασίζεται στην εκπομπή φθορισμού που επάγεται από την αλληλεπίδραση των πρωτεϊνών. |

| | |
|---|---|
| **Label Transfer** | Σήμανση γνωστής πρωτεΐνης με «ετικέτα» και κατόπιν αναζήτηση της «ετικέτας» στα υποψήφια για αλληλεπίδραση μόρια. |
| **Far Western** | Ταυτοποίηση αλληλεπιδρώντων πρωτεϊνών με τεχνικές ανοσοαποτύπωσης |
| **Isothermal Titration Calorimetry (ITC)** | Ταυτοποίηση αλληλεπιδρώντων πρωτεϊνών μέσω άμεσης μέτρησης της ενθαλπίας σύνδεσης. |

#### 1.3.2.2. Μεγάλης κλίμακας

Οι πιο συχνά χρησιμοποιούμενες μέθοδοι μεγάλης κλίμακας για τον προσδιορισμό αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών είναι:

Μέθοδος διυβριδισμού σε σύστημα σακχαρομύκητα: εκμεταλλεύεται τη δομή αλλά και τη λειτουργία των μεταγραφικών παραγόντων(Ito, Chiba, & Yoshida, 2001; Walhout & Vidal, 2001) Οι μεταγραφικοί παράγοντες διαθέτουν μια δομική αυτοτελή περιοχή που προσδένεται στο DNA (Binding Domain, BD) και μια άλλη δομικά αυτοτελή περιοχή (Activation Domain, AD) που αλληλεπιδρά με άλλα μόρια της μεταγραφικής μηχανής και είναι υπεύθυνη για την ενεργοποίηση τους μεταγραφικού παράγοντα. Για την ενεργοποίηση της μεταγραφής απαιτείται η παρουσία και των δυο. Για να εξεταστεί μια πιθανή αλληλεπίδραση μεταξύ δυο πρωτεϊνών Χ, Υ εκφράζεται η μία (δόλωμα-bait) σε ένα στέλεχος σακχαρομύκητα, συντηγμένη με το Binding Domain ενός μεταγραφικού παράγοντα και η άλλη (θήραμα-pray) συντηγμένη με το Activation Domain του ίδιου μεταγραφικού παράγοντα σε ένα άλλο διακριτό στέλεχος σακχαρομύκητα. Έπειτα, οι δυο συντηγμένες πρωτεΐνες συνεκφράζονται σε ένα στέλεχος σακχαρομύκητα, που προκύπτει από την σύντηξη των δυο μεμονωμένων στελεχών, και αν υπάρχει αλληλεπίδραση μεταξύ των πρωτεϊνών Χ και Υ, δημιουργείται ένας λειτουργικός μεταγραφικός παράγοντας, ο οποίος επάγει την έκφραση ενός γονιδίου αναφοράς. Κατά την εφαρμογή της συγκεκριμένης μεθόδου σε πειράματα μεγάλης κλίμακας, εφαρμόζονται δυο κύριοι τύποι προσεγγίσεων(Rajagopala & Uetz, 2009; Rual et al., 2005; Stelzl et al., 2005; Uetz et al., 2000):

**Προσέγγιση βασιζόμενη σε μήτρα (matrix-based approach) :** Κατά την προσέγγιση αυτή εισάγεται στο σακχαρομύκητα ένα σύνολο από «πρωτεΐνες – θηράματα» με γνωστή αμινοξική αλληλουχία. Στέλεχος σακχαρομύκητα το οποίο φέρει την «πρωτεΐνη – δόλωμα» συζευγνύεται με κάθε ένα από τα μετασχηματισμένα στελέχη που φέρουν την «πρωτεΐνη – θήραμα». Εάν το διπλοειδές στέλεχος αναπτυχθεί στο θρεπτικό, οι πρωτεΐνες «θήραμα» και «δόλωμα» αλληλεπιδρούν και η θέση στην οποία αναπτύσσεται κάθε στέλεχος αντικατοπτρίζει και την «πρωτεΐνη – θήραμα» την οποία φέρει.

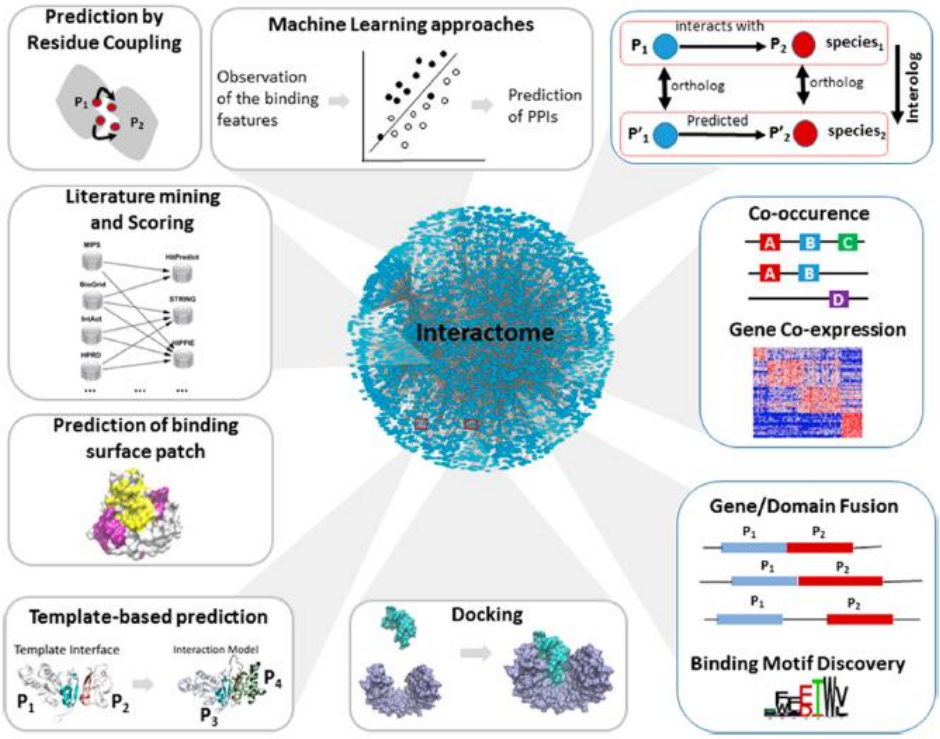
**Προσέγγιση βασιζόμενη σε βιβλιοθήκη (library based approach) :** Κατά την προσέγγιση αυτή μια «πρωτεΐνη – δόλωμα» εκφράζεται σε ένα στέλεχος μύκητα (στέλεχος αναφοράς) ενώ μια συλλογή από «πρωτεΐνες - θηράματα» (η βιβλιοθήκη) μετασχηματίζεται σε ένα στέλεχος μύκητα αντίθετου συζευκτικού τύπου. Τα στελέχη συζευγνύονται και οι κλώνοι οι οποίοι εκφράζουν το αλληλεπιδρόν ζεύγος, και κατ'επέκταση εμφανίζουν ενεργοποίηση των γονιδίων αναφοράς, απομονώνονται σε επιλεκτικό μέσο.

Φασματομετρία μάζας «καθαρισμένων» πρωτεϊνικών συμπλόκων: Ανεξάρτητες πρωτεΐνες επισημαίνονται και χρησιμοποιούνται ως «δολώματα» για τον βιοχημικό καθαρισμό ολόκληρων πρωτεϊνικών συμπλόκων. Στη συνέχεια οι πρωτεΐνες του συμπλόκου διαχωρίζονται και αναγνωρίζονται από φασματομετρία μάζας. Δυο πρωτόκολλα υπάρχουν: διαδοχικός συγγενικός καθαρισμός (tandem affinity purification-TAP), και η αναγνώριση πρωτεϊνικών συμπλόκων με φασματομετρία μάζας (mass-spectrometric protein complex identification- HMSPCI).

Μικροσυστοιχίες πρωτεϊνών: Στοχεύουν στην εύρεση των επιπέδων έκφρασης μιας πρωτεΐνης σε ένα βιολογικό δείγμα ή στην εύρεση των αλληλεπιδράσεων που μπορεί να έχει με άλλα βιομόρια, όπως άλλες πρωτεΐνες, αντισώματα ή φάρμακα(Conrads, Zhou, Petricoin, Liotta, & Veenstra, 2003).

### 1.3.3. Υπολογιστικές μέθοδοι για την μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών

Λόγω μιας σειράς εγγενών τεχνικών περιορισμών οι παραπάνω πειραματικές μέθοδοι παράγουν ένα μεγάλο ποσοστό ψευδών αλληλεπιδράσεων το οποίο εκτιμάται ότι ανέρχεται σε 70-90%(Qi, Bar-Joseph, & Klein-Seetharaman, 2006; Rual et al., 2005). Η μικρή επικάλυψη αλλά και η χαμηλή ποιότητα των παραγόμενων από πειραματικές μεθόδους δεδομένων έχει οδηγήσει στην ανάπτυξη μιας σειράς υπολογιστικών μεθόδων για την αξιολόγηση των καταγεγραμμένων αλλά και την πρόβλεψη νέων πρωτεϊνικών αλληλεπιδράσεων (Chen, Chua, et al., 2006; Chen, Hsu, Lee, & Ng, 2006; Skrabanek, Saini, Bader, & Enright, 2008). Σε γενικές γραμμές οι μέθοδοι που βασίζονται στη μήτρα εμφανίζουν καλύτερη επαναληψιμότητα ενώ οι μέθοδοι που βασίζονται στα ίδια τα σύμπλοκα μεγαλύτερη ακρίβεια (προβλέπουν λιγότερα ορθά αλληλεπιδρώντα κατάλοιπα με μεγαλύτερη σιγουριά).




**Εικόνα 1.47** Εναλλακτική κατηγοριοποίηση των υπολογιστικών μεθόδων πρόβλεψης PPIs. (i) Πρόβλεψη κατά ζεύγη: μέθοδοι μηχανικής μάθησης, μέθοδοι που βασίζονται σε βιβλιογραφικά δεδομένα, μέθοδοι που βασίζονται στις αλληλεπιδράσεις ορθόλογων πρωτεϊνών (interolog search), μέθοδοι που βασίζονται στη σύντηξη γονιδίων ή δομικών αυτοτελών περιοχών, στην συνέκφραση ή την συνύπαρξη των γονιδίων. (ii) Πρόβλεψη της περιοχής αλληλεπίδρασης: μέθοδοι αναζήτησης μοτίβων, μέθοδοι πρόβλεψης της περιοχής αλληλεπίδρασης (iii) Πρόβλεψη της αλληλεπίδρασης προς σχηματισμό συμπλόκων: μέθοδοι αγκυροβόλησης (docking), μέθοδοι βασιζόμενες στη μήτρα (template-based) (Keskin, Tuncbag, & Gursoy, 2016).

Οι μέθοδοι κατηγοριοποιούνται σε διαφορετικές ομάδες με βάση διαφορετικά κριτήρια. Στην παρούσα περιγραφή θα ακολουθηθεί η παρακάτω κατηγοριοποίηση (Rao, Srinivas, Sujini, & Kumar, 2014):

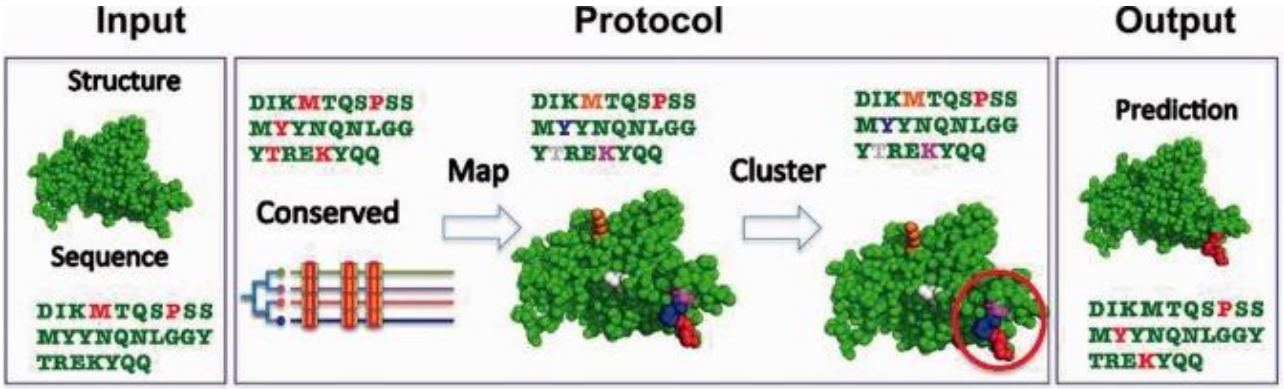
1. Προσεγγίσεις που βασίζονται στη δομή των πρωτεϊνών
2. Προσεγγίσεις που βασίζονται στην αμινοξική ακολουθία των πρωτεϊνών
    a. Προσεγγίσεις που βασίζονται στις αλληλεπιδράσεις ορθόλογων πρωτεϊνών
    b. Προσεγγίσεις που βασίζονται στις αλληλεπιδράσεις μεταξύ δομικά αυτοτελών περιοχών
3. Προσεγγίσεις που βασίζονται σε πληροφορία που προκύπτει από μελέτη των γονιδίων
    a. Προσεγγίσεις που βασίζονται σε χαρακτηριστικά γονιδιακής κλίμακας
    b. Προσεγγίσεις που βασίζονται στο φαινόμενο της γονιδιακής σύντηξης

c. Προσεγγίσεις που βασίζονται σε δεδομένα γονιδιακής έκφρασης

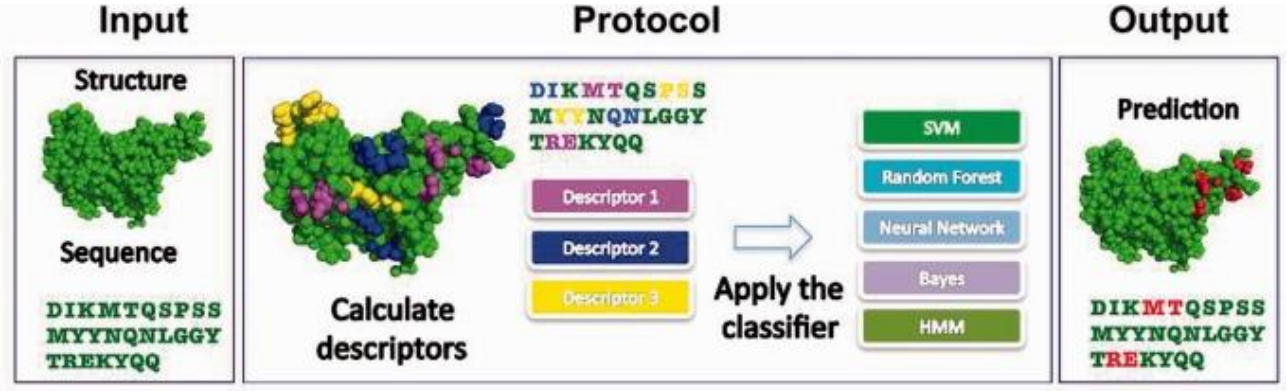
4. Προσεγγίσεις που βασίζονται σε εξελικτική πληροφορία
   a. Προσεγγίσεις που βασίζονται στα φυλογενετικά δένδρα
   b. Προσεγγίσεις που βασίζονται στα φυλογενετικά προφίλ
5. Προσεγγίσεις που βασίζονται στην τοπολογία των πρωτεϊνικών δικτύων
6. Προσεγγίσεις που βασίζονται στην αμινοξική ακολουθία των πρωτεϊνών αξιοποιώντας την εξελικτική πληροφορία.

### 1.3.3.1. Μελέτες σε επίπεδο δομής πρωτεϊνών

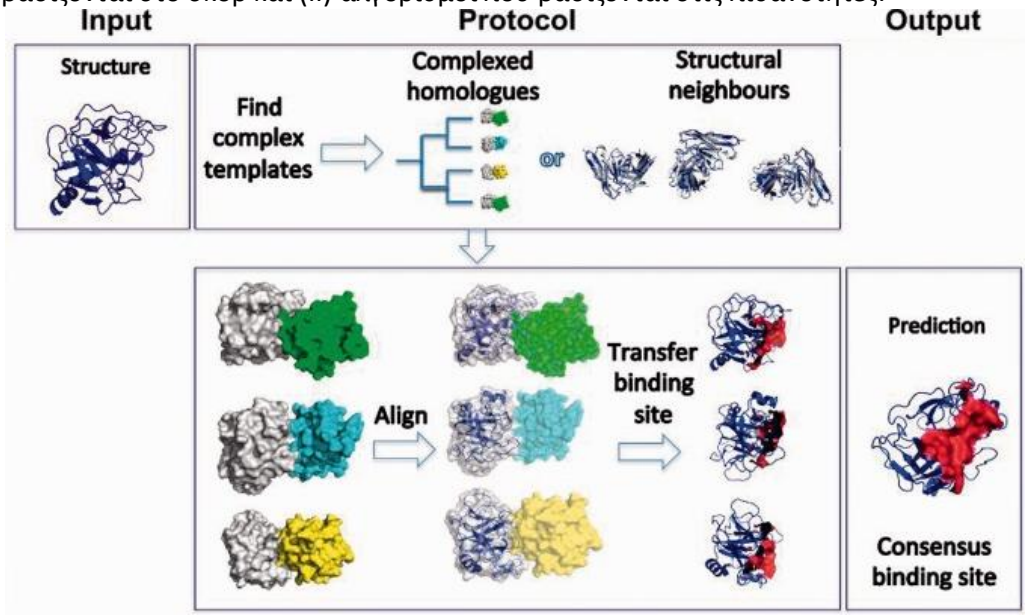
Η βασική ιδέα των προσεγγίσεων που βασίζονται στη τρισδιάστατη δομή των πρωτεϊνικών μορίων, είναι η πρόβλεψη της αλληλεπίδρασης μεταξύ δυο άγνωστων πρωτεϊνών αν είναι γνωστό ότι αλληλεπιδρούν δυο άλλες γνωστές πρωτεΐνες με παρόμοια δομή. Αναλυτικότερα, αν δυο πρωτεΐνες Α και Β αλληλεπιδρούν, και υπάρχουν δυο άλλες πρωτεΐνες Α' και Β' των οποίων οι δομές είναι παρόμοιες με εκείνες των πρωτεϊνών Α και Β, μπορεί να εξαχθεί το συμπέρασμα ότι και η Α', Β' αλληλεπιδρούν. Στις περιπτώσεις που δεν υπάρχει διαθέσιμη κρυσταλλογραφικά προσδιορισμένη δομή, το πρώτο βήμα στις προσεγγίσεις αυτές είναι η πρόγνωση της δομή με βάση την αμινοξική ακολουθία των πρωτεϊνών. Οι διάφορες μέθοδοι παρουσιάζονται στις παρακάτω εικόνες (Esmaielbeiki, Krawczyk, Knapp, Nebel, & Deane, 2016)




**Εικόνα 1.48** Μέθοδοι βασιζόμενοι στην χαρτογράφηση καταλοίπων σε κρυσταλλογραφικά προσδιορισμένες δομές. Εντοπίζονται αρχικά τα συντηρημένα κατάλοιπα στις δύο ομάδες πρωτεϊνών που θεωρητικά αλληλεπιδρούν. Χαρτογραφούνται στις διαθέσιμες δομές και κατηγοριοποιούνται σε ομάδες γειτονικών καταλοίπων. Η ακρίβεια της πρόβλεψης βελτιστοποιείται με την προσθήκη φυσικοχημικών παραμέτρων κατά την στοίχιση.



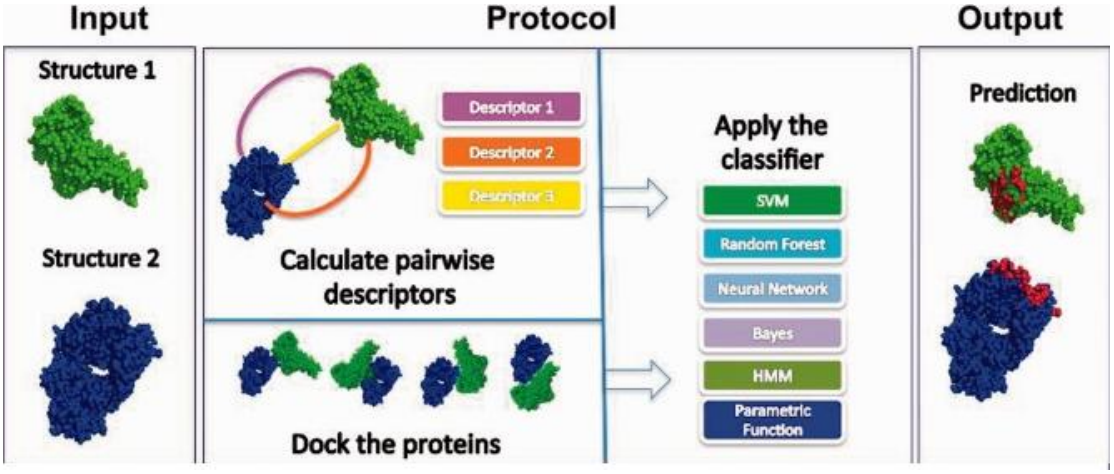

**Εικόνα 1.49** Μέθοδοι που βασίζονται στα χαρακτηριστικά των τρισδιάστατων δομών για τον εντοπισμό επιφανειών διεπαφής. Εκμεταλλεύονται το γεγονός πως στην επιφάνεια διεπαφής εντοπίζονται διαφορετικά δομικά χαρακτηριστικά σε σχέση με την υπόλοιπη πρωτεΐνη (υδρόφοβες επιφάνειες κ.α.). Οι μέθοδοι πρόβλεψης δέχονται ως είσοδο και συνδυάζουν πολλαπλές ιδιότητες των καταλοίπων, και διαφέρουν σημαντικά μεταξύ τους ως προς αυτό. Θα μπορούσαν να κατηγοριοποιηθούν σε δύο ομάδες : (i) αλγόριθμοι που βασίζονται στο σκορ και (ii) αλγόριθμοι που βασίζονται στις πιθανότητες.




**Εικόνα 1.50** Μέθοδοι βασιζόμενες σε μήτρα, ομόλογα σύμπλοκα. Οι επιφάνειες διεπαφής μεταξύ ομόλογων συμπλόκων είναι συντηρημένες. Όμως τέτοιες ομόλογες δομές δεν είναι διαθέσιμες. Έτσι η παρούσα μέθοδος θεωρεί μήτρα δομικά, αλλά όχι απαραίτητα λόγω κοινής εξελικτικής πορείας, παρόμοια σύμπλοκα. Χρησιμοποιεί γνωστά σύμπλοκα που μόνο η μια από τις αλληλεπιδρούσες πρωτεΐνες είναι ομόλογη της προς αναζήτηση πρωτεΐνης. Η επιφάνεια διεπαφής θεωρείται πως θα εντοπίζεται και στην προς αναζήτηση πρωτεΐνη. Αυτό είναι εφικτό στην περίπτωση που ομόλογες πρωτεΐνες αλληλεπιδρούν με το άλλο μέλος του συμπλόκου με παρόμοιο προσανατολισμό και η επιφάνεια διεπαφής εντοπίζεται συντηρημένη ανεξάρτητα από την ομοιότητα των πρωτεϊνών μεταξύ τους. Οι φυσικοχημικές ιδιότητες των καταλοίπων που συμμετέχουν στην επιφάνεια διεπαφής εμφανίζουν υψηλότερη ομοιότητα μεταξύ ομόλογων πρωτεϊνών, από ότι μεταξύ μη ομόλογων. Πρωτεΐνες που διαθέτουν παρόμοια στερεοδιάταξη με την προς αναζήτηση πρωτεΐνη, ακόμα και αν δεν σχετίζονται εξελικτικά, μπορούν να συνεισφέρουν στην πρόβλεψη. Μπορεί να εντοπιστεί λειτουργική σχέση με τη βοήθεια μακρινών δομικών γειτόνων. Πρωτεΐνες με παρόμοιες στερεοδιατάξεις αλλά χαμηλή ομοιότητα ακολουθιών τείνουν να αλληλεπιδρούν με το άλλο μέλος του συμπλόκου μέσω της ίδιας περιοχής. Αυτοί

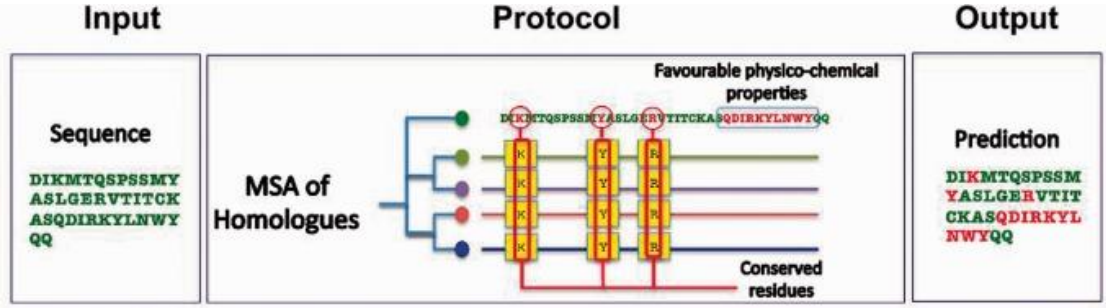
οι δομικοί γείτονες εξερευνούνται ως μήτρες για την πρόβλεψη της επιφάνειας διεπαφής. Το σύνολο των δομών που προέκυψαν, είτε με βάση ομόλογα σύμπλοκα, είτε με βάση δομικούς γείτονες, στοιχίζονται δομικά με την προς αναζήτηση πρωτεΐνη. Τα κατάλοιπα που στοιχίζονται με μεγαλύτερη συχνότητα θεωρούνται ως αλληλεπιδρώντα κατάλοιπα.

Οι μέθοδοι που περιγράφηκαν παραπάνω προβλέπουν τις επιφάνειες διεπαφής στην προς αναζήτηση πρωτεΐνη. Οι περισσότερες πρωτεΐνες όμως εμφανίζουν διαφορετικά πρότυπα αλληλεπίδρασης αναλόγως με την πρωτεΐνη που αλληλεπιδρούν. Αυτό καθιστά αναγκαία την ύπαρξη μεθόδων πρόβλεψης της επιφάνειας διεπαφής εξειδικευμένων ως προς την άλλη πρωτεΐνη-μέλος του συμπλόκου. Οι μέθοδοι αυτοί εντοπίζουν αλληλεπιδρώντα κατάλοιπα μεταξύ των δύο προς αναζήτηση πρωτεϊνών που έχει υποτεθεί πως αλληλεπιδρούν. Η απόδοση των μεθόδων αυτών μειώνεται με την αύξηση των στερεοδιατακτικών αλλαγών των αλληλεπιδρώντων πρωτεϊνών.




**Εικόνα 1.51** Μέθοδος βασισμένη στην αγκυροβόληση πρωτεΐνης- πρωτεΐνης. Οι μέθοδοι αυτοί διερευνούν επιφάνειες διεπαφής με βάση ενεργειακούς υπολογισμούς και σκορ συχνότητας εμφάνισης. Μεταξύ των μειονεκτημάτων εντοπίζονται: η απαίτηση δύο πρωτεϊνικών δομών και η ταχύτητά τους.

### 1.3.3.2. Μελέτες σε επίπεδο αμινοξικής ακολουθίας πρωτεϊνών




**Εικόνα 1.52** Μέθοδοι που λαμβάνουν ως είσοδο αμινοξικές ακολουθίες. Έπειτα επεξεργάζονται τις ακολουθίες με βάση χαρακτηριστικά που θα μπορούσαν να συνεισφέρουν στο να αποτελεί ένα κατάλοιπο μέρος μιας επιφάνειας διεπαφής, όπως η συντήρηση ή οι φυσικοχημικές ιδιότητες.

Α) Προσεγγίσεις που βασίζονται στις αλληλεπιδράσεις ορθόλογων πρωτεϊνών

Μεταφέρεται ο σχολιασμός μιας πρωτεϊνικής ακολουθίας με καθορισμένη λειτουργία σε μια ακολουθία στόχο με βάση την επί της εκατό ομοιότητα των αμινοξικών τους ακολουθιών. Ο σχολιασμός με βάση την ομοιότητα βασίζεται στην εύρεση ομολόγων της προς αναζήτηση πρωτεΐνης σε βάσεις δεδομένων καλά σχολιασμένων ακολουθιών μέσω τοπικής στοίχισης των ακολουθιών (S.-A. Lee et al., 2008).

Με τη μεθοδολογία αυτή θα εντοπιστούν αρκετές πρωτεΐνες του υπό μελέτη οργανισμού οι οποίες θα μοιράζονται σημαντικά επίπεδα ομοιότητας με πρωτεΐνες που εμπλέκονται στο σχηματισμό συμπλόκων σε άλλους οργανισμούς. Όταν τελικά σχολιαστεί η πιθανή λειτουργία της προς αναζήτησης πρωτεΐνης πρέπει να γίνει διάκριση μεταξύ ορθόλογων και παράλογων πρωτεϊνών.

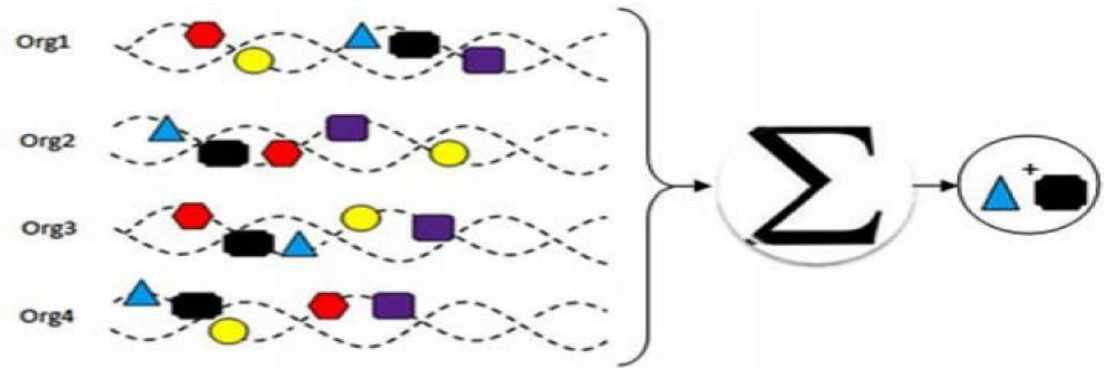
Ομόλογες πρωτεΐνες μπορεί να είναι ορθόλογες ή παράλογες. Με τον όρο ορθόλογα, αναφερόμαστε σε γονίδια και κατ' επέκταση και σε πρωτεΐνες που εντοπίζονται σε διαφορετικά είδη τα οποία έχουν εξελιχθεί από ένα κοινό προγονικό γονίδιο ενώ αντίθετα με τον όρο παράλογα αναφερόμαστε σε γονίδια και κατ' επέκταση και σε πρωτεΐνες που είναι προϊόντα γονιδιακού διπλασιασμού (Tatusov, Koonin, & Lipman, 1997). Οι ορθόλογες πρωτεΐνες διατηρούν τη λειτουργικότητά τους κατά την πάροδο των εξελικτικών χρόνων, ενώ οι παράλογες είναι πολύ πιθανό να αποκτήσουν νέες λειτουργίες. Ως εκ τούτου, αν δυο πρωτεΐνες Α και Β αλληλεπιδρούν μεταξύ τους, τότε και οι ορθόλογές τους Α και Β σε ένα άλλο είδος είναι πιθανό να αλληλεπιδρούν.

Β) Προσεγγίσεις που βασίζονται στις αλληλεπιδράσεις μεταξύ δομικά αυτοτελών περιοχών

Αρχικά ας ορίσουμε τι εννοούμε όταν αναφερόμαστε σε μια δομικά αυτοτελή περιοχή. Εξετάζοντας τις τρισδιάστατες δομές πρωτεϊνών, παρατηρείται ότι σε αρκετές περιπτώσεις κάποια τμήματα της δομής ξεχωρίζουν μεταξύ τους. Αυτό το γεγονός υποδηλώνει δομική και, συχνά, λειτουργική ανεξαρτησία και αυτοτέλεια. Αυτές οι περιοχές, οι οποίες είναι κατά κανόνα συμπαγείς και δομικά ανεξάρτητες ονομάζονται δομικά αυτοτελείς περιοχές. Η διάκριση των περιοχών αυτών βοηθά ιδιαίτερα στη μελέτη των διαθέσιμων πρωτεϊνικών δομών, καθώς θεωρούνται δομικά αυθύπαρκτες, δηλαδή αποκτούν το τελικό τους δίπλωμα ανεξάρτητα από το υπόλοιπο τμήμα της πολυπεπτιδικής αλυσίδας, ενώ επίσης παίζουν καθοριστικό ρόλο και στην ανάπτυξη μεθόδων πρόβλεψης της δομικής κατηγοριοποίησης των πρωτεϊνών, της υποκυτταρικής τους θέσης, της τοπολογίας τους (όσον αφορά στις μεμβρανικές πρωτεΐνες κ.α. Καθώς οι δομικά αυτοτελείς περιοχές εμπλέκονται άμεσα στην ανάπτυξη διαμοριακών αλληλεπιδράσεων, είναι θεμελιώδους σημασίας και για τις αλληλεπιδράσεις πρωτεϊνών – πρωτεϊνών γενικότερα. Διάφορες μελέτες έχουν δείξει ότι οι αλληλεπιδράσεις μεταξύ δομικά αυτοτελών περιοχών (Domain – Domain Interactions DDIs) αποτελούν ένα ασφαλές κριτήριο για την πρόβλεψη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών και το αντίστροφο(Memišević, Wallqvist, & Reifman, 2013; Promponas, Ouzounis, & Iliopoulos, 2014; Wojcik & Schächter, 2001).

### 1.3.3.3. Μελέτες σε επίπεδο γονιδίων

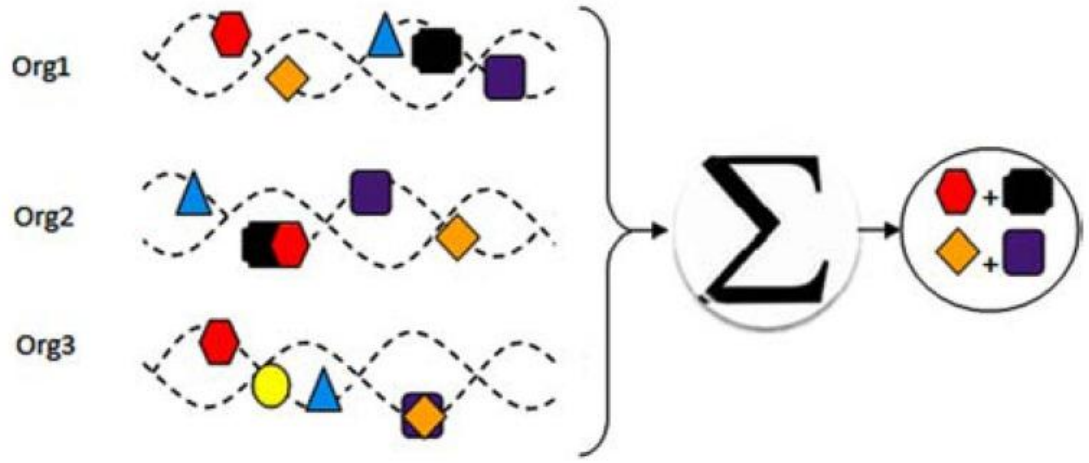
Α) Προσεγγίσεις που βασίζονται σε χαρακτηριστικά γονιδιακής κλίμακας




**Εικόνα 1.53** Υπολογιστικές προσεγγίσεις που βασίζονται στη γειτνίαση των γονιδίων στο χρωμόσωμα. Η βασική ιδέα των προσεγγίσεων αυτών είναι ότι τα λειτουργικά συσχετιζόμενα γονίδια βρίσκονται σε γειτονικές θέσεις στο γονιδίωμα. Έτσι για παράδειγμα, οι πρωτεΐνες που κωδικοποιούνται από το μαύρο γονίδιο και από το μπλε γονίδιο πιθανά αλληλεπιδρούν (Zahiri, Bozorgmehr, & Masoudi-Nejad, 2013).

Οι υπολογιστικές προσεγγίσεις για τη μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών που βασίζονται σε χαρακτηριστικά γονιδιακής κλίμακας όπως είναι η «γειτνίαση» των γονιδίων στο χρωμόσωμα (gene neighbourhood) είναι από τις προσεγγίσεις που εφαρμόστηκαν για την πρόβλεψη πρωτεϊνικών αλληλεπιδράσεων(Bowers et al., 2004; Dandekar, Snel, Huynen, & Bork, 1998; Galperin & Koonin, 2000). Οι προσεγγίσεις αυτές βρίσκουν κυρίως εφαρμογή σε προκαρυωτικά γονιδιώματα, όπου τα γονίδια που κωδικοποιούν για λειτουργικά αλληλεπιδρώντα πρωτεϊνικά μόρια, τείνουν να οργανώνονται σε γειτονικές περιοχές του γονιδιώματος, που ονομάζονται οπερόνια και μεταγράφονται ως ένα πολυστιστρονικό mRNA. Αν η «σχέση γειτνίασης» διατηρείται εξελικτικά σε πολλά γονιδιώματα, τότε μπορεί να εξαχθεί το συμπέρασμα της λειτουργικής σύνδεσης μεταξύ των κωδικοποιούμενων –από τα γειτονικά γονίδια πρωτεϊνών(Wojcik & Schächter, 2001; Yamada, Kabir, & Tsunedomi, 2003) . Η γειτνίαση των γονιδίων χρησιμοποιείται για τη μελέτη λειτουργικών σχέσεων μεταξύ των αντίστοιχων πρωτεϊνών και έχει αποδειχθεί ότι είναι μια πιο ανεξάρτητη μέθοδος και αποδίδει καλύτερα σε σχέση με μεθόδους που λαμβάνουν ως κριτήριο το σχετικό προσανατολισμό των γονιδίων στο χρωμόσωμα.

Β) Προσεγγίσεις που βασίζονται στο φαινόμενο της γονιδιακής σύντηξης (gene fusion)




**Εικόνα 1.54** Υπολογιστικές προσεγγίσεις που βασίζονται στο φαινόμενο της γονιδιακής σύντηξης. Εάν δυο ξεχωριστές πρωτεΐνες σε ένα οργανισμό, εμφανίζονται ως «συντηγμένες» σε έναν άλλο, μπορούμε να υποθέσουμε ότι αλληλεπιδρούν.(Zahiri et al., 2013)

Η βασική ιδέα των υπολογιστικών προσεγγίσεων που βασίζονται στο φαινόμενο της γονιδιακής σύντηξης είναι ότι πρωτεΐνες αποτελούμενες από μια δομικά αυτοτελή περιοχή σε έναν οργανισμό μπορούν να «συντηχθούν» δημιουργώντας μια πρωτεΐνη με περισσότερες της μιας δομικά αυτοτελούς περιοχής σε άλλους οργανισμούς. Η «σύντηξη» των δομικά αυτοτελών περιοχών υποδηλώνει την ύπαρξη λειτουργικής συσχέτισης μεταξύ των δυο πρωτεϊνών, οι

οποίες είναι πιθανό να σχηματίσουν ένα πρωτεϊνικό σύμπλοκο(Enright, Iliopoulos, Kyrpides, & Ouzounis, 1999; Marcotte et al., 1999). Φαινόμενα «σύντηξης» δομικά αυτοτελών περιοχών είναι πολύ κοινά, σε πρωτεΐνες που συμμετέχουν σε μεταβολικά μονοπάτια(Freiberg, 2001; Overbeek, Fonstein, D'Souza, Pusch, & Maltsev, 1999). Η υπολογιστική αυτή προσέγγιση μπορεί να χρησιμοποιηθεί για την πρόβλεψη των αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών αξιοποιώντας την πληροφορία σχετικά με τη διάταξη των περιοχών στο γονιδίωμα που κωδικοποιούν για τις δομικά αυτοτελείς περιοχές, μεταξύ διαφορετικών γονιδιωμάτων.

Γ) Προσεγγίσεις που βασίζονται σε δεδομένα γονιδιακής έκφρασης

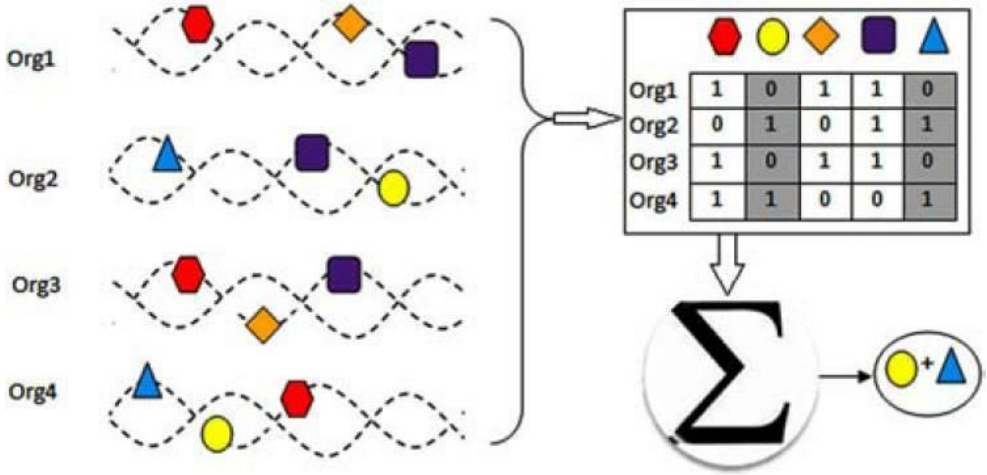
Οι προσεγγίσεις αυτές έχουν το πλεονέκτημα ταυτόχρονης μελέτης των επιπέδων γονιδιακής έκφρασης ενός οργανισμού. Με τον όρο γονιδιακή έκφραση αναφερόμαστε στην ποσοτικοποίηση των επιπέδων έκφρασης ενός συγκεκριμένου γονιδίου σε ένα κύτταρο, σε ένα ιστό ή σε έναν οργανισμό υπό την επίδραση διαφορετικών πειραματικών συνθηκών ή σε συγκεκριμένα χρονικά διαστήματα. Εφαρμόζοντας συγκεκριμένους αλγορίθμους ομαδοποίησης, τα γονίδια μπορούν να ομαδοποιηθούν ανάλογα με τα επίπεδα έκφρασής τους και το προφίλ της γονιδιακής έκφρασης που τελικά προκύπτει, κάτω από καθορισμένες πειραματικές συνθήκες κάθε φορά, μπορεί να βοηθήσει ώστε να εξαχθούν λειτουργικές σχέσεις μεταξύ των γονιδίων αλλά και των πρωτεϊνών τις οποίες κωδικοποιούν. Τέλος, πολλές έρευνες, εστιάζουν επίσης στην διερεύνηση της σχέσης συνέκφρασης γονιδίων και αλληλεπίδρασης των κωδικοποιούμενων πρωτεϊνών(Grigoriev, 2001).

#### 1.3.3.4. Μελέτες σε επίπεδο εξελικτικών σχέσεων πρωτεϊνών

Α) Προσεγγίσεις που βασίζονται στα φυλογενετικά δένδρα

Βασική αρχή της σημερινής βιολογίας είναι ότι όλοι οι οργανισμοί έχουν προέλθει από ένα κοινό πρόγονο μέσω της εξέλιξης. Ο πιο ευρέως διαδεδομένος τρόπος εικονικής αναπαράστασης της εξελικτικής ιστορίας των οργανισμών είναι μέσω της χρήσης φυλογενετικών δένδρων. Οι υπολογιστικές προσεγγίσεις για την πρόβλεψη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών που βασίζονται στα φυλογενετικά δένδρα, βασίζονται στην υπόθεση ότι οι πρωτεΐνες που αλληλεπιδρούν έχουν παρόμοια μοριακά φυλογενετικά δένδρα λόγω της συνεξέλιξής τους με στόχο να διατηρηθεί η ευνοϊκή αλληλεπίδρασή τους κατά την πάροδο των εξελικτικών χρόνων(Ochoa & Pazos, 2010). Η βασική αρχή στην οποία και στηρίζονται οι υπολογιστικές προσεγγίσεις αυτού του είδους είναι ότι η συνεξέλιξη των αλληλεπιδρώντων πρωτεϊνών μπορεί να αντανακλάται από το βαθμό ομοιότητας των φυλογενετικών δένδρων των υπό εξέταση πρωτεϊνών(Craig & Liao, 2007).

Β) Προσεγγίσεις που βασίζονται στα φυλογενετικά προφίλ




**Εικόνα 1.55** Υπολογιστικές μέθοδοι που βασίζονται στα φυλογενετικά προφίλ. Το φυλογενετικό προφίλ για κάθε πρωτεΐνη είναι ένα δυαδικό διάνυσμα που αντανακλά την απουσία ή την παρουσία της υπό μελέτη πρωτεΐνης σε ένα σύνολο οργανισμών (Zahiri et al., 2013).

Η βασική ιδέα των προσεγγίσεων αυτών είναι ότι οι πρωτεΐνες που συνδέονται λειτουργικά έχουν την τάση να συνυπάρχουν κατά τη διάρκεια της εξέλιξης ενός οργανισμού(Srinivas, Rao, Sridhar, & Gedela, 2008). Με άλλα λόγια, εάν δύο πρωτεΐνες είναι λειτουργικά συνδεδεμένες στο γονιδίωμα ενός οργανισμού, θα υπάρχει ισχυρή εξελικτική πίεση για να κληρονομούνται μαζί(Freiberg, 2001). Έτσι, αντίστοιχες ορθόλογες πρωτεΐνες σε άλλα γονιδιώματα με την πάροδο των εξελικτικών χρόνων ή θα διατηρηθούν ή θα χαθούν. Ως εκ τούτου, μπορούμε να ανιχνεύσουμε την παρουσία ή την απουσία των αντίστοιχων πρωτεϊνών στο φυλογενετικά προφίλ των υπό μελέτη γονιδιωμάτων. Τα φυλογενετικά προφίλ χρησιμοποιούνται για να περιγράψουν την ταυτόχρονη ύπαρξη μιας πρωτεΐνης σε ένα σετ αναλυόμενων γονιδιωμάτων. Έτσι, αν δυο πρωτεΐνες «μοιράζονται» ένα κοινό φυλογενετικό προφίλ μπορούμε να υποθέσουμε ότι έχουν και κάποια λειτουργική συσχέτιση. Αξίζει να σημειωθεί ότι τέτοιου είδους προσεγγίσεις δίνουν ικανοποιητικά αποτελέσματα όσον αφορά στην πρόβλεψη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών μόνο σε προκαρυωτικούς οργανισμούς(Lin, Wu, & Chang, 2013).

### 1.3.3.5. Μελέτες σε επίπεδο τοπολογίας πρωτεϊνικών δικτύων.

Η υπολογιστική ανάλυση των δικτύων αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών ξεκινάει με την αναπαράσταση της δομής του δικτύου. Η απλούστερη αναπαράσταση είναι με τη μορφή ενός μαθηματικού γράφου G=(V,E) αποτελούμενου από V κόμβους και E ακμές. Σε ένα τέτοιο γράφο οι πρωτεΐνες αναπαρίστανται ως κόμβοι, ενώ δυο πρωτεΐνες που αλληλεπιδρούν αναπαρίστανται ως γειτονικοί κόμβοι συνδεδεμένοι με μια ακμή (Wagner,

2003).Μια λεπτομερής ανάλυση ενός δικτύου αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών, δηλαδή η «ανάλυση» της διάταξης των αλληλεπιδράσεων μεταξύ των πρωτεϊνών, βοηθά ώστε να προκύψουν πολυάριθμα δεδομένα. Για παράδειγμα, γειτονικές πρωτεΐνες στο δίκτυο μπορεί να μοιράζονται την ίδια λειτουργία, ενώ πυκνά συνδεδεμένα υποδίκτυα πρωτεϊνών είναι πιθανό να σχηματίζουν πρωτεϊνικά σύμπλοκα σε ορισμένες βιολογικές διεργασίες. Έτσι, συμπεράσματα όσον αφορά στη λειτουργία μιας πρωτεΐνης μπορούν να εξαχθούν εάν εντοπιστούν οι πρωτεΐνες με τις οποίες αλληλεπιδρά και τα πρωτεϊνικά σύμπλοκα στα οποία συμμετέχει. Η πρόβλεψη νέων αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών με βάση την τοπολογία των πρωτεϊνικών δικτύων είναι μια σχετικά νέα προσέγγιση που βασίζεται αποκλειστικά στη δομική πληροφορία που παρέχεται από το δίκτυο των πρωτεϊνικών αλληλεπιδράσεων(Cannistraci, Alanis-Lobato, & Ravasi, 2013; Guimarães, Jothi, Zotenko, & Przytycka, 2006; Lehner & Fraser, 2004; Rhodes et al., 2005) και βασίζεται σε μετρικές των δικτύων όπως είναι η συνδεσιμότητα ενός κόμβου η οποία αποκαλύπτει τον αριθμό των γειτόνων του.

### 1.3.4. Υπολογιστικές μέθοδοι για την μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών βασισμένες σε δεδομένα συνεξέλιξης

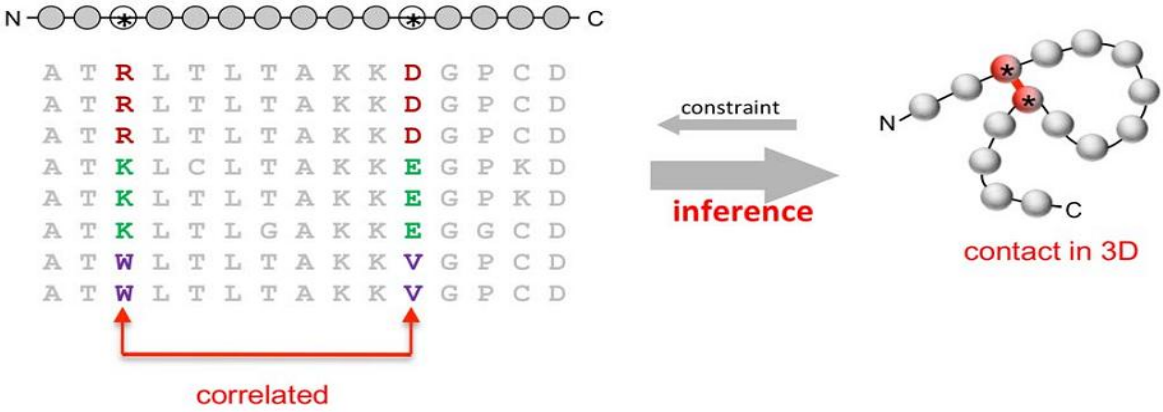
Με βάση όλα όσα έχουν αναφερθεί παραπάνω όσον αφορά στις υπολογιστικές προσεγγίσεις για τη μελέτη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών γίνεται αντιληπτό ότι οι περισσότερες από αυτές βασίζονται σε δομικά χαρακτηριστικά των υπό μελέτη πρωτεϊνών ή σε χαρακτηριστικά γονιδιακής κλίμακας που βρίσκουν εφαρμογή κυρίως σε προκαρυωτικούς οργανισμούς. Στην παρούσα ενότητα θα γίνει μια αναλυτικότερη περιγραφή των υπολογιστικών προσεγγίσεων που έχουν αναπτυχθεί για την πρόβλεψη αλληλεπιδράσεων πρωτεϊνών – πρωτεϊνών που βασίζονται στην αμινοξική ακολουθία των πρωτεϊνών αξιοποιώντας την συν-εξελικτική πληροφορία η οποία μπορεί να αποκαλύψει κατάλοιπα με δομικό και λειτουργικό ρόλο αλλά και κατάλοιπα που είναι υπεύθυνα για την αλληλεπίδραση μεταξύ δυο πρωτεϊνικών μορίων.

Η συνεξέλιξη, η οποία μπορεί να οριστεί ως η αλληλεξάρτηση μεταξύ των εξελικτικών αλλαγών δυο οντοτήτων, παίζει καθοριστικό ρόλο σε όλα τα βιολογικά επίπεδα, από το επίπεδο του οικοσυστήματος ως το επίπεδων των μορίων. Η πρώτη περιγραφή της συνεξέλιξης αφορούσε το βιολογικό επίπεδο των ειδών και έγινε από τον Κάρολο Δαρβίνο, κατά τη μελέτη των ορχιδέων και των επικονιαστών τους όπου και παρατήρησε ότι το μήκος της προβοσκίδας των επικονιαστών σχετιζόταν με το μέγεθος της στεφάνης του άνθους της ορχιδέας (Darwin, 1862). Κατά το πρώτο μισό του 20$^{ου}$ αιώνα, αρκετοί βιολόγοι συνέχισαν τη μελέτη του φαινομένου

της συνεξέλιξης με αποτέλεσμα να εγκαθιδρυθεί και η γενετική του βάση (DOBZHANSKY, 1950).

Ο όρος συνεξέλιξη αποδίδεται στον P.Ehrlich, που μελέτησε το φαινόμενο στο επίπεδο των ειδών(Ehrlich & Raven, 1964), ενώ ο πιο αποδεκτός ορισμός του φαινομένου αυτού είναι ότι πρόκειται για τις αμοιβαίες εξελικτικές αλλαγές μεταξύ αλληλεπιδρώντων ειδών. Ακόμα και από αυτές τις πρώιμες μελέτες, γίνεται αντιληπτό ότι **η έννοια της αλληλεπίδρασης είναι στενά συνδεδεμένη με αυτή της συνεξέλιξης**. Το γεγονός ότι συσχετιζόμενες οντότητες ή οντότητες που αλληλεπιδρούν συνεξελίσσονται είναι και ο λόγος για τον οποίο είναι χρήσιμη η μελέτη της συνεξέλιξης και σε μοριακό επίπεδο.

Η μελέτη της συνεξέλιξης σε μοριακό επίπεδο είναι πολύ πιο πρόσφατη (de Juan, Pazos, & Valencia, 2013)και είναι εμφανής μεταξύ αμινοξικών καταλοίπων πρωτεϊνικών ακολουθιών, όπου μεταλλαγές σε θέσεις της αμινοξικής ακολουθίας που παίζουν σημαντικό δομικό ή λειτουργικό ρόλο συχνά συσχετίζονται μεταξύ τους. Μια πρακτική εφαρμογή της παρατήρησης αυτής είναι η πρόβλεψη επαφών μεταξύ αμινοξικών καταλοίπων σε πρωτεϊνικές δομές με μόνη πηγή πληροφορίας την πρωτεϊνική ακολουθία και πιο συγκεκριμένα μέσω της πολλαπλής στοίχισης ομόλογων πρωτεϊνών της υπό μελέτη πρωτεΐνης (de Juan et al., 2013).

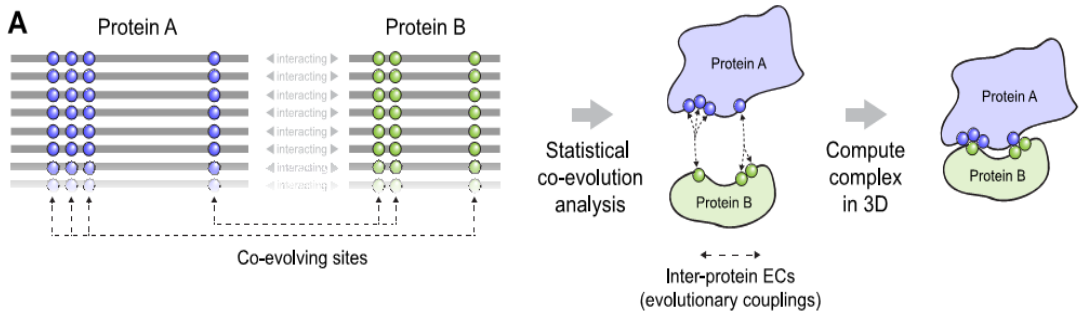



**Εικόνα 1.56** Συνεξέλιξη μεταξύ αμινοξικών καταλοίπων που παίζουν σημαντικό δομικό ή λειτουργικό ρόλο σε μια οικογένεια ομόλογων πρωτεϊνών. Κύρια πηγή συνεξελικτικής πληροφορίας μεταξύ αμινοξικών καταλοίπων είναι η πολλαπλή στοίχιση μιας οικογένειας ομόλογων πρωτεϊνών. Στην πολλαπλή στοίχιση μπορούμε να παρατηρήσουμε ότι αμινοξικά κατάλοιπα σε συγκεκριμένες θέσεις συμμεταβάλονται και να εξάγουμε το συμπέρασμα ότι είναι υπεύθυνα για τη διατήρηση της δομικής και λειτουργικής ακεραιότητας της υπό μελέτη πρωτεϊνικής ακολουθίας (Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. 2011)

Αν «ανέβουμε» ένα «επίπεδο μοριακής ιεραρχίας», η συνεξέλιξη είναι επίσης εμφανής και μεταξύ αλληλεπιδρώντων ή λειτουργικά συσχετιζόμενων πρωτεϊνών. Πολλά ζεύγη αλληλεπιδρώντων πρωτεϊνών μοιράζονται κοινή εξελικτική ιστορία, δηλαδή υπόκεινται σε κοινό εξελικτικό έλεγχο όπως περιγράφηκε και στην ενότητα των υπολογιστικών προσεγγίσεων που βασίζονται στα φυλογενετικά

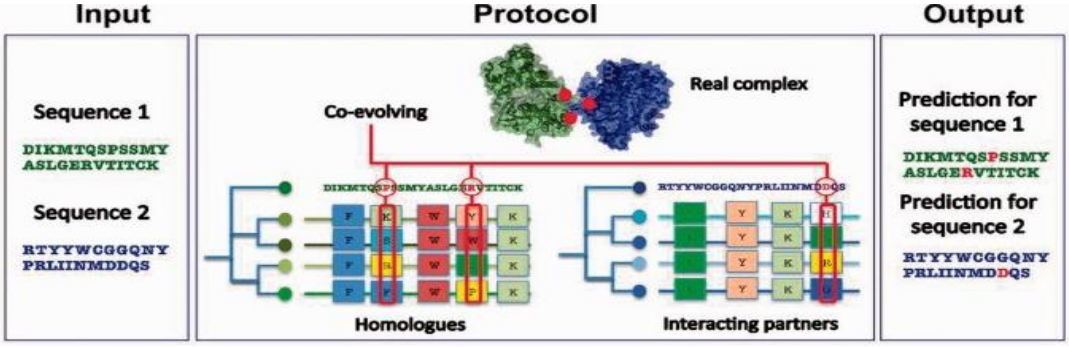
προφίλ. Η μελέτη της συνεξέλιξης στο επίπεδο των δυο αλληλεπιδρώντων πρωτεϊνών είναι εμφανής επίσης και μεταξύ ζευγαριών αμινοξικών καταλοίπων ή και ολόκληρων ομάδων αμινοξικών καταλοίπων τα οποία συμμεταβάλονται και είναι υπεύθυνα για τις αλληλεπιδράσεις μεταξύ των πρωτεϊνικών μορίων. Έτσι η συνεξέλιξη στο επίπεδο των αλληλεπιδρώντων πρωτεϊνών αφορά τις **αμοιβαίες αλλαγές σε αμινοξικά κατάλοιπα που είναι υπεύθυνα για την αλληλεπίδραση μεταξύ των πρωτεϊνικών μορίων**.

Μια πρακτική εφαρμογή της συνεξέλιξης στο επίπεδο των αλληλεπιδρώντων πρωτεϊνών είναι η πρόβλεψη αλληλεπιδρώντων καταλοίπων ή και καταλοίπων που έρχονται σε επαφή σε πρωτεϊνικά σύμπλοκα με μόνη πηγή πληροφορίας τις πρωτεϊνικές ακολουθίες των αλληλεπιδρώντων πρωτεϊνών και πιο συγκεκριμένα εξετάζοντας «συνδυαστικά» τις πολλαπλές στοιχίσεις ομόλογων πρωτεϊνών τους.




**Εικόνα 1.57** Συνεξέλιξη μεταξύ ζευγαριών αμινοξικών καταλοίπων που είναι υπεύθυνα για τις αλληλεπιδράσεις μεταξύ δυο πρωτεϊνικών μορίων. Κύρια πηγή συνεξελικτικής πληροφορίας μεταξύ ζευγαριών αμινοξικών καταλοίπων δυο αλληλεπιδρώντων πρωτεϊνών είναι οι πολλαπλές στοιχίσεις των ομόλογων πρωτεϊνών τους. Εξετάζοντας συνδυαστικά τις δυο πολλαπλές στοιχίσεις μπορούμε να παρατηρήσουμε ότι ζευγάρια αμινοξικών καταλοίπων σε συγκεκριμένες θέσεις των πολλαπλών στοιχίσεων συμμεταβάλονται και να εξάγουμε το συμπέρασμα ότι είναι υπεύθυνα για την αλληλεπίδραση μεταξύ των δυο πρωτεϊνικών μορίων. Με βάση τα προβλεπόμενα αλληλεπιδρώντα κατάλοιπα και με τη βοήθεια μεθόδων αγκυροβόλησης μπορούμε να προβλέψουμε ακόμη και τη δομή του πρωτεϊνικού συμπλόκου(Hopf et al., 2014).

Η συνεξελικτική πληροφορία μπορεί να συνδυαστεί και με δομική πληροφορία, όπως στην περίπτωση του αλγορίθμου Coev2Net (Hosur, Xu, Bienkowska, & Berger, 2011). Ο αλγόριθμος ακολουθεί μια πορεία τριών βημάτων, που περιλαμβάνει την πρόβλεψη της επιφάνειας αλληλεπίδρασης, την αξιολόγηση της πρόβλεψης και την απόδοση ενός σκορ αξιοπιστίας στην προβλεπόμενη αλληλεπίδραση.




**Εικόνα 1.58** Σχηματική αναπαράσταση του πρωτοκόλλου που ακολουθείται στις μεθόδους πρόβλεψης αλληλεπιδράσεων πρωτεϊνών-πρωτεϊνών με τη βοήθεια των μεθόδων συνεξέλιξης.

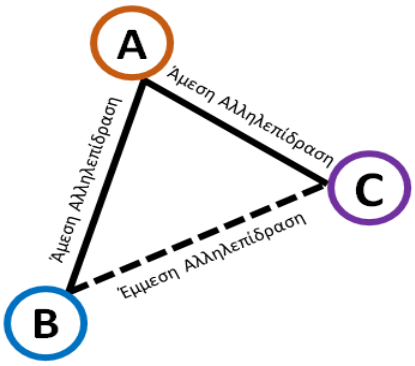
## 1.3.5. Στατιστικά μοντέλα για την πρόβλεψη αλληλεπιδρώντων καταλοίπων μέσω της συνεξελικτικής πληροφορία

Υπάρχουν διάφορα στατιστικά μοντέλα για την πρόβλεψη αλληλεπιδρώντων καταλοίπων μέσω της συνεξελικτικής πληροφορίας, είτε αναφερόμαστε στο επίπεδο μιας πρωτεΐνης είτε αναφερόμαστε στο επίπεδο δυο αλληλεπιδρώντων πρωτεϊνών. Οι διαθέσιμές υπολογιστικές προσεγγίσεις μπορούν να χωριστούν σε δυο μεγάλες κατηγορίες: τα τοπικά στατιστικά μοντέλα (**local statistical models**) και τα ολικά στατιστικά μοντέλα (**glodal statistical models**). Ο διαχωρισμός αυτός των υπολογιστικών προσεγγίσεων εξαρτάται και βασίζεται στις θεμελιώδεις τους αρχές.

Η πραγματική εξελικτική συσχέτιση μεταξύ δυο αμινοξικών καταλοίπων μπορεί να «καλυφθεί» από έμμεσες συσχετίσεις μεταξύ αμινοξικών καταλοίπων και μερικές φορές οι συσχετίσεις αυτές είναι πιο ισχυρές από τις άμεσες (Marks, Hopf, & Sander, 2012).

Ένας λόγος εμφάνισης έμμεσων συσχετίσεων μεταξύ δυο μη αλληλεπιδρώντων αμινοξικών καταλοίπων θα μπορούσε να είναι η παρουσία κοινών γειτονικών καταλοίπων. Η παραδοχή στην οποία βασίζονται τα τοπικά στατιστικά μοντέλα είναι ότι κάθε ζεύγος θέσεων σε μια πολλαπλή στοίχιση, δηλαδή κάθε ζεύγος αμινοξικών καταλοίπων είναι στατιστικά ανεξάρτητο από όλες τις υπόλοιπες θέσεις. Αυτό καθιστά τα τοπικά στατιστικά μοντέλα λιγότερο αποτελεσματικά στον προσδιορισμό πιθανά αλληλεπιδρώντων καταλοίπων. Σε αντίθεση με τα τοπικά στατιστικά μοντέλα, τα ολικά στατιστικά μοντέλα βασίζονται στην παραδοχή ότι κάθε ζευγάρι θέσεων σε μια πολλαπλή στοίχιση εξαρτάται και από όλες τις υπόλοιπες, με αποτέλεσμα τα ζεύγη των αμινοξικών καταλοίπων που προβλέπονται με βάση μοντέλα αυτά να είναι πιο πιθανό να είναι και πραγματικά ζεύγη αλληλεπιδρώντων καταλοίπων. Στους παρακάτω πίνακες παρατίθενται οι

κυριότερες μέθοδοι των τοπικών στατιστικών μοντέλων και των ολικών στατιστικών μοντέλων και οι κύριες εφαρμογές τους.




**Εικόνα 1.59** Έμμεση αλληλεπίδραση μεταξύ των καταλοίπων B και C. Συγκεκριμένα τα κατάλοιπα B και C συνεξελίσσονται με το A, έτσι ώστε να διατηρηθούν οι αλληλεπιδράσεις μεταξύ των τριών καταλοίπων. Ωστόσο, δεν υπάρχει φυσική επαφή μεταξύ του B και του C αλλά έμμεση αλληλεπίδραση η οποία παρατηρείται εξαιτίας της παρέμβασης του A.

**Πίνακας 1.7** Τοπικά στατιστικά μοντέλα και οι κύριες εφαρμογές τους.

| **Τοπικά στατιστικά μοντέλα** | | |
|---|---|---|
| **Στατιστικό κριτήριο** | Κύρια εφαρμογή | Αναφορές |
| **Mutual Information (Αμοιβαία πληροφορία)** | Πρόβλεψη επαφών μεταξύ αμινοξικών καταλοίπων | (Korber, Farber, Wolpert, & Lapedes, 1993) |
| **Corrected Mutual Information (Διορθωμένη αμοιβαία πληροφορία)** | Πρόβλεψη επαφών μεταξύ αμινοξικών καταλοίπων | (Dunn, Wahl, & Gloor, 2008) |
| **McLachlan based Substitution Correlation(McBasc)** | Πρόβλεψη επαφών μεταξύ αμινοξικών καταλοίπων | (Göbel, Sander, Schneider, & Valencia, 1994) |
| **Scatistical Coupling Analysis** | Πρόβλεψη διαφορετικών στερεοδιατάξεων και αλλοστερικής ρύθμισης πρωτεϊνών | (Lockless & Ranganathan, 1999; Reynolds, McLaughlin, & Ranganathan, 2011) |

**Πίνακας 1.8** Ολικά στατιστικά μοντέλα και οι κύριες εφαρμογές τους.

| **Ολικά στατιστικά μοντέλα** | | |
|---|---|---|
| **Στατιστικό κριτήριο** | **Κύρια εφαρμογή** | **Αναφορές** |
| **plmDCA (pseudo-likelihood maximization Direct Coupling Analysis)** | Πρόβλεψη αλληλεπιδρώντων καταλοίπων μεταξύ δυο πρωτεϊνών, πρόβλεψη πρωτεϊνικού διπλώματος (μόνο για πρωτεΐνες που αποτελούνται από μια δομικά αυτοτελή περιοχή) | (Ekeberg, Hartonen, & Aurell, 2014; Ekeberg, Lövkvist, Lan, Weigt, & Aurell, 2012; Feinauer, Skwark, Pagnani, & Aurell, 2014) |

| | | |
|---|---|---|
| **PSICOV (Protein Sparse Inverse COVariance)** | Πρόβλεψη αμινοξικών καταλοίπων με σημαντικό δομικό ή λειτουργικό ρόλο σε μια πρωτεΐνη. | (Jones, Buchan, Cozzetto, & Pontil, 2012) |
| **mfDCA (mean-field Direct Coupling Analysis)** | Πρόβλεψη αλληλεπιδρώντων καταλοίπων μεταξύ δυο πρωτεϊνών | (Marks et al., 2012; Morcos et al., 2011) |

#### 1.3.5.1. Μέθοδος cMI (Corrected Mutual Information)

**Εισαγωγικά** : Μέθοδος Αμοιβαίας Πληροφορίας (Mutual Information, MI)

Η Αμοιβαία Πληροφορία (Mutual Information, MI) είναι ένα στατιστικό μέτρο που βρίσκει εφαρμογές κυρίως στη Θεωρία της Πληροφορίας (Information Theory). Στο πλαίσιο των βιολογικών δεδομένων, και ειδικότερα στο πλαίσιο της εκτίμησης της συσχέτισης μεταξύ των θέσεων μιας πολλαπλής στοίχισης ακολουθιών, αποτελεί ένα ελκυστικό στατιστικό μέτρο καθώς μετράει ρητά την εξάρτηση μιας θέσης της πολλαπλής στοίχισης από μια άλλη. Οι εφαρμογές του στατιστικού μέτρου της MI σε πολλαπλές στοιχίσεις ακολουθιών με στόχο να εξαχθούν συμπεράσματα σχετικά με αλληλεπιδρώντα αμινοξικά κατάλοιπα , περιορίζονται κυρίαρχα από τη φύση του στατιστικού μοντέλου. Πρόκειται για ένα τοπικό στατιστικό μοντέλο, δηλαδή κάθε θέση της πολλαπλής στοίχισης θεωρείται ανεξάρτητη από όλες τις υπόλοιπες, με αποτέλεσμα να μην μπορεί να γίνει διάκριση μεταξύ άμεσα και έμμεσα αλληλεπιδρώντων καταλοίπων.

Για τον υπολογισμό της Αμοιβαίας Πληροφορίας μεταξύ δυο θέσεων σε μια πολλαπλή στοίχιση ακολουθιών πρέπει αρχικά να υπολογιστεί η εντροπία του Shannon. Η **εντροπία του Shannon** (H) , για μια θέση α της πολλαπλής στοίχισης είναι ένα μέτρο της τυχαιότητας των αμινοξικών καταλοίπων στη συγκεκριμένη θέση (Cover and Thomas, 1991).

Υπολογίζεται με βάση τον τύπο:

$$\boldsymbol{H(\alpha) = -\sum_{x} p(x,\alpha)\log_{20} p(x,\alpha)}$$

όπου $p(x,\alpha)$ είναι η συχνότητα εμφάνισης του εκάστοτε αμινοξικού καταλοίπου $x$στη θέση $a$. Η τιμή της εντροπίας του Shannon κυμαίνεται από μηδέν (0), στην περίπτωση της πλήρους συντήρησης των αμινοξικών καταλοίπων στην συγκεκριμένη θέση της πολλαπλής στοίχισης έως ένα (1) στην περίπτωση που και τα 20 αμινοξικά κατάλοιπα είναι ισοκατανεμημένα στη συγκεκριμένη θέση.

Η **συνδυασμένη πληροφορία/εντροπία** $\boldsymbol{H(\alpha, b)}$ κάθε ζεύγους θέσεων της πολλαπλής στοίχισης υπολογίζεται ως εξής

$$\boldsymbol{H(\alpha, b) = -\sum_{x}^{20}\sum_{y}^{20} p(x, y)\log_{20} p(x, y)}$$

όπου υπολογίζονται οι τιμές της εντροπίας για όλα τα ζεύγη αμινοξικών καταλοίπων $(x, y)$μεταξύ των θέσεων $a$ και b της πολλαπλής στοίχισης. Οι τιμές της συνδυασμένης εντροπίας κυμαίνονται από μηδέν (0) έως δυο(2).

Η **Αμοιβαία Πληροφορία**, στην ουσία μετράει τη μείωση της αβεβαιότητας για μια θέση της πολλαπλής στοίχισης με δεδομένη την πληροφορία που έχουμε για μια άλλη θέση της ίδιας πολλαπλής στοίχισης (Cover and Thomas, 1991). Στην πραγματικότητα, αντανακλά το βαθμό στον οποίο αν γνωρίζουμε την ταυτότητα του αμινοξικού καταλοίπου $x$ στη θέση $a$ μιας πολλαπλής στοίχισης, μπορούμε να προβλέψουμε την ταυτότητα ενός άλλου αμινοξικού καταλοίπου y στη θέση b μια πολλαπλής στοίχισης. Έτσι μπορούμε να πούμε ότι αντικατοπτρίζει τον βαθμό συσχέτισης μεταξύ των θέσεων α και b σε μια πολλαπλή στοίχιση.

Η Αμοιβαία Πληροφορία υπολογίζεται με τον παρακάτω τύπο

$$\boldsymbol{MI(\alpha, b) = H(\alpha) + H(b) - H(\alpha, b)}$$

και οι τιμές της κυμαίνονται από 0 έως 1 με τις μεγαλύτερες τιμές να υποδηλώνουν μεγαλύτερη αλληλοεξάρτηση μεταξύ των θέσεων α και b.

Μέθοδος διορθωμένης αμοιβαίας πληροφορίας – cMI (**C**orrected **M**utal **I**nformation)

Οι εφαρμογές της Αμοιβαίας Πληροφορίας περιορίζονται σε βιολογικά δεδομένα και πιο συγκεκριμένα σε πολλαπλές στοιχίσεις ακολουθιών από τρεις παράγοντες:

Οι θέσεις με υψηλότερη εντροπία, τείνουν να έχουν υψηλότερα ποσοστά τόσο τυχαίας όσο και μη τυχαίας Αμοιβαίας Πληροφορίας σε σχέση με τις θέσεις με χαμηλότερη εντροπία. (Fodor and Aldrich, 2004a; Martin et al., 2005).

Τυχαίες τιμές Αμοιβαίας Πληροφορίας προκύπτουν όταν οι πολλαπλές στοιχίσεις δεν περιέχουν αρκετές ακολουθίες έτσι ώστε ο θόρυβος του υποβάθρου να θεωρείται αμελητέος.

Όλα τα ζεύγη των θέσεων των πολλαπλών στοιχίσεων χαρακτηρίζονται από τιμές Αμοιβαίας Πληροφορίας λόγω των φυλογενετικών σχέσεων των αλληλουχιών της πολλαπλής στοίχισης (Wollenberg and Atchley, 2000). Ο συγκεκριμένος περιορισμός μπορεί να αρθεί, σε έναν βαθμό αν από τις πολλαπλές στοιχίσεις εξαιρεθούν οι αλληλουχίες με πολύ υψηλά ποσοστά ομοιότητας και από στενά συγγενικούς οργανισμούς.

Όπως αναφέρθηκε και παραπάνω, η Αμοιβαία Πληροφορία συσχετίζεται ισχυρά με την εντροπία των αντίστοιχων θέσεων (Fodor and Aldrich, 2004a; Martin et al., 2005). Η επιρροή αυτή μπορεί να αρθεί μερικώς μέσω της διόρθωσης (Martin et al., 2005):

$$\boldsymbol{MI_r = MI(\alpha, b)/H(\alpha, b)}$$

Επιπλέον, Αμοιβαία Πληροφορία μεταξύ δυο θέσεων μιας πολλαπλής στοίχισης προκύπτει λόγω του **θορύβου υποβάθρου** και λόγω της **κοινής εξελικτικής καταγωγής** των ακολουθιών της πολλαπλής στοίχισης.

Με στόχο να διαχωριστεί το «σήμα» που προκαλείται από **«πραγματικές»** αλληλεπιδράσεις μεταξύ αμινοξικών καταλοίπων, $\boldsymbol{MI_{sf}}$ , από το σήμα του υποβάθρου $\boldsymbol{MI_b}$ που είναι το άθροισμα τυχαίων θορύβων και της κοινής εξελικτικής καταγωγής γίνονται δυο υποθέσεις :

Κάθε θέση στην πολλαπλή στοίχιση έχει μια συγκεκριμένη τάση προς την $\boldsymbol{MI_b}$, που σχετίζεται με την εντροπία της και τη φυλογενετική της ιστορία.

Η $\boldsymbol{MI_b}$ μεταξύ 2 θέσεων α και b της πολλαπλής στοίχισης μπορεί να εκφραστεί ως το πηλίκο των **μέσων τιμών** $\boldsymbol{MI_b}$ των θέσεων α και β προς τη μέση τιμή $MI_b$ όλων των θέσεων της πολλαπλής στοίχισης.

Ο όρος αυτός ονομάζεται **Average Product Correction** (APC) και υπολογίζεται με βάση τον παρακάτω τύπο

$$\boldsymbol{APC(\alpha, b) = \frac{MI(\alpha, \overline{x})MI(b, \overline{x})}{\overline{MI}}}$$

Έτσι η $\boldsymbol{MI_{sf}}$ μπορεί να υπολογιστεί με βάση τις παρακάτω εξισώσεις :

$$\boldsymbol{MI_{sf} = MI(\alpha, b) - APC(a, b)}$$

$$\boldsymbol{MI_{sf} = \sum_{x,y} P(x_a, y_b) \log_{20}\left(\frac{P(x_a, y_b)}{P(x_a)P(y_b)}\right) - \frac{MI(\alpha, \overline{x})MI(b, \overline{x})}{\overline{MI}}}$$

Τέλος, με στόχο να γίνει διόρθωση ως προς τον μικρό αριθμό των πρωτεϊνικών ακολουθιών που στην πολλαπλή στοίχιση, οι πιθανότητες των ζευγαριών των αμινοξικών καταλοίπων $P(x, y)$ κανονικοποιούνται με βάση έναν παράγοντα $N(x, y)$, που αντιπροσωπεύει, τον αριθμό των παρατηρήσεων του ζεύγους $(x, y)$ στις θέσεις α και b της πολλαπλής στοίχισης.

Ο παράγοντας $N(x, y)$οποίος υπολογίζεται με βάση τον τύπο :

$$\boldsymbol{N(x, y) = \sum_{(x,y)}\left(\lambda + N(x, y)\right), P(x_a) = \sum_y P(x_a, y_b) \wedge P(y_b) = \sum_x P(x_a, y_b)}$$

και αρχικά σε όλα τα ζεύγη των αμινοξικών ανατίθεται η τιμή $N(x_a, y_b) = \lambda$, της οποίας το εύρος των τιμών μετά από δοκιμές σε διάφορα σετ δεδομένων

ορίζεται ως: $0.025 \leq \lambda \leq 0.075$. Το συγκεκριμένο εύρος τιμών της παραμέτρου λ μπορεί να εφαρμοστεί σε οποιοδήποτε σύνολο δεδομένων ανεξαρτήτου μεγέθους, εξελικτικού μοντέλου και ρυθμού.

### 1.3.5.2. Μέθοδος mfDCA (mean-field Direct Coupling Analysis)

**Εισαγωγικά** : Μέθοδος Άμεσης Συσχέτισης Ζευγαριών Αμινοξικών Καταλοίπων (Direct Coupling Analysis, DCA)

Όπως αναφέρθηκε και εισαγωγικά το κύριο πλεονέκτημα των ολικών στατιστικών μοντέλων είναι η επιτυχής διάκριση άμεσων και έμμεσων συσχετίσεων αμινοξικών καταλοίπων. Τα στατιστικά ολικά στατιστικά μοντέλα χρησιμοποιούν ως συνάρτηση βαθμολόγησης των συσχετίσεων (scoring function) την «Άμεση Συσχέτιση Ζευγαριών Αμινοξικών Καταλοίπων»(Direct Coupling Analysis, DCA).

Μέθοδος Mean Field DCA

Η πολλαπλή στοίχιση των ακολουθιών οργανώνεται ως ένας πίνακας *M x L* , $\{A_i^m\}$ αμινοξικών καταλοίπων. Το κάθε αμινοξικό κατάλοιπο χαρακτηρίζεται από τον αριθμό $m = 1, M$ που αντιπροσωπεύει τις σειρές του πίνακα και άρα την κάθε ακολουθία στην πολλαπλή στοίχιση και από τον αριθμό $i = 1, L$ που αντιπροσωπεύει τις στήλες του πίνακα και άρα την θέση του αμινοξικού καταλοίπου στην πολλαπλή στοίχιση. Κάθε στοιχείο του πίνακα $A \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y, -\}$ μπορεί να πάρει $q = 21$ τιμές , μια για κάθε ένα από τα αμινοξικά κατάλοιπα και μια για τα κενά της πολλαπλής στοίχισης. Αρχικά υπολογίζονται οι απλές πιθανότητες και τις πιθανότητες για τα ζεύγη των αμινοξικών καταλοίπων ως εξής:

$$\boldsymbol{f_i(A) \equiv \frac{1}{M} \sum_{m=1}^{M} \delta(A_i^m, A)}$$

$$\boldsymbol{f_{ij}(A, B) \equiv \frac{1}{M} \sum_{m=1}^{M} \delta(A_i^m, A) \delta(A_j^m, B)}$$

για τις θέσεις στην ακολουθία $1 \leq i, j \leq L$, όπου *A* είναι μια μεταβλητή για το αμινοξικό κατάλοιπο στη θέση *i*, και *B* μια μεταβλητή για το αμινοξικό κατάλοιπο στη θέση *j*.

Τα αθροίσματα τυποποιούνται χρησιμοποιώντας μια συνάρτηση δέλτα, *δ(α,b)* που παίρνει την τιμή 1 όταν α=b και 0 σε κάθε άλλη περίπτωση.

Αν οι στήλες i και j ήταν στατιστικώς ανεξάρτητες, η από κοινού εμπειρική συχνότητα της κατανομής $f_{ij}(A,B)$ θα ήταν κατά προσέγγιση ίση με τις ανεξάρτητες συχνότητες κατανομής $f_i\,(A)\cdot f_j\,(B)$. Γενικότερα, η απομάκρυνση των τιμών από την ισότητα μετρά τη στατιστική συσχέτιση μεταξύ των θέσεων της ακολουθίας. Ψευδείς συσχετίσεις προκύπτουν και στα ολικά στατιστικά μοντέλα λόγω δυο κυρίως παραγόντων:

«Άνισης δειγματοληψίας» μεταξύ των πρωτεϊνικών αλληλουχιών. Στις περισσότερες περιπτώσεις γίνεται αρχικά η αλληλούχιση των γονιδιωμάτων και κατόπιν η μελέτη του πρωτεώματος εξελικτικά κοντινών οργανισμών με αποτέλεσμα να υπάρχουν μεγαλύτερος αριθμός διαθέσιμων ακολουθιών πρωτεϊνών από «συγκεκριμένους» οργανισμούς (υπερεκπροσώπηση οργανισμών στο δείγμα) .

Έμμεσων συσχετίσεων που προκύπτουν ως αποτέλεσμα άμεσων συσχετίσεων.

Με στόχο να μειωθούν οι ψευδείς συσχετίσεις λόγω της «μεροληψίας» όσον αφορά στις διαθέσιμες ακολουθίες, γίνεται στάθμιση των πρωτεϊνικών ακολουθιών και ανατίθεται χαμηλότερη τιμή βάρους σε ακολουθίες με υψηλά ποσοστά ομοιότητας και υψηλότερη τιμή βάρους σε ακολουθίες με χαμηλότερα ποσοστά ομοιότητας μεταξύ των πρωτεϊνών της υπό μελέτη οικογένειας. Αν $L$ είναι το μήκος της ακολουθίας, ορίζεται ένα κατώφλι ομοιότητας $x$, όπου $0 \le x \le 1$, και ομαδοποιούνται οι ακολουθίες που έχουν πάνω από $xL$ ταυτόσημα αμινοξικά κατάλοιπα.

Αναλυτικότερα, για κάθε ακολουθία $m$ στην πολλαπλή στοίχιση, υπολογίζεται ο αριθμός των αλληλουχιών $k_m$ των οποίων η ομοιότητα σε σχέση με την αλληλουχία $m$ -η οποία μετράται ως ο συνολικός αριθμός των ταυτόσημων αμινοξικών καταλοίπων σε σχέση με την m – είναι μεγαλύτερη από $xL$:

$$\boldsymbol{k_m \equiv \sum_{n-1}^{M} \theta \left( \sum_{i=1}^{L} \delta(A_i^m, A_i^n) - xL \right)}$$

Όπου $\vartheta$ είναι η συνάρτηση μοναδιαίου βήματος. Έτσι μπορούν να οριστούν εκ νέου οι συχνότητες των εξισώσεων 9 και 10 σταθμίζοντας κάθε αλληλουχία $m$ κατά την αντίστροφη γειτονική πυκνότητα $\frac{1}{K_m}$ του «χώρου» των διαθέσιμων αλληλουχιών ως εξής:

$$\boldsymbol{f_i(A) \equiv \frac{1}{\lambda + M_{eff}} \left( \frac{\lambda}{q} + \sum_{m=1}^{M} \frac{1}{k_m} \delta(A_i^m, A) \right)}$$

$$\boldsymbol{f_{ij}(A,B) \equiv \frac{1}{\lambda + M_{eff}} \left(\frac{\lambda}{q^2} + \sum_{m=1}^{M} \frac{1}{k_m} \delta(A_i^m, A) \delta(A_j^m, B)\right)}$$

όπου

$$M_{eff} = \sum_{m=1}^{M} \frac{1}{k_m}$$

$M_{eff}$ είναι «αποτελεσματικός» αριθμός των αλληλουχιών στην πολλαπλή στοίχιση μετά την στάθμιση στις κατάλληλες θέσεις. Ο όρος $\lambda$ χρησιμοποιείται για την κανονικοποίηση των δεδομένων σε περιπτώσεις πεπερασμένων συνόλων δεδομένων. Αριθμητικές δοκιμές σε μια πληθώρα διαφορετικών πολλαπλών στοιχίσεων οδήγησαν στο να οριστούν τιμές για το $x{\sim}0.7$ και για το $\lambda{\sim}0.5$ (Marks et al., 2011). (Η απόδοση της τιμής στη παράμετρο $x{\sim}0.7$ πρακτικά σημαίνει ότι σε όλες οι ακολουθίες σε μια πολλαπλή στοίχιση που έχουν ποσοστό ομοιότητας μικρότερο από 30% σε σχέση με μια ακολουθία αναφοράς *m,* αποδίδεται τιμή βάρους ίση με 1).

Η στατιστική μέθοδος mfDCA, βασίζεται στο μοντέλο της μέγιστης εντροπίας (maximum entropy) με στόχο να ταυτοποιηθεί ένα ελάχιστο σύνολο συζευγμένων ζευγαριών, ως απόρροια των υπολογισμών των συσχετίσεων μεταξύ των ζευγών των αμινοξικών καταλοίπων του ολικού στατιστικού μοντέλου. Στη συνέχεια από το σύνολο των ζευγαριών αυτών εξάγονται πληροφορίες σχετικά με την αλληλεπίδραση των καταλοίπων. Η θεώρηση αυτή είναι εννοιολογικά διαφορετική σε σχέση με άλλες προσεγγίσεις που υπολογίζουν τη συσχέτιση κάθε ζεύγους αμινοξικών καταλοίπων ανεξάρτητα.

Το στατιστικό μοντέλο για κάθε οικογένεια πρωτεϊνών περιγράφει την πιθανότητα ύπαρξης της αμινοξικής ακολουθίας οποιουδήποτε μέλους της οικογένειας ως την από κοινού πιθανότητα κατανομής $P(A_1, \dots, A_L)$. Στην πραγματικότητα η εκτίμηση μιας τέτοιας συνάρτησης, δηλαδή ο προσδιορισμός της πιθανότερης κατάστασης του συστήματος με βάση τα δεδομένα που έχουν ήδη υπολογιστεί είναι πρακτικά αδύνατος καθώς ο αριθμός των παραμέτρων του πρέπει να εκτιμηθούν είναι $q^L$.

Το πρόβλημα αυτό απλοποιείται αν ο στόχος περιοριστεί στο να γίνει εκτίμηση 2 πιθανοτήτων κατανομής, μιας που να περιγράφει τις συχνότητες των μεμονωμένων αμινοξικών καταλοίπων και μιας που να περιγράφει της συχνότητες των ζευγαριών των αμινοξικών καταλοίπων που παρατηρούνται στην πολλαπλή στοίχιση. Έτσι προκύπτουν δυο κατανομές πιθανοτήτων:

$$\boldsymbol{P_i(A_i) \equiv \sum_{\{A_k = 1, \dots, q\} | k \neq i} P(A_1, \dots, A_L) = f_i(A_i)}$$

$$\boldsymbol{P_{ij}(A_i, A_j)} \equiv \sum_{\{\boldsymbol{A_k=1,\dots,q}\}|\boldsymbol{k\neq i,j}} \boldsymbol{P(A_1, \dots, A_L) = f_{ij}(A_i, A_j)}$$

όπου το άθροισμα υπολογίζεται με βάση όλες τις πιθανές ακολουθίες, για παράδειγμα με βάση όλες της πιθανές τιμές $A_k$ για τα αμινοξικά κατάλοιπα στη θέση *k*, εκτός από αυτά που παραλείπονται με βάση τους περιορισμούς που προκύπτουν από το αριστερό μέλος της εξίσωσης κάθε φορά. Επί της αρχής, με τη μεθοδολογία αυτή θα μπορούσαν να υπολογιστούν και συσχετίσεις μεταξύ 3άδων καταλοίπων, ή γενικότερα μεταξύ k-καταλοίπων. Παρόλα αυτά, ακόμα και για $k = 3$, ο αριθμός των παραμέτρων που πρέπει να εκτιμηθούν αυξάνεται δραματικά, με βάση την εξίσωση $\sim\binom{L}{3}q^3$, όπου L είναι το μήκος της ακολουθίας.

Υπάρχει ένας μεγάλος αριθμός κατανομών πιθανοτήτων, που συνάδουν με τα δεδομένα αυτά, και στην υλοποίηση της μεθόδου mfDCA επιλέχθηκε το μοντέλο της **μέγιστης εντροπίας *S***.

Η εντροπία μιας τυχαίας μεταβλητής υποδηλώνει πόσο αβέβαιοι είμαστε για την τιμή της ή ισοδύναμα πόση είναι η αναμενόμενη ελάχιστη ποσότητα πληροφορίας που πρέπει να μας δοθεί για να γνωρίζουμε με βεβαιότητα την τιμή μιας τυχαίας μεταβλητής. Η εντροπία υπολογίζεται από την εξίσωση:

$$\boldsymbol{S} = -\sum_{\{\boldsymbol{A_i}|\boldsymbol{i=1,\dots,L}\}} \boldsymbol{P(A_{1,\dots,}\, A_L)\ln\,(A_1, \dots, A_L)}$$

Η από κοινού κατανομή πιθανοτήτων δίνεται από την εξίσωση:

$$\boldsymbol{P(A_1, \dots, A_L) = \frac{1}{Z} exp\left\{\sum_{1\leq i\leq j\leq L} e_{ij}\ A_i, A_j + \sum_{1\leq i\leq L} h_i\,(A_i)\right\}}$$

όπου

$$\boldsymbol{Z = \sum_{\{A_i|i=1,\dots,L\}} exp\left\{\sum_{1\leq i\leq j\leq L} e_{ij}\,(A_i, A_j) + \sum_{1\leq i\leq L} h_i\,(A_i)\right\}}$$

Οι πολλαπλασιαστές Lagrange $h_i(A_i)$ και $e_{ij}(A_i, A_j)$ εισάγονται έτσι ώστε να συμβαδίζει η από κοινού κατανομή πιθανοτήτων με τις εμπειρικές κατανομές συχνοτήτων για κάθε ένα αμινοξικό κατάλοιπο ξεχωριστά και για τα ζεύγη των καταλοίπων. Οι τιμές των πολλαπλασιαστών Lagrange $h_i(A_i)$ και $e_{ij}(A_i, A_j)$ επιλέγονται έτσι ώστε να ικανοποιούνται οι περιορισμοί των εξισώσεων 15 και 16 . Η παράμετρος Z, είναι μια σταθερά κανονικοποίησης που εξαρτάται μόνο από τις παραμέτρους του εφαρμοζόμενου στατιστικού μοντέλου. Η πρόκληση έτσι ώστε να ικανοποιούνται οι περιορισμοί που αναφέρθηκαν είναι να εκτιμηθούν ικανοποιητικά οι περιθώριες συναρτήσεις για τα μεμονωμένα κατάλοιπα και για τα

ζευγάρια των αμινοξικών καταλοίπων και η συνάρτηση διαμέρισης Z. Οι περιθώριες συναρτήσεις της συγκεκριμένης κατανομής δίνονται από τους τύπους :

$$\frac{\partial lnZ}{\partial h_i(A_i)} = -P_i(A_i)$$

και

$$\frac{\partial^2 lnZ}{\partial h_i(A_i)\partial h_j(A_j)} = -P_{ij}(A_i, A_j) + P_i(A_i)P_j(A_j)$$

Όμως, ο άμεσος υπολογισμός των δυο παραπάνω εξισώσεων εξακολουθεί να είναι υπολογιστικά απαγορευτικός και γι' αυτό γίνονται κάποιες απλοποιήσεις χωρίς το μοντέλο να χάνει τον ολικό του χαρακτήρα. Οι απλοποιήσεις αφορούν στους πολλαπλασιαστές Lagrange $h_i(q)$, $e_{ij}(A,q)$και $e_{ij}(q,A)$ , στους οποίους και ανατίθεται η τιμή μηδέν για τις θέσεις στην ακολουθία $1 \leq i \leq j \leq L$ και για τα αμινοξικά κατάλοιπα A={1…q}.

Στην πραγματικότητα για να υπολογιστούν τα ζευγάρια των αμινοξικών καταλοίπων που συσχετίζονται πραγματικά με βάση το μοντέλο της **μέγιστης εντροπίας** , η πρώτη προϋπόθεση είναι να υπάρχει μέγιστη συμφωνία μεταξύ των αναμενόμενων τιμών των συχνοτήτων των ζευγαριών των αμινοξικών καταλοίπων (όπως αυτές υπολογίζονται από τις περιθώριες συναρτήσεις) με τις παρατηρούμενες συχνότητες. Η δεύτερη προϋπόθεση είναι ότι η μέγιστη εντροπία της ολικής κατανομής πιθανότητας του μοντέλου, η οποία εξασφαλίζει τη μέγιστη ομοιομορφία του υπολογιζόμενου μοντέλου, να μπορεί να ικανοποιηθεί χωρίς να παραβιάζεται η πρώτη προϋπόθεση. Στην πραγματικότητα, ένα τέτοιο στατιστικό μοντέλο είναι παρόμοιο με τα στατιστικά μοντέλα που χρησιμοποιούνται στη στατιστική φυσική με στόχο να εκφράσουν την πιθανότητα να διαμορφωθεί ένα σύστημα πολλαπλών σωματιδίων, η οποία και προσεγγίζεται με βάση τη συνάρτηση του Hamilton ως το άθροισμα των ενεργειών αλληλεπίδρασης μεταξύ ενός ζεύγους σωματιδίων και των ενεργειών των μεμονωμένων σωματιδίων όταν σε αυτά εφαρμόζεται ένα εξωτερικό πεδίο. Αν κάνουμε την αναλογία στο επίπεδο των πολλαπλών στοιχίσεων, μπορούμε να πούμε ότι μια θέση στην ακολουθία i αντιστοιχεί σε ένα σωματίδιο και μπορεί να λάβει μια από τις 21 επιτρεπόμενες τιμές, ενώ το ζεύγος των αμινοξικών καταλοίπων i,j αντιστοιχεί σε ένα ζεύγος αλληλεπιδρώντων σωματιδίων.

Επειδή, όπως αναφέρθηκε και παραπάνω, ο υπολογιστικός προσδιορισμός του μεγάλου αριθμού των παραμέτρων $e_{ij}(A_i, A_j)$ και $h_i(A_i)$ που να ικανοποιούν τις προϋποθέσεις του ολικού στατιστικού μοντέλου παραμένει δύσκολος, στη στατιστική μέθοδο mfDCA, εφαρμόζεται η προσέγγιση του μέσου πεδίου (**mean – field approximation**).

Με βάση την προσέγγιση αυτή, τα πραγματικά συσχετιζόμενα κατάλοιπα δίνονται από τον την αντίστροφή του πίνακα συνδιακύμανσης :

$$\boldsymbol{C_{ij}(A_i, A_j) = f_{ij}(A_i, A_j) - f_i(A_i) f_j(A_j)}$$

ως εξής

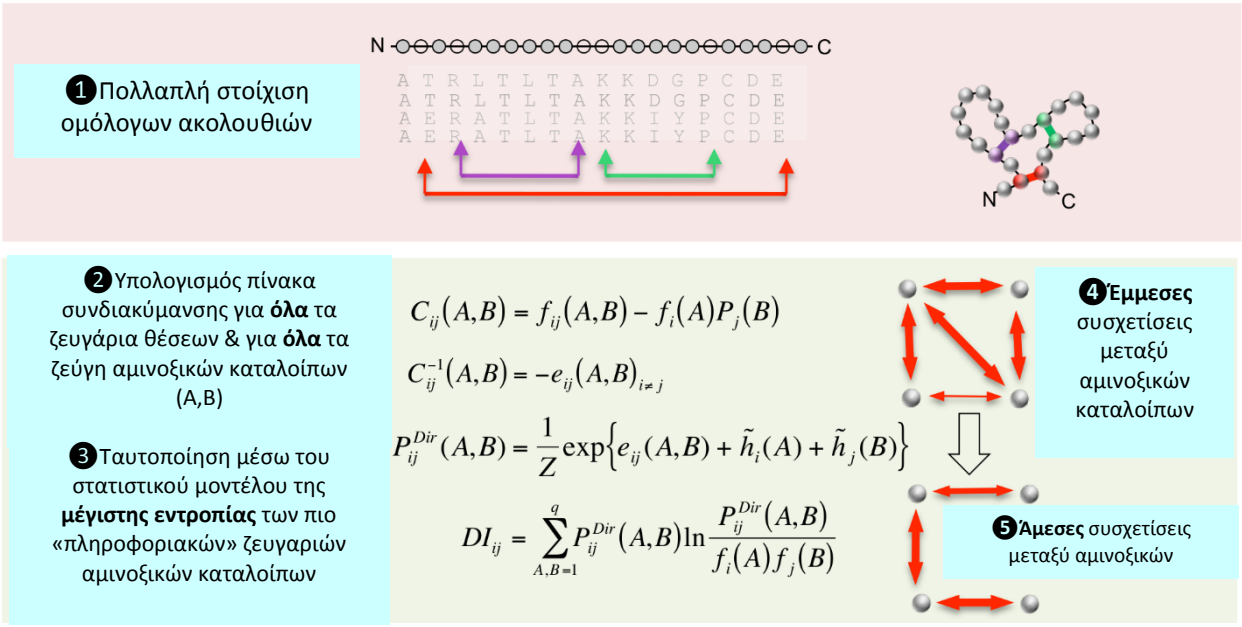
$$\boldsymbol{e_{ij}(A_i, A_j) = -(C^{-1})ij(A_i, A_j)}$$

Η εκτίμηση της παραμέτρου $e_{ij}(A_i, A_j)$ είναι αυτή που τελικά επιτρέπει την κατάταξη των ζευγαριών των αμινοξικών καταλοίπων ανάλογα με τις τιμές της συσχέτισής τους μέσω της εκτίμησης της πιθανότητας κατανομής της **άμεσης πληροφορίας** (Direct Information)

$$\boldsymbol{P_{ij}^{Dir}(A_i, A_j) = \frac{1}{Z} exp\{e_{ij}(A_i, A_j + \hbar_i(A_i) + \hbar_j(A_j)\}}$$

Τελικά, ο αλγόριθμος υπολογίζει την **άμεση πληροφορία** μεταξύ 2 καταλοίπων i και j ως τη σχετική εντροπία μεταξύ των κατανομών $P_{ij}^{Dir}$ και της κατανομής $f_i f_j$ για κάθε ανεξάρτητη θέση ως εξής :

$$\boldsymbol{DI_{ij} = \sum_{A_i, A_j=1}^{q} P_{ij}^{Dir}(A_i, A_j) ln\left(\frac{P_{ij}^{Dir}(A_i, A_j)}{f_i(A_i) f_j(A_j)}\right)}$$




**Εικόνα 1.60** Συνοπτική σχηματική παρουσίαση της στατιστικής μεθόδου mfDCA (Marks et al., 2011)

### 1.3.5.3. Μέθοδος plmDCA (Pseudo-likelihood Direct Coupling Analysis)

Ο σκοπός της εκτίμησης κατάστασης των στατιστικών μοντέλων είναι να προσδιοριστεί η πιθανότερη κατάσταση του συστήματος με βάση τις παρατηρηθείσες τιμές. Ένας τρόπος να επιτευχθεί αυτό, είναι χρησιμοποιώντας την Εκτίμηση Μέγιστης Πιθανοφάνειας (Maximum Likelihood Estimation ή MLE), μιας μεθόδου που εφαρμόζεται ευρύτατα στη στατιστική. Τα σφάλματα των μετρήσεων, θεωρείται ότι ακολουθούν μια γνωστή κατανομή πιθανότητας με άγνωστες παραμέτρους. Η κοινή συνάρτηση πυκνότητας πιθανότητας όλων των μετρήσεων μπορεί να γραφεί σαν συνάρτηση παραμέτρων. Αυτή η συνάρτηση ονομάζεται συνάρτηση πιθανοφάνειας και μεγιστοποιείται όταν οι άγνωστοι παράμετροι βρίσκονται όσο πιο κοντά γίνεται στις πραγματικές τους τιμές. Ως εκ τούτου, μπορεί να οριστεί ένα πρόβλημα βελτιστοποίησης ώστε να μεγιστοποιηθεί η συνάρτηση πιθανοφάνειας. Η λύση θα δώσει την εκτίμηση μέγιστης πιθανοφάνειας για τις ζητούμενες παραμέτρους. Το συγκεκριμένο ολικό στατιστικό μοντέλο που εφαρμόζεται από τη στατιστική μέθοδο plmDCA, αντικαθιστά την από κοινού κατανομή πιθανοτήτων με την δεσμευμένη/υπό συνθήκες πιθανότητα να παρατηρήσουμε μια μεταβλητή $A_r$ δεδομένων των παρατηρήσεων για όλες τις υπόλοιπες μεταβλητές $A_{\setminus r}$ με βάση την παρακάτω εξίσωση:

$$\boldsymbol{P(A_r = A_r^m | A_{\setminus r} = A_{\setminus r}^m) = \frac{\exp\left(h_r(A_r^m + \sum_{i=1, i\neq r}^{L} e_{ri}(A_r^m, A_i^m)\right)}{\sum_{A_j=1}^{q} h_r\left(A_j + \sum_{i=1, i\neq r}^{L} e_{ri}(A_j^m, A_i^m)\right)}}$$

όπου για λόγους απλοποίησης τίθεται $e_{ri}(A_r^m, A_i^m) = e_{ri}(A_j^m, A_i^m)$ όταν $i < r$. Έτσι, μπορούμε να μεγιστοποιήσουμε την υπό συνθήκες πιθανοφάνεια ελαχιστοποιώντας τη συνάρτηση :

$$\boldsymbol{g_r(h_r, e_r) = -\frac{1}{M}\sum_{m=0}^{M} log\left[P\{h_r, e_r\}(A_r = A_r^m | A_{\setminus r} = A_{\setminus r}^m)\right]}$$

Η αντικατάσταση της από κοινού κατανομής πιθανότητας από τη δεσμευμένη πιθανότητα επιλύει την υπολογιστικά δύσκολη εκτίμηση του μεγάλου αριθμού παραμέτρων του μοντέλου της μέγιστης εντροπίας, με αποτέλεσμα να μην χρειάζεται ο υπολογισμός της συνάρτησης διαμέρισης Z και των περιθώριων συναρτήσεων, αλλά να είναι απαραίτητη μόνο η βελτιστοποίηση της υπό συνθήκες πιθανότητας με βάση τις 21 πιθανές τιμές που μπορεί να λάβει μια μεταβλητή σε μια πολλαπλή στοίχιση. Τελικά ο αλγόριθμος υπολογίζει την **άμεση πληροφορία** μεταξύ 2 καταλοίπων i και j ως τη σχετική εντροπία μεταξύ των κατανομών $P_{ij}^{Dir}$ και της κατανομής $f_i f_j$ για κάθε ανεξάρτητη θέση.

Στο σημείο αυτό, κρίνεται σκόπιμο να αναφερθεί, ότι καθώς και η μέθοδος mfDCA και η μέθοδος plmDCA, είναι στατιστικές μέθοδοι, βιβλιογραφικά προτείνεται ότι πρέπει να αξιολογούνται ως «αληθώς θετικά» το ~3% των συνολικών ζευγαριών των αλληλεπιδρώντων καταλοίπων (Ekeberg et al., 2013; Hopf et al., 2014, 2015; Marks et al., 2012).

### 1.3.5.4. Μέθοδος PSICOV

Η μέθοδος PSICOV (**P**resice **S**tructural Contact Prediction using sparse **I**nverse **Cov**ariance estimation) ανήκει στα ολικά στατιστικά μοντέλα και η εκτίμηση των αλληλεπιδρώντων καταλοίπων μέσω του μοντέλου της αντιστροφής του πίνακα συνδιασποράς των αμινοξικών καταλοίπων.

Δημιουργείται ένα πίνακας $21m\ x\ 21\ n$, όπου m είναι ο αριθμός των στηλών, δηλαδή των θέσεων των αμινοξικών καταλοίπων και n ο αριθμός των γραμμών, δηλαδή των πρωτεϊνικών ακολουθιών της πολλαπλής στοίχισης.

Από τον αρχικό πίνακα αυτό μπορεί να υπολογιστεί ένας **πίνακας συνδιασποράς** (covariance matrix). Ο πίνακας συνδιασποράς δείχνει την τάση των ζευγαριών των διάφορων στοιχείων του διανύσματος να συν-μεταβάλλονται. Ο πίνακας συνδιασποράς ορίζεται ως εξής :

$$\boldsymbol{S_{ij}^{ab}} = \frac{\mathbf{1}}{\boldsymbol{n}}\sum_{\boldsymbol{k=1}}^{\boldsymbol{n}}(\boldsymbol{x_i^{ak}} - \overline{\boldsymbol{x_i^{a}}})\,(\boldsymbol{x_j^{bk}} - \overline{\boldsymbol{x_j^{b}}})$$

όπου $\boldsymbol{x_i^{ak}}$ είναι μια δυαδική μεταβλητή $(x \in \{0,1\})$ που υποδυκνύει την παρουσία ή την απουσία του αμινοξέος α στη θέση i και στην ακολουθία k και αντίστοιχα η μεταβλητή $\boldsymbol{x_j^{bk}}$ υποδυκνεί την παρουσία ή απουσία του αμινοξέος b στη θέση j και στην ακολουθία k. Αυτός ο υπολογισμός της συνδιασποράς, που βασίζεται σε δυαδικές μεταβλητές για τα αμινοξικά κατάλοιπα είναι παρόμοιος με αυτόν που χρησιμοποιήθηκε από τους Halabi et al. (2009) για να εντοπίσουν ανεξάρτητες εξελικτικές μονάδες. Με βάση τον κλασικό ορισμό της συνδιασποράς

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

και με βάση ότι η αναμενόμενη μέση τιμή μιας δυαδικής μεταβλητής $E(X)$ θα είναι ίση με την πιθανότητα αληθούς παρατήρησης $p(x = 1)$, απλοποιείται ο παραπάνω τύπος με βάση τις παρατηρούμενες περιθώριες συχνότητες των αμινοξέων $(f(A_i)$ και $f(B_j))$ και τις παρατηρούμενες περιθώριες συχνότητες των ζευγών των αμινοξικών καταλοίπων $(f(A_iB_j))$ για δεδομένες θέσεις τις πολλαπλής στοίχισης ως εξής :

$$\boldsymbol{S_{ij}^{ab}} = \boldsymbol{E}\left(\boldsymbol{x_i^a x_j^b}\right) - \boldsymbol{E}\left(\boldsymbol{x_i^a}\right)\boldsymbol{E}\left(\boldsymbol{x_j^b}\right) = \boldsymbol{f}\left(\boldsymbol{A_iB_j}\right) - \boldsymbol{f}(\boldsymbol{A_i})(\boldsymbol{B_j})$$

Κάθε μεμονωμένο στοιχείο του παραπάνω πίνακα δίνει τη συνδιασπορά του αμινοξικού καταλοίπου α στη θέση i με το αμινοξικό κατάλοιπο b στη θέση j. Υπολογίζοντας τον αντίστροφο πίνακα του πίνακα συνδιασποράς, καταλήγουμε

στον πίνακα Θ, (precision ή concentration matrix), από τον οποίο και μπορούν να εξαχθούν μερικοί συντελεστές συσχέτισης (partial correlation coefficients) για όλα τα ζευγάρια μεταβλητών ως εξής:

$$\boldsymbol{\rho_{ij}} = -\frac{\boldsymbol{\Theta_{ij}}}{\sqrt{\boldsymbol{\Theta_{ii}\Theta_{jj}}}}$$

Στη συγκεκριμένη περίπτωση ο πίνακας των μερικών συντελεστών συσχέτισης, δίνει τη συσχέτιση μεταξύ κάθε ζευγαριού αμινοξικών καταλοίπων σε δυο οποιεσδήποτε θέσεις της πολλαπλής στοίχισης με βάση και σε συμφωνία με τις παρατηρούμενες συχνότητες των αμινοξέων σε όλες τις υπόλοιπες θέσεις της πολλαπλής στοίχισης. Έτσι, υποθέτοντας ότι ο πίνακας συνδιασποράς μπορεί να αντιστραφεί, ο αντίστροφος πίνακας συνδιασποράς παρέχει πληροφορίες σχετικά με το βαθμό της άμεσης συσχέτισης ζευγαριών αμινοξικών καταλοίπων στις θέσεις της πολλαπλής στοίχισης. Εξαιρουμένων των στοιχείων της διαγωνίου του πίνακα, οι τιμές των υπόλοιπων στοιχείων που αποκλίνουν σημαντικά από το μηδέν υποδεικνύουν ζευγάρια αμινοξικών καταλοίπων που αλληλεπιδρούν. Ένα σημαντικό μειονέκτημα της συγκεκριμένης μεθόδου είναι ότι απαιτείται πολύ μεγάλος αριθμός αλληλουχιών στις πολλαπλές στοιχίσεις (>1.000) (Jones et al., 2012).

# 1.4. Σκοπός

Στόχος της παρούσας διπλωματικής εργασίας είναι η υπολογιστική ανάλυση των αλληλεπιδράσεων των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης με μεθόδους που εκμεταλλεύονται τη συνεξελικτική πληροφορία. Ο ακριβής **μηχανισμός αλληλεπίδρασης μεταξύ των κινασών και της CaM κατά την αποδέσμευση της αυτοτελούς δομικής περιοχής της κινάσης από το ενεργό κέντρο αυτής και την επακόλουθη ενεργοποίηση των κινασών** παραμένει σήμερα ένα άλυτος γρίφος καθώς τα δομικά δεδομένα που αφορούν την ομάδα των μορίων αυτών είναι φτωχά.

Για να ενεργοποιηθούν οι εξαρτώμενες από καλμοντουλίνη κινάσες απαιτούν την πρόσδεση, ακόμα και παροδικά, της καλμοντουλίνης, ώστε να απελευθερωθεί η αυτοτελής δομική περιοχή αυτοαναστολής. Τα σύμπλοκα $Ca^{2+}$-CaM-κινάσες ρυθμίζουν πολύ διαφορετικές κυτταρικές λειτουργίες, όπως η έκφραση γονιδίων, ο κυτταρικός θάνατο (απόπτωση), η αναδιοργάνωση του κυτταροσκελετού, η μάθηση, η μνήμη και πολλές ακόμα. Η μελέτη του τρόπου αλληλεπίδρασης των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης μπορεί να παίξει καθοριστικό ρόλο στην αποσαφήνιση των μηχανισμών απόκρισης των κυττάρων σε αλλαγές της συγκέντρωσης του ασβεστίου αλλά και της ρύθμισης των λειτουργιών του οργανισμού γενικότερα.

Επιπλέον όπως αναφέρθηκε και εισαγωγικά, δεν έχει καταγραφεί κάποια συντηρημένη ακολουθία αναγνώρισης των πρωτεϊνών στόχων από την καλμοντουλίνη, όμως τα σύμπλοκα που δημιουργούνται εμφανίζουν υψηλή εξειδίκευση και συγγένεια αλληλεπίδρασης. Η μεγάλη στερεοδιατακτική διαφορά της $Ca^{2+}$-CaM σε σχέση με την apo-CaM αυξάνει την πολυπλοκότητα του τρόπου με τον οποίο η καλμοντουλίνη αναγνωρίζει την αυτοτελή δομική περιοχή πρόσδεσής της στις κινάσες, καθώς για τη δημιουργία του συμπλόκου απαιτούνται σημαντικές στερεοδιατακτικές αλλαγές τόσο στην CaM όσο και στις κινάσες ώστε να επιτευχθεί ενεργοποίηση των τελευταίων. Επιπλέον, είτε επειδή αρκετές από τις κινάσες είναι είτε μεγάλα μόρια, είτε επειδή σχηματίζουν πολυμερή είναι δύσκολη η μελέτη των συμπλόκων σε μοριακό επίπεδο. Κατά συνέπεια, η διαθέσιμη δομική πληροφορία έχει αποκτηθεί κυρίως από σύμπλοκα της CaM με μικρά πεπτίδια (α-έλικες) των κινασών.

# 2.Υλικά και μέθοδοι

## 2.1. Κατασκευή του συνόλου δεδομένων

UniprotKB: Οι πρωτεϊνικές ακολουθίες για τις εξαρτώμενες από καλμοντουλίνη κινάσες και τις CaM συλλέχθηκαν από τη βάση δεδομένων πρωτεϊνικών ακολουθιών UniProtKB/SwissProt (https://www.uniprot.org/)

Η UniprotKB (Uniprot Knowledgebase, http://www.uniprot.org/), αποτελείται από δύο υποσύνολα:

Uniprot/SwissProt : περιέχει 558,125 ακολουθίες (Rel. 2018_08) οι οποίες έχουν περάσει από κάποιου είδους έλεγχο και συνοδεύονται από συμπληρωματικά σχόλια όπως, βιβλιογραφικές αναφορές, γενικά στοιχεία δευτεροταγούς δομής, σύνδεσμοι σε άλλες βάσεις δεδομένων σχετικές με κάθε εγγραφή καθώς και σημειώσεις για τη βιολογική λειτουργία (αν είναι γνωστές) καθώς και άλλες χρήσιμες πληροφορίες και

Uniprot/TrEMBL : περιέχει σήμερα (Rel. 2018_08) 124,797,108 ακολουθίες η οποίες όμως δεν έχουν υποστεί ανθρώπινο σχολιασμό και έχουν προκύψει από αυτόματη (ηλεκτρονική) μετάφραση γονιδιωματικών αλληλουχιών.

Η συλλογή του συνόλου δεδομένων για τις κινάσες έγινε με βάση τις οικογένειες των κινασών, όπως βρίσκονται κατατεθειμένες στη βάση δεδομένων kinase.com (Manning et al., 2002). Όπως αναφέρθηκε και εισαγωγικά, οι ευκαρυωτικές κινάσες έχουν κατηγοριοποιηθεί σε οικογένειες με βάση την ομοιότητα της αμινοξικής τους ακολουθίας, επιπλέον πληροφορίες από δομικές αυτοτελείς περιοχές εκτός του καταλυτικού κέντρου, την εξελικτική τους συντήρηση και όλες τις γνωστές λειτουργίες στις οποίες συμμετέχουν. Αυτή η υβριδική κατηγοριοποίηση χωρίζει τις κινάσες σε 11 υπεροικογένειες (AGC, CMGC, CAMK, CK1, STE, Tyrosine Kinase (TK), Tyrosine Kinase-Like (TKL), RGC, PKL, AtypicalKinases και άλλες κινάσες.

Με βάση την κατηγοριοποίηση αυτή η υπεροικογένεια των CaMK αποτελείται από τα εξής μέλη:

τις CAMKI, CAMKII, MLCK, DAPK και PHK (Phosphorylase Kinase) που αποτέλεσαν μέρος του συνόλου δεδομένων και αναλύθηκαν εκτενώς εισαγωγικά,

και τις CAMKL (μια εκτενής οικογένεια που περιλαμβάνει αρκετές συντηρημένες υποοικογένειες όπως οι AMPK, MARK, CHK1 και LKB), τις MAPKAPK (MAPK-Activated Protein Kinase), Trb (Tribbles kinases, έχουν όλες προβλεφθεί ως καταλυτικά ανενεργές), PKD (Protein Kinase D), CAMK-Unique (περιλαμβάνει κινάσες που δε μπορούν να καταταχθούν σε κάποια από τις υπόλοιπες οικογένειες), CASK, DCAMKL, PIM, PSK, RAD53 (γνωστή και ως CHK2, συμμετέχει στα σημεία

ελέγχου του κυτταρικού κύκλου), RSKb (η δεύτερη δομικά αυτοτελής περιοχή των RSK κινασών), Trio (κινάσης που σχετίζονται με τη λειτουργία των μυών), TSSK (κινάση καταλοίπου σερίνης που εμφανίζεται αποκλειστικά στους όρχεις), CDPK (οι εξαρτώμενες από καλμοντουλίνη κινάσες που εντοπίζονται στα φυτά), CAMK-Tt (εντοπίζονται στην Tetrahymena και σε άλλα βλεφαριδοφόρα πρωτόζωα), NNK1 (Nitrogen Network Kinase 1 - εντοπίζονται στους μύκητες), PRR1 (Pheromone Response Regulator 1 – εντοπίζονται στους μύκητες).

Συνεπώς η πρώτη αναζήτηση πραγματοποιήθηκε έναντι της Swiss-Prot με βάση τα κριτήρια κατάταξης των κινασών σε οικογένειες και αναζήτηση με family name : CAMK Ser/Thr protein kinase family

Ακολούθησε **δεύτερη αναζήτηση στην Swiss-Prot με βάση το E.C. – Enzyme Commission number** με όνομα κατηγορίας Calcium/calmodulin-dependent protein kinase και κωδικό 2.7.11.17 (enzyme.expasy.org/EC/2.7.11.17) που αποτελεί κατηγοριοποίηση των ενζύμων με βάση τις αντιδράσεις που καταλύουν (αν διαφορετικά ένζυμα καταλύουν την ίδια αντίδραση θα διαθέτουν τον ίδιο E.C.).

Τα αποτελέσματα **φιλτραρίστηκαν με την βοήθεια της βάσης δεδομένων Gene Ontology**, με χρήση της κλάσης Molecular Function και του όρου Calmodulin binding (GO:0005516).

**Gene ontology** (GO): αποτελεί μια πρωτοβουλία με στόχο την ενιαία παρουσίαση των γονιδίων και των γονιδιακών προϊόντων σε όλα τα βιολογικά είδη οργανισμών. Το πρόγραμμα αποσκοπεί: 1) στη διατήρηση και ανάπτυξη ελεγχόμενου λεξιλογίου για τα χαρακτηριστικά των γονιδίων και των γονιδιακών προϊόντων, 2) στο σχολιασμό των γονιδίων και των γονιδιακών προϊόντων και στην αφομοίωση και τη διάδοση των δεδομένων που αφορούν το σχολιασμό και 3) στην ανάπτυξη εργαλείων για εύκολη πρόσβαση στα δεδομένα που παρέχονται και για να καταστεί δυνατή η λειτουργική ερμηνεία των πειραματικών δεδομένων με χρήση της GO.

Αν και η ίδια η ονοματολογία των γονιδίων στοχεύει στη διατήρηση και την ανάπτυξη ελεγχόμενου λεξιλογίου γονιδίων και γονιδιακών προϊόντων, η GO επεκτείνει την προσπάθεια χρησιμοποιώντας σημασμένη γλώσσα ώστε τα δεδομένα (όχι μόνο αυτά που αφορούν γονίδια και τα προϊόντα τους αλλά και το σύνολο των χαρακτηριστικών τους) να είναι αναγνώσιμα από μηχανή. Αυτό επιτυγχάνεται με έναν τρόπο ομοιόμορφο για όλα τα είδη (ενώ συχνά η ονοματολογία των γονιδίων ποικίλλει ανάλογα με την ταξινομική βαθμίδα). Οι οντολογίες αποτελούν αναπαραστάσεις καταστάσεων που είναι ανιχνεύσιμες ή άμεσα παρατηρήσιμες και των σχέσεων μεταξύ αυτών των καταστάσεων. Ελλείψει παγκόσμιας καθιερωμένης ορολογίας στην επιστήμη της βιολογίας η επικοινωνία και η ανταλλαγή δεδομένων καθίσταται δυσκολότερη. Συχνά η ορολογία είναι συγκεκριμένη για ένα είδος, έναν τομέα έρευνας, ακόμα και μιας ερευνητικής

ομάδας. Το πρόγραμμα GO παρέχει μια οντολογία καθορισμένων όρων που αντιπροσωπεύουν τις ιδιότητες των γονιδιακών προϊόντων. Η οντολογία καλύπτει τρεις τομείς:

I. cellular component: οι περιοχές ενός κυττάρου ή το εξωκυτταρικό του περιβάλλον.

II. molecular function: οι στοιχειώδεις δραστηριότητες ενός γονιδιακού προϊόντος σε μοριακό επίπεδο, όπως η αλληλεπίδραση με άλλα μόρια ή η κατάλυση και

III. biological process: λειτουργίες ή σύνολα μοριακών γεγονότων με καθορισμένη αρχή και τέλος, που σχετίζονται με τη λειτουργία ολοκληρωμένων μονάδων που εμφανίζουν το φαινόμενο της ζωής, όπως κύτταρα, ιστοί, όργανα και οργανισμοί.

Κάθε όρος της GO εντός της κάθε οντολογίας έχει ένα όνομα, το οποίο μπορεί να είναι μια λέξη ή μια σειρά λέξεων, ένα μοναδικό αλφαριθμητικό αναγνωριστικό, έναν ορισμό με αναφερόμενες πηγές και με τη μορφή π.χ. molecular_function υποδεικνύεται ο τομέας στον οποίο ανήκει. Οι όροι μπορεί επίσης να έχουν συνώνυμους όρους, οι οποίοι ταξινομούνται ως ακριβώς ισοδύναμοι με το όνομα, ευρύτεροι, στενότεροι ή συγγενικοί και παρέχονται παραπομπές άλλες βάσεις δεδομένων. Η GO είναι δομημένη ως κατευθυνόμενος ακυκλικός γράφος και κάθε όρος έχει καθορισμένες σχέσεις με έναν ή περισσότερους όρους στον ίδιο τομέα και μερικές φορές σε άλλους τομείς. Το λεξιλόγιο της GO έχει σχεδιαστεί για να είναι ουδέτερο για όλα είδη και οι όροι του μπορούν να εφαρμοστούν τόσο για προκαρυωτικούς όσο και για ευκαρυωτικούς οργανισμούς, μονοκύτταρους ή πολυκύτταρους (Ashburner et al., 2000).

Οι ακολουθίες των κινασών επεξεργάστηκαν περαιτέρω ώστε να απορριφθούν από την ανάλυση περιοχές των κινασών που δε δύναται να συμμετέχουν στην αλληλεπίδρασή τους με τις CaM.

Ακολούθησε **κατασκευή του συνόλου δεδομένων για τις CaM.** Η αναζήτηση πραγματοποιήθηκε με βάση protein family: calmodulin και protein name: calmodulin. Πραγματοποιήθηκε **περαιτέρω επεξεργασία του συνόλου με βάση τον οργανισμό**.

**CD-HIT** : Τέλος, οι ακολουθίες φορτώθηκαν σε fasta format στην πλατφόρμα CD-HIT (W. Li & Godzik, 2006). Εφαρμόστηκε το πρόγραμμα CD-HIT που ομαδοποιεί πρωτεϊνικές ακολουθίες με βάση ένα κατώφλι ομοιότητας των αμινοξικών καταλοίπων (ορίστηκε σε 0.2). Σε κάθε ομάδα εντοπίζεται μια αντιπροσωπευτική ακολουθία. Ως είσοδο δέχεται ένα σύνολο δεδομένων πρωτεϊνικών ακολουθιών σε fasta μορφή. Ως έξοδο παρέχει ένα fasta αρχείο με τις αντιπροσωπευτικές ακολουθίες και ένα αρχείο κειμένου με την λίστα των ομάδων. Για τις υπόλοιπες παραμέτρους επιλέχθηκαν οι default τιμές.

## 2.2. Πολλαπλές στοιχίσεις ακολουθιών

Το επόμενο βήμα είναι **η πολλαπλή στοίχιση των κινασών και των καλμοντουλινών, με χρήση του προγράμματος Clustal Omega** (Sievers et al., 2011).

Πολλαπλή στοίχιση στην ουσία είναι η στοίχιση ομάδων ακολουθιών. Οι πολλαπλές στοιχίσεις διαθέτουν, όπως είναι λογικό, μεγαλύτερη ευαισθησία από ότι η στοίχιση ακολουθιών κατά ζεύγη, στην εύρεση μακρινών ομολόγων σε βάσεις δεδομένων. Οι πιο πρακτικές και ακριβείς μέθοδοι που χρησιμοποιούνται συνήθως σε προβλήματα πολλαπλής στοίχισης είναι οι ιεραρχικές μέθοδοι οι οποίες ολοκληρώνονται σε τρία βήματα. Αρχικά πραγματοποιείται στοίχιση κατά ζεύγη για όλες τις ακολουθίες οι οποίες πρόκειται να στοιχηθούν, συνήθως με τον αλγόριθμο Needleman και Wunsch. Με βάση τα αποτελέσματα του αλγόριθμου δημιουργείται ένα δενδρικό διάγραμμα (guide tree) με όλες τις ακολουθίες. Τέλος η πολλαπλή στοίχιση πραγματοποιείται με διαδοχικά βήματα ανάλογα με το τι θέση καταλαμβάνουν οι ακολουθίες στο δενδρικό διάγραμμα.

Ο αλγόριθμος Feng και Doolitle, στον οποίο βασίζεται το υπολογιστικό πρόγραμμα Clustal Omega, βασίζεται σε ιεραρχικές μεθόδους.

**Clustal Omega**: είναι ένα ιδιαίτερα διαδεδομένο πρόγραμμα πολλαπλής στοίχισης βιολογικών ακολουθιών. Συγκεκριμένα, αποτελεί την πιο πρόσφατη έκδοση της οικογένειας Clustal και συγκριτικά με προηγούμενες εκδόσεις δίνει την δυνατότητα να στοιχηθούν εκατοντάδες χιλιάδες ακολουθίες σε λίγες μόνο ώρες. Το πρόγραμμα χρησιμοποιεί έναν ιδιαίτερα πολύπλοκο αλγόριθμο προοδευτικής στοίχισης (progressive alignment) για να κάνει σταδιακή στοίχιση πολλαπλών πρωτεϊνικών ή νουκλεοτιδικών ακολουθιών. Το πρόγραμμα βρίσκεται στο EBI (European Bioinformatics Institute) και είναι διαθέσιμο στην παρακάτω ιστοσελίδα: https://www.ebi.ac.uk/Tools/msa/clustalo/.

Η χειροκίνητη επεξεργασία των στοιχίσεων, κυρίως στα τμήματα που συμμετέχουν στην αλληλεπίδραση, και η οπτική παρουσίαση των αποτελεσμάτων της στοίχισης έγινε με το πρόγραμμα επεξεργασίας πολλαπλών στοιχίσεων Jalview v. 2.10 (Waterhouse, Procter, Martin, Clamp, & Barton, 2009).

Τέλος διαμορφώθηκε η διάταξη των ακολουθιών στις στοιχίσεις με στόχο να αποτελούνται από τον ίδιο αριθμό ακολουθιών και για κάθε ακολουθία της πολλαπλής στοίχισης των κινασών να εντοπίζεται στην ίδια ακριβώς θέση η καλμοντουλίνη με την οποία αλληλεπιδρά (απαραίτητες προϋποθέσεις για τις πολλαπλές στοιχίσεις που δέχεται ως είσοδο ο i-coms).

## 2.3. Επιλογή server συνεξέλιξης

### 2.3.1. Εισαγωγή

Όπως αναφέρθηκε και εισαγωγικά οι υπολογιστικές μέθοδοι για τη μελέτη αλληλεπιδρώντων καταλοίπων, έχουν ως «είσοδο» τις πολλαπλές στοιχίσεις των ακολουθιών. Οι περισσότεροι διαθέσιμοι servers με ενσωματωμένες μεθόδους για τη μελέτη αλληλεπιδράσεων μεταξύ αμινοξικών καταλοίπων μέσω της συνεξελικτικής πληροφορίας, «ειδικεύονται» στις περιπτώσεις μελέτης αλληλεπιδράσεων μεταξύ αμινοξικών καταλοίπων ενός πρωτεϊνικού μορίου, όπως οι : visualCMAT (Suplatov, Sharapova, Timonina, & Kopylov, 2018), εντοπίζει συνεξελισσόμενα κατάλοιπα στο εσωτερικό ενός πρωτεϊνικού μορίου και βασίζεται και σε διαθέσιμη δομή του μορίου αυτού (αρχείο PDB ως είσοδο). Λόγω του περιορισμού αυτού δε δέχεται ως είσοδο ενωμένες στοιχίσεις και

BIS2Analyzer (Oteri, Nadalin, Champeimont, & Carbone, 2017), που επίσης εντοπίζει συνεξελισσόμενα κατάλοιπα μιας ομάδας πρωτεϊνών. Ο λόγος που δεν ήταν δυνατή η χρήση του προγράμματος αυτού είναι πως απαιτεί μικρό αριθμό ακολουθιών (δεκάδες) ή, για μεγαλύτερα σύνολα δεδομένων, υψηλά συντηρημένες ακολουθίες (περίπου 70% ομοιότητα).

Δύο από τους διαθέσιμους servers για τη μελέτη συνεξέλιξης μεταξύ αμινοξικών καταλοίπων δυο αλληλεπιδρώντων πρωτεϊνών είναι ο *GREMLIN* (Ovchinnikov, Kamisetty, & Baker, 2014), και ο *EVComplex* (T. A. Hopf et al., 2014) που όμως οι μέθοδοι που χρησιμοποιούν έχουν τροποποιηθεί κατάλληλα για τη μελέτη προκαρυωτικών αλληλεπιδρώντων πρωτεϊνών και απαιτούν μεγάλο σύνολο δεδομένων.

### 2.3.2. Ο server i-coms

Ο i-COMS, είναι ένας server που αναπτύχθηκε από την ομάδα ερευνητών της ινστιτούτου Βιοπληροφορικής (Iserte, Simonetti, Zea, Teppa, & Marino-buslje, 2015), του ινστιτούτου Leloir της Αργεντινής (Fundación Instituto Leloir-FIL). Ο συγκεκριμένος server δίνει τη δυνατότητα στο χρήστη να χρησιμοποιήσει μία ή περισσότερες από τις στατιστικές μεθόδους που είναι διαθέσιμες σε αυτόν για τη μελέτη αλληλεπιδρώντων καταλοίπων μέσω συνεξελικτικής πληροφορίας. Οι στατιστικές μέθοδοι που είναι διαθέσιμες στον συγκεκριμένο server είναι οι :

- cMI(**C**orrected **M**utal **I**nformation)
- mfDCA(**m**ean-**f**ield **D**irect **C**oupling **A**nalysis)
- plmDCA(**p**seudo-**l**ikelihood **D**irect **C**oupling **A**nalysis)

Δέχεται την πρώτη ακολουθία σε κάθε πολλαπλή στοίχιση ως ακολουθία αναφοράς και με βάση αυτήν κάνει και την απεικόνιση των αποτελεσμάτων. Επεξεργάζεται τις πολλαπλές στοιχίσεις και αφαιρεί τα τμήματα που εμφανίζουν κενά στην ακολουθία αναφοράς. Επίσης αφαιρεί από το σύνολο δεδομένων ακολουθίες που καλύπτουν <50% της ακολουθίας αναφοράς (από άποψη μήκους).

Παρέχει τη δυνατότητα στον χρήστη να κατεβάσει τις πολλαπλές στοιχίσεις, έπειτα από την επεξεργασία που έχουν υποστεί, καθώς και την «ενωμένη» πολλαπλή στοίχιση και σε αρχεία κειμένου τα αποτελέσματα της κάθε μεθόδου. Επίσης παρέχει πληροφορίες που αφορούν το σύνολο των στηλών των πολλαπλών στοιχίσεων, το σύνολο των ακολουθιών καθώς και τον αριθμό των ομάδων στις οποίες έχουν κατηγοριοποιηθεί οι ακολουθίες, με χρήση του αλγορίθμου Hobohm-1 (Hobohm, Scharf, Schneider, & Sander, 1992).

Τέλος παρέχει 3 τρόπους απεικόνισης των αποτελεσμάτων, σε κύκλο (στην περίμετρο απεικονίζονται οι ακολουθίες αναφοράς των 2 πρωτεϊνών, ενώ με γραμμές στον εσωτερικό το σύνολο των συνεξελισσόμενων ζευγαριών καταλοίπων, για συγκεκριμένο ποσοστό των τιμών που έδωσε η κάθε μέθοδος), σε γράφημα πυκνότητας των σκορ των ζευγαριών που υπολογίστηκαν με χρήση των μεθόδων είτε στο εσωτερικό της κάθε πρωτεΐνης είτε μεταξύ των πρωτεϊνών και σε πίνακα, όπου στους 2 άξονες εντοπίζονται οι ακολουθίες αναφοράς και των 2 πρωτεϊνών και με σημεία τα ζευγάρια των καταλοίπων.

## 2.4. Επιλογή στατιστικής μεθόδου - Οπτικοποίηση των δομών

Η επιλογή της στατιστικής μεθόδου έγινε με σύγκριση των συνεξελισσόμενων καταλοίπων που πρόβλεψαν οι 3 στατιστικές μέθοδοι του i-coms σε σχέση με τα κατάλοιπα που φαίνεται να αλληλεπιδρούν στις πειραματικά προσδιορισμένες δομές των παρακάτω συμπλόκων :

CaMK Iα σε σύμπλοκο με CaM (κωδικός PDB: 1MXE)

CaMK IIδ σε σύμπλοκο με CaM2 (κωδικός PDB: 2WEL)

MYLK σε σύμπλοκο με CaM1 (κωδικός PDB: 5JQA)

DAPK1 σε σύμπλοκο με CaM2 (κωδικός PDB: 2X0G)

DAPK2 σε σύμπλοκο με CaM2 (κωδικός PDB: 1WRZ)

Το **σύνολο των αλληλεπιδράσεων των πειραματικά προσδιορισμένων δομών έγινε με τη βοήθεια του πρόγραμμα RING** - Residue Interaction Network Generator (Martin et al., 2011; Piovesan, Minervini, & Tosatto, 2016) που δημιουργήθηκε από το εργαστήριο BiocomputingUP στο πανεπιστήμιο της Πάντοβα στην Ιταλία.

Το δίκτυο αλληλεπιδράσεων κατασκευάζεται σε δύο βήματα. Στο πρώτο βήμα ο αλγόριθμος εντοπίζει τα ζευγάρια καταλοίπων που δύναται να αλληλεπιδρούν με βάση τις αποστάσεις τους. Στο δεύτερο βήμα χαρακτηρίζει κάθε αλληλεπίδραση (δεσμοί υδρογόνου, αλληλεπιδράσεις Van der Waals, δισουλφιδικοί δεσμοί, γέφυρες άλατος και αλληλεπιδράσεις πι ηλεκτρονίων). Για το σύνολο των δομών χρησιμοποιήθηκαν οι default επιλογές. Ως έξοδο παρέχει, εκτός από την απεικόνιση του δικτύου με ακμές διαφορετικού χρώματος για την κάθε αλληλεπίδραση, ένα αρχείο κειμένου που περιλαμβάνει τα ζευγάρια των καταλοίπων και τον χαρακτηρισμό της κάθε αλληλεπίδρασης.

Η αξιολόγηση των αποτελεσμάτων του RING, η οπτικοποίηση των δομών και ο σχολιασμός των συνεξελισσόμενων καταλοίπων έγινε με τη βοήθεια του PyMol v. 1.2. Το λογισμικό PyMOL, είναι ένα ανοιχτού κώδικα πρόγραμμα οπτικοποίησης μοριακών γραφικών, το οποίο αναπτύχθηκε από τον Warren Lyford DeLano (DeLano, 2002). Μπορεί να παράγει υψηλής ποιότητας τρισδιάστατες εικόνες μικρών μορίων ή βιολογικών μακρομορίων. Παρέχει στο χρήστη πλήθος επιλογών επεξεργασίας των δομών, όπως μετρήσεις ενδομοριακών και διαμοριακών αποστάσεων, μέτρηση μέσης τετραγωνικής απόκλισης (RMSD) μεταξύ δύο δομών κ.α.

Και για τις 3 μεθόδους θεωρήθηκαν αποδεκτά τα ζευγάρια καταλοίπων που εμφάνιζαν τιμές μεγαλύτερες τις κατώτερης τιμής που εντοπίζεται στον υπολογισμό συνεξελισσόμενων καταλοίπων στο εσωτερικό της κάθε πρωτεΐνης (Iserte et al.,

2015). Για τη μέθοδο cMI αλληλεπιδρώντα ζευγάρια θεωρούνται όσα έχουν λάβει θετικές τιμές. Ενώ για τις μεθόδους mfDCA και plmDCA αξιολογήθηκαν ως «αληθώς θετικά» ~3% των συνολικών ζευγαριών των αλληλεπιδρώντων καταλοίπων, όπως προτείνεται και βιβλιογραφικά (Ekeberg et al., 2013; Hopf et al., 2014, 2015; Marks et al., 2012).

## 2.5. Κατασκευή δομικών μοντέλων των συμπλόκων με τη βοήθεια μεθόδων αγκυροβόλησης

Για την αγκυροβόληση πρωτεϊνών – πρωτεϊνών ελέγχθηκαν, ως προς τις επιδόσεις τους, **η μέθοδος SwarmDock και η μέθοδος HADDOCK** (High Ambiguity Data-driven Docking).

Η μέθοδος SwarmDock (Torchala & Bates, 2014; Torchala, Moal, Chaleil, Agius, & Bates, 2013; Torchala, Moal, Chaleil, Fernandez-Recio, & Bates, 2013) είναι μια μέθοδος αγκυροβόλησης πρωτεΐνης-πρωτεΐνης με ευκαμψία δομών.

Δεδομένα εισαγωγής αποτελούν οι τρισδιάστατες δομές των πρωτεϊνών, ενώ ο χρήστης έχει την δυνατότητα να ορίσει αλληλεπιδρώντα κατάλοιπα και να καθορήσει παραμέτρους, όπως ο αριθμός κανονικών τρόπων ταλάντωσης. Ο αλγόριθμος βασίζεται στην αναπαράσταση των πρωτεϊνών από ένα δίκτυο ελαστικών ατόμων με κανονικούς τρόπους ταλάντωσης που μπορεί να χαρτογραφήσει τις διαφορές συνιστώσες των αδέσμευτων δομών σε αντιστοιχία με τη δέσμια δομή του σύμπλοκο. Η αναπαράσταση αυτή έχει ως πλεονέκτημα την αποτύπωση των στερεοδιατακτικών αλλαγών στις οποίες υπόκεινται κατά τη δημιουργία του το σύμπλοκο, όχι μόνο ως προς τις πλευρικές του ομάδες αλλά και ως προς τον σκελετό.

Αποτελεί έναν καινοτόμο επαναληπτικό μιμητικό αλγόριθμο βασισμένο στον πληθυσμό. Κάθε μέλος του πληθυσμού αποτελεί ένα σωματίδιο με καθορισμένη θέση και ταχύτητα στον χώρο αναζήτησης. Σε κάθε επανάληψη, η ενέργεια κάθε σωματιδίου επαναπροσδιορίζεται και αξιολογούνται η θέση και η ταχύτητα του. Ο χώρος αναζήτησης καθορίζεται από το κέντρο μάζας του προσδέτη και την τετραδιάστατη αναπαράσταση του. Ο υποδοχέας κρατείται σταθερός ενώ τα σωματίδια κινούνται τυχαία στο χώρο αναζήτησης.

Η διάδοση του αλγορίθμου επιτυγχάνεται μέσω δύο συντελεστών, του γνωσιακού που μετράει την ικανότητα του σωματιδίου να κινηθεί προς μία θέση χαμηλής ενέργειας στην οποία έχει ξαναβρεθεί, και του κοινωνικού συντελεστή που εκφράζει την ικανότητα ένα σωματίδιο να κινηθεί προς μία θέση χαμηλής ενέργειας

την οποία έχει βρει κάποιο γειτονικό του σωματίδιο. Μετά από έναν αριθμό επαναλήψεων, η ενεργειακά χαμηλότερη στερεοδιάταξη αποθηκεύεται, και στο σύνολο των επαναλήψεων ομαδοποιούνται οι αποθηκευμένες στερεοδιατάξεις. Η ανάλυση των αγκυροβολημένων στερεοδιατάξεων γίνεται μέσω της μέσης τετραγωνικής απόκλισης (RMSD) των ζευγαριών των ατόμων υποδοχέα- προσδέτη που αλληλεπιδρούν.

Η ενεργειακή συνάρτηση αποτελείται από έναν όρο van der Waals κι έναν ηλεκτροστατικό όρο, ενώ προστίθεται κι ένας όρος εξομάλυνσης της ενέργειας προς αποφυγή υπολογισμού μακρινών αλληλεπιδράσεων. Στη γενική του μορφή :

$$E_{int} = \sum_{i}^{atoms} \sum_{j}^{atoms} E_{ij}$$

Το HADDOCK είναι μια κατευθυνόμενη από δεδομένα μέθοδος αγκυροβόλησης μεταξύ πρωτεϊνών που αξιοποιεί δομικές πληροφορίες (Ambiguous Interactions Restraints, AIRs), προερχόμενες από διάφορες βιοχημικές ή / και υπολογιστικές μελέτες, σαν περιορισμούς για να κατευθύνει την πορεία της αγκυροβόλησης (de Vries et al., 2007; Dominguez, Boelens, & Bonvin, 2003).

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 δίεδρων γωνιών σε συνθήκες κενού και μια τελική προσομοίωση ανόπτησης (simulated annealing) σε συνθήκες διαλύτη. Οι ενεργειακοί υπολογισμοί και οι προσομοιώσεις εκτελούνται μέσω του Crystallographic & NMR System (CNS) v. 1.2 (Brunger, 2007; Brünger et al., 1998), χρησιμοποιώντας το πεδίο δυνάμεων OPLS (Jorgensen & Tirado-Rives, 1988). Οι μη ομοιοπολικές αλληλεπιδράσεις υπολογίζονται στα όρια 6.5 – 8.5 Å, χρησιμοποιώντας συναρτήσεις μετασχηματισμού για το δυναμικό Lennard – Jones και το ηλεκτροστατικό δυναμικό. Το τελικό αποτέλεσμα του HADDOCK περιλαμβάνει το σύνολο των προβλεφθέντων συμπλόκων, ομαδοποιημένα σε clusters με βάση τη δομική ομοιότητά τους, ενώ κάθε cluster βαθμολογείται με ένα κριτήριο (HADDOCK Score), το οποίο προκύπτει σαν σταθμισμένο άθροισμα των τιμών της ενέργειας αλληλεπίδρασης για τις αλληλεπιδράσεις van der Waals ($\Delta G_{vdw}$), τις ηλεκτροστατικές αλληλεπιδράσεις ($\Delta G_{ele}$) και την επίδραση του διαλύτη ($\Delta G_{desolv}$), της θαμμένης επιφάνειας ή επιφάνειας αλληλεπίδρασης (BSA) και των παραβιάσεων των αρχικών περιορισμών (AIRs).

Και στις δύο μεθόδους για τα 5 ζευγάρια αλληλεπιδρώντων πρωτεϊνών, ορίστηκαν ως γνωστά κατάλοιπα αλληλεπίδρασης τα κατάλοιπα των δύο πρωτεϊνών που προβλέφθηκαν ως συνεξελισσόμενα με τη μέθοδο plmDCA.

Για την ομαδοποίηση (clustering) των αποτελεσμάτων στο τελικό στάδιο του HADDOCK, ορίστηκε τιμή κατωφλίου για το RMSD κάθε cluster στα 7.5 Å.

## 2.6. Δομική στοίχιση

Η υπέρθεση δομών (Structure Superposition), ή δομική στοίχιση (Structural Alignment), επιδιώκει την εύρεση της ομολογίας μεταξύ δύο ή περισσότερων δομών με βάση το σχήμα και την τρισδιάστατη στερεοδιάταξή τους. Σημαντικό στοιχείο για τη χρήση της μεθόδου είναι να υπάρχει η ελάχιστη έστω δομική ομοιότητα. Οι δομικές στοιχίσεις μπορούν να φανούν ιδιαίτερα χρήσιμες σε μελέτες δομικής γονιδιωματικής (structural genomics) ή πρωτεωμικής (proteomics), αλλά και σαν μέτρο αξιολόγησης στοιχίσεων βασισμένων αποκλειστικά στην αμινοξική ακολουθία (Maiorov & Crippen, 1994). Παράλληλα, η σύγκριση διαφορετικών δομών της ίδιας πρωτεΐνης , μπορεί να δώσει στοιχεία για τις μεταβολές στη στερεοδιάταξη που μπορεί να υιοθετήσει μια πολυπεπτιδική αλυσίδα. Απαραίτητο στοιχείο για μια υπέρθεση δομών είναι οι συντεταγμένες των μορίων στο χώρο, που παρέχονται από την λύση της τρισδιάστατης δομής είτε πειραματικά, μέσω κρυσταλλογραφίας ακτινών Χ ή πυρηνικού μαγνητικού συντονισμού (NMR), είτε θεωρητικά από μεθόδους πρόγνωσης όπως η προτυποποίηση μέσω ομολογίας ή η αγκυροβόληση. Κατά συνέπεια, η ποιότητα και η αξιοπιστία των αποτελεσμάτων μιας δομικής στοίχισης εξαρτώνται άμεσα από την ποιότητα των δομών που συγκρίνονται. Τα αποτελέσματα μιας δομικής στοίχισης περιλαμβάνουν την υπέρθεση των ατομικών συντεταγμένων των συγκρινόμενων μορίων καθώς και την τιμή της ρίζας των μέσων τετραγώνων των αποκλίσεων (Root Mean Square Deviation, RMSD). Το RMSD δύο στοιχημένων δομών αποτελεί μέτρο της μέσης απόκλισης μεταξύ των ατόμων, συνήθως του ανθρακικού σκελετού των πρωτεϊνών, και υπολογίζεται ως εξής:

$$RMSD(v,w) = \sqrt{\frac{\sum_{i=1}^{n} ||v_i - w_i||^2}{n}}$$

$$RMSD(v,w) = \sqrt{\frac{\sum_{i=1}^{n} ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}{n}}$$

όπου v, w τα σημεία στο χώρο μεταξύ των οποίων μετράται η απόσταση (συνήθως άτομα Cα).

Η υπέρθεση των μοντέλων και των κρυσταλλογραφικά λυμένων δομών πραγματοποιήθηκε με τη βοήθεια του λογισμικού PyMol v. 1.2, συγκεκριμένα της λειτουργίας align. Η λειτουργία align πραγματοποιεί αρχικά μία BLASTlike στοίχιση των αλληλουχιών των δύο συγκρινόμενων μορίων (χρησιμοποιώντας ως πίνακα τον BLOSUM62) ακολουθούμενη από στοίχιση των δομών τους, και στη συνέχεια

εκτελεί μια σειρά από κύκλους βελτιστοποίησης προκειμένου να απορρίψει δομικές ακραίες τιμές που μπορεί να προκύψουν κατά το “ζευγάρωμα” των ατόμων (DeLano, 2002). Το τελικό αποτέλεσμα δίνεται γραφικά με την απεικόνιση των ταιριασμένων δομών, ενώ παρατίθενται επίσης ο αριθμός των ατόμων που στοιχήθηκαν, η τιμή RMSD κάθε κύκλου βελτιστοποίησης, ο αριθμός των ατόμων που απορρίφθηκαν και η τελική τιμή RMSD της υπέρθεσης.

# 3.Αποτελέσματα και συζήτηση

## 3.1. Κατασκευή του συνόλου δεδομένων

Η αναζήτηση εξαρτώμενων από καλμοντουλίνη κινασών στην Swiss-Prot με βάση το κριτήρια κατάταξης των κινασών σε οικογένειες και την αναζήτηση με λέξεις κλειδιά το **family name : CAMK Ser/Thr protein kinase family** επέστρεψε 458 καταχωρήσεις. Το σύνολο δεδομένων στη συνέχεια εμπλουτίστηκε με βάση το E.C. – Enzyme Commission number με όνομα κατηγορίας **Calcium/calmodulin-dependent protein kinase και κωδικό 2.7.11.17** (enzyme.expasy.org/EC/2.7.11.17). Η αναζήτηση επέστρεψε 64 καταχωρήσεις, από τις οποίες οι 55 πληρούσαν και τα δύο κριτήρια. Οι υπόλοιπες 9 καταχωρήσεις αφορούν εξαρτώμενες από καλμοντουλίνη κινάσες του ρυζιού και τις α και β ισομορφές της CaMKK και κρατήθηκαν για περαιτέρω επεξεργασία.

Ο όγκος της πληροφορίας στην Uniprot/SwissProt είναι τεράστιος, όπως επίσης και η πιθανότητα (παρά τις προσπάθειες), η πληροφορία αυτή να είναι λανθασμένη ή απλά ελλιπής. Για το λόγο αυτό **εισήχθησαν στην αναζήτηση πολλά διαφορετικά κριτήρια**. Επίσης η αναζήτηση με βάση την υπεροικογένεια και όχι την οικογένεια των CaMK (που θα έδινε ακολουθίες με υψηλή ομολογία) έγινε με στόχο τον εμπλουτισμό του συνόλου δεδομένων για την καλύτερη δυνατή απόδοση των αλγορίθμων συνεξέλιξης.

Αναφέρεται πως το σύνολο των ακολουθιών πρέπει να ξεπερνά τις 100. Επίσης οι ακολουθίες πρέπει να διαφέρουν μεταξύ τους ώστε να εξασφαλίζεται η αυξημένη ποικιλία αμινοξικών καταλοίπων σε κάθε στήλη της πολλαπλής στοίχισης. Θέσεις που δε μεταβάλλονται ποτέ, όπως τα συντηρημένα κατάλοιπα του ενεργού κέντρου των κινασών θα «αντιστέκονται» στις διορθώσεις των μεθόδων συνεξέλιξης. Επιπροσθέτως, ορθόλογες ακολουθίες από συγγενικά είδη μπορεί να εμφανίζουν ελάχιστες αλλαγές, ακόμα και σε λιγότερο κρίσιμες θέσεις. Η μεγάλη ομοιότητα των αμινοξικών ακολουθιών μπορεί να διαταράξει σοβαρά την ανάλυση, μειώνοντας την συνολική ποικιλομορφία των ακολουθιών και, σε κάποιες περιπτώσεις, αυξάνοντας τον φυλογενετικό θόρυβο. Από την άλλη πλευρά όμως, οι αλληλουχίες δεν πρέπει να διαφέρουν σε τέτοιο βαθμό ώστε να χαθεί η εμπιστοσύνη ως προς την λειτουργική τους ομολογία. Έτσι ένα τυπικό κατώτερο όριο ομολογίας θεωρείται η 20-30% ομοιότητα των αμινοξικών ακολουθιών. Αν στο σύνολο δεδομένων συμπεριληφθούν και παράλογες πρωτεΐνες μπορεί να επιτευχθεί η ζητούμενη ποικιλία αλληλουχιών και έτσι να βελτιωθεί η απόδοση των μεθόδων συνεξέλιξης. Προϋπόθεση αποτελεί οι παράλογες πρωτεΐνες να υπόκεινται σε κοινούς λειτουργικούς περιορισμούς με τις ορθόλογες (Ashenberg & Laub, 2013).

Στο σύνολο των 467 καταχωρήσεων εφαρμόστηκε ένα επιπλέον κριτήριο σχετικά με την αλληλεπίδρασή τους με την καλμοντουλίνη. Αυτό πραγματοποιήθηκε με την βοήθεια της βάσης δεδομένων Gene Ontology. Το βήμα αυτό κρίθηκε απαραίτητο καθώς αρκετές από τις καταχωρημένες πρωτεΐνες μόνο κατ' όνομα είναι εξαρτώμενες από καλμοντουλίνη. Χαρακτηριστικό παράδειγμα αποτελούν οι CDPKs των φυτών. Οι κινάσες αυτές διαθέτουν μοτίβα EF-hands και είναι ικανές να προσδένουν ιόντα ασβεστίου χωρίς τη μεσολάβηση της καλμοντουλίνης (θα μπορούσαν να θεωρηθούν CaM-like κινάσες)(Klimecka & Muszyńska, 2007). Η αναζήτηση αυτή επέστρεψε 125 καταχωρήσεις οι οποίες και αποτέλεσαν το τελικό σύνολο δεδομένων για τις κινάσες (βλ. Πίνακα 3.1).

**Πίνακας 3.1.** Πίνακας συνόλου δεδομένων εξαρτώμενων από καλμοντουλίνη κινασών με βάση τον οργανισμό.

| Οργανισμός | Σύνολο εξαρτώμενων από CaM κινασών |
|---|---|
| **Mus musculus (Mouse)** | 24 |
| **Homo sapiens (Human)** | 24 |
| **Rattus norvegicus (Rat)** | 18 |
| **Arabidopsis thaliana (Mouse-ear cress)** | 8 |
| **Bos taurus (Bovine)** | 7 |
| **Caenorhabditis elegans** | 6 |
| **Oryctolagus cuniculus (Rabbit)** | 4 |
| **Saccharomyces cerevisiae (Baker's yeast)** | 3 |
| **Danio rerio (Zebrafish)** | 3 |
| **Drosophila melanogaster (Fruit fly)** | 3 |
| **Schizosaccharomyces pombe (Fission yeast)** | 3 |
| **Gallus gallus (Chicken)** | 2 |
| **Caenorhabditis briggsae** | 2 |
| **Oryza sativa subsp. japonica (Rice)** | 2 |
| **Xenopus tropicalis (Western clawed frog)** | 2 |
| **Xenopus laevis (African clawed frog)** | 2 |
| **Lilium longiflorum (Trumpet lily)** | 1 |
| **Emericella nidulans (Aspergillus nidulans)** | 2 |
| **Lotus japonicus** | 1 |
| **Takifugu rubripes (Japanese pufferfish)** | 1 |
| **Ovis aries (Sheep)** | 1 |
| **Sus scrofa (Pig)** | 1 |
| **Pisum sativum (Garden pea)** | 1 |
| **Pongo abelii (Sumatran orangutan)** | 1 |
| **Macrobrachium nipponense (Oriental river shrimp)** | 1 |
| **Malus domestica (Apple) (Pyrus malus)** | 1 |
| **Medicago truncatula (Barrel medic)** | 1 |

Με στόχο να ικανοποιείται η 20% ομοιότητα των πρωτεϊνικών ακολουθιών που προαναφέρθηκε, έγινε επεξεργασία των αμινοξικών ακολουθιών. Ως αποτέλεσμα ήταν η **απομάκρυνση τμημάτων των ακολουθιών που με βάση τη θέση τους στη δομή της κινάσης ή τις καταγεγραμμένες στη βιβλιογραφία λειτουργίες τους**, δεν μπορούσαν να λαμβάνουν μέρος στην αλληλεπίδραση κινάσης-καλμοντουλίνης. Χαρακτηριστικό παράδειγμα αποτελούν οι αυτοτελείς δομικές περιοχές πρόσδεσης ακτίνης, οι δομικές αυτοτελείς περιοχές πρόσδεσης μυοσίνης καθώς και οι IG δομικές αυτοτελείς περιοχές των MLCKs (για περισσότερες λεπτομέρειες βλ.1.2.3.6). Στο τελικό σύνολο δεδομένων οι ακολουθίες αποτελούνταν από **1) το καταλυτικό κέντρο με ενεργότητα κινάσης, τη ρυθμιστική περιοχή που αποτελείται από: 2) τη αυτοτελή δομική περιοχή αυτοαναστολής και 3) τη αυτοτελή δομική περιοχή πρόσδεσης καλμοντουλίνης, καθώς και 5) τμήματα προς το αμινοτελικό ή καρβοξυτελικό άκρο αυτών άγνωστης λειτουργίας.**

Έπειτα με τη βοήθειά του προγράμματος CD-HIT κατηγοριοποιήθηκαν οι επεξεργασμένες ακολουθίες σε ομάδες με βάση την 20% ομοιότητά τους. Το σύνολο των ακολουθιών κατηγοριοποιήθηκε σε μια ομάδα. Αξίζει να σημειωθεί πως με κατώφλι 30% ομοιότητα οι ακολουθίες κατηγοριοποιήθηκαν σε **7 ομάδες**. Στην πρώτη ομάδα περιέχονται οι : CAMKIV DAPK2, DAPK1, CaMKIβ, CaMKIδ, CaMKIα, CaMKIγ, CaMKIIβ, CaMKIIδ, CaMKIIα και MLCK1. Στη δεύτερη ομάδα όλες οι κινάσες του Arabidopsis thaliana, η CaMK1 του Oryza sativa και η MLCK2. Στην τρίτη ομάδα οι : υπόλοιπες κινάσες των φυτών, PHKG1, CSKP (Peripheral plasma membrane protein CASK) και CaMKIIγ. Στην τέταρτη ομάδα οι : MKNK1 και 2 (MAP kinase-interacting serine/threonine-protein kinase), MAPK5 (Mitogen Activated Protein Kinase 5). Στην πέμπτη ομάδα οι : CaMKIIδ του Xenopus laevis, CaMKII του Schizosaccharomyces pombe και PHKG2. Στην έκτη ομάδα οι : CaMKK1 και 2. Στην έβδομη ομάδα οι : MAPK2 και MAPK3.

Η συλλογή του συνόλου δεδομένων για τις CaM από την Swiss-Prot έγινε **με βάση την οικογένεια των CaM (Calmodulin family)**. Η αναζήτηση επέστρεψε 172 καταχωρίσεις, 41 από τις οποίες ήταν Calmodulin-like, ή CaM-related, ή Colglandulin, ή Squidulin και αφαιρέθηκαν. Για τις υπόλοιπες 131 καταχωρήσεις ακολούθησε δεύτερη αναζήτηση με στόχο να αφαιρεθούν ακολουθίες από οργανισμούς στους οποίους δεν υπάρχει κατατεθειμένη κινάση, στους ίδιους ή σε συγγενικούς οργανισμούς. Ο βαθμός συγγένειας τοποθετήθηκε σε επίπεδο τάξης με την εξής λογική : αν για κάποιο είδος υπήρχε στο σύνολο δεδομένων καταχώρηση κινάσης αλλά όχι καλμοντουλίνης, με δεδομένο πως οι καλμοντουλίνες είναι υψηλά συντηρημένες, κρατήθηκε στο σύνολο δεδομένων η ακολουθία καλμοντουλίνης συγγενικού είδους. Αυτό έγινε σαν μια προσπάθεια, σε μια τόσο συντηρημένη ομάδα πρωτεϊνών, η όποια ποικιλία εντοπίζεται σε επίπεδο ακολουθίας να συμπεριληφθεί στο σύνολο δεδομένων λόγω των περιορισμών των μεθόδων

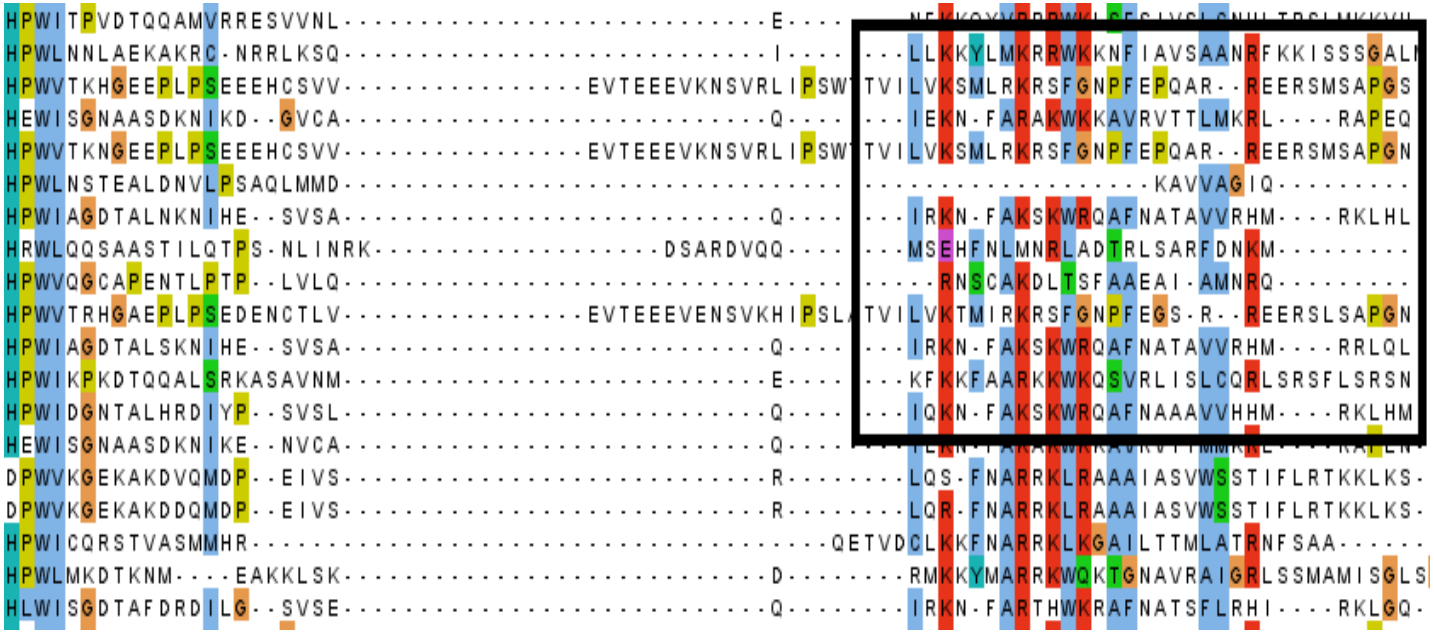
συνεξέλιξης που προαναφέρθηκαν. Χαρακτηριστικό παράδειγμα οι κινάσες των Pisum sativum και Lotus japonicus για τις οποίες κρατήθηκε η CaM της Glycine max, η κινάση του Medicago truncatula για την οποία κρατήθηκε η CaM του Medicago sativa, η κινάση του Macrobrachium nipponense για την οποία κρατήθηκε η CaM της Drosophila melanogaster και η κινάση του Takifugu rubripes για την οποία κρατήθηκαν οι CaM των Epinephelus akaara και Perca flavescens. Το σύνολο δεδομένων των CaM, τελικά, αποτελείται από 42 καταχωρίσεις και παρουσιάζεται στον πίνακα 3.2.

**Πίνακας 3.2.** Πίνακας συνόλου δεδομένων για τις CaM με βάση τον οργανισμό.

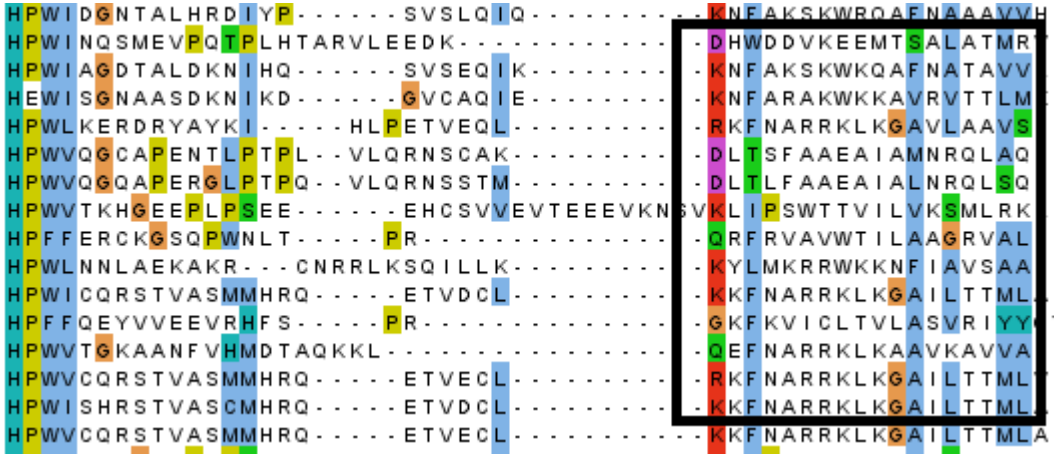
| Οργανισμός | Σύνολο CaM πρωτεϊνών |
|---|---|
| **Arabidopsis thaliana (Mouse-ear cress)** | 7 |
| **Bos taurus (Bovine)** | 1 |
| **Caenorhabditis elegans** | 1 |
| **Danio rerio (Zebrafish)** | 1 |
| **Drosophila melanogaster (Fruit fly)** | 1 |
| **Emericella nidulans (Aspergillus nidulans)** | 1 |
| **Epinephelus akaara (Hong Kong grouper)** | 1 |
| **Gallus gallus (Chicken)** | 3 |
| **Glycine max (Soybean)** | 1 |
| **Homo sapiens (Human)** | 3 |
| **Lilium longiflorum (Trumpet lily)** | 1 |
| **Malus domestica (Apple)** | 1 |
| **Medicago sativa (Alfalfa)** | 1 |
| **Mus musculus (Mouse)** | 3 |
| **Oryctolagus cuniculus (Rabbit)** | 1 |
| **Oryza sativa subsp. japonica (Rice)** | 3 |
| **Ovis aries (Sheep)** | 1 |
| **Perca flavescens (American yellow perch)** | 1 |
| **Pongo abelii (Sumatran orangutan)** | 1 |
| **Rattus norvegicus (Rat)** | 3 |
| **Saccharomyces cerevisiae (Baker's yeast)** | 1 |
| **Schizosaccharomyces pombe (Fission yeast)** | 2 |
| **Xenopus laevis (African clawed frog)** | 3 |

## 3.2. Πολλαπλές στοιχίσεις ακολουθιών

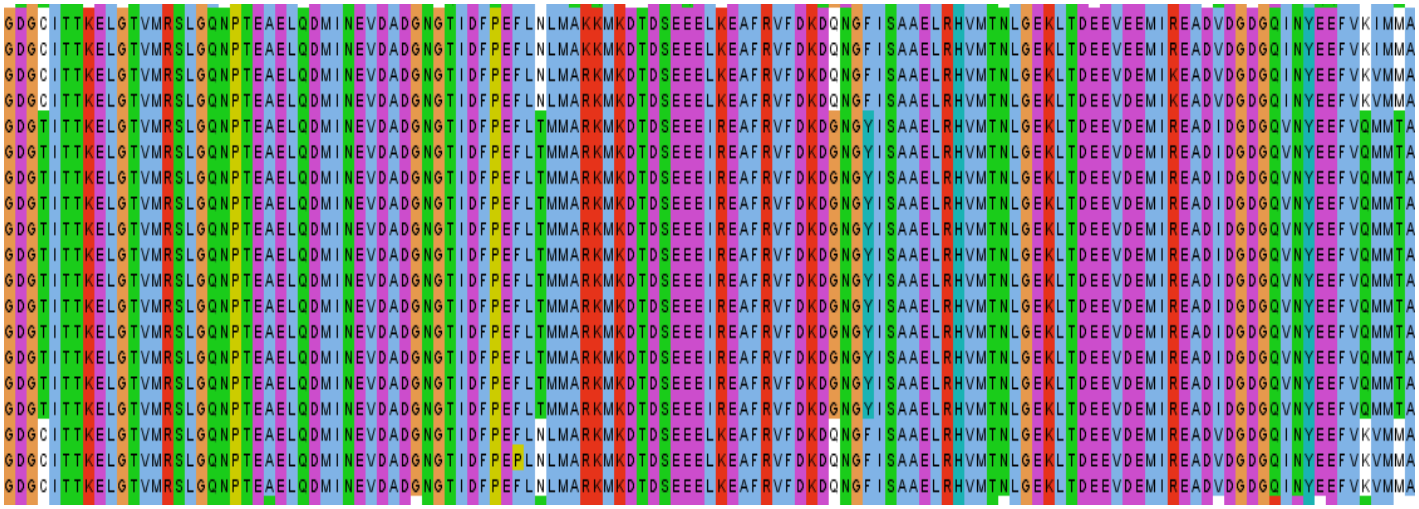
Τα προγράμματα στοίχισης που χρησιμοποιήθηκαν, ενώ έδιναν αποτελεσματικές στοιχίσεις των ακολουθιών στο καταλυτικό κέντρο των κινασών καθώς και στις CaM, στις περιοχές που είναι πειραματικά αποδεδειγμένο πως προσδένεται η CaM δεν ήταν το ίδιο αποτελεσματικά. Αυτό πιθανά οφείλεται στη τεράστια ποικιλία που εμφανίζουν μεταξύ τους οι περιοχές αναγνώρισης της CaM, όπως αναλύθηκε και εισαγωγικά (O'Neil & DeGrado, 1990). Έγινε προσπάθεια χειροκίνητης στοίχισης των τμημάτων αυτών, με βάση τη δομική πληροφορία που υπάρχει κατατεθειμένη και πειραματικά δεδομένα που έχουν προκύψει από κατευθυνόμενες σημειακές μεταλλάξεις στις αντίστοιχες περιοχές. Δόθηκε ιδιαίτερη βάση στα κατάλοιπα αγκυροβόλησης στις ακολουθίες που υπήρχε σχετική πληροφορία και με βάση αυτές στοιχήθηκαν και οι υπόλοιπες. Οι διαφορές των δύο στοιχίσεων, καθώς και η τελική στοίχιση των CaM, παρουσιάζονται στις παρακάτω εικόνες.



**Εικόνα 3.1** Τμήμα της πολλαπλής στοίχισης των ακολουθιών των εξαρτώμενων από CaM κινασών, όπως προέκυψε από το πρόγραμμα ClustalO.



**Εικόνα 3.2** Τμήμα της πολλαπλής στοίχισης των εξαρτώμενων από καλμοντουλίνη κινασών έπειτα από επεξεργασία της δομικής αυτοτελούς περιοχής πρόσδεσης της CaM με βάση πειραματικά δεδομένα (στο πλαίσιο η αντίστοιχη περιοχή).



**Εικόνα 3.3** Τμήμα της πολλαπλής στοίχισης των CaM (είναι εμφανής η υψηλή συντήρηση των μελών της οικογένειας).

Οι πολλαπλές στοιχίσεις των ακολουθιών επεξεργάστηκαν περαιτέρω ώστε **να εντοπίζεται στην ίδια θέση τόσο η εξαρτώμενη από καλμοντουλίνη κινάση, όσο και η καλμοντουλίνη με την οποία αλληλεπιδρά**. Η διάταξη αυτή έγινε χειροκίνητα,(όπου χρειάστηκε εισήχθησαν κενά με στόχο στην τελική πολλαπλή στοίχιση όλες οι ακολουθίες να έχουν το ίδιο τελικό μήκος), ενώ συνέβαλλε καθοριστικά ώστε να αρθεί και ο περιοριστικός παράγοντας του μικρού αριθμού των CaM ακολουθιών, καθώς όπως έχει ήδη αναφερθεί υπολογιστικές μέθοδοι αυτές για να δώσουν αξιόπιστα αποτελέσματα απαιτούν ένα μεγάλο αριθμό πρωτεϊνικών αλληλουχιών. Δόθηκε ιδιαίτερη προσοχή, όσο αυτό ήταν εφικτό, στην υψηλή συντήρηση των καλμοντουλινών. Έτσι παρά το γεγονός πως οι 3 ισομορφές της καλμοντουλίνης στον άνθρωπο, για παράδειγμα, υπόκεινται σε ιστοειδική έκφραση, είναι πιθανό και οι 3, σε διαφορετικούς ιστούς, να αλληλεπιδρούν με την ίδια κινάση. Για να μην υπερεκπροσωπούνται στο σύνολο δεδομένων, και με δεδομένο ότι οι πρωτεϊνικές τους ακολουθίες είναι ταυτόσημες, **κάθε κινάση στα θηλαστικά και τα πτηνά διατάχθηκε στην ίδια θέση με μια μόνο καλμοντουλίνη**. Οι ακολουθίες των καλμοντουλινών των υπόλοιπων taxa κρατήθηκαν με στόχο να αυξηθεί, όσο γίνεται η πληροφορία στην κάθε στήλη της πολλαπλής στοίχισης. Συνολικά η «ενωμένη» στοίχιση αποτελείται από 183 ζευγάρια **αλληλεπιδρώντων πρωτεϊνών**.

## 3.3. Επιλογή στατιστικής μεθόδου πρόβλεψης συνεξελισσόμενων καταλοίπων

Εξετάστηκαν τα αποτελέσματα των στατιστικών μεθόδων και για τα πέντε σύμπλοκα αναφοράς. Οι μέθοδοι cMI και mfDCA έδωσαν αξιοσημείωτα χαμηλά ποσοστό αληθώς θετικών αλληλεπιδράσεων, με την mfDCA να δίνει 2,3% και την cMI 10.3% στο σύμπλοκο CaMK IIδ – CaM2. Έτσι, **αφού τα προβλεπόμενα ως συνεξελισσόμενα κατάλοιπα από την plmDCA είχαν μεγαλύτερη επικάλυψη με τα «πραγματικά» αλληλεπιδρώντα κατάλοιπα σε σχέση με τις άλλες δυο μεθόδους, για τους σκοπούς της παρούσας διπλωματικής εργασίας επιλέχθηκε ως η καταλληλότερη εκ των τριών διαθέσιμων μεθόδων**.

Αξίζει όμως να σχολιάσουμε, όσον αφορά τις άλλες δύο μεθόδους, πως η μέθοδος της διορθωμένης αμοιβαίας πληροφορίας (cMI), ενώ παρουσίασε μεγαλύτερο ποσοστό αληθώς θετικών αμινοξικών καταλοίπων σε σύγκριση με τη μέθοδο mfDCA, προέβλεψε και πολλά αλληλεπιδρώντα κατάλοιπα σε περιοχές που πρακτικά είναι αδύνατο να υπάρχουν αλληλεπιδράσεις μεταξύ κινασών και CaM. Χαρακτηριστικό παράδειγμα αποτελούν κατάλοιπα που συμμετέχουν σε αλληλεπιδράσεις των συντηρημένων αυτοτελών δομικών περιοχών των κινασών και δεν εντοπίζονται εκτεθειμένα στο διαλύτη ώστε να είναι εφικτή η αλληλεπίδρασή τους με κατάλοιπα της CaM. Το γεγονός, αυτό θα μπορούσε να θεωρηθεί αναμενόμενο, καθώς η cMI, από τη φύση της ως τοπικό στατιστικό μοντέλο, δεν μπορεί να διακρίνει άμεσες και έμμεσες συσχετίσεις αμινοξικών καταλοίπων με αποτέλεσμα να προβλέπεται μεγάλος αριθμός αληθώς ψευδών συσχετίσεων, περιορισμός που μπορεί να ξεπεραστεί μερικώς από τη χρήση ολικών στατιστικών μοντέλων.

Αντίθετα, η μέθοδος mfDCA, όπως και η plmDCA, προέβλεψε αρκετά «γειτονικά» κατάλοιπα (που θα μπορούσαν να ειδωθούν ως «ψευδώς αρνητικά»), τόσο ως προς τα αλληλεπιδρώντα κατάλοιπα στην κρυσταλλογραφικά προσδιορισμένη δομή, όσο και ως προς τα κατάλοιπα εκείνα που είναι εφικτό να αλληλεπιδρούν, όπως κατάλοιπα των οποίων οι πλευρικές αλυσίδες βρίσκονται εκτεθειμένες και τα οποία εντοπίζονται κυρίως στις θηλιές μεταξύ των β-κλώνων του N-λοβού ή μεταξύ των α-ελίκων του C-λοβού. Το γεγονός αυτό είναι αυξημένης σημασίας στην προσπάθεια διατύπωσης ενός πιθανού μηχανισμού αλληλεπίδρασης μεταξύ των κινασών και της CaM κατά την αποδέσμευση της αυτοτελούς δομικής περιοχής της κινάσης από το ενεργό κέντρο αυτής (τουλάχιστον σε θεωρητικό επίπεδο, καθώς δεν υπάρχουν επαρκή δεδομένα για να ελεγχθεί η υπόθεση και, στις περιπτώσεις που είναι γνωστός ο μηχανισμός, διαφέρει μεταξύ των διαφορετικών υπό-οικογενειών των εξαρτώμενων από καλμοντουλίνη κινασών), με δεδομένο πως, όπως αναφέρθηκε εισαγωγικά, και τα

δυο μόρια υπόκεινται σε σημαντικές στερεοδιατακτικές αλλαγές κατά την αλληλεπίδραση.

**Πίνακας 3.3.** Αποτελέσματα της στατιστικής μεθόδου plmDCA του i-coms.

| Σύμπλοκο κινάσης – καλμοντουλίνης | True Positive (plmDCA) | False Positive (plmDCA) | Σύνολο συνεξελισσόμενων ζευγαριών καταλοίπων | Συνεξελισσόμενα κατάλοιπα στην πειραματικά προσδιορισμένη περιοχή αλληλεπίδρασης |
|---|---|---|---|---|
| **CaMK Iα – CaM** | 27.8% | 3.87% | 6,389 | 131 |
| **CaMK IIδ – CaM2** | 34.8% | 6,76% | 7,521 | 240 |
| **MLCK – CaM1** | 32.9% | 4.19% | 6,701 | 142 |
| **DAPK1 – CaM2** | 34.9% | 3.63% | 6,039 | 282 (στην επιφάνεια του καταλυτικού κέντρου)<br>143 (στην δομική αυτοτελή περιοχή πρόσδεσης CaM) |
| **DAPK2 – CaM2** | 32.7% | 3.73% | 6,131 | 118 |

Σε περιπτώσεις ενζύμων αναφέρεται στη βιβλιογραφία πως αν ένα κατάλοιπο εντοπίζεται στον χώρο να περιστοιχίζεται από συνεξελισσόμενα κατάλοιπα, μπορούμε να συμπεράνουμε πως πρόκειται για κατάλοιπο που συμμετέχει στην κατάλυση (Buslje, Teppa, Di Domenico, Delfino, & Nielsen, 2010). Αν και τα συνεξελισσόμενα κατάλοιπα συχνά πραγματοποιούν άμεσες επαφές, αυτό δεν ισχύει πάντα. Αυτές οι αλληλεπιδράσεις δεν πρέπει να αντιμετωπίζονται ως ψευδώς θετικά αποτελέσματα. Μπορεί να συμβαίνουν παροδικά κατά τις στερεοδιατακτικές αλλαγές που συμβαίνουν πριν την σταθεροποίηση του συμπλόκου και δεν έχουν αποτυπωθεί στην κρυσταλλικά προσδιορισμένη δομή του συμπλόκου (Dago et al., 2012; Marks et al., 2011). Επίσης μπορεί να προβλεφθούν κατάλοιπα τα οποία επηρεάζουν την ίδια λειτουργία (π.χ. αλληλεπίδραση με άλλες πρωτεΐνες, αναγνώριση υποστρώματος) αλλά από διαφορετικές περιοχές στο μόριο μιας πρωτεΐνης ως ισχυρά συνεξελισσόμενα κατάλοιπα (Chakrabarti & Panchenko, 2009; Lunzer, Golding, & Dean, 2010). Για παράδειγμα, κατάλοιπα που προωθούν μια συγκεκριμένη στερεοδιάταξη μπορεί να εντοπίζονται σε απομακρυσμένα σημεία στη δομή μιας πρωτεΐνης, όπως στις πρωτεΐνες που υπόκεινται σε αλλοστερική ρύθμιση.

## 3.4. Αξιολόγηση αποτελεσμάτων του server i-coms

Ακολούθησε **αξιολόγηση των συνεξελισσόμενων ζευγαριών καταλοίπων που προέκυψαν από τη μέθοδο plmDCA του i-coms**. Αυτό πραγματοποιήθηκε τόσο με σχολιασμό των ζευγαριών πάνω στις πειραματικά προσδιορισμένες δομές των συμπλόκων, όσο και με τη βοήθεια προγραμμάτων αγκυροβόλησης πρωτεΐνης-πρωτεΐνης.

Η αξιολόγηση των αποτελεσμάτων έγινε με τη βοήθεια του SwarmDock. Σε αυτό το σημείο πρέπει να αναφερθεί προς το πρόγραμμα αγκυροβόλησης HADDOCK απέτυχε στην τοποθέτηση της α-έλικας της κινάσης στον εσωτερικό της καλμοντουλίνης. Έτσι, ενώ το συνολικό RMSD για τις ομάδες καλμοντουλίνης που προέκυψαν από την αγκυροβόληση ήταν ικανοποιητικό (περίπου 1.000), οι έλικες των αντίστοιχων κινασών δεν κατάφερναν να προσομοιώσουν τις πειραματικά προσδιορισμένες δομές. Αυτό οφείλεται πιθανά είτε στους αυστηρούς περιορισμούς τους αλγορίθμου HADDOCK (κυρίως κατά την ελαχιστοποίηση ενέργειας των στατικών (rigid body) δομών), είτε στην υψηλά εξειδικευμένη αλληλεπίδραση μεταξύ κινασών και καλμοντουλινών, η οποία επηρεάζεται σημαντικά από τη σύσταση του περιβάλλοντος, στο οποίο λαμβάνει χώρα η αλληλεπίδραση και δύσκολα προσδιορίζεται σε συνθήκες υπολογιστικής αγκυροβόλησης. Η επιτυχία του SwarmDock, από την άλλη, πιθανά οφείλεται στην ευελιξία που παρουσιάζει ο αλγόριθμος. Τα κατάλοιπα που εισήχθησαν και στις δύο μεθόδους ως περιορισμοί απεικονίζονται στις πειραματικά προσδιορισμένες δομές των συμπλόκων.

Στο σύνολο των δομών εξαιρέθηκαν από τα προγράμματα αγκυροβόλησης τα κατάλοιπα που συμμετέχουν στη δέσμευση των ιόντων ασβεστίου και εντοπίζονται στις θηλιές κάθε EF-hand μοτίβου. Αυτό δε σημαίνει πως δεν μπορούμε να εξάγουμε χρήσιμη πληροφορία από το γεγονός ότι προβλέφθηκαν ως αλληλεπιδρώντα και στα 5 σύμπλοκα, αλλά σχετικός σχολιασμός θα ακολουθήσει μετά το σχολιασμό της κάθε δομής.

Σημειώνεται σε αυτό το σημείο πως η απεικόνιση των δομικών αυτοτελών περιοχών πρόσδεσης της CaM (αμφιπαθείς α-έλικες) σε helical wheel πραγματοποιήθηκε με τη βοήθεια του heliQuest (Gautier, Douguet, Antonny, & Drin, 2008). Στις απεικονίσεις σημειώνονται τα κατάλοιπα αγκυροβόλησης στην CaM, καθώς και τα συνεξελισσόμενα κατάλοιπα που έλαβαν τα πρώτα 20 υψηλότερα σκορ κατά τη μέθοδο plmDCA, όπως προτείνουν οι δημιουργοί του i-coms (Iserte et al., 2015).

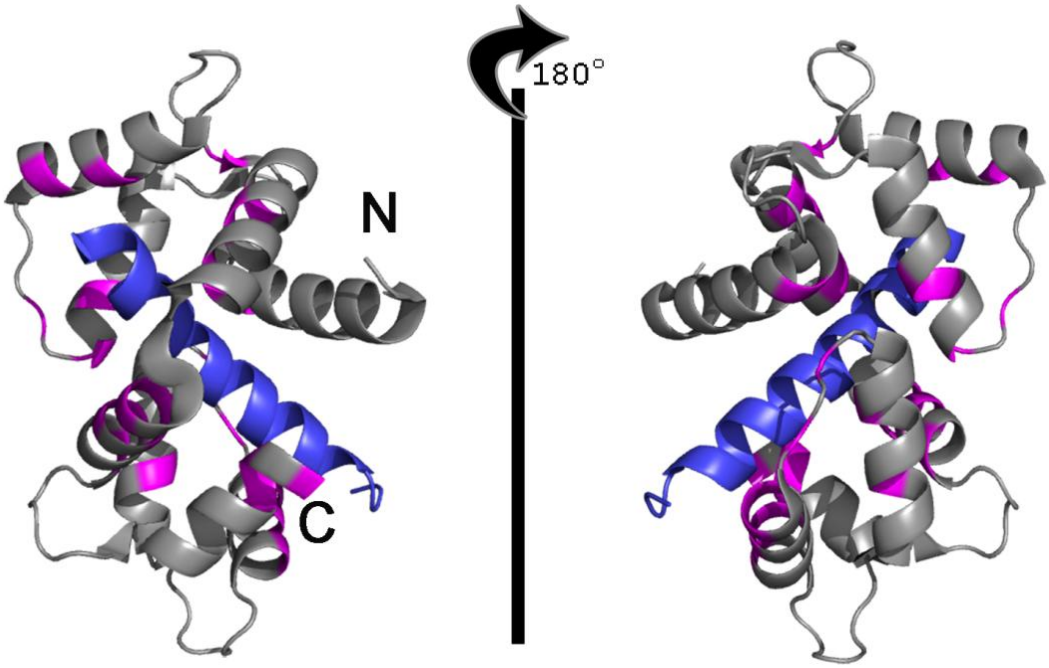
Ο χρωματικός κώδικας στις δομές που παρουσιάζονται ακολουθεί το παρακάτω σχήμα:

| | | | |
|---|---|---|---|
| Πειραματικά επιβεβαιωμένα κατάλοιπα | Αποδεκτά συνεξελισσόμενα κατάλοιπα | Συνεξελισσόμενα κατάλοιπα που απορρίπτονται | Κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα |

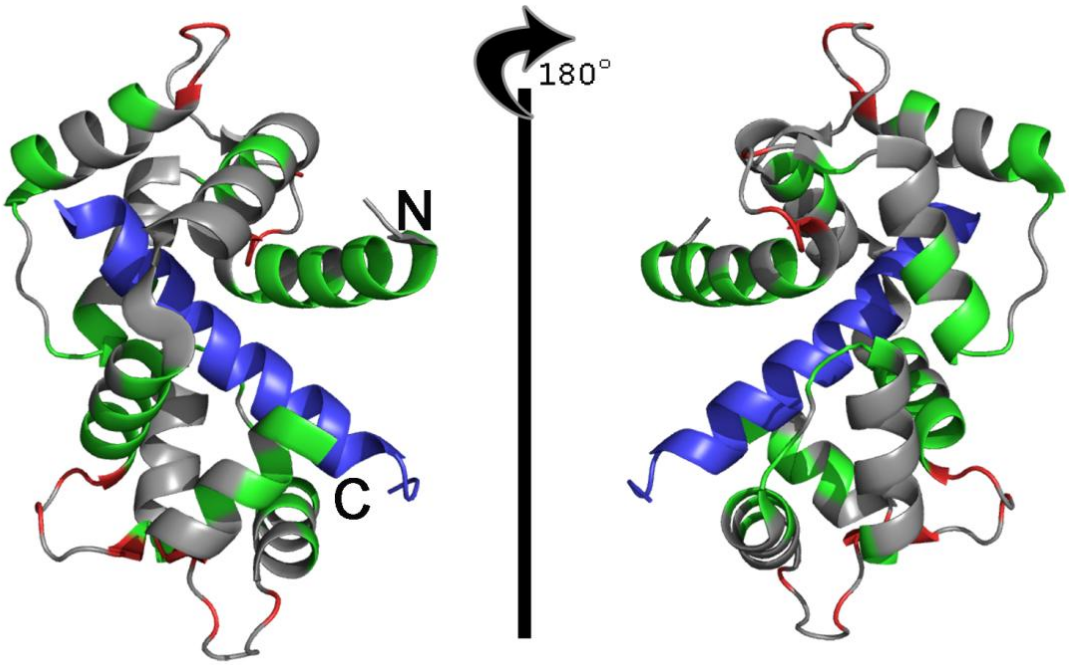
### 3.4.1. Μελέτη συμπλόκου CaMK Ια – $Ca^{2+}$-CaM και συμπεράσματα

Στην εικόνα 3.4 απεικονίζεται με αναπαράσταση cartoon το σύμπλοκο CaMK Ια - $Ca^{2+}$-CaM (κωδικός PDB: 1MXE). Με γκρι απεικονίζεται η καλμοντουλίνη CaM, ενώ με μπλε απεικονίζεται η κινάση CaMK Ια. Επισημαίνονται με magenta οι θέσεις των καταλοίπων που συμμετέχουν στην αλληλεπίδραση των δύο μορίων με βάση τις default παραμέτρους του προγράμματος RING, από όπου συγκεντρώθηκαν τα ζευγάρια των καταλοίπων που αλληλεπιδρούν καθώς και ο χαρακτηρισμός της φύσης της κάθε αλληλεπίδρασης.

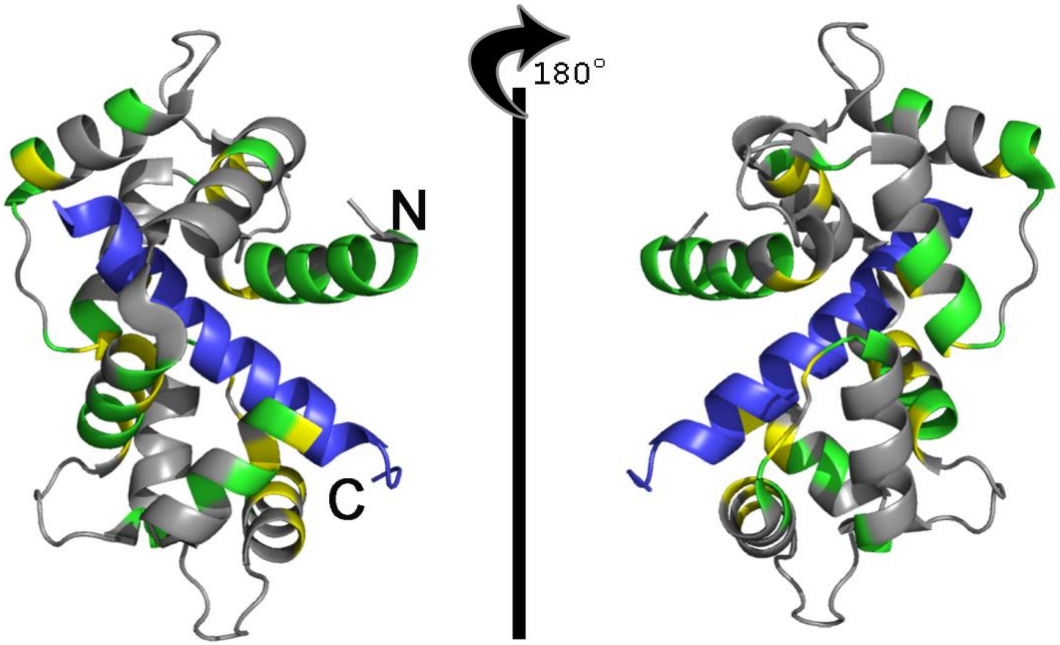
Αντίστοιχα με την εικόνα 3.4, στην εικόνα 3.5 απεικονίζεται το σύνολο των συνεξελισσόμενων καταλοίπων, όπως προέκυψαν από τη μέθοδο plmDCA. Με πράσινο απεικονίζονται τα προβλεπόμενα αποδεκτά συνεξελισσόμενα κατάλοιπα, ενώ με κόκκινο τα κατάλοιπα που απορρίφθηκαν, καθώς εντοπίζονται στην περιοχή πρόσδεσης των ιόντων ασβεστίου. Τα αποδεκτά κατάλοιπα (πράσινο χρώμα) χρησιμοποιήθηκαν για την περαιτέρω αξιολόγηση των αποτελεσμάτων με μεθόδους αγκυροβόλησης. Τα κατάλοιπα που ανήκουν ταυτόχρονα σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα δίνονται στην εικόνα 3.6.




**Εικόνα 3.4** Πειραματικά επιβεβαιωμένα κατάλοιπα αλληλεπίδρασης μεταξύ CaMK Ια - CaM (κωδικός PDB: 1MXE).




**Εικόνα 3.5** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο CaMK Ια - $Ca^{2+}$-CaM. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κόκκινο τα κατάλοιπα που απορρίφθηκαν.

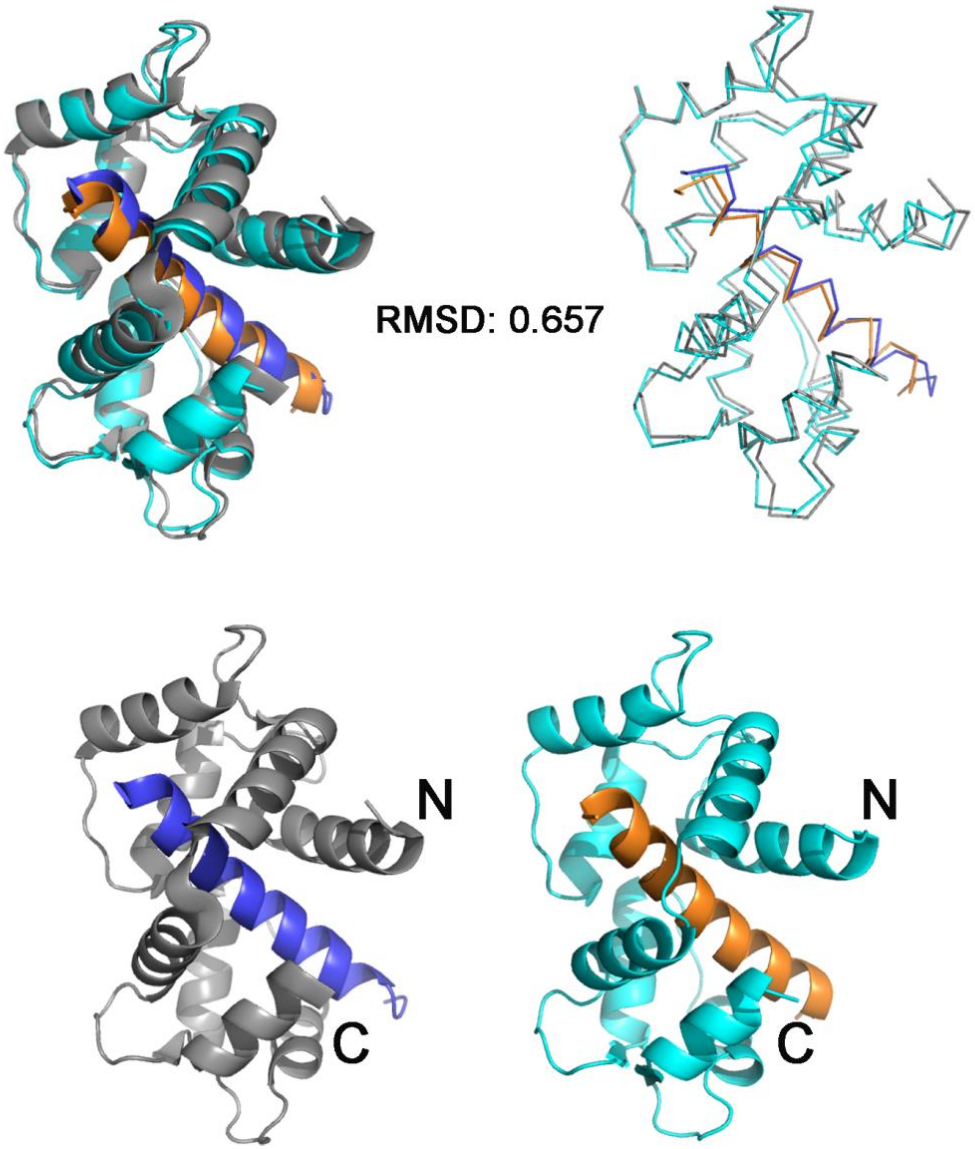



**Εικόνα 3.6** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο CaMK Ια - $Ca^{2+}$-CaM. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κίτρινο τα κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα (αληθώς θετική πρόβλεψη).

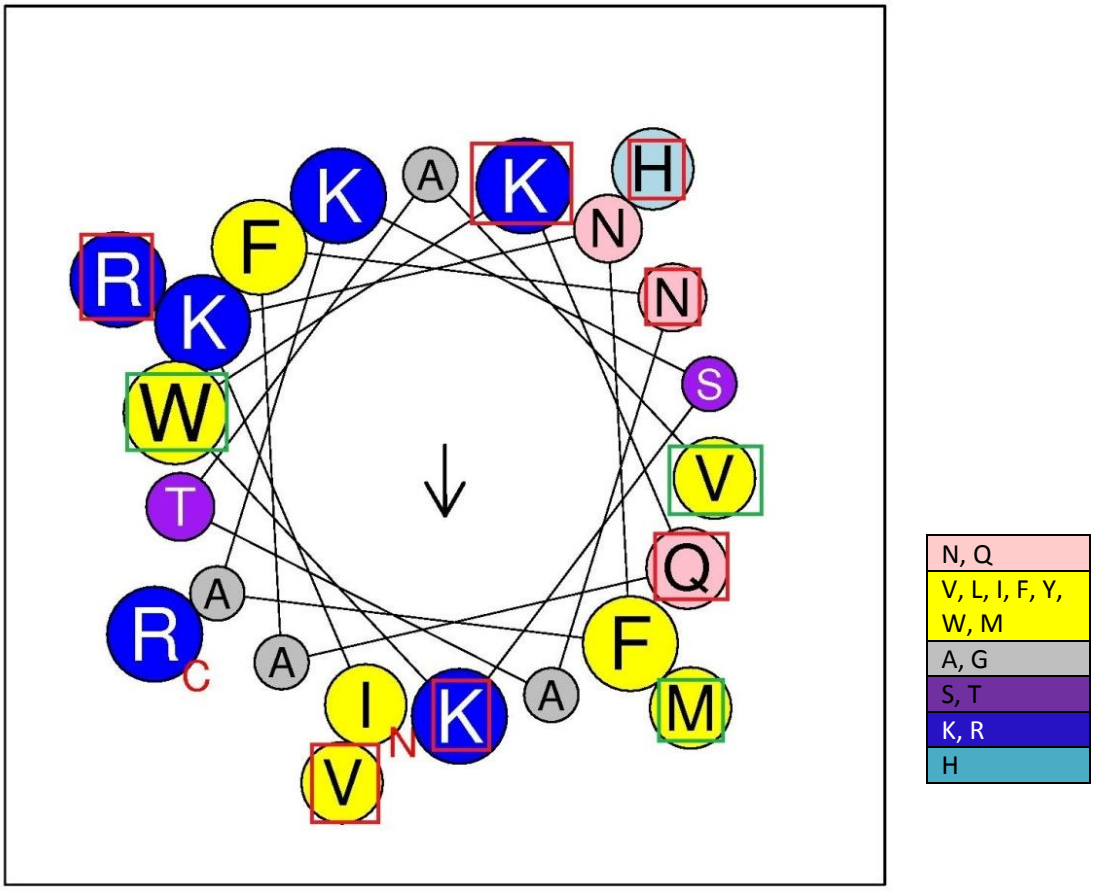
Η κατευθυνόμενη αγκυροβόληση βασισμένη στα αποτελέσματα του προγράμματος icoms επιβεβαίωσε την ορθή πρόβλεψη των συνεξελισσόμενων καταλοίπων, καθώς η καλύτερη δομή της πρώτης ομάδας αγκυροβόλησης με βάση τις συνολικές επαφές στην αύλακα αλληλεπίδρασης κινάσης-καλμοντουλίνης έχει RMSD μικρότερο του 1 (RMSD=0.657), όταν συγκρίνεται με την πειραματικά προσδιορισμένη δομή 1MXE. Η εικόνα 3.7 απεικονίζει τη δομική στοίχιση (αναπαράσταση καρτούν - επάνω αριστερά, αναπαράσταση Cα ανθράκων – επάνω δεξιά) μεταξύ της δομής του πρώτου cluster και της πειραματικά προσδιορισμένης δομής 1MXE. Αξίζει να σημειωθεί ότι στις ομάδες αγκυροβόλησης εντοπίστηκαν και άλλες ομάδες με πολύ καλύτερες ενέργειες αλληλεπίδρασης (όπως το 8$^{ο}$ cluster με ενέργεια καλύτερης δομής -32.35) όμως ο αριθμός των συνολικών επαφών δεν αφορούσαν την υπό μελέτη περιοχή ενδιαφέροντός μας μεταξύ κινάσης και καλμοντουλίνης (κριτήριο αποτέλεσαν οι συνολικές επαφές των καταλοίπων της κινάσης). Ιδιαίτερα ενδιαφέρον είναι πως δομές στις οποίες η α-έλικα δεν βρισκόταν ολόκληρη στο εσωτερικό του τούνελ της καλμοντουλίνης, αλλά ένα μέρος της βρισκόταν εκτός αυτού, αν και προφανώς δεν έδιναν ικανοποιητικό RMSD εμφάνιζαν καλύτερες ενέργειες αλληλεπίδρασης. Τα αποτελέσματα του για το σύμπλοκο CaMK Iα – $Ca^{2+}$-CaM παρουσιάζονται περιληπτικά στον παρακάτω πίνακα:

**Πίνακας 3.4.** Αποτελέσματα SwarmDock για το σύμπλοκο CaMK Iα – $Ca^{2+}$-CaM.

| **Κριτήριο αξιολόγησης** | **Τιμή** |
|---|---|
| **Clusters με ιεραρχική ομαδοποίηση στα 3 Å** | 126 |
| **Σύνολο δομών 1$^{ου}$ cluster** | 14 |
| **Μέση ενέργεια δομών του 1$^{ου}$ cluster** | -8.465 |
| **Τυπική απόκλιση της μέσης ενέργειας** | 5.136 |
| **Ενέργεια καλύτερης δομής του 1$^{ου}$ cluster** | -12.76 |
| **Συνολικές αλληλεπιδράσεις** | 719 |
| **Αλληλεπιδράσεις καταλοίπων CaM** | 386 |
| **Αλληλεπιδράσεις καταλοίπων κινάσης** | 719 |




**Εικόνα 3.7** Δομική στοίχιση με τη ρουτίνα align του προγράμματος μοριακών γραφικών PYMOL μεταξύ της κρυσταλλογραφικά προσδιορισμένης δομής του συμπλόκου CaMK Ια – $Ca^{2+}$-CaM (κινάση μπλε/καλμοντουλίνη γκρι) και της καλύτερης δομής του $1^{ου}$ cluster (κινάση πορτοκαλί/καλμοντουλίνη κυανό) έπειτα από την κατευθυνόμενη αγκυροβόληση.




**Εικόνα 3.8** Σε πράσινο τετράγωνο βρίσκονται τα κατάλοιπα που βιβλιογραφικά αναφέρονται ως κατάλοιπα αγκυροβόλησης, ενώ σε κόκκινο αυτά που έλαβαν τα πρώτα 20 υψηλότερα σκορ με βάση τη μέθοδο plmDCA.

Βιβλιογραφικά αναφέρεται το κατάλοιπο W303 ως κατάλοιπο απαραίτητο για την πρόσδεση της CaM (Swulius & Waxham, 2008). Σε κρυσταλλογραφικά προσδιορισμένες δομές της CaM KI (που όμως όπως αναφέρθηκε και εισαγωγικά δεν εντοπίζεται λυμένη ολόκληρη η αυτοτελής δομική περιοχή πρόσδεσης της CaM) το κατάλοιπο αυτό βρίσκεται σε μια θηλιά που συνδέει την α-έλικα της αυτοτελούς δομικής περιοχής αυτοαναστολής (aR1) με την α-έλικα της αυτοτελούς δομικής περιοχής πρόσδεσης της CaM (aR2) και κατά συνέπεια εντοπίζεται τελείως εκτεθειμένο. Έτσι αναφέρεται πως αποτελεί κατάλοιπο αγκυροβόλησης κατά την αλληλεπίδραση της C-τελικής περιοχής της CaM (κατάλοιπο M124) με την κινάση και την επακόλουθη αποδέσμευση του καταλυτικού κέντρου της κινάσης.

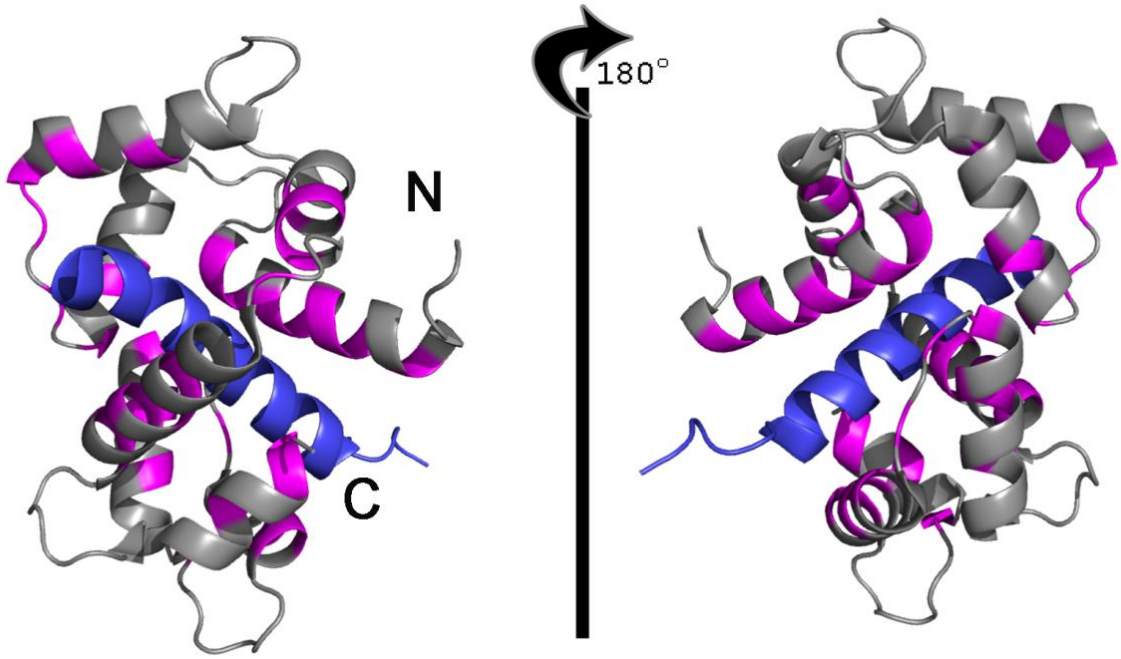
Παρατηρούμε πως και σε αυτήν την περίπτωση έχουν προβλεφθεί ως ισχυρά συνεξελισσόμενα τα γειτονικά της θρεονίνης, σε επίπεδο ακολουθίας, κατάλοιπα λυσίνης, καθώς και γειτονικά, στον χώρο, των καταλοίπων αγκυροβόλησης βαλίνη και μεθειονίνη. Το γεγονός αυτό ενισχύει τη σημασία πρόβλεψης γειτονικών κατάλοιπων των σημαντικών για την αλληλεπίδραση καταλοίπων.

Υπενθυμίζουμε σε αυτό το σημείο πως κατά την ανενεργή κατάσταση της κινάσης η αR1 αλληλεπιδρά με αμινοξικά κατάλοιπα της περιοχής πρόσδεσης του υποστρώματος ενώ η αR2 με κατάλοιπα της περιοχής πρόσδεσης του ATP, ενώ μεταξύ τους εντοπίζεται η θηλιά που προαναφέρθηκε. Στην αR2 εντοπίζονται και τα κατάλοιπα αγκυροβόλησης V και M όπως απεικονίζονται στην εικόνα 3.8. Το κατάλοιπο μεθειονίνης στην ανενεργή κατάσταση της κινάσης βρίσκεται θαμμένο στην περιοχή πρόσδεσης ATP. Συμπεραίνουμε πως μετά την πρόσδεση του C-τελικού άκρου της CaM στο N-τελικό κατάλοιπο αγκυροβόλησης W (καθώς η αλληλεπίδραση συμβαίνει με αντιπαράλληλο προσανατολισμό) κομβική, για την περαιτέρω ενεργοποίηση της κινάσης, είναι η αλληλεπίδραση της N-τελικής περιοχής της CaM με το αντίστοιχο C-τελικό κατάλοιπο αγκυροβόλησης M της κινάσης. Στις στερεοδιατακτικές αλλαγές που απαιτούνται για να πραγματοποιηθεί η αλληλεπίδραση αυτή και η απομάκρυνση του C-τελικού άκρου της aR2 από την περιοχή πρόσδεσης του ATP μπορεί να συνεισφέρουν τα συνεξελισσόμενα κατάλοιπα που εντοπίζονται προς το N-τελικό άκρο της M και πιθανά να εντοπίζονται με τις πλευρικές αλυσίδες τους εκτεθειμένες στην επιφάνεια του καταλυτικού κέντρου της κινάσης. Η υπόθεση αυτή δε μπορεί να επιβεβαιωθεί, καθώς όπως προαναφέρθηκε στο σύνολο των λυμένων δομών δεν εντοπίζεται ολόκληρη η α-έλικα πρόσδεσης της $Ca^{2+}$-CaM ή ακόμα και αν εντοπίζεται (PDB 4FG9) η κινάση βρίσκεται στην ενεργή κατάσταση με το ATP προσδεμένο και δε μπορούμε να γνωρίζουμε την ακριβή θέση των καταλοίπων στην apo στερεοδιάταξη της κινάσης.

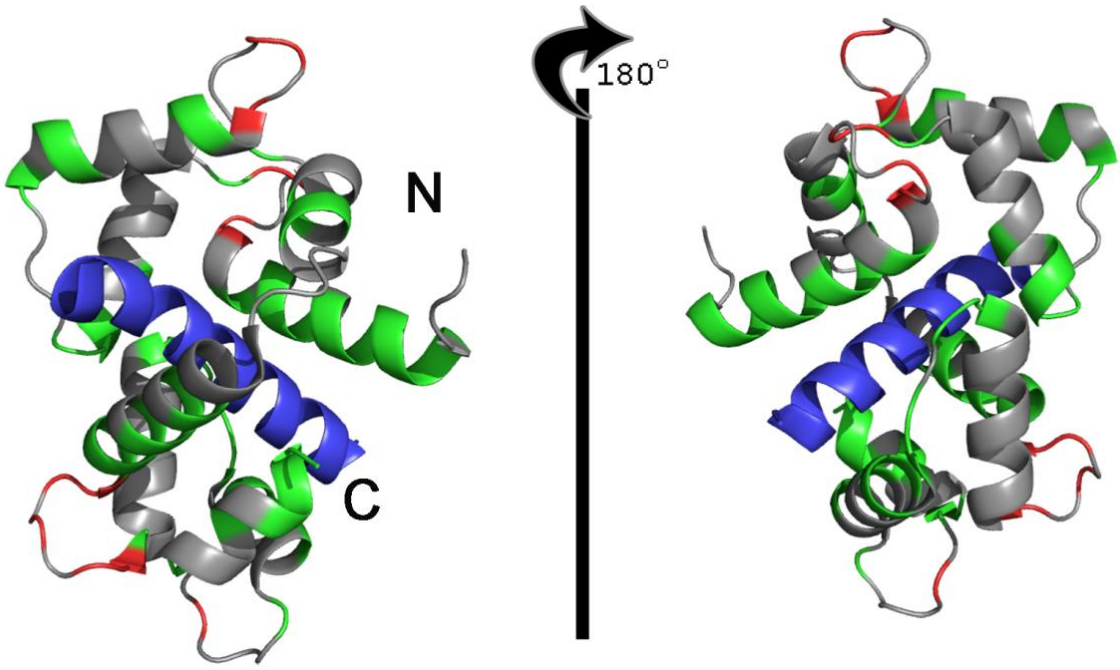
### 3.4.2. Μελέτη συμπλόκου CaMK IIδ – $Ca^{2+}$-CaM2 και συμπεράσματα

Στην εικόνα 3.9 απεικονίζεται με αναπαράσταση cartoon το σύμπλοκο CaMK IIδ – $Ca^{2+}$-CaM2 (κωδικός PDB: 2WEL). Με γκρι απεικονίζεται η καλμοντουλίνη CaM2, ενώ με μπλε απεικονίζεται η κινάση CaMK IIδ. Επισημαίνονται με magenta οι θέσεις των καταλοίπων που συμμετέχουν στην αλληλεπίδραση των δύο μορίων με βάση τις default παραμέτρους του προγράμματος RING, από όπου συγκεντρώθηκαν τα ζευγάρια των καταλοίπων που αλληλεπιδρούν καθώς και ο χαρακτηρισμός της φύσης της κάθε αλληλεπίδρασης.

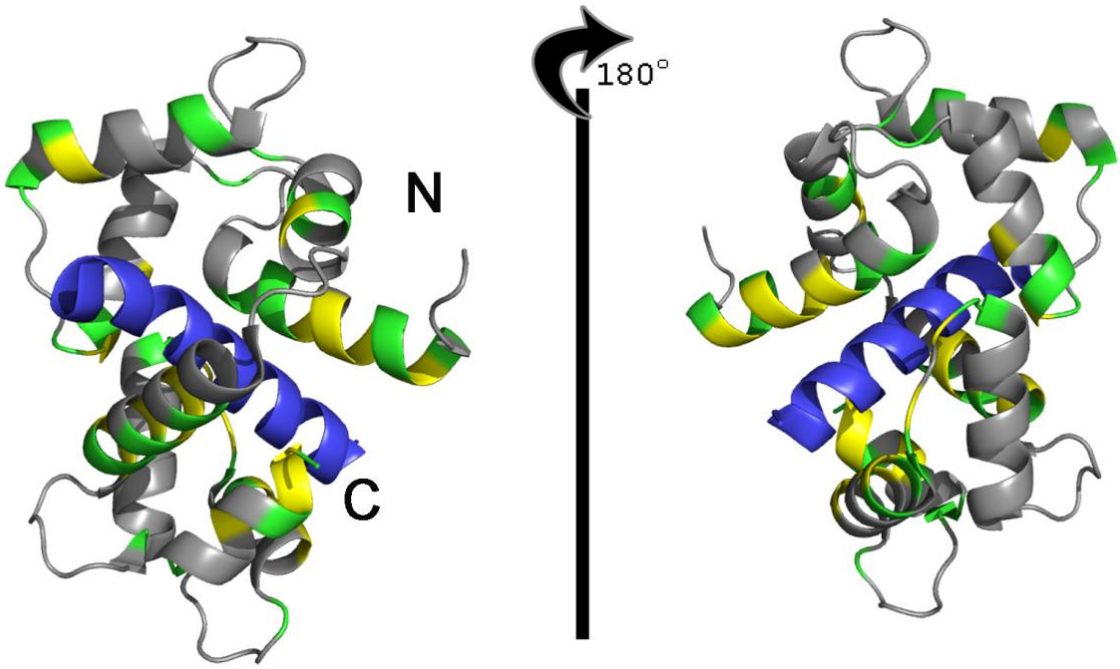
Αντίστοιχα με την εικόνα 3.9, στην εικόνα 3.10 απεικονίζεται το σύνολο των συνεξελισσόμενων καταλοίπων όπως προέκυψαν από τη μέθοδο plmDCA. Με πράσινο απεικονίζονται τα προβλεπόμενα αποδεκτά συνεξελισσόμενα κατάλοιπα, ενώ με κόκκινο τα κατάλοιπα που απορρίφθηκαν, καθώς εντοπίζονται στην περιοχή πρόσδεσης των ιόντων ασβεστίου. Τα αποδεκτά κατάλοιπα (πράσινο χρώμα) χρησιμοποιήθηκαν για την περαιτέρω αξιολόγηση των αποτελεσμάτων με μεθόδους αγκυροβόλησης. Τα κατάλοιπα που ανήκουν ταυτόχρονα σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα δίνονται στην εικόνα 3.11.




**Εικόνα 3.9** Πειραματικά επιβεβαιωμένα κατάλοιπα αλληλεπίδρασης μεταξύ CaMK IIδ – Ca2+-CaM (κωδικός PDB: 2WEL).




**Εικόνα 3.10** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο CaMK IIδ - CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κόκκινο τα κατάλοιπα που απορρίφθηκαν.

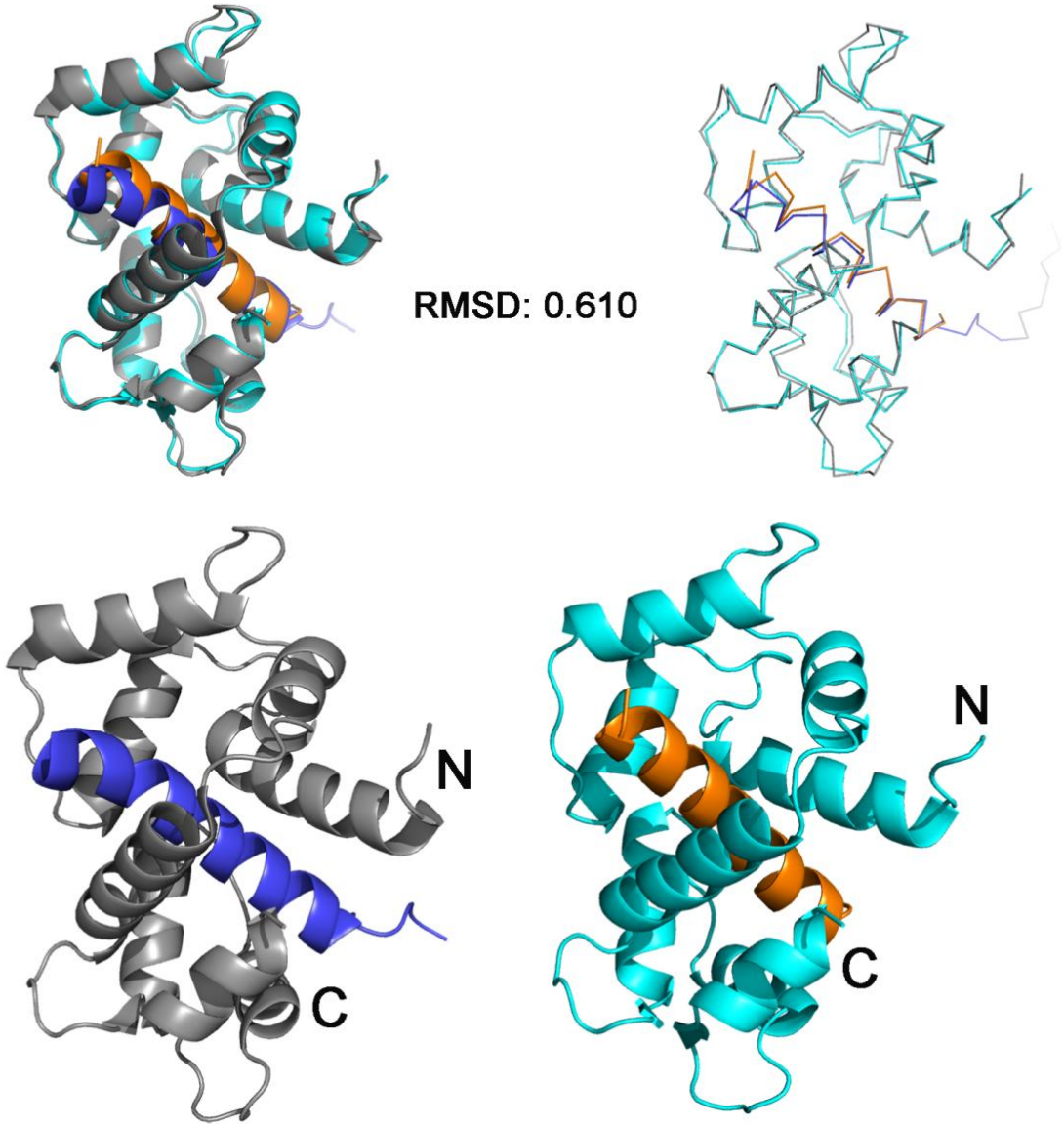



**Εικόνα 3.11** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο CaMK IIδ - CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κίτρινο τα κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα (αληθώς θετική πρόβλεψη).

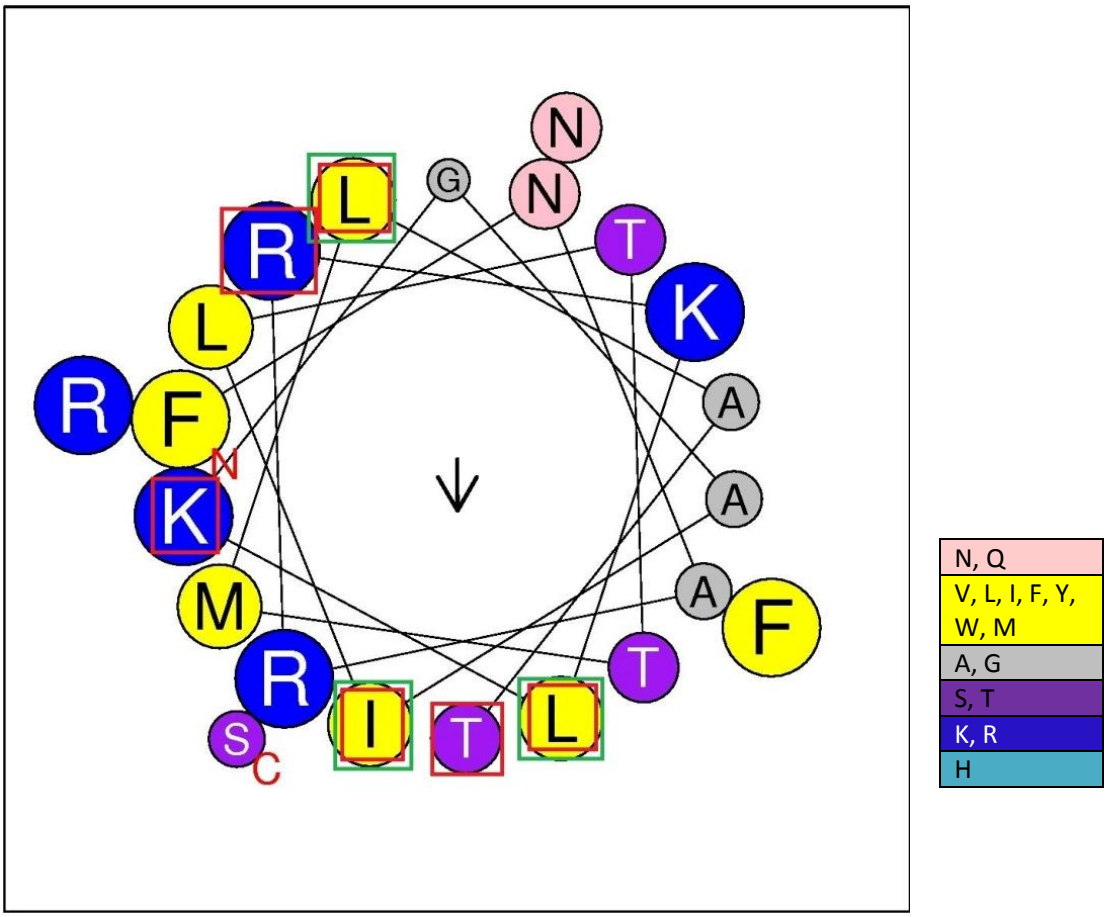
Η κατευθυνόμενη αγκυροβόληση βασισμένη στα αποτελέσματα του προγράμματος i-coms επιβεβαίωσε την ορθή πρόβλεψη των συνεξελισσόμενων καταλοίπων, καθώς η καλύτερη δομή της πρώτης ομάδας αγκυροβόλησης με βάση τις συνολικές επαφές στην αύλακα αλληλεπίδρασης κινάσης-καλμοντουλίνης έχει RMSD μικρότερο του 1 (RMSD=0.610), όταν συγκρίνεται με την πειραματικά προσδιορισμένη δομή 2WEL. Η εικόνα 3.12 απεικονίζει τη δομική στοίχιση (αναπαράσταση καρτούν - επάνω αριστερά, αναπαράσταση Cα ανθράκων – επάνω δεξιά) μεταξύ της δομής του πρώτου cluster και της πειραματικά προσδιορισμένης δομής 2WEL. Χρησιμοποιήθηκε το ίδιο κριτήριο (συνολικές επαφές καταλοίπων κινάσης), όμως σε αυτή την περίπτωση η δομή αυτή παρουσίαζε την καλύτερη δυνατή στερεοχημική ποιότητα και τη βέλτιστη ενέργεια αλληλεπίδρασης. Τα αποτελέσματα του για το σύμπλοκο CaMK IIδ - $Ca^{2+}$-CaM2 παρουσιάζονται περιληπτικά στον παρακάτω πίνακα:

**Πίνακας 3.5.** Αποτελέσματα SwarmDock για το σύμπλοκο CaMK IIδ - $Ca^{2+}$-CaM2.

| **Κριτήριο αξιολόγησης** | **Τιμή** |
|---|---|
| **Clusters με ιεραρχική ομαδοποίηση στα 3 Å** | 135 |
| **Σύνολο δομών 1$^{ου}$ cluster** | 31 |
| **Μέση ενέργεια δομών του 1$^{ου}$ cluster** | -51.859 |
| **Τυπική απόκλιση της μέσης ενέργειας** | 5.058 |
| **Ενέργεια καλύτερης δομής του 1$^{ου}$ cluster** | -62.20 |
| **Συνολικές αλληλεπιδράσεις** | 696 |
| **Αλληλεπιδράσεις καταλοίπων CaM** | 385 |
| **Αλληλεπιδράσεις καταλοίπων κινάσης** | 643 |




**Εικόνα 3.12** Δομική στοίχιση με τη ρουτίνα align του προγράμματος μοριακών γραφικών PYMOL μεταξύ της κρυσταλλογραφικά προσδιορισμένης δομής του συμπλόκου CaMK IIδ - $Ca^{2+}$-CaM2 (κινάση μπλε/καλμοντουλίνη γκρι) και της καλύτερης δομής του $1^{ου}$ cluster (κινάση πορτοκαλί/καλμοντουλίνη κυανό) έπειτα από την κατευθυνόμενη αγκυροβόληση.



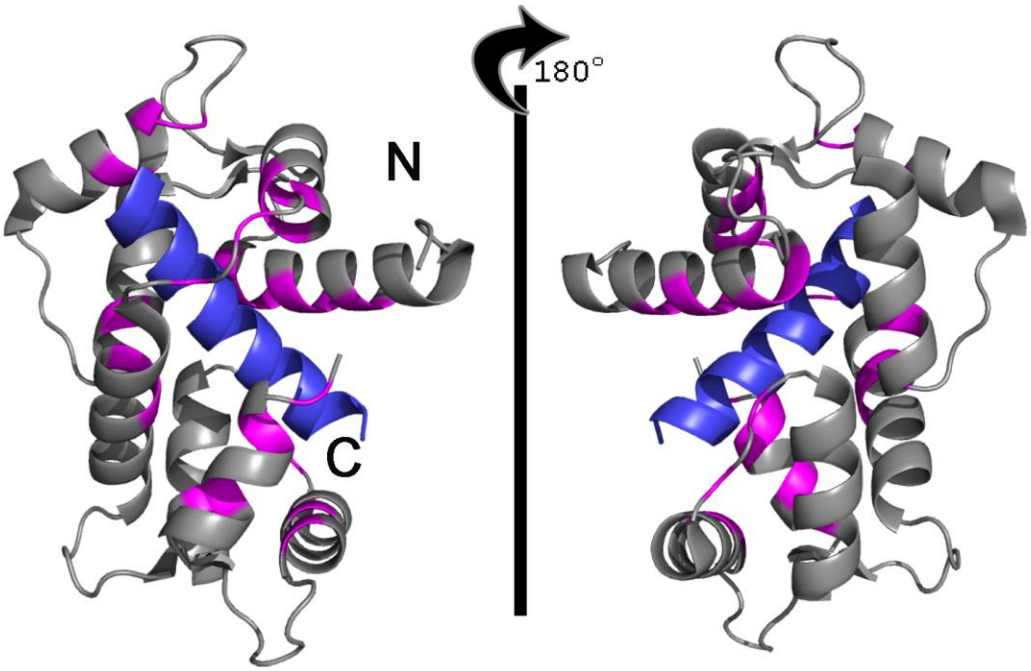

**Εικόνα 3.13** Σε πράσινο τετράγωνο βρίσκονται τα κατάλοιπα που βιβλιογραφικά αναφέρονται ως κατάλοιπα αγκυροβόλησης, ενώ σε κόκκινο αυτά που έλαβαν τα πρώτα 20 υψηλότερα σκορ με βάση τη μέθοδο plmDCA.

Στην περίπτωση αυτή, που η ακολουθία των πεπτιδίων-στόχων ακολουθεί την λίγο διαφορετική τοποθέτηση καταλοίπων 1-5-10 προβλέφθηκε το σύνολο των καταλοίπων αγκυροβόλησης. Προβλέφθηκαν επίσης, όπως στις υπόλοιπες περιπτώσεις, γειτονικά στο χώρο κατάλοιπα ως συνεξελισσόμενα. Η ύπαρξη καταλοίπων αγκυροβόλησης στις συγκεκριμένες θέσεις έχει ως αποτέλεσμα τη δημιουργία πιο συμπαγούς συμπλόκου, με τις 2 αυτοτελείς δομικές περιοχές της $Ca^{2+}$-CaM να εντοπίζονται κατά μια στροφή της έλικας πιο κοντά στο χώρο, όπως γίνεται αντιληπτό και από τις απεικονίσεις των δομών που παρατίθενται, γεγονός που πιθανά διευκολύνει τη συνολική στερεοδιάταξη του δωδεκαμερούς.

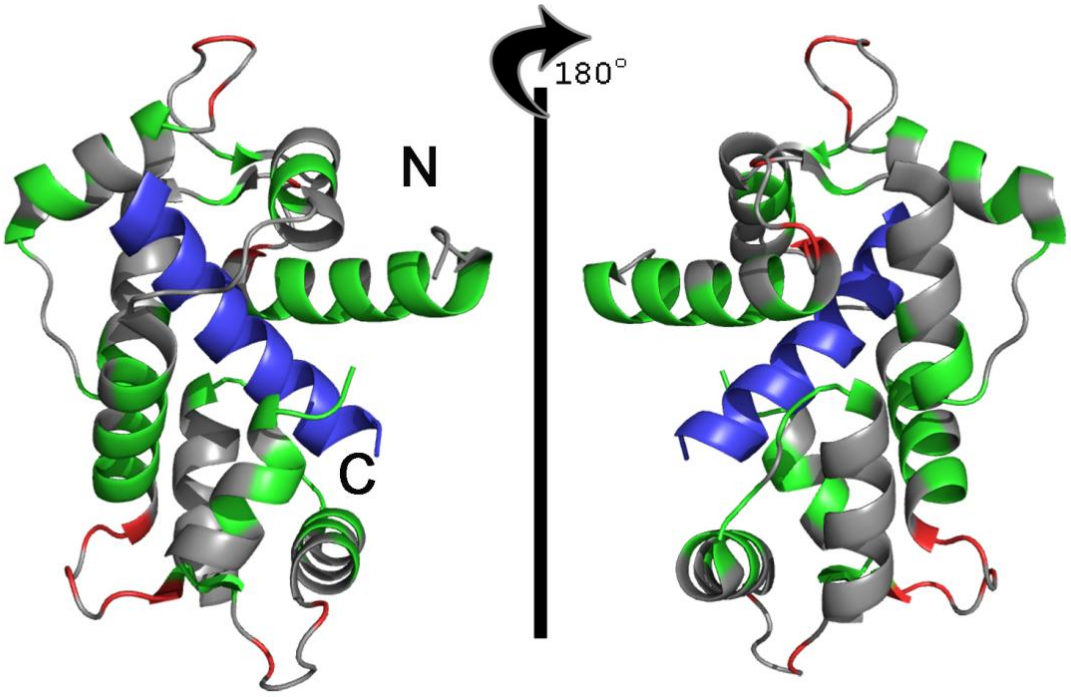
### 3.4.3. Μελέτη του συμπλόκου MLCK I – $Ca^{2+}$-CaM και συμπεράσματα

Στην εικόνα 3.14 απεικονίζεται με αναπαράσταση cartoon το σύμπλοκο MLCK I- $Ca^{2+}$-CaM1 (κωδικός PDB: 5JQA). Με γκρι απεικονίζεται η καλμοντουλίνη CaM1, ενώ με μπλε απεικονίζεται η κινάση MLCK. Επισημαίνονται με magenta οι θέσεις των καταλοίπων που συμμετέχουν στην αλληλεπίδραση των δύο μορίων με βάση τις default παραμέτρους του προγράμματος RING, από όπου συγκεντρώθηκαν τα ζευγάρια των καταλοίπων που αλληλεπιδρούν καθώς και ο χαρακτηρισμός της φύσης της κάθε αλληλεπίδρασης.

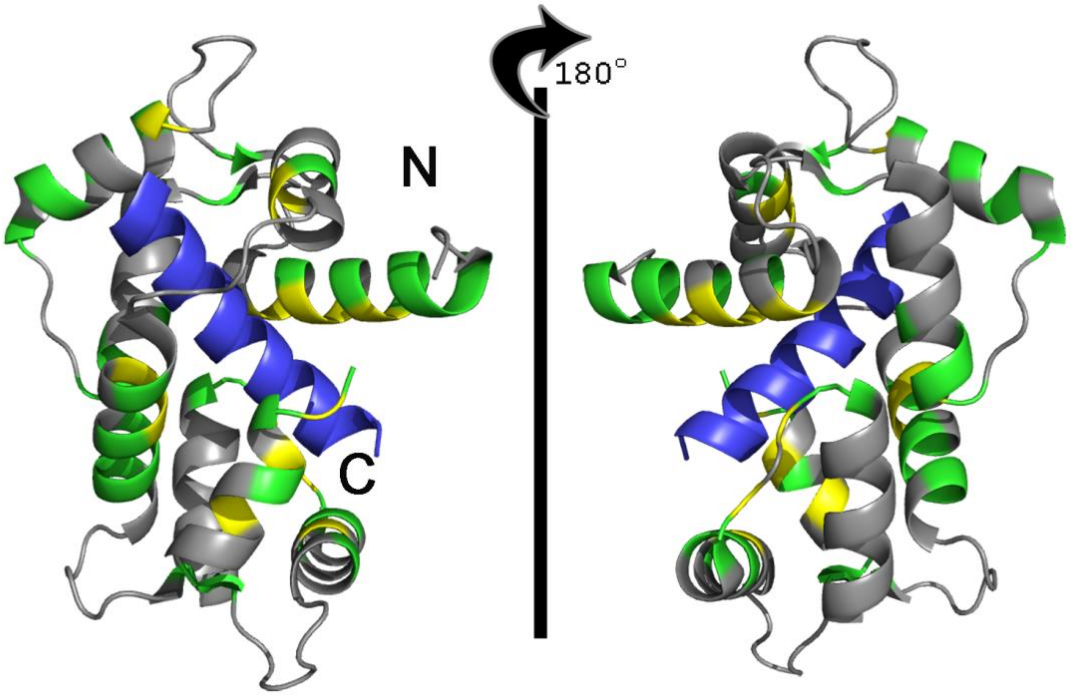
Αντίστοιχα με την εικόνα 3.14, στην εικόνα 3.15 απεικονίζεται το σύνολο των συνεξελισσόμενων καταλοίπων όπως προέκυψαν από τη μέθοδο plmDCA. Με πράσινο απεικονίζονται τα προβλεπόμενα αποδεκτά συνεξελισσόμενα κατάλοιπα, ενώ με κόκκινο τα κατάλοιπα που απορρίφθηκαν, καθώς εντοπίζονται στην περιοχή πρόσδεσης των ιόντων ασβεστίου. Τα αποδεκτά κατάλοιπα (πράσινο χρώμα) χρησιμοποιήθηκαν για την περαιτέρω αξιολόγηση των αποτελεσμάτων με μεθόδους αγκυροβόλησης. Τα κατάλοιπα που ανήκουν ταυτόχρονα σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα δίνονται στην εικόνα 3.16.




**Εικόνα 3.14** Πειραματικά επιβεβαιωμένα κατάλοιπα αλληλεπίδρασης μεταξύ MLCK - Ca2+-CaM1 (κωδικός PDB: 5JQA).




**Εικόνα 3.15** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο MLCK I – $Ca^{2+}$-CaM1. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κόκκινο τα κατάλοιπα που απορρίφθηκαν.

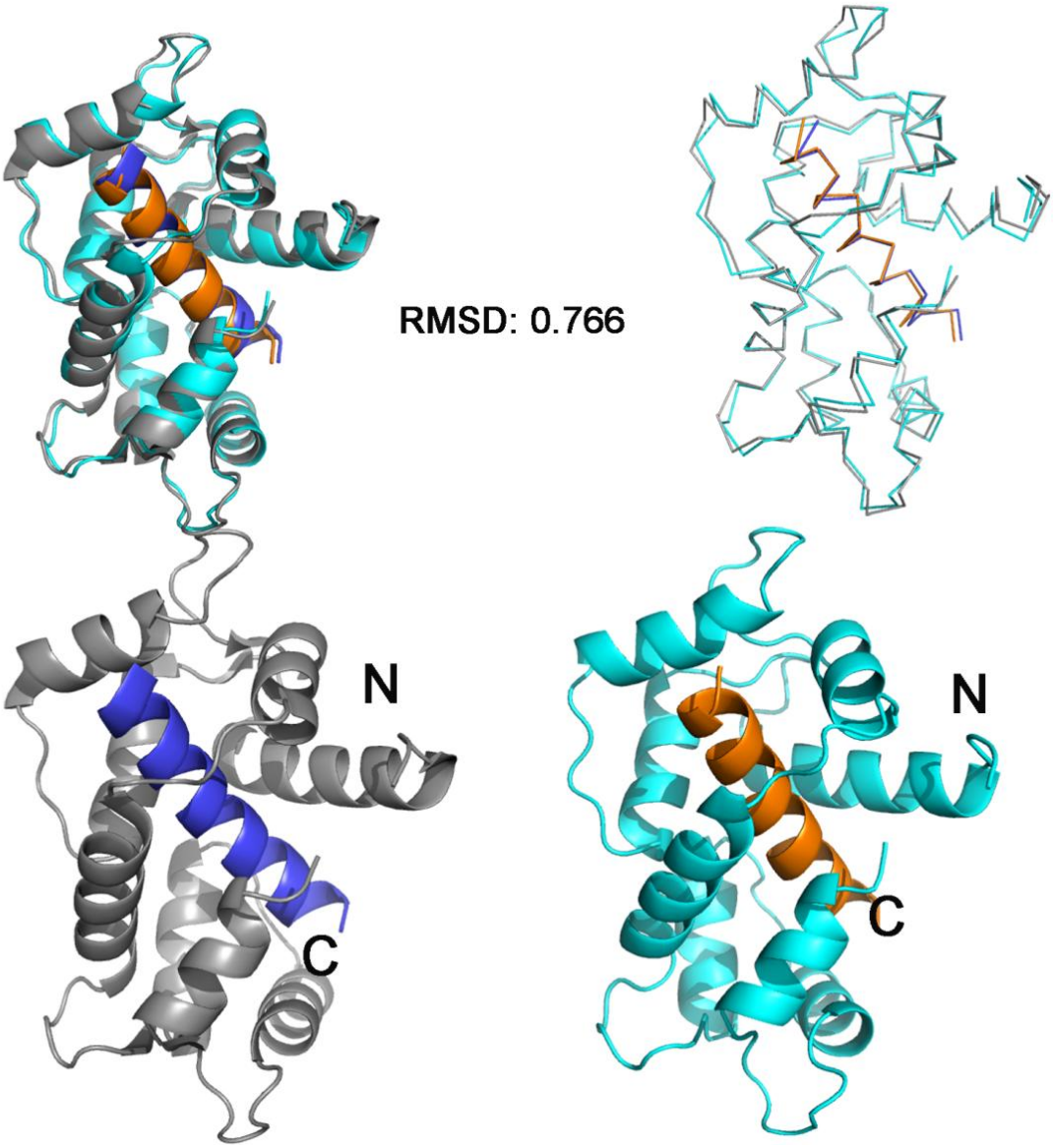



**Εικόνα 3.16** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο MLCK I – Ca2+-CaM1. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κίτρινο τα κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα (αληθώς θετική πρόβλεψη).

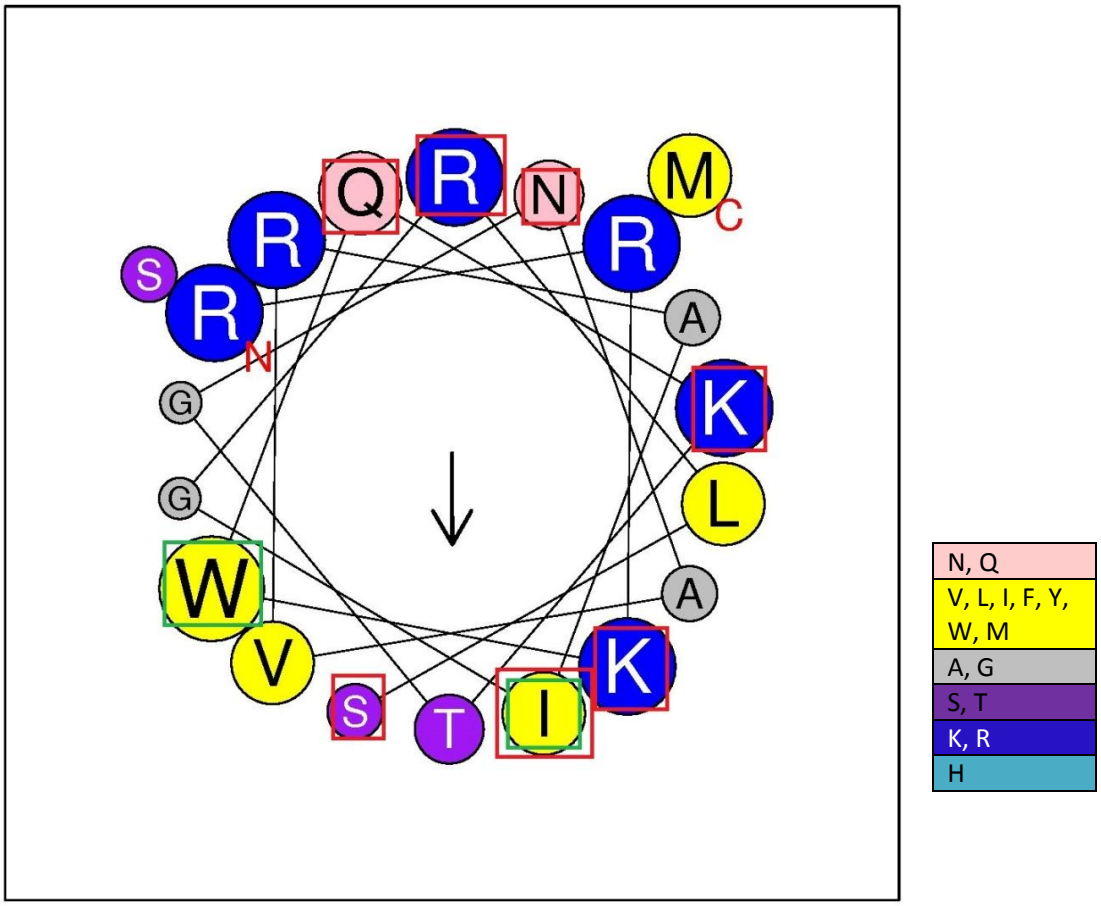
Η κατευθυνόμενη αγκυροβόληση βασισμένη στα αποτελέσματα του προγράμματος icoms επιβεβαίωσε την ορθή πρόβλεψη των συνεξελισσόμενων καταλοίπων, καθώς η καλύτερη δομή της πρώτης ομάδας αγκυροβόλησης με βάση τις συνολικές επαφές στην αύλακα αλληλεπίδρασης κινάσης-καλμοντουλίνης έχει RMSD μικρότερο του 1 (RMSD=0.766), όταν συγκρίνεται με την πειραματικά προσδιορισμένη δομή 5JQA. Η εικόνα 3.17 απεικονίζει τη δομική στοίχιση (αναπαράσταση καρτούν - επάνω αριστερά, αναπαράσταση Cα ανθράκων – επάνω δεξιά) μεταξύ της δομής του πρώτου cluster και της πειραματικά προσδιορισμένης δομής 5JQA. Χρησιμοποιήθηκε το ίδιο κριτήριο (συνολικές επαφές καταλοίπων κινάσης), όμως σε αυτή την περίπτωση η δομή αυτή παρουσίαζε την καλύτερη δυνατή στερεοχημική ποιότητα και τη βέλτιστη ενέργεια αλληλεπίδρασης. Αυτό μπορεί να οφείλεται στην ελαφρά πιο ανοιχτή στερεοδιάταξη που εμφανίζει στην συγκεκριμένη περίπτωση η καλμοντουλίνη. Τα αποτελέσματα του για το σύμπλοκο MLCK - Ca2+-CaM1 παρουσιάζονται περιληπτικά στον παρακάτω πίνακα:

**Πίνακας 3.6.** Αποτελέσματα SwarmDock για το σύμπλοκο MLCK - Ca2+-CaM1.

| **Κριτήριο αξιολόγησης** | **Τιμή** |
|---|---|
| **Clusters με ιεραρχική ομαδοποίηση στα 3 Å** | 97 |
| **Σύνολο δομών 1$^{ου}$ cluster** | 4 |
| **Μέση ενέργεια δομών του 1$^{ου}$ cluster** | -35.995 |
| **Τυπική απόκλιση της μέσης ενέργειας** | 1.709 |
| **Ενέργεια καλύτερης δομής του 1$^{ου}$ cluster** | -36.99 |
| **Συνολικές αλληλεπιδράσεις** | 831 |
| **Αλληλεπιδράσεις καταλοίπων CaM** | 351 |
| **Αλληλεπιδράσεις καταλοίπων κινάσης** | 831 |




**Εικόνα 3.17** Δομική στοίχιση με τη ρουτίνα align του προγράμματος μοριακών γραφικών PYMOL μεταξύ της κρυσταλλογραφικά προσδιορισμένης δομής του συμπλόκου MLCK I – Ca2+-CaM1 (κινάση μπλε/καλμοντουλίνη γκρι) και της καλύτερης δομής του 1ου cluster (κινάση πορτοκαλί/καλμοντουλίνη κυανό) έπειτα από την κατευθυνόμενη αγκυροβόληση.



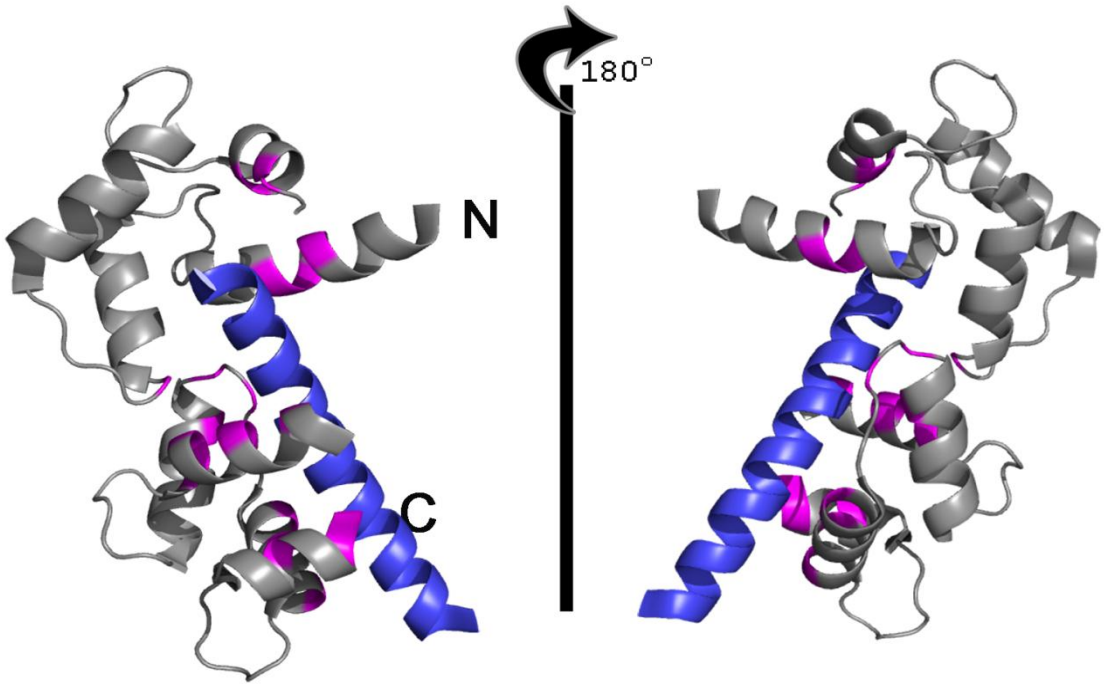

**Εικόνα 3.18** Σε πράσινο τετράγωνο βρίσκονται τα κατάλοιπα που βιβλιογραφικά αναφέρονται ως κατάλοιπα αγκυροβόλησης, ενώ σε κόκκινο αυτά που έλαβαν τα πρώτα 20 υψηλότερα σκορ με βάση τη μέθοδο plmDCA.

Σε αυτήν την περίπτωση παρατηρούμε να προβλέπεται ως συνεξελισσόμενα το κατάλοιπο αγκυροβόλησης I. Εκτός από αυτό για το κατάλοιπο αγκυροβόλησης του αμινοτελικού άκρου W προβλέπονται γειτονικά κατάλοιπα σε επίπεδο ακολουθίας, ενώ για το κατάλοιπο I του καρβοξυτελικού άκρου προβλέπονται γειτονικά κατάλοιπα σε επίπεδο δομής.

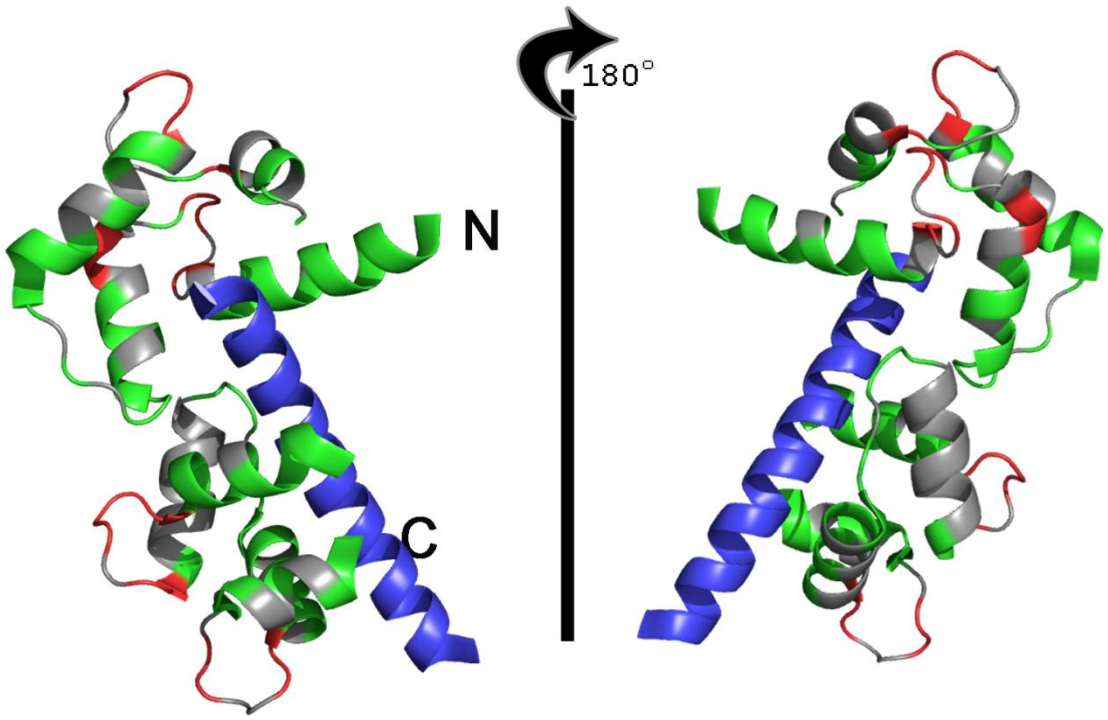
### 3.4.4. Μελέτη συμπλόκου DAPK1 – $Ca^{2+}$-CaM και συμπεράσματα

Στην εικόνα 3.19 απεικονίζεται με αναπαράσταση cartoon το σύμπλοκο DAPK1 - $Ca^{2+}$-CaM2 (κωδικός PDB: 2X0G). Με γκρι απεικονίζεται η καλμοντουλίνη CaM2, ενώ με μπλε απεικονίζεται η κινάση DAPK1. Επισημαίνονται με magenta οι θέσεις των καταλοίπων που συμμετέχουν στην αλληλεπίδραση των δύο μορίων με βάση τις default παραμέτρους του προγράμματος RING, από όπου συγκεντρώθηκαν τα ζευγάρια των καταλοίπων που αλληλεπιδρούν καθώς και ο χαρακτηρισμός της φύσης της κάθε αλληλεπίδρασης.

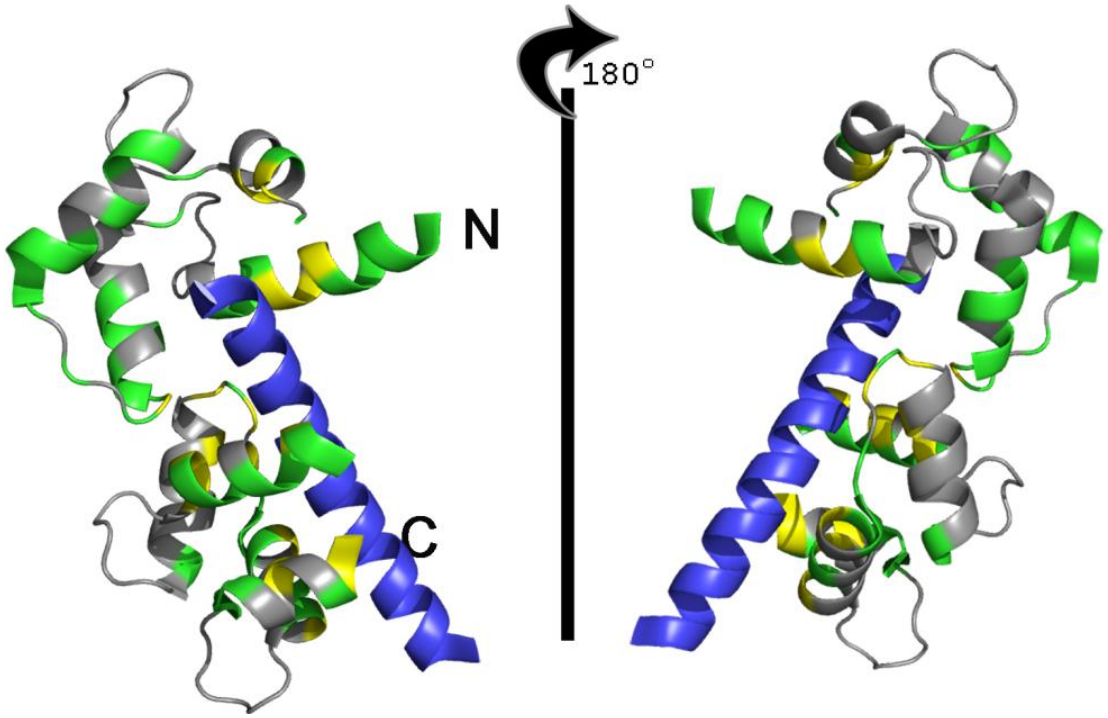
Αντίστοιχα με την εικόνα 3.19, στην εικόνα 3.20 απεικονίζεται το σύνολο των συνεξελισσόμενων καταλοίπων όπως προέκυψαν από τη μέθοδο plmDCA. Με πράσινο απεικονίζονται τα προβλεπόμενα αποδεκτά συνεξελισσόμενα κατάλοιπα, ενώ με κόκκινο τα κατάλοιπα που απορρίφθηκαν, καθώς εντοπίζονται στην περιοχή πρόσδεσης των ιόντων ασβεστίου. Τα αποδεκτά κατάλοιπα (πράσινο χρώμα) χρησιμοποιήθηκαν για την περαιτέρω αξιολόγηση των αποτελεσμάτων με μεθόδους αγκυροβόλησης. Τα κατάλοιπα που ανήκουν ταυτόχρονα σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα δίνονται στην εικόνα 3.21.




**Εικόνα 3.19** Πειραματικά επιβεβαιωμένα κατάλοιπα αλληλεπίδρασης μεταξύ DAPK1 - $Ca^{2+}$-CaM2 (κωδικός PDB: 2X0G).




**Εικόνα 3.20** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο DAPK1 - $Ca^{2+}$-CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κόκκινο τα κατάλοιπα που απορρίφθηκαν.

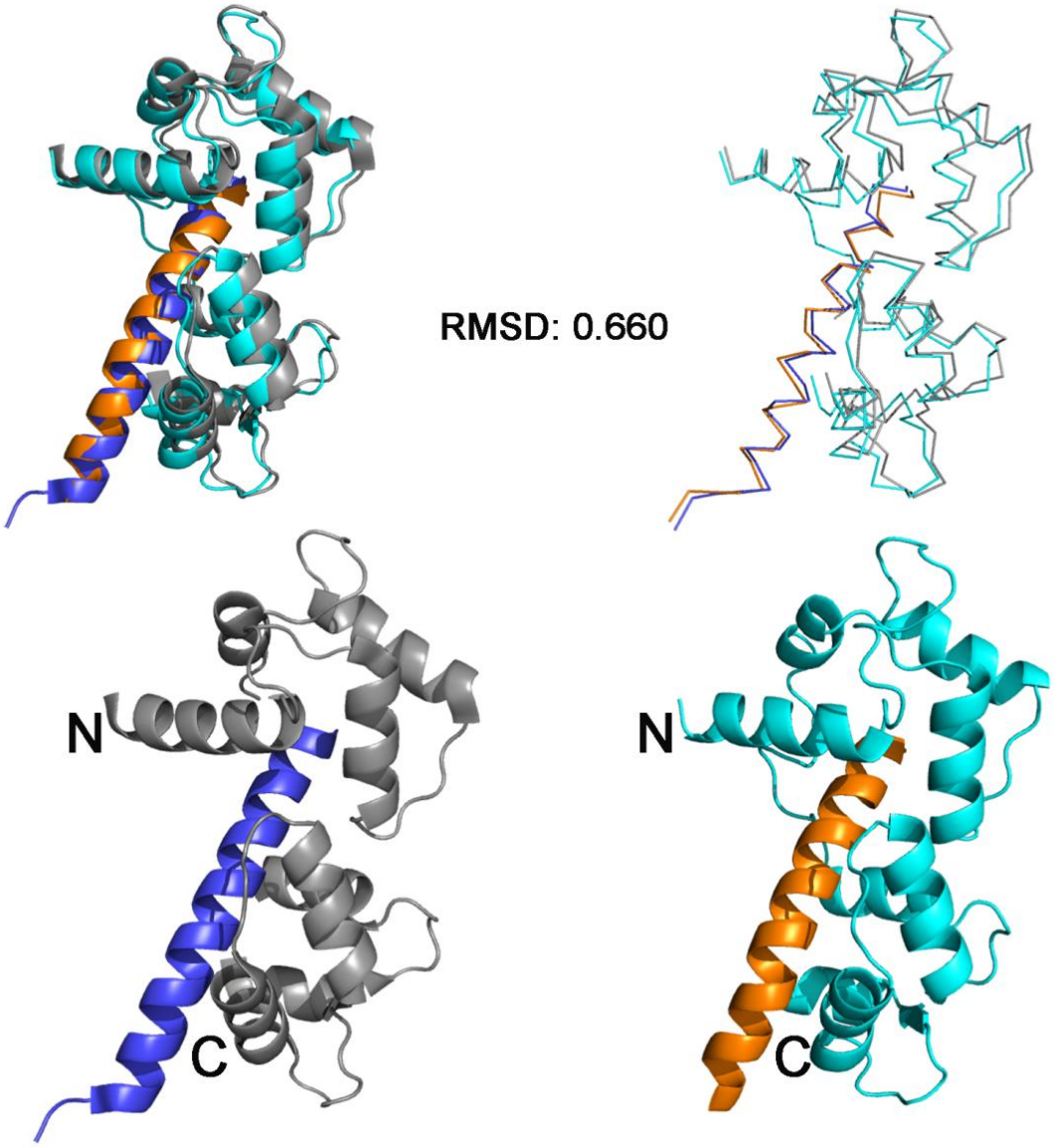



**Εικόνα 3.21** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο DAPK1 - $Ca^{2+}$-CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κίτρινο τα κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα (αληθώς θετική πρόβλεψη).

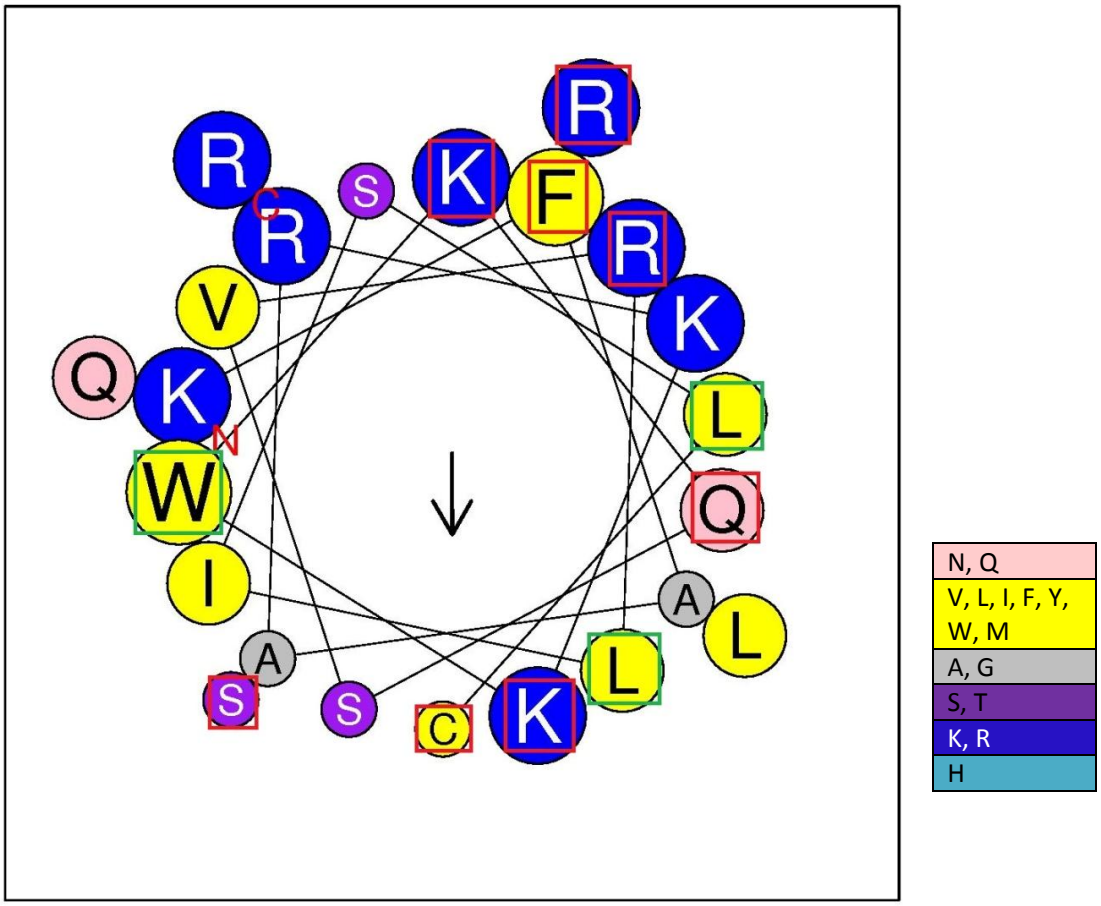
Η κατευθυνόμενη αγκυροβόληση βασισμένη στο αποτέλεσμα του προγράμματος icoms επιβεβαίωση την ορθή πρόβλεψη των συνεξελισσόμενων καταλοίπων, καθώς η καλύτερη δομή της πρώτης ομάδας αγκυροβόλησης με βάση τις συνολικές επαφές στην αύλακα αλληλεπίδρασης κινάσης-καλμοντουλίνης έχει RMSD μικρότερο του 1 (RMSD=0.660), όταν συγκρίνεται με την πειραματικά προσδιορισμένη δομή 2X0G. Η εικόνα 3.22 απεικονίζει τη δομική στοίχιση (αναπαράσταση καρτούν - επάνω αριστερά, αναπαράσταση Cα ανθράκων – επάνω δεξιά) μεταξύ της δομής του πρώτου cluster και της πειραματικά προσδιορισμένης δομής 2X0G. Αξίζει να σημειωθεί ότι στις ομάδες αγκυροβόλησης εντοπίστηκαν και άλλες ομάδες με πολύ καλύτερες ενέργειες αλληλεπίδρασης (βέλτιστη ενέργεια -44.88) όμως ο αριθμός των συνολικών επαφών δεν αφορούσαν την υπό μελέτη περιοχή ενδιαφέροντός μας μεταξύ κινάσης και καλμοντουλίνης (κριτήριο αποτέλεσαν οι συνολικές επαφές των καταλοίπων της κινάσης). Τα αποτελέσματα του για το σύμπλοκο DAPK1 - $Ca^{2+}$-CaM2 παρουσιάζονται περιληπτικά στον παρακάτω πίνακα:

**Πίνακας 3.7.** Αποτελέσματα SwarmDock για το σύμπλοκο DAPK1 - $Ca^{2+}$-CaM2.

| Κριτήριο αξιολόγησης | Τιμή |
|---|---|
| **Clusters με ιεραρχική ομαδοποίηση στα 3 Å** | 180 |
| **Σύνολο δομών 1$^{ου}$ cluster** | 4 |
| **Μέση ενέργεια δομών του 1$^{ου}$ cluster** | -46.265 |
| **Τυπική απόκλιση της μέσης ενέργειας** | 2.059 |
| **Ενέργεια καλύτερης δομής του 1$^{ου}$ cluster** | -42.95 |
| **Συνολικές αλληλεπιδράσεις** | 1024 |
| **Αλληλεπιδράσεις καταλοίπων CaM** | 399 |
| **Αλληλεπιδράσεις καταλοίπων κινάσης** | 805 |




**Εικόνα 3.22** Δομική στοίχιση με τη ρουτίνα align του προγράμματος μοριακών γραφικών PYMOL μεταξύ της κρυσταλλογραφικά προσδιορισμένης δομής του συμπλόκου DAPK1 - $Ca^{2+}$-CaM2 (κινάση μπλε/καλμοντουλίνη γκρι) και της καλύτερης δομής του $1^{ου}$ cluster (κινάση πορτοκαλί/καλμοντουλίνη κυανό) έπειτα από την κατευθυνόμενη αγκυροβόληση.




**Εικόνα 3.23** Σε πράσινο τετράγωνο βρίσκονται τα κατάλοιπα που βιβλιογραφικά αναφέρονται ως κατάλοιπα αγκυροβόλησης, ενώ σε κόκκινο αυτά που έλαβαν τα πρώτα 20 υψηλότερα σκορ με βάση τη μέθοδο plmDCA.

Παρατηρείται το εξής ενδιαφέρον χαρακτηριστικό. Σύμπλοκο της CaM με πεπτίδιο που προέρχεται από μια DAPK εμφανίζει τη χαρακτηριστική στερεοδιάταξη, με τις 2 περιοχές της CaM να τυλίγονται γύρω από το πεπτίδιο και τοποθέτηση καταλοίπων αγκυροβόλησης 1-(10)-14 (με το κατάλοιπο στη θέση 1 να είναι η Trp305) – σημειώνεται πως και σε αυτήν την περίπτωση εντοπίζονται γειτονικά σε επίπεδο δομής αυτή τη φορά κατάλοιπα των καταλοίπων αγκυροβόλησης.

Αντιθέτως, σύμπλοκο της CaM με ολόκληρη την DAPK1, όπως στην δομή που μελετήθηκε, η N-τελική περιοχή της CaM έχει μεταφερθεί μέχρι και 20Å και περιστραφεί κατά 40°, με αποτέλεσμα η CaM να έχει μια πολύ πιο εκτεταμένη στερεοδιάταξη. Αξίζει σε αυτό το σημείο να αναφερθεί πως, για τη συγκεκριμένη δομή, το ποσοστό των αληθώς θετικών καταλοίπων ήταν σημαντικά υψηλό (19 στα 20 κατάλοιπα). Αυτό μπορεί να μας οδηγήσει στο συμπέρασμα πως τα «ψευδώς θετικά» κατάλοιπα που σταθερά προβλέφθηκαν σε όλες τις καλμοντουλίνες ως συνεξελισσόμενα με τη μέθοδο plmDCA, συμμετέχουν στις αλληλεπιδράσεις που αναπτύσσονται με κατάλοιπα της α-έλικας των κινασών, βοηθώντας τις δυο πρωτεΐνες να λάβουν τις τελικές μορφές τους στο χώρο, κατά τη διάρκεια δηλαδή

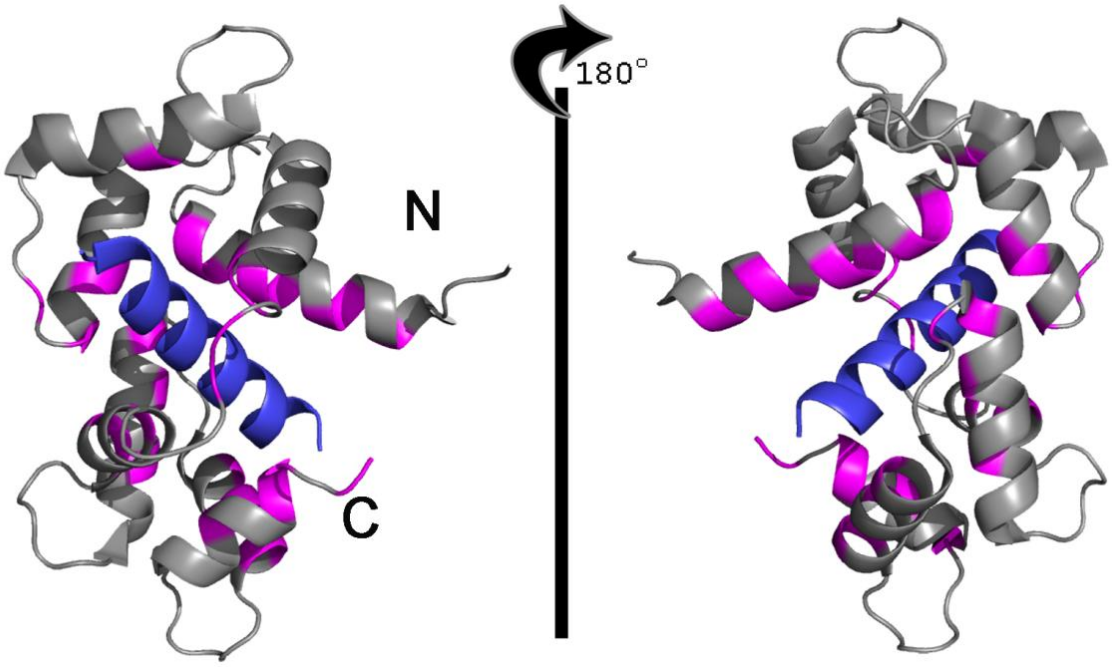
των στερεοδιατακτικών αλλαγών που συμβαίνουν με στόχο την αποδέσμευση της αυτοτελούς δομικής περιοχής και την επακόλουθη ενεργοποίηση της κινάσης.

Έκτος από τις αλληλεπιδράσεις που παρατηρούνται μεταξύ της καλμοντουλίνης και της έλικας της περιοχής αυτοαναστολής της DAPK1, και οι 2 περιοχές της CaM και κυρίως η C-τελική συμμετέχουν σε αλληλεπιδράσεις με την καταλυτική δομικά αυτοτελή περιοχή της κινάσης. Αυτό έχει ως αποτέλεσμα την περιορισμένη προσβασιμότητα του ενεργού κέντρου της κινάσης στο ρυθμιστικό σύμπλοκο DAPK1- $Ca^{2+}$-CaM2. Αυτό αντικατοπτρίζει την τεράστια σημασία της χρήσης ολόκληρων πρωτεϊνών, καθώς η κλειστή στερεοδιάταξη που λαμβάνει η CaM σε σύμπλοκο με μικρά πεπτίδια θα μπορούσε να προληφθεί λόγω στερεοχημικής παρεμπόδισης, όπως στο συγκεκριμένο παράδειγμα ή θα απαιτούσε μια περαιτέρω μετατόπιση της δομικά αυτοτελούς περιοχής υπεύθυνης για αυτορρύθμιση σε σχέση με το καταλυτικό κέντρο.

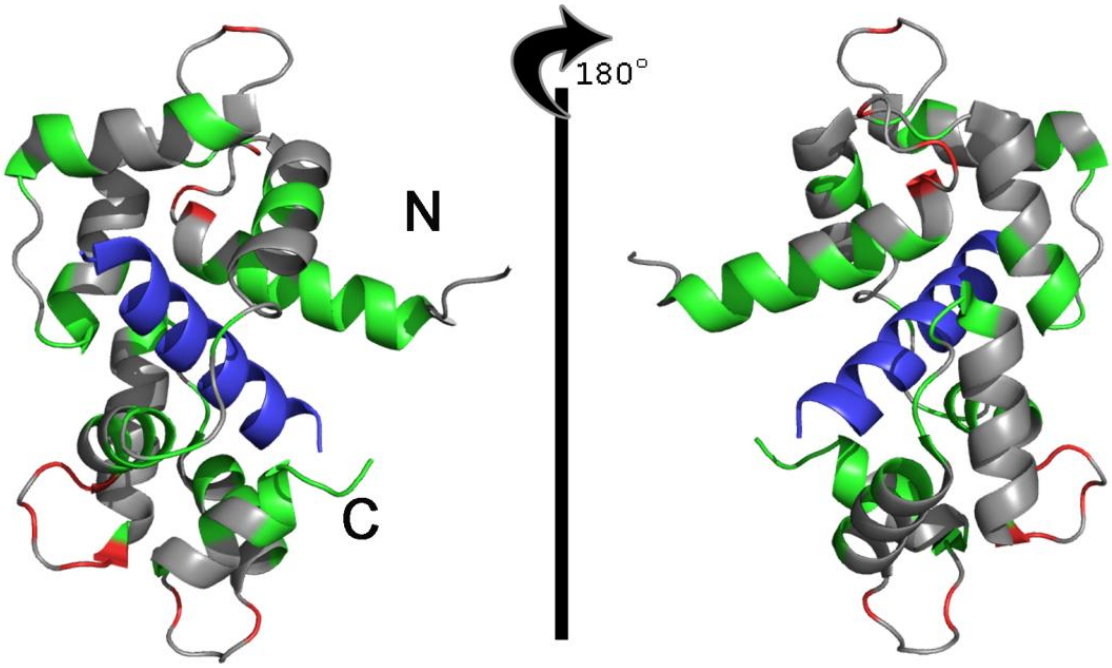
### 3.4.5. Μελέτη του συμπλόκου DAPK2 – $Ca^{2+}$-CaM και συμπεράσματα

Στην εικόνα 3.24 απεικονίζεται με αναπαράσταση cartoon το σύμπλοκο DAPK2 - $Ca^{2+}$-CaM2 (κωδικός PDB: 1WRZ). Με γκρι απεικονίζεται η καλμοντουλίνη $Ca^{2+}$-CaM2, ενώ με μπλε απεικονίζεται η κινάση DAPK2. Επισημαίνονται με magenta οι θέσεις των καταλοίπων που συμμετέχουν στην αλληλεπίδραση των δύο μορίων με βάση τις default παραμέτρους του προγράμματος RING, από όπου συγκεντρώθηκαν τα ζευγάρια των καταλοίπων που αλληλεπιδρούν καθώς και ο χαρακτηρισμός της φύσης της κάθε αλληλεπίδρασης.

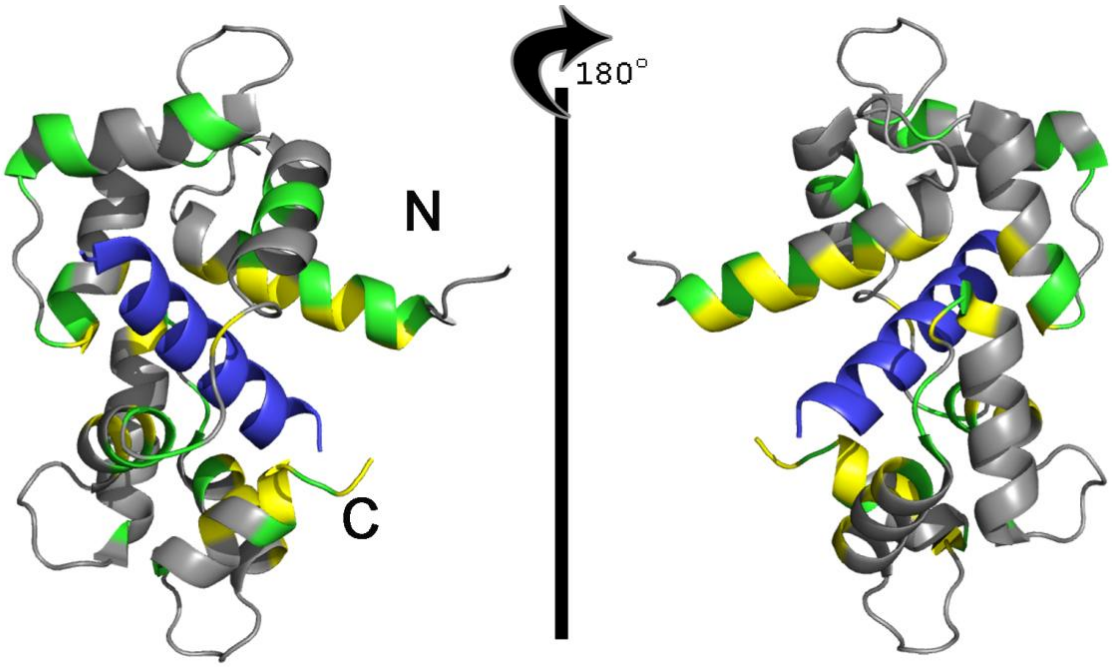
Αντίστοιχα με την εικόνα 3.24, στην εικόνα 3.25 απεικονίζεται το σύνολο των συνεξελισσόμενων καταλοίπων όπως προέκυψαν από τη μέθοδο plmDCA. Με πράσινο απεικονίζονται τα προβλεπόμενα αποδεκτά συνεξελισσόμενα κατάλοιπα, ενώ με κόκκινο τα κατάλοιπα που απορρίφθηκαν, καθώς εντοπίζονται στην περιοχή πρόσδεσης των ιόντων ασβεστίου. Τα αποδεκτά κατάλοιπα (πράσινο χρώμα) χρησιμοποιήθηκαν για την περαιτέρω αξιολόγηση των αποτελεσμάτων με μεθόδους αγκυροβόλησης. Τα κατάλοιπα που ανήκουν ταυτόχρονα σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα δίνονται στην εικόνα 3.26.




**Εικόνα 3.24** Πειραματικά επιβεβαιωμένα κατάλοιπα αλληλεπίδρασης μεταξύ DAPK2 - $Ca^{2+}$-CaM2 (κωδικός PDB: 1WRZ).




**Εικόνα 3.25** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο DAPK2 - $Ca^{2+}$-CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κόκκινο τα κατάλοιπα που απορρίφθηκαν.

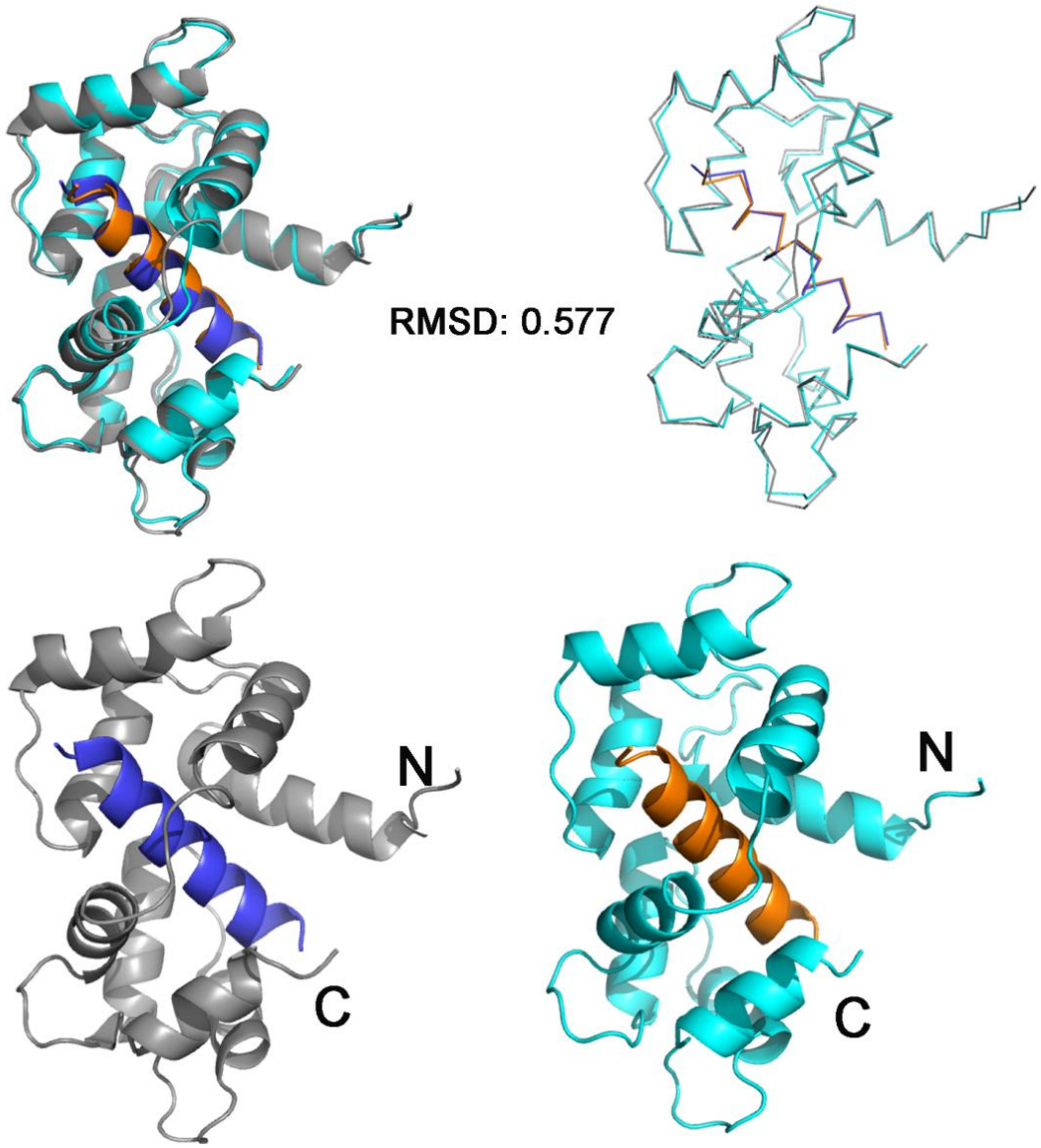



**Εικόνα 3.26** Πρόβλεψη συνεξελισσόμενων καταλοίπων με τη μέθοδο plmDCA για το σύμπλοκο DAPK2 - $Ca^{2+}$-CaM2. Με πράσινο τα αποδεκτά συνεξελισσόμενα κατάλοιπα και με κίτρινο τα κατάλοιπα που ανήκουν σε πειραματικά επιβεβαιωμένες περιοχές με συνεξελισσόμενα κατάλοιπα (αληθώς θετική πρόβλεψη).

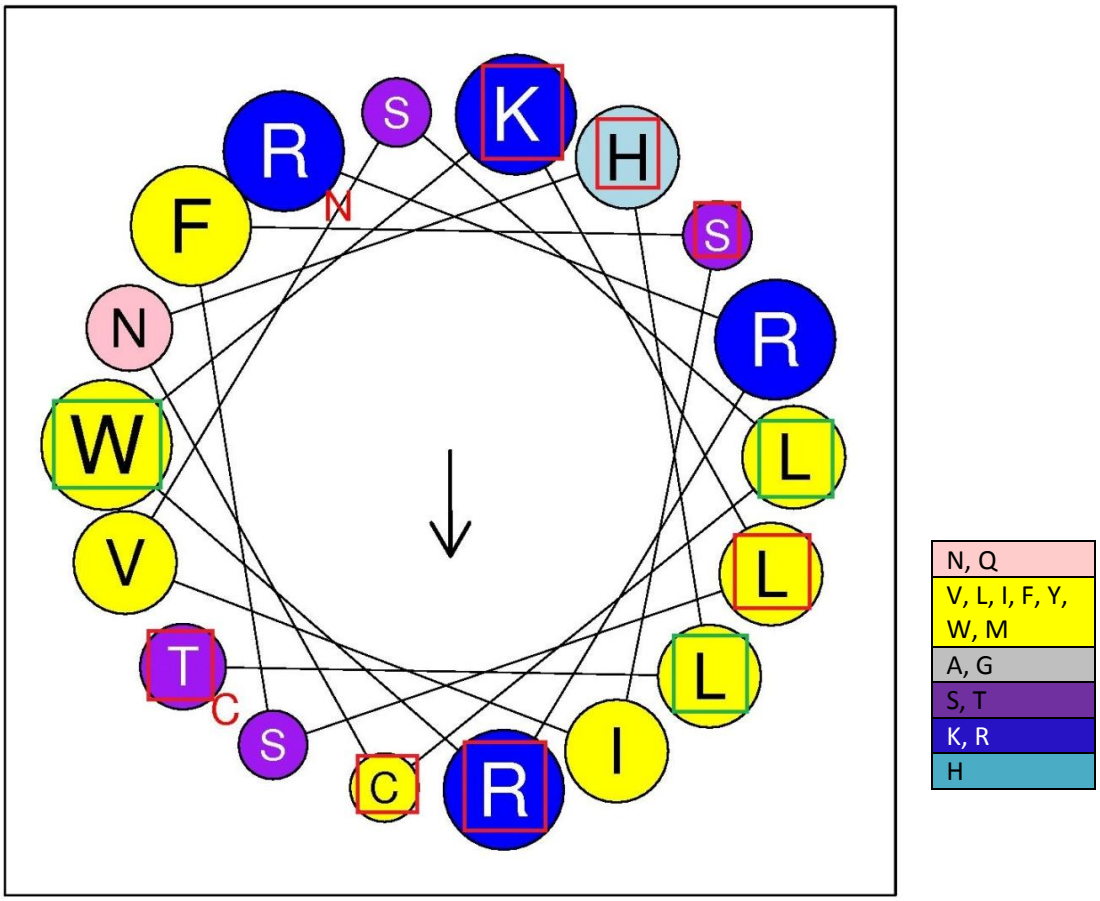
Η κατευθυνόμενη αγκυροβόληση βασισμένη στο αποτέλεσμα του προγράμματος icoms επιβεβαίωση την ορθή πρόβλεψη των συνεξελισσόμενων καταλοίπων, καθώς η καλύτερη δομή της πρώτης ομάδας αγκυροβόλησης με βάση τις συνολικές επαφές στην αύλακα αλληλεπίδρασης κινάσης-καλμοντουλίνης έχει RMSD μικρότερο του 1 (RMSD=0.577), όταν συγκρίνεται με την πειραματικά προσδιορισμένη δομή 1WRZ. Η εικόνα 3.27 απεικονίζει τη δομική στοίχιση (αναπαράσταση καρτούν - επάνω αριστερά, αναπαράσταση Cα ανθράκων – επάνω δεξιά) μεταξύ της δομής του πρώτου cluster και της πειραματικά προσδιορισμένης δομής 1WRZ. Αξίζει να σημειωθεί ότι στις ομάδες αγκυροβόλησης εντοπίστηκαν και άλλες ομάδες με πολύ καλύτερες ενέργειες αλληλεπίδρασης (βέλτιστη ενέργεια -66.70) όμως ο αριθμός των συνολικών επαφών δεν αφορούσαν την υπό μελέτη περιοχή ενδιαφέροντός μας μεταξύ κινάσης και καλμοντουλίνης (κριτήριο αποτέλεσαν οι συνολικές επαφές των καταλοίπων της κινάσης). Τα αποτελέσματα του για το σύμπλοκο DAPK2 - $Ca^{2+}$-CaM2 παρουσιάζονται περιληπτικά στον παρακάτω πίνακα:

**Πίνακας 3.8.** Αποτελέσματα SwarmDock για το σύμπλοκο DAPK2 - $Ca^{2+}$-CaM2.

| **Κριτήριο αξιολόγησης** | **Τιμή** |
|---|---|
| **Clusters με ιεραρχική ομαδοποίηση στα 3 Å** | 111 |
| **Σύνολο δομών 1$^{ου}$ cluster** | 26 |
| **Μέση ενέργεια δομών του 1$^{ου}$ cluster** | -58.114 |
| **Τυπική απόκλιση της μέσης ενέργειας** | 4.666 |
| **Ενέργεια καλύτερης δομής του 1$^{ου}$ cluster** | -62.39 |
| **Συνολικές αλληλεπιδράσεις** | 672 |
| **Αλληλεπιδράσεις καταλοίπων CaM** | 253 |
| **Αλληλεπιδράσεις καταλοίπων κινάσης** | 672 |




**Εικόνα 3.27** Δομική στοίχιση με τη ρουτίνα align του προγράμματος μοριακών γραφικών PYMOL μεταξύ της κρυσταλλογραφικά προσδιορισμένης δομής του συμπλόκου DAPK2 - $Ca^{2+}$-CaM2 (κινάση μπλε/καλμοντουλίνη γκρι) και της καλύτερης δομής του $1^{ου}$ cluster (κινάση πορτοκαλί/καλμοντουλίνη κυανό) έπειτα από την κατευθυνόμενη αγκυροβόληση.




**Εικόνα 3.28** Σε πράσινο τετράγωνο βρίσκονται τα κατάλοιπα που βιβλιογραφικά αναφέρονται ως κατάλοιπα αγκυροβόλησης, ενώ σε κόκκινο αυτά που έλαβαν τα πρώτα 20 υψηλότερα σκορ με βάση τη μέθοδο plmDCA.

Και σε αυτήν την περίπτωση εντοπίζονται για το πρώτο κατάλοιπο αγκυροβόλησης γειτονικά σε επίπεδο ακολουθίας, ενώ για τα κατάλοιπα αγκυροβόλησης του C-τελικού άκρου γειτονικά σε επίπεδο δομής.

## 3.4.6. Συγκριτική μελέτη των $Ca^{2+}$-CaM πρωτεϊνών και συμπεράσματα

Η συγκριτική και λειτουργική μελέτη των $Ca^{2+}$-CaM πραγματοποιήθηκε βασισμένη σε δυο κυρίως κριτήρια : 1) τις εμφανώς διαφορετικές στερεοδιατάξεις που αποκτούν όταν αλληλεπιδρούν με τις αμφιπαθείς α-έλικες διαφορετικών κινασών και 2) τα κατάλοιπα που προβλέφθηκαν ως ισχυρά συνεξελισσόμενα Τα αποτελέσματα που προέκυψαν δίνουν πιθανή ερμηνεία για το ρόλο των συνεξελισσόμενων αμινοξικών καταλοίπων.

**1) Διαφορετικές στερεοδιατάξεις των $Ca^{2+}$-CaM.** Η δομική στοίχιση των $Ca^{2+}$-CaM κατά ζεύγη με την αυτοματοποιημένη ρουτίνα «Align» του προγράμματος μοριακών γραφικών PyMol, απέδειξε την ποικιλία των στερεοδιατάξεων που αποκτά το μόριο της καλμοντουλίνης, όταν αλληλεπιδρά με διαφορετικές α-έλικες διαφορετικών κινασών. Όπως αποδεικνύεται από τη συγκριτική ανάλυση των RMSD (πίνακας 3.9 -3.13) η καλμοντουλίνη αποκτά την πλέον «ανοιχτή» ή εκτεταμένη στερεοδιάταξη της σε σύμπλοκο με την DAPK1 (PDB: 2X0G) και την πλέον «κλειστή» σε σύμπλοκο με την CaMK IIδ (PDB: 2WEL), όπως έχει ήδη αναλυθεί εκτενώς στα προηγούμενα κεφάλαια.

**Πίνακας 3.9.** Πίνακας RMSD της $Ca^{2+}$-CaM του συμπλόκου CaMK Iα - $Ca^{2+}$-CaM (1mxe) σε σχέση με τις $Ca^{2+}$-CaM των υπόλοιπων συμπλόκων.

| RMSD | 1wrz | 1mxe | 2x0g | 5jqa | 2wel |
|---|---|---|---|---|---|
| **1mxe** | **1.285** | **-** | **3.536** | **1.936** | **1.307** |

**Πίνακας 3.10.** Πίνακας RMSD της $Ca^{2+}$-CaM2 του συμπλόκου CaMK IIδ - $Ca^{2+}$-CaM2 (2wel) σε σχέση με τις $Ca^{2+}$-CaM των υπόλοιπων συμπλόκων.

| RMSD | 1wrz | 1mxe | 2x0g | 5jqa | 2wel |
|---|---|---|---|---|---|
| **2wel** | **1.455** | **1.307** | **>10** | **1.835** | **-** |

**Πίνακας 3.11.** Πίνακας RMSD της $Ca^{2+}$-CaM1 του συμπλόκου MLCK - $Ca^{2+}$-CaM1 (5jqa) σε σχέση με τις $Ca^{2+}$-CaM των υπόλοιπων συμπλόκων.

| RMSD | 1wrz | 1mxe | 2x0g | 5jqa | 2wel |
|---|---|---|---|---|---|
| **5jqa** | **1.596** | **1.936** | **2.529** | **-** | **1.835** |

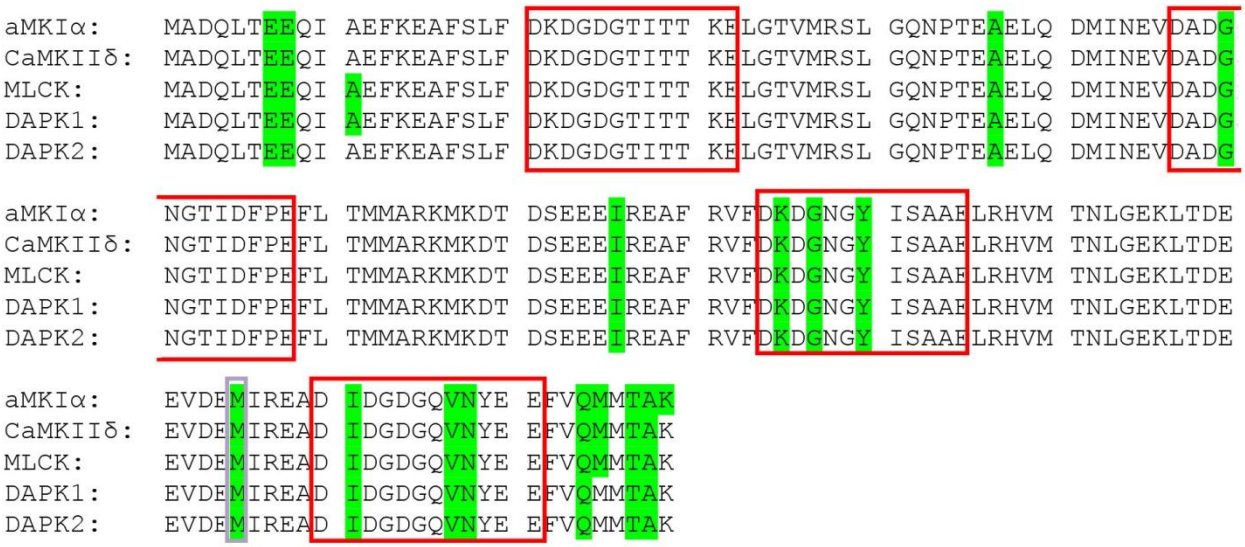
**Πίνακας 3.12.** Πίνακας RMSD της $Ca^{2+}$-CaM2 του συμπλόκου DAPK1 - $Ca^{2+}$-CaM2 (2x0g) σε σχέση με τις $Ca^{2+}$-CaM των υπόλοιπων συμπλόκων.

| RMSD | 1wrz | 1mxe | 2x0g | 5jqa | 2wel |
|---|---|---|---|---|---|
| **2x0g** | **3.669** | **3.536** | **-** | **2.529** | **>10** |

**Πίνακας 3.13.** Πίνακας RMSD της $Ca^{2+}$-CaM2 του συμπλόκου DAPK2 - $Ca^{2+}$-CaM2 (1wrz) σε σχέση με τις $Ca^{2+}$-CaM των υπόλοιπων συμπλόκων.

| RMSD | 1wrz | 1mxe | 2x0g | 5jqa | 2wel |
|---|---|---|---|---|---|
| **1wrz** | **-** | **1.285** | **3.669** | **1.596** | **1.455** |

**2) Προβλεπόμενα συνεξελισσόμενα αμινοξικά κατάλοιπα.** Στην εικόνα 3.29 παρατίθενται οι ακολουθίες των καλμοντουλινών που μελετήθηκαν σε σύμπλοκο με τις αντίστοιχες κινάσες. Με πράσινο χρώμα επισημαίνονται τα κατάλοιπα που έλαβαν τα 20 μεγαλύτερα σκορ έπειτα από την υπολογιστική ανάλυση (Iserte et al., 2015).




**Εικόνα 3.29** Πολλαπλή στοίχιση ακολουθιών καλμοντουλινών.

Σε μοβ πλαίσιο απεικονίζεται το κατάλοιπο μεθειονίνης που αποτελεί το κατάλοιπο της C-τελικής αυτοτελούς δομικής περιοχής της καλμοντουλίνης που αλληλεπιδρά με τα κατάλοιπα αγκυροβόλησης του N-τελικού άκρου των αυτοτελών δομικών περιοχών πρόσδεσης της CaM των κινασών. Το γεγονός ότι προβλέφθηκε ως ισχυρά συνεξελισσόμενο κατάλοιπο, και στα 5 σύμπλοκα, έχει ιδιαίτερη σημασία. Αποτέλεσε αρχικά επιβεβαίωση ενός ήδη γνωστού μηχανισμού για κάποιες οικογένειες εξαρτώμενων από καλμοντουλίνη κινασών. Επίσης, αποτέλεσε επιβεβαίωση της μεθόδου plmDCA ως μεθόδου πρόβλεψης αληθώς αλληλεπιδρώντων καταλοίπων ή συμμετεχόντων στην αλληλεπίδραση καταλοίπων, ενώ τέλος αποτέλεσε ένδειξη για το μηχανισμό αλληλεπίδρασης των υπόλοιπων οικογενειών, όπου η α-έλικα προσδένεται στην CaM με αντιπαράλληλο τρόπο.

Σε κόκκινο πλαίσιο απεικονίζονται οι θηλιές του κάθε μοτίβου EF-hand στις οποίες εντοπίζονται τα 4 προσδεμένα ιόντα ασβεστίου. Όπως αναφέρθηκε και εισαγωγικά, η δέσμευση των ιόντων ασβεστίου είναι μια συνεργατική διαδικασία. Αρχικά προσδένονται στα EF-hands που εντοπίζονται στο C-τελικό άκρο και λόγω αυτού αυξάνεται η συγγένεια πρόσδεσής τους για τα μοτίβα του N-τελικού άκρου. Η συνολική συγγένεια πρόσδεσης αυξάνεται περαιτέρω με την πρόσδεση της αντίστοιχης αυτοτελούς δομικής περιοχής και τις επακόλουθες στερεοδιατακτικές αλλαγές της καλμοντουλίνης.

Σε πολλές περιπτώσεις η πρόσδεση των ιόντων ασβεστίου στην Ν-τελική περιοχή συμβαίνει μετά την πρόσδεση της α-έλικας. Παρατηρούμε πως εντοπίζονται ισχυρά συνεξελισσόμενα κατάλοιπα κυρίως στις θηλιές των EF-hand μοτίβων της καρβοξυτελικής αυτοτελούς δομικής περιοχής. Επίσης ως συνεξελισσόμενα δεν προβλέφθηκαν τα κατάλοιπα που συμμετέχουν ενεργά στην πρόσδεση του ασβεστίου (κατάλοιπα ασπαρτικού και γλουταμικού οξέος), αλλά τα γειτονικά αυτών υδρόφοβα κατάλοιπα G, Y, I και V και υδρόφιλα κατάλοιπα K και N. Έτσι μπορούμε να υποθέσουμε τη συμμετοχή των καταλοίπων αυτών στο συνεργατικό φαινόμενο πρόσδεσης των ιόντων ασβεστίου. Επιπλέον, στη δομή του συμπλόκου της DAPK1 με την καλμοντουλίνη τα κατάλοιπα της πρώτης και της τελευταίας θηλιάς των EF-hand μοτίβων εντοπίζονται να αλληλεπιδρούν με κατάλοιπα θηλιών (όπως η βασική θηλιά) του καταλυτικού κέντρου της κινάσης.

Με αυτό ως δεδομένο μπορούμε να εξάγουμε το συμπέρασμα πως πιθανά συμμετέχουν στις αλληλεπιδράσεις που αναπτύσσονται μεταξύ της καλμοντουλίνης και της κινάσης πριν η καλμοντουλίνη πάρει την τελική της κλειστή στερεοδιάταξη με την α-έλικα να εντοπίζεται στο εσωτερικό της αύλακάς της και η κινάση την τελική ενεργή της στερεοδιάταξη με το ενεργό της κέντρο απελευθερωμένο και ικανό για κατάλυση της αντίδρασης φωσφορυλίωσης του εκάστοτε υποστρώματος.

## 3.5. Γενικά συμπεράσματα

Η υπολογιστική ανάλυση των αλληλεπιδράσεων των εξαρτώμενων από την καλμοντουλίνη κινασών και της καλμοντουλίνης με μεθόδους συνεξέλιξης αποτέλεσε ένα σημαντικό βήμα για το εμπλουτισμό των γνωστών μέχρι σήμερα δεδομένων που αφορούν την μεταξύ τους αλληλεπίδραση. Τα συμπεράσματα της υπολογιστής μελέτης συνοψίζονται ακολούθως.

Αρχικά, όσον αφορά το σύνολο δεδομένων, δηλαδή το σύνολο των πρωτεϊνικών μορίων που αποτέλεσαν και το πεδίο μελέτης ο αριθμός των αμινοξικών ακολουθιών των ευκαρυωτικών κινασών που βρίσκονται κατατεθειμένες και καλά σχολιασμένες στη βάση δεδομένων SwissProt είναι σχετικά μικρός. Η απουσία πλήθους εγγραφών από τη βάση αυτή είναι αξιοσημείωτη και έρχεται σε αντιδιαστολή με το γεγονός ότι οι κινάσες αποτελούν τη δεύτερη μεγαλύτερη οικογένεια πρωτεϊνών. Ο μικρός αριθμός ακολουθιών σε συνδυασμό με την υψηλή συντήρηση του καταλυτικού τους κέντρου καθιστούν τη μελέτη των αλληλεπιδράσεων των μορίων αυτών με τη βοήθεια μεθόδων συνεξέλιξης δύσκολη, λόγω της χαμηλής πολυπλοκότητας που δίνεται ως αρχικό αρχείο εισόδου στους αλγορίθμους συνεξέλιξης.

Ωστόσο, η μέθοδος plmDCA έδωσε ικανοποιητικά αποτελέσματα, όσον αφορά τις κινάσες, παρά την απουσία πληθώρας δεδομένων στο αρχικό υπό μελέτη σύνολο. Η ύπαρξη πειραματικών δεδομένων που αφορούν τις περιοχές αλληλεπίδρασης ήταν αναγκαία για να γίνει η διάκριση μεταξύ των αλληλεπιδρώντων καταλοίπων και των καταλοίπων που προβλέφθηκαν ως αλληλεπιδρώντα λόγω της υψηλής συντήρησης. Κομβικής σημασίας είναι η πρόβλεψη καταλοίπων που γειτνιάζουν με τα σημαντικά για την αλληλεπίδραση κατάλοιπα, καθώς τυχόν αλλαγές σε αυτά (θετικής ή αρνητικής επίδρασης για την εν λόγω αλληλεπίδραση) μπορεί να επάγουν εξελικτική πίεση σε βαθμό που να επηρεάσει την αλληλεπίδραση. Ως προς τις υπολογιστικές μεθόδους συνεξέλιξης αξίζει να αναφερθεί ότι, σημαντική πληροφορία είναι πιθανό να χάθηκε καθώς σε αρκετές περιπτώσεις η αλληλεπίδραση μεταξύ των μορίων είναι παροδική και οι μέθοδοι συνεξέλιξης αναφέρεται πως αδυνατούν να δώσουν το ίδιο ικανοποιητικά αποτελέσματα με περιπτώσεις μελέτης μόνιμων συμπλόκων ή μελέτης αλληλεπιδρώντων κατάλοιπων στο εσωτερικό μιας πρωτεΐνης.

Τα συμπεράσματα προκύπτουν και από τη συγκριτική μελέτη του συνόλου των ακολουθιών των CaM και των εξαρτώμενων από CaM κινασών, σε συνδυασμό με τη μελέτη των διαθέσιμων δομών και αναλύθηκαν στην προηγούμενη ενότητα, μας οδηγούν στο συμπέρασμα πως οι μέθοδοι συνεξέλιξης προβλέπουν επιτυχώς σημαντικά κατάλοιπα των αλληλεπιδράσεων μεταξύ πρωτεϊνών. Ελλείψει

κρυσταλλογραφικά προσδιορισμένων συμπλόκων στις οποίες να εντοπίζεται, εκτός της αυτοτελούς δομικής περιοχής πρόσδεσης της καλμοντουλίνης, το καταλυτικό κέντρο της κινάσης μπορούμε να βασιστούμε στις μεθόδους συνεξέλιξης για τον εντοπισμό σημαντικών για την αλληλεπίδραση κατάλοιπων αλλά και την διελεύκανση του μηχανισμού αλληλεπίδρασης. Όλα αυτά μπορούν να αποτελέσουν τη βάση για μελλοντικά πειράματα που αφορούν την αλληλεπίδραση κινασών-καλμοντουλινών.

Όπως έχει ήδη αναφερθεί η **έννοια της αλληλεπίδρασης είναι στενά συνδεδεμένη με αυτή της συνεξέλιξης** και ο λόγος για τον οποίο είναι χρήσιμη η μελέτη της συνεξέλιξης και σε μοριακό επίπεδο. Έτσι, το πρωτόκολλο της υπολογιστικής ανάλυσης που παρουσιάζεται στην παρούσα διπλωματική εργασία θα μπορούσε να χρησιμοποιηθεί κατά τη μελέτη των αλληλεπιδράσεων σημαντικών ομάδων συμπλόκων μορίων καθώς συμβάλει καθοριστικά στο λειτουργικό σχολιασμό αλληλεπιδράσεων προσθέτοντας την άγνωστη μέχρι σήμερα πληροφορία της συνεξέλιξης.

Συνολικά, τα δεδομένα συνεξέλιξης που αφορούν την καλμοντουλίνη και τις εξαρτώμενες από την καλμοντουλίνη κινάσες εμπλουτίζουν τις προσπάθειες μελέτης του μηχανισμού και της φύσης των μεταξύ τους αλληλεπιδράσεων, ένα ερευνητικό πεδίο που αποτελεί στόχο την ανάπτυξη και το σχεδιασμό φαρμάκων για πληθώρα παθολογικών καταστάσεων, μεταξύ άλλων παθήσεις του νευρικού συστήματος ή διαφορετικές μορφές καρκίνου.

# 4.Βιβλιογραφία


Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Michael Cherry, J., … Genet. (2000). Gene Ontology: tool for the unification of biology. *Nat Genet*, *25*(1), 25–29. https://doi.org/10.1038/75556

Ashenberg, O., & Laub, M. T. (2013). Using analyses of amino acid coevolution to understand protein structure and function. *Methods in Enzymology*, *523*, 191–212. https://doi.org/10.1016/B978-0-12-394292-0.00009-6

Berchtold, M. W., Egli, R., Rhyner, J. A., Hameister, H., & Strehler, E. E. (1993). Localization of the Human Bona Fide Calmodulin Genes CALM1, CALM2, and CALM3 to Chromosomes 14q24-q31, 2p21.1-p21.3, and 19q13.2-q13.3. *Genomics*, *16*(2), 461–465. https://doi.org/10.1006/geno.1993.1211

Bialik, S., & Kimchi, A. (2014). The DAP-kinase interactome. *Apoptosis*, *19*(2), 316–328. https://doi.org/10.1007/s10495-013-0926-3

Biology, C., Suplatov, D., Sharapova, Y., Timonina, D., & Kopylov, K. (2018). The visualCMAT: A web-server to select and interpret correlated mutations/co-evolving residues in protein families, *16*(2), 1–15. https://doi.org/10.1142/S021972001840005X

Biswas, A., Noman, N., & Sikder, A. (2010). Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information. *BMC Bioinformatics*, *11*(1), 273. https://doi.org/10.1186/1471-2105-11-273

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*, *294*(5), 1351–1362. https://doi.org/10.1006/jmbi.1999.3310

Blom, N., Sicheritz-Pontén, 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

Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K., Lee, I., & Marcotte, E. M. (2004). Protein interaction networks from yeast to human. *Current Opinion in Structural Biology*, *14*(3), 292–299. https://doi.org/10.1016/j.sbi.2004.05.003

Bowers, P. M., Pellegrini, M., Thompson, M. J., Fierro, J., Yeates, T. O., & Eisenberg, D. (2004). Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biology*, *5*(5), R35. https://doi.org/10.1186/gb-2004-5-5-r35

Brunger, A. T. (2007). Version 1.2 of the Crystallography and NMR system. *Nature Protocols*, *2*(11), 2728–2733. https://doi.org/10.1038/nprot.2007.406

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., … Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica. Section D, Biological Crystallography*, *54*(Pt 5), 905–921. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9757107

Cannistraci, C. V., Alanis-Lobato, G., & Ravasi, T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding, 29 Bioinformatics § (2013). https://doi.org/10.1093/bioinformatics/btt208

Chen, J., Chua, H. N., Hsu, W., Lee, M.-L., Ng, S.-K., Saito, R., … Wong, L. (2006). Increasing confidence of protein-protein interactomes., *17*(2), 284–297. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17514831

Chen, J., Hsu, W., Lee, M. L. L., & Ng, S.-K. S. K. (2006). Increasing confidence of protein interactomes using network topological metrics. *Bioinformatics*, *22*(16), 1998–2004. https://doi.org/10.1093/bioinformatics/btl335

Conrads, T. P., Zhou, M., Petricoin, E. F., Liotta, L., & Veenstra, T. D. (2003). Cancer diagnosis using proteomic patterns. *Expert Review of Molecular Diagnostics*, *3*, 411–420. https://doi.org/10.1586/14737159.3.4.411

Coultrap, S. J., & Bayer, K. U. (2012). CaMKII regulation in information processing and storage. *Trends in Neurosciences*, *35*(10), 607–618. https://doi.org/10.1016/j.tins.2012.05.003

Craig, R. A., & Liao, L. (2007). Phylogenetic tree information aids supervised learning for predicting protein-protein interaction based on distance matrices. *BMC Bioinformatics*, *8*(1), 6. https://doi.org/10.1186/1471-2105-8-6

Dandekar, T., Snel, B., Huynen, M., & Bork, P. Conservation of gene order: A fingerprint of proteins that physically interact, 23 Trends in Biochemical Sciences § (1998). https://doi.org/10.1016/S0968-0004(98)01274-2

Darwin, C. (1862). On the contrivances by which British and foreign orchids are fertilized by insects, and on the good effects of intercrossing. *John Murray*, 365. https://doi.org/10.1017/CBO9780511910197

de Juan, D., Pazos, F., & Valencia, A. (2013). Emerging methods in protein co-evolution. *Nature Reviews Genetics*, *14*(4), 249–261. https://doi.org/10.1038/nrg3414

de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., … Bonvin, A. M. J. J. (2007). HADDOCK versus HADDOCK: New features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins: Structure, Function, and Bioinformatics*, *69*(4), 726–733. https://doi.org/10.1002/prot.21723

DeLano, W. L. (2002). *PyMOL: An open-source molecular graphics tool*. *CCP4 Newsletter On Protein Crystallography* (Vol. 40). Retrieved from pymol.org

Dhanasekaran, N., & Reddy, E. P. (1998). Signaling by dual specificity kinases. *Oncogene*, *17*(11), 1447–1455. https://doi.org/10.1038/sj.onc.1202251

DOBZHANSKY, T. (1950). Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. *Genetics*, *35*(3), 288–302. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15414931

Dominguez, C., Boelens, R., & Bonvin, A. M. J. J. (2003). HADDOCK: A Protein−Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society*, *125*(7), 1731–1737. https://doi.org/10.1021/ja026939x

Dunn, S. D. D., Wahl, L. M. M., & Gloor, G. B. B. (2008). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics*, *24*(3), 333–340. https://doi.org/10.1093/bioinformatics/btm604

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants: A Study in Coevolution, (4). https://doi.org/10.2307/2406212

Ekeberg, M., Hartonen, T., & Aurell, E. (2014). Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences. *Journal of Computational Physics*, *276*, 341–356. https://doi.org/10.1016/J.JCP.2014.07.024

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., & Aurell, E. (2012). Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. https://doi.org/10.1103/PhysRevE.87.012707

Enright, A. J., Iliopoulos, I., Kyrpides, N. C., & Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*, *402*(6757), 86–90. https://doi.org/10.1038/47056

Esmaielbeiki, R., Krawczyk, K., Knapp, B., Nebel, J. C., & Deane, C. M. (2016). Progress and challenges in predicting protein interfaces. *Briefings in Bioinformatics*, *17*(1), 117–131. https://doi.org/10.1093/bib/bbv027

Farag, A. K., & Roh, E. J. (2018). Death-associated protein kinase (DAPK) family modulators: Current and future therapeutic outcomes. *Medicinal Research Reviews*, (March), 1–37.

https://doi.org/10.1002/med.21518
Feinauer, C., Skwark, M. J., Pagnani, A., & Aurell, E. (2014). Improving Contact Prediction along Three Dimensions. *PLoS Computational Biology*, *10*(10), e1003847. https://doi.org/10.1371/journal.pcbi.1003847
Freiberg, C. (2001). Novel computational methods in anti-microbial target identification. *Drug Discovery Today*, *6*(01), 72–80. https://doi.org/10.1016/S1359-6446(01)00167-2
Galperin, M. Y., & Koonin, E. V. (2000). Who's your neighbor? New computational approaches for functional genomics. *Nature Biotechnology*, *18*(6), 609–613. https://doi.org/10.1038/76443
Gautier, R., Douguet, D., Antonny, B., & Drin, G. (2008). HELIQUEST: a web server to screen sequences with specific α-helical properties. *BIOINFORMATICS APPLICATIONS NOTE*, *24*(18), 2101–2102. https://doi.org/10.1093/bioinformatics/btn392
Gnad, F., Ren, S., Cox, J., Olsen, J. V, Macek, B., Oroshi, M., & Mann, M. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology*, *8*(11), R250. https://doi.org/10.1186/gb-2007-8-11-r250
Göbel, U., Sander, C., Schneider, R., & Valencia, A. (1994). Correlated mutations and residue contacts in proteins. *Proteins: Structure, Function, and Genetics*, *18*(4), 309–317. https://doi.org/10.1002/prot.340180402
Goldberg, J., Nairn, A. C., & Kuriyan, J. (1996). Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I. *Cell*, *84*(6), 875–887. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8601311
Grigoriev, A. (2001). A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. *Nucleic Acids Research*, *29*(17). https://doi.org/10.1093/nar/29.17.3513
Guimarães, K. S., Jothi, R., Zotenko, E., & Przytycka, T. M. (2006). Predicting domain-domain interactions using a parsimony approach. *Genome Biology*, *7*(11), R104. https://doi.org/10.1186/gb-2006-7-11-r104
Herce, H. D., Deng, W., Helma, J., Leonhardt, H., & Cardoso, M. C. (2013). Visualization and targeted disruption of protein interactions in living cells. *Nature Communications*, *4*(1), 2660. https://doi.org/10.1038/ncomms3660
Hobohm, U., Scharf, M., Schneider, R., & Sander, C. (1992). Selection of representative protein data sets. *Protein Science : A Publication of the Protein Society*, *1*(3), 409–417. https://doi.org/10.1002/pro.5560010313
Hong, F., Haldeman, B. D., Jackson, D., Carter, M., & Baker, J. E. (2012). Biochemistry of Smooth Muscle Myosin Light Chain Kinase. *Arch Biochem Biophys*, *510*(2), 135–146. https://doi.org/10.1016/j.abb.2011.04.018.Biochemistry
Hopf, T. a., Morinaga, S., Ihara, S., Touhara, K., Marks, D. S., & Benton, R. (2015). Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. *Nature Communications*, *6*, 6077. https://doi.org/10.1038/ncomms7077
Hopf, T. A., Schärfe, C. P. I., Rodrigues, J. P. G. L. M., Green, A. G., Kohlbacher, O., Sander, C., … Marks, D. S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *ELife*, *3*. https://doi.org/10.7554/eLife.03430
Hosur, R., Xu, J., Bienkowska, J., & Berger, B. (2011). IWRAP: An interface threading approach with application to prediction of cancer-related protein-protein interactions. *Journal of Molecular Biology*, *405*(5), 1295–1310. https://doi.org/10.1016/j.jmb.2010.11.025
Huang, H.-D., Lee, T.-Y., Tzeng, S.-W., & Horng, J.-T. (2005). KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Research*, *33*(Web Server), W226–W229. https://doi.org/10.1093/nar/gki471

Huse, M., & Kuriyan, J. (2002). The Conformational Plasticity of Protein Kinases. *Cell*, *109*(3), 275–282. https://doi.org/10.1016/S0092-8674(02)00741-9

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, *32*(3), 1037–1049. https://doi.org/10.1093/nar/gkh253

Iserte, J., Simonetti, F. L., Zea, D. J., Teppa, E., & Marino-buslje, C. (2015). I-COMS : Interprotein-COrrelated Mutations Server, *43*(June), 320–325. https://doi.org/10.1093/nar/gkv572

Ito, T., Chiba, T., & Yoshida, M. (2001). Exploring the protein interactome using comprehensive two-hybrid projects. *Trends in Biotechnology*, *19*(10 Suppl). https://doi.org/10.1016/S0167-7799(01)01790-5

J Tavares, C. D., Ferguson, S. B., Giles, D. H., Wang, Q., Wellmann, R. M., O, J. P., … Dalby, K. N. (2014). The Molecular Mechanism of Eukaryotic Elongation Factor 2 Kinase Activation * From the ‡ Graduate Program in Cell and Molecular Biology, the. https://doi.org/10.1074/jbc.M114.577148

James T. Stull, Kristine E. Kamm, and R. V. (2012). Myosin Light Chain Kinase and the Role of Myosin Light Chain Phosphorylation in Skeletal Muscle. *National Institue of Health*, *510*(2), 120–128. https://doi.org/10.1016/j.abb.2011.01.017.Myosin

Johnson, L. N., & Lewis, R. J. (2001). Structural Basis for Control by Phosphorylation. *Chemical Reviews*, *101*(8), 2209–2242. https://doi.org/10.1021/cr000225s

Jones, D. T., Buchan, D. W. A., Cozzetto, D., & Pontil, M. (2012). PSICOV: precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments. *Bioinformatics*, *28*(2), 184–190. https://doi.org/10.1093/bioinformatics/btr638

Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, *110*(6), 1657–1666. https://doi.org/10.1021/ja00214a001

Jung, I., Matsuyama, A., Yoshida, M., & Kim, D. (2010). PostMod: sequence based prediction of kinase-specific phosphorylation sites with indirect relationship. *BMC Bioinformatics*, *11*(Suppl 1), S10. https://doi.org/10.1186/1471-2105-11-S1-S10

Kathrina L. Marcelo, A. R. M. and B. Y. (2017). The Ca2+ /Calmodulin/CaMKK2 Axis: Nature's Metabolic CaMshaft, *27*(10), 706–718. https://doi.org/10.1016/j.tem.2016.06.001.The

Kennelly, P. J., & Krebs, E. G. (1991). Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. *The Journal of Biological Chemistry*, *266*(24), 15555–15558. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1651913

Kennelly, P. J., & Potts, M. (1996). Fancy meeting you here! A fresh look at "prokaryotic" protein phosphorylation. *Journal of Bacteriology*, *178*(16), 4759–4764. https://doi.org/10.1128/JB.178.16.4759-4764.1996

Keskin, O., Tuncbag, N., & Gursoy, A. (2016). Predicting Protein–Protein Interactions from the Molecular to the Proteome Level. *Chemical Reviews*, *116*(8), 4884–4909. https://doi.org/10.1021/acs.chemrev.5b00683

Khapchaev, A. Y., & Shirinsky, V. P. (2016). Myosin Light Chain Kinase MYLK1: Anatomy, Interactions, Functions, and Regulation. *Biochemistry. Biokhimiia*, *81*(13), 1676–1697. https://doi.org/10.1134/S000629791613006X

Klimecka, M., & Muszyńska, G. (2007). Structure and functions of plant calcium-dependent protein kinases. *Acta Biochimica Polonica*, *54*(2), 219–233.

Korber, B. T., Farber, R. M., Wolpert, D. H., & Lapedes, a S. (1993). Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proceedings of the National Academy of Sciences of the*

*United States of America*, *90*(15), 7176–7180.
Kylarova, S., Psenakova, K., Herman, P., & Obsil, T. (2018). CaMKK2 kinase domain interacts with the autoinhibitory region through the N-terminal lobe including the RP insert. *BBA - General Subjects*, #pagerange#. https://doi.org/10.1016/j.bbagen.2018.07.025
Lee, K., Alphonse, S., Piserchio, A., Tavares, C. D. J., Giles, D. H., Wellmann, R. M., … Ghose, R. (2016). Structural Basis for the Recognition of Eukaryotic Elongation Factor 2 Kinase by Calmodulin. *Structure*, *24*(9), 1441–1451. https://doi.org/10.1016/j.str.2016.06.015
Lee, S.-A., Chan, C., Tsai, C.-H., Lai, J.-M., Wang, F.-S., Kao, C.-Y., & Huang, C.-Y. F. (2008). Ortholog-based protein-protein interaction prediction and its application to inter-species interactions. *BMC Bioinformatics*, *9 Suppl 12*(Suppl 12), S11. https://doi.org/10.1186/1471-2105-9-S12-S11
Lehner, B., & Fraser, A. G. (2004). A first-draft human protein-interaction map., *5*(9). https://doi.org/10.1186/gb-2004-5-9-r63
Letovsky, S., & Kasif, S. (2003). Predicting protein function from protein/protein interaction data: A probabilistic approach. *Bioinformatics*, *19*(SUPPL. 1). https://doi.org/10.1093/bioinformatics/btg1026
Li, L., Shakhnovich, E. I., & Mirny, L. a. (2003). Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(8), 4463–4468. https://doi.org/10.1073/pnas.0737647100
Li, L., Zhang, Y., Zou, L., Li, C., Yu, B., Zheng, X., & Zhou, Y. (2012). An Ensemble Classifier for Eukaryotic Protein Subcellular Location Prediction Using Gene Ontology Categories and Amino Acid Hydrophobicity. *PLoS ONE*, *7*(1), e31057. https://doi.org/10.1371/journal.pone.0031057
Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158
Lin, T.-W. W., Wu, J.-W. W., & Chang, D. T.-H. H. (2013). Combining Phylogenetic Profiling-Based and Machine Learning-Based Techniques to Predict Functional Related Proteins. *PLoS ONE*, *8*(9), e75940. https://doi.org/10.1371/journal.pone.0075940
Lisman, J., Yasuda, R., & Raghavachari, S. (2012). Mechanisms of CaMKII action in long-term potentiation. *Nature Reviews. Neuroscience*, *13*(3), 169–182. https://doi.org/10.1038/nrn3192
Lockless, S. W., & Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, *286*(5438), 295–299. https://doi.org/10.1126/science.286.5438.295
MacDonald, J. A., Mackey, A. J., Pearson, W. R., & Haystead, T. A. J. (2002). A strategy for the rapid identification of phosphorylation sites in the phosphoproteome. *Molecular & Cellular Proteomics : MCP*, *1*(4), 314–322. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12096113
Maiorov, V. N., & Crippen, G. M. (1994). Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, *235*(2), 625–634. https://doi.org/10.1006/jmbi.1994.1017
Manning, G., Whyte, D. B., Martinez, R., & Hunter, T. (2002). The Protein Kinase Complement of the Human Genome, *298*(December).
Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science (New York, N.Y.)*, *285*(5428). https://doi.org/10.1126/science.285.5428.751
Marina, A., Waldburger, C. D., & Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *The EMBO Journal*, *24*(24), 4247–4259. https://doi.org/10.1038/sj.emboj.7600886

Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., & Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE*, *6*(12), e28766. https://doi.org/10.1371/journal.pone.0028766

Marks, D. S., Hopf, T. A., & Sander, C. (2012). Protein structure prediction from sequence variation. *Nature Publishing Group*, *30*(11), 1072–1080. https://doi.org/10.1038/nbt.2419

Martin, A. J. M., Vidotto, M., Boscariol, F., Di Domenico, T., Walsh, I., & Tosatto, S. C. E. (2011). RING: networking interacting residues, evolutionary information and energetics in protein structures. *Bioinformatics*, *27*(14), 2003–2005. https://doi.org/10.1093/bioinformatics/btr191

Memišević, V., Wallqvist, A., & Reifman, J. (2013). Reconstituting protein interaction networks using parameter-dependent domain-domain interactions. *BMC Bioinformatics*, *14*(1), 154. https://doi.org/10.1186/1471-2105-14-154

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., … Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, *108*(49), E1293–E1301. https://doi.org/10.1073/pnas.1111471108

O'Neil, K. T., & DeGrado, W. F. (1990). How calmodulin binds its targets: sequence independent recognition of amphiphilic α-helices. *Trends in Biochemical Sciences*, *15*(2), 59–64. https://doi.org/10.1016/0968-0004(90)90177-D

Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research*, *31*(13), 3635–3641. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12824383

Ochoa, D., & Pazos, F. (2010). Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics*, *26*(10), 1370–1371. https://doi.org/10.1093/bioinformatics/btq137

Osawa, M., Tokumitsu, H., Swindells, M. B., Kurihara, H., Orita, M., Shibanuma, T., … Ikura, M. (1999). letters A novel target recognition revealed by calmodulin in complex with kinase kinase, *75*, 819–824.

Oteri, F., Nadalin, F., Champeimont, R., & Carbone, A. (2017). BIS2Analyzer: A server for co-evolution analysis of conserved protein families. *Nucleic Acids Research*, *45*(W1), W307–W314. https://doi.org/10.1093/nar/gkx336

Ovchinnikov, S., Kamisetty, H., & Baker, D. (2014). Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *ELife*, *3*, e02030. https://doi.org/10.7554/eLife.02030

Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., & Maltsev, N. (1999). The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences*, *96*(6), 2896–2901. https://doi.org/10.1073/pnas.96.6.2896

Pearson, R. B., & Kemp, B. E. (1991). Protein kinase phosphorylation site sequences and consensus specificity motifs: tabulations. *Methods in Enzymology*, *200*, 62–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1956339

Pellicena, P., & Kuriyan, J. (2006). Protein-protein interactions in the allosteric regulation of protein kinases. *Current Opinion in Structural Biology*, *16*(6), 702–709. https://doi.org/10.1016/j.sbi.2006.10.007

Piovesan, D., Minervini, G., & Tosatto, S. C. E. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Research*, *44*, 367–374. https://doi.org/10.1093/nar/gkw315

Plewczynski, D., Basu, S., & Saha, I. (2012). AMS 4.0: consensus prediction of post-translational modifications in protein sequences. *Amino Acids*, *43*(2), 573–582. https://doi.org/10.1007/s00726-012-1290-2

Plewczynski, D., Tkacz, A., Wyrwicz, L. S., & Rychlewski, L. (2005). AutoMotif server:

prediction of single residue post-translational modifications in proteins. *Bioinformatics*, *21*(10), 2525–2527. https://doi.org/10.1093/bioinformatics/bti333

Promponas, V. J., Ouzounis, C. A., & Iliopoulos, I. (2014). Experimental evidence validating the computational inference of functional associations from gene fusion events: A critical survey. *Briefings in Bioinformatics*, *15*(3), 443–454. https://doi.org/10.1093/bib/bbs072

Qi, Y., Bar-Joseph, Z., & Klein-Seetharaman, J. (2006). Evaluation of different biological data and computational classification methods for use in protein interaction prediction, *63*(3), 490–500. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16450363

Qian Xu, Q., Pan, S. J., Xue, H. H., & Qiang Yang, Q. (2011). Multitask Learning for Protein Subcellular Location Prediction. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *8*(3), 748–759. https://doi.org/10.1109/TCBB.2010.22

Racioppi, L., ¶1, §, & Means, A. R. (2012). Calcium/Calmodulin-dependent Protein Kinase Kinase 2: Roles in Signaling and Pathophysiology *. https://doi.org/10.1074/jbc.R112.356485

Rajagopala, S. V, & Uetz, P. (2009). Analysis of protein-protein interactions using array-based yeast two-hybrid screens. *Methods Mol Biol*, *548*, 223–245. https://doi.org/10.1007/978-1-59745-540-4_13

Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. S. (2014). Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics*, *2014*(ii), 1–12. https://doi.org/10.1155/2014/147648

Reynolds, K. A., McLaughlin, R. N., & Ranganathan, R. (2011). Hot spots for allosteric regulation on protein surfaces. *Cell*, *147*(7), 1564–1575. https://doi.org/10.1016/j.cell.2011.10.049

Rhodes, D. R., Tomlins, S. A., Varambally, S., Mahavisno, V., Barrette, T., Kalyana-Sundaram, S., … Chinnaiyan, A. M. (2005). Probabilistic model of the human protein-protein interaction network. *Nature Biotechnology*, *23*(8), 951–959. https://doi.org/10.1038/nbt1103

Rual, J.-F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., … Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature News*, *437*(7062), 1173–1178. https://doi.org/10.1038/nature04209

Senga, Y., Ishida, A., Shigeri, Y., Kameshita, I., & Sueyoshi, N. (2015). The Phosphatase-Resistant Isoform of CaMKI, Ca 2+ /Calmodulin- Dependent Protein Kinase I δ (CaMKI δ ), Remains in Its " Primed " Form without Ca 2+ Stimulation. https://doi.org/10.1021/bi5012139

Shiloh, R., Bialik, S., & Kimchi, A. (2014). The DAPK family: A structure-function analysis. *Apoptosis*, *19*(2), 286–297. https://doi.org/10.1007/s10495-013-0924-5

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., … Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1). https://doi.org/10.1038/msb.2011.75

Skrabanek, L., Saini, H. K., Bader, G. D., & Enright, A. J. (2008). Computational prediction of protein-protein interactions. *Molecular Biotechnology*. https://doi.org/10.1007/s12033-007-0069-2

Soderling, T. R., & Stull, J. T. (2001). Structure and Regulation of Calcium / Calmodulin-Dependent Protein Kinases.

Srinivas, K., Rao, A. A., Sridhar, G. R., & Gedela, S. (2008). Methodology for Phylogenetic Tree Construction Journal of Proteomics & Bioinformatics, *011*, 5–11. https://doi.org/10.4172/jpb.s1000002

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler, H., … Wanker, E. E. (2005). A human protein-protein interaction network: A resource for annotating the proteome. *Cell*, *122*(6), 957–968. https://doi.org/10.1016/j.cell.2005.08.029

Swulius, M. T., & Waxham, M. N. (2008). Ca(2+)/calmodulin-dependent protein kinases. *Cellular and Molecular Life Sciences : CMLS*, *65*(17), 2637–2657. https://doi.org/10.1007/s00018-008-8086-2

Taniguchi, M., Okamoto, R., Ito, M., Goto, I., Fujita, S., Konishi, K., … Itoh, T. (2015). New isoform of cardiac myosin light chain kinase and the role of cardiac myosin phosphorylation in α1-adrenoceptor mediated inotropic response. *PLoS ONE*, *10*(10), 1–17. https://doi.org/10.1371/journal.pone.0141130

Tatusov, R. L., Koonin, E. V, & Lipman, D. J. (1997). A genomic perspective on protein families. *Science (New York, N.Y.)*, *278*(5338). https://doi.org/10.1126/science.278.5338.631

Taylor, I. W., Linding, R., Warde-Farley, D., Liu, Y., Pesquita, C., Faria, D., … Wrana, J. L. (2009). Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnology*, *27*(2), 199–204. https://doi.org/10.1038/nbt.1522

Temmerman, K., Simon, B., & Wilmanns, M. (2013). Structural and functional diversity in the activity and regulation of DAPK-related protein kinases. *FEBS Journal*, *280*(21), 5533–5550. https://doi.org/10.1111/febs.12384

Thaimattam, R., Banerjee, R., Miglani, R., & Iqbal, J. (2007). Protein kinase inhibitors: structural insights into selectivity. *Curr.Pharm.Des*, *13*(27), 2751–2765. https://doi.org/10.2174/138161207781757042

Titz, B. B., Schlesner, M., & Uetz, P. (2004). What do we learn from high-throughput protein interaction data?, *1*(1), 111–121. https://doi.org/10.1586/14789450.1.1.111

Torchala, M., & Bates, P. A. (2014). Predicting the Structure of Protein–Protein Complexes Using the SwarmDock Web Server. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1137, pp. 181–197). https://doi.org/10.1007/978-1-4939-0366-5_13

Torchala, M., Moal, I. H., Chaleil, R. A. G., Agius, R., & Bates, P. A. (2013). A Markov-chain model description of binding funnels to enhance the ranking of docked solutions. *Proteins: Structure, Function, and Bioinformatics*, *81*(12), 2143–2149. https://doi.org/10.1002/prot.24369

Torchala, M., Moal, I. H., Chaleil, R. A. G., Fernandez-Recio, J., & Bates, P. A. (2013). SwarmDock: a server for flexible protein–protein docking. *Bioinformatics*, *29*(6), 807–809. https://doi.org/10.1093/bioinformatics/btt038

Uetz, P., Giot, L., Cagney, G., Mansfield, T. a, Judson, R. S., Knight, J. R., … Rothberg, J. M. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, *403*(6770), 623–627. https://doi.org/10.1038/35001009

Vénien-Bryan, C., Jonic, S., Skamnaki, V., Brown, N., Bischler, N., Oikonomakos, N. G., … Johnson, L. N. (2009). The structure of phosphorylase kinase holoenzyme at 9.9 angstroms resolution and location of the catalytic subunit and the substrate glycogen phosphorylase. *Structure (London, England : 1993)*, *17*(1), 117–127. https://doi.org/10.1016/j.str.2008.10.013

Wagner, A. (2003). How the global structure of protein interaction networks evolves. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1514), 457–466. https://doi.org/10.1098/rspb.2002.2269

Walhout, A. J. M., & Vidal, M. (2001). High-Throughput Yeast Two-Hybrid Assays for Large-Scale Protein Interaction Mapping. *Methods*, *24*(3), 297–306. https://doi.org/10.1006/meth.2001.1190

Wan, J., Kang, S., Tang, C., Yan, J., Ren, Y., Liu, J., … Li, T. (2008). Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Research*, *36*(4), e22–e22. https://doi.org/10.1093/nar/gkm848

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wazen, R. M., Kuroda, S., Nishio, C., Sellin, K., Brunski, J. B., & Nanci, A. (2014). Identification of Cardiac-Specific Myosin Light Chain Kinase, *8*(9), 1385–1395. https://doi.org/10.2217/nnm.12.167.Gene

Will, N., Piserchio, A., Snyder, I., Ferguson, S. B., Giles, D. H., Dalby, K. N., & Ghose, R. (2016). Structure of the C-terminal helical repeat domain of eukaryotic elongation factor 2 kinase. *Biochemistry*, *55*(38), 5377–5386. https://doi.org/10.1021/acs.biochem.6b00711

Wojcik, J., & Schächter, V. (2001). Protein-protein interaction map inference using interacting domain profile pairs. *Bioinformatics (Oxford, England)*, *17 Suppl 1*. https://doi.org/10.1093/bioinformatics/17.suppl_1.S296

Wolanin, P. M., Thomason, P. A., & Stock, J. B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. *Genome Biology*, *3*(10), REVIEWS3013. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12372152

Xue, Y., Li, A., Wang, L., Feng, H., & Yao, X. (2006). PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics*, *7*, 163. https://doi.org/10.1186/1471-2105-7-163

Xue, Y., Liu, Z., Cao, J., Ma, Q., Gao, X., Wang, Q., … Ren, J. (2011). GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Engineering Design and Selection*, *24*(3), 255–260. https://doi.org/10.1093/protein/gzq094

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., & Yao, X. (2008). GPS 2.0, a Tool to Predict Kinase-specific Phosphorylation Sites in Hierarchy. *Molecular & Cellular Proteomics*, *7*(9), 1598–1608. https://doi.org/10.1074/mcp.M700574-MCP200

Yamada, M., Kabir, M. S., & Tsunedomi, R. (2003). Divergent promoter organization may be a preferred structure for gene control in Escherichia coli. *Journal of Molecular Microbiology and Biotechnology*, *6*(3–4), 206–210. https://doi.org/10.1159/000077251

Yoshikawa, S., Takagi, T., Ohsawa, N., Tomabechi, Y., Terada, T., Shirouzu, M., … Yokoyama, S. (2011). Crystal Structure of the Ca 2 /Calmodulin-dependent Protein Kinase Kinase in Complex with the Inhibitor STO-609 * □ S Mutsuko Kukimoto-Niino. https://doi.org/10.1074/jbc.M111.251710

Yu, Z., Deng, Z., Wong, H.-S., & Tan, L. (2010). Identifying Protein-Kinase-Specific Phosphorylation Sites Based on the Bagging–AdaBoost Ensemble Approach. *IEEE Transactions on NanoBioscience*, *9*(2), 132–143. https://doi.org/10.1109/TNB.2010.2043682

Zahiri, J., Bozorgmehr, J. H., & Masoudi-Nejad, A. (2013). Computational Prediction of Protein-Protein Interaction Networks: Algo-rithms and Resources. *Current Genomics*, *14*(6), 397–414. https://doi.org/10.2174/1389202911314060004

Zha, M., Zhong, C., Ou, Y., Han, L., Wang, J., & Ding, J. (2012). Crystal Structures of Human CaMKI a Reveal Insights into the Regulation Mechanism of CaMKI, *7*(9). https://doi.org/10.1371/journal.pone.0044828

Zhao, X., Zhang, W., Xu, X., Ma, Z., & Yin, M. (2012). Prediction of Protein Phosphorylation Sites by Using the Composition of k-Spaced Amino Acid Pairs. *PLoS ONE*, *7*(10), e46302. https://doi.org/10.1371/journal.pone.0046302