



Εθνικό και Καποδιστριακό

Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

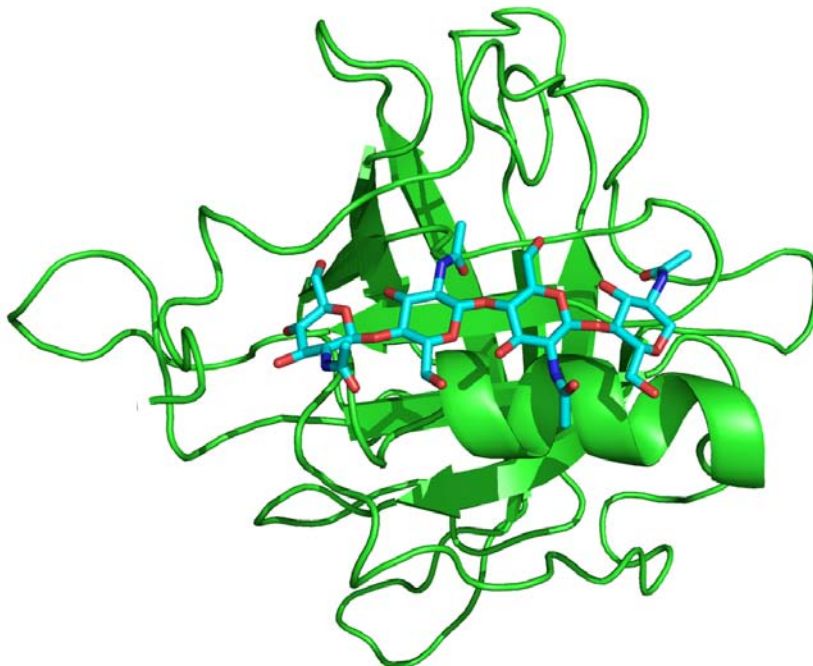
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΤΟΜΕΑΣ ΒΙΟΛΟΓΙΑΣ ΚΥΤΤΑΡΟΥ & ΒΙΟΦΥΣΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΕΙΔΙΚΕΥΣΗΣ ΣΤΗ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Διπλωματική Εργασία

Υπολογιστικές μελέτες των πρωτεϊνών του δερματίου των αρθροπόδων και οι αλληλεπιδράσεις τους με χιτίνη



Γιαννακούρας Θεοχάρης – Ηλίας

Αθήνα, Ιούνιος 2014

Στους γονείς μου,
Βαγγέλη και Κωνσταντίνα

Τριμελής Εξεταστική Επιτροπή

1. Καθηγητής (επιβλέπων) Σταύρος Ι. Χαμόδρακας

Τμήμα Βιολογίας

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

2. Καθηγητής Κωνσταντίνος Ε. Βοργιάς

Τμήμα Βιολογίας

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

3. Επίκουρη Καθηγήτρια Βασιλική Α. Οικονομίδου

Τμήμα Βιολογίας

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Περιεχόμενα

Πρόλογος - Ευχαριστίες	6
Περίληψη	9
Abstract	11
Κεφάλαιο 1 ^ο : Εισαγωγή	13
1.1 Αρθρόποδα	13
1.2 Δερμάτιο αρθροπόδων	15
1.3 Ελικοειδής αρχιτεκτονική	18
1.4 Χιτίνη	20
1.4.1 Χημική δομή χιτίνης.....	21
1.4.2 α-χιτίνη	22
1.4.3 β-χιτίνη	24
1.4.4 γ-χιτίνη.....	26
1.4.5 Μοντέλο Atkins.....	27
1.5 Σύμπλοκο χιτίνης-πρωτεϊνών (μοντέλο Blackwell).....	28
1.6 Πρωτεΐνες του δερματίου	29
1.6.1 Μη δομικές πρωτεΐνες	30
1.6.1.1 Χρωστικές	31
1.6.1.2 Ένζυμα	32
1.6.1.3 Αμυντικές πρωτεΐνες.....	33
1.6.1.4 Αρυλφορίνες (Arylphorins)	34
1.6.2 Δομικές πρωτεΐνες	36
1.6.2.1 Γενική επισκόπηση	36
1.6.2.2 Οικογένειες δομικών πρωτεϊνών	39
1.6.2.2.1 CPR	39
1.6.2.2.2 CPF.....	43
1.6.2.2.3 CPFL	44
1.6.2.2.4 Tweedle (TWDL).....	45
1.6.2.2.5 CPLCG	48
1.6.2.2.6 CPLCW	49
1.6.2.2.7 CPLCA	50

1.6.2.2.8 CPLCP	51
1.6.2.2.9 CPG	51
1.6.2.2.10 Apidermin.....	52
1.6.2.2.11 CPAP3	53
1.6.2.2.12 CPAP1	54
1.6.2.2.13 CPCFC	55
1.6.2.2.14 Δομικές πρωτεΐνες που δεν ανήκουν σε οικογένεια.....	56
1.6.2.3 Συχνά επαναλαμβανόμενα μοτίβα	56
1.6.3 Ομοιότητα με λιποκαλίνες.....	58
1.6.3.1 Λιποκαλίνες	58
1.6.3.2 Η λιποκαλίνη bovine Retinol Binding Protein (PDB ID: 1FEN)	62
1.6.4 Προγνώσεις δομών	66
1.6.4.1 Πρόγνωση δομής CPR (RR-1).....	67
1.6.4.2 Πρόγνωση δομής CPR (RR-2).....	73
1.6.4.3 Πρόγνωση δομής CPF	77
1.6.5 Αλληλεπιδράσεις με χιτίνη (CPR RR-1, CPR RR-2, CPF).....	79
1.7 Στόχοι Διπλωματικής	83
Κεφάλαιο 2 ^ο : Υλικά και Μέθοδοι.....	85
2.1 Πρωτεΐνες – ακολουθίες Tweedle και CPAP3.....	85
2.2 Στοίχιση ακολουθιών	89
2.3 Προτυποποίηση με βάση την ομολογία	90
2.4 Δομική στοίχιση	90
2.5 Δομική ανάλυση και άλλες υπολογιστικές μελέτες	91
2.6 Πειράματα αγκυροβόλησης (Docking)	93
Κεφάλαιο 3 ^ο : Αποτελέσματα	96
3.1 Στοίχιση ακολουθιών	96
3.2 Προτυποποίηση με βάση την ομολογία	111
3.3 Δομική στοίχιση	113
3.4 Δομική ανάλυση και άλλες υπολογιστικές μελέτες	117
3.5 Πειράματα αγκυροβόλησης (Docking)	122
Κεφάλαιο 4 ^ο : Συμπεράσματα – Συζήτηση	130
4.1 Γενική Περιγραφή	130
4.2 Συμπεράσματα – Πρωτεΐνες Tweedle.....	131

4.3 Συμπεράσματα – Πρωτεΐνες CPAP3.....	134
4.4 Επίλογος – Μελλοντικές Προκλήσεις.....	135
Παράρτημα.....	136
1. Δομικά μοντέλα πρωτεϊνών της οικογένειας Tweedle, ύστερα από προτυποποίηση με βάση την ομολογία	137
2. Abstract και Poster στο 35 ^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Βιολογικών Επιστημών (ΕΕΒΕ), 23-25 Μαΐου 2013	140
3. Abstract και Poster στο 8 ^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Υπολογιστικής Βιολογίας και Βιοπληροφορικής, 22-24 Νοεμβρίου 2013	142
Βιβλιογραφία	145

Πρόλογος - Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη Βιοπληροφορική, στον Τομέα Βιολογίας Κυττάρου και Βιοφυσικής, του Τμήματος Βιολογίας, του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Η εκπόνηση της διπλωματικής εργασίας έγινε υπό την επίβλεψη του Καθηγητή Σταύρου Ι. Χαμόδρακα, της Επίκουρης Καθηγήτριας Βασιλικής Α. Οικονομίδου και του Διδάκτορα Νίκου Χ. Παπανδρέου.

Αρχικά, θα επιθυμούσα να ευχαριστήσω ιδιαίτερα τον επιβλέποντα της διπλωματικής μου εργασίας και επιστημονικό υπεύθυνο του Μεταπτυχιακού Διπλώματος Ειδίκευσης, Καθηγητή Σταύρο Χαμόδρακα. Είμαι ιδιαίτερα ευγνώμων που με επέλεξε να συμμετάσχω στο συγκεκριμένο μεταπτυχιακό δίπλωμα ειδίκευσης καθώς και για την ευκαιρία που μου έδωσε να είμαι μέλος της ερευνητικής του ομάδας. Καθ' όλη τη διάρκεια της εκπόνησης της διπλωματικής μου εργασίας με καθοδηγούσε στο ερευνητικό μου έργο, παρέχοντας μου απλόχερα τις επιστημονικές του γνώσεις σε ένα ευρύ φάσμα των επιστημών της Βιοπληροφορικής, της Βιοφυσικής και της Βιοχημείας. Οι συμβουλές του σε κρίσιμα σημεία της εκπόνησης της παρούσας διπλωματικής, αποτέλεσαν για μένα τον ιδανικό οδηγό προς την επιτυχή ολοκλήρωσή της. Επίσης, μέσα και από τις διαλέξεις των μαθημάτων του κατά την διάρκεια του μεταπτυχιακού, μου έδωσε το έναυσμα να γνωρίσω και να εντρυφήσω πάνω σε πρωτόγνωρα για μένα επιστημονικά πεδία, καθώς και να αποκτήσω μια ώριμη και σωστά δομημένη επιστημονική σκέψη. Συνοψίζοντας, το σημαντικότερο ήταν ότι, μέσα από την επικοινωνία και αλληλεπίδραση μαζί του, είχα το ιδανικότερο επιστημονικό και λειτουργικό περιβάλλον, που είχε ως αποτέλεσμα να αγαπήσω, τόσο τις επιστήμες της βιοπληροφορικής και της βιοφυσικής, όσο και ειδικότερα το μοναδικά ενδιαφέρον βιολογικό σύστημα του δερματίου των αρθροπόδων.

Επίσης, θα ήθελα να ευχαριστήσω την Επίκουρη Καθηγήτρια Βασιλική Οικονομίδου που μου κάνει την τιμή να είναι μέλος στην τριμελή εξεταστική επιτροπή της διπλωματικής μου εργασίας. Είναι τιμή για μένα, που πρακτικά συνέχισα το ερευνητικό της έργο πάνω στις πρωτεΐνες του δερματίου, καθώς πολλά επιστημονικά δεδομένα που χρησιμοποίησα, βασίστηκαν σε δικές της μελέτες. Κατά

την διάρκεια της μελέτης μου, πολλές απορίες λυνόντουσαν άμεσα όταν ανέτρεχα σε δικές της προγενέστερες εργασίες. Ακόμα, με την ενεργή της συμμετοχή, βοήθησε στο ευχάριστο κλίμα του εργαστηρίου. Τέλος, την ευχαριστώ και για τις χρήσιμες επιστημονικές διαλέξεις της κατά την διάρκεια των μαθημάτων.

Ακόμα, ευχαριστώ ιδιαίτερα τον καθηγητή Κωνσταντίνο Βοργιά που μου έκανε την τιμή να είναι μέλος της εξεταστικής μου επιτροπής, καθώς και για τις πλούσιες σε γνώσεις διαλέξεις του κατά την διάρκεια των μαθημάτων του μεταπτυχιακού.

Οι έννοιες της λέξης «ευχαριστώ» είναι πολύ μικρές για να περιγράψουν το μέγεθος της ευγνωμοσύνης που νιώθω για όσα μου προσέφερε ο διδάκτορας Νίκος Παπανδρέου κατά την εκπόνηση της διπλωματικής μου εργασίας και συνολικά κατά την διάρκεια της παρουσίας μου στο μεταπτυχιακό δίπλωμα ειδίκευσης. Ήταν δίπλα μου σε οτιδήποτε χρειαζόμουν, ανά πάσα στιγμή. Η βοήθειά του ήταν απεριόριστη σε πολλούς τομείς της παρουσίας μου στο εργαστήριο. Η συμβολή του στην εκπόνηση της παρούσας διπλωματικής ήταν καθοριστική, με την καθημερινή του επίβλεψη και καθοδήγηση, τις επιστημονικές του συμβουλές, την επίλυση των κάθε λογής αποριών μου και με τον συνολικά απεριόριστο χρόνο και την απεριόριστη ενέργεια που επένδυσε ώστε να επιτευχθεί το αποτέλεσμα που βρίσκεται σε αυτές τις σελίδες. Η συνεργασία μου μαζί του ήταν άψογη, χάρη και στην υπομονή που επέδειξε αρκετές φορές, και ειλικρινά τον ευγνωμονώ για αυτό.

Επιπροσθέτως, θα ήθελα να πω ένα μεγάλο ευχαριστώ στην συμφοιτήτριά μου στο μεταπτυχιακό Ζωή Ιωαννίδου. Από την πρώτη στιγμή με βοήθησε να κατανοήσω πάρα πολλές έννοιες σχετικές με το δερμάτιο των αρθροπόδων και τις πρωτεΐνες του, καθώς αποτελούσαν ερευνητικό πεδίο και της δικής της διπλωματικής εργασίας. Με βοήθησε απεριόριστα σε πάρα πολλές στιγμές, κατά την διάρκεια της εκπόνησης της παρούσα μελέτης, προσφέροντας μου χρήσιμες συμβουλές και πολλές εξειδικευμένες γνώσεις.

Εν συνεχεία, θα ήθελα να ευχαριστήσω τους υποψήφιους διδάκτορες Εβίτα Τσιολάκη και Νίκο Λούρο, καθώς με βοήθησαν πάρα πολύ, προσφέροντας μου καθημερινά εξειδικευμένες γνώσεις και συμβουλές τόσο σε θέματα βιολογικού ενδιαφέροντος όσο και σε διάφορα πρακτικά θέματα.

Για την εύρυθμη λειτουργία του μεταπτυχιακού διπλώματος ειδίκευσης, καθώς και του εργαστηρίου βιοφυσικής και βιοπληροφορικής, δουλεύουν καθημερινά πολλοί άνθρωποι, τους οποίους νιώθω την ανάγκη να τους ευχαριστήσω. Ιδιαίτερη αναφορά οφείλω να κάνω στην Διδάκτορα Ζωή Λίτου, καθώς και πάλι στον Διδάκτορα Νίκο Παπανδρέου, των οποίων η συνεισφορά στην ομαλή λειτουργία του μεταπτυχιακού είναι σημαντικότερη, λύνοντας πολλά καθημερινά προβλήματα και βοηθώντας με όλους τους τρόπους τα υπόλοιπα μέλη του εργαστηρίου.

Επίσης, θα ήθελα να ευχαριστήσω όλα τα μέλη του εργαστηρίου και τους συμφοιτητές μου στο μεταπτυχιακό, που δημιούργησαν για μένα ένα πολύ ευχάριστο περιβάλλον με πολλές όμορφες στιγμές, τόσο μέσα στο εργαστήριο, όσο και έξω από αυτό.

Κλείνοντας, ειδικές ευχαριστίες οφείλω να αποδώσω στην οικογένεια μου, που με στήριξε, και συνεχίζει να με στηρίζει, σε όλες τις αποφάσεις που πήρα στη διάρκεια της ζωής μου, συμβουλευοντάς με, έχοντας πάντα ως γνώμονα το καλύτερο για μένα.

«Στους γονείς μου οφείλω το ζην και στους δάσκαλούς μου το εύ ζην!»

Μέγας Αλέξανδρος

Υ.Γ. Η παρούσα εργασία παρουσιάστηκε:

1. Στο 35^ο ετήσιο επιστημονικό συνέδριο της Ελληνικής Εταιρίας Βιολογικών Επιστημών (ΕΕΒΕ), στο Ναύπλιο, στις 23-25 Μαΐου 2013.
2. Στο 8^ο ετήσιο επιστημονικό συνέδριο της Ελληνικής Εταιρίας Υπολογιστικής Βιολογίας και Βιοπληροφορικής (HSCBB), στη Λαμία, στις 22-24 Νοεμβρίου 2013.

Περίληψη

Το δερμάτιο είναι ένα διμερές, σύνθετο υλικό που παρέχει δομική και μηχανική υποστήριξη δρώντας λειτουργικά τόσο ως δέρμα όσο και ως σκελετός στα αρθρόποδα. Αποτελείται από ίνες χιτίνης ενσωματωμένες σε μία πρωτεϊνική 'μήτρα'. Οι φυσικές του ιδιότητες καθορίζονται από την δομή των δυο βασικών δομικών στοιχείων του, των πρωτεϊνών του δερματίου (CP) και της χιτίνης, καθώς και από τις αλληλεπιδράσεις μεταξύ τους. Η κύρια οικογένεια πρωτεϊνών του δερματίου είναι η CPR (πρωτεΐνες που εντοπίζεται στην αμινοξική τους ακολουθία η εκτεταμένη R&R συναινετική ακολουθία) οι οποίες έχει αποδειχθεί πειραματικά ότι δεσμεύουν χιτίνη. Μελέτες Συγκριτικής Προτυποποίησης και πειράματα Αγκυροβόλησης οδήγησαν στην πρόταση ότι το βασικό δομικό μοτίβο των πρωτεϊνών που φέρουν την R&R συναινετική ακολουθία είναι ένα αντιπαράλληλο β-πτυχωτό φύλλο που σχηματίζει ένα μισό 'β-βαρέλι', ενώ προτάθηκαν πιθανοί τρόποι αλληλεπίδρασης τους με τη χιτίνη. Οι μελέτες βασίστηκαν στην παρατήρηση ότι οι CPRs ομοιάζουν, σε επίπεδο αλληλουχίας, με την πρωτεΐνη του πλάσματος που δεσμεύει ρετινόλη (RBP) και ανήκει στην οικογένεια των λιποκαλινών. Το επόμενο βήμα ήταν να μελετηθούν και άλλες οικογένειες δομικών πρωτεϊνών του δερματίου και οι αλληλεπιδράσεις τους με χιτίνη. Για παράδειγμα, η οικογένεια Tweedle αποτελεί μια μεγάλη οικογένεια πρωτεϊνών του δερματίου για την οποία πρόσφατα πειραματικά δεδομένα αποδεικνύουν ότι τα μέλη της αλληλεπιδρούν με χιτίνη. Μέχρι σήμερα, 236 ακολουθίες από 17 διαφορετικά είδη έχουν χαρακτηριστεί ως μέλη της. Όπως οι CPRs έτσι και οι πρωτεΐνες Tweedle εμφανίζουν ομοιότητα σε επίπεδο αλληλουχίας με την λιποκαλίνη bovine RBP. Με τη μέθοδο της συγκριτικής προτυποποίησης κατασκευάστηκαν μοντέλα για μέλη της οικογένειας Tweedle. Από την ανάλυσή τους προτείνεται ότι το βασικό δομικό μοτίβο τους είναι ένα 'β-βαρέλι', που απαρτίζεται από 8 αντιπαράλληλους β-κλώνους και μία α-έλικα παράλληλη προς τους β-κλώνους. Δομική ανάλυση των μοντέλων έδειξε κοινά χαρακτηριστικά και ιδιότητες με τις πρωτεΐνες της οικογένειας των λιποκαλινών. Επίσης, πραγματοποιήθηκαν πειράματα αγκυροβόλησης για να διαπιστωθεί πως τα μέλη των Tweedle αλληλεπιδρούν με χιτίνη. Μια άλλη οικογένεια δομικών πρωτεϊνών είναι η CPAP3, η οποία περιέχει domain (ChtBD2) το οποίο έχει αποδειχθεί σε άλλες πρωτεΐνες εκτός δερματίου ότι αλληλεπιδρά με χιτίνη. Μέχρι σήμερα, 63 ακολουθίες

από 10 διαφορετικά είδη έχουν χαρακτηριστεί ως μέλη της. Οι πρωτεΐνες της οικογένειας CPAP3 (όπως οι CPRs και οι Tweedle) εμφανίζουν ομοιότητα σε επίπεδο αλληλουχίας με την λιποκαλίνη bovine RBP. Με τη μέθοδο της συγκριτικής προτυποποίησης κατασκευάστηκαν μοντέλα για τα μέλη της οικογένειας CPAP3. Από την ανάλυσή τους, προτείνεται ότι το βασικό δομικό τους μοτίβο είναι ένα ολόκληρο β-βαρέλι, που αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους, μια (1) α-έλικα παράλληλη προς αυτούς και μια (1) μικρή α-έλικα στο ένα άνοιγμα του βαρελιού. Συμπερασματικά, τα δομικά μοντέλα των οικογενειών Tweedle και CPAP3 παρουσιάζουν ομοιότητες με τις λιποκαλίνες και κρίνονται κατάλληλα για αλληλεπίδραση με χιτίνη.

Abstract

The physical properties of cuticle are determined by the structure of its two major components, cuticular proteins (CPs) and chitin, and, also, by their interactions. The major family of cuticular proteins is CPR (the extended R&R consensus motif is present in their aminoacid sequences) and it has been experimentally proved that its members bind chitin. Homology modeling studies and docking experiments suggested an antiparallel β -sheet half-barrel structure as the basic folding motif of the R&R Consensus and provided insights for the modes of CPRs interactions with chitin. These attempts were based on the observation that CPR protein sequences share sequence similarity to bovine plasma retinol binding (RBP) protein, which belongs to the lipocalin family. The next step was to study other families of structural cuticular proteins and their interactions with chitin. For example, Tweedle (TWDL) is a large family of structural cuticular proteins and it has been experimentally proved that bind to chitin. To date, 236 sequences from 17 different species have been identified as members of this family. Like CPRs, Tweedle proteins display remote sequence similarity with bovine RBP. Homology modeling was used to build structural models of the Tweedle proteins. Analysis of their structural characteristics led us to propose that their basic folding motif is a β -barrel, which consists mainly of eight antiparallel β -strands and an α -helix parallel to the strands. Furthermore, the structural characteristics of the models analyzed and revealed structural similarities with lipocalins. Finally, docking experiments were performed, to investigate how members of the Tweedle family may interact with chitin chains in order to certify previous experimental results. Another interesting family of structural cuticular proteins is CPAP3, that includes a chitin binding domain in non cuticular proteins. To date, 62 sequences from 10 different species have been identified as members of this family. CPAP3 proteins (like CPRs and Tweedle) display remote sequence similarity with bovine RBP. Homology modeling was used to build structural models of the CPAP3 proteins. Analysis of their structural characteristics led us to propose that their basic folding motif is a β -barrel, which consists mainly of eight (8) antiparallel β -strands, an α -helix parallel to the strands and a small α -helix near one extremity of the barrel. In conclusion, the protein models of Tweedle and CPAP3 families display similarities with the lipocalin family and they considered suitable for chitin binding.

Εισαγωγή



«Οι αιτίες είναι τέσσερις. Μία από αυτές είναι η ουσία ή ουσιαστική φύση ενός πράγματος (γιατί ο λόγος ύπαρξης ενός πράγματος ανάγεται τελικά στον τύπο του), άλλη αιτία είναι η ύλη ή αυτό που υπόκειται, το υποκείμενο, τρίτη αιτία είναι η πηγή της κίνησης και τέταρτη είναι η αιτία που στέκει αντίκρυ της, ο σκοπός και το «αγαθό» (γιατί αυτό είναι ο τελικός σκοπός κάθε κίνησης και γένεσης)».
(*Φυσικά Ι*)

Αριστοτέλης (384 π.Χ. – 322 π.Χ.)

Κεφάλαιο 1^ο : Εισαγωγή

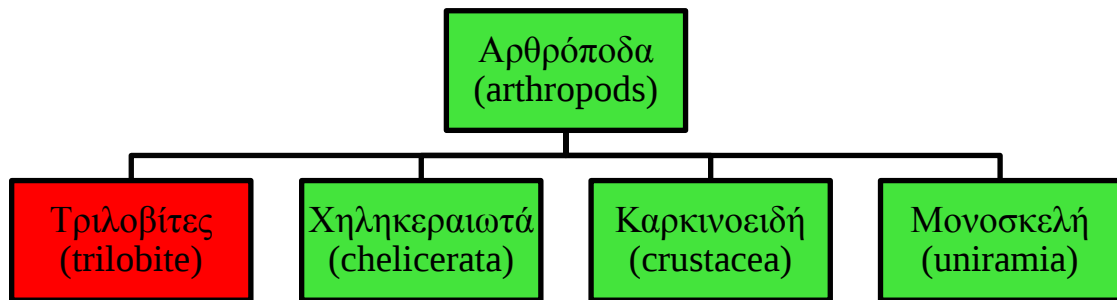
1.1 Αρθρόποδα

Το φύλο (phylum) των αρθροπόδων είναι το πιο άφθονο και το πιο ποικίλο φύλο στο ζωικό βασίλειο και αποτελεί από τα 3/4 περίπου από όλων τα γνωστών ειδών. Περίπου 900.000 είδη αρθροπόδων έχουν καταγραφεί και πιθανότατα αρκετά ακόμα θα καταγραφούν στο μέλλον (Hickman, Roberts et al. 2001).

Τα αρθρόποδα είναι μεταμερικά, ευκοιλωματικά πρωτοστόμια με καλά ανεπτυγμένο σύστημα οργάνων. Τα περισσότερα εμφανίζουν μεταμερισμό (tagmatization). Υπάρχουν σε όλα τα ενδιαιτήματα που είναι ικανά να υποστηρίξουν έμβιους οργανισμούς. Ίσως περισσότερο από οποιοδήποτε άλλο παράγοντα, η επικράτηση των αρθροπόδων οφείλεται στις προσαρμογές που έγιναν δυνατές μέσω του δερμάτιού τους (που όπως θα αναφερθεί παρακάτω είναι ένα σύστημα πρωτεϊνών – χιτίνης). Άλλα σημαντικά στοιχεία των αρθροπόδων είναι η αναπνοή μέσω της τραχείας, πολύ αποτελεσματικά αισθητήρια όργανα, σύνθετη συμπεριφορά και η ικανότητα της μεταμόρφωσης (metamorphosis) (Hickman, Roberts et al. 2001).

Ελάχιστα αρθρόποδα ξεπερνούν τα 60 cm σε μήκος, ενώ τα περισσότερα είναι πολύ πιο μικρά σε μέγεθος. Το μεγαλύτερο είναι ένα γιαπωνέζικο καβούρι, που ονομάζεται *Macrocheira kaempferi*, που φτάνει σε μήκος τα 3.8 m και μικρότερο ένα παρασιτικό ζώο, που ονομάζεται *Demodex*, που είναι μικρότερο από 0.1 mm σε μήκος (Hickman, Roberts et al. 2001).

Το φύλο (phylum) των αρθροπόδων χωρίζεται σε τέσσερα (4) υπόφυλλα (subphylums). Τα υπόφυλλα είναι: οι τριλοβίτες (trilobite), τα χηληκεραιωτά (chelicerata), τα καρκινοειδή (crustacea) και τα μονοσκελή (uniramia) (σχήμα 1). Οι τριλοβίτες ήταν κυρίαρχοι στην παλαιοζωική περίοδο, ενώ σήμερα έχουν εξαφανιστεί. Στα χηληκεραιωτά ανήκουν είδη όπως οι αράχνες και τα τσιμπούρια. Στα καρκινοειδή ανήκουν τα καβούρια. Στα μονοσκελή ανήκουν τα έντομα (insects) και τα μυριάποδα (myriapods) (Hickman, Roberts et al. 2001). Κάποια παραδείγματα παρουσιάζονται στην εικόνα 1.



Σχήμα 1: Το φύλο (phylum) των αρθροπόδων χωρίζεται σε τέσσερα (4) υπόφυλλα (subphylums). Τα υπόφυλλα είναι: οι τριλοβίτες (trilobite), τα χηληκεραιωτά (chelicerata), τα καρκινοειδή (crustacea) και τα μονοσκελή (uniramia). Οι τριλοβίτες είναι χρωματισμένοι με κόκκινο χρώμα καθώς έχουν πλέον εξαφανιστεί.



Dalmanites limulurus *Brachypelma vagans* *Grapsus grapsus* *Nasonia vitripennis*

Εικόνα 1: Ενδεικτικά παραδείγματα αρθροπόδων από κάθε υπόφυλλο (subphylum). Στο 1 είναι το *Dalmanites limulurus*, εκπρόσωπος των εξαφανισμένων τριλοβιτών (trilobite). Στο 2 είναι η αράχνη *Brachypelma vagans*, εκπρόσωπος των χηληκεραιωτών (chelicerata). Στο 3 είναι το καβούρι *Grapsus grapsus*, εκπρόσωπος των καρκινοειδών (crustacea). Τέλος, στο 4 είναι η σφήκα *Nasonia Vitripennis* που ανήκει στα μονοσκελή (uniramia).

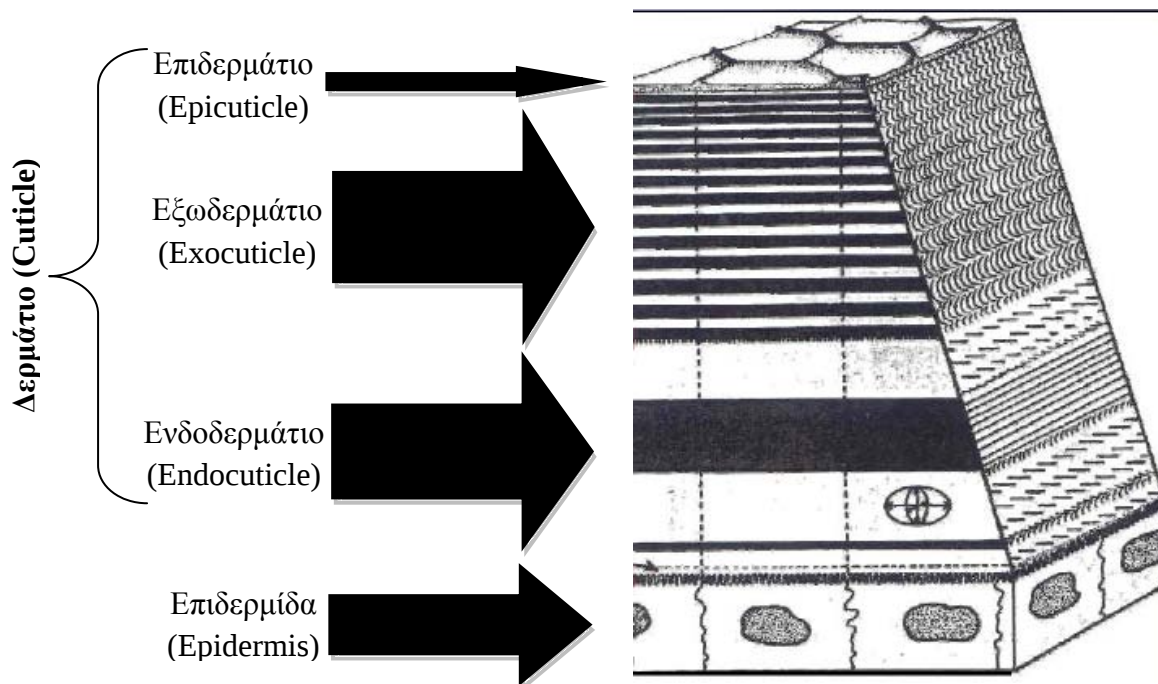
1.2 Δερμάτιο αρθροπόδων

Το δερμάτιο (cuticle) είναι ένα διμερές, σύνθετο υλικό που παρέχει δομική και μηχανική υποστήριξη, δρώντας λειτουργικά τόσο ως δέρμα όσο και ως σκελετός, στα αρθρόποδα. Το δερμάτιο προστατεύει το αρθρόποδο, του δίνει το σχήμα του, το βοηθάει στις κινήσεις του και του προσδίδει κάποιες τοπικές μηχανικές ιδιότητες όπως μεγάλη ελαστικότητα, πρόσφυση, αντοχή στη φθορά και έλεγχος της διάχυσης. Μπορεί επίσης να χρησιμοποιηθεί ως προσωρινή «αποθήκη» τροφίμων και είναι ένα σημαντικό εμπόδιο για τα παράσιτα και τις ασθένειες που πιθανά προκαλούν (Neville 1975; Vincent and Wegst 2004).

Το δερμάτιο αποτελείται από ίνες χιτίνης ενσωματωμένες σε μία πρωτεϊνική “μήτρα”. Οι φυσικές του ιδιότητες καθορίζονται κυρίως από την δομή των δυο βασικών δομικών στοιχείων του, των πρωτεϊνών του δερματίου (cuticular proteins - CP) και της χιτίνης (chitin), από τις αλληλεπιδράσεις μεταξύ τους και επίσης από τον βαθμό σκλήρυνσης (sclerotization) του, την παρουσία λιπιδίων, την παρουσία μεταλλικών ιόντων (Zn, Mn, Fe) και τον βαθμό ενυδάτωσής του (Neville 1975; Vincent and Wegst 2004).

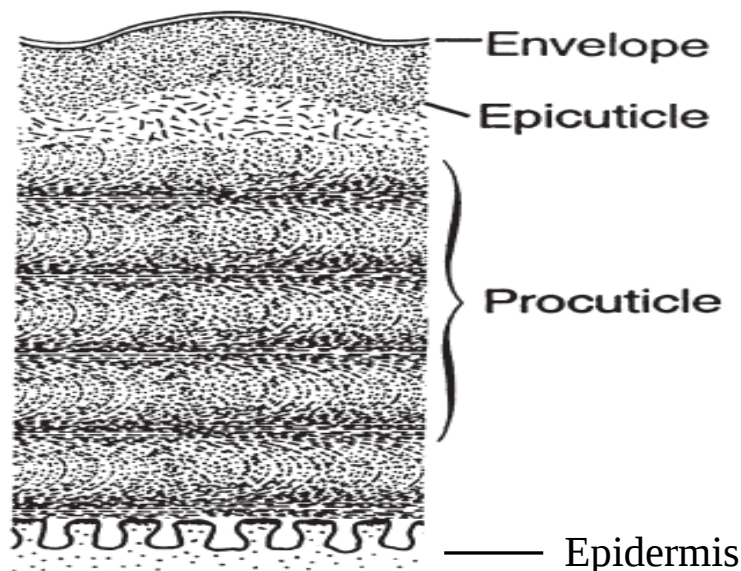
Το σχήμα του δερματίου είναι ένα επίπεδο φύλλο. Είναι κυρτό και κυματοειδές, και πολλές περιοχές του έχουν μια «κλειστή» μορφή με κυκλική διατομή. Ακόμα και τα φτερά και τα έλυτρα έχουν σκληρυνθεί ως αποτέλεσμα του κυρτού σχήματος (Vincent and Wegst 2004).

Αρχικά το δερμάτιο χωρίστηκε σε τρία (3) επίπεδα (layers). Πρόκειται για τα (από μέσα προς τα έξω): ενδοδερμάτιο (endocuticle), εξωδερμάτιο (exocuticle) και επιδερμάτιο (epicuticle) (εικόνα 2) (Neville 1975). Σήμερα, (Willis, Papandreou et al. 2012) έχει επικρατήσει ο διαχωρισμός του δερματίου σε τρία (3) επίπεδα (layers) σύμφωνα με τις παρακάτω αλλαγές. Τα επίπεδα ενδοδερμάτιο και εξωδερμάτιο έχουν συνενωθεί και ονομάζονται προδερμάτιο (procuticle), το επίπεδο του επιδερματίου παρέμεινε ως έχει και τέλος προστέθηκε ένα νέο επίπεδο εξωτερικά που ονομάζεται φάκελος (envelope - παλιότερα ονομαζόταν cuticulin) (εικόνα 3) (Locke 2001).



Εικόνα 2: Τα τρία (3) επίπεδα (layers) του δερματίου, όπως είχαν χωριστεί αρχικά. Πρόκειται για τα (από μέσα προς τα έξω): ενδοδερμάτιο, εξωδερμάτιο και επιδερμάτιο. Στην εικόνα φαίνεται και η επιδερμίδα (epidermis) όπου μαζί με το δερμάτιο (cuticle) αποτελούν το εξωτερικό περίβλημα των αρθροπόδων (Neville 1975).

The Three Layers of Cuticle



Εικόνα 3: Τα τρία (3) επίπεδα (layers) που χωρίζεται σήμερα το δερμάτιο. Πρόκειται για τα (από μέσα προς τα έξω): προδερμάτιο (procuticle), επιδερμάτιο (epicuticle) και φάκελος (envelope) (Locke 2001). Στην εικόνα φαίνεται και η επιδερμίδα όπου μαζί με το δερμάτιο (cuticle) αποτελούν το εξωτερικό περίβλημα των αρθροπόδων.

Το επίπεδο προδερμάτιο είναι η ένωση των παλιότερων επιπέδων εξωδερμάτιο και ενδοδερμάτιο. Είναι με μεγάλη διάφορα το μεγαλύτερο σε μέγεθος επίπεδο του δερματίου. Σχηματίζεται από φύλλα, τα οποία είναι ενιαία στρώματα ινών με παράλληλο προσανατολισμό το ένα στο άλλο μέσα σε κάθε επίπεδο, με μια μικρή αλλά σταθερή αλλαγή στον προσανατολισμό από φύλλο σε φύλλο για να δώσει το γνωστό μοτίβο ελικοειδής αρχιτεκτονικής που συναντάμε στο δερμάτιο (θα αναλυθεί στην ενότητα 1.3). Το προδερμάτιο είναι η περιοχή που συνδυάζει την χιτίνη με τις πρωτεΐνες σε ποικίλους συνδυασμούς, σκληροποιείται (sclerotized) και χρωματίζεται σε διάφορους βαθμούς (Locke 2001; Willis, Papandreou et al. 2012).

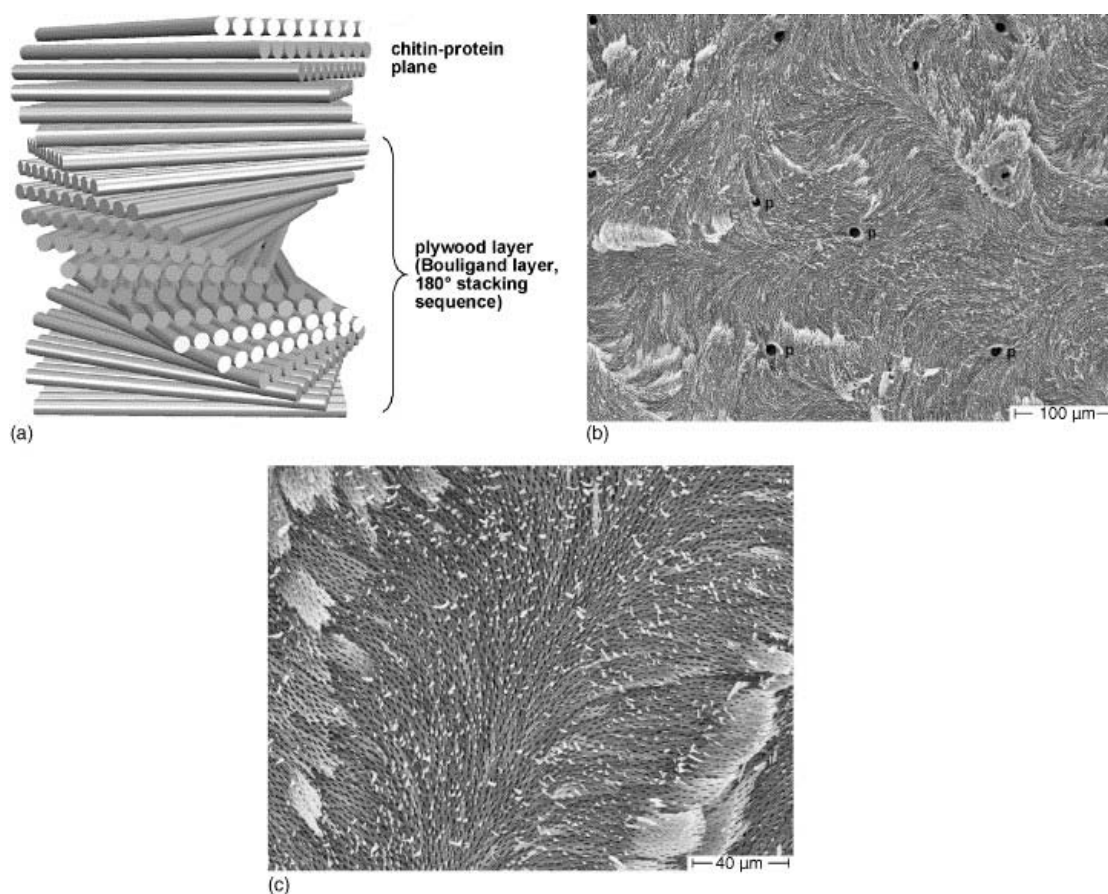
Το επόμενο επίπεδο είναι το επιδερμάτιο που έχει πάχος περίπου 1μm. Είναι ένα επίπεδο χωρίς χιτίνη το οποίο σταθεροποιείται από κινόνες (quinones). Παλιότερα ονομαζόταν και “εσωτερικό επιδερμάτιο”, με το cuticulin να ήταν το εξωτερικό (Locke 2001; Willis, Papandreou et al. 2012).

Στην αρχή κάθε κύκλου έκδυσης (molt cycle) στο δερμάτιο, δημιουργούνται μικρολάχνες με πλάκες στις άκρες τους, οι οποίες δημιουργούν τον φάκελο (envelope). Το διακριτό επίπεδο του φακέλου (envelope) έχει μέγεθος 10-30 nm, χρησιμεύει όχι μόνο στην προστασία της βαθύτερης επιδερμίδας από τα ένζυμα που προέρχονται από το υγρό έκδυσης όταν αρχίζουν να «πέπτουν» το παλιό δερμάτιο, αλλά επηρεάζει την αντοχή στην τριβή και τη μόλυνση, την διείσδυση των εντομοκτόνων, την διαπερατότητα, την ανακλαστικότητα της επιφάνειας και τα φυσικά χρώματα. Οι ακολουθίες και οι ιδιότητες από τις πρωτεΐνες που το αποτελούν παραμένουν άγνωστες (Locke 2001; Willis, Papandreou et al. 2012).

Υπάρχουν δύο (2) είδη δερματίου, το μαλακό (soft/flexible) δερμάτιο και το σκληρό (hard/rigid) δερμάτιο. Το μαλακό δερμάτιο περιέχει σε βάρος ίσα κλάσματα σε χιτίνη και πρωτεΐνες, ενώ το υπόλοιπο 40% - 75% αποτελείται από νερό. Το σκληρό δερμάτιο περιέχει σε βάρος 15-30% σε χιτίνη και μόνο 12% νερό (Reynolds 1975; Vincent and Wegst 2004).

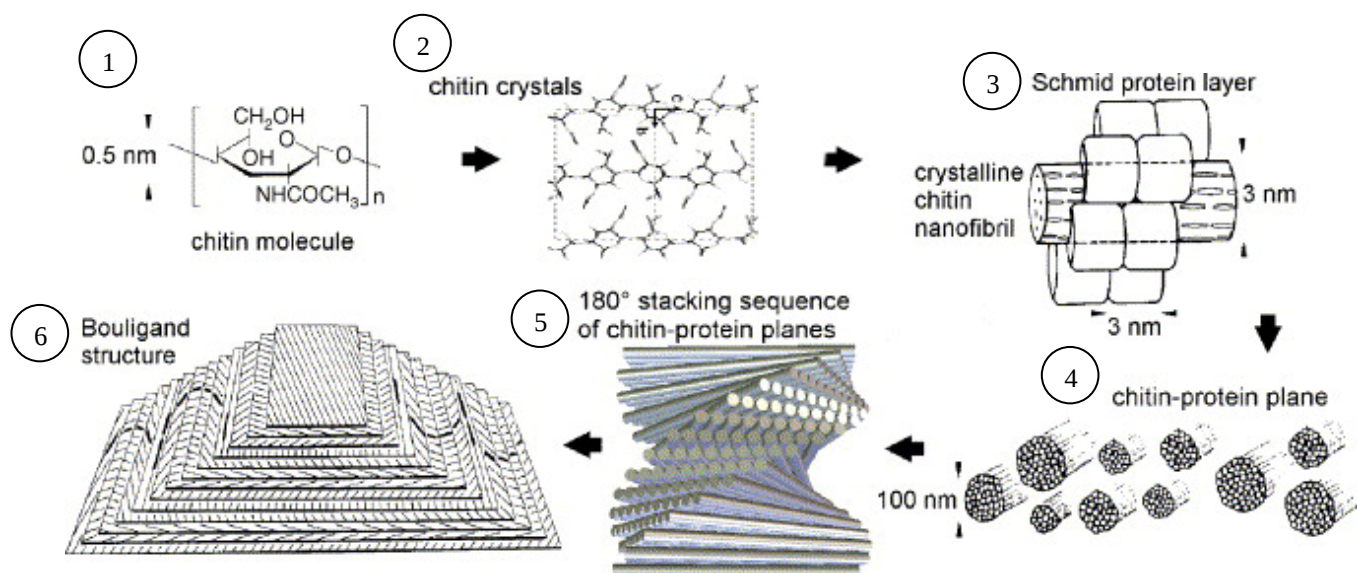
1.3 Ελικοειδής αρχιτεκτονική

Το δερμάτιο παρουσιάζει δομή με την μορφή ελικοειδούς αρχιτεκτονικής. Η ελικοειδής αρχιτεκτονική μπορεί να περιγραφεί ως ελικοειδώς διευθετημένα παράλληλα επίπεδα από ινίδια, που είναι προσανατολισμένα παράλληλα μεταξύ τους σε κάθε επίπεδο. Μεταξύ διαδοχικών επιπέδων, η διεύθυνση των ινιδίων αλλάζει κατεύθυνση καθώς τα ινίδια περιστρέφονται προοδευτικά κατά μικρή και σταθερή γωνία, δημιουργώντας μια αριστερόστροφη έλικα με τον άξονα κάθετο στο επίπεδο (Hamodrakas 1992). Στην εικόνα 4 παρουσιάζονται μια σχηματική παρουσίαση και δυο πειραματικές εικόνες της ελικοειδούς αρχιτεκτονικής στα αρθρόποδα.



Εικόνα 4: Σχηματική απεικόνιση (a) και πειραματικές (b,c) εικόνες της ελικοειδής αρχιτεκτονικής που δημιουργείται στο σύστημα πρωτεϊνών-χιτίνης στο δερμάτιο των αρθροπόδων. Οι SEM (Scanning Electron Microscopy – Ηλεκτρονικό μικροσκόπιο σάρωσης) μικρογραφίες (b,c) του ενδοδερμάτιου (που ανήκει στο επίπεδο προδερμάτιο), παράλληλα προς την επιφάνεια, δείχνουν την σταδιακή περιστροφή στη δομή. Οι τρύπες (p) είναι πόροι του δερμάτιου (Raabe, Romano et al. 2006).

Υπάρχουν έξι (6) ιεραρχικά χωρισμένα επίπεδα, τα οποία οδηγούν στην ελικοειδή αρχιτεκτονική στο δερμάτιο. Το πρώτο επίπεδο είναι το μόριο της χιτίνης. Η αντιπαράλληλη ευθυγράμμιση αυτών των μορίων οδηγεί σε κρυστάλλους χιτίνης, οι οποίοι μπορούν να θεωρηθούν το δεύτερο επίπεδο στην ιεραρχία. Το τρίτο δομικό επίπεδο είναι η διάταξη 18-25 αλυσίδων χιτίνης, σε μορφή στενόμακρων κρυσταλλικών μονάδων, που τυλιγμένες με πρωτεΐνες σχηματίζουν νανο-ινίδια με διάμετρο περίπου 2-5 nm και μήκος περίπου 300 nm (τα τρία πρώτα επίπεδα θα παρουσιαστούν αναλυτικά στις ενότητες 1.4 και 1.5). Το τέταρτο επίπεδο είναι η ομαδοποίηση αυτών το νανο-ινιδίων σε μακριές ίνες χιτίνης-πρωτεϊνών με διάμετρο περίπου 50-250 nm. Το πέμπτο επίπεδο στην ιεραρχία προκύπτει από τον σχηματισμό ενός επιπέδου περιπλεγμένου και διακλαδισμένου δικτύου ινών χιτίνης-πρωτεϊνών. Το διάστημα ανάμεσα τους είναι γεμάτο με μια ποικιλία πρωτεϊνών και διαφόρων μετάλλων. Το έκτο επίπεδο, το οποίο είναι ορατό σε ένα φωτονικό μικροσκόπιο, είναι το μοτίβο με τα «στριμένα φύλλα». Αυτό το επίπεδο δημιουργείται από τα επίπεδα χιτίνης-πρωτεϊνών. Η σταδιακή περιστροφή αυτών των επιπέδων από στρώμα σε στρώμα δημιουργεί την τελική ελικοειδή αρχιτεκτονική όπως περιγράφηκε (Raabe, Romano et al. 2006). Τα επίπεδα που μόλις περιγράφηκαν παρουσιάζονται στην εικόνα 5.



Εικόνα 5: Τα έξι (6) ιεραρχικά επίπεδα της «μήτρας» χιτίνης – πρωτεϊνών που δημιουργούν την δομή ελικοειδούς αρχιτεκτονικής στο δερμάτιο των αρθροπόδων (Raabe, Romano et al. 2006).

1.4 Χιτίνη

Η χιτίνη είναι το πιο άφθονο δομικό βιομόριο στη φύση. Συναντάται σε ένα πλήθος οργανισμών όπως τα βακτήρια, τους μύκητες, τα μαλάκια, αλλά είναι σίγουρα πιο εμφανής στο μεγαλύτερο και πιο διαφοροποιημένο φύλο του ζωικού βασιλείου, τα αρθρόποδα (Gupta, Fabritius et al. 2011).

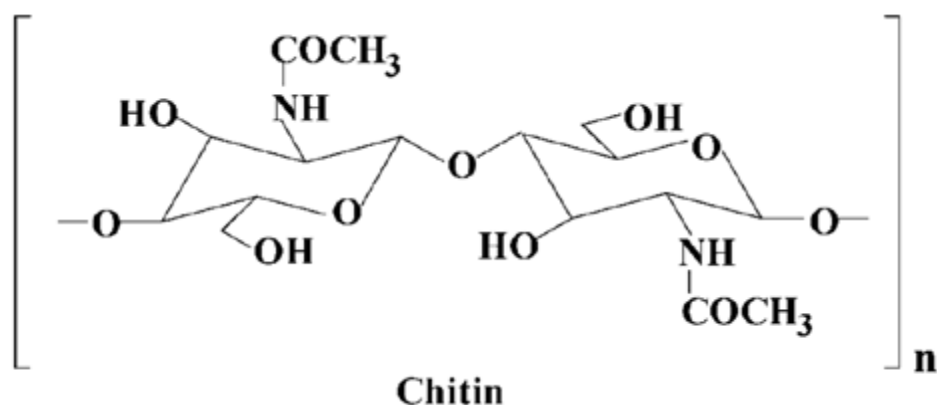
Είναι το κύριο συστατικό των κυτταρικών τοιχωμάτων των μυκήτων, του εξωσκελετού των αρθροπόδων (π.χ. τα οστρακόδερμα), των εντόμων, το εσωτερικό όστρακο των κεφαλοπόδων και των χταποδιών γι' αυτό και εντοπίζεται σε μεγάλη αφθονία. Δομικά η χιτίνη μπορεί να συγκριθεί με την κυτταρίνη ενώ λειτουργικά μπορεί να συσχετιστεί με το ρόλο των κερατινών. Η χιτίνη έχει τη δυνατότητα να σχηματίζει μικροϊνίδια με τη βοήθεια των πρωτεϊνών και άλλων πολυσακχαριτών, που μπορούν να έχουν μεγάλη δύναμη και τους επιτρέπει να τεντώνονται και να αρθρώνουν. Η κατανόηση του πώς η φύση καταφέρνει να παράγει ένα τέτοιο ελαφρύ, ανθεκτικό, και ταυτόχρονα πολυ-λειτουργικό υλικό είναι πολύ σημαντική ώστε να σχεδιαστούν νέα βιο-υλικά. Σήμερα βρίσκει ευρύτατη εφαρμογή στην ιατρική και τη βιομηχανία (Cohen 2001; Gupta, Fabritius et al. 2011).

Η σύνθεση γενικά περιλαμβάνει τα ακόλουθα βασικά στάδια: 1) σύνθεση της χιτίνης ενδοκυτταρικά ή εξωκυτταρικά 2) μεταφορά των μορίων της χιτίνης στον εξωκυττάριο χώρο 3) χημικές τροποποιήσεις σε σημεία μη κρυσταλλικής χιτίνης σε συνδυασμό με άλλα μόρια 4) κρυστάλλωση της μη τροποποιημένης χιτίνης που καλύπτεται από τα υπόλοιπα συστατικά. Η υπερδομή που δημιουργείται έχει ελαστικές και κολλώδεις μηχανικές ιδιότητες 5) ωρίμανση του συστατικού μέσω σχηματισμού δευτεροταγών ομοιοπολικών δεσμών των συστατικών του και εναπόθεση σε διαφορετικά υποστρώματα (Cohen 2001).

Η χιτίνη συναντάται στη φύση σε τρεις (3) διαφορετικές πολυμορφικές κρυσταλλικές δομές, που θα αναλυθούν στις παρακάτω ενότητες. Πρόκειται για την α-χιτίνη (α-chitin), την β-χιτίνη (β-chitin) και την γ-χιτίνη (γ-chitin) (Rudall, J.W.L. Beament et al. 1963). Οι δομές της α-χιτίνης και της β-χιτίνης έχουν μελετηθεί με περίθλαση ακτίνων X (Gardner and Blackwell 1975; Minke and Blackwell 1978; Blackwell and Weih 1980).

1.4.1 Χημική δομή χιτίνης

Η χιτίνη είναι ένας γραμμικός πολυσακχαρίτης. Συγκεκριμένα είναι ένα πολυμερές της N-ακέτυλο-γλυκοζαμίνης (NAG), παράγωγο της γλυκόζης που εντοπίζεται στη φύση. Σε πειράματα υδρόλυσης της χιτίνης με ένζυμα, που ονομάζονται χιτινάσες, προέκυψε γλυκοζαμίνη και N-ακέτυλογλυκοζαμίνη ενώ βρέθηκε ότι η βασική δομή της χιτίνης απαρτίζεται από εναλλασσόμενα μόρια N-ακέτυλογλυκοζαμίνης που συνδέονται με β-1,4-γλυκοζιτικούς δεσμούς και δεσμούς υδρογόνου (Rudall, J.W.L. Beament et al. 1963). Η χημική δομή της χιτίνης παρουσιάζεται στην εικόνα 6.

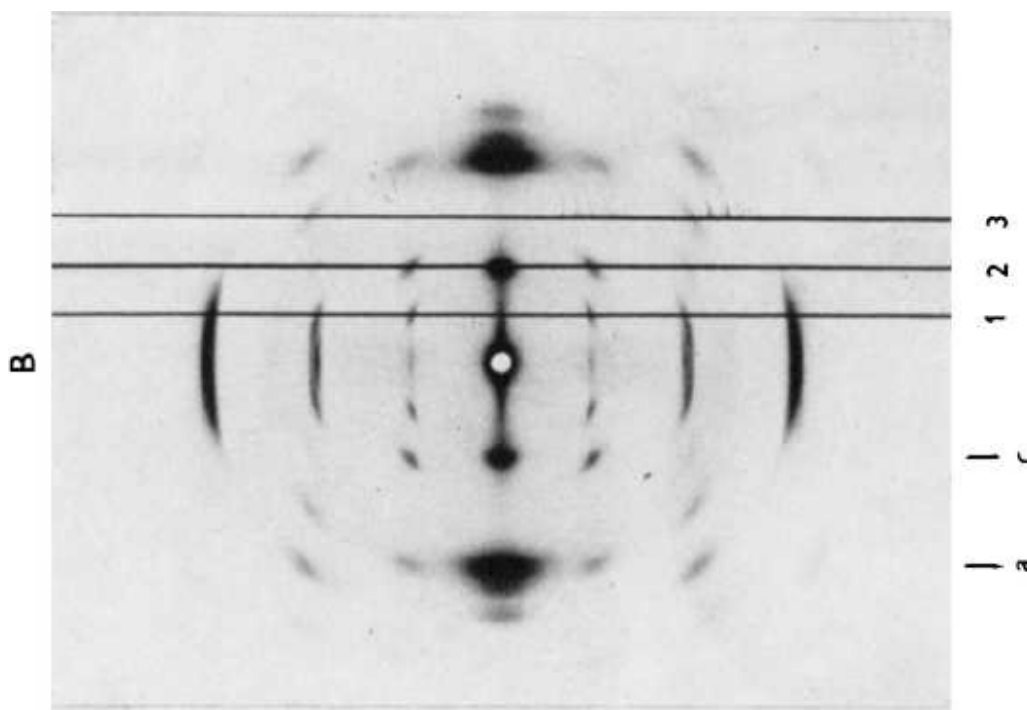


Εικόνα 6: Η χημική δομή της χιτίνης ($C_8H_{13}NO_5$)_n.

Ακολουθώντας λοιπόν τις κρυσταλλογραφικά λυμένες δομές, η χιτίνη δημιουργεί αλυσίδες πολυσακχαριτών που διευθετούνται η μία μετά την άλλη με δημιουργία δεσμών υδρογόνου. Ανάλογα λοιπόν με το δίκτυο των δεσμών υδρογόνου που δομείται προκύπτουν 3 στερεοϊσομερείς δομές: οι α, β και γ, με την α να εμφανίζει την πιο σταθερή δομή. Τα μικροϊνίδια διατάσσονται παράλληλα σε επίπεδα (10μm). Στο επόμενο επίπεδο, τα επίπεδα διατάσσονται το ένα κάτω από το άλλο με σταθερή γωνία στροφής, δημιουργώντας διευθετημένα στρώματα (1mm). Είναι αξιοσημείωτο ότι τα μικροϊνίδια περιλαμβάνουν νανοϊνίδια των 300nm σε μήκος και 10nm σε πλάτος και χαρακτηρίζονται από υψηλή δομική και κρυσταλλική ακρίβεια (Morganti and Morganti 2008).

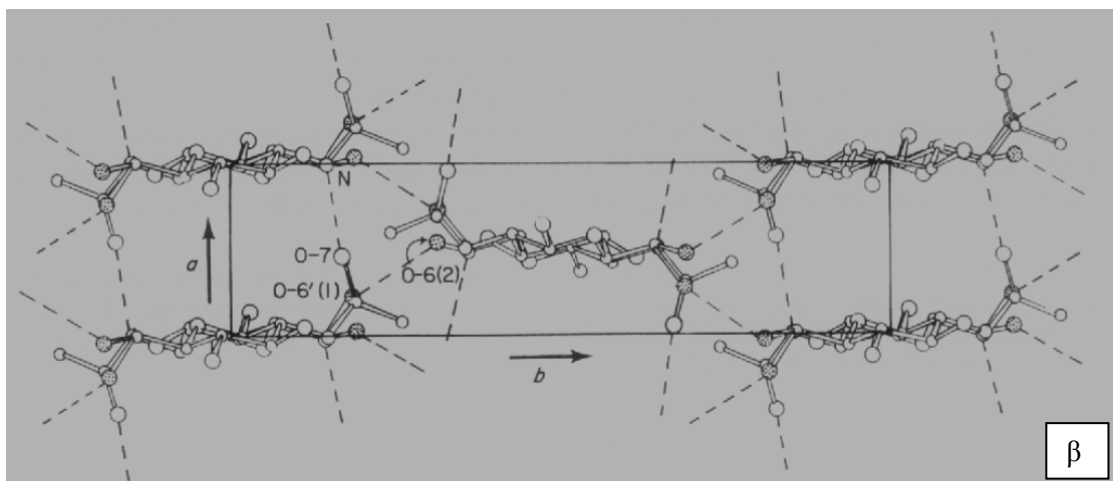
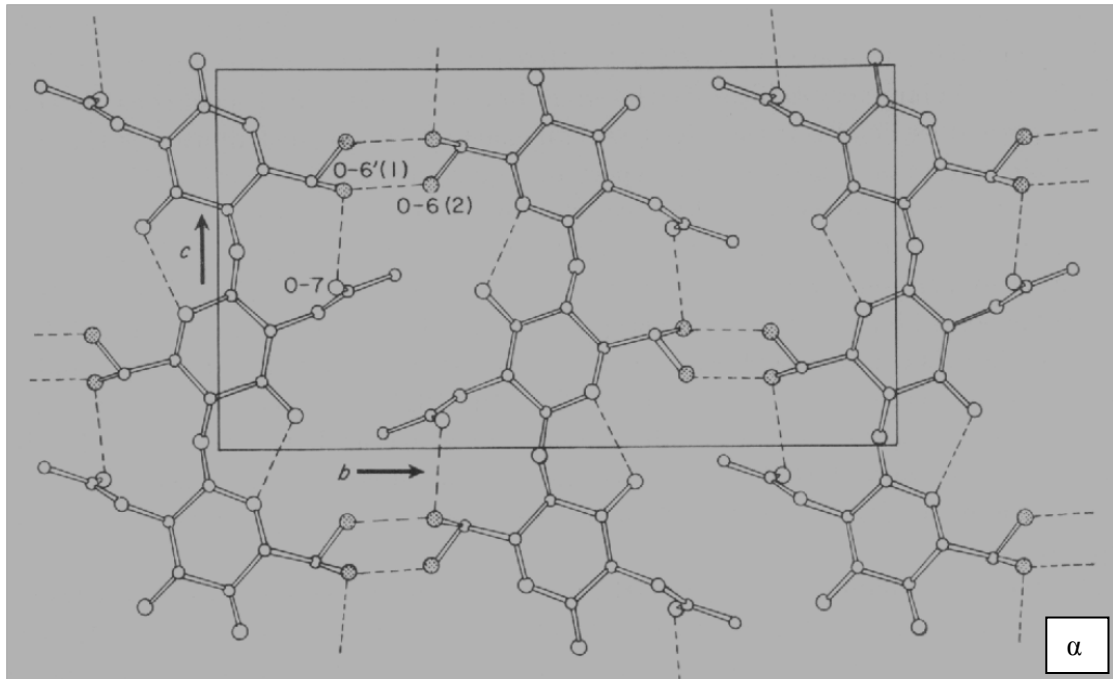
1.4.2 α-χιτίνη

Οι αλυσίδες της α-χιτίνης έχουν βρεθεί στο δερμάτιο των αρθροπόδων, στο κυτταρικό τοίχωμα των μυκήτων και στα κελύφη των μαλάκιων. Είναι η πιο διαδεδομένη και πιο σταθερή δομή χιτίνης σε σχέση με τις β- και γ- μορφές (Gupta, Fabritius et al. 2011). Η δομή της α-χιτίνης έχει προσδιοριστεί με περίθλαση ακτίνων Χ (εικόνα 7) (Minke and Blackwell 1978).



Εικόνα 7: Το περιθλασίγραμμα ακτίνων Χ που μέσω αυτού προσδιορίστηκε η δομή της α-χιτίνης (Minke and Blackwell 1978).

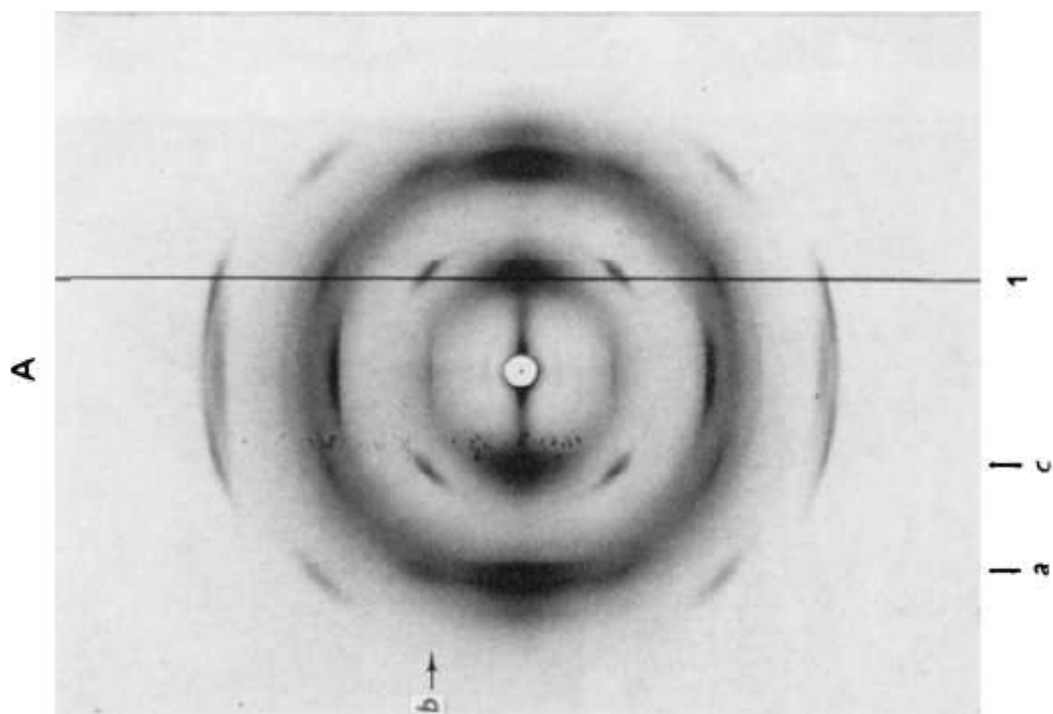
Από την ανάλυση περιθλασιγραμμάτων της α-χιτίνης προκύπτει ότι οι αλυσίδες έχουν αντίθετη κατεύθυνση και το σύστημα περιγράφεται ως αντιπαράλληλο και οι αλυσίδες up και down ($\uparrow\downarrow$). Στην α-χιτίνη, οι αλυσίδες οργανώνονται σε επίπεδα τα οποία συγκρατούνται σταθερά από ένα ισχυρό δίκτυο υδρογονικών δεσμών, τόσο εσωτερικά στο κάθε επίπεδο όσο και ανάμεσα σε διαφορετικά επίπεδα (εικόνα 8) (Minke and Blackwell 1978).



Εικόνα 8: Η δομή της α-χιτίνης από δυο διαφορετικές προβολές (α,β). Φαίνονται οι αντιπαράλληλες αλυσίδες που συγκρατούνται από ένα ισχυρό δίκτυο υδρογονικών δεσμών (Minke and Blackwell 1978).

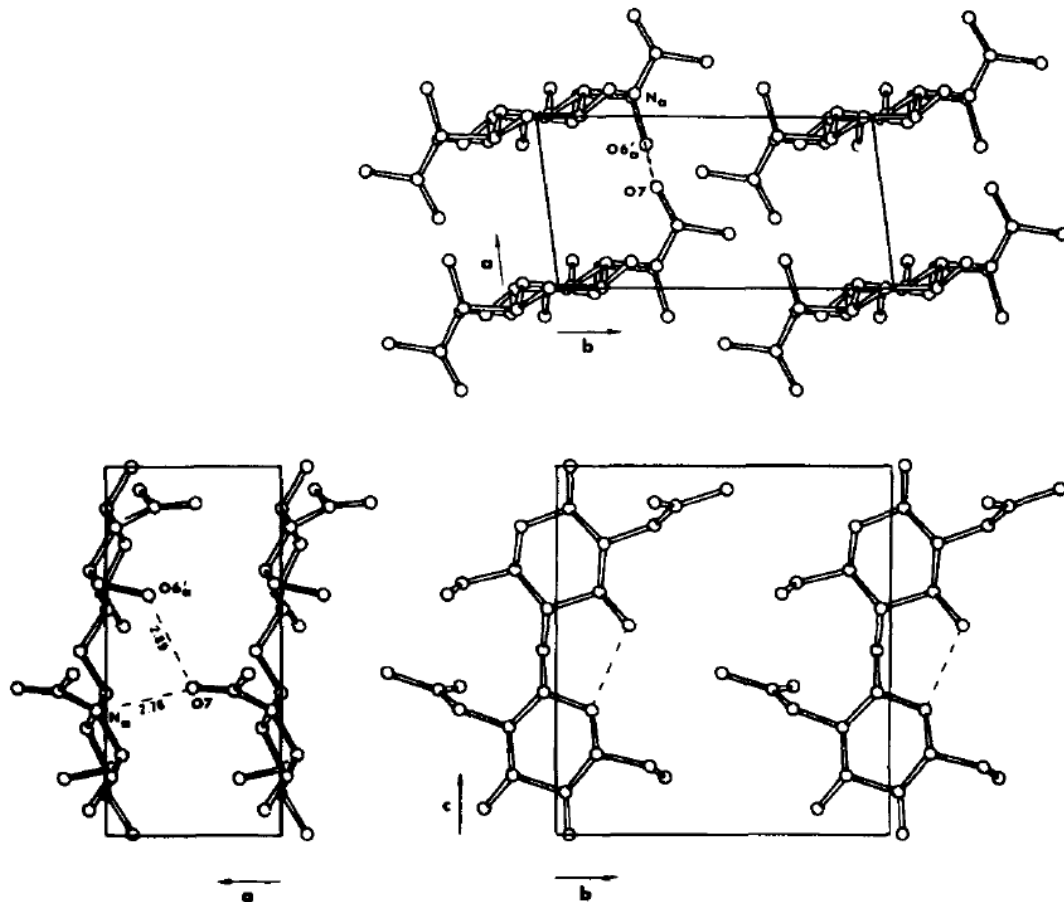
1.4.3 β-χιτίνη

Οι αλυσίδες της β-χιτίνης έχουν εντοπιστεί κυρίως στα καλαμάρια (Gupta, Fabritius et al. 2011). Βρίσκεται στα εξωκυτταρικά νημάτια των διατόμων, στα πωγωνοφόρα και τους σωλήνες των δακτυλιοσκωλήκων (Sugiyama, Boisset et al. 1999). Η δομή της β-χιτίνης έχει προσδιοριστεί με περίθλαση ακτίνων X (εικόνα 9) (Gardner and Blackwell 1975).



Εικόνα 9: Περίθλασίγραμμα ακτίνων X που προσδιορίζει την δομή της β-χιτίνης (Sawada, Nishiyama et al. 2012).

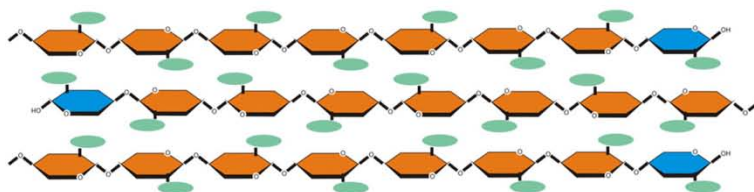
Η δομή της β-χιτίνης διαφέρει από αυτή της α-χιτίνης. Στη β-χιτίνη οι αλυσίδες διευθετούνται παράλληλα μεταξύ τους, δηλαδή είναι της μορφής . Η δομή σταθεροποιείται σε όλες τις κατευθύνσεις με αλληλεπιδράσεις Van der Waals και υδροφοβικές αλληλεπιδράσεις αλλά δεσμοί υδρογόνου δεν αναπτύσσονται κατά τον b άξονα (εικόνες 10 και 11) (Gardner and Blackwell 1975).



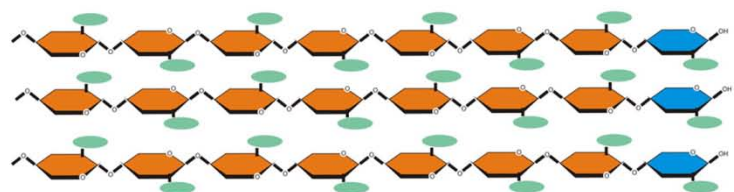
Εικόνα 10: Το δομικό μοντέλο της β-χιτίνης σε τρεις διαφορετικές προβολές. Φαίνεται η παράλληλη διεύθυνση των αλυσίδων και οι εσωτερικοί υδρογονικοί δεσμοί (Gardner and Blackwell 1975).

Chitin conformations

α-Chitin (antiparallel chains)



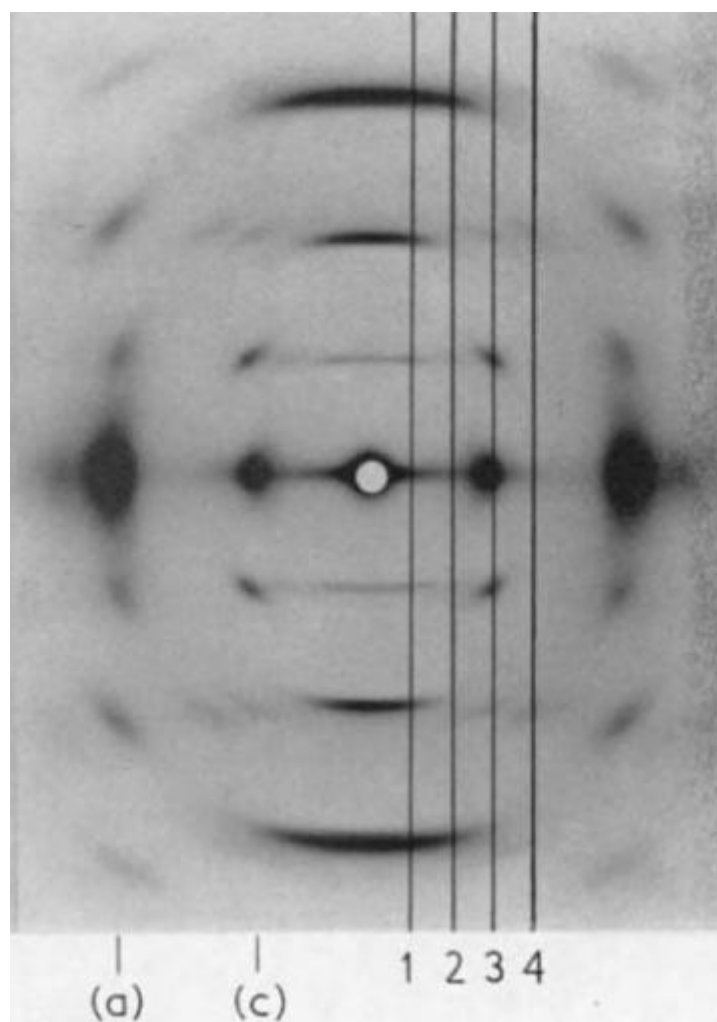
β-Chitin (parallel chains)



Εικόνα 11: Οι διαφορές στη δομή της α-χιτίνης με την δομή της β-χιτίνης. Αριστερά είναι η διάταξη των αλυσίδων στην α-χιτίνη όπου είναι αντιπαράλληλες και δεξιά είναι η διάταξη των αλυσίδων στη β-χιτίνη όπου είναι παράλληλες (Seidl 2008).

1.4.4 γ-χιτίνη

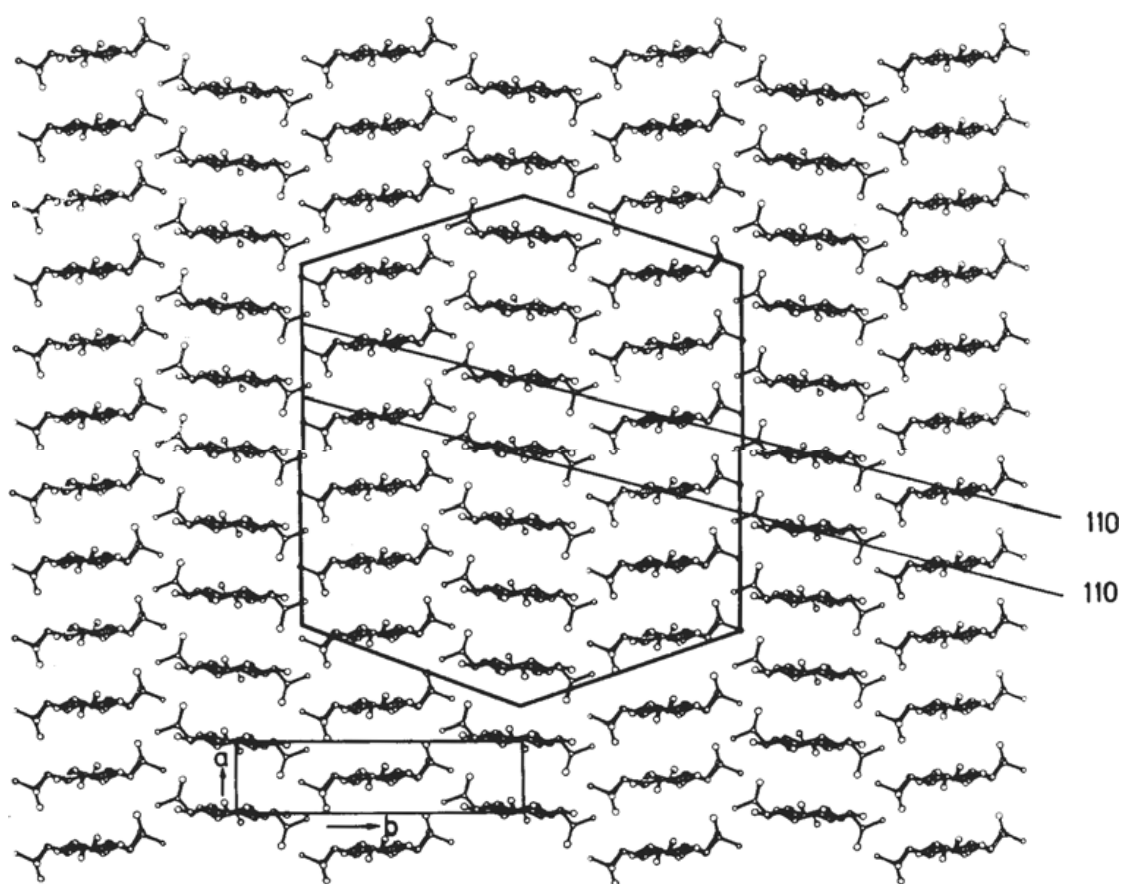
Οι αλυσίδες της γ-χιτίνης έχουν βρεθεί στις ίνες των κουκουλιών στα σκαθάρια *Ptinus* και στο στομάχι του καλαμαριού *Loligo* (Jang, Kong et al. 2004). Η γ-χιτίνη είναι λιγότερο μελετημένη από τις α-χιτίνη και β-χιτίνη, ωστόσο η δομή της έχει συζητηθεί και μελετηθεί (εικόνα 12) (Rudall, J.W.L. Beament et al. 1963; Atkins 1985). Στη δομή της, δυο αλυσίδες διατάσσονται παράλληλα και στη συνέχεια μια τρίτη διατάσσεται αντιπαράλληλα, δηλαδή είναι της μορφής $\uparrow\downarrow$ (Jang, Kong et al. 2004).



Εικόνα 12: Το περιθλασίγραμμα ακτίνων X της γ-χιτίνης (Rudall, J.W.L. Beament et al. 1963).

1.4.5 Μοντέλο Atkins

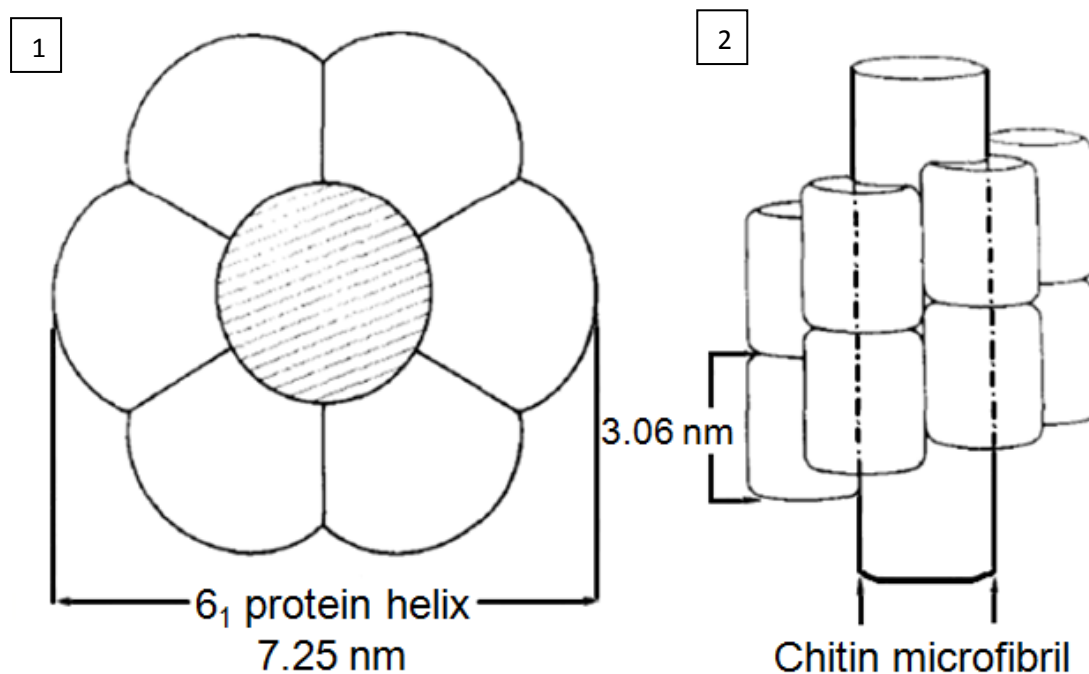
Λαμβάνοντας υπόψη το συνολικό μέγεθος των κρυστάλλων, τόσο από ηλεκτρονική μικροσκοπία όσο και από μετρήσεις με άλλες μεθόδους (measurements from line broadening measurements) αποδείχτηκε ότι η χιτίνη οργανώνεται σε ινίδια (nanofibres) χιτίνης, όπου το κάθε ινίδιο χιτίνης αποτελείται από 19 αλυσίδες του πολυσακχαρίτη. Το κάθε ινίδιο χιτίνης έχει διάμετρο περίπου 3 nm και μήκος περίπου 300 nm (εικόνα 13) (Atkins 1985).



Εικόνα 13: Ύστερα από μελέτες περιθλασιγραμμάτων αποδείχτηκε ότι κάθε ινίδιο χιτίνης που αποτελείται από 19 αλυσίδες χιτίνης. Τα επίπεδα 110 αντιστοιχούν σε περιοχές από το περιθλασίγραμμα που αποδεικνύουν την ύπαρξη αυτής της κρυσταλλικής δομής (Atkins 1985).

1.5 Σύμπλοκο χιτίνης-πρωτεϊνών (μοντέλο Blackwell)

Υστερα από πειράματα περίθλασης ακτίνων-X στη σφήκα *Megarhyssa* προτάθηκε ότι δημιουργείται ένα σύμπλοκο, όπου οι πρωτεΐνες περιβάλλουν τα ινίδια χιτίνης (microfibrils). Δημιουργείται στο σύμπλοκο ένας «πυρήνας» χιτίνης που περιβάλλεται από πρωτεΐνες σε εξαγωνική διευθέτηση, σε μορφή έλικας, διαμέτρου 7.25 nm (εικόνα 14) (Blackwell and Weih 1980).



Εικόνα 14: Το σύμπλοκο χιτίνης-πρωτεϊνών σε δυο διαφορετικές προβολές. Η πρώτη (1) είναι κάθετα στον άξονα του ινιδίου χιτίνης, ενώ η δεύτερη (2) είναι παράλληλα στον άξονα του. Φαίνεται ο τρόπος με τον οποίο οι πρωτεΐνες περιβάλλουν τα ινίδια χιτίνης, δημιουργώντας μια 6_1 έλικα (Blackwell and Weih 1980).

1.6 Πρωτεΐνες του δερματίου

Για χρόνια ένα πρόβλημα που απασχολούσε τους ερευνητές ήταν σε ποια περιοχή γίνεται η σύνθεση των πρωτεϊνών του δερματίου. Υπάρχουν αναφορές στην διεθνή βιβλιογραφία ότι πρωτεΐνες που βρέθηκαν στην αιμόλεμφο ήταν παρούσες και στο δερμάτιο καθώς και ακόμη ότι σημασμένες πρωτεΐνες που είχαν εγχυθεί στην αιμόλεμφο εμφανιζόταν στο δερμάτιο. Σήμερα είναι πλέον γνωστό ότι η επιδερμίδα (epidermis) μπορεί να συνθέσει τόσο τις πρωτεΐνες του δερματίου όσο και αυτές της αιμόλεμφο. Μπορεί επίσης να μεταφέρει πρωτεΐνες που δημιουργούνται σε κύτταρα εκτός της επιδερμίδας, από την αιμόλεμφο στο δερμάτιο (Willis, Papandreou et al. 2012).

Οι πρωτεΐνες του δερματίου (cuticular proteins – CP) χωρίζονται σε δύο (2) μεγάλες οικογένειες: τις μη δομικές πρωτεΐνες (non-structural proteins) και τις δομικές πρωτεΐνες (structural proteins) (Willis, Papandreou et al. 2012).

1.6.1 Μη δομικές πρωτεΐνες

Οι μη δομικές πρωτεΐνες του δερματίου χωρίζονται σε τέσσερις (4) κατηγορίες: τις χρωστικές, διάφορα ένζυμα, τις αμυντικές και τις αρυλφορίνες (arylphorins) (Willis, Papandreou et al. 2012). Στον πίνακα 1 παρουσιάζονται κάποιες αντιπροσωπευτικές μη δομικές πρωτεΐνες που έχουν εντοπιστεί στο δερμάτιο και πληροφορίες για αυτές.

Οργανισμός	Όνομα πρωτεΐνης	Αρ. αμινοξέων	Λειτουργία
<i>Schistocerca gregaria</i>	Putative carotene binding protein	250	Μεταφέρει καροτίνη στο δερμάτιο
<i>Caliphora vicina</i>	Arylphorin A4	743	Έχει βρεθεί στο δερμάτιο
<i>Caliphora vicina</i>	Arylphorin C223	743	Έχει βρεθεί στο δερμάτιο
<i>Drosophila melanogaster</i>	Yellow	520	Εναποθέτει μελανίνη στο δερμάτιο
<i>Bombyx mori</i>	Cecropin A	41	Αμυντική πρωτεΐνη
<i>Bombyx mori</i>	Cecropin B	41	Αμυντική πρωτεΐνη
<i>Bombyx mori</i>	Prophenoloxidase	675	Ένζυμο στη μελανογένεση
<i>Calpodes ethilus</i>	CECP 22	169	«Πέψη» παλιού δερματίου
<i>Manduca sexta</i>	Arylphorin α	684	Έχει βρεθεί στο δερμάτιο
<i>Manduca sexta</i>	Arylphorin β	687	Έχει βρεθεί στο δερμάτιο
<i>Manduca sexta</i>	Insecticyanin A	189	Μπλε χρωστική
<i>Manduca sexta</i>	Insecticyanin B	189	Μπλε χρωστική
<i>Manduca sexta</i>	Scolexin A	279	Πρωτεάση σερίνης
<i>Manduca sexta</i>	Scolexin B	279	Πρωτεάση σερίνης

Πίνακας 1: Στην πρώτη στήλη παρουσιάζεται ο οργανισμός που προέρχεται η πρωτεΐνη, στην δεύτερη στήλη το όνομα της πρωτεΐνης, στην τρίτη στήλη ο αριθμός των αμινοξικών καταλοίπων της “ώριμης” πρωτεΐνης και στην τέταρτη στήλη η λειτουργία της πρωτεΐνης ή κάποιο ιδιαίτερο χαρακτηριστικό της (Willis, Papandreou et al. 2012).

1.6.1.1 Χρωστικές

Τρεις (3) διαφορετικές χρωστικές έχουν εντοπιστεί στο δερμάτιο. Πρόκειται για τις insecticyanins και για δυο (2) διαφορετικές yellow πρωτεΐνες (Willis, Papandreou et al. 2012).

Οι insecticyanins είναι μπλε χρωστικές που δημιουργούνται στην επιδερμίδα (epidermis) και εκκρίνονται στην αιμόλεμφο (hemolymph) και στο δερμάτιο (cuticle). Ανήκουν στην οικογένεια των λιποκαλινών και εμφανίζονται σαν τετραμερές μαζί με το γ ισομερές της biliverdin IX, η οποία βρίσκεται μέσα στην υδρόφοβη εσοχή (hydrophobic pocket) του βαρελιού. Στο δερμάτιο, σε συνεργασία με τις καροτίνες (carotenes), δίνουν πράσινο χρωματισμό. Η δομή τους (από τον οργανισμό *Manduca sexta*) έχει προσδιοριστεί κρυσταλλογραφικά με ακτίνες X, σε διακριτικότητα 2.6 Å, έχοντας καταχωρηθεί στην βάση δεδομένων Protein Data Bank (Berman, Westbrook et al. 2000) με τον κωδικό 1Z24 (εικόνα 15) (Holden, Rypniewski et al. 1987). Θεωρούνται οι καλύτερα δομικά χαρακτηρισμένες πρωτεΐνες του δερματίου (Willis, Papandreou et al. 2012)



Εικόνα 15: Η προσδιορισμένη κρυσταλλογραφικά δομή της χρωστικής insecticyanin (με πράσινο χρώμα) από τον οργανισμό *Manduca sexta* (PDB ID: 1Z24) (με μήκος 189 αμινοξικά κατάλοιπα), σε διακριτικότητα 2.6 Å. Με μπλε χρώμα είναι το γ ισομερές της biliverdin IX, η οποία βρίσκεται μέσα στην υδρόφοβη εσοχή (hydrophobic pocket) του βαρελιού (Holden, Rypniewski et al. 1987).

Η πρωτεΐνη yellow έχει εντοπιστεί στην *Drosophila melanogaster*. Αποτελείται από 520 αμινοξικά κατάλοιπα και κωδικοποιείται από το γονίδιο y (CG3757) (Kornezos and Chia 1992). Είναι δομικό συστατικό του δερματίου που αλληλεπιδρά με προϊόντα του γονιδίου ebony και βοηθάει στην εναπόθεση μελανίνης. Η αναλυτική εξέταση της έκφραση του γονιδίου y αποκάλυψε ότι και τα cis και τα trans ρυθμιστικά στοιχεία είναι υπεύθυνα για τις διαφοροποιήσεις στα χρώματα του δερματίου ανάμεσα σε διαφορετικά είδη (Wittkopp, Vaccaro et al. 2002; Wittkopp, Stewart et al. 2009). Δεν υπάρχει κάποιο δεδομένο για κάποιο γνωστό δομικό αυτοτελές στοιχείο που να αλληλεπιδρά με χιτίνη (chitin binding domain) στη πρωτεΐνη yellow. Το μόνο δομικό αυτοτελές στοιχείο (domain) που έχει αναγνωριστεί είναι καταχωρημένο στην βάση δεδομένων Pfam (Punta, Coggill et al. 2012), με το όνομα pfam03022 (Han, Fang et al. 2002; Willis, Papandreou et al. 2012).

Μια ακόμα μη δομική πρωτεΐνη εμπλέκεται στον χρωματισμό του δερματίου. Πρόκειται για την πρωτεΐνη yellow, με κωδικό στην βάση δεδομένων UniProt (Arweiler, Bairoch et al. 2004) P82886.1, η οποία είναι πρωτεΐνη που αλληλεπιδρά με το β-καροτένιο (beta-carotene binding). Αποτελείται από 250 αμινοξικά κατάλοιπα. Έχει απομονωθεί από το δερμάτιο της ενήλικης *Schistocerca gregaria* (ακρίδα της ερήμου). Πρόκειται για κίτρινη χρωστική. Τέλος, έχει προταθεί ότι συμμετέχει στην μεταφορά της καροτίνης στην επιδερμίδα (epidermis) και μετά στο δερμάτιο (Wybrandt and Andersen 2001; Willis, Papandreou et al. 2012).

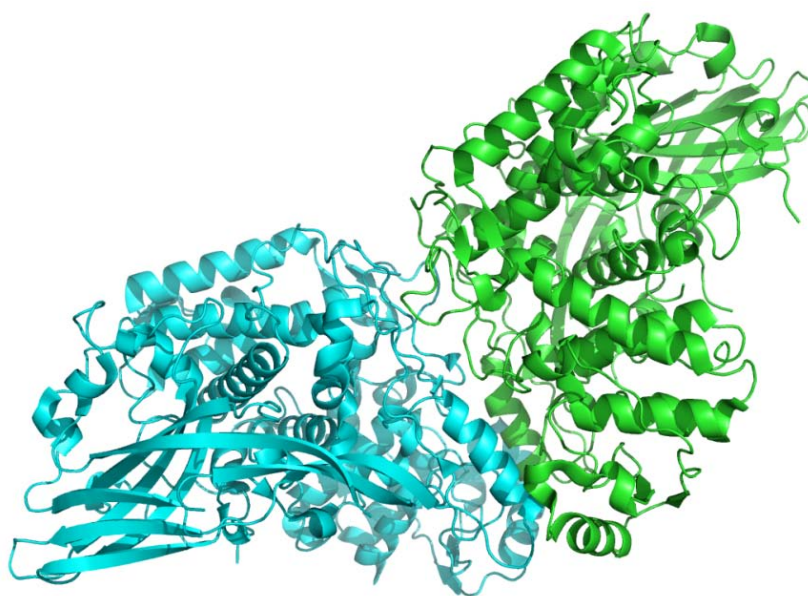
1.6.1.2 Ένζυμα

Στο δερμάτιο έχουν βρεθεί διάφορα ένζυμα τα οποία συμμετέχουν στην διαδικασία της σκλήρυνσης (sclerotization) (Andersen 2010). Επίσης εμφανίζονται ένζυμα που σχετίζονται με την έκδυση (molting fluid), όπως η πρωτεΐνη CEPP22 που παρατηρήθηκε στο γένος *Calpodes*, που πιθανότατα λειτουργεί σαν αμιδάση (Willis, Papandreou et al. 2012). Περαιτέρω ανάλυση έδειξε ότι τέτοιες πρωτεΐνες υπάρχουν στο δερμάτιο και πριν την διαδικασία της έκδυσης (Marcu and Locke 1998). Τέλος, ένζυμα που συμμετέχουν στην πέψη του παλιού δερματίου εντοπίζονται προσωρινά στο δερμάτιο. Σε αυτά περιλαμβάνονται πρωτεάσες και χιτινάσες (Marcu and Locke 1998).

1.6.1.3 Αμυντικές πρωτεΐνες

Στο δερμάτιο βρίσκονται και συστατικά του αμυντικού συστήματος του εντόμου. Στο δερμάτιο του *Bombyx mori* έχει βρεθεί το αντιβακτηριακό πεπτίδιο cecropin (Lee and Brey 1994). Επίσης, στο δερμάτιο του *Bombyx mori* βρέθηκαν μια προφαινολοξειδάση και μια πρωτεάση σερίνης. Η πρώτη συμμετέχει στην μελανογένεση που πραγματοποιείται ως αποτέλεσμα ενός τραυματισμού, ενώ η δεύτερη ενεργοποιεί την πρώτη (Ashida and Brey 1995).

Η δομή της προφαινολοξειδάσης από τον οργανισμό *Manduca sexta* (PDB ID: 3HHS) έχει προσδιοριστεί σε ατομική διακριτικότητα (εικόνα 16) (Li, Wang et al. 2009). Η κρυσταλλογραφικά προσδιορισμένη δομή της προφαινολοξειδάσης είναι ένα διμερές. Το κάθε μονομερές αποτελείται από δύο (2) αυτοτελείς δομικές περιοχές (domains), οι οποίες κατατάσσονται κατά CATH (Sillitoe, Cuff et al. 2013) στις «κυρίως άλφα» (Mainly Alpha) και «κυρίως βήτα» (Mainly Beta) αντίστοιχα.



Εικόνα 16: Η ασύμμετρη μονάδα της προσδιορισμένης κρυσταλλογραφικά δομής της προφαινολοξειδάσης, ως διμερές, από τον οργανισμό *Manduca sexta* (PDB ID: 3HHS) (Li, Wang et al. 2009).

Ακόμα, στο δερμάτιο της *Manduca sexta* παρατηρήθηκε η παρουσία της αμυντικής πρωτεΐνης scolexin, σε δυο διαφορετικές μορφές, χωρίς να γίνει εφικτή η διάκριση μεταξύ τους (Molnar, Borhegyi et al. 2001).

Τέλος, το δερμάτιο εμφανίζεται να αποτελεί αποθηκευτικό χώρο για ένα πεπτίδιο (HCP; GI:240104242; 2RPS_A) το οποίο τονώνει τη συγκέντρωση και τη διακίνηση των αιμοκυττάρων στο σκώρο *Pseudaletia separate* (*Mythimna separate*) (Nakatogawa, Oda et al. 2009).

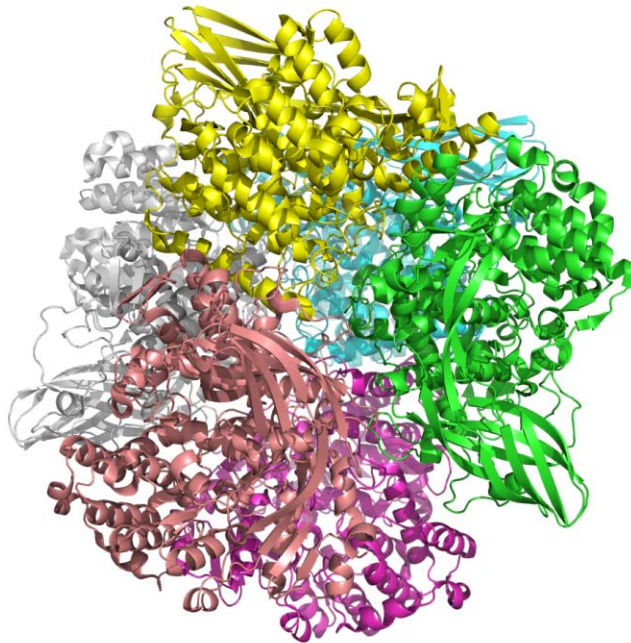
1.6.1.4 Αρυλφορίνες (Arylphorins)

Η τελευταία κατηγορία μη δομικών πρωτεϊνών του δερματίου είναι οι αρυλφορίνες (arylphorins). Οι αρυλφορίνες περιέχουν μεγάλη ποσότητα αρωματικών αμινοξικών καταλοίπων και μερικά λιπίδια (Willis, Papandreou et al. 2012).

Αρχικά θεωρήθηκε ότι συνθέτονται μόνο στην αιμόλεμφο και από εκεί εν συνεχεία καταλήγουν στο δερμάτιο (Scheller 1980). Ωστόσο, υπάρχουν δεδομένα ότι και η επιδερμίδα (epidermis) είναι ικανή να συνθέσει αρυλφορίνες (Riddiford and Hice 1985). Αυτές οι μελέτες κατέδειξαν επίσης ότι μια δεδομένη πρωτεΐνη μπορεί να συντεθεί σε πολλά διαφορετικά κύτταρα. Ακόμα όμως δεν έχει προσδιοριστεί αν είναι το ίδιο γονίδιο υπεύθυνο για τη σύνθεση σε όλα τα κύτταρα (Willis, Papandreou et al. 2012).

Ο ρόλος των αρυλφορίνων παραμένει άγνωστος. Είναι γενικά παραδεκτό ότι συμμετέχουν στη σκλήρυνση (sclerotization) του δερματίου, λόγω της υψηλής περιεκτικότητας τους σε τυροσίνη (T) (Willis, Papandreou et al. 2012). Σύμφωνα με την UniProt (Arweiler, Bairoch et al. 2004), οι αρυλφορίνες είναι αποθηκευτικές πρωτεΐνες, στις προνύμφες (larval), που χρησιμοποιούνται σαν πηγή αρωματικών αμινοξικών καταλοίπων για την πρωτεϊνική σύνθεση, κατά τη διάρκεια της μεταμόρφωσης του αρθροπόδου (Willott, Wang et al. 1989). Επίσης, η μία από τις έξι υπομονάδες των αρυλφορίνων, όταν δρα σαν μονομερές, έχει δειχθεί *in vitro* και *in vivo*, ότι διεγείρει τον κυτταρικό πολλαπλασιασμό (Hakim, Blackburn et al. 2007).

Η δομή της αρυλφορίνης από τον οργανισμό *Antheraea pernyi* (PDB ID: 3GWJ) έχει προσδιοριστεί σε ατομική διακριτικότητα (εικόνα 17) (Ryu, Lee et al. 2009). Η κρυσταλλογραφικά προσδιορισμένη δομή της αρυλφορίνης είναι ένα εξαμερές όμοιων μεταξύ τους πολυπεπτιδικών αλυσίδων.



Εικόνα 17: Η ασύμμετρη μονάδα της προσδιορισμένης κρυσταλλογραφικά δομής της αρυλφορίνης, ως εξαμερές, από τον οργανισμό *Antheraea pernyi* (PDB ID: 3GWJ) (Ryu, Lee et al. 2009).

1.6.2 Δομικές πρωτεΐνες

1.6.2.1 Γενική επισκόπηση

Σήμερα, ο αριθμός των δομικών πρωτεϊνών του δερματίου που έχουν βρεθεί (με διάφορες μεθόδους) είναι μερικές χιλιάδες. Η ανάλυση των δομικών πρωτεϊνών, οδήγησε στην ταυτοποίηση των ειδικών, κοινών, συντηρημένων χαρακτηριστικών και μοτίβων τους. Έχουν ταξινομηθεί σε δεκατρείς (13) διακριτές οικογένειες, σύμφωνα με τις αμινοξικές τους αλληλουχίες (Willis 2010; Willis, Papandreou et al. 2012). Στον πίνακα 2 παρουσιάζεται ο αριθμός των πρωτεϊνών του δερματίου που έχουν βρεθεί σε έξι (6) καλά σχολιασμένους οργανισμούς, κατηγοριοποιημένες ανά οικογένειες και οργανισμό.

Οι περισσότερες δομικές πρωτεΐνες του δερματίου είναι σχετικά μικρές, περίπου 200 αμινοξικά κατάλοιπα. Ωστόσο, αυτό το μέγεθος δεν είναι χαρακτηριστικό όλων των πρωτεϊνών του δερματίου καθώς έχουν παρατηρηθεί και ακολουθίες διαφορετικού μήκους. Η μεγαλύτερη που έχει βρεθεί αποτελείται από 22.971 αμινοξικά κατάλοιπα, ονομάζεται dumpy, (GenBank (Benson, Cavanaugh et al. 2013): CG33196), προέρχεται από την *Drosophila melanogaster* και ευθύνεται για λειτουργίες σχετικά με τους μυς στο δερμάτιό της (Willis, Papandreou et al. 2012).

Είναι πολύ ενδιαφέρον ότι οι πρωτεΐνες των περισσότερων οικογενειών του δερματίου βρίσκονται μόνο στα αρθρόποδα, με μερικές να είναι μόνο σε μια ή δυο τάξεις εντόμων και κάποιες άλλες να είναι πιο διαδεδομένες (πίνακας 3) (Willis, Papandreou et al. 2012).

Οι πιο πολλές πρωτεΐνες του δερματίου έχουν καταχωρηθεί στην βάση δεδομένων cuticleDB, η οποία παρέχει και χρήσιμα εργαλεία για μια ποικιλία μελετών (Magkrioti, Spyropoulos et al. 2004).

Ένας χρήσιμος τρόπος για να απεικονιστούν τα διαγνωστικά χαρακτηριστικά και τα μοτίβα των οικογενειών των πρωτεϊνών των αρθροπόδων είναι με την μέθοδο των WebLogos (Schneider and Stephens 1990; Crooks, Hon et al. 2004). Στη συνέχεια θα εμφανιστούν απεικονίσεις για τις οικογένειες με αυτή τη μέθοδο.

	CPR	CPF+ CPFL	TWDL	CPLCG	CPLCW	CPLCA	CPLCP	CPG	Apidermin	CPAP1	CPAP3	CPCFC	Άλλες	Σύνολο
<i>An. gambiae</i>	156	11	12	27	9	3	4 +23?	0	0	0	7	1	10	240+
<i>B. mori</i>	148	5	4	0	0	0	7	18	0	0	1	1	33	217
<i>D. melanogaster</i>	101	3	27	3	0	11	5	0	0	2	6	1	?	159
<i>A. mellifera</i>	32	3	2	0	0	0	2	0	3	0	5	0	?	47
<i>N. vitripennis</i>	62	4	2	0	0	0	3	0	3	0	6	0	?	80
<i>T. castaneum</i>	101	8	3	2	0	0	4	0	0	10	7	2	?	137
Σύνολο	600	34	50	32	9	14	25	18	6	12	32	5	43+?	880+

Πίνακας 2: Ο αριθμός των γονιδίων, για διαφορετικές οικογένειες πρωτεϊνών του δερματίου, σε οργανισμούς με σχολιασμό των πρωτεϊνών του δερματίου, σε ολόκληρο το γονιδίωμα. Αριστερά εμφανίζεται το όνομα του οργανισμού και στη συνέχεια τα γονίδια που έχουν βρεθεί για κάθε οικογένεια πρωτεϊνών του δερματίου. Η στήλη «άλλες» αφορά πρωτεΐνες με απουσία κάποιου χαρακτηριστικού που να τις κατατάσσει σε κάποια οικογένεια. Το «Σύνολο», αφορά το σύνολο των σχολιασμένων πρωτεϊνών, που έχουν βρεθεί, για κάθε αρθρόποδο (Willis, Papandreou et al. 2012).

	Diptera		Lepidoptera	Coleoptera	Hymenoptera	Hemiptera	Orthoptera	Dictyoptera	Phthiraptera	Collembola
	Brachycera	Nematocera								
CPR	+	+	+	+	+	+	+	+	+	+
CPE/CPFL	+	+	+	+	+	+	+	+	+	ΔΔ
TWDL	+	+	+	+	+	+	+	+	+	ΔΔ
CPLCA	+	+	Όχι	Όχι	Όχι	Όχι	Όχι	Όχι	Όχι	ΔΔ
CPLCG	+	+	Όχι	+	Όχι	Όχι	Όχι	+	Όχι	ΔΔ
CPLCW	Όχι	+	Όχι	Όχι	Όχι	Όχι	Όχι	Όχι	Όχι	ΔΔ
CPLCP	+	+	+	+	+				+	
CPG			+							
Apidermin	Όχι	Όχι	Όχι	Όχι	+	Όχι	ΔΔ	ΔΔ	Όχι	ΔΔ
CPAP1	+	+	+	+	+	+	+	+	+	+
CPAP3	+	+	+	+	+	+	+	+	+	+
CPCFC	+	+	+	+	Όχι	+	+	+	+	+
18 aa motif	+	+	+	+	+	+	+	+	+	ΔΔ
CP με >3 AAP[AVL]	+	+	+	+	+	+	+		+	ΔΔ

Πίνακας 3: Η παρουσία των οικογενειών των δομικών πρωτεϊνών του δερματίου, καθώς και κάποιων χαρακτηριστικών μοτίβων (motifs), σε διαφορετικές τάξεις (orders) εντόμων (insects). Η κάθε στήλη αντιστοιχεί στην τάξη εντόμων που φέρει σαν όνομα. Η κάθε γραμμή αντιστοιχεί στην κάθε οικογένεια (ή μοτίβο στο τέλος) που φαίνεται το όνομα της στην πρώτη στήλη. Με «+» δηλώνεται η παρουσία της οικογένειας, με «Όχι» η απουσία της οικογένειας, με «ΔΔ» δηλώνεται η μη ύπαρξη ικανών δεδομένων για να καταγραφεί η απουσία και τέλος με « » (κενό) δηλώνεται ότι τα μοτίβα δεν είναι επαρκώς καθορισμένα στις γνωστές πρωτεΐνες για να γίνει ο έλεγχος (Willis, Papandreou et al. 2012).

1.6.2.2 Οικογένειες δομικών πρωτεϊνών

Οι δεκατρείς (13) δομικές οικογένειες των πρωτεϊνών του δερματίου των αρθροπόδων παρουσιάζονται αναλυτικά, η κάθε μια ξεχωριστά, στις παρακάτω ενότητες.

1.6.2.2.1 CPR

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins with R&R consensus», δηλαδή «πρωτεΐνες του δερματίου που έχουν την R&R συναινετική ακολουθία» (Willis 2010).

Η CPR οικογένεια δομικών πρωτεϊνών του δερματίου είναι η πιο διαδεδομένη οικογένεια με τα περισσότερα ταυτοποιημένα μέλη (Willis 2010; Willis, Papandreou et al. 2012). Χαρακτηρίζεται από την ύπαρξη της R&R συναινετικής ακολουθίας (R&R consensus) που αρχικά περιείχε ένα μοτίβο 28 αμινοξικών καταλοίπων (Rebers and Riddiford 1988). Το αρχικό μοτίβο αναγνωρίστηκε από τους Rebers και Riddiford το 1988 και ήταν το: G-x(8)-G-x(6)-Y-x(2)-A-x-E-x-G-F-x(7)-P-x-P (Rebers and Riddiford 1988; Willis 2010). Σήμερα, θεωρείται αποδεκτή η εκτεταμένη R&R συναινετική ακολουθία (extended R&R consensus) που αποτελείται από 63 αμινοξικά κατάλοιπα και είναι η: G - x(7) - [DEN] - G - x(6) - [FY] - x - A - [DGN] - x(2,3) - G - [FY] - x - [AP] - x(6) (Iconomidou, Willis et al. 1999; Willis 1999; Willis 2010). Στην βάση δεδομένων Pfam (Punta, Coggill et al. 2012) η συναινετική ακολουθία R&R είναι καταχωρημένη με τις ονομασίες pfam00379 και chitin_bind_4 και αποτελείται από 68 αμινοξικά κατάλοιπα (Willis 2010).

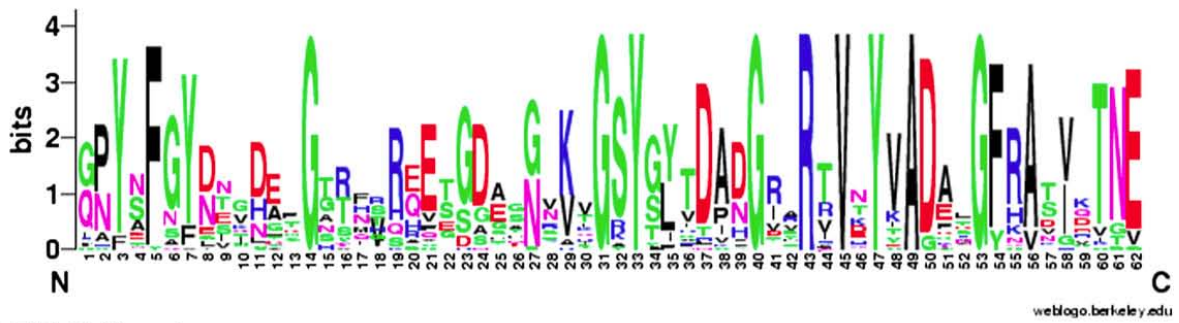
Η σημασία της R&R συναινετικής ακολουθίας φαίνεται στον ιστότοπο της βάσης δεδομένων της Pfam (Punta, Coggill et al. 2012), όπου αναφέρει την ύπαρξη 2456 μοναδικών πρωτεϊνών που περιέχουν την R&R συναινετική ακολουθία, από 67 διαφορετικά είδη αρθροπόδων (Willis, Papandreou et al. 2012). Η CPR οικογένεια έχει παρατηρηθεί μόνο στα αρθρόποδα. Γενικά, παρ' όλες τις μικρές διαφοροποιήσεις στην R&R συναινετική ακολουθία που έχουν παρατηρηθεί (Futahashi, Okamoto et al. 2008), είναι το πιο εύκολα παρατηρήσιμο μοτίβο και το πιο διαδεδομένο ανάμεσα στις πρωτεΐνες του δερματίου (Willis 2010; Willis, Papandreou et al. 2012).

Έχουν παρατηρηθεί και ταυτοποιηθεί τρεις (3) διακριτές μορφές της συναινετικής ακολουθίας R&R, και έχουν ονομαστεί RR-1, RR-2 και RR-3 (Andersen 1998; Andersen 2000). Η μορφή RR-1 αφορά πρωτεΐνες που έχουν απομονωθεί από μαλακά δερμάτια (soft/flexible cuticles), η μορφή RR-2 αφορά πρωτεΐνες που σχετίζονται με τα σκληρά δερμάτια (hard/rigid cuticles) (Andersen 2000), ενώ η μορφή RR-3 αφορά μια μικρή ομάδα πρωτεϊνών. Οι RR-2 πρωτεΐνες εμφανίζονται κυρίως στο εξωδερμάτιο (exocuticle), ενώ οι RR-1 πρωτεΐνες έχουν βρεθεί κυρίως στο ενδοδερμάτιο (endocuticle) (Andersen 2000). Η RR-2 συναινετική ακολουθία εμφανίζει μεγαλύτερη συντήρηση από την RR-1, όπως φαίνεται και στις εικόνες 18 και 19 (Willis, Papandreou et al. 2012). Οι RR-3 πρωτεΐνες περιέχουν ένα μοτίβο 75 αμινοξικών καταλοίπων που δεν έχουν οι άλλες δυο υποοικογένειες (Andersen 2000).

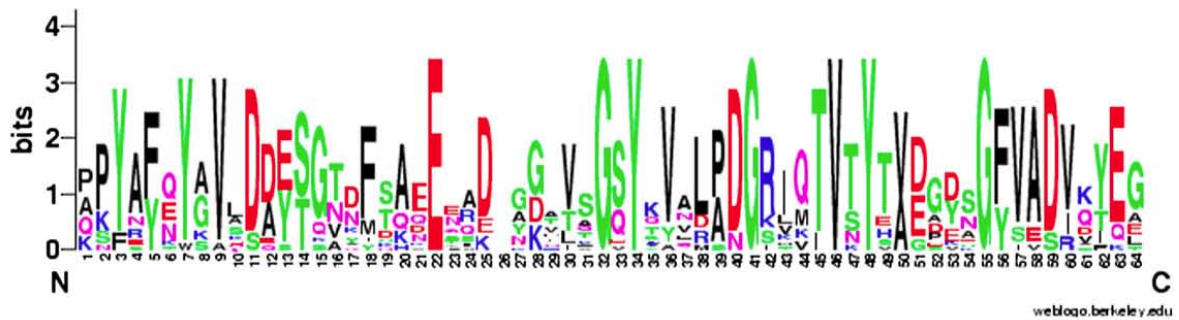
Αξίζει να αναφερθεί ότι ο ιστότοπος της CuticleDB (Magkrioti, Spyropoulos et al. 2004) παρέχει ένα χρήσιμο εργαλείο (Karouzou, Spyropoulos et al. 2007), το οποίο αναγνωρίζει αν μια πρωτεΐνη ανήκει στην RR-1 ή στην RR-2 υποοικογένεια.

Από όταν πρωτοπαρατηρήθηκε η R&R συναινετική ακολουθία, είχε προταθεί ότι παίζει κάποιο θεμελιώδη ρόλο στην δομή και λειτουργία του δερματίου των αρθροπόδων (Rebers and Riddiford 1988). Αργότερα, προτάθηκε ότι η συναινετική ακολουθία έχει ιδιότητες ικανές για αλληλεπίδραση των πρωτεϊνών με χιτίνη (Bouhin, Charles et al. 1992; Charles, Bouhin et al. 1992). Εν συνεχεία, παρατηρήθηκε ότι μια CPR πρωτεΐνη του *Anopheles gambiae* (αρχικά ήταν γνωστή ως Agcp2b - ENTREZ accession number: 2961110, ενώ τώρα ονομάζεται AgamCPR97) μπορεί να αλληλεπιδρά με χιτίνη (Rebers and Willis 2001). Πλέον, είναι αποδεδειγμένο ότι οι CPR πρωτεΐνες έχουν την ικανότητα να αλληλεπιδρούν με χιτίνη (Togawa, Nakato et al. 2004; Togawa, Augustine Dunn et al. 2007; Qin, Lapidot et al. 2009; Willis 2010).

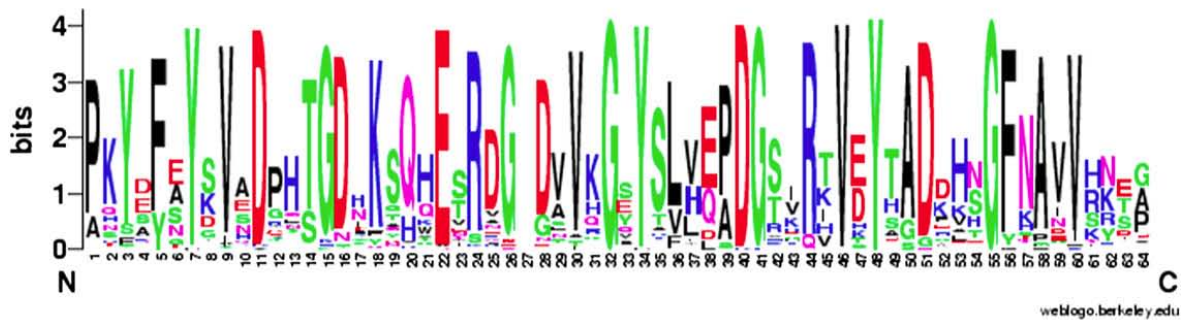
A RR-2 Chelicerata



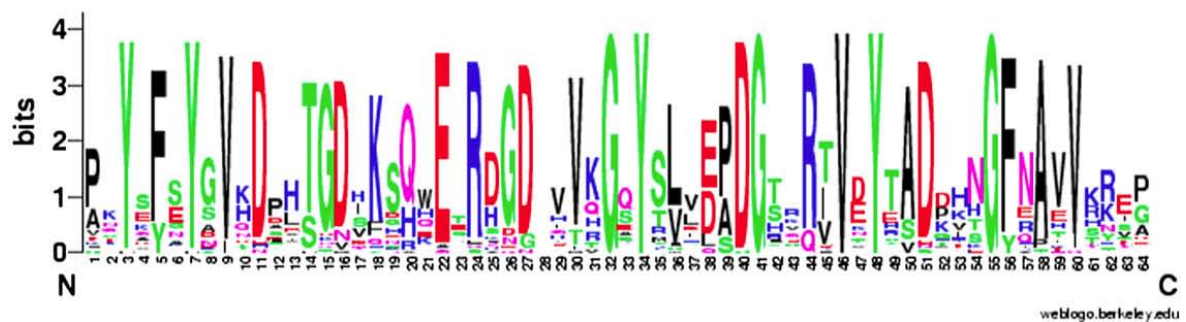
B RR-2 Crustacea



C RR-2 Lepidoptera

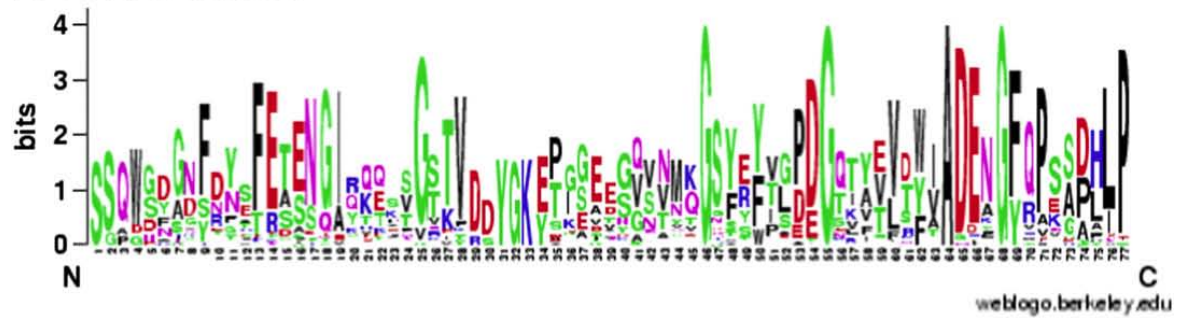


D RR-2 Diptera

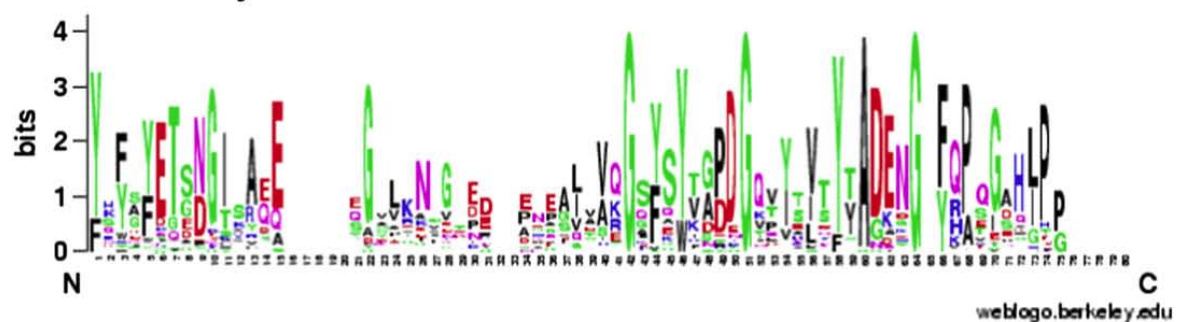


Εικόνα 18: Σύγκριση των WebLogos των RR-2 συναινετικών ακολουθιών από τέσσερις (4) τάξεις αρθροπόδων. Στα Χηληκεραιωτά (Chelicerata) (A) βασίστηκε σε 33 ακολουθίες. Στα Καρκινοειδή (Crustacea) (B) βασίστηκε σε 16 ακολουθίες. Στα Λεπιδόπτερα (Lepidoptera) (C) βασίστηκε σε 87 ακολουθίες από τον *Bombyx mori*. Στα Δίπτερα (Diptera) (D) βασίστηκε σε 101 ακολουθίες από τον *Anopheles gambiae* (Willis 2010).

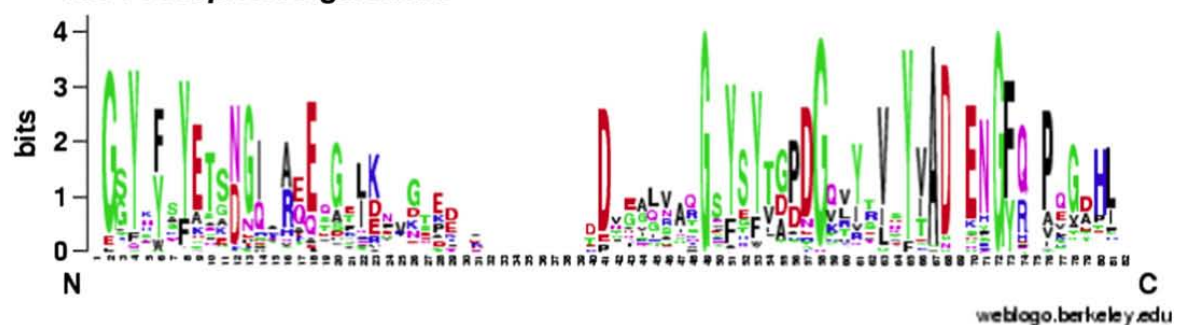
A RR-1 Crustacea



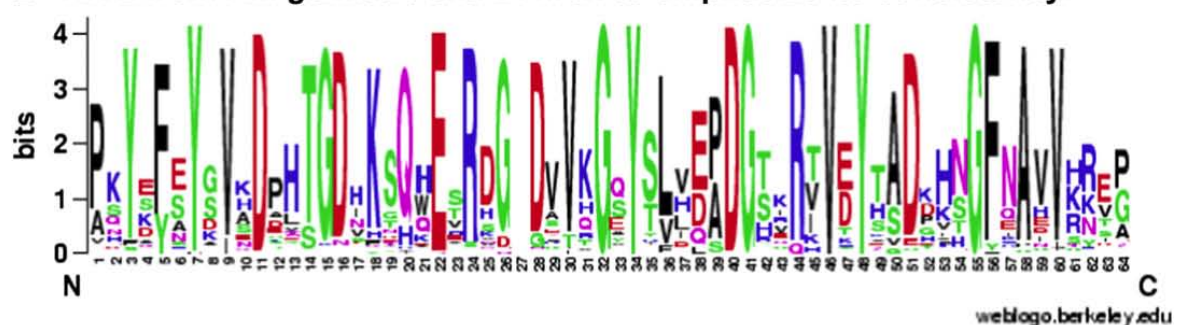
B RR-1 *Bombyx mori*



C RR-1 *Anopheles gambiae*



D RR-2 from *An. gambiae* and *B. mori* to emphasize its consistency.



Εικόνα 19: Τα WebLogos των πρωτεϊνών της RR-1 σε σύγκριση με την RR-2 συναινετική ακολουθία.

Το WebLogo Των καρκινοειδών (Crustacea) (A) βασίζεται σε 48 ακολουθίες, του λεπιδόπτερου (B) *Bombyx mori*, βασίζεται σε 52 ακολουθίες και του δίπτερου (C) *Anopheles gambiae* βασίζεται σε 51 ακολουθίες. Η σύνθεση των RR-2 ακολουθιών (D) βασίζεται σε συνδυασμό των δεδομένων των (C) και (D) της εικόνας 18 (Willis 2010).

1.6.2.2.2 CPF

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins with Forty residues repeat», δηλαδή «πρωτεΐνες του δερματίου με επανάληψη σαράντα αμινοξικών καταλοίπων» (Willis 2010).

Αρχικά, το “F” στο ακρωνύμιο σήμαινε πενήντα (Fifty) καθώς είχε αναγνωριστεί μια ακολουθία 51 αμινοξικών καταλοίπων, σε έξι (6) πρωτεΐνες του δερματίου από τους οργανισμούς *Tenebrio molitor* και *Locusta migratoria* (Andersen, Rafn et al. 1997). Η αρχική συναινετική ακολουθία ήταν η ακόλουθη: (AY)-(AP)-x(2)-(PA)-(PA)-//A-(LIV)-x-(SA)-(QS)-x-(SQ)-x-(IV)-(LV)-R-S-x-G-(NG)-x(3)-V-S-x-Y(ST)-K-(TA)-(VI)-D-(TS)-(PA)-(YF)-S-S-V-x-K-x-D-x-R-(VI)-(TS)-N-x-(GA)//(IVL) (Andersen, Rafn et al. 1997).

Σήμερα, η συναινετική ακολουθία αποτελείται πλέον από 42-44 αμινοξικά κατάλοιπα (Togawa, Augustine Dunn et al. 2007; Willis 2010). Επιπλέον, εκτός του βασικού μοτίβου, έχει παρατηρηθεί ότι υπάρχουν αρκετές σημαντικές ομοιότητες στα αμινοξικά κατάλοιπα των CPF πρωτεϊνών που βρίσκονται κοντά στο καρβοξυτελικό άκρο (Togawa, Augustine Dunn et al. 2007; Willis 2010).

Ακόμα δεν γνωρίζουμε ποιά ακριβώς είναι η λειτουργία των πρωτεϊνών αυτής της οικογένειας στο δερμάτιο (Willis 2010), ωστόσο είναι σίγουρο ότι είναι πρωτεΐνες του δερματίου καθώς έχουν παρατηρηθεί μοναδικά πεπτίδια σε πρωτεομική ανάλυση στο δερμάτιο (He, Botelho et al. 2007). Έχει προταθεί ότι αλληλεπιδρούν με συστατικά του δερματίου, όπως με τις φερομόνες φύλου (sex pheromones) (Hall 1994; Greenspan and Ferveur 2000) ή με λιπίδια του δερματίου (Papandreou, Iconomidou et al. 2010). Επίσης, τα μέχρι στιγμής δεδομένα δείχνουν ότι οι CPF πρωτεΐνες δεν αλληλεπιδρούν με χιτίνη, όμως δεν μπορεί να αποκλειστεί τελείως για όλα τα στάδια ζωής των αρθροπόδων (Rebers and Willis 2001; Willis, Iconomidou et al. 2005; Togawa, Augustine Dunn et al. 2007; Papandreou, Iconomidou et al. 2010; Willis 2010).

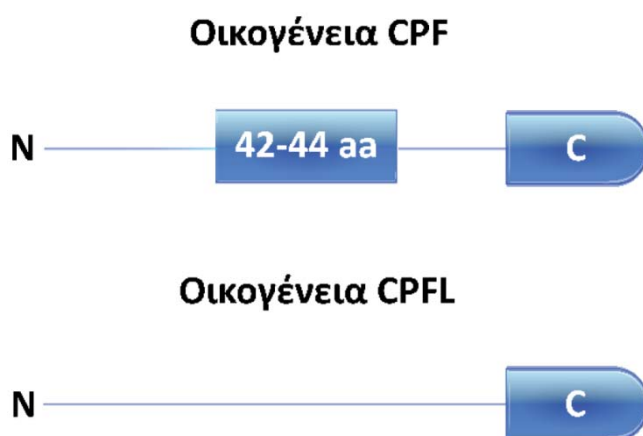
Τέλος, οι πρωτεΐνες της οικογένειας CPF έχουν παρατηρηθεί στα εξάποδα (hexapoda), συμπεριλαμβανομένων των κολλέμβολα (collembola) και των δίπλουρων (diplura), αλλά όχι ακόμα στα καρκινοειδή (crustacea) ή στα χηληκεραιωτά (chelicerata) (Willis, Papandreou et al. 2012).

1.6.2.2.3 CPFL

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins with Forty residues repeat - Like», δηλαδή «πρωτεΐνες του δερματίου που μοιάζουν με αυτές της οικογένειας με την επανάληψη σαράντα αμινοξικών καταλοίπων» (Willis 2010).

Οι πρωτεΐνες της οικογένειας CPFL μοιάζουν με αυτές της οικογένειας CPF. Συγκεκριμένα, παρουσιάζουν το ίδιο μοτίβο στα αμινοξικά κατάλοιπα που βρίσκονται κοντά στο καρβοξυτελικό άκρο, όμως δεν έχουν ομοιότητα στο βασικό μοτίβο 42 αμινοξικών καταλοίπων των CPF (Togawa, Augustine Dunn et al. 2007; Willis 2010).

Οι πρωτεΐνες CPFL, όπως και της οικογένειας CPF, έχουν ταυτοποιηθεί πειραματικά σαν πρωτεΐνες του δερματίου (He, Botelho et al. 2007; Willis 2010). Έχουν βρεθεί σε επτά (7) τάξεις εντόμων (Willis 2010).

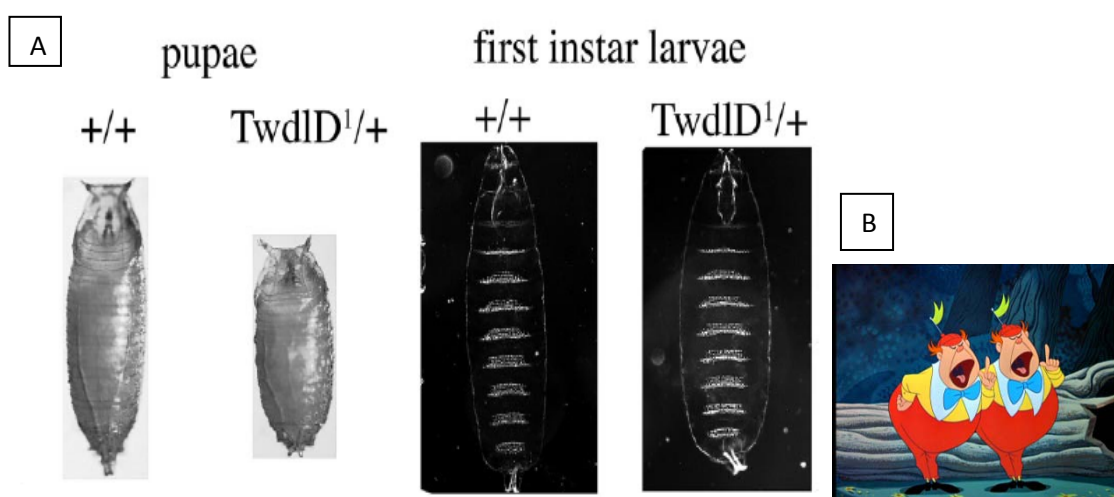


Σχήμα 2: Σχηματική απεικόνιση των οικογενειών CPF και CPFL. Η οικογένεια CPFL, έχει το συντηρημένο καρβοξυτελικό άκρο που έχουν και οι πρωτεΐνες που ανήκουν στην οικογένεια CPF, αλλά δεν φέρει την χαρακτηριστική κεντρική περιοχή των 42-44 αμινοξικών καταλοίπων.

1.6.2.2.4 Tweedle (TWDL)

Η Tweedle (TWDL) είναι μια μεγάλη οικογένεια δομικών πρωτεϊνών του δερματίου. Μέχρι σήμερα έχουν ταυτοποιηθεί ως μέλη της οικογένειας Tweedle 236 μοναδικές πρωτεϊνικές ακολουθίες, από 17 διαφορετικούς οργανισμούς αρθροπόδων (Cornman 2009; Cornman and Willis 2009; Willis 2010; Zhan, Merlin et al. 2011; Willis, Papandreou et al. 2012).

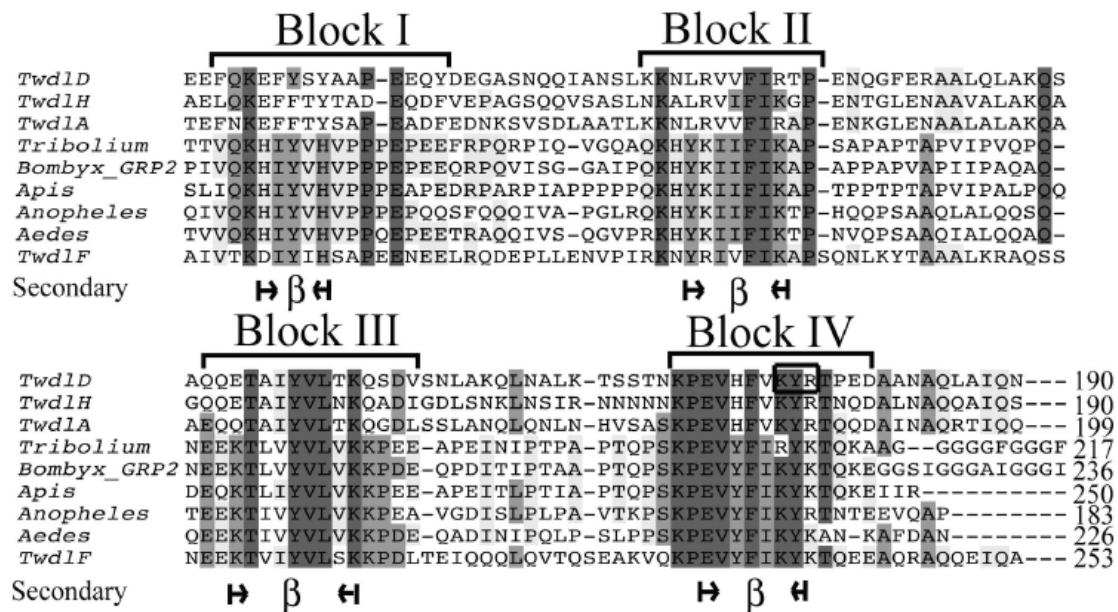
Η οικογένεια παρατηρήθηκε πρώτη φορά στον οργανισμό *Drosophila melanogaster* και πήρε το όνομά της λόγω της ομοιότητας που εμφάνιζε μια μετάλλαξη ενός μέλους της (TweedleD, UniProt accession number: Q9VBD2) με τον χαρακτήρα “Tweedledee” από το κλασικό βιβλίο του Lewis Carroll “Alice through the Looking-glass” (Guan, Middlebrooks et al. 2006) (εικόνα 20).



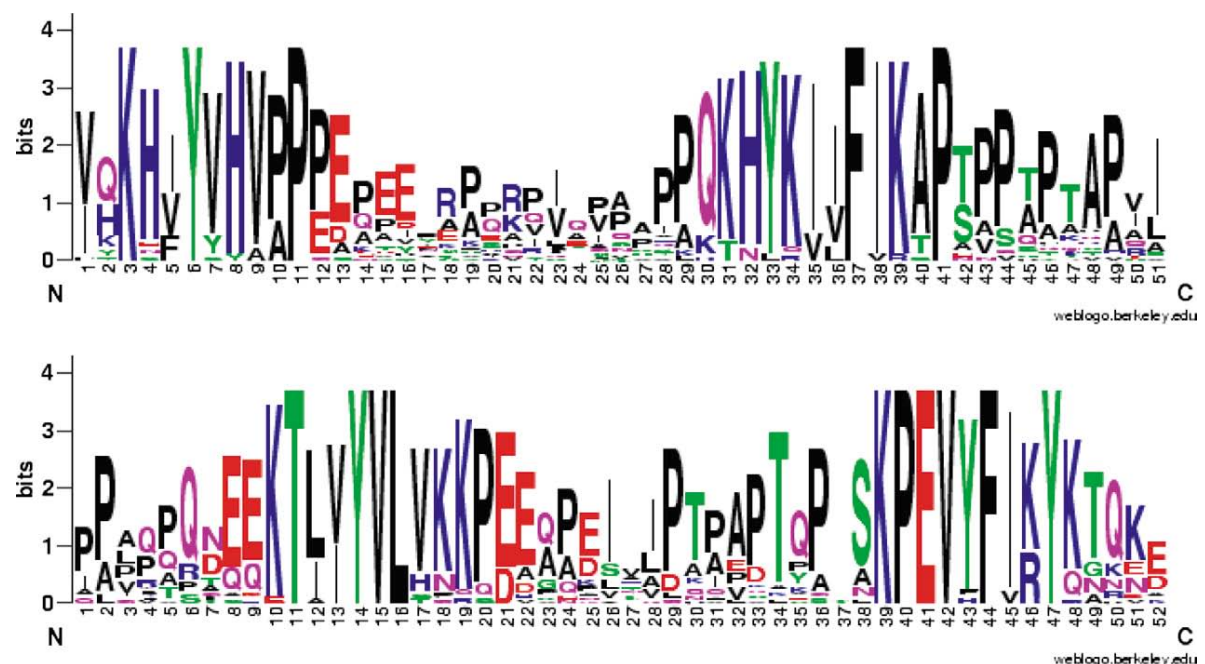
Εικόνα 20: Στο Α μέρος της εικόνας παρουσιάζεται ο αρχικός φαινότυπος της πρωτεΐνης TweedleD και εν συνεχεία ο μεταλλαγμένος φαινότυπος (Guan, Middlebrooks et al. 2006). Ο μεταλλαγμένος φαινότυπος, ο οποίος έδωσε και την ιδέα για το όνομα της ομάδας, θύμισε στους ερευνητές τον χαρακτήρα “Tweedledee” από το κλασικό βιβλίο του Lewis Carroll “Alice through the Looking-glass”, που εμφανίζεται στο Β μέρος της εικόνας.

Η οικογένεια είναι ευρέως διαδεδομένη, με ταυτοποιημένη παρουσία σε επτά (7) τάξεις εντόμων (Diptera, Lepidoptera, Coleoptera, Hymenoptera, Hemiptera, Phthiraptera, Dictyoptera), για τις οποίες έχουμε ικανοποιητικά δεδομένα. Δεν έχει ταυτοποιηθεί η παρουσία της στα υπόφυλα των καρκινοειδών (Crustacea) και των χηληκεραιωτών (Chelicerata) (Willis 2010; Willis, Papandreou et al. 2012).

Έχουν παρατηρηθεί τέσσερις (4) περιοχές με συντηρημένα αμινοξικά κατάλοιπα στην αλληλουχία, οι οποίες είναι διαγνωστικές για τις πρωτεΐνες αυτές στην τάξη των εντόμων (Guan, Middlebrooks et al. 2006). Με το λογισμικό PhD (Rost 1996) έχει προβλεφθεί ότι οι β-κλώνοι είναι το βασικό στοιχείο δευτεροταγούς δομής για τις τέσσερις αυτές συντηρημένες περιοχές (Guan, Middlebrooks et al. 2006) (εικόνες 21 και 22).



Εικόνα 21: Οι αρχικές πειραματικές στοιχίσεις ακολουθιών της οικογένειας Tweedle όπου εμφανίστηκαν οι τέσσερις (4) περιοχές με τα συντηρημένα αμινοξικά κατάλοιπα (blocks) (Guan, Middlebrooks et al. 2006).



Εικόνα 22: Το Web Logo των πρωτεϊνών της οικογένειας Tweedle. Η ακολουθία είναι συνεχόμενη για τα 103 αμινοξικά κατάλοιπα που φαίνονται στην εικόνα και κόπηκε σε δύο μέρη για να είναι εμφανείς οι τέσσερις (4) συντηρημένες περιοχές. Το Web Logo δημιουργήθηκε από είκοσι-τέσσερις (24) ακολουθίες προερχόμενες οκτώ (8) διαφορετικούς οργανισμούς (Willis 2010; Willis, Papandreou et al. 2012).

Επίσης, η παρουσία αρωματικών αμινοξικών καταλοίπων μέσα στους προβλεφθέντες β-κλώνους (Guan, Middlebrooks et al. 2006), έχει προταθεί σαν ένα ιδανικό περιβάλλον για δομές β-βαρελιών που μπορούν να αλληλεπιδρούν με χιτίνη (Iconomidou, Willis et al. 1999; Hamodrakas, Willis et al. 2002). Έχει ακόμα προταθεί ότι η συντήρηση της συναινετικής ακολουθίας R&R (στις πρωτεΐνες της οικογένειας CPR) θα μπορούσε κάλλιστα να οφείλεται στην ανάγκη να διατηρηθεί μια στερεοδιάταξη ικανή να αλληλεπιδρά με χιτίνη (chitin-binding domain), λόγω της απουσίας σταθεροποιητικών δισουλφιδικών δεσμών (Rebers and Willis 2001). Η ίδια αιτία μπορεί να εφαρμοστεί στις πρωτεΐνες της οικογένειας Tweedle, όπου παρουσιάζονται και σε αυτές τις πρωτεΐνες περιοχές με συντηρημένες συναινετικές ακολουθίες (Guan, Middlebrooks et al. 2006; Willis, Papandreou et al. 2012).

Πλέον, τα μέλη της οικογένειας TWDL θεωρούνται (μαζί με τις πρωτεΐνες της οικογένειας CPR) ότι μπορούν να αλληλεπιδρούν με χιτίνη. Αυτή η σημαντική ανακάλυψη βασίστηκε στην παρουσία της Tweedle πρωτεΐνης BmorCPT1 (UniProt accession number: C0H6Z4) μετά από αναλύσεις πρωτεομικής, σε πρωτεΐνες προνύμφων του οργανισμού *Bombyx mori*, όπου διαπιστώθηκε ότι αλληλεπιδρά με χιτίνη. Επιπλέον, με τη χρήση μιας ανασυνδιασμένης μορφή της, διαπιστώθηκε ότι αλληλεπιδρά με χιτίνη σε *in vitro* δοκιμές. (Tang, Liang et al. 2010; Willis, Papandreou et al. 2012).

1.6.2.2.5 CPLCG

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins of Low-Complexity with conserved Glycine residues», δηλαδή «χαμηλής πολυπλοκότητας πρωτεΐνες του δερματίου με συντηρημένα κατάλοιπα γλυκίνης (G)» (Cornman and Willis 2009).

Η οικογένεια CPLCG είναι η μεγαλύτερη από τις «καινούριες» δομικές οικογένειες του δερματίου. Αρχικά αναγνωρίστηκε η οικογένεια από δυο (2) πρωτεΐνες που εντοπίστηκαν στη *Drosophila melanogaster* (Qiu and Hardin 1995). Σήμερα, έχουν εντοπιστεί τρεις (3) στη *Drosophila melanogaster* και έχουν εντοπιστεί ακόμα 27 πρωτεΐνες στον *Anopheles gambiae* (Cornman and Willis 2009; Willis 2010).

Το χαρακτηριστικό της οικογένειας είναι το μοτίβο G-x(2)-H-x-A-P-x(2)-G-H (όπου τα αμινοξικά κατάλοιπα x περιορίζονται σε ελάχιστα αμινοξέα) το οποίο βρίσκεται κοντά στο καρβοξυτελικό άκρο (Cornman and Willis 2009; Willis 2010). Το μοτίβο παρουσιάζεται στο WebLogo της οικογένειας, στην εικόνα 23.

Επιπλέον, η οικογένεια δεν εμφανίζεται μόνο στα δίπτερα (diptera), αλλά και σε άλλες τάξεις (orders) εντόμων, καθώς και στα καρκινοειδή (π.χ. στη *Daphnia*) (Willis 2010; Willis, Papandreou et al. 2012).



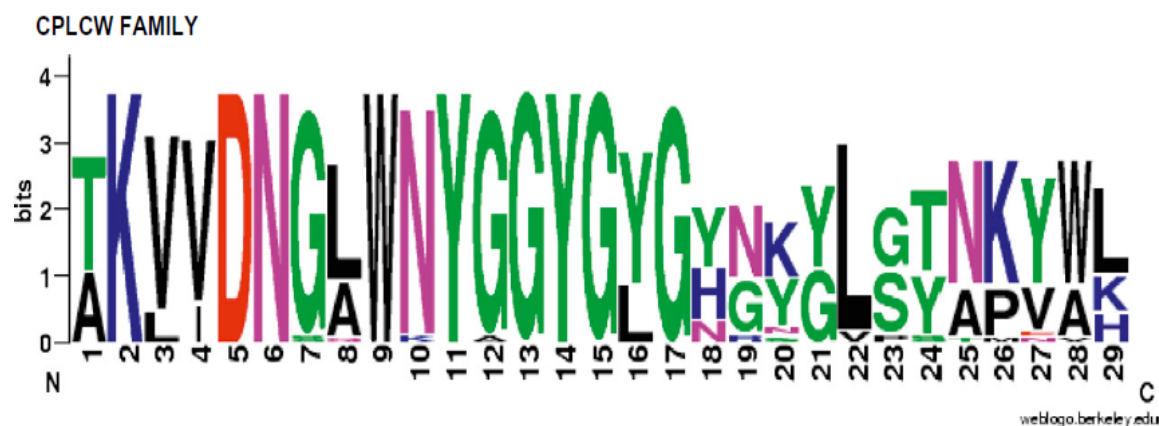
Εικόνα 23: Το WebLogo της οικογένειας CPLCG. Για να δημιουργηθεί χρησιμοποιήθηκαν 86 ακολουθίες από δίπτερα. Φαίνεται έντονα, η υψηλή συντήρηση στα αμινοξικά κατάλοιπα GHPG, στις θέσεις 5, 8, 11 και 14 (Willis 2010).

1.6.2.2.6 CPLCW

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins of Low-Complexity with invariant W residue», δηλαδή «χαμηλής πολυπλοκότητας πρωτεΐνες του δερματίου με αμετάβλητα κατάλοιπα τρυπτοφάνης (W)» (Willis 2010).

Πρόκειται για μια ακόμα μικρή οικογένεια δομικών πρωτεϊνών, από την οποία εννέα (9) μέλη έχουν εντοπιστεί στον *Anopheles gambiae*. Η οικογένεια φαίνεται να εμφανίζεται μόνο στα κουνούπια (Willis 2010; Willis, Papandreou et al. 2012).

Την οικογένεια χαρακτηρίζει το αμινοξικό κατάλοιπο τρυπτοφάνης (W) που εμφανίζεται στην θέση εννέα (9), καθώς και μια περιοχή 29 αμινοξικών καταλοίπων που εμφανίζει μεγάλη ομοιότητα, όπως φαίνεται και στο παρακάτω WebLogo (εικόνα 24) (Willis, Papandreou et al. 2012). Επίσης, τα μέλη αυτής της οικογένειας κυμαίνονται σε μήκος από 106 μέχρι 175 αμινοξικά κατάλοιπα και έχουν αμινοξική ταυτότητα της τάξης του 92% - 100% (Cornman and Willis 2009; Willis 2010).



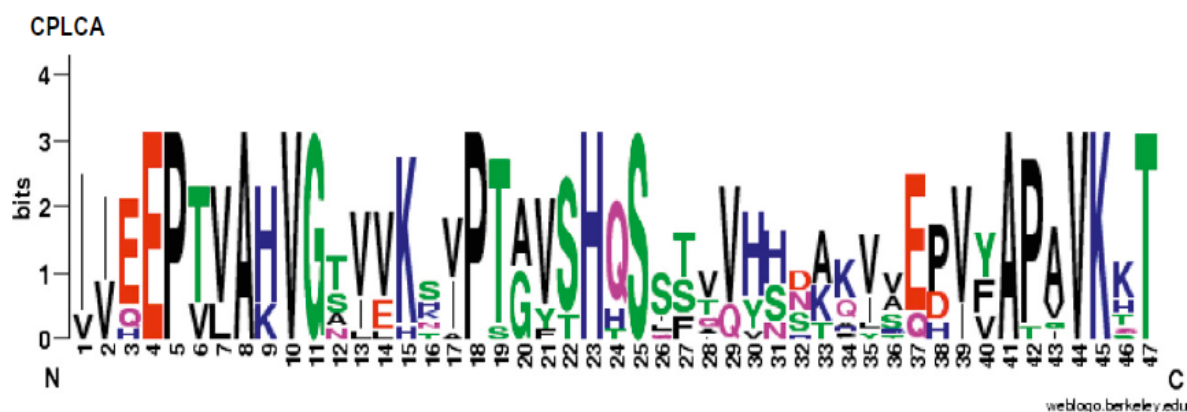
Εικόνα 24: Το WebLogo της οικογένειας CPLCW. Για την δημιουργία του χρησιμοποιήθηκαν οι 26 ακολουθίες της οικογένειας (οι οποίες υπάρχουν μόνο σε κουνούπια) (Willis 2010).

1.6.2.2.7 CPLCA

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins of Low Complexity with Alanine residues», δηλαδή «χαμηλής πολυπλοκότητας πρωτεΐνες του δερματίου με κατάλοιπα αλανίνης» (Willis 2010).

Τα μέλη της οικογένειας περιέχουν 13% - 26% κατάλοιπα αλανίνης, που είναι χαμηλότερο ποσοστό από ότι μερικές πρωτεΐνες που ανήκουν σε άλλες οικογένειες. Τα μέλη της οικογένειας ταυτοποιούνται από την παρουσία ενός retinin domain (pfam04527 / IPR007614) (Cornman and Willis 2009; Willis 2010; Willis, Papandreou et al. 2012). Η πρωτεΐνη ρετινίνη (retinin) της *Drosophila melanogaster* (με μόλις 10.6% κατάλοιπα αλανίνης), που εκφράζεται στον κερατοειδή, αποτελεί το φυλογενετικό όριο της οικογένειας (Kim, Choi et al. 2008).

Αρχικά φαινόταν ότι η οικογένεια αυτή συναντάται μόνο στα δίπτερα (diptera), όμως πλέον έχει βρεθεί και στο γένος *Daphnia* που ανήκει στην τάξη των καρκινοειδών (crustacea) (Cornman and Willis 2009; Willis 2010; Willis, Papandreou et al. 2012).



Εικόνα 25: Το WebLogo της οικογένειας CPLCA. Βασίστηκε σε τρεις ακολουθίες από καθένα από τα τέσσερα είδη των *Anopheles gambiae*, *Aedes aegypti*, *Culex pipiens* και *Drosophila melanogaster* και είχαν την περισσότερη ταύτιση με την πρωτεΐνη AgamCPLCA1. Η συγκεκριμένη περιοχή αντιστοιχεί στο retinin domain (Willis 2010).

1.6.2.2.8 CPLCP

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins of Low-Complexity with conserved Proline residues», δηλαδή «χαμηλής πολυπλοκότητας πρωτεΐνες του δερματίου με διατηρημένα κατάλοιπα προλίνης (P)» (Cornman and Willis 2009).

Η οικογένεια CPLCP είναι η πιο προβληματική στον χαρακτηρισμό της καθώς δεν εμφανίζει ικανοποιητικά δεδομένα σε πρωτεομικές αναλύσεις (Cornman and Willis 2009; Willis, Papandreou et al. 2012). Τα μέλη της οικογένειας εμφανίζουν σε μεγάλη συχνότητα ζευγάρια προλίνης-βαλίνης (P-V) και προλίνης-τυροσίνης (P-Y). Επίσης, εμφανίζονται επιπλέον μοτίβα, αλλά μόνο σε συγκεκριμένα είδη. Για παράδειγμα το μοτίβο GLW[DE] που εμφανίζεται μόνο στα κουνούπια (Cornman and Willis 2009; Willis 2010; Willis, Papandreou et al. 2012). Τέλος, αυτή η οικογένεια εμφανίζεται σε αρκετά διαφορετικά έντομα, ωστόσο στα κουνούπια (π.χ. στον *Aedes aegypti* με 19 μέλη και στον *Culex pipiens* με 24 μέλη) παρατηρείται ο μεγαλύτερος αριθμός αυτών των πρωτεϊνών (Cornman and Willis 2009; Willis 2010; Willis, Papandreou et al. 2012).

1.6.2.2.9 CPG

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins Glycine-rich», δηλαδή «πρωτεΐνες του δερματίου πλούσιες σε γλυκίνη (G)» (Willis 2010).

Μια ομάδα 28 γονιδίων που κωδικοποιεί αλληλουχίες πλούσιες σε επαναλήψεις GGGG και GGxGG παρατηρήθηκε αρχικά στον *Bombyx mori* (Futahashi, Okamoto et al. 2008). Η οικογένεια αυτή εμφανίζει μεγάλη ετερογένεια, καθώς πρωτεΐνες που έχουν αυτά τα χαρακτηριστικά φαίνεται να ανήκουν σε άλλες οικογένειες (π.χ. στην οικογένεια CPLCP) (Willis 2010; Willis, Papandreou et al. 2012). Τέλος, φαίνεται η οικογένεια αυτή να εμφανίζεται μόνο στα λεπιδόπτερα, καθώς έχουν βρεθεί πολλές πρωτεΐνες (18) μόνο στον *Bombyx mori* (Willis 2010).

1.6.2.2.10 Apidermin

Αρχικά, τρεις (3) πρωτεΐνες της οικογένειας αυτής βρέθηκαν στην *Apis mellifera*. Πρόκειται για μικρές σε μέγεθος (6.1 - 9.2 kDa), υψηλής υδροφοβικότητας και με τουλάχιστον 30% περιεκτικότητα στο αμινοξικό κατάλοιπο της αλανίνης (A) (Kucharski, Maleszka et al. 2007). Πλέον, τρεις (3) παρόμοιες πρωτεΐνες έχουν βρεθεί στην *Nasonia vitripennis*, οι οποίες όμως είναι σαφώς μεγαλύτερες σε μέγεθος (21.6 - 37 kDa) (Willis 2010; Willis, Papandreou et al. 2012).

Ουσιαστικά, πρωτεΐνες αυτής της οικογένειας έχουν βρεθεί μόνο στα υμενόπτερα (*Hymenoptera*) (Willis 2010; Willis, Papandreou et al. 2012). Προσπάθειες να βρεθούν ομόλογα σε άλλα είδη δεν έφεραν επιτυχία, λόγω της απουσίας μοναδικών χαρακτηριστικών σε επίπεδο ακολουθίας που να καθορίζουν αυτή την οικογένεια (Willis 2010). Τα μέλη της έχουν αναγνωριστεί με chromosomal linkage (Willis, Papandreou et al. 2012).

1.6.2.2.11 CPAP3

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins Analogous to Peritrophins – 3 domain», δηλαδή «πρωτεΐνες του δερματίου ανάλογες με περιτροφίνες με τρεις (3) επαναλήψεις του domain ChtBD2» (Willis 2010).

Ένα βασικό χαρακτηριστικό για τις πρωτεΐνες του δερματίου είναι η απουσία του αμινοξικού κατάλοιπου της κυστεΐνης (C) (Andersen, Hojrup et al. 1995; Willis, Iconomidou et al. 2005). Έτσι, ήταν αναπάντεχη η παρατήρηση μια πρωτεΐνης του δερματίου στην *Drosophila melanogaster*, η οποία έχει 18 αμινοξικά κατάλοιπα κυστεΐνης, οργανωμένα σε τρεις (3) επαναλήψεις των έξι (6) κυστεϊνών και ενώνονται μεταξύ τους με τρεις (3) δισουλφιδικούς δεσμούς. Η κάθε επανάληψη αποτελεί ένα domain, το οποίο έχει την ικανότητα αλληλεπίδρασης με χιτίνη και ονομάζεται ChtBD2. Το ChtBD2 domain αρχικά είχε βρεθεί μόνο στην οικογένεια των περιτροφινών (peritrophins) και εκεί οφείλεται η ονομασία της οικογένειας (peritrophins-like). Το αντίστοιχο γονίδιο φαίνεται να εκφράζεται στην εμβρυϊκή τραχεία (embryonic tracheae) και ονομάστηκε gasp (Barry, Triplett et al. 1999). Στη συνέχεια βρέθηκαν και άλλες παρόμοιες πρωτεΐνες στην *Drosophila melanogaster* και ταξινομήθηκαν σαν μέλη της οικογένειας με την ονομασία obrtructor (Behr and Hoch 2005). Σήμερα, ύστερα και από νεότερη ανάλυση σε 42 πρωτεΐνες που βρέθηκαν στο *Tribolium castaneum*, η οικογένεια που φέρει τρεις (3) φορές ChtBD2 domain ονομάζεται CPAP3 και το domain έχει καθοριστεί και είναι το: C – x(13,24) – C – x(5) – C – x(9,10) – C – x(12,16) – C – x(7,8) – C (Jasrapuria, Arakane et al. 2010).

Υποτίθεται ότι οι πρωτεΐνες που περιέχουν το ChtBD2 domain μπορούν να αλληλεπιδρούν με χιτίνη. Όμως αυτό έχει αποδειχθεί μόνο σε μια χιτινάση (Arakane, Zhu et al. 2003) και σε μια ανάλογη πρωτεΐνη από άλλο είδος (Nisole, Stewart et al. 2010). Γενικά το μοτίβο αυτό είναι αρκετά διαδεδομένο και σε οργανισμούς εκτός αρθροπόδων (Jasrapuria, Arakane et al. 2010).

Αξίζει να σημειωθεί ότι έχει προσδιοριστεί κρυσταλλογραφικά πρωτεϊνική δομή που περιέχει αυτό το ChtBD2 domain. Πρόκειται για την πρωτεΐνη tachycitin με PDB ID: 1DQC (Suetake, Tsuda et al. 2000).

1.6.2.2.12 CPAP1

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins Analogous to Peritrophins – 1 domain repeat», δηλαδή «πρωτεΐνες του δερματίου ανάλογες με περιτροφίνες με μια (1) εμφάνιση του domain ChtBD2» (Willis 2010).

Η οικογένεια αυτή παρουσιάζει ακριβώς τα ίδια χαρακτηριστικά με την αμέσως προηγούμενη (CPAP3). Η μόνη διαφορά είναι ότι το μοτίβο ChtBD2 με τις έξι (6) κυστεΐνες εμφανίζεται μόνο μια φορά αντί για τρεις (3). Προφανώς και εδώ οι κυστεΐνες ενώνονται με τρεις (3) δισουλφιδικούς δεσμούς (Jasrapuria, Arakane et al. 2010; Willis 2010). Το μοτίβο στις πρωτεΐνες της οικογένεια CPAP1 έχει την μορφή: C – x(14,16) – C – x(5) – C – x(9,13) – C – x(12) – C – x(7,8) – C (Jasrapuria, Arakane et al. 2010). Τέλος, η οικογένεια CPAP1 είναι λιγότερο διαδεδομένη με την CPAP3, ενώ έχει παρατηρηθεί μόνο στα σκαθάρια (beetles) (Willis, Papandreou et al. 2012).

CPAP3

CPAP3-B CP-KVNETVA---ATHPRYADEDDCQYFVVCINGDTPRRSGCKLGQVFD---DVGKKCDWVRNVPEC : 60

CPAP3-A1 CP-KDGTQDANGQLVHPKYAHPTDCQRFVYVCLNGQEPDLGCQVGEVYN---EESQRCDAPENVPGC : 64

CPAP3-A2 CP-DEEILGPGRKLPHTFAHEDCGKFYICRNGVMPQKGQCVKCLVYN---EETFTCDPKNVPGC : 64

CPAP3-C CP-APGEVNSNG---SFSRHAHEDDCRKYVICLEG-TAREYGCPICTVFKIGDADGTGNCEDPEDVPGC : 64

CPAP3-E CP-NDGRSFGLG-EAEFRFRFSENDGCQRYFVCVNG-RPRLYNCGEGRAFN---DLIGACDGVENVTCG : 62

CPAP3-D1 CP-TKVPSNSPAAKFWPYPRFAVPGDCQRLITCVNG-FPRLISCGECKAFD---QHSLTCEEPELVPHC : 64

CPAP3-D2 CNDDPNGNVPLG-----KSNRRYQCQGG-YPRLQRCPAMLVFD---RRSLRCVVPPTEDC : 52

Consensus CP q P DC RFW C NG PR C G L FD Cd p Vp C

CPAP1

[illegible]

Εικόνα 26: Χαρακτηριστικές στοιχίσεις πρωτεϊνών του δερματίου από τις οικογένειες CPAP1 και CPAP3. Φαίνεται το μοτίβο που εμφανίζεται σε κάθε μια στο κάτω μέρος της εκάστοτε στοίχισης. Στην οικογένεια CPAP3 η παραπάνω στοίχιση αφορά την 3^η επανάληψη του μοτίβου (Jasrapuria, Arakane et al. 2010).

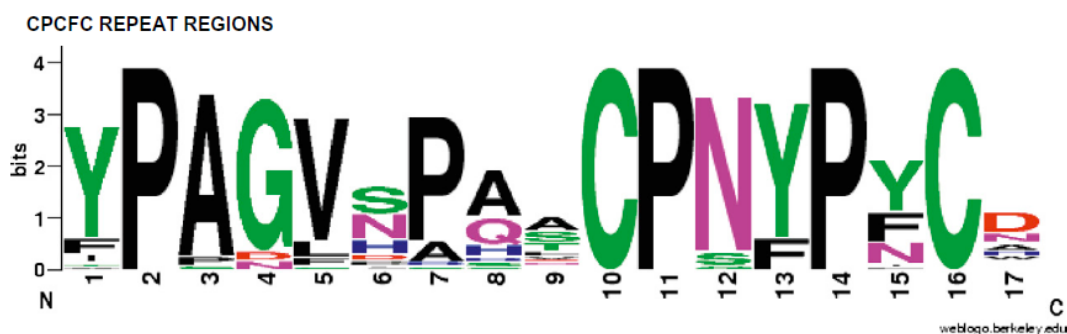
1.6.2.2.13 CPCFC

Το όνομα-ακρωνύμιο της οικογένειας προέρχεται από το «Cuticular Proteins with Cysteine - Five residues - Cysteine», δηλαδή «πρωτεΐνες του δερματίου με επαναλαμβανόμενες ακολουθίες ενός ζευγαριού κυστεϊνών (C) με πέντε αμινοξέα ανάμεσά τους» (Willis, Papandreou et al. 2012).

Το πρώτο μέλος της οικογένειας είναι η πρωτεΐνη BcNCP1 που εντοπίστηκε στον οργανισμό *Blaberus craniifer* (Jensen, Rothmann et al. 1997). Η οικογένεια χαρακτηρίζεται από την παρουσία τριών (3) επαναλαμβανόμενων περιοχών, όπου σε κάθε μία υπάρχει ένα ζευγάρι κυστεϊνών (C), το οποίο χωρίζουν πέντε (5) άλλα αμινοξικά κατάλοιπα, όπου το πρώτο και το τέταρτο αμινοξικό κατάλοιπο σε κάθε επανάληψη είναι προλίνη (P) (εικόνες 27 και 28) (Willis, Papandreou et al. 2012).

Bc-NCP1, no.1	5-Y P A G L N P A L C P N Y P N C-20
no.2	43-Y P A G V S P A A C P N Y P F C-58
no.3	71-Y P A G V H P A A C P N Y P Y C-86
	* * * * * * * * * * *

Εικόνα 27: Χαρακτηριστικό παράδειγμα της οικογένειας CPCFC, και μάλιστα από την πρώτη πρωτεΐνη της οικογενείας (BcNCP1), όπου φαίνονται οι τρεις (3) επαναλαμβανόμενες περιοχές καθώς και τα ζευγάρια κυστεϊνών (C). Τα ταυτόσημα αμινοξικά κατάλοιπα συμβολίζονται με * (Jensen, Rothmann et al. 1997).



Εικόνα 28: Το WebLogo της οικογένειας CPCFC, το οποίο βασίστηκε σε οχτώ (8) διαφορετικά έντομα (Willis 2010).

1.6.2.2.14 Δομικές πρωτεΐνες που δεν ανήκουν σε οικογένεια

Υπάρχουν κάποιες δομικές πρωτεΐνες του δερματίου που δεν πληρούν τα κριτήρια για να κατηγοριοποιηθούν σε κάποια οικογένεια. Ανάμεσα σε αυτές είναι τρεις (3) πρωτεΐνες που έχουν βρεθεί στον *Anopheles gambiae* (Cornman and Willis 2009) και μια ομάδα πρωτεϊνών που ονομάζεται CPH (cuticular protein hypothetical – υποθετικές πρωτεΐνες του δερματίου) και έχουν βρεθεί στον *Bombyx mori* (Futahashi, Okamoto et al. 2008).

1.6.2.3 Συχνά επαναλαμβανόμενα μοτίβα

Στις πρωτεΐνες του δερματίου έχουν παρατηρηθεί μερικά συχνά επαναλαμβανόμενα μοτίβα, τα οποία βρίσκονται σε πολλές οικογένειες. Δυστυχώς, οι λειτουργίες τους παραμένουν ακόμα άγνωστες (Willis 2010).

Το πρώτο μοτίβο που παρατηρήθηκε είναι και το πιο συχνό στις πρωτεΐνες του δερματίου. Πρόκειται για το μοτίβο A-A-P-[AVL] (αρχικά ήταν A-A-P-(A/V), αλλά πήρε την σημερινή του μορφή ύστερα από τον σχολιασμό των πρωτεϊνών του δερματίου του *Anopheles gambiae*) (Andersen, Hojrup et al. 1995; Willis, Papandreou et al. 2012). Το μοτίβο αυτό, αν και αρχικά προτάθηκε σαν ένα διαγνωστικό χαρακτηριστικό για τις πρωτεΐνες του δερματίου – κάτι το οποίο δεν ισχύει, έχει παρατηρηθεί και στις πρωτεΐνες του χορίου καθώς και σε μια πρωτεΐνη του δερματίου στον *Caenorhabditis elegans* (Andersen, Hojrup et al. 1995). Παρ' όλα αυτά έχει προταθεί ότι η πολλαπλή παρουσία αυτού του μοτίβου, σε μια πρωτεΐνη, μπορεί να είναι διαγνωστικό χαρακτηριστικό για τις πρωτεΐνες του δερματίου (Willis 2010). Τέλος, έχει προταθεί ότι το μοτίβο αυτό φαίνεται να παρουσιάζει ισχυρή τάση στην δημιουργία στροφών (turns) (Andersen, Hojrup et al. 1995).

Ένα μοτίβο 18 αμινοξικών καταλοίπων παρατηρήθηκε αρχικά τρεις φορές σε μια πρωτεΐνη του δερματίου (την BmorCPH31) από τον οργανισμό *Bombyx mori* (Nakato, Izumi et al. 1992). Το μοτίβο, σύμφωνα με την βάση δεδομένων cuticleDB (Magkrioti, Spyropoulos et al. 2004), είναι το: V-x-D-T-P-E-V-A-A-A-x-A-A-H-x-A-A-H (Willis 2010), όπως φαίνεται στην εικόνα 29. Στη συνέχεια το μοτίβο παρατηρήθηκε σε έναν αριθμό πρωτεϊνών από τέσσερις (4) τάξεις (orders) εντόμων και σε δυο (2) καρκινοειδή (Andersen 2000). Το μοτίβο αυτό εμφανίζεται σχεδόν σε

όλες τις πρωτεΐνες της υποοικογένειας RR-3 των CPR (Futahashi, Okamoto et al. 2008; Willis 2010). Τέλος, το μοτίβο δεν εμφανίζεται σε καμία πρωτεΐνη του δερματίου που προέρχεται από τα χηληκεραιωτά (από τις ~250 κατατεθειμένες ακολουθίες) (Willis 2010). Η λειτουργία του παραμένει ακόμη άγνωστη (Willis 2010).



Εικόνα 29: Το WebLogo του επαναλαμβανόμενου μοτίβου των 18 αμινοξικών καταλοίπων. Το WebLogo προέρχεται από 26 πρωτεΐνες από 5 τάξεις εντόμων και από 2 καρκινοειδή (Willis 2010).

Επίσης, έχουν παρατηρηθεί μοτίβα που περιέχουν τα αμινοξικά κατάλοιπα της γλυκίνης (G), της τυροσίνης (Y) και της λευκίνης (L) (Andersen, Hojrup et al. 1995; Willis, Papandreou et al. 2012). Τα μοτίβα αυτά συνήθως ξεκινάνε με τις μορφές: G-Y-G-L ή G-L-L-G (Andersen, Hojrup et al. 1995). Ακόμα, άλλες πρωτεΐνες του δερματίου είναι πλούσιες στο αμινοξικό κατάλοιπο της γλυκίνης (G), αλλά με την μορφή συνεχόμενων γλυκίνων στην ακολουθία (Andersen, Hojrup et al. 1995). Πρωτεΐνες πλούσιες σε γλυκίνη έχουν βρεθεί σε μια ποικιλία δομών όπως στα κυτταρικά τοιχώματα των φυτών, σε ωοθήκες κατσαρίδων και στο μετάξι (Bouhin, Charles et al. 1992).

Πρόσφατα παρατηρήθηκαν και αναλύθηκαν δυο (2) μικρά μοτίβα, που βρέθηκαν σε πολλούς οργανισμούς του είδους *Drosophila*, τα οποία είναι τα GYR και YLP (Cornman 2010).

Τέλος, ένα κοινό χαρακτηριστικό στις πρωτεΐνες του δερματίου (εκτός από τις οικογένειες CPAP1, CPAP3, CPCFC) είναι η απουσία του αμινοξικού καταλοίπου της κυστεΐνης (C) στις εκκρινόμενες μορφές τους (Andersen, Hojrup et al. 1995; Willis, Iconomidou et al. 2005).

1.6.3 Ομοιότητα με λιποκαλίνες

Αναπάντεχα, παρατηρήθηκαν μακρινές ομοιότητες σε επίπεδο ακολουθίας (Ταυτότητα ακολουθίας: ~20%, Ομοιότητα ακολουθίας (συντήρηση): ~60%) μεταξύ της εκτεταμένης R&R συναινετικής ακολουθίας αρκετών CPR πρωτεϊνών και της bovine plasma retinol binding protein (PDB ID: 1FEN), που ανήκει στην οικογένεια των λιποκαλινών (Hamodrakas, Willis et al. 2002). Εν συνεχεία, παρατηρήθηκε παρόμοια μακρινή ομοιότητα σε επίπεδο ακολουθίας (Ταυτότητα ακολουθίας: <20%, Ομοιότητα ακολουθίας: ~50%) και μεταξύ των ακολουθιών των πρωτεϊνών της οικογένειας CPF, με την bovine plasma retinol binding protein (PDB ID: 1FEN) (Papandreou, Iconomidou et al. 2010).

Οικογένεια	Sequence Identity	Sequence Similarity	Reference
CPR	~ 20%	~ 60%	(Hamodrakas, Willis et al. 2002)
CPF	< 20%	~ 50%	(Papandreou, Iconomidou et al. 2010)

Πίνακας 4: Τα ποσοστά ταυτότητας ακολουθίας (Sequence Identity) και ομοιότητας ακολουθίας (Sequence Similarity), των δομικών πρωτεϊνικών οικογενειών του δερματίου CPR και CPF, με την ακολουθία της λιποκαλίνης bovine plasma retinol binding protein (PDB ID: 1FEN). Στην τελευταία στήλη παρουσιάζεται και η βιβλιογραφία (reference) από όπου προκύπτουν τα αντίστοιχα δεδομένα.

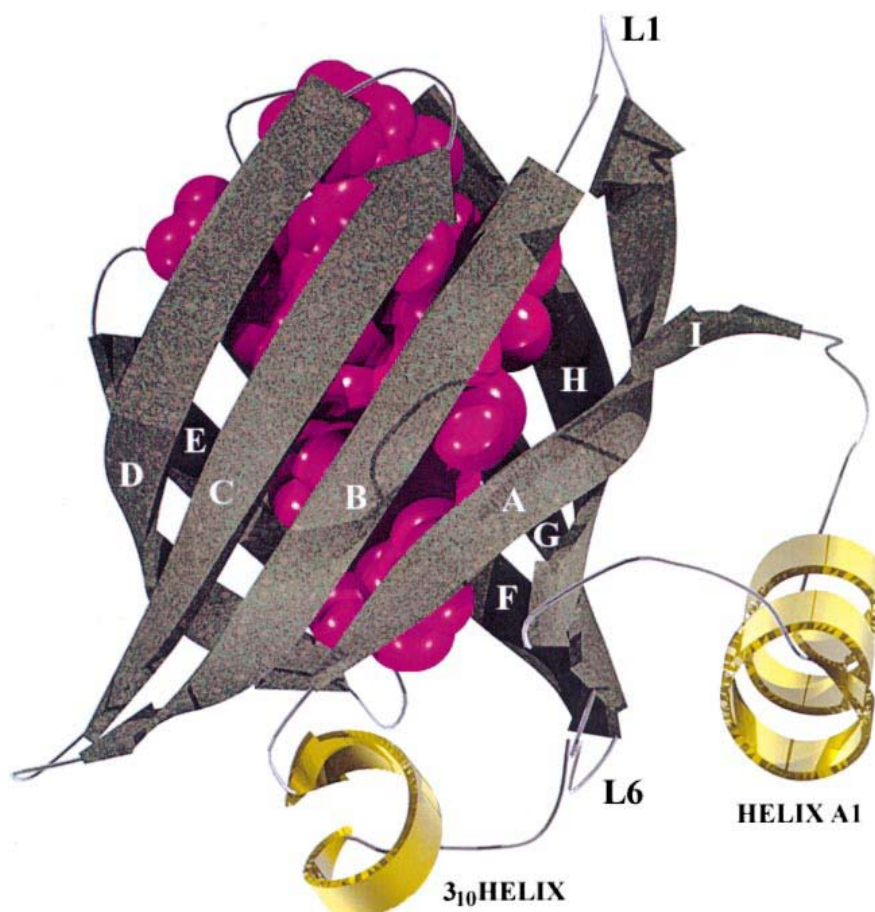
1.6.3.1 Λιποκαλίνες

Τα μέλη της οικογένειας των λιποκαλινών αποτελούν μικρές (περίπου 160-200 αμινοξικά κατάλοιπα) εκκρινόμενες πρωτεΐνες, οι οποίες μοιράζονται αρκετές κοινές συνηθισμένες μοριακές ιδιότητες, όπως είναι: η δέσμευση μικρών, βασικά υδρόφοβων μορίων (όπως η ρετινόλη), η δέσμευση σε ειδικούς υποδοχείς στην κυτταρική επιφάνεια και ο σχηματισμός ομοιοπολικών και μη ομοιοπολικών συμπλόκων με άλλα διαλυτά μακρομόρια (Flower 1996; Flower, North et al. 2000; Adam, Charloteaux et al. 2008).

Οι ποικίλες βιολογικές λειτουργίες των λιποκαλινών προκαλούνται από μία ή περισσότερες από τις παραπάνω ιδιότητες. Οι λιποκαλίνες εμφανίζουν μεγάλη λειτουργική ποικιλομορφία, με συμμετοχή στη μεταφορά ρετινόλης, στο χρωματισμό στα ασπόνδυλα, στην όσφρηση και στη μεταφορά φερομονών καθώς και στη σύνθεση προσταγλανδίνης. Οι λιποκαλίνες επίσης συμμετέχουν στη ρύθμιση της κυτταρικής ομοιόστασης και στη διαμόρφωση της ανοσολογικής απόκρισης, επίσης ως πρωτεΐνες μεταφορείς που ενεργούν για το γενικό καθαρισμό από ενδογενείς και εξωγενείς ενώσεις (Flower 1996; Flower, North et al. 2000; Adam, Charloteaux et al. 2008).

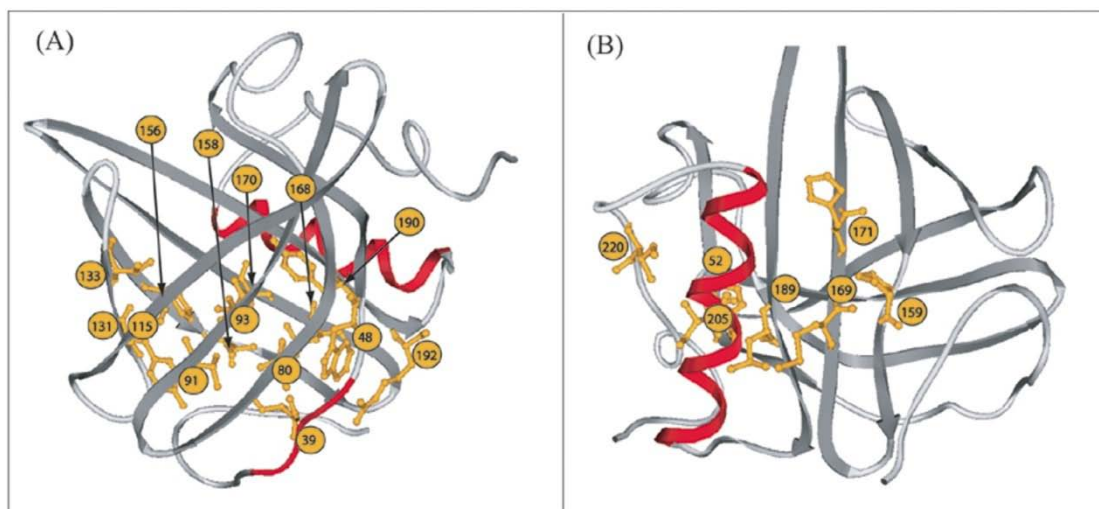
Οι λιποκαλίνες είναι ευρέως διαδεδομένες στη φύση και βρίσκονται σε βακτήρια, φυτά, αρθρόποδα και σπονδυλωτά. Σε αιματοφάγα αρθρόποδα, οι λιποκαλίνες εμπλέκονται στην επιτυχή ολοκλήρωση του γεύματος του αίματος, παρεμβαίνοντας με συσσωμάτωση αιμοπεταλίων, στην πήξη του αίματος και στη φλεγμονή. Επίσης, είναι υπεύθυνες για τη μετάδοση ασθενειών από παράσιτα, όπως το *Trypanosoma cruzi* και το *Borrelia burgdorferi* (Adam, Charloteaux et al. 2008).

Παρά τις κοινές λειτουργίες, οι λιποκαλίνες έχουν χαμηλή ταυτότητα σε επίπεδο ακολουθίας (περίπου 20%), μεταξύ των μελών της οικογένειας. Ωστόσο, μοιράζονται την ίδια δομή, η οποία αποτελείται κυρίως από οκτώ αντιπαράλληλους (8) β-κλώνους που σχηματίζουν ένα β-βαρέλι. Επίσης, μια 3_{10} έλικα βρίσκεται στο ένα άνοιγμα του β-βαρελίου (H1) και μια δεύτερη α-έλικα είναι παράλληλη με την επιφάνεια του β-βαρελίου (H2). Το εσωτερικό της κοιλότητας είναι αρκετά μεγάλο ώστε να επιτρέπει τη σύνδεση ενός μικρού, συνήθως υδρόφοβου, μορίου (εικόνα 30) (Flower 1996; Flower, North et al. 2000; Adam, Charloteaux et al. 2008).



Εικόνα 30: Σχηματική απεικόνιση τυπικής δομής των πρωτεϊνών της οικογένειας των λιποκαλινών. Φαίνεται το β-βαρέλι που αποτελείται από οκτώ (8) (Α-Η) αντιπαράλληλους β-κλώνους (χρωματισμένοι με γκρι χρώμα), η 3_{10} έλικα που βρίσκεται στο ένα άνοιγμα του β-βαρελίου (με χρυσό χρώμα), η α-έλικα (με χρυσό χρώμα δεξιά) καθώς και ένας ένατος (I) κλώνος που υπάρχει σε αρκετές λιποκαλίνες. Οι βρόχοι L1 και L6 είναι και αυτοί σημειωμένοι. Επίσης, η περιοχή δέσμευσης του μορίου παρουσιάζεται με μωβ χρωματισμό (Flower, North et al. 2000).

Επιπλέον, η παρουσία δύο συστάδων (inner and outer clusters) με υδρόφοβα (κυρίως) κατάλοιπα έχει ταυτοποιηθεί. Η μια συστάδα είναι στο εσωτερικό του β-βαρελίου (inner cluster), περιλαμβάνει όλους τους β-κλώνους και την 3_{10} έλικα. Η άλλη συστάδα είναι εξωτερική (outer cluster), περιλαμβάνει τέσσερις β-κλώνους, καθώς και την α-έλικα που τοποθετείται παράλληλα προς την επιφάνεια του β-βαρελίου (εικόνα 31) (Adam, Charlotiaux et al. 2008).



Εικόνα 31: Στις εικόνες (A) και (B) φαίνονται τα αμινοξικά κατάλοιπα (με τις γενικές τους θέσεις) που βρίσκονται στις δύο υδρόφοβες συστάδες των πρωτεϊνών της οικογένειας των λιποκαλινών (απεικονίζονται ως balls and sticks και με πορτοκαλί χρώμα). Στην εικόνα (A) φαίνονται τα αμινοξικά κατάλοιπα που ανήκουν στην εσωτερική συστάδα (inner cluster) και βρίσκονται στις γενικές θέσεις: (39), (48), (80), (91), (93), (115), (131), (133), (156), (158), (168), (170), (190) και (192). Στην εικόνα (B) φαίνονται τα αμινοξικά κατάλοιπα που ανήκουν στην εξωτερική συστάδα (outer cluster) και βρίσκονται στις γενικές θέσεις: (52), (159), (169), (171), (189), (205) και (220) (Adam, Charloteaux et al. 2008).

1.6.3.2 Η λιποκαλίνη bovine Retinol Binding Protein (PDB ID: 1FEN)

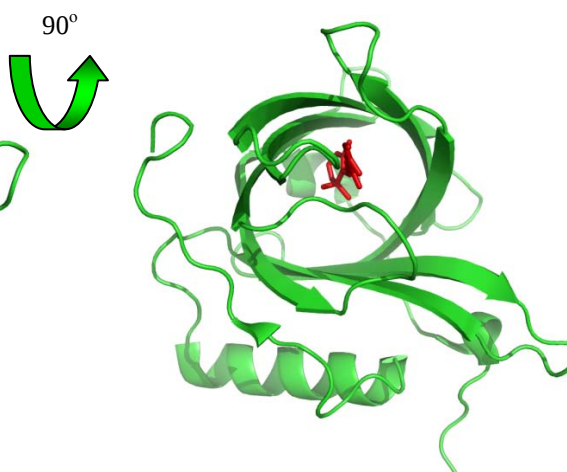
Η λιποκαλίνη όπου στηρίχθηκε η εκπόνηση της παρούσας διπλωματικής είναι η bovine Retinol Binding Protein (PDB ID: 1FEN). Προέρχεται από τον οργανισμό *Bos taurus*. Η ακολουθία της πρωτεΐνης είναι κατατεθειμένη στη βάση δεδομένων UniProt (Arweiler, Bairoch et al. 2004) με κωδικό P18902 (Berni, Stoppini et al. 1990). Η δομή της προσδιορίστηκε με κρυσταλλογραφία ακτίνων-X το 1994, σε διακριτικότητα 1.9Å (Zanotti, Marcello et al. 1994). Είναι κατατεθειμένη στην βάση δεδομένων Protein Data Bank (PDB) (Berman, Westbrook et al. 2000) με κωδικό 1FEN.

Η δομή της λιποκαλίνης bovine RBP (PDB ID: 1FEN) αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους (A-H) που σχηματίζουν ένα β-βαρέλι, έναν (1) ακόμα μικρό β-κλώνο στο τέλος της ακολουθίας, δύο (2) α-έλικες (μια μεγάλη η οποία είναι παράλληλη στους β-κλώνους και μια μικρή η οποία βρίσκεται στο ένα άνοιγμα του β-βαρελιού), μια (1) μικρή 3_{10} έλικα στην αρχή της ακολουθίας καθώς και δέκα (10) στροφές (εικόνες 32 και 33) (Zanotti, Marcello et al. 1994).

(A) Πλαϊνή όψη

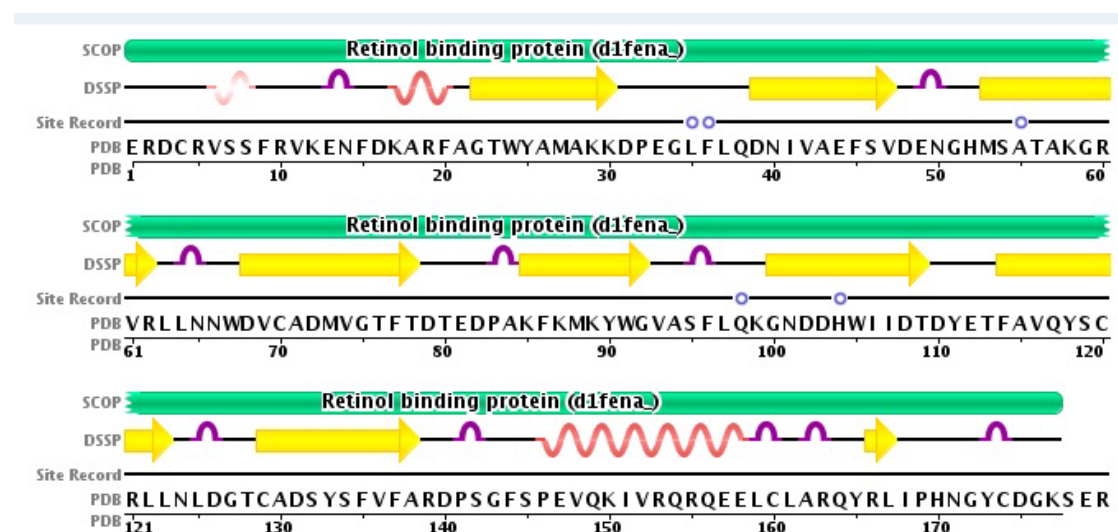


(B) Κάθετη όψη



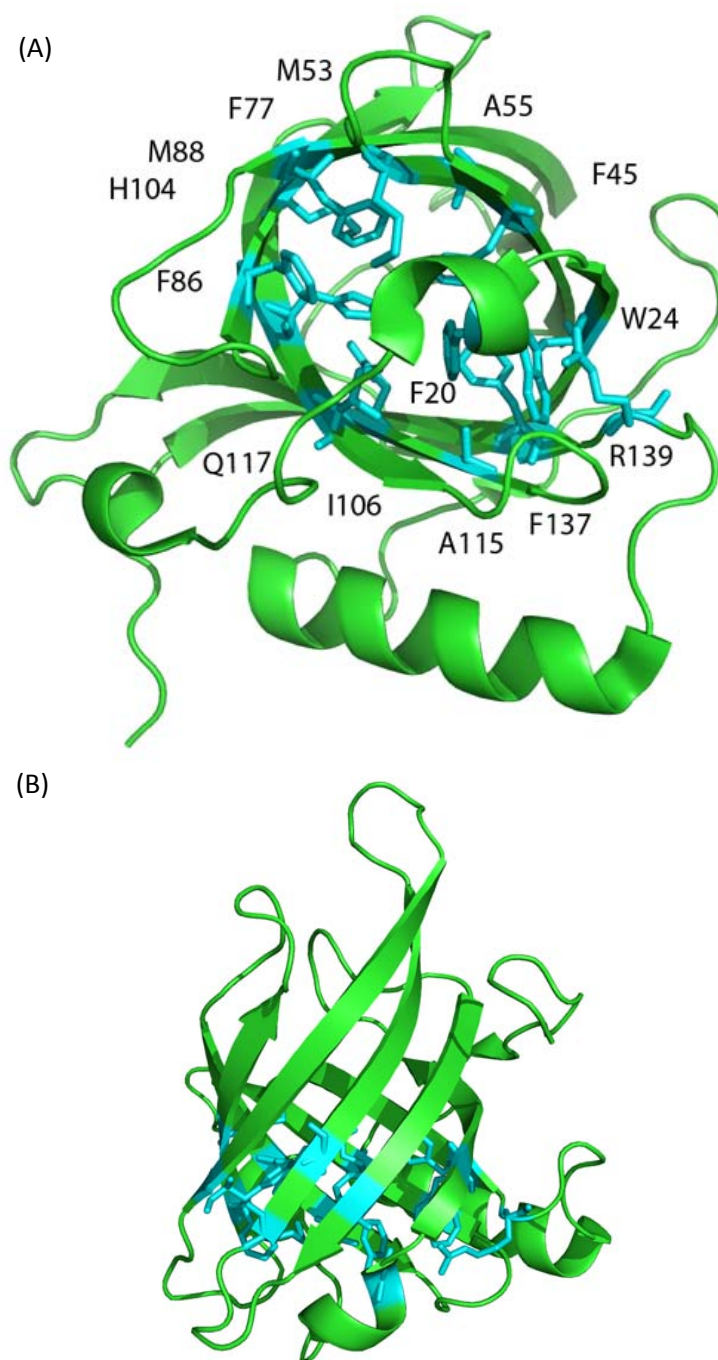
Εικόνα 32: Η προσδιορισμένη κρυσταλλογραφικά δομή της λιποκαλίνης bovine retinol binding protein (PDB ID: 1FEN) με διακριτικότητα 1.9Å (Zanotti, Marcello et al. 1994). Αριστερά (A) είναι η πλαϊνή όψη, ενώ δεξιά (B) είναι η όψη κοιτώντας κάθετα πάνω από το β-βαρέλι. Στις εικόνες φαίνονται τα δομικά χαρακτηριστικά της πρωτεΐνης και κυρίως τα βασικά της, δηλαδή το β-βαρέλι και η α-έλικα που είναι παράλληλη προς τους β-κλώνους. Η πρωτεΐνη είναι χρωματισμένη με ανοιχτό

πράσινο, ενώ με κόκκινο είναι χρωματισμένη η ουσία All-Trans Axerophthene με τύπο $C_{20}H_{30}$ που κρυσταλλώθηκε μαζί με την λιποκαλίνη bovine RBP.

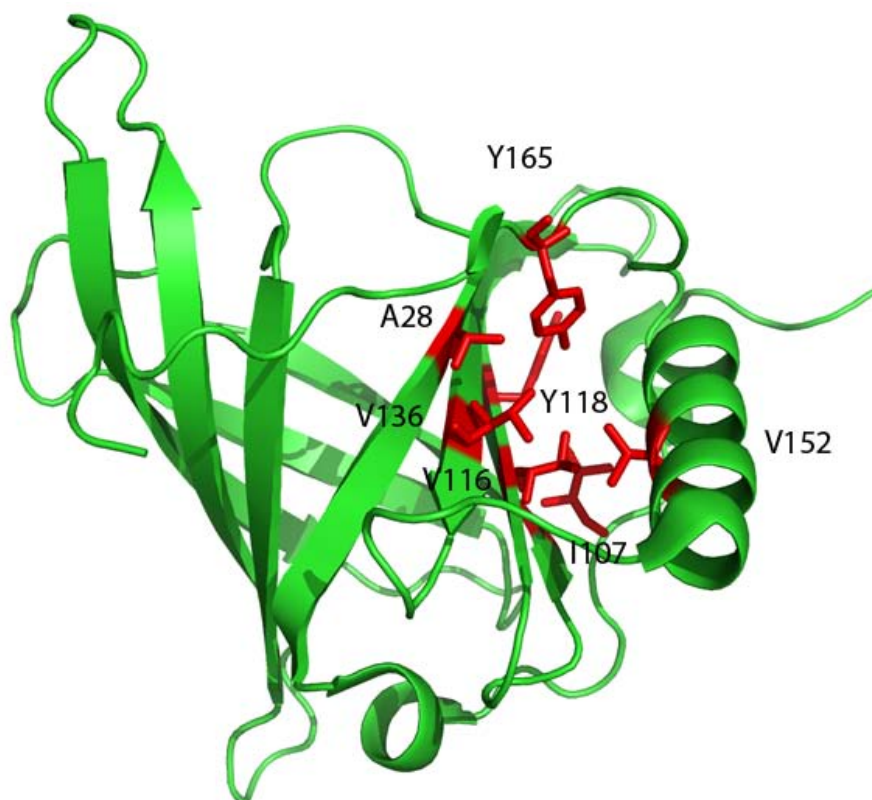


Εικόνα 33: Τα δομικά χαρακτηριστικά της λιποκαλίνης bovine retinol binding protein (PDB ID: 1FEN) όπως είναι κατατεθειμένα στη βάση δεδομένων Protein Data Bank (PDB) (Berman, Westbrook et al. 2000). Στην παρούσα εικόνα εμφανίζονται τα στοιχεία δευτεροταγούς δομής (κατά DSSP) και που αντιστοιχεί κάθε αμινοξικό κατάλοιπο. Με κίτρινα βέλη είναι οι β-κλώνοι, με έντονο ροζ είναι χρωματισμένες οι α-έλικες, με αχνό ροζ είναι η 3_{10} έλικα και με μοβ οι στροφές.

Στην λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN) υπάρχουν δύο (2) συστάδες αποτελούμενες κυρίως από υδρόφοβα αμινοξικά κατάλοιπα (κοινό χαρακτηριστικό των πρωτεϊνών της οικογένειας των λιποκαλινών (Adam, Charleoteaux et al. 2008)). Πρόκειται για μια εσωτερική συστάδα (inner cluster) και μια εξωτερική συστάδα (outer cluster). Την εσωτερική συστάδα αποτελούν τα αμινοξικά κατάλοιπα F20, W24, F45, M53, A55, F77, F86, M88, H104, I106, A115, Q117, F137 και R139 (Papandreou, Iconomidou et al. 2010). Την εξωτερική συστάδα αποτελούν τα αμινοξικά κατάλοιπα A28, I107, V116, Y118, V136, V152 και Y165 (Papandreou, Iconomidou et al. 2010) (εικόνες 34 και 35).



Εικόνα 34: Η εσωτερική συστάδα (inner cluster) της λιποκαλίνης bovine retinol binding protein (PDB ID: 1FEN) από δύο οπτικές γωνίες. Στην (Α) φαίνεται κάθετα στο β-βαρέλι, ενώ στη (Β) φαίνεται η πλαϊνή όψη. Παρουσιάζονται τα αμινοξικά κατάλοιπα και οι θέσεις τους, τόσο στην αμινοξική ακολουθία, όσο και πάνω στην δομή της πρωτεΐνης (χρωματισμένα με γαλάζιο). Την εσωτερική συστάδα αποτελούν τα αμινοξικά κατάλοιπα F20, W24, F45, M53, A55, F77, F86, M88, H104, I106, A115, Q117, F137 και R139.



Εικόνα 35: Η εξωτερική συστάδα (outer cluster) της λιποκαλίνης bovine retinol binding protein (PDB ID: 1FEN). Φαίνονται τα αμινοξικά κατάλοιπα και οι θέσεις τους, τόσο στην αμινοξική ακολουθία, όσο και πάνω στην δομή της πρωτεΐνης (χρωματισμένα με κόκκινο). Την εξωτερική συστάδα αποτελούν τα αμινοξικά κατάλοιπα A28, I107, V116, Y118, V136, V152 και Y165.

1.6.4 Προγνώσεις δομών

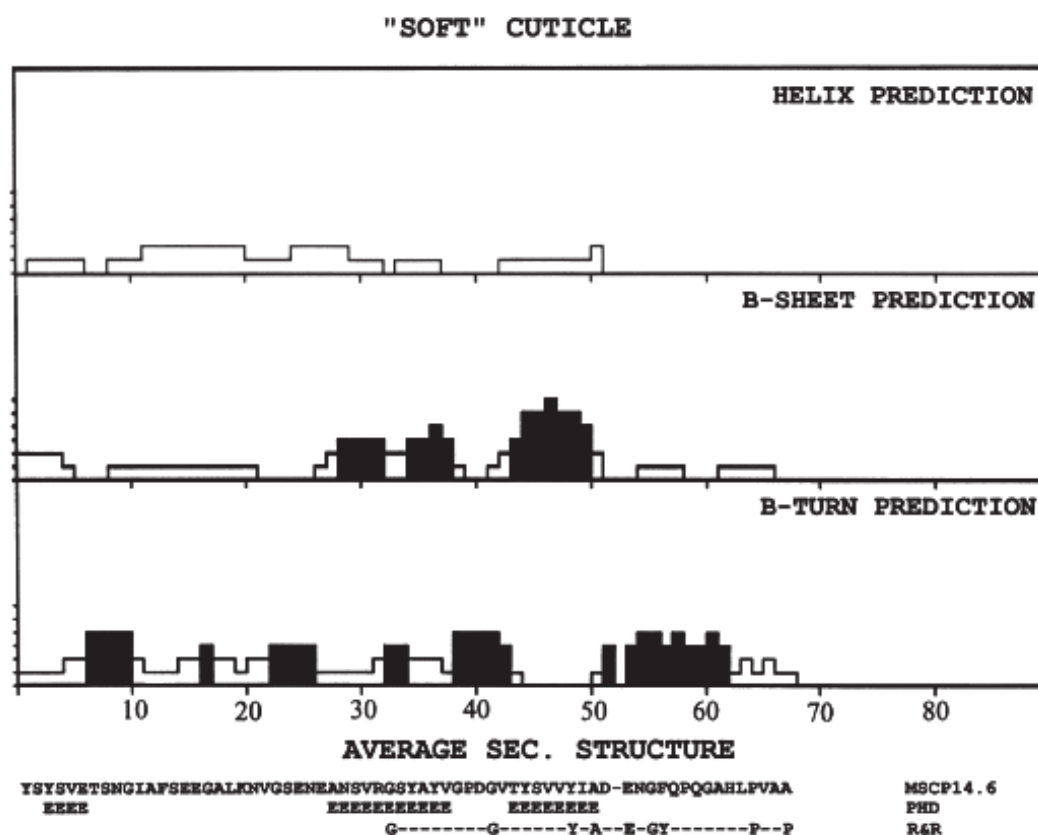
Μια από τις πιο δύσκολες πλευρές στη μελέτη του δερματίου είναι η διαλεύκανση των αλληλεπιδράσεων μεταξύ των πρωτεϊνών του δερματίου και των μη πρωτεϊνικών συστατικών του δερματίου. Για να γίνει εφικτή αυτή η μελέτη πρέπει να γνωρίζουμε την δομή των πρωτεϊνών του δερματίου. Μέχρι σήμερα καμία πρωτεΐνη του δερματίου δεν έχει προσδιορισμένη δομή. Οπότε πρέπει να ακολουθηθεί μια πειραματική διαδικασία πρόγνωσης δομών των πρωτεϊνών του δερματίου (Willis, Papandreou et al. 2012).

Διάφορες προσεγγίσεις έχουν πραγματοποιηθεί για να συγκεντρωθούν πληροφορίες σχετικά με τις αλληλεπιδράσεις των πρωτεϊνών του δερματίου με άλλα συστατικά του. Η πρώτη ήταν να αναλυθούν ακολουθίες πρωτεϊνών του δερματίου, με κατάλληλο λογισμικό, για να προβλεφθεί η δευτεροταγής δομή τους. Η δεύτερη προσέγγιση ήταν να χρησιμοποιηθούν φασματοσκοπικές τεχνικές στα συστατικά του δερματίου για να αντληθούν πληροφορίες σχετικά με τη στερεοδιάταξη των πρωτεϊνικών συστατικών τους *in situ*, και να συγκριθούν πειραματικές πληροφορίες με προβλέψεις. Τρίτον, οι τριτοταγείς δομές των πρωτεϊνών του δερματίου να μοντελοποιηθούν και η τέταρτη προσέγγιση ήταν μια άμεση (*in silico*) πειραματική προσέγγιση (μέσω πειραμάτων αγκυροβόλησης) για να ελεγχθεί αν οι πρωτεΐνες μπορούν (και με ποιόν τρόπο) να αλληλεπιδράσουν με χιτίνη (Willis, Papandreou et al. 2012).

Μέχρι σήμερα προγνώσεις δομών για πρωτεΐνες του δερματίου έχουν πραγματοποιηθεί για πρωτεΐνες της οικογένειας CPR – για τις υποοικογένειες RR-1 (Iconomidou, Willis et al. 1999; Iconomidou, Chrysikos et al. 2001; Hamodrakas, Willis et al. 2002) και RR-2 (Iconomidou, Willis et al. 1999; Iconomidou, Willis et al. 2005) – και για τις πρωτεΐνες της οικογένειας CPF (Papandreou, Iconomidou et al. 2010).

1.6.4.1 Πρόγνωση δομής CPR (RR-1)

Αρχικά, έγινε στοίχιση των πρωτεϊνών της υποοικογένειας RR-1 (πρωτεΐνες CPR του μαλακού δερμάτιου), ώστε να προσδιοριστεί ακριβώς η περιοχή της R&R συναινετικής ακολουθίας (Iconomidou, Willis et al. 1999). Εν συνεχεία, σύμφωνα με τον μέσο όρο των εκάστοτε score του κάθε αμινοξικού καταλοίπου από προηγούμενες προβλέψεις δευτεροταγούς δομής (Chou and Fasman 1974a; Lim 1974a; Chou and Fasman 1974b; Lim 1974b; Dufton and Hider 1977; Nagano 1977a; Nagano 1977b; Garnier, Osguthorpe et al. 1978), προσδιορίστηκε ποιά είναι η πιθανή δομή των πρωτεϊνών (Iconomidou, Willis et al. 1999). Επίσης, έγινε χρήση και του λογισμικού πρόγνωσης δομής PhD (Rost 1996) για επιπλέον πληροφορίες (Iconomidou, Willis et al. 1999). Τα αποτελέσματα όλων των παραπάνω έδειξαν την ύπαρξη τριών (3) β-κλώνων, καμία α-έλικα και τριών (3) στροφών (Iconomidou, Willis et al. 1999). Όλα αυτά τα αποτελέσματα φαίνονται στην εικόνα 36.

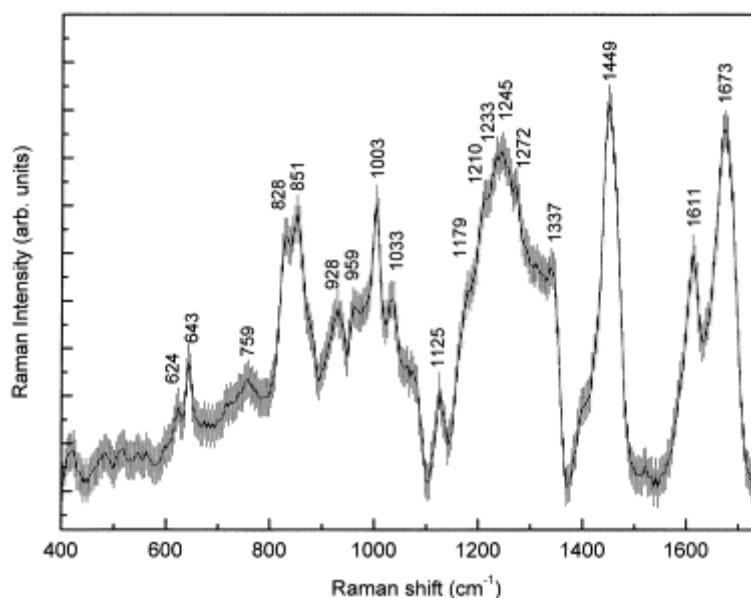


Εικόνα 36: Το σχήμα παρουσιάζει τις προβλέψεις της δευτεροταγούς δομής των πρωτεϊνών CPR RR-1, χρησιμοποιώντας τον μέσο όρο των score του κάθε καταλοίπου από τις ξεχωριστές προηγούμενες προβλέψεις (Chou and Fasman 1974a; Lim 1974a; Chou and Fasman 1974b; Lim 1974b; Dufton and

Hider 1977; Nagano 1977a; Nagano 1977b; Garnier, Osguthorpe et al. 1978) καθώς και του λογισμικού πρόγνωσης PhD (Rost 1996) (κάτω, όπου E είναι β-κλώνος). Τα αποτελέσματα δείχνουν την ύπαρξη τριών (3) β-κλώνων, καμία α-έλικα και τρεις (3) στροφές (Iconomidou, Willis et al. 1999).

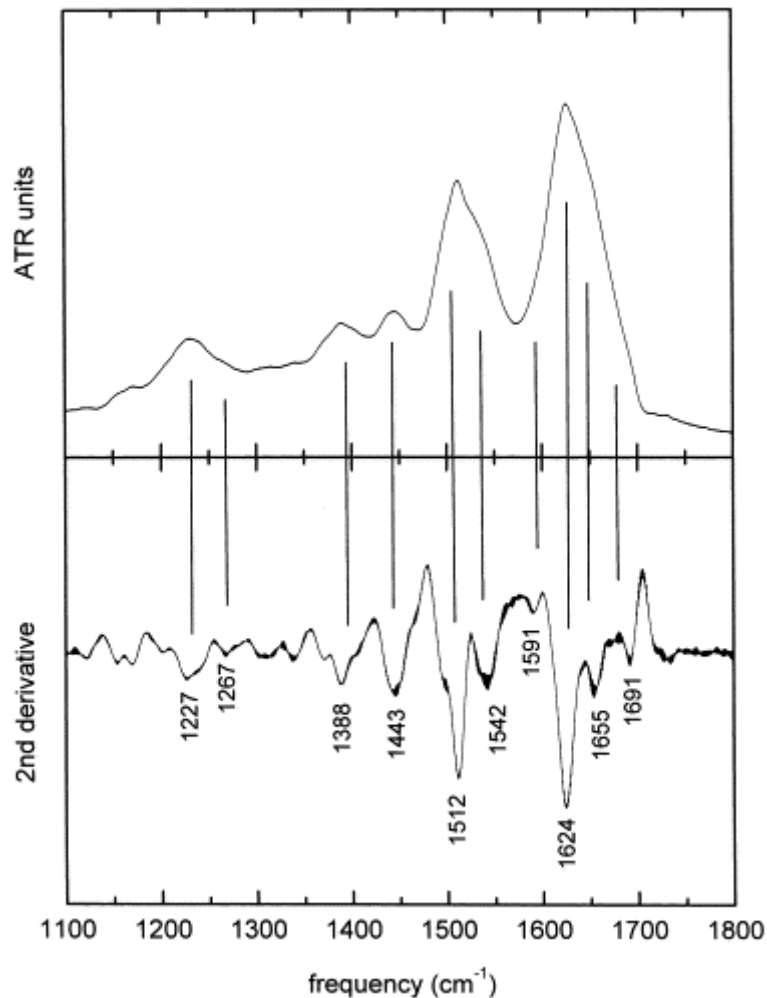
Αξίζει να σημειωθεί ότι η πρόταση ότι οι πρωτεΐνες του δερματίου υιοθετούν μια στερεοδιάταξη β-φύλλου δεν είναι νέα. Υπάρχουν διαθέσιμα πειραματικά δεδομένα, από περίθλαση ακτίνων-X, ότι στην πρωτεΐνη που αλληλεπιδρά με χιτίνη στο δερμάτιο κυριαρχεί η παρουσία β-κλώνων (Fraenkel and Rudall 1947).

Ύστερα, πραγματοποιήθηκαν πειράματα φασματοσκοπίας laser - Raman (FT-Raman), φασματοσκοπίας υπερερύθρου (ATR FT-IR) και κυκλικού διχρωϊσμού (CD) (Iconomidou, Chryssikos et al. 2001). Για τα πειράματα το δερμάτιο προερχόταν από τον οργανισμό *Hyalophora cecropia* και η πρωτεΐνη ήταν η HCCP12. Τα αποτελέσματα έδειξαν ότι η πρωτεΐνη αποτελείται κυρίως από αντιπαράλληλα β-πτυχωτά φύλλα και ότι δεν υπάρχει α-έλικα (Iconomidou, Chryssikos et al. 2001). Ωστόσο, στα αποτελέσματα του κυκλικού διχρωϊσμού η πρόγνωση ήταν 54% β-φύλλο, ~13% α-έλικα, 24% β-στροφές και 9% τυχαία δομή (Iconomidou, Chryssikos et al. 2001). Τα αποτελέσματα των παραπάνω πειραμάτων παρουσιάζονται στις εικόνες 37, 38 και 39.

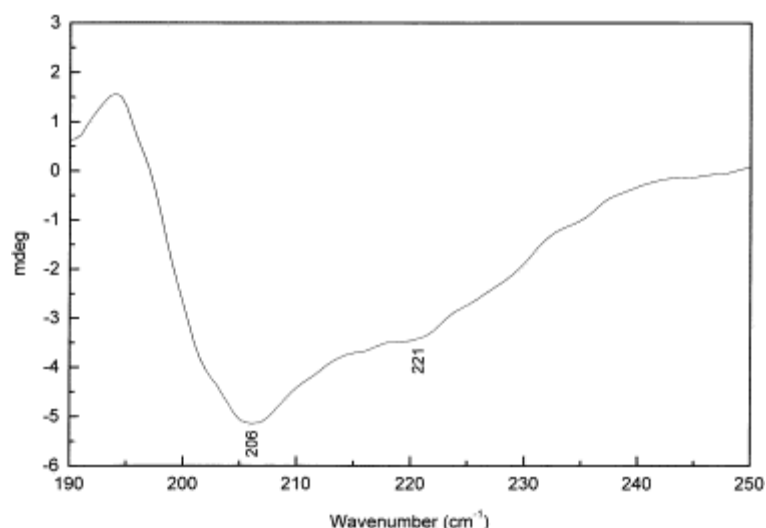


Εικόνα 37: Αποτέλεσμα φασματοσκοπίας laser - Raman (FT-Raman). Η εικόνα δείχνει το φάσμα που παρουσιάζει τις διαφορές μεταξύ των φασμάτων του «φυσικού» δερματίου και του δερματίου από το οποίο έχουν αφαιρεθεί οι πρωτεΐνες. Τα φάσματα πριν από την αφαίρεση κανονικοποιήθηκαν στα 952

cm^{-1} στην ένταση φασματικής υπογραφής λόγω της χιτίνης. Ένα σχεδόν ταυτόσημο φάσμα διαφοράς προέκυψε όταν η κανονικοποίηση έγινε πριν από την αφαίρεση, με βάση την ένταση των 709 cm^{-1} στην φασματική υπογραφή, η οποία επίσης οφείλεται στην χιτίνη. Το καθαρό αποτέλεσμα, σε αμφότερες τις περιπτώσεις, είναι ένα φάσμα διαφοράς σχεδόν εξ ολοκλήρου λόγω της πρωτεΐνης (Iconomidou, Chryssikos et al. 2001).



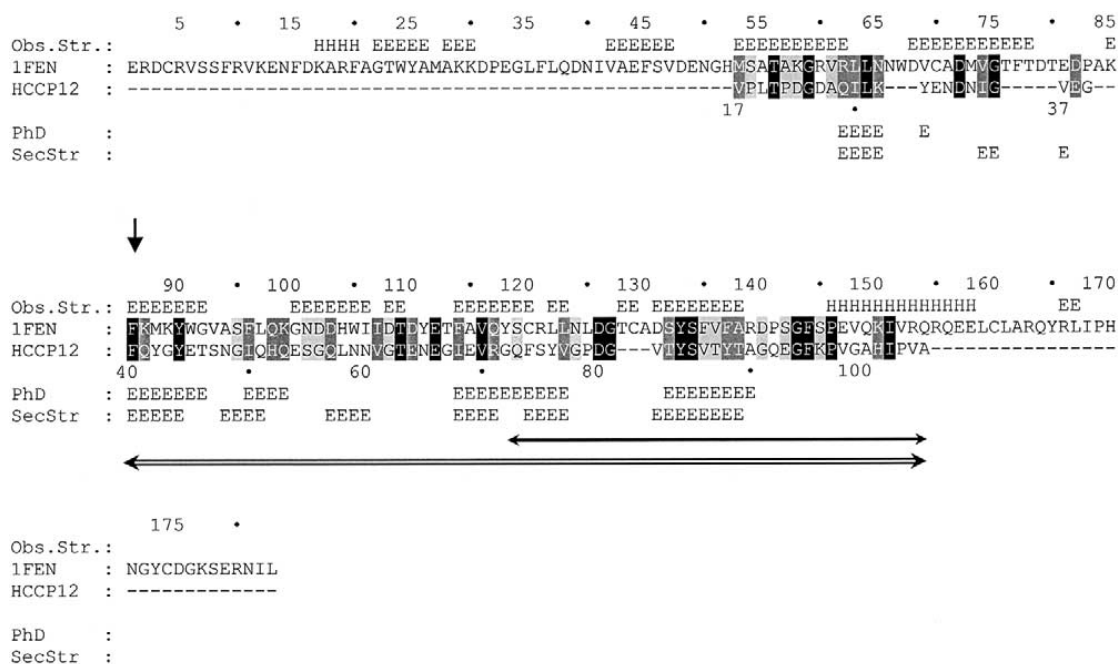
Εικόνα 38: Εξασθενημένο φάσμα ολικής ανάκλασης FT-υπερύθρου (ATR FT-IR) των πρωτεϊνών του δερματίου που προέρχονται από τον *Hyalophora cecropia*. Η δεύτερη παράγωγος του φάσματος περιλαμβάνεται στην εικόνα (Iconomidou, Chryssikos et al. 2001).



Εικόνα 39: Το φάσμα που προέκυψε από το πείραμα του κυκλικού διχρωϊσμού (CD) με πρωτεΐνες του δερματίου του οργανισμού *Hyalophora cecropia*. Το αποτέλεσμα δείχνει ότι η δευτεροταγής δομής των πρωτεϊνών είναι 54% β-φύλλο, ~13% α-έλικα, 24% β-στροφές και 9% τυχαία δομή (Iconomidou, Chryssikos et al. 2001).

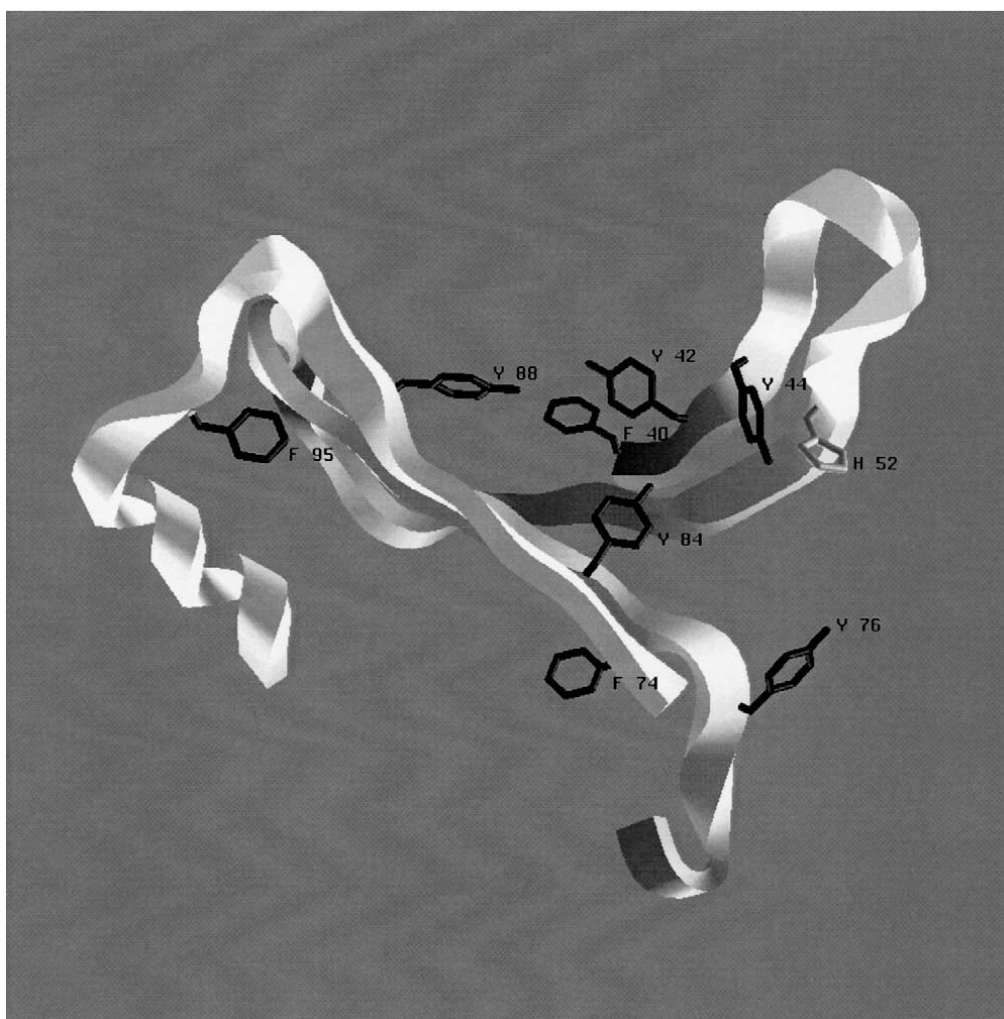
Όλα τα παραπάνω αποτελέσματα είναι σε συμφωνία με προηγούμενα πειραματικά δεδομένα (Iconomidou, Chryssikos et al. 2001), με τα οποία είχε προβλεφθεί ότι μια δομή αντιπαράλληλου β-πτυχωτού φύλλου είναι η επικρατέστερη δομή για αλληλεπίδραση των πρωτεϊνών του δερματίου με α-χιτίνη (Atkins 1985).

Αργότερα, παρατηρήθηκε αναπάντεχα (όπως έχει αναφερθεί και στην ενότητα 1.6.3) ομοιότητα των πρωτεϊνών της οικογένειας CPR με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN) (Hamodrakas, Willis et al. 2002). Η ομοιότητα βασίστηκε στην στοίχιση της CPR RR-1 πρωτεΐνης του δερματίου HCCP12 (από τον οργανισμό *Hyalophora cecropia*) με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN) (εικόνα 40) (Hamodrakas, Willis et al. 2002).



Εικόνα 40: Η στοίχιση της CPR RR-1 πρωτεΐνης του δερματίου HCCP12 (από τον οργανισμό *Hyalophora cecropia*) με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN) με την χρήση του ClustalW (Thompson, Higgins et al. 1994). Κάτω από τη στοίχιση φαίνονται και τα αποτελέσματα των προγραμμάτων πρόγνωσης PhD (Rost 1996) και SecStr (Hamodrakas 1988), όπου με E είναι η πρόγνωση για β-κλώνο. Με μονή υπογράμμιση φαίνεται η R&R συναινετική ακολουθία (R&R consensus), ενώ με διπλή υπογράμμιση φαίνεται η εκτεταμένη R&R συναινετική ακολουθία (extended R&R consensus) της CPR πρωτεΐνης. Με βέλος σημειώνεται το πρώτο αμινοξικό κατάλοιπο όπου ξεκινάει το δομικό μοντέλο που φτιάχτηκε ως αποτέλεσμα της ομοιότητας των δυο ακολουθιών (Hamodrakas, Willis et al. 2002).

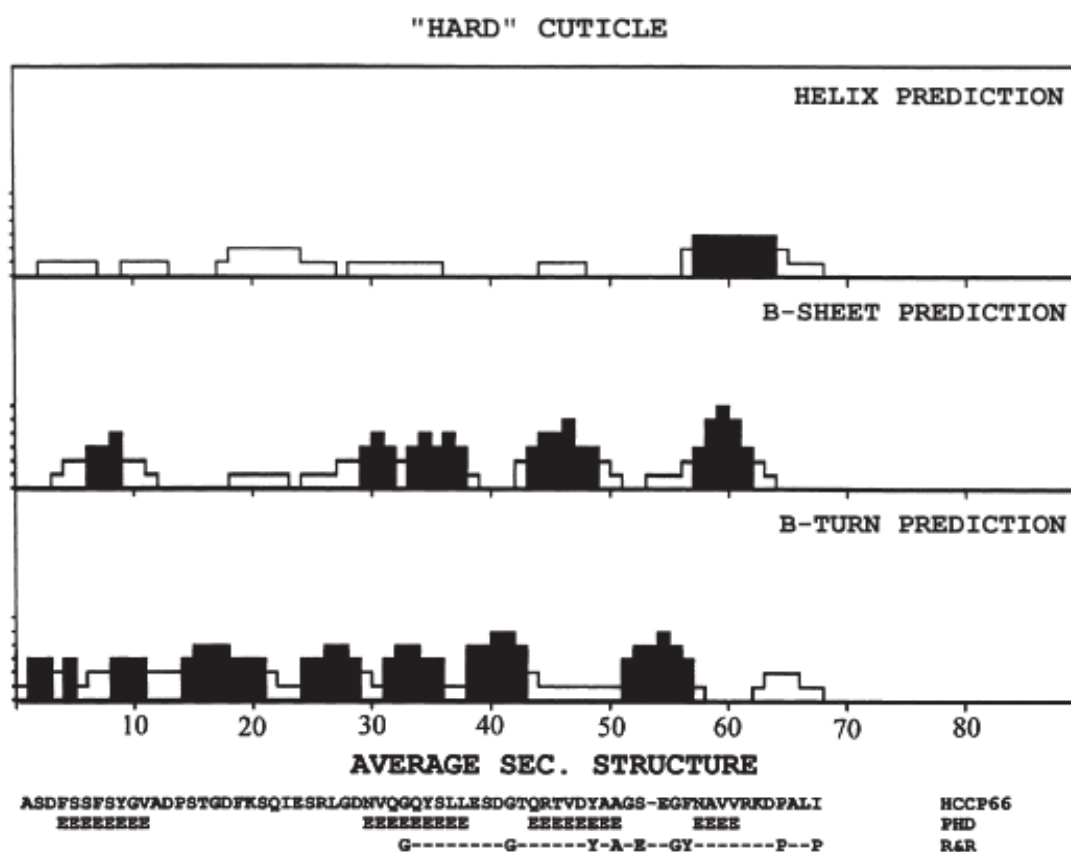
Ως αποτέλεσμα της παραπάνω ομοιότητας, με χρήση της μεθόδου προτυποποίησης με βάση την ομολογία και χρησιμοποιώντας το λογισμικό WHAT IF (Vriend 1990), δημιουργήθηκε το δομικό μοντέλο της πρωτεΐνης HCCP12 της οικογένειας CPR RR-1 (εικόνα 41) (Hamodrakas, Willis et al. 2002). Πρόκειται για δομή μισού βαρελιού με αντιπαράλληλους β-κλώνους που περιέχει χαρακτηριστικά κατάλληλα για αλληλεπίδραση με χιτίνη (Hamodrakas, Willis et al. 2002). Στη σχισμή του μισού βαρελιού υπάρχει μια πληθώρα αρωματικών αμινοξικών καταλοίπων, τα οποία δημιουργούν ιδανικές συνθήκες για αλληλεπίδραση πρωτεΐνης – χιτίνης (Hamodrakas, Willis et al. 2002).



Εικόνα 41: Το δομικό μοντέλο της πρωτεΐνης HCCP12 που ανήκει στην οικογένεια CPR RR-1 μετά από προτυποποίηση με βάση την ομολογία. Είναι μια δομή μισού βαρελιού με αντιπαράλληλους β-κλώνους. Επίσης, είναι εμφανής η παρουσία αρωματικών αμινοξικών καταλοίπων (Y – τυροσίνη, F – φαινυλαλανίνη, H - ιστιδίνη) στη σχισμή του μισού βαρελιού, που παρέχουν ένα ιδανικό περιβάλλον για αλληλεπίδραση πρωτεΐνης – χιτίνης (Hamodrakas, Willis et al. 2002).

1.6.4.2 Πρόγνωση δομής CPR (RR-2)

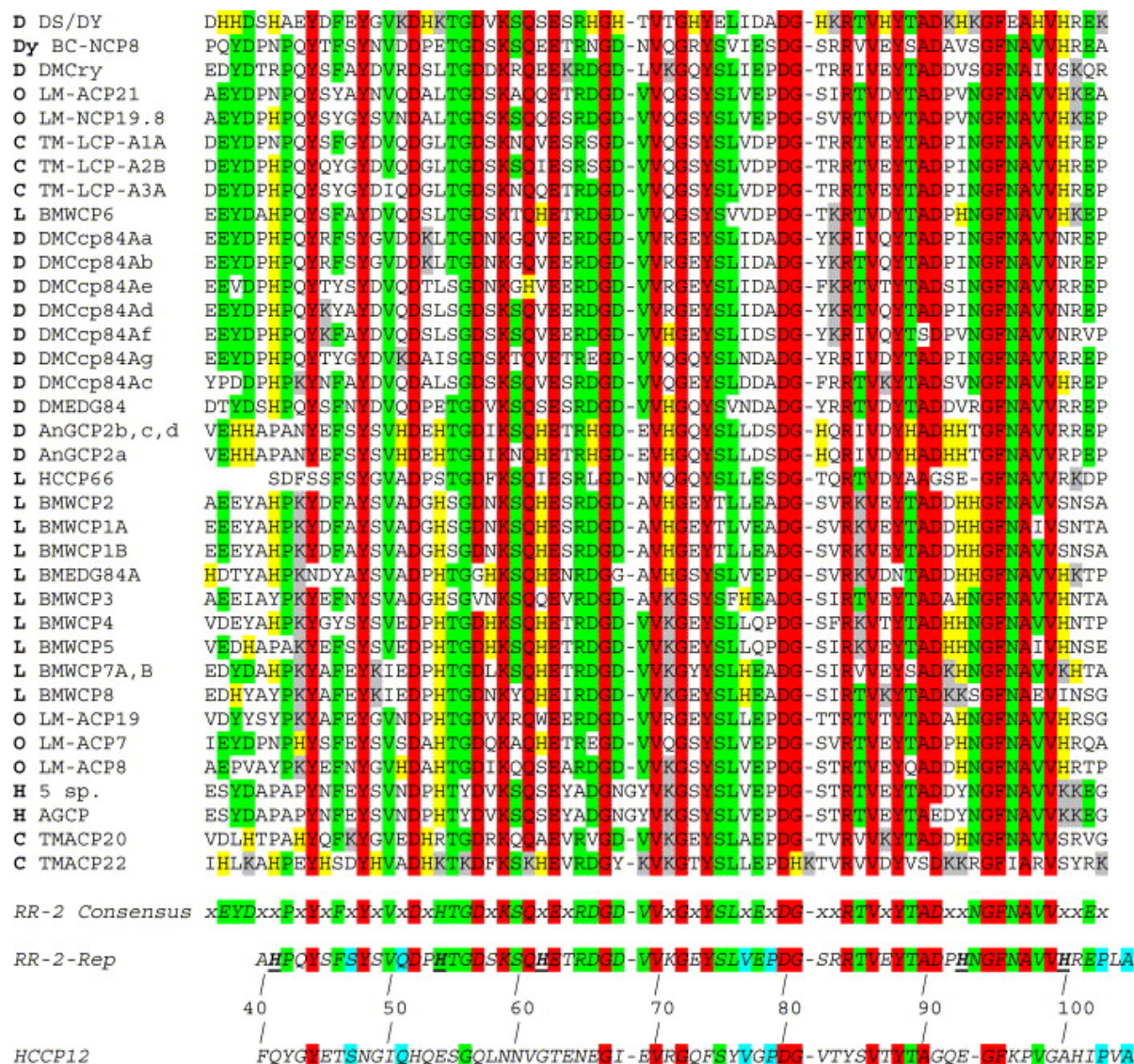
Η διαδικασία που ακολουθήθηκε για τις RR-1 πραγματοποιήθηκε και για τις πρωτεΐνες της RR-2 υποοικογένειας (οι οποίες όπως έχει αναφερθεί αφορά πρωτεΐνες CPR που υπάρχουν στο σκληρό δερμάτιο). Αρχικά, έγινε στοίχιση των πρωτεϊνών της υποοικογένειας RR-2 ώστε να προσδιοριστεί ακριβώς η περιοχή της R&R συναινετικής ακολουθίας (Iconomidou, Willis et al. 1999). Εν συνεχεία, σύμφωνα με τον μέσο όρο των εκάστοτε score του κάθε αμινοξικού καταλοίπου από προηγούμενες προβλέψεις της δευτεροταγούς δομής (Chou and Fasman 1974a; Lim 1974a; Chou and Fasman 1974b; Lim 1974b; Dufton and Hider 1977; Nagano 1977a; Nagano 1977b; Garnier, Osguthorpe et al. 1978), προσδιορίστηκε ποιά είναι η πιθανή δομή των πρωτεϊνών CPR RR-2 (Iconomidou, Willis et al. 1999). Επίσης, έγινε χρήση και του λογισμικού πρόγνωσης δομής PhD (Rost 1996) για επιπλέον πληροφορίες (Iconomidou, Willis et al. 1999). Τα αποτελέσματα όλων των παραπάνω έδειξαν την ύπαρξη τεσσάρων (4) β-κλώνων, καμία α-έλικα και τεσσάρων (4) στροφών (Iconomidou, Willis et al. 1999). Όλα αυτά τα αποτελέσματα φαίνονται στην εικόνα 42.



Εικόνα 42: Το σχήμα παρουσιάζει τις προβλέψεις της δευτεροταγούς δομής, χρησιμοποιώντας τον μέσο όρο των score του κάθε καταλοίπου από τις ξεχωριστές προηγούμενες προβλέψεις (Chou and Fasman 1974a; Lim 1974a; Chou and Fasman 1974b; Lim 1974b; Dufton and Hider 1977; Nagano 1977a; Nagano 1977b; Garnier, Osguthorpe et al. 1978) καθώς και του λογισμικού πρόγνωσης PhD (Rost 1996) (κάτω, όπου E είναι β-κλώνος). Τα αποτελέσματα δείχνουν την ύπαρξη τεσσάρων (4) β-κλώνων, καμία α-έλικα και τεσσάρων (3) στροφών (Iconomidou, Willis et al. 1999).

Όπως και στις πρωτεΐνες CPR RR-1, έτσι και στις πρωτεΐνες CPR RR-2 τα πειραματικά δεδομένα που υιοθετούν μια στερεοδιάταξη β-φύλλου δεν είναι νέα και είναι σε συμφωνία με προηγούμενα πειραματικά αποτελέσματα. Υπάρχουν διαθέσιμα αποδεικτικά πειραματικά στοιχεία, από περίθλαση ακτίνων-X, ότι στην πρωτεΐνη που αλληλεπιδρά με χιτίνη στο δερμάτιο κυριαρχεί η παρουσία β-κλώνων (Fraenkel and Rudall 1947).

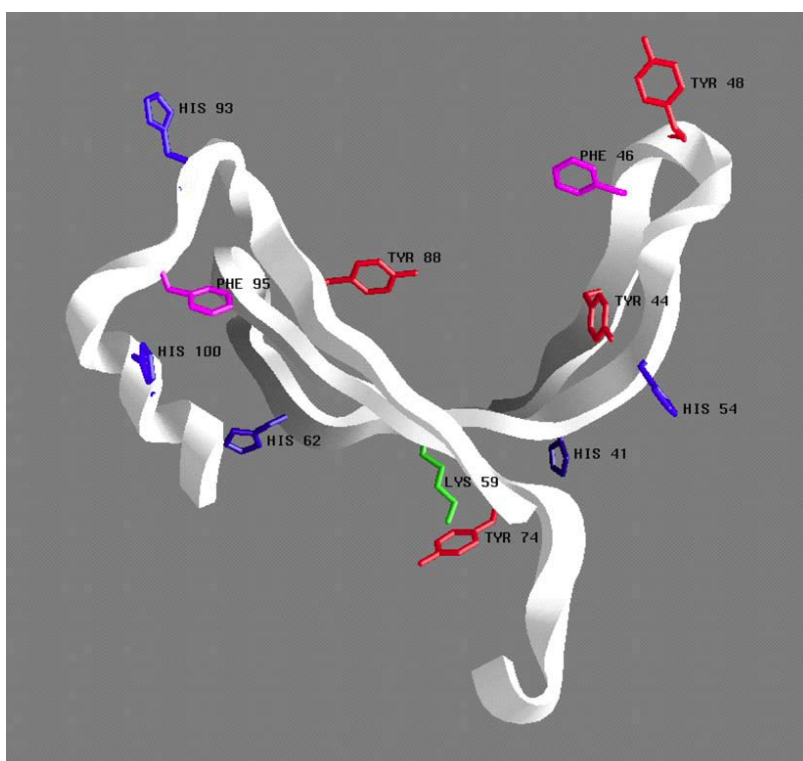
Όπως έχει ειπωθεί, στη βάση δεδομένων Pfam (Punta, Coggill et al. 2012) η συναινετική ακολουθία R&R είναι καταχωρημένη με τις ονομασίες pfam00379 και chitin_bind_4 και αποτελείται από 68 αμινοξικά κατάλοιπα (Willis 2010). Πραγματοποιήθηκε στοίχιση της περιοχής που αντιστοιχεί στο pfam00379, χρησιμοποιώντας 44 πρωτεΐνες του δερματίου με την RR-2 συναινετική ακολουθία, και έτσι «δημιουργήθηκε» η RR-2 - Rep. Εν συνεχεία, η RR-2 - Rep στοιχίστηκε με την πρωτεΐνη HCCP12 της υποοικογένειας RR-1 (Iconomidou, Willis et al. 2005). Τα αποτελέσματα φαίνονται στην εικόνα 43.



Εικόνα 43: Τα αποτελέσματα της στοίχισης, με χρήση του ClustalW (Thompson, Higgins et al. 1994), της περιοχής pfam00379, 44 πρωτεϊνών της υποοικογένειας RR-2 των CPR. Αριστερά από τα ονόματα των πρωτεϊνών συμβολίζεται η τάξη αρθροπόδων από όπου προέρχεται (D – δίπτερα, O – ορθόπτερα, L – λεπιδόπτερα, H – ημίπτερα, C – κολεόπτερα και Dy - δικτυόπτερα). Με κόκκινο χρώμα συμβολίζονται αμινοξικά κατάλοιπα που υπάρχουν τουλάχιστον στο 95% των περιπτώσεων, με πράσινο συμβολίζεται το αμινοξικό κατάλοιπο που υπερτερεί στη θέση αυτή. Με κίτρινο είναι χρωματισμένες οι ιστιδίνες (I) και με ανοιχτό γκρι οι λυσίνες (K). Η RR-2 συναινετική ακολουθία, καθώς και η RR-2 – Rep που δημιουργήθηκε, βρίσκονται κάτω από τις στοιχισμένες πρωτεΐνες. Τα x αμινοξικά κατάλοιπα της RR-2 έχουν αντικατασταθεί στην RR-2 - Rep με το επικρατέστερο. Τελευταία κάτω εμφανίζεται η ακολουθία της HCCP12 (RR-1, που προέρχεται από το μαλακό δερμάτιο) και με γαλάζιο έχουν χρωματιστεί τα κοινά αμινοξικά κατάλοιπα αυτής με την RR-2 - Rep (Iconomidou, Willis et al. 2005).

Εν συνεχεία, έχοντας ως δομή-οδηγό (template) την προβλεφθείσα δομή της HCCP12 (Hamodrakas, Willis et al. 2002), προβλέφθηκε η δομή της RR-2 – Rep

(Iconomidou, Willis et al. 2005). Η πρόγνωση έγινε με την μέθοδο της προτυποποίησης με βάση την ομολογία και με χρήση του λογισμικού WHAT IF (Vriend 1990). Το μοντέλο της RR-2 – Rep που δημιουργήθηκε είναι μια δομή αντιπαράλληλων β-κλώνων που δημιουργούν μια δομή μισού βαρελιού (Iconomidou, Willis et al. 2005). Στη σχισμή του βαρελιού υπάρχουν πολλά αρωματικά αμινοξικά κατάλοιπα (Y – τυροσίνη, F – φαινυλαλανίνη, H - ιστιδίνη) που δημιουργούν ιδανικές συνθήκες για αλληλεπίδραση πρωτεΐνης – χιτίνης (Iconomidou, Willis et al. 2005) (εικόνα 44). Επίσης, η παρουσία πολλών ιστιδινών (I) είναι χαρακτηριστικό των πρωτεϊνών της RR-2 υποοικογένειας, που πιθανότατα σχετίζεται με την σκλήρυνση (sclerotization) του δερμάτιου. Εναλλακτικά, θα μπορούσαν να εμπλέκονται σε μεταβολές της ικανότητας δέσμευσης ύδατος στο δερμάτιο, επειδή μικρές μεταβολές του pH μπορούν να επηρεάσουν τον ιονισμό των ομάδων ιμιδαζολίου (Andersen, Hojrup et al. 1995; Iconomidou, Willis et al. 2005; Willis, Papandreou et al. 2012).

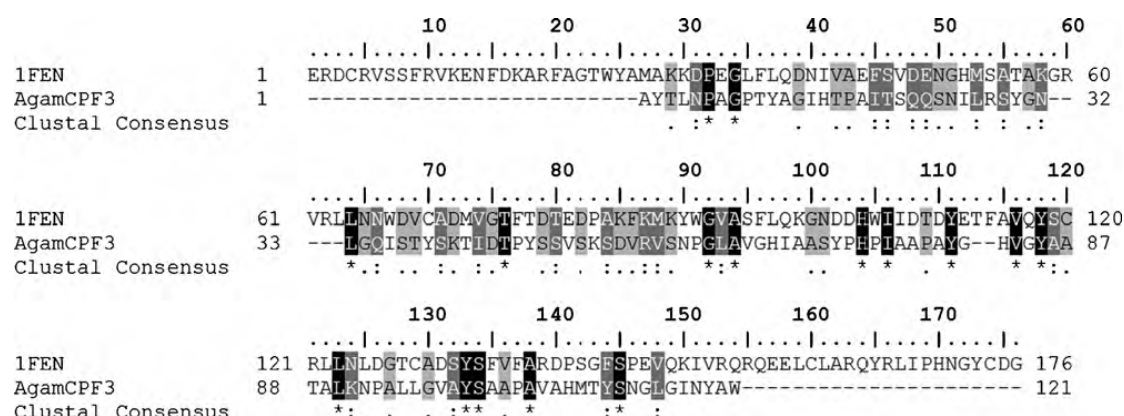


Εικόνα 44: Το δομικό μοντέλο της RR-2 – Rep, δηλαδή των πρωτεϊνών της υποοικογένειας RR-2 (σκλήρο δερμάτιο) των CPR. Η δομή αποτελείται από αντιπαράλληλους β-κλώνους που σχηματίζουν ένα μισό βαρέλι. Στη σχισμή του βαρελιού υπάρχουν πολλά αρωματικά αμινοξικά κατάλοιπα (Y – τυροσίνη, F – φαινυλαλανίνη, H - ιστιδίνη) που δημιουργούν ιδανικές συνθήκες για αλληλεπίδραση πρωτεΐνης – χιτίνης. Επίσης, η παρουσία πολλών ιστιδινών (I) είναι χαρακτηριστικό των πρωτεϊνών της RR-2 υποοικογένειας που πιθανότατα σχετίζεται με την σκλήρυνση (sclerotization) του δερμάτιου (Iconomidou, Willis et al. 2005).

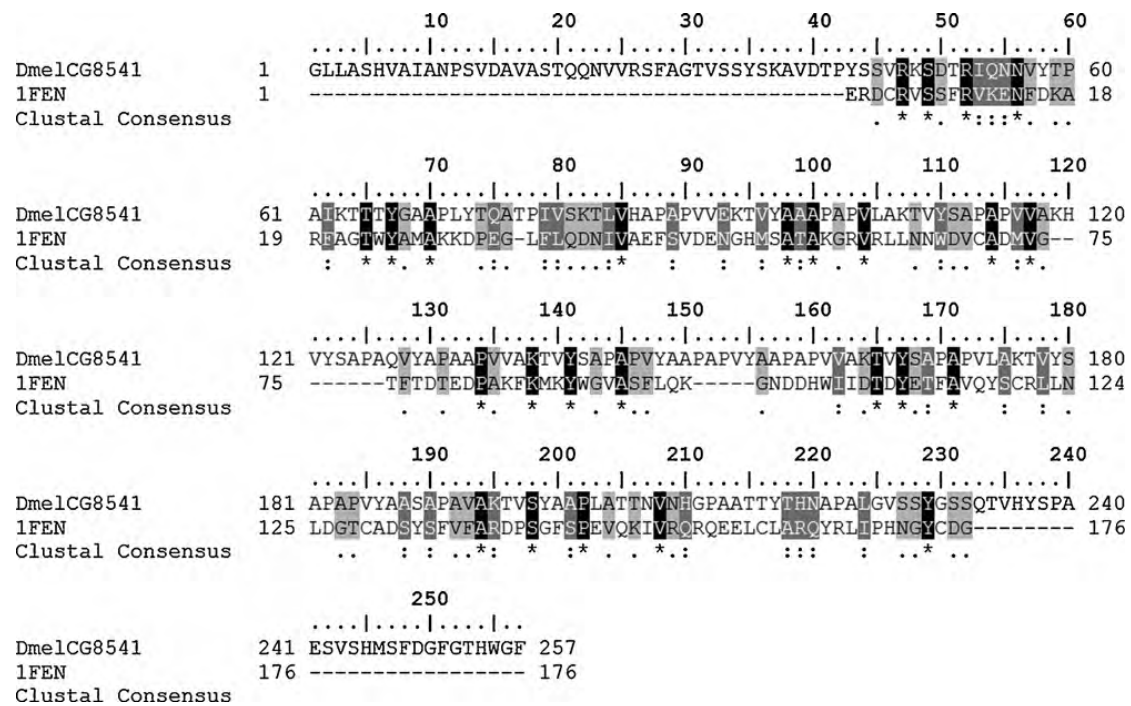
1.6.4.3 Πρόγνωση δομής CPF

Επόμενη οικογένεια που έγινε πρόγνωση δομής ήταν η CPF. Οι πρωτεΐνες της οικογένειας CPF παρουσιάζουν και αυτές μεγάλη ομοιότητα με την λιποκαλίνη bovine retinol binding protein (RBP) (PDB ID: 1FEN) (Papandreou, Iconomidou et al. 2010).

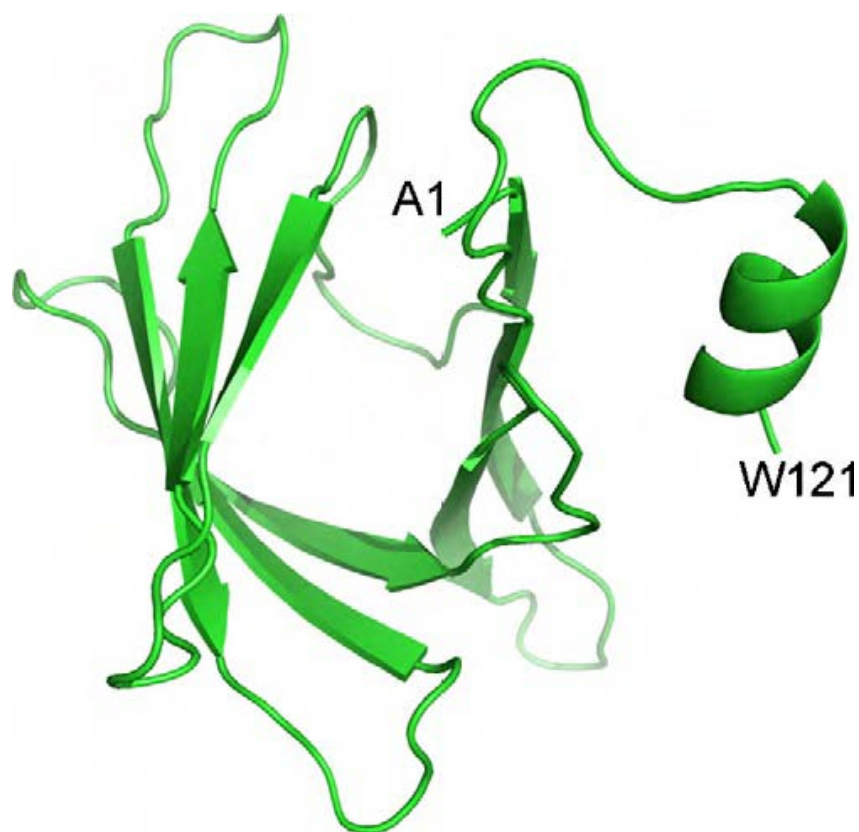
Τα πρωτεϊνικά μοντέλα που δημιουργήθηκαν βασίστηκαν στην ομοιότητα της λιποκαλίνης bovine RBP (PDB ID: 1FEN) με τις πρωτεΐνες AgamCPF3 (ENTREZ accession number: 118790289) (Togawa, Augustine Dunn et al. 2007), από τον *Anopheles gambiae*, και μιας ομόλογης (homolog) CPF, της CG8541 (ENTREZ accession number: 7295230) (Togawa, Augustine Dunn et al. 2007), από την *Drosophila melanogaster* (Papandreou, Iconomidou et al. 2010). Τα αποτελέσματα των στοιχίσεών τους παρουσιάζονται στις εικόνες 45 και 46.



Εικόνα 45: Τα αποτελέσματα της στοιχίσης, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), της CPF πρωτεΐνης AgamCPF3 (από τον *Anopheles gambiae*) (Togawa, Augustine Dunn et al. 2007) με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN). Η ταυτότητα των δυο ακολουθιών είναι 13%, ενώ η ομοιότητα είναι περίπου 50% (Papandreou, Iconomidou et al. 2010).



Στη συνέχεια, με τη μέθοδο της προτυποποίησης με βάση την ομολογία, χρησιμοποιώντας το λογισμικό Modeller (Sali and Blundell 1993), δημιουργήθηκε το δομικό μοντέλο της CPF πρωτεΐνης AgamCPF3, έχοντας ως δομή-οδηγό (template) την προσδιορισμένη κρυσταλλογραφικά δομή της λιποκαλίνης bovine retinol binding protein (RBP) (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)) (Papandreou, Iconomidou et al. 2010). Το μοντέλο της δομής των CPF πρωτεϊνών προγνώσθηκε ότι είναι ένα ολόκληρο β-βαρέλι με οχτώ (8) αντιπαράλληλους β-κλώνους (εικόνα 47) (Papandreou, Iconomidou et al. 2010).



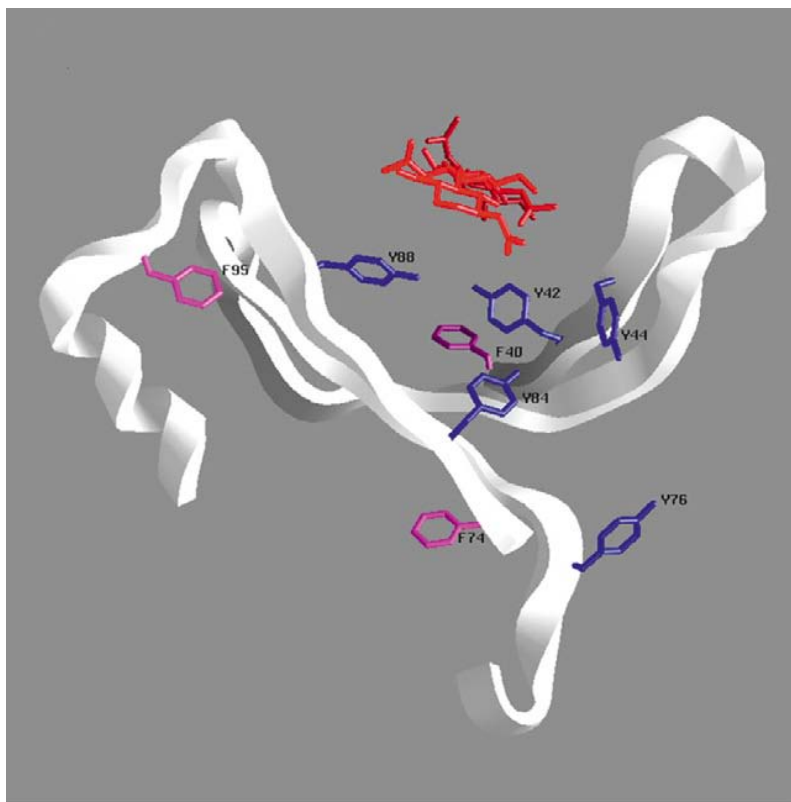
Εικόνα 47: Το δομικό μοντέλο της CPF πρωτεΐνης AgamCPF3. Πρόκειται κυρίως για οχτώ (8) αντιπαράλληλους β-κλώνους που σχηματίζουν ένα ολόκληρο άνω-κάτω β-βαρέλι (Papandreou, Iconomidou et al. 2010).

1.6.5 Αλληλεπιδράσεις με χιτίνη (CPR RR-1, CPR RR-2, CPF)

Το επόμενο βήμα ήταν πειράματα αγκυροβόλησης (docking) ώστε να μελετηθεί αν τα πρωτεϊνικά μοντέλα αλληλεπιδρούν με τη χιτίνη, καθώς και με ποιό τρόπο γίνεται αυτή η αλληλεπίδραση.

Τα πρώτα πειράματα πραγματοποιήθηκαν μεταξύ του μοντέλου της πρωτεΐνης HCCP12 (από τον *Hyalophora cecropia*), που ανήκει στην οικογένεια CPR RR-1 (μαλακό δερμάτιο), και αλυσίδων χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG) (Hamodrakas, Willis et al. 2002). Το πείραμα πραγματοποιήθηκε με χρήση του λογισμικού GRAMM (Vakser 1996) ήταν χαμηλής ανάλυσης (low resolution) αγκυροβόληση με τις προκαθορισμένες (default) παραμέτρους του προγράμματος. Το

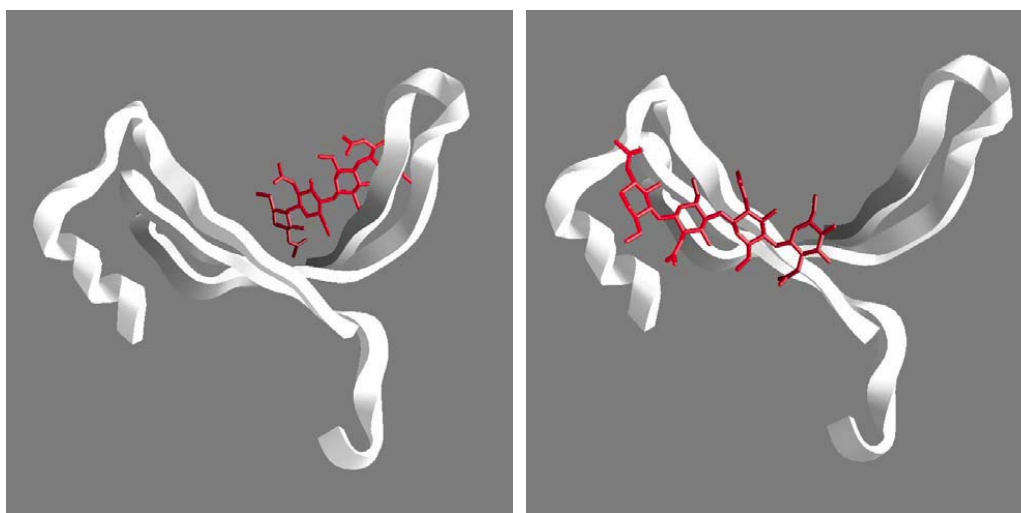
αποτέλεσμα έδειξε ότι το μοντέλο μπορεί να αλληλεπιδράσει με τουλάχιστον μία εκτεταμένη αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG), η οποία αλληλεπιδρά κάθετα προς τους β-κλώνους του μισού βαρελιού (εικόνα 48) (Hamodrakas, Willis et al. 2002).



Εικόνα 48: Το αποτέλεσμα του πειράματος αγκυροβόλησης (docking), με χρήση του λογισμικού GRAMM (Vakser 1996), μεταξύ του μοντέλου της πρωτεΐνης HCCP12 (που ανήκει στις CPR RR-1) και αλυσίδας χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG). Το μοντέλο της πρωτεΐνης είναι χρωματισμένο με λευκό χρώμα ενώ η αλυσίδα χιτίνης 4NAG με κόκκινο. Όπως φαίνεται η αλυσίδα χιτίνης αλληλεπιδρά κάθετα προς τους β-κλώνους του μισού βαρελιού. Επίσης στην εικόνα φαίνονται τα αρωματικά αμινοξικά κατάλοιπα (Y – τυροσίνη με μπλε, F – φαινυλαλανίνη με ροζ) που υπάρχουν στη σχισμή του μισού βαρελιού, τα οποία δημιουργούν ιδανικό περιβάλλον για αλληλεπίδραση με χιτίνη (Hamodrakas, Willis et al. 2002; Willis, Papandreou et al. 2012).

Στη συνέχεια πραγματοποιήθηκαν πειράματα αγκυροβόλησης (docking) μεταξύ του μοντέλου της υποθετικής πρωτεΐνης RR-2 – Rep, που «εκπροσωπεί» τις πρωτεΐνες της οικογένειας CPR RR-2, και αλυσίδων χιτίνης (ομοτετραμερών N-ακετυλογλυκοζαμίνης - 4NAG) (Iconomidou, Willis et al. 2005). Το πείραμα πραγματοποιήθηκε με χρήση του λογισμικού GRAMM (Vakser 1996). Το αποτέλεσμα έδειξε ότι οι αλυσίδες χιτίνης (4NAG) μπορούν να αλληλεπιδράσουν παράλληλα στους β-κλώνους του μισού β-βαρελιού (εικόνα 49) (Iconomidou, Willis

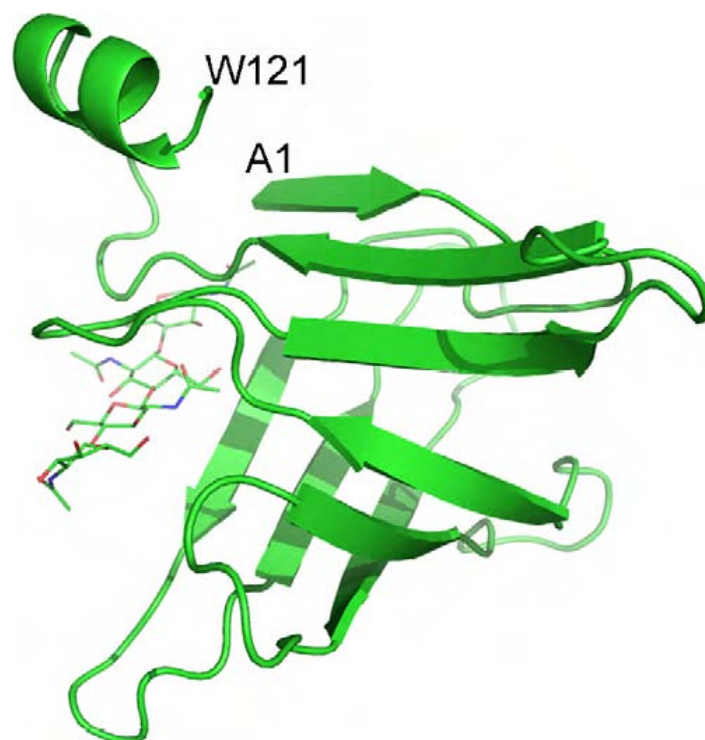
et al. 2005). Έτσι, τα β-βαρέλια των πρωτεϊνών του δερματίου μπορούν να παρεμβαίνουν μεταξύ των μεγάλων αλυσίδων χιτίνης στο δερμάτιο, χωρίς να διαταράσσεται η συνέχεια (Iconomidou, Willis et al. 2005; Willis, Papandreou et al. 2012). Αυτή η παράλληλη διάταξη των β-κλώνων των πρωτεϊνών του δερματίου με τις αλυσίδες χιτίνης είναι σε συμφωνία με προηγούμενες παρατηρήσεις από περιθλασιγράμματα ακτίνων X (Atkins 1985).



Εικόνα 49: Δύο (2) αποτελέσματα του πειράματος αγκυροβόλησης (docking), μεταξύ του μοντέλου της υποθετικής πρωτεΐνης RR-2 – Rep, που «εκπροσωπεί» τις πρωτεΐνες της οικογένειας CPR RR-2, και αλυσίδων χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG). Το πείραμα πραγματοποιήθηκε με χρήση του λογισμικού GRAMM (Vakser 1996). Το μοντέλο της RR-2 – Rep είναι χρωματισμένο με λευκό ενώ η αλυσίδα χιτίνης είναι χρωματισμένη με κόκκινο χρώμα. Τα αποτελέσματα δείχνουν ότι οι αλυσίδες χιτίνης (4NAG) μπορούν να αλληλεπιδράσουν παράλληλα στους β-κλώνους του μισού β-βαρελιού (Iconomidou, Willis et al. 2005).

Τέλος, πραγματοποιήθηκαν πειράματα αγκυροβόλησης μεταξύ του μοντέλου της πρωτεΐνης AgamCPF3 (από τον *Anopheles gambiae*), που ανήκει στην οικογένεια CPF και αλυσίδων χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG) (Papandreou, Iconomidou et al. 2010). Το πείραμα πραγματοποιήθηκε με χρήση του λογισμικού GRAMM (Vakser 1996) ήταν υψηλής ανάλυσης (high resolution) αγκυροβόληση με τις προκαθορισμένες (default) παραμέτρους του προγράμματος. Τα αποτελέσματα έδειξαν ότι το ομοτετραμερές χιτίνης (4NAG) δεν προσαρμόζεται μέσα στο θύλακα αλληλεπίδραση (binding pocket) των CPF, μάλλον, οι πρωτεΐνες CPF μπορούν να αλληλεπιδρούν «χαλαρά» με τις αλυσίδες χιτίνης, με τους β-κλώνους να εκτείνονται παράλληλα προς τις αλυσίδες χιτίνης (εικόνα 50) (Papandreou, Iconomidou et al. 2010), σε συμφωνία με προηγούμενα πειραματικά

δεδομένα (Atkins 1985). Περαιτέρω απόδειξη ότι οι πρωτεΐνες CPF δεν μπορούν να αλληλεπιδράσουν άμεσα με χιτίνη προέρχεται από την αποτυχία ανασυνδυασμένων πρωτεϊνών της οικογένειας CPF να αλληλεπιδράσουν με χιτίνη (Togawa, Augustine Dunn et al. 2007).



Εικόνα 50: Αποτέλεσμα του πειράματος αγκυροβόλησης (docking) μεταξύ του μοντέλου της πρωτεΐνης AgamCPF3 (από τον *Anopheles gambiae*), που ανήκει στην οικογένεια CPF και αλυσίδων χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG). Το πείραμα πραγματοποιήθηκε με χρήση του λογισμικού GRAMM (Vakser 1996) ήταν υψηλής ανάλυσης (high resolution) αγκυροβόληση με τις προκαθορισμένες (default) παραμέτρους του προγράμματος. Η πρωτεΐνη AgamCPF3 είναι χρωματισμένη με πράσινο χρώμα με μορφή κορδέλας ενώ η αλυσίδα χιτίνης (4NAG) είναι με την μορφή ball and stick. Τα αποτελέσματα έδειξαν ότι το τετραμερές χιτίνης (4NAG) δεν προσαρμόζεται μέσα στο θύλακα αλληλεπίδραση (binding pocket) των CPF, μάλλον, οι πρωτεΐνες CPF μπορούν να αλληλεπιδρούν «χαλαρά» με τις αλυσίδες χιτίνης, με τους β-κλώνους να εκτείνονται παράλληλα προς τις αλυσίδες χιτίνης (Papandreou, Iconomidou et al. 2010).

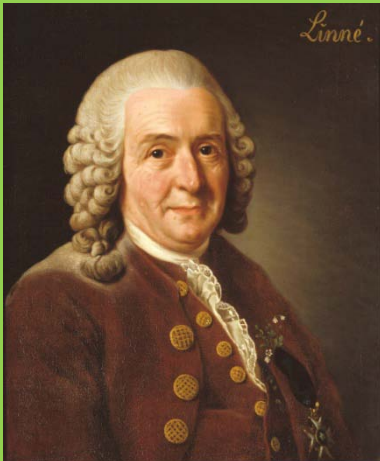
1.7 Στόχοι Διπλωματικής

Οι στόχοι της παρούσας διπλωματικής ήταν:

Η μελέτη των δομικών πρωτεϊνών του δερματίου των αρθροπόδων και οι αλληλεπιδράσεις αυτών με χιτίνη. Αναλυτικά:

1. Εύρεση αλληλουχιών από βιβλιογραφία και βάσεις δεδομένων και ανάλυση αυτών και των χαρακτηριστικών τους.
2. Εύρεση ομοιοτήτων σε επίπεδο ακολουθίας με προσδιορισμένες δομικά πρωτεΐνες (κυρίως με την οικογένεια των λιποκαλινών και ιδιαίτερα με την λιποκαλίνη bovine Retinol Binding Protein – PDB ID: 1FEN).
3. Κατασκευή δομικών μοντέλων των πρωτεϊνών με την μέθοδο της προτυποποίησης με βάση την ομολογία (homology modeling).
4. Ανάλυση των δομικών μοντέλων και κατανόηση των χαρακτηριστικών τους (πιθανή εύρεση κοινών χαρακτηριστικών με τις πρωτεΐνες της οικογένειας των λιποκαλινών).
5. Πραγματοποίηση πειραμάτων αγκυροβόλησης (docking) με τα δομικά μοντέλα των πρωτεϊνών και γραμμικών τμημάτων αλυσίδων χιτίνης.
6. Παρουσίαση ενός ολοκληρωμένου πιθανού μοντέλου της δομής των πρωτεϊνών του δερματίου των αρθροπόδων καθώς και του τρόπου αλληλεπίδρασης αυτών με την χιτίνη, σε συμφωνία με προηγούμενα πειραματικά αποτελέσματα.

Υλικά και Μέθοδοι



«In natural science the principles of truth ought to be confirmed by observation».

(Philosophia Botanica)

Carl Linnaeus (1707-1778)

Κεφάλαιο 2^ο : Υλικά και Μέθοδοι

2.1 Πρωτεΐνες – ακολουθίες Tweedle και CPAP3

Ύστερα από αναζήτηση στη σχετική βιβλιογραφία (Willis 2010; Willis, Papandreou et al. 2012) καθώς και σε έγκυρες βάσεις δεδομένων όπως UniProt (Arweiler, Bairoch et al. 2004) και cuticleDB (Magkrioti, Spyropoulos et al. 2004) συγκεντρώθηκαν οι αμινοξικές αλληλουχίες όλων των γνωστών πρωτεϊνών των οικογενειών Tweedle και CPAP3.

Οι γνωστές πρωτεΐνες που ανήκουν στην οικογένεια Tweedle είναι 236 από 17 είδη αρthropόδων όπως φαίνεται αναλυτικά στον πίνακα 5.

	Οργανισμός	Αριθμός πρωτεϊνών Tweedle
1	<i>Acyrtosiphon pisum</i>	3
2	<i>Aedes aegypti</i>	9
3	<i>Anopheles gambiae</i>	12
4	<i>Apis mellifera</i>	2
5	<i>Bombyx mori</i>	4
6	<i>Culex pipiens</i>	10
7	<i>Danaus plexippus</i>	2
8	<i>Drosophila ananassae</i>	27
9	<i>Drosophila grimshawi</i>	27
10	<i>Drosophila melanogaster</i>	27
11	<i>Drosophila mojavensis</i>	26
12	<i>Drosophila pseudoobscura</i>	28
13	<i>Drosophila virilis</i>	26
14	<i>Drosophila willistoni</i>	28
15	<i>Nasonia vitripennis</i>	2
16	<i>Pediculus humanus corporis</i>	1
17	<i>Tribolium castaneum</i>	2
	Σύνολο	236

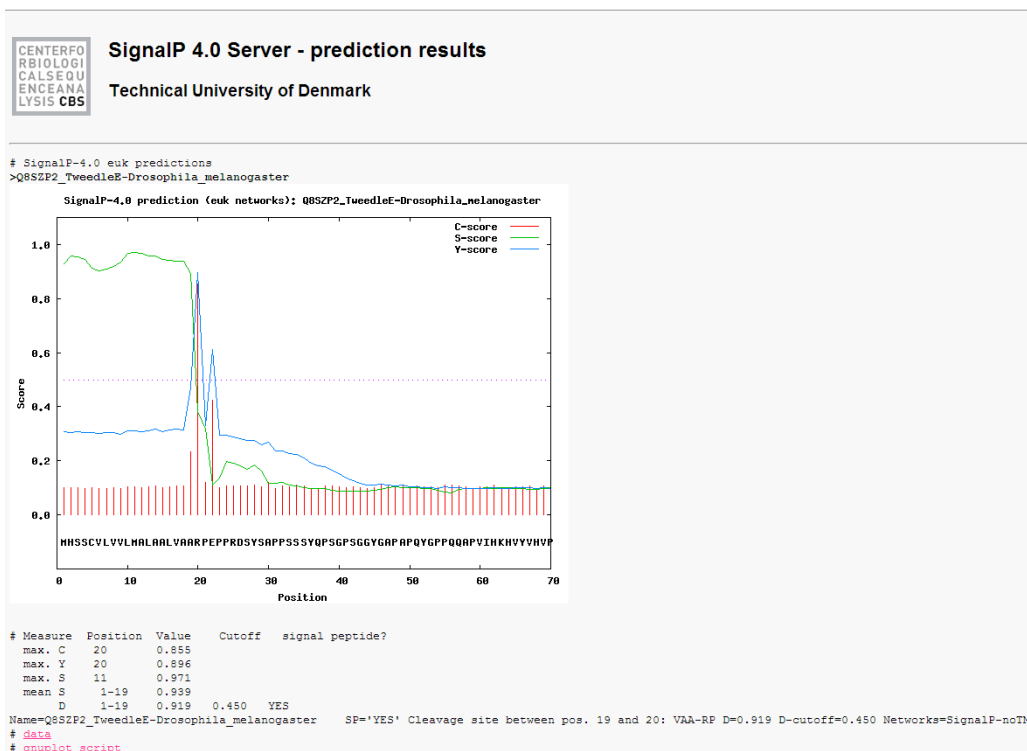
Πίνακας 5: Οι γνωστές πρωτεΐνες της οικογένειας Tweedle ανά οργανισμό.

Οι γνωστές πρωτεΐνες που ανήκουν στην οικογένεια CPAP3 είναι 63 από 10 διαφορετικά είδη όπως φαίνεται στον πίνακα 6.

	Οργανισμός	Αριθμός πρωτεϊνών CPAP3
1	<i>Acyrtosiphon pisum</i>	6
2	<i>Aedes aegypti</i>	7
3	<i>Anopheles gambiae</i>	7
4	<i>Apis mellifera</i>	5
5	<i>Culex pipiens</i>	7
6	<i>Danaus plexippus</i>	5
7	<i>Drosophila melanogaster</i>	6
8	<i>Nasonia vitripennis</i>	6
9	<i>Pediculus humanus corporis</i>	6
10	<i>Tribolium castaneum</i>	8
	Σύνολο	63

Πίνακας 6: Οι γνωστές πρωτεΐνες της οικογένειας CPAP3 ανά οργανισμό.

Στην συνέχεια από τις ακολουθίες αφαιρέθηκε το πεπτίδιο οδηγητής (σε όσες υπήρχε) σύμφωνα με τις πληροφορίες από την UniProt ή σύμφωνα με τα αποτελέσματα του προγράμματος SignalP 4.0 Server (Petersen, Brunak et al. 2011) Technical University of Denmark (DTU). Το SignalP χρησιμοποιεί τεχνητά νευρωνικά δίκτυα και προβλέπει την ύπαρξη (ή όχι) καθώς και το ακριβές σημείο του πεπτιδίου οδηγητή στην αμινοξική αλληλουχία.



Εικόνα 51: Παράδειγμα εξόδου του προγράμματος SignalP 4.0 Server (Petersen, Brunak et al. 2011). Το πρόγραμμα δέχεται ως είσοδο την αμινοξική αλληλουχία και στην έξοδο φαίνεται η ύπαρξη ή όχι του πεπτιδίου οδηγητή, καθώς και σε ποιο σημείο της ακολουθίας βρίσκεται. Η συγκεκριμένη περίπτωση αφορά την ακολουθία της πρωτεΐνης TweedleE (UniProt accession number: Q8SZP2) του οργανισμού *Drosophila melanogaster*. Στην εικόνα φαίνεται ότι το πεπτίδιο οδηγητής βρίσκεται στις θέσεις 1-19 της ακολουθίας.

Οι υπολογιστικές μελέτες που πραγματοποιήθηκαν στο πλαίσιο της παρούσας διπλωματικής εργασίας έγιναν στο σύνολο των παραπάνω πρωτεϊνών, όμως για τον καλύτερο σχολιασμό των αποτελεσμάτων θα εστιάσουμε σε ένα αντιπροσωπευτικό δείγμα από κάθε οικογένεια.

Για την οικογένεια Tweedle, το δείγμα αυτό αποτελείται από δέκα (10) διαφορετικές πρωτεΐνες Tweedle προερχόμενες από τέσσερα (4) διαφορετικά είδη αρθροπόδων και παρουσιάζονται στον πίνακα 7. Τα κριτήρια επιλογής τους ήταν κυρίως να προέρχονται από πολλούς οργανισμούς (προφανώς επελέχθησαν πολλές από τον οργανισμό *Drosophila melanogaster* καθώς είναι ο καλύτερα μελετημένος) και ότι τα μεγέθη των αμινοξικών αλληλουχιών (ύστερα από την αφαίρεση του πεπτιδίου οδηγητή) κυμαίνονται από 163-256 που είναι περίπου όσο της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN), που είναι πολύ σημαντική για τις υπολογιστικές μελέτες που πραγματοποιήθηκαν.

Πρωτεΐνη	UniProt	Αρ. αμινοξέων	Οργανισμός
TweedleZ	Q8IR09	192	<i>Drosophila melanogaster</i>
TweedleQ	Q9VBC8	230	<i>Drosophila melanogaster</i>
TweedleJ	Q9VBD6	256	<i>Drosophila melanogaster</i>
TweedleE	Q8SZP2	178	<i>Drosophila melanogaster</i>
TweedleBeta	A1Z8Q0	179	<i>Drosophila melanogaster</i>
TweedleS	Q9VBD3	199	<i>Drosophila melanogaster</i>
AgamTWDL8	Q7QFK3	199	<i>Anopheles gambiae</i>
AgamTWDL12	Q7Q9V7	185	<i>Anopheles gambiae</i>
BmorCPT4	C0H6Z4	163	<i>Bombyx mori</i>
CPTmotif4	G6D0I0	213	<i>Danaus Plexippus</i>

Πίνακας 7: Τα χαρακτηριστικά των δέκα (10) πρωτεϊνών του αντιπροσωπευτικού δείγματος Tweedle. Στην πρώτη στήλη είναι το όνομα της πρωτεΐνης, στη δεύτερη στήλη είναι ο κωδικός της στην βάση δεδομένων UniProt, στην τρίτη στήλη είναι ο αριθμός των αμινοξικών καταλοίπων (ύστερα από την αφαίρεση του πεπτιδίου οδηγητή) και στην τέταρτη στήλη είναι το αρθρόποδο από το οποίο προέρχεται η πρωτεΐνη.

Για την οικογένεια CPAP3, το δείγμα αυτό αποτελείται από οχτώ (8) διαφορετικές πρωτεΐνες Tweedle προερχόμενες από επτά (7) διαφορετικά είδη αρθροπόδων και παρουσιάζονται στον πίνακα 8. Τα κριτήρια επιλογής τους ήταν κυρίως να προέρχονται από πολλούς οργανισμούς και ότι τα μεγέθη των αμινοξικών αλληλουχιών (ύστερα από την αφαίρεση του πεπτιδίου οδηγητή) κυμαίνονται από 194-226 που είναι περίπου όσο της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN), που είναι πολύ σημαντική για τις υπολογιστικές μελέτες που πραγματοποιήθηκαν.

UniProt	Αρ. αμινοξέων	Οργανισμός
A1YR29	218	<i>Tribolium castaneum</i>
E0VTP0	200	<i>Pediculus humanus</i>
Q5TWB5	207	<i>Anopheles gambiae</i>
G6D9I3	221	<i>Danaus plexippus</i>
J9JZY1	226	<i>Acyrtosiphon pisum</i>
Q16VF6	207	<i>Aedes aegypti</i>
Q8T0V6	218	<i>Drosophila melanogaster</i>
Q16NQ0	194	<i>Aedes aegypti</i>

Πίνακας 8: Τα χαρακτηριστικά των οχτώ (8) πρωτεϊνών του αντιπροσωπευτικού δείγματος CPAP3. Στην πρώτη στήλη είναι ο κωδικός της στην βάση δεδομένων UniProt, στην δεύτερη στήλη είναι ο αριθμός των αμινοξικών καταλοίπων (ύστερα από την αφαίρεση του πεπτιδίου οδηγητή) και στην τρίτη στήλη είναι το αρθρόποδο από το οποίο προέρχεται η πρωτεΐνη.

2.2 Στοιχίση ακολουθιών

Οι αμινοξικές αλληλουχίες των πρωτεϊνών των οικογενειών Tweedle και CPAP3 στοιχίστηκαν με μια ποικιλία πρωτεϊνών της οικογένειας των λιποκαλινών (με προσδιορισμένες κρυσταλλογραφικά δομές), εστιάζοντας κυρίως στις στοιχίσεις με την Retinol Binding Protein (RBP) του οργανισμού *Bos Taurus* (UniProt: P18902 (Berni, Stoppini et al. 1990)) (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)). Οι λιποκαλίνες που χρησιμοποιήθηκαν φαίνονται στον πίνακα 9.

Λιποκαλίνη	PDB ID	Οργανισμός	Reference
Retinol Binding Protein (RBP)	1FEN	<i>Bos taurus</i>	(Zanotti, Marcello et al. 1994)
Retinol Binding Protein (RBP) from pig plasma	1AQB	<i>Sus scrofa domesticus</i>	(Zanotti, Panzalorto et al. 1998)
Bilin Binding Protein (BBP)	1BBP	<i>Pieris brassicae</i>	(Huber, Schneider et al. 1987)
Bovine lipocalin allergen	1BJ7	<i>Bos taurus</i>	(Rouvinen, Rautiainen et al. 1999)
Human Neutrophil Gelatinase-Associated Lipocalin (HNGAL)	1NGL	<i>Homo sapiens</i>	(Coles, Diercks et al. 1999)
Nitrophorin 2	1PEE	<i>Rhodnius prolixus</i>	(Weihsel, A. and Montfort, W.R to be published)
Histamine Binding Protein	1QFT	<i>Rhipicephalus appendiculatus</i>	(Paesen, Adams et al. 1999)
Retinol Binding Protein (RBP)	1RBP	<i>Homo sapiens</i>	(Cowan, Newcomer et al. 1990)

Πίνακας 9: Οι λιποκαλίνες που χρησιμοποιήθηκαν στα πειράματα στοιχίσης. Στην πρώτη στήλη είναι το όνομα της λιποκαλίνης (σύμφωνα με την PDB), στην δεύτερη στήλη είναι ο κωδικός της PDB, στην τρίτη στήλη είναι ο οργανισμός που προέρχεται και στην τέταρτη στήλη είναι η βιβλιογραφική αναφορά της κρυσταλλογραφικά προσδιορισμένης δομής της πρωτεΐνης.

Επίσης οι πρωτεΐνες της οικογένειας Tweedle στοιχίστηκαν και μεταξύ τους, ώστε να μελετήσουμε τα κοινά χαρακτηριστικά των ακολουθιών τους και κυρίως τις τέσσερις (4) συντηρημένες ομάδες (blocks) που τις ταυτοποιεί σαν οικογένεια δομικών πρωτεϊνών (Guan, Middlebrooks et al. 2006; Willis 2010).

Οι πολλαπλές στοιχίσεις πραγματοποιήθηκαν χρησιμοποιώντας το πρόγραμμα ClustalW (Thompson, Higgins et al. 1994), διατηρώντας τις προκαθορισμένες παραμέτρους.

2.3 Προτυποποίηση με βάση την ομολογία

Λόγω της σημαντικής ομοιότητας που προέκυψε ύστερα από τις πολλαπλές στοιχίσεις (θα αναλυθεί στο Κεφάλαιο 3) των ακολουθιών των πρωτεϊνών των οικογενειών Tweedle και CPAP3 με την αμινοξική αλληλουχία της bovine Retinol Binding Protein καθώς και ότι η bovine Retinol Binding Protein λιποκαλίνη έχει προσδιορισμένη κρυσταλλογραφικά δομή (με PDB ID: 1FEN) (Zanotti, Marcello et al. 1994), πραγματοποιήθηκαν πειράματα προτυποποίησης με βάση την ομολογία.

Τα πειράματα αυτά πραγματοποιήθηκαν χρησιμοποιώντας το πρόγραμμα Modeller9v2 (Sali and Blundell 1993). Το λογισμικό Modeller μοντελοποιεί τρισδιάστατες δομές πρωτεϊνών ικανοποιώντας τους περιορισμούς στο χώρο. Ο χρήστης παρέχει μια στοίχιση της ακολουθίας που προορίζεται για μοντελοποίηση με την ακολουθία κάποιας γνωστής δομής που καλείται δομή-οδηγός (σε μορφή .pir), καθώς επίσης και τις συντεταγμένες των ατόμων της δομής-οδηγού. Το λογισμικό αυτόματα υπολογίζει το νέο μοντέλο με όλα τα άτομα εκτός από τα υδρογόνα. (Sali and Blundell 1993).

Τα πρωτεϊνικά μοντέλα βελτιστοποιήθηκαν δομικά χρησιμοποιώντας το λογισμικό UCSF Chimera 1.7 (Pettersen, Goddard et al. 2004). Συγκεκριμένα, χρησιμοποιήθηκε η λειτουργία «Minimize structure» η οποία ελαχιστοποιεί την ενέργεια των μοριακών μοντέλων. Εξ ορισμού, η ελαχιστοποίηση ενέργειας απλά μετακινεί το σύστημα προς ένα τοπικό ελάχιστο και δεν ψάχνει για το ολικό ελάχιστο. Ενώ, η εγκυρότητα των μοντέλων ελέγχθηκε με τη μέθοδο Protein Model Check του WHAT IF web server (Hooft, Vriend et al. 1996).

2.4 Δομική στοίχιση

Πραγματοποιήθηκε δομική στοίχιση (structure alignment) μεταξύ των δημιουργηθέντων μοντέλων των πρωτεϊνών της οικογένειας Tweedle (ύστερα από την προτυποποίηση με βάση την ομολογία) με την λυμένη κρυσταλλογραφικά δομή της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN). Η στοίχιση έγινε ώστε να εξεταστεί η ενδεχόμενη ύπαρξη των εσωτερικών και εξωτερικών υδρόφοβων συστάδων (inner and outer clusters) στα μοντέλα, όπως υπάρχουν και στην Retinol Binding Protein της οικογένειας των λιποκαλινών (Adam, Charloteaux et al. 2008).

Για την δομική στοίχιση χρησιμοποιήθηκε το πρόγραμμα Combinatorial Extension method (Shindyalov and Bourne 1998) το οποίο βρίσκεται στον ιστότοπο της βάσης δεδομένων PDB (Berman, Westbrook et al. 2000), στην διαδικτυακή διεύθυνση (URL) <http://source.rcsb.org/jfatcatserver/ceHome.jsp>. Ο αλγόριθμος Combinatorial Extension χρησιμοποιεί μια υπολογιστική μέθοδο που εκτείνει παρόμοια ζεύγη αμινοξικών καταλοίπων. Το καλύτερο μονοπάτι (path) βελτιστοποιείται με δυναμικό προγραμματισμό και με την μέθοδο Monte Carlo.

Ο πιο διαδεδομένος δείκτης βαθμολόγησης των δομικών στοιχίσεων είναι ο RMSD (Root Mean Square Deviation) (Sippl 1990):

$$\begin{aligned} \text{RMSD}(\mathbf{v}, \mathbf{w}) &= \sqrt{\frac{1}{n} \sum_{i=1}^n \|\mathbf{v}_i - \mathbf{w}_i\|^2} \\ &= \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)} \end{aligned}$$

$\mathbf{v}, \mathbf{w}$: Το σύνολο των ατόμων C_α , με τις συντεταγμένες τους, στις δυο δομές (x, y, z).

n : Ο αριθμός των ζευγαριών ατόμων C_α .

Ο δείκτης RMSD μετριέται σε Ångström (Å). $1\text{Å} = 10^{-10}\text{m}$.

2.5 Δομική ανάλυση και άλλες υπολογιστικές μελέτες

Στο πλαίσιο περαιτέρω ανάλυσης των χαρακτηριστικών των πρωτεϊνών των οικογενειών Tweedle και CPAP3, πραγματοποιήθηκαν και μια ποικιλία άλλων μελετών με την χρήση πληθώρας προγραμμάτων.

Αρχικά, όλες οι οπτικοποιήσεις των διαφόρων μορίων της παρούσας διπλωματικής πραγματοποιήθηκαν με την χρήση του προγράμματος PyMOL Molecular Graphics System (Delano 2005).

Μέσω της οπτικοποίησης των πρωτεϊνικών μοντέλων, έγινε έλεγχος με τι δομικά χαρακτηριστικά ταυτίζονται οι τέσσερις συντηρημένες ομάδες (blocks) στην αμινοξική ακολουθία των πρωτεϊνών, που χαρακτηρίζουν την οικογένεια Tweedle.

Χρησιμοποιώντας την οπτικοποίηση αυτή, έγιναν δομικές μελέτες πάνω στα πρωτεϊνικά μοντέλα της οικογένειας Tweedle. Οι μελέτες αυτές αφορούσαν την εξέταση του περιβάλλοντος στο εσωτερικό των β-βαρελιών και στην υποθετική εσωτερική συστάδα (inner cluster) στα πρωτεϊνικά μοντέλα, ανάλογα με την αμινοξική τους σύνθεση (υδρόφοβα, πολικά, φορτισμένα). Ακόμα εξετάστηκε η παρουσία αρωματικών αμινοξέων στο εσωτερικό των β-βαρελιών.

Επιπροσθέτως, έγινε χρήση του προγράμματος LipocalinPred (Ramana and Gupta 2009) το οποίο δέχεται μια αμινοξική ακολουθία και προβλέπει αν είναι ή όχι λιποκαλίνη καθώς και το πόσο πιθανό είναι αυτό. Συγκεκριμένα χρησιμοποιήθηκε η μέθοδος SVM (Support Vector Machines) prediction – PSSM model (Position Specific Substitution Matrix). Η λειτουργία του βασίζεται στην εύρεση μακρινών ομολογιών μέσω του PSI-BLAST. Η χρήση του LipocalinPred πραγματοποιήθηκε σαν επιπρόσθετος έλεγχος ομοιότητας των πρωτεϊνών Tweedle και CPAP3 με πρωτεΐνες της οικογένειας των λιποκαλινών.

Εικόνα 52: Η βασική σελίδα του προγράμματος LipocalinPred όπου τοποθετείται η αμινοξική αλληλουχία της πρωτεΐνης για να γίνει η πρόβλεψη αν είναι ή όχι λιποκαλίνη. Για την πρόβλεψη γίνεται χρήση support vector machines (SVM).

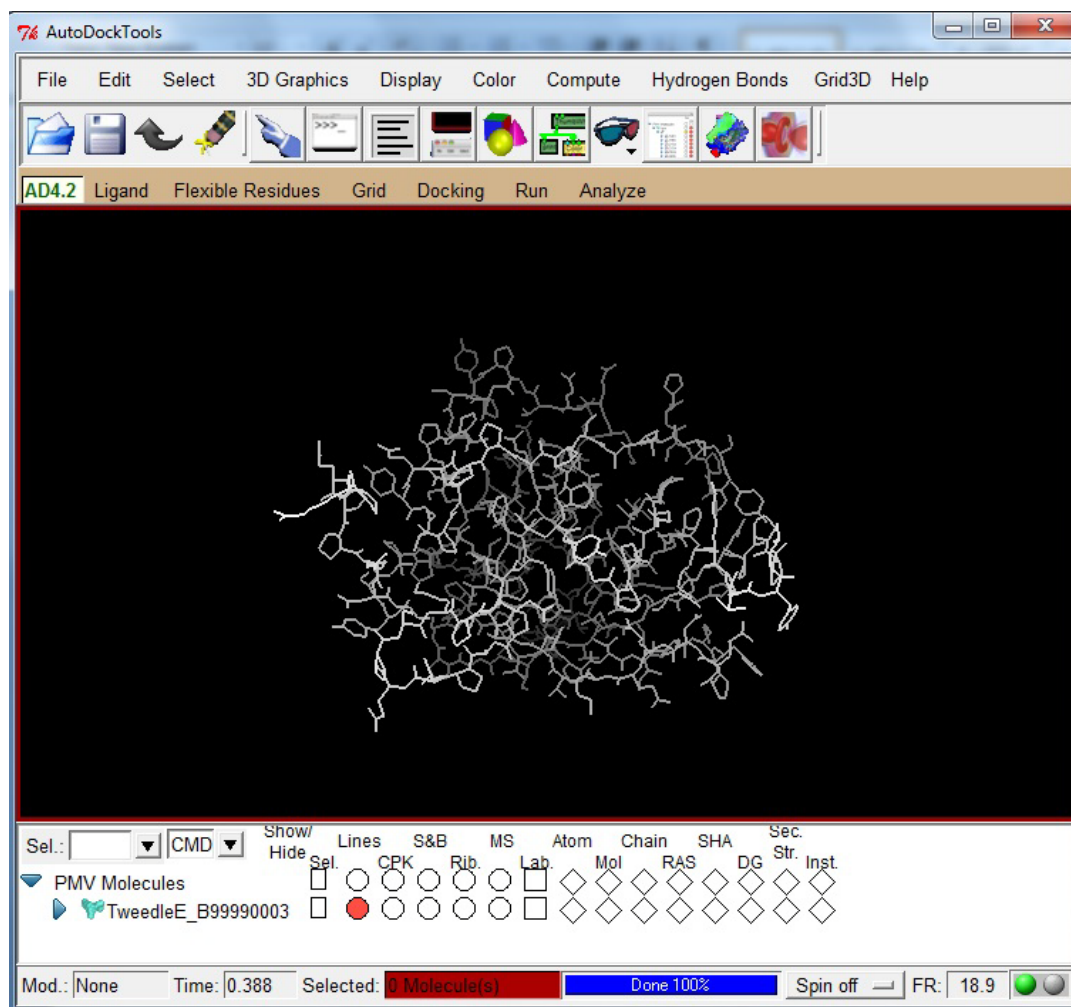
2.6 Πειράματα αγκυροβόλησης (Docking)

Πραγματοποιήθηκαν πειράματα αγκυροβόλησης (Docking) ώστε να παραχθεί ένα πιθανό μοντέλο αλληλεπίδρασης μεταξύ των μοντέλων των πρωτεϊνών της οικογένειας Tweedle και των αλυσίδων χιτίνης. Τα πειράματα πραγματοποιήθηκαν χρησιμοποιώντας τα ευρέως χρησιμοποιούμενα προγράμματα αγκυροβόλησης GRAMM (Vakser 1996) και AutoDock4.2 (Morris, Huey et al. 2009), το οποίο είναι και το πρόγραμμα με τις περισσότερες αναφορές στην σχετική βιβλιογραφία (Sousa, Fernandes et al. 2006). Σαν πρωτεΐνη-στόχος θεωρήθηκαν τα πρωτεϊνικά μοντέλα της οικογένειας Tweedle, ενώ μόρια προσδέτες (ligand) ήταν μια ποικιλία γραμμικών τμημάτων αλυσίδων χιτίνης (NAG, N-ακετυλογλυκοζαμίνης).

Αναλυτικότερα, με το πρόγραμμα GRAMM πραγματοποιήθηκαν πειράματα αγκυροβόλησης των πρωτεϊνικών μοντέλων Tweedle με αλυσίδες χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG), ομοεξαμερών N-ακετυλογλυκοζαμίνης (6NAG), ομοδεκαοχταμερών N-ακετυλογλυκοζαμίνης (18NAG) καθώς και πολλαπλές αλυσίδες 18NAG. Η τεχνική της αγκυροβόλησης ήταν υψηλής ανάλυσης γενική αγκυροβόληση (High-resolution generic docking), χρησιμοποιώντας τον αλγόριθμο Geometric docking II (Katchalski-Katzir, Shariv et al. 1992) και προσαρμόζοντας το βήμα πλέγματος (grid step) στο ελάχιστο δυνατό.

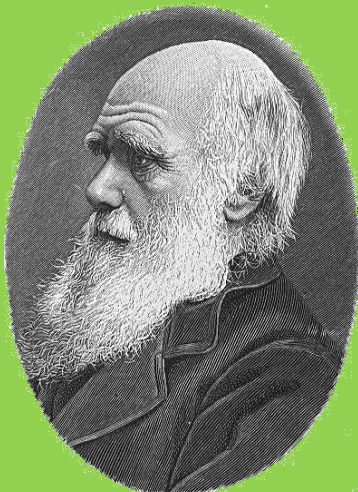
Για το πρόγραμμα AutoDock4.2 τα πειράματα πραγματοποιήθηκαν μεταξύ των πρωτεϊνικών μοντέλων και μιας αλυσίδας ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG), θεωρώντας την αλυσίδα άκαμπτη (rigid), ενώ όλοι οι περιστρεφόμενοι δεσμοί αφέθηκαν ελεύθεροι. Οι πλευρικές αλυσίδες της πρωτεΐνης παρέμειναν σταθερές. Η εφαρμογή AutoDockTools1.5.4 χρησιμοποιήθηκε για να παραχθούν τα εισερχόμενα αρχεία των πειραμάτων αγκυροβόλησης. Σε όλα τα πειράματα ένας χάρτης πλέγματος (grid map) με χώρο πλέγματος (grid-spacing) 1.0Å δημιουργήθηκε για να καλύψει το χώρο που καταλαμβάνουν οι πρωτεΐνες Tweedle (οι διαστάσεις x, y, z προσαρμόστηκαν ανάλογα με το μέγεθος του κάθε πρωτεϊνικού μοντέλου – ενδεικτικά οι τιμές ήταν 55 x 55 x 53 για την TweedleZ, 59 x 57 x 57 για την TweedleE και 51 x 49 x 49 για την Cuticular protein tweedle motif4 κτλ). Δέκα «τρεξίματα» πραγματοποιήθηκαν με χρήση του αλγορίθμου genetic Lamarckian (Morris, Goodsell et al. 1998). Οι ρυθμίσεις ήταν οι προκαθορισμένες (default) του

προγράμματος, με 2.5×10^6 μέγιστες εκτιμήσεις ενέργειας και 2.7×10^4 μέγιστο αριθμό παραγώγων. Επιλέχθησαν ώστε η ρύθμιση «mutation rate» να είναι 0.02 και η ρύθμιση «crossover rate» να είναι 0.8. Αποτελέσματα που διέφεραν κατά $<2\text{\AA}$ στον δείκτη θέσης rmsd (root mean square deviation) ομαδοποιήθηκαν μαζί.



Εικόνα 53: Η βασική εικόνα της εφαρμογής AutoDockTools του προγράμματος AutoDock όπου φαίνονται κάποιες βασικές επιλογές καθώς και ένα ενδεικτικό παράδειγμα προβολής μιας πρωτεΐνης (TweedleE) στο κέντρο (πριν την διεξαγωγή του πειράματος).

Αποτελέσματα



«It has often and confidently been asserted, that man's origin can never be known: but ignorance more frequently begets confidence than does knowledge: it is those who know little, and not those who know much, who so positively assert that this or that problem will never be solved by science.»

(The Descent of Man)

Charles Robert Darwin (1809-1882)

Κεφάλαιο 3^ο : Αποτελέσματα

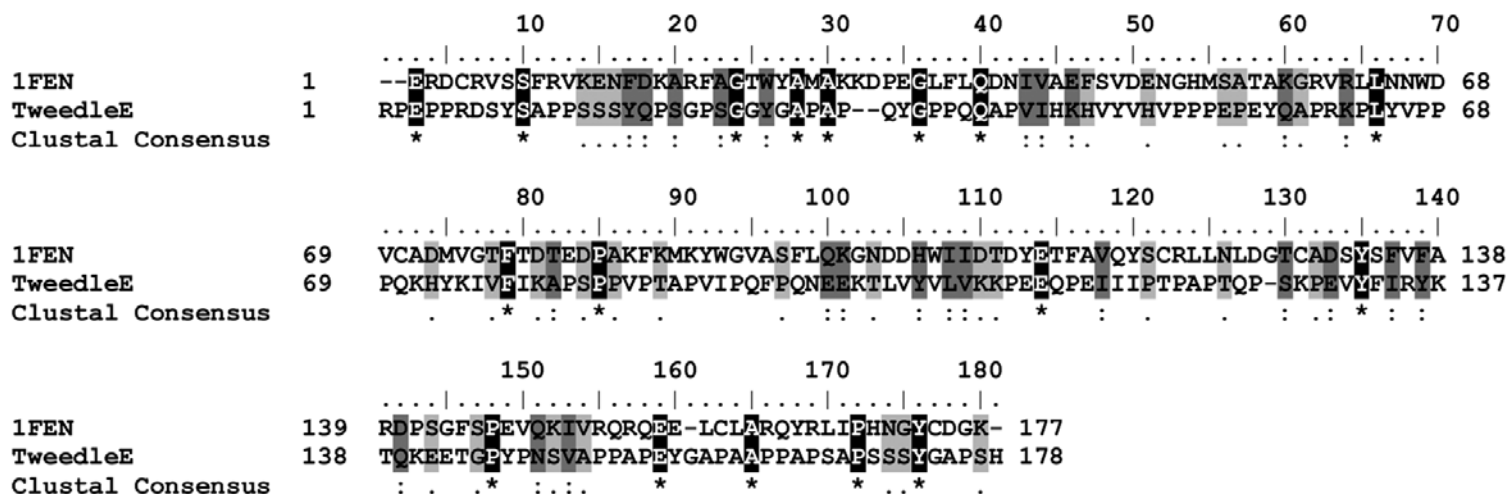
3.1 Στοιχίση ακολουθιών

Οι ακολουθίες των πρωτεϊνών της οικογένειας Tweedle στοιχήθηκαν αρχικά με την ακολουθία της λιποκαλίνης Retinol Binding Protein (RBP) (UniProt: P18902 (Berni, Stoppini et al. 1990)) (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)), εν συνεχεία με ακολουθίες άλλων λιποκαλινών με προσδιορισμένες κρυσταλλογραφικά δομές και τέλος μεταξύ τους ώστε να παρατηρήσουμε την συντήρηση των τεσσάρων (4) συντηρημένων ομάδων (blocks) που τις ταυτοποιούν σαν οικογένεια δομικών πρωτεϊνών (Guan, Middlebrooks et al. 2006; Willis 2010).

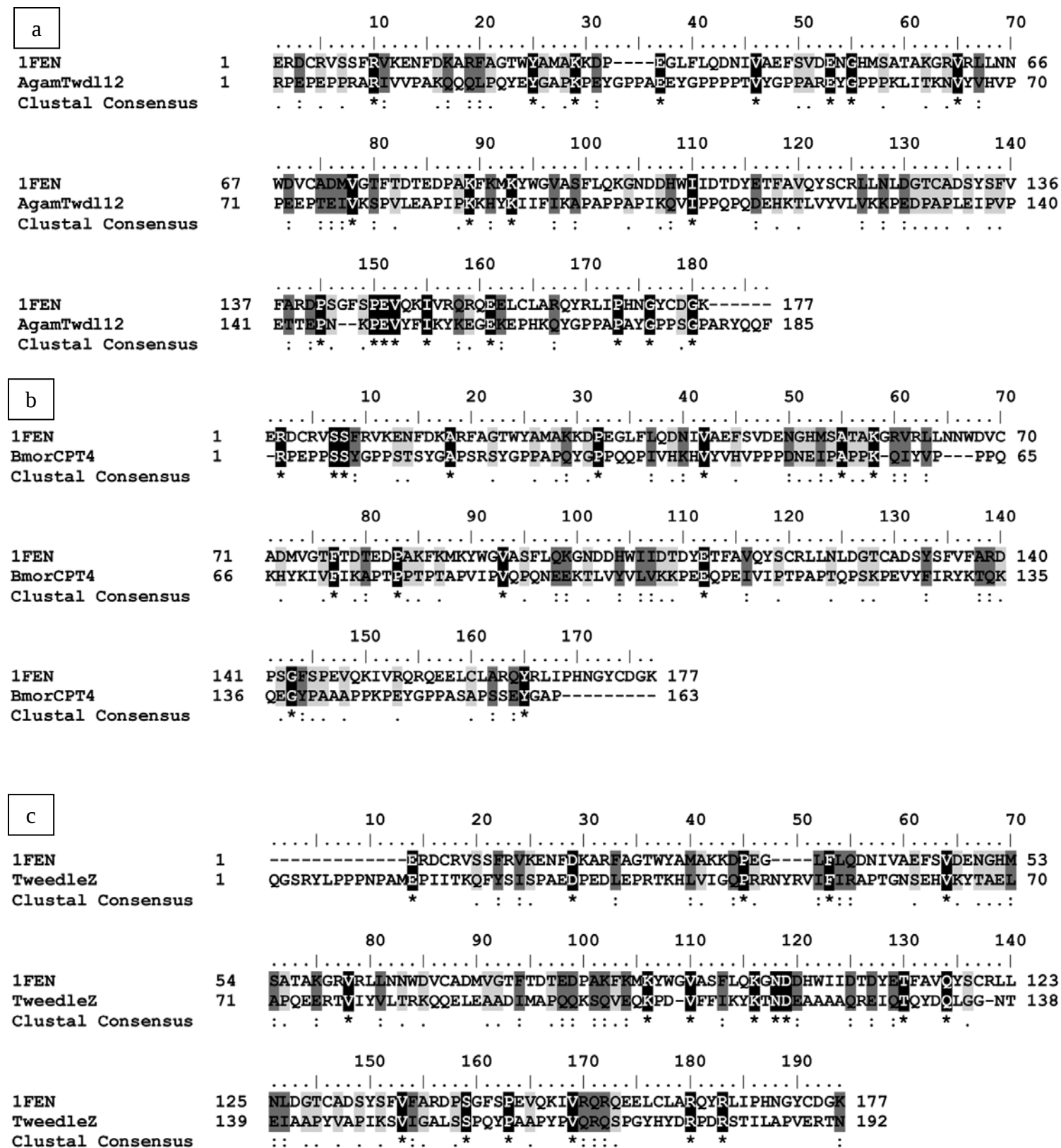
Στον πίνακα 10 παρουσιάζονται τα αποτελέσματα των στοιχίσεων των πρωτεϊνών του αντιπροσωπευτικού δείγματος της οικογένειας Tweedle με την ακολουθία της λιποκαλίνης bovine Retinol Binding Protein (RBP) (PDB ID: 1FEN). Είναι εμφανές ότι όλες οι πρωτεΐνες του δείγματος εμφανίζουν μια ταυτότητα σε επίπεδο ακολουθίας, με επικρατούσες τιμές περίπου 10% - 11%. Η ομοιότητα σε επίπεδο ακολουθίας είναι σαφώς υψηλότερη πλησιάζοντας το 50% ή και ξεπερνώντας το σε κάποιες περιπτώσεις. Αξίζει να σημειωθεί ότι ο αριθμός των αμινοξέων στις πρωτεΐνες του αντιπροσωπευτικού δείγματος είναι περίπου όσο της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN), που είναι 177, το οποίο βοηθάει στην ποιοτική ανάλυση των αποτελεσμάτων. Επιπροσθέτως, η ποιότητα των αποτελεσμάτων στις στοιχίσεις μεταξύ των δυο οικογενειών αποδεικνύεται από την σχεδόν συνεχόμενη στοιχίση σε όλο το μήκος των ακολουθιών (χωρίς μεγάλα κενά ενδιάμεσα στη δομή ή και καθόλου) όπως παρουσιάζεται στις εικόνες 54,55,56,57 και 58.

Πρωτεΐνη	UniProt	Αρ. αμινοξέων	Sequence Identity	Sequence Similarity
TweedleZ	Q8IR09	192	10.7%	42.4%
TweedleQ	Q9VBC8	230	14.7%	54.8%
TweedleJ	Q9VBD6	256	17.5%	57.6%
TweedleE	Q8SZP2	178	9.6%	39%
TweedleBeta	A1Z8Q0	179	10.2%	43.5%
TweedleS	Q9VBD3	199	11.8%	49.2%
AgamTWDL8	Q7QFK3	199	11.2%	42.9%
AgamTWDL12	Q7Q9V7	185	11.8%	42.4%
BmorCPT4	C0H6Z4	163	8.6%	43%
CPTmotif4	G6D0I0	213	10.7%	50.8%

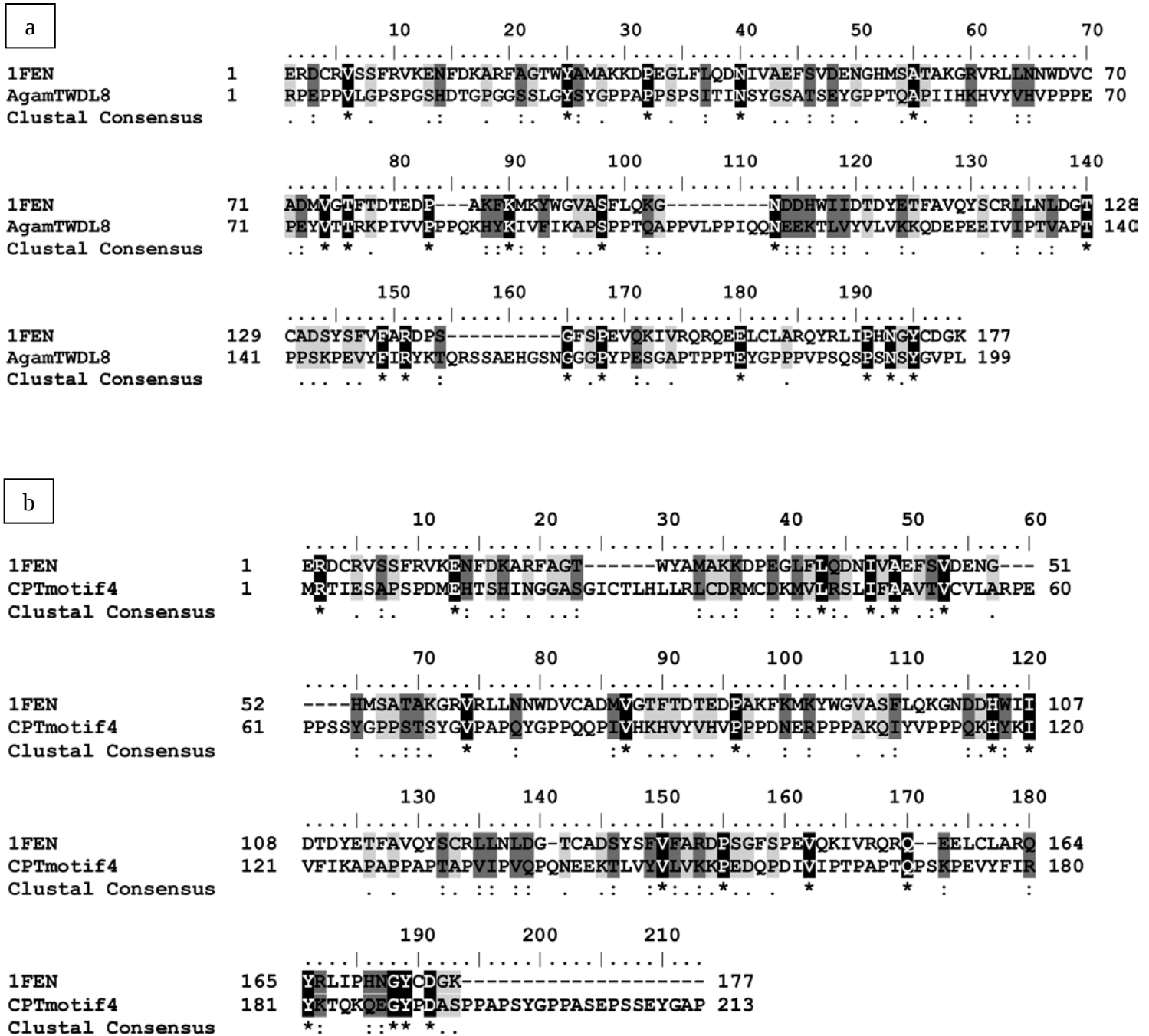
Πίνακας 10: Αποτελέσματα μετά από στοίχιση των αμινοξικών αλληλουχιών του αντιπροσωπευτικού δείγματος των πρωτεϊνών της οικογένειας Tweedle με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN). Στην πρώτη στήλη είναι το όνομα της πρωτεΐνης, στην δεύτερη στήλη είναι ο κωδικός της UniProt, στην τρίτη στήλη είναι ο αριθμός των αμινοξέων (μετά την αφαίρεση του πεπτιδίου οδηγητή), στην τέταρτη στήλη είναι η ταυτότητα (%) μεταξύ των συγκρινόμενων ακολουθιών (Sequence Identity – SI) και στην πέμπτη στήλη είναι η ομοιότητα (%) (Sequence Similarity – SS) μεταξύ των συγκρινόμενων ακολουθιών. Τα αποτελέσματα αυτά προήλθαν χρησιμοποιώντας το πρόγραμμα ClustalW (Thompson, Higgins et al. 1994).



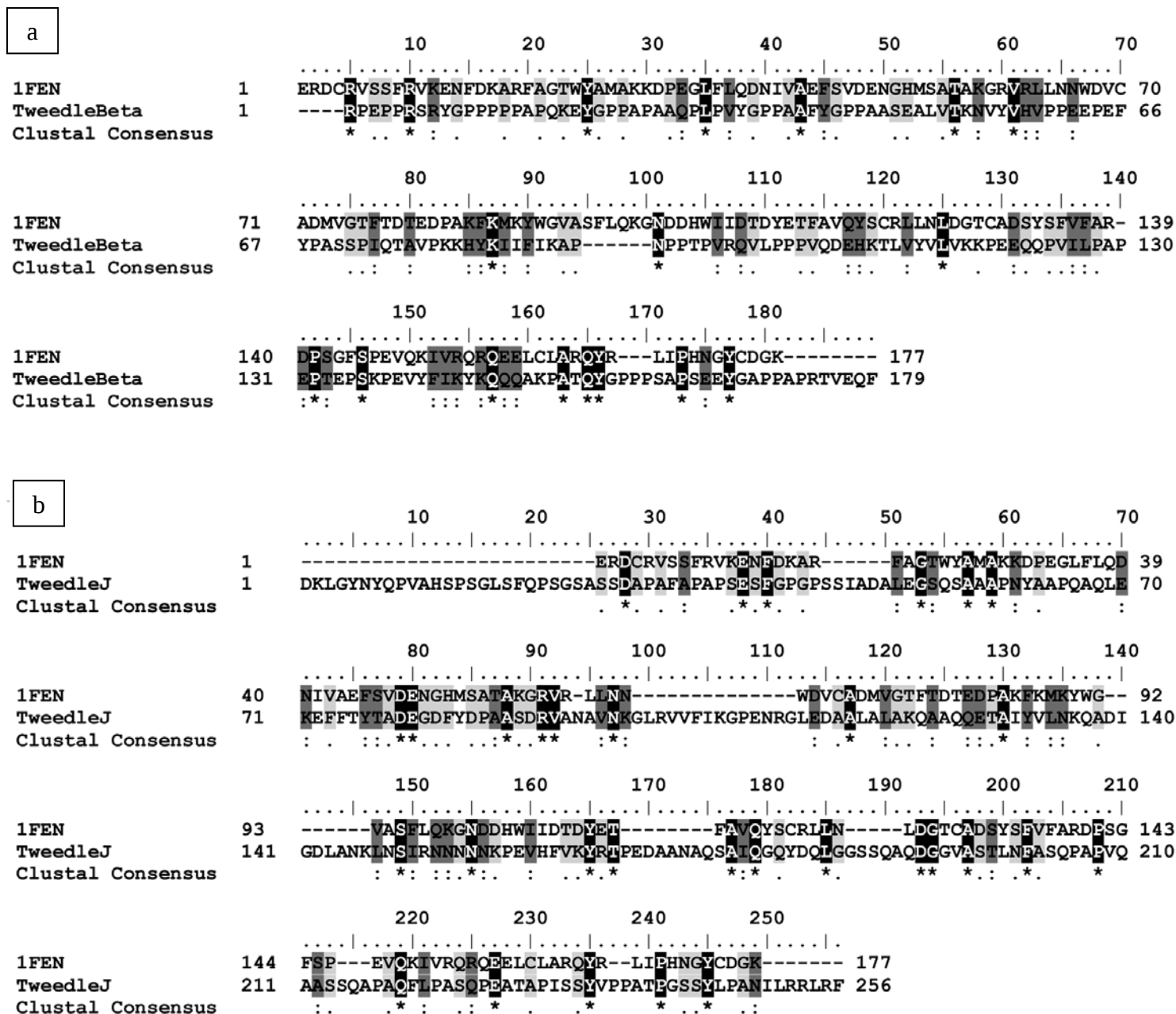
Εικόνα 54: Αποτελέσματα στοίχισης, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης bovine RBP (PDB ID: 1FEN) και της πρωτεΐνης TweedleE. Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από την ακολουθία έχει αφαιρεθεί το πεπτιδίο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.



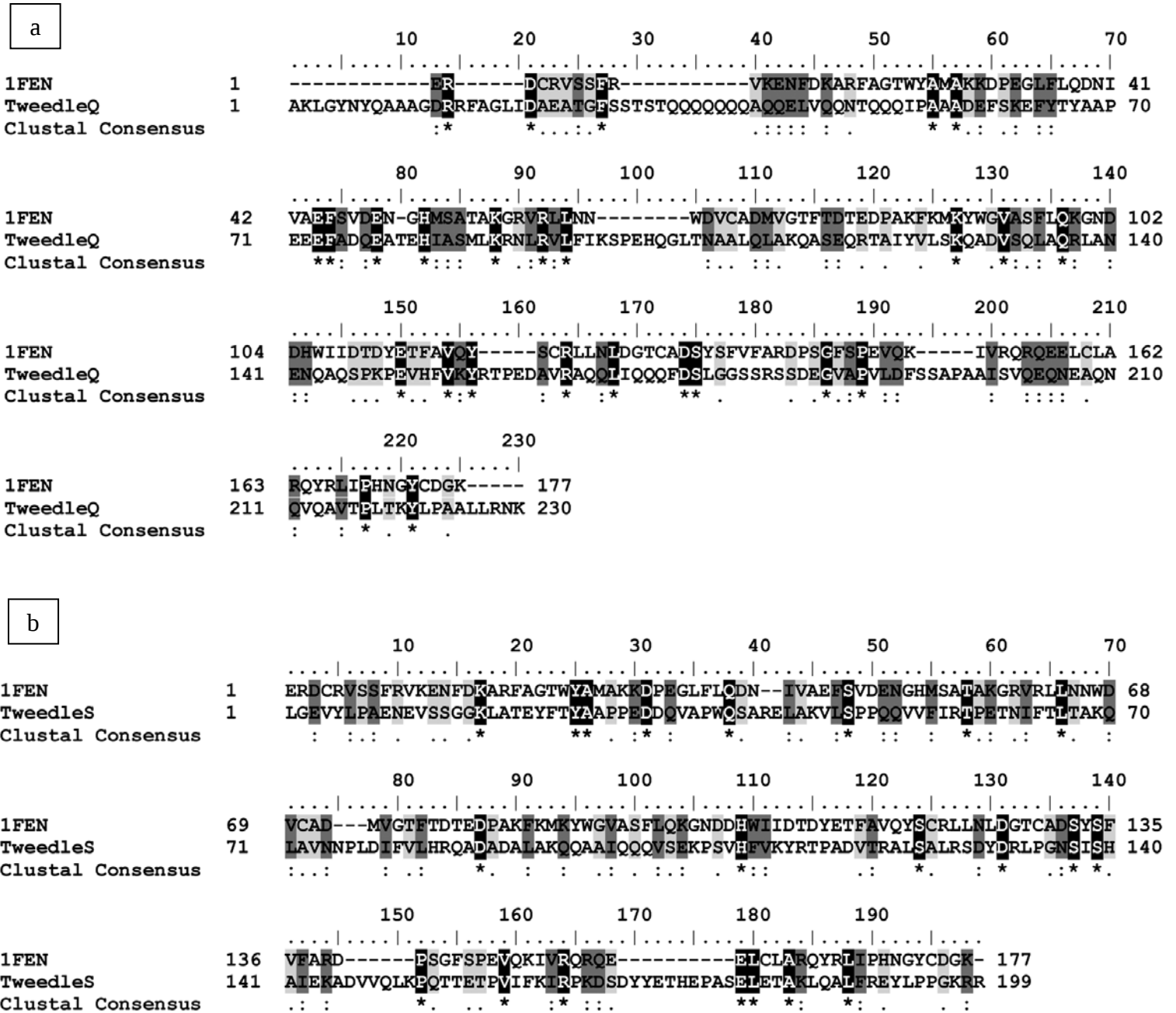
Εικόνα 55: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης bovine RBP (PDB ID: 1FEN) και πρωτεϊνών της οικογένειας Tweedle (a. AgamTWDL12, b. BmorCPT4, c. TweedleZ). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτιδίο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.



Εικόνα 56: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης bovine RBP (PDB ID: 1FEN) και πρωτεϊνών της οικογένειας Tweedle (a. AgamTWDL8, b. CPTmotif4). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτιδίο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.



Εικόνα 57: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης bovine RBP (PDB ID: 1FEN) και πρωτεϊνών της οικογένειας Tweedle (a. TweedleBeta, b. TweedleJ). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτίδιο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.



Εικόνα 58: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης bovine RBP (PDB ID: 1FEN) και πρωτεϊνών της οικογένειας Tweedle (a. TweedleQ, b. TweedleS). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτίδιο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.

Στη συνέχεια παρατίθενται τα αποτελέσματα των στοιχίσεων μεταξύ της ακολουθίας της πρωτεΐνης TweedleE (UniProt: Q8SZP2) του οργανισμού *Drosophila melanogaster* και ακολουθιών από ένα δείγμα λιποκαλινών με προσδιορισμένες κρυσταλλογραφικά δομές (πίνακας 11). Είναι εμφανές ότι η συγκεκριμένη πρωτεΐνη της οικογένειας Tweedle (καθώς και πολλές άλλες της ίδιας οικογένειας) εμφανίζουν αρκετά σημαντικά ποσοστά ταυτότητας (περίπου 11%) και ομοιότητας (περίπου 45%) σε επίπεδο ακολουθιών με τις πρωτεΐνες της οικογένειας των λιποκαλινών. Επίσης στις στοιχίσεις, δεν παρατηρούνται μεγάλα κενά (σε ορισμένες περιπτώσεις κανένα κενό) κάτι που καταδεικνύει ότι τα αποτελέσματα είναι αρκετά πιο ικανοποιητικά (εικόνα 59 a-g).

Λιποκαλίνη	PDB	Οργανισμός	Αρ. Αμινοξ.	Sequence Identity	Sequence Similarity
Retinol Binding Protein (RBP) from pig plasma	1AQB	<i>Sus scrofa domestica</i>	175	11.4%	45.1%
Bilin Binding Protein (BBP)	1BBP	<i>Pieris brassicae</i>	173	11.6%	43.9%
Bovine lipocalin allergen	1BJ7	<i>Bos taurus</i>	150	10%	46.7%
Plasma Retinol Binding Protein (RBP)	1FEN	<i>Bos taurus</i>	177	9.6%	39%
Human Neutrophil Gelatinase-Associated Lipocalin (HNGAL)	1NGL	<i>Homo sapiens</i>	179	14%	47.7%
Nitrophorin 2	1PEE	<i>Rhodnius prolixus</i>	180	7.9%	44.4%
Histamine Binding Protein	1QFT	<i>Rhipicephalus appendiculatus</i>	175	10.8%	47.4%
Retinol Binding Protein (RBP)	1RBP	<i>Homo sapiens</i>	175	10.8%	44%

Πίνακας 11: Αποτελέσματα μετά από στοίχιση των αμινοξικών αλληλουχιών της πρωτεΐνης TweedleE (UniProt: Q8SZP2), του οργανισμού *Drosophila melanogaster*, με αμινοξικές αλληλουχίες λιποκαλινών με προσδιορισμένες κρυσταλλογραφικά δομές. Στον πίνακα φαίνονται: στην πρώτη στήλη το όνομα της λιποκαλίνης (σύμφωνα με την PDB), στην δεύτερη στήλη ο κωδικός της λιποκαλίνης στην PDB, στην τρίτη στήλη ο οργανισμός που προέρχεται η πρωτεΐνη, στην τέταρτη στήλη ο αριθμός των αμινοξέων που συνεισφέρουν στην κρυσταλλογραφικά λυμένη δομή, στην πέμπτη στήλη η ποσοστιαία ταυτότητα μεταξύ των συγκρινόμενων ακολουθιών (Sequence Identity – SI) και στην έκτη στήλη είναι η ποσοστιαία ομοιότητα (Sequence Similarity – SS) μεταξύ των συγκρινόμενων ακολουθιών. Τα αποτελέσματα αυτά προήλθαν χρησιμοποιώντας το πρόγραμμα ClustalW.

a

		10	20	30	40	50	60	70										
1AQB	1	ERDCRVSS	FRVKENFD	KARFSGT	WYAMAKD	PEGLFLQ	DNIVA	EFSD	ENGHMS	ATAKGR	VR-LL	64						
TweedleE	1	RPEPPRDS	YSAPPSS	SYQPSG	FGSGGYG	APAPQY	GPPQAP	VIH	KHVYV	HVPPE	PEYQAP	RKPLYV	PPPPQ	70				
Clustal Consensus		*	*	...	:	**	:	*	*	:	:	*	:	*				
		80	90	100	110	120	130	140										
1AQB	65	NNWDVC	ADMVGT	FTDTE	EAKFKM	KYWG	---V	ASFLO	KGNDD	-HWI	TD	TDYD	TYAV	QYSC	RLQ	NLDG	TTC	129
TweedleE	71	KHYKIV	FIKAPS	PFVPTA	PVPIQ	FPONE	EKTLV	YVLV	KKPEE	QPEI	ITPT	PAPT	QPSK	PEVY	FIRY	KTK	Q	140
Clustal Consensus		:	:	:	:	*	:	:	*	:	:	*	*	*	:	:	:	:
		150	160	170	180													
1AQB	130	ADSYST	FVFARD	PHGFS	PEVQK	IVRQR	QEELC	LARQY	RIITH	NGYCD		175						
TweedleE	141	EETGP	YPNSV	APP--	APYGA	PAAP-	PAPS	APSS	SYGA	PSH--		178						
Clustal Consensus		:	:	:	*	:	**	.	.	:	*	*						

b

		10	20	30	40	50	60	70												
1BBP	1	-----	NVYHD	GACPE	VKPEV	DNFD	WSNY	HGKW	IEVAK	YPNS	VEKY	GKCG	WAEY	TP	EGK	SVK	VS	NYH	VIHG	64
TweedleE	1	RPEPPRDS	YSAPPSS	SYQPSG	FGSGGYG	APAPQY	GPPQAP	VIH	KHVYV	HVPPE	PEPE	PEYQ	APR	KPLY	VPP	PPQ	70			
Clustal Consensus		:	*	...	:	*	.	.	:	:	:	*	*	:	*	:	:	*	:	:
		80	90	100	110	120	130	140												
1BBP	65	KEYF	TEGT	AYEV	GDSK	IGKI	YHKL	TYGG	-V	KEN	VENV	LSTD	NKNY	ITIG	YCKY	DE	DKGH	QDFV	VLSR	133
TweedleE	71	KHYKIV	FIKAPS	PFVPTA	PVPIQ	FPONE	EKTLV	YVLV	KKPEE	QPEI	ITPT	PAPT	QPSK	PEVY	FIRY	KTK	Q	140		
Clustal Consensus		*	*	*	*	.	:	:	*	*	:	:	*	*	.	:	*	:	:	:
		150	160	170	180															
1BBP	134	SKVL	TGEAK	ITAV	ENYLI	GS	SEV	VD	SQ	KL	VYS	DF	SE	AACK	VN	173				
TweedleE	141	EETGP	YPNSV	APPAP	YGA	PAAP	PAPS	APSS	SYGA	PSH--	178									
Clustal Consensus		:	.	.	*	.	*	:	*	*	:									

1BJ7 1 -IDPSK--IPGEWRIIYAAADNKKDKIVEGGPLRNYYRRIECINDCESLSITFYLKDQGTG 57
TweedleE 1 RPEPRDSYSAPPSSSYQPSGGYGAPAPQYGPFOQAPVTHKHVYVHVPPPEPEYQAP 60
Clustal Consensus *: : .. * . : . . . * . : : * : : : : : : :
70 80 90 100 110 120
1BJ7 58 LLLTEVAKROEGYVYVLEFYGTNTLEVIHVS-----ENMLVTYVENYDGERITKMT 109
TweedleE 61 RKPLYVPEPKHYKIVFIKAPSPFPVPTAPVPIQFPQNEEKTIVYVIVKKPEEQPEIIIPT 120
Clustal Consensus * . * : * : : : . * : : : : : : : :
130 140 150 160 170
1BJ7 110 GLAKGTSFTPEELEKYQQLNSERGVPNENIENLIKTDNCP----- 150
TweedleE 121 PAPTQPSKPEVYFIRYKTKQKEETGPYPNSVAPPAPEYGAEAAAPPAPSAPSSSYGAPSH 178
Clustal Consensus * : : : * : : * : : : : * : : : : : : : :

```

1NGL      1  MDSTSDLIAPPTISKVPLQONFQDNQFQCKYVVGLAGNAILREDKDPQRYATIIYELKEDKSYNVTSV 70
TweedleE  1  RPEFPRDSYSAPPSS-----SYQPSGPGSGYGAPAPQYGPFPQAPVIHKHVYVHVPPP-EPEYQAPRKP 63
Clustal Consensus      ... *   *** *       :*: .  * : . .  . . :   :*: . :   * :   .

                        80      90      100      110      120      130      140
1NGL      71  ITRKKKKCDYWIRTFVPGCEGEFTLGNIKSYE-GLTSYLVRVVSNTYNQHAMVFFKKVS---QNREYFKI 136
TweedleE  64  IYVPPPPQKHVKIVEIKAPSEPVPTAFVLPQEQNEEKLTVYVLVKKPEEQPEIIIPPTAPTQPSKPEVYF 133
Clustal Consensus      *:       .::  .*: . . *   *   * .: *   . ** *: . : ::: ::: . :   . :   .

                        150      160      170      180
1NGL      137  TLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVIDQCIDG-- 179
TweedleE  134  IRYKTKKEETGPYPNSVAPPAPGYGAPAAPPSAPSSSYGAPSH 178
Clustal Consensus      *   ** *   : : :   *   *   *

```

```

1PEE      1 MDCSTNISPKQGLDKAKYFSGKWYVTHFLDKDPOVTDQYCSSFTPRES DGTVKEALYHYANAKTSEYNI 70
TweedleE 1 --RPEPPRDSYSAPPSSSYQPSGPGGYGAPAPQYGGPQQAQFVIHKHVYVHVPPPEPEYCAPKELYVPP 68
Clustal Consensus      .      . .      :. :. :      :      **      :.. :.      *      .      .*: * :*. :

                        80      90      100      110      120      130      140
1PEE      71 GEGKLESSGLQYTAKYKTVDKKKAVLKEADEKNSYTTIVLEADDSSALVHICLREGSKDLGDIYTVLTHQ 140
TweedleE 69 POKHYKIVFIKAPSPFVPTAPVIPQFPONEKTLVYVIVKKPEEQPEIIIPTPAPTQPSKPEVVFIRYKT 138
Clustal Consensus      : : :      :: :      .      . : : : **      : * : : : : : :      .      . : : * : :

                        150      160      170      180
1PEE      141 KDAEPSAKVRSAVTQAGLQLSQFVGTKDLGCOYDDQFTSL 180
TweedleE 139 QKEETGPYPNSVAPPAPEYGAPAAPPAAPSAPSSSYGAPSH 178
Clustal Consensus      : * : : : * : :      : : : : : : :      *

```

f

		10	20	30	40	50	60	70	
1QFT	1	NCPDWADEAANGAHQDAWKSLKADVENVYVMVKATYKNDPVWGNDETCVGVMANDVNEDEKSIQAEFLFM	70						
TweedleE	1	-RPEPPRDSYSAPPSSSYQPSGPGSGG--YGAPAPQYGPQQAAPVTHKHVYVHVPPPEPEYQAPRKPLYV	66						
Clustal Consensus		:*: . : : *	: *	: *	: *	: *	: *	: *	: *
		80	90	100	110	120	130	140	
1QFT	71	NNADTNMQEATEKVTAVKMYGYNRENAFRYETEDGQVFTDVLAYSDDNCDVIYVEGTDGNEE-----GY	134						
TweedleE	67	PPPQKHVKIVFIKAPSPVPTAPVIFQEPONEEKTLLVYVLVKKPEEQPEIIPTEAPTQPSKPEVYFIRY	136						
Clustal Consensus		: : : : : *	: :	: *	: *	: *	: *	: *	: *
		150	160	170	180				
1QFT	135	ELWTTDYDNIEANCLNKFNEYAVGRETRDVFTSACLEIAAA-	175						
TweedleE	137	KTQKEETGPYPNSVAPPAPETYGAPAAPAPSPSSSYGAPSH	178						
Clustal Consensus		: : : : *	: **	: *	: *	: *	: *	: *	: *

g

5

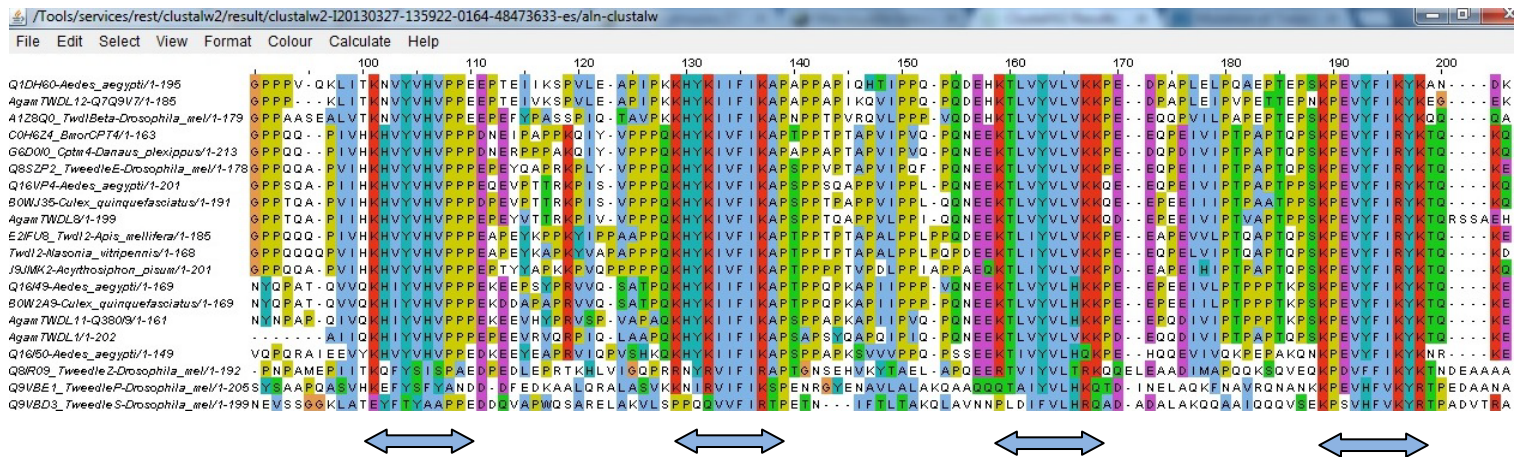
		10	20	30	40	50	60	70	
1RBP	1	--ERDCRVSSFRVKENFDKARFSGTWYAMAKKD---PEGLFLQDNIVAEFSVDETGMQSATAKGRVR-LL	64						
TweedleE	1	RPEPPRDSYSAPPSSSYQPSGPGSGGYGAPAPQYGPQAPVHKHVYVHVPPPEPEYQAPRKPLYVPPQ	70						
Clustal Consensus	1	* * : : ** : * * : : : : : *	19						

		80	90	100	110	120	130	140	
1RBP	65	NNWDVVCADMVGTFDTDEPAKFKMKYWG---VASFLQKGNDD-HWIVDDYDITYAVQYSCRLNLDGTC	129						
TweedleE	71	KHYKIVFIKAPSPFVPTAPVIFQEPONEEKTIVYVLVKKPEEQPEIIPTEAPTQPSKPEVYFIRYKTQK	140						
Clustal Consensus	20	: : : : * : : * : * : * : * : : : *	42						

		150	160	170	180				
1RBP	130	ADSYSTVFSTRDENGLEPEAQKIVRQEQEELCLARQYRLVINGYCD	175						
TweedleE	141	EETGPYPNSVAP---PAPEYGAPAAPAPSPSSSYGAPSH-----	178						
Clustal Consensus	42	: : : * * * : : * *	51						

Εικόνα 59: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της πρωτεΐνης της οικογένειας Tweedle, TweedleE (UniProt: Q8SZP2), από τον οργανισμό *Drosophila melanogaster* και πρωτεϊνών της οικογένειας των λιποκαλινών (a. 1AQB, b. 1BBP, c. 1BJ7, d. 1NGL, e. 1PEE, f. 1QFT και g. 1RBP). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτίδιο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.

Τέλος, οι πρωτεΐνες της οικογένειας Tweedle στοιχίστηκαν μεταξύ τους ώστε να παρατηρήσουμε τις τέσσερις (4) συντηρημένες ομάδες (blocks) που τις ταυτοποιούν σαν δομική οικογένεια πρωτεϊνών (Guan, Middlebrooks et al. 2006; Willis 2010). Επιλέχθηκαν δειγματολογικά όσες πρωτεΐνες είχαν μήκος μέχρι διακόσια (200) αμινοξικά κατάλοιπα (συνολικά είκοσι πρωτεΐνες) για την πραγματοποίηση του πειράματος. Το πείραμα πραγματοποιήθηκε χρησιμοποιώντας το πρόγραμμα ClustalW και τα αποτελέσματά του φαίνονται στην εικόνα 60. Είναι εμφανές ότι υπάρχει μεγάλη συντήρηση στις γνωστές συντηρημένες ομάδες σε ένα μεγάλο αριθμό πρωτεϊνών.



Εικόνα 60: Αποτέλεσμα της στοιχίσης είκοσι (20) πρωτεϊνών Tweedle από το πρόγραμμα ClustalW (με τη χρήση του Jalview). Αριστερά φαίνονται τα ονόματα των πρωτεϊνών μαζί με τους κωδικούς UniProt και στη συνέχεια φαίνονται οι τέσσερις διατηρημένες ομάδες (blocks) σημειωμένες με βέλη.

Οι πρωτεΐνες της οικογένειας CPAP3 στοιχήθηκαν αρχικά με την λιποκαλίνη bovine Retinol Binding Protein (PDB ID: 1FEN). Είναι εμφανές ότι όλες οι πρωτεΐνες του δείγματος εμφανίζουν μια ταυτότητα σε επίπεδο ακολουθίας, με επικρατούσες τιμές περίπου 14%. Η ομοιότητα σε επίπεδο ακολουθίας είναι σαφώς υψηλότερη περίπου στο 50% και ψηλότερα (πίνακας 12). Οι στοιχίσεις φαίνονται στην εικόνα 61 (a-h).

UniProt	Οργανισμός	Αρ. αμινοξέων	Sequence Identity	Sequence Similarity
A1YR29	<i>Tribolium castaneum</i>	218	16.4%	49.2%
E0VTP0	<i>Pediculus humanus</i>	200	10.2%	49.7%
Q5TWB5	<i>Anopheles gambiae</i>	207	11.3%	54.8%
G6D9I3	<i>Danaus plexippus</i>	221	13.6%	54.2%
J9JZY1	<i>Acyrtosiphon pisum</i>	226	16.9%	57.6%
Q16VF6	<i>Aedes aegypti</i>	207	11.9%	48.0%
Q8T0V6	<i>Drosophila melanogaster</i>	218	16.4%	46.3%
Q16NQ0	<i>Aedes aegypti</i>	194	15.3%	48.6%

Πίνακας 12: Αποτελέσματα μετά από στοίχιση των αμινοξικών αλληλουχιών του αντιπροσωπευτικού δείγματος των πρωτεϊνών της οικογένειας CPAP3 με την λιποκαλίνη bovine retinol binding protein (PDB ID: 1FEN). Στην πρώτη στήλη είναι ο κωδικός της UniProt, στην δεύτερη στήλη είναι ο οργανισμός που προέρχεται η πρωτεΐνη, στην τρίτη στήλη είναι ο αριθμός των αμινοξέων (μετά την αφαίρεση του πεπτιδίου οδηγητή), στην τέταρτη στήλη είναι η ποσοστιαία ταυτότητα μεταξύ των συγκρινόμενων ακολουθιών (Sequence Identity – SI) και στην πέμπτη στήλη είναι η ποσοστιαία ομοιότητα (Sequence Similarity – SS) μεταξύ των συγκρινόμενων ακολουθιών. Τα αποτελέσματα αυτά προήλθαν χρησιμοποιώντας το πρόγραμμα ClustalW (Thompson, Higgins et al. 1994).

a		10	20	30	40	50	60	
	1FEN	1	ERDCRVSS---	FRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVA	EFSDENGHMSAT	56		
	A1YR29	1	QENFKCPDDFGFYPHHTSCDKYWKCDNNVAELKTCGNGLAFDASDPKFLTENC	DIYHNVD	60			
	Clustal Consensus		::: : ..	: : *	... *	* : ** ::	: : : :	..
		70	80	90	100	110	120	
	1FEN	57	AKGRVRLIN--	NWDVCADMVGTFDTEDPAKFKMKYWG	VASFLOKGNDDHWIIDTDYET	114		
	A1YR29	61	CGDRTQLEPPISSPHCERLYGIEADESKCDVFNWCNNGEASRYQCSPGLAYDREARVCMW	120				
	Clustal Consensus		. *. : *	: * * : *	.. *	: * ** *	. . : :	:
		130	140	150	160	170	180	
	1FEN	115	AVQY-SCRLNLNLDG-----	TCADSYSFVFARDPSGFSPEVQKIVRQROEELC-LARQ	164			
	A1YR29	121	ADOVPECKNEEVAGGFTCPAPGEVSNSSGFSRHAHPDDCRKYICLEGTA	REYGCPIGT	180			
	Clustal Consensus		* * . * : : *	: : * **	.*..	:	: * * :	..
		190	200	210				
	1FEN	165	YRLIPHN--CYCDGK-----	177				
	A1YR29	181	FKIGDADGTGNCEDPEDVPGCEDYYKDVDLKALKKLG	218				
	Clustal Consensus		::: : * * :					
b		10	20	30	40	50	60	70
	1FEN	1	ERDCRVSSFRVKENFDKARFAGTWYAMAK-----	KDPEGLFLQDN-----	IVA	EFSDENGHMS	54	
	E0VTP0	1	-QNFQCPGLGVFT-YPDPGLCDHFLYLCVNGTLTYEQCENG	LFLDGTGNVHNHCNYNWAVNCGDRKPDYTP	68			
	Clustal Consensus		:: : : : *	: . . . : *	:	: ** : : .	. . . : : :	
		80	90	100	110	120	130	140
	1FEN	55	ATAKRVRLINNWDVCADMVGTFDTEDPAKFKMKYWG	VASFLOKGNDDHW--IIDTDYETFAVQYS	121			
	E0VTP0	69	ISSPGCEYQFGIYPIGQECNTNEVKCEFGVPLPEPCLPGLVYDERIHTCNWPDMLLEKCNPEEII	GFKCP	138			
	Clustal Consensus		:: * : : : :	. * . *	:	: : : : *	: : .	: : *
		150	160	170	180	190	200	
	1FEN	122	LLNLDTGTCADSYSFVFARDPS-GFSPEVQKIVRQROEELCLARQYRLIPHNGYCDGK-----	177				
	E0VTP0	139	TNPVEDSVAARY-WPEPRFPVPDCGRLITCVNGYPRLITCGDGNVYDETTGLCSSPEHVP	PHW	200			
	Clustal Consensus		: : * * : * * *	. . . : *	:	. * * .		
c		10	20	30	40	50	60	70
	1FEN	1	-----ERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVA	EFSDENGHMSAT	AKGRVRL	64		
	Q5TWB5	1	AAAPSCPEPYGEQAYLHPDHCDOFFLCTNGTLTLETCE	NGLLFDGKGAVHNHCNYNWAVDCGRKFATDQ	70			
	Clustal Consensus		. : : : *	: . .	: : ** : : .	: * : . : :		
		80	90	100	110	120	130	140
	1FEN	65	N--NWDVCADMVGTFDTEDPAK--FKMKYWG	VASFLOK-----	NDDHWIIDTDYETFAVQYS	119		
	Q5TWB5	71	TPLSTPGCEYLFGVYPDAECSTSYHKCAFGVHQLCEPGLVYDHRHGCNWPDL	LDSCNPEAVVG	140			
	Clustal Consensus		. . * : * . : * : :	* : * * :	*	. : : *	. * :	
		150	160	170	180	190	200	
	1FEN	120	CRLLNLDTGTCADSYSFVFARDPSGFSPEVQKIVRQROEELCLARQYRLIPHNGYCDGK-----	177				
	Q5TWB5	141	CPQSVPSGTVAARFWPYPRYPVPGDCHRLITCVDGHPRLITCGEGKVNEESLT	CENPEDVPHACRH	207			
	Clustal Consensus		* . ** * :	. * . : * : .	: . : .	: . * :		

d

		10	20	30	40	50	60	
1FEN	1	--ERDCRVSSSERVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVAEFSDENGHMSATAK						58
G6D9I3	1	QFKCPAKDGOYEDDRQCKFYECVDGAATTKLCPDGLVFDPTIRKINKCDQPFNVDCGDR						60
Clustal Consensus		: : : : : : * : * : * : * : : *						

		70	80	90	100	110	120	
1FEN	59	GRVRLNNWDVCAVMVGTFTDTE-----PAKFKMKYWGVAFLQK						99
G6D9I3	61	TELQPPKPSPICPRRNCFEAFHPDPSVCNVFMNCIEGEAIEVKCTAGLHFDDEYSGTCVWPD						120
Clustal Consensus		: : : : * : * : : : * : : : : :						

		130	140	150	160	170	180	
1FEN	100	GNDDHWIIDTDYETFAVQYSCRLLNLDG-----TCADSYSFVFARDPSGFS						145
G6D9I3	121	AAGRSGCAETEEKTKDGFECPKNQVDPOGQAVAHPKFPHPNDCQRFYVCLNGVEPRDLG						180
Clustal Consensus		: : : * : * : : : * : * : : *						

		190	200	210	220			
1FEN	146	PEVKIVRQROEELCLARQYRLIPHNGYCDGK-----				177		
G6D9I3	181	CTVGEVYNE-ESOKCDAPENVRGCEWYKDAEDAPPAPKGRS				221		
Clustal Consensus		* : : : : * : : : : * : *						

e

		10	20	30	40	50	60	
J9JZY1	1	QQNNRGEFCQPKPGFYADQIQCDLYYHCSVDGELTEKLCPDGLFLDSSPSHEKODTS						60
1FEN	1	ERDCRVSSSERVKEN--DKARFAGTWYAMAKKD-----PEGLFLQDNIVAEFSDEN						50
Clustal Consensus		: : : : * : : : * : : * : * : * : : *						

		70	80	90	100	110	120	
J9JZY1	61	VNVDGQRTVQOEPKPSKGCPRANGYYRHWDGVCDEKVNVDGNANEMPCPPGLVYDDS						120
1FEN	51	G-----HMSATAKGRVR--LLNNWD--VCADMVGTFT--DTEDPAKFKMKYWG--						92
Clustal Consensus		: : : * : : : * : * : * : * : : *						

		130	140	150	160	170	180	
J9JZY1	121	TSSCAWATDSKRQCTTKRDALTDGFTCPDGDVVGPNRILPHPTFAHPDDCKEYICRN						180
1FEN	93	VASFLQKGNDDHWIIDTDYETFAVQYSCR--LLNLDGTCADSYSFVFARDPSGFSPEVQ						149
Clustal Consensus		: * : : : * : : : : : * : : : * : *						

		190	200	210	220			
J9JZY1	181	GVIPQYSGSCSAGTVYNDVSFKCDDPENVPGCENYENEDKKGGKN				226		
1FEN	150	KIVROR-----QEELCLARQYRLIPHNGYCDGK-----				177		
Clustal Consensus		: : * : : : : * : * : :						

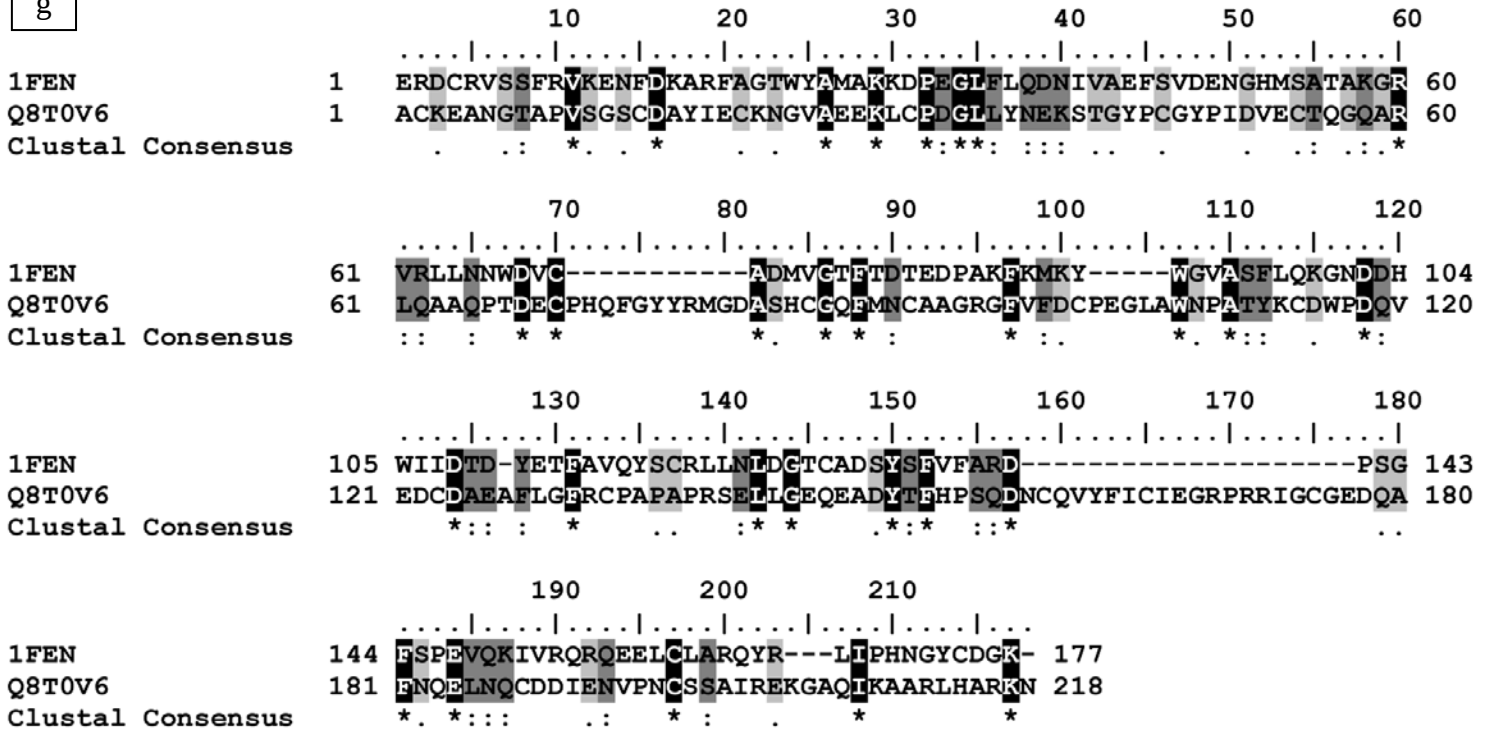
f

		10	20	30	40	50	60	70	
Q16VF6	1	AAVPQCPEPYGEQAYLHPDHCDQFFLCTNGTLTLETCENGLEFDGKGAVHNHCNYNWAVDCGNRKFVTDQ							70
1FEN	1	-----ERDCRVSSSERVKENFDKARFAGTWYAMAKKDPEGLFLQDNIVAEFSDENGHMSATAKGRVRL							64
Clustal Consensus		: : : : * : : : : : * : : : *							

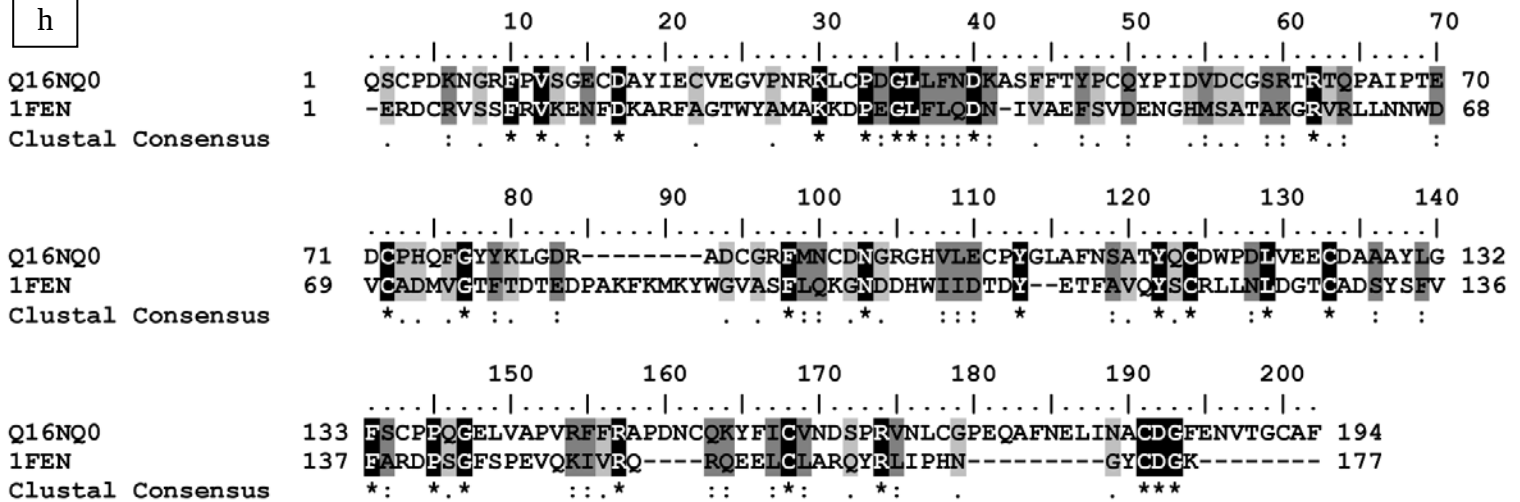
		80	90	100	110	120	130	140	
Q16VF6	71	TPISTPGCEYLEGCIYPSHSCSTTYNKCAYGEPHPEHCEAGLVYDHRHIGCQWPDKMLETCPNPDVAVGFK							140
1FEN	65	N--NWDVCAVMVGTFTDTEPAKFKMKYWGVAFLQKGN--DHWIIDTDYETFAVQYSCRLLNLDGT							128
Clustal Consensus		: : * : * : * : : * : * : : : *							

		150	160	170	180	190	200	
Q16VF6	141	CPDSVPSNSINHRFWPYPRYALPGDCHRLITCVDGHPRLITCCEGKVFNEESLTCEDPEDAPYACRH						207
1FEN	129	CADSYSFVFARDPSGFSPEVQKIVRQROEELCLARQYRLIPHNG-----YCDGK-----						177
Clustal Consensus		* : * : : * : : : * : : * : *						

g



h

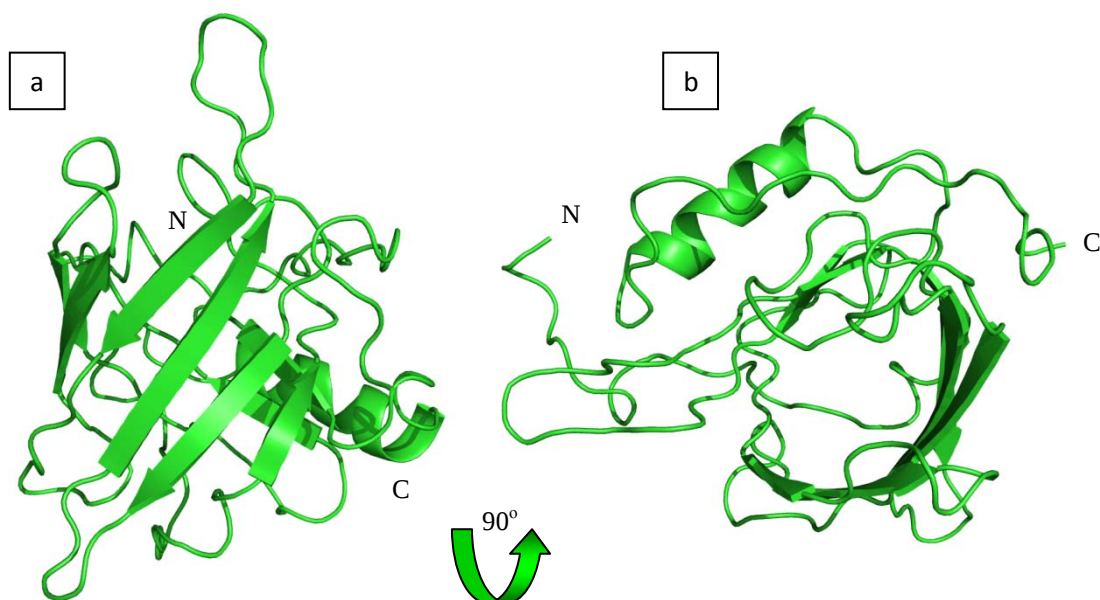


Εικόνα 61: Αποτελέσματα στοιχίσεων, με χρήση του προγράμματος ClustalW (Thompson, Higgins et al. 1994), μεταξύ της λιποκαλίνης RBP (PDB ID: 1FEN) και πρωτεϊνών της οικογένειας CPAP3 (a. A1YR29, b. E0VTPO, c. Q5TWB5, d. G6D9I3, e. J9JZY1, f. Q16VF6 g. Q8T0V6, h. Q16NQ0). Τα αμινοξέα δίνονται με τον κώδικα του ενός γράμματος. Από τις ακολουθίες έχει αφαιρεθεί το πεπτιδίο οδηγητής. Οι σκούροι μαύροι χρωματισμοί αντιστοιχούν σε ταυτόσημα αμινοξέα, ενώ οι σκούροι γκρι και ανοιχτοί γκρι χρωματισμοί αντιστοιχούν σε συντηρητικές αλλαγές 60% και 80% σε επίπεδο σημαντικότητας. Αντίστοιχα συμβολίζονται με (*) η ταυτότητα και με (:) και (.) η ομοιότητα.

3.2 Προτυποποίηση με βάση την ομολογία

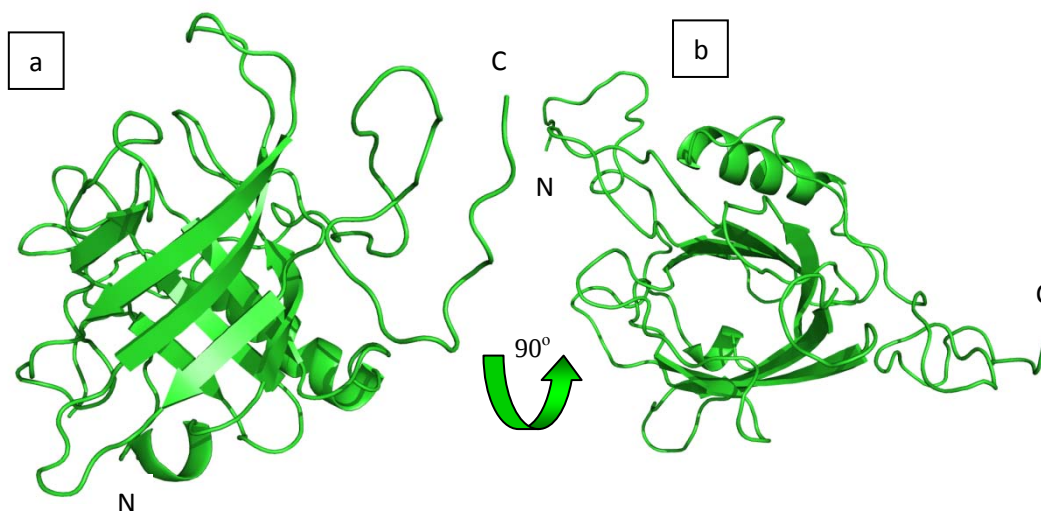
Παρόλο που η ταυτότητα σε επίπεδο ακολουθίας είναι σχετικά μικρή (κάτω από 20%) σε όλες τις στοιχίσεις, η ομοιότητα όμως σε επίπεδο ακολουθίας είναι επαρκής (περίπου 50%) για την κατασκευή δομικών μοντέλων των πρωτεϊνών με την μέθοδο της προτυποποίησης με βάση την ομολογία (Homology modeling). Χρησιμοποιώντας σαν δομή-οδηγό (template) την κρυσταλλογραφικά προσδιορισμένη δομή της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN) (Zanotti, Marcello et al. 1994), κατασκευάστηκαν δομικά μοντέλα των πρωτεϊνών των οικογενειών Tweedle και CPAP3. Το πείραμα πραγματοποιήθηκε με την χρήση του προγράμματος Modeller9v2 (Sali and Blundell 1993).

Στην εικόνα 62 παρουσιάζεται το χαρακτηριστικό μοντέλο της πρωτεϊνικής δομής της TweedleE, στο οποίο φαίνονται τα κοινά δομικά χαρακτηριστικά που έχουν και τα υπόλοιπα μοντέλα της οικογένειας Tweedle (αρκετά από αυτά βρίσκονται στο παράρτημα). Ύστερα από ανάλυση των δομικών χαρακτηριστικών των μοντέλων, φαίνεται ότι το βασικό δομικό μοτίβο των πρωτεϊνών της οικογένειας Tweedle είναι ένα ολόκληρο β-βαρέλι (β -barrel), που αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους (β -strands) καθώς και μια (1) α -έλικα (α -helix) παράλληλη προς αυτούς.



Εικόνα 62: Το δομικό μοντέλο της TweedleE (UniProt: Q8SZP2), του οργανισμού *Drosophila melanogaster*. Στην (a) φαίνεται η δομή από το πλάι ενώ στην (b) φαίνεται από πάνω. Τα N και C αντιστοιχούν στο αμινικό και καρβοξυτελικό άκρο αντίστοιχα. Το βασικό δομικό μοτίβο των πρωτεϊνών της οικογένειας Tweedle είναι ένα ολόκληρο β-βαρέλι (β-barrel), που αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους (β-strands) καθώς και μια (1) α-έλικα (α-helix) παράλληλη προς αυτούς. Η οπτικοποίηση έγινε με το πρόγραμμα PyMOL Molecular Graphics System(Delano 2005).

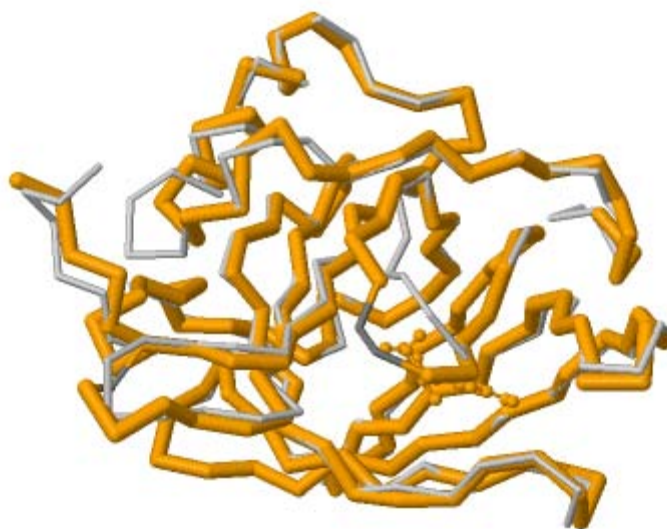
Ομοίως, δημιουργήθηκε το πρωτεϊνικό μοντέλο της πρωτεΐνης Q16VF6 (Nene, Wortman et al. 2007) της οικογένειας CPAP3. Στην εικόνα 63 φαίνονται τα δομικά χαρακτηριστικά του μοντέλου, το οποίο εμφανίζει παρόμοια με των μοντέλων της οικογένειας Tweedle. Δηλαδή, φαίνεται ότι το βασικό δομικό μοτίβο των πρωτεϊνών της οικογένειας CPAP3 είναι ένα ολόκληρο β-βαρέλι (β-barrel), που αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους (β-strands), μια α-έλικα (α-helix) παράλληλη προς αυτούς και μια μικρή α-έλικα στη βάση του βαρελιού.



Εικόνα 63: Το πρωτεϊνικό μοντέλο της πρωτεΐνης Q16VF6, του οργανισμού *Aedes aegypti*. Στην (a) φαίνεται η δομή από το πλάι ενώ στην (b) φαίνεται από πάνω. Η οπτικοποίηση έγινε με το πρόγραμμα PyMOL Molecular Graphics System(Delano 2005).

3.3 Δομική στοίχιση

Πραγματοποιήθηκε δομική στοίχιση (structure alignment) μεταξύ των δημιουργηθέντων μοντέλων των πρωτεϊνών της οικογένειας Tweedle (ύστερα από την προτυποποίηση με βάση την ομολογία) με την λυμένη κρυσταλλογραφικά δομή της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN) (εικόνες 64 και 65). Η στοίχιση έγινε ώστε να εξεταστεί η ενδεχόμενη ύπαρξη των εσωτερικών και εξωτερικών υδρόφοβων συστάδων (inner and outer clusters) στα μοντέλα, όπως υπάρχουν και στην Retinol Binding Protein της οικογένειας των λιποκαλινών (Adam, Charloteaux et al. 2008).



Εικόνα 64: Το αποτέλεσμα της δομικής στοίχισης μεταξύ της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN) και του μοντέλου της πρωτεΐνης TweedleE (UniProt: Q8SZP2), που προέρχεται από τον οργανισμό *Drosophila melanogaster*, σύμφωνα με το πρόγραμμα Combinatorial Extension method (Shindyalov and Bourne 1998). Με γκρι και πορτοκαλί χρώμα φαίνονται οι δυο πρωτεϊνικές δομές.

■ EQR:173Len1:176Len2:178score: 395,51Z-score:7,02RMSD:0,85SeqID:10%SeqSim:23%Cov1:98%Cov2:97%

1:A	ERDCRVSSFRVKENFDKARFAGTWYAMAKKDPEGLFLQDNI VAEFSVDENGHMSATAKGRVRLNNWDVC	70:A
3:	EPPRDSYSAPPSSSYQPSGPGSGGYGAPAPQY- - GPPQQAPVI HKHVVYVHVPPPEPEYQAPRKPLYVPPPQ	70:
71:A	ADMVGTFTDTEPAKFMMKYWGVASFLOKGNDDHWI DTDYETF AVQYSCRLNLNDGTCADSYSFVFARD	140:A
71:	KHYKIVFVKAPSPVPPTAPVIPQFPQNEEKLTVYVLVKKPEEQPEI I I PTPAPTQP- SKPEVYFIRYKTQ	139:
141:A	PSGFSPEVQKIVRQRQEE- LCLARQYRLI PHNGYCDG	176:A
140:	KEETGPYPNSVAPPAPPEYGAPAAPPAPSAPSSSYGAP	176:

Εικόνα 65: Το αποτέλεσμα της δομικής στοίχισης μεταξύ των αμινοξικών καταλοίπων μεταξύ της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN) και του μοντέλου της πρωτεΐνης TweedleE (UniProt: Q8SZP2), που προέρχεται από τον οργανισμό *Drosophila melanogaster*, σύμφωνα με το πρόγραμμα Combinatorial Extension method (Shindyalov and Bourne 1998). Με υπογράμμιση δίνεται η τιμή του δείκτη RMSD.

Σύμφωνα με τα αποτελέσματα του πειράματος, σημειώθηκαν ποιά αμινοξικά κατάλοιπα είναι υδρόφοβα ώστε να συγκριθούν με τις αντίστοιχες υδρόφοβες εσωτερικές (inner) και εξωτερικές (outer) συστάδες (clusters) της λιποκαλίνης bovine Retinol Binding Protein (PDB ID: 1FEN).

Τα αποτελέσματα δείχνουν υψηλή συντήρηση υδρόφοβων αμινοξικών καταλοίπων στο εσωτερικό των υποθετικών εσωτερικών και εξωτερικών συστάδων (inner and outer clusters) στα δομικά μοντέλα της οικογένειας Tweedle, όπως ισχύει και στην οικογένεια των λιποκαλινών (πίνακες 13 και 14).

Γενικές θέσεις λιποκαλινών	1FEN	<i>Anopheles gambiae</i>		<i>Drosophila melanogaster</i>						<i>Danaus plexippus</i>	<i>Bombyx mori</i>
		Agam TWDL8	Agam TWDL12	TweedleE	TweedleJ	TweedleQ	TweedleS	TweedleZ	Tweedle Beta	CPTmotif4	BmorCPT4
39	F20	G20	L20	P22	L51	T49	T20	L33	A16	G20	S19
48	W24	G24	E24	Y26	Q55	I53	T24	T37	E20	L30	G23
80	F45	A45	P49	H45	Y76	F74	L47	E62	Y41	V51	H44
91	M53	T53	P57	P53	Y84	I83	F55	L70	A49	G66	I52
93	A55	A55	K59	P55	P86	S85	R57	P72	V51	P68	A54
115	F77	R77	P81	F77	Q123	A115	L82	I94	I73	H90	F72
131	F86	Y89	K90	P86	Y132	V124	L91	V103	Y82	D99	P81
133	M88	I91	Y92	A88	L134	S126	K93	Q105	I84	E101	A83
156	H104	K116	Q108	Y104	P158	N142	H109	A120	T94	H117	Y99
158	I106	L118	I110	L106	V160	A144	V111	A122	V96	K119	L101
168	A115	D127	K119	E115	A177	F153	T120	Y131	Q105	P128	E110
170	Q117	P129	L121	I117	Q179	K155	A122	Q133	E107	A130	V112
190	F137	F149	E141	Y136	S204	S180	I142	I152	L127	L151	K132
192	R139	R151	T143	T138	P206	S182	K144	A154	A129	K153	Q134

Πίνακας 13: Εσωτερική συστάδα (Inner cluster). Οι αλλαγές των αμινοξικών καταλοίπων σε σύγκριση με τα αμινοξικά κατάλοιπα που βρίσκονται στην εσωτερική υδρόφοβη συστάδα της λιποκαλίνης RBP (PDB ID: 1FEN). Η πρώτη στήλη περιέχει τις γενικές θέσεις των αμινοξέων στην εσωτερική συστάδα των λιποκαλινών (Adam, Charloteaux et al. 2008). Η δεύτερη στήλη δείχνει τα αμινοξικά κατάλοιπα της εσωτερικής συστάδα στην λιποκαλίνη bovine RBP (PDB ID: 1FEN). Οι υπόλοιπες στήλες παρουσιάζουν, ύστερα από την δομική στοίχιση με την μέθοδο Combinatorial Extension (Shindyalov and Bourne 1998), τα αμινοξικά κατάλοιπα των μοντέλων που βρίσκονται στο εσωτερικό της υποθετικής εσωτερικής συστάδας (στις κορυφές των στηλών φαίνονται οι οργανισμοί που προέρχονται καθώς και τα ονόματα των πρωτεϊνών). Τα υδρόφοβα κατάλοιπα (A, V, I, L, M, F, P, Y, W, G) είναι χρωματισμένα με γκρι. Είναι προφανές ότι οι υποθετικές εσωτερικές συστάδες των μοντέλων παρουσιάζουν το ίδιο υδρόφοβο περιβάλλον με τις εσωτερικές συστάδες των λιποκαλινών.

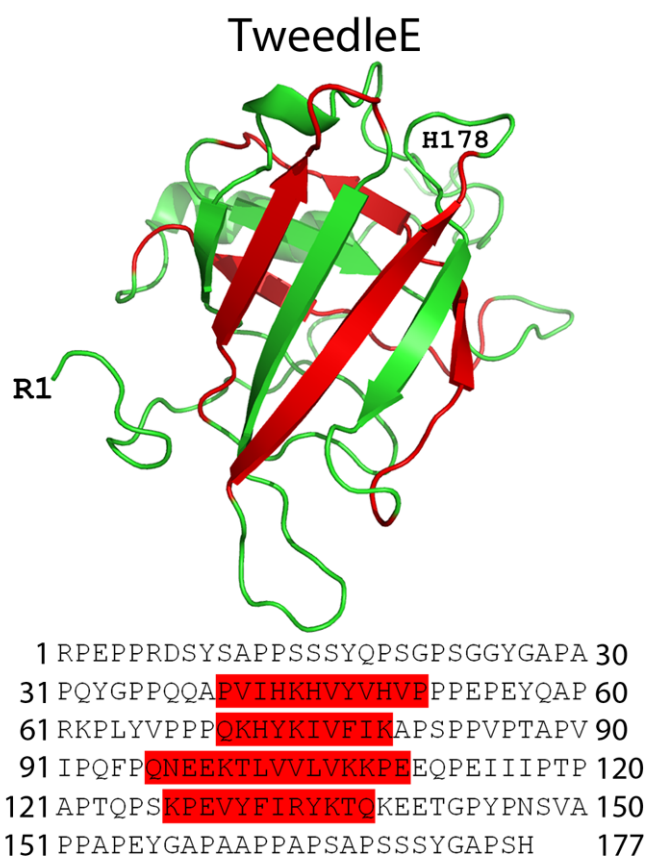
Γενικές θέσεις λιποκαλινών	1FEN	<i>Anopheles gambiae</i>		<i>Drosophila melanogaster</i>						<i>Danaus plexippus</i>	<i>Bombyx mori</i>
		Agam TWDL8	Agam TWDL12	TweedleE	TweedleJ	TweedleQ	TweedleS	TweedleZ	Tweedle Beta	CPTmotif4	BmorCPT4
52	A28	G28	P28	A30	A59	A57	P28	V41	P24	C34	P27
159	I107	V119	P111	V107	H161	Q145	K112	A123	R97	I120	V102
169	V116	E128	T120	I116	I178	V154	R121	D132	D106	P129	I111
171	Y118	E130	V122	I118	G180	Y156	L123	L134	H108	P131	I113
189	V136	Y148	P140	R135	A203	S179	V151	V151	I126	V150	Y131
205	V152	A174	K154	A151	P222	I200	V167	V167	I143	P166	P147
220	Y165	P187	G167	A165	Y235	Q213	D180	D180	Y156	Y181	Y160

Πίνακας 14: Εξωτερική συστάδα (Outer cluster). Οι αλλαγές των αμινοξικών καταλοίπων σε σύγκριση με τα αμινοξικά κατάλοιπα που βρίσκονται στην εξωτερική υδρόφοβη συστάδα της λιποκαλίνης RBP (PDB ID: 1FEN). Η πρώτη στήλη περιέχει τις γενικές θέσεις των αμινοξέων στην εξωτερική συστάδα των λιποκαλινών (Adam, Charloreaux et al. 2008). Η δεύτερη στήλη δείχνει τα αμινοξικά κατάλοιπα της εξωτερικής συστάδας στην λιποκαλίνη bovine RBP (PDB ID: 1FEN). Οι υπόλοιπες στήλες παρουσιάζουν, ύστερα από την δομική στοίχιση με την μέθοδο Combinatorial Extension (Shindyalov and Bourne 1998), τα αμινοξικά κατάλοιπα των μοντέλων που βρίσκονται στο εσωτερικό της υποθετικής εξωτερικής συστάδας (στις κορυφές των στηλών φαίνονται οι οργανισμοί που προέρχονται καθώς και τα ονόματα των πρωτεϊνών). Τα υδρόφοβα κατάλοιπα (A, V, I, L, M, F, P, Y, W, G) είναι χρωματισμένα με γκρι. Είναι προφανές ότι οι υποθετικές εξωτερικές συστάδες των μοντέλων παρουσιάζουν παρόμοιο υδρόφοβο περιβάλλον με τις εξωτερικές συστάδες των λιποκαλινών.

3.4 Δομική ανάλυση και άλλες υπολογιστικές μελέτες

Χρησιμοποιώντας το πρόγραμμα οπτικοποίησης PyMOL Molecular Graphics System (Delano 2005), πραγματοποιήθηκαν δομικές αναλύσεις στα πρωτεϊνικά μοντέλα της οικογένειας Tweedle.

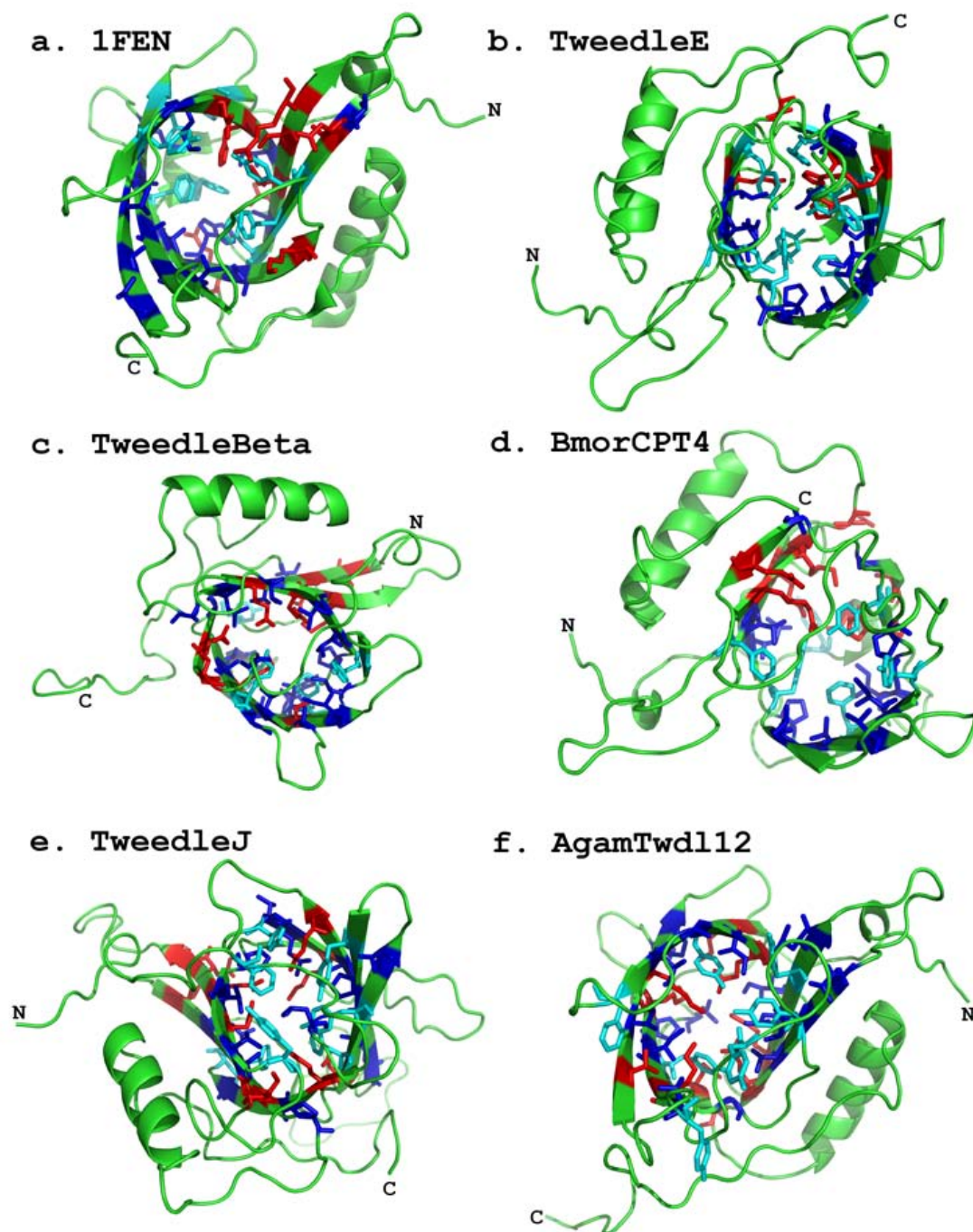
Συγκεκριμένα, παρατηρήθηκε με ποιές δομικές περιοχές ταυτίζονται οι τέσσερις (4) συντηρημένες αμινοξικές περιοχές (blocks) των πρωτεϊνών της οικογένειας Tweedle (Guan, Middlebrooks et al. 2006). Στην μεγάλη πλειοψηφία των πρωτεϊνικών μοντέλων, αυτές οι αμινοξικές περιοχές συμπίπτουν με τα τέσσερα (4) παράλληλα πτυχωτά β-φύλλα (εικόνα 65) και είναι σε συμφωνία με τις προβλέψεις (Guan, Middlebrooks et al. 2006) που είχαν πραγματοποιηθεί με το λογισμικό PhD (Rost 1996).



Εικόνα 65: Οι τέσσερις (4) διατηρημένες αμινοξικά περιοχές (blocks) των πρωτεϊνών της οικογένειας Tweedle (στην εικόνα είναι χρωματισμένες με κόκκινο χρώμα), φαίνεται να διπλώνουν σε β-πτυχωτά φύλλα, σε συμφωνία με τις προηγούμενες πειραματικές προβλέψεις (Guan, Middlebrooks et al. 2006) του λογισμικού PhD (Rost 1996). Η παραπάνω ακολουθία και το παραπάνω πρωτεϊνικό μοντέλο προέρχονται από την πρωτεΐνη TweedleE (κωδικός UniProt: Q8SZP2), του οργανισμού *Drosophila melanogaster*.

Εν συνεχεία, σύμφωνα με την οπτικοποίηση του λογισμικού PyMOL (Delano 2005), έγιναν μελέτες που αφορούσαν την εξέταση του περιβάλλοντος στο εσωτερικό των β-βαρελιών και στην υποθετική εσωτερική συστάδα (inner cluster), στα πρωτεϊνικά μοντέλα, ανάλογα με την αμινοξική τους σύνθεση (υδρόφοβα, πολικά, φορτισμένα). Ακόμα εξετάστηκε η παρουσία αρωματικών αμινοξέων στο εσωτερικό των β-βαρελιών.

Τα αποτελέσματα έδειξαν ότι στα πρωτεϊνικά μοντέλα της οικογένειας Tweedle, το εσωτερικό των β-βαρελιών αποτελείται κυρίως από υδρόφοβα αμινοξικά κατάλοιπα (εικόνα 66). Επίσης, παρατηρήθηκε μια αξιοσημείωτη παρουσία αρωματικών αμινοξικών καταλοίπων (φαινυλαλανίνες – F, τυροσίνες Y) στο εσωτερικό των β-βαρελιών καθώς και στις σχισμές αυτών (εικόνα 66). Φαίνεται ότι δημιουργούν μια επιφάνεια από αρωματικούς δακτυλίους προσαρμοσμένους ειδικά για αλληλεπιδράσεις μεταξύ πρωτεϊνών-χιτίνης στο δερμάτιο (Quirocho 1989; Vyas 1991; Elgavish and Shaanan 1997; Hamodrakas, Kanellopoulos et al. 1997; Svitil and Kirchman 1998; Shen and Jacobs-Lorena 1999).



Εικόνα 66: Τα υδρόφοβα αμινοξικά κατάλοιπα (A, V, I, L, M, F, P, Y, W, G) περιέχονται σε μεγάλη ποσότητα στο εσωτερικό των β-βαρελίων και στην υποθετική εσωτερική συστάδα. Επίσης βρέθηκε ένας αξιοσημείωτος αριθμός από αρωματικά αμινοξικά κατάλοιπα (F, Y) στο εσωτερικό των β-βαρελίων και στις σχισμές αυτών. Στις παραπάνω εικόνες παρουσιάζονται τα υδρόφοβα κατάλοιπα με σκούρο μπλε χρώμα, τα αρωματικά (και υδρόφοβα) με γαλάζιο χρώμα και τα πολικά με κόκκινο χρώμα (b), (c), (d), (e), (f). Επίσης παρουσιάζονται τα αμινοξικά κατάλοιπα της λιποκαλίνης Retinol Binding Protein (PDB ID: 1FEN) στο εσωτερικό του β-βαρελίου και στην εσωτερική συστάδα (a), ώστε να γίνει η σύγκριση με τα πρωτεϊνικά μοντέλα.

Επιπροσθέτως, έγινε χρήση του προγράμματος LipocalinPred (Ramana and Gupta 2009) το οποίο δέχεται μια αμινοξική ακολουθία και προβλέπει αν είναι ή όχι λιποκαλίνη καθώς και το πόσο πιθανό είναι αυτό. Συγκεκριμένα χρησιμοποιήθηκε η μέθοδος SVM (Support Vector Machines) prediction – PSSM model (Position Specific Substitution Matrix). Το λογισμικό θεωρεί τις πρωτεΐνες ως λιποκαλίνες όταν το SVM Score > 0 (όσο μεγαλύτερο είναι τόσο ανεβαίνει η πιθανότητα να είναι λιποκαλίνη). Αξίζει να σημειωθεί ότι η παρούσα εργασία αφορά δομικές πρωτεΐνες του δερματίου οι οποίες φαίνεται να παρουσιάζουν δομικές ομοιότητες με λιποκαλίνες και δεν αφορά λιποκαλίνες, ώστε να έχουμε μόνο θετικά αποτελέσματα. Τα αποτελέσματα του προγράμματος δείχνουν ότι αρκετές πρωτεΐνες της οικογένειας Tweedle θα μπορούσαν να είναι λιποκαλίνες.

Στις πρωτεΐνες του αντιπροσωπευτικού δείγματος της οικογένειας Tweedle, το λογισμικό LipocalinPred θεωρεί όλες τις πρωτεϊνικές ακολουθίες ότι είναι λιποκαλίνες (10/10 – 100%) (εικόνα 67).



PREDICTION RESULT

Your Job ID is : 177307
 Your input seq(s) : Input file

Note: Higher the SVM score, better is the confidence level of prediction. The confidence can also be taken care by using a high value of 'threshold' at submission page.

S.No.	Sequence IDs	SVM Score	Decision
PSSM model			
1	AgamTWDL8-Anopheles_gambiae	0.18988037	Lipocalin
2	AgamTWDL12-Anopheles_gambiae	0.30847747	Lipocalin
3	C0H6Z4_BmorCPT4-Bombyx_mori	0.45445678	Lipocalin
4	A1Z8Q0_TwdlBeta-Drosophila_melanogaster	0.3628591	Lipocalin
5	Q8SZP2_TweedleE-Drosophila_melanogaster	0.64699746	Lipocalin
6	Q9VBD6_TweedleJ-Drosophila_melanogaster	0.10943567	Lipocalin
7	Q9VBC8_TweedleQ-Drosophila_melanogaster	0.030340762	Lipocalin
8	Q9VBD3_TweedleS-Drosophila_melanogaster	0.32354228	Lipocalin
9	Q8IR09_TweedleZ-Drosophila_melanogaster	0.26855372	Lipocalin
10	G6D0I0_Cptm4-Danaus_plexippus	0.91554738	Lipocalin

Εικόνα 67: Το αποτέλεσμα του προγράμματος LipocalinPred (Ramana and Gupta 2009) για τις πρωτεΐνες Tweedle του αντιπροσωπευτικού δείγματος, όπου τις θεωρεί όλες λιποκαλίνες.

Στη συνέχεια έγινε χρήση του προγράμματος και για τις πρωτεΐνες του αντιπροσωπευτικού δείγματος των πρωτεϊνών της οικογένειας CPAP3. Όπως φαίνεται στην εικόνα 68, τις περισσότερες πρωτεΐνες τις χαρακτηρίζει λιποκαλίνες (6/8 – 75%).

LipocalinPred



**PREDICTION
RESULT**

Your Job ID is : 216828

Your input seq(s) : [Input file](#)

Note: Higher the SVM score, better is the confidence level of prediction. The confidence can also be taken care by using a high value of 'threshold' at submission page.

S.No.	Sequence IDs	SVM Score	Decision
PSSM model			
1	A1YR29-Tribolium_castaneum	1.3381772	Lipocalin
2	E0VTP0-Pediculus_humanus_corporis	0.80659421	Lipocalin
3	Q5TWB5-Anopheles_gambiae	1.0397488	Lipocalin
4	G6D9I3-Danaus_plexippus	-0.1644363	Non-lipocalin
5	J9JZY1-Acyrtosiphon_pisum	0.0266151	Lipocalin
6	Q16VF6-Aedes_aegypti	1.8380541	Lipocalin
7	Q8T0V6-Drosophila_melanogaster	-0.29020577	Non-lipocalin
8	Q16NQ0-Aedes_aegypti	0.03682638	Lipocalin

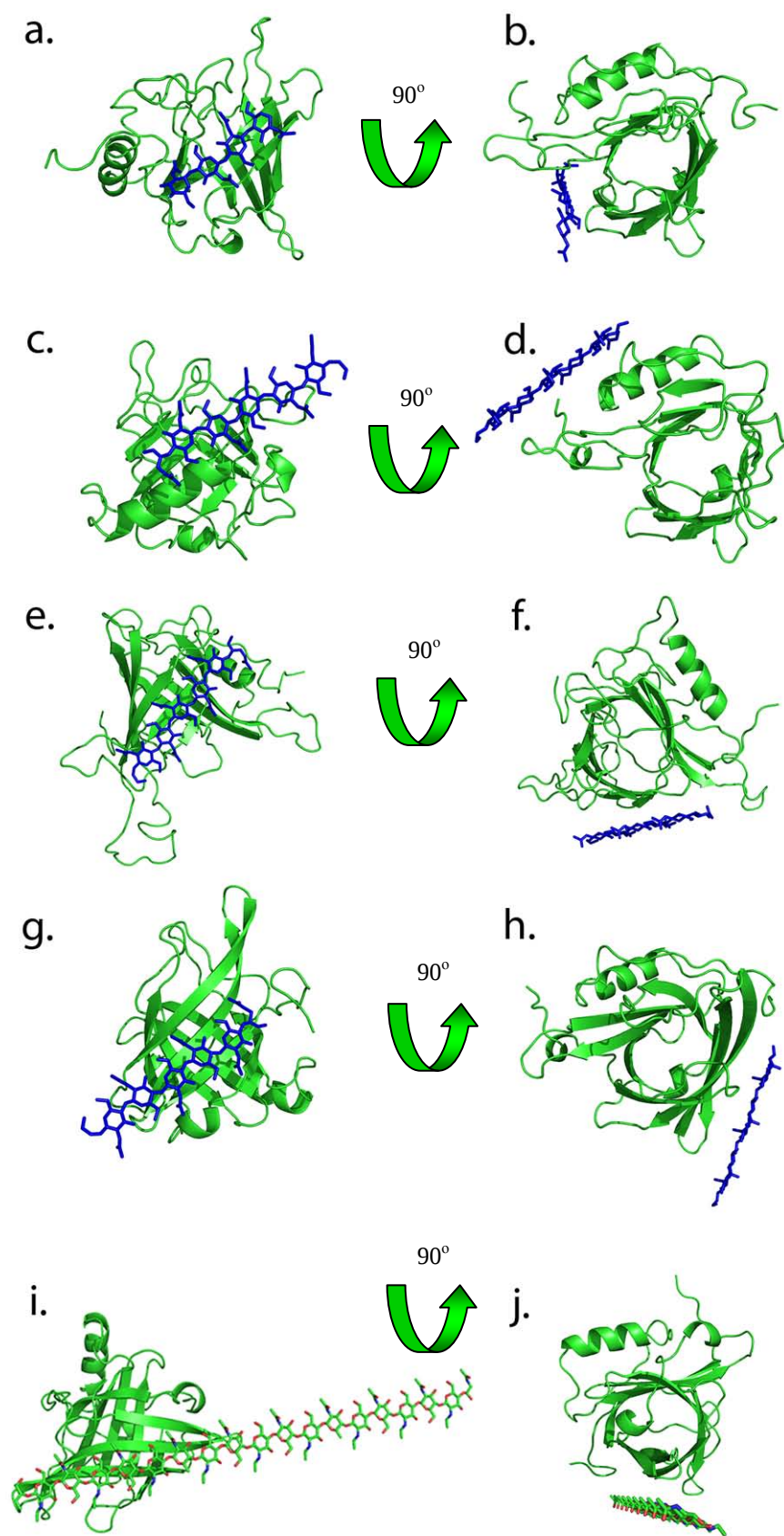
Εικόνα 68: Το αποτέλεσμα του προγράμματος LipocalinPred (Ramana and Gupta 2009) για τις πρωτεΐνες της του αντιπροσωπευτικού δείγματος της οικογένειας CPAP3. Φαίνεται ότι το πρόγραμμα θεωρεί σχεδόν τις έξι από τις οχτώ ως λιποκαλίνες, με μεγαλύτερη πιθανότητα (μεγάλο SVM Score) την πρωτεΐνης Q16VF6, με την οποία γίνανε και τα προηγούμενα δομικά μοντέλα.

3.5 Πειράματα αγκυροβόλησης (Docking)

Πραγματοποιήθηκαν πειράματα αγκυροβόλησης (docking) χρησιμοποιώντας τα γνωστά προγράμματα αγκυροβόλησης GRAMM (Vakser 1996) και Autodock4.2 (Morris, Huey et al. 2009). Μάλιστα το λογισμικό Autodock4.2 έχει τις πιο πολλές αναφορές (citations) στην συγκεκριμένη επιστημονική βιβλιογραφία από οποιοδήποτε άλλο αντίστοιχο λογισμικό αγκυροβόλησης (Sousa, Fernandes et al. 2006).

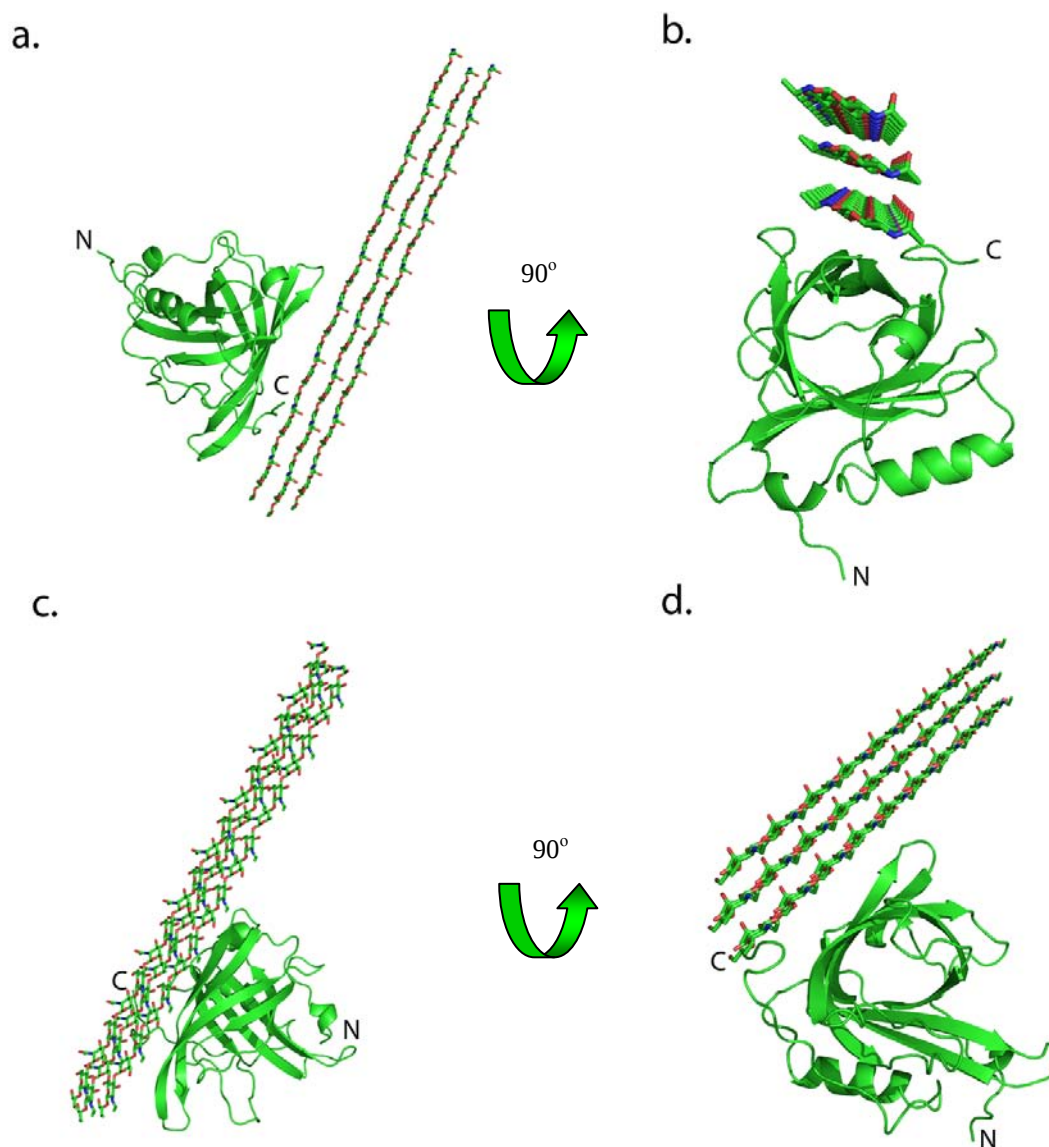
Αρχικά, χρησιμοποιώντας το πρόγραμμα GRAMM (Vakser 1996), πραγματοποιήθηκαν πειράματα αγκυροβόλησης (docking), έχοντας ως πρωτεΐνη-στόχος (target protein) τα δομικά μοντέλα των πρωτεϊνών της οικογένειας Tweedle και χρησιμοποιώντας ως προσδέτη/υποκαταστάτη (ligand) μια ποικιλία από αλυσίδες χιτίνης (συγκεκριμένα: μια αλυσίδα ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG), μια αλυσίδα ομοεξαμερούς N-ακετυλογλυκοζαμίνης (6NAG), μια αλυσίδα ομοδεκαοχταμερούς N-ακετυλογλυκοζαμίνης (18NAG) και επίσης πολλαπλές 18NAG αλυσίδες – όπου NAG: μόριο N-ακετυλογλυκοζαμίνης). Η τεχνική της αγκυροβόλησης ήταν υψηλής ανάλυσης γενική αγκυροβόληση (High-resolution generic docking), χρησιμοποιώντας τον αλγόριθμο Geometric docking II (Katchalski-Katzir, Shariv et al. 1992) και προσαρμόζοντας το βήμα πλέγματος (grid step) στο ελάχιστο δυνατό.

Η πλειοψηφία των πειραμάτων αγκυροβόλησης παρείχαν μια ποικιλία από διαφορετικές στερεοδιατάξεις οι οποίες είναι θερμοδυναμικά ευνοϊκές. Τα αποτελέσματα δείχνουν ότι όλες αλυσίδες χιτίνης που χρησιμοποιήθηκαν στα πειράματα μπορούν να αλληλεπιδράσουν με τα πρωτεϊνικά μοντέλα παράλληλα στους β-κλώνους, σε συμφωνία με προηγούμενες πειραματικές παρατηρήσεις από περιθλασιγράμματα ακτίνων X (Atkins 1985) (εικόνα 69).



Εικόνα 69: Πειραματικά αποτελέσματα του προγράμματος αγκυροβόλησης GRAMM (Vakser 1996) με πρωτεΐνη-στόχο (target protein) διάφορα μοντέλα των πρωτεϊνών της οικογένειας Tweedle χρησιμοποιώντας ως προσδέτη/υποκαταστάτη (ligand) αλυσίδες χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG), ομοεξαμερών N-ακετυλογλυκοζαμίνης (6NAG) και ομοδεκαοχταμερών N-ακετυλογλυκοζαμίνης (18NAG). Σε όλα τα θερμοδυναμικά ευνοϊκά αποτελέσματα, οι αλυσίδες χιτίνης μπορούν να αλληλεπιδράσουν παράλληλα στους β-κλώνους. (a), (b) TweedleE (*Drosophila melanogaster*) με 4NAG. (c), (d) BmorCPT4 (*Bombyx mori*) με 6NAG. (e), (f) TweedleJ (*Drosophila melanogaster*) με 6NAG. (g), (h) TweedleZ (*Drosophila melanogaster*) με 6NAG. (i), (j) TweedleZ (*Drosophila melanogaster*) με 18NAG. Οι αλυσίδες χιτίνης 4NAG και 6NAG είναι χρωματισμένες με μπλε, ενώ στα (i) και (j) είναι με πολλά χρώματα ανάλογα με το είδος του εκάστοτε ατόμου.

Επίσης, παρουσιάζει ιδιαίτερο ενδιαφέρον το αποτέλεσμα του πειράματος αγκυροβόλησης του μοντέλου της πρωτεΐνης TweedleZ με τρεις (3) αλυσίδες ομοδεκαοχταμερών N-ακετυλογλυκοζαμίνης (18NAG) (με βήμα πλέγματος 3.0). Φαίνεται ότι οι ίνες χιτίνης μπορούν να αλληλεπιδράσουν παράλληλα στους β-κλώνους καθώς και παράλληλα μεταξύ τους (εικόνα 70). Αυτό το εντυπωσιακό αποτέλεσμα είναι σε συμφωνία με το προτεινόμενο μοντέλο κρυστάλλου χιτίνης (Atkins 1985) καθώς και με την προτεινόμενη δομή του συμπλόκου χιτίνης-πρωτεϊνών (Blackwell and Weih 1980).

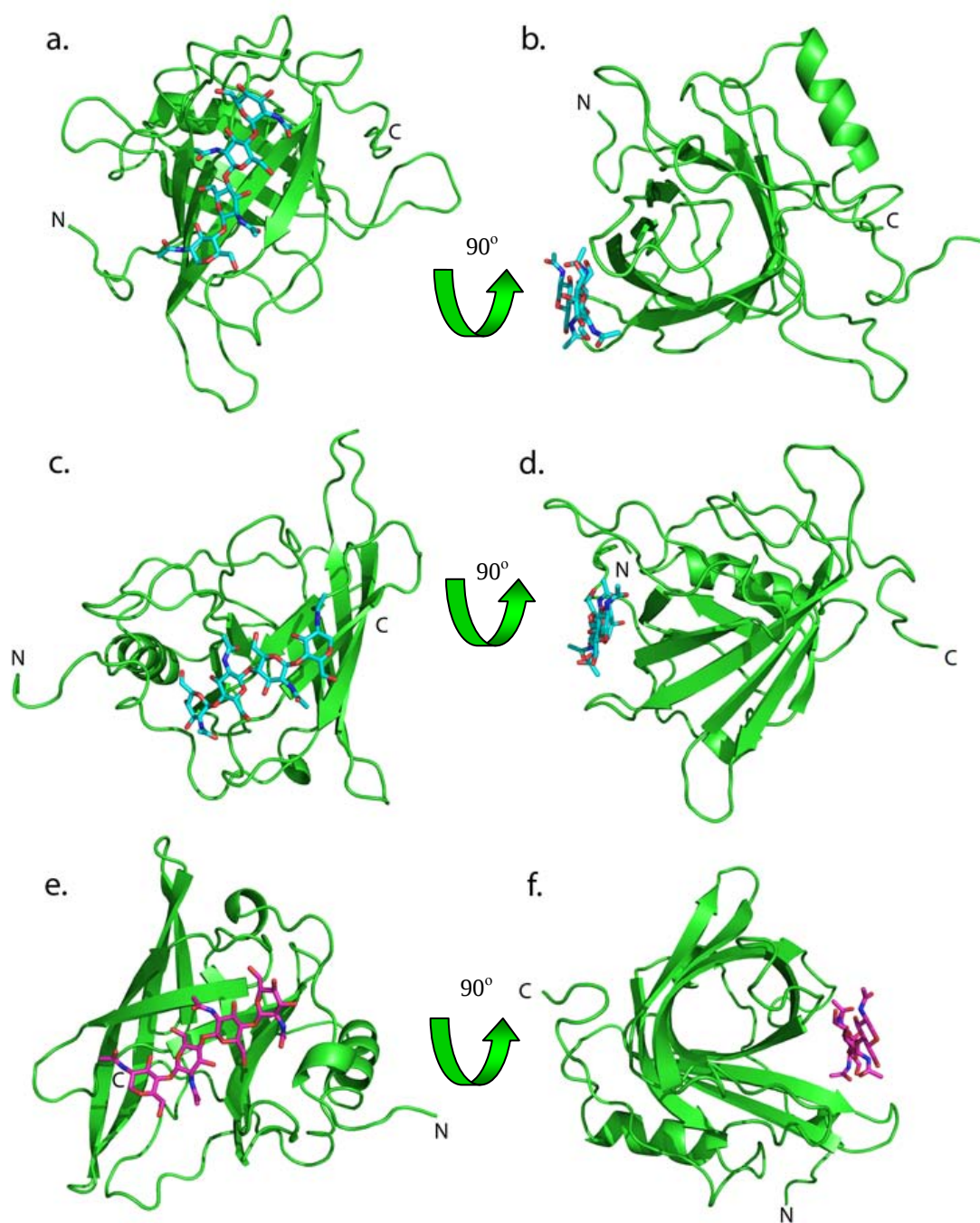


Εικόνα 70: Πειραματικά αποτελέσματα του προγράμματος αγκυροβόλησης GRAMM (Vakser 1996) με πρωτεΐνη-στόχο (target protein) το μοντέλο της πρωτεΐνης TweedleZ (*Drosophila melanogaster*) χρησιμοποιώντας ως προσδέτη/υποκαταστάτη (ligand) τρεις (3) αλυσίδες χιτίνης ομοδεκαοχταμερών N-ακετυλογλυκοζαμίνης (18NAG) (με βήμα πλέγματος 3.0). Το αποτέλεσμα δείχνει ότι οι τρεις (3) αλυσίδες χιτίνης ομοδεκαοχταμερών N-ακετυλογλυκοζαμίνης (18NAG) μπορούν να αλληλεπιδράσουν παράλληλα στους β-κλώνους καθώς και στον εαυτό τους. Τα (a), (b), (c), (d) είναι το ίδιο αποτέλεσμα από διαφορετικές οπτικές γωνίες.

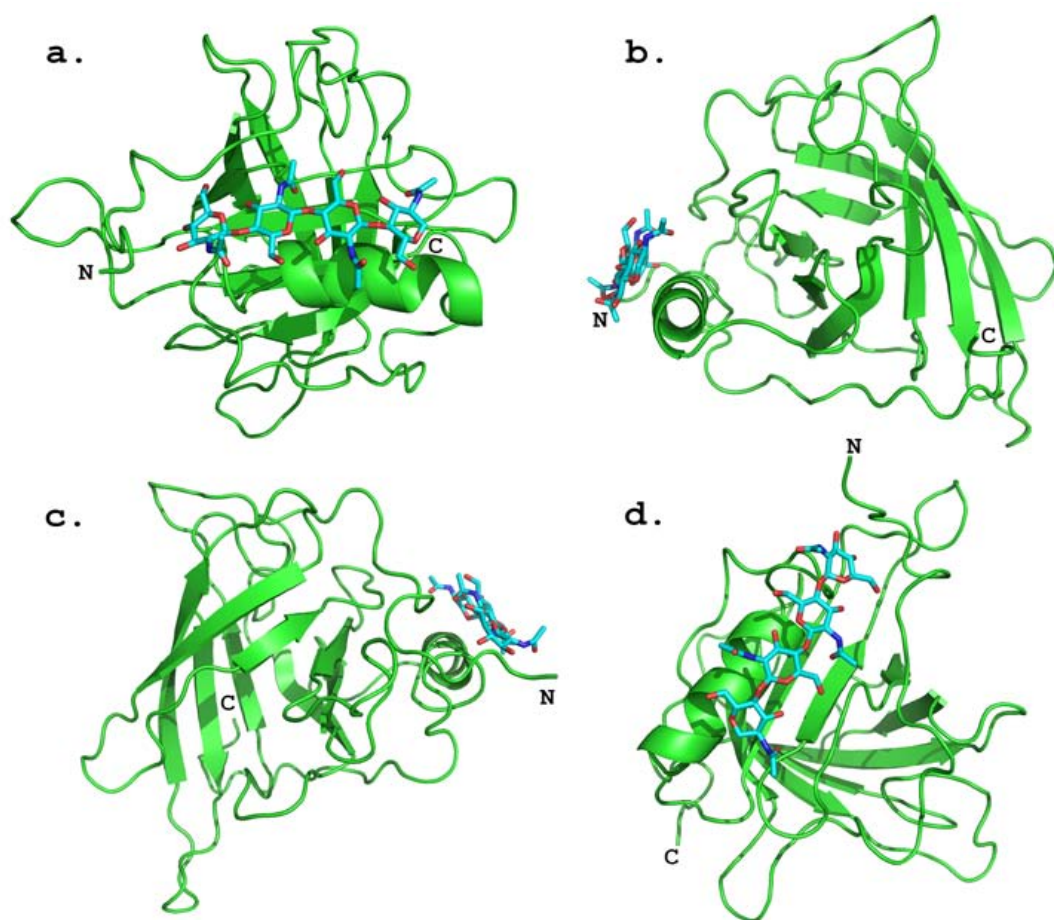
Εν συνεχεία, διεξήχθησαν πειράματα αγκυροβόλησης (docking) χρησιμοποιώντας το λογισμικό Autodock4.2 (Morris, Huey et al. 2009). Για πρωτεΐνη-στόχος χρησιμοποιήθηκαν τα μοντέλα των Tweedle πρωτεϊνών και για προσδέτη/υποκαταστάτη (ligand) χρησιμοποιήθηκε μια αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG). Σε όλα τα πειράματα, δημιουργήθηκαν πλέγματα (grid maps) με μέγεθος (grid-spacing) 1.0Å, ώστε να περιβάλλουν τα μόρια των πρωτεϊνών Tweedle, και χρησιμοποιήθηκε ο αλγόριθμος Lamarckian genetic algorithm searches (Morris, Goodsell et al. 1998).

Τα αποτελέσματα δείχνουν ότι η αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG) μπορεί να αλληλεπιδρά με τα πρωτεϊνικά μοντέλα με πολλαπλούς ευνοϊκά θερμοδυναμικά τρόπους. Είναι εμφανές ότι οι αλυσίδες χιτίνης ομοτετραμερών N-ακετυλογλυκοζαμίνης (4NAG) αλληλεπιδράνε με τα πρωτεϊνικά μοντέλα παράλληλα στους β-κλώνους (εικόνα 71), σε συμφωνία με προηγούμενες πειραματικές παρατηρήσεις από περιθλασιγράμματα ακτίνων X (Atkins 1985).

Ένα ακόμα ενδιαφέρον αποτέλεσμα είναι ότι η αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG) μπορεί να αλληλεπιδράσει με τα πρωτεϊνικά μοντέλα παράλληλα στους β-κλώνους καθώς και στην α-έλικα (εικόνα 72).



Εικόνα 71: Πειραματικά αποτελέσματα του προγράμματος αγκυροβόλησης AutoDock4.2 (Morris, Huey et al. 2009) με πρωτεΐνη-στόχο τα μοντέλα των πρωτεϊνών Tweedle και προσδέτη/υποκαταστάτη (ligand) μια αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG). Τα αποτελέσματα δείχνουν ότι, με πολλούς ευνοϊκούς θερμοδυναμικά τρόπους, οι αλυσίδες χιτίνης 4NAG αλληλεπιδρούν με τα πρωτεϊνικά μοντέλα παράλληλα στους β-κλώνους. Στα (a), (b) είναι το μοντέλο της CPTmotif4 (από *Danaus plexippus*). Στα (c), (d) είναι το μοντέλο της TweedleE (από *Drosophila melanogaster*). Στα (e), (f) είναι το μοντέλο της TweedleZ (από *Drosophila melanogaster*). Οι αλυσίδες χιτίνης 4NAG φαίνονται με γαλάζιο ή ροζ χρώμα.



Εικόνα 72: Πειραματικό αποτέλεσμα του προγράμματος αγκυροβόλησης AutoDock4.2 (Morris, Huey et al. 2009) με πρωτεΐνη-στόχο το μοντέλο της CPTmotif4 (*Danaus plexippus*) πρωτεΐνης Tweedle και προσδέτη/υποκαταστάτη (ligand) μια αλυσίδα χιτίνης ομοτετραμερούς N-ακετυλογλυκοζαμίνης (4NAG). Το αποτέλεσμα δείχνει ότι η αλυσίδα χιτίνης 4NAG μπορεί να αλληλεπιδρά, με τα πρωτεϊνικά μοντέλα, παράλληλα τόσο στους β-κλώνους όσο και στην α-έλικα. Τα (a), (b), (c), (d) είναι το ίδιο αποτέλεσμα από διαφορετικές οπτικές γωνίες. Η αλυσίδα χιτίνης 4NAG είναι χρωματισμένη με γαλάζιο.

Συμπεράσματα Συζήτηση



«Science and everyday life cannot and should not be separated.»

Rosalind Elsie Franklin (1920-1958)

Κεφάλαιο 4^ο : Συμπεράσματα – Συζήτηση

4.1 Γενική Περιγραφή

Πλέον, είναι γνωστός ένας μεγάλος αριθμός από ακολουθίες δομικών πρωτεϊνών του δερματίου (Willis 2010). Ωστόσο, καμία δομική πρωτεΐνη του δερματίου δεν έχει πειραματικά προσδιορισμένη δομή ακόμα. Επίσης, παρ' όλο που οι δομικές πρωτεΐνες του δερματίου μπορούν να αλληλεπιδρούν με χιτίνη, η ακριβής φύση των αλληλεπιδράσεων αυτών είναι πειραματικά άγνωστη (Willis, Papandreou et al. 2012). Πρόσφατες μελέτες έχουν προτείνει δομικά μοτίβα για τις πρωτεΐνες των οικογενειών CPR RR-1 (Iconomidou, Willis et al. 1999; Iconomidou, Chryssikos et al. 2001; Hamodrakas, Willis et al. 2002), CPR RR-2 (Iconomidou, Willis et al. 1999; Iconomidou, Willis et al. 2005) και CPF (Papandreou, Iconomidou et al. 2010), καθώς και θεωρητικά μοντέλα αλληλεπιδράσεων αυτών με χιτίνη. Αυτές οι μελέτες βασίστηκαν στην μακρινή ομοιότητα μεταξύ των CPRs και των λιποκαλινών (Iconomidou, Willis et al. 1999; Hamodrakas, Willis et al. 2002), που βασίστηκε στην πρόβλεψη του λογισμικού PhD (Rost 1996).

Το επόμενο λογικό βήμα ήταν να μελετηθούν οι πρωτεΐνες της οικογένειας Tweedle των δομικών πρωτεϊνών του δερματίου. Τα αίτια που οδήγησαν σε αυτήν την επιλογή ήταν ότι έχει εντοπιστεί και ταυτοποιηθεί ένας επαρκής αριθμός πρωτεϊνών ως μέλη στην οικογένεια Tweedle (Willis 2010), καθώς και ότι υπάρχουν πειραματικές αποδείξεις ότι μπορούν να αλληλεπιδρούν με χιτίνη (Tang, Liang et al. 2010). Η παρούσα μελέτη αφορά τα δομικά χαρακτηριστικά της οικογένειας Tweedle, εστιάζοντας στις ομοιότητες με τις λιποκαλίνες, και επίσης ερευνητικές προσεγγίσεις σχετικά με τις αλληλεπιδράσεις των πρωτεϊνών Tweedle με αλυσίδες χιτίνης με πειράματα αγκυροβόλησης (Docking).

Επιπροσθέτως, η παρούσα εργασία μελέτησε τις δομικές πρωτεΐνες του δερματίου από τις οικογένειες CPAP3. Οι πρωτεΐνες της οικογένειας CPAP3 έχουν ένα πολύ σημαντικό χαρακτηριστικό. Περιέχουν το ChtBD2 domain (Willis 2010), το οποίο έχει αποδειχθεί ότι έχει την ικανότητα αλληλεπίδρασης με χιτίνη (σε μια χιτινάση (Arakane, Zhu et al. 2003) – καθώς το παρόν domain είναι διαδεδομένο και σε

οργανισμούς εκτός αρθροπόδων). Επίσης, έχει πλέον ταυτοποιηθεί ένας σημαντικός αριθμός των συγκεκριμένων πρωτεϊνών και άρα είναι δυνατή η μελέτη τους (Jasrapuria, Arakane et al. 2010; Willis 2010).

4.2 Συμπεράσματα – Πρωτεΐνες Tweedle

Τα αποτελέσματα των στοιχίσεων μεταξύ των πρωτεϊνών της οικογένειας Tweedle και της λιποκαλίνης bovine Retinol Binding Protein (RBP) (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)) απέδειξαν ομοιότητα σε επίπεδο ακολουθίας. Συγκεκριμένα, η ταυτότητα των στοιχιζόμενων ακολουθιών ήταν περίπου στο 10-11% ενώ η ομοιότητα ήταν εξαιρετικά υψηλή, περίπου στο 50%. Επίσης, εντυπωσιακό είναι και η ποιότητα των στοιχίσεων καθώς είναι συνεχόμενες με ελάχιστα κενά (πίνακας 10, εικόνες 54,55,56,57 και 58).

Ανάλογη ομοιότητα σε επίπεδο ακολουθίας έδειξαν και οι στοιχίσεις ενός χαρακτηριστικού εκπροσώπου της οικογένειας Tweedle (συγκεκριμένα της TweedleE) με ένα επιλεγμένο σύνολο λιποκαλινών. Συγκεκριμένα, η ταυτότητα των στοιχιζόμενων ακολουθιών ήταν περίπου στο 11% ενώ η ομοιότητα ήταν αρκετά σημαντική, περίπου 45%. Ομοίως, η ποιότητα των στοιχίσεων ήταν ικανοποιητική, καθώς ήταν σχεδόν συνεχόμενη και χωρίς μεγάλα κενά (πίνακας 11 και εικόνα 59a-g). Το παραπάνω είναι ένα πρωτόγνωρο αποτέλεσμα καθώς μέχρι τώρα δεν είχαν πραγματοποιηθεί μελέτες ομοιότητας δομικών πρωτεϊνών του δερματίου με λιποκαλίνες εκτός της bovine RBP (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)). Αποδεικνύεται έτσι η ομοιότητα των πρωτεϊνών Tweedle σε επίπεδο ακολουθίας, με το σύνολο της οικογένειας των λιποκαλινών.

Επίσης, ύστερα από πολλαπλές στοιχίσεις επιβεβαιώθηκε η ύπαρξη των τεσσάρων (4) συντηρημένων ομάδων (blocks) των πρωτεϊνών της οικογένειας Tweedle, στο σύνολο των μέχρι σήμερα ταυτοποιημένων πρωτεϊνών της οικογένειας (ένα ενδεικτικό σύνολο παρουσιάζεται στην εικόνα 60).

Σύμφωνα με τα παραπάνω αποτελέσματα, κρίθηκε δυνατή η προτυποποίηση με βάση την ομολογία των πρωτεϊνών Tweedle, χρησιμοποιώντας ως δομή-οδηγό (template) την προσδιορισμένη κρυσταλλογραφικά δομή της λιποκαλίνης RBP (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)).

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο (ενότητα 3.2), το προτεινόμενο δομικό μοτίβο για τις πρωτεΐνες της οικογένειας Tweedle είναι ένα ολόκληρο β-βαρέλι, το οποίο αποτελείται κυρίως από οκτώ (8) αντιπαράλληλους β-κλώνους και μια (1) α-έλικα παράλληλη στους κλώνους (εικόνα 62).

Εντυπωσιακό είναι το γεγονός ότι, στην πλειοψηφία των δομικών μοντέλων οι τέσσερις (4) συντηρημένες ομάδες των πρωτεϊνών Tweedle συμπίπτουν με τα τέσσερα (4) παράλληλα πτυχωτά β-φύλλα του β-βαρελιού (ενδεικτικό παράδειγμα η πρωτεΐνη TweedleE που παρουσιάζεται στην εικόνα 65). Αυτό το πειραματικό αποτέλεσμα είναι σε απόλυτη συμφωνία με τις πειραματικές προβλέψεις (Guan, Middlebrooks et al. 2006) που είχαν πραγματοποιηθεί με το λογισμικό PhD (Rost 1996). Αυτό το αποτέλεσμα αποτελεί μια σημαντική απόδειξη της ρεαλιστικότητας των δομικών μοντέλων.

Εν συνεχεία, επιβεβαιώθηκαν κοινά δομικά χαρακτηριστικά μεταξύ των δομικών μοντέλων των πρωτεϊνών Tweedle και των πρωτεϊνών της οικογένειας των λιποκαλινών (ιδιαίτερα της λιποκαλίνης bovine Retinol Binding Protein, PDB ID: 1FEN). Η δομική στοίχιση των δομικών μοντέλων των πρωτεϊνών Tweedle με την λιποκαλίνη bovine Retinol Binding Protein (PDB ID: 1FEN), αποκάλυψε υψηλή συντήρηση υδρόφοβων αμινοξικών καταλοίπων στο εσωτερικό των υποθετικών εσωτερικών και εξωτερικών συστάδων (inner and outer clusters) στα δομικά μοντέλα των πρωτεϊνών Tweedle (πίνακες 13 και 14), το οποίο αποτελεί κοινό χαρακτηριστικό με την οικογένεια των λιποκαλινών (Adam, Charlotiaux et al. 2008). Ειδικότερα, η συντήρηση είναι υψηλότερη στην υποθετική εξωτερική συστάδα (outer cluster) των πρωτεϊνικών μοντέλων (πίνακας 14). Επίσης, αποδείχτηκε ότι το εσωτερικό των β-βαρελίων των δομικών μοντέλων Tweedle αποτελείται κυρίως από υδρόφοβα αμινοξικά κατάλοιπα (εικόνα 66), όπως ακριβώς και στη δομή της λιποκαλίνης bovine Retinol Binding Protein (PDB ID: 1FEN) (Zanotti, Marcello et al. 1994).

Αξίζει να σημειωθεί ότι το λογισμικό πρόγνωσης λιποκαλινών LipocalinPred (Ramana and Gupta 2009) θεωρεί και τις δέκα (10) πρωτεΐνες Tweedle του αντιπροσωπευτικού δείγματος ως λιποκαλίνες (10/10 – 100%) (εικόνα 67). Αυτό το αποτέλεσμα αποτελεί μια επιπλέον ένδειξη ομοιότητας των πρωτεϊνών της οικογένειας των λιποκαλινών με τις πρωτεΐνες της οικογένειας Tweedle.

Η παρατήρηση των δομικών χαρακτηριστικών των δομικών μοντέλων των πρωτεϊνών Tweedle φανέρωσε μια αξιοσημείωτη παρουσία αρωματικών αμινοξικών καταλοίπων (φαινυλαλανίνες – F, τυροσίνες – Y) στο εσωτερικό των β-βαρελιών καθώς και στις σχισμές αυτών (εικόνα 66). Φαίνεται ότι δημιουργούν μια επιφάνεια από αρωματικούς δακτυλίους προσαρμοσμένους ειδικά για αλληλεπιδράσεις μεταξύ πρωτεϊνών-χιτίνης στο δερμάτιο (Quiocho 1989; Vyas 1991; Elgavish and Shaanan 1997; Hamodrakas, Kanellopoulos et al. 1997; Svitil and Kirchman 1998; Shen and Jacobs-Lorena 1999). Επιπροσθέτως, το δομικό μοτίβο των αντιπαράλληλων β-κλώνων συνοδεύόμενο με την παρουσία ενός σημαντικού αριθμού αρωματικών αμινοξικών καταλοίπων έχει προταθεί ότι αποτελεί δομικό μοτίβο που συμμετέχει σε αλληλεπιδράσεις πρωτεϊνών – υδατανθράκων (Suetake, Tsuda et al. 2000; Watanabe, Ariga et al. 2003; Katouno, Taguchi et al. 2004). Ένα ακόμα ενδιαφέρον στοιχείο στα δομικά μοντέλα των πρωτεϊνών Tweedle, είναι η μικρή α-έλικα κοντά στο καρβοξυτελικό άκρο. Αυτή η περιοχή κοντά στο καρβοξυτελικό άκρο των δομικών μοντέλων μοιάζει με το αυτοτελές δομικό στοιχείο (domain) στην πρωτεΐνη tachycitin που αλληλεπιδρά με χιτίνη στα ασπόνδυλα και στην πρωτεΐνη hevein που αλληλεπιδρά με χιτίνη στα φυτά, όπου τους δυο (2) β-κλώνους ακολουθεί μια μικρή «ελικοειδής» στροφή (helical turn) (Suetake, Tsuda et al. 2000). Όλες αυτές οι ενδείξεις συνηγορούν ότι τα δομικά μοντέλα των πρωτεϊνών Tweedle έχουν τα κατάλληλα δομικά χαρακτηριστικά ώστε να αλληλεπιδρούν με χιτίνη.

Το τελευταίο στάδιο της μελέτης των πρωτεϊνών Tweedle ήταν πειράματα αγκυροβόλησης (docking) με στόχο να ερευνηθεί το μοντέλο αλληλεπίδρασης τους στο δερμάτιο, αφού όπως έχει αναφερθεί έχει αποδειχθεί ότι οι πρωτεΐνες Tweedle αλληλεπιδρούν με χιτίνη (Tang, Liang et al. 2010). Στα πειράματα, ως πρωτεΐνη – στόχος (target protein) ήταν τα δομικά μοντέλα των πρωτεϊνών Tweedle ενώ ως μόριο – προσδέτης (ligand) ήταν μια ποικιλία από αλυσίδες χιτίνης (ενότητα 3.5). Τα αποτελέσματα έδωσαν επιπλέον στοιχεία για την αξιοπιστία των δομικών μοντέλων των πρωτεϊνών Tweedle. Στην πλειοψηφία όσων αποτελεσμάτων ήταν

θερμοδυναμικά ευνοϊκά (τόσο με το λογισμικό GRAMM όσο και με το λογισμικό Autodock4.2), φαίνεται ότι οι ποικίλες αλυσίδες χιτίνης μπορούν να αλληλεπιδρούν παράλληλα στους β-κλώνους των δομικών μοντέλων (εικόνες 69,70,71 και 72) και αυτό είναι σε συμφωνία με το προτεινόμενο κρυσταλλικό μοντέλο της χιτίνης (Atkins 1985) (ενότητα 1.4.5) και την προτεινόμενη δομή του συμπλόκου χιτίνης – πρωτεϊνών (Blackwell and Weih 1980) (ενότητα 1.5). Άρα, συμπερασματικά είναι προφανές ότι τα δομικά μοντέλα των πρωτεϊνών της οικογένειας Tweedle μπορούν να αλληλεπιδρούν με ινίδια χιτίνης με ρεαλιστικό τρόπο και επιβεβαιώνουν προηγούμενες προτάσεις και πειραματικά αποτελέσματα.

Επιλογικά, ύστερα από τα παραπάνω πειραματικά συμπεράσματα, οι πρωτεΐνες της οικογένειας Tweedle φαίνεται να παρουσιάζουν δομικά χαρακτηριστικά και ιδιότητες όμοιες με τις πρωτεΐνες της οικογένειας των λιποκαλινών. Επίσης, τα επιτυχημένα πειράματα αγκυροβόλησης με γραμμικά τμήματα αλυσίδων χιτίνης, αποδεικνύουν ότι το προτεινόμενο μοντέλο δομής – αλληλεπίδρασης είναι ρεαλιστικό και επιβεβαιώνει προηγούμενα πειραματικά αποτελέσματα.

4.3 Συμπεράσματα – Πρωτεΐνες CPAP3

Τα αποτελέσματα των στοιχίσεων μεταξύ των πρωτεϊνών της οικογένειας CPAP3 και της λιποκαλίνης bovine Retinol Binding Protein (RBP) (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)) απέδειξαν ομοιότητα σε επίπεδο ακολουθίας. Συγκεκριμένα, η ταυτότητα των στοιχιζόμενων ακολουθιών ήταν περίπου στο 14% ενώ η ομοιότητα ήταν εξαιρετικά υψηλή, περίπου στο 50%. Επίσης, εντυπωσιακό είναι και η ποιότητα των στοιχίσεων καθώς είναι σχεδόν συνεχόμενες με ελάχιστα κενά (πίνακας 12 και εικόνα 61 a-h).

Αξίζει να σημειωθεί ότι το λογισμικό πρόγνωσης λιποκαλινών LipocalinPred (Ramana and Gupta 2009) θεωρεί τις έξι από τις οχτώ (6/8 - 75%) πρωτεΐνες του αντιπροσωπευτικού δείγματος των CPAP3, ως λιποκαλίνες (εικόνα 68). Αυτό το αποτέλεσμα αποτελεί μια επιπλέον ένδειξη ομοιότητας των πρωτεϊνών της οικογένειας των λιποκαλινών με τις πρωτεΐνες της οικογένειας CPAP3.

Σύμφωνα με τα παραπάνω αποτελέσματα, κρίθηκε δυνατή η προτυποποίηση με βάση την ομολογία των πρωτεϊνών CPAP3, χρησιμοποιώντας ως δομή-οδηγό (template) την προσδιορισμένη κρυσταλλογραφικά δομή της λιποκαλίνης RBP (PDB ID: 1FEN (Zanotti, Marcello et al. 1994)).

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο (ενότητα 3.2), το προτεινόμενο δομικό μοτίβο για τις πρωτεΐνες της οικογένειας CPAP3 Το βασικό δομικό μοτίβο των πρωτεϊνών της οικογένειας CPAP3 είναι ένα ολόκληρο β-βαρέλι, που αποτελείται από οχτώ (8) αντιπαράλληλους β-κλώνους, μια (1) α-έλικα παράλληλη προς αυτούς και μια (1) μικρή α-έλικα στο ένα άνοιγμα του βαρελιού (εικόνα 63).

Συμπερασματικά, οι πρωτεΐνες της οικογένειας CPAP3 παρουσιάζουν πολλές ομοιότητες σε επίπεδο ακολουθίας με τις πρωτεΐνες της οικογένειας των λιποκαλινών. Επίσης, τα δομικά πρωτεϊνικά μοντέλα που δημιουργήθηκαν παρουσιάζουν όμοια δομικά χαρακτηριστικά με τα δομικά μοντέλα των πρωτεϊνών Tweedle που αναλύθηκαν παραπάνω. Παρουσιάζουν λοιπόν ιδιαίτερο ενδιαφέρον και σε συνδυασμό με το ότι το βασικό τους domain (ChtBD2) έχει αποδειχθεί ότι αλληλεπιδρά με χιτίνη σε άλλες πρωτεΐνες εκτός δερματίου (π.χ. χιτινάση (Arakane, Zhu et al. 2003) κρίνεται σημαντική η περαιτέρω πειραματική τους διερεύνηση τόσο με μεθόδους αγκυροβόλησης (docking) όσο και με άλλες ερευνητικές προσεγγίσεις.

4.4 Επίλογος – Μελλοντικές Προκλήσεις

Εν κατακλείδι, είναι σημαντικό να τονιστεί ότι τα δομικά μοντέλα των πρωτεϊνών Tweedle και CPAP3 προτείνονται με το σκοπό να πραγματοποιηθούν και άλλα περαιτέρω πειράματα. Είναι πολύ σημαντικό να προσδιοριστεί πειραματικά η δομή των δομικών πρωτεϊνών του δερματίου με κρυσταλλογραφία ακτίνων X ή με φασματογραφία NMR, είτε μόνες τους, είτε ως σύμπλοκα με ολιγοσακχαρίτες ανάλογους της χιτίνης ή με λιπίδια του δερματίου. Ο προσδιορισμός αυτών των πρωτεϊνικών δομών θα παράσχει επιπλέον πληροφορίες σχετικά με την πολύπλοκη δομή του δερματίου των αρθροπόδων, με τις μοναδικές φυσικές του ιδιότητες. Θα βοηθήσει επίσης να διαφωτιστεί η φύση των πρωτεϊνικών αλληλεπιδράσεων με χιτίνη και άλλα λειτουργικά μόρια, που έχουν πολύ σημαντικό ρόλο στη δομή και στη φυσιολογία του δερματίου.

Παράρτημα

1. Πρωτεϊνικά μοντέλα της οικογένειας Tweedle, ύστερα από προτυποποίηση με βάση την ομολογία.

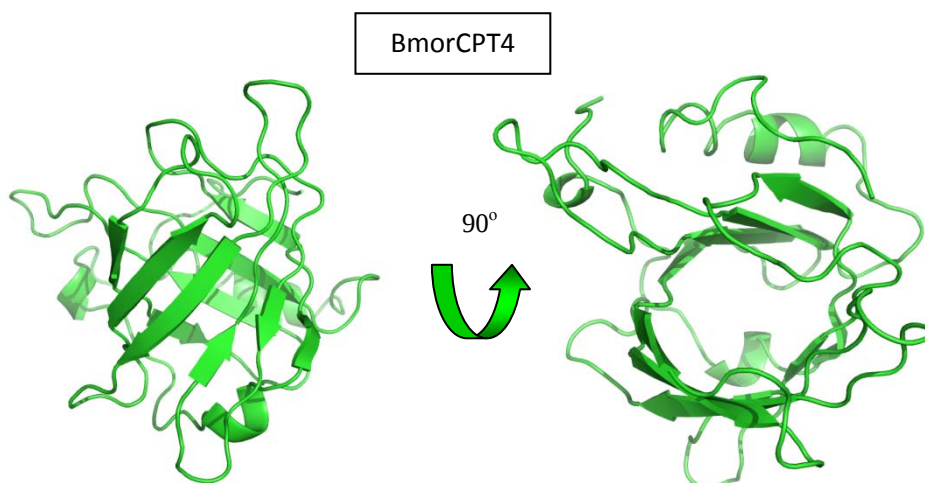
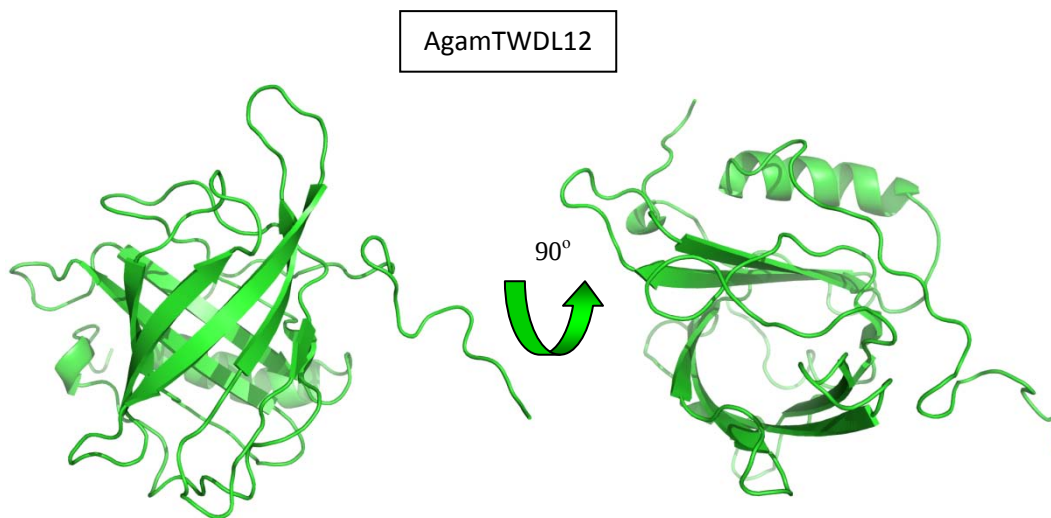
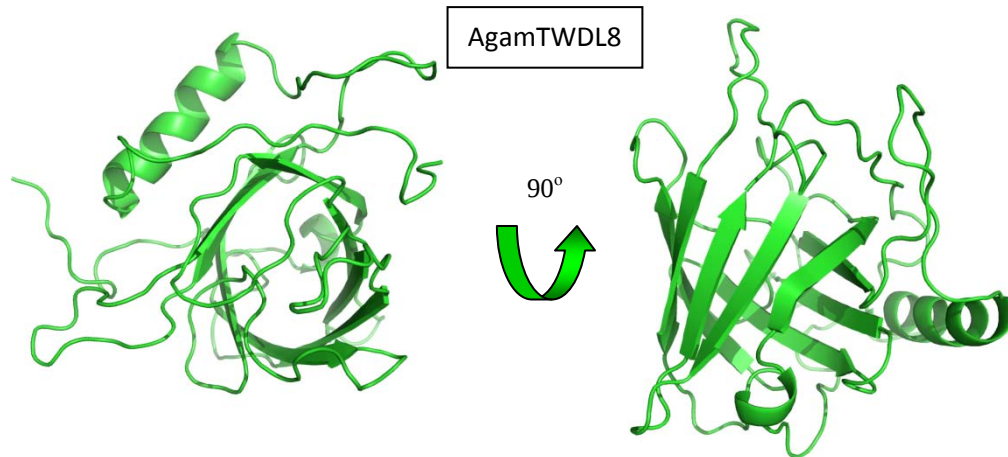
2. Abstract και Poster στο 35^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Βιολογικών Επιστημών (EEBE), 23-25 Μαΐου 2013

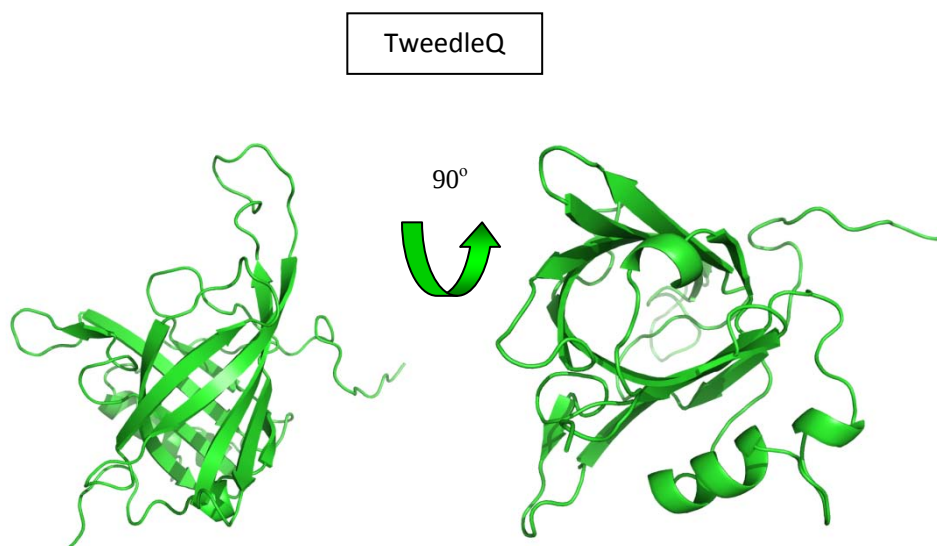
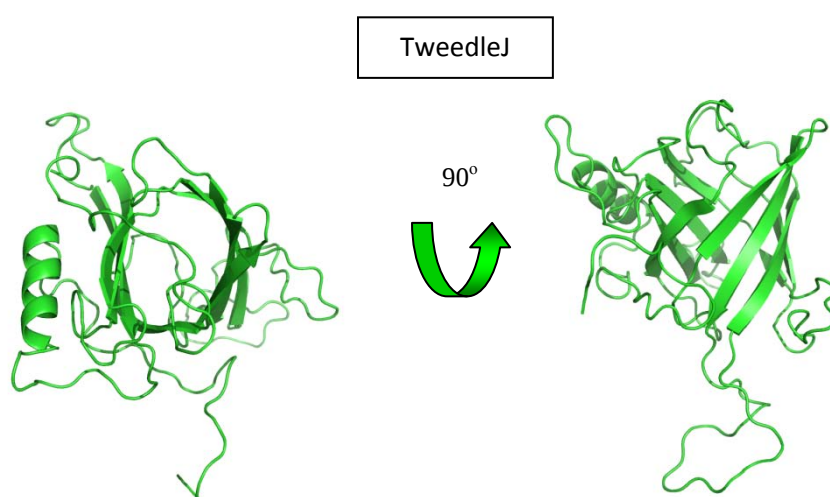
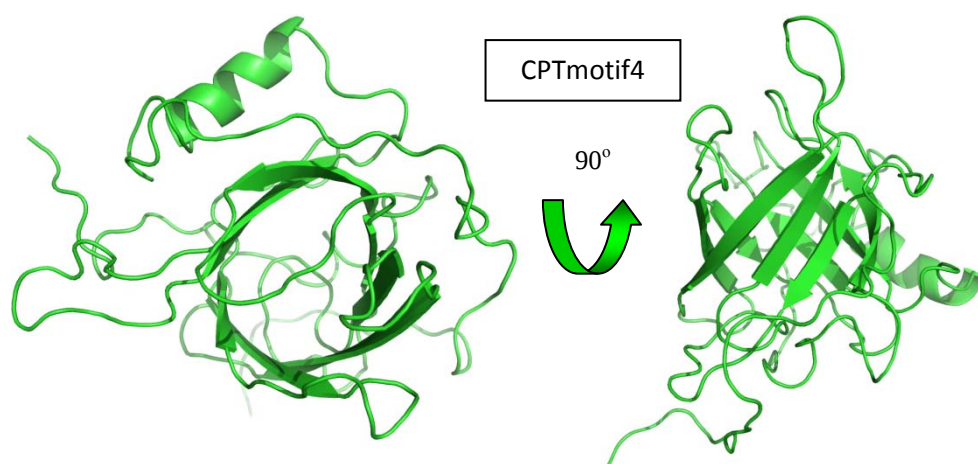
Θέμα: «Υπολογιστικές μελέτες αλληλεπιδράσεων των πρωτεϊνών του δερματίου των αρθροπόδων, που ανήκουν στην οικογένεια Tweedle, με χιτίνη»

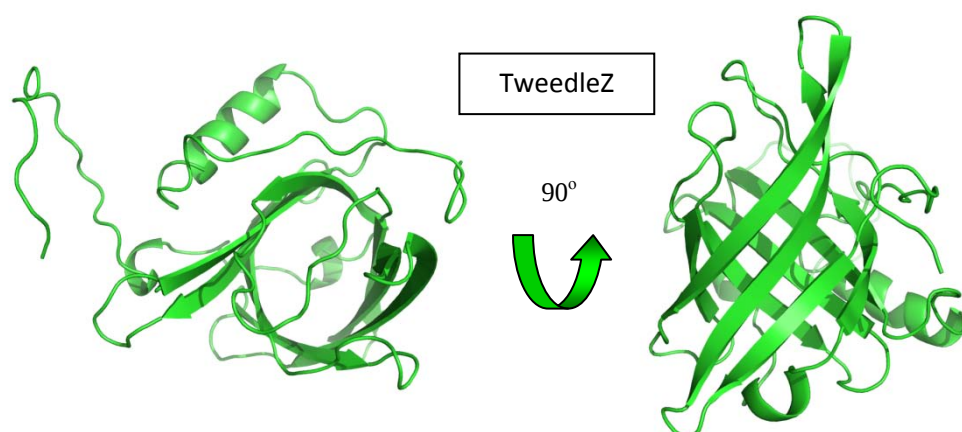
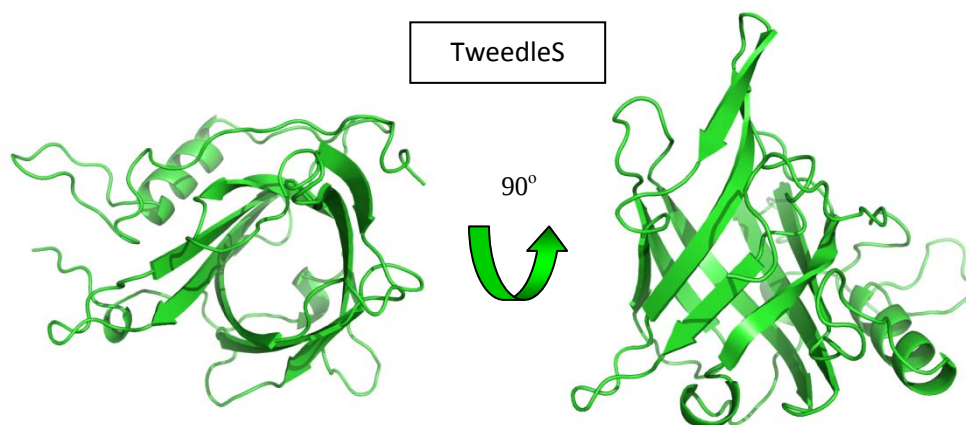
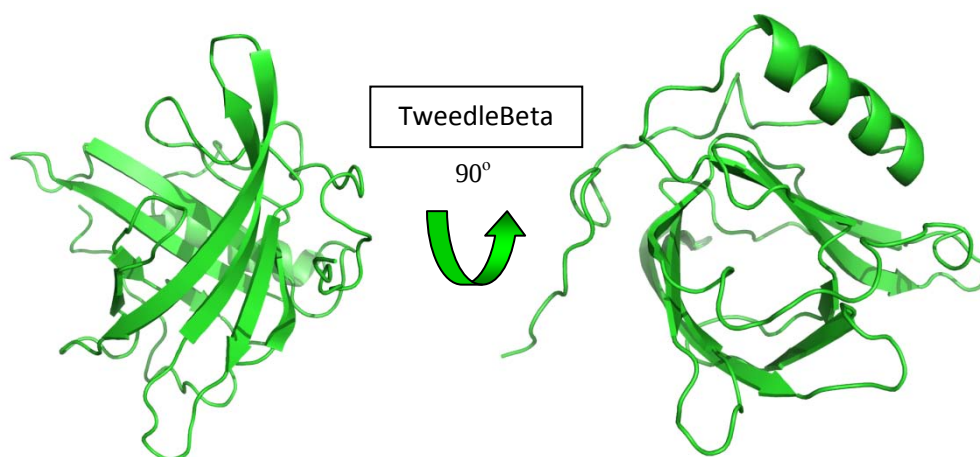
3. Abstract και Poster στο 8^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Υπολογιστικής Βιολογίας και Βιοπληροφορικής (EEYBB) – Hellenic Society for Computational Biology and Bioinformatics (HSCBB), 22-24 Νοεμβρίου 2013

Θέμα: «Computational studies of the Tweedle family cuticular proteins: analysis of putative structure and interactions with chitin filaments»

1. Δομικά μοντέλα πρωτεϊνών της οικογένειας Tweedle, ύστερα από προτυποποίηση με βάση την ομολογία







2. Abstract και Poster στο 35^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Βιολογικών Επιστημών (ΕΕΒΕ), 23-25 Μαΐου 2013



Πρακτικά 35ου Επιστημονικού Συνεδρίου της Ελληνικής Εταιρείας Βιολογικών Επιστημών
Ναύπλιο, 23-25 Μαΐου 2013

ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΜΕΛΕΤΕΣ ΑΛΛΗΛΕΠΙΔΡΑΣΕΩΝ ΤΩΝ ΠΡΩΤΕΪΝΩΝ ΤΟΥ ΔΕΡΜΑΤΙΟΥ ΤΩΝ ΑΡΘΡΟΠΟΔΩΝ, ΠΟΥ ΑΝΗΚΟΥΝ ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ TWEEDLE, ΜΕ ΧΙΤΙΝΗ

Γιαννακούρας Θεοχάρης-Ηλίας, Παπανδρέου Νίκος, Οικονομίδου Βασιλική, Χαμόδρακας Σταύρος

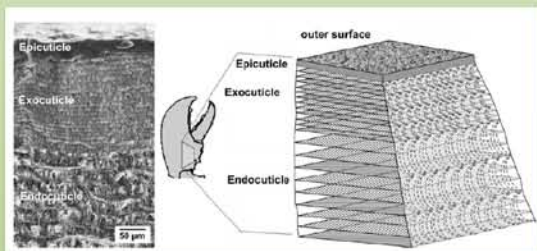
Τμήμα Βιολογίας, Τομέας Βιολογίας Κυττάρων και Βιοφυσικής, Πανεπιστήμιο Αθηνών, Πανεπιστημιόπολις, Αθήνα 157 01

Το δερμάτιο είναι ένα διμερές, σύνθετο υλικό που παρέχει δομική και μηχανική υποστήριξη δρώντας λειτουργικά τόσο ως δέρμα όσο και ως σκελετός στα αρθρόποδα. Αποτελείται από ίνες χιτίνης ενσωματωμένες σε μία πρωτεϊνική 'μήτρα'. Οι φυσικές του ιδιότητες καθορίζονται από την δομή των δυο βασικών δομικών στοιχείων του, των πρωτεϊνών του δερματίου (CP) και της χιτίνης, καθώς και από τις αλληλεπιδράσεις μεταξύ τους. Η κύρια οικογένεια πρωτεϊνών του δερματίου είναι η CPR (πρωτεΐνες που εντοπίζεται στην αμινοξική τους ακολουθία η εκτεταμένη R&R συναινετική ακολουθία) οι οποίες έχει αποδειχθεί πειραματικά ότι δεσμεύουν χιτίνη. Μελέτες Συγκριτικής Προτυποποίησης και πειράματα Αγκυροβόλησης οδήγησαν στην πρόταση ότι το βασικό δομικό μοτίβο των πρωτεϊνών που φέρουν την R&R συναινετική ακολουθία είναι ένα αντιπαράλληλο β-πτυχωτό φύλλο που σχηματίζει ένα μισό 'β-βαρέλι', ενώ προτάθηκαν πιθανοί τρόποι αλληλεπίδρασης τους με τη χιτίνη. Οι μελέτες βασίστηκαν στην παρατήρηση ότι οι CPRs ομοιάζουν, σε επίπεδο αλληλουχίας, με την πρωτεΐνη του πλάσματος που δεσμεύει ρετινόλη (RBP) και ανήκει στην οικογένεια των λιποκαλινών. Το επόμενο βήμα ήταν να μελετηθούν και άλλες οικογένειες δομικών πρωτεϊνών του δερματίου και οι αλληλεπιδράσεις τους με χιτίνη. Για παράδειγμα, η οικογένεια Tweedle αποτελεί μια μεγάλη οικογένεια πρωτεϊνών του δερματίου. Μέχρι σήμερα, 236 ακολουθίες από 17 διαφορετικά είδη έχουν χαρακτηριστεί ως μέλη της. Όπως οι CPRs έτσι και οι πρωτεΐνες Tweedle εμφανίζουν ομοιότητα σε επίπεδο αλληλουχίας με την RBP. Με τη μέθοδο της συγκριτικής προτυποποίησης κατασκευάστηκαν μοντέλα για μέλη της οικογένειας Tweedle. Από την ανάλυσή τους προτείνουμε ότι το βασικό δομικό μοτίβο τους είναι ένα 'β-βαρέλι', που απαρτίζεται από 8 αντιπαράλληλους β-κλώνους και μία α-έλικα παράλληλη προς τους β-κλώνους. Επίσης, πραγματοποιήθηκαν πειράματα αγκυροβόλησης για να διαπιστώσουμε πως τα μέλη των Tweedle αλληλεπιδρούν με χιτίνη.

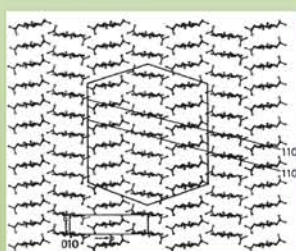
ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΜΕΛΕΤΕΣ ΑΛΛΗΛΕΠΙΔΡΑΣΕΩΝ ΧΙΤΙΝΗΣ - ΠΡΩΤΕΪΝΩΝ, ΠΟΥ ΑΝΗΚΟΥΝ ΣΤΗΝ ΟΙΚΟΓΕΝΕΙΑ TWEEDLE, ΣΤΟ ΔΕΡΜΑΤΙΟ ΤΩΝ ΑΡΘΡΟΠΟΔΩΝ

Γιαννακούρας, Θ. Ε., Παπανδρέου, Ν. Χ., Οικονομίδου, Β.Α. και Χαμόδρακας, Σ. Ι.

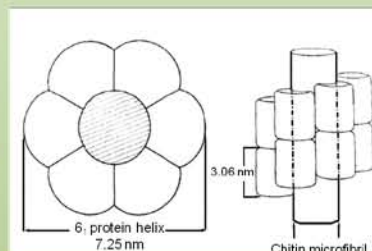
Τομέας Βιολογίας Κυπάρου και Βιοφυσικής, Τμήμα Βιολογίας, Πανεπιστήμιο Αθηνών, Αθήνα 157 01



Το **δεράτιο** (cuticle) είναι ένα διμερές, σύνθετο υλικό που παρέχει δομική και μηχανική υποστήριξη δρώντας λειτουργικά τόσο ως δέρμα όσο και ως σκελετός στα αρθρόποδα. Οι φυσικές του ιδιότητες καθορίζονται από την δομή των δυο βασικών δομικών στοιχείων του, των πρωτεϊνών του δερματίου (Cuticular Proteins) και της χιτίνης (chitin), καθώς και από τις αλληλεπιδράσεις μεταξύ τους.



Η δομή της **α-χιτίνης** που προτάθηκε από τον Atkins το 1985. Από την ανάλυση περιθλασμογράμματος, προτάθηκε ότι οι ίνες χιτίνης απαρτίζονται από 19 αλυσίδες.



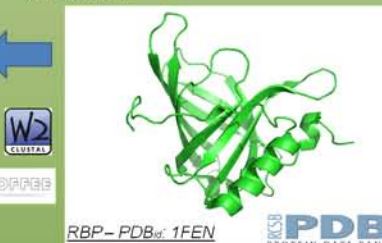
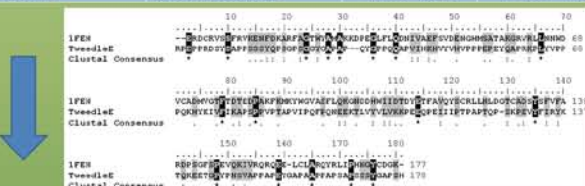
Ίνες χιτίνης ενσωματωμένες σε μία πρωτεϊνική 'μήτρα' σύμφωνα με το μοντέλο των Blackwell και Weih (1980).

Είδος	Πρωτεΐνη (Uniprot)	Sequence identity	Sequence similarity
<i>Anopheles gambiae</i>	AgamTWDL8 (819-CuticleDB)	20/177 = 11.2%	76/177 = 42.9%
<i>Anopheles gambiae</i>	AgamTWDL12 (Q7Q9V7)	21/177 = 11.8%	75/177 = 42.4%
<i>Drosophila melanogaster</i>	TweedleE (Q8SZP2)	17/177 = 9.6%	69/177 = 39%
<i>Drosophila melanogaster</i>	TweedleJ (Q9VBD6)	31/177 = 17.5%	102/177 = 57.6%
<i>Drosophila melanogaster</i>	TweedleQ (Q9VBC8)	26/177 = 14.7%	97/177 = 54.8%
<i>Drosophila melanogaster</i>	TweedleS (Q9VBD3)	21/177 = 11.8%	87/177 = 49.2%
<i>Drosophila melanogaster</i>	TweedleZ (Q8IR09)	19/177 = 10.7%	75/177 = 42.4%
<i>Drosophila melanogaster</i>	TweedleBeta (A1Z8Q0)	18/177 = 10.2%	77/177 = 43.5%
<i>Danaus plexippus</i>	CP tweedle motif 4 (G6D0I0)	19/177 = 10.7%	90/177 = 50.8%
<i>Bombyx mori</i>	BmorCPT1 (C0H6Z1)	30/177 = 17%	93/177 = 52.5%
<i>Bombyx mori</i>	BmorCPT4 (C0H6Z4)	14/163 = 8.6%	70/163 = 43%

Οικογένεια Tweedle
Αποτελεί μια μεγάλη οικογένεια πρωτεϊνών του δερματίου. Μέχρι σήμερα, 236 ακολουθίες από 17 διαφορετικά είδη έχουν χαρακτηριστεί ως μέλη της (Uniprot, βιβλιογραφία). Ύστερα από στοιχίσεις ακολουθιών, διαπιστώθηκε ότι εμφανίζουν ομοιότητα με την bovine plasma Retinol Binding Protein – RBP (PDBα: 1FEN).

Πλέον, έχει προσδιοριστεί η ακολουθία μερικών χιλιών δομικών πρωτεϊνών του δερματίου. Είναι οργανωμένες σε 13 οικογένειες, ανάλογα με τις αμινοξικές τους ακολουθίες.

A/A	Όνομα Οικογένειας
1.	CPR
2.	CPF
3.	CPFL
4.	TWEEDLE (TWDL)
5.	CPLCG
6.	CPLCW
7.	CPLCA
8.	CPLCP
9.	CPG
10.	Apidermin
11.	CPAP1
12.	CPAP3
13.	CPCFC

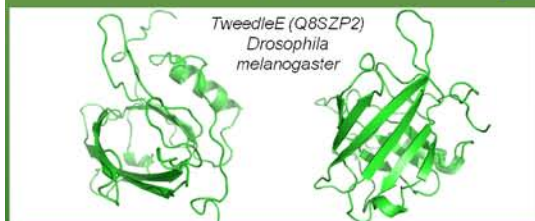


X-Ray Diffraction, Resolution: 1.9 Å, Zanotti et al. (1994)

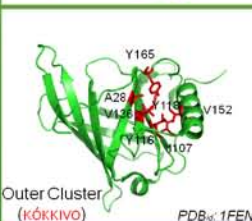
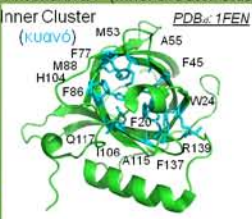
Σε επίπεδο ακολουθίας υπάρχει:
Sequence Identity < 20% και **Sequence Similarity ≈ 50%**

Κατασκευή δομικού μοντέλου, εφαρμόζοντας προτυποποίηση με βάση την ομολογία (**Homology modelling**), χρησιμοποιώντας ως δομή οδηγό, την δομή της λιποκαλίνης RBP (1FEN). Χρησιμοποιήθηκε το λογισμικό modeller9v2.

Στα μοντέλα που κατασκευάστηκαν, παρατηρούμε ότι το βασικό δομικό τους μοτίβο είναι ένα **β-βαρέλι** που απαρτίζεται από οχτώ (8) αντισυμπάλληλους β-κλώνους και μία (1) α-έλικα παράλληλη προς τους β-κλώνους.



Υδρόφοβα Clusters Λιποκαλίνων
Έχει παρατηρηθεί η παρουσία συστάδων (Clusters) υδρόφωβων αμινοξικών καταλοίπων στον πυρήνα και εξωτερικά των δομών των λιποκαλίνων (Inner & Outer Cluster).



Ανάλογες συστάδες εντοπίζονται και στη δομή των πρωτεϊνών της οικογένειας Tweedle. Στους παρακάτω πίνακες παρουσιάζονται τα κατάλογοι που αντιστοιχούν στις συστάδες όπως προέκυψαν μετά από δομική στοιχίση με την RBP (1FEN). Με **πράσινο** είναι χρωματισμένα τα υδρόφοβα κατάλογοι.

Inner Cluster												
Anopheles gambiae				Drosophila melanogaster						Danaus p.	Bombyx mori	
Lipocalms	1FEN	AgamTWDL8	AgamTWDL12	TweedleE	TweedleJ	TweedleQ	TweedleS	TweedleZ	G6D0I0	BmorCPT1	BmorCPT4	BmorCPT1
39	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20
48	T24	T24	T24	T24	T24	T24	T24	T24	T24	T24	T24	T24
80	T45	T45	T45	T45	T45	T45	T45	T45	T45	T45	T45	T45
91	T52	T52	T52	T52	T52	T52	T52	T52	T52	T52	T52	T52
93	T53	T53	T53	T53	T53	T53	T53	T53	T53	T53	T53	T53
115	T72	T72	T72	T72	T72	T72	T72	T72	T72	T72	T72	T72
131	T86	T86	T86	T86	T86	T86	T86	T86	T86	T86	T86	T86
153	T99	T99	T99	T99	T99	T99	T99	T99	T99	T99	T99	T99
156	T104	T104	T104	T104	T104	T104	T104	T104	T104	T104	T104	T104
158	T106	T106	T106	T106	T106	T106	T106	T106	T106	T106	T106	T106
160	T110	T110	T110	T110	T110	T110	T110	T110	T110	T110	T110	T110
170	T117	T117	T117	T117	T117	T117	T117	T117	T117	T117	T117	T117
190	T131	T131	T131	T131	T131	T131	T131	T131	T131	T131	T131	T131
192	T139	T139	T139	T139	T139	T139	T139	T139	T139	T139	T139	T139
Outer Cluster												
Anopheles gambiae				Drosophila melanogaster						Danaus p.	Bombyx mori	
Lipocalms	1FEN	AgamTWDL8	AgamTWDL12	TweedleE	TweedleJ	TweedleQ	TweedleS	TweedleZ	G6D0I0	BmorCPT1	BmorCPT4	BmorCPT1
52	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20	T20
159	T107	T107	T107	T107	T107	T107	T107	T107	T107	T107	T107	T107
169	T116	T116	T116	T116	T116	T116	T116	T116	T116	T116	T116	T116
171	T118	T118	T118	T118	T118	T118	T118	T118	T118	T118	T118	T118
189	T136	T136	T136	T136	T136	T136	T136	T136	T136	T136	T136	T136
205	T162	T162	T162	T162	T162	T162	T162	T162	T162	T162	T162	T162
220	T182	T182	T182	T182	T182	T182	T182	T182	T182	T182	T182	T182

Συμπεράσματα

- ✓ Οι πρωτεΐνες της οικογένειας Tweedle εμφανίζουν κοινά χαρακτηριστικά με τις πρωτεΐνες της οικογένειας των λιποκαλίνων
- ✓ Το βασικό δομικό τους μοτίβο είναι ένα **β-βαρέλι** που απαρτίζεται από οχτώ (8) αντισυμπάλληλους β-κλώνους και μία (1) α-έλικα παράλληλη προς τους β-κλώνους.
- ✓ Τα πειράματα αγκυροβόλησης που έχουν πραγματοποιηθεί ως τώρα υποδηλώνουν ότι οι πρωτεΐνες της οικογένειας Tweedle είναι δυνατόν να αλληλεπιδρούν με χιτίνη. Πρέπει να πραγματοποιηθούν πειράματα και άλλες δοκιμές με αγκυροβόληση, προκειμένου να καταλήξουμε σε ένα μοντέλο αλληλεπιδράσεως τους με τις ίνες χιτίνης

Η χιτίνη αλληλεπιδρά παράλληλα στους β-κλώνους και διαπερνάει το εξωτερικό υδρόφοβο cluster, παράλληλα στην α-έλικα



High resolution Docking grid: 3.0, GRAMM

Πειράματα αγκυροβόλησης (Docking)

Πραγματοποιήθηκαν πειράματα αγκυροβόλησης (Docking) των δομικών μοντέλων με αλυσίδες χιτίνης. Χρησιμοποιήθηκαν τα λογισμικά GRAMM και AutoDock 4.



Αποτέλεσμα αγκυροβόλησης του μοντέλου της TweedleE (Q8SZP2) με μια αλυσίδα χιτίνης

3. Abstract και Poster στο 8^ο επιστημονικό συνέδριο της Ελληνικής Εταιρείας Υπολογιστικής Βιολογίας και Βιοπληροφορικής, 22-24 Νοεμβρίου 2013

*Proceedings of the 8th conference of the Hellenic Society for Computational Biology & Bioinformatics
HSCBB13 – Department of computer Science and Biomedical Informatics
University of Thessaly, Lamia, Greece, 22-11-2013*



Computational studies of the Tweedle family cuticular proteins: analysis of putative structure and interactions with chitin filaments

Theocharis Giannakouras, Nikos Papandreou*, Vassiliki Iconomidou, Stavros Hamodrakas

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis 157 01, Athens, Greece

*Correspondence to: npapand@biol.uoa.gr

I. INTRODUCTION

Cuticle is a complex, bipartite, composite material that provides structural and mechanical support, acting functionally both as skin and as skeleton to arthropods. It is composed of chitin filaments embedded in a proteinaceous matrix. The extraordinary physical properties of cuticle are determined by the structure of its two major components, cuticular proteins (CPs) and chitin, and, also, by their interactions [1], [2].

Homology modeling studies and docking experiments suggested an antiparallel β -sheet half-barrel structure as the basic folding motif of the R&R Consensus and provided insights for the modes of CPRs interactions with chitin. These attempts were based on the observation that CPR protein sequences share sequence similarity to bovine plasma retinol binding (RBP) protein, which belongs to the lipocalin family [3], [4], [5], [6].

The next step was to study other families of structural cuticular proteins and their interactions with chitin. For example, Tweedle (TWDL) is a large family of structural cuticular proteins and it has, experimentally, been proved that they bind to chitin. To date, 236 sequences from 17 different species have been identified as members of this family. The family is widespread, with representatives already identified in seven insect orders [7].

In this work, we present sequence similarities between the proteins of the Tweedle family and bovine Retinol Binding Protein (RBP) lipocalin (PDB accession number: 1FEN). We propose that the basic folding motif of Tweedle proteins is a full β -barrel, which consists mainly of eight (8) antiparallel β -strands and an (1) α -helix. Also, structural similarities are displayed, especially the conservation of the hydrophobic "inner" and "outer" clusters, between Tweedle proteins and bovine RBP lipocalin. Furthermore, the model's structural characteristics were analyzed and their hypothetical binding properties were certified. Finally, docking experiments were carried out, in order to provide a possible interaction model of Tweedle proteins with chitin filaments.

II. MATERIALS AND METHODS

A set of protein sequences, belonging to the Tweedle family, was gathered. The set consists of ten (10) Tweedle protein sequences from four (4) different arthropods. Pairwise

alignments of these sequences with the sequence of bovine RBP (Retinol Binding Protein – PDB accession number: 1FEN) lipocalin were produced with CLUSTALW.

Comparative modeling experiments were carried out with Modeller9v2, based on the pairwise alignments and using as template the experimentally determined structure of bovine plasma retinol binding protein, (1FEN), which belongs to the class of lipocalins.

Structural alignments of the Tweedle protein models with the structure of bovine RBP were produced with the Combinatorial Extension method.

Docking experiments were performed, independently, utilizing the program GRAMM and the program Autodock4.2, with target protein structure a Tweedle protein model, utilizing as ligand a variety of chitin chains (NAG, N-acetylglucosamine etc).

III. RESULTS - DISCUSSION

Although sequence identity is rather low (below 20%) in all comparisons, levels of sequence similarity (ca. 50% conservative substitutions) are adequate for the construction of structural models of these proteins by comparative modeling (approximately 50%). Analysis of the structural characteristics of the models, led us to propose that their basic folding motif is a full β -sheet barrel, which consists mainly of eight (8) antiparallel β -strands and an (1) α -helix.

Furthermore, structural alignments of the Tweedle protein models with the structure of bovine RBP were performed. The results show a remarkable conservation of the hydrophobic residues into the inner and outer hydrophobic clusters of the lipocalin-like protein models.

Also, the presence of hydrophobic residues inside the barrels was checked. The results show that the inner part of the barrels consists mainly of hydrophobic residues. A significant high percentage of aromatic residues (phenylalanines – F, tyrosines – Y) inside the barrels were noted.

All these results reveal that the Tweedle proteins share the same structural characteristics with the lipocalin family ([8], [9]).

Finally, docking experiments were carried out, with target protein structures the Tweedle protein models, utilizing as ligands a variety of chitin chains, in order to find plausible ways of interaction between cuticular proteins and chitin chains. In the majority of the thermodynamically favored cases, it seems that the chitin filaments may interact with the Tweedle protein beta barrels, arranged parallel to the β -sheet strands, which is in agreement to the proposed chitin crystal model [10] and the long-ago proposed structure of chitin-protein complexes [1].

Further, more refined experimental work is needed, to find at atomic detail interactions of chitin with cuticular proteins and reveal the secrets of the molecular architecture of arthropod cuticle. This can be achieved, perhaps, co-crystallizing Tweedle cuticular proteins with chitin chains and solving the resulting crystal structures.

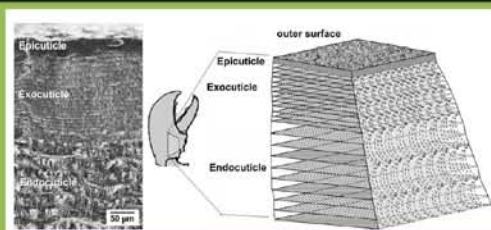
REFERENCES

- [1] Blackwell, J., Weih, M.A., 1980. Structure of chitin-protein complexes: ovipositor of the ichneumon fly *Megarhyssa*. *J Mol Biol* 137, 49-60.
- [2] Vincent, J.F., Wegst, U.G., 2004. Design and mechanical properties of insect cuticle. *Arthropod Struct Dev* 33, 187-199.
- [3] Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 1999. Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle? *Insect Biochem Mol Biol* 29, 285-292.
- [4] Iconomidou, V.A., Chrysos, G.D., Gionis, V., Willis, J.H., Hamodrakas, S.J., 2001. "Soft"-cuticle protein secondary structure as revealed by FT-Raman, ATR FT-IR and CD spectroscopy. *Insect Biochem Mol Biol* 31, 877-885.
- [5] Hamodrakas, S.J., Willis, J.H., Iconomidou, V.A., 2002. A structural model of the chitin-binding domain of cuticle proteins. *Insect Biochem Mol Biol* 32, 1577-1583.
- [6] Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 2005. Unique features of the structural model of 'hard' cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle. *Insect Biochem Mol Biol* 35, 553-560.
- [7] Willis, J.H., 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem Mol Biol* 40, 189-204.
- [8] Flower, D.R., North, A.C., Sansom, C.E., 2000. The lipocalin protein family: structural and sequence overview. *Biochim Biophys Acta* 1482, 9-24.
- [9] Adam, B., Charlotiaux, B., Beaufays, J., Vanhamme, L., Godfroid, E., Brasseur, R., Lins, L., 2008. Distantly related lipocalins share two conserved clusters of hydrophobic residues: use in homology modeling. *BMC Struct Biol* 8, 1.
- [10] Atkins, E., 1985. Conformations in polysaccharides and complex carbohydrates. *Proceeding of the International Symposium on Biomolecular Structure and Interactions. Journal of Biosciences* 8, 375-387.

COMPUTATIONAL STUDIES OF THE TWEEDELE FAMILY CUTICULAR PROTEINS: ANALYSIS OF PUTATIVE STRUCTURE AND INTERACTIONS WITH CHITIN FILAMENTS

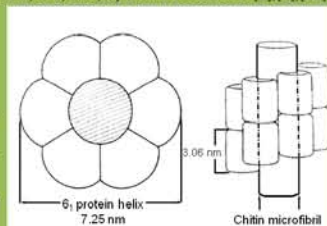
Theocharis Giannakouras, Nikos Papandreou, Vassiliki Iconomidou, Stavros Hamodrakas

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis 157 01, Athens, Greece



Cuticle is a complex, bipartite, composite material that provides structural and mechanical support, acting functionally both as skin and as skeleton to arthropods. It is composed of chitin filaments embedded in a proteinaceous matrix. The extraordinary physical properties of cuticle are determined by the structure of its two major components, cuticular proteins (CPs) and chitin, and, also, by their interactions. [1],[2],[12]

Chitin filaments embedded in a proteinaceous matrix, according to the model of Blackwell and Weih. [1]



Purpose

In this work, we present sequence similarities between the proteins of Tweedle family and the bovine Retinol Binding Protein (RBP) lipocalin (PDB accession number: 1FEN). We propose that the basic folding motif of Tweedle proteins is a full β -barrel, which consists mainly of eight (8) antiparallel β -strands and an (1) α -helix parallel to the strands. Also, structural similarities are displayed, between Tweedle proteins and bovine RBP. Finally, docking experiments were carried out, in order to provide a possible interaction model of Tweedle proteins with chitin filaments.

Materials and methods

A set of protein sequences, belonging to the Tweedle family, was collected from UniProt and cuticleDB. The set consists of ten (10) Tweedle protein sequences from four (4) different arthropods.

Pairwise alignments of these sequences with the sequence of bovine RBP (Retinol Binding Protein – PDB accession number: 1FEN) were produced with CLUSTALW.

Comparative modeling experiments were carried out with Modeller9v2, based on the pairwise alignments and using as template the experimentally determined structure of bovine plasma retinol binding protein RBP, (1FEN), which belongs to the class of lipocalins.

Structural alignments of the Tweedle protein models with the structure of bovine RBP were produced by the Combinatorial Extension method.

Docking experiments were performed, independently, utilizing the program GRAMM and the program Autodock4.2, with target protein structure a Tweedle protein model, utilizing as ligand a variety of chitin chains (NAG, N-acetylglucosamine etc).

Docking experiments were carried out, with target protein structures the Tweedle protein models, utilizing as ligands a variety of chitin chains, in order to find plausible ways of interaction between cuticular proteins and chitin chains. In the majority of the thermodynamically favored cases, it seems that the chitin filaments may interact with the Tweedle protein beta barrels, arranged parallel to the β -sheet strands, which is in agreement to the proposed chitin crystal model [10] and the long-ago proposed structure of chitin-protein complexes [1].



In conclusion, more refined experimental work is needed, to find at atomic detail interactions of chitin with cuticular proteins and reveal the secrets of the molecular architecture of arthropod cuticle. This can be achieved, perhaps, co-crystallizing Tweedle cuticular proteins with chitin chains and solving the resulting crystal structures.

Introduction

Cuticular proteins are classified into thirteen distinct families according to their amino acid sequences. [7]

Cuticular proteins families	
1.	CPR
2.	CPF
3.	CPFL
4.	TWEEDLE (TWDL)
5.	CPLCG
6.	CPLCW
7.	CPLCA
8.	CPLCP
9.	CPG
10.	Apidermin
11.	CPAP1
12.	CPAP3
13.	CPCFC

Tweedle (TWDL) is a large family of structural cuticular proteins and it has experimentally been proved that they bind to chitin. To date, 236 sequences from 17 different species have been identified as members of this family. The family is widespread, with representatives already identified in seven insect orders. [7]

Homology modeling studies and docking experiments suggested an antiparallel β -sheet half-barrel structure as the basic folding motif of the R&R Consensus and provided insights for the modes of CPRs interactions with chitin. These attempts were based on the observation that CPR protein sequences share sequence similarity to bovine plasma retinol binding (RBP - PDB accession number: 1FEN) protein, which belongs to the lipocalin family. [3],[4],[5],[6]



Lipocalins protein family members are typically small (160-200 residues) secreted proteins sharing several common molecular recognition properties. Despite common functions, lipocalins share low pairwise sequence identity (approximately 20%) among family members. However, they share the same structure, which consists mainly of eight (8) strands forming an anti-parallel up and down β -barrel. Also, a 3_1 helix caps one extremity of the barrel and a second α -helix is parallel to its surface. Moreover, the presence of two hydrophobic clusters of residues has been identified. One cluster is internal to the barrel (inner cluster), involving all strands and the 3_1 helix. The other is external, involving four strands and the helix lying parallel to the barrel surface (outer cluster). [8],[9]

Results - Discussion

Pairwise alignments results of Tweedle family members set against bovine RBP (1FEN).

Although sequence identity is rather low (below 20%) in all comparisons, levels of sequence similarity (ca. 50% conservative substitutions) are adequate for the construction of structural models of these proteins by comparative modeling (approximately 50%).

Organism	Protein (UniProt)	Sequence identity	Sequence similarity
<i>Anopheles gambiae</i>	AgamTWDL8 (819-CuticleDB)	20/177 = 11.2%	76/177 = 42.9%
<i>Anopheles gambiae</i>	AgamTWDL12 (Q709V7)	21/177 = 11.8%	75/177 = 42.4%
<i>Drosophila melanogaster</i>	TweedleE (Q8SP2)	17/177 = 9.6%	69/177 = 39%
<i>Drosophila melanogaster</i>	TweedleJ (Q9VBD6)	31/177 = 17.5%	102/177 = 57.6%
<i>Drosophila melanogaster</i>	TweedleD (Q9VBC8)	26/177 = 14.7%	87/177 = 54.8%
<i>Drosophila melanogaster</i>	TweedleS (Q9VBD3)	21/177 = 11.8%	87/177 = 49.2%
<i>Drosophila melanogaster</i>	TweedleZ (Q8IR09)	19/177 = 10.7%	75/177 = 42.4%
<i>Drosophila melanogaster</i>	TweedleBeta (A128Q0)	18/177 = 10.2%	77/177 = 43.5%
<i>Danaus plexipus</i>	CP tweedle motif 4 (G6D0I0)	19/177 = 10.7%	90/177 = 50.8%
<i>Bombyx mori</i>	BmorCPT1 (C0H6Z1)	30/177 = 17%	93/177 = 52.5%
<i>Bombyx mori</i>	BmorCPT4 (C0H6Z4)	14/183 = 8.6%	70/183 = 43%



Structural alignments of the Tweedle protein models with the structure of bovine RBP were performed. The results show a remarkable conservation of the hydrophobic residues into the inner and outer hydrophobic clusters of the lipocalin-like protein models.

		Inner Cluster											
		Anopheles gambiae		Drosophila melanogaster								Ornith. pil.	Bombus morio
Lipocalin	1FEN	AgamTWDL8	AgamTWDL12	TweedleE	TweedleJ	TweedleD	TweedleS	TweedleZ	TweedleBeta	TweedleDelta	Ornith. pil.	Bombus morio	
20	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
85	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
86	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
91	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
93	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
115	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
132	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
133	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
158	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
168	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
170	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
190	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
192	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
Outer Cluster													
Lipocalin	1FEN	Anopheles gambiae		Drosophila melanogaster								Ornith. pil.	Bombus morio
Lipocalin	1FEN	AgamTWDL8	AgamTWDL12	TweedleE	TweedleJ	TweedleD	TweedleS	TweedleZ	TweedleBeta	TweedleDelta	Ornith. pil.	Bombus morio	
52	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
159	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
168	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
171	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
198	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	
205	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	1.04	

Βιβλιογραφία

- Adam, B., B. Charletoaux, et al. (2008). "Distantly related lipocalins share two conserved clusters of hydrophobic residues: use in homology modeling." BMC Struct Biol **8**: 1.
- Andersen, S. O. (1998). "Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*." Insect Biochem Mol Biol **28**(5-6): 421-34.
- Andersen, S. O. (2000). "Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*." Insect Biochem Mol Biol **30**(7): 569-77.
- Andersen, S. O. (2010). "Insect cuticular sclerotization: a review." Insect Biochem Mol Biol **40**(3): 166-78.
- Andersen, S. O., P. Hojrup, et al. (1995). "Insect cuticular proteins." Insect Biochem Mol Biol **25**(2): 153-76.
- Andersen, S. O., K. Rafn, et al. (1997). "Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*." Insect Biochem Mol Biol **27**(2): 121-31.
- Apweiler, R., A. Bairoch, et al. (2004). "UniProt: the Universal Protein knowledgebase." Nucleic Acids Res **32**(Database issue): D115-9.
- Arakane, Y., Q. Zhu, et al. (2003). "Properties of catalytic, linker and chitin-binding domains of insect chitinase." Insect Biochem Mol Biol **33**(6): 631-48.
- Ashida, M. and P. T. Brey (1995). "Role of the integument in insect defense: pro-phenol oxidase cascade in the cuticular matrix." Proc Natl Acad Sci U S A **92**(23): 10698-702.
- Atkins, E. (1985). "Conformations in polysaccharides and complex carbohydrates." Proceeding of the International Symposium on Biomolecular Structure and Interactions. Journal of Biosciences **8**(Suppl.): 375-387.
- Barry, M. K., A. A. Triplett, et al. (1999). "A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*." Insect Biochem Mol Biol **29**(4): 319-27.
- Behr, M. and M. Hoch (2005). "Identification of the novel evolutionary conserved obstructor multigene family in invertebrates." FEBS Lett **579**(30): 6827-33.
- Benson, D. A., M. Cavanaugh, et al. (2013). "GenBank." Nucleic Acids Res **41**(Database issue): D36-42.
- Berman, H. M., J. Westbrook, et al. (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-42.
- Berni, R., M. Stoppini, et al. (1990). "The bovine plasma retinol-binding protein. Amino acid sequence, interaction with transthyretin, crystallization and preliminary X-ray data." Eur J Biochem **192**(2): 507-13.

- Blackwell, J. and M. A. Weih (1980). "Structure of chitin-protein complexes: ovipositor of the ichneumon fly *Megarhyssa*." J Mol Biol **137**(1): 49-60.
- Bouhin, H., J. P. Charles, et al. (1992). "Characterization of a cDNA clone encoding a glycine-rich cuticular protein of *Tenebrio molitor*: developmental expression and effect of a juvenile hormone analogue." Insect Mol Biol **1**(2): 53-62.
- Bouhin, H., J. P. Charles, et al. (1992). "Developmental profiles of epidermal mRNAs during the pupal-adult molt of *Tenebrio molitor* and isolation of a cDNA clone encoding an adult cuticular protein: effects of a juvenile hormone analogue." Dev Biol **149**(1): 112-22.
- Charles, J. P., H. Bouhin, et al. (1992). "cDNA cloning and deduced amino acid sequence of a major, glycine-rich cuticular protein from the coleopteran *Tenebrio molitor*. Temporal and spatial distribution of the transcript during metamorphosis." Eur J Biochem **206**(3): 813-9.
- Chou, P. Y. and G. D. Fasman (1974a). "Prediction of protein conformation." Biochemistry **13**(2): 222-45.
- Chou, P. Y. and G. D. Fasman (1974b). "Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins." Biochemistry **13**(2): 211-22.
- Cohen, E. (2001). "Chitin synthesis and inhibition: a revisit." Pest Manag Sci **57**(10): 946-50.
- Coles, M., T. Diercks, et al. (1999). "The solution structure and dynamics of human neutrophil gelatinase-associated lipocalin." J Mol Biol **289**(1): 139-57.
- Cornman, R. S. (2009). "Molecular evolution of *Drosophila* cuticular protein genes." PLoS One **4**(12): e8345.
- Cornman, R. S. (2010). "The distribution of GYR- and YLP-like motifs in *Drosophila* suggests a general role in cuticle assembly and other protein-protein interactions." PLoS One **5**(9).
- Cornman, R. S. and J. H. Willis (2009). "Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle." Insect Mol Biol **18**(5): 607-22.
- Cowan, S. W., M. E. Newcomer, et al. (1990). "Crystallographic refinement of human serum retinol binding protein at 2A resolution." Proteins **8**(1): 44-61.
- Crooks, G. E., G. Hon, et al. (2004). "WebLogo: a sequence logo generator." Genome Res **14**(6): 1188-90.
- Delano, W. L. (2005). The PyMOL Molecular Graphics System. 400, Oyster Point Blvd., Suite 213, South San Francisco, CA 94080-1918, USA, DeLano Scientific LLC.
- Dufton, M. J. and R. C. Hider (1977). "Snake toxin secondary structure predictions. Structure activity relationships." J Mol Biol **115**(2): 177-93.

- Elgavish, S. and B. Shaanan (1997). "Lectin-carbohydrate interactions: different folds, common recognition principles." Trends Biochem Sci **22**(12): 462-7.
- Flower, D. R. (1996). "The lipocalin protein family: structure and function." Biochem J **318** (Pt 1): 1-14.
- Flower, D. R., A. C. North, et al. (2000). "The lipocalin protein family: structural and sequence overview." Biochim Biophys Acta **1482**(1-2): 9-24.
- Fraenkel, G. and K. M. Rudall (1947). "The structure of insect cuticles." Proc R Soc Med **134**(874): 111-43.
- Futahashi, R., S. Okamoto, et al. (2008). "Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*." Insect Biochem Mol Biol **38**(12): 1138-46.
- Gardner, K. H. and J. Blackwell (1975). "Refinement of the structure of beta-chitin." Biopolymers **14**(8): 1581-95.
- Garnier, J., D. J. Osguthorpe, et al. (1978). "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins." J Mol Biol **120**(1): 97-120.
- Greenspan, R. J. and J. F. Ferveur (2000). "Courtship in *Drosophila*." Annu Rev Genet **34**: 205-232.
- Guan, X., B. W. Middlebrooks, et al. (2006). "Mutation of TweedleD, a member of an unconventional cuticle protein family, alters body shape in *Drosophila*." Proc Natl Acad Sci U S A **103**(45): 16794-9.
- Gupta, N. S., H. Fabritius, et al. (2011). Chitin in the Exoskeletons of Arthropoda: From Ancient Design to Novel Materials Science. Chitin, Springer Netherlands. **34**: 35-60.
- Hakim, R. S., M. B. Blackburn, et al. (2007). "Growth and mitogenic effects of arylphorin in vivo and in vitro." Arch Insect Biochem Physiol **64**(2): 63-73.
- Hall, J. C. (1994). "The mating of a fly." Science **264**(5166): 1702-14.
- Hamodrakas, S. J. (1988). "A protein secondary structure prediction scheme for the IBM PC and compatibles." Comput Appl Biosci **4**(4): 473-7.
- Hamodrakas, S. J. (1992). Molecular architecture of helicoidal proteinaceous eggshells. Results and problems in cell differentiation. Berlin-Heidelberg, SpringerVerlag. **19**: 116-186.
- Hamodrakas, S. J., P. N. Kanellopoulos, et al. (1997). "The crystal structure of the complex of concanavalin A with 4'-methylumbelliferyl- α -D-glucopyranoside." J Struct Biol **118**(1): 23-30.
- Hamodrakas, S. J., J. H. Willis, et al. (2002). "A structural model of the chitin-binding domain of cuticle proteins." Insect Biochem Mol Biol **32**(11): 1577-83.

- Han, Q., J. Fang, et al. (2002). "Identification of *Drosophila melanogaster* yellow-f and yellow-f2 proteins as dopachrome-conversion enzymes." Biochem J **368**(Pt 1): 333-40.
- He, N., J. M. Botelho, et al. (2007). "Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry." Insect Biochem Mol Biol **37**(2): 135-46.
- Hickman, C. P., L. S. Roberts, et al. (2001). Integrated principles of zoology. Boston, McGraw-Hill.
- Holden, H. M., W. R. Rypniewski, et al. (1987). "The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution." EMBO J **6**(6): 1565-70.
- Hooft, R. W., G. Vriend, et al. (1996). "Errors in protein structures." Nature **381**(6580): 272.
- Huber, R., M. Schneider, et al. (1987). "Molecular structure of the bilin binding protein (BBP) from *Pieris brassicae* after refinement at 2.0 Å resolution." J Mol Biol **198**(3): 499-513.
- Iconomidou, V. A., G. D. Chrysikos, et al. (2001). "'Soft'-cuticle protein secondary structure as revealed by FT-Raman, ATR FT-IR and CD spectroscopy." Insect Biochem Mol Biol **31**(9): 877-85.
- Iconomidou, V. A., J. H. Willis, et al. (1999). "Is beta-pleated sheet the molecular conformation which dictates formation of helicoidal cuticle?" Insect Biochem Mol Biol **29**(3): 285-92.
- Iconomidou, V. A., J. H. Willis, et al. (2005). "Unique features of the structural model of 'hard' cuticle proteins: implications for chitin-protein interactions and cross-linking in cuticle." Insect Biochem Mol Biol **35**(6): 553-60.
- Jang, M.-K., B.-G. Kong, et al. (2004). "Physicochemical characterization of α -chitin, β -chitin, and γ -chitin separated from natural resources." Journal of Polymer Science Part A: Polymer Chemistry **42**(14): 3423-3432.
- Jasrapuria, S., Y. Arakane, et al. (2010). "Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, expression and function." Insect Biochem Mol Biol **40**(3): 214-27.
- Jensen, U. G., A. Rothmann, et al. (1997). "Cuticular proteins from the giant cockroach, *Blaberus craniifer*." Insect Biochem Mol Biol **27**(2): 109-20.
- Karouzou, M. V., Y. Spyropoulos, et al. (2007). "Drosophila cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences." Insect Biochem Mol Biol **37**(8): 754-60.
- Katchalski-Katzir, E., I. Shariv, et al. (1992). "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques." Proc Natl Acad Sci U S A **89**(6): 2195-9.

- Katouno, F., M. Taguchi, et al. (2004). "Importance of exposed aromatic residues in chitinase B from *Serratia marcescens* 2170 for crystalline chitin hydrolysis." J Biochem **136**(2): 163-8.
- Kim, E., Y. Choi, et al. (2008). "Characterization of the *Drosophila melanogaster* retinin gene encoding a cornea-specific protein." Insect Mol Biol **17**(5): 537-43.
- Kornezos, A. and W. Chia (1992). "Apical secretion and association of the *Drosophila* yellow gene product with developing larval cuticle structures during embryogenesis." Mol Gen Genet **235**(2-3): 397-405.
- Kucharski, R., J. Maleszka, et al. (2007). "Novel cuticular proteins revealed by the honey bee genome." Insect Biochem Mol Biol **37**(2): 128-34.
- Lee, W. J. and P. T. Brey (1994). "Isolation and identification of cecropin antibacterial peptides from the extracellular matrix of the insect integument." Anal Biochem **217**(2): 231-5.
- Li, Y., Y. Wang, et al. (2009). "Crystal structure of *Manduca sexta* prophenoloxidase provides insights into the mechanism of type 3 copper enzymes." Proc Natl Acad Sci U S A **106**(40): 17002-6.
- Lim, V. I. (1974a). "Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure." J Mol Biol **88**(4): 857-72.
- Lim, V. I. (1974b). "Algorithms for prediction of alpha-helical and beta-structural regions in globular proteins." J Mol Biol **88**(4): 873-94.
- Locke, M. (2001). "The Wigglesworth Lecture: Insects for studying fundamental problems in biology." J Insect Physiol **47**(4-5): 495-507.
- Magkrioti, C. K., I. C. Spyropoulos, et al. (2004). "cuticleDB: a relational database of Arthropod cuticular proteins." BMC Bioinformatics **5**: 138.
- Marcu, O. and M. Locke (1998). "A cuticular protein from the moulting stages of an insect." Insect Biochem Mol Biol **28**(9): 659-69.
- Minke, R. and J. Blackwell (1978). "The structure of alpha-chitin." J Mol Biol **120**(2): 167-81.
- Molnar, K., N. H. Borhegyi, et al. (2001). "The immunoprotein scolexin and its synthesizing sites--the midgut epithelium and the epidermis." Acta Biol Hung **52**(4): 473-84.
- Morganti, P. and G. Morganti (2008). "Chitin nanofibrils for advanced cosmeceuticals." Clin Dermatol **26**(4): 334-40.
- Morris, G. M., D. S. Goodsell, et al. (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of Computational Chemistry **19**(14): 1639-1662.
- Morris, G. M., R. Huey, et al. (2009). "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility." J Comput Chem **30**(16): 2785-91.

- Nagano, K. (1977a). "Logical analysis of the mechanism of protein folding. IV. Super-secondary structures." J Mol Biol **109**(2): 235-50.
- Nagano, K. (1977b). "Triplet information in helix prediction applied to the analysis of super-secondary structures." J Mol Biol **109**(2): 251-74.
- Nakato, H., S. Izumi, et al. (1992). "Structure and expression of gene coding for a pupal cuticle protein of *Bombyx mori*." Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1132**(2): 161-167.
- Nakatogawa, S., Y. Oda, et al. (2009). "A novel peptide mediates aggregation and migration of hemocytes from an insect." Curr Biol **19**(9): 779-85.
- Nene, V., J. R. Wortman, et al. (2007). "Genome sequence of *Aedes aegypti*, a major arbovirus vector." Science **316**(5832): 1718-23.
- Neville, A. C. (1975). Biology of the arthropod cuticle. Berlin ; New York, Springer-Verlag.
- Nisole, A., D. Stewart, et al. (2010). "Cloning and characterization of a Gasp homolog from the spruce budworm, *Choristoneura fumiferana*, and its putative role in cuticle formation." J Insect Physiol **56**(10): 1427-35.
- Paesen, G. C., P. L. Adams, et al. (1999). "Tick histamine-binding proteins: isolation, cloning, and three-dimensional structure." Mol Cell **3**(5): 661-71.
- Papandreou, N. C., V. A. Iconomidou, et al. (2010). "A possible structural model of members of the CPF family of cuticular proteins implicating binding to components other than chitin." J Insect Physiol **56**(10): 1420-6.
- Petersen, T. N., S. Brunak, et al. (2011). "SignalP 4.0: discriminating signal peptides from transmembrane regions." Nat Methods **8**(10): 785-6.
- Pettersen, E. F., T. D. Goddard, et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-12.
- Punta, M., P. C. Coghill, et al. (2012). "The Pfam protein families database." Nucleic Acids Res **40**(Database issue): D290-301.
- Qin, G., S. Lapidot, et al. (2009). "Expression, cross-linking, and characterization of recombinant chitin binding resilin." Biomacromolecules **10**(12): 3227-34.
- Qiu, J. and P. E. Hardin (1995). "Temporal and spatial expression of an adult cuticle protein gene from *Drosophila* suggests that its protein product may impart some specialized cuticle function." Dev Biol **167**(2): 416-25.
- Quiocho, F. A. (1989). "Protein-carbohydrate interactions: basic molecular features." Pure and Applied Chemistry **61**: 1293-1306.
- Raabe, D., P. Romano, et al. (2006). "Microstructure and crystallographic texture of the chitin-protein network in the biological composite material of the exoskeleton of the lobster *Homarus americanus*." Materials Science and Engineering: A **421**(1-2): 143-153.

- Ramana, J. and D. Gupta (2009). "LipocalinPred: a SVM-based method for prediction of lipocalins." BMC Bioinformatics **10**: 445.
- Rebers, J. E. and L. M. Riddiford (1988). "Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes." J Mol Biol **203**(2): 411-23.
- Rebers, J. E. and J. H. Willis (2001). "A conserved domain in arthropod cuticular proteins binds chitin." Insect Biochem Mol Biol **31**(11): 1083-93.
- Reynolds, S. E. (1975). "The mechanical properties of the abdominal cuticle of *Rhodnius* larvae." J Exp Biol **62**(1): 69-80.
- Riddiford, L. M. and R. H. Hice (1985). "Developmental profiles of the mRNAs for *Manduca* arylphorin and two other storage proteins during the final larval instar of *Manduca sexta*." Insect Biochemistry **15**(4): 489-502.
- Rost, B. (1996). "PHD: predicting one-dimensional protein structure by profile-based neural networks." Methods Enzymol **266**: 525-39.
- Rouvinen, J., J. Rautiainen, et al. (1999). "Probing the molecular basis of allergy. three-dimensional structure of the bovine lipocalin allergen Bos d 2." J Biol Chem **274**(4): 2337-43.
- Rudall, K. M., J. E. T. J.W.L. Beament, et al. (1963). The Chitin/Protein Complexes of Insect Cuticles. Advances in Insect Physiology, Academic Press. **Volume 1**: 257-313.
- Ryu, K. S., J. O. Lee, et al. (2009). "The presence of monoglucosylated N196-glycan is important for the structural stability of storage protein, arylphorin." Biochem J **421**(1): 87-96.
- Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.
- Sawada, D., Y. Nishiyama, et al. (2012). "Water in crystalline fibers of dihydrate beta-chitin results in unexpected absence of intramolecular hydrogen bonding." PLoS One **7**(6): e39376.
- Scheller, K., Zimmermann, H. -P., Sekeris, C. E. (1980). "Calliphorin, a protein involved in the cuticle formation of the blowfly, *Calliphora vicina*." Z. Naturforsch **35**(c): 387-389.
- Schneider, T. D. and R. M. Stephens (1990). "Sequence logos: a new way to display consensus sequences." Nucleic Acids Res **18**(20): 6097-100.
- Seidl, V. (2008). "Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions." Fungal Biology Reviews **22**(1): 36-42.
- Shen, Z. and M. Jacobs-Lorena (1999). "Evolution of chitin-binding proteins in invertebrates." J Mol Evol **48**(3): 341-7.
- Shindyalov, I. N. and P. E. Bourne (1998). "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path." Protein Eng **11**(9): 739-47.

- Sillitoe, I., A. L. Cuff, et al. (2013). "New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures." Nucleic Acids Res **41**(Database issue): D490-8.
- Sippl, M. J. (1990). "Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins." J Mol Biol **213**(4): 859-83.
- Sousa, S. F., P. A. Fernandes, et al. (2006). "Protein-ligand docking: current status and future challenges." Proteins **65**(1): 15-26.
- Suetake, T., S. Tsuda, et al. (2000). "Chitin-binding proteins in invertebrates and plants comprise a common chitin-binding structural motif." J Biol Chem **275**(24): 17929-32.
- Sugiyama, J., C. Boisset, et al. (1999). "Molecular directionality of beta-chitin biosynthesis." J Mol Biol **286**(1): 247-55.
- Svitil, A. L. and D. L. Kirchman (1998). "A chitin-binding domain in a marine bacterial chitinase and other microbial chitinases: implications for the ecology and evolution of 1,4-beta-glycanases." Microbiology **144** (Pt 5): 1299-308.
- Tang, L., J. Liang, et al. (2010). "Identification of the chitin-binding proteins from the larval proteins of silkworm, Bombyx mori." Insect Biochem Mol Biol **40**(3): 228-34.
- Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.
- Togawa, T., W. Augustine Dunn, et al. (2007). "CPF and CPFL, two related gene families encoding cuticular proteins of Anopheles gambiae and other insects." Insect Biochem Mol Biol **37**(7): 675-88.
- Togawa, T., H. Nakato, et al. (2004). "Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, Bombyx mori." Insect Biochem Mol Biol **34**(10): 1059-67.
- Vakser, I. A. (1996). "Low-resolution docking: prediction of complexes for underdetermined structures." Biopolymers **39**(3): 455-64.
- Vincent, J. F. and U. G. Wegst (2004). "Design and mechanical properties of insect cuticle." Arthropod Struct Dev **33**(3): 187-99.
- Vriend, G. (1990). "WHAT IF: a molecular modeling and drug design program." J Mol Graph **8**(1): 52-6, 29.
- Vyas, N. K. (1991). "Atomic features of protein-carbohydrate interactions." Current Opinion in Structural Biology **1**(5): 732-740.
- Watanabe, T., Y. Ariga, et al. (2003). "Aromatic residues within the substrate-binding cleft of Bacillus circulans chitinase A1 are essential for hydrolysis of crystalline chitin." Biochem J **376**(Pt 1): 237-44.

- Willis, J. H. (1999). "Cuticular Proteins in Insects and Crustaceans." American Zoologist **39**(3): 600-609.
- Willis, J. H. (2010). "Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era." Insect Biochem Mol Biol **40**(3): 189-204.
- Willis, J. H., V. A. Iconomidou, et al. (2005). 4.2 - Cuticular Proteins. Comprehensive Molecular Insect Science. Amsterdam, Elsevier: 79-109.
- Willis, J. H., N. C. Papandreou, et al. (2012). 5 - Cuticular Proteins. Insect Molecular Biology and Biochemistry. San Diego, Academic Press: 134-166.
- Willott, E., X. Y. Wang, et al. (1989). "cDNA and gene sequence of Manduca sexta arylphorin, an aromatic amino acid-rich larval serum protein. Homology to arthropod hemocyanins." J Biol Chem **264**(32): 19052-9.
- Wittkopp, P. J., E. E. Stewart, et al. (2009). "Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila." Science **326**(5952): 540-4.
- Wittkopp, P. J., K. Vaccaro, et al. (2002). "Evolution of yellow gene regulation and pigmentation in Drosophila." Curr Biol **12**(18): 1547-56.
- Wybrandt, G. B. and S. O. Andersen (2001). "Purification and sequence determination of a yellow protein from sexually mature males of the desert locust, Schistocerca gregaria." Insect Biochem Mol Biol **31**(12): 1183-9.
- Zanotti, G., M. Marcello, et al. (1994). "Crystallographic studies on complexes between retinoids and plasma retinol-binding protein." J Biol Chem **269**(47): 29613-20.
- Zanotti, G., M. Panzalorto, et al. (1998). "Structure of pig plasma retinol-binding protein at 1.65 Å resolution." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 1049-52.
- Zhan, S., C. Merlin, et al. (2011). "The monarch butterfly genome yields insights into long-distance migration." Cell **147**(5): 1171-85.