

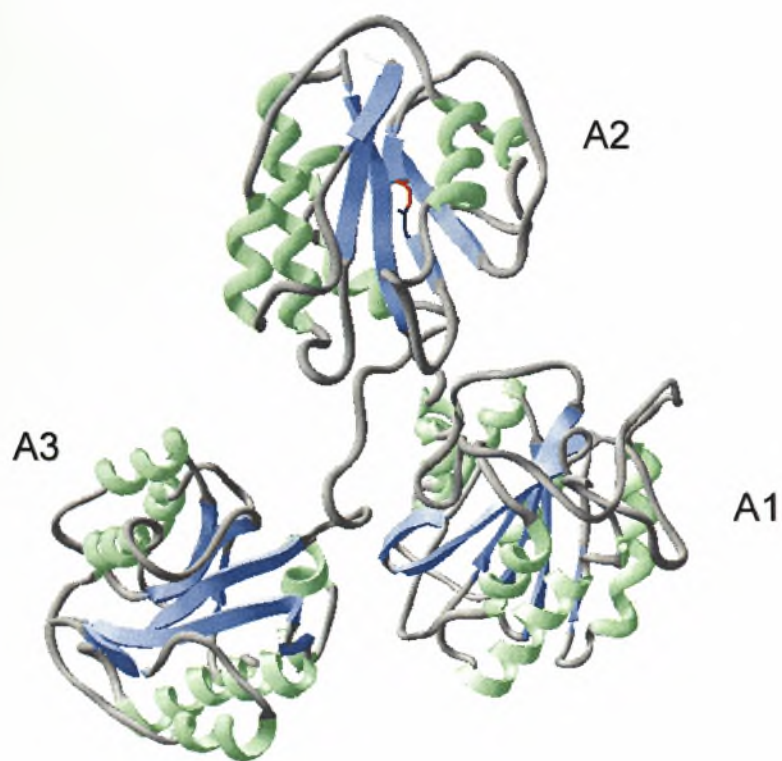


Πανεπιστήμιο Θεσσαλίας

Σχολή Επιστημών Υγείας

Τμήμα Βιοχημείας – Βιοτεχνολογίας

“Εγκατάσταση του πακέτου σχεδιασμού και πρόβλεψης δομής πρωτεϊνών MODELLER και εφαρμογή του στην πρωτεΐνη Von Willebrand factor”



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΖΙΑΣΤΟΥΔΗ ΜΑΡΙΑ

ΛΑΡΙΣΑ 2008



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΥΠΗΡΕΣΙΑ ΒΙΒΛΙΟΘΗΚΗΣ & ΠΛΗΡΟΦΟΡΗΣΗΣ
ΕΙΔΙΚΗ ΣΥΛΛΟΓΗ «ΓΚΡΙΖΑ ΒΙΒΛΙΟΓΡΑΦΙΑ»

Αριθ. Εισ.: 6601/1
Ημερ. Εισ.: 12-05-2009
Δωρεά: Π.Θ.
Ταξιδετικός Κωδικός: ΠΤ - ΒΒ
2008
ΤΖΙ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ



004000087297

Στοιχεία διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε στο εργαστήριο Πληροφορικής του τμήματος Βιοχημείας – Βιοτεχνολογίας το χρονικό διάστημα Φεβρουάριος 2008 – Ιούνιος 2008 υπό την επίβλεψη του κ. Γεώργιου Παπαδόπουλου.

Τριμελής συμβουλευτική επιτροπή:

Δρ. Παπαδόπουλος Γεώργιος (επιβλέπων)

ΠΔ 407/80

Δρ. Κοντού Μαρία

Λέκτορας Κλινικής Χημείας

Δρ. Σταθόπουλος Κωνσταντίνος

Επίκουρος Καθηγητής Βιοχημείας

Ευχαριστίες

Πολλοί είναι οι άνθρωποι που συνεισέφεραν στην ολοκλήρωση της εργασίας αυτής και προεξέχουσα θέση ανάμεσα σε αυτούς κατέχει ο επιβλέπων καθηγητής μου κ. Γ. Παπαδόπουλος, τον οποίο ευχαριστώ θερμά για την άριστη συνεργασία μας, την ανεξάντλητη υπομονή του, τις πολύτιμες συμβουλές και διορθώσεις του, καθώς και για το χρόνο που αφιέρωσε για να μου εξηγήσει και την τελευταία απορία μου. Αποτέλεσε για μένα αστείρευτη πηγή γνώσεων, ενθάρρυνσης και αμείωτου ενδιαφέροντος. Εκτός από πρότυπο επιστήμονα αποτέλεσε για μένα πρότυπο ακέραιου ανθρώπου και χαρακτήρα. Θα ήθελα, επίσης, να ευχαριστήσω θερμά τη λέκτορα κλινικής χημείας κ. Μ. Κοντού, της οποίας η βοήθεια ήταν καθοριστικής σημασίας για την περάτωση των σπουδών μου γενικότερα, καθώς και τον επίκουρο καθηγητή Βιοχημείας κ. Κ. Σταθόπουλο.

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ.....	1
1.1 Σκοπός	2
1.2 Περίληψη.....	2
2. ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ.....	
2.1 Ο παράγοντας von Willebrand	3
2.2 Μέθοδοι Προσδιορισμού Δομής Πρωτεϊνών	8
A) Πειραματικές Μέθοδοι	8
A1) Κρυσταλλογραφία	8
A2) Πυρηνικός Μαγνητικός Συντονισμός.....	10
A3) Κυκλικός Διχρωϊσμός.....	10
A4) Φασματοσκοπία Laser-Raman.....	11
A5) Φασματοσκοπία Υπερύθρου	11
B) Θεωρητικές Μέθοδοι	11
B1) Συγκριτικός Σχεδιασμός.....	12
B2) Αναγνώριση Αναδίπλωσης.....	19
B3) Απλοποιημένες προβλέψεις σε μία ή δύο διαστάσεις	23
B4) Ab initio Μέθοδοι	26
2.3 Εξυπηρέτης Phyre.....	26
2.4 Υπολογιστικές Προσομοιώσεις.....	27
2.4.1 Οι στόχοι των προσομοιώσεων Μοριακής Δυναμικής	29
2.4.2 Ο αλγόριθμος των προσομοιώσεων Μοριακής Δυναμικής	30
2.4.3 Οι περιορισμοί των προσομοιώσεων.....	34
2.5 Το πρόγραμμα Modeller.....	35
2.5.1 Οδηγίες εγκατάστασης του προγράμματος	35
2.5.2 Ο αλγόριθμος του προγράμματος	35
2.5.3 Εφαρμογή του προγράμματος.....	36
3. ΠΡΑΚΤΙΚΟ ΜΕΡΟΣ.....	47
3.1 Μοντελοποίηση τμημάτων του παράγοντα von Willebrand	47
3.1.1 Εφαρμογή του προγράμματος Modeller	47
3.1.2 Εφαρμογή της μεθόδου αναγνώρισης αναδίπλωσης	53
3.2 Προσομοίωση των θεωρητικά προβλεπόμενων δομών	54
3.2.1 Προϋποθέσεις λειτουργίας του προγράμματος NAMD	54
3.2.2 Τα 4 βήματα που προηγούνται των προσομοιώσεων.....	55
3.2.3 Εξισορρόπηση: το στάδιο όπου λαμβάνει χώρα η προσομοίωση	59

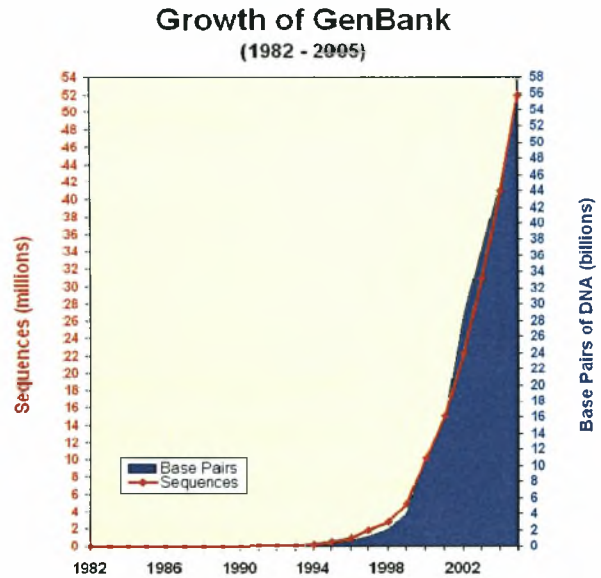
4. ΑΠΟΤΕΛΕΣΜΑΤΑ.....	60
4.1 Η δομή του πρώτου συνδετικού τμήματος.....	60
4.2 Η δομή του δεύτερου συνδετικού τμήματος.....	62
4.3 Η δομή του τμήματος Α1-Α2-Α3	63
5. ΣΥΖΗΤΗΣΗ.....	65
6. ΠΑΡΑΡΤΗΜΑ	66
7. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	70



Εισαγωγή

1. Εισαγωγή

Τα τελευταία χρόνια το χάσμα μεταξύ των πρωτεϊνικών ακολουθιών που βρίσκονται κατατεθειμένες σε ελεύθερα προσβάσιμες τράπεζες δεδομένων και των λυμένων κρυσταλλογραφικά πρωτεϊνικών δομών διαρκώς αυξάνεται. Αυτό οφείλεται τόσο στην ταχύτατη αποκρυπτογράφηση νέων γονιδιωμάτων (εικόνα 1) όσο και στους περιορισμούς στους οποίους υπόκεινται οι πειραματικές μέθοδοι προσδιορισμού δομής. Στην προσπάθεια να γεφυρωθεί αυτό το χάσμα επιστρατεύονται διάφορες θεωρητικές μέθοδοι. Μέχρι στιγμής βέβαια, πρόγνωση δομής μόνο από την ακολουθία δεν έχει επιτευχθεί με πολύ ικανοποιητικό τρόπο, αφού δεν έχει καταστεί δυνατή η απόλυτη συσχέτιση ακολουθίας-δομής. Παρά το γεγονός αυτό, αποκαλύπτονται διαρκώς κανόνες από την αξιοποίηση πληροφοριών πειραματικά προσδιορισμένων δομών. Η πιο επιτυχής θεωρητική προσέγγιση είναι ο συγκριτικός σχεδιασμός (Homology/Comparative Modelling), ενώ πολύ ελπιδοφόρα για το μέλλον είναι και η μέθοδος της αναγνώρισης αναδίπλωσης (Threading ή fold recognition). Το θεωρητικό σχεδιασμό των πρωτεϊνών ακολουθούν οι προσομοιώσεις, οι οποίες προσπαθούν να ανακαλύψουν την πιο σταθερή και συγχρόνως την πιο σωστή δομή. Στην παρούσα εργασία χρησιμοποιήθηκε το πρόγραμμα Modeller [1] για το συγκριτικό σχεδιασμό τμήματος του παράγοντα von Willebrand, το πρόγραμμα Phyre [2] που υποστηρίζει τη μέθοδο της αναγνώρισης αναδίπλωσης για το σχεδιασμό τμήματος του παραπάνω παράγοντα, ενώ τις προσομοιώσεις εκτέλεσε το πρόγραμμα NAMD [3]. Για την προετοιμασία των αρχείων PDB για τις προσομοιώσεις χρησιμοποιήθηκε το πρόγραμμα VMD [4].



Εικόνα 1: Η εκθετική αύξηση της GenBank.

1.1 Σκοπός

Στο πλαίσιο της ανά χείρας εργασίας έγινε μοντελοποίηση της δομής των συνδετικών τμημάτων [A1-A2] και [A2-A3] του παράγοντα von Willebrand μεταξύ των τριών επικρατειών, καθώς και της συνολικής στερεοδιάταξης του τμήματος [A1-A2-A3]. Να σημειωθεί ότι οι επικρατείες A1 και A3 έχουν κρυσταλλογραφηθεί, ενώ η A2 όχι.

1.2 Περίληψη

Για τις ανάγκες της εργασίας χρησιμοποιήθηκαν οι κρυσταλλογραφικά προσδιορισμένες δομές για τις A1 και A3 υπομονάδες και ένα θεωρητικά υπολογισμένο μοντέλο για την υπομονάδα A2, εφόσον η δομή της δεν έχει ακόμα προσδιοριστεί πειραματικά. Η διαμόρφωση των συνδετικών κομματιών μεταξύ των τριών παραπάνω υπομονάδων δεν έχει προσδιοριστεί πειραματικά. Στόχος αυτής της εργασίας ήταν η πρόβλεψη της δομής αυτών των τμημάτων. Μετά την επίτευξη του παραπάνω βήματος, έγινε η σύνδεση των επιμέρους τμημάτων και, τέλος, εξισορροπήθηκε το τμήμα A1-A2-A3. Σαν αποτέλεσμα προέκυψε ένα μοντέλο-χίμαιρα, το οποίο περιέχει τόσο πειραματικά όσο και θεωρητικά προσδιορισμένα τμήματα. Το συγκεκριμένο μοντέλο μπορεί να χρησιμοποιηθεί για την κατανόηση του τρόπου με τον οποίο αλληλεπιδρά ο vWF με παράγοντες όπως η ηπαρίνη, ο grIb, το κολλαγόνο και ιδίως με την πρωτεάση ADAMTS13.



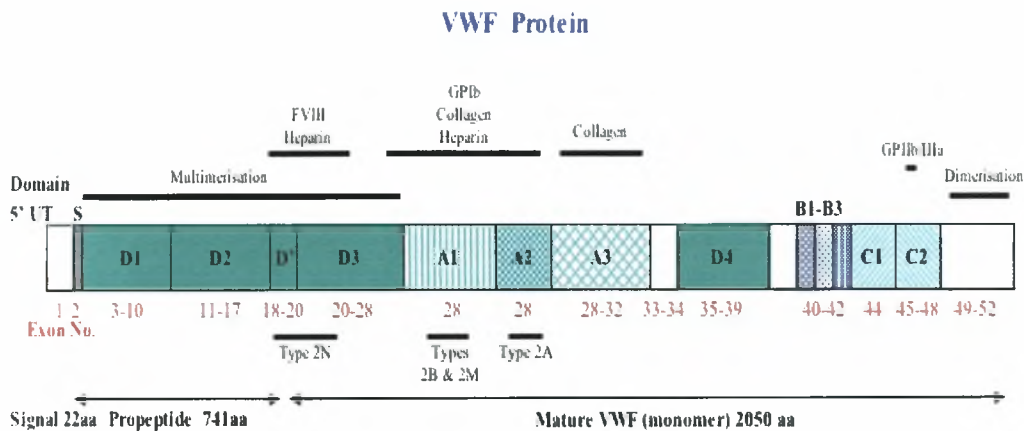
**Θεωρητικό
Μέρος**

2. Θεωρητικό μέρος

2.1 Ο παράγοντας Von Willebrand

Από όλους τους παράγοντες που συμμετέχουν στο μηχανισμό της πήξης, η παρούσα εργασία εστιάζεται στον παράγοντα von Willebrand, ο οποίος είναι μία μεγάλη πολυμερής και με πολλές επικράτειες γλυκοπρωτεΐνη που διαδραματίζει δύο, τουλάχιστον, σημαντικούς ρόλους στα αρχικά στάδια της αιμόστασης. Συγκεκριμένα, συμμετέχει στην πρόσδεση και προσκόλληση των κυκλοφορούντων αιμοπεταλίων στο κατεστραμμένο αγγειακό ενδοθήλιο, στο συνδετικό ιστό και σε άλλα προσκολλημένα αιμοπετάλια, καθώς και στη σταθεροποίηση του παράγοντα VIII στο πλάσμα δρώντας ως το μόριο που τον μεταφέρει. Εντοπίζεται στα αιμοπετάλια, τα ενδοθηλιακά κύτταρα και το πλάσμα και παράγεται, κυρίως, από το ενδοθήλιο (στα σωματίδια Weibel-Palade), από τα μεγακαρυοκύτταρα (στα α-κοκκία των αιμοπεταλίων) και από τον υπενδοθηλιακό συνδετικό ιστό [5, 6].

Το γονίδιο του vWF βρίσκεται στο χρωμόσωμα 12 (12p 13.2). Ο πρόδρομος vWF περιέχει 13 επικράτειες στη μονομερή του δομή, που είναι πολλαπλάσιες των τεσσάρων κύριων τύπων από το A έως το D με την εξής διεύθυνση: D1- D2- D'- D3 - A1 - A2 -A3 – D4- B1 - B2 -B3 - C1- C2 (εικόνα 2.1.1) [7].



Εικόνα 2.1.1: Η δομή ενός μονομερούς του vWF

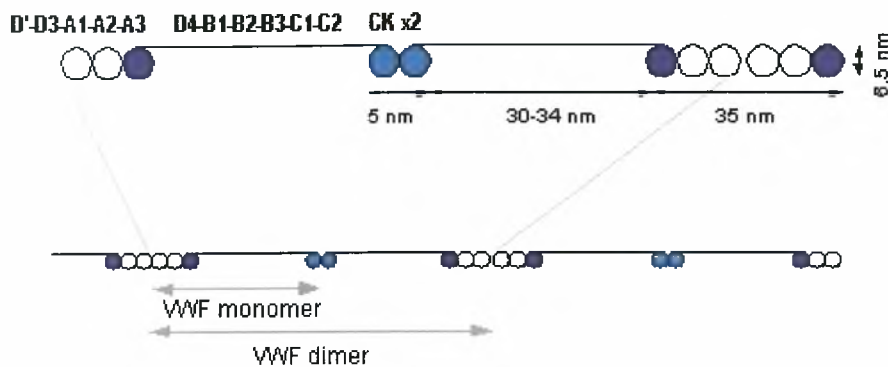
Συντίθεται ως μονομερή των 250kDa. Το βασικό μονομερές του είναι μία πρωτεΐνη μοριακού βάρους 220.000. Ο πολυμερής vWF περιέχει 2 έως 100 υπομονάδες με μοριακά βάρη που κυμαίνονται μέχρι $20 \cdot 10^6$. Κάθε μονομερές περιέχει έναν αριθμό ειδικών περιοχών εξειδικευμένης λειτουργίας, στοιχεία αυτών είναι:

- η D'/D3 επικράτεια, που προσδένεται στον παράγοντα VIII
- η A1 επικράτεια, που προσδένεται:
 - στον υποδοχέα gplb των αιμοπεταλίων
 - στην ηπαρίνη
 - πιθανώς στο κολλαγόνο
- η A3 επικράτεια, που προσδένεται στο κολλαγόνο

- η C1 επικράτεια, στην οποία η περιοχή RGD προσδένεται στην ιντεγρίνη των αιμοπεταλίων $\alpha_{IIb}\beta_3$ όταν είναι ενεργή
- η επικράτεια 'κόμπος κυστεΐνης' στο καρβοξυτελικό άκρο της πρωτεΐνης.

Μετά τη σύνθεσή του το πρόδρομο μόριο υφίσταται εκτεταμένη μετα-μεταφραστική τροποποίηση, κατά την οποία διμερίζεται μέσω πολλαπλών διαμοριακών δισουλφιδικών δεσμών μεταξύ των καρβοξυτελικών άκρων των πρωτεϊνών, αποκόπτονται τα προπολυπεπτίδια D1- D2 και, εν συνεχεία, μόλις μεταφερθεί στο σύμπλεγμα Golgi, πολυμερίζεται μέσω των δισουλφιδικών δεσμών που βρίσκονται ανάμεσα από τα διμερή μεταξύ των αμινοτελικών τους άκρων [8, 9, 10].

Τα τελικά πολυμερή κυμαίνονται σε μέγεθος μεταξύ 500 και 20.000 kDa και είναι τα πολυμερή με το υψηλότερο μοριακό βάρος και τα περισσότερα λειτουργικά ταυτόχρονα [11, 12]. Διαφέρουν μόνο ως προς τον αριθμό των μονομερών από τα οποία συνίστανται (εικόνα 2.1.2).



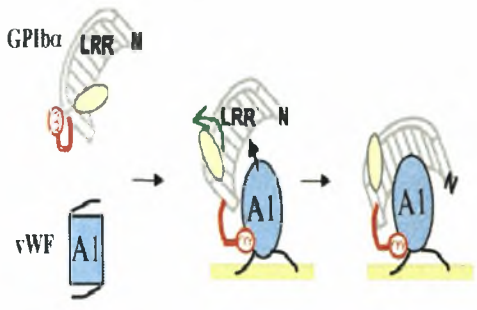
Εικόνα 2.1.2: Το μονομερές του vWF και τα πολυμερή.

Τα μεγαλύτερα πολυμερή ξεδιπλώνονται πιο άμεσα και πιο εύκολα και περιέχουν, επιπλέον, πιο πολλές θέσεις πρόσδεσης των αιμοπεταλίων [13]. Τα συγκεκριμένα πολυμερή διακρίνονται για τη μεγαλύτερη ικανότητά τους να σχηματίζουν θρόμβους. Το μέγεθος του παράγοντα von Willebrand και συνεπώς η ικανότητα προσκόλλησης των αιμοπεταλίων ρυθμίζονται ώστε να εμποδιστεί ο παρεκκλίνων σχηματισμός θρόμβων πλούσιων σε αιμοπετάλια. Πιο αναλυτικά, η ρύθμιση του μεγέθους του πολυμερούς είναι επιβεβλημένη για τη σωστή αιμόσταση, επειδή η κυκλοφορία των πολύ μεγάλων πολυμερών του καταλήγει σε θρομβωτική θρομβοπενική πορφύρα (προκαλείται από ανεπάρκεια της πρωτεάσης), μία ασθένεια που απειλεί τη ζωή και χαρακτηρίζεται από διάχυτη μικροαγγειακή θρόμβωση [14, 15, 16]. Αντίστροφα, μεταλλάξεις στην επικράτεια A2 οδηγούν σε μείωση στο μέγεθος των πολυμερών λόγω υπερβολικής πρωτεόλυσης σε μια αιμορραγική διαταραχή, γνωστή ως νόσος von Willebrand τύπου 2 [17].

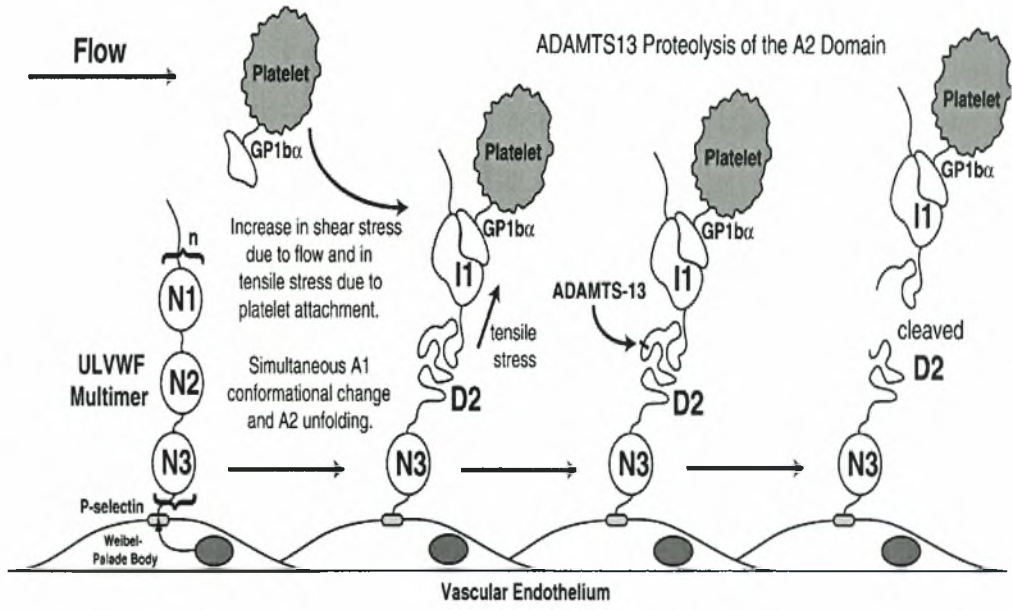
Όταν ένα αιμοφόρο αγγείο καταστραφεί, τα υποκείμενα κύτταρα που βρίσκονται στο υποενδοθήλιο εκτίθενται στο αίμα. Ο vWF μπορεί να προσδένεται στο κολλαγόνο του υποενδοθηλίου μέσω των A1 και A3 υπομονάδων του και, λόγω αυτής της πρόσδεσης και λόγω των δυνάμεων που ασκεί το ρέον αίμα, επάγεται μία δομική αλλαγή στον vWF που οδηγεί στο ξεδίπλωμά του. Αυτή η δομική αλλαγή εκθέτει με τη σειρά της τις θέσεις πρόσδεσης της γλυκοπρωτεΐνης (Gp)Iba, με συνέπεια την εμφάνιση των θέσεων πρόσδεσης των αιμοπεταλίων των επικρατειών A1 και C1 (εικόνα 2.1.3).

Εικόνα 2.1.3: Υποθετικό μοντέλο της ενεργοποίησης του (Gp)Iba και της πρόσδεσης στο vWF.

Κάτω από της επίδρασης ισχυρών διατμητικών τάσεων, ο (Gp)Iba και ο vWF ενεργοποιούνται καταλήγοντας σε μία υψηλής συνάφειας αλληλεπίδραση. Δομικές αλλαγές υφίστανται τόσο ο (Gp)Iba όσο και ο vWF. 1) Ο vWF του πλάσματος ξεδιπλώνεται και εκθέτεται η A1 επικράτεια η οποία και προσκολλάται στο υπενδοθήλιο (το οποίο συμβολίζεται με πράσινο) όπου και ενεργοποιείται. 2) Η κινητή ανιοντική περιοχή του (Gp)Iba αλληλεπιδρά αρχικά με την A1 επικράτεια, αλλά η υψηλής συγγένειας πρόσδεση εμφανίζεται μόνον εφόσον ενεργοποιηθεί και ο (Gp)Iba και όταν ο βρόγχος-R εκθέτεται αποκαλύπτοντας τη δεύτερη θέση πρόσδεσης της A1 στην κοίλη πλευρά του (Gp)Iba.



Μόλις προσδεθούν, τα αιμοπετάλια γίνονται ενεργά και εκθέτουν στην επιφάνειά τους το GpIb/IIIa, ο οποίος δημιουργεί μία πιο στενή και πιο σταθερή αλληλεπίδραση και με τον vWF και με το ινωδογόνο. Τα ενεργοποιημένα αιμοπετάλια παρέχουν την πλούσια σε φωσφατιδυλοσερίνη επιφάνειά τους, η οποία είναι καθοριστική για τη συγκρότηση των συμπλεγμάτων των προθρομβωτικών ενζύμων που οδηγούν στην παραγωγή της θρομβίνης (εικόνα 2.1.4).



Εικόνα 2.1.4: Απεικόνιση του μηχανισμού μέσω του οποίου ο vWF προσδένεται στα αιμοπετάλια κάτω από ισχυρές δυνάμεις συνάφειας.

Οι επικράτειες A1, A2 και A3 μεσολαβούν στις μακρομοριακές αλληλεπιδράσεις με τον παράγοντα. Όσον αφορά την πρώτη επικράτεια, της οποίας το αμινο-τελικό και το καρβοξυ-τελικό άκρο ενώνονται με ένα δισουλφιδικό δεσμό, συμμετέχει στην πρόσδεση του vWF στον γλυκοπρωτεϊνικό υποδοχέα των αιμοπεταλίων Ib (GpIb). Η συγκεκριμένη επικράτεια αντιστοιχεί στα κατάλοιπα 1261-1468. Αυτός είναι ο πρωταρχικός μηχανισμός μέσω του οποίου ο vWF προσδένεται στα αιμοπετάλια

κάτω από ισχυρές δυνάμεις συνάφειας. Ο vWF δεν προσδένεται αμέσως αυθόρμητα στον συγκεκριμένο υποδοχέα. Πιστεύεται ότι καθοριστικό ρόλο διαδραματίζουν οι δομικές αλλαγές που επέρχονται μετά την πρόσδεση του παράγοντα στο κολλαγόνο του υποενδοθηλίου. Πειράματα *in vitro* απέδειξαν ότι η πρόσδεση του vWF στα αιμοπετάλια μπορεί να επαχθεί από το αντιβιοτικό ριστοσετίνη και δύο τύποι παραγώγων της βοτροσετίνης από το δηλητήριο του φιδιού, παρόλο που οι δύο αυτές ενώσεις αντιδρούν με διαφορετικά κατάλοιπα της επικράτειας A1 [18]. Και ενώ η πρόσδεση με το κολλαγόνο επάγει την αλληλεπίδραση με τον υποδοχέα των αιμοπεταλίων, η σύνδεση με την ηπαρίνη, που είναι ο τρίτος προσδέτης της A1, την αναστέλλει. Η A2 επικράτεια, που αντιστοιχεί στα κατάλοιπα 1496-1669, περιέχει μία φυσιολογική θέση κοπής για τη μεταλλοπρωτεάση η οποία είναι συνδεδεμένη με την πρωτεάση υπό συνθήκες κανονικής κυκλοφορίας. Η A3 επικράτεια, η οποία οριοθετείται από τα 1685-1873 κατάλοιπα και τα άκρα της οποίας ενώνονται με ένα δισουλφιδικό δεσμό, περιέχει την κύρια θέση πρόσδεσης του ινώδους κολλαγόνου [19].

Το μέγεθος του vWF ρυθμίζεται από την πρόσφατα χαρακτηρισμένη πρωτεΐνη του πλάσματος ADAMTS13. Η ADAMTS13 (μία ιντεγρίνη και μεταλλοπρωτεάση με θρομβοσπονδίνη) είναι μία πολυεπικρατής μεταλλοπρωτεάση που περιορίζει την προσκόλληση των αιμοπεταλίων μέσω ενός μηχανισμού αρνητικής ανάδρασης στον οποίο ισχυρές δυνάμεις συνάφειας επάγουν την πρωτεόλυση του παράγοντα και εμποδίζουν την διάχυτη μικροαγγειακή θρόμβωση [20]. Ο συγκεκριμένος μηχανισμός ρύθμισης είναι αποτέλεσμα γεγονότων που έχουν μεταξύ τους σχέση αιτίου-αποτελέσματος. Πιο συγκεκριμένα, το ρέον αίμα ασκεί μία δύναμη στον όλο και αυξανόμενο θρόμβο πλούσιο σε αιμοπετάλια. Αυτήν η δύναμη ξεδιπλώνει τον vWF και εμφανίζεται τελικά η θέση κοπής για τη μεταλλοπρωτεάση. Η συγκεκριμένη πρωτεάση διασπά τον vWF μεταξύ της τυροσίνης 1605 και της μεθειονίνης 1606 μέσα στην A2 επικράτεια, με αποτέλεσμα να τον κόβει σε μικρότερα πεπτίδια και να απελευθερώνονται προσκολλημένα αιμοπετάλια [21]. Η πρωτεόλυση μπορεί να συμβεί μόνον εάν ο vWF έχει ξεδιπλωθεί, είτε εξαιτίας των ρεολογικών δυνάμεων είτε λόγω της παρουσίας αποδιατακτικών παραγόντων, και οι δύο λόγοι από τους οποίους έχουν ως αποτέλεσμα την έκθεση του εύθραυστου δεσμού της επικράτειας A2 [22]. Φυσιολογικά, αυτό συμβαίνει λόγω της έκκρισης του παράγοντα από τα ενδοθηλιακά κύτταρα περιορίζοντας την ικανότητα σχηματισμού θρόμβων των πολύ μεγάλων vWF που έχουν πρόσφατα εκκριθεί μετατρέποντάς τα σε μικρότερα πολυμερή. Η φύση της αναγνώρισης της ADAMTS13 είναι φτωχά κατανοημένη. Εν μέρει, αυτό οφείλεται στην πολύπλοκη αναδίπλωση και αλληλεξάρτηση των A1-A2-A3 επικρατειών. Κάτω από στατικές ή χαμηλές συνθήκες συνάφειας η πρωτεόλυση είναι αργή εξαιτίας της σφαιρικής διαμόρφωσης που αποκτά ο πολυμερής παράγοντας. Επιπλέον, η πολύπλοκη αναδίπλωση της A2 επικράτειας και των περιβαλλόντων επικρατειών κρύβει πιο πολύ τον εύθραυστο δεσμό.

Η πρόσδεση της μεταλλοπρωτεάσης στην A2 επικράτεια, καθώς και σε τμήματα αυτής, έχει αποδειχτεί πως είναι υψηλής συγγένειας [23]. Φυσιολογικά, φαίνεται ότι η πρωτεάση κυκλοφορεί σε σύμπλεγμα με τον πολυμερή παράγοντα. Όμως, η πρόσδεσή της δε χρειάζεται απαραίτητα να συνοδεύεται από πρωτεόλυση [24]. Η A1 επικράτεια έχει πρόσφατα αποδειχτεί ότι αναστέλλει την πρόσβαση της πρωτεάσης στη θέση κοπής στην A2 επικράτεια. Αυτήν η αναστολή εξαλείφθηκε στην παρουσία του Gr1ba, του κύριου προσδέτη της A1 επικράτειας, υποδεικνύοντας ότι κάτω από φυσιολογικές συνθήκες η πρόσδεση των αιμοπεταλίων μπορεί να αυξάνει την

πρωτεόλυση είτε αυξάνοντας την πρόσβαση στον εύθραυστο δεσμό είτε αποκαλύπτοντας τη θέση πρόσβασης της πρωτεάσης.

Το να επιτευχθεί η ισορροπημένη ρύθμιση της προσκόλλησης των αιμοπεταλίων είναι μία αξιοσημείωτη πρόκληση κάτω από τις συνθήκες που κυριαρχούν στο αίμα. Ο vWF και η ADAMTS13 εμφανίζονται σε χαμηλές συγκεντρώσεις των 10 μg/ml και του 1 μg/ml, αντίστοιχα, σε σύγκριση με τη συνολική συγκέντρωση πρωτεΐνης του πλάσματος των 80000 μg/ml. Ωστόσο, ο vWF είναι το μοναδικό γνωστό υπόστρωμα της ADAMTS13 στο πλάσμα, αν και η ADAMTS13 είναι ιδιοστατικά ενεργή [25]. Επιπλέον, ο vWF είναι ανθεκτικός στην ADAMTS13 μέχρι να υποβληθεί σε διατμητικές δυνάμεις ροής, να προσροφηθεί σε μία επιφάνεια ή να μεταχειρισθεί με χαστροπικά αντιγόνα, όπως η ουρία ή η υδροχλωρική γουανιδίνη. Διάφοροι μηχανισμοί συνεισφέρουν στην αξιοσημείωτη ειδικότητα της πρωτεάσης.

Η ποιοτική ή ποσοτική έλλειψη του παράγοντα οδηγεί στη νόσο Von Willebrand, η οποία είναι η πιο κοινή κληρονομική διαταραχή της πήξεως. Υπάρχουν τέσσερις τύποι της κληρονομούμενης νόσου (η νόσος μπορεί να είναι και επίκτητη):

- 1) Τύπου 1: το 60-80% των περιπτώσεων. Οφείλεται σε ποσοτική ανεπάρκεια (ετεροζυγωτία για το γονίδιο που προκαλεί την ασθένεια). Τα επίπεδα του vWF που ανιχνεύονται είναι μειωμένα.
- 2) Τύπου 2: το 20-30% των περιπτώσεων. Οφείλεται σε ποιοτική ανεπάρκεια. Τα επίπεδα του vWF είναι φυσιολογικά, αλλά τα πολυμερή είναι δομικά μη φυσιολογικά ή υποομάδες (subgroups) μικρών ή μεγάλων πολυμερών απουσιάζουν. Ο συγκεκριμένος τύπος της νόσου εμφανίζει τέσσερις υποτύπους: 2A, 2B, 2M και 2N (Normandy)
- 3) Τύπου 3: είναι η πιο σοβαρή μορφή της ασθένειας (ομοζυγωτία του επιβλαβούς γονιδίου).
- 4) Αιμοπεταλιακού τύπου: κληρονομείται με αυτοσωμικό επικρατή τρόπο και προκαλείται από μεταλλάξεις κέρδους λειτουργίας του υποδοχέα του vWF που βρίσκεται πάνω στα αιμοπετάλια και συγκεκριμένα της α-αλυσίδας του GpIb. Η απώλεια των μεγάλων πολυμερών εμφανίζεται και στον υπότυπο 2B στον οποίο όμως δεν ανιχνεύονται καθόλου μεταλλάξεις.

Οι τύποι 1 και 2 κληρονομούνται με αυτοσωμικό επικρατή τρόπο, ενώ ο τύπος 3 (και μερικές φορές και ο τύπος 2) με αυτοσωμικό υπολειπόμενο τρόπο.

2.2. Μέθοδοι προσδιορισμού της δομής των πρωτεϊνών

A) ΠΕΙΡΑΜΑΤΙΚΕΣ ΜΕΘΟΔΟΙ ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΠΡΩΤΕΪΝΙΚΗΣ ΔΟΜΗΣ

Η λειτουργικότητα μιας πρωτεΐνης και η αλληλεπίδρασή της με άλλα γειτονικά μόρια δεν είναι δυνατόν να κατανοηθούν από τη χημική της σύσταση μόνο, καθώς εξαρτώνται κυρίως από την τρισδιάστατη δομή της. Οι πρωτεΐνες είναι το μοναδικό βιομόριο που επιτελεί τόσες διαφορετικές λειτουργίες. Προκειμένου να γίνουν κατανοητές όλες οι λειτουργίες τους χρησιμοποιούνται ποικίλες βιοφυσικές μεθοδολογίες. Οι πιο σημαντικές σήμερα:

- ✓ μέθοδοι προσδιορισμού της στερεοδιάταξης βιομορίων και βιολογικών δομών είναι η περίθλαση ακτίνων-Χ και νετρονίων, η μέθοδος της τρισδιάστατης επανασύστασης, ο πυρηνικός μαγνητικός συντονισμός (NMR) και ESR, καθώς και
- ✓ μέθοδοι που δίνουν χρήσιμες, συμπληρωματικές συνήθως πληροφορίες, όπως είναι οι φασματοσκοπίες Laser-Raman και υπερύθρου, ο οπτικός στροφικός διασκεδασμός (O.R.D.) και ο κυκλικός διχρωσμός (C.D.).

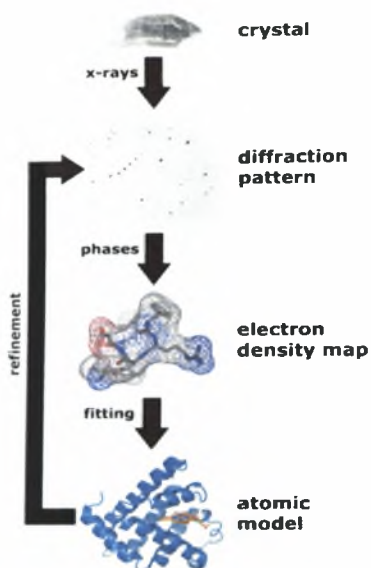
Η πιο αποδοτική βιοφυσική μέθοδος για την παρατήρηση ατομικών λεπτομερειών είναι η περίθλαση ακτίνων-Χ [26]. Παρακάτω αναλύονται συνοπτικά οι πιο σημαντικές μόνο μέθοδοι.

A1) Κρυσταλλογραφία

Για να διακρίνουμε λεπτομέρειες της μοριακής δομής πρέπει να χρησιμοποιήσουμε φως με μήκη κύματος της τάξης του 1Å, δηλαδή ακτίνες-Χ. Ένα χαρακτηριστικό των συγκεκριμένων ακτίνων είναι το γεγονός ότι σκεδάζονται ασθενώς μόνο από ένα μόριο και για αυτόν το λόγο χρησιμοποιούνται οι κρύσταλλοι, οι οποίοι αποτελούνται από ένα μεγάλο αριθμό μορίων διευθετημένων κανονικά. Ένας κρύσταλλος σκεδάζει τις ακτίνες-Χ ισχυρά, σαν ένα γιγαντιαίο μόριο. Οι ακτίνες περιθλώνται από τα ηλεκτρονιακά νέφη των μορίων, αφού τόσο τα μήκη κύματος των ακτίνων-Χ όσο και η διάμετρος του νέφους είναι της τάξης του Angstrom. Τα βιολογικά μακρομόρια, όπως οι πρωτεΐνες και τα νουκλεϊκά οξέα, έχουν πολύπλοκα σχήματα, ώστε να μη μπορούν να πακεταριστούν καλά, αλλά είναι δυνατόν να σχηματίσουν καλούς κρυστάλλους, με την προϋπόθεση ότι ένα υγρό συμπληρώνει τα κενά μεταξύ τους.

Κρυσταλλογραφία ονομάζεται η μελέτη του κρυσταλλικού πλέγματος, με τη βοήθεια της οποίας ανακαλύπτεται η γεωμετρική διάταξη των σωματιδίων που συγκροτούν τους κρυστάλλους. Το ζητούμενο της μεθόδου είναι να μπορέσει να παρατηρήσει κανείς κάποιο βιομόριο με διακριτική ικανότητα. Δέσμη ακτίνων Χ περνά μέσα από τα μικροσκοπικά σωματίδια του κρυστάλλου και δημιουργεί ένα σχήμα από κηλίδες πάνω σε φιλμ. Από το σχήμα που αποτυπώνεται στο φιλμ οι κρυσταλλογράφοι αναγνωρίζουν τη γεωμετρική δόμηση.

Οι εφαρμογές της μεθόδου είναι πολλές. Μία από αυτές είναι η εφαρμογή της μεθόδου στη διερεύνηση της δομής του DNA με καθοριστικό μάλιστα ρόλο στη Γενετική. Επιπλέον, εφαρμόζεται στην προσπάθεια προσδιορισμού της δομής αρκετά μεγάλων πρωτεϊνών, στην προσπάθεια να εξακριβωθεί πώς μία φαρμακευτική ουσία αλληλεπιδρά με έναν πρωτεϊνικό στόχο, καθώς και ποιες αλλαγές μπορούν να γίνουν ώστε να βελτιωθεί. Παρόλα αυτά, ο καθορισμός της δομής μεμβρανικών πρωτεϊνών παραμένει πρόκληση, καθώς απαιτούνται απορρυπαντικά ή άλλα μέσα ώστε να αυξηθεί η διαλυτότητά τους. Επιπλέον, η κρυσταλλογραφία δεν ανιχνεύει τυπικά τα άτομα υδρογόνου, καθώς είναι πολύ μικρά για να αλληλεπιδράσουν με την ακτινοβολία και επίσης, επειδή περιέχουν μόνο ένα ηλεκτρόνιο.



Εικόνα 2.2.1: Επισκόπηση των σταδίων της κρυσταλλογραφίας

Τα στάδια που ακολουθούνται για τον προσδιορισμό της κρυσταλλικής δομής μιας πρωτεΐνης είναι τα εξής (εικόνα 2.2.1):

1. Απομόνωση του βιομορίου
2. Κρυστάλλωση (και των ισόμορφων παραγώγων)
3. Περίθλαση ακτίνων-Χ (συλλογή δεδομένων)
4. Λύση του προβλήματος φάσης
5. Παραγωγή χαρτών ηλεκτρονικής πυκνότητας
6. Ερμηνεία χαρτών (χρήση μοριακών μοντέλων)
7. Βελτίωση δομής

Συνηθισμένα προβλήματα που απαντώνται, εάν και εφόσον έχουν επιτευχθεί τα δύο πρώτα στάδια, είναι:

- 1) Η αταξία των κρυστάλλων
- 2) Η δημιουργία όχι ακριβώς ισόμορφων παραγώγων
- 3) Οι δυσκολίες στην μέτρηση ασθενών ανακλάσεων
- 4) Ο μεγάλος αριθμός ανακλάσεων
- 5) Το γεγονός ότι παρατεταμένη έκθεση των κρυστάλλων στη δέσμη των ακτίνων-Χ οδηγεί στην καταστροφή των κρυστάλλων [26].

A2) Πυρηνικός Μαγνητικός Συντονισμός

Ένας από τους τομείς στους οποίους η τεχνική MNR βρίσκει εφαρμογή είναι στον καθορισμό της δομής των μακρομορίων. Συγκεκριμένα, παράγεται ένα σύνολο δομών. Οι δομές θα συγκλίνουν μόνο εάν οι πληροφορίες είναι επαρκείς ώστε να υπαγορεύσουν μία συγκεκριμένη αναδίπλωση. Πολύ σημαντική είναι η συνεισφορά των περιορισμών που εξάγονται από πειραματικά προσδιορισμένες δομές και αφορούν συνήθως μήκη δεσμών και γωνίες. Οι αλγόριθμοι μετατρέπουν τους περιορισμούς και τις γενικές ιδιότητες των πρωτεϊνών σε ενεργειακούς όρους και επιχειρείται η ελαχιστοποίηση της ενέργειας.

Εκτός από τον προσδιορισμό δομών, εξάγονται πληροφορίες και για τη δυναμική διαφόρων τμημάτων της πρωτεΐνης. Μπορούν, λοιπόν, να εξαχθούν πληροφορίες όσον αφορά τις ενδομοριακές κινήσεις σε ατομικό επίπεδο. Τα είδη των κινήσεων που μπορούν να εντοπιστούν είναι κινήσεις που συμβαίνουν σε μία χρονική κλίμακα που κυμαίνεται από 10 picoseconds έως 10 nanoseconds περίπου. Επιπλέον, πιο αργές κινήσεις, που λαμβάνουν χώρα σε μία χρονική κλίμακα που κυμαίνεται από 10 microseconds έως 100 milliseconds επίσης μπορούν να μελετηθούν. Τα αποτελέσματα συνήθως απεικονίζουν κινήσεις του σκελετού της πρωτεΐνης που είναι και το πιο άκαμπτο τμήμα της πρωτεΐνης [27].

A3) Κυκλικός διχρωσμός

Μία από τις οπτικές ιδιότητες ορισμένων υλικών είναι και ο διχρωσμός, δηλαδή η επιλεκτική απορρόφηση που εξαρτάται από την πόλωση. Ο κυκλικός διχρωσμός μετρά τη διαφορά μεταξύ της απορρόφησης του δεξιά και αριστερά κυκλικά πολωμένου φωτός. Μετά το πέρασμα του φωτός από το δείγμα, το γραμμικά πολωμένο φως μετατρέπεται τελικά σε ελλειπτικά πολωμένο. Η ελλειπτικότητα μπορεί να οριστεί σαν ο λόγος των μηκών των ημιαξόνων της έλλειψης e ή σαν θ , όπου $e = e\theta$. Η μεταβολή της γωνίας θ με το μήκος κύματος της ακτινοβολίας είναι γνωστή σε φάσμα κυκλικού διχρωσμού (C.D).

Στην πρωτοταγή δομή η ασυμμετρία που επιφέρει οπτική ενεργότητα είναι η ύπαρξη C^* , ενώ στη δευτεροταγή δομή των βιοπολυμερών οφείλεται στην ύπαρξη έλικας. Διακριτά φάσματα O.R.D και C.D χαρακτηρίζουν πολυπεπτίδια με α -έλικες, β -πτυχωτές επιφάνειες ή ακανόνιστη δομή. Επίσης, είναι δυνατόν από τα φάσματα O.R.D και C.D να προσδιοριστεί η ποσοστιαία αναλογία δευτεροταγών δομών μιας πρωτεΐνης. Πολύτιμες πληροφορίες εξάγονται, επίσης, από τα O.R.D και C.D μεμβρανών (κυρίως των πρωτεϊνικών τους συνιστωσών), νουκλεϊκών οξέων, ιών κ.λπ.

Οι εφαρμογές του κυκλικού διχρωσμού είναι οι ακόλουθες:

1. Προσδιορισμός δευτεροταγούς δομής πρωτεϊνών
2. Προσδιορισμός τριτοταγούς δομής πρωτεϊνών
Στο εγγύς υπεριώδες (250 – 300nm) τα χρωμοφόρα είναι τα αρωματικά αμινοξέα και οι δισουλφιδικοί δεσμοί.

Στη συγκεκριμένη περιοχή τα φάσματα θεωρούνται ενδεικτικά της τριτοταγούς δομής.

250-270 nm : φαινολαλανίνη
270-290 nm : τυροσίνη
280-300 nm : τρυπτοφάνη

A4) Φασματοσκοπία Laser-Raman

Η τεχνική αυτή στηρίζεται στο οπτικό φαινόμενο Raman. Με τη τεχνική αυτή είναι δυνατός ο προσδιορισμός των διαφόρων χαρακτηριστικών ομάδων των οργανικών χημικών ουσιών. Τα φάσματα Laser-Raman δίνουν πληροφορίες για μοριακές ταλαντώσεις. Οι αρχές της μεθόδου είναι απλές: όταν μονοχρωματικό φως προσπέσει σε ένα δείγμα ή περάσει από ένα δείγμα, εκτός από το φως που σκεδάζεται στην ίδια συχνότητα με την αρχική (σκέδαση Rayleigh), ένα μέρος του μπορεί να σκεδαστεί σε συχνότητες μετατοπισμένες (σκέδαση Raman, μη ελαστική σκέδαση). Αυτές οι μετατοπίσεις συχνότητας (θετικές ή αρνητικές) είναι ίσες με τις συχνότητες των μοριακών ταλαντώσεων του δείγματος.

Η εκπομπή μιας ταινίας Raman, προέρχεται από αλλαγή της πολωσιμότητας (επαγόμενης διπολικής ροπής) του μορίου, σαν αποτέλεσμα της προσπίπτουσας ακτινοβολίας στο δείγμα, πράγμα που αποτελεί και το γενικό κανόνα επιλογής. Φάσματα Raman παράγονται εύκολα και από δείγματα που περιέχουν νερό. Μειονέκτημα της μεθόδου είναι το υπόβαθρο φθορισμού που συνοδεύει την έντονη ακτινοβολία πολλών βιολογικών υλικών [26].

A5) Φασματοσκοπία υπερύθρου

Η φασματοσκοπία υπερύθρου στηρίζεται στην αλληλεπίδραση της ύλης με το υπέρυθρο φως. Ο βασικός κανόνας για να είναι μία ταλάντωση ενεργή στο υπέρυθρο είναι ότι η ταλάντωση θα πρέπει να έχει σαν αποτέλεσμα μια αλλαγή στη διπολική ροπή του μορίου. Συνήθως μετράται η απορρόφηση του φωτός από το δείγμα σε σχέση με συχνότητα, η οποία εκφράζεται από το νόμο των Beer-Lambert. Η συγκεκριμένη μέθοδος δίνει πληροφορίες για μοριακές ταλαντώσεις καταγράφοντας την αλληλεπίδραση των μορίων με την ηλεκτρομαγνητική ενέργεια. Τη φυσική βάση για την κατανόηση των μοριακών ταλαντώσεων συνιστά η καμπύλη δυναμικής ενέργειας Morse.

B) Θεωρητικές Μέθοδοι Πρόβλεψης Δομής

Οι θεωρητικές μέθοδοι στηρίζονται η καθεμία σε κάποιους κανόνες και παρατηρήσεις που προκύπτουν από την αξιοποίηση των ολοένα και αυξανόμενων πληροφοριών από πειραματικά προσδιορισμένες πρωτεϊνικές δομές. Παρόλα αυτά, είναι δυνατόν πρωτεΐνες όμοιες σε μεγάλο βαθμό να έχουν εντελώς διαφορετικούς

βιολογικούς ρόλους. Συνεπώς, οφείλει κανείς να είναι ιδιαίτερα προσεκτικός όταν προσπαθεί να επεκτείνει τα συμπεράσματά του και στη λειτουργία των πρωτεϊνών.

Οι μέθοδοι πρόβλεψης της δομής των πρωτεϊνών είναι τέσσερις:

- ✓ Συγκριτικός σχεδιασμός (comparative/homology modeling)
- ✓ Αναγνώριση αναδίπλωσης (fold recognition/threading)
- ✓ *Ab initio* προσεγγίσεις
- ✓ Πρόβλεψη δευτεροταγούς δομής

Στη συνέχεια, παρουσιάζονται οι βασικές αρχές των παραπάνω μεθόδων και τα αποτελέσματά τους.

B1) Συγκριτικός σχεδιασμός

Η συγκεκριμένη μέθοδος στηρίζεται σε προηγούμενη γνώση. Προσπαθεί να προσδιορίσει θεωρητικά την άγνωστη δομή μιας πρωτεΐνης, εφόσον βρεθεί ομολογή της με λυμένη δομή (εικόνα 2.2.2).

The slide titled "Overview of Homology Modeling" illustrates the workflow from a raw amino acid sequence to a 3D structure. It features a screenshot of a software interface showing two protein sequences: LYSOZY (1) KYVGRCEBLAAAMKKEHCLDNYEGT (25) and LYSOZY (26) GHVVCFAKTEERHNTQATNRRTEDS (50). Below the sequences are options for "Residue ID", "Mode" (Seq, Box, Tick Marks, Font Size, Help), and "Color by" (C. alpha, p. value, Contents). Three arrows point from the "Raw Sequence" towards the "Modeling" stage: "Comparative Modeling Knowledge-Based" (orange), "Physicochemical Simulations" (purple), and "Experiments" (blue). The "Experiments" box is linked to "X-ray, NMR, e-Diffraction". The "Modeling" box is dated "06/14/06". A 3D ribbon diagram of a protein structure is shown on the right. At the bottom, it credits "S. Ravichandran, Ph.D., ABCC, NCI-Frederick" and "3D-Structure Accelrys: Homology".

Εικόνα 2.2.2: Η εικόνα περιγράφει την πορεία που ακολουθείται κατά την πρόβλεψη της τριτοταγούς δομής μιας πρωτεΐνης με τη μέθοδο του συγκριτικού σχεδιασμού. Αξιοποιούνται τόσο πληροφορίες που έχουν εξαχθεί από πειράματα κατά τον προσδιορισμό της δομής των προτύπων όσο και πληροφορίες που προκύπτουν από τις προσομοιώσεις.

Ανάλυση των στοιχίσεων μεταξύ ακολουθιών με γνωστές δομές αναδεικνύει ότι ζευγάρια πρωτεϊνών με μεγαλύτερη ομοιότητα από 30% (για μήκη στοίχισης > 80 καταλοίπων) έχουν ομόλογες 3D-δομές, δηλαδή το βασικό δίπλωμά τους είναι όμοιο και μόνο περιοχές μη καθορισμένης δευτεροταγούς δομής ίσως διαφέρουν. Αν, λοιπόν, διαπιστωθεί ομολογία ανάμεσα σε μια πρωτεΐνη (U) άγνωστης 3D-δομής με μία γνωστής (T) τότε μπορεί να μοντελοποιηθεί μια δομή για την (U) με 'εκμαγείο' την (T). Η βασική προϋπόθεση είναι ότι η (U) και η (T) έχουν ταυτόσημες κύριες αλυσίδες. Η δυσκολία βρίσκεται στη σωστή προσαρμογή των πλευρικών αλυσίδων,

πρόβλημα του οποίου η επίλυση επιβάλλει τη χρήση ενεργειακών υπολογισμών, καθώς και δεδομένων για όλες τις διευθετήσεις πλευρικών αλυσίδων σε λυμένες πρωτεϊνικές δομές. Ακολουθίες με ποσοστά ομοιότητας 15-25% μπορεί να είναι συγγενικές, ενώ ακολουθίες με ποσοστό ομοιότητας μικρότερο του 15% πιθανόν δεν είναι συγγενικές. Τουλάχιστον αυτό ισχύει για τις περιοχές συγκεκριμένης δευτεροταγούς δομής. Η πρωτεΐνη γνωστής δομής καλείται πρότυπο (template), ενώ η πρωτεΐνη άγνωστης δομής ονομάζεται στόχος ή μοντέλο (model). Με τον αυτόν τον τρόπο έχουν διαλευκανθεί αρκετές στερεοδιατάξεις. Η μέθοδος βασίζεται σε δύο σημαντικές παρατηρήσεις:

- ✓ Η δομή μιας πρωτεΐνης είναι μοναδικά καθοριζόμενη από την αλληλουχία των αμινοξέων της. Επομένως, η γνώση της αλληλουχίας θα έπρεπε να αρκεί, τουλάχιστον θεωρητικά στον προσδιορισμό της δομής.

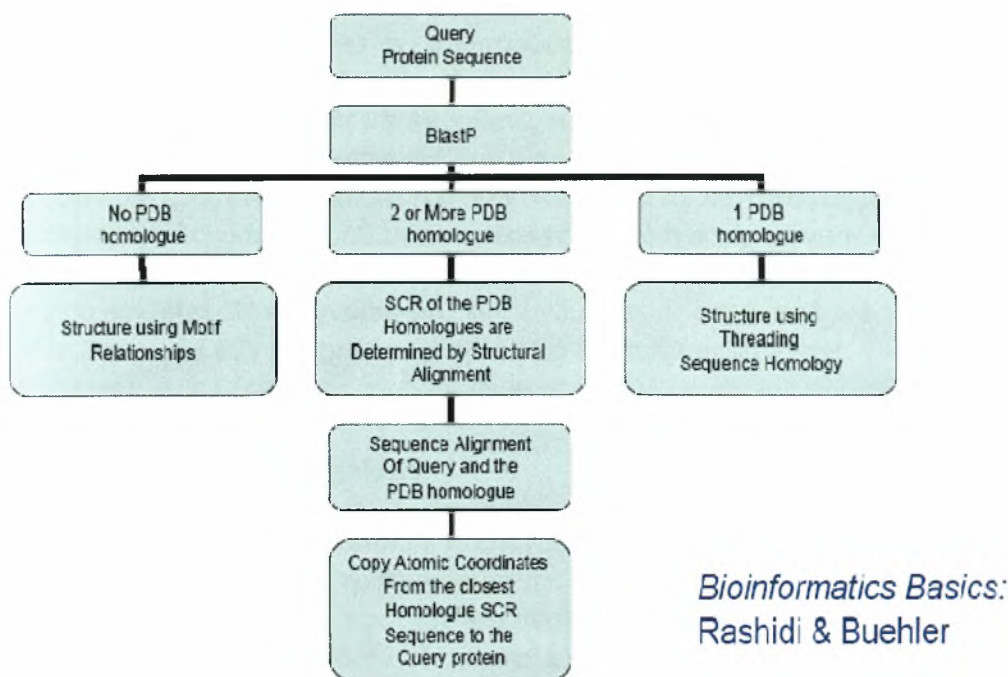
Στην πράξη, η μεγάλη πλειοψηφία ζευγών πρωτεϊνών με παρόμοιες δομές παρουσιάζουν σχετικά χαμηλή ομοιότητα στην πρωτοταγή τους δομή. Κατά τη διάρκεια της εξέλιξης, λοιπόν, η δομή κατέληξε να είναι πιο σταθερή και να αλλάζει πιο αργά από ό,τι η ακολουθία. Έτσι, παρόμοιες ακολουθίες υιοθετούν πρακτικά πανομοιότυπες δομές, ενώ και μακριά σχετιζόμενες ακολουθίες αποκτούν, συχνά, παρόμοιες δομές.

- ✓ Οι πυρηνικές περιοχές παρέχουν ένα άριστο πρότυπο για το σχεδιασμό της πρωτεΐνης-στόχου. Εάν οι πυρηνικές περιοχές παρουσιάζουν 50% ομοιότητα, τότε οι δύο πρωτεΐνες μπορούν σχεδόν πάντα να υπερτεθούν με $RMSD \leq 1 \text{ \AA}$.

Ο συγκριτικός σχεδιασμός είναι μία διαδικασία πολλών σταδίων, η οποία μπορεί να συμπεριληφθεί σε επτά στάδια (εικόνα 9). Σχεδόν σε όλα τα βήματα πρέπει να γίνουν επιλογές. Ο σχεδιαστής δεν μπορεί να είναι ποτέ σίγουρος ότι έκανε τις καλύτερες επιλογές και έτσι ένα μεγάλο μέρος του σχεδιασμού απαιτεί σοβαρή σκέψη λόγω των πολλών φαινομενικά παρόμοιων επιλογών. Πολλές ώρες μελέτης έχουν περάσει διδάσκοντας τον υπολογιστή πώς να λαμβάνει τέτοιου είδους αποφάσεις και τώρα οι σχεδιαστές μπορούν πλέον να

μοντελοποιούν σχεδόν αυτόματα. Έτσι, μπορούν να κατασκευαστούν μοντέλα για περίπου το 25% των αμινοξέων ενός γενώματος και, ως εκ τούτου, να συμπληρωθούν οι προσπάθειες των σχεδίων της δομικής γενωμικής. Το συγκεκριμένο ποσοστό ποικίλει μεταξύ των οργανισμών και αυξάνεται σταθερά λόγω της συνεχούς αύξησης της PDB. Για το υπόλοιπο 75% του γενώματος δεν υπάρχει διαθέσιμο κάποιο πρότυπο και η αμέσως επόμενη αξιόπιστη λύση είναι η μέθοδος αναγνώρισης αναδίπλωσης (fold recognition). Αν και η αυτοματοποιημένη κατασκευή μοντέλων παρέχει υψηλή παραγωγή, η αποτίμηση αυτών των μεθόδων κατά την διάρκεια του CASP δείχνει ότι η ανθρώπινη παρέμβαση είναι ακόμη απαραίτητη, ιδιαίτερα όταν η βαθμολόγηση της ομοπαράθεσης είναι στο οριακό σημείο.

Overview of Homology Modeling



06/14/06

S. Ravichandran, Ph.D., ABCC,
NCI-Frederick

Εικόνα 2.2.3: Απεικόνιση των βημάτων του συγκριτικού σχεδιασμού.

ΤΑ ΕΠΤΑ ΒΗΜΑΤΑ

Βήμα 1: Καθορισμός του προτύπου και αρχική ομοπαράθεση

Στην ζώνη του ασφαλούς συγκριτικού σχεδιασμού το ποσοστό ομοιότητας μεταξύ της ακολουθίας στόχου και του προτύπου είναι αρκετά υψηλό. Το κατάλληλο πρότυπο μπορεί να εντοπιστεί με ένα απλό πρόγραμμα εύρεσης ομοπαράθεσης όπως το BLAST [30] ή το FASTA [31].

Συγκεκριμένα, το πρόγραμμα συγκρίνει την αλληλουχία-στόχο με όλες τις αλληλουχίες γνωστών δομών στην PDB χρησιμοποιώντας κυρίως δύο πίνακες:

- 1) Έναν πίνακα υποκατάστασης καταλοίπων &
- 2) Έναν πίνακα ομοπαράθεσης

Το πρώτο μέλημα, λοιπόν, είναι να βρεθεί μία κατάλληλη πρωτεΐνη που να ικανοποιεί τις απαιτήσεις ομολογίας και να ακολουθήσει έλεγχος της αξιοπιστίας του μοντέλου. Τέλος, ομοπαράτιθεται η αλληλουχία στόχος με τις αλληλουχίες γνωστών δομών που είναι καταχωρημένες στην PDB και προκύπτει μία λίστα με τις καλύτερες βαθμολογήσεις.

Βήμα 2: Διόρθωση της ομοπαράθεσης

Έχοντας προσδιορίσει ένα ή περισσότερα πιθανά πρότυπα μοντελοποίησης επιστρατεύονται πλέον πιο εξεζητημένες μέθοδοι για να επιτευχθεί μια καλύτερη ομοπαράθεση. Μερικές φορές μπορεί να είναι δύσκολο να ομοπαρατεθούν δύο ακολουθίες σε μία περιοχή, όπου το ποσοστό ομοιότητας είναι πολύ χαμηλό. Τότε, μπορούν να χρησιμοποιηθούν άλλες αλληλουχίες από ομόλογες πρωτεΐνες. Πολλά προγράμματα είναι διαθέσιμα για την ομοπαράθεση ενός αριθμού από σχετιζόμενες μεταξύ τους ακολουθίες όπως για παράδειγμα το CLUSTALW [32] και η τελική ακολουθία περιλαμβάνει πολλές επιπρόσθετες πληροφορίες. Είναι σημαντικό το γεγονός ότι οι ομοπαράθεσεις απεικονίζουν καταρχήν τη φυλογενετική συσχέτιση μεταξύ δύο αμινοξικών ακολουθιών, η οποία δεν ταυτίζεται κατ' ανάγκη με τη δομική συσχέτιση. Για το λόγο αυτό, επιβάλλεται μερικές φορές η στοίχιση να τροποποιείται στα σημεία παρεμβολών / διαγραφών, ώστε να είναι δυνατή η δημιουργία ρεαλιστικού μοντέλου. Η πρόγνωση δευτεροταγούς δομής στην ακολουθία του μοντέλου, μπορεί να δώσει επιπλέον βοήθεια για την τελειοποίηση της ομοπαράθεσης, ταιριάζοντας τα στοιχεία δευτεροταγούς δομής της πρόγνωσης με αυτά του προτύπου.

Βήμα 3: Κατασκευή του σκελετού

Όταν η ομοπαράθεση είναι έτοιμη, η πραγματική κατασκευή του μοντέλου μπορεί να ξεκινήσει. Είναι λοιπόν εφικτή η παραγωγή του σκελετού, με ειδικό λογισμικό το οποίο δέχεται ως δεδομένα τη δομή-πρότυπο και τη στοίχιση των ακολουθιών. Η δημιουργία του σκελετού είναι μία συνήθης διαδικασία για το μεγαλύτερο μέρος του μοντέλου: απλώς αντιγράφει κανείς τις συντεταγμένες από τα κατάλοιπα εκείνου του προτύπου που συγκέντρωσε τη μεγαλύτερη βαθμολογία κατά την ομοπαράθεση. Αν δύο στοιχιζόμενα κατάλοιπα διαφέρουν μεταξύ τους, μόνο οι συντεταγμένες του σκελετού (N, Ca, C και O) μπορούν να αντιγραφούν. Αν είναι ίδια, τότε μπορεί να συμπεριληφθεί και η πλευρική αλυσίδα (τουλάχιστον η πιο άκαμπτη πλευρική αλυσίδα, αφού τα ισομερή εκ περιστροφής τείνουν να συντηρούνται). Στη συνέχεια, ακολουθεί οπτική παρατήρηση του μοντέλου και έλεγχος για πιθανές μη κανονικές αποστάσεις μεταξύ των ατόμων, καθώς και ενδεχόμενη βελτίωση της ομοπαράθεσης των ακολουθιών με σκοπό τη βελτίωση του μοντέλου.

Η οπτική παρατήρηση του μοντέλου πραγματοποιείται με προγράμματα μοριακών γραφικών. Το απλούστερο πρόγραμμα του είδους, ιδιαίτερα πρακτικό και εύχρηστο είναι το Rasmol/Chime.

Οι πειραματικά προσδιορισμένες πρωτεϊνικές δομές δεν είναι τέλειες, αλλά κάπως καλύτερες από τα μοντέλα. Υπάρχουν αναρίθμητες πηγές σφαλμάτων, από φτωχή ηλεκτρονιακή πυκνότητα στον χάρτη των ακτίνων-Χ έως απλά ανθρώπινα λάθη κατά την προετοιμασία του φακέλου PDB προκειμένου να γίνει η στοίχιση. Πολλές ώρες δουλειάς έχουν αναλωθεί για να γραφούν προγράμματα που να εντοπίζουν τέτοια λάθη. Είναι φανερό ότι ο άμεσος τρόπος για την κατασκευή ενός καλού μοντέλου είναι η επιλογή εκείνου του μοντέλου με τα λιγότερα λάθη. Στην περίπτωση που υπάρχουν δύο πρότυπα και καθένα έχει μία φτωχά καθορισμένη περιοχή, διαφορετική μεταξύ των προτύπων, τότε πρέπει να συνδυαστούν τα καλά τμήματα και των δύο σε ένα μοντέλο. Η παραπάνω προσέγγιση είναι γνωστή ως «σχεδιασμός με πολλαπλά πρότυπα». Η ίδια διαδικασία ακολουθείται και αν η στοίχιση μεταξύ στόχου και προτύπου δείχνει καλή βαθμολογία σε διαφορετικά όμως σημεία. Είναι όμως πολύ δύσκολο με την παραπάνω διαδικασία να επιτευχθούν

αποτελέσματα που είναι πραγματικά κοντά στην αληθινή δομή. Παρόλα αυτά είναι πιθανό και έχει αποδειχθεί από την ομάδα του AndrejSalis στο CASP4 [33].

Βήμα 4: σχεδιασμός βρόγχων

Στην πλειοψηφία των περιπτώσεων, η ομοπαράθεση περιέχει κενά είτε στην αλληλουχία στόχο (deletions) είτε στην αλληλουχία του προτύπου (insertions). Στην πρώτη περίπτωση, απλώς παραλείπονται τα κατάλοιπα του προτύπου και δημιουργείται ένα κενό στο μοντέλο που πρέπει να κλείσει. Στην δεύτερη περίπτωση, κάποιος διατηρεί τον συνεχή σκελετό του προτύπου, τον κόβει και προσθέτει τα κατάλοιπα που λείπουν. Και οι δύο περιπτώσεις συνεπάγονται δομικές αλλαγές του σκελετού. Επειδή, βέβαια, δομικές αλλαγές δεν μπορούν να συμβούν μέσα σε κανονικά στοιχεία δευτεροταγούς δομής, είναι ασφαλές να μετατοπίζονται όλες οι παρεμβολές ή τα κενά στην ομοπαράθεση έξω από τις έλικες και τα β-πτυχωτά φύλλα και να τοποθετούνται σε βρόγχους και στροφές. Τα άσχημα νέα είναι ότι αυτές οι αλλαγές στην δομή του βρόγχου είναι εξαιρετικά δύσκολες να προβλεφθούν, ένα μεγάλο άλυτο πρόβλημα στο συγκριτικό σχεδιασμό.

Υπάρχουν δύο βασικές προσεγγίσεις στον σχεδιασμό των βρόγχων:

- 1) Η μία βασίζεται στην γνώση. Συγκεκριμένα, κάποιος ψάχνει στην PDB για γνωστούς βρόγχους με άκρα που ταιριάζουν τα κατάλοιπα μεταξύ του βρόγχου που πρέπει να εισαχθούν και απλά αντιγράφει κανείς την δομή του βρόγχου. Όλα τα σημαντικά προγράμματα σχεδιασμού υποστηρίζουν αυτήν την μέθοδο (για παράδειγμα 3D-Jigsaw [34], Modeller [35], Swiss-Model [36], ή WHAT IF [37]).
- 2) Η άλλη βασίζεται στην ενέργεια. Όπως στην πραγματική *ab initio* πρόβλεψη της αναδίπλωσης, μία συνάρτηση ενέργειας χρησιμοποιείται για να εκτιμηθεί η ποιότητα του βρόγχου. Μετά, αυτή η συνάρτηση ελαχιστοποιείται χρησιμοποιώντας τις προσομοιώσεις Monte Carlo [38] ή τεχνικές μοριακής δυναμικής [39] για να επιτευχθεί η καλύτερη δομή του βρόγχου. Συχνά, η ενεργειακή συνάρτηση τροποποιείται για να διευκολυνθεί η μελέτη [40].

Τουλάχιστον στους μικρούς βρόγχους (μέχρι 5-8 κατάλοιπα), οι διάφορες μέθοδοι έχουν μία σημαντική ευκαιρία να προβλέψουν τη διεύθυνση του βρόγχου που υπερθέτεται σωστά στην αληθινή δομή. Όπως τονίσθηκε παραπάνω, οι επιφανειακοί βρόγχοι τείνουν να αλλάζουν την δομή τους εξαιτίας κρυσταλλικών επιδράσεων. Επομένως, εάν η πρόβλεψη γίνει για μία μεμονωμένη πρωτεΐνη και μετά βρεθεί να διαφέρει από την κρυσταλλική δομή, μπορεί ακόμα να είναι σωστή.

Βήμα 5: Σχεδιασμός πλευρικών αλυσίδων

Όταν συγκρίθηκαν οι διατάξεις των πλευρικών αλυσίδων των καταλοίπων που είναι συντηρημένα σε δομικά όμοιες πρωτεΐνες, βρέθηκε ότι τα συγκεκριμένα κατάλοιπα έχουν συχνά παρόμοιες γωνίες χ_1 (δηλαδή η γωνία μεταξύ του δεσμού $C_\alpha - C_\beta$). Είναι, συνεπώς, δυνατό να αντιγραφούν μόνο τα συντηρημένα κατάλοιπα εξ ολοκλήρου από το πρότυπο στο μοντέλο και να επιτύχουμε μια υψηλότερη ακρίβεια από ό,τι θα αποκτούσαμε εάν αντιγράφαμε μόνο το σκελετό και ξαναπροβλέπαμε τις πλευρικές αλυσίδες. Πρακτικά, αυτός ο πρακτικός κανόνας ισχύει μόνο σε υψηλά ποσοστά ομοιότητας των ακολουθιών, όταν τα συντηρημένα κατάλοιπα σχηματίζουν δίκτυα από τις αλληλεπιδράσεις. Όταν είναι απομονωμένα (<35% ομοιότητα), τα

ισομερή εκ περιστροφής των συντηρημένων καταλοίπων μπορεί να διαφέρουν μέχρι και στις 45% των περιπτώσεων [41].

Στην πράξη, όλες οι επιτυχημένες προσεγγίσεις στην τοποθέτηση πλευρικών αλυσίδων είναι τουλάχιστον μερικώς βασιζόμενες στην γνώση. Χρησιμοποιούνται βιβλιοθήκες συνηθισμένων ισομερών εκ περιστροφής που έχουν εξαχθεί από υψηλής ανάλυσης δομές ακτίνων-Χ. Τα διάφορα αυτά ισομερή δοκιμάζονται επιτυχώς και βαθμολογούνται με μία ποικιλία ενεργειακών συναρτήσεων. Διαισθητικά, κάποιος θα μπορούσε να περιμένει ότι η πρόβλεψη ισομερών εκ περιστροφής θα είναι υπολογιστικά απαιτητική εξαιτίας της συνδυαστικής έκρηξης-η επιλογή ενός συγκεκριμένου ισομερούς εκ περιστροφής αυτόματα επηρεάζει τα ισομερή εκ περιστροφής των γειτονικών καταλοίπων, που με τη σειρά τους επηρεάζουν τα γειτονικά τους και ούτω καθεξής. Με 100 κατάλοιπα και περίπου 5 ισομερή εκ περιστροφής ανά κατάλοιπο, κάποιος θα έπρεπε κιόλας να βαθμολογήσει 5^{100} διαφορετικούς συνδυασμούς. Πολλές ώρες έρευνας έχουν αναλωθεί, προκειμένου να αναπτυχθούν μέθοδοι που να κάνουν αυτό το τεράστιο πεδίο έρευνας πιο ελκυστικό [42]. Ο αριθμός των συνδυασμών είναι πράγματι πολύ μεγάλος, που ακόμα και η φύση δεν θα τους δοκίμαζε όλους κατά την διάρκεια του διπλώματος, πράγμα το οποίο υποδεικνύει πως υπάρχουν μηχανισμοί που συρρικνώνουν το πεδίο έρευνας.

Το κλειδί για να αντιμετωπίσει κανείς την συνδυαστική έκρηξη υπάρχει στον σκελετό της πρωτεΐνης. Συγκεκριμένες διαμορφώσεις του σκελετού απαιτούν συγκεκριμένα ισομερή εκ περιστροφής (επιτρέποντας, λόγω χάρη, έναν δεσμό υδρογόνου μεταξύ πλευρικής αλυσίδας και σκελετού) και έτσι μειώνεται δραματικά το πλήθος των υποψήφιων ισομερών εκ περιστροφής. Για μία δεδομένη διεύθυνση του σκελετού μπορεί να υπάρχει ένα μόνο ισχυρά τοποθετημένο ισομερές εκ περιστροφής που μπορεί να μοντελοποιηθεί ακριβώς, και ως εκ τούτου λειτουργεί ως άγκυρα για τις παρακείμενες και πιο ευλύγιστες πλευρικές αλυσίδες. Αυτές οι βιβλιοθήκες θέσεων-εξαρτώμενων ισομερών εκ περιστροφής χρησιμοποιούνται ευρέως σήμερα [43, 44].

Επιπλέον αποδείξεις, για το ό,τι το συνδυαστικό πρόβλημα της πρόβλεψης των ισομερών εκ περιστροφής είναι πολύ μικρότερο από ό,τι αρχικά πιστευόταν, βρέθηκαν πρόσφατα. Οι Xiang και Honig [45] μετακίνησαν πρώτα μόνο μία πλευρική αλυσίδα από γνωστές δομές και την ξαναπροέβλεψαν. Σε ένα δεύτερο βήμα, μετακίνησαν όλες τις πλευρικές αλυσίδες και τις ξαναπρόσθεσαν μετά χρησιμοποιώντας την ίδια απλή στρατηγική. Έκπληκτοι διαπίστωσαν ότι η ακρίβεια ήταν οριακά μόνο υψηλότερη στην πρώτη και πιο απλή περίπτωση.

Η ακρίβεια στην πρόβλεψη είναι συνήθως αρκετά υψηλή για κατάλοιπα που βρίσκονται σε υδρόφοβο πυρήνα, όπου περισσότερο από το 90% όλων των χ_1 γωνιών υπολογίζονται περίπου με $\pm 20^\circ$ απόκλιση από τις πειραματικές τιμές, αλλά πολύ χαμηλότερα για κατάλοιπα που βρίσκονται στην επιφάνεια, όπου το ποσοστό είναι συχνά ακόμα και κάτω από 50%. Δύο είναι οι λόγοι στους οποίους οφείλεται αυτό:

- 1) Πειραματικοί λόγοι: ευέλικτες πλευρικές αλυσίδες που βρίσκονται στην επιφάνεια τείνουν να υιοθετούν πολλαπλές διαμορφώσεις, οι οποίες επίσης επηρεάζονται από τις κρυσταλλικές επαφές. Συνεπώς, ακόμα και το πείραμα δεν μπορεί να παρέχει μία μόνο σωστή απάντηση.
- 2) Θεωρητικοί λόγοι: οι ενεργειακές συναρτήσεις που χρησιμοποιούνται για να βαθμολογήσουν τα ισομερή εκ περιστροφής μπορούν εύκολα να

διαχειριστούν το υδρόφοβο πακετάρισμα στον πυρήνα (κυρίως αλληλεπιδράσεις Van der Waals), αλλά δεν είναι αρκετά ακριβείς για να υπολογίσουν σωστά τις πολύπλοκες ηλεκτροστατικές αλληλεπιδράσεις στην επιφάνεια, συμπεριλαμβάνοντας δεσμούς υδρογόνου με μόρια νερού και να τα συσχετίσουν με τα αποτελέσματα της εντροπίας.

Βήμα 6: Ενεργειακή βελτιστοποίηση του μοντέλου με χρήση κατάλληλου λογισμικού

Η βελτιστοποίηση ενός μοντέλου γίνεται με μεθόδους ενεργειακών υπολογισμών. Στην απλούστερη περίπτωση, κάνουμε ενεργειακή ελαχιστοποίηση του μοντέλου, με αποτέλεσμα να το φέρουμε στο πλησιέστερο ενεργειακό ελάχιστο. Σε άλλες περιπτώσεις, χρησιμοποιούμε αλγορίθμους που υλοποιούν τεχνικές μοριακής δυναμικής (Molecular Dynamics), όπου κάνοντας προσομοίωση των αλληλεπιδράσεων μεταξύ όλων των ατόμων της δομής, έχουμε τη δυνατότητα να "διαταράξουμε" το μοντέλο, και να το οδηγήσουμε σε ένα περισσότερο απομακρυσμένο ενεργειακό ελάχιστο.

Η πιο άμεση προσέγγιση βελτιστοποίησης μοντέλου είναι απλά να τρέξουμε μία προσομοίωση μοριακής δυναμικής για το μοντέλο. Μία τέτοια προσομοίωση ακολουθεί τις κινήσεις της πρωτεΐνης σε femto δευτερόλεπτα (10^{-15} s) και μιμείται την πραγματική διαδικασία αναδίπλωσης. Κάποιος έτσι ελπίζει ότι το μοντέλο θα ολοκληρώσει την αναδίπλωσή του και θα προσαρμοστεί στην αληθινή δομή κατά τη διάρκεια της προσομοίωσης. Το πλεονέκτημα είναι ότι μία προσομοίωση μοριακής δυναμικής αναμφίβολα περιέχει αποτελέσματα εντροπίας, που διαφορετικά είναι δύσκολο να μεταχειριστούν. Το μειονέκτημα, όμως, είναι ότι οι συναρτήσεις δυναμικής δεν είναι αρκετά ακριβείς, ώστε η προσομοίωση να τρέξει επιτυχώς. Παρόλα αυτά, ένας από τους κύριους στόχους του Blue Gene, του ταχύτερου αυτή τη στιγμή υπολογιστή στον κόσμο, θα είναι να τρέξει ακριβώς αυτόν τον τύπο προσομοιώσεων μοριακής δυναμικής [46].

Βήμα 7: έλεγχος εγκυρότητας του μοντέλου

Το μοντέλο σίγουρα θα περιέχει κάποια λάθη. Το τελευταίο βήμα δεν θα μπορούσε παρά να περιλαμβάνει έλεγχο του βελτιστοποιημένου μοντέλου, αξιολόγηση αυτού ταιριάζοντάς το στο χώρο με κάποια πειραματικά προσδιορισμένη δομή (τη δομή-πρότυπο ή άλλη) και τελικά επικύρωση ή όχι αυτού.

Ο αριθμός των λαθών (για μία δεδομένη μέθοδο) εξαρτάται από δύο τιμές:

1) *Το ποσοστό ομοιότητας των αλληλουχιών του στόχου και του προτύπου.*

Εάν είναι υψηλότερη από 90%, η ακρίβεια του μοντέλου μπορεί να συγκριθεί με κρυσταλλογραφικά προσδιορισμένες δομές, εκτός από μερικές ανεξάρτητες πλευρικές αλυσίδες [47, 48]. Από 50% έως 90% ομοιότητα, το λάθος rms των συντεταγμένων του μοντέλου μπορεί να φθάνει τα 1.5 Å, με αξιοσημείωτα μεγαλύτερα τοπικά λάθη. Εάν το ποσοστό ομοιότητας πέφτει στο 25%, η ομοπαράθεση που προκύπτει να είναι η κύρια ραχοκοκαλιά για το homology modeling, συχνά οδηγεί σε πολύ μεγάλα λάθη.

2) *Ο αριθμός των λαθών στο πρότυπο.*

Τα λάθη σε ένα μοντέλο δεν αποτελούν τόσο μεγάλο πρόβλημα, όταν μπορούν να εντοπιστούν και να περιοριστούν. Δεν είναι πολύ σημαντικό, για παράδειγμα, εάν ένας βρόγχος, που βρίσκεται μακριά από το ενεργό κέντρο ενός ενζύμου, έχει

λανθασμένη τοποθέτηση. Ένα σημαντικό βήμα στη διαδικασία του συγκριτικού σχεδιασμού είναι, επομένως, η επαλήθευση ενός μοντέλου. Υπάρχουν δύο κυρίως διαφορετικοί τρόποι για να εκτιμήσουμε λάθη σε μία δομή:

1) *Υπολογίζοντας την ενέργεια του μοντέλου βάσει συνάρτησης δυναμικού:*

Αυτήν η μέθοδος εξετάζει εάν τα μήκη των δεσμών και οι γωνίες των δεσμών είναι μεταξύ κανονικών ορίων και εάν υπάρχουν αρκετές απότομες προεξοχές στο μοντέλο (αναφορικά με την υψηλή ενέργεια Van der Waals). Ουσιώδεις ερωτήσεις, όπως 'Έχει αναδιπλωθεί το μοντέλο σωστά;', δεν μπορούν ακόμα να απαντηθούν, επειδή τελείως κακοδιπλωμένα αλλά καλά ελαχιστοποιημένα μοντέλα συχνά φθάνουν σε ίδιες δυναμικές ενέργειες, όπως η δομή στόχος [49]. Αυτό το συναρπαστικό εύρημα οφείλεται στο γεγονός ότι οι συναρτήσεις μοριακής δυναμικής δεν περιέχουν ρητά εντροπικούς όρους (όπως είναι το φαινόμενο της υδροφοβικότητας), αλλά βασίζονται στην προσομοίωση για να τους δημιουργήσουν. Αν και αυτό το πρόβλημα μπορεί να απαντηθεί διευρύνοντας τη συνάρτηση δυναμικού και προσθέτοντας, για παράδειγμα, έναν όρο διαλυτοποίησης, το κύριο μειονέκτημα είναι το γεγονός ότι πάντα θα αποκτάται ένας μόνο αριθμός για ολόκληρη την πρωτεΐνη και δε θα είναι εύκολο να ανακαλυφθούν λάθη σε ξεχωριστά κατάλοιπα.

2) *Καθορισμός των δεικτών κανονικότητας που περιγράφουν πόσο καλά ένα δεδομένο χαρακτηριστικό του μοντέλου μοιάζει με το ίδιο χαρακτηριστικό στις αληθινές δομές.* Πολλές μορφές πρωτεϊνικών δομών ταιριάζουν πολύ καλά για ανάλυση κανονικότητας. Οι περισσότερες από αυτές είναι άμεσα ή έμμεσα βασιζόμενες στην ανάλυση διατομικών αποστάσεων και επαφών.

Συνοπτικά, μπορεί να ειπωθεί ότι ο συγκριτικός σχεδιασμός είναι δυστυχώς δύσκολος. Ίδανικά, ο συγκριτικός σχεδιασμός χρησιμοποιεί αναγνώριση αναδίπλωσης (threading), για να βελτιώσει την ομοπαράθεση και ab initio δίπλωμα για να προβλέψει τους βρόγχους, καθώς και προσομοιώσεις μοριακής δυναμικής για να οδηγηθεί στη σωστή δομή.

B2) Αναγνώριση αναδίπλωσης

Στην περίπτωση που δεν εμφανίζεται υψηλή ομολογία, τεχνικές 'Threading' (ή Remote Homology Modelling, όπως αλλιώς αναφέρονται) είναι δυνατόν να αποκαλύψουν 'απομακρυσμένες' ομολογίες, οι οποίες δεν ανιχνεύονται με τη χρήση συμβατικών μεθόδων στοίχισης ακολουθιών.

Με δεδομένη την ύπαρξη πολλών πρωτεϊνών στην PDB με ομολογίες μικρότερες από 25% αλλά 'ομόλογες' δομές, ορίστηκε η έννοια της 'απομακρυσμένης ομολογίας'. Η ιδέα είναι να ανιχνευθεί αυτή η απομακρυσμένη ομολογία και κατά τον ίδιο τρόπο με τη μέθοδο του Homology Modelling να μοντελοποιηθεί μια δομή για την άγνωστη πρωτεΐνη. Υπάρχουν διαφορετικά είδη τέτοιων μεθόδων, που ξεχωρίζουν μεταξύ τους κυρίως ως προς τον τρόπο προσδιορισμού της 'ομολογίας'.

Για να επιτύχει η διαδικασία του Threading είναι αναγκαίο να ξεπεραστούν 3 κύριες δυσκολίες:

1. Διαπίστωση της απομακρυσμένης ομολογίας
2. Ορθή στοίχιση των ακολουθιών
3. Χαμηλή πραγματική ομολογία, που δυσκολεύει την εύρεση της σωστής δομής, ιδίως στα τμήματα με μη κανονική δευτεροταγή δομή.

Απομακρυσμένες ομολογίες ανιχνεύονται πολύ συχνά, όμως καμία μέθοδος δεν κατορθώνει να εντοπίσει περισσότερες από τις μισές από αυτές τις περιπτώσεις σε μια προεπιλεγμένη ομάδα πρωτεϊνών που χρησιμοποιείται για αξιολόγηση των μεθόδων.

Η προσπάθεια της επαρκούς και ακριβούς πρόγνωσης της δομής, με βάση την ακολουθία, έχει τις ρίζες της στη γνώση ότι, ενώ η ακολουθία μπορεί να προσδιορίσει τη δομή, η ίδια στερεοδιάταξη μπορεί να προκύψει από πολλές διαφορετικές ακολουθίες. Οι ιδέες των Chothia και Lesk [47] ότι η δομή συντηρείται σε μεγαλύτερο βαθμό από ό,τι η ακολουθία και ότι υπάρχει ένας περιορισμένος αριθμός μοτίβων των α-ανθράκων του σκελετού, δείχνουν ότι οι ομοιότητες μεταξύ των πρωτεϊνών μπορεί να μην ανευρεθούν απαιτητικώς μόνο μέσω των παραδοσιακών μεθόδων που βασίζονται στην ακολουθία. Συμπεραίνοντας ότι η σχέση μεταξύ ακολουθίας και δομής βρίσκεται στη ρίζα "του προβλήματος του πρωτεϊνικού διπλώματος" και εφόσον η παρούσα γνώση δεν είναι αρκετή ώστε να χρησιμοποιηθεί για την απευθείας πρόβλεψη της πρωτεϊνικής δομής από τις βασικές αρχές, όπως είναι οι θεμελιώδεις εξισώσεις της φυσικής, πιο εύκολο είναι να αναγνωριστεί ποια από τις γνωστές αναδίπλωσεις των πρωτεϊνών είναι πιθανό να είναι παρόμοια με την (άγνωστη) αναδίπλωση μιας καινούριας πρωτεΐνης, όταν μόνο η αμινοξική της ακολουθία είναι γνωστή. Αυτό καλείται ποικιλοτρόπως αντίστροφο πρόβλημα αναδίπλωσης (βρίσκεται μια ακολουθία που ταιριάζει με μια δομή), threading (όταν μια ακολουθία εξετάζεται πάνω σε μία γνωστή δομή) και, τελικά, πρόβλημα αναγνώρισης αναδίπλωσης. Η λύση του παραπάνω προβλήματος είναι απαραίτητη, ώστε να επιλυθεί το γενικότερο πρόβλημα αναδίπλωσης [57]. Ο όρος threading πρώτα επινοήθηκε από τους Jones, Taylor και Thornton το 1992. Τότε ο συγκεκριμένος όρος αναφερόταν ειδικά στη χρησιμοποίηση μιας τρισδιάστατης ατομικής απεικόνισης του προτύπου στην αναγνώριση αναδίπλωσης. Σήμερα, οι όροι threading και αναγνώριση αναδίπλωσης χρησιμοποιούνται συχνά (αν και λανθασμένα κατά κάποιους) εναλλακτικά.

Η μέθοδος αναγνώρισης αναδίπλωσης έχει έναν ενδιάμεσο ρόλο στην πρόβλεψη της πρωτεϊνικής δομής μεταξύ του συγκριτικού σχεδιασμού και της *ab initio* πρόβλεψης, καθώς και κάποια κοινά σημεία με τις δύο συγκεκριμένες τεχνικές.. Αναλυτικότερα, όπως ο συγκριτικός σχεδιασμός, έτσι και η αναγνώριση αναδίπλωσης χρησιμοποιεί γνωστές δομές ως πρότυπα για ακολουθίες άγνωστης δομής. Από την άλλη μεριά, όπως και η *ab initio* πρόβλεψη, επιδιώκει να βελτιστοποιήσει μία συνάρτηση δυναμικού (ένα αντικείμενο ή μια συνάρτηση βαθμολόγησης), ώστε να εκτιμήσει την καταλληλότητα της ομοπαράθεσης της ακολουθίας με μία συγκεκριμένη χωρική διαμόρφωση [58].

Σκοπός της τεχνικής είναι η αναζήτηση δομών που έχουν παρόμοια αναδίπλωση, χωρίς αναγνωρίσιμη ομοιότητα μεταξύ των ακολουθιών τους. Αυτή η μέθοδος παίρνει μια ακολουθία επερώτησης, της οποίας η δομή δεν είναι γνωστή, και την ταιριάζει-την "περνά" (thread) στις συντεταγμένες πρωτεϊνών, των οποίων η δομή έχει λυθεί είτε με κρυσταλλογραφία ακτίνων-Χ είτε με φασματοσκοπία NMR. Η ακολουθία μετακινείται από θέση σε θέση δια μέσω της δομής, υποβαλλόμενη σε μερικούς προκαθορισμένους φυσικούς περιορισμούς: παραδείγματος χάριν, τα μήκη

των στοιχείων δευτεροταγούς δομής και οι περιοχές των βρόγχων να μπορούν να είναι είτε ορισμένοι είτε να ποικίλλουν μέσα σε ορισμένο εύρος. Για κάθε τοποθέτηση της ακολουθίας έναντι της δομής καθορίζονται μεταξύ μη τοπικών καταλοίπων, ανά ζεύγη και υδρόφοβες αλληλεπιδράσεις. Αυτοί οι θερμοδυναμικοί υπολογισμοί χρησιμοποιούνται για να καθορίσουν την ενεργειακά πιο ευνοϊκή και στερεοδιαταξικά σταθερή ομοπαράθεση της ακολουθίας, έναντι της γνωστής δομής που τοποθετεί τα αμινοξέα στα προτιμώμενα δομικά περιβάλλοντα και κοντά σε άλλους προτιμώμενους τύπους αμινοξέων. Μόλις επιλεγεί, η καλύτερη ομοπαράθεση, εγκαθιδρύει μια αντιστοιχία μεταξύ των αμινοξέων και των ειδικών θέσεων στο μοντέλο. Η επιτυχία της μεθόδου, λοιπόν, εξαρτάται και από την επιλογή του σωστού δομικού μοντέλου από μια βιβλιοθήκη δομών και της σωστής ομοπαράθεσης από όλες τις πιθανές [59].

Αν και είναι μια τεχνική εμφανώς πολύ ισχυρή, οι απαιτήσεις σε υπολογιστική ισχύ, καθώς και η αρτιότητα και η επάρκεια γνώσεων μπορεί να αποδειχτούν εμπόδια για τους περισσότερους βιολόγους. Σε μία προσπάθεια να μειωθεί ο βαθμός δυσκολίας, αναπτύχθηκαν προγράμματα φιλικά προς το χρήστη, για να δώσουν στο μέσο βιολόγο μια καλή πρώτη προσέγγιση για την πρωτεϊνική μοντελοποίηση.

Η συγκεκριμένη μέθοδος χρησιμοποιείται όταν:

- Η αμινοξική ακολουθία έχει ελάχιστη ή καθόλου ομοιότητα με κάθε ακολουθία αμινοξέων γνωστής δομής. Στην περίπτωση αυτή ο συγκριτικός σχεδιασμός είναι αδύνατος, και
- Όταν κάποιο μοντέλο από τη βιβλιοθήκη, στην οποία είναι κατατεθειμένες όλες οι γνωστές δομές, απεικονίζει τη σωστή αναδίπλωση για αυτήν την ακολουθία.

Το threading απαιτεί (1) μία αναπαράσταση της ακολουθίας, (2) μία βιβλιοθήκη δομών, (3) μια αντικειμενική συνάρτηση που βαθμολογεί τις ομοπαράθεσεις ακολουθίας-δομής, (4) μία μέθοδο ομοπαράθεσης της ακολουθίας πάνω στη δομή, και (5) μία μέθοδο επιλογής του μοντέλου από τη βιβλιοθήκη [58].

Η πρακτική σημασία της συγκεκριμένης μεθόδου έγκειται στο γεγονός ότι φαινομενικά άσχετες πρωτεΐνες υιοθετούν παρόμοιες δομές. Γίνεται, λοιπόν, εμφανές ότι ο αριθμός των διακεκριμένων προτύπων για το δίπλωμα είναι πεπερασμένος: μάλιστα είναι εξαιρετικά μικρότερος από αυτόν που θα ανέμενε κανείς, λόγω της αφθονίας των διαφορετικών συνδυασμών με τους οποίους μπορούν αμινοξικά κατάλοιπα να συνθέσουν πολυπεπτιδικές αλυσίδες. Συγκεκριμένα, περισσότερες από τις μισές από τις πρόσφατα επιλυμένες πρωτεΐνες, που υπολογιζόταν να είναι άσχετες με κάθε άλλη γνωστή πρωτεΐνη, αποδείχτηκε τελικά ότι είχαν μια γνωστή αναδίπλωση. Μερικά χρόνια πριν, τέτοιες περιπτώσεις θεωρούνταν φαινόμενο αξιοπερίεργο. Τώρα, πλέον, έχουν γίνει σχεδόν κανόνας. Γενικά, μόνο 100 αναδιπλώσεις αντιστοιχούν στις περισσότερες από τις μισές πρωτεϊνικές υπεροικογένειες σε μία από τις πιο δημοφιλείς κατηγοριοποιήσεις πρωτεϊνικών δομών, την SCOP [57].

Θεωρητικό υπόβαθρο της αναγνώρισης αναδίπλωσης

Δύο διαφορετικές οπτικές γωνίες κυριαρχούν στις θεωρητικές μελέτες των πρωτεϊνών και ως αποτέλεσμα υπάρχουν δύο διαφορετικές κατηγορίες αλγορίθμων. Πρόχειρα, αυτοί οι αλγόριθμοι θα μπορούσαν να ονομάζονται βιολογικοί και φυσικοί εξαιτίας του βαθμού των δύο αρχών που ενσωματώνουν οι ερευνητικές φιλοσοφίες.

Ένας βιολόγος προσπαθεί να εξηγήσει το φυσικό κόσμο σύμφωνα με την εξέλιξη. Οι βιολόγοι αναγνωρίζουν, περιγράφουν και ταξινομούν την ποικιλότητα της ζωής, ώστε να αποκαλύψουν τα σχέδια της εξέλιξης. Ακολουθώντας αυτό το πνεύμα, οι μοριακοί βιολόγοι μελετούν τις πρωτεΐνες όπως οι φυσιοδίφες του 18 και 19 αιώνα που μελετούσαν τα είδη των φυτών και των ζώων. Οι πρωτεΐνες περιγράφονται και ταξινομούνται σε οικογένειες και αναλύονται με τη βοήθεια μεταλλάξεων σε διάφορες θέ

σεις κατά μήκος της ακολουθίας. Η ομοιότητα στις ακολουθίες μεταξύ πρωτεϊνών που ανήκουν σε διαφορετικά είδη σχηματίζουν τη βάση της μοριακής φυλογένεσης, που τώρα συναγωνίζεται την παραδοσιακή μορφολογική φυλογενετική στα πλαίσια της ανάλυσης των σχέσεων μεταξύ των ειδών. Υπάρχει επίσης η τάση να μελετώνται οι σχέσεις μεταξύ ολόκληρων διαδικασιών, όπως τα ρυθμιστικά δίκτυα και τα μεταβολικά μονοπάτια ή μεταξύ υποπληθυσμών μέσα στα είδη. Αντίθετα, ένας φυσικός ψάχνει να εξηγήσει τη φύση στα πλαίσια των θεμελιωδών νόμων. Σε αυτό το πνεύμα, η πρωτεϊνική δομή θεωρείται ένα περίπλοκο σχήμα που καθορίζεται από ειδικές αλληλεπιδράσεις μεταξύ των αμινοξέων κατά μήκος της αλυσίδας. Οι διαφορετικές αλληλουχίες που υιοθετούν παρόμοιες δομές και που θεωρούνται πολλαπλές λύσεις στο ίδιο πρόβλημα ελαχιστοποίησης και αναγνώρισης διπλώματος μπορούν να σχηματιστούν ως μία αναγκαστική ελαχιστοποίηση, όπου μόνο μερικά σημεία στο εύρος των διαμορφώσεων λαμβάνονται υπόψη [57].

Μερικά προβλήματα κατανοούνται και μελετώνται μόνο προσεγγιστικά. Για στενά σχετιζόμενες πρωτεΐνες είναι πιθανό να κατασκευαστούν αξιόπιστα εξελικτικά δέντρα και να αναλυθούν οι σχέσεις μεταξύ των οργανισμών από τους οποίους προέρχονται χωρίς καμία αναφορά στο γεγονός ότι αυτές οι πρωτεΐνες αναδιπλώνονται στην ίδια δομή υιοθετώντας μία ελάχιστη ελεύθερη ενέργεια σαν λύση. Για τη μελέτη της δυναμικής περιορισμένου χρονικού εύρους των κινήσεων των πλευρικών αλυσίδων ή για την πρόβλεψη των αποτελεσμάτων μιας σημειακής μετάλλαξης μπορεί να μη νοιάζεται κανείς για το πώς το ένζυμο εξελίχτηκε. Όμως, πολλά προβλήματα απαιτούν επίγνωση και των δύο οπτικών γωνιών.

➤ **Οι πρωτεΐνες από τη σκοπιά ενός βιολόγου**

Από τη στιγμή που οι πρωτεϊνικές ακολουθίες έγιναν για πρώτη φορά διαθέσιμες, οι ερευνητές ανακάλυψαν ότι ακολουθίες ομόλογων πρωτεϊνών από σχετιζόμενους οργανισμούς είναι πολύ όμοιες. Η διαφορά αυξάνει με την αυξανόμενη εξελικτική απόσταση μεταξύ των ειδών, το οποίο συμφωνεί με τους γνωστούς μηχανισμούς αντιγραφής και επιδιόρθωσης του DNA. Αυτήν η παρατήρηση οδήγησε στην ανάπτυξη μεθόδων ομοπαράθεσης των ακολουθιών, οι οποίες προσπαθούν να βρουν την κατάλληλη ομοπαράθεση μεταξύ των ακολουθιών δύο ή περισσότερων συγκρινόμενων πρωτεϊνών. Αφού οι ομόλογες πρωτεΐνες εξ ορισμού προέκυψαν από έναν κοινό πρόγονο, είναι αναμενόμενο ότι μια σειρά μεταλλάξεων, αφαιρέσεων και ενθέσεων οδήγησαν από τον κοινό πρόγονο στις δύο σύγχρονες ακολουθίες. Μία λίστα τέτοιων στοιχειωδών βημάτων είναι ισοδύναμη με μία μοναδική ομοπαράθεση. Εάν υποθεθεί ομολογία μεταξύ δύο πρωτεϊνών, μια μηδενική υπόθεση είναι ότι οι δύο ακολουθίες δε προέρχονται από κοινό πρόγονο. Η σύγκριση των ακολουθιών τους θα έπρεπε να είναι παρόμοια με αυτήν της σύγκρισης δύο τυχαίων συμβολοσειρών. Συνεπώς, εάν βρεθεί μεγαλύτερη από την τυχαία ομοιότητα, γίνεται γενικά αποδεκτό ότι τέτοιες πρωτεΐνες είναι ομόλογες. Αυτή η σύνδεση είναι τόσο δυνατή που η ομοιότητα ακολουθίας χρησιμοποιείται οριστικά ως συνώνυμο της ομολογίας, παρά το γεγονός ότι η ομολογία είναι ένα πιο δυνατό σενάριο και τα δύο δεν είναι

ισοδύναμα. Την ίδια στιγμή, μπορεί δύο πρωτεΐνες να είναι ομόλογες παρά την έλλειψη μιας εύκολα αναγνωρίσιμης ομοιότητας στην ακολουθία.

Όταν η ομολογία διαπιστωθεί, μπορεί κανείς να προβλέψει τη δομή και τη λειτουργία νέων πρωτεϊνών υποθέτοντας ότι στην εξέλιξη τουλάχιστον κάποιες μορφές της λειτουργίας συντηρούνται. Ο κανόνας ότι η ισχυρή ομοιότητα στις ακολουθίες είναι ισοδύναμη με την ισχυρή δομική και λειτουργική ομοιότητα είναι ο μόνος αξιόπιστος κανόνας πρόβλεψης που ανακαλύφθηκε μέχρι τώρα. Ο συγκριτικός σχεδιασμός έχει μετατρέψει αυτόν το γενικό κανόνα σε μία δυναμική μέθοδο πρόβλεψης. Από το 1960, οι προσπάθειες των κρυσταλλογράφων και, πιο πρόσφατα, οι φασματοσκοπίες NMR απέδωσαν χιλιάδες πρωτεϊνικές δομές. Οι τεχνικές που χρησιμοποιούνται στην αναγνώριση αναδίπλωσης με τη βοήθεια σύγκρισης των ακολουθιών μπορούν επίσης να χρησιμοποιηθούν για το γενικότερο πρόβλημα αναδίπλωσης πιο απομακρυσμένων ομολόγων είτε έχουν είτε δεν έχουν τα μακρινά ομόλογα γνωστές δομές [57].

➤ **Οι πρωτεΐνες από τη σκοπιά ενός φυσικού**

Σύμφωνα με τη θερμοδυναμική θεωρία, η φυσική διαμόρφωση μιας πρωτεΐνης αντιστοιχεί σε ένα ολικό ελάχιστο της ελεύθερης ενέργειας από το σύστημα πρωτεΐνης/διαλύτη [60, 61]. Συνεπώς, έχοντας μία σωστή συνάρτηση δυναμικού, θα μπορούσε κανείς να χρησιμοποιήσει μεθόδους υπολογιστικής φυσικής για να ψάξει για τη φυσική διαμόρφωση σε ένα εύρος πιθανών. Παρά τις πολλές σημαντικές προόδους [62], αυτή η προσέγγιση δεν είναι ακόμη ικανή να προβλέψει αξιόπιστα μία προηγουμένως γνωστή δομή για την οποία μόνο η ακολουθία είναι γνωστή. Δύο κύρια προβλήματα που αντιμετωπίζει η *ab initio* πρόβλεψη της πρωτεϊνικής δομής είναι η έλλειψη κατάλληλων συναρτήσεων δυναμικού και το μεγάλο εύρος των πιθανών διευθετήσεων, ακόμη και των πιο μικρών πρωτεϊνών.

B3) ΑΠΛΟΠΟΙΗΜΕΝΕΣ ΠΡΟΓΝΩΣΕΙΣ ΣΕ ΜΙΑ Ή ΔΥΟ ΔΙΑΣΤΑΣΕΙΣ: Μια ικανοποιητική προσέγγιση στο πρόβλημα του 'διπλώματος'

Για τις περισσότερες πρωτεΐνες δεν είναι εφαρμόσιμες οι μέθοδοι του 'Homology Modelling' και του 'Threading' και η πρόγνωση που επιθυμούμε πρέπει να γίνει σε ένα απλοποιημένο επίπεδο: Μονοδιάστατες (1D) προγνώσεις (δευτεροταγούς δομής, διαμεμβρανικών τμημάτων, έκθεσης στο διαλύτη) είναι δυνατές, με αξιόλογα πολλές φορές αποτελέσματα και με τη χρήση 'εξελικτικής' πληροφορίας η απόδοσή τους βελτιώνεται αισθητά. Διδιάστατες (2D) προγνώσεις (αλληλεπίδρασης μεταξύ απομακρυσμένων στην πρωτοταγή δομή καταλοίπων, δισουλφιδικών δεσμών) έχουν, προς το παρόν, μικρή ακρίβεια.

Πρόβλεψη σε μία διάσταση: δευτεροταγής δομή και προσβασιμότητα

Δεν είναι, όμως, πάντα δυνατός ο συγκριτικός σχεδιασμός. Στις περιπτώσεις αυτές, το πρώτο βήμα είναι η πρόβλεψη της δευτεροταγούς δομής. Η **βασική ιδέα** πάνω στην οποία στηρίζεται είναι το γεγονός ότι τμήματα διαδοχικών καταλοίπων

προτιμούν να βρίσκονται σε συγκεκριμένες καταστάσεις δευτεροταγούς δομής [50,51]. Επομένως, το πρόβλημα της πρόβλεψης μετατράπηκε σε πρόβλημα αναγνώρισης προτύπων και αντιμετωπίζεται με τη χρήση κατάλληλων αλγορίθμων. Μία συνηθισμένη προσέγγιση είναι να επιτευχθεί πρόβλεψη για το κεντρικό κατάλοιπο μιας περιοχής με τυπικό εύρος 11-25 γειτονικών καταλοίπων εάν ανήκει μεταξύ των τριών παρακάτω καταστάσεων:

1. α-έλικα (α-helix => H)
2. β-πτυχωτή επιφάνεια (β-sheet => E)
3. τίποτε από τα παραπάνω ('coil' ή 'loop'=> L)

Σε αντίθεση με την πρόβλεψη της τριτοταγούς δομής, στην πρόβλεψη της δευτεροταγούς δομής έχουν γίνει σημαντικές πρόοδοι. Οι μέθοδοι αξιολογούνται ως προς την αξιοπιστία τους με βάση λυμένες δομές πρωτεϊνών, κατατεθειμένων στην PDB. Ο προσδιορισμός της δευτεροταγούς δομής μπορεί να πραγματοποιηθεί για γνωστές 3D-δομές λαμβάνοντας υπ' όψη τη θέση στο χώρο των ατόμων που την απαρτίζουν, καθώς και το 'πρότυπο σχηματισμού Η-δεσμών' του καρβονυλίου (C=O) της κύριας αλυσίδας και των αμιδικών ομάδων (NH).

Πρότυπα υπερ-δευτεροταγούς δομής είναι δυνατόν να καθορίσουν τη βιολογική λειτουργία

Συνδυασμοί στοιχείων δευτεροταγούς δομής, με συγκεκριμένη γεωμετρική διεύθυνση, εμφανίζονται συχνά σε δομές πρωτεϊνών. Τέτοιοι συνδυασμοί αποτελούν την υπερ-δευτεροταγή δομή και ορισμένοι από αυτούς συσχετίζονται άμεσα με συγκεκριμένες λειτουργίες. Για παράδειγμα, ο συνδυασμός helix-loop-helix αποτελεί ένα μοτίβο για πρόσδεση σε DNA (Gibson et al., 1993). Έτσι, επιτυχής πρόγνωση των στοιχείων δευτεροταγούς δομής δύναται να βοηθήσει στην αποκάλυψη τέτοιων σχηματισμών ανώτερης τάξης που παραπέμπει, συχνά, σε καθορισμένους βιολογικούς ρόλους.

Πρόγνωση Δομής Διαμεμβρανικών Πρωτεϊνών

Ακόμη και με το υπέρ-αισιόδοξο σενάριο ότι στο άμεσο μέλλον ο πειραματικός προσδιορισμός της δομής των πρωτεϊνών θα αποτελεί υπόθεση ρουτίνας, για την τάξη των διαμεμβρανικών πρωτεϊνών δεν αναμένεται ριζική διαφοροποίηση από τα όσα μέχρι σήμερα ισχύουν. Είναι πολύ δύσκολο, για τη συντριπτική τους πλειοψηφία, να δώσουν μονοκρυστάλλους κατάλληλους για κρυσταλλογραφία ακτίνων-Χ, ενώ δίνουν συνήθως φτωχές πληροφορίες με τη φασματοσκοπία NMR. Συνεπώς, για την κατηγορία αυτών των πρωτεϊνών οι προγνωστικές μέθοδοι είναι ακόμη πιο αναγκαίες από ό,τι για σφαιρικές-υδατοδιαλυτές πρωτεΐνες. Αν λάβουμε υπόψη τον περιορισμό στους βαθμούς ελευθερίας που επιβάλλουν τα διαμεμβρανικά τμήματα, γίνεται αυτονόητο πως στη συνέχεια η διαδικασία πρόγνωσης του διπλώματος είναι απλούστερη από ότι στις σφαιρικές πρωτεΐνες.

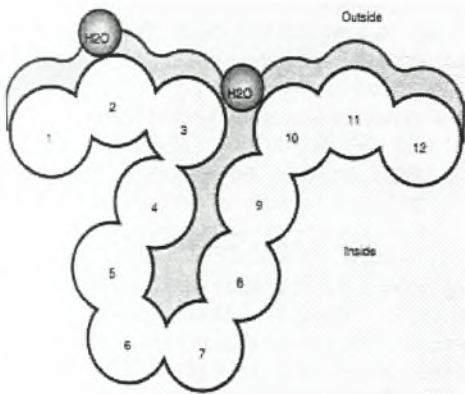
Οι πρώτες προσπάθειες στον τομέα αυτό, βασίζονταν σε υδροφοβική ανάλυση των ακολουθιών, για να εκμεταλλευτούν την προτίμηση που υπάρχει σε υδρόφοβα αμινοξικά κατάλοιπα μέσα στη λιπιδική διπλοστοιβάδα, εξαιτίας των ευνοϊκών

αλληλεπιδράσεών τους με τις αλειφατικές περιοχές των λιπιδίων. Στην περίπτωση αυτή, ενώ οι περισσότερες διαμεμβρανικές περιοχές εντοπίζονται, έχουμε και αρκετές λανθασμένες προβλέψεις, οι οποίες αντιστοιχούν σε μη διαμεμβρανικά τμήματα με υψηλή υδροφοβικότητα.

Νεώτερες μέθοδοι, που βασίζονται σε πολλαπλές στοιχίσεις ομόλογων πρωτεϊνών δίνουν αισθητά βελτιωμένα αποτελέσματα, αποτυγχάνουν όμως όταν δεν βρίσκονται ομολογίες στις βάσεις δεδομένων. Στατιστικοί αλγόριθμοι, μπορούν να βοηθήσουν στην απόρριψη των ψευδών σημάνσεων, εντοπίζοντας πρότυπα που χαρακτηρίζουν τις διαμεμβρανικές περιοχές. Συνδυαστικοί αλγόριθμοι πρόγνωσης, δίνουν και εδώ αποτελέσματα με μεγαλύτερα ποσοστά αξιοπιστίας. Εξαιρετικό ενδιαφέρον έχουν επίσης αλγόριθμοι που αποσκοπούν στην πρόγνωση του προσανατολισμού των μη διαμεμβρανικών τμημάτων, στον ενδοκυττάριο ή εξωκυττάριο χώρο (για πρωτεΐνες της κυτταρικής μεμβράνης) [57].

Μέθοδοι πρόβλεψης προσβασιμότητας διαλύτη

Εάν τμήματα δευτεροταγούς δομής μπορούσαν να προβλεφθούν ακριβώς, τότε η τρισδιάστατη δομή τους θα μπορούσε να προβλεφθεί απλά δοκιμάζοντας διαφορετικές διευθετήσεις αυτών των τμημάτων στο χώρο [52, 53, 54, 55]. Η παραπάνω άποψη αποτελεί σημείο μακράς διαφωνίας. Ένα κριτήριο για την αξιολόγηση κάθε διευθέτησης θα μπορούσε να είναι η εφαρμογή μεθόδων πρόβλεψης της προσβασιμότητας ενός διαλύτη (εικόνα 10). Ο κύριος σκοπός είναι η πρόβλεψη της έκτασης στην οποία ένα 'κρυμμένο' κατάλοιπο είναι προσβάσιμο στο διαλύτη. Μία από τις πιο γρήγορες μεθόδους είναι εκείνη που υπολογίζει την προσβασιμότητα εκτιμώντας τον όγκο του 'κρυμμένου' καταλοίπου που εκτίθεται στο διαλύτη [56]. Διαφορετικά κατάλοιπα έχουν διαφορετική περιοχή πιθανής προσβασιμότητας.



Εικόνα 2.2.4: Εκτίμηση προσβασιμότητας. Συνήθως υπολογίζεται κυλώντας ένα σφαιρικό μόριο νερού στην επιφάνεια της πρωτεΐνης και αθροίζοντας την περιοχή που μπορεί να προσεγγιστεί από το μόριο του νερού σε κάθε κατάλοιπο.

Η μέθοδος κατατάσσει τα αμινοξέα σε δύο κατηγορίες:

- 1) Σε αυτά που είναι 'κρυμμένα' (προσβασιμότητα διαλύτη < 16%)
- 2) Σε αυτά που εκτίθενται στο διαλύτη (προσβασιμότητα διαλύτη \geq 16%)

Η ακριβής επιλογή του ορίου δεν είναι καλά καθορισμένη. Ο παραπάνω διαχωρισμός βασίζεται στην υδροφοβικότητα κάθε καταλοίπου, δηλαδή πολύ υδρόφοβα αμινοξέα

προβλέπονται έτσι ώστε να είναι κρυμμένα και να μην εκτίθενται καθόλου στο διαλύτη.

Άλλες πιο εξελιγμένες μέθοδοι χρησιμοποιούν ως κριτήριο και τις θέσεις γλυκοζυλίωσης. Γενικότερα, όλες οι μέθοδοι βελτιώνονται με την εφαρμογή εξελικτικών πληροφοριών. Η πρόσβαση του διαλύτη σε κάθε σημείο της δομής είναι εξελικτικά συντηρημένη στις πρωτεϊνικές οικογένειες. Για αυτόν το λόγο αξιοποιούνται πληροφορίες πολλαπλής ομοπαράθεσης [57].

B4) Ab initio μέθοδοι

Οι *ab initio* προβλέψεις προσπαθούν να προβλέψουν τη φυσική διαμόρφωση μιας πρωτεΐνης μόνο από την πρωτοταγή της δομή. Ο Anfinsen έδειξε πριν από 40 χρόνια ότι όλη η πληροφορία που χρειάζεται μία πρωτεΐνη για να αναδιπλωθεί περιέχεται στην ακολουθία της [63,60]. Μεγάλος αριθμός ερευνητών, μεταξύ των οποίων και ο Anfinsen, υποστηρίζουν ότι οι φυσικές διαμορφώσεις των περισσότερων πρωτεϊνών είναι οι χαμηλότερες ενεργειακά διαμορφώσεις των αλληλουχιών τους [για κάποιες αξιοσημείωτες εξαιρέσεις παραπέμπουμε στην αναφορά 64]. Η επιτυχής δομική πρόβλεψη προϋποθέτει δύο πράγματα:

- Μία συνάρτηση ελεύθερης ενέργειας
- Μία μέθοδο η οποία να ψάχνει το πεδίο των διαμορφώσεων για χαμηλά ελάχιστα ενέργειας.

Πολλές μέθοδοι χρησιμοποιούν απλοποιημένα μοντέλα, κατάλληλα διαμορφωμένα δυναμικά και συνηθισμένες στρατηγικές έρευνας [65, 66]. Ενθαρρυντικά, αυτές οι απλοποιημένες μέθοδοι αρχίζουν να δείχνουν κάποια επιτυχία στην πρόβλεψη της πρωτεϊνικής δομής [67, 68].

Η πρόβλεψη γίνεται με βάση πρώτες αρχές χωρίς να χρησιμοποιείται καμία εμπειρική πληροφορία. Χρησιμοποιούνται όλα τα είδη ενεργειακών αλληλεπιδράσεων που μεσολαβούν στη διαδικασία αναδίπλωσης και τελικά βρίσκεται η δομή με τη χαμηλότερη ελεύθερη ενέργεια, όπως υποστηρίζει και η θερμοδυναμική προσέγγιση. Μπορεί η συγκεκριμένη τεχνική να είναι η πιο δύσκολη, είναι όμως η περιεκτικότερη σε πληροφορία μέθοδος [69, 70].

2.3 Εξυπηρέτης Phyre

Ο εξυπηρέτης Phyre χρησιμοποιείται για το σχεδιασμό τρισδιάστατων μοντέλων μιας ακολουθίας επερώτησης με βάση κοντινή ή απομακρυσμένη ομολογία με κάποια γνωστή πειραματικά προσδιορισμένη πρωτεϊνική δομή. Χρησιμοποιεί ένα συνδυασμό πληροφοριών πρωτοταγούς και δευτεροταγούς δομής, προκειμένου να ψάξει μία βάση δεδομένων γνωστών δομών χρησιμοποιώντας βελτιστοποιημένους αλγόριθμους συγκρίσεων προφίλ. Το πρόγραμμα μπορεί να δεχθεί είτε μία ακολουθία είτε μία ακολουθία μαζί με μία πολλαπλή ομοπαράθεση είτε ένα αρχείο δέσμης απλών ακολουθιών (batch file of single sequences).

Πολύ συνοπτικά αναφέρονται τα βήματα που ακολουθούνται:

1. Επιλογή των ομόλογων ακολουθιών
2. Πρόβλεψη της δευτεροταγούς δομής
3. Σάρωση της βιβλιοθήκης όπου βρίσκονται καταχωρημένες οι στερεοδιατάξεις
4. Σχεδιασμός του μοντέλου
5. Domaining (εξηγείται αναλυτικά παρακάτω)

Το πρώτο βήμα αποσκοπεί στην επιλογή των ομόλογων ακολουθιών, η οποία πραγματοποιείται μετά από μία έρευνα PSI-BLAST που υπολογίζει ένα στατιστικό προφίλ των πιθανοτήτων μεταλλαξιγένεσης για κάθε θέση. Το συγκεκριμένο προφίλ ονομάζεται PSSM και χρησιμοποιείται από το πρόγραμμα PSIPRED, το οποίο προβλέπει δευτεροταγείς δομές. Το συγκεκριμένο πρόγραμμα χρησιμοποιεί τεχνητά νευρωνικά δίκτυα για την ανάλυση τοπικών τμημάτων του προφίλ. Στη συνέχεια, η δευτεροταγής δομής που προέκυψε σε συνδυασμό με το στατιστικό προφίλ PSSM χρησιμοποιούνται στη σάρωση μιας βιβλιοθήκης όπου βρίσκονται καταχωρημένες οι δομικές στερεοδιατάξεις. Η συγκεκριμένη βιβλιοθήκη περιέχει προφίλ των ακολουθιών (PSSM) και προφίλ των δευτεροταγών δομών όλων των γνωστών δομών. Δύο είναι οι κύριες βιβλιοθήκες: η SCOP (SCOP95) και η PDB. Ο σκοπός αυτής της σάρωσης είναι η ιδανική ομοπαράθεση της ακολουθίας με κάθε δομή στη βιβλιοθήκη. Οι δέκα καλύτερες βαθμολογικά ομοπαράθεσεις χρησιμοποιούνται στο σχεδιασμό του μοντέλου. Επιπλέον, κενά ή παρεμβολές υφίστανται μία αναβαθμισμένη διαδικασία σχεδιασμού των βρόγχων. Αυτό συνεπάγεται έρευνα μίας βιβλιοθήκης μικρών τμημάτων (μικρότερα των 15 καταλοίπων) με διάφορα κριτήρια, όπως ομοιότητα ακολουθίας και ποιότητα του γεωμετρικού ταιριάσματος στο κενό. Πολλαπλοί βρόγχοι υψηλής βαθμολογίας δοκιμάζονται με τη βοήθεια ενός αλγορίθμου που καλείται Cyclic Coordinate Descent (CCD). Ο καλύτερος βρόγχος επιλέγεται να συμπεριληφθεί στη δομή. Τέλος, χρησιμοποιείται μία βιβλιοθήκη ισομερών εκ περιστροφής, η οποία έχει δημιουργηθεί από τους Dunbrack και Cohen, για τη σύνδεση των πλευρικών αλυσίδων στο σκελετό προκειμένου να αποφευχθούν συγκρούσεις και ταυτόχρονα αποκτάται μία ενεργειακά αποδεκτή και συμπαγής στερεοδιαμόρφωση. Επιπλέον, εάν το πρόγραμμα ανιχνεύσει μία αξιόλογη αντιστοιχία με ένα σχετικά μικρό τμήμα της ακολουθίας-στόχου, τα υπόλοιπα τμήματα της ακολουθίας 'στέλνονται' πίσω για περαιτέρω αποτίμηση, κάτι πολύ χρήσιμο σε περίπτωση που η ακολουθία περιέχει πολλές επικράτειες (το λεγόμενο domaining) [2]. Το αποτέλεσμα στέλνεται με μήνυμα ηλεκτρονικού ταχυδρομείου.

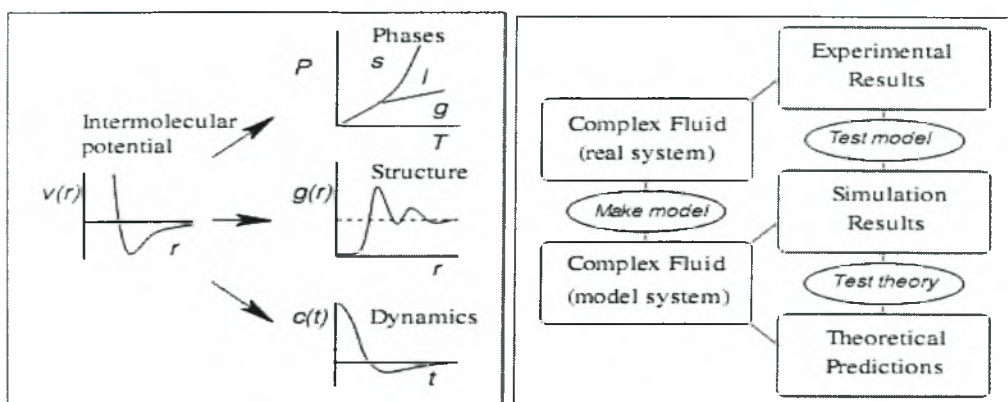
2.4 Υπολογιστικές προσομοιώσεις:

ένα σημαντικό όπλο στα χέρια των ερευνητών

Οι προσομοιώσεις βρίσκονται μεταξύ θεωρίας και πειράματος δρώντας ως γέφυρες μεταξύ του μικροσκοπικού και του μακροσκοπικού κόσμου. Διεξάγονται με

την ελπίδα να γίνουν κατανοητές οι ιδιότητες της συναρμολόγησης των μακρομορίων όσον αφορά τη δομή τους και τις μικροσκοπικές αλληλεπιδράσεις μεταξύ τους λειτουργώντας ως συμπλήρωμα στα συμβατικά πειράματα. Δίνουν, λοιπόν, τη δυνατότητα στους ερευνητές να μάθουν κάτι καινούριο, κάτι που δε μπορεί να βρεθεί με άλλον τρόπο (εικόνα 2.4.1).

Οι ηλεκτρονικοί υπολογιστές χρησιμοποιούνται προκειμένου να σχεδιαστεί ένα φυσικό σύστημα. Οι υπολογισμοί εκτελούνται από ένα μαθηματικό μοντέλο που εκτελεί και επιλύει ο υπολογιστής και τα αποτελέσματα ερμηνεύονται με βάση τις φυσικές ιδιότητες. Εφόσον οι προσομοιώσεις ασχολούνται με μοντέλα, μπορούν να ταξινομηθούν στις θεωρητικές μεθόδους. Από την άλλη μεριά, τα φυσικά μεγέθη μπορούν να υπολογιστούν στον υπολογιστή δικαιολογώντας συνεπώς και τον όρο υπολογιστικό πείραμα (computer experiment).



Εικόνα 2.4.1: Οι προσομοιώσεις ως γέφυρα μεταξύ α) μικροσκοπικού και μακροσκοπικού κόσμου και β) θεωρίας και πειράματος.

Το αποφασιστικής σημασίας πλεονέκτημά τους είναι η ικανότητά τους να αμβλύνουν το 'επιλύσιμο' από το 'μη επιλύσιμο', καθώς μπορούν να διεξάγονται προσομοιώσεις πειραμάτων που είναι δύσκολο ή αδύνατο να πραγματοποιηθούν στο εργαστήριο, πειραμάτων λόγω χάρη υπό ακραίες θερμοκρασίες ή πιέσεις. Δυστυχώς, εξαιτίας της αυξανόμενης πολυπλοκότητας, λιγότερο ακριβείς προσεγγίσεις είναι αναπόφευκτες.

Από μια πιο πρακτική άποψη, οι υπολογιστικές προσομοιώσεις μπορεί να χρησιμοποιηθούν για την ανακάλυψη και το σχεδιασμό νέων μορίων. Εξετάζοντας τις ιδιότητες ενός μορίου το οποίο σχεδιάζεται μέσω υπολογιστή είναι κάτι πολύ γρηγορότερο και λιγότερο δαπανηρό από τη σύνθεσή του σε ένα πραγματικό πείραμα. Ο σχεδιασμός φαρμάκων μέσω υπολογιστή είναι ευρέως χρησιμοποιούμενος στις φαρμακοβιομηχανίες [72].

Εξαιτίας της πολυπλοκότητας των βιολογικών συστημάτων, υπολογιστικές (με τη βοήθεια Η/Υ) μέθοδοι έχουν γίνει πλέον απαραίτητες στις επιστήμες υγείας. Με γρηγορότερους και πιο αποτελεσματικούς υπολογιστές όλο και πιο πολύπλοκα συστήματα θα είναι δυνατό να μελετώνται εφαρμόζοντας τον υπολογιστικό σχεδιασμό (computer modeling) ή υπολογιστικές προσομοιώσεις.

2.4.1 ΟΙ ΣΤΟΧΟΙ ΤΩΝ ΠΡΟΣΟΜΟΙΩΣΕΩΝ ΜΟΡΙΑΚΗΣ ΔΥΝΑΜΙΚΗΣ

Οι προσομοιώσεις μοριακής δυναμικής μπορούν να βοηθήσουν στη μελέτη τόσο στατικών ιδιοτήτων ισορροπίας όσο και δυναμικών. Ειδικότερα, οι μακροσκοπικές φυσικές ιδιότητες μπορούν να διακριθούν σε:

1. Στατικές ιδιότητες ισορροπίας, όπως είναι η σταθερά πρόσδεσης ενός αναστολέα σε ένα ένζυμο, το μέσο δυναμικό ενέργειας ενός συστήματος ή η συνάρτηση ακτινωτής κατανομής σε ένα υγρό, και
2. Δυναμικές ή ιδιότητες μη ισορροπίας, όπως το ιξώδες ενός υγρού, οι διαδικασίες διάχυσης στις μεμβράνες, η δυναμική των αλλαγών φάσεων, η κινητική των αντιδράσεων ή η δυναμική των ατελειών στους κρυστάλλους.

Η επιλογή της κατάλληλης τεχνικής εξαρτάται από το ερώτημα που κάθε φορά τίθεται προς απάντηση, καθώς και από την επιτευξιμότητα της μεθόδου να αποδώσει αξιόπιστα αποτελέσματα. Όσο μεγαλύτερη η πολυπλοκότητα ενός συστήματος και όσο μεγαλύτερη η χρονική διάρκεια των διαδικασιών που είναι στο επίκεντρο του ενδιαφέροντος κάθε φορά, όσο και πιο μεγαλύτερες προσεγγίσεις απαιτούνται [73].

Οι δύο κύριες κατηγορίες τεχνικών προσομοίωσης που είναι αντιπροσωπευτικές για την κατάσταση ισορροπίας είναι οι ακόλουθες:

- Μοριακή δυναμική
- Monte Carlo

Υπάρχει, βέβαια, και μία ποικιλία υβριδικών τεχνικών που συνδυάζουν χαρακτηριστικά και των δύο αναφερθέντων μεθόδων. Το εμφανές πλεονέκτημα της πρώτης έναντι της δεύτερης κατηγορίας τεχνικών είναι το γεγονός ότι η μοριακή δυναμική μπορεί να αποδώσει δυναμικές ιδιότητες του συστήματος, όπως συντελεστές μεταφοράς, χρονο-εξαρτώμενες αποκρίσεις σε διαταραχές, ρεολογικές ιδιότητες και φάσματα ταλαντώσεων.

Για την αναπαράσταση συστημάτων που δε βρίσκονται σε κατάσταση ισορροπίας και για την ανάλυση γεγονότων που συνεχώς εξελίσσονται, μόνο η πρώτη μέθοδος είναι κατάλληλη. Μπορεί οι προσομοιώσεις Monte Carlo να είναι πιο απλές (δεν απαιτούν τον υπολογισμό των δυνάμεων), δεν αποδίδουν όμως σημαντικά καλύτερες στατιστικές από τις προσομοιώσεις μοριακής δυναμικής σε δεδομένο υπολογιστικό χρόνο. Συνεπώς, η μοριακή δυναμική έχει καθιερωθεί ως η επικρατούσα τεχνική.

Οι μέθοδοι μοριακής δυναμικής μπορούν να χωριστούν σε δύο κύριες οικογένειες. Στην *κλασική μηχανική* προσέγγιση το μόρια μεταχειρίζονται σαν να είναι κλασικά αντικείμενα μοιάζοντας πολύ με το μοντέλο 'ball and stick'. Τα άτομα αντιστοιχούν στις μπάλες και οι ελαστικές ράβδοι στους δεσμούς. Οι νόμοι της κλασικής μηχανικής καθορίζουν τη δυναμική του συστήματος. Στις *κβαντικές* ('quantum' or 'first-principles') προσεγγίσεις, οι οποίες ήρθαν στο προσκήνιο τη δεκαετία του 1980 με τη δημιουργική δουλειά των Car και Parinello, λαμβάνουν ιδιαίτερως υπόψη την κβαντική φύση του χημικού δεσμού [72].

2.4.2 Ο ΑΛΓΟΡΙΘΜΟΣ ΤΩΝ ΠΡΟΣΟΜΟΙΩΣΕΩΝ ΜD

Στις προσομοιώσεις μοριακής δυναμικής η χρονική εξέλιξη ενός συνόλου αλληλεπιδρώντων μορίων δίνεται από την επίλυση των εξισώσεων κίνησης του Νεύτωνα για ένα σύστημα N αλληλεπιδρώντων ατόμων:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i, \quad i = 1 \dots N.$$

όπου F: η δύναμη

m: η μάζα

\mathbf{r}_i : η θέση του κάθε ατόμου

N: το πλήθος των αλληλεπιδρώντων ατόμων του συστήματος

Οι δυνάμεις είναι οι αρνητικές παράγωγοι μιας συνάρτησης δυναμικού V ($\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$):

$$\mathbf{F}_i = - \frac{\partial V}{\partial \mathbf{r}_i}$$

Όπου V: η δυναμική ενέργεια

Το πρόβλημα εύρεσης μιας ρεαλιστικής συνάρτησης δυναμικού που θα μιμείται επαρκώς τις διαδικασίες στο φυσικό κόσμο οδηγεί σε υπολογιστικές απλοποιήσεις. Η κλασική μοριακή δυναμική βασίζεται σε εμπειρικά δυναμικά μίας συγκεκριμένης συναρτησιακής σχέσης που απεικονίζουν τη φυσική και τη χημεία των συστημάτων. Μία τυπική συνάρτηση δυναμικού που χρησιμοποιείται στις προσομοιώσεις των βιολογικών συστημάτων είναι το εξής:

$$\begin{aligned} U(\mathbf{r}_1, \dots, \mathbf{r}_N) = & \sum_{\text{bonds}} \frac{a_i}{2} (l_i - l_{i0})^2 + \sum_{\text{angles}} \frac{b_i}{2} (\theta_i - \theta_{i0})^2 \\ & + \sum_{\text{torsions}} \frac{c_i}{2} [1 + \cos(n\omega_i - \gamma_i)] \\ & + \sum_{\text{atom pairs}} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\ & + \sum_{\text{atom pairs}} k \frac{q_i q_j}{r_{ij}} \end{aligned}$$

U: η δυναμική ενέργεια σαν συνάρτηση της θέσης \mathbf{r}_i των ατόμων

l_i : το μήκος του δεσμού

θ_i : η γωνία που σχηματίζουν 3 διαδοχικά άτομα συνδεδεμένα με ομοιοπολικούς δεσμούς

ε_{ij} : η διηλεκτρική σταθερά

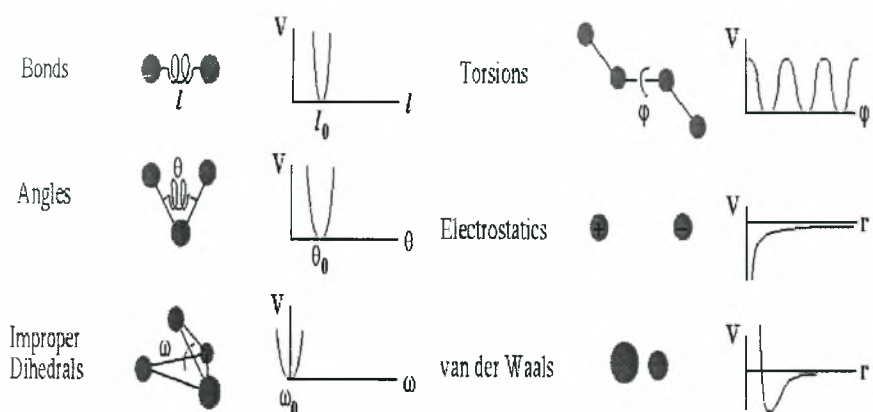
ω_i : η γωνία στροφής γύρω από δεσμό

n: ακέραιος αριθμός

Τα a_i , l_{i0} , b_i , θ_{i0} , c_i , ϵ_{ij} , σ_{ij} , k , q_i , q αποτελούν σταθερές παραμέτρους, οι τιμές των οποίων είναι καθορισμένες για κάθε τύπο ατόμων.

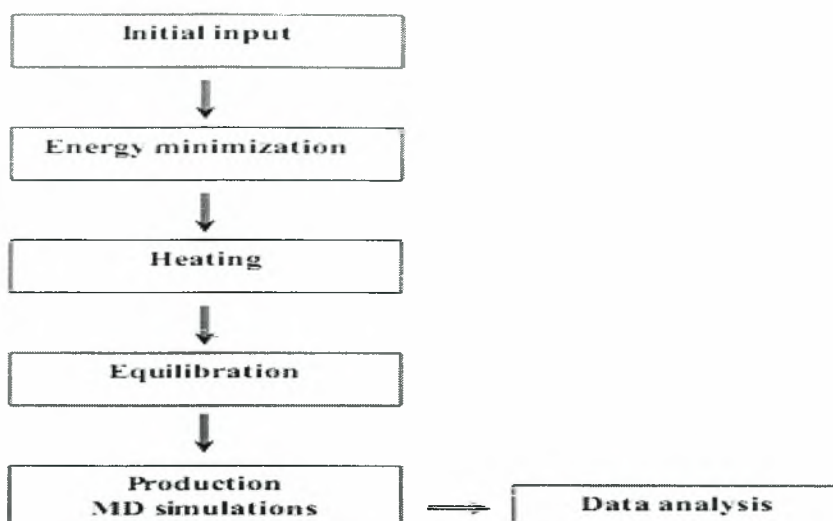
Όπως φαίνεται και από τον παραπάνω τύπο, στη δυναμική ενέργεια του συστήματος συνεισφέρουν όλα τα άτομα, ακόμη και τα άτομα του διαλύτη. Οι συνεισφορές είναι τόσο ενδομοριακές όσο και διαμοριακές.

Empirical Potential Energy Function



Εικόνα 2.4.2: Απεικονίζονται οι μορφές δυναμικής ενέργειας της κάθε συνεισφοράς της συνάρτησης δυναμικού.

Πριν ξεκινήσει η προσομοίωση του συστήματος, λαμβάνει χώρα μία ελαχιστοποίηση ενέργειας, ο στόχος της οποίας είναι η εύρεση της δομής και της ενέργειας εκείνης στο απόλυτο μηδέν. Εάν μια αρχική διαμόρφωση ενός μακρομορίου απέχει πολύ από την κατάσταση ισορροπίας, οι δυνάμεις μπορεί να είναι υπερβολικά μεγάλες και η προσομοίωση μπορεί να αποτύχει. Σε τέτοιες περιπτώσεις απαιτείται μια καλή ελαχιστοποίηση ενέργειας. Ένας άλλος λόγος για να εκτελέσει κανείς μια ελαχιστοποίηση ενέργειας είναι η απομάκρυνση όλης της κινητικής ενέργειας από το σύστημα: εάν μερικά στιγμιότυπα από τις δυναμικές προσομοιώσεις πρέπει να συγκριθούν, η ελαχιστοποίηση ενέργειας μειώνει το θερμικό θόρυβο στις δομές και τις ενέργειες δυναμικών και επομένως μπορούν να συγκριθούν καλύτερα [73]. Στη συνέχεια, προσφέρουμε στο σύστημα θερμότητα. Το στάδιο στο οποίο λαμβάνει χώρα πραγματικά η προσομοίωση είναι η εξισορρόπηση, όπου οι συνθήκες διατηρούνται σταθερές και αφήνεται το σύστημα να εξελιχθεί.



Εικόνα 2.4.3: Τα βήματα μίας προσομοίωσης.

THE GLOBAL MD ALGORITHM

1. Input initial conditions

Potential interaction V as a function of atom positions

Positions r of all atoms in the system

Velocities v of all atoms in the system

↓

repeat 2,3,4 for the required number of steps:

2. Compute forces

The force on any atom

$$F_i = -\frac{\partial V}{\partial r_i}$$

is computed by calculating the force between non-bonded atom pairs:

$$F_i = \sum_j F_{ij}$$

plus the forces due to bonded interactions (which may depend on 1, 2, 3, or 4 atoms), plus restraining and/or external forces.

The potential and kinetic energies and the pressure tensor are computed.

↓

3. Update configuration

The movement of the atoms is simulated by numerically solving Newton's equations of motion

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i}$$

or

$$\frac{dr_i}{dt} = v_i; \quad \frac{dv_i}{dt} = \frac{F_i}{m_i}$$

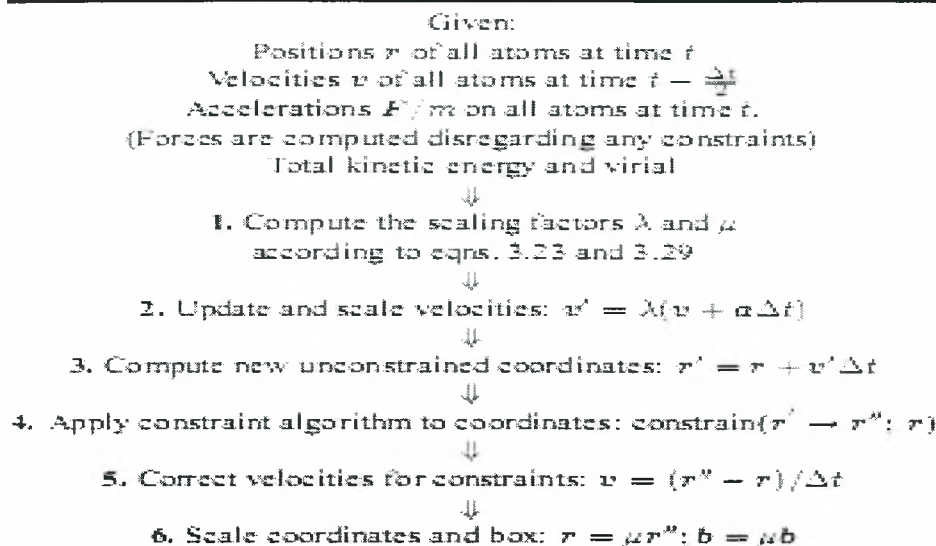
↓

4. if required: Output step

write positions, velocities, energies, temperature, pressure, etc.

Εικόνα 2.4.4: Σχηματική αναπαράσταση του αλγορίθμου MD

THE UPDATE ALGORITHM



Εικόνα 2.4.5: Ο αλγόριθμος επαναληπτικών υπολογισμών MD

Προκειμένου να ολοκληρωθούν οι δευτέρας τάξεως διαφορικές εξισώσεις, οι στιγμιαίες δυνάμεις που δρουν πάνω στα άτομα, καθώς και οι αρχικές θέσεις και ταχύτητες αυτών πρέπει να οριστούν. Οι τροχιές MD ορίζονται και από τα διανύσματα θέσης και από τα διανύσματα ταχύτητας και περιγράφουν την εξέλιξη του συστήματος στο χρόνο στο χώρο των φάσεων. Οι ταχύτητες $v_i(t)$ καθορίζουν την κινητική ενέργεια και τη θερμοκρασία του συστήματος. Καθώς τα μόρια ακολουθούν τις τροχιές τους, μπορούν να οπτικοποιηθούν και να αναλυθούν παρέχοντας τις κατά μέσο όρο ιδιότητες. Τα δυναμικά γεγονότα που μπορεί να επηρεάσουν τις λειτουργικές ιδιότητες του συστήματος μπορούν ευθέως να ανιχνευτούν σε ατομικό επίπεδο κάνοντας τη μοριακή δυναμική άκρως σημαντική στη μοριακή βιολογία.

Το σύστημα παρακολουθείται για ένα διάστημα, λαμβάνοντας υπόψη ότι η θερμοκρασία και η πίεση ή ο όγκος διατηρούνται σε επιτρεπτές τιμές και οι συντεταγμένες καταγράφονται σε ένα αρχείο σε κανονικά διαστήματα. Αυτό το αρχείο είναι αυτό που εξάγεται από το πρόγραμμα. Οι συντεταγμένες σαν συνάρτηση του χρόνου αποτελούν την τροχιά του συστήματος. Μετά τις αρχικές μεταβολές το σύστημα συνήθως φθάνει σε μία κατάσταση ισορροπίας. Πολλές μακροσκοπικές ιδιότητες μπορούν να εξαχθούν από το αρχείο εξόδου.

2.4.3 Οι περιορισμοί των προσομοιώσεων

Είναι χρήσιμο σε αυτό το σημείο να λάβουμε υπόψη τους περιορισμούς των προσομοιώσεων μοριακής δυναμικής. Ο χρήστης πρέπει να είναι γνώστης αυτών των περιορισμών και πάντα να ελέγχει τα αποτελέσματα με βάση γνωστές πειραματικές ιδιότητες ώστε να αποτιμά κάθε φορά την ακρίβεια της προσομοίωσης. Οι περιορισμοί παρατίθενται παρακάτω:

➤ **Οι προσομοιώσεις είναι κλασικές.**

Χρησιμοποιώντας τις εξισώσεις κίνησης του Νεύτωνα αυτόματα συνεπάγεται η εφαρμογή της κλασικής μηχανικής για να περιγραφεί η κίνηση των ατόμων. Αυτό είναι σωστό για τα περισσότερα άτομα σε κανονικές θερμοκρασίες αλλά υπάρχουν και εξαιρέσεις. Τα άτομα υδρογόνου είναι αρκετά ελαφριά και η κίνηση των πρωτονίων είναι μερικές φορές σημαντικού κβαντομηχανικού χαρακτήρα. Για παράδειγμα, ένα πρωτόνιο μπορεί να υπερπηδήσει ένα φράγμα δυναμικού και να βρεθεί σε περιοχές του χώρου απαγορευμένες από την κλασική μηχανική (φαινόμενο σήραγγας) κατά τη διάρκεια της μεταβίβασής του σε ένα δεσμό υδρογόνου. Τέτοιου είδους διαδικασίες δε μπορούν να αντιμετωπιστούν κατάλληλα από την κλασική δυναμική.

➤ **Τα ηλεκτρόνια είναι στη θεμελιώδη κατάσταση.**

Στη μοριακή δυναμική χρησιμοποιείται ένα συντηρητικό πεδίο δυνάμεων που είναι συνάρτηση των θέσεων των ατόμων μόνο. Αυτό σημαίνει ότι οι κινήσεις των ηλεκτρονίων δε λαμβάνονται υπόψη. Τα ηλεκτρόνια υποτίθεται ότι προσαρμόζουν τη δυναμική τους αμέσως μόλις οι θέσεις των ατόμων αλλάξουν (the Born-Oppenheimer προσέγγιση) και παραμένουν στη θεμελιώδη κατάσταση. Αυτό είναι πραγματικά σωστό σχεδόν πάντα. Όμως, τόσο οι διαδικασίες μεταφοράς των ηλεκτρονίων όσο και οι χημικές αντιδράσεις δε μπορούν να επεξεργαστούν σωστά.

➤ **Οι συναρτήσεις δυναμικού είναι προσεγγιστικές.**

Οι συναρτήσεις δυναμικού παρέχουν τις δυνάμεις. Δεν είναι πραγματικά μέρος της μεθόδου προσομοίωσης και οι παράμετροί τους μπορούν να τροποποιηθούν από το χρήστη, καθώς οι ανάγκες εμφανίζονται ή οι γνώσεις βελτιώνονται.

➤ **Η συνάρτηση δυναμικού εμφανίζει την ιδιότητα της αθροιστικότητας κατά ζεύγη (pair-additive).**

Αυτό σημαίνει ότι όλες οι δυνάμεις που δεν προέρχονται από τους δεσμούς είναι αποτέλεσμα του συνόλου ζευγών αλληλεπιδράσεων. Αθροίζονται όλα τα πιθανά ζεύγη και υπολογίζεται η συνολική ενέργεια.

➤ **Οι μεγάλες εμβέλειες αλληλεπιδράσεις από ένα σημείο και μετά δε λαμβάνονται υπόψη.**

➤ **Οι οριακές συνθήκες δεν είναι φυσικές.**

2.5 MODELLER

2.5.1 Τρόπος εγκατάστασης του προγράμματος

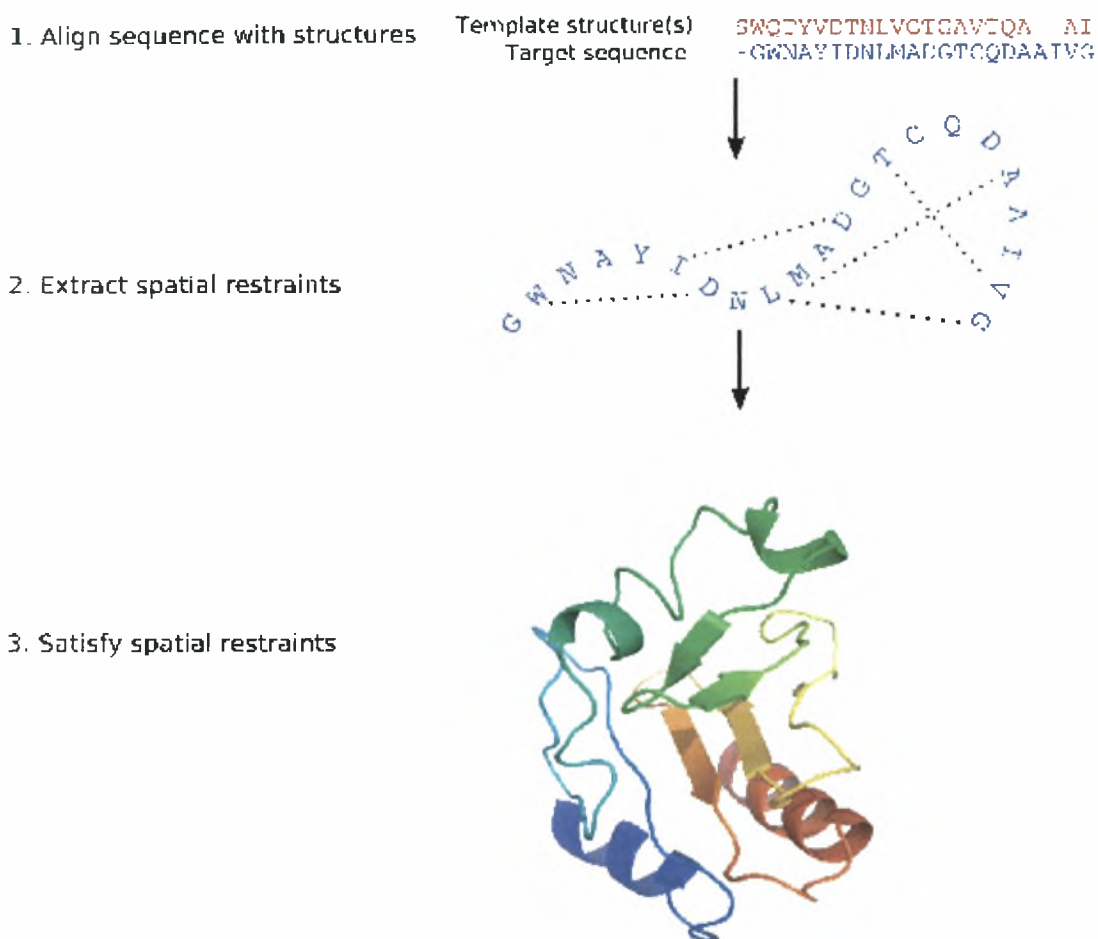
Για την εγκατάσταση του προγράμματος απαιτείται ένας κωδικός, ο οποίος αποκτάται μετά από αίτηση για άδεια. Μαζί με το πρόγραμμα πρέπει να αποθηκευτούν οι φάκελοι οι οποίοι περιέχουν τις ακολουθίες γνωστών δομών (PDB), καθώς και η βάση δεδομένων η οποία περιέχει τα προφίλ των ακολουθιών της PDB. Ο φάκελος είναι σε μορφή RPM. Με διπλό κλικ εξάγονται τα αρχεία που περιέχει σε έναν καινούριο φάκελο. Να σημειωθεί πως όλες οι ενέργειες πρέπει να γίνονται από τον χρήστη root στο Linux. Απαιτείται να πληκτρολογηθεί η εξής εντολή στην κονσόλα: `env KEY_MODELLER9v4=Κωδικός rpm -Uvh modeller-9v4-1.i386.rpm`. Εάν η εγκατάσταση είναι επιτυχής, τότε το πρόγραμμα θα πρέπει να τρέξει εάν πληκτρολογήσουμε `mod9v4`.

2.5.2 Ο αλγόριθμος του προγράμματος

Το Modeller είναι ένα πρόγραμμα το οποίο υπολογίζει χρησιμοποιώντας τη μέθοδο του συγκριτικού σχεδιασμού τριδιάστατες πρωτεϊνικές δομές μέσω ικανοποίησης ορισμένων χωρικών περιορισμών. Η ικανοποίηση των χωρικών περιορισμών είναι μία τεχνική εμπνευσμένη από τη μέθοδο NMR, μέσω της οποίας ένα σύνολο γεωμετρικών κριτηρίων χρησιμοποιείται για τη δημιουργία μίας συνάρτησης πιθανοτήτων για τη θέση του κάθε ατόμου στην πρωτεΐνη. Ο χρήστης παρέχει στο πρόγραμμα μία ομοπαράθεση της ακολουθίας που πρόκειται να σχεδιαστεί με βάση γνωστές συγγενικές δομές και το πρόγραμμα υπολογίζει αυτόματα ένα μοντέλο της πρωτεΐνης, το οποίο περιέχει όλα τα άτομα εκτός των υδρογόνων. Όπως προαναφέρθηκε, το πρόγραμμα υποστηρίζει τη μέθοδο του συγκριτικού σχεδιασμού μέσω ικανοποίησης χωρικών περιορισμών. Το πρόγραμμα έχει, επίσης, τη δυνατότητα να εκτελεί επιπρόσθετες ενέργειες, όπως για παράδειγμα *de novo* σχεδιασμό των πρωτεϊνικών δομών, βελτιστοποίηση ποικίλων μοντέλων πρωτεϊνικής δομής, πολλαπλή ομοπαράθεση ακολουθιών ή/και δομών, σάρωση βάσεων δεδομένων ακολουθιών, σύγκριση πρωτεϊνικών δομών κλπ. Το πρόγραμμα, επιπλέον, παρέχει τη δυνατότητα για *ab initio* πρόβλεψη δομής ή βρόγχων των πρωτεϊνών, που συχνά διαφέρουν σημαντικά ακόμα και μεταξύ ομόλογων πρωτεϊνών και συνεπώς είναι δύσκολο να προβλεφθούν μέσω συγκριτικού σχεδιασμού. Το πρόγραμμα γράφτηκε από τον Andrej Sali στο πανεπιστήμιο της Καλιφόρνιας στο Σαν Φρανσίσκο.

Τα αρχεία εισόδου στο πρόγραμμα περιέχουν περιορισμούς στη δομή ακολουθιών αμινοξέων και προσδετών που πρόκειται να μοντελοποιηθούν. Το αρχείο εξόδου περιέχει τις συντεταγμένες της τριδιάστατης δομής που ικανοποιεί τους περιορισμούς όσο το δυνατόν περισσότερο. Οι περιορισμοί μπορούν να προκύψουν από ένα διαφορετικό αριθμό πηγών. Αυτές περιλαμβάνουν ομόλογες πρωτεϊνικές δομές (συγκριτικός σχεδιασμός), πειράματα NMR, κανόνες σχηματισμού δευτεροταγούς δομής (combinatorial modeling), επανασύσταση εικόνας σε ηλεκτρονικό μικροσκόπιο, θεσεο-ελεγχόμενη μεταλλαξιγένεση, μέσα πεδία δυναμικών καταλοίπου-καταλοίπου και ατόμου-ατόμου κλπ. Οι περιορισμοί μπορούν

να ενεργούν πάνω σε αποστάσεις, γωνίες, διεδρες γωνίες, ζεύγη διεδρων γωνιών και μερικά άλλα χαρακτηριστικά που αφορούν το χώρο και που ορίζονται από τα άτομα ή τα ψευδοάτομα.



Εικόνα 2.5.1: Ο αλγόριθμος του Modeller.

Το πρόγραμμα είναι διαθέσιμο για τα περισσότερα λειτουργικά συστήματα, μεταξύ των οποίων Windows και Linux. Είναι γραμμένο στη γλώσσα προγραμματισμού FORTRAN 90. Χρησιμοποιεί μία scripting γλώσσα προγραμματισμού, ενώ δε συνοδεύεται από γραφικό περιβάλλον στη δωρεάν έκδοση. Το πρόγραμμα διατίθεται δωρεάν για ακαδημαϊκή χρήση. Στην παρούσα εργασία χρησιμοποιήθηκε το MODELLER 9v4, το οποίο κυκλοφόρησε στις 15 Μαΐου 2008.

2.5.3 Εφαρμογή του προγράμματος

Μοντελοποίηση της γαλακτικής αφυδρογονάσης από το μικροοργανισμό *Trichomonas vaginalis* με βάση μόνο ένα πρότυπο.

Ένα καινούριο γονίδιο γαλακτικής αφυδρογονάσης προσδιορίστηκε στη γενωμική αλληλουχία του *Trichomonas vaginalis* (TvLDH). Η συγκεκριμένη πρωτεΐνη είχε μεγαλύτερη συγγένεια με τη μηλική δεϋδρογονάση από τα ίδια είδη (TvMDH) σε σύγκριση με οποιαδήποτε άλλη LDH. Έγινε η υπόθεση ότι η TvLDH προέκυψε από

την TvMDH με συγκλίνουσα εξέλιξη σχετικά πρόσφατα. Μοντέλα με βάση την ομολογία κατασκευάστηκαν για την TvLDH και την TvMDH με σκοπό να μελετηθούν οι αλληλουχίες στο δομικό περιεχόμενο και να προταθούν πειράματα κατευθυνόμενης μεταλλαξιγένεσης ώστε να διασαφηνιστούν αλλαγές ειδικότητας σε αυτήν την ολοφάνερη περίπτωση συγκλίνουσας εξέλιξης ενζυμικής ειδικότητας. Τα αγρίου τύπου καθώς και τα μεταλλαγμένα ένζυμα εκφράστηκαν και συγκρίθηκαν οι ενεργότητές τους.

Τα βήματα εξηγούνται παρακάτω. Σε κάθε βήμα μόνο η αλληλουχία είναι γνωστή.

1. Έρευνα για δομές συγγενικές στην TvLDH

Πρώτα, πρέπει η αλληλουχία-στόχος, που στο συγκεκριμένο παράδειγμα είναι η TvLDH, να μετατραπεί σε PIR μορφή, γιατί μόνον τη συγκεκριμένη μορφή μπορεί και διαβάσει το πρόγραμμα. Το αρχείο θα έχει την κατάληξη .ali (επειδή περιέχει την ακολουθία) και θα ονομάζεται 'TvLDH.ali'.

```
>P1 ; TvLDH
sequence: TvLDH: ::::: 0.00: 0.00
MSEAAHVLIITGAAGQIGYILSHWIASGELYGDRQVYLHLLDIPPAMNRLTALTMELEDCAFPHLAGFVATTPKA
AFKIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKPSVKVLVIGNPDNTNCEIAMLHAKNLKPEN
FSSLMLDQNRAYYEVASKLGVNDVKDVHDIIVWGNHGESMVADLTQATFTKEGKTQKVVVDVLDHDIYVFDTFFKKI
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNPYGIKPGVVVFSFPCNVNDEKGIHVV
EGFKVNDWLRKLDFTKDLFHEKEIALNHLAQQG*
```

Εικόνα 2.5.2: Η ακολουθία που πρόκειται να μοντελοποιηθεί σε μορφή FASTA. Το όνομα του αρχείου είναι TvLDH.ali.

Η πρώτη σειρά περιέχει τον κωδικό της αλληλουχίας, στη μορφή **>P1: κωδικός**. Η δεύτερη σειρά με τα δέκα πεδία που χωρίζονται με άνω και κάτω τελείες περιέχει συνήθως πληροφορίες για το αρχείο, εάν υπάρχουν. Μόνο δύο από τα πεδία χρησιμοποιούνται για τις ακολουθίες, η λέξη 'sequence' δηλώνει ότι το αρχείο περιέχει μία αλληλουχία της οποίας η δομή είναι άγνωστη και η λέξη TvLDH είναι το όνομα του αρχείου που περιέχει το μοντέλο. Το υπόλοιπο του αρχείου περιέχει την αλληλουχία και το σύμβολο * δηλώνει το τέλος της ακολουθίας. Σε περίπτωση που παραλειφθεί αυτό το σύμβολο το script δε θεωρείται σωστό και το πρόγραμμα δε μπορεί να τρέξει. Δεν πρέπει να ξεχνιέται ποτέ, όπως ποτέ δεν πρέπει επίσης να παραλείπονται τα δέκα πεδία της δεύτερης σειράς. Τα αμινοξέα πάντα πρέπει να γράφονται με ένα γράμμα και με κεφαλαία. Με μικρά γράμματα γράφονται μόνο τα τροποποιημένα αμινοξέα, εκείνα δηλαδή που δεν περιλαμβάνονται στα 20 βασικά.

Η έρευνα για δυνητικά συγγενικές ακολουθίες γνωστής δομής εκτελείται με την εντολή **profile.build()**. Το script που αντιστοιχεί στην παραπάνω εντολή είναι το ακόλουθο:


```

from modeller import *

log.verbose()
env = environ()

#-- Prepare the input files

#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
         chains_list='ALL', minmax_db_seq_len=(30, 4000), clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Now, read in the binary database
sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='TvLDH.ali', alignment_format='PIR', align_codes='ALL')

#-- Convert the input sequence/alignment into
#   profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
         gap_penalties_1d=(-500, -50), n_prof_iterations=1,
         check_profile=False, max_aln_evalue=0.01)

#-- Write out the profile in text format
prf.write(file='build_profile.prf', profile_format='TEXT')

#-- Convert the profile back to alignment format
aln = prf.to_alignment()

#-- Write out the alignment file
aln.write(file='build_profile.ali', alignment_format='PIR')

```

Εικόνα 2.5.3: Το script που αντιστοιχεί στην εντολή `profile.build()`. Το όνομα του αρχείου είναι `build_profile.py`.

Το παραπάνω script εκτελεί τα ακόλουθα:

- 1) Εκκινεί το 'περιβάλλον' για να τρέξει το πρόγραμμα δημιουργώντας ένα καινούριο αντικείμενο 'environ'. Σχεδόν όλα τα scripts του MODELLER απαιτούν αυτό το βήμα, καθώς το νέο αντικείμενο χρειάζεται για τα πιο πολλά χρήσιμα αντικείμενα.
- 2) Δημιουργεί ένα άλλο καινούριο αντικείμενο 'sequence_db', το οποίο καλείται 'sdb'. Τα συγκεκριμένα αντικείμενα χρησιμοποιούνται για να συμπεριλάβουν μεγάλες βάσεις δεδομένων πρωτεϊνικών ακολουθιών.
- 3) Διαβάζει ένα αρχείο κειμένου που περιέχει απαραίτητες αλληλουχίες PDB με 95% ομοιότητα στην βάση δεδομένων sdb. Οι αλληλουχίες μπορούν να βρεθούν στο αρχείο 'pdb_95.pir'. και αυτός ο φάκελος περιέχει τις αλληλουχίες σε μορφή PIR. Αλληλουχίες που έχουν λιγότερα από 30 και περισσότερα από 4000 κατάλοιπα απορρίπτονται και μη τυπικά κατάλοιπα αφαιρούνται.

- 4) Γράφει ένα δυαδικό ειδικό αρχείο με όλες τις αλληλουχίες που διαβάστηκαν στο προηγούμενο βήμα.
- 5) Διαβάζει το δυαδικό αρχείο ξανά. Να σημειωθεί ότι εάν πρόκειται να χρησιμοποιηθεί η ίδια βάση δεδομένων πολλές φορές, τότε τα δύο προηγούμενα βήματα πρέπει να χρησιμοποιηθούν μόνο την πρώτη φορά ώστε να δημιουργηθεί η δυαδική βάση δεδομένων. Τις επόμενες φορές που θα τρέξει το πρόγραμμα αυτά τα δύο βήματα μπορούν να παραλειφθούν και να χρησιμοποιηθεί το δυαδικό αρχείο, αφού το διάβασμα του συγκεκριμένου φακέλου είναι πολύ πιο γρήγορο από το διάβασμα ενός φακέλου γραμμένου σε μορφή PIR.
- 6) Δημιουργεί ένα καινούριο αντικείμενο με το όνομα 'aln', δηλαδή 'alignment' και τη μετατρέπει σε προφίλ 'prf'. Τα προφίλ περιέχουν ίδιες πληροφορίες με τις αλληλουχίες, μόνο που η δομή τους είναι πιο συμπαγής και η εύρεση μιας αλληλουχίας είναι ευκολότερη.
- 7) Στη συνέχεια, αφού προηγήθηκαν τα παραπάνω βήματα που ήταν απαραίτητα, ψάχνει πλέον στη βάση δεδομένων 'sdb' για την αλληλουχία-στόχο που πλέον έχει ονομαστεί από προηγούμενο βήμα 'prf'. Συμφωνίες από τη βάση δεδομένων των ακολουθιών προσθέτονται στο προφίλ.
- 8) Γράφει ένα προφίλ της αλληλουχίας-στόχου και των ομόλογών της στο φάκελο με την ονομασία 'build_profile.prf'. Η ίδια πληροφορία γράφεται επίσης σε τυπική μορφή ομοπαράθεσης.
- 9) Το script γράφεται στη γλώσσα προγραμματισμού Python, περιέχει εντολές ειδικές του Modeller και τρέχει με την εξής εντολή:

mod9v4 build_profile.py

(η Python μπορεί να χρησιμοποιηθεί μόνον εάν είναι σωστά εγκατεστημένη στο σύστημα)

Η επιλογή profile.build() έχει πολλές επιλογές. Στο συγκεκριμένο παράδειγμα *rr_file* έχει καθοριστεί να χρησιμοποιεί τον πίνακα BLOSUM62 (το αρχείο 'blosum62.sim.mat' παρέχεται από το πρόγραμμα). Ανάλογα, οι παράμετροι *matrix_offset* και *gap_penalties_1d* έχουν ρυθμιστεί στις κατάλληλες τιμές για τη BLOSUM62 matrix. Στο συγκεκριμένο παράδειγμα, η αναζήτηση εκτελείται μία φορά και για αυτόν το λόγο η παράμετρος *n_prof_iterations* έχει οριστεί με την τιμή 1. Τέλος, η παράμετρος *max_aln_value* ορίζεται να έχει την τιμή 0,01, υποδεικνύοντας ότι μόνο αλληλουχίες με e-values μικρότερα ή ισοδύναμα με 0,01 θα συμπεριλαμβάνονται στο τελικό προφίλ.

2. Επιλογή ενός προτύπου

Το αρχείο εξόδου του πρώτου script 'build_profile.prf' είναι γραμμένο στο αρχείο 'build_profile.log'. Το MODELLER πάντα παράγει ένα αρχείο log. Τα λάθη και οι προειδοποιήσεις στα αρχεία log μπορούν να βρεθούν ψάχνοντας για τους χαρακτήρες 'E' και 'W', αντίστοιχα. Το MODELLER γράφει το προφίλ και σε μορφή κειμένου στο αρχείο με την ονομασία 'build_profile.prf'. ένα τμήμα του (έχουν παραλειφθεί οι στοιχισμένες αλληλουχίες) παρατίθεται παρακάτω. Οι πρώτες έξι γραμμές είναι σχόλια και υποδεικνύουν τις παραμέτρους εισόδου που χρησιμοποίησε το πρόγραμμα για να δημιουργήσει το προφίλ. Οι επόμενες σειρές περιλαμβάνουν τις ομοιότητες που βρέθηκαν από την εντολή **profile.build()**. Παρακάτω παρατίθεται το αρχείο 'build_profile.prf':

```

# Number of sequences:      30
# Length of profile       :   335
# N_PROF_ITERATIONS      :    1
# GAP_PENALTIES_1D       : -900.0  -50.0
# MATRIX_OFFSET           :    0.0
# RR_FILE                 : ${MODINSTALL8v1}/modlib//as1.sim.mat
 1 TvLDH                S    0  335    1  335    0    0    0    0.  0.0
 2 la5z                  X    1  312   75  242   63  229  164  28.  0.83E-08
 3 lb8pA                  X    1  327    7  331    6  325  316  42.  0.0
 4 lbcmA                  X    1  318    1  325    1  310  309  45.  0.0
 5 lt2dA                  X    1  315    5  256    4  250  238  25.  0.66E-04
 6 lciVA                  X    1  374    6  334   33  358  325  35.  0.0
 7 2cmd                   X    1  312    7  320    3  303  289  27.  0.16E-05
 8 lo6zA                  X    1  303    7  320    3  287  278  26.  0.27E-05
 9 lur5A                  X    1  299   13  191    9  171  158  31.  0.25E-02
10 lguzA                  X    1  305   13  301    8  280  265  25.  0.28E-08
11 lgvOA                  X    1  301   13  323    8  289  274  26.  0.28E-04
12 lhyeA                  X    1  307    7  191    3  183  173  29.  0.14E-07
13 li0zA                  X    1  332   85  300   94  304  207  25.  0.66E-05
14 li10A                  X    1  331   85  295   93  298  196  26.  0.86E-05
15 li1dnA                 X    1  316   78  298   73  301  214  26.  0.19E-03
16 6ldh                   X    1  329   47  301   56  302  244  23.  0.17E-02
17 2ldx                   X    1  331   66  306   67  306  227  26.  0.25E-04
18 5ldh                   X    1  333   85  300   94  304  207  26.  0.30E-05
19 9ldtA                  X    1  331   85  301   93  304  207  26.  0.10E-05
20 lllc                   X    1  321   64  239   53  234  164  26.  0.20E-03
21 llldA                  X    1  313   13  242    9  233  216  31.  0.31E-07
22 5mdhA                  X    1  333    2  332    1  331  328  44.  0.0
23 7mdhA                  X    1  351    6  334   14  339  325  34.  0.0
24 1mldA                  X    1  313    5  198    1  189  183  26.  0.13E-05
25 loc4A                  X    1  315    5  191    4  186  174  28.  0.18E-04
26 lojwA                  X    1  294   78  320   68  285  218  28.  0.43E-05
27 lpzgA                  X    1  327   74  191   71  190  114  30.  0.16E-06
28 lsmkA                  X    1  313    7  202    4  198  188  34.  0.0
29 lsovA                  X    1  316   81  256   76  248  160  27.  0.93E-03
30 ly6jA                  X    1  289   77  191   58  167  109  33.  0.32E-05

```

Εικόνα 2.5.4: Αφού δώσουμε την πρώτη εντολή, μεταξύ των αρχείων που προκύπτουν είναι και το παραπάνω. Ονομάζεται 'build_profile.prf' και περιέχει τις ακολουθίες πρότυπα.

Οι πιο σημαντικές στήλες στο αρχείο εξόδου του *profile.build()* είναι η δεύτερη, δέκατη, ενδέκατη και δωδέκατη στήλη. Η δεύτερη στήλη αναφέρει τον κωδικό της PDB αλληλουχίας που συγκρίθηκε με την αλληλουχία-στόχο. Ο κωδικός PDB σε κάθε σειρά είναι αντιπροσωπευτικός μιας ομάδας αλληλουχιών που μοιράζονται 95% ή και περισσότερη ομοιότητα ή έχουν λιγότερα από 30 κατάλοιπα ή λιγότερο από 30% διαφορά στο μήκος. Η ενδέκατη στήλη αναφέρει τα ποσοστά ομοιότητας κανονικοποιημένα ως προς τα μήκη της ομοπαράθεσης, τα οποία φαίνονται στη δέκατη στήλη. Γενικά, ποσοστό ομοιότητας πάνω από 25% περίπου υποδεικνύει ένα δυνητικό πρότυπο εκτός εάν η ομοπαράθεση είναι περιορισμένη. Ένα καλύτερο μέτρο της αξιοπιστίας της ομοπαράθεσης δίνεται στη δωδέκατη στήλη από την τιμή της e-value της ομοπαράθεσης. Στο συγκεκριμένο παράδειγμα, έξι αλληλουχίες δείχνουν πολύ σημαντικές ομοιότητες με e-value ίση με 0. Όπως αναμενόταν, όλες αντιστοιχούν σε μηλικές αφυδρογονάσεις (*1bdm:A*, *5mdh:A*, *1b8p:A*, *1civ:A*, *7mdh:A*, και *1smk:A*). Προκειμένου να επιλεγεί το πιο κατάλληλο πρότυπο από αυτές τις έξι ακολουθίες, θα χρησιμοποιηθεί η εντολή *alignment.compare_structures()*, ώστε να εκτιμηθεί η δομική και αμινοξική ομοιότητα μεταξύ των πιθανών προτύπων (αρχείο 'compare.py').

```

from modeller import *

env = environ()
aln = alignment(env)
for (pdb, chain) in (('1b8p', 'A'), ('1bdm', 'A'), ('1civ', 'A'),
                    ('5mdh', 'A'), ('7mdh', 'A'), ('1smk', 'A')):
    m = model(env, file=pdb, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(m, atom_files=pdb, align_codes=pdb+chain)
aln.malign()
aln.malign3d()
aln.compare_structures()
aln.id_table(matrix_file='family.mat')
env.dendrogram(matrix_file='family.mat', cluster_cut=-1.0)

```

Εικόνα 2.5.5: Το όνομα του script είναι compare.py.

Στη συγκεκριμένη περίπτωση, δημιουργήσαμε ένα αρχικό κενό alignment object 'aln' και στη συνέχεια χρησιμοποιήσαμε το βρόγχο for της Python, ώστε να δώσουμε οδηγίες στο πρόγραμμα να διαβάσει το κάθε αρχείο PDB. Να σημειωθεί ότι όλα τα αρχεία PDB πρέπει να βρίσκονται στον ίδιο κατάλογο που βρίσκεται και το script. Χρησιμοποιούμε το όρισμα *model_segment* για να ζητήσουμε να διαβαστεί μόνο μία αλυσίδα από το κάθε PDB. Καθώς διαβάζεται κάθε δομή, χρησιμοποιούμε τη μέθοδο *append_model* για να προσθέσουμε τη δομή στην ομοπαράθεση. Στο τέλος του βρόγχου όλες οι δομές βρίσκονται στην ομοπαράθεση, αλλά δεν είναι κατάλληλα στοιχισμένες η μία με την άλλη. Συνεπώς, η ομοπαράθεση βελτιώνεται με *malign* για να υπολογιστεί μία πολλαπλή ομοπαράθεση. Ακολούθως, η εντολή *malign3d* εκτελεί μία υπέρθεση διαδοχικών βελτιώσεων ελαχίστων τετραγώνων των έξι τρισδιάστατων δομών, χρησιμοποιώντας την πολλαπλή ομοπαράθεση αλληλουχιών ως αρχή. Η εντολή *compare_structures* συγκρίνει τις δομές με βάση την ομοπαράθεση που κατασκευάστηκε από την εντολή *malign3d*. Δεν κάνει ομοπαράθεση, αλλά υπολογίζει RMS και DRMS αποκλίσεις μεταξύ των θέσεων των ατόμων και των αποστάσεων, διαφορές μεταξύ των διέδρων γωνιών της κύριας αλυσίδας και των πλευρικών αλυσίδων και μερικά άλλα μέτρα. Τελικά, η εντολή *id_table* γράφει έναν φάκελο με κατά ζεύγη αποστάσεις που μπορεί να χρησιμοποιηθεί αμέσως ως το αρχείο εισόδου για την εντολή *dendrogram*. Η συγκεκριμένη εντολή υπολογίζει ένα δενδρόγραμμα από τις κατά ζεύγη αποστάσεις, που βοηθά στην απεικόνιση των διαφορών μεταξύ των υποψήφιων προτύπων, όπως φαίνεται από απόσπασμα του αρχείου (αρχείο 'compare.log').

Sequence identity comparison (ID_TABLE):

```

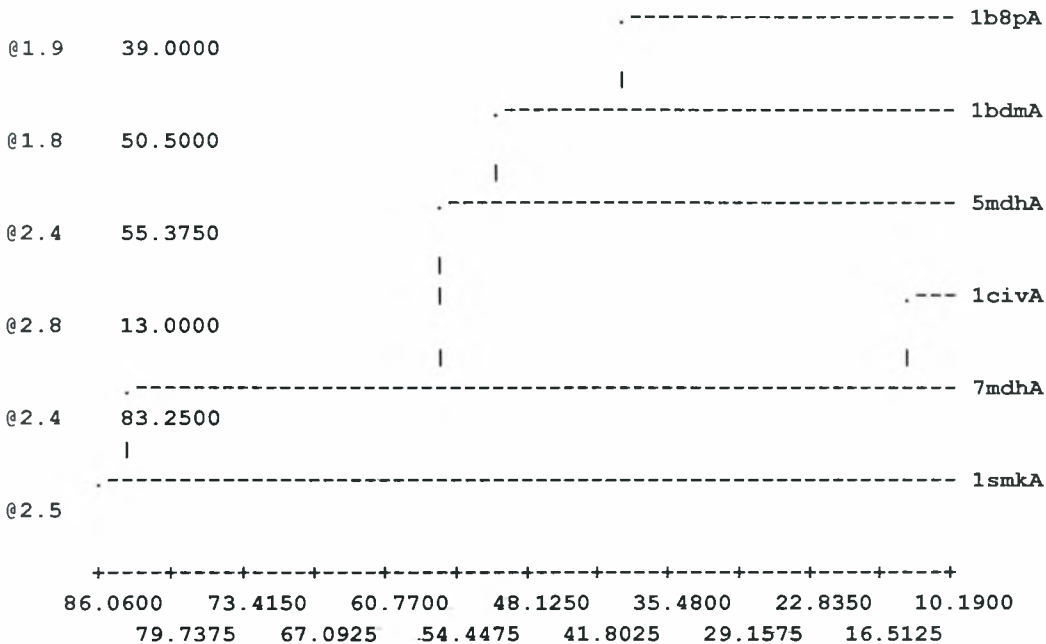
Diagonal      ... number of residues;
Upper triangle ... number of identical residues;
Lower triangle ... % sequence identity, id/min(length).

      1b8pA @11bdmA @11civA @25mdhA @27mdhA @21smkA @2
1b8pA @1      327      194      147      151      153      49
1bdmA @1      61      318      152      167      155      56

```

1civA @2	45	48	374	139	304	53
5mdhA @2	46	53	42	333	139	57
7mdhA @2	47	49	87	42	351	48
1smkA @2	16	18	17	18	15	313

Weighted pair-group average clustering based on a distance matrix:



Εικόνα 2.5.6: Απόσπασμα του αρχείου compare.log.

Η παραπάνω στοίχιση δείχνει ότι οι ακολουθίες *1civ:A* και *7mdh:A* είναι σχεδόν πανομοιότυπες τόσο δομικά όσο και όσον αφορά τις αλληλουχίες τους. Όμως, η *7mdh:A* διακρίνεται για την καλύτερη κρυσταλλογραφική της ανάλυση, αποκλείοντας την *1civ:A*. Μία δεύτερη ομάδα δομών (*5mdh:A*, *1bdm:A*, and *1b8p:A*) μοιράζονται κάποιες ομοιότητες. Από αυτήν την ομάδα, η *5mdh:A* έχει τη μικρότερη ανάλυση επιτρέποντας συνεπώς μόνο τις *1bdm:A* and *1b8p:A* να ληφθούν υπόψη. Η *1smk:A* είναι η πιο αποκλίνουσα δομή από όλα τα πιθανά πρότυπα. Τελικά, επιλέγεται η *1bdm:A* έναντι των *1b8p:A* και *5mdh:A* λόγω του καλύτερου κρυσταλλογραφικού παράγοντα R και της υψηλότερης ομοιότητας στην ακολουθία με την ακολουθία στόχο (45%).

3. Ομοπαράθεση της TvLDH με το πρότυπο

Ένας καλός τρόπος ομοπαράθεσης της ακολουθίας με τη δομή *1bdm:A* είναι η εντολή *align2d()*. Αν και η εντολή βασίζεται σε έναν αλγόριθμο dynamic programming, είναι διαφορετική από τις συνηθισμένες μεθόδους ομοπαράθεσης ακολουθιών, επειδή λαμβάνει υπόψη διάφορα δομικά χαρακτηριστικά του προτύπου. Η συγκεκριμένη δυνατότητα επιτυγχάνεται με τη βοήθεια μιας συνάρτησης που επιβάλλει διάφορες ποινές στα κενά που τείνει να τοποθετεί τα κενά σε διαλύτη που εκτίθεται και σε ελικωμένες περιοχές έξω από στοιχεία δευτεροταγούς δομής και

μεταξύ δύο θέσεων που είναι κοντά στο χώρο. Συνεπώς, τα λάθη της ομοπαράθεσης μειώνονται περίπου κατά το ένα τρίτο αναφορικά με εκείνα που προκύπτουν με τις συνηθισμένες τεχνικές ομοπαράθεσης ακολουθιών. Αυτή η βελτίωση γίνεται ακόμη πιο σημαντική καθώς η ομοιότητα μεταξύ των ακολουθιών μειώνεται και ο αριθμός των κενών αυξάνεται. Στο συγκεκριμένο παράδειγμα, η ομοιότητα μεταξύ του προτύπου και του στόχου είναι τόσο υψηλή που σχεδόν κάθε μέθοδος ομοπαράθεσης με λογικές παραμέτρους θα κατέληγε στην ίδια ομοπαράθεση. Το ακόλουθο script ομοπαράθετει την TvLDH ακολουθία του φακέλου 'TvLDH.ali' με τη δομή *1bdm*:A στο PDB φάκελο '1bdm.pdb' (αρχείο 'align2d.py').

```
from modeller import *

env = environ()
aln = alignment(env)
mdl = model(env, file='1bdm', model_segment=('FIRST:A','LAST:A'))
aln.append_model(mdl, align_codes='1bdmA', atom_files='1bdm.pdb')
aln.append(file='TvLDH.ali', align_codes='TvLDH')
aln.align2d()
aln.write(file='TvLDH-1bdmA.ali', alignment_format='PIR')
aln.write(file='TvLDH-1bdmA.pap', alignment_format='PAP')
```

Εικόνα 2.5.7: Το script align2d.py.

Στο παραπάνω script ξαναδημιουργείται ένα αντικείμενο με την ονομασία 'environ' για να χρησιμοποιήσει ως σημείο εισόδου στις επόμενες εντολές. Δημιουργούμε μία άδεια ομοπαράθεση με το όνομα 'aln' και τελικά ένα νέο πρωτεϊνικό μοντέλο 'mdl' στο οποίο διαβάζουμε την αλυσίδα A της δομής *1bmd*. Η εντολή *append_model()* μεταφέρει την PDB ακολουθία αυτού του μοντέλου στην ομοπαράθεση και της δίνει το όνομα '1bdmA' (*align_codes*). Στη συνέχεια, προσθέτουμε την ακολουθία 'TvLDH' από το φάκελο 'TvLDH.seq' στην ομοπαράθεση χρησιμοποιώντας την εντολή *append()*. Η εντολή *align2d()* εκτελείται, ώστε να ομοπαρατεθούν οι ακολουθίες. Τελικά, η ομοπαράθεση γράφεται σε δύο μορφές, στην PIR ('TvLDH-1bdmA.ali') και την PAP ('TvLDH-1bdmA.pap'). η πρώτη μορφή χρησιμοποιείται από το πρόγραμμα στο επόμενο βήμα που σχεδιάζεται το μοντέλο, ενώ η μορφή PAP είναι ευκολότερο να παρατηρηθεί οπτικά. Εξαιτίας της υψηλής ομοιότητας στόχου-προτύπου, υπάρχουν μόνο μερικά κενά στην ομοπαράθεση. Στη μορφή PAP όλες οι ταυτόσημες θέσεις σημαίνονται με a '*' (αρχείο 'TvLDH-1bdmA.pap').

```

_aln.pos       10         20         30         40         50         60
1bdmA   MKAPVRVAVTGAAGQIGYSLLFRIAAGEMLGKDPVILQELLEIPQAMKALEGVVMELEDCAFPLLAGL
TvLDH   MSEAABVLIITGAAGQIGYIILSHWIASGELYG-DRQVYLHLLDIPAMNRLTALTMELEDCAFPHLAGF
_consrvd *      *  ***** *    ** * * *   * * * * * * * *  ***** *

_aln.p         70         80         90         100        110        120        130
1bdmA   EATDDPFDVAFKADADYALLVGAAPRL-----QVNGKIFTEQGRALAEVAKDKVKVLVGNPNANTN
TvLDH   VATTDPKAAPFKIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKPSVKVLVI GNPDNTN
_consrvd ** **  ***** * * *  *          * **  * *   ** ***** ** **

_aln.pos       140        150        160        170        180        190        200
1bdmA   ALIAYKNAPGLNPRNFTAMTRLDHNRKAQLAKKTGTGVDIRRMTVWGNHSSIMFPDLFHAEVD---
TvLDH   CEIAMLHAKNLPENFSSLSMLDQNRAYEYEVASKLGVVDVVKVDVHDIIIVWGNHGESMVADLTQATFTKEG
_consrvd **  * * * * *  ** ***   * * * *   ***** * * * *
```

```

_aln.pcs      210      220      230      240      250      260      270
1bdmA      -GRPALELVDMEWYEKVFIP TVAQRGAATI QARGASSAASAANA AIEHIRDWALGTPEGDWVSM AVPS
TvLDH      KTQKVVDVLDHDYVFD TFFKKIGHRAW DILEHRGFTS AASPTKAAIQHMKAWLFGTAPGEVLSMGIPV
_consrvd      *      *      *      *      **      ****      ***      *      **      *      **      *

_aln.pcs      280      290      300      310      320      330
1bdmA      Q--GEYGIPEGIVYSFPVTAK-DGAYRVVEGLEINEFARKRMEIT AQELLDEME QVKAL--GLI
TvLDH      PEGNPGYGIKPGVVFSPFCNV DKEGKIHVVEGFKNVDWLREK LDFTEKDLFHEKEIALNH LAQGG
_consrvd      ***      *      *      *      *      *      *      *      *      *      *      *

```

Εικόνα 2.5.8: Το αρχείο TvLDH-1bdmA.pap.

4. σχεδιασμός του μοντέλου

Εφόσον η ομοπαράθεση στόχου-προτύπου κατασκευαστεί, το πρόγραμμα υπολογίζει ένα τρισδιάστατο μοντέλο του στόχου τελείως αυτόματα χρησιμοποιώντας την **automodel** class. Το επόμενο script θα δημιουργήσει πέντε παρόμοια μοντέλα της TvLDH με βάση την πρότυπη δομή **1bdm:A** και την ομοπαράθεση στο φάκελο 'TvLDH-1bdmA.ali' (file 'model-single.py').

```

from modeller import *
from modeller.automodel import *

env = environ()
a = automodel(env, alnfile='TvLDH-1bdmA.ali',
              knowns='1bdmA', sequence='TvLDH')
a.starting_model = 1
a.ending_model = 5
a.make()

```

Εικόνα 2.5.9: Το script model-single.py.

Η πρώτη σειρά φορτώνει την *automodel* class. Στη συνέχεια, δημιουργείται ένα αντικείμενο *automodel* που ονομάζεται 'a' και τίθενται οι παράμετροι που θα καθοδηγήσουν τη διαδικασία του σχεδιασμού του μοντέλου. Το όνομα *Alnfile* δίνεται στο φάκελο που περιέχει την ομοπαράθεση προτύπου-στόχου σε μορφή PIR. Η παράμετρος *Knowns* καθορίζει τη γνωστή/ές πρότυπη/ες δομή/ές στο *alnfile* ('TvLDH-1bdmA.ali'). Η παράμετρος *sequence* καθορίζει το όνομα της πρότυπης ακολουθίας στο αρχείο *alnfile*, ενώ οι όροι *starting_model* και *ending_model* καθορίζουν τον αριθμό των μοντέλων που υπολογίζονται με δείκτες από το 1 έως το 5. Η τελευταία σειρά στο φάκελο καλεί τη μέθοδο *make* που υπολογίζει τα μοντέλα. Τα πιο σημαντικά αρχεία που εξάγονται είναι 'model-single.log', το οποίο αναφέρει προειδοποιήσεις, λάθη και άλλες χρήσιμες πληροφορίες που περιλαμβάνουν τους εισαχθέντες περιορισμούς που χρησιμοποιούνται για το σχεδιασμό και που καταπατούνται στο τελικό μοντέλο, καθώς και το αρχείο 'TvLDH.B9999000*.pdb' που περιέχει τις συντεταγμένες από τα πέντε μοντέλα στη μορφή PDB. Τα μοντέλα μπορούν να οπτικοποιηθούν από οποιοδήποτε πρόγραμμα διαβάζει φακέλους στη μορφή PDB, όπως το Chimera.

5. Εκτίμηση του μοντέλου

Εάν για τον ίδιο στόχο υπολογιστούν διάφορα μοντέλα, πρέπει να επιλεγεί το καλύτερο μοντέλο επιλέγοντας το μοντέλο με τη χαμηλότερη τιμή της αντικειμενικής συνάρτησης του προγράμματος, η οποία αναφέρεται;ι στη δεύτερη σειρά του μοντέλου PDB. Η τιμή της αντικειμενικής συνάρτησης του προγράμματος σεν είναι ένα απόλυτο μέτρο, με την έννοια ότι μπορεί να χρησιμοποιηθεί μόνο για να κατατάξει μοντέλα που έχουν υπολογιστεί με βάση την ίδια ομοπαράθεση.

Μόλις επιλεγεί ένα τελικό μοντέλο, υπάρχουν διάφοροι τρόποι για να αξιολογηθεί. Στο συγκεκριμένο παράδειγμα, το δυναμικό DOPE χρησιμοποιείται για να εκτιμηθεί η αναδίπλωση του μοντέλου. Πριν, βέβαια, από οποιαδήποτε εκτίμηση του μοντέλου, πρέπει κανείς να ελέγξει το αρχείο log μήπως προέκυψαν κάποια λάθη καθώς έτρεχε το πρόγραμμα ('model-single.log') και επιπλέον πρέπει να ελεγχθεί και εάν κάποιος περιορισμός καταπατήθηκε. Το αρχείο 'evaluate_model.py' εκτιμά ένα μοντέλο που εισάγεται με το δυναμικό DOPE.

```
from modeller import *
from modeller.scripts import complete_pdb

log.verbose() # request verbose output
env = environ()
env.libs.topology.read(file='${LIB}/top_heav.lib') # read topology
env.libs.parameters.read(file='${LIB}/par.lib') # read parameters

# read model file
mdl = complete_pdb(env, 'TvLDH.B99990001.pdb')

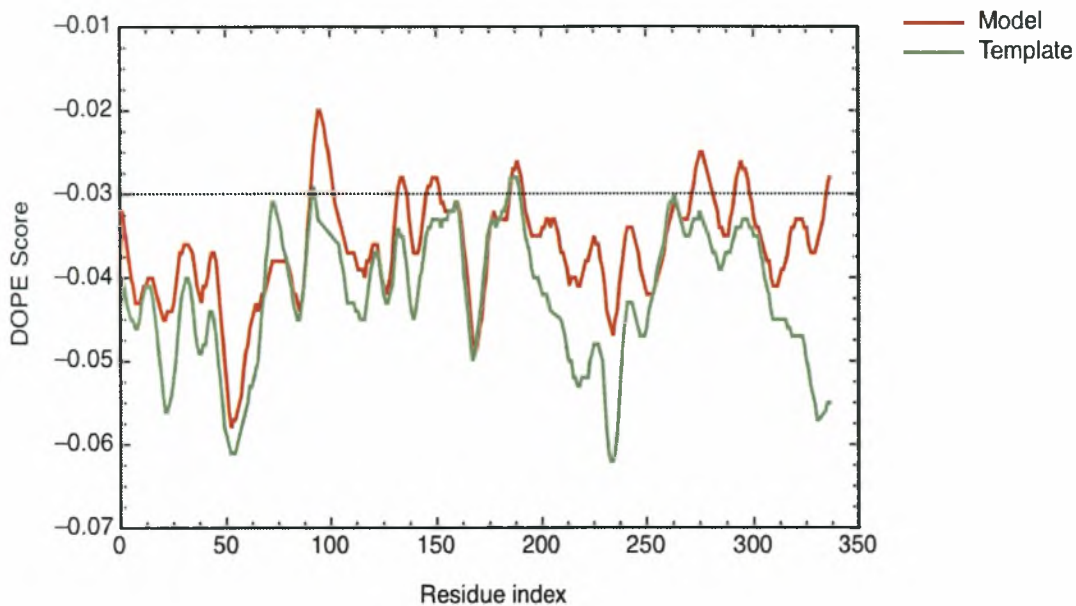
# Assess with DOPE:
s = selection(mdl) # all atom selection
s.assess_dope(output='ENERGY_PROFILE NO_REPORT', file='TvLDH.profile',
              normalize_profile=True, smoothing_window=15)
```

Εικόνα 2.5.10: Το script evaluate_model.py.

Στο παραπάνω script χρησιμοποιείται το *complete_pdb* script για να διαβάσει ένα φάκελο PDB και να τον προετοιμάσει για τους ενεργειακούς υπολογισμούς (αυτό αυτόματα λαμβάνει υπόψη την πιθανότητα ο φάκελος PDB να περιέχει άτομα σε μη καθορισμένη διάταξη ή να έχει διαφορετικά υποσύνολα ατόμων, όπως όλα τα άτομα που περιλαμβάνουν υδρογόνα, ενώ το πρόγραμμα χρησιμοποιεί μόνο βαριά άτομα ή αντίστροφα. Η ενέργεια DOPE ακολούθως υπολογίζεται με την εντολή *assess_dope* και επιπλέον ζητείται ένα ενεργειακό προφίλ, με εξομάλυνση σε παράθυρο 15 καταλοίπων, και κανονικοποιείται με τη βοήθεια διαφόρων περιορισμών που ασκούνται πάνω σε κάθε κατάλοιπο. Αυτό το προφίλ καταγράφεται στο αρχείο 'TvLDH.profile', το οποίο μπορεί να χρησιμοποιηθεί ως αρχείο εισόδου σε ένα πρόγραμμα γραφικών

Τόσο το DOPE όσο και το PROCHECK ΕΠΙΒΕΒΑΙΩΝΟΥΝ ΟΤΙ ΘΑ ΑΠΟΚΤΗΘΕΙ ΕΝΑ ΛΟΓΙΚΟ ΜΟΝΤΕΛΟ, ΜΕ ΜΙΑ ΕΝΕΡΓΕΙΑ ΣΥΓΚΡΙΣΙΜΗ ΜΕ ΕΚΕΙΝΗ ΤΟΥ ΠΡΟΤΥΠΟΥ (-38999.7) ΚΑΙ ΤΟΥ ΣΤΟΧΟΥ (-40407.7). όμως, το βαθμολογικό προφίλ DOPE δείχνει καθαρές διαφορές μεταξύ των δύο προφίλ όσον αφορά τον βρόγχο του ενεργού κέντρου

ΚΑΙ ΤΟΥ ΣΤΟΧΟΥ (-40407.7). όμως, το βαθμολογικό προφίλ DOPE δείχνει καθαρές διαφορές μεταξύ των δύο προφίλ όσον αφορά τον βρόγχο του ενεργού κέντρου μεταξύ των καταλοίπων 90 και 100 και των ελίκων στο καρβοξυτελικό άκρο της ακολουθίας στόχου.



Εικόνα 2.5.11: Η βαθμολόγηση DOPE.

Αυτός ο βρόγχος αλληλεπιδρά με τα κατάλοιπα 220-250, που σχηματίζουν το υπόλοιπο μισό του ενεργού κέντρου. Το δεύτερο αυτό κομμάτι είναι πολύ καλά προσδιορισμένο στο πρότυπο και πιθανώς σωστά σχεδιασμένο στη δομή του στόχου, αλλά εξαιτίας των μη επιθυμητών μη δεσμικών αλληλεπιδράσεων με τα κατάλοιπα 90-100 αναφέρεται επίσης ως υψηλής ενέργειας από το DOPE. Γενικά, ένα πιθανό λάθος που υποδεικνύεται από το DOPE μπορεί να μην είναι πραγματικά λάθος, ειδικά εάν αναφέρεται σε ενεργό κέντρο ή σε περιοχή αλληλεπίδρασης πρωτεϊνών. Όμως, στη συγκεκριμένη περίπτωση, οι ίδιοι βρόγχοι του ενεργού κέντρου έχουν ένα καλύτερο προφίλ στην πρότυπη δομή, γεγονός που ενισχύει τον ισχυρισμό ότι πιθανώς το μοντέλο είναι λανθασμένο στην περιοχή του ενεργού κέντρου.



**Πρακτικό
Μέρος**

3. Πρακτικό Μέρος

3.1 Μοντελοποίηση τμημάτων του παράγοντα von Willebrand

Σκοπός της εργασίας είναι η εύρεση της τριτοταγούς δομής των τμημάτων που συνδέουν τις τρεις επικράτειες του vWF και κατ' επέκταση η πρόβλεψη της τριτοταγούς δομής του τμήματος του vWF που περιλαμβάνει τις επικράτειες A1-A2-A3. Οι υπομονάδες A1 και A3 έχουν ήδη κρυσταλλογραφηθεί και επομένως οι δομές τους είναι γνωστές. Αντίθετα, η δομή της υπομονάδας A2 παραμένει άγνωστη, καθώς πειραματικά δεν έχει ακόμη προσδιοριστεί. Εντούτοις, υπάρχει μοντέλο για την A2 που έχει υπολογιστεί θεωρητικά [75].

Όσον αφορά την A1 υπομονάδα οι δομές που βρίσκονται καταχωρημένες στη βάση δεδομένων PDB είναι οι: 1OAK, 1AUQ, 1SQ0. Προκειμένου να επιλεγεί η καταλληλότερη συγκρίνεται το μήκος της ακολουθίας και η διακριτική ικανότητα (resolution). Επίσης, ελέγχεται η ύπαρξη κενών στη δομή. Προκρίνεται η δομή με τη μεγαλύτερη έκταση, τα λιγότερα ή καθόλου κενά και την καλύτερη ανάλυση. Συγκεκριμένα, για να επικυρωθεί σωστά η ακολουθία μιας εγγραφής PDB, πρέπει πρώτα να συγκριθεί η *προκύπτουσα ακολουθία* από τις εγγραφές ATOM με αυτή από τις εγγραφές SEQRES. Η σύγκριση των αλληλουχιών έγινε με τη βοήθεια του προγράμματος LALIGN. Από τα αποτελέσματα εξακριβώθηκε η ορθότητα και των τριών κρυσταλλογραφημένων δομών.

Η ίδια διαδικασία έγινε και για την υπομονάδα A3, για την οποία οι δομές που είναι κατατεθειμένες στη βάση δεδομένων PDB είναι δύο και συγκεκριμένα η 1AO3 και η 1ATZ. Εφόσον, πάλι εξακριβώθηκε με τη βοήθεια του προγράμματος LALIGN ότι δεν υπάρχουν κενά, ακολούθησε μελέτη των αρχείων ώστε να γίνει η επιλογή ενός. Και οι δύο προαναφερθείσες δομές περιέχουν μόνο την A3 υπομονάδα, με τη διαφορά ότι η 1ATZ την περιέχει δύο φορές.

Λαμβάνοντας υπόψη τα παραπάνω, αποφασίστηκε να χρησιμοποιηθεί η 1AUQ για την A1 επικράτεια και η 1ATZ για την A3. Η επιλογή των κατάλληλων δομών έπρεπε να γίνει, ώστε να οριοθετηθούν οι επικράτειες και επομένως και τα συνδεδεμένα τμήματα που μας ενδιαφέρουν.

Η δομή της επικράτειας A2 δεν έχει επιλυθεί πειραματικά και για αυτόν το λόγο χρησιμοποιήθηκε το θεωρητικά προσδιορισμένο μοντέλο. Το πρώτο συνδεδεμένο τμήμα οριοθετείται μεταξύ των καταλοίπων 1469 και 1495 του ώριμου vWF, ενώ το δεύτερο τμήμα μεταξύ των καταλοίπων 1670 και 1684. Όπως φαίνεται το πρώτο κομμάτι αποτελείται από 27 αμινοξέα, ενώ το δεύτερο από μόλις 15 αμινοξέα. Ο συνολικός αριθμός αμινοξέων ολόκληρης της πρωτεΐνης είναι 2813.

3.1.1 Εφαρμογή του Modeller

Αρχικά, κατεβλήθη προσπάθεια να προσδιοριστεί η δομή των συνδεδεμένων κομματιών μέσω συγκριτικού σχεδιασμού με τη βοήθεια του προγράμματος MODELLER9v4. Η διαδικασία περιγράφεται αναλυτικά. Το script με το οποίο το

πρόγραμμα εκτελεί την αναζήτηση δομών των οποίων οι ακολουθίες μοιάζουν με των αλληλουχία-στόχο, και η αλληλουχία-στόχος είναι πρώτα το πρώτο συνδετικό κομμάτι, ονομάζεται `build_profile.py` και είναι το εξής:

```
from modeller import *
log.verbose()
env = environ()

#-- Prepare the input files

#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdb95.fsa', seq_database_format='FASTA',
        chains_list='ALL', minmax_db_seq_len=(1, 40000), clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb95.bin', seq_database_format='BINARY',
        chains_list='ALL')

#-- Now, read in the binary database
sdb.read(seq_database_file='pdb95.bin', seq_database_format='BINARY',
        chains_list='ALL')

#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='link1/2_vWF.ali', alignment_format='PIR', align_codes='ALL')

#-- Convert the input sequence/alignment into
# profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',
        gap_penalties_1d=(-500, -50), n_prof_iterations=5,
        check_profile=False, max_aln_evalue=0.01, gaps_in_target=False)

#-- Write out the profile
prf.write(file='buildprofile.prf', profile_format='TEXT')

#-- Convert the profile back to alignment format
aln = prf.to_alignment()

#-- Write out the alignment file
aln.write(file='buildprofile.ali', alignment_format='PIR')
```

Εικόνα 3.1: Το script της εικόνας 2.5.2 μορφοποιημένο κατάλληλα στην πρωτεΐνη vWF.

Το παραπάνω script, που αφορά το πρώτο συνδετικό κομμάτι, είναι γραμμένο στη γλώσσα προγραμματισμού Python και εκτελείται στο λειτουργικό σύστημα Linux. Οι ίδιες ακριβώς ενέργειες γίνονται και στο λειτουργικό σύστημα Windows. Η εντολή που τρέχει το script είναι: `mod9v4 build_profile.py`.

Η απάντηση του παραπάνω script είναι ένα άλλο αρχείο, το περιεχόμενο του οποίου φαίνεται παρακάτω:

MODELLER 9v4, 2008/05/09, r6267

PROTEIN STRUCTURE MODELLING BY SATISFACTION OF SPATIAL RESTRAINTS

Copyright(c) 1989-2008 Andrej Sali
All Rights Reserved

Written by A. Sali
with help from

B. Webb, M.S. Madhusudhan, M-Y. Shen, M.A. Marti-Renom,
N. Eswar, F. Alber, M. Topf, B. Oliva, A. Fiser,
R. Sanchez, B. Yerkovich, A. Badretdinov,
F. Melo, J.P. Overington, E. Feyfant
University of California, San Francisco, USA
Rockefeller University, New York, USA
Harvard University, Cambridge, USA
Imperial Cancer Research Fund, London, UK
Birkbeck College, University of London, London, UK

Kind, OS, HostName, Kernel, Processor: 4, Linux suse 2.6.22.17-0.1-default i686
Date and time of compilation : 2008/05/09 13:05:53
MODELLER executable type : i386-intel8
Job starting time (YY/MM/DD HH:MM:SS): 2008/07/31 09:48:33

openf__224_> Open \$(LIB)/restyp.lib
openf__224_> Open \${MODINSTALL9v4}/modlib/resgrp.lib
rdresgr_266_> Number of residue groups: 2
openf__224_> Open \${MODINSTALL9v4}/modlib/sstruc.lib

Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3234070 3158.271 3.084
Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3234598 3158.787 3.085

openf__224_> Open \${MODINSTALL9v4}/modlib/resdih.lib
Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3283198 3206.248 3.131
rdrdih_263_> Number of dihedral angle types : 9

Maximal number of dihedral angle optima: 3

Dihedral angle names : Alph Phi Psi Omeg chi1 chi2 chi3 chi4 chi5

openf__224_> Open \${MODINSTALL9v4}/modlib/radii.lib
Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3292174 3215.014 3.140

openf__224_> Open \${MODINSTALL9v4}/modlib/radii14.lib
openf__224_> Open \${MODINSTALL9v4}/modlib/af_mnchdef.lib

rdwilmo_274_> Mainchain residue conformation classes: APBLE
openf__224_> Open \${MODINSTALL9v4}/modlib/mnch.lib

rdclass_257_> Number of classes: 5
openf__224_> Open \${MODINSTALL9v4}/modlib/mnch1.lib

openf__224_> Open \${MODINSTALL9v4}/modlib/mnch2.lib
openf__224_> Open \${MODINSTALL9v4}/modlib/mnch3.lib

openf__224_> Open \${MODINSTALL9v4}/modlib/xs4.mat
rdrrwgh_268_> Number of residue types: 21

openf__224_> Open pdb95.fsa
Dynamically allocated memory at amaxsequence_db [B,KiB,MiB]: 3584463 3500.452 3.418

openf__224_> Open pdb95.fsa
openf__224_> Open pdb95.fsa

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 3586281 3502.228 3.420

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 3587731 3503.644 3.422

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 3590631 3506.476 3.424

.....
SEQ_DATABASE_FILE : pdb95.fsa
SEQ_DATABASE_FORMAT : FASTA
CHAINS_LIST : ALL
CLEAN_SEQUENCES : T
MINMAX_DB_SEQ_LEN : 1 40000
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

SEQ_DATABASE_FILE : pdb95.bin
SEQ_DATABASE_FORMAT : BINARY
SEARCH_CHAINS_LIST : ALL
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

SEQ_DATABASE_FILE : pdb95.bin
SEQ_DATABASE_FORMAT : BINARY
CHAINS_LIST : ALL
CLEAN_SEQUENCES : T
MINMAX_DB_SEQ_LEN : 0 999999
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

openf__224_> Open link1_vWF.ali

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 5238836 5116.051 4.996
Dynamically allocated memory at amaxsequence [B,KiB,MiB]: 5238964 5116.176 4.996

Read the alignment from file : link1_vWF.ali
Total number of alignment positions: 33

# Code	#_Res	#_Segm	PDB_code	Name
--------	-------	--------	----------	------

1	link1_vWF	33	1	vWF
---	-----------	----	---	-----

Dynamically allocated memory at amaxprofile [B,KiB,MiB]: 5239397 5116.599 4.997

openf__224_> Open \${LIB}/blosum62.sim.mat

rdrwrgh_268_> Number of residue types: 21

profile_iteration_> processing sequence: 1 33 1 0.000000 0.000000 0.000000 1

profile_iteration_> processing sequence: 552 33 1 0.0800000 0.0001449 0.0000006 552

profile_iteration_> processing sequence: 1103 33 1 0.1700000 0.0001541 0.0000007 1103

AVERAGE RESIDUAL STANDARD ERROR: 24.12477

AVERAGE RESIDUAL STANDARD DEVIATION: 183.1201

profile_660_> Quality of the fit between observed/expected score distributions.

[Iteration, Chi2, nbins, KS-Stat, nscores] = 1, 0.1788, 121, 0.0226, 5508

HISTOGRAM OF Z-SCORES:

Z: 1 -3.95000 0.00000 0.00000

Z: 1 -3.85000 0.00000 0.00000

Z: 1 -3.75000 0.00000 0.00000

Z: 1 -3.65000 0.00000 0.00000

HITS FOUND IN ITERATION: 1

profile_iteration_> Weight Matrix [sec]: 0.000000

profile_iteration_> Score Collection [sec]: 0.870000

profile_iteration_> Score Collection/Seq [sec]: 0.000158

profile_iteration_> DP Steps per sec [ali/sec]: 6404.651367

profile_iteration_> Statistics [sec]: 0.010000

profile_iteration_> Add to profile [sec]: 0.000000

profile_iteration_> Overall performance [/seq]: 0.000162

profile_iteration_> Overall performance [seq/sec]: 6188.764160

openf__224_> Open buildprofile.prf

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 5240878 5118.045 4.998

Dynamically allocated memory at amaxsequence [B,KiB,MiB]: 5241006 5118.170 4.998

openf__224_> Open buildprofile.ali

```

Dynamically allocated memory at finish [B,KiB,MiB]: 5238904 5116.117 4.996
Starting time : 2008/07/31 09:48:33
Closing time : 2008/07/31 09:48:35
Total CPU time [seconds] : 1.83

```

Εικόνα 3.2: Το αρχείο προέκυψε αφού τρέξαμε την πρώτη εντολή. Σύμφωνα με την εικόνα 2.5.3, δεν βρέθηκε κανένα πρότυπο.

Η διαδικασία είναι ίδια και για το δεύτερο συνδεδετικό κομμάτι, η απάντηση για το οποίο είναι η ακόλουθη:

MODELLER 9v4, 2008/05/09, r6267

PROTEIN STRUCTURE MODELLING BY SATISFACTION OF SPATIAL RESTRAINTS

Job starting time (YY/MM/DD HH:MM:SS): 2008/07/31 09:52:39

```

openf__224_> Open $(LIB)/restyp.lib
openf__224_> Open ${MODINSTALL9v4}/modlib/resgrp.lib
rdresgr_266_> Number of residue groups: 2
openf__224_> Open ${MODINSTALL9v4}/modlib/sstruc.lib

```

Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3234070 3158.271 3.084

Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3234598 3158.787 3.085

```
openf__224_> Open ${MODINSTALL9v4}/modlib/resdih.lib
```

Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3283198 3206.248 3.131

```
rdrdih_263_> Number of dihedral angle types : 9
```

```
Maximal number of dihedral angle optima: 3
```

```
Dihedral angle names : Alph Phi Psi Omeg chi1 chi2 chi3 chi4 chi5
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/radii.lib
```

Dynamically allocated memory at amaxlibraries [B,KiB,MiB]: 3292174 3215.014 3.140

```
openf__224_> Open ${MODINSTALL9v4}/modlib/radii14.lib
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/af_mnchdef.lib
```

```
rdwilmo_274_> Mainchain residue conformation classes: APBLE
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/mnch.lib
```

```
rdclass_257_> Number of classes: 5
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/mnch1.lib
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/mnch2.lib
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/mnch3.lib
```

```
openf__224_> Open ${MODINSTALL9v4}/modlib/xs4.mat
```

```
rdrrwgh_268_> Number of residue types: 21
```

```
openf__224_> Open pdb95.fsa
```

Dynamically allocated memory at amaxsequence_db [B,KiB,MiB]: 3584463 3500.452 3.418

```
openf__224_> Open pdb95.fsa
```

```
openf__224_> Open pdb95.fsa
```

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 3586281 3502.228 3.420

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 3587731 3503.644 3.422

```

.....
SEQ_DATABASE_FILE : pdb95.fsa
SEQ_DATABASE_FORMAT : FASTA
CHAINS_LIST : ALL
CLEAN_SEQUENCES : T
MINMAX_DB_SEQ_LEN : 1 40000
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

```


SEQ_DATABASE_FILE : pdb95.bin
SEQ_DATABASE_FORMAT : BINARY
SEARCH_CHAINS_LIST : ALL
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

SEQ_DATABASE_FILE : pdb95.bin
SEQ_DATABASE_FORMAT : BINARY
CHAINS_LIST : ALL
CLEAN_SEQUENCES : T
MINMAX_DB_SEQ_LEN : 0 999999
Number of sequences : 5508
Number of residues : 1268520
Length of longest sequence: 1287

openf__224_> Open link2_vWF.ali

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 5238836 5116.051 4.996

Dynamically allocated memory at amaxsequence [B,KiB,MiB]: 5238952 5116.164 4.996

Read the alignment from file : link2_vWF.ali
Total number of alignment positions: 30

# Code	#_Res	#_Segm	PDB_code	Name
--------	-------	--------	----------	------

1	link1_vWF	30	1	vWF
---	-----------	----	---	-----

Dynamically allocated memory at amaxprofile [B,KiB,MiB]: 5239373 5116.575 4.997

openf__224_> Open \${LIB}/blosum62.sim.mat

rdrwrgh_268_> Number of residue types: 21

profile_iteration_> processing sequence: 1 30 1 0.0000000 0.0000000 0.0000000 1

profile_iteration_> processing sequence: 552 30 1 0.0800000 0.0001449 0.0000006 552

profile_iteration_> processing sequence: 1103 30 1 0.1600000 0.0001451 0.0000006 1103

AVERAGE RESIDUAL STANDARD ERROR: 35.02277

AVERAGE RESIDUAL STANDARD DEVIATION: 190.4418

profile_660_> Quality of the fit between observed/expected score distributions.

[Iteration, Chi2, nbins, KS-Stat, nscores]= 1, 0.1692, 121, 0.0218, 5508

HISTOGRAM OF Z-SCORES:

Z: 1 -3.95000 0.00000 0.00000

Z: 1 -3.85000 0.00000 0.00000

Z: 1 -3.75000 0.00000 0.00000

HITS FOUND IN ITERATION: 1

profile_iteration_> Weight Matrix [sec]: 0.000000

profile_iteration_> Score Collection [sec]: 0.800000

profile_iteration_> Score Collection/Seq [sec]: 0.000145

profile_iteration_> DP Steps per sec [ali/sec]: 6885.000000

profile_iteration_> Statistics [sec]: 0.000000

profile_iteration_> Add to profile [sec]: 0.000000

profile_iteration_> Overall performance [/seq]: 0.000147

profile_iteration_> Overall performance [seq/sec]: 6800.000000

openf__224_> Open buildprofile.prf

Dynamically allocated memory at amaxalignment [B,KiB,MiB]: 5240767 5117.937 4.998

Dynamically allocated memory at amaxsequence [B,KiB,MiB]: 5240883 5118.050 4.998

openf__224_> Open buildprofile.ali

Dynamically allocated memory at	finish [B,KiB,MiB]:	5238793	5116.009	4.996
Starting time	:	2008/07/31 09:52:39		
Closing time	:	2008/07/31 09:52:41		
Total CPU time [seconds]	:	1.74		

Εικόνα 3.3: Το αρχείο μας πληροφορεί πως δε βρέθηκε κανένα πρότυπο.

Όπως φαίνεται παραπάνω από τα scripts, δεν βρέθηκε κανένα πρότυπο σύμφωνα με την εικόνα 2.5.3, ώστε να υιοθετηθεί η δομή του και να γίνει η πρόβλεψη της δομής με βάση την ομολογία.

3.1.2 Εφαρμογή της μεθόδου αναγνώρισης αναδίπλωσης

Εφόσον, λοιπόν, ο προσδιορισμός με το συγκριτικό σχεδιασμό δεν ήταν δυνατός για κανένα από τα δύο κομμάτια, χρησιμοποιήθηκε η μέθοδος της αναγνώρισης αναδίπλωσης (threading). Για αυτόν το σκοπό επιλέχθηκε ο εξυπηρέτης E-mail (server) Phyre από το διαδίκτυο, ο οποίος υποστηρίζει το threading ή fold recognition, όπως αποκαλείται επίσης η μέθοδος.

Η διαδικασία είναι η εξής: στέλνεται η ακολουθία της οποίας η δομή είναι άγνωστη και το αποτέλεσμα της αναζήτησης λαμβάνεται με μήνυμα ηλεκτρονικού ταχυδρομείου. Στέλνεται δηλαδή στο χρήστη ένας φάκελος με όλα τα πιθανά πρότυπα από τα οποία επιλέγεται πάλι ένα μόνο. Οι δομές διαφέρουν ως προς την έκταση της ομοπαράθεσης και ως προς το βαθμό αξιοπιστίας της πρόβλεψης.

Η αλληλουχία που επιλέχθηκε ως πρότυπο για το σχεδιασμό του πρώτου συνδετικού τμήματος είναι η d1dcla1, η ομοπαράθεση της οποίας παρατίθεται στο παράρτημα. Συγκεκριμένα, τα αμινοξέα 3 – 36 του προτύπου στοιχήθηκαν με τα αμινοξέα της αλληλουχίας-στόχου και υιοθετήθηκε η δομή του προτύπου. Στη συνέχεια, η πρότυπη δομή εξισορροπήθηκε με τη βοήθεια προσομοιώσεων μοριακής δυναμικής και η εξισορροπημένη δομή χρησιμοποιήθηκε τελικά.

Όσον αφορά το δεύτερο κομμάτι, η διαδικασία που ακολουθήθηκε είναι η ίδια. Το πρότυπο που αντιστοιχεί στο δεύτερο συνδετικό κομμάτι είναι η ακολουθία d1nlta3. Τα κατάλοιπα 5-25 του προτύπου ομοπαράτεθηκαν με τα κατάλοιπα της πρωτεΐνης στόχου. Η συγκεκριμένη ομοπαράθεση, επίσης, παρατίθεται στο παράρτημα.

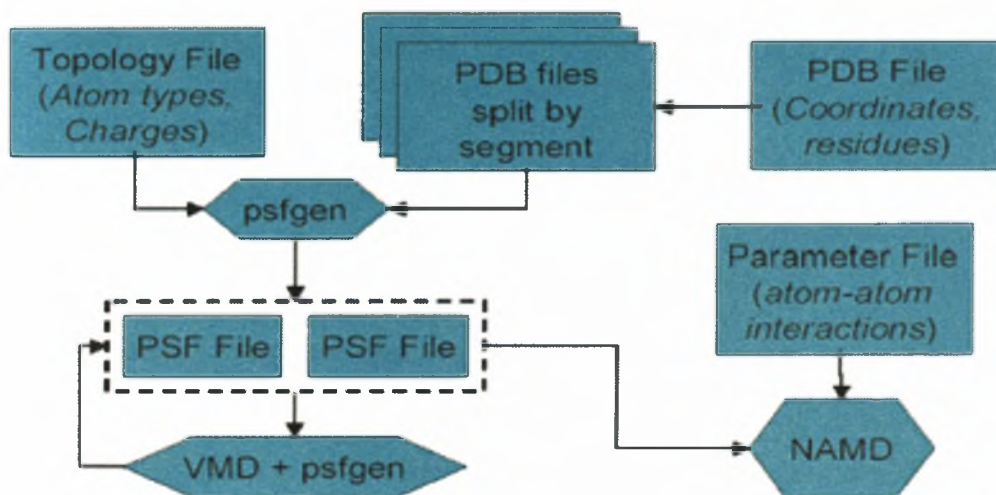
Το πρώτο μέρος της πειραματικής διαδικασίας τελείωσε με την εύρεση των προτύπων και την πρόβλεψη της δομής των δύο συνδετικών κομματιών. Το δεύτερο μέρος αποσκοπεί στην εξισορρόπηση των παραπάνω δομών, ώστε να βελτιωθεί η δομή. Για να επιτευχθεί ο παραπάνω σκοπός, πολλές προσομοιώσεις μοριακής δυναμικής έλαβαν χώρα μέχρι το σύστημα να φθάσει στην κατά το δυνατόν 'πλήρη' εξισορρόπηση. Οι προσομοιώσεις εκτελέστηκαν από δύο προγράμματα, το NAMM [72] και το VMD [73].

3.2 Προσομοίωση των θεωρητικά προβλεπόμενων δομών

3.2.1 Προϋποθέσεις λειτουργίας του προγράμματος NAMD

Το NAMD για να λειτουργήσει χρειάζεται τουλάχιστον τέσσερα πράγματα:

- Ένα αρχείο *PBD*, ένα αρχείο δηλαδή που αποθηκεύει, μεταξύ άλλων, τις συντεταγμένες των ατόμων ή/και τις ταχύτητες του συστήματος.
- Ένα αρχείο *PSF*, το οποίο αποθηκεύει δομικές πληροφορίες της πρωτεΐνης, όπως διάφορους τύπους αλληλεπιδράσεων δεσμών, γωνίες, φορτία, μάζες.
- Ένα αρχείο με τις παραμέτρους της συνάρτησης δυναμικού. Το force field είναι η μαθηματική έκφραση του δυναμικού (potential) των ατόμων και των αλληλεπιδράσεών τους στο σύστημα. CHARM, X-PLOR, AMBER, και GROMACS είναι τέσσερις τύποι πεδίων δυνάμεων και το NAMD μπορεί να χρησιμοποιήσει και τους τέσσερις. Στην παρούσα εργασία χρησιμοποιήθηκε η συνάρτηση δυναμικού CHARM. Οι παράμετροι υπαγορεύουν την ισχύ των δεσμών (bond strengths), τα μήκη ισορροπίας (equilibrium lengths) κτλ.
- Ένα αρχείο *configuration*, στο οποίο ο χρήστης καθορίζει όλες τις επιλογές που το πρόγραμμα θα έπρεπε να υιοθετήσει κατά τη διάρκεια μιας προσομοίωσης. Το συγκεκριμένο αρχείο πληροφορεί το πρόγραμμα για το πώς η προσομοίωση πρέπει να τρέξει.



Εικόνα 3.4: Σχηματική απεικόνιση του πρώτου βήματος

3.2.2 Τα 4 βήματα που προηγούνται των προσομοιώσεων

1. Παραγωγή του αρχείου PSF (Protein Structure File)

Από τους τέσσερις φακέλους που αναφέρθηκαν παραπάνω, ένα αρχικό αρχείο PBD μπορεί να αποκτηθεί από την πρωτεϊνική βάση δεδομένων και τα αρχεία με τις παραμέτρους και την τοπολογία μπορούν να αποκτηθούν από το διαδίκτυο από την ιστοσελίδα

<http://www.pharmacy.umaryland.edu/~alex/research.html>. Το τέταρτο με τη σειρά της παραπάνω αναφοράς αρχείο του προγράμματος (αρχείο configuration) δημιουργείται από το χρήστη. Το αρχείο PSF πρέπει να δημιουργηθεί από το χρήστη από το αρχικό αρχείο PBD και τα αρχεία που περιέχουν τις παραμέτρους.

Πρώτα, πρέπει από ένα αρχείο PBD να αφαιρεθούν τα μόρια του νερού και να δημιουργηθεί ένα καινούριο αρχείο PBD που θα περιέχει μόνο την πρωτεΐνη. Στη συνέχεια, εκκινείται το πρόγραμμα VMD. Στην οθόνη που δίνονται οι εντολές πρέπει με την εντολή `cd` να μετακινηθούμε στον κατάλογο που υπάρχει το αρχείο `pdb`. Γράφουμε:

```
( vmd > ) cd C:/Users/user/Desktop/όνομα φακέλου.
```

Για να τρέξει χρειάζεται τα αρχεία `par_all27_prot_na.inp`, `top_all27_prot_na.inp`, `psfgen` και το script `psf_όνομα αρχείου PBD.txt`.

Ακολούθως, από το μενού του προγράμματος φορτώνεται το παραπάνω αρχείο PBD με `File, New molecule`, καθορισμός τύπου αρχείου: `PBD`, `Browse` (ώστε να βρεθεί το αρχείο μέσα από τους καταλόγους), επιλογή του επιθυμητού αρχείου PBD και φόρτωμα (`load`).

Πρέπει να σημειωθεί το γεγονός ότι η κρυσταλλογραφική δομή που είναι κατατεθειμένη στην PBD δεν περιέχει τα άτομα υδρογόνου της πρωτεΐνης. Αυτό συμβαίνει γιατί η κρυσταλλογραφία ακτίνων-Χ συνήθως δε μπορεί να αναλύσει / ξεκαθαρίσει τα άτομα υδρογόνου. Το αρχείο PBD που περιέχει μόνο την πρωτεΐνη μαζί με το `psf` θα περιέχει συντεταγμένες για τα υδρογονικά άτομα που έχουν προκύψει μετά από υπόθεση. Αργότερα, η ελαχιστοποίηση ενέργειας της πρωτεΐνης θα εξασφαλίσει τις θέσεις που είναι ενεργειακά εύλογες.

Στη συνέχεια, στην κονσόλα VMD main επιλέγουμε `Extentions, Tk Console` και στην κονσόλα που μόλις εμφανίστηκε πληκτρολογούμε `play <script psf>`. Προκύπτουν δύο αρχεία με καταλήξεις `.pdb` και `.psf`.

2. Διαλυτοποίηση της ουσίας

Τώρα η πρωτεΐνη πρέπει να διαλυτοποιηθεί, ώστε να προσομοιαστεί καλύτερα το κυτταρικό περιβάλλον. Αυτό θα γίνει σε δύο βήματα τοποθετώντας την πρωτεΐνη σε μία σφαίρα νερού, ώστε να προετοιμαστεί για ελαχιστοποίηση και εξισορρόπηση χωρίς περιοδικές οριακές συνθήκες.

Διαμορφώνουμε το script `solvate_όνομα πρωτεΐνης.txt` κάνοντας τις εξής μεταβολές:

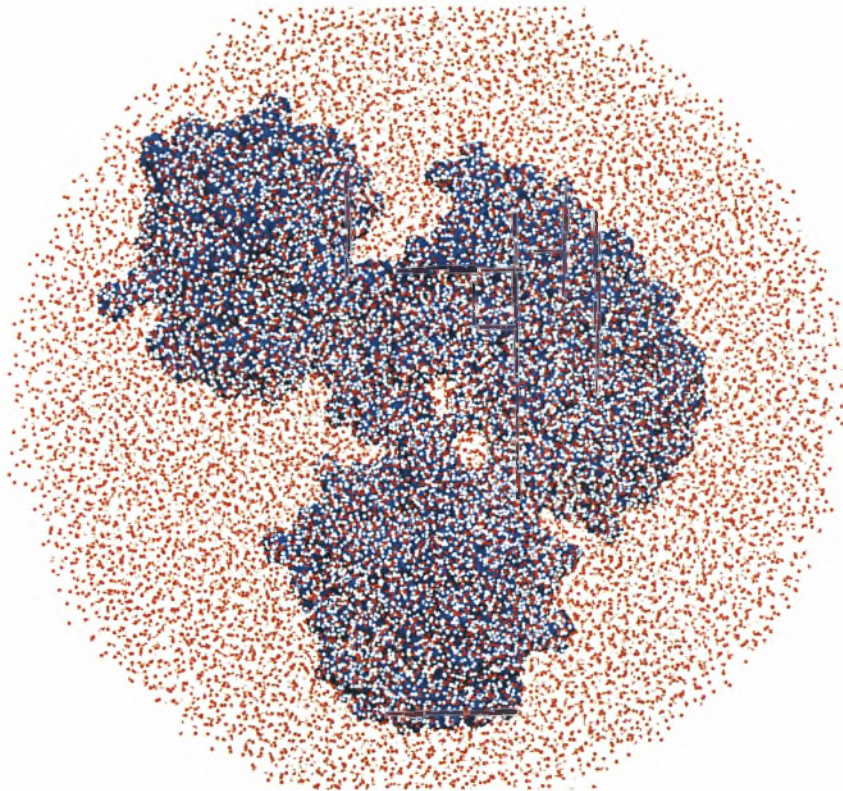
```
.....  
set psf      1auq.psf  
set pdb      1auq.pdb  
set box      1auq_box  
set psfDrop  1auq_ws.psf  
set pdbDrop  1auq_ws.pdb  
set all 6  
.....
```

.Εικόνα 3.5: Απόσπασμα του script `solvate`.

Τα ονόματα των παραπάνω αρχείων κάθε φορά διαμορφώνονται κατάλληλα. Στο τέλος αυτού του βήματος, τα δύο τελευταία αρχεία θα προκύψουν. Μετά τη φράση `set all` πληκτρολογείται ένας αριθμός που δηλώνει την επιθυμητή απόσταση του πιο ακραίου μέρους της πρωτεΐνης από την επιφάνεια της υδάτινης σφαίρας. Θα πρέπει, όμως, να λαμβάνεται υπόψη το γεγονός ότι όσο πιο μεγάλη είναι η ακτίνα, τόσο πιο πολλά μόρια νερού προστίθενται και τόσο περισσότερο χρόνο επομένως η προσομοίωση απαιτεί να τρέξει.

Τα βήματα είναι τα ακόλουθα: στο VMD φορτώνουμε το αρχείο PDB που προέκυψε από το πρόγραμμα `psfgen`. Στη συνέχεια, επιλέγουμε `Extensions, Tk console`. Στην οθόνη που εμφανίζεται πρώτα πληκτρολογούμε την εντολή `cd C:/όνομα φακέλου μέσα από τον οποίο θα τρέξει το solvate`, μετά `enter` και τέλος `play solvate_όνομα.txt`. Οι δύο τελευταίες σειρές που εμφανίζονται στην Tk console αποθηκεύονται για να χρησιμοποιηθούν αργότερα. Δίνουν τις συντεταγμένες του κέντρου της υδάτινης σφαίρας και την ακτίνα της.

Αφού τρέξει το πρόγραμμα προκύπτουν ένα αρχείο με την κατάληξη `.pdb` και ένα με την κατάληξη `.psf`, τα οποία αναφέρονται στη διαλυτοποιημένη πλέον πρωτεΐνη όπως ορίστηκαν στο script `solvate`.



Εικόνα 3.6: Η διαλυτοποιημένη πρωτεΐνη. Με μπλε απεικονίζεται η πρωτεΐνη και με κόκκινο τα μόρια του νερού.

3. Προσθήκη ιόντων

Ιόντα θα πρέπει να προστεθούν στο νερό για να εξασφαλιστεί η ηλεκτρική ουδετερότητα του συστήματος. Είναι ιδιαίτερα απαραίτητα όταν η πρωτεΐνη που μελετάται φέρει σημαντικό ηλεκτρικό φορτίο. Σε μια τέτοια περίπτωση, ο αριθμός των ιόντων θα επιλεγεί έτσι ώστε το σύστημα μετά την προσθήκη τους να καθίσταται ουδέτερο. Τα ιόντα αυτά θα προστατεύουν τις περιοχές της πρωτεΐνης που φέρουν φορτίο και θα κάνουν όλο το σύστημα πιο σταθερό. Το αρχείο `psf` περιέχει το φορτίο κάθε ατόμου και μπορεί να χρησιμοποιηθεί για να καθοριστεί το φορτίο ολόκληρου του συστήματος ή μέρος αυτού.

Προκειμένου να τρέξει αυτό το βήμα, χρειάζεται το αρχείο `math.tcl`. Από την κονσόλα του προγράμματος VMD, κι ενώ έχουμε καθορίσει ως τύπο αρχείου `automatically`, φορτώνουμε πρώτα το αρχείο με κατάληξη `ws.psf` και μετά επιλέγουμε αυτό το αρχείο στην ίδια κονσόλα και με `File, Load Data Into Molecule` φορτώνουμε και το αρχείο με κατάληξη `ws.pdb`.

Στο φάκελο που βρίσκονται όλα τα παραπάνω αρχεία εισάγουμε και το αρχείο `addions_όνομα.txt`, του οποίου η τελευταία σειρά παρατίθεται παρακάτω:

```
eval addions neutr_1auq_ws.psf neutr_1auq_ws.pdb.
```

Στη συγκεκριμένη σειρά καθορίζονται τα ονόματα των αρχείων που θα εξαχθούν και θα αναφέρονται πλέον στο ουδέτερο σύστημα. Στην Tk console, αφού προσέξουμε ότι είμαστε μέσα στο σωστό φάκελο, πληκτρολογούμε `play addions_όνομα.txt`, και

προκύπτουν τα δύο αρχεία που αναφέρθηκαν στο παραπάνω script με την ονομασία additions_όνομα.txt.

4. Ελαχιστοποίηση δυναμικής ενέργειας

Το συγκεκριμένο στάδιο προετοιμάζει το σύστημα για την προσομοίωση. Ο σκοπός αυτού του βήματος δεν είναι να βρεθεί ένα πραγματικά ολικό ενεργειακό ελάχιστο, αλλά να προσαρμοστεί η δομή στη συνάρτηση δυναμικού, να κατανεμηθούν τα μόρια του διαλύτη και να αμβλύνει πιθανές στερικές συγκρούσεις, οι οποίες δημιουργούνται από τις συντεταγμένες των ατόμων κατά τη διάρκεια δημιουργίας του αρχικού αρχείου συντεταγμένων. Τρεις τύποι αρχείων δημιουργούνται από κάθε προσομοίωση που εκτελείται από το πρόγραμμα NAMD:

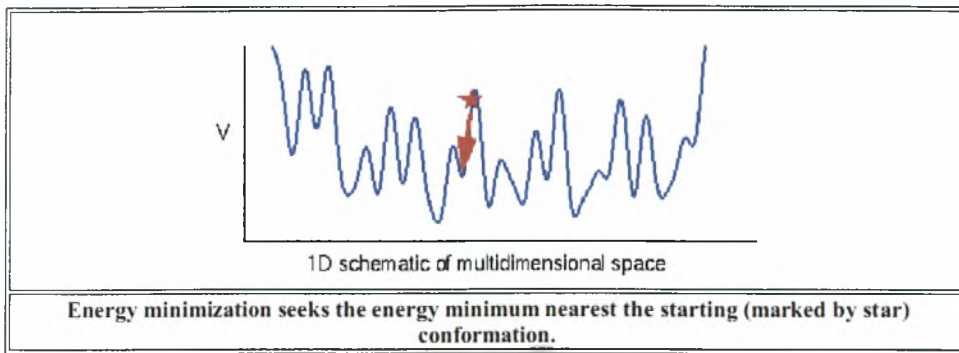
- 1) Οι τελικές συντεταγμένες και ταχύτητες σώζονται σε ένα αρχείο PDB ή σε δυαδική μορφή (οι καταλήξεις των αρχείων είναι coor και vel αντίστοιχα).
- 2) Οι συντεταγμένες επανεκκίνησης και οι ταχύτητες επανεκκίνησης αποθηκεύονται σε ένα αρχείο PDB ή σε δυαδική μορφή που σώζονται περιοδικά και χρησιμοποιούνται στην επανεκκίνηση των προσομοιώσεων σε περίπτωση μη κανονικού τερματισμού (οι καταλήξεις των αρχείων είναι coor και vel αντίστοιχα).
- 3) Ένα αρχείο τροχιάς (η κατάληξή του είναι dcd), το οποίο περιέχει τις περιοδικά αποθηκευμένες συντεταγμένες όλων των ατόμων.

Για το συγκεκριμένο βήμα απαιτείται το script min_ion_όνομα.txt και το πρόγραμμα namd2.exe στον ίδιο φάκελο με όλα τα προηγούμενα αρχεία. Αφού μετακινηθούμε στο φάκελο μέσα από τον οποίο τρέχουμε τα scripts, πληκτρολογείται η εξής εντολή: namd2.exe +p2 min_όνομα αρχείου.txt > log_min_όνομα αρχείου.txt.

```
Info: NONZERO IMPRECISION IN COULOMB...
)
Info: Entering startup phase 8 with 35208 kB of memory in use.
Info: Finished startup with 35208 kB of memory in use.
TCL: Minimizing for 100 steps
PRESSURE: 0 0 0 0 0 0 0 0 0 0
CPRESSURE: 0 0 0 0 0 0 0 0 0 0
ETITLE:      TS      BOND      ANGLE      DIHED      IMPRP  \
            ELECT      VDW      BOUNDARY    MISC      KINETIC \
            TOTAL      TEMP      TOTAL2     TOTAL3    TEMPAV\
G
ENERGY:      0      1913.4792      3008.0536      358.3486      10.7302 \
            -18281.1014      5217.3380      0.0000      0.0000      0.0000 \
            -7773.1518      0.0000      -7773.1518      -7773.1518      0.0000\
0
INITIAL STEP: 1e-06
GRADIENT TOLERANCE: 15332.5
PRESSURE: 1 0 0 0 0 0 0 0 0 0
CPRESSURE: 1 0 0 0 0 0 0 0 0 0
```

Εικόνα 3.7: Τυπικό αρχείο εξόδου log της ελαχιστοποίησης.

Ο στόχος αυτού του βήματος είναι η τοπική εύρεση της ελάχιστης δυναμικής ενέργειας (εικόνα 30). Μία ελαχιστοποίηση θεωρείται επιτυχής όταν ο αριθμός μετά τον όρο gradient tolerance γίνει μονοψήφιος αριθμός (εικόνα 3.6). Στη δυναμική ενέργεια του συστήματος συνεισφέρουν ενδομοριακές και διαμοριακές αλληλεπιδράσεις μεταξύ όλων των ατόμων, των αμινοξέων και μορίων νερού, δηλαδή ηλεκτροστατικές δυνάμεις και δυνάμεις van der Waals.



Εικόνα 3.8: Κατά τη διάρκεια της ελαχιστοποίησης αναζητείται το ενεργειακό ελάχιστο που βρίσκεται πιο κοντά στην αρχική στερεοδιαμόρφωση.

3.2.3 Εξισορρόπηση: το στάδιο όπου λαμβάνει χώρα η προσομοίωση

Το στάδιο αυτό χρησιμοποιείται προκειμένου να εξισορροπηθεί η κινητική και δυναμική ενέργεια του συστήματος. Για κάθε άτομο επιλύονται οι νόμοι του Νεύτωνα, προκειμένου να καθοριστούν οι τροχιές των ατόμων. Το σύστημα αφήνεται να εξελιχθεί, μέχρι να αποκτήσει την ελάχιστη δυναμική ενέργεια κάτω από συγκεκριμένες συνθήκες και όχι την ελάχιστη δυνατή. Οι συνθήκες, καθώς και το πλήθος των ατόμων κατά τη διάρκεια της εξισορρόπησης παραμένουν σταθερές.

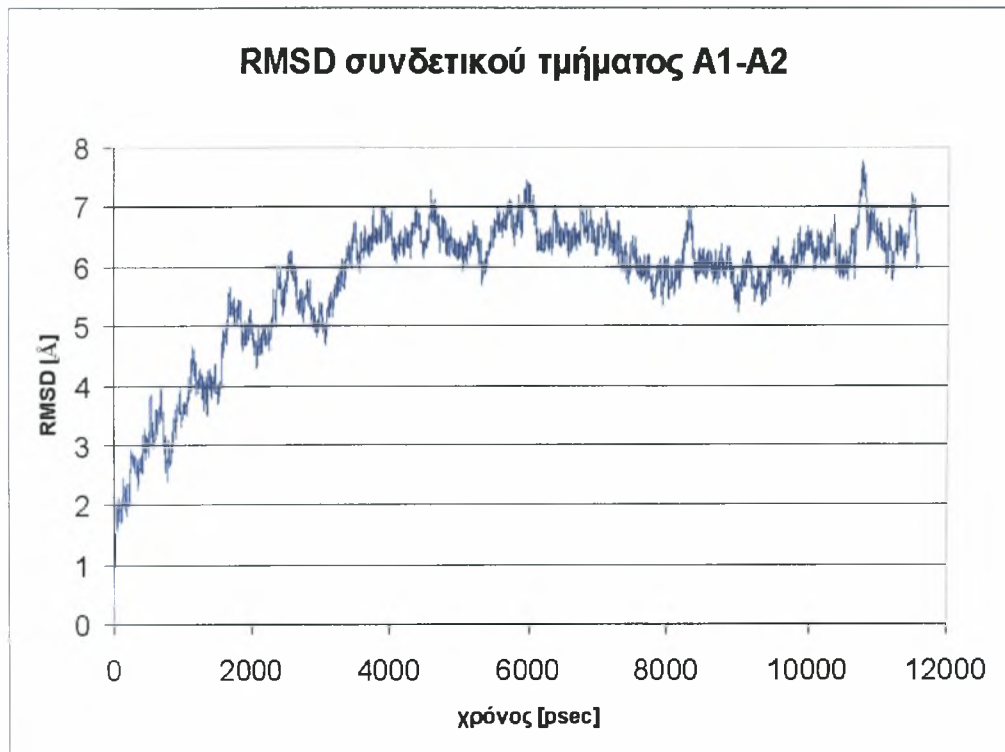


Αποτελέσματα

4. Αποτελέσματα

4.1 Η δομή του πρώτου συνδετικού τμήματος

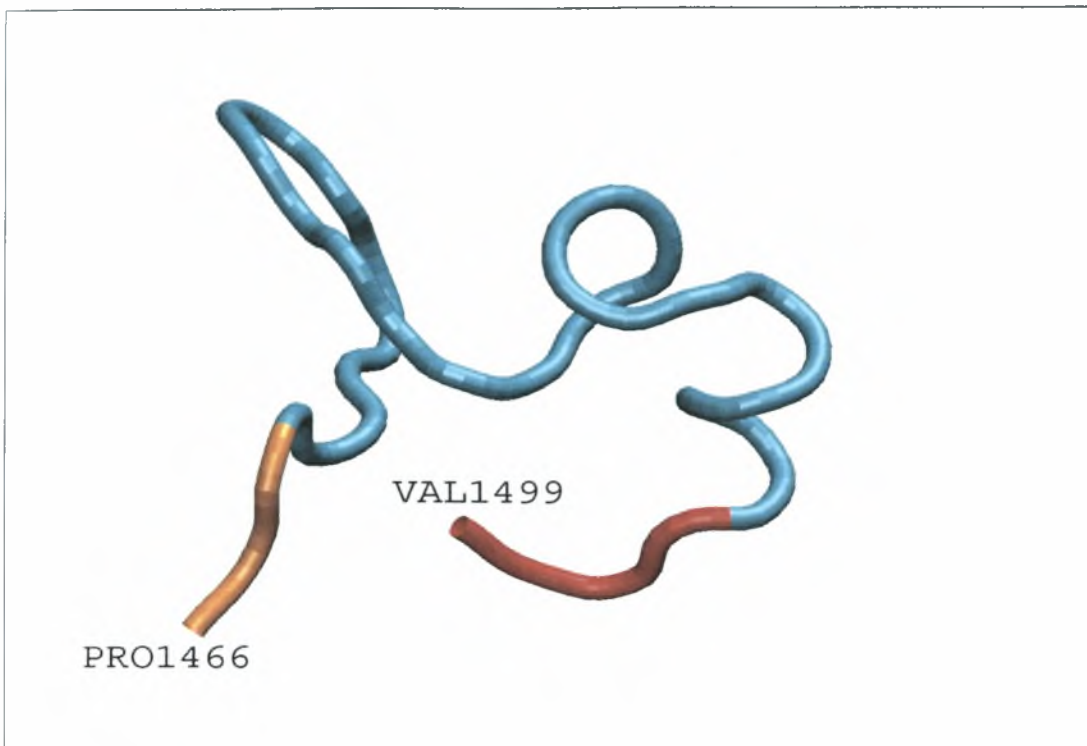
Οι προσομοιώσεις Μοριακής Δυναμικής συνεχίστηκαν μέχρι το σύστημα να σταθεροποιηθεί. Χρειάστηκαν προσομοιώσεις συνολικής διάρκειας 14,4 psec. Μία προσομοίωση θεωρείται επιτυχής όταν το RMSD (τετραγωνική ρίζα της μέσης τετραγωνικής απόκλισης) παύει να μεταβάλλεται συστηματικά και εμφανίζει μόνο στατιστικές διακυμάνσεις. Το RMSD του πρώτου συνδετικού τμήματος παρατίθεται στην εικόνα 4.1:



Εικόνα 4.1: Το διάγραμμα απεικονίζει το RMSD του τμήματος A1-A2

Κατά τη διάρκεια των πρώτων 6000 psec το RMSD συνεχώς αυξάνει. Στο τέλος, όμως, οι μεταβολές είναι μόλις της τάξεως του 1 Å.

Η δομή του πρώτου συνδετικού τμήματος του vWF που προέκυψε μετά από τις προσομοιώσεις φαίνεται στην εικόνα 4.2:



Εικόνα 4.2: Η δομή του σκελετού του πρώτου συνδετικού τμήματος.

Το αμινοτελικό άκρο, το οποίο χρωματίστηκε πορτοκαλί, περιλαμβάνει τα τρία τελευταία κατάλοιπα της A1 επικράτειας, ενώ το καρβοξυτελικό, το οποίο χρωματίστηκε κόκκινο, περιλαμβάνει τα τέσσερα πρώτα κατάλοιπα της A2 επικράτειας. Τα κατάλοιπα των επικρατειών συμπεριελήφθησαν στο συνδετικό τμήμα, γιατί χρησιμεύουν στην σωστή σύνδεση των συνδετικών τμημάτων με τις επικράτειες. Το αμινοτελικό άκρο δεν εμφανίζει μία συγκεκριμένη δευτεροταγή δομή. Τα αμέσως επόμενα τέσσερα κατάλοιπα (1469-1472) είναι υποψήφια για μία β-πτύχωση. Τα κατάλοιπα 1473-1485 δεν σχηματίζουν συγκεκριμένο τύπο δευτεροταγούς δομής. Τα αμέσως δύο επόμενα κατάλοιπα (1486-1487) είναι, επίσης, υποψήφια να σχηματίζουν β-πτύχωση. Αντίθετα, τα κατάλοιπα 1488-1490 δεν σχηματίζουν κάποια συγκεκριμένη δευτεροταγή δομή. Στη συνέχεια, τα κατάλοιπα 1491-1495 σχηματίζουν μία α-έλικα. Τέλος, τα αμέσως δύο επόμενα κατάλοιπα (1496-1497) δεν σχηματίζουν συγκεκριμένο τύπο δευτεροταγούς δομής, ενώ τα δύο τελευταία κατάλοιπα 1498-1499 μπορεί να μετέχουν σε μία β-πτύχωση η οποία συνεχίζει μέχρι το κατάλοιπο 1505, αλλά δε φαίνεται στην εικόνα καθώς περιλαμβάνεται στη δομή της A2 επικράτειας. Τα αποτελέσματα αυτά συμφωνούν με πειράματα κυκλικού διχρωϊσμού, όπου μελετήθηκε η ξεδίπλωση του A1-A2-A3 τμήματος με τη βοήθεια ουρίας [73].

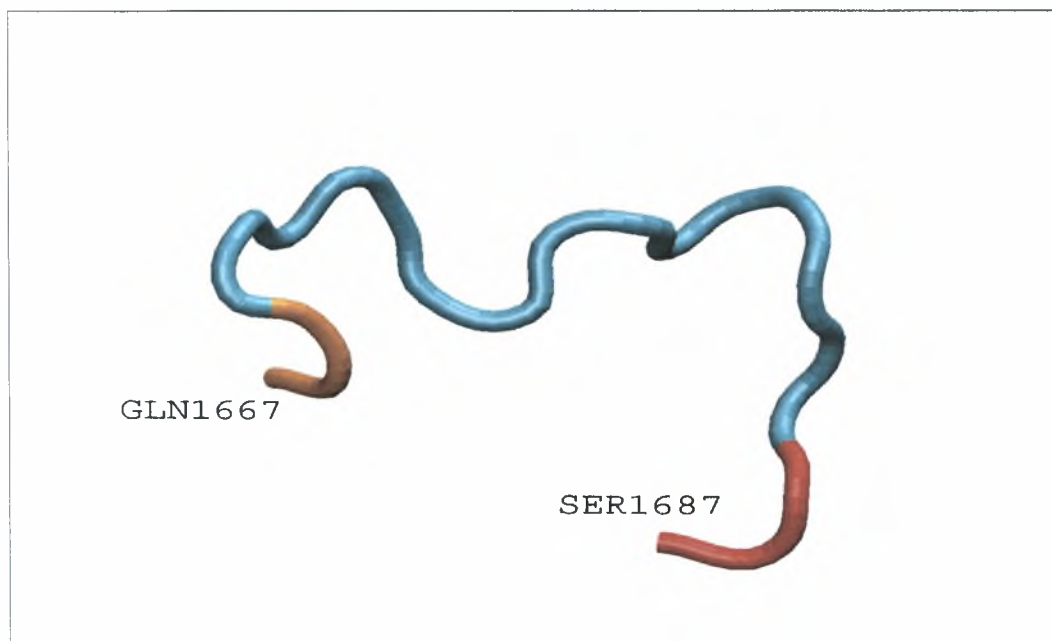
4.2 Η δομή του δεύτερου συνδετικού τμήματος

Η ίδια διαδικασία ακολουθήθηκε και για το δεύτερο συνδετικό τμήμα. Για το συγκεκριμένο τμήμα πραγματοποιήθηκαν προσομοιώσεις συνολικής διάρκειας 16 nsec. Το παρακάτω διάγραμμα δείχνει πώς εξελίχθηκε το σύστημα σε συνάρτηση με το χρόνο.



Εικόνα 4.3: Το διάγραμμα απεικονίζει την χρονική εξέλιξη του RMSD κατά τη διάρκεια της εξισορρόπησης του συνδετικού τμήματος A2-A3.

Η δομή του δεύτερου συνδετικού τμήματος απεικονίζεται στην εικόνα 4.4:

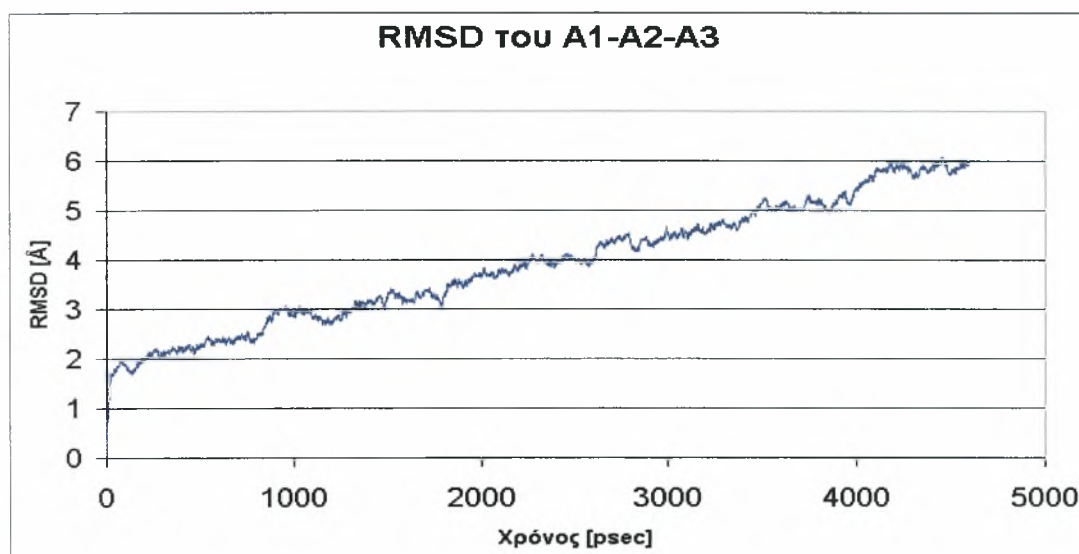


Εικόνα 4.4: Η δομή του σκελετού του δεύτερου συνδετικού τμήματος.

Όπως φαίνεται στην εικόνα 4.4, τα παραπάνω κατάλοιπα δεν σχηματίζουν καμία συγκεκριμένη δευτεροταγή δομή.

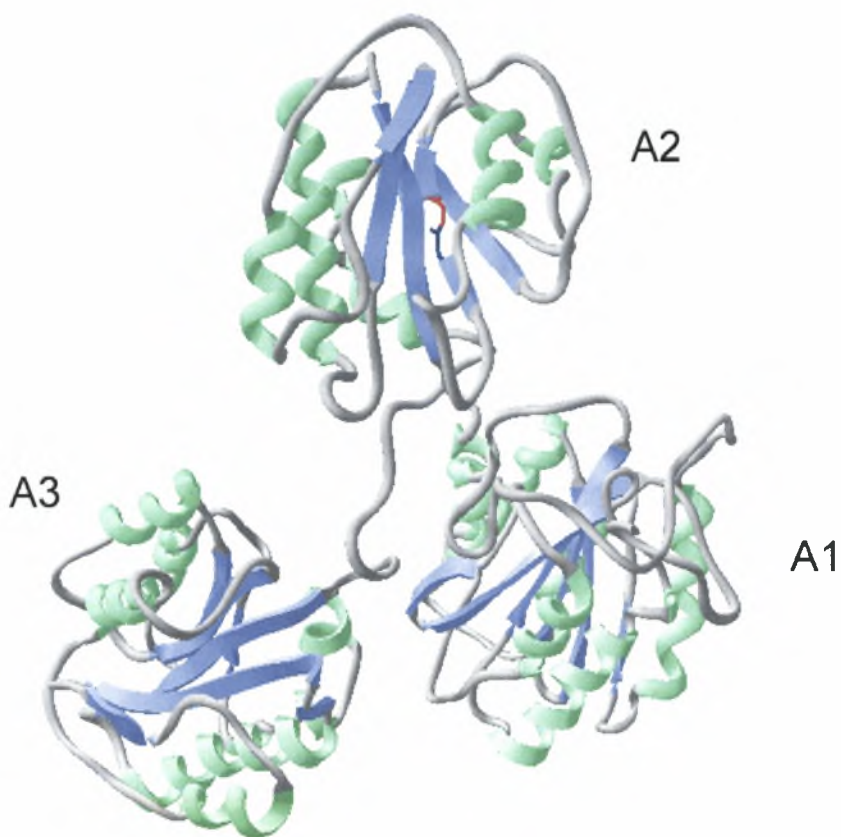
4.3 Η δομή του τμήματος A1-A2-A3

Στη συνέχεια, τα δύο συνδεδεμένα τμήματα ενώνονται με τις τρεις επικράτειες και αφήνεται το σύστημα να εξισορροπηθεί. Χρειάστηκαν προσομοιώσεις συνολικής διάρκειας 4,6 nsec. Το RMSD του τμήματος παρατίθεται στην εικόνα 4.5.



Εικόνα 4.5: Το διάγραμμα απεικονίζει το RMSD του τμήματος A1-A2-A3.

Η δομή της συνολικής πρωτεΐνης παρατίθεται στην εικόνα 4.6:



Εικόνα 4.6: Η δομή του τμήματος A1-A2-A3 μετά τις προσομοιώσεις.

Στην παραπάνω εικόνα οι α-έλικες χρωματίστηκαν πράσινες, οι β-πτυχώσεις μπλε και με γκρι χρωματίστηκαν τα τμήματα που δεν εμφανίζουν συγκεκριμένο τύπο δευτεροταγούς δομής. Στην A2 υπομονάδα διακρίνονται με μπλε και κόκκινο η Tyg1605 και η Met1606, αντίστοιχα. Ο δεσμός που ενώνει τα δύο αυτά κατάλοιπα είναι εύθραυστος και αποτελεί τη θέση κοπής της μεταλλοπρωτεάσης ADAMTS13. Παρατηρώντας την εικόνα, είναι εμφανής η ομοιότητα των τριών υπομονάδων μεταξύ τους. Ήταν, βέβαια, αναμενόμενο, καθώς μοιάζουν πολύ και οι γενωμικές τους ακολουθίες.



Συζήτηση

5. Συζήτηση

Οι πειραματικές μέθοδοι προσδιορισμού της δομής των πρωτεϊνών είναι πολλές φορές δαπανηρές, επίπονες και χρονοβόρες. Επομένως, προέκυψε η ανάγκη να ληφθούν πληροφορίες για το πρωτεϊνικό δίπλωμα με άλλες μεθόδους. Σε αυτό βοήθησε και η τεχνολογική επανάσταση στον τομέα των υπολογιστικών συστημάτων και της Πληροφορικής παρέχοντας σήμερα αξιόλογα εφόδια για τη δημιουργία θεωρητικών μεθόδων πρόβλεψης της δομής.

Τα πειραματικά δεδομένα σαφέστατα δείχνουν ότι όλη η αναγκαία πληροφορία ώστε μια πρωτεΐνη να διπλωθεί στην φυσική της στερεοδομή είναι κωδικοποιημένη στην γραμμική αμινοξική της ακολουθία. Για αυτόν το λόγο, έγιναν πολλές προσπάθειες ώστε να προβλεφθεί η δομή μια πρωτεΐνης από την αμινοξική της ακολουθία με περιορισμένη, όμως, επιτυχία μόνο. Μεταξύ των θεωρητικών μεθόδων, οι πιο αξιόπιστες μέθοδοι είναι ο συγκριτικός σχεδιασμός και η αναγνώριση αναδίπλωσης. Παρόλα αυτά δεν πρέπει σε καμία περίπτωση να θεωρούμε τα αποτελέσματα απολύτως σωστά. Οι πηγές λαθών είναι αρκετές. Ένα λάθος μπορεί για παράδειγμα να προκύπτει από μία λανθασμένη ομοπαράθεση και όλη η υπόλοιπη διαδικασία σχεδιασμού να ακυρώνεται. Οι δομές, λοιπόν, που προκύπτουν πρέπει να αντιμετωπίζονται με επιφύλαξη.

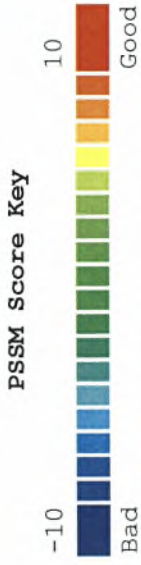
Ένα από τα πλέον αξιόλογα πακέτα λογισμικού για την πρόβλεψη της δομής των πρωτεϊνών με τη μέθοδο του συγκριτικού σχεδιασμού (homology modeling) είναι το 'Modeller'. Μετά την εγκατάστασή του κατεβλήθη προσπάθεια να προσδιοριστεί η δομή των συνδετικών τμημάτων μεταξύ των επικρατειών A1, A2 και A2-A3 του vWF. Η προσπάθεια απέτυχε τόσο λόγω του μικρού μήκους των αλυσίδων όσο και λόγω της πτωχής ομοιότητας με πρότυπα της βάσης δεδομένων PDB. Η προσπάθεια συνεχίστηκε καταφεύγοντας στη μέθοδο της αναγνώρισης αναδίπλωσης (threading) με τη βοήθεια του web server Phyre. Τα μοντέλα που πρότεινε ο εν λόγω server εξισορροπήθηκαν με προσομοιώσεις μοριακής δυναμικής και ακολούθως συνενώθηκαν με τις επικράτειες A1, A2 και A3. Το συνολικό μοντέλο εξισορροπήθηκε για περίπου 4,5 nsec. Το χρονικό αυτό διάστημα είναι μόλις ικανοποιητικό και απαιτείται μεγαλύτερος χρόνος εξισορρόπησης για να επιτευχθεί ένα πιο αξιόπιστο μοντέλο. Παρόλο που ο διαθέσιμος υπολογιστικός χρόνος ήταν περιορισμένος, η δομή που προέκυψε είναι μία εικόνα της διάταξης των επικρατειών A1, A2, A3 που συμφωνεί με πειραματικές παρατηρήσεις μελέτης δευτεροταγούς δομής. Συνεπώς, μπορεί να χρησιμοποιηθεί σαν αφετηρία για την μελέτη ενδομοριακών και διαμοριακών αλληλεπιδράσεων, με απώτερο σκοπό την κατανόηση της λειτουργίας του vWF ως παράγοντα του μηχανισμού της αιμόστασης.



Παράρτημα

QuickPhyre Alignment of query sequence to fold library protein d1n1ta3

FASTA version of alignment



Secondary Structure Key

- █ beta strand
- █ C Coil
- █ alpha helix

Query Index

Query Sequence Conservation 30 %

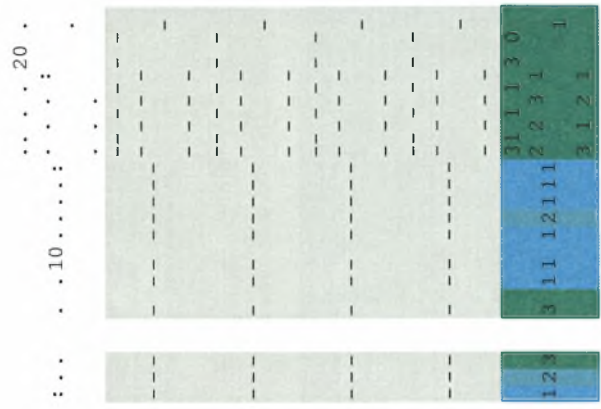
Query Sequence Conservation 40 %

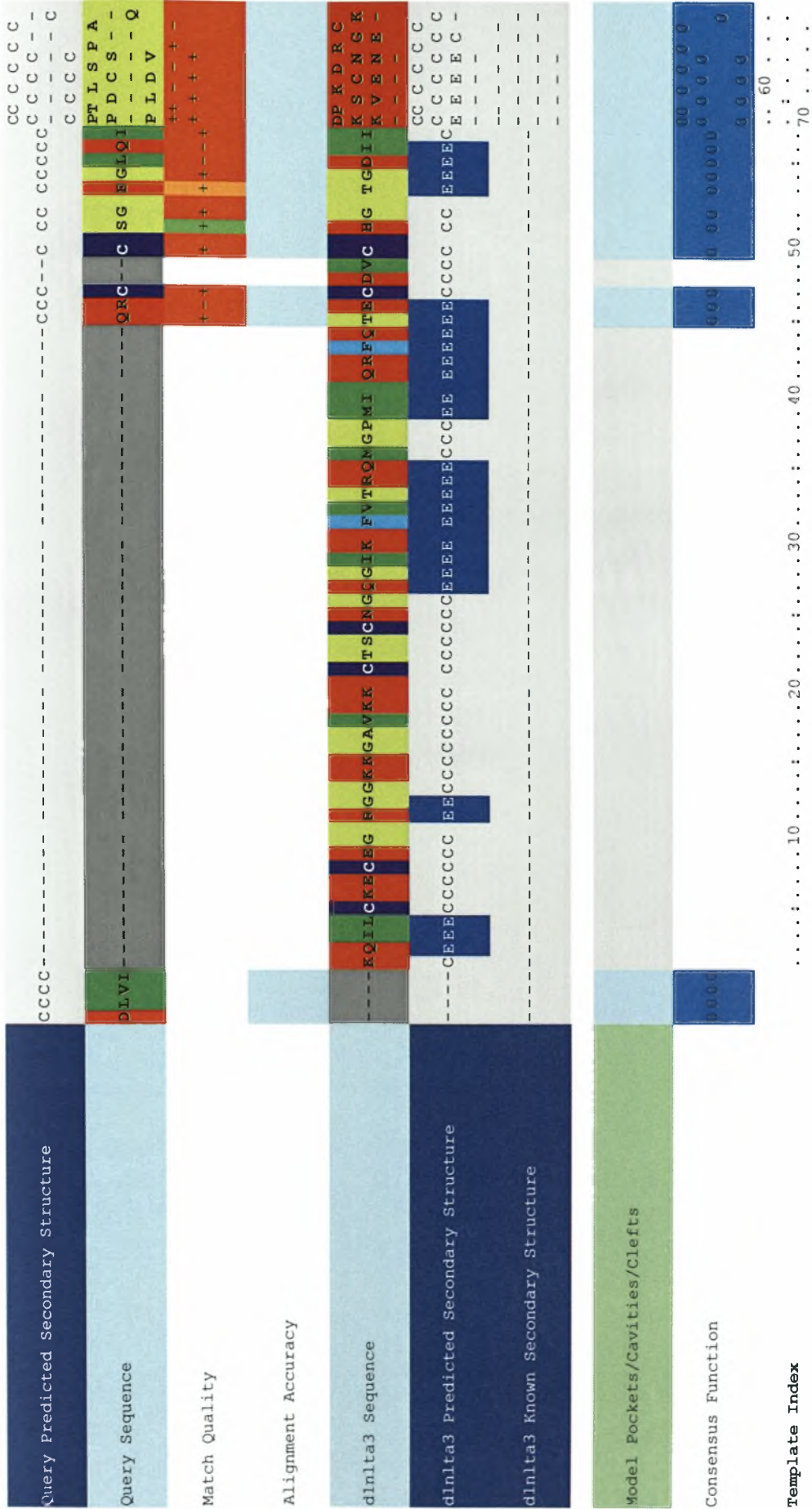
Query Sequence Conservation 50 %

Query Sequence Conservation 60 %

Query Sequence Entropy (Normalised)

Query Sequence Evolutionary Trace ?







Βιβλιογραφία

7. Βιβλιογραφία

1. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234, 779-815, 1993.
2. Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA. Proteins: Structure, Function, Bioinformatics, vol 70(3) 611-625 (2008).
3. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. & Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics, *J Comput Phys.* 151, 283-312.
4. Humphrey, W., Dalke, A. & Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph.* 14, 33-8, 27-8
5. Bajaj, M. S., Kuppaswamy, M. N., Manepalli, A. N., and Bajaj, S. P. (1999) *Thromb. Haemostasis* 82, 1047–1052
6. de Wit, T. R., and van Mourik, J. A. (2001) *Best Pract. Res. Clin. Haematol.* 14, 241–255
7. Sadler JE: vonWillebrand factor. *J Biol Chem* 266:22777, 1991
8. Voorberg J, Fontijn R, Calafat J, Janssen H, van Mourik JA, Pannekoek H (1991) *J Cell Biol* 113:195–205
9. Marti T, Rosselet SJ, Titani K, Walsh KA (1987) *Biochemistry* 26:8099–8109
10. Sadler JE (1998). "Biochemistry and genetics of von Willebrand factor". *Annu. Rev. Biochem.* 67: 395–424. doi:10.1146/annurev.biochem.67.1.395. PMID 9759493.
11. Hoyer LW, Shainoff JR (1980) *Blood* 55:1056–1059
12. Federici AB, Bader R, Pagani S, Colibretti ML, De Marco L, Mannucci PM (1989) *Br J Haematol* 73:93–99
13. S. Zanardelli, J. T. B. Crawley, Chan K. N. Chan Kwo Chion, J. K. Lam, R. J. S. Preston, and D. A. Lane, 2006, ADAMTS13 Substrate Recognition of vonWillebrand Factor A2 Domain, *Journal of Biological Chemistry*, Vol 231, pp: 1555-1563
14. Levy GG, Nichols WC, Lian EC, Foroud T, McClintick JN, McGee BM, Yang AY, Siemieniak DR, Stark KR, Gruppo R, et al. (2001) *Nature* 413:488–494.
15. Furlan M, Robles R, Galbusera M, Remuzzi G, Kyrle PA, Brenner B, Krause M, Scharrer I, Aumann V, Mittler U, et al. (1998) *N Engl J Med* 339:1578–1584.
16. Tsai HM, Lian EC (1998) *N Engl J Med* 339:1585–1594.
17. Sadler JE (1998) *Annu Rev Biochem* 67:395–424.
18. Sugimoto M, Mohri H, McClintock RA, Ruggeri ZM: Identification of discontinuous von Willebrand factor sequences involved in complex formation with botrocetin. *J Biol Chem* 266:18172, 1991
19. V. Jenkins, K. John Pasi, S. J. Perkins, Molecular modeling of ligand and mutation sites of the type A domains of human von Willebrand Factor and their relevance to von Willebrand's disease, *Blood*, Vol 91, No 6 (March 15), 1998:pp 2032-2044
20. Weiqiang Gao, P. J. Anderson, E. M. Majerus, E. A. Tuley, and J. E. Sadler, Exosite interactions contribute to tension-induced cleavage of von Willebrand factor by the antithrombotic ADAMTS13 metalloprotease, *PNAS*, December 12, 2006, vol. 103, no. 50, 19099–19104

21. Furlan, M., Robles, R., and Lammle, B. (1996) *Blood* 87, 4223–4234
22. Tsai, H. M. (1996) *Blood* 87, 4235–4244
23. Majerus, E. M., Anderson, P. J., and Sadler, J. E. (2005) *J. Biol. Chem.* 280, 21773–21778
24. Ai, J., Smith, P., Wang, S., Zhang, P., and Zheng, X. L. (2005) *J. Biol. Chem.* 280, 29428–29434
25. Majerus EM, Zheng X, Tuley EA, Sadler JE (2003) *J Biol Chem* 278:46643–46648.
26. Σταύρου Ι. Χαμόδρακα, 1993, Θέματα Μοριακής Βιοφυσικής, Εκδόσεις Συμμετρία: Αθήνα
27. K. E van Holde, C. Johnson, Pui Shing Ho, Principles of Physical Biochemistry, Prentice Hall, 2006
28. alpha.physics.uoi.gr/optlab/PF1/pol1.ppt
29. mde-katsor.biol.uoa.gr/mde4/stefanidi/cd.htm
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. *JMolBiol* 215:403-10.
31. Pearson WR (1990): Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol* 183:63-98.
32. Thompson JD, Higgins DG, Gibson TJ (1994): ClustalW: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-80.
33. CASP4 (Forthcoming) Results from the Comparative Assessment of Techniques for Protein Structure Prediction. *Proteins* 45(S5):98-162.
34. Bates PA, Steinberg MJE (1999): Model building by comparison at CASP3: using expert knowledge and computer automation. *Proteins* (Suppl. 3):47-54.
35. Sali A, Blundell TL (1993): Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779-815.
36. Peitsch MC, Schwede T, Guex N (2000): Automated protein modelling—the proteome in 3D. *Pharmacogenomics* 1:257-66.
37. Vriend G (1990): WHAT IF—A molecular modeling and drug design program. *J Molec Graphics* 8:52-6.
38. Simons KT, Bonneau R, Ruczinski I, Baker D (1999): Ab initio structure prediction of CASP ΙΠ targets using ROSETTA. *Proteins* (Suppl. 3):171-6.
39. Fiser A, Do RK, Sali A (2000): Modeling of loops in protein structures. *Protein Sci* 9:1753-73. Hooft RWW, Vriend G, Sander C, Abola EE (1996): Errors in protein structures. *Nature* 381:272.
40. Tappura K (2001): Influence of rotational energy barriers to the conformational search of protein loops in molecular dynamics and ranking the conformations. *Proteins* 44:167-79.
41. Sanchez R, Sali A (1997): Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins* (Suppl. 1):50-8.
42. Desmet J, De Maeyer M, Hazes B, Lasters I (1992): The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 356:539-42.
43. de Filippis V, Sander C, Vriend G (1994): Predicting local structural changes that result from point mutations. *Protein Eng* 7:1203-8.
44. Stites WE, Meeker AK, Shortle D (1994): Evidence for strained interactions between side-chains and the polypeptide backbone. *J Mol Biol* 235:27-32.

45. Xiang Z, Honig B (2001): Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol* 311:421 -30.
46. IBM Blue Gene team (2001): Blue Gene: a vision for protein science using a petaflop supercomputer. *IBM Sys J* 40:310-27.
47. Chothia C, Lesk AM (1986): The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823-36.
48. Sippl MJ (1993): Recognition of errors in three dimensional structures of proteins. *Proteins* 17:355-62.
49. Novotny J, Rashin AA, Bruccoleri RE (1988): Criteria that discriminate between native proteins and incorrectly folded models. *Proteins* 4:19-30.
50. Branden C, Tooze J (1991): Introduction to Protein Structure. New York: Garland Publishing
51. Rost B (1996): PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol* 266:525-39
52. Cohen FE, Sternberg MJE, Taylor WR (1981): Analysis of the tertiary structure of protein β -sheet sandwiches. *J Mol Biol* 148:253-72
53. Monge, Friesner RA, Honig B (1994): An algorithm to generate low-resolution protein tertiary structures from knowledge of secondary structure. *Proc Natl Acad Sci USA* 91:5027-9
54. Mumenthaler C, Braun W (1995): Predicting the helix packing of globular proteins by self-correcting distance geometry. *Protein Sci* 4:863-71.
55. Cohen FE, Presnell SR (1996): The combinatorial approach. In: Stenberg MJE, editor. *Protein Structure Prediction*. Oxford University Press, pp 207-28.
56. Conolly ML (1983): Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709-13.
57. Philip E. Bourne, Helge Weissig, 2003, *Structural Bioinformatics*, Wiley-Liss, Inc., Hoboken, New Jersey, Canada
58. Jadwiga Bienkowska and Rick Lathrop, Threading algorithms
59. Richard h. Lathrop, Temple F. Smith, Global optimum protein threading with gapped alignment and empirical pair score functions.
60. Anfinsen CB (1973): Principles that govern the folding of protein chains. *Science* 181:223-30.
61. Privalov PL, Gill SJ (1988): Stability of Protein Structure and Hydrophobic Interaction. *Adv Protein Chem* 39:191-235.
62. Bonneau R, Tsai J, Ruczinski I, Chivian D, Rohl C, Strauss CE, Baker D (2001): Rosetta in CASP4: progress in *ab initio* protein structure prediction. *Proteins* 45(Suppl 5): 119-26.
63. Anfinsen CB, Haber E, Sela M, White FW Jr (1961): The kinetics of the formation of native ribonuclease during oxidation of the reduced polypeptide domain. *Proc Nad Acad Sci USA* 47:1309-14.
64. Baker D, Agard DA (1994): Kinetics versus thermodynamics in protein folding. *Biochemistry* 33:7505-9.
65. Ortiz AR, Kolinski A, Rotkiewicz P, Bkowski B, Skolnick J (1999): Ab initio folding of proteins using restraints derived from evolutionary information. *Proteins* 37(S3): 177-85.
66. Pillardy J, Czaplowski C, Liwo A, Lee J, Ripoll DR, Kazmieraewicz R, Oldziej S, Wede-^ meyer WJ, Gibson KD, Amautova YA, Saunders J, Ye YJ, Sheraga

- HA (2001): Recent improvements in prediction of protein structure by global optimization of a potential energy function. *Proc Natl Acad Sci USA* 98:2329-33.
67. Murzin AG (2001): Progress in protein structure prediction. *Nat Struct Biol* 8:110-2. Obmolova G, Ban C, Hsieh P, Yang W (2000): Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. *Nature* 407:703-10.
68. Lesk AM, Lo Conte L, Hubbard T (2001): Assessment of novel fold targets in CASP4: predictions of three-dimensional structures, secondary structures, and interresidue contacts. *Proteins* 45(S5):98-118.
69. Jarosaw Meller, Molecular Dynamics, 2001, Encyclopedia of life sciences, Nature Publishing Group
70. Gromacs User Manual
71. M. P. Allen, Introduction to Molecular Dynamics Simulation, 2004, Computational Soft Matter: From Synthetic Polymers to Proteins, Vol. 23, pp: 1-28
72. Jeffrey J. Sutherland · Lee A. O'Brien, David Lillicrap · Donald F. Weaver, Molecular modeling of the von Willebrand factor A2 Domain and the effects of associated type 2A von Willebrand disease mutations, 2004, *J Mol Model* 10:259-270
73. M. Anton et al., *J. Mol. Biol* (2007) 366, 966-1000.

