



Δημοκρίτειο Πανεπιστήμιο Θράκης

Τμήμα Μοριακής Βιολογίας και Γενετικής

Πτυχιακή εργασία

Υπολογιστικές μελέτες και μοριακή μοντελοποίηση μιας  
παραλλαγής της πρωτεΐνης Rop με αποκαταστημένη  
περιοδικότητα επτάδας

Κοκκινίδου Μαρία

Επιβλέπων καθηγητής: κ. Γλυκός Νικόλαος

Αλεξανδρούπολη 2010



Δημοκρίτειο Πανεπιστήμιο Θράκης  
Τμήμα Μοριακής Βιολογίας και Γενετικής

Πτυχιακή εργασία

Υπολογιστικές μελέτες και μοριακή μοντελοποίηση μιας παραλλαγής της  
πρωτεΐνης Ror με αποκατεστημένη περιοδικότητα επτάδας

Επιβλέπων καθηγητής: κ. Γλυκός Νικόλαος

Κοικινίδου Μαρία  
Αλεξανδρούπολη 2010

# Ευχαριστίες

Η παρούσα πτυχιακή εργασία εκπονήθηκε στο εργαστήριο υπολογιστικής βιολογίας του τμήματος Μοριακής βιολογίας και γενετικής, στην Αλεξανδρούπολη.

Θα ήθελα λοιπόν να ευχαριστήσω τον καθηγητή μου κ. Νικόλαο Γλυκό ο οποίος με δέχτηκε στο εργαστήριο του, μου μετέδωσε τις γνώσεις του και με βοήθησε να εναρμονιστώ στο χώρο, στο σύνολο και στη δουλειά του εργαστηρίου με τον καλύτερο δυνατό τρόπο.

Επίσης, να ευχαριστήσω την Παναγιώτα Γεωργούλια για την βοήθεια της και όλα τα μέλη του εργαστηρίου για την αρμονική συνεργασία.

Ακόμη, θα ήθελα να ευχαριστήσω την οικογένεια μου που στέκεται στο πλευρό μου όλα αυτά τα χρόνια και υποστηρίζει με αγάπη την κάθε επιλογή μου.

Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου που ήταν δίπλα μου τόσο στις εύκολες όσο και στις δύσκολες στιγμές της ζωής μου.

# Περιεχόμενα

<b>Περίληψη</b>	6
<b>Abstract</b>	7
<b>Κεφάλαιο 1ο: Εισαγωγή</b>	
1.1 Στόχος	8
1.2 Η Rop ως μοντέλο και η πορεία προς την Rop2αL (QL)	9
1.3 Χαρακτηριστικά του μεταλλάγματος Rop2αL (QL) από πειραματικά δεδομένα του εργαστηρίου Κρήτης	14
<b>Κεφάλαιο 2ο: Ομόλογη Μοντελοποίηση</b>	
2.1 Εισαγωγή	18
2.2 Ομόλογη μοντελοποίηση και η διαδικασία της	19
2.3 Ενδεδειγμένα προγράμματα ομόλογης μοντελοποίησης	23
2.4 Swiss Model και τα χαρακτηριστικά του	23
2.5 Εφαρμογές του Swiss Model και τα αποτελέσματα του	24
2.6 Το Modeller και τα χαρακτηριστικά του	29
2.7 Εφαρμογές του Modeller και τα αποτελέσματά τους	31
2.8 Αρχικές εφαρμογές και παραγωγή του πρώτου μοντέλου (πλεξούδα)	32
2.9 Βελτίωση του πρώτου μοντέλου και παραγωγή της τελικής δομής της Rop2ααL(QL)	41
2.10 Έλεγχος δομών	44
<b>Κεφάλαιο 3ο: Προσομοιώσεις Μοριακής Δυναμικής</b>	
3.1 Εισαγωγή	45
3.2 Προετοιμασία συστημάτων	50
3.2.1 Προετοιμασία του συστήματος της πλεξούδας	51
3.2.2 Προετοιμασία του συστήματος του sigmoidal και οι απόπειρες εφαρμογής του	53

## Κεφάλαιο 4ο: Ανάλυση προσομοιώσεων

4.1 Ανάλυση προσομοίωσης του μοντέλου πλεξούδας	
4.1.1 Σταθερότητα αρχικής δομής	59
4.1.2 Σταθερότητα δομών από το trajectory και υπολογισμός μέσης δομής	62
4.1.3 Ανάλυση της κινητικότητας των ατόμων (rmsf)	64
4.1.4 Εξέλιξη του Radius of gyration (Rg)	67
4.1.5 Χάρτης αποστάσεων Ca-Ca ατόμων και ανάλυση των διακυμάνσεων	68
4.1.6 Υπολογισμός cross-correlation matrix και ανάλυση	70
4.1.7 Rmsd μεταξύ όλων των πιθανών συνδυασμών δομών από το τροχιακό	72
4.1.8 Ομαδοποίηση των δομών του trajectory (cluster analysis)	74
4.1.9 Υπέρθεση των δομών με τα αντίστοιχα rmsf fluctuations του κυρίαρχου cluster τους.	76
4.1.10 Anisotropic fluctuations	78
4.1.11 Principal component analysis	80
4.1.12 Σύγκριση των μέσων δομών από τα κυρίαρχα cluster (cartesian & dihedral)	94
4.1.13 Συντήρηση δευτεροταγούς δομής	96
4.1.14 Ανάλυση διεδρων γωνιών ( $\chi$ -value)	97
4.2 Ανάλυση προσομοίωσης μοντέλου Sigmoidal	
4.2.1 Σταθερότητα αρχικής δομής	105
4.2.2 Σταθερότητα δομών από το trajectory και υπολογισμός μέσης δομής	107
4.2.3 Ανάλυση της κινητικότητας των ατόμων (rmsf)	109
4.2.4 Εξέλιξη του Radius of gyration (Rg)	110
4.2.5 Χάρτης αποστάσεων Ca-Ca ατόμων και ανάλυση των διακυμάνσεων	111
4.2.6 Υπολογισμός cross-correlation matrix και ανάλυση	113
4.2.7 Rmsd μεταξύ όλων των πιθανών συνδυασμών δομών από το τροχιακό	115
4.2.8 Ομαδοποίηση των δομών του trajectory (cluster analysis)	116
4.2.9 Υπέρθεση των δομών με τα αντίστοιχα rmsf fluctuations του κυρίαρχου cluster τους.	118
4.2.10 Anisotropic fluctuations	119
4.2.11 Principal component analysis	121
4.2.12 Σύγκριση των μέσων δομών από τα κυρίαρχα cluster (cartesian & dihedral)	131

4.2.13 Συντήρηση δευτεροταγούς δομής	132
4.2.14 Ανάλυση διεδρων γωνιών ( $\chi$ -value)	133

## **Κεφάλαιο 5ο: Σύγκριση αποτελεσμάτων ανάλυσης μεταξύ των δύο μοντέλων πλεξούδα και sigmoidal**

**140**

Συμπεράσματα	146
Βιβλιογραφία	157
Παράρτημα	157

# Περίληψη

Η Rop είναι ένα ομοδιμερές, αντιπαράλληλο, 4-α-ελικοειδές δεμάτιο που συμμετέχει στην αντιγραφή των πλασμιδίων ColE1 της E.coli. Το κάθε μονομερές έχει την τοπολογία έλικα-στροφή-έλικα. Μέχρι σήμερα έχουν μελετηθεί εκτενώς ο ρόλος της στροφής και του μοτίβου επανάληψης επτάδας του υδρόφοβου πυρήνα στην αναδίπλωση και σταθερότητα του μορίου. Όσο αφορά το μοτίβο της επτάδας δημιουργήθηκαν δύο σημαντικά μεταλλάγματα η RM6 και η 2αα. Στην παρούσα εργασία μελετάται μια σημειακή μεταλλαγή (Rop2ααL, Q->L) της παραλλαγής 2αα της πρωτεΐνης Rop. Προηγούμενες μελέτες (Kyriazidis master thesis, 2006) έδειξαν ότι η μεταλλαγή αυτή προκαλεί μετατροπή του διμερούς σε τετραμερές με ταυτόχρονη αλλαγή του μοτίβου αναδίπλωσης.

Μέσω της τεχνικής της ομόλογης μοντελοποίησης κατασκευάστηκαν δύο μοντέλα για αυτήν την μεταλλαγή, η πλεξούδα και το sigmoidal, χρησιμοποιώντας ως πρότυπη αλληλουχία τη δομή της RM6. Με την βοήθεια των προσομοιώσεων ερευνήσαμε τη σταθερότητα και τη δυναμική των δύο μοντέλων και συμπεράναμε ότι εμφανίζει μεγαλύτερη σταθερότητα η δομή του sigmoidal απ'ότι η δομή της πλεξούδας όχι όμως με σημαντικές διαφορές. Ωστόσο η δομή του sigmoidal μοιάζει περισσότερο απ' ότι της πλεξούδας με την πρότυπη δομή τους, την RM6.

# Abstract

Rop is a homodimeric, antiparallel, 4- $\alpha$ -helix-bundle that takes part in replicating of plasmids ColE1 of E.coli. Every monomer has the topology helix-turn-helix. Up until now researches have extensively studied the role of heptat in the folding and the stability of the particle. As far as the pattern of the heptat two important mutants have been created, RM6 and 2 $\alpha\alpha$ . In this project we are examining a mutation (Rop2 $\alpha\alpha$ L, Q-> L) of the already mutated 2 $\alpha\alpha$  structure of Rop. Previous studies showed that this mutation causes the conversion of the dimer to tetramer with simultaneous change of the folding pattern.

Through the technique of Homology modeling we created 2 models for this mutation and we used the structure of RM6 as template sequence. With the help of molecular dynamics simulations we investigated the stability and the dynamics of both models and we reached to the conclusion that the sigmoidal structure, seems to be more stable than the structure of pleksouda. However, we did not find any significant differences. Nevertheless the structure of sigmoidal resembles to the template structure of RM6, more than pleksouda.



# Κεφάλαιο 1

## Εισαγωγή

Η μοριακή μοντελοποίηση είναι η επιστήμη και η τέχνη της μελέτης μοριακών δομών και λειτουργίας αυτών, μέσω κατασκευής υπολογιστικών μοντέλων. Αποτελεί έναν σημαντικό στόχο της συστηματικής και μαθηματικής βιολογίας. Η υπολογιστική βιολογία στοχεύει να αναπτύξει και να χρησιμοποιήσει αποδοτικούς αλγορίθμους, βάσεις δεδομένων δομών και εργαλεία μοριακής απεικόνισης, ώστε να ενορχηστρώσει την ολοκλήρωση της μεγάλης ποσότητας βιολογικών στοιχείων, με στόχο την υπολογιστική μοντελοποίηση. Περιλαμβάνει τη χρήση των υπολογιστικών προσομοιώσεων βιολογικών συστημάτων, για κυτταρικά υποσυστήματα, ώστε να αναλύσουν και να απεικονίσουν τις σύνθετες συνδέσεις αυτών των κυτταρικών διαδικασιών και να κατανοήσουν εξελικτικές διαδικασίες.

### 1.1 Στόχος

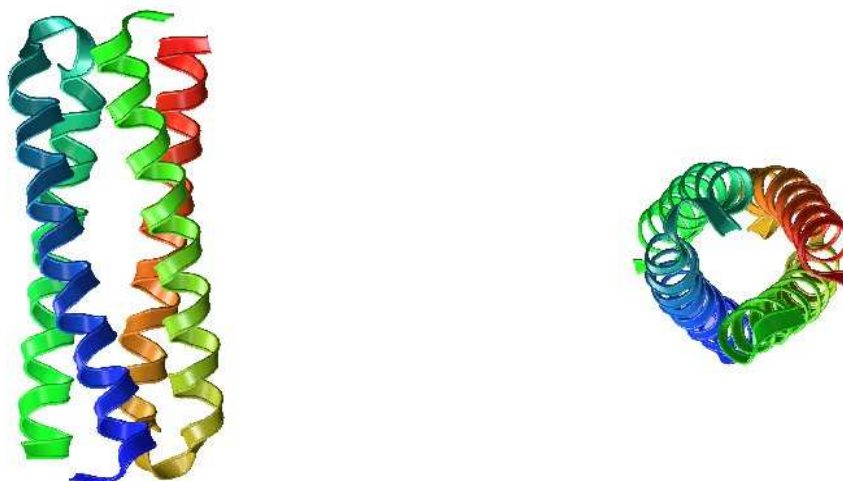
Στόχος αυτής της εργασίας είναι η δημιουργία και η μελέτη της δομής-μοντέλου ενός συγκεκριμένου μεταλλάγματος της πρωτεΐνης Ror. Η μελέτη αυτής της μετάλλαξης έχει ενδιαφέρον γιατί αν τα αποτελέσματα επαληθευθούν πειραματικά, θα είναι η πρώτη φορά που θα δειχθεί ότι μια απλή αμινοξική αντικατάσταση μπορεί να μετατρέψει ένα διμερές σε τετραμερές με ταυτόχρονη αλλαγή του μοτίβου αναδίπλωσης.

## 1.2 Η Rop ως μοντέλο και η πορεία προς την Rop2aL (QL)

Η δομική απλότητα του μοτίβου αναδίπλωσης της Rop, είναι αυτό που οδηγεί πολλές ερευνητικές ομάδες να την χρησιμοποιούν ως σύστημα μοντέλο στην έρευνα των σχέσεων δομής - αλληλουχίας για την αναδίπλωση και δυναμική των 4-α-ελικοειδών δεματιών (Cohen & Parry, 1990; Chou et al., 1988). Αυτό συμβαίνει διότι η απλή δομή της που έχει μελετηθεί διεξοδικά τόσο με κρυσταλλογραφία ακτίνων Χ (Banner et al., 1987) όσο και με φασματοσκοπία NMR (Eberle et al., 1991) αποτελεί, χαρακτηριστικό παράδειγμα ενός κανονικού, ομοδιμερούς, αντιπαράλληλου 4-α-ελικοειδούς δεματίου.

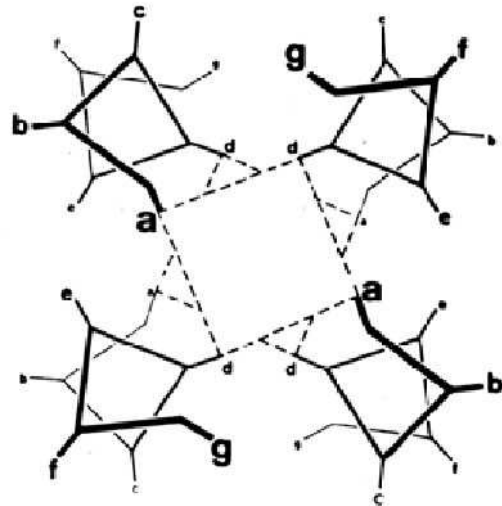
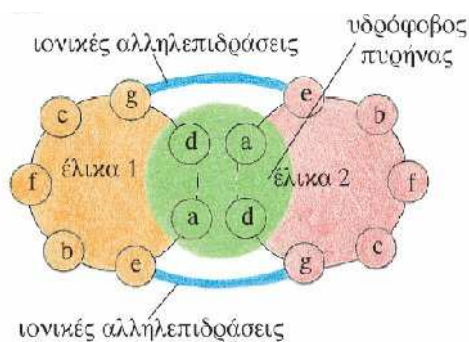
Επιπλέον, εξαιτίας της δομής της, η Rop εμφανίζει σημεία κλειδιά τα οποία προσφέρονται για μεταλλάξεις (Magliery and Regan, 2004) προκειμένου να μελετηθούν αλλαγές στην τοπολογία και τη βιολογική λειτουργία της πρωτεΐνης καθώς επίσης για γενετική και δομική ανάλυση (Castagnoli et al., 1989). Τα σημεία στα οποία πραγματοποιήθηκαν συστηματικές μεταλλάξεις είναι: στα απαραίτητα για τη λειτουργία και το βιολογικό της ρόλο κατάλοιπα (Predki et al., 1995), τα κατάλοιπα του υδρόφοβου πυρήνα (Munson et al., 1994a, 1996) και τα κατάλοιπα της στροφής (Nagi & Regan, 1997; Nagi et al., 1999; Castagnoli et al., 1994). Μέχρι σήμερα έχουν γίνει πολλές μεταλλάξεις σημειακές αλλά και πιο εκτενείς στις οποίες παρόμοιες αμινοξικές αλληλουχίες της πρωτεΐνης οδηγούν σε παρόμοιες (Vlassi et al., 1994) ή και σε τελείως διαφορετικές δομές (Glykos et al., 1999; 2004) με διαφορετικές τοπολογίες, διαφορετικούς υδρόφοβους πυρήνες και διαφορετικές ιδιότητες επιφάνειας.

Ο κύριος ρόλος της πρωτεΐνης Rop είναι να ρυθμίζει την αντιγραφή των πλασμιδίων, ColE1, της E.coli, λειτουργώντας έμμεσα ως αναστολέας, καθώς βοηθά στη σωστή τοποθέτηση των ρυθμιστικών μορίων RNA του συμπλόκου (Polisky et al., 1988; Castagnoli et al., 1989). Η Rop είναι ένα διμερές όπου κάθε μονομερής υπομονάδα της αποτελείται από δύο α-έλικες, οργανωμένες σε μια δομή ενός αντιπαράλληλου σπειρωμένου σπειράματος, στο οποίο οι υδρόφοβες πλευρικές αλυσίδες πακετάρονται σύμφωνα με το μοντέλο “εσοχών-εξοχών” και ενώνονται με μια στροφή τριών αμινοξέων (Banner et al., 1986). Τα δυο αυτά μονομερή πακετάρονται σύμφωνα με το μοντέλο “προεξοχές στις αύλακες” (Tooze 1999) και σχηματίζουν το διμερές μόριο της Rop, στο οποίο διευθετούνται ως ένα αριστερόστροφο 4-α-ελικοειδές δεμάτιο (Cohen & Parry, 1990) (εικόνα1.1).



**Εικόνα 1.1** Η δομή της πρωτεΐνης Rop, χαρακτηριστικό παράδειγμα 4- $\alpha$ -ελικοειδούς δεματίου. Τα χρώματα από το μπλε στο πράσινο δηλώνουν την πορεία της πρωτεΐνης από το αμινοτελικό στο καρβοξυτελικό άκρο του  $\alpha$  μονομερούς και από το πράσινο στο κόκκινο την πορεία από το αμινοτελικό στο καρβοξυτελικό άκρο του μονομερούς  $\beta$ . Δεξιά, βλέπουμε την πρωτεΐνη μετά από στροφή  $90^\circ$  γύρω από έναν άξονα κάθετο στον  $\gamma$  άξονα του δεματίου και στο επίπεδο. Τα προγράμματα που χρησιμοποιήθηκαν για την μοριακή απεικόνιση ήταν τα: Rasmol (Bernstein 1999), VMD (Humphrey et al., 1996) και Raster3D (Merritt & Bacon 1997).

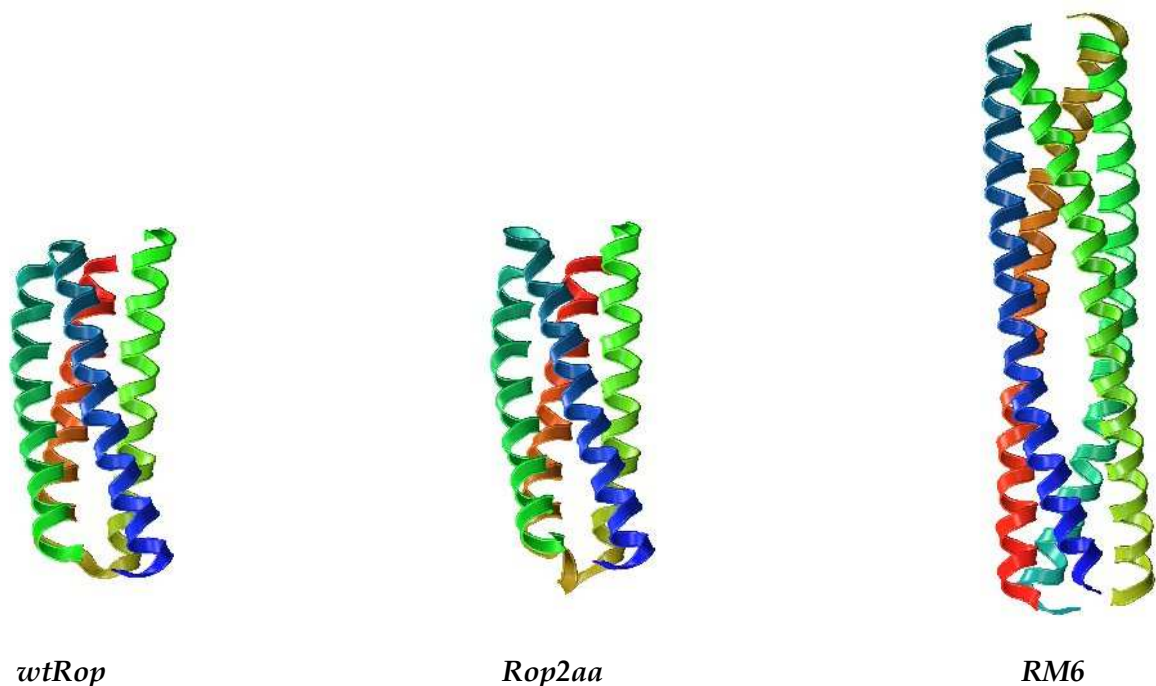
Όπως σε κάθε πρωτεΐνη με μορφή σπειρωμένου σπειράματος έτσι και στην αλληλουχία της Rop ισχύει ότι μειώνεται ο αριθμός των καταλοίπων ανά στροφή σε κάθε έλικα από 3.6 σε 3.5 (Crick, 1953b) οπότε το πρότυπο των αλληλεπιδράσεων των πλευρικών αλυσίδων ανάμεσα στις έλικες επαναλαμβάνεται κάθε επτά κατάλοιπα, δηλαδή κάθε 2 στροφές. Επιπλέον οι  $\alpha$ -έλικες εμφανίζουν αμφιπαθή χαρακτήρα, δηλαδή παρατηρείται ένα πρότυπο υδρόφοβων και υδρόφιλων αμινοξέων. Αυτή, λοιπόν, η περιοδικότητα των επτά καταλοίπων, μπορεί να περιγραφεί ως επανάληψη επτάδας (Munson et al., 1994) και τα αντίστοιχα αμινοξέα συμβολίζονται ως (a, b, c, d, e, f, g)<sub>n</sub> (Crick, 1953; Cohen, 1986)(εικόνα 1.2). Οι θέσεις a και d καταλαμβάνονται πάντα από υδρόφοβα κατάλοιπα και βρίσκονται στο εσωτερικό του δεματίου (ελάχιστα προσβάσιμες στο διαλύτη), ενώ η θέση f είναι περισσότερο εκτεθειμένη στο εξωτερικό περιβάλλον του δεματίου και αρκετά συχνά καταλαμβάνεται από υδρόφιλο αμινοξύ (Paliakasis and Kokkinidis, 1992). Μελετώντας συγκεκριμένα το μοτίβο της επτάδας στην αλληλουχία της Rop, παρατηρήθηκε μια διαταραχή στην περιοχή της στροφής που ανάγεται στην απουσία δύο καταλοίπων. Δημιουργήθηκε το ερώτημα εάν και κατά πόσο η στροφή της πρωτεΐνης οφείλεται στη διακοπή της περιοδικότητας της επτάδας (Banner et al., 1987; Glykos et al., 2006).



**Εικόνα 1.2** Αριστερά : Σχηματική αναπαράσταση (σε αξονική προβολή) μιας επανάληψης επτάδας και οι επτά χαρακτηριστικές θέσεις των καταλοίπων, (η εικόνα αναπαράγεται άνευ άδειας από τους Branden & Tooze)

Δεξιά:: Η διεθέτιση των επτά θέσεων a-g στη δομή της πρωτεΐνης Rop. Σχηματική αναπαράσταση μόνο των ατόμων των a και β ανθράκων (Ca, Cβ), με πάχος δύο στροφές για κάθε έλικα. Με τις διακοπτόμενες γραμμές ορίζονται κατά προσέγγιση τα επίπεδα (layers) όπου πετυχαίνεται βέλτιστο πακετάρισμα των εσωτερικών υδρόφοβων αμινοξέων (a, d θέσεις), (η εικόνα αναπαράγεται χωρίς άδεια από τους Paliakasis & Kokkinidis, 1992)

Προκειμένου να απαντηθεί αυτό, σχεδιάστηκαν δυο μεταλλάγματα. Το πρώτο βασίστηκε στην εισαγωγή δύο καταλοίπων στην περιοχή της στροφής και συγκεκριμένα προστέθηκαν δυο κατάλοιπα αλανίνης εκατέρωθεν του καταλοίπου ασπαρτικού οξέως 30 και ονομάστηκε Rop2αα (2αα) (Castagnoli et al., 1989, Vlasi et al., 1994; Steif et al., 1995). Το δεύτερο, προέκυψε από την απαλοιφή των πέντε καταλοίπων (Asp30-Gln34) της περιοχής της στροφής με αποτέλεσμα τη δημιουργία της δομής της RM6 (Castagnoli et al., 1989; Lassalle & Hinz, 1998; Glykos et al., 2006). Το προσδοκώμενο ήταν, με την αποκατάσταση της περιοδικότητας της επτάδας να εξαλειφθεί η στροφή και το μονομερές της κάθε έλικας να είναι πλέον μια εκτεταμένη α-έλικα. Παρόλο που και στις δυο δομές επιτεύχθηκε η αποκατάσταση της περιοδικότητας, με την 2αα να είναι κατά επτά κατάλοιπα μακρύτερη, διαπιστώθηκε μετά από κρυσταλλογραφικό προσδιορισμό των δομών ότι η RM6 επιβεβαίωσε την αρχική υπόθεση (ότι η στροφή οφείλεται στην διακοπή της περιοδικότητας της επτάδας) σε αντίθεση με την 2αα (Εικόνα 1.3)



**Εικόνα 1.3** Παρουσιάζονται αριστερά η δομή της φυσικής Rop (PDB id: 1rop), στο κέντρο η μετάλλαξη της 2aa (PDB id:1nkd) και δεξιά η μετάλλαξη της RM6 (PDB id:1qx8), ώστε να διακρίνονται οι διαμορφώσεις των δομών κατά την προσπάθεια αποκατάστασης του μοτίβου της επτάδας της φυσικής Rop. Η 2aa δεν εμφανίζει πολλές αλλαγές στην περιοχή της στροφής και στη σχετική διεθέτηση των ελίκων της δομή της σε σχέση με την φυσική Rop, όμως στη δομή της RM6 έχει αποκατασταθεί το μοτίβο της επτάδας, με αποτέλεσμα την εξάλειψη των στροφών, την επιμήκυνση των ελίκων και την μετατροπή του μορίου από διμερές σε τετραμερές. Ισχύει ο ίδιος χρωματικός κώδικας για τις δομές *wtRop* *Rop2aa* με της εικόνας 1.1. όμως στην RM6 το μπλε δηλώνει το αμινοτελικό άκρο του πρώτου μονομερούς (A) και έπειτα απεικονίζονται με τη σειρά τα υπόλοιπα μονομερή B, C, D μέχρι το κόκκινο όπου αντιστοιχεί στο καρβοξυτελικό άκρο του D.

Συγκεκριμένα η RM6 με την αφαίρεση των πέντε καταλοίπων παρουσίασε μια αναδιοργάνωση της δομής της φυσικής Rop. Προέκυψε ένα τετραμερές, αντιπαράλληλο, αριστερόστροφο 4-α-ελικοειδές δεμάτιο με το κάθε μονομερές της να είναι μια εκτεταμένη α-έλικα αντί του μοτίβου έλικα-στροφή-έλικα των μονομερών της διμερούς Rop. Εμφανίζει πολύ αυξημένη θερμοδυναμική σταθερότητα με τη θερμοκασία τήξης ( $T_m$ ) να αυξάνεται στους 101° C από 71° C που παρουσιάζει η φυσική Rop (Lassalle et al., 1998). Αυτή η αυξημένη σταθερότητάς της οφείλεται κυρίως στον καλά πακεταρισμένο υδρόφοβο πυρήνα ο οποίος αποτελείται από δεκαέξι επίπεδα, διπλάσια από τα οκτώ της φυσικής Rop. Τέλος, ο κρυσταλλογραφικός προσδιορισμός της δομής ότι η RM6 μπορεί να περιγραφεί ως ένα διμερές διμερών (Glykos et al., 2006).

Από την άλλη, στο μοντέλο της 2aa παρατηρήθηκε ότι με την συμπλήρωση της επτάδας -εισαγωγή των δυο καταλοίπων αλανίνης- ο στόχος δεν επιτεύχθει. Προέκυψε

διμερές το οποίο ήταν ένα αριστερόστροφο, αντιπαράλληλο 4-α-ελικοειδές δεμάτιο σχεδόν πανομοιότυπο με τη φυσική Rop (Presmell & Cohen 1989), εμφανίζοντας και παρόμοιες θερμοδυναμικές ιδιότητες (Vlassi et al., 1994).

Παρατηρώντας την αλληλουχία της 2αα διαπιστώθηκε ότι, όταν τα δύο κατάλοιπα αλανίνης προστέθηκαν, γύρω από το ασπαρτικό οξύ 30 (Asp 30), κατέλαβαν τις θέσεις e και g του μοτίβου της επτάδας και ότι με την εισαγωγή αυτή μετατοπίστηκαν οι θέσεις των υπόλοιπων καταλοίπων. Στην επτάδα που προέκυψε (ADA~~A~~DEQ) τα τρία πρώτα κατάλοιπα (ADA) κατέλαβαν τις θέσεις e-g και τα τέσσερα τελευταία κατάλοιπα (A~~A~~DEQ) μπήκαν στις a έως d θέσεις αντίστοιχα. Δηλαδή στη θέση d μια καθαρά υδρόφοβη θέση, τοποθετήθηκε η γλουταμίνη (Q), ένα καθαρά υδρόφιλο κατάλοιπο. Οπότε προέκυψε ένα λανθασμένο πρότυπο υδροφοβικότητας. Προκειμένου να αποκατασταθεί το πρότυπο αυτό της 2αα, διατυπώθηκε η ιδέα (Glykos et al., 2006) της αντικατάστασης της γλουταμίνης (θέση 36 στην παραλλαγή 2αα, θέση 34 για την wtRop) με ένα πιο συμβατό με την θέση d της επτάδας υδρόφοβο κατάλοιπο, όπως η λευκίνη, η οποία και προτιμάται για το εσωτερικό των υδρόφοβων πυρήνων (Paliakasis & Kokkinidis, 1992; Fadouloglou et al., 2001). Το προσδωκόμενο θα ήταν να επιτευχθεί η εξάλειψη της στροφής ώστε το μονομερές της 2αα να έχει πλέον μορφή εκτεταμένης α-έλικας, που ήταν και ο αρχικός στόχος δημιουργίας της (δομή παρόμοια με την RM6). Από εδώ και στο εξής θα αναφερόμαστε σε αυτήν την μεταλλαγή ως Rop2aL ή QL (Glykos et al., 2006).

### Αλληλουχίες μεταλλαγμάτων της πρωτεΐνης Rop

<b>wtRop</b>	MTKQEK <del>T</del> ALNMARFIRSQT <del>L</del> TLLEKLNEL	<b>DADEQ AD</b>	ICESLHDHADELYRSCLARF ( GDDGENL )
<b>RM6</b>	MTKQEK <del>T</del> ALNMARFIRSQT <del>L</del> TLLEKLNEL	<b>----- AD</b>	ICESLHDHADELYRSCLARF ( GDDGENL )
<b>Rop2aa</b>	MTKQEK <del>T</del> ALNMARFIRSQT <del>L</del> TLLEKLNEL	<b>ADA<del>A</del>DEQAD</b>	ICESLHDHADELYRSCLARF ( GDDGENL )
<b>Rop2aL (QL)</b>	MTKQEK <del>T</del> ALNMARFIRSQT <del>L</del> TLLEKLNEL	<b>ADA<del>A</del>DELAD</b>	ICESLHDHADELYRSCLARF ( GDDGENL )

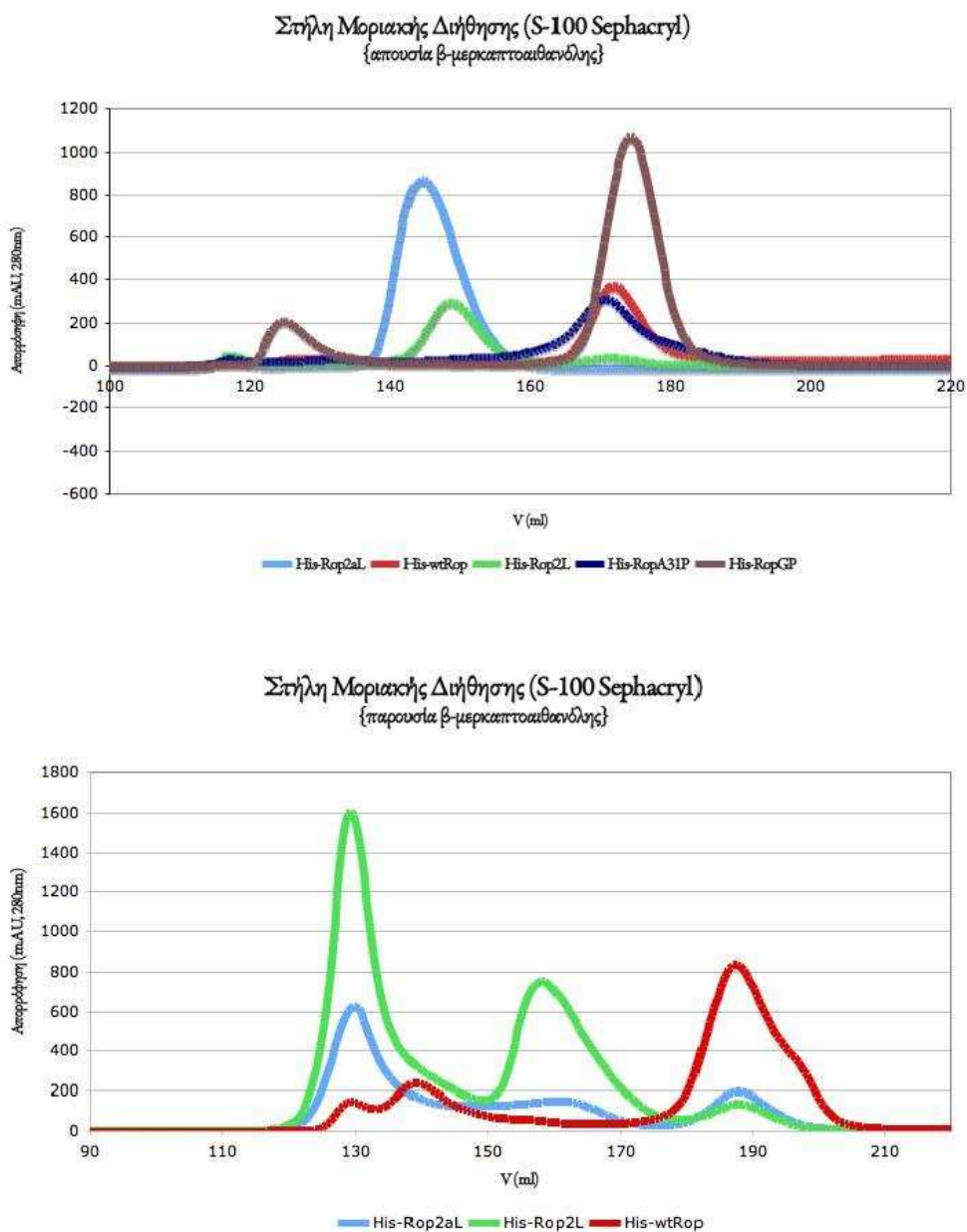
Εικόνα 1.3 Πίνακας όπου παρουσιάζονται οι αλληλουχίες της φυσικής Rop (wtRop) και των αντιστοίχων μεταλλαγμάτων της. Στην τρίτη στήλη παρουσιάζεται το σημείο της Rop όπου πραγματοποιούνται οι αλλαγές, ενώ η δεύτερη και τέταρτη στήλη παραμένουν ίδιες. Με μωβ χρώμα είναι τα κατάλοιπα της στροφής της Rop τα οποία στην RM6 αφαιρούνται. Με πράσινο χρώμα απεικονίζονται οι αλανίνες που προστέθηκαν εκατέρωθεν του ασπαρτικού οξέος 30 για να κατασκευαστεί το μόριο της 2αα. Ενώ, με γαλάζιο βλέπουμε στο μέταλλαγμα της Rop2aL(QL) την λευκίνη που αντικατέστησε την γλουταμίνη στην θέση 36.

### 1.3 Χαρακτηριστικά του μεταλλάγματος Rop2aL (QL) από πειραματικά δεδομένα του εργαστηρίου Κρήτης

Την αμινοξική αντικατάσταση της γλουταμίνης από την λευκίνη επιμελήθηκαν πειραματικά στο βιοτεχνολογικό εργαστήριο της Κρήτης οι κύριοι Κυριαζίδης και Κοκκινίδης στα πλαίσια της εργασίας τους πάνω στο ρόλο της στροφής και του μοτίβου της επτάδας στην αναδίπλωση και σταθερότητα της πρωτεΐνης Rop μαζί και με άλλα παραπλήσια μεταλλάγματα (Kyriazidis, Master Thesis, 2006).

Αρχικά, για τη δημιουργία του γονιδίου της Rop2aL (QL) χρησιμοποίησαν ένα σύστημα κατευθυνόμενης μεταλλαξιγένεσης το οποίο επιτρέπει την απευθείας μετάλλαξη του γονιδίου-στόχου. Έπειτα κλωνοποίησαν το γονίδιο, σε πλασμιδιακούς φορείς έκφρασης, που χρησιμοποιούν το σύστημα Lac για υπερέκφραση του γονιδίου-στόχου και στη συνέχεια κάνοντας επαγωγή, πέτυχαν υπερπαραγωγή της αντίστοιχης πρωτεΐνης. Ο πλασμιδιακός φορέας προσθέτει επιπλέον στο καρβοξυτελικό άκρο των πρωτεϊνών έναν επίτοπο έξι ιστιδινών (His-tag) ο οποίος είναι απαραίτητος στο πρώτο στάδιο καθαρισμού των μεταλλαγμάτων με χρωματογραφία συγγένειας και έπειτα αποκόπτονται για την κρυστάλλωση. Στη συνέχεια απομόνωσαν τα μεταλλάγματα σε δυο στάδια καθαρισμού χρησιμοποιώντας χρωματογραφία συγγένειας και αργότερα χρωματογραφία μοριακής διήθησης. Τέλος τα κλάσματα που περιείχαν το εκάστοτε μετάλλαγμα σε καθαρότητα >95% συλλέχθηκαν και χρησιμοποιήθηκαν σε πειράματα κυκλικού διχρωϊσμού για να προσδιορίσουν την δευτεροταγή δομή της πρωτεΐνης καθώς και σε πειράματα κρυστάλλωσης.

Τα αποτελέσματα από τα πειράματα βοήθησαν στην εύρεση χαρακτηριστικών της δομής του κάθε μεταλλάγματος. Αρχικά, από την χρωματογραφία μοριακής διήθησης υπολόγισαν τα φαινομενικά μοριακά τους βάρη συμπεραίνοντας στοιχεία για τον ολιγομερισμό τους. Η φυσική Rop εκλουόταν σε όγκο που αντιστοιχεί σε διμερή μορφή του μορίου ενώ το μετάλλαγμα της Rop2aL εκλουόταν σε όγκο που αντιστοιχούσε σε υψηλότερο μοριακό βάρος από την φυσική Rop συγκεκριμένα διπλάσιο και σχεδόν παρόμοιο με της RM6. Τα αποτελέσματα αυτά διαφοροποιούνταν από την ύπαρξη ή όχι αναγωγικού (β-μερκαπτοαιθανόλης). Στην περίπτωση της Rop2aL, παρουσία αναγωγικού, η πρωτεΐνη εκλουόταν σε δύο πληθυσμούς που αντιστοιχούσαν σε διμερή (188ml) και τετραμερή (161ml) μορφή, ενώ απουσία του αναγωγικού εμφανιζόταν ένας και μόνο πληθυσμός στα 145 ml δηλαδή τετραμερή μορφή (εικόνα 1.4)

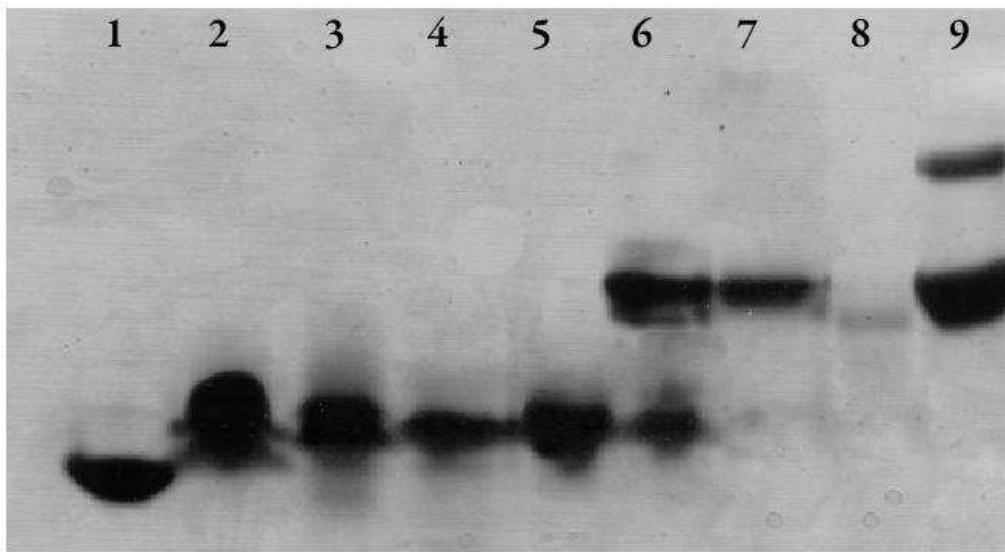


**Εικόνα 1.4** Αποτελέσματα από την χρωματογραφία μοριακής διήθησης. Παρουσιάζονται οι όγκοι έκλουσης των διαφόρων μεταλλαγμάτων που μελετήθηκαν στην Κρήτη. Ορισμένα μεταλλάγματα εκλούνται στον ίδιο όγκο με τη φυσική Rop το οποίο δηλώνει διμερή μορφή ενώ άλλα φαίνεται να εκλούνται σε μικρότερο όγκο άρα μεγαλύτερο μοριακό βάρος. Ωστόσο, τα δεδομένα διαφοροποιούνται με την παρουσία ή όχι αναγωγικού (β-μερκαπτοαιθανόλη). (Αναπαράγεται με άδεια από τους Kyriazidis et al. Master Thesis, 2006)

Έπειτα πραγματοποιήθηκε ηλεκτροφόρηση των μεταλλαγμάτων κάτω από μη αποδιατακτικές συνθήκες (native PAGE), για να ελέγξουν εάν έχει επιτευχθεί η αποκατάσταση του μοτίβου της επιτάδας χρησιμοποιώντας ως μάρτυρα τη φυσική Rop.



Παρατήρησαν, (εικόνα 1.5) ότι στην native ηλεκτροφόρηση η τετραμερής μορφή της Rop2aL (στήλη 7, εικόνας 1.5) έτρεχε υψηλότερα από την φυσική Rop (σημείο διαφοροποίησης για την αποκατάσταση της επτάδας), ενώ η διμερής (στήλη 6) που εμφάνιζε δύο πληθυσμούς ο ένας έτρεχε κοντά στο ίδιο επίπεδο με την Rop, ενώ ο άλλος στο ίδιο με την τετραμερή μορφή της. Το γεγονός ότι στην στήλη 6 Rop2aL(2mer) εμφανίζεται και διμερής και τετραμερής πληθυσμός δηλώνει ότι υπάρχει ισοροπία και η δομή «τείνει» να γίνει τετραμερής. Οι πληθυσμοί αυτοί προκύπτουν από την ύπαρξη β-μερκαπτοαιθανόλης στο ρυθμιστικό διάλυμα της στήλης μοριακής διήθησης.



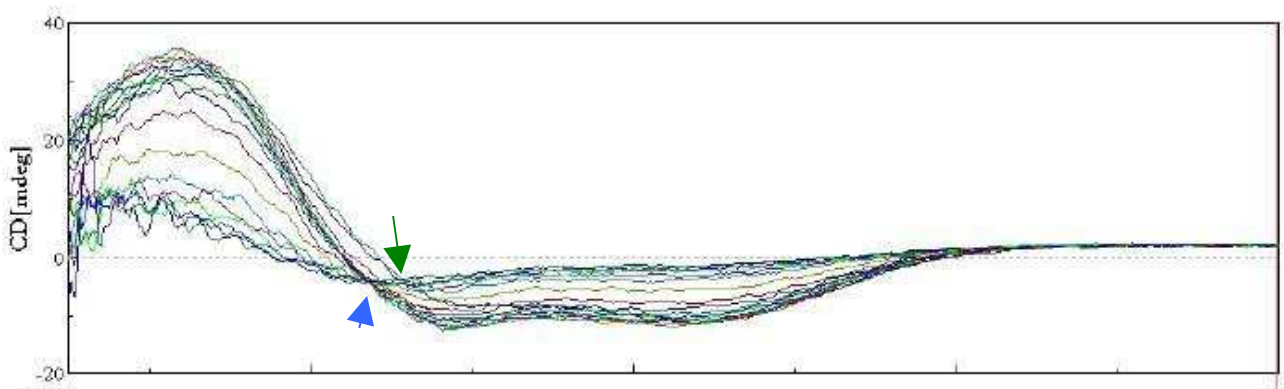
**Εικόνα 1.5** Εγγενή ηλεκτροφόρηση Native-PAGE διαφόρων μεταλλαγμάτων της Rop.

1: Η φυσική wtRop που χρησιμοποιήθηκε ως μάρτυρας, 2:His-RopA31P, 3:His-RopPP, 4:His-RopGP, 5:His-RopPG, 6:His-Rop2aL(2mer), 7:His-Rop2aL(4mer), 8:His-Rop2L(4mer), 9:His-Rop2L(6mer).

Τα μεταλλάγματα στις στήλες 2,3,4,5,7,8,9 μελετήθηκαν στα πλαίσια της μεταπτυχιακής εργασίας του Κυριαζίδη, 2006. Το ενδιαφέρον της παρούσας εργασίας περιορίζεται στα μεταλλάγματα των 6 και 7 στηλών. Παρατηρούμε ότι τα μεταλλάγματα που είχαν ως αρχικό τους στόχο την αποκατάσταση του μοτίβου της επτάδας (μεταλλάγματα στις 6,7,8,9) εμφάνιζαν πληθυσμούς με μεγαλύτερο μοριακό βάρος (MB) από τη διμερή μορφή του μάρτυρα His-wtRop. (Αναπαράγεται με άδεια από Kyriazidis, Master Thesis, 2006)

Από τα πειράματα του κυκλικού διχρωισμού με θερμική αποδιάταξη βρέθηκε το σημείο τήξης της πρωτεΐνης Rop2aL και ταυτόχρονα έγινε καταγραφή φασμάτων στο άπω υπεριώδες ώστε να παρακολουθήσουν την πορεία αναδιπλώσης της πρωτεΐνης. Το σημείο τήξης ήταν στους 62°C δηλαδή υψηλότερο από της φυσικής Rop (45°C) γεγονός που μπορεί να οφείλεται στη δημιουργία ενός νέου πιο εκτεταμένου υδρόφοβου πυρήνα (υψηλότερη θερμοκρασταθερότητα των μορίων αυτών). Από το φάσμα της

θερμοεπαγώμενης αποδιάταξης ενισχύεται η υπόθεση για τη τιμή της θερμοκρασία τήξης ( $T_m$ ) που χρησιμοποιήθηκε αρχικά, όμως δεν επαληθεύθηκε με βεβαιότητα το σημείο αποδιάταξης καθώς εμφανιζόταν δύο θερμοκρασίες  $60^\circ\text{C}$  και  $65^\circ\text{C}$  των οποίων οι καμπύλες στο φάσμα απέκλιναν από το ισοδιχρωϊκό σημείο, χαρακτηριστικό των ελικοειδών δεματίων. Η καμπύλη ήταν σιγμοειδής γεγονός που υποδήλωνε ομαλή μετάβαση, από την αναδιπλωμένη στην αποδιαταγμένη μορφή του μεταλλάγματος (εικόνα 1.6) ενώ εμφάνιζε δύο αρνητικές κορυφές (222nm και 208nm) και μια θετική κορυφή (190nm), χαρακτηριστικές τιμές φάσματος για τις  $\alpha$ -έλικες.



**Εικόνα 1.6** Το φάσμα του κυκλικού διχρωισμού για την Rop2aL. Η καμπύλη η οποία είναι σιγμοειδής μας πληροφορεί ότι υπάρχουν δυο καταστάσεις, όμως η μετάβαση του μορίου στην μετουσιωμένη του μορφή δεν είναι ξεκάθαρη. Παρατηρούμε ότι τρεις καμπύλες αποκλίνουν (πράσινο βέλος) από το ισοδιχρωϊκό σημείο (μπλε βέλος) το οποίο είναι χαρακτηριστικό των ελικοειδών δεματίων. (Αναπαράγεται με άδεια από Kyriazidis, Master thesis, 2006)

Τα παραπάνω πειράματα είναι συμβατά με την υπόθεση ότι η δομή της Rop2aL (QL) είναι τετραμερής. Για την εργασία αυτή στόχος είναι να με την τεχνική της ομόλογης μοντελοποίηση να κατασκευή το τετραμερές και με τις προσομοιώσεις μοριακής δυναμικής να ελεγχθεί (χαρακτηρισμός δομής και δυναμικής) η αξιοπιστία του μοντέλου.

# Κεφάλαιο 2

## Ομόλογη μοντελοποίηση

### 2.1 Εισαγωγή

Κατά τη διάρκεια της εξέλιξης, παρατηρούμε ότι η δομή μιας πρωτεΐνης αλλάζει με πολύ πιο αργούς ρυθμούς από ότι η αντίστοιχη αλληλουχία της (Rost, B). Αυτό έχει ως αποτέλεσμα παρόμοιες αλληλουχίες να υιοθετούν πρακτικά ίδιες δομές, αλλά και μακριά συσχετιζόμενες αλληλουχίες να αναδιπλώνονται σε παρόμοιες δομές (Sanchez & Sali, 1997). Η υπόθεση αυτή ενισχύεται και από τις ιδέες των Chothia και Lesk 1986,1992. Έτσι καταλήγουμε στο συμπέρασμα ότι η δομή μιας πρωτεΐνης συντηρείται σε πολύ μεγαλύτερο βαθμό από ότι η αλληλουχία της.

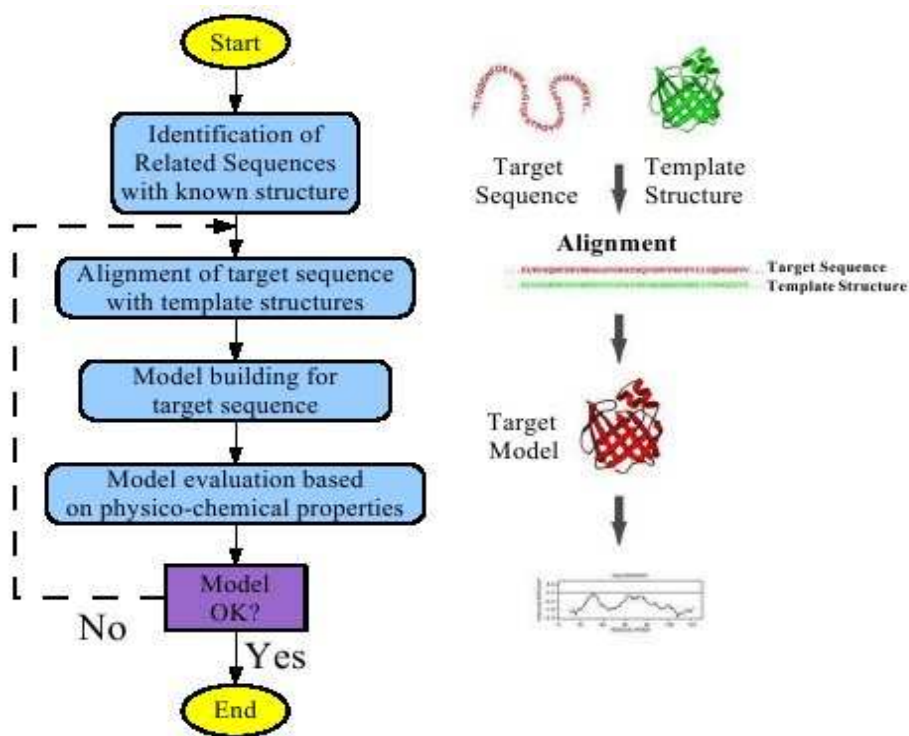
Παρόλο που η πρώτη κρυσταλλογραφικά προσδιορισμένη δομή δημοσιεύτηκε δεκαετίες πριν την πρώτη αλληλούχιση του DNA, ο ρυθμός αύξησης των πειραματικά προσδιορισμένων πρωτεϊνικών δομών ήταν πολύ πιο αργός από τον αντίστοιχο ρυθμό των διαθέσιμων πρωτεϊνικών αλληλουχιών (Sanchez & Sali, 1997). Το γεγονός αυτό οφείλεται στις δυσκολίες και τις απαιτήσεις που εμφανίζουν (όπως χρόνος, πόροι και πειραματικές προκλήσεις) οι πειραματικές μέθοδοι κρυσταλλογραφίας ακτινών-Χ (X-ray) και NMR φασματοσκοπίας σε σχέση με τις αυτοματοποιημένες μεθόδους αλληλούχισης. Κατά συνέπεια δημιουργήθηκε ένα μεγάλο κενό της πληροφορίας στις συγκρίσεις του μεγέθους των αντίστοιχων βάσεων δεδομένων (Rost, B). Προκειμένου να γεφυρωθεί αυτό το χάσμα έγιναν προσπάθειες οι οποίες επικεντρώθηκαν σε θεωρητικές προσεγγίσεις για την πρόγνωση δομής και λειτουργίας των πρωτεϊνών χρησιμοποιώντας υπολογιστικές μεθόδους (Eswar et al., 2006).

Συγκεκριμένα, η πρόγνωση τριτοταγούς δομής είναι μια δύσκολη διαδικασία και δεν υπάρχει ακόμα μια και μόνο μέθοδος η οποία να πληρεί όλα εκείνα τα χαρακτηριστικά τα οποία θα είναι ικανά να προσδιορίσουν μια επιτυχημένη πρόγνωση. Πρόκειται για μια πολύπλοκη και τεχνικά απαιτητική μέθοδο, απαιτεί καθαρά υπολογιστικές τεχνικές και για το λόγο αυτό έχουν αναπτυχθεί τρεις βασικές κατηγορίες αλγορίθμων. Η πρώτη, η ομολογη μοντελοποίηση (Homology modeling), χρησιμοποιεί γνωστές πρωτεϊνικές δομές ως πρότυπα για τη δημιουργία μοντέλων των άγνωστων δομών. Η δεύτερη, γνωστή ως Threading (fold recognition) αναζητά δομές με παρόμοια αναδίπλωση, χωρίς προφανή ομοιότητα ακολουθίας, ενώ η τρίτη είναι αυτοούσιο το υπολογιστικό τμήμα του προβλήματος της αναδίπλωσης των πρωτεϊνών (Ab initio structure prediction) (Xiang, 2006). Στην προκειμένη περίπτωση, για την εύρεση της τριτοταγούς δομής του μεταλλάγματος της Rop, το Rop2aL (QL), θα εφαρμόσουμε την τεχνική της ομολογης μοντελοποίησης όπου για την κατασκευή της δομής της θα χρησιμοποιήσουμε ως πρότυπη δομή, την γνωστή δομή της RM6.

## 2.2 Ομολογη μοντελοποίηση και η διαδικασία της

Μια από τις πιο ισχυρές τεχνικές πρόγνωσης τριτοταγούς δομής είναι η ομολογη μοντελοποίηση. Η επιτυχία της μεθόδου -αλλά και η αποτυχία της- βασίζεται όπως προαναφέρθηκε, στο γεγονός ότι η δομή της πρωτεΐνης είναι περισσότερο συντηρημένη από την αλληλουχία της. Ωστόσο, έχουν αναφερθεί περιπτώσεις στις οποίες παρόλο που υπήρχε υψηλή ταυτότητα αλληλουχίας, μια σημειακή μετάλλαξη μπορεί να προκαλέσει μεγάλη αλλαγή στην τοπολογία της πρωτεΐνης (Glykos et al., 1999). Η βαρύτητα της ομολογης μοντελοποίησης ως μεθόδου πρόγνωσης τριτοταγών πρωτεϊνικών δομών αυξάνεται σταθερά, λόγω του χάσματος που υπάρχει μεταξύ του μεγάλου αριθμού διαθέσιμων πρωτεϊνικών αλληλουχιών σε σχέση με τις πειραματικά λυμένες δομές και φυσικά χάρη στη συνεχώς αυξανόμενη αξιοπιστία και ακρίβεια της μεθόδου (Xiang et al., 2006). Η τεχνική αυτή αφορά την κατασκευή και ανάλυση πρωτεϊνικών μοντέλων και συγκεκριμένα προβλέπει την τριτοταγή δομή μιας πρωτεϊνικής αλληλουχίας στόχου (target) βασιζόμενη στις δομικές πληροφορίες ομολογων πρωτεϊνών που αντλούνται από τη στοιχισή του στόχου με μια ή περισσότερες πρωτεΐνες γνωστής δομής (template) (Sali, 1995)

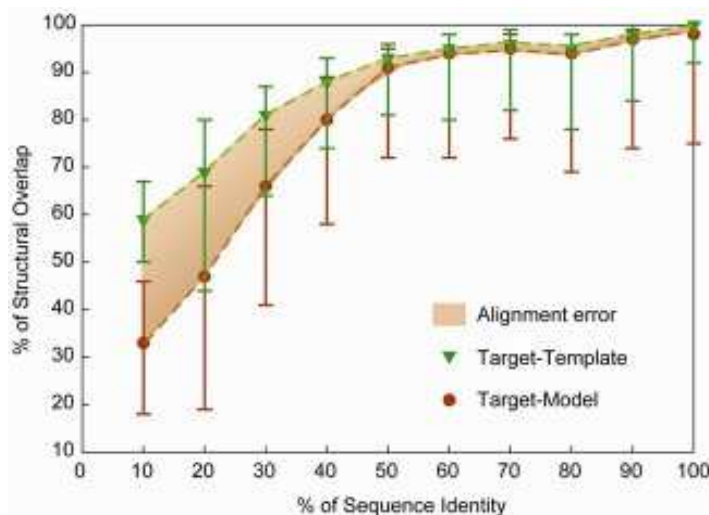
Όσον αφορά τη διαδικασία της τεχνικής αυτής, όπως απεικονίζεται και στην εικόνα 2.1, αποτελείται από τέσσερα βασικά διαδοχικά βήματα (Sanchez et., 1997). Αρχικά, απαιτείται η εύρεση όσο το δυνατόν περισσότερων πρωτεϊνών με γνωστή δομή οι οποίες είναι ομόλογες με την αλληλουχία της πρωτεΐνης στόχου. Στη συνέχεια, παράγεται η καλύτερη δυνατή στοίχιση μεταξύ της άγνωστης αλληλουχίας-στόχου και των αλληλουχιών των πρότυπων δομών. Έπειτα, δημιουργείται το μοντέλο του πρωτεϊνικού σκελετού της πρωτεΐνης-στόχου χρησιμοποιώντας ως μήτρα τις ήδη γνωστές δομές. Στις περιοχές που δεν μπόρεσε να επιτευχθεί ακριβής μοντελοποίηση, δηλαδή όπου υπάρχουν κενά είτε στο στόχο είτε στο πρότυπο (με αποτέλεσμα την εμφάνιση εισαγωγών-διαγραφών), τα κενά αυτά μοντελοποιούνται με τη μορφή βρόχων αντίστοιχου μήκους (loop modeling). Ακολούθως, προστίθενται οι πλευρικές ομάδες των αμινοξέων στον πρωτεϊνικό σκελετό, βελτιστοποιείται η γεωμετρία τους και κατ' επέκταση και ολόκληρη η δομή με μεθόδους ελαχιστοποίησης ενέργειας (energy minimization) (Sanchez & Sali, 1997; Fiser & Sali, 2003).



**Εικόνα 2.1** Στην εικόνα αριστερά παρουσιάζονται τα 4 βασικά βήματα της ομόλογης μοντελοποίησης με διάγραμμα ροής, ενώ δεξιά βλέπουμε τη διαδικασία αυτή σχηματικά. (Αναπαράγεται άνευ αδειάς από τους Sali et al., 1993).

Παρόλο που αυτή η ιδέα είναι ξεκάθαρη και έχει εφαρμοστεί με επιτυχία σε πολλές πρωτεΐνες (π.χ. Blundell et al., 1983; Chothia et al., 1986; Greer, 1985a; Palmer et al., 1986;), δεν υπάρχει μια γενική ομοφωνία για το ποιος είναι ο καλύτερος τρόπος, ώστε να τεθεί σε εφαρμογή (Havel et al., 1990).

Το πρώτο πρόβλημα είναι να αποφασίσουμε με δεδομένη μια στοίχιση, ποια είναι η «καλύτερη», -πλέον πιθανή-, δομή για τη νέα αλληλουχία η οποία θα βασίζεται σε αυτή τη στοίχιση και στις δομές που υπάρχουν. Για την επίτευξη αυτού του στόχου απαιτείται να προσέξουμε το μοτίβο αναδίπλωσης της νέας αλληλουχίας να είναι το ίδιο με αυτό των αλληλουχιών γνωστής δομής. Επιπλέον, επιβάλλεται, όπου υπάρχουν μεταβολές στις αποστάσεις μεταξύ των στοιχείων δευτεροταγούς δομής (αφαιρέσεις ή προσθέσεις), να εισάγονται βρόχοι με κατάλληλο αριθμό καταλοιπών, ώστε να αποκατασταθεί το μήκος της αλληλουχίας. Τέλος, απαιτείται να προσέξουμε την ταυτότητα αλληλουχίας των δομών, για την οποία γνωρίζουμε, σύμφωνα με την εικόνα 2.2, ότι όσο υψηλότερη είναι η ταυτότητα αλληλουχίας με την πρότυπη, τόσο μεγαλύτερη ακρίβεια εμφανίζει το μοντέλο (Madhusundhan et al., 2005). Από την άλλη, όταν μειώνεται η ταυτότητα αλληλουχίας μεταξύ της αλληλουχίας στόχου και της πρότυπης δομής, η μέση ομοιότητα δομών μεταξύ προτύπου και στόχου μειώνεται κι αυτή (Sanchez & Sali 2000). Από την βιβλιογραφία προκύπτει ότι το ποσοστό ταυτότητας αλληλουχίας, τουλάχιστον 30%, αποτελεί ένα καλό δείκτη για την αναμενόμενη ακρίβεια του προγνωστικού μοντέλου (Marti-Renom, et al 2000).



Εικόνα 2.2 Η μέση ακρίβεια του υπολογιζόμενου μοντέλου ως συνάρτηση της ταυτότητας αλληλουχίας προτύπων-στόχου. Η μέση επικάλυψη της πειραματικά προσδιορισμένης πρωτεϊνικής δομής (Target) με το υπολογιστικό μοντέλο (Model) (κάτω γραμμή) και με την πρότυπη δομή (Template) στην οποία το μοντέλο ήταν βασισμένο (πάνω γραμμή), παρουσιάζεται σαν συνάρτηση της ταυτότητας αλληλουχίας, στόχου-προτύπων.

Αυτή η ταυτότητα αλληλουχίας υπολογίζεται από την στοίχιση των δομών. Η επικάλυψη δομών ορίζεται ως το μέρος των ισοδύναμων ατόμων Ca μετά από άκαμπτη υπέρθεση των δύο δομών. Δύο Ca άτομα θεωρούνται ισοδύναμα εάν βρίσκονται μέσα σε 3.5Å σε απόσταση το ένα από το άλλο. Τα σημεία των καμπύλων αντιστοιχούν στις μεσαίες τιμές, και οι μπάρες λάθους στις θετικές και αρνητικές κατευθύνσεις αντιστοιχούν στις μέσες θετικές και αρνητικές διαφορές από τη διάμεσο, αντιστοίχα. Η σκιασμένη περιοχή μεταξύ των δύο καμπύλων αντιστοιχεί περίπου στο πρότυπο λάθος που προκύπτει από λάθη στοίχισεων. (Αναπαράγεται χωρίς άδεια από M.S. Madhusundhan et al., 2005)

Εκτός από τους παραπάνω παράγοντες οι οποίοι έχουν πολύ μεγάλη επιρροή στην ακρίβεια του μοντέλου, σημαντικό ρόλο παίζει και ο τρόπος κατασκευής του μοντέλου. Χρειάζεται να αναφέρουμε ότι υπάρχουν τρεις διαφορετικές μέθοδοι με τις οποίες υπολογίζεται η τρισδιάστατη δομή μέσω της στοιχίσης. Η πλέον συνηθισμένη και αρχική μέθοδος είναι η μοντελοποίηση με συναρμολόγηση-συνάθροιση στερεών σωμάτων (*rigid bodies assembly*) (Johnson et al., 1994). Η μέθοδος αυτή κατασκευάζει τη δομή μιας πρωτεΐνης διατηρώντας σταθερά κάποια τμήματα της στα οποία συμπεριλαμβάνονται περιοχές του πυρήνα, βρόχοι και πλευρικές ομάδες, που προκύπτουν από την κατάτμηση σχετικών πρωτεϊνικών δομών.

Η άλλη ομάδα μεθόδων μοντελοποίησης είναι ο συνδυασμός τμημάτων (*segment matching* ή *coordinate reconstruction*), που βασίζεται στις κατά προσέγγιση θέσεις των συντηρημένων ατόμων από την πρότυπη δομή, για να υπολογιστούν οι συντεταγμένες των υπόλοιπων ατόμων. Αυτό επιτυγχάνεται με τη χρήση μιας βάσης δεδομένων με μικρά τμήματα πρωτεϊνικών δομών, με ενεργειακούς και γεωμετρικούς κανόνες ή με το συνδυασμό αυτών των κριτηρίων (Levitt et al., 1992).

Η τρίτη ομάδα των μεθόδων είναι η μοντελοποίηση μέσω εκπλήρωσης χωροταξικών περιορισμών χρησιμοποιώντας είτε γεωμετρικές αποστάσεις είτε τεχνικές βελτιστοποίησης, ώστε να πληρούνται οι χωρικοί περιορισμοί οι οποίοι αποκτώνται από τη στοιχίση της αλληλουχίας-στόχου με τις σχετικές πρότυπες γνωστές δομές (Sali & Blundell, 1993). Δεδομένου ότι αυτή η μοντελοποίηση βασίζεται σε περιορισμούς, είναι δυνατόν να χρησιμοποιηθούν πολλοί διαφορετικοί τύποι πληροφοριών για την αλληλουχία στόχο γεγονός που αποδεικνύει ότι είναι ίσως η πιο πολλά υποσχόμενη όλων των τεχνικών ομόλογης μοντελοποίησης (Sali, 1995).

Τελικά, κύριος στόχος της ομόλογης μοντελοποίησης είναι να βρεθεί η δομή εκείνη ή μια οικογένεια συγγενών δομών που θα συμφωνεί καλύτερα με το σύνολο των περιορισμών. Ωστόσο, στη δική μας περίπτωση δε χρησιμοποιήσαμε ως πρότυπη δομή για την αλληλουχία μας τη δομή με την υψηλότερη ομοιότητα (*sequence identity*) και αυτό γιατί, αυτή είναι η δομή της 2αα (1nkd: pdb code) η οποία εμφανίζει δομή παρόμοια με τη φυσική Rop. Αντίθετα, με την πρωτεΐνη μας, η οποία όπως αναφέραμε στο πρώτο κεφάλαιο, φαίνεται να είναι ένα ομοτετραμερές 4-α-ελικοειδές δεμάτιο με το κάθε μονομερές της να είναι μια επιμηκυμένη α-έλικα που θα αντιστοιχεί στο μονομερές της φυσικής Rop χωρίς τον χαρακτηριστικό του βρόχο. Η δομή που ταιριάζει καλύτερα στην τοπολογία αυτή είναι της RM6, την οποία χρησιμοποιήσαμε και ως πρότυπη δομή.

## 2.3 Ενδεδειγμένα προγράμματα ομολογής μοντελοποίησης

Για την πρόγνωση πρωτεϊνικής δομής από μια πρωτεϊνική αλληλουχία υπάρχει μεγάλος αριθμός εργαλείων που μπορούν να προβλέψουν την τριτοταγή δομή σε ικανοποιητικό βαθμό. Στα πλαίσια της παρούσας εργασίας χρησιμοποιήθηκαν δύο κυρίαρχα προγράμματα που αφορούν την ομολογική μοντελοποίηση (Homology modeling), το διαδικτυακό Swiss-model και το Modeller.

## 2.4 Swiss-model και τα χαρακτηριστικά του

Αρχικά, είναι αναγκαίο να αναφέρουμε ότι χρειάζεται να διαθέτει κανείς σημαντική εμπειρία στη δομική βιολογία και ικανότητα χειρισμού ιδιαίτερα εξειδικευμένων προγραμμάτων υπολογιστών για κάθε ένα από τα μεμονωμένα βήματα της διαδικασίας της μοντελοποίησης, ώστε να είναι σε θέση να χειρίζεται πρωτεϊνικές δομές (Tramontano et al., 2001). Γι' αυτό το λόγο αναπτύχθηκε η ιδέα μιας εύχρηστης αυτοματοποιημένης μηχανής μοντελοποίησης από τους Peitsch και συνεργάτες (Peitsch et al., 1993; 1995; 1996), η οποία εφαρμόστηκε αρχικά πριν από μία δεκαπενταετία θέτοντας τις βάσεις για τη δημιουργία του Swiss Model. Από τότε, το Swiss-Model αναπτύσσεται συνεχώς και σήμερα συνιστά ένα πρόγραμμα ομολογής μοντελοποίησης, το οποίο χρησιμοποιείται ευρύτατα (Geux et al., 1999; Geux & Peitsch, 1997; Schwede et al., 2003) και διατίθενται ελεύθερα για χρήση στο διαδίκτυο.

Η αυτοματοποιημένη διαδικασία που χρησιμοποιείται από το Swiss-Model μιμείται τα βασικά βήματα του homology modeling (Peitsch, 1996) και είναι σχεδιασμένο να λειτουργεί με τον ελάχιστο αριθμό αρχείων εισαγωγής από το χρήστη. Το Swiss-Model δίνει στο χρήστη τη δυνατότητα επιλογής μεταξύ των τριών τρόπων λειτουργίας που διαθέτει (Schwede et al., 2003; Arnold, 2006) και παρουσιάζονται στη συνέχεια.

Στην πρώτη προσέγγιση, γνωστή ως *'first approach mode'*, χρειάζεται να εισαχθεί μόνο η αμινοξική αλληλουχία της πρωτεΐνης στόχου, ώστε να κατασκευαστεί το τριδιάστατο μοντέλο. Η επιλογή των πρότυπων δομών, η στοιχισή των αλληλουχιών και η κατασκευή του μοντέλου γίνονται αυτόματα από το πρόγραμμα (Schwede et al., 2003). Συγκεκριμένα, όταν υποβληθεί μια αλληλουχία, το Swiss-Model τη συγκρίνει με τα περιεχόμενα της πρωτεϊνικής βάσης δεδομένων PDB και η μοντελοποίηση επιχειρείται



μόνο εάν υπάρχει μια πρωτεΐνη ομόλογη της πρωτεΐνης στόχου με τουλάχιστον 25% ομοιότητα αλληλουχίας σε μια περιοχή μήκους μεγαλύτερου από 20 κατάλοιπα. Στην συνέχεια δημιουργούνται τα ατομικά μοντέλα και γίνεται ελαχιστοποίηση ενέργειας, για να φτιαχτεί το καλύτερο μοντέλο (Gibas & Jambeck 2001).

Το *'alignment mode'*, είναι μια ακόμη διαδικασία μοντελοποίησης η οποία βασίζεται σε μια προκαθορισμένη από το χρήστη στοιχισή μεταξύ στόχου και προτύπων. Ο χρήστης διευκρινίζει ποια αλληλουχία στη δεδομένη στοιχισή είναι η αλληλουχία στόχος και ποια αντιστοιχεί στη δομικά γνωστή πρωτεϊνική αλυσίδα από τη βάση δεδομένων PDB, ώστε το πρόγραμμα να κατασκευάσει το μοντέλο (Schwede et al., 2003).

Στις περιπτώσεις όπου υπάρχουν σύνθετοι στόχοι μοντελοποίησης χρησιμοποιούμε το *'project mode'* το οποίο απαιτεί για τη λειτουργία του το ειδικό πρόγραμμα DeepView (Swiss-PdbViewer). Το project mode επιτρέπει στο χρήστη να υποβάλλει στο Swiss-Model χειροκίνητα το βελτιστοποιημένο αίτημα μοντελοποίησης. Η αφετηρία γι' αυτόν τον τρόπο είναι ένα αρχείο προγράμματος DeepView το οποίο περιέχει τις δομές προτύπων σε υπέρθεση και τη στοιχισή μεταξύ του στόχου και των προτύπων. Επιπλέον, το project mode μπορεί να χρησιμοποιηθεί και επαναληπτικά, ώστε να βελτιώσει την παραγόμενη από το *'first approach mode'* δομή (Schwede et al., 2003). Ειδικότερα, το DeepView υποστηρίζει διάφορες πλατφόρμες και έχει τη δυνατότητα δημιουργίας μοριακών επιφανειών, πολλαπλής στοιχισής ακολουθιών, χρήσης συναρτήσεων βαθμολόγησης (scoring), όπως και πραγματοποίηση απλής και γρήγορης μοντελοποίησης συμπεριλαμβάνοντας και την τοπο-εξαρτώμενη μεταλλαξιγένεση αλλά και την πολύπλοκη μοντελοποίηση, όπως η "επανασύσταση" των βρόχων (Guex & Peitsch, 1997).

## 2.5 Εφαρμογές του Swiss-Model και τα αποτελέσματά τους

Προκειμένου να κατασκευαστεί το μοντέλο της QL με τη διαδικασία της ομόλογης μοντελοποίησης, χρησιμοποιήθηκε σε πρώτη φάση το αυτοματοποιημένο διαδικτυακό πρόγραμμα Swiss-Model. Στη συνέχεια θα παρουσιαστούν ορισμένες εφαρμογές με τη σειρά που πραγματοποιήθηκαν προκειμένου να επιτευχθεί η δημιουργία του μοντέλου. Αρχικά, δοκιμάστηκε η πρώτη λειτουργία του, το *'first approach mode'* στο οποίο αρκεί μονάχα να υποβάλουμε την αμινοξική αλληλουχία της πρωτεΐνης της οποίας την τριδιάστατη δομή επιθυμούμε να κατασκευάσουμε. Επιδιώξαμε, δηλαδή, να δούμε εάν

μπορεί πραγματικά να κατασκευάσει τη δομή της δικής μας ομόλογης πρωτεΐνης, πρώτα όμως με μονομερή μορφή. Αρχικά, υποβάλαμε στο πρόγραμμα την αλληλουχία της QL συμπεριλαμβανομένου και των έξι τελικών καταλοίπων της πρωτεΐνης και μετά από έρευνα των βάσεων δεδομένων πρωτεϊνικών αλληλουχιών, το πρόγραμμα επέλεξε ως κατάλληλη ομόλογη δομή αυτή με τη μεγαλύτερη ομοιότητα συγκριτικά με την αλληλουχία στόχο, την 1rpoA (προσθήκη 2 Ala εκατέρωθεν του Asp31) και κατασκεύασε με ομόλογη μοντελοποίηση την παρακάτω δομή.



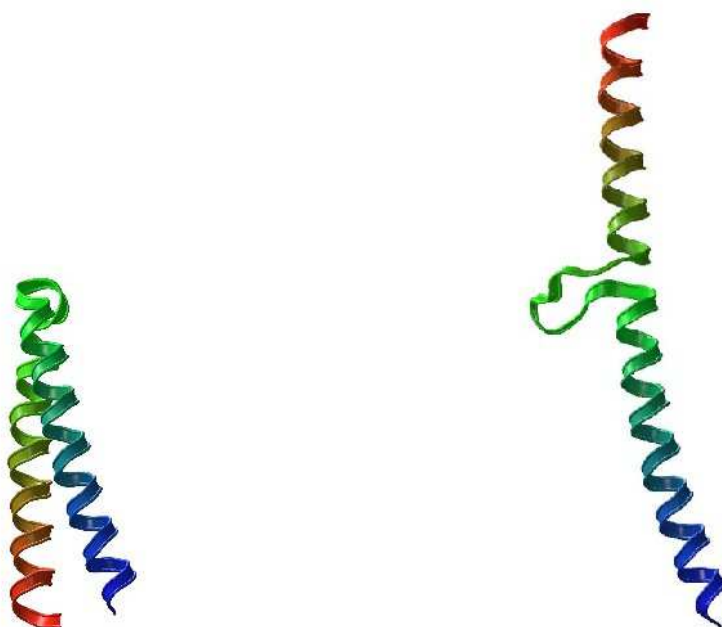
**Εικόνα 2.3** Η πρώτη απόπειρα για την κατασκευή της δομής του μονομερούς της QL με την απλούστερη τεχνική του Swiss Model, μοιάζει με το μονομερές της Rop και όχι με της RM6. (Όλες οι εικόνες έχουν προετοιμαστεί με υπέρθεση στα κατάλοιπα 1 με 15 του αμινοτελικού άκρου της αλυσίδας.

Τα αποτελέσματα της πρώτης έρευνας παρουσιάζονται στην εικόνα 2.3 όπου διαπιστώνουμε ότι η δομή που προέκυψε από το πρόγραμμα μοιάζει αρκετά με τη μονομερή μορφή της Rop. Αυτό όμως δεν είναι ο επιδιωκόμενος στόχος μας. Από πειράματα που έγιναν στο βιοτεχνολογικό εργαστήριο της Κρήτης από τους Κοκκινίδη και συνεργάτες (παράγραφος 1.3), πιθανολογείται ότι στη δομή της QL αποκαθίσταται η περιοδικότητα επτάδας και δεν υπάρχει στροφή στην αλυσίδα αλλά μια συνεχή, ενιαία α-έλικα. Οπότε η δομή που επέλεξε το πρόγραμμα, ενώ έχει μεγάλη ομοιότητα στην αλληλουχία, δεν ικανοποιεί την αλλαγή στη μορφολογία του μονομερούς της πρωτεΐνης που επιδιώκουμε να κατασκευάσουμε.

Στη συνέχεια, αφαιρέσαμε τα επτά τελικά (και ιδιαίτερα ευκίνητα) κατάλοιπα με το σκεπτικό ότι είναι παράλογο να προσπαθούμε να μοντελοποιήσουμε υπολογιστικά κάτι το οποίο δεν προσδιορίστηκε με κρυσταλλογραφία. Σε αυτήν την περίπτωση, το πρόγραμμά επέλεξε ως πρότυπη αλληλουχία για τη μοντελοποίηση τη δομή 1nkd η οποία είναι η Rop2a δηλαδή η αλληλουχία μας, αλλά αντίστροφα στη θέση του L βρίσκεται το Q που

αλλάξαμε σκόπιμα, όπως προαναφέρθηκε (παράγραφος 1.2). Όπως ήταν αναμενόμενο η δομή που προέκυψε από το πρόγραμμα εμφανίζει στροφή, όπως φαίνεται και στην εικόνα 2.4.

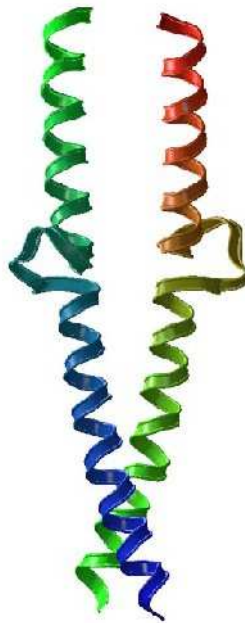
Για το λόγο αυτό στη συνέχεια ορίσαμε στο πρόγραμμα μια συγκεκριμένη δομή, την RM6, την οποία θέλουμε να χρησιμοποιήσει ως οδηγό, για να στηριχθεί και να φτιάξει το μοντέλο μας. Εφαρμόστηκε, δηλαδή, το στάδιο του alignment mode, στο οποίο καθορίζεται από το χρήστη τόσο η στοιχισή όσο και η πρότυπη δομή.



*Εικόνα 2.4 Αριστερά: Προσπάθεια κατασκευής του μονομερούς της QL χωρίς τα επτά τελικά κατάλοιπα της αλυσίδας  
Δεξιά :Με το alignment mode στο οποίο καθορίσαμε ως πρότυπη δομή την RM6, η δομή έγινε επιμηκρωμένη αλλά όχι ενιαία αλυσίδα με δευτεροταγή δομή α-έλικας και η εισαγωγή των επτά καταλοίπων έχει διαμόρφωση τυχαίου σπειράματος.*

Το αποτέλεσμα, όπως καταδεικνύεται στην εικόνα 2.4 (δεξιά), ήταν η μονομερής μορφή της πρωτεΐνης να είναι μια ευθεία αλυσίδα, όπως η RM6, όμως και πάλι το μοντέλο που προέκυψε δεν ήταν το επιθυμητό, γιατί στην περιοχή της εισαγωγής των επτά καταλοίπων δεν εμφανίστηκαν στοιχεία κανονικής δευτεροταγούς δομής, όπως α-έλικα αλλά ένας βρόχος, αντίθετα με το προσδοκώμενο μοντέλο, αλλά σχετικά σύμφωνο με τον τρόπο λειτουργίας του προγράμματος, δηλαδή όταν δεν υπάρχουν στοιχεία για την δομή που θα μοντελοποιήσει, να κατασκευάζει βρόχο (loop modelling).

Σε μια επόμενη προσπάθεια, κρίθηκε σκόπιμο να προχωρήσουμε στο επίπεδο του project mode, για να μπορέσουμε να δώσουμε τους δικούς μας πιο σύνθετους περιορισμούς. Ειδικότερα, επιδιώξαμε το μοντέλο μας να παρουσιάσει τετραμερή μορφή με ενιαίες και τις τέσσερις αλυσίδες χωρίς στροφές και επιπλέον η δευτεροταγής δομή να είναι α-έλικα ακόμη και στην περιοχή της πρώην στροφής, με σκοπό την αποκατάσταση της επτάδας. Για αυτό το στάδιο του project mode, χρησιμοποιήθηκε επικουρικά το συμπληρωματικό πρόγραμμα DeepView, όπου μέσω του περιβάλλοντος εργασίας του, μας δίνεται η δυνατότητα να παρέμβουμε απευθείας στη στοίχιση των αλληλουχιών και να την βελτιστοποιήσουμε. Στο σημείο αυτό είναι αναγκαίο να προσέξουμε να μη γίνεται στοίχιση μεταξύ καταλοίπων διαφορετικών αλυσίδων και να είναι καθορισμένες οι θέσεις των εισαγωγών-διαγραφών. Έπειτα, αποθηκεύσαμε το αρχείο προγράμματος που προέκυψε και το υποβάλαμε στο Swiss-Model.

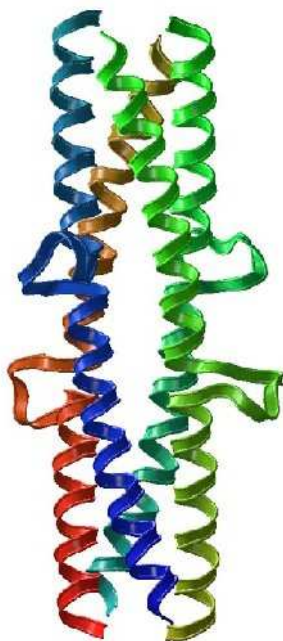


*Εικόνα 2.5 Η δομή που απεικονίζεται είναι διμερής και όχι τετραμερής και τα κατάλοιπα εισαγωγής εμφανίζονται ως βρόχοι, ώστε να διατηρηθεί το μήκος της δομής.*

Παρατηρώντας τα αποτελέσματά του προγράμματος στην εικόνα 2.5, βλέπουμε ότι δημιουργεί μια δομή διμερούς και όχι τετραμερούς μορφής και επιπλέον δεν κατασκευάζει τα κατάλοιπα της εισαγωγής ως α-έλικα αλλά ως βρόχο. Αυτό συμβαίνει ίσως, γιατί η RM6 δεν αναγνωρίζεται από το πρόγραμμα ως τετραμερές μοντέλο, αλλά σύμφωνα με τα στοιχεία που περιλαμβάνονται στο αρχείο pdb της κρυσταλλογραφικά προσδιορισμένης δομής της RM6, αυτή εμφανίζεται ως διμερές διμερών καθώς πειραματικά προσδιορίστηκε το ένα διμερές, ενώ το δεύτερο παρήχθη με εφαρμογή

κρυσταλλογραφικής συμμετρίας. Οπότε το συγκεκριμένο πρόγραμμα χρησιμοποιεί με βάση τη στοιχίση μόνο τις δύο πρώτες αλυσίδες της αλληλουχίας στόχου και δημιουργεί διμερές, σύμφωνα με τα στοιχεία που είχε ως δεδομένα από την πρότυπη δομή.

Αναλογιζόμενοι τα δεδομένα αυτά, αλλάξαμε τον τρόπο εισαγωγής της πρότυπης δομής ώστε να την αναγνωρίζει το πρόγραμμα ως τετραμερές. Γι' αυτό έχουμε επέμβει στο αρχείο το οποίο περιέχει τα δεδομένα και τις συντεταγμένες της RM6.



*Εικόνα 2.6 Η τετραμερής δομή της QL, όπως κατασκευάστηκε από την λειτουργία του Project Mode με την βοήθεια του προγράμματος DeepView, η οποία εξακολουθεί να έχει τα κατάλοιπα εισαγωγής ως βρόχους.*

Το αποτέλεσμα, όπως παρουσιάζεται και στην εικόνα 2.6, είναι ότι το πρόγραμμα κατάφερε με την ομόλογη μοντελοποίηση να κατασκευάσει τη δομή της QL ως τετραμερές, που ήταν ο αρχικός μας στόχος σύμφωνα με την πρότυπη δομή RM6, όμως αδυνατεί και πάλι στο σημείο της εισαγωγής να σχηματίσει τα κατάλοιπα ως α-έλικα.

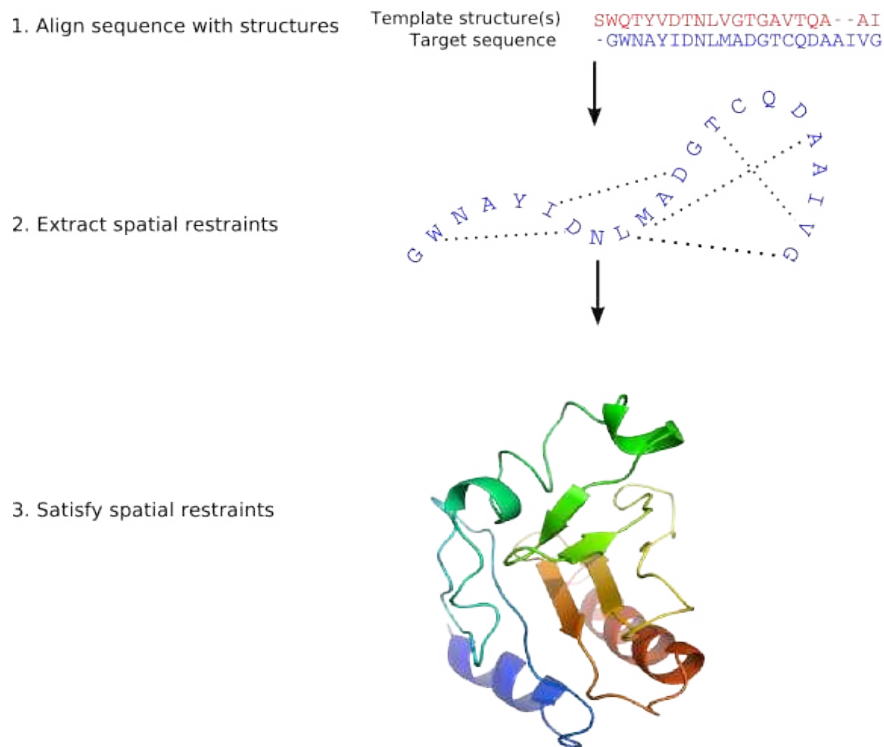
Απόρροια των παραπάνω εφαρμογών είναι ότι το πρόγραμμα Swiss-Model δεν είναι ικανό να κατασκευάσει την επιθυμητή δομή (τουλάχιστον όσο αφορά τα δεδομένα αυτής της εργασίας). Παρόλο που η χρήση του προγράμματος αυτού για ομόλογη μοντελοποίηση φαίνεται απλή, η αυτοματοποίηση προκαλεί στατιστικά λάθη. Μοντελοποίησε βέβαια το τετραμερές, αλλά δεν μας παρείχε τη δυνατότητα να του δώσουμε επιπλέον στοιχεία, ώστε να ορίσουμε τη δευτεροταγή δομή των καταλοίπων στις θέσεις ένθεσης, σύμφωνα με τα οποία θα συγκεκριμενοποιούσαμε τη δευτεροταγή δομή του. Για το λόγο αυτό, λοιπόν, κρίθηκε αναγκαίο να προχωρήσουμε στο επόμενο πρόγραμμα, το Modeller το οποίο, σε αντίθεση με το προηγούμενο, μας επιτρέπει να

επέμβουμε σε οποιοδήποτε στάδιο της διαδικασίας της ομόλογης μοντελοποίησης επιθυμούμε μέσω ενός εύχρηστου εγχειριδίου εντολών (scripting mode).

## 2.6 Το Modeller και τα χαρακτηριστικά του

Το Modeller είναι ένα υπολογιστικό πρόγραμμα το οποίο παρέχει μια αυτοματοποιημένη προσέγγιση για ομόλογη και συγκριτική πρωτεϊνική μοντελοποίηση δομών, διαμέσου της ικανοποίησης χωροταξικών περιορισμών (Sali & Blundell, 1993; Fiser et al., 2000).

Όπως φαίνεται και στην εικόνα 2.7, με την εκτέλεση του προγράμματος αυτού αρχικά πραγματοποιείται η στοίχιση της αλληλουχίας στόχου με την αλληλουχία πρότυπης δομής και ο υπολογισμός της τριτοταγούς δομής της πρωτεΐνης στόχου αυτοματοποιημένα (Sanchez & Sali, 2000). Κατασκευάζεται η δομή της πρωτεΐνης, για την οποία γνωρίζουμε μονάχα την αλληλουχία της, χρησιμοποιώντας ως πρότυπο τα δομικά χαρακτηριστικά από ομόλογες πρωτεϊνικές αλληλουχίες με γνωστή δομή και βασιζόμενο στη στοίχιση τους αποφασίζεται ποια μέρη από την πρότυπη δομή είναι πιθανό να συντηρούνται στην αλληλουχία (Sali & Blundell, 1993). Έπειτα, χρησιμοποιείται η στοίχιση και η πρότυπη δομή, για την εξαγωγή ενός σετ από χωρικούς περιορισμούς (Gibas & Jambeck, 2001) που αφορούν απόστασης μεταξύ ατόμων και διεδρών γωνιών για το στόχο, οι οποίοι στη συνέχεια συνδυάζονται με τους γενικούς δομικούς κανόνες των πρωτεϊνών, όπως μήκοι δεσμών και προτιμήσεις γωνιών. Τελικά, το μοντέλο υπολογίζεται με τη διαδικασία της βελτιστοποίησης κατά την οποία ελαχιστοποιούνται οι αποκλίσεις από τους περιορισμούς βάση ομολογίας (Sali & Blundell 1993; Sanchez & Sali 2000). Με αυτόν τον τρόπο, κατασκευάζεται η τρισδιάστατη δομή της αλληλουχίας στόχου η οποία περιέχει όλα τα βαριά άτομα -και όχι τα υδρογόνα- του σκελετού της πρωτεΐνης και των πλευρικών αλυσίδων των αμινοξέων ικανοποιώντας τους περιορισμούς που αρχικά τέθηκαν βάση ομολογίας (Martin-Remon et al., 2000).



**Εικόνα 2.7** Ομόλογη μοντελοποίηση πρωτεϊνών με ικανοποίηση χωροταξικών περιορισμών. Πρώτον, οι γνωστές πρότυπες τριτοταγείς δομές στοιχίζονται με την αλληλουχία στόχο που θα μοντελοποιηθεί. Δεύτερον, τα χωρικά χαρακτηριστικά, όπως αποστάσεις Ca-Ca, υδρογονοδεσμοί, διεδρες γωνίες κύριας αλυσίδας και πλευρικών αλυσίδων μεταφέρονται από τη δομή οδηγό στην αλληλουχία στόχο. Τρίτον, η τρισδιάστατη δομή κατασκευάζεται με την ικανοποίηση όσο των δυνατών περισσότερων περιορισμών. (Αναπαράγεται χωρίς άδεια από τους Sali & Blundell, 1993).

Οι ομόλογα παραγόμενοι περιορισμοί, που μόλις αναφέραμε, συνδυάζονται έπειτα με στερεοχημικούς περιορισμούς, όπως αποστάσεις δεσμών, γωνίες δεσμών, διεδρες γωνίες και μη δεσμικές επαφές ατόμου με άτομο, που λαμβάνονται σύμφωνα με το μοριακό δυναμικό CHARMM-22 (MacKerell & Karplus 1998; Fiser & Sali 2003). Επιπλέον, υπολογίζουμε στατιστικούς περιορισμούς οι οποίοι προκύπτουν εμπειρικά από το σύνολο των γνωστών πρωτεϊνικών δομών που εμπεριέχονται σε βάσεις δεδομένων (Sali & Overington, 1994; Shen & Sali, 2006). Τέλος, επιπρόσθετοι περιορισμοί προέρχονται από ένα μεγάλο αριθμό διαφορετικών πηγών οι οποίες μπορούν εύκολα να προστεθούν στους ομόλογα παραγόμενους περιορισμούς. Τέτοιου είδους περιορισμοί μπορεί να παραχθούν από πειράματα NMR φασματοσκοπίας (Sutcliffe et al., 1992), κανόνες για πακετάρισμα δευτεροταγούς δομής (Taylor, 1993), ανάλυση υδροφοβικότητας (Aszodi, 1996) και σχετικές μεταλλάξεις (Gobel et al., 1994), πειράματα διασταυρώσεων (Cross-linking) (Rossi et al., 1995), φασματοσκοπία φθορισμού (fluorescence spectroscopy) και πολλοί άλλοι τρόποι (Sali, 1995).

Το ιδιαίτερο χαρακτηριστικό του προγράμματος αυτού είναι πως δεν περιέχει γραφικό περιβάλλον, αλλά απαιτείται η γνώση χρήσης γραμμής εντολών

χρησιμοποιώντας την γλώσσα προγραμματισμού python. Χρειάζεται ως βασικά αρχεία εισαγωγής τη στοιχισή της αλληλουχίας που επρόκειτο να μοντελοποιηθεί με τη δομή οδηγό (αρχείο: alignment.ali), την αλληλουχία στόχο, τις ατομικές συντεταγμένες των πρότυπων δομών και ένα απλό αρχείο script (default.py). Τότε το modeller αυτόματα υπολογίζει το δομικό μοντέλο της πρωτεΐνης, χωρίς την επέμβαση του χρήστη. Στο τέλος το πρόγραμμα μας παρέχει το pdb αρχείο που περιέχει την προβλεπόμενη δομή, καθώς και άλλα αρχεία στα οποία περιγράφονται τα δεδομένα της διαδικασίας της ομόλογης μοντελοποίησης (Gibas & Jambeck, 2001).

Αξίζει να σημειωθεί ότι το Modeller μπορεί να χρησιμοποιηθεί σε όλα τα στάδια της ομόλογης μοντελοποίησης, διαμέσου των ενσωματωμένων εργαλείων του και εκτός από τη μοντελοποίηση γενικά μπορεί να διεκπεραιώσει και βοηθητικά-ειδικότερα μέρη της, όπως εύρεση πρότυπων δομών, πολλαπλές στοιχίσεις, ομαδοποίηση δομών και αλληλουχιών και τέλος μοντελοποίηση βρόχων.

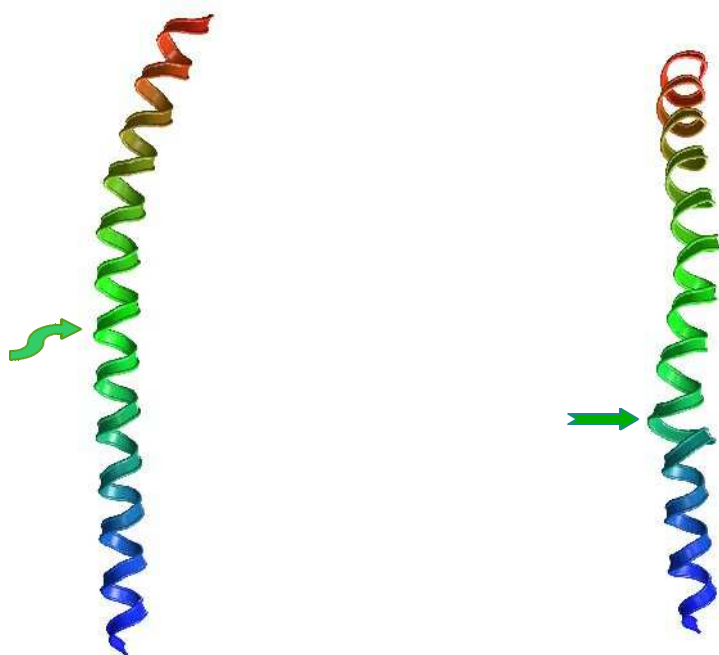
## **2.7 Εφαρμογές του Modeller και τα αποτελέσματά τους**

Για την επόμενη απόπειρα κατασκευής της ομοτετραμερούς QL πρωτεΐνης χρησιμοποιήσαμε το Modeller. Όπως ήδη έχουμε αναφέρει, το πρόγραμμα αυτό δουλεύει μέσω γραμμής εντολών και δεν έχει γραφικό περιβάλλον οπότε δίνουμε στο πρόγραμμα τα επιθυμητά στοιχεία μέσω script. Για να μελετήσουμε τις δομές που μας ετοιμάζει, χρησιμοποιούμε κάποιο πρόγραμμα μοριακής απεικόνισης, όπως το VMD ή το RasMol.



## 2.8 Αρχικές εφαρμογές και παραγωγή του πρώτου μοντέλου (πλεξούδα)

Ξεκινήσαμε να δοκιμάζουμε τις δυνατότητες του Modeller στην ομόλογη μοντελοποίηση, αρχικά με απλές αντικαταστάσεις αμινοξέων, με εισαγωγές δύο έως επτά καταλοίπων στο μονομερές της RM6, ώστε να δούμε τις μεταβολές της δομής σε κάθε περίπτωση. Μελετώντας τις δομές που κατασκευάζει το πρόγραμμα παρατηρούμε ότι σε σημειακές μεταλλάξεις, η δευτεροταγής δομή της αλυσίδας δεν επηρεάζεται από την απλή αντικατάσταση, όπως δείχνει και η εικόνα 2.8α. Όμως στις περιπτώσεις εισαγωγής καταλοίπων το πρόγραμμα προσθέτει τα κατάλοιπα στην αλυσίδα χωρίς τη δημιουργία κανονικών στοιχείων δευτεροταγούς δομής α-έλικας αλλά ως βρόχους δημιουργώντας μια μικρή κάμψη της αλυσίδας, όπως φαίνεται και στην εικόνα 2.8β. Η κάμψη αυτή οφείλεται στη τήρηση των περιορισμών αποστάσεων από το πρόγραμμα.



*Εικόνα 2.8* Συγκριτική απεικόνιση του μονομερούς της RM6 με την αλλαγή Leu->Arg στη θέση 23 αριστερά και δεξιά το μοντέλο μετά την εισαγωγή των 2 καταλοίπων. Οι αλλαγές έγιναν στα σημεία όπου υποδεικνύουν τα βελάκια .

Όπως αναφέραμε, η δομή της QL εμφανίζεται με δύο στροφές παραπάνω από την πρότυπη δομή της RM6 λόγω των επτά επιπλέον καταλοίπων. Έτσι, οι επόμενες μελέτες μας κατευθύνθηκαν προς την εισαγωγή αυτών των επτά καταλοίπων. Ξεκινώντας τις προσπάθειες μοντελοποίησης, αρχίσαμε τη μορφοποίηση του μονομερούς και στη συνέχεια προχωρήσαμε στο τετραμερές. Χρησιμοποιήσαμε ως πρότυπη δομή το μονομερές

της RM6 (1QX8a), ενώ η αλληλουχία στόχος μας είναι το μονομερές της QL (2aaQLa) οι αλληλουχίες των οποίων παρουσιάζονται στο πρώτο κεφάλαιο (σχεδιάγραμμα 1.1). Στοιχίζουμε τις δύο αλληλουχίες (εικόνα 2.9) και τρέχουμε το απλό script (εικόνα 2.10) στο οποίο καθορίζουμε τα αρχεία εισαγωγής ώστε να κατασκευαστεί η δομή.

```
>P1;1QX8a
structureX:1QX8a:5      :A:58  :A::  :
EKTALNMARFIRSQTLTLLEKLNELAD-----ICESLHDHADELYRSCLARF*
```

```
>P1;2aaQLa
sequence:2aaQLa:1      : :54   :A::  :
EKTALNMARFIRSQTLTLLEKLNELADAADELADICESLHDHADELYRSCLARF*
END
```

**Εικόνα 2.9** Εδώ παραθέτουμε το αρχείο 'alignment\_aQL.ali' που περιέχει τη στοίχιση του μονομερούς της RM6(1QX8a) και του μονομερούς της αλληλουχίας που θέλουμε να μοντελοποιήσουμε της 2aaQL.

```
# Homology modeling by the automodel class
from modeller import *          # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class

log.verbose() # request verbose output
env = environ() # create a new MODELLER environment to build this model in

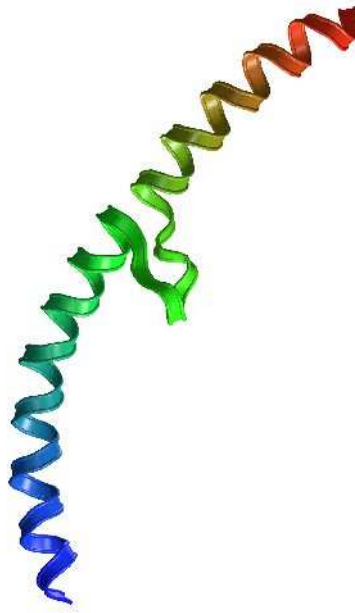
# directories for input RM6-tests files
env.io.RM6_test_directory = ['.', 'RM6_test/']

a = automodel(env,
              alnfile = 'alignment_aQL.ali', # alignment filename
              knowns = '1QX8a', # codes of the templates
              sequence = '2aaQLa') # code of the target

a.starting_model= 1 # index of the first model
a.ending_model = 1 # index of the last model
# (determines how many models to calculate)
a.make() # do the actual homology modeling
```

**Εικόνα 2.10** Το script βάση του οποίου γίνεται η μοντελοποίηση της αλληλουχίας 2aaQLa με βάση τη γνωστή δομή 1QX8a γραμμένο σε γλώσσα προγραμματισμού python. (script A)

Έτσι, το πρόγραμμα μας δίνει τη δομή της εικόνας 2.11 η οποία δείχνει ότι πρόσθεσε τα επτά κατάλοιπα στη θέση που του ορίσαμε. Όμως, επειδή δεν είχε δεδομένα από την πρότυπη δομή για τα συγκεκριμένα κατάλοιπα, τα πρόσθεσε ως βρόχο. Η συγκεκριμένη δομή εμφανίζει μια κάμψη διότι προσπαθεί να τηρεί τους χωρικούς περιορισμούς της. Παρατηρούμε και σε αυτήν την περίπτωση αυτήν την έντονη κάμψη της αλυσίδας.



**Εικόνα 2.11** Το μονομερές της QL με τα 7 επιπλέον κατάλοιπα τα οποία προστίθενται ως βρόχος.

Το επόμενο, λοιπόν, βήμα μας είναι να προσθέσουμε επιπλέον δεδομένα, ώστε να κάνουμε α-έλικα τα κατάλοιπα εισαγωγής. Αυτό θα το πετυχαίναμε θέτοντας συγκεκριμένους περιορισμούς για δευτεροταγή δομή. Για το λόγο αυτό προσθέσαμε στο αρχικό μας script τις εντολές μορφοποίησης δευτεροταγούς δομής, ώστε να προσδιορίσουμε τα κατάλοιπα της εισαγωγής 28-34 να είναι α-έλικα και ταυτόχρονα τα υπόλοιπα κατάλοιπα πριν και μετά την προσθήκη να διατηρούνται σταθερά (εικόνα 2.12), όπως ήταν στην πρότυπη δομή. Τα κατάλοιπα καθίστανται σταθερά με τη χρήση της εντολής στερεών σωμάτων (rigid bodies).

```
# Homology modeling by the automodel class
#
# Demonstrates how to refine only a part of the model.
#
# You may want to use the more exhaustive "loop" modeling routines instead.

from modeller import *          # Load standard Modeller classes
from modeller.automodel import * # Load the automodel class

log.verbose()

env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']
# selected atoms do not feel the neighborhood
env.edat.nonbonded_sel_atoms = 2

class MyModel(automodel):
    def special_restraints(self, aln):
        rsr = self.restraints
        at = self.atoms
#
# Add some restraints from a file:
#
        rsr.append(file='my_rsrsl.rsr')
#
# Residues 28 through 34 should be an alpha helix:
```

```

rsr.add(secondary_structure.alpha(self.residue_range('28:', '34:')))

# Restrain the specified CA-CA distance to 10 angstroms (st. dev.=0.1)
# Use a harmonic potential and X-Y distance group.
rsr.add(forms.gaussian(group=physical.xy_distance,
                       feature=features.distance(at['CA:35'],
                                                  at['CA:40']),
                       mean=10.0, stdev=0.1))

mdl = model(env, file='1QX8a')

# Keep residues 5-27 in chain A rigid:
r = rigid_body(mdl.residue_range('5:A', '27:A'))
mdl.restraints.rigid_bodies.append(r)

# Keep residues 35-54 in chain A rigid:
r = rigid_body(mdl.residue_range('35:A', '51:A'))
mdl.restraints.rigid_bodies.append(r)

a = MyModel(env,
            alnfile = 'alignment_aQL.ali',      # alignment filename
            knowns   = '1QX8a',                # codes of the templates
            sequence = '2aaQLa')                # code of the target
a.starting_model= 3                            # index of the first model
a.ending_model   = 3                            # index of the last model
# (determines how many models to calculate)
a.make()                                         # do homology modeling

```

**Εικόνα 2.12** Το script το οποίο δηλώνει ποια κατάλοιπα θα γίνουν α-έλικα και ποια θα κρατηθούν σταθερά. (script B)

Παρατηρώντας, όμως, τα αποτελέσματα της μοντελοποίησης στην εικόνα 2.13, το πρόγραμμα έκανε α-έλικα τα κατάλοιπα που του ορίσαμε, όμως για να το πετύχει αυτό δεν κράτησε σταθερά στις θέσεις τους τα κατάλοιπα της υπόλοιπης αλυσίδας αλλά χρησιμοποίησε τα γειτονικά κατάλοιπα της εισαγωγής μετατρέποντας τα σε βρόχο. Αυτό δείχνει ότι προσπαθεί να κρατήσει τους χωρικούς του περιορισμούς σταθερούς.



**Εικόνα 2.13** Το μονομερές με τα κατάλοιπα εισαγωγής σε μορφή α-έλικας. Για να επιτευχθεί αυτό χρησιμοποιήθηκαν τα γειτονικά κατάλοιπα της εισαγωγής, τα οποία μετατράπηκαν σε βρόχους.

Συνεχίζοντας την προσπάθεια για τη δημιουργία μιας ενιαίας α-έλικας, δώσαμε ως εντολή να γίνουν α-έλικα όλα τα κατάλοιπα της αλυσίδας, δηλαδή 1-54 και όχι μόνο τα κατάλοιπα της εισαγωγής (εικόνα 2.14). Ως αποτέλεσμα το πρόγραμμα δημιούργησε ολόκληρη την αλυσίδα ως α-έλικα όμως δεν είναι ευθεία, αλλά εμφανίζει μια γωνία, όπως φαίνεται στην εικόνα 2.15, λόγω των χωρικών περιορισμών που τηρεί μεταξύ των ατόμων κατά μήκος της αλυσίδας.

```
# Residues 1 through 54 should be an alpha helix:  
rsr.add(secondary_structure.alpha(self.residue_range('1:', 54:')))
```

**Εικόνα 2.14** Η τροποποιημένη εντολή για το script B που θα καθορίζει όλα τα κατάλοιπα της αλυσίδας να φέρουν μορφή α-έλικας.



**Εικόνα 2.15** Η μονομερής μορφή ως ενιαία α-έλικα με κάμψη.

Αφού καταφέραμε να δώσουμε δομή α-έλικας στο μονομερές θα περάσουμε στην μοντελοποίηση της κανονικής δομής της πρωτεΐνης μας με την τετραμερή της μορφή. Όσο αφορά την κάμψη του μονομερούς η οποία οφείλεται στους περιορισμούς αποστάσεων κατά μήκος της αλυσίδας ίσως να διορθωθεί κατά τον σχηματισμό του τετραμερούς όπου δημιουργούνται κατάλληλες αλληλεπιδράσεις μεταξύ των μονομερών.

Στη συνέχεια προσπαθήσαμε να δώσουμε στη δομή τετραμερή μορφή, γι' αυτό χρησιμοποιήσαμε το script: model-multichain.py. Όμως είναι απαραίτητο να προσέξουμε τη συμμετρία του μοντέλου γι' αυτό ανατρέξαμε στη σχετική βιβλιογραφία για την

πρότυπη δομή της RM6, για να πληροφορηθούμε τη συμμετρία της. Μελετώντας τη δομή στο πρόγραμμα μοριακής απεικόνισης Rasmol (Bernstein, 1999) και ερευνώντας τη βιβλιογραφία συμπεράναμε ότι οι αλυσίδες που εμφανίζουν συμμετρία είναι οι αλυσίδες A-C και οι B-D (Glykos et al., 2006). Για να πραγματοποιήσουμε, λοιπόν, τη μοντελοποίηση χρησιμοποιούμε ως πρότυπη δομή το τετραμερές της RM6, 1QX8. Η στοίχιση του με την αλληλουχία στόχο καθώς και το ανάλογο script με το οποίο θα γίνει η μοντελοποίηση παρουσιάζονται στην εικόνα 2.16 και 2.17 αντίστοιχα.

```
>P1;1QX81
structureX:1QX81:5      ::220  :::      :
EKTALNMARFIRSQTLTLLEKLNELAD-----ICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELAD-----ICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELAD-----ICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELAD-----ICESLHDHADELYRSCLARF*
```

```
>P1;2aaQL
sequence:2aaQL:5      : :216  : :::      :
EKTALNMARFIRSQTLTLLEKLNELADAADELADICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELADAADELADICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELADAADELADICESLHDHADELYRSCLARF/
EKTALNMARFIRSQTLTLLEKLNELADAADELADICESLHDHADELYRSCLARF*
END
```

**Εικόνα 2.16** Η στοίχιση των τετραμερών, πάνω η αλληλουχία της RM6 και κάτω η αλληλουχία στόχος Rop2aL (2aaQL).

```
# Homology modeling by the automodel class
#
# Demonstrates how to build multi-chain models, and symmetry restraints
#
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()

# Override the 'special_restraints' and 'user_after_single_model' methods:
class MyModel(automodel):
    def special_restraints(self, aln):
        # Constrain the A and B chains to be identical (but only restrain
        # the C-alpha atoms, to reduce the number of interatomic distances
        # that need to be calculated):
        s1 = selection(self.chains['A']).only_atom_types('CA')
        s2 = selection(self.chains['B']).only_atom_types('CA')
        s3 = selection(self.chains['C']).only_atom_types('CA')
        s4 = selection(self.chains['D']).only_atom_types('CA')
        self.restraints.symmetry.append(symmetry(s1, s3, 1.0))
        self.restraints.symmetry.append(symmetry(s2, s4, 1.0))

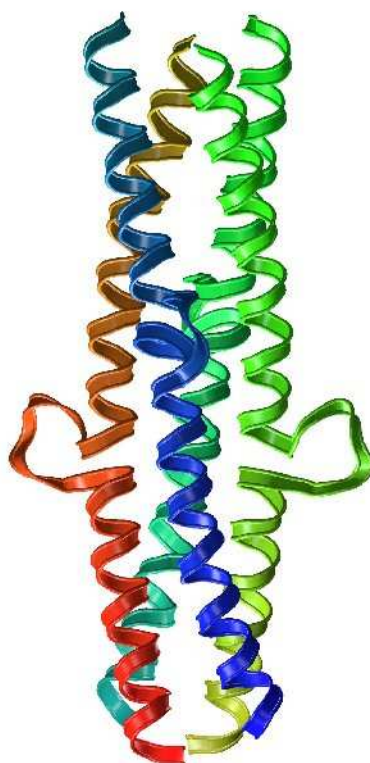
env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', './atom_files']

# Be sure to use 'MyModel' rather than 'automodel' here!
a = MyModel(env,
             alnfile = 'fourchain.ali' ,      # alignment filename
             knowns  = '1QX81',              # codes of the templates
             sequence = '2aaQL')              # code of the target

a.starting_model= 1                          # index of the first model
a.ending_model  = 1                          # index of the last model
                                                    # (determines how many models to calculate)
a.make()                                       # do homology modeling
```

**Εικόνα 2.17** Το script με το οποίο μέσω του υποσυνόλου MyModel προσδιορίζει την τετραμερή μορφή της πρωτεΐνης με συμμετρία A C και B D. (script C).

Το πρόγραμμα, αν και κατασκεύασε την τετραμερή δομή για την πρωτεΐνη, έχει τη μορφή στην οποία είχαμε φτάσει και με το αυτοματοποιημένο πρόγραμμα Swiss-Model, δηλαδή τετραμερές με τα κατάλοιπα εισαγωγής ως βρόχους (εικόνα 2.18). Λαμβάνοντας υπόψη την υπέρθεση των δυο δομών μεταξύ όλων των ατόμων του πρωτεϊνικού σκελετού (216 κατάλοιπα), η τοπική απόκλιση (Rmsd) είναι μόλις 2.2 Å το οποίο δείχνει ότι οι δομές πράγματι είναι ίδιες.

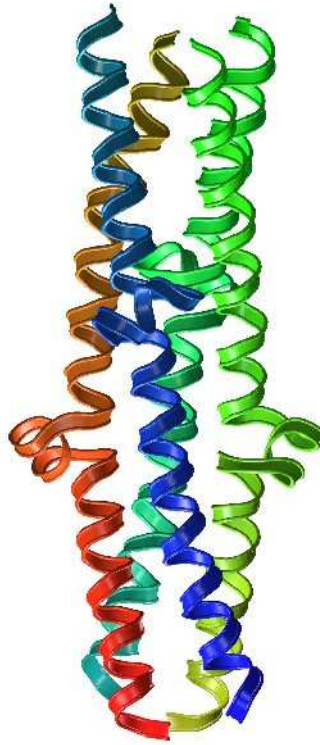


*Εικόνα 2.18 Το τετραμερές με τα κατάλοιπα εισαγωγής ως βρόχους*

Το επόμενο στάδιο είναι να μετατρέψουμε τα κατάλοιπα της εισαγωγής από βρόχο σε α-έλικα και για τις τέσσερις αλυσίδες μέσω των εντολών που προστίθενται στο script C, (εικόνα 2.19).

```
#residues should be an alpha helix:
rsr = self.restraints
at = self.atoms
rsr.add(secondary_structure.alpha(self.residue_range('28:A', '34:A')))
rsr.add(secondary_structure.alpha(self.residue_range('82:B', '88:B')))
rsr.add(secondary_structure.alpha(self.residue_range('136:C', '142:C')))
rsr.add(secondary_structure.alpha(self.residue_range('190:D', '196:D')))
def user_after_single_model(self):
    # Report on symmetry violations greater than 1Å after building
    # each model:
    self.restraints.symmetry.report(1.0)
```

*Εικόνα 2.19 Οι εντολές που προστίθενται στο script C για επιτύχουμε (τόσο τετραμερή δομή όσο και κάθε μονομερές να έχει τη δομή ενιαίας α-έλικας) ταυτόχρονα με την τετραμερή δομή, τα κατάλοιπα εισαγωγής της κάθε αλυσίδα α-έλικα.*



*Εικόνα 2.20 Η τετραμερής δομή όπου τα κατάλοιπα της εισαγωγής δεν ευθυγραμμίζονται με την υπόλοιπη έλικα.*

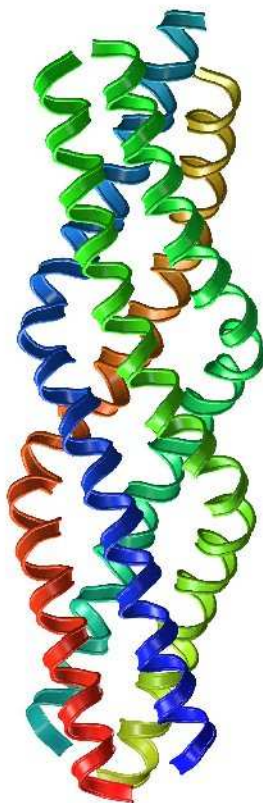
Το αποτέλεσμα, όπως βλέπουμε στην εικόνα 2.20, είναι ένα τετραμερές, όπου τα κατάλοιπα της εισαγωγής είναι α-έλικα, αλλά τα τοποθέτησε εκτός από την ευθεία της αλυσίδας και χρησιμοποίησε τα γειτονικά ως βρόχους, για να το πετύχει. Γι' αυτό τροποποιήσαμε κατάλληλα το αρχείο script C δίνοντας εντολές, όπως φαίνεται στην εικόνα 2.21, ώστε όλο το μήκος της κάθε αλυσίδας να γίνει α-έλικα.

```
#residues should be an alpha helix:  
rsr = self.restraints  
at = self.atoms  
rsr.add(secondary_structure.alpha(self.residue_range('1:A', '54:A')))  
rsr.add(secondary_structure.alpha(self.residue_range('55:B', '108:B')))  
rsr.add(secondary_structure.alpha(self.residue_range('109:C', '162:C')))  
rsr.add(secondary_structure.alpha(self.residue_range('163:D', '216:D')))
```

*Εικόνα 2.21 Οι εντολές με τις οποίες καθορίζονται τα κατάλοιπα κατά μήκος των αλυσίδων να γίνουν α-έλικα (script D)*



Αποτέλεσμα αυτής της διεργασίας ήταν ένα άτοπο τετραμερές δεμάτιο (εικόνα 2.22), με όλες τις αλυσίδες του να εμφανίζουν δευτεροταγή δομή  $\alpha$ -έλικας. Όμως η δομή αυτή δε μοιάζει καθόλου με τα συνηθισμένα 4- $\alpha$ -ελικοειδή δεμάτια, γιατί δεν εμφανίζει τα βασικά χαρακτηριστικά τους (παράγραφος 1.4) συνεπώς, η πιθανότητα να ομοιάζει με την πραγματική δομή της QL είναι μάλλον πολύ μικρή.



*Εικόνα 2.22 Η τελική δομή για την QL, όπως προέκυψε από τις πρώτες προσπάθειες του Modeller.*

Παρότι η δομή αυτή, την οποία θα ονομάσουμε προσωρινά «πλεξούδα», φαίνεται να είναι μάλλον απίθανη, επιδέχεται βελτιστοποίηση μέσω των προσομοιώσεων μοριακής δυναμικής δίνοντας έτσι και την δυνατότητα να ελέγξουμε την ευαισθησία των προσομοιώσεων μοριακής δυναμικής σε ό, τι αφορά την ακρίβεια της αρχικής δομής. Με αυτόν τον τρόπο έχουμε τη δυνατότητα να μελετήσουμε τη σταθερότητα της δομής, τις πιθανές μεταβολές που θα προκύψουν και το γεγονός αν οι προσομοιώσεις καταφέρουν να «διορθώσουν» τη δομή του μοντέλου με βάση τα εμπειρικά μοντέλα που χρησιμοποιούν.

## 2.9 Βελτιστοποίηση πρώτου μοντέλου και παραγωγή της τελικής δομής της QL

Για τη βελτιστοποίηση της δομής της πλεξούδας χρησιμοποιήσαμε το βασικό τρόπο λειτουργίας του προγράμματος, δηλαδή τη μοντελοποίηση μέσω ικανοποίησης συγκεκριμένων χωρικών περιορισμών, ώστε να βελτιώσουμε το τελικό μας μοντέλο.

Όπως αναφέραμε στην παράγραφο 2.6 του κεφαλαίου αυτού, το Modeller για την λειτουργία του χρησιμοποιεί περιορισμούς οι οποίοι είναι πολλοί και παράγονται από διαφορετικές πηγές. Δεν τροποποιήσαμε, όμως, αυθαίρετα περιορισμούς αποστάσεων και γωνιών μεταξύ ατόμων αλλά, συγκεκριμένα μετά από έρευνα. Το modeler, όταν κατασκευάζει μια δομή, παράγει και αντίστοιχα αρχεία που αναφέρουν στοιχεία της διαδικασίας, διάφορες μετρήσεις και χαρακτηριστικά της δομής. Το αρχείο με κατάληξη .rsr είναι αυτό που εμπεριέχει τους περιορισμούς, τους οποίους πληρεί το πρόγραμμα κατά τη διάρκεια μοντελοποίησης της δομής. Κατά συνέπεια, χρειάστηκε τροποποίηση κατάλληλη ώστε να επιλέξουμε και συνακόλουθα να αλλάξουμε αυτούς τους περιορισμούς, ώστε να βελτιστοποιήσουμε το μοντέλο.

Αρχική μας σκέψη ήταν, αφού η δομή μας αναδιπλώνεται με τη μορφή σπειροειδούς σπειράματος, είναι αναγκαία η αλλαγή των τιμών των διεδρων γωνιών της πρωτεΐνης μας, ώστε οι τιμές να είναι σύμφωνες με τις τιμές των αντίστοιχων γωνιών σε σπειροειδή σπειράματα. Όμως, μετά από έρευνα του αρχείου rsr, διαπιστώσαμε ότι οι τιμές των διεδρων γωνιών ήταν μέσα στα επιτρεπόμενα όρια, σύμφωνα με το Ramachandran plot.

Επόμενο βήμα μας ήταν να μελετήσουμε τους περιορισμούς αποστάσεων. Παρατηρώντας το rsr αρχείο του μοντέλου πλεξούδας, διαπιστώσαμε ότι το Modeller θέτει περιορισμούς αποστάσεων μεταξύ ατόμων κατά μήκος όλης της αλυσίδας. Στόχος μας ήταν να αφαιρέσουμε τους περιορισμούς αυτούς που ενώνουν άτομα που βρίσκονται πριν και μετά την εισαγωγή, ώστε να δημιουργηθεί κενό και να μπουν στη θέση αυτή τα καινούργια κατάλοιπα.

Αυτό θα το πετυχαίναμε, εάν τροποποιούσαμε συγκεκριμένα σημεία του αρχείου rsr. Από το αρχείο αυτό μας ενδιαφέρει η γραμμή που προσδιορίζει τους περιορισμούς (δηλώνεται με το R στην πρώτη στήλη). Στην τέταρτη στήλη της γραμμής αυτής παρουσιάζονται οι φυσικοί περιορισμοί των αποστάσεων κωδικοποιημένα με αριθμούς από τους οποίους οι οκτώ πρώτοι αφορούν τους στερεοχημικούς περιορισμούς, οι αριθμοί από 9-27 είναι οι γεωμετρικοί και οι υπόλοιποι από 28 έως 39 προέρχονται από

πειραματικά στοιχεία. Από αυτούς τους περιορισμούς μας ενδιαφέρουν στην προκειμένη περίπτωση α) οι αποστάσεις των Ca ατόμων μεταξύ των καταλοίπων (δηλώνεται με τον αριθμό 9), β) οι αποστάσεις των ατόμων N (άζωτο) ενός καταλοίπου με το O (οξυγόνο) άλλου καταλοίπου (δηλώνεται με τον αριθμό 10), γ) αποστάσεις μεταξύ πλευρικών αλυσίδων του σκελετού (δηλώνεται με τον αριθμό 26 ) (Στο παράρτημα φαίνεται η μορφή του αρχείου rsr και η πίνακας με τους τύπους των περιορισμών).

Η πρωτεΐνη μας μπορεί να διακριθεί σε τρία τμήματα ανάλογα με την αλληλουχία της. Έχουμε δυο τμήματα πριν και μετά την εισαγωγή των επτά καταλοίπων και το τμήμα της εισαγωγής. Στο τμήμα αυτό συμπεριλαμβάνονται εκτός από τα επτά κατάλοιπα της εισαγωγής, τα κατάλοιπα που αντιστοιχούν στα κατάλοιπα της εισαγωγής της γειτονικής πρωτεΐνης (βλέπε πίνακα 2.1) λόγω του ότι οι έλικες είναι αντιπαράλληλες. Επιλέχθηκαν, λοιπόν, στο από το αρχείο rsr, οι αποστάσεις που ένωναν άτομα των δύο τμημάτων πριν και μετά την εισαγωγή και αφαιρέθηκαν.

**Πίνακας 2.1** Καθορισμός ατόμων του μορίου που ανήκουν στα κατάλοιπα εισαγωγής. Με κόκκινο χρώμα σημειώνονται τα κατάλοιπα και ο αντίστοιχος αριθμός ατόμων τους, τα οποία δεν ανήκουν στην εισαγωγή. Τα σημειωμένα με γαλάζιο κατάλοιπα είναι τα κατάλοιπα που εισήχθησαν κατά την μοντελοποίηση της δομής. Όμως, ως κατάλοιπα εισαγωγής θεωρούμε και τα επτά κατάλοιπα πριν την εισαγωγή που αντιστοιχούν στην εισαγωγή των γειτονικών καταλοίπων λόγω του ότι οι έλικες είναι αντιπαράλληλες. Η φορά των ελίκων φαίνεται από την αρίθμηση των καταλοίπων.

Rop2aL(QL)				
Μονομερή	A	D	B	C
Άτομα καταλοίπων πριν την εισαγωγή.	Glu 1 (1 atom)	Phe 216 (1724)	Glu 55 (432)	Phe 162 (1293)
	Leu 19 (153 atom)	197 (1559)	Leu 73 (584)	Ile 143 (1128)
Εισαγωγή	20 (154)	Asp 196 (1558)	74 (585)	Asp 142 (1127)
	27 (217)	Ala 190 (1511)	81 (648)	Ala 136 (1087)
	Ala 28 (218)	189 (1510)	Ala 82 (649)	135 (1086)
	Asp 34 (265)	182 (1447)	Asp 88 (696)	128 (1016)
Άτομα καταλοίπων μετά την εισαγωγή	Ile 35 (266)	Leu 181 (1446)	Ile 89 (697)	Leu 127 (1015)
	Phe 54 (431)	Glu 163 (1294)	Phe 108 (862)	Glu 109 (863)

Αφού έχουμε τροποποιήσει το αρχείο rsr, σύμφωνα με τους περιορισμούς μας, χρησιμοποιήσαμε για τη μοντελοποίηση το script model-mysr.py (εικόνα 2.23) σύμφωνα με το οποίο το πρόγραμμα μας θα βασιστεί μονάχα στο rsr αρχείο που θα του υποβάλουμε, για να κατασκευάσει τη δομή χωρίς καμιά άλλη εντολή.

```

# Modeling using a provided restraints file (csrfile)
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()
env = environ()

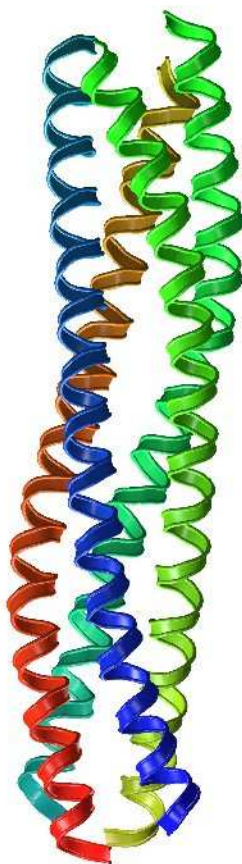
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

a = automodel(env,
             alnfile = 'fourchain.ali',      # alignment filename
              knowns = '1QX81',            # codes of the templates
              sequence = '2aaQL',          # code of the target
              csrfile = 'my.pleksouda.rsr')  # use 'my' restraints file
a.starting_model= 1                       # index of the first model
a.ending_model = 1                         # index of the last model
# (determines how many models to calculate)
a.make()                                   # do homology modeling

```

**Εικόνα 2.23** Μοντελοποίηση της δομής βάσης του αρχείου χωρικών περιορισμών (rsr) το οποίο θέσαμε εμείς. (Script E)

Αποτέλεσμα των παραπάνω διεργασιών είναι μια δομή η οποία μοιάζει πολύ με την επιθυμητή, διότι διατηρεί τα κριτήρια που τέθηκαν αρχικά και ταυτόχρονα εμφανίζει τα βασικά χαρακτηριστικά ενός α-ελικοειδούς δεματίου. Η δομή αυτή φαίνεται στην εικόνα 2.24 και θα αναφέρεται ως «sigmoidal».



**Εικόνα 2.24** Η τελική μορφή που προτείνουμε για την QL κατασκευασμένη με τη βοήθεια του Modeller

## 2.10 Έλεγχος δομών

Το τελευταίο στάδιο της ερευνάς μας ήταν να ελέγξουμε τη δομή που κατασκευάσαμε με τη διαδικασία της ομόλογης μοντελοποίησης. Απαιτήθηκε να μελετήσουμε, εάν η δομή μοιάζει με την πρότυπη δομή της RM6, αν είναι συμβατή με τους κανόνες αναδίπλωσης δεματιών και να εξασφαλιστεί ότι δεν υπήρχαν εμφανείς στερεοχημικές παρεμποδίσεις. Συγκεκριμένα, συγκρίναμε τα υδρόφοβα επίπεδα του πυρήνα της πρωτεΐνης μας με τα αντίστοιχα της RM6 και βρήκαμε ότι έχουν διαφορά δύο επιπέδων, λόγω της εισαγωγής των επτά καταλοίπων, ενώ όλα τα υπόλοιπα παρέμειναν ίδια και καλά σχηματισμένα. Γι' αυτό συμπεράναμε ότι διατηρήθηκε ο υδρόφοβος πυρήνας της. Ελέγξαμε, επίσης, να μην υπάρχει κανένα υδρόφιλο αμινοξύ μέσα στον υδρόφοβο πυρήνα και αν υπάρχει να αλλάξουμε τις διεδρες γωνίες της πλευρικής του αλυσίδας, ώστε να είναι στραμμένη προς το εξωτερικό της δομής. Τέλος, διερευνήσαμε τα άκρα αν αυτά είναι καλά πακεταρισμένα, ώστε να μην υπάρχει μεγάλο κενό το οποίο θα επέτρεπε τη διέλευση μορίων νερού στο πρώτο στάδιο της προσομοίωσης με κίνδυνο να διαλυθεί το υδρόφοβο πακετάρισμα και κατ' επέκταση ολόκληρο το σύστημα της προσομοίωσης.

Συνοψίζοντας, καταφέραμε με την ομόλογη μοντελοποίηση να κατασκευάσουμε δύο δομές για την αλληλουχία στόχο, την πλεξούδα (εικόνα 2.24) και το sigmoidal (εικόνα 2.26), που θα χρησιμοποιήσουμε στις προσομοιώσεις μοριακής δυναμικής για να μελετήσουμε τη σταθερότητα και τη δυναμική της δομής τους. Επιπλέον, με την πλεξούδα θα ελέγξουμε την δυνατότητα των προσομοιώσεων να βελτιστοποιούν-τροποποιούν της πρωτεϊνικές δομές με βάση τα εμπειρικά μοντέλα που χρησιμοποιούν.

# Κεφάλαιο 3

## Προσομοιώσεις μοριακής δυναμικής

### 3.1 Εισαγωγή

Οι προσομοιώσεις μοριακής δυναμικής αποτελούν τη ρεαλιστική αναπαράσταση του περιβάλλοντος των βιολογικών μακρομορίων σε ατομικό επίπεδο. Αναφέρονται στην έρευνα της δομής και δυναμικής των πρωτεϊνών, νουκλεϊκών οξέων και λιπιδίων ή άλλων βιολογικών μορίων χρησιμοποιώντας θεωρητικές τεχνικές (Cheatham & Brooks, 1998). Λειτουργούν σαν μια γέφυρα μεταξύ του μικροσκοπικού και μακροσκοπικού κόσμου του εργαστηρίου, δηλαδή ξεκινάμε με μια πρόβλεψη για τις αλληλεπιδράσεις μεταξύ των μορίων και καταλήγουμε σε “ακριβείς” προβλέψεις για την πρωτεΐνη (Allen, 2004).

Δύο είναι οι κύριες, μεταξύ άλλων υβριδικών, τεχνικές προσομοιώσεων: οι προσομοιώσεις μοριακής δυναμικής (MD) και οι προσομοιώσεις Monte Carlo (MC) (Allen, 2004). Υπάρχουν τρεις κύριες κατηγορίες εφαρμογών των μεθόδων προσομοίωσης. Η πρώτη χρησιμοποιείται για τον καθορισμό και τη βελτιστοποίηση των πρωτεϊνικών δομών καθώς και για τη μελέτη και παρατήρηση διαμορφώσεων στο χώρο (sampling configuration space). Η δεύτερη, προσφέρεται για περιγραφή του συστήματος που μελετάμε σε κατάσταση ισορροπίας δίνοντας, τη δυνατότητα μελέτης δομικών, κινητικών ιδιοτήτων, σταθερότητας του μοντέλου της πρωτεΐνης και θερμοδυναμικών παραμέτρων. Στην περίπτωση αυτή είναι απαραίτητο να υπολογιστεί ο χώρος που διαμορφώνεται με ακρίβεια και κάθε σημείο να σταθμίζεται με τον κατάλληλο παράγοντα Boltzmann. Στην τρίτη, εξετάζεται η πραγματική δυναμική του συστήματος υπολογίζοντας το χώρο και σταθμίζοντας με τον κατάλληλο παράγοντα Boltzmann, αλλά σε συνάρτηση με το χρόνο για την σωστή αναπαράσταση της εξέλιξης του συστήματος. Για τις δυο πρώτες κατηγορίες

μπορούν να χρησιμοποιηθούν τόσο οι προσομοιώσεις μοριακής δυναμικής (MD) όσο και οι προσομοιώσεις Monte Carlo (MC). Στην τρίτη όμως, όπου το πρωταρχικό μέλημα είναι οι κινήσεις και η εξέλιξη τους στο χρόνο, μόνο οι προσομοιώσεις μοριακής δυναμικής παρέχουν τις απαραίτητες πληροφορίες (Karplus, 2003).

Θα επικεντρωθούμε στην τεχνική της μοριακής δυναμικής, μια ντετερμινιστική μέθοδο η οποία είναι ένα εργαλείο για θεωρητική μελέτη των βιολογικών μορίων. Αυτή η υπολογιστική μέθοδος μελετά την συμπεριφορά του μοριακού συστήματος σε συνάρτηση με το χρόνο. Παρέχει λεπτομερείς πληροφορίες όσον αφορά τις διακυμάνσεις και τις διαμορφώσεις των πρωτεϊνικών και νουκλεϊκών οξέων<sup>1</sup> και γενικότερα της δομής, της λειτουργίας, της δυναμικής και θερμοδυναμικής των βιολογικών μακρομορίων σε ατομικό επίπεδο (Karplus & McCammon, 2002). Η μέθοδος επιτρέπει την πρόβλεψη των στατικών και δυναμικών ιδιοτήτων των μορίων χρησιμοποιώντας υπάρχουσες αλληλεπιδράσεις μεταξύ των μορίων.

Οι προσομοιώσεις MD για την μελέτη των βιολογικών συστημάτων, βασίζονται στους θεμελιώδους σημασίας νόμους της στατιστικής και της κλασικής μηχανικής. Με τις εξισώσεις κίνησης (2<sup>ος</sup> νόμος του Νεύτωνα) που εφαρμόζονται στην κλασική μηχανική, επιτυγχάνεται η παραγωγή ενός τροχιακού (trajectory) το οποίο περιγράφει τις θέσεις, τις ταχύτητες και τις επιταχύνσεις τμημάτων του βιολογικού συστήματος σε μικροσκοπικό επίπεδο. Από την άλλη, με την στατιστική μηχανική μέσω μαθηματικών εκφράσεων, μελετώνται οι μακροσκοπικές ιδιότητες του συστήματος, όπως η πίεση, η ενέργεια, η θερμοκρασία κι άλλες. (<sup>1</sup> ; Ercolessi, 1997)

Στην ολική ενέργεια του συστήματος, εκτός από την κινητική ενέργεια που αναφέραμε, συμπεριλαμβάνεται και η δυναμική. Η δυναμική ενέργεια υπολογίζεται με βάση τα δυναμικά πεδία (force fields) -ή αλλιώς εμπειρικές συναρτήσεις δυναμικής ενέργειας-, τα οποία αποτελούνται από ένα σει παραμέτρων και επιχειρούν να αναπαράγουν εμπειρικά φυσικές ιδιότητες, όπως δομικά, δυναμικά και θερμοδυναμικά δεδομένα. Υπάρχουν διάφορα force fields για πρωτεΐνες όπως AMBER, CHARMM, OPLSAA. Η δυναμική ενέργεια υπολογίζεται από το άθροισμα της ενέργειας των εσωτερικών αλληλεπιδράσεων δεσμικών ατόμων (γωνίες, μήκη και περιστροφές δεσμών) με την ενέργεια εξωτερικών αλληλεπιδράσεων (van der Waals και ηλεκτροστατικές) των μη δεσμικών ατόμων (<sup>1</sup>)(εικόνα 3.1)..

<sup>1</sup> [http://www.ch.embnet.org/MD\\_tutorial/](http://www.ch.embnet.org/MD_tutorial/)

$$V(R) = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{rotate-along-bond}$$

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

*Εικόνα 3.1 Η συνάρτηση της δυναμικής ενέργειας του συστήματος που υπολογίζεται από τα εμπειρικά δυναμικά πεδία.*

Κατά πόσο οι προσομοιώσεις μοριακής δυναμικής είναι αξιόπιστες και τα μοντέλα τους εμφανίζουν μεγάλη ακρίβεια διαπιστώνεται από τη σύγκριση μεταξύ των αποτελεσμάτων τους και των αποτελεσμάτων άλλων πειραματικών δεδομένων. Χαρακτηριστικό παράδειγμα αποτελεί η υψηλή πιστότητα των αποτελεσμάτων στο μόριο της μυσσοφαιρίνης από τη συνδυασμένη ανάλυση time-resolved x-ray δομών και προσομοιώσεων MD (Hummer et al., 2004). Αξίζει να σημειωθεί ότι ο πυρηνικός μαγνητικός συντονισμός NMR συμβάλλει στην ανάδειξη της δυναμικής και θερμοδυναμικής των μακρομορίων. Πάντως και στις δύο περιπτώσεις -NMR και προσομοιώσεις- δεν υπάρχει σημαντική απόκλιση στα συμπεράσματα· ενισχύοντας έτσι την αξιοπιστία των προσομοιώσεων (Arthanari et al., 2003).

Έχουν επιτευχθεί σημαντικές βελτιώσεις στο χώρο των προσομοιώσεων, τις βάσεις των οποίων, έθεσε ο Karplus. Ανέπτυξε ένα πρόγραμμα (Pre-Charmm), με τη συμβολή του Bruce, το οποίο υπολόγιζε χρησιμοποιώντας ως δεδομένες πληροφορίες: την αμινοξική αλληλουχία, ένα σετ συντεταγμένων, την ενέργεια του συστήματος και των παραγώγων του ως συνάρτηση των ατομικών θέσεων. Υπολόγισαν έτσι τις δυνάμεις των ατόμων μιας πρωτεΐνης, προκειμένου να επιτευχθεί η ελαχιστοποίηση της ενέργειας (οι οποίες δυνάμεις χρησιμοποιήθηκαν στη συνέχεια στις εξισώσεις του Νεύτωνα), ώστε να υπολογιστεί η δυναμική του εκάστοτε συστήματος (Karplus & McCammon, 2002). Μια από τις επόμενες βελτιώσεις ήταν ο υπολογισμός σε ουδέτερο σύστημα της ενέργειας που αναπτύσσεται λόγω των ηλεκτροστατικών αλληλεπιδράσεων, με τη μέθοδο particle mesh Ewald (PME) (Darden et al., 1993), καθώς και πολλές άλλες. Οι προσομοιώσεις μοριακής δυναμικής έχουν εμφανίσει σημαντική πρόοδο η οποία οφείλεται στη διαθεσιμότητα των αξιόπιστων δυναμικών πεδίων μοριακής μηχανικής (force field), την ανάπτυξη καλύτερων πρωτοκόλλων προσομοίωσης, την χρήση γρηγορότερων υπολογιστών και τέλος τον υπολογισμό long-range electrostatic forces (Cheatham & Brooks, 1998).



Στις προσομοιώσεις MD ακολουθούνται τα παρακάτω βήματα:

- Χρησιμοποιούνται οι αρχικές συντεταγμένες της προς μελέτη δομής η οποία προέρχεται είτε από κρυσταλλογραφία ακτινών Χ, από NMR είτε από ομόλογη μοντελοποίηση.
- Ενυδατώνεται η πρωτεΐνη σε ένα κουτί με νερά και ιόντα διαμορφώνοντας έτσι κατά προσέγγιση πραγματικές συνθήκες φυσιολογικού (in vitro) υδάτινου περιβάλλοντος. Στο σύστημα χρησιμοποιούνται συγκεκριμένα όρια που εμφανίζουν περιοδικότητα (periodic boundary conditions, PBC), έτσι ώστε η επιφανειακή αλληλεπίδραση των μορίων νερού να ελαττωθεί αποτελεσματικά (Haile 1997). Με την παρουσία των ιόντων επιτυγχάνεται η εξουδετέρωση των φορτίων και έτσι συμβάλει στην σταθερότητα του συστήματος. Στην περίπτωση αυτή (PBC) η ενέργεια λόγω ηλεκτροστατικών αλληλεπιδράσεων υπολογίζεται με την μέθοδο PME η οποία απαιτεί την ύπαρξη ουδέτερου συστήματος (Darden et al., 1993).
- Ελαχιστοποιείται η ενέργεια του συστήματος η οποία βασίζεται στην εναρμόνιση της αρχικής δομής με τα εμπειρικά πεδία που χρησιμοποιούνται. Διαφορετικά μπορεί να προκύψει μια μη σταθερή προσομοίωση, γιατί θα έχει διαστρεβλωθεί έστω και τοπικά η δομή. Η γεωμετρία και τα μήκη δεσμών πρέπει επίσης να είναι σύμφωνα με το δυναμικό πεδίο που ορίσαμε για να αποφευχθεί οποιοδήποτε ενεργειακό πρόβλημα. (minimization)
- Σταδιακή θέρμανση του συστήματος (Heating), μέχρι την επιθυμητή θερμοκρασία φροντίζοντας να μην αποδιοργανωθεί η δομή ώστε να μετατραπεί η θερμική ενέργεια σε κινητική και να παρακολουθήσουμε τις αλλαγές πιο γρήγορα.
- Εξισορροπείται (equilibration) το σύστημα και ελέγχονται ιδιότητες όπως η δομή, η πίεση η θερμοκρασία και η ενέργεια. Σκοπός της εξισορρόπησης είναι να “τρέχει” η προσομοίωση μέχρι αυτές οι ιδιότητες να γίνουν σταθερές ώστε να επιτευχθούν συνθήκες ισοβαρείς και ισόθερμες (NpT) (Jorgensen et al., 1983). Η πίεση και η θερμοκρασία διατηρούνται σταθερές χρησιμοποιώντας τα Nose-Hoover Langevin Dynamics με τη μέθοδο ελέγχου Langevin piston barostat.
- Παραγωγική φάση (τελικό βήμα). Παρατηρούνται οι διαμορφώσεις της πρωτεΐνης, οι θέσεις, οι ταχύτητες και οι επιταχύνσεις των ατόμων τα οποία περιγράφονται σε ένα trajectory, ενώ προσδιορίζονται με την ολοκλήρωση των εξισώσεων κίνησης βάση του αλγόριθμου Verlet-I (Allen, 2004).

Κατά την εφαρμογή των δυναμικών πεδίων, λαμβάνονται υπόψη πολλές προσεγγίσεις προκειμένου να μειωθεί το υπολογιστικό κόστος. Για να αντιπαρέλθουμε

όμως τα ενδεχόμενα λάθη που προκύπτουν από τις προσεγγίσεις αυτές, χρειάζεται να ορίσουμε περιορισμούς οι οποίοι στη συνέχεια μπορούν να ξεπεραστούν με τη χρήση νέων μικτών δυναμικών πεδίων κβαντομηχανικής-μοριακής μηχανικής.<sup>1</sup>

Οι προσομοιώσεις χρησιμοποιούνται για να αξιολογήσουν την ισορροπία και τις δυναμικές ιδιοότητες ενός περίπλοκου συστήματος το οποίο δεν μπορεί να υπολογιστεί αναλυτικά. Η αποδοτικότητα τους εξαρτάται από την ικανότητα των δυναμικών πεδίων να αναπαριστούν την συμπεριφορά των ατόμων στο εκάστοτε σύστημα. Παρόλο που είναι μια πολύ ισχυρή τεχνική υπάρχουν και κάποιοι άλλοι περιορισμοί εκτός από αυτόν που προαναφέραμε, όπως ο χρόνος. Οι προσομοιώσεις πραγματοποιούνται σε χρονικό διάστημα nanosecond ενώ οι πρωτεΐνες αναδιπλώνονται σε χιλιοστά του δευτερολέπτου ή σε δευτερόλεπτα. Ταυτόχρονα απαιτείται πολύ μεγάλο χρονικό διάστημα υπολογισμών για την αναπαράσταση της εξέλιξης του συστήματος ακόμα και στους πιο γρήγορους υπολογιστές λόγω του μεγάλου αριθμού ατόμων του συστήματος. Αυτό μπορεί να αντιμετωπιστεί με την μέθοδο του κατανεμημένου υπολογισμού (Larson et al., 2003). Ωστόσο προκύπτει νέο πρόβλημα, το πως πρακτικά θα γίνει αυτός ο καταμερισμός. Ακόμα και αν ξεπεραστεί αυτό, δημιουργείται η πρόκληση της ανάλυσης του τεράστιου όγκου δεδομένων ώστε να βγουν συμπεράσματα (Larson et al., 2003). Μια εναλλακτική μέθοδος δόθηκε από τους Izaguire et al 1999, με τον αλγόριθμο Verlet-I/r-RESPA, ο οποίος δίνει τη δυνατότητα πραγματοποίησης προσομοιώσεων μεγαλύτερης διάρκειας, πραγματοποιώντας μεγαλύτερα βήματα (timesteps).

Οι προσομοιώσεις μοριακής δυναμικής είναι ένα από τα σημαντικά εργαλεία για την μελέτη της διαμόρφωσης των πρωτεϊνών, των νουκλεϊνικών οξέων και των συμπλοκών τους. Ο κύριος ρόλος τους είναι η μελέτη της σταθερότητας και της αναδίπλωσης των πρωτεϊνών. Επίσης, σε μεγάλες αλλαγές με την χρήση των προσομοιώσεων μελετάτε η στερεοδιαμόρφωση των μακρομορίων και η μοριακή αναγνώριση από: πρωτεΐνες, DNA, RNA, λιπίδια, ορμόνες STP κ.λ.π. Επιπλέον, χρησιμοποιούνται στη μελέτη ενζυμικών αντιδράσεων και δυναμικών διαδικασιών, όπως η ιοντική μεταφορά και η δέσμευση-αποδέσμευση προσδετών. Επιπροσθέτως, με την χρήση των προσομοιώσεων δίνεται η δυνατότητα προσδιορισμού και μελέτης τρισδιάστατων δομών (έπειτα από ομολογη μοντελοποίηση, κρυσταλλογραφία ακτινών Χ και NMR). Τέλος, εφαρμόζονται στο σχεδιασμό βιολογικά ενεργών μορίων όπως φάρμακα.<sup>1</sup>

Οι προσομοιώσεις μοριακής δυναμικής έχουν εφαρμοστεί σε ένα ευρύ φάσμα συστημάτων. Χρησιμοποιήθηκαν για να βρεθεί το μονοπάτι μεταξύ της ανοιχτής και

κλειστής διαμόρφωσης πρωτεϊνών, που ήταν αδύνατο να καθοριστεί από πειραματικές τεχνικές. Χαρακτηριστικό παράδειγμα ήταν η χρήση τους για να καθοριστεί το μεταβατικό μονοπάτι της τσαπερονίνης GroEL (Sigler et al., 2000). Πρόσφατα η κρύο-ηλεκτρονική μικροσκοπία επιβεβαίωσε την πρόβλεψη των προσομοιώσεων (Ranson, et al., 2001).

Μελετήθηκε επιπλέον, η λειτουργία της πρωτεϊνικής κινάσης τυροσίνης της οικογένειας Src, η οποία επηρεάζεται από την ευελιξία του μικρού πεπτιδικού προσδέτη μεταξύ των επικρατειών SH2 και SH3 της πρωτεΐνης (Young et al., 2001). Η σημαντικότητα αυτής της σύνδεσης επιβεβαιώθηκε και από συμπληρωματικά πειράματα.

Η χρήση των MD βοήθησε, επίσης, στην αποκάλυψη μιας γέφυρας άλατος η οποία δημιουργεί μια κινητική παγίδα (Kinetic trap) στην αποδιάταξη της S<sub>nase</sub> (truncated staphylococcal nuclease). Η γέφυρα αυτή σχηματίζεται μεταξύ της αργινίνης 105 και του γλουταμινικού 135 (Gruia et al., 2003)

Αναλύσεις προσομοιώσεων μεγάλης διάρκειας για την ακετυλοχολινεστεραση έχουν αποκαλύψει δυναμικά χαρακτηριστικά του διαύλου αυτού, απαραίτητα για τη λειτουργία του, τα οποία δεν είχαν προβλεφθεί από πειράματα (Tai et al., 2001). Ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός ότι αυτές οι διακυμάνσεις της δομής του ενζύμου οδηγούν σε ένα ασυνεχές ανοιγοκλείσιμο του διαύλου από την επιφάνεια του ενζύμου μέχρι βαθιά την ενεργή περιοχή, μετά από τη σύνδεση της ακετυλοχολίνης στο διάυλο.

Η στενή αλληλεπίδραση μεταξύ των περίπλοκων πειραμάτων και των εξειδικευμένων προσομοιώσεων έχει οδηγήσει σε μια γενική κατανόηση του μηχανισμού του πρωτεϊνικού διπλώματος. Ωστόσο, παραμένουν ακόμα πολλά που πρέπει να μάθουμε σχετικά με την αναδίπλωση μεμονωμένων πρωτεϊνών, συμπεριλαμβανομένου της πρόγνωσης της δομής από την αλληλουχία της. Τελευταία, έχει δοθεί η δυνατότητα να παρακολουθήσουμε την αναδίπλωση και αποδιάταξη πεπτιδίων (Cavalli et al., 2002)

### **3.2 Προετοιμασία συστημάτων**

Για να ελέγξουμε, να επαληθεύσουμε καθώς ακόμη και να βελτιστοποιήσουμε τις δομές του τετραμερούς της QL που κατασκευάσαμε με την ομόλογη μοντελοποίηση (κεφάλαιο 2) μελετάμε τις δομικές, κινητικές ιδιότητες, τη σταθερότητα και τη δυναμική αναδίπλωσης των μοντέλων με τις προσομοιώσεις μοριακής δυναμικής. Οι προσομοιώσεις πραγματοποιήθηκαν με το πρόγραμμα NAMD (Kale et al., 1999) χρησιμοποιώντας το

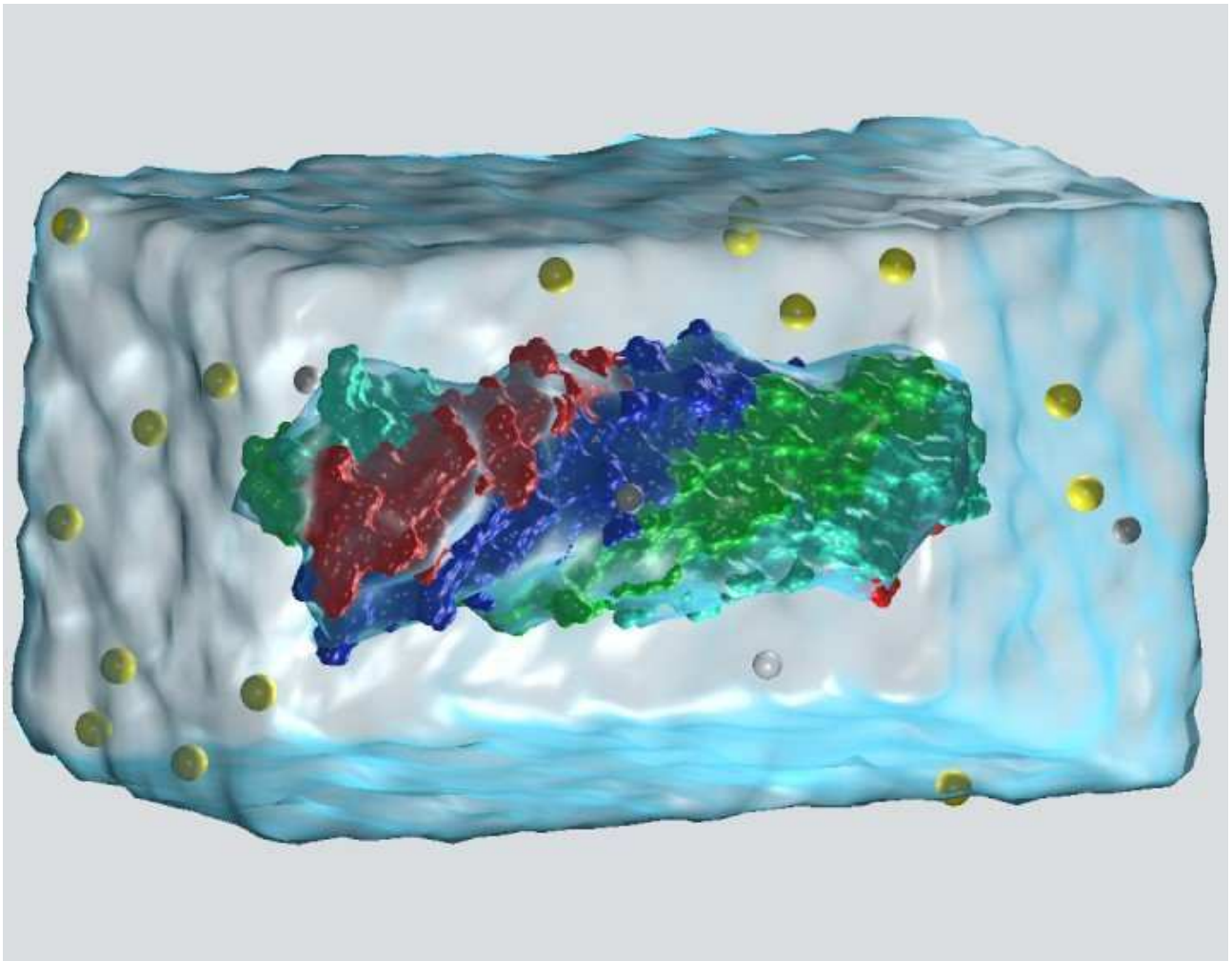
CHARMM27 δυναμικό πεδίο (MacKerell et al., 1998).

### 3.2.1. Προετοιμασία του συστήματος της πλεξούδας

Για την προετοιμασία του συστήματος της πλεξούδας (κεφάλαιο 2) ακολουθήσαμε τα παρακάτω βήματα

**A.** Στην προετοιμασία, τροφοδοτήσαμε το σύστημα με τα απαραίτητα αρχεία για την πρωτεΐνη:

- το αρχείο παραμέτρων με πληροφορίες για force field Charmm27,
- το αρχείο τοπολογίας του δυναμικού πεδίου Charmm27,
- το αρχείο pdb της πλεξούδας με τις συντεταγμένες των ατόμων που πήραμε με ομόλογη μοντελοποίηση,
- ευθυγραμμίσαμε το μόριο με το ορθοκανονικό σύστημα αξόνων και μετατοπίσαμε το κέντρο βάρους του στην αρχή των αξόνων με τη βοήθεια του προγράμματος moleman (Kleywegt & Jones, 1995). Επειδή, όμως, η προσομοίωση πραγματοποιούνταν σε όξινες και όχι ουδέτερες συνθήκες, χρησιμοποιήσαμε την πλήρως πρωτονιομένη μορφή της ιστιδίνης (δηλαδή και τα δυο άτομα αζώτου να είναι πρωτονιωμένα, αντί HIS->HSP).
- διαμέσου του προγράμματος psfgen (τμήμα του προγράμματος NAMD), το οποίο έχει την ικανότητα να προσθέτει τα άτομα της δομής που απουσιάζουν, συνήθως πρωτόνια και πλευρικές ομάδες “μαντεύοντας” τις ατομικές τους συντεταγμένες, πήραμε το psf αρχείο με τις δομικές πληροφορίες της πρωτεΐνης και το αντίστοιχο pdb με τις αρχικές συντεταγμένες στις οποίες προστέθηκαν και αυτές που έλειπαν από το αρχικό pdb αρχείο.
- ενυδατώσαμε την πρωτεΐνη με τη βοήθεια του προγράμματος VMD (Humphrey et al., 1996) τοποθετώντας την πρωτεΐνη σε ένα κουτί διαστάσεων  $109 \times 60 \times 58 \text{ \AA}^3$  με νερά (τύπου TIP3, (Jorgensen et al., 1983; Zeilkiewicz et al., 2005)) και ιόντα (Cl, Na) δημιουργώντας ένα στρώμα νερού της τάξης των  $15.0 \text{ \AA}$  γύρω από τα πρωτεϊνικά μόρια και αφαιρώντας όλα τα άτομα νερού γύρω από την πρωτεΐνη σε απόσταση μικρότερη από  $1.8 \text{ \AA}$ . Τελικά, το σύστημα περιείχε συνολικά 37.515 άτομα από τα οποία 3428 είναι άτομα πρωτεΐνης, 34066 άτομα νερού, 4 άτομα Cl και 17 άτομα Na. Τα συγκεκριμένα ιόντα προστέθηκαν για να επιτευχθεί η εξισορρόπηση του ηλεκτρικού φορτίου (παράγραφος 3.1) με τη βοήθεια του προγράμματος VMD. Τα νερά ήταν τύπου TIP3, γιατί από τους υπάρχοντες τύπους (SPC, SPCE, TIP4P, TIP5P και TIP3), είναι αυτός με τον οποίο είχε παραμετροποιηθεί το δυναμικό πεδίο CHARMM27.



Εικόνα 3.2 Η στοιχειώδης κυψελίδα (ορθογώνιο κουτί) που αναπαριστά το σύστημα προσομοίωσης της πλεξούδας. Στο κέντρο φαίνεται το μόριο της πρωτεΐνης με τη μορφή επιφανείας όπου για την κάθε αλυσίδα έχει εφαρμοστεί το ίδιο χρωματικό πρότυπο όπως και στο κεφάλαιο 2. Τα μόρια του νερού είναι σχεδιασμένα ως διαφανής επιφάνεια πυκνότητας ατόμων (*weighted atomic density grid map*) και τα ιόντα έχουν σχεδιαστεί με βάση την ακτίνα Van der Waals, με κίτρινο χρώμα τα ιόντα Na και με μπλε τα ιόντα Cl.

- καθορίσαμε τα όρια του συστήματος (κατά μήκος του ορθογώνιου πλαισίου) με το πρόγραμμα `rdbset`. Οι τιμές αυτές (μέσω του script) χρησιμοποιούνται από το NAMD για την αρχικοποίηση των διαστάσεων του κουτιού. Οι συντεταγμένες των ορίων του πλαισίου δίνονται παρακάτω:

Orthogonal Coordinate limits in output file:

	Minimum	Maximum	Centre	Range
On X :	-54.37	55.59	0.61	109.97
On Y :	-30.56	30.31	-0.13	60.86
On Z :	-30.20	28.45	-0.88	58.65

**B.** Ελαχιστοποίησαμε την ενέργεια (energy minimization, Heat.namd) των ατόμων της πρωτεΐνης, σύμφωνα με το δυναμικό πεδίου Charmm27 όταν στο σύστημα τηρούνται NvT συνθήκες δηλαδή, σταθερός όγκος και μεταβλητή πίεση,

- για 2000 βήματα ( $4 \cdot 10^{-3}$  ns) για τα άτομα των πλευρικών αλυσίδων και έπειτα
- για 2000 βήματα για όλα τα άτομα μαζί

**Γ.** Σταδιακή θέρμανση του συστήματος (heating up, Heat.namd) μέχρι την τελική θερμοκρασία των 320 K για 2000 βήματα (η αύξηση γίνεται με  $\Delta T=20$  K)

**Δ.** Εξισορρόπηση του συστήματος κατά τη διάρκεια της θέρμανσης με ισοβαρείς-ισόθερμες συνθήκες NpT (heat.namd):

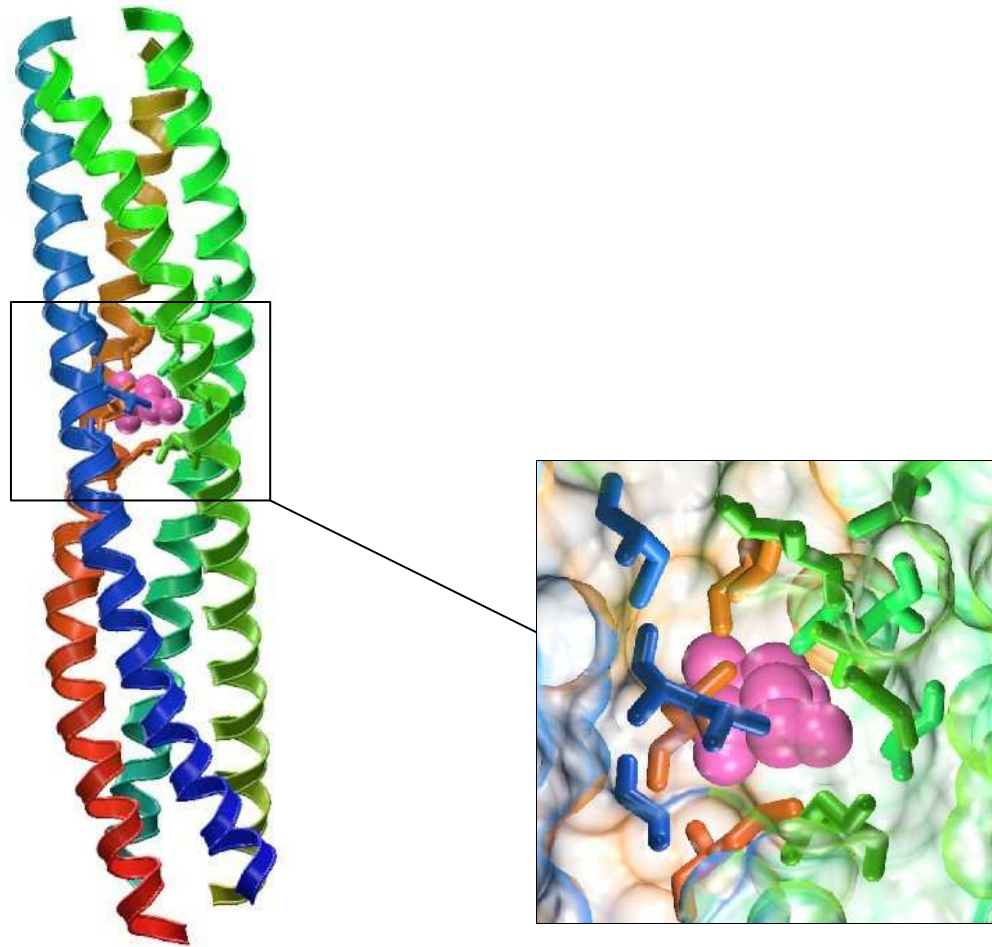
- για 2.000 βήματα διατηρώντας σταθερά τα Ca άτομα του σκελετού και,
- για 4.000 βήματα χωρίς περιορισμούς.

**Ε.** Παραγωγική φάση με τη χρήση του equilibration script (equi.namd)

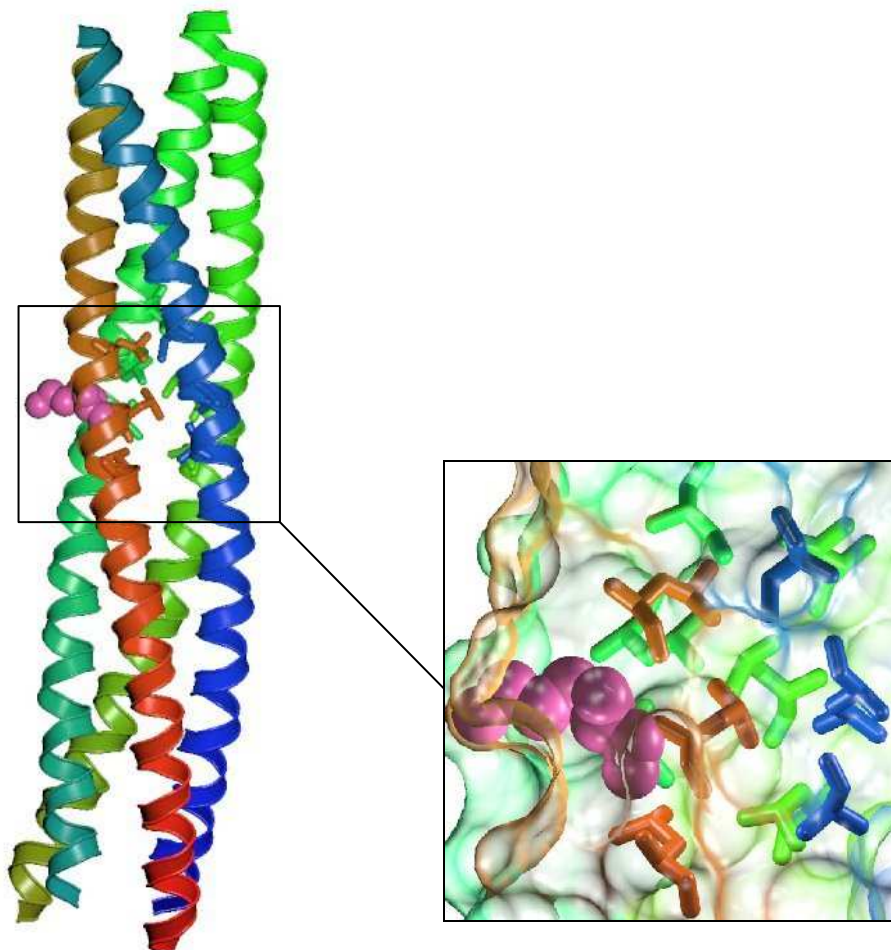
- για 50.000.000 βήματα (100ns), όπου καταγράφονται σε trajectory τα δεδομένα και οι συντεταγμένες των ατόμων κάθε 200 βήματα (0.4 ps). Ο χρόνος για κάθε βήμα είναι 2fs, οι μη-δεσμικές αλληλεπιδράσεις υπολογίζονται κάθε δύο βήματα, ενώ οι ηλεκτροστατικές αλληλεπιδράσεις κάθε τέσσερα βήματα χρησιμοποιώντας την PME μέθοδο.

### 3.2.2. Προετοιμασία του συστήματος του sigmoidal και οι απόπειρες εφαρμογής του

Ακολουθήσαμε την ίδια διαδικασία προετοιμασίας του συστήματος του μοντέλου sigmoidal, χρησιμοποιώντας παρόμοιες παραμέτρους με το σύστημα προσομοίωσης της πλεξούδας. Κατά την διάρκεια της προσομοίωσης, η φάση της σταδιακής θέρμανσης δεν ολοκληρώθηκε καθώς σταματήσαμε την διαδικασία. Αυτό συνέβη γιατί στον υδρόφοβο πυρήνα της, είχε απομείνει ένα υδρόφιλο κατάλοιπο λυσίνης (lys), με αποτέλεσμα το διαχωρισμό των αλυσίδων και την αποδιοργάνωση της δομής, εικόνα 3.3. Για το λόγο αυτό, αλλάξαμε τον προσανατολισμό των διεδρων γωνιών της πλευρικής ομάδας της λυσίνης, εικόνα 3.4. Το κενό που δημιουργήθηκε το καλύψαμε αναμορφώνοντας τις πλευρικές αλυσίδες των γειτονικών καταλοίπων κατάλληλα χρησιμοποιώντας το πρόγραμμα xfit (McRee, 1992).



*Εικόνα 3.3* Το υδρόφιλο κατάλοιπο της λυσίνης 183 που έχει εισέλθει στον υδρόφοβο πορήνα, απεικονίζεται με βάση την ακτίνα Van der Waals με μωβ χρώμα. Για λόγους κατανόησης έχουν απεικονιστεί με τη μορφή licorice τα 3 γειτονικά της επίπεδα. Δεξιά παρατηρούμε σε μεγέθυνση την παρεμβολή την λυσίνης στα υδρόφοβα επίπεδα του πορήνα, με την υπόλοιπη πρωτεΐνη να αναπαρίσταται με τη μορφή διαφανούς επιφάνειας (διατηρώντας το χρωματισμό των αλυσίδων).



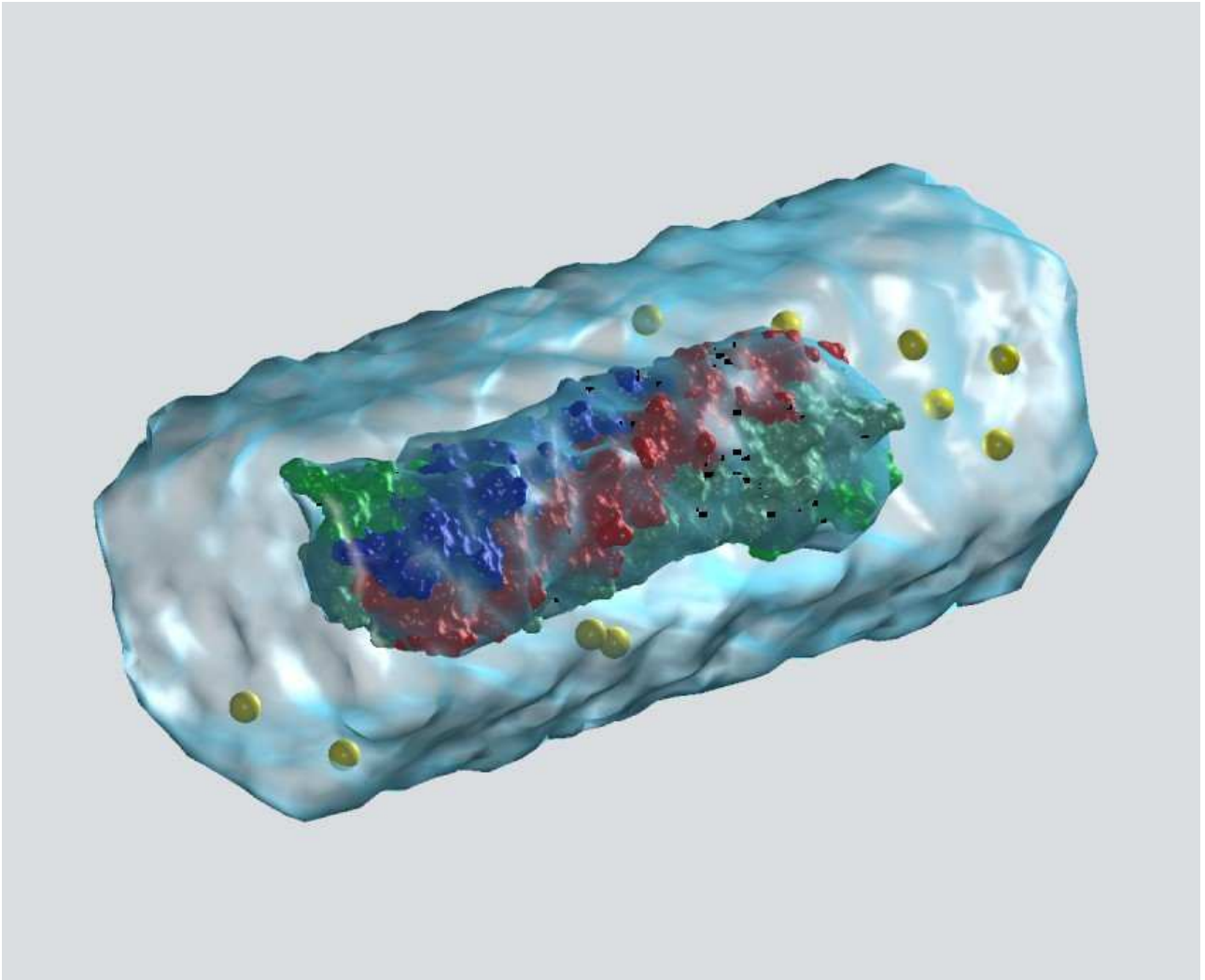
*Εικόνα 3.4* Απεικονίζεται το υδρόφιλο κατάλοιπο της λυσίνης 183 (με βάση την ακτίνα Van der Waals με μωβ χρώμα) με αλλαγή στον προσανατολισμό της πλευρικής της αλυσίδας ώστε να βρίσκεται εκτός του υδρόφοβο πυρήνα (αριστερά). Προκειμένου να φανεί αυτή η αλλαγή, ολόκληρο το μόριο της πρωτεΐνης έχει στραφεί κατά 90 μοίρες, στον άξονα y, σε σχέση με την προηγούμενη εικόνα. Δεξιά παρατηρούμε τη σχετική θέση της λυσίνης σε σχέση με τα 3 γειτονικά επίπεδα μετά την αλλαγή στις διέδρες γωνίες των πλευρικών ομάδων.

Στη συνέχεια, αφού ελέγξαμε τη δομή για πιθανή πρόσβαση νερού, επανεκκινήσαμε τις προσομοιώσεις. Αυτή τη φορά, στο στάδιο της ενυδάτωσης χρησιμοποιήσαμε εξαγωνικό κουτί, επειδή η πρωτεΐνη έχει κυλινδρικό σχήμα, και με το εξαγωνικό κουτί έχουμε λιγότερα μόρια νερού και κατά συνέπεια αυξάνεται η ταχύτητα της προσομοίωσης. Από τα 22800 συνολικά άτομα του συστήματος, τα 3428 είναι τα άτομα της πρωτεΐνης, 12 είναι Na και 19440 είναι άτομα νερού (εικόνα 3.5).

Στο εξαγωνικό κουτί υπήρχαν συνολικά 22.880 άτομα (εικόνα 3.5) και καθορίσαμε τα όρια ως εξής :

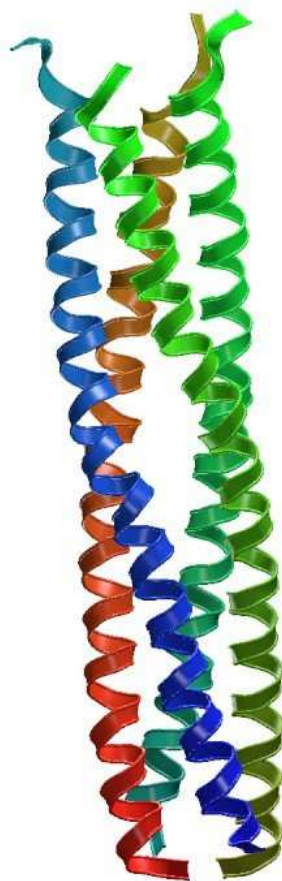
	Minimum	Maximum	Centre	Range
On X :	-57,19	59,58	1,19	116,78
On Y :	-29,02	29,42	0,20	58,44
On Z :	-25,77	25,58	-0,09	51,36





Εικόνα 3.5 Πλαινή όψη του πλαισίου της στοιχειώδους κυψελίδας όπου αναπαριστάνται το μόριο του μοντέλου sigmoidal, ενυδατωμένο με ιόντα σε εξαγωνικό κουτί. Στο κέντρο φαίνεται το μόριο της πρωτεΐνης με τη μορφή επιφανείας όπου για την κάθε αλυσίδα έχει εφαρμοστεί το ίδιο χρωματικό πρότυπο όπως και στο κεφάλαιο 2. Τα μόρια του νερού είναι σχεδιασμένα ως διαφανής επιφάνεια πυκνότητας ατόμων (*weighted atomic density grid map*) και τα ιόντα έχουν σχεδιαστεί με βάση την ακτίνα Van der Waals, με κίτρινο χρώμα

Η νέα προσομοίωση προχώρησε κανονικά μέχρι το στάδιο της παραγωγικής φάσης όπου συγκεκριμένα στο στιγμιότυπο 8900, παρατηρήσαμε πως η δομή της πρωτεΐνης στο κάτω μέρος της ήταν καλά πακεταρισμένη, όπως και στη μέση, όμως αυτό δε συνέβαινε στο πάνω μέρος. Τα δύο τελευταία επίπεδα του δεματίου δεν ήταν πλέον σχηματισμένα καθώς τα τελικά κατάλοιπα είχαν ξεδιπλώθει. (Τα δύο τελευταία επίπεδα αποτελούνται από τα εξής κατάλοιπα: Phe 54,A -Glu 109,D -Phe 184,B -Glu 163,C και Cys 50,A -Ala 112,D -Cys 104,B -Ala 166,C) (εικόνα 3.6)



*Εικόνα 3.6* Απεικονίζεται η αποδιάταξη της δομής του επάνω μέρους της πρωτεΐνης .

Ελέγχοντας τη δομή, βάση του προγράμματος VMD, ανακαλύψαμε ότι η αποδιάταξη οφειλόταν στην εισαγωγή ενός μορίου νερού στο επάνω μέρος του πυρήνα. Γι' αυτό αποφασίσαμε να αλλάξουμε την διάταξη των πλευρικών αλυσίδων στο επάνω μέρος της δομής, ώστε να είναι παρόμοια με αυτήν του καλοσχηματισμένου κάτω μέρους, ώστε αυτό να γίνει συμπαγές και σταθερό και να μην υπάρχει χώρος να διεισδύσουν νερά.

Αφήσαμε την προσομοίωση να προχωρήσει κανονικά μέχρι το στιγμιότυπο 25.760 (~10ns). Στο σημείο αυτό, και εκμεταλλευόμενοι τον χρόνο που κερδίσαμε όσο πραγματοποιούνταν οι προσομοιώσεις, δημιουργήσαμε ένα νέο μοντέλο εφαρμόζοντας και πάλι τη διαδικασία της μοντελοποίησης ελευθερώνοντας κι άλλα restraints (στο προηγούμενο αρχείο rsr αφαιρέσαμε όλους τους περιορισμούς που αφορούσαν τις αποστάσεις που έχουμε είδη αναφέρει), ώστε καταλήξαμε σε μια πιο βελτιωμένη δομή η οποία δεν έφερε σημαντικές αλλαγές από την προηγούμενη δομή. Μετά από σύγκριση των δύο δομών που πραγματοποιήθηκε, έδειξε ότι το rmsd για τα άτομα του σκελετού ήταν 0.5Å και χρησιμοποιώντας και τις πλευρικές αλυσίδες 1.3Å. Συμπεραίνουμε, λοιπόν, ότι η

δομή είναι καθόλα ίδια με την δομή της εικόνας 2.26, οπότε δεν ωφελεί να δείξουμε ξανά μια ίδια δομή.

Αυτή τη βελτιωμένη διαμόρφωση της δομής του sigmoidal που κατασκευάστηκε στην πορεία, την μελετήσαμε σε ένα τρίτο (και τελευταίο) σύστημα προσομοίωσης, χρησιμοποιώντας όμοιες παραμέτρους με την αρχική sigmoidal έχοντας συνολικά 21782 άτομα και τα όρια του κουτιού διαμορφώθηκαν ως εξής:

Orthogonal Coordinate limits in output file:				
	Minimum	Maximum	Centre	Range
On X :	-59.42	59.82	0.20	119.24
On Y :	-28.50	28.63	0.07	57.13
On Z :	-24.92	25.00	0.04	49.92

Η προσομοίωση πραγματοποίησε και τα 50.000.000 βήματα (100 ns) όπως ήταν προγραμματισμένη και θα αναλυθεί λεπτομερώς στο επόμενο κεφάλαιο.

Επαγωγικά σκεπτόμενοι και με βάση την προσωπική μας εμπειρία, θα μπορούσαμε να υποστηρίξουμε ότι οι προσομοιώσεις είναι ένα αποτελεσματικό εργαλείο για τον έλεγχο της δομής των πρωτεϊνών. Έχουν την δυνατότητα και κατά τη διάρκεια της λειτουργίας τους, να δίνουν λεπτομερείς πληροφορίες για το σύστημα, βασιζόμενοι στους θεμελιώδους νόμους φυσικής. Συνιστούν έναν ενδεδειγμένο τρόπο μελέτης, διότι συμβάλλουν στην μελλοντική διεξαγωγή πειραμάτων. Τα αποτελέσματα των προσομοιώσεων μπορούν να δώσουν σημαντικές πληροφορίες για την σταθερότητα και τη δυναμική των μορίων. Ανάλογα με αυτά μπορούμε να συνεχίσουμε σε επόμενα πειράματα ( π.χ. κρυσταλλογραφία ακτίνων Χ) ή να τροποποιήσουμε τις παραμέτρους κατάλληλα ώστε να λάβουμε καλύτερα αποτελέσματα.

# Κεφάλαιο 4

## Ανάλυση προσομοιώσεων

### 4.1 Ανάλυση προσομοίωσης του μοντέλου πλεξούδας

Στο στάδιο της ανάλυσης των προσομοιώσεων θα μελετήσουμε τη σταθερότητα της προς μελέτη πρωτεΐνης καθώς και τις κινητικές και δυναμικές ιδιότητες της. Οι αναλύσεις θα πραγματοποιηθούν με τη βοήθεια του `carna`, ενός προγράμματος ανάλυσης μοριακής δυναμικής (Glykos 2006).

#### 4.1.1 Σταθερότητα της αρχικής δομής

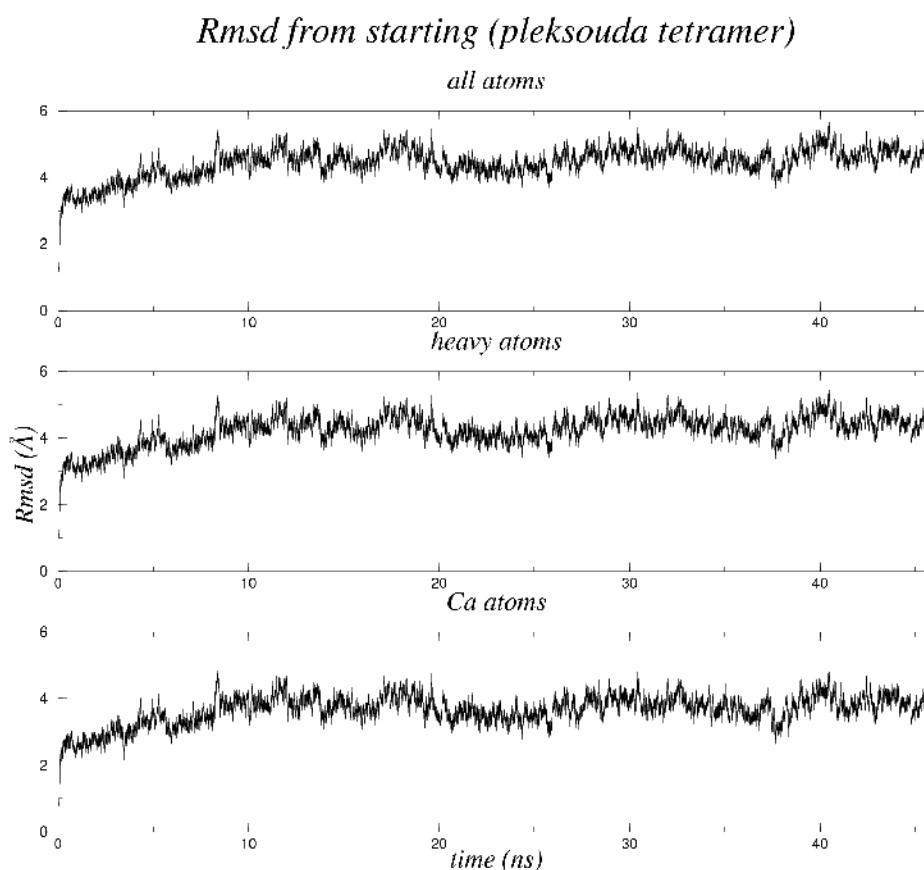
Θα ξεκινήσουμε υπολογίζοντας πρώτα την τιμή του `Rmsd` (root-mean-square-deviation) από την αρχική δομή (Garemyr & Elofsson, 1999) η οποία υπολογίζεται με βάση τον τύπο:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{i=N} \delta_i^2} \quad (1)$$

όπου  $N$  ο αριθμός των ατόμων του μορίου,  $\delta$  είναι η απόσταση μεταξύ ζευγών ισοδύναμων ατόμων, η οποία δίνεται από την διαφορά  $(x_i - x)$  όπου  $(x)$  είναι οι συντεταγμένες των ατόμων στην εκάστοτε διαμόρφωση και  $(x_i)$  οι συντεταγμένες από την αρχική του θέση. Με το `Rmsd` μελετάμε τη σταθερότητα της πρωτεΐνης καθώς επίσης

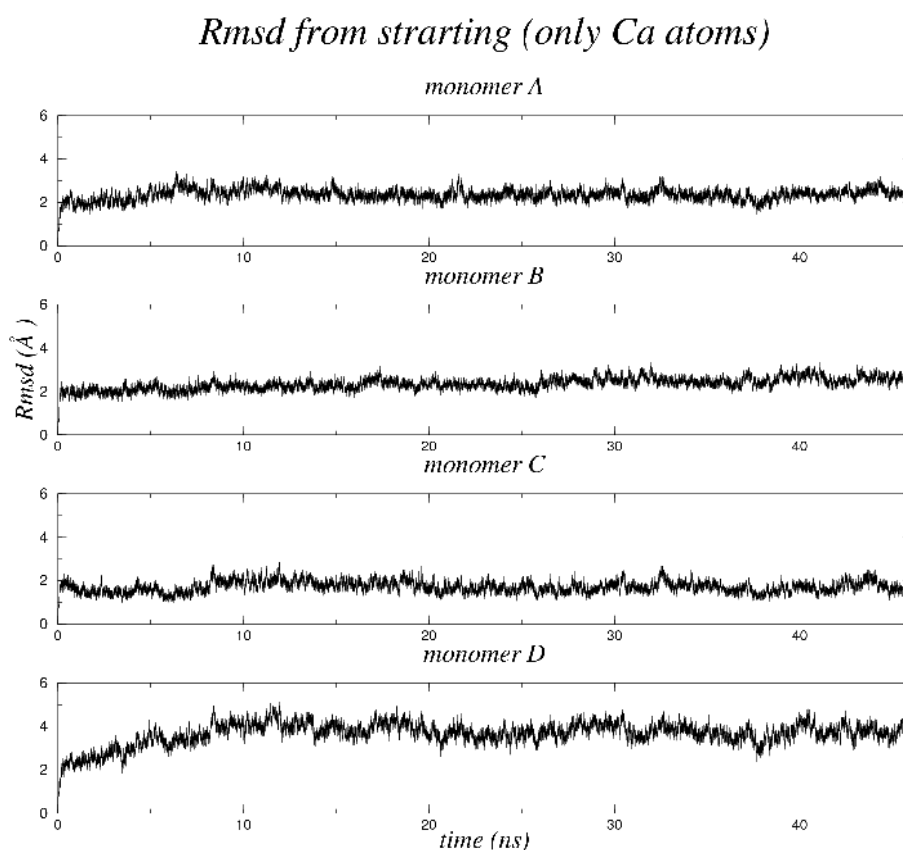
μετράμε και την απόκλιση της κάθε διαμόρφωσης του μορίου από την αρχική του δομή (την δομή που κατασκευάσαμε από το Modeller, κεφάλαιο 2) κατά τη διάρκεια της προσομοίωσης.

Στην αρχή της προσομοίωσης παρατηρούμε (εικόνα 4.1) απότομη αύξηση του rmsd για τα πρώτα 0.084ns (210frames) που αντιστοιχούν στο στάδιο της θέρμανσης (heating). Η συνολική πρωτεΐνη (all atoms) έχει μεγάλο rmsd 4.43Å κατά τη διάρκεια της προσομοίωσης και μεγάλη διακύμανση 0.44 Å. Όμως στην περίπτωση όπου υπολογίζουμε rmsd μόνο για τα άτομα του σκελετού (Ca atoms), τότε εμφανίζεται μικρότερο, 3.66 Å αλλά οι διακυμάνσεις επιμένουν να είναι έντονες 0.44Å . Με τον υπολογισμό μόνο των Ca ατόμων μειώνεται ο αριθμός ατόμων που λαμβάνεται υπόψη στους υπολογισμούς του Rmsd. Στην περίπτωση αυτή, χρησιμοποιούνται μόνο τα άτομα του σταθερού τμήματος της δομής, δηλαδή του σκελετού, ενώ αγνοούνται οι πλευρικές ομάδες των καταλοίπων. Έτσι, τυχόν ευκίνητα άτομα στα άκρα επιδρούν αρνητικά στη σταθερότητα ολόκληρου του μορίου, γι' αυτό εξακολουθούν να υπάρχουν έντονες διακυμάνσεις.



**Εικόνα 4.1** Rmsd από την αρχική δομή για όλα τα άτομα της πρωτεΐνης(all atoms), όλα τα άτομα εκτός των πρωτονίων(heavy atoms), μόνο τα Ca άτομα του πρωτεϊνικού σκελετού(Ca atoms) σε συνάρτηση με το χρόνο(ns). Το γράφημα αυτό, όπως και τα υπόλοιπα μέχρι και την παράγραφο 4.1.4, σχεδιάστηκε με το πρόγραμμα Xmgr (Turner, P.J. 1995).

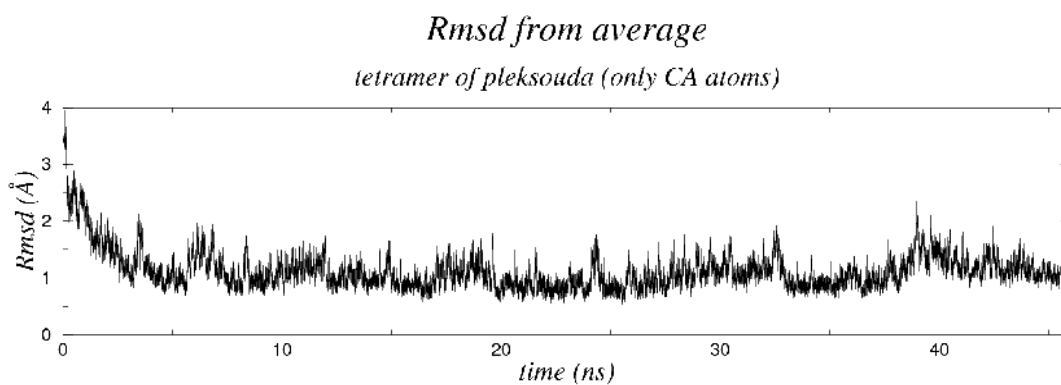
Παρατηρώντας το rmsd από την αρχική δομή για κάθε μονομερές ξεχωριστά, (εικόνα 4.2) και λαμβάνοντας υπόψη μόνο τα Ca άτομα, συμπεραίνουμε ότι τα μονομερή A, B, C παρέμειναν αρκετά όμοια με την αρχική τους δομή αφού έχουν σχετικά σταθερό και μικρό rmsd 2.33Å, 2.34Å 1.7Å αντίστοιχα, με μικρές διακυμάνσεις στην τιμή του, γύρω στο 0.26Å. Αντίθετα, η δομή του μονομερούς D, η οποία κατά τη διάρκεια του trajectory απέκλινε αρκετά από την αρχική της διαμόρφωση, είχε υψηλή τιμή rmsd 3.65 Å και υψηλή διακύμανση 0.52 Å. Υποθέτουμε, λοιπόν, ότι το μονομερές D επηρεάζει σε μεγάλο βαθμό τη σταθερότητα και τη διακύμανση της πρωτεΐνης η οποία εμφανίζει την συμπεριφορά της εικόνας 4.1.



*Εικόνα 4.2 Rmsd από την αρχική δομή των 4 μονομερών, για τα άτομα του πρωτεϊνικού σκελετού.*

#### 4.1.2 Σταθερότητα δομών από το trajectory και υπολογισμός μέσης δομής

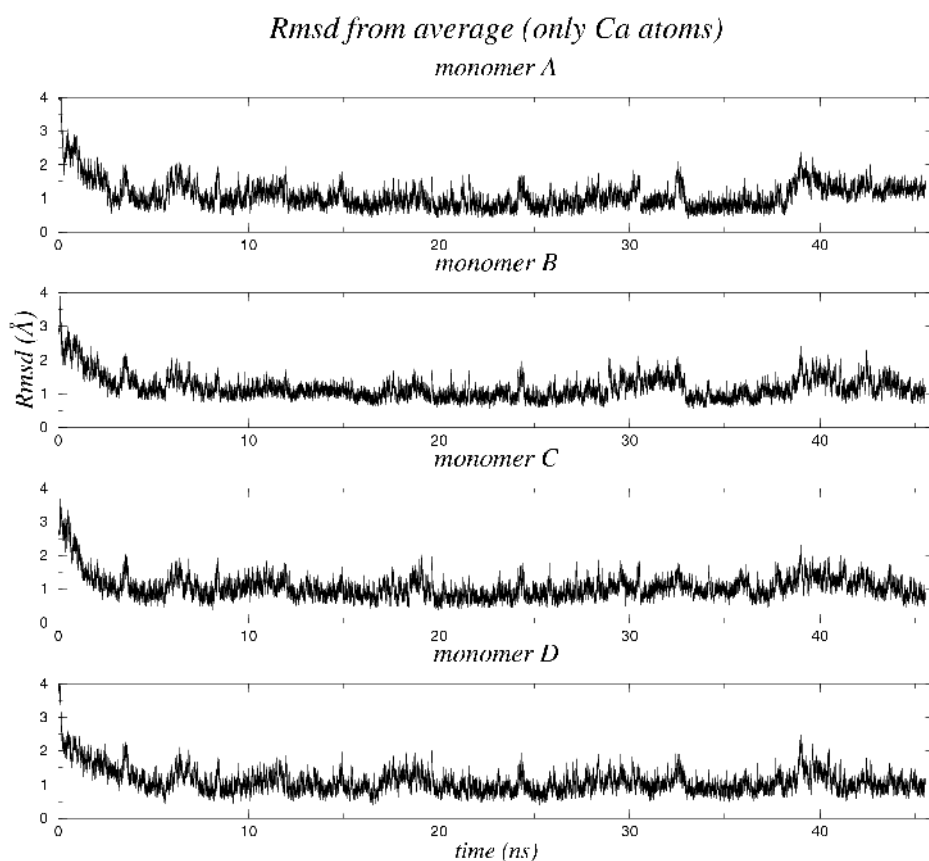
Ο επόμενος υπολογισμός είναι το Rmsd από τη μέση δομή, το οποίο μας πληροφορεί για το πόσο σταθερό είναι το μόριο, αν σταθεροποιείται σε μια δομή ή αν συγκλίνει σε πολλές. Υπολογίζουμε την τιμή του βάσει του τύπου 1 μόνο που το (x) αντιπροσωπεύει τις συντεταγμένες των ατόμων από την μέση δομή και όχι από την αρχική. Επιπλέον μέσω αυτού του υπολογισμού παίρνουμε και τη μέση δομή του trajectory η οποία δεν είναι μια πραγματική δομή που εμφανίζεται στο trajectory κατά τη διάρκεια της προσομοίωσης, αλλά την υπολογίζουμε με βάση το σύνολο των διαμορφώσεων που λαμβάνει το μόριο κατά τη διάρκεια αυτού του trajectory (τροχιά).



Εικόνα 4.3 Το Rmsd από την μέση δομή για τα Ca άτομα του πρωτεϊνικού σκελετού της πρωτεΐνης.

Στην εικόνα 4.3 απεικονίζεται το rmsd από τη μέση δομή για τα άτομα του σκελετού του μορίου της πλεξούδας. Βλέπουμε ότι σταθεροποιήθηκε σε μια μέση δομή σχετικά γρήγορα με rmsd σε μικρή τιμή  $1,12\text{Å}$ , οπότε συμπεραίνουμε ότι το μόριο είναι αρκετά σταθερό, με διακύμανση  $0,34\text{Å}$ .

Αν όμως κοιτάσουμε μεμονωμένα τα rmsd από την μέση δομή για κάθε μονομερές της πρωτεΐνης (εικόνα 4.4), θα δούμε ότι το rmsd για το μονομερές A, είναι αρκετά σταθερό, διακύμανση της τάξης του  $0,41\text{Å}$  και τιμή γύρω στο  $1,09\text{Å}$  με μια μικρή αύξηση περίπου στα 38ns. Το μονομερές B, έχει σταθερό rmsd  $1,18\text{Å}$  με διακύμανση  $0,34\text{Å}$  μέχρι τα 33ns, όπου εμφανίζει μια απότομη αλλαγή. Το μονομερές C φαίνεται να είναι το πιο σταθερό, καθώς εμφανίζει rmsd  $1,07\text{Å}$  με  $0,35\text{Å}$  διακύμανση, ενώ το μονομερές D έχει επίσης μικρό rmsd  $1,10\text{Å}$ , με διακύμανση  $0,36\text{Å}$ .



**Εικόνα 4.4** Το *Rmsd* από την μέση δομή για τα άτομα του πρωτεϊνικού σκελετού των μονομερών της πρωτεΐνης

Παρατηρώντας συγκριτικά τις αποκλίσεις από τη μέση και από την αρχική δομή των μονομερών, βλέπουμε ότι τα μονομερή A, B και C διαφέρουν αρκετά από την αρχική τους δομή, ωστόσο το *rmsd* μέσης είναι ιδιαίτερα σταθερό, το οποίο σημαίνει ότι παρόλο που απέχουν από την αρχική του δομή, βρίσκουν πολύ γρήγορα μια μέση και σταθερή δομή. Η διαφορετική συμπεριφορά των μονομερών εξετάζεται λεπτομερώς σε επόμενες αναλύσεις.

**Πίνακας 4.1** Συγκεντρωτικός πίνακας των μέσων τιμών και τοπικών αποκλίσεων των *Rmsd* από την αρχική και τη μέση δομή σε Å.

	RMSD FROM STARTING		RMSD FROM AVERAGE	
	Μέση τιμή	Διακύμανση	Μέση τιμή	Διακύμανση
<b>A monomer</b>	2.33	0.26	1.09	0.41
<b>B monomer</b>	2.34	0.27	1.18	0.34
<b>C monomer</b>	1.70	0.26	1.07	0.35
<b>D monomer</b>	3.65	0.52	1.10	0.36
<b>All atoms</b>	4.43	0.44	1.12	0.34
<b>Heavy atoms</b>	4.22	0.46		
<b>Ca atoms</b>	3.66	0.44		



### 4.1.3 Ανάλυση της κινητικότητας των ατόμων (rmsf)

Συνεχίζοντας την ανάλυση του μοντέλου «πλεξούδα» θα μελετήσουμε την κινητικότητα κάθε καταλοίπου της πρωτεΐνης γι' αυτό χρειαζόμαστε την παράμετρο rmsf (root-mean-square-fluctuations), η οποία μας πληροφορεί για τις διακυμάνσεις των ατόμων σε σχέση με την θέση που είχαν στη μέση δομή και τονίζει τα ευκίνητα κατάλοιπα. Ο τύπος υπολογισμού του είναι:

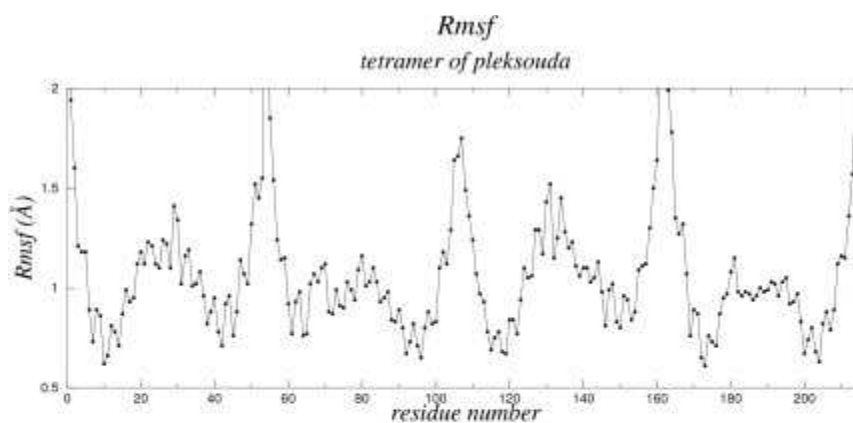
$$RMS_i^{fluct} = \sqrt{\frac{1}{N_f} \sum_f (r_i^f - r_i^{ave})^2} \quad (2)$$

όπου  $N$  ο συνολικός αριθμός των ατόμων,  $r_i^{ave}$  είναι η μέση θέση του κάθε ατόμου  $i$  και  $r_i^f$  η θέση του ατόμου  $i$  στο στιγμιότυπο  $f$ . Τα πλεονέκτημα της παραμέτρου αυτής είναι η δυνατότητα άμεσης σύγκρισης αυτών των τιμών με τις πειραματικά προσδιορισμένες τιμές των  $B$  παραγόντων ( $B$ -factors) από κρυσταλλογραφικές δομές, στην περίπτωση όπου οι διακυμάνσεις είναι ισοτροπικές (Cregut et al., 1998; Garemyr & Elofsson, 1999). Ο τύπος για τη σύγκριση τους είναι:

$$B_i = \frac{8}{3} \pi^2 (RMS_i^{fluct})^2 \quad (3)$$

Η μεγαλύτερη κινητικότητα συνεπώς και οι μεγαλύτερες διακυμάνσεις, παρατηρούνται (εικόνα 4.5) στα άκρα των τεσσάρων ελίκων, ενώ υπάρχει και μια υψηλότερη-από-τη-μέση κινητικότητα γύρω από το κέντρο του δεματίου, όπως συμβαίνει και με τα rmsf των καταλοίπων της πρότυπης δομής της πλεξούδας, την RM6 (Glykos et al., 2006).

Στο γράφημα (εικόνα 4.5) απεικονίζονται διαδοχικά τα κατάλοιπα των τεσσάρων αλυσίδων, όπου μεγαλύτερη κινητικότητα εμφανίζουν τα τελευταία κατάλοιπα της αλυσίδας A και ακολουθούν της D με τιμές από 1.5 - 4.8Å. Στα κατάλοιπα γύρω από το κέντρο του δεματίου εμφανίζονται επαναλαμβανόμενες αυξομειώσεις της τιμής του rmsf τους που αντιστοιχεί στην κατανομή των καταλοίπων σύμφωνα με το μοτίβο της επτάδας. Κανονικά τα υδρόφοβα κατάλοιπα θα πρέπει να έχουν μικρό rmsf στον πυρήνα ενώ τα εκτεθειμένα μεγαλύτερο. Ωστόσο, στην προκειμένη περίπτωση λόγω της δομής της πλεξούδας στο εσωτερικό της υπάρχει μεγάλη διακύμανση των καταλοίπων διότι το πακετάρισμα δεν φαίνεται να είναι ισχυρό (δείτε την εικόνα 4.32 για την κινητικότητα των υδρόφοβων καταλοίπων).



**Εικόνα 4.5** Απεικόνιση της κινητικότητας των Ca ατόμων της πρωτεΐνης σύμφωνα με την παράμετρο Rmsf. Για λόγους ενκρίνειας έχουν αφαιρεθεί από την εικόνα τα κατάλοιπα εκείνα τα οποία είχαν υψηλό rmsf (πάνω από 2 Å), τα οποία είναι: Glu(E) 1 της A αλυσίδας, Glu(E) 55 της B αλυσίδας, Phe(F) 162 της C αλυσίδας, Glu(E) 163, Arg(R) 215 και Phe(F) 216 της αλυσίδας D.

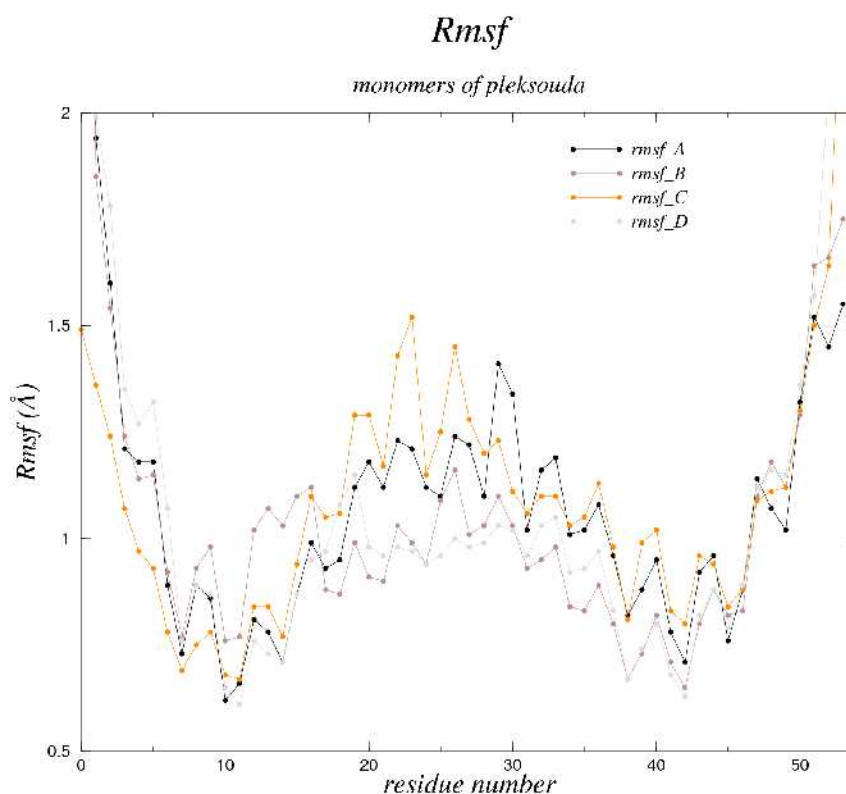
Για να ποσοτικοποιήσουμε τις παραπάνω παρατηρήσεις ελέγξαμε τον γραμμικό συντελεστή συσχέτισης μεταξύ των μονομερών όπου χρησιμοποιήθηκε ένα μικρό πρόγραμμα σε γλώσσα προγραμματισμού C (παράρτημα 4.1) προκειμένου να αλλάξει η φορά των C και D αλυσίδων οι οποίες είναι αντίρροπες από τις A και B. Κανονικά θα πρέπει να έχουν μεγάλη συσχέτιση, από τα δεδομένα, τα μονομερή A με D και B με C.

**Πίνακας 4.2**

Γραμμικός συντελεστής συσχέτισης μεταξύ των μονομερών A, B, C, D:

-	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>A</b>	1	0.915 (0.820)	0.858 (0,760)	<b>0.870 (0,777)</b>
<b>B</b>		1	<b>0.870 (0,630)</b>	0.849 (0,799)
<b>C</b>			1	0.791 (0,452 )
<b>D</b>				1

Με βάση όμως την λεπτομερή παρατήρηση της κινητικότητας του κάθε μονομερούς συγκριτικά με τα υπόλοιπα (εικόνα 4.6) και βασιζόμενοι στον συντελεστή συσχέτισης τους φαίνεται ότι κάτι τέτοιο δεν ισχύει όπως δείχνουν οι τιμές του πίνακα 4.2. Αυτό ίσως συμβαίνει γιατί κάποια κατάλοιπα στην αρχή ή στο τέλος βρίσκονται στο ίδιο σημείο με υψηλή τιμή με αποτέλεσμα να μεγαλώνει ο συντελεστής παρόλο που στην υπόλοιπη αλληλουχία δεν έχουν εμφανή συσχέτιση.



Εικόνα 4.6 Rmsf των καταλοίπων των μονομερών της πρωτεΐνης.

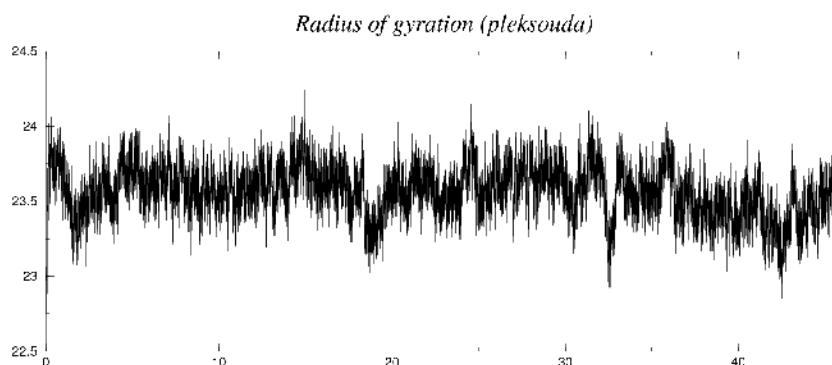
Γι' αυτό το λόγο κάνουμε ξανά τους ίδιους υπολογισμούς αλλά με cutoff  $2\text{\AA}$  ώστε μην υπολογίζονται τα υπερκινητικά κατάλοιπα και να δίνεται σημασία στο σύνολο των διακομάνσεων των καταλοίπων των αλληλουχιών όμως χωρίς κάποια σημαντική διαφορά. Μια υπόθεση που μπορούμε να κάνουμε και να βασιστούμε στην παράγραφο 4.1.5 είναι ότι δεν υπάρχει υψηλή συσχέτιση, ίσως λόγω της υψηλής κινητικότητας της αλυσίδας C και γι' αυτό δεν εμφανίζει με καμιά αλυσίδα υψηλή συσχέτιση και επηρεάζει όλο το τετραμερές. Στην περίπτωση όπου, τα ζεύγη των μονομερών που αναφέρθηκαν, εμφάνιζαν μεγάλη συσχέτιση, θα μπορούσαμε να ισχυριστούμε ότι το μοντέλο της πλεξούδας είναι διμερές διμερών, όπως και η πρότυπη δομή του RM6 (κεφάλαιο 2).

#### 4.1.4 Εξέλιξη του Radius of gyration (Rg)

Το Rg μιας ομάδας ατόμων ορίζεται ως η απόσταση rms για κάθε άτομο του μορίου από το κέντρο βάρους του μορίου (Kuszewski et al., 1998). Γι' αυτό το λόγο είναι ενδεικτικό του μεγέθους και του κατά πόσο συμπαγές είναι το μόριο. Έτσι ελέγχουμε την σταθερότητα αλλά και το βαθμό διατήρησης της δομής του. Ο τύπος του είναι:

$$RadiusGyration = \sqrt{\frac{1}{Ni} \sum_i (r_i - r_{cm})^2} \quad (4)$$

όπου N ο συνολικός αριθμός ατόμων,  $r_i$  είναι η απόσταση του κάθε ατόμου  $i$  από το κέντρο βάρους του μορίου και  $r_{cm}$  είναι η μέση απόσταση του ατόμου από το κέντρο βάρους του.



Εικόνα 4.7 Εξέλιξη του Rg για το μόριο της πλεξούδας.

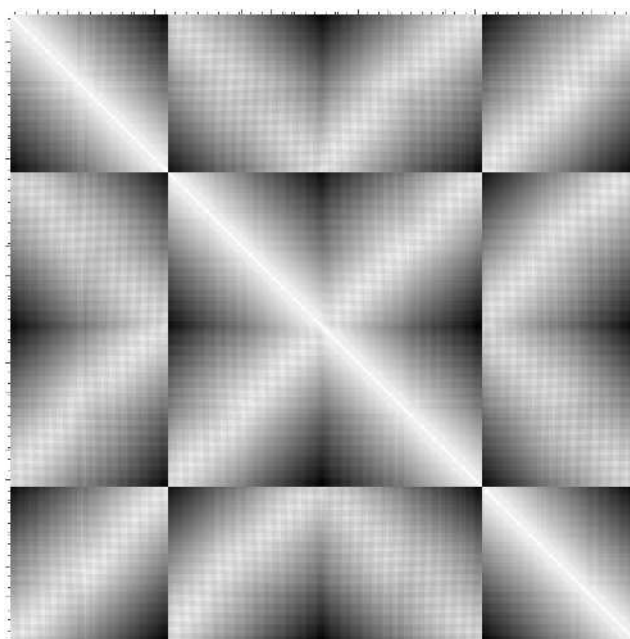
Παρατηρούμε στην εικόνα 4.7 ότι η τιμή του Rg για το μόριο της πλεξούδας κομμάινεται γύρω από μια μέση τιμή 23.5Å, με μεγάλη διακύμανση. Φαίνεται αρκετά σταθερό στα σημεία πριν τα 20 ns όπου πέφτει, το ίδιο και μετά τα 30ns, ενώ μετά τα 35ns εμφανίζει μια φθίνουσα πορεία με μικρή κλίση.

#### 4.1.5 Χάρτης αποστάσεων Ca-Ca ατόμων και ανάλυση των διακυμάνσεων του.

Μελετώντας με μεγαλύτερη προσοχή τη δυναμική της «πλεξούδας» θα εξετάσουμε την μεταβλητότητα των αποστάσεων μεταξύ των ατόμων, η οποία μας πληροφορεί για την κινητικότητα της πρωτεΐνης κατά τη διάρκεια της προσομοίωσης.

- Χάρτης Ca-Ca distance map

Ο χάρτης Ca-Ca distance map (εικόνα 4.8), προκύπτει από έναν δισδιάστατο πίνακα ο οποίος περιλαμβάνει τις μέσες αποστάσεις μεταξύ των ατόμων Ca. Η πάνω οριζόντια γραμμή και η αριστερή κάθετη γραμμή αντιστοιχούν στα άτομα του κάθε μονομερούς A, B, C, D. Ο πίνακας είναι συμμετρικός με άξονα συμμετρίας την διαγώνιο η οποία αντιπροσωπεύει την απόσταση του κάθε Ca από τον εαυτό του. Δεν χρησιμοποιούμε όλα τα στιγμιότυπα (113803 frames) στον πίνακα αλλά ένα πιο μικρό αντιπροσωπευτικό trajectory ανά 100 frames το οποίο θα περιλαμβάνει μονάχα τα Ca άτομα και θα έχει διαστάσεις 1139x1139 ώστε να είναι πιο αποτελεσματικοί οι υπολογισμοί.



Εικόνα 4.8 Ca-Ca distance map

Η διαβάθμιση του χρώματος (εικόνα 4.8) από λευκό σε μαύρο ακολουθεί την αύξηση της απόστασης μεταξύ των ατόμων όπου με λευκό απεικονίζονται οι μικρότερες αποστάσεις και με μαύρο οι μεγαλύτερες, γι' αυτό και η διαγώνιος είναι λευκή. Η μέγιστη απόσταση που εμφανίζεται ανάμεσα στα άτομα είναι 77,3 Å, ενώ επιβεβαιώνεται ότι τα μονομερή C

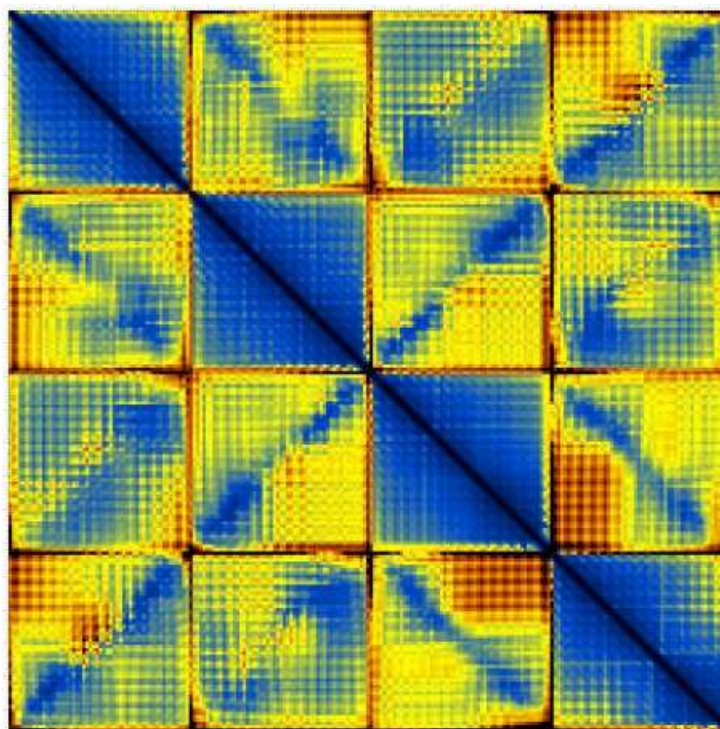
και D είναι αντιπαράλληλα σε σχέση με τα A και B. Μεγάλες αποστάσεις εμφανίζονται μεταξύ ατόμων των άκρων του ίδιου μονομερούς και άκρων αντίρροπων μονομερών.

- **Rmsd από τις μέσες αποστάσεις**

Ο χάρτης rmsd, αντιστοιχεί στις αποκλίσεις των μέσων αποστάσεων των ατόμων του προηγούμενου πίνακα (εικόνα 4.8) δίνοντας πληροφορίες για τη διακίνηση τους. Συγκεκριμένα βλέπουμε ποια κατάλοιπα εμφανίζουν τη μεγαλύτερη κινητικότητα και βρίσκουμε τα ευκίνητα και τα σταθερά τμήματα της πρωτεΐνης. Περιέχει πληροφορία σχετικά με την σταθερότητα της πρωτεΐνης και απεικονίζει διαφορές μεταξύ των σχετικά στερεών τμημάτων της δομής και των υψηλά ευκίνητων τμημάτων της. Τμήματα της δομής τα οποία φέρουν χαμηλές τιμές rmsd μεταξύ των Ca ατόμων τους μπορεί να θεωρηθούν ως στερεά σώματα (rigid body).

Στην εικόνα 4.9 ισχύει ό, τι προαναφέρθηκε για τις σχέσεις μεταξύ των μονομερών, ενώ η κινητικότητα των ατόμων απεικονίζεται ως εξής: με μπλε χαμηλή κινητικότητα άρα σταθερότητα, με κόκκινο υψηλή κινητικότητα και με κίτρινο μια ενδιάμεση κατάσταση. Παρατηρώντας το γράφημα βλέπουμε ότι τα τετράγωνα της διαγώνιου που αντιστοιχούν στο κάθε μονομερές εμφανίζουν αρκετό μπλε, άρα αρκετά κατάλοιπα είναι σταθερά σε κάποια τμήματα της εκάστοτε αλυσίδας. Ειδικότερα για τα μονομερή A και C θα μπορούσαμε να θεωρήσουμε ότι αποτελούνται από 2 τμήματα (μπλε τετράγωνα), όπου οι αποκλίσεις των Ca είναι σχετικά ίδιες, το D εμφανίζει 3 τμήματα σταθερά, ενώ στο B τα τμήματα είναι άνισα κατανομημένα. Τον διαχωρισμό αυτό της κάθε αλυσίδας σε 2 ή 3 τμήματα τον περιμέναμε λόγω της δομής της πρωτεΐνης η οποία φαίνεται να αποτελείται από 3 κομμάτια, δηλαδή το τμήμα που περιλαμβάνει τα κατάλοιπα εισαγωγής και τα άλλα 2 τμήματα πριν και μετά την εισαγωγή. Επιπλέον, σχηματίζεται μια σκακιέρα από μαύρες γραμμές στις άκρες των τετραγώνων οι οποίες αντιστοιχούν στα αμινοτελικά (N-τελικό) και καρβοξυτελικά (C-τελικό) τελικά άκρα τα οποία φέρουν κατάλοιπα με μεγάλη κινητικότητα.

Συγκεκριμένα η πρώτη μαύρη γραμμή είναι στο N τελικό άκρο το κατάλοιπο Glu 1 του A μονομερούς, η δεύτερη στο N τελικό είναι το Glu 55 το οποίο είναι το 1ο κατάλοιπο της B αλυσίδας. Οι επόμενες αντιστοιχούν στο C τελικό, το αμινοξύ φαινυλαλανίνη 108 της B αλυσίδας, ακολουθεί στο C τελικό άκρο η φαινυλαλανίνη 162 της C αλυσίδας και η φαινυλαλανίνη 216 στο C τελικό άκρο της D αλυσίδας.



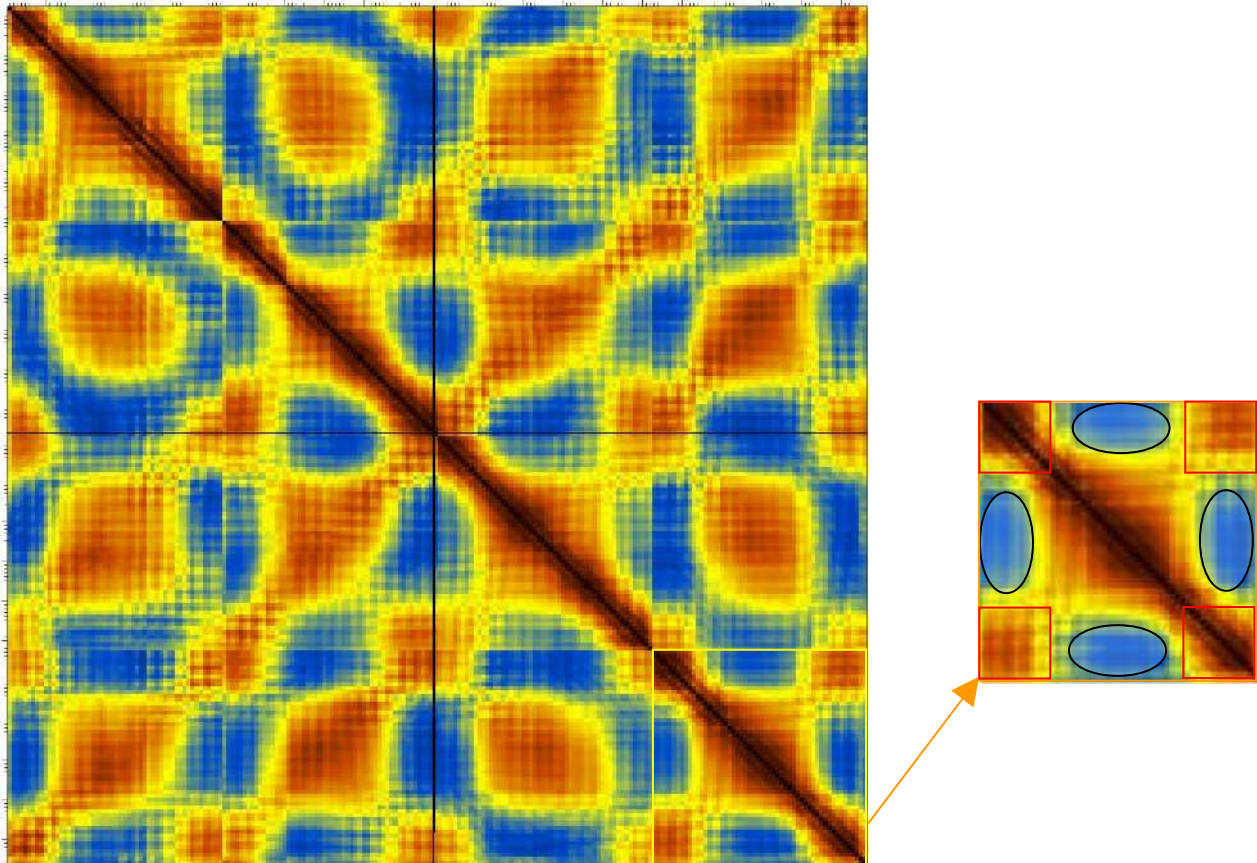
Εικόνα 4.9 Rmsd του χάρτη Ca-Ca distance map

Επίσης, το γράφημα εμφανίζει χαρακτηριστικά αντιπαράλληλου δεματίου. Αν δούμε και τις ατομικές κινητικότητες των Ca από το rmsf (εικόνα 4.5) βλέπουμε ότι τα πιο κινητικά κατάλοιπα είναι το 55(E), 108(F), 162(F), 216(F), οπότε συμφωνούν τα αποτελέσματα του rmsf με το χάρτη rmsd Ca-Ca. Η κινητικότητα αυτών των καταλοίπων δικαιολογείται διότι βρίσκονται στα άκρα του δεματίου.

#### 4.1.6 Υπολογισμός cross correlation και ανάλυση

Η γνώση των ατομικών κινήσεων και της μεταξύ τους συσχέτισης, στις πρωτεΐνες παίζει κεντρικό ρόλο στην κατανόηση της βιολογικής λειτουργίας των πρωτεϊνών (Huber & Bennett, 1983). Για την ανάλυση, λοιπόν, της κίνησης ομάδων ατόμων δημιουργούμε τον πίνακα variance-covariance των ατομικών μετατοπίσεων (Ichiye & Karplus, 1991) και έπειτα με την κανονικοποίηση των τιμών του, κατασκευάζεται το γράφημα cross correlation το οποίο δίνει πληροφορίες για την συσχέτιση των διαφόρων κινήσεων που πραγματοποιεί κάθε τμήμα του μορίου (Lange et al., 2006). Μας δείχνει δηλαδή ποια τμήματα του μορίου κινούνται προς την ίδια ή προς την αντίθετη κατεύθυνση. Δυο τμήματα μπορεί να εμφανίσουν θετική συσχέτιση +1 (κόκκινο χρώμα) όταν κινούνται μαζί, αρνητική συσχέτιση -1 (μπλε) όταν κινούνται αντίθετα ή να μην σχετίζονται μεταξύ

τους 0 (κίτρινο) (Hunenberger et al., 1995).



*Εικόνα 4.10 Ο πίνακας variance-covariance, cross-correlation. Στα δεξιά το χαρακτηριστικό μοτίβο που εμφανίζεται και στα 4 μονομερή. Η διαγώνιος κόκκινη, τα κατάλοιπα των άκρων κινούνται μαζί σχεδιάζονται ως κόκκινο τετράγωνα, ενώ τα τμήματα που δεν κουνιούνται μαζί, τα άκρα με το κεντρικό τμήμα, σχηματίζονται ως μπλε ελλειψοειδή.*

Η έντονη διαγώνια γραμμή (βαθύ κόκκινο χρώμα) (εικόνα 4.10) δηλώνει την κίνηση των ατόμων Ca με τον εαυτό τους τα οποία εμφανίζουν πλήρη συσχέτιση (+1) άρα βαθύ κόκκινο χρώμα. Διακρίνουμε όμως δύο ακόμη γραμμές μαζί με την κεντρική διαγώνιο το οποίο σημαίνει ότι τα γειτονικά κατάλοιπα κινούνται μαζί πιθανόν λόγω ομοιοπολικής σύνδεσης αυτών. Παρατηρώντας το κάθε μονομερές βλέπουμε να εμφανίζεται ένα μοτίβο κίνησης (δεξιά εικόνα) για τα επιμέρους τμήματα του που αντιστοιχεί στη δομή της πλεξούδας η οποία αποτελείται από 3 τμήματα, το κεντρικό της εισαγωγής των καταλοίπων (περιλαμβάνει) και 2 τμήματα πριν και μετά την εισαγωγή. Το ίδιο μοτίβο συσχέτισης καταλοίπων εμφανίζεται και στα τέσσερα μονομερή του μορίου μας αποκαλύπτοντας ότι τα κατάλοιπα ανά περιοχές εμφανίζουν διαφορετική συμπεριφορά στην κινητικότητα τους. Συγκεκριμένα όπως είναι σχεδιασμένο στην εικόνα 4.10 τα άκρα του κάθε μονομερούς κινούνται προς την ίδια κατεύθυνση (κόκκινο χρώμα) και σε αντίθετη κατεύθυνση από το κεντρικό τμήμα του μονομερούς (μπλε), ενώ με κίτρινο



χρώμα είναι τα κατάλοιπα τα οποία κινούνται μόνα τους. Τα στοιχεία που συλλέγουμε για τις κινήσεις του μορίου θα πρέπει να συμφωνούν και με την ανάλυση των PCA (παράγραφος 4.1.11)

Επιπλέον, τα τελικά κατάλοιπα των άκρων των μονομερών θα έπρεπε να είναι ζωγραφισμένα ως κίτρινες γραμμές αφού είναι ευκίνητα και κάνουνε μια κίνηση ξεχωριστή σε σχέση με το υπόλοιπο μόριο, όμως αυτό δεν συμβαίνει στην περίπτωση μεταξύ των άκρων B και C όπου τα όρια δεν είναι ευδιάκριτα και σχεδιάστηκαν μετέπειτα (μαύρες γραμμές). Στην προκειμένη περίπτωση τα άκρα όχι μόνο δεν είναι κίτρινα, αλλά φαίνεται το τέλος του ενός να είναι συνέχεια του άλλου. Ελέγχοντας την αντιπροσωπευτική δομή από το trajectory, διαπιστώνουμε ότι στο άκρο όπου βρίσκονται τα καρβοξυτελικά άκρα των C και D με τα αμινοτελικά άκρα των A και B δημιουργούνται ισχυρές ιοντικές αλληλεπιδράσεις μεταξύ των πλευρικών ομάδων των καταλοίπων των άκρων και συγκεκριμένα:

- μεταξύ του Glu109 και της Ala112 από την αλυσίδα C με απόσταση δεσμού 3.289Å,
- μεταξύ της Phe108 της αλυσίδας B με την Arg53 της A αλυσίδας με απόσταση δεσμού 3.301Å και
- η ισχυρότερη σύνδεση μεταξύ της Arg107, B αλυσίδα, με το Glu163 από την D αλυσίδα και απόσταση δεσμού 2.610Å,

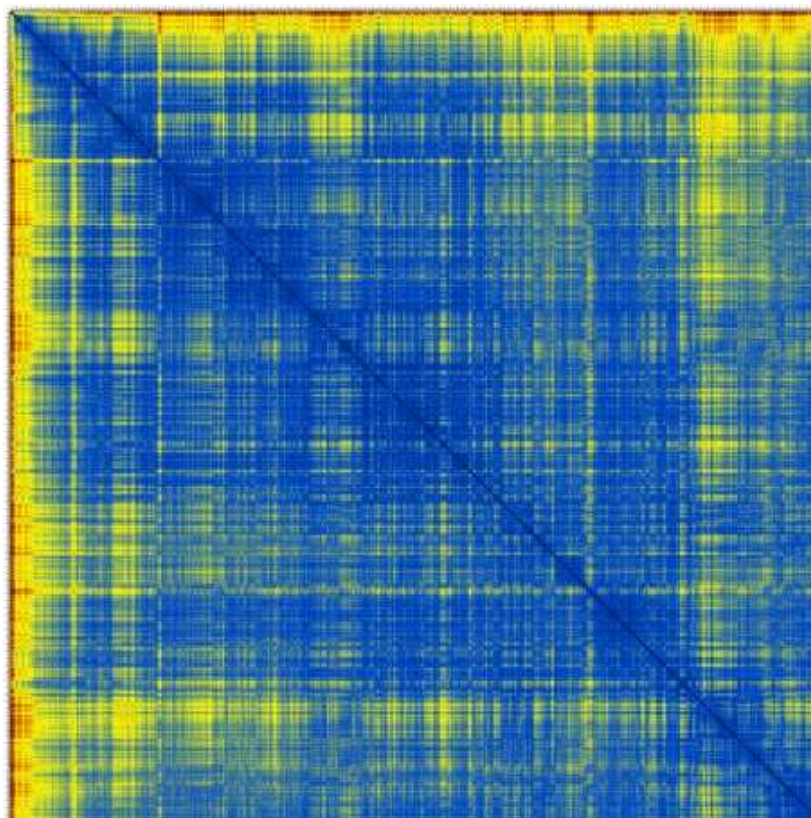
ενώ στο άλλο άκρο δεν συνέβη κάτι ανάλογο, το οποίο οφείλεται ίσως στον τρόπο μοντελοποίησης της δομής.

Οπότε αφού τα τελικά κατάλοιπα των αλυσίδων B και C έχουν αναπτύξει ισχυρές ιοντικές αλληλεπιδράσεις με γειτονικά τους, έχουν πακεταριστεί καλά ώστε σε όλη την προσομοίωση παρέμειναν συνδεδεμένα και κουνιούνται μαζί. Για το γεγονός αυτό μας είχε πληροφορήσει και η παράμετρος rmsf (εικόνα 4.5), η οποία έδινε πολύ μικρές κινητικότητες για τα κατάλοιπα αυτά σε σχέση με τα υπόλοιπα κατάλοιπα των άκρων.

#### **4.1.7 RMSD μεταξύ όλων των πιθανών συνδυασμών δομών από το τροχιακό**

Υπολογίζοντας το rmsd μεταξύ των δομών για κάθε frame (πίνακας frame-to-frame) από τα υπόλοιπα, μας δίνεται η δυνατότητα μελέτης της κινητικότητας του μορίου και των δομών που υιοθετεί και επιπλέον διακρίνεται μια πρώτη ομαδοποίηση των δομών. Οι δομές που συναντάμε σε κάθε frame δεν είναι πανομοιότυπες αλλά εάν το

rmsd της δομής κάποιου frame σε σχέση με το rmsd δομής άλλου frame είναι μικρό τότε οι δομές μπορεί να θεωρεί ότι ανήκουν στην ίδια ομάδα. Έτσι, μπορούμε να γνωρίζουμε με ακρίβεια σε πιο σημείο και πότε συμβαίνουν αλλαγές στο μόριο. Ο πίνακας θα είναι διοδιάστατος και θα περιέχει όλες τις δομές του trajectory που έχει πάρει η πρωτεΐνη κατά τη διάρκεια της προσομοίωσης ανά 100 στιγμιότυπα.

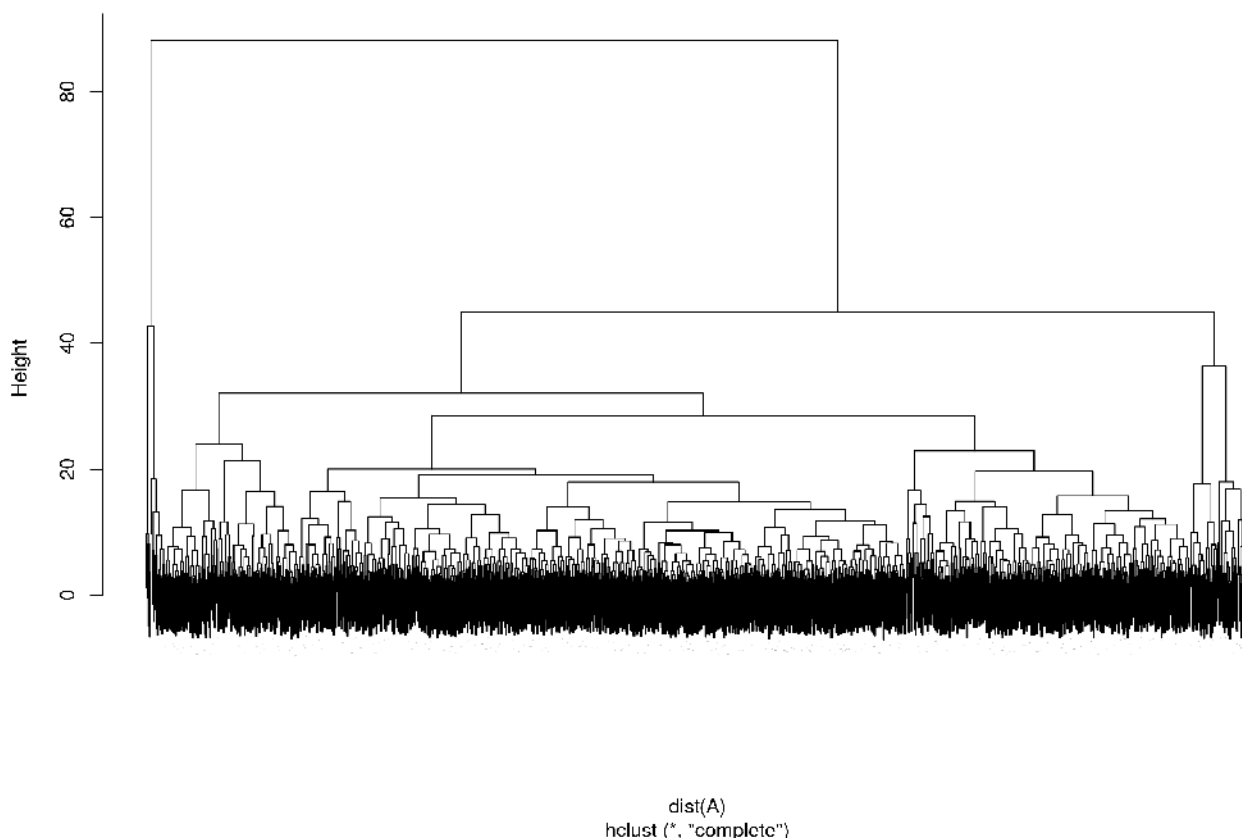


*εικόνα 4.11 Frame-to-frame rmsd (crossDCD.matrix).*

Στην εικόνα 4.11 παρατηρούμε ότι στην αρχή του trajectory (πρώτα 210 frames) η δομή ξεκινάει από μια διαμόρφωση (μικρό μπλε τετραγωνάκι στην αρχή) την οποία εγκαταλείπει σχετικά γρήγορα και παίρνει μια άλλη την οποία διατηρεί για το υπόλοιπο της προσομοίωσης (ο πίνακας είναι όλος μπλε). Η μεγαλύτερη απόσταση που παρατηρείται είναι  $1,5\text{\AA}$ . Συμπεραίνουμε, λοιπόν, ότι υπάρχει ένα μεγάλο cluster και ένα πολύ μικρό. Στοιχείο το οποίο υποδείκνυε και το γράφημα του rmsd από τη μέση δομή (εικόνα 4.3) όπου γρήγορα σταθεροποιήθηκε σε μια δομή.

#### 4.1.8 Ομαδοποίηση των δομών του trajectory (cluster analysis).

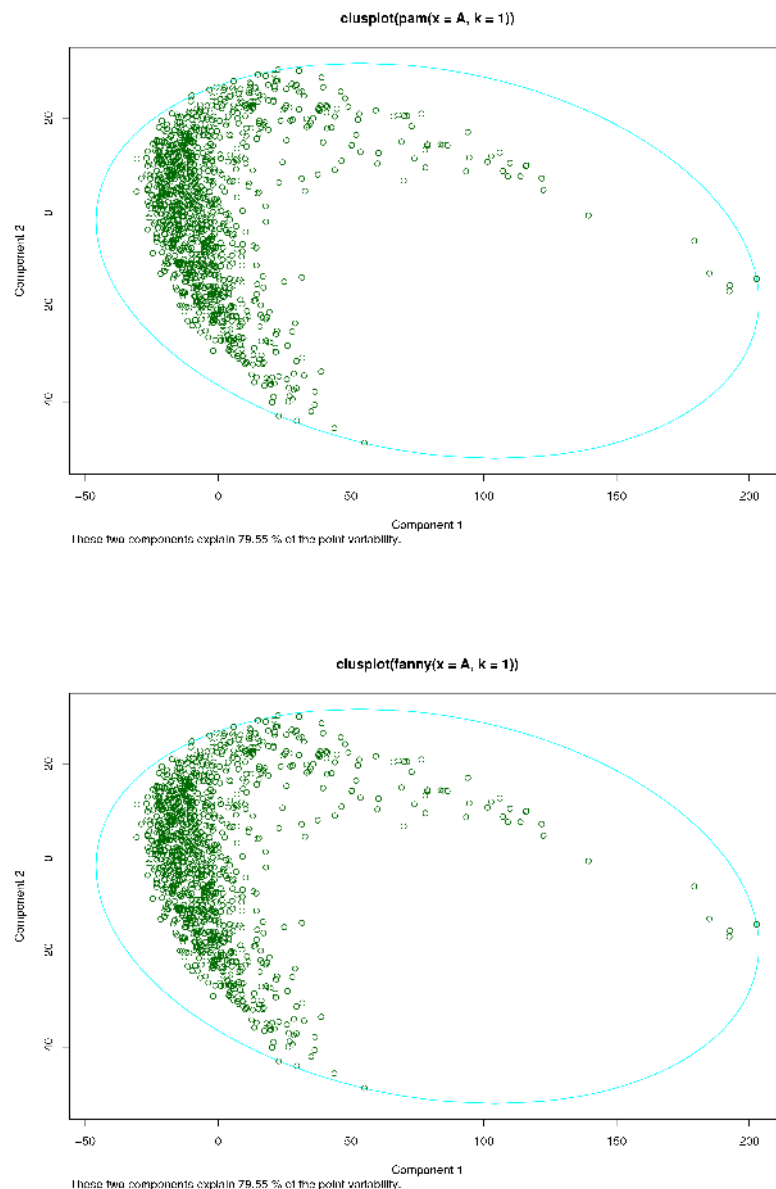
Το cluster analysis είναι μια ανάλυση με ιδιαίτερο ενδιαφέρον διότι ομαδοποιούνται οι δομές που παρατηρούνται κατά τη διάρκεια της προσομοίωσης σε clusters (ομάδες), σύμφωνα με το βαθμό με τον οποίο οι δομές αυτές ομοιάζουν μεταξύ τους ή όχι. Το πρώτο στάδιο είναι να κατασκευαστεί ένα δενδρόγραμμα, ανάλογο των φυλογενετικών δένδρων, βάση του πίνακα αποστάσεων rmsd (εικόνα 4.11) και έπειτα να προσδιοριστούν οι δομές για κάθε cluster. Τα clusters που θα προκύψουν φανερώνουν τις κύριες δομές που υιοθετεί το μόριο, ενώ το ποσοστό των δομών που περιέχει το κάθε cluster δείχνει το χρονικό διάστημα παραμονής του μορίου στη συγκεκριμένη διαμόρφωση.



Εικόνα 4.12 Δενδρόγραμμα των δομών της πλεξούδας που παρατηρούνται κατά τη διάρκεια της προσομοίωσης χρησιμοποιώντας τον πίνακα αποστάσεων rmsd.

Όλοι οι υπολογισμοί γίνονται στατιστικά με το πρόγραμμα R (R Development Core Team, 2004). Σύμφωνα με τον ιεραρχικό τρόπο απεικόνισης τα δεδομένα αναπαρίσταται σε ένα

δενδρογράμμα (hierarchical clustering) όπως στην εικόνα 4.12 όπου φαίνεται να υπάρχουν δυο cluster από τα οποία το ένα είναι πολύ μικρό και αντιστοιχεί στο στάδιο του heating, ενώ το δεύτερο είναι πολύ μεγαλύτερο και είναι και το κυρίαρχο cluster, έτσι επιβεβαιώνεται η υπόθεση από τον frame-to-frame πίνακα (εικόνα 4.11). Υπάρχει ένα κυρίαρχο cluster μέσα στο οποίο φαίνονται και άλλα μικρότερα τα οποία όμως επικαλύπτονται άρα οι δομές μοιάζουν αρκετά.



**Εικόνα 4.13** Προσδιορισμός των cluster σύμφωνα με τους αλγόριθμους pam (επάνω) και fanny (κάτω). Είναι φανερό ότι οι δομές ομαδοποιούνται σε ένα κυρίαρχο cluster.

Με βάση τον αριθμό από cluster που υπολογίσαμε βάση του δενδρογράμματος θα χρησιμοποιήσουμε τους αλγόριθμους pam και funny (partitional clustering) για

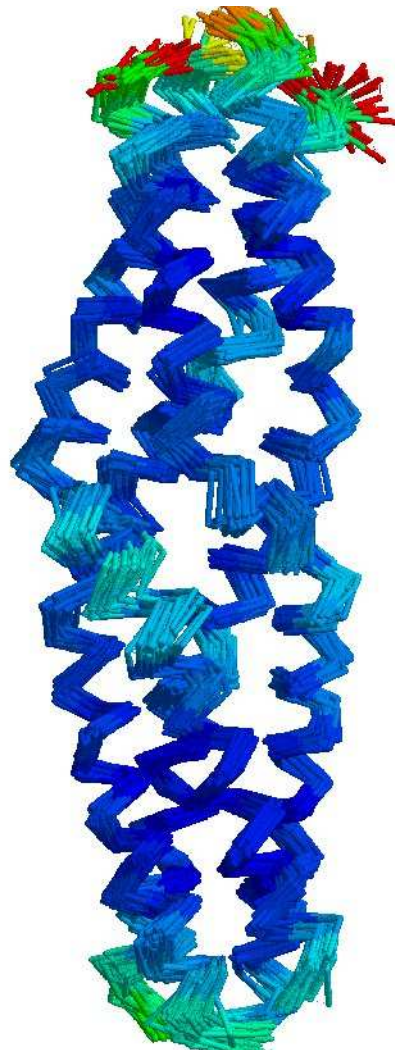
στατιστική ανάλυση και να προσδιορίσουμε το ποσοστό δομών του κάθε cluster ώστε να γνωρίζουμε με ακρίβεια σε ποιο cluster δομών ανήκει το κάθε frame. Από τους αλγόριθμους αυτούς προκύπτουν δύο γραφήματα το πρώτο, cluster plot, είναι αυτό που προσδιορίζει τα cluster και το δεύτερο, silhouette plot, δείχνει την κατανομή των δομών ανά cluster σε ποσοστά.

Οι δυο αλγόριθμοι pam και fanny έδειξαν ότι υπάρχει 1 κυρίαρχο cluster στο οποίο συμπεριλαμβάνονται και τα frames του heating που στο δένδρογραμμα είναι το πρώτο μικρό cluster και στην εικόνα 4.13 οι απομακρυσμένες τιμές στα δεξιά και των δύο γραφημάτων. Δεν χρειάστηκε να σχεδιάσει γράφημα silhouette plot αφού έχουμε 1 cluster στο οποίο ανήκουν όλες οι δομές.

#### **4.1.9 Υπέρθωση των δομών με τα αντιστοιχα rmsf fluctuations του κυρίαρχου cluster τους.**

Κατά τη διαδικασία της ομαδοποίησης των δομών του trajectory μέσω του στατιστικού προγράμματος R αλλά και μέσω του προγράμματος carma με την ανάλυση των κύριων συνιστωσών (Principal Component Analysis, PCA) που ακολουθεί αναλυτικά στην παράγραφο 4.1.11, πληροφορούμαστε για την ύπαρξη ενός μοναδικού κυρίαρχου cluster όπως επίσης δείχνει και ο πίνακας frame to frame (εικόνα 4.11).

Στην εικόνα 4.14 φαίνεται η υπέρθεση των δομών του trajectory με βήμα 1000 frames, χρησιμοποιώντας στη θέση των ατομικών θερμικών παραγόντων (B factors), τα rmsf από την μέση δομή. Η εικόνα αναπαριστά την κινητικότητα των καταλοίπων (του κυρίαρχου cluster) κατά την διάρκεια του trajectory σύμφωνα με μια χρωματική κλίμακα από το μπλε (μικρή κινητικότητα) στο κόκκινο (μεγάλη κινητικότητα). Παρατηρούμε ότι μεγάλη κινητικότητα εμφανίζουν τα κατάλοιπα στα άκρα του δεματίου και στις τέσσερις αλυσίδες και ιδιαίτερα στο επάνω άκρο της δομής που αντιστοιχεί στο N τελικό άκρο των A και B αλυσίδων και στο C τελικό των C και D με ιδιαίτερα ευκίνητα κατάλοιπα τα Glu 1 της A και Glu 55 της B (κόκκινο χρώμα), Phe 162 της C (κίτρινο) και Phe 216 της D (πορτοκαλί) με rmsf μεγαλύτερο από 2Å.

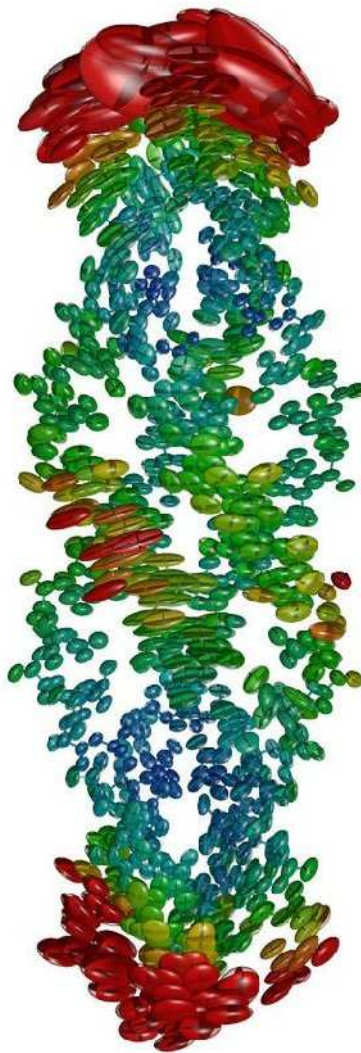


**Εικόνα 4.14** Υπέρθυση των δομών του κορίαρχου cluster κάθε 1000 frame, χρησιμοποιώντας τα *rmsf* της μέσης δομής στη θέση των ατομικών θερμικών παραγόντων (*B factors*). Αναπαράσταση της κινητικότητας των κατάλοιπων σύμφωνα με τη χρωματική κλίμακα από το μπλε (μικρή κινητικότητα) στο κόκκινο (μεγάλη κινητικότητα). Τα *rmsf* τα οποία έχουν χρησιμοποιηθεί στη θέση των ατομικών θερμικών παραγόντων (*B factors*) για την απόδοση της χρωματικής κλίμακας έχουν τροποποιηθεί με κοινή κλίμακα από μπλε (0.47Å) - κόκκινο (2.82Å) για λόγους σύγκρισης με το δεύτερο μοντέλο *sigmoidal*.

Επιπλέον, με πράσινο χρώμα απεικονίζονται τα κατάλοιπα της A αλυσίδας Lys 2 και Thr 3, της B αλυσίδας Lys 65 και Thr 57, της C αλυσίδας Arg 161 και Ala 160, της D αλυσίδας Arg215 και Ala 214 με *rmsf* μεγαλύτερο από 1.5 Å. Τέλος, παρατηρούμε μια ελαφριά αύξηση της κινητικότητας (γαλάζιο) των καταλοίπων στο ενδιάμεσο τμήμα των αλυσίδων και συγκεκριμένα στα κατάλοιπα Asp 135, Glu 132, Glu 128 της C αλυσίδας, στα κατάλοιπα Asp30 της A και Thr 71 της B και Glu 182 της D.

#### 4.1.10 Anisotropic fluctuations

Όπως αναφέραμε οι πληροφορίες οι οποίες φέρουν οι χάρτες των μέσων αποστάσεων (Ca-Ca rmsd) είναι περιορισμένες και αγνοούν παντελώς την δυναμική του trajectory. Γι' αυτό τον λόγο χρησιμοποιούμε μια γραφική αναπαράσταση που βασίζεται στον χάρτη αυτόν και αναπαριστά την μέση δομή της πλεξούδας χρησιμοποιώντας τους θερμικούς ανισότροπους παράγοντες ως ελλειψοειδή.



*Εικόνα 4.15* Απεικόνιση της μέσης δομής της πλεξούδας χρησιμοποιώντας τους θερμικούς ανισότροπους ελλειψοειδείς παράγοντες (*thermal anisotropic ellipsoids*) υπολογιζόμενοι από την κίνηση του σκελετού και των C $\beta$  ατόμων κατά τη διάρκεια της προσομοίωσης με πιθανότητα κατανομής της πυκνότητας των ατομικών θέσεων κατά 90%. Τα rmsf τα οποία έχουν χρησιμοποιηθεί στη θέση των ατομικών θερμικών παραγόντων (*B factors*) για την απόδοση της χρωματικής κλίμακας έχουν τροποποιηθεί με κοινή κλίμακα από μπλε (0.47Å) - κόκκινο (2.82Å) όπως και στην εικόνα 4.10. Αυτή η εικόνα ετοιμάστηκε με το πρόγραμμα *Raster3D* (Merritt & Bacon, 1997)

Τα ελλειψοειδή υπολογίζονται με βάση την κίνηση του σκελετού και των C $\beta$  ατόμων κατά τη διάρκεια της προσομοίωσης. Οι ατομικές κινητικότητες είναι αποτέλεσμα της θερμικής κίνησης των ατόμων βάση της μέσης δύναμης η οποία κυμαίνεται σε ένα τοπικό δυναμικό ελάχιστο (York et al., 1993). Η διαθεσιμότητα από ένα trajectory των ατομικών κινήσεων επιτρέπει τον υπολογισμό των ανισότροπων ελλειψοειδών περιγράφοντας έτσι την πιθανότητα κατανομής της πυκνότητας των ατομικών θέσεων, με αρκετά όμοιο τρόπο με τα κρυσταλλογραφικά ανισότροπα θερμικά ελλειψοειδή. Η κατανομή των ελλειψοειδών αποκαλύπτει την δομικά σημαντική κινητικότητα των καταλοίπων. Με βάση το μέγεθος και το χρώμα των ελλειψοειδών συμπεραίνουμε την κίνηση του σκελετού και των C $\beta$  ατόμων του, δηλαδή αποκαλύπτει την ευελιξία σημείων και τμημάτων των δομών.

Συγκεκριμένα, στην εικόνα 4.15, παρατηρούμε εκτεταμένη κίνηση στα άκρα του δεματίου και κυρίως στο πάνω άκρο, στο N τελικό άκρο των A, B αλυσίδων και στο C τελικό των C, D που αντιστοιχούν στα κατάλοιπα Glu 1 της A και Glu 55 της B, Phe 162 της C και Phe 216 της D ως μεγαλύτερα ελλειψοειδή. Με μικρότερη κινητικότητα, σχηματίζονται ως μικρότερα ελλειψοειδή, τα κατάλοιπα Lys 2 και Thr 3 της A αλυσίδας, της B αλυσίδας Lys 56 και Thr 57, της C αλυσίδας Arg 161 και Ala 160, της D αλυσίδας Arg215 και Ala 214. Τέλος, εμφανίζονται και αρκετά εύελικτα άτομα στο ενδιάμεσο τμήμα των αλυσίδων με την μεγαλύτερη κινητικότητα να εμφανίζονται στην C αλυσίδα τα άτομα των Asp 135, Glu 132, Glu 128 καταλοίπων όπως επίσης και στα κατάλοιπα Asp30 της A και Thr 71 της B. Συμπεραίνουμε, λοιπόν, από την εικόνα των δομών σε υπέρθεση (4.14), από τη μέση δομή με βάση τους ανισότροπους παράγοντες (4.15) αλλά και από τα rmsf (4.5) ότι υπάρχουν ιδιαίτερα ευκίνητα κατάλοιπα (ουρές) τα οποία μπορούν να επηρεάσουν σημαντικά την αξιοπιστία των αποτελεσμάτων κατά την εξέλιξη των αναλύσεων της προσομοίωσης εξαιτίας της τυχαίας και ακανόνιστης κίνησης τους μιας και δεν συμμετέχουν σε κάποια σταθερή δομή. Συνήθως τα κατάλοιπα αυτά δεν φέρουν κάποια σημαντική πληροφορία όταν μελετάμε τις διαμορφώσεις της μέσης δομής, διότι μας ενδιαφέρει ο σκελετός της δομής του μορίου. Ο υπολογισμός των anisotropic fluctuations ετοιμάστηκε με την εντολή `g_rmsf` από το πρόγραμμα GROMACS (Lindahl et al., 2001).



#### 4.1.11 Principal component analysis.

Ένας ακόμη τρόπος στατιστικής μελέτης και ομαδοποίησης τους δεδομένων είναι το Principal Component analysis δηλαδή η ανάλυση των κύριων συνιστωσών (Amadei et al., 1993; Garcia 1992; Ichiye & Karplus, 1991). Το PCA δίνει την δυνατότητα στον ερευνητή να εξάγει συμπέρασμα μελετώντας μερικές μόνο μεταβλητές αντί του συνόλου και μάλιστα εκείνες που θα έχουν τις πλέον αντιπροσωπευτικές ιδιότητες για τη συγκεκριμένη έρευνα. Είναι σχεδιασμένη έτσι ώστε να μειώνει συστηματικά τον αριθμό των μεταβλητών που πρέπει να εξεταστούν, σε μικρότερο αριθμό «νέων» μεταβλητών που ονομάζονται κύριες συνιστώσες οι οποίες είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών και δεν συσχετίζονται μεταξύ τους (Jolliffe 2002). Η διαδικασία υπολογισμών μπορεί να ερμηνευθεί είτε γεωμετρικά είτε αλγεβρικά. Συνοπτικά δημιουργείται ένας διδιάστατος πίνακας συνδιακύμανσης (ή ο πίνακας συσχέτισης ) (π.χ. Πίνακας variance-covariance) μεταξύ των μεταβλητών από τον οποίο υπολογίζονται οι χαρακτηριστικές τιμές (eigenvalues) και τα αντίστοιχα χαρακτηριστικά ανύσματα (eigenvectors) όπου ικανοποιείται η εξίσωση:

$$Ax = \lambda x$$

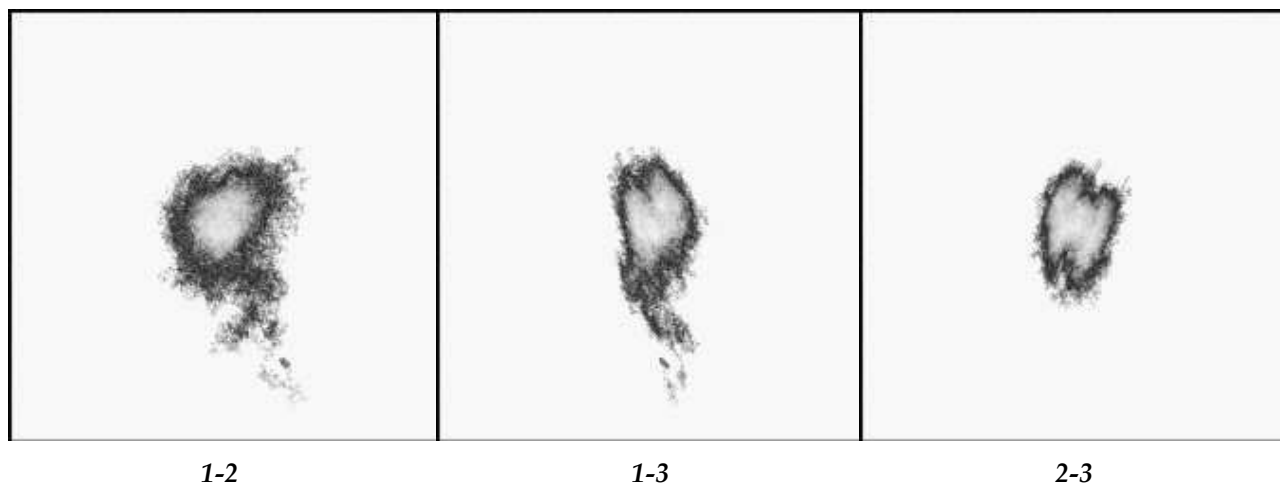
με  $A$  αναπαρίσταται ο διδιάστατος πίνακας,  $\lambda$  είναι η χαρακτηριστική τιμή και  $x$  το χαρακτηριστικό άνυσμα, το οποίο είναι μη μηδενικό και αντιστοιχεί στην χαρακτηριστική τιμή. Οι χαρακτηριστικές τιμές διαταράσσονται με φθίνουσα τάξη και αποτελούν τις διακυμάνσεις των κύριων συνιστωσών (PCA) (Amadei et al., 1993; Cregut et al., 1998). Με κάθε ζεύγος eigenvectors - eigenvalues γίνεται εφικτό να χαρακτηριστούν οι επιμέρους κινήσεις της σύνθετης κίνησης του μορίου. Επίσης μας επιτρέπει να εξάγουμε ποσοτικά συμπεράσματα για το εάν ήταν αρκετή η διάρκεια της προσομοίωσης ώστε να έχουμε ικανοποιητικό δείγμα (sufficient sampling) της κίνησης του μορίου (Grossfield et al., 2007).

Ωστόσο ακόμα και όταν υπάρχει sufficient sampling για το τμήμα της προσομοίωσης που μελετάμε δεν θα πρέπει να αποκλείουμε τυχόν επιπλέον διαμορφώσεις τις οποίες δεν προλάβουμε να δούμε. Ένας τρόπος για να εξετάσουμε το δείγμα μας είναι να πάρουμε την προβολή των ατομικών διακυμάνσεων για συγκεκριμένα χαρακτηριστικά ανύσματα (eigenvectors). Έτσι παράγονται  $\Delta G$  γραφήματα (energy landscape) για όλους τους συνδυασμούς των επιλεγμένων eigenvectors, που αναπαριστούν την ελεύθερη

ενέργεια. Τέλος, πραγματοποιεί cluster analysis με βάση τον αριθμό των κύριων συνιστωσών που θα του οριστεί από τον χρήστη.

- **Cartesian PCA**

Το Principal component analysis πραγματοποιείται μέσω του προγράμματος μοριακής δυναμικής carma. Αρχικά, για να πραγματοποιηθεί το Cartesian-PCA, σε αντίθεση με το dPCA πρέπει να αφαιρεθούν όλες οι μετατοπίσεις και οι περιστροφές που συνέβησαν κατά τη διάρκεια των προσομοιώσεων, ώστε να πραγματοποιηθεί η ανάλυση των κύριων συνιστωσών σύμφωνα με το καρτεσιανό σύστημα χωρίς προβλήματα. Το πρόγραμμα εκτελεί cluster analysis στον χώρο όπου είναι καθορισμένο για τους 3 eigenvectors που αντιστοιχούν στις 3 μεγαλύτερες eigenvalues και υπολόγισε 1 cluster, όπως και το στατιστικό πρόγραμμα R, το οποίο αποτελείται από 51753 frames από τα συνολικά του trajectory. Τα ΔG plots (εικόνα 4.16) που σχεδίασε δηλώνουν επιπλέον, ότι υπήρξε sufficient sampling, δηλαδή ήταν αρκετός ο χρόνος της προσομοίωσης για να δούμε ικανοποιητικά τις χαρακτηριστικές κινήσεις του μορίου.



*Εικόνα 4.16* Γραφήματα της ελεύθερης ενέργειας για τις κύριες συνιστώσες 1, 2, 3, οι οποίες καθορίζονται από τα χαρακτηριστικά ανόσματα (eigenvectors) σε ζευγάρια 1-2, 1-3, 2-3 σε κλίμακα -61,31 μέχρι +61,31 με ενεργειακό ελάχιστο ΔG 2.13 kcal/mol.

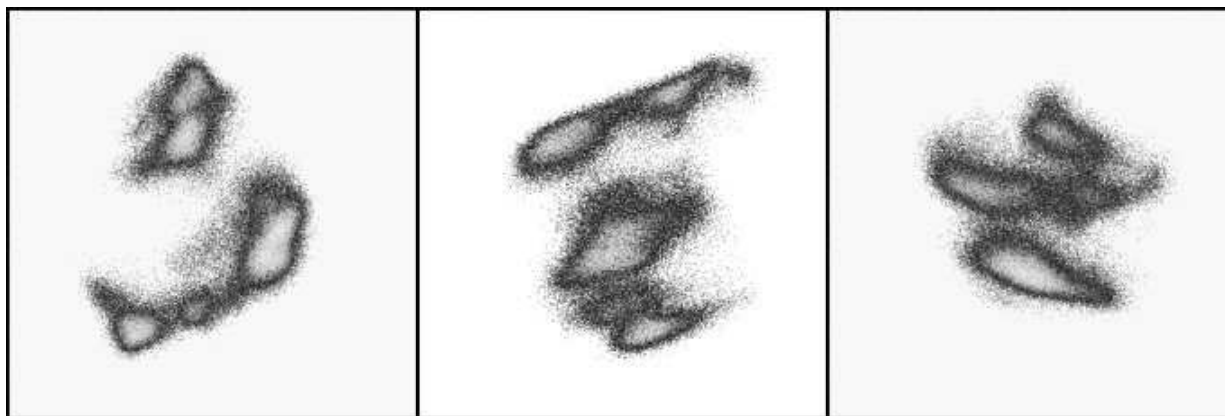
- **Dihedral PCA**

Η ανάλυση των κύριων συνιστωσών (PCA) μπορεί επίσης να πραγματοποιηθεί όπως πρόσφατα πρότειναν οι Mu et al., 2005, χρησιμοποιώντας τις διεδρες γωνίες του σκελετού της πρωτεΐνης αντί των καρτεσιανών συντεταγμένων. Οι διεδρές γωνίες πλεονεκτούν επειδή ως εσωτερικές συντεταγμένες που είναι φυσιολογικά παρέχουν σωστό

διαχωρισμό των εσωτερικών και ολικών κινήσεων του μορίου, κάτι το οποίο είναι θεμελιώδες για την κατασκευή και την ερμηνεία των  $\Delta G$  γραφημάτων ελεύθερης ενέργειας των βιομορίων όταν αυτά υφίστανται μεγάλες δομικές διαμορφώσεις (Altis et al., 2007). Με το PCA-dihedral, λοιπόν, πραγματοποιείται ανάλυση κύριων συνιστωσών σύμφωνα με τις διεδρες γωνίες ( $\phi, \psi$ ) της πρωτεΐνης κατά την διάρκεια του trajectory. Το πρόγραμμα χρησιμοποίησε για cluster analysis τους 3 πρώτους eigenvectors που αντιστοιχούν στις 3 μεγαλύτερες eigenvalues και υπολόγισε 5 clusters τα οποία αποτελούνται από:

1o cluster -->	5562 frames με ποσοστό	occupancy	17,7%
2o cluster -->	15415 frames με ποσοστό	"	49.1%
3o cluster -->	7380 frames με ποσοστό	"	23.5%
4o cluster -->	2544 frames με ποσοστό	"	8.1%
5o cluster -->	514 frames με ποσοστό	"	1.6%

Ως κυρίαρχο cluster θεωρούμε το cluster με τα περισσότερα frames δηλαδή το 2, το οποίο καταλαμβάνει το 49.1% των frames που ανήκουν σε cluster.



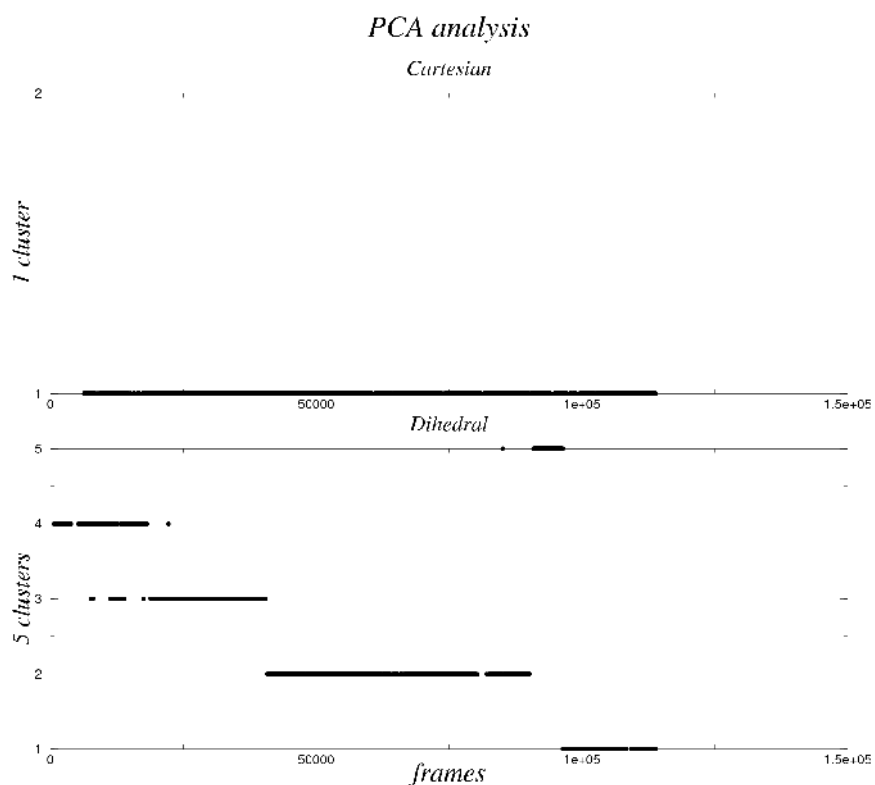
1-2

1-3

2-3

**Εικόνα 4.17** Τα γραφήματα ελεύθερης ενέργειας δείχνουν τα 5 cluster όμως δεν είναι ευδιάκριτα γι' αυτό υποψιαζόμαστε ότι δεν έχουμε sufficient sampling (κλίμακα σχεδίασης -3,50 μέχρι +3,50 και ενεργειακό ελάχιστο kcal/mol).

Το γεγονός ότι οι δύο τρόποι ανάλυσης κύριων συνιστωσών Cartesian και Dihedral έδωσαν διαφορετικό αριθμό από cluster οφείλεται στα διαφορετικά στοιχεία τα οποία χρησιμοποιούν στους υπολογισμούς τους. Ωστόσο εάν ερευνήσουμε το περιεχόμενο του κάθε cluster μέσω της εικόνας 4.18 συμπεραίνουμε ότι το 1 cluster του Cartesian, το dihedral το αναγνωρίζει ως 5 διαφορετικά, τα οποία όμως περιέχουν συνολικά τον ίδιο αριθμό frames.

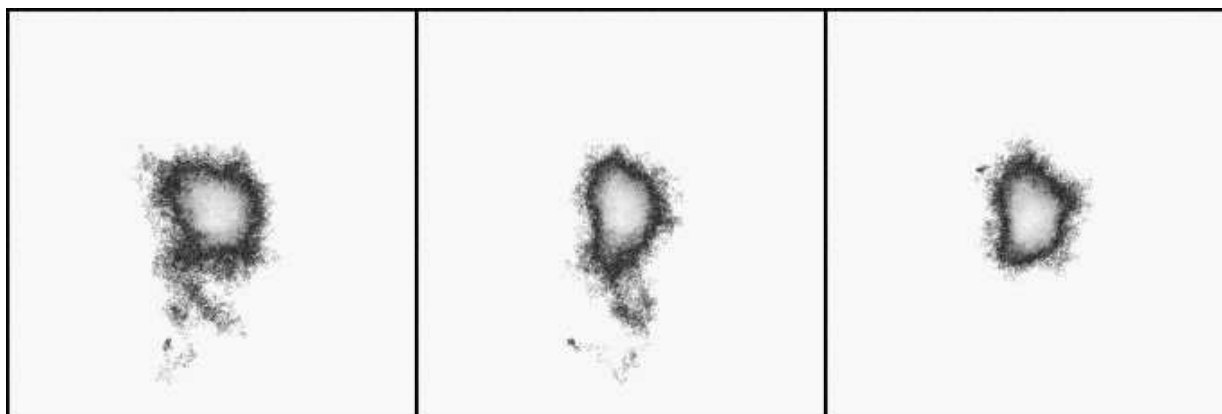


**Εικόνα 4.18** Βλέπουμε την κατανομή των *frames* του κάθε *cluster* μέσω *cartesian* και *dihedral* PCA ανάλυση.

Ένας λόγος ίσως που το *dihedral* υπολογίζει 5 *cluster* και όχι 1 όπως το *cartesian* μπορεί να είναι η ύπαρξη των ευκίνητων κατάλοιπων στα άκρα της δομής. Γι' αυτό θα υπολογίσουμε PCA ξανά με περισσότερη έμφαση στα κύρια τμήματα της δομής του μορίου χωρίς τα υπερκινητικά κατάλοιπα, των οποίων η συμπεριφορά έχει μελετηθεί λεπτομερώς (παράγραφοι 4.1.9, 4.1.10), ώστε να δούμε τις διαμορφώσεις των δομών κατά τη διάρκεια του *trajectory*. Οπότε αποκόβουμε τα κατάλοιπα εκείνα που έχουν *rmsf* μεγαλύτερο από 1,5 Å (20 κατάλοιπα, περίπου δύο κατάλοιπα από κάθε άκρο), διότι αλλοιώνουν τα αποτελέσματα.

- **Cartesian PCA (χωρίς υπερκινητικά κατάλοιπα)**

Έπειτα από δοκιμές κρίναμε ότι θα πρέπει να εξαιρέσουμε από τις μελέτες μας τα κατάλοιπα με *rmsf* άνω των 1,5 Å γιατί αλλιώς τα αποτελέσματα μας δεν είναι ξεκάθαρα. Το PCA, έπειτα από την αφαίρεση αυτών των 20 υπερκινητικών καταλοίπων, υπολόγισε 2 *cluster* από τα οποία το κυρίαρχο είχε 45303 *frames* και κάλυπτε το 99,93% (*occupancy*) των συνολικών *frames* που ανήκουν σε *cluster*, και το δεύτερο είχε μόλις 33 *frames* με 0.07%. Στα διαγράμματα ΔG plots φαίνεται σχετικά ευδιάκριτα ότι υπάρχει ένα κυρίαρχο *cluster* ικανοποιητικό *sufficient sampling*.



1-2

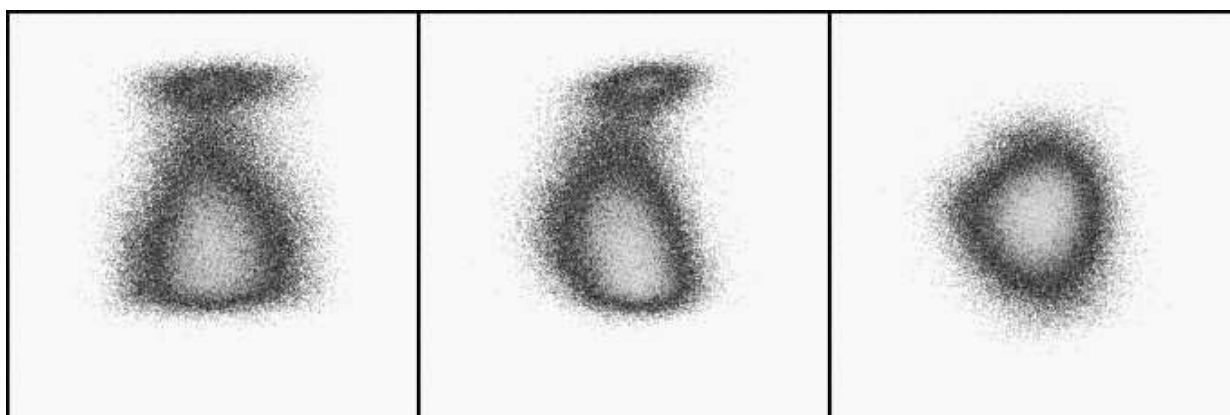
1-3

2-3

**Εικόνα 4.19** Γραφήματα της ελεύθερης ενέργειας για τις κύριες συνιστώσες 1, 2, 3, οι οποίες καθορίζονται από τα χαρακτηριστικά ανύσματα (eigenvectors) σε ζευγάρια 1-2, 1-3, 2-3 με κλίμακα -52.88 μέχρι +52.88 με ενεργειακό ελάχιστο 2.82 kcal/mol.

- **Dihedral PCA (χωρίς υπερκινητικά κατάλοιπα)**

Το Dihedral PCA έδειξε ότι υπάρχουν 11 cluster από τα οποία το κυρίαρχο είναι το πρώτο και αυτό με το οποίο θα ασχοληθούμε, αποτελείται από 15613 και έχει occupancy 99,6%, ενώ τα υπόλοιπα 10 cluster αποτελούνται από πολύ λίγα frames το καθένα, συνολικά είχαν μόλις 62 το οποίο αποτελεί το 0.4% του συνολικού αριθμού. Οπότε είναι φανερό το 1 και κυρίαρχο cluster. Επίσης είναι φανερό ότι τα κυρίαρχα cluster από τις 2 μεθόδους μοιάζουν παρά πολύ διότι ερευνώντας στο γράφημα της εικόνας 4.21 συμπεραίνουμε ότι αποτελούνται από παρόμοια frames.

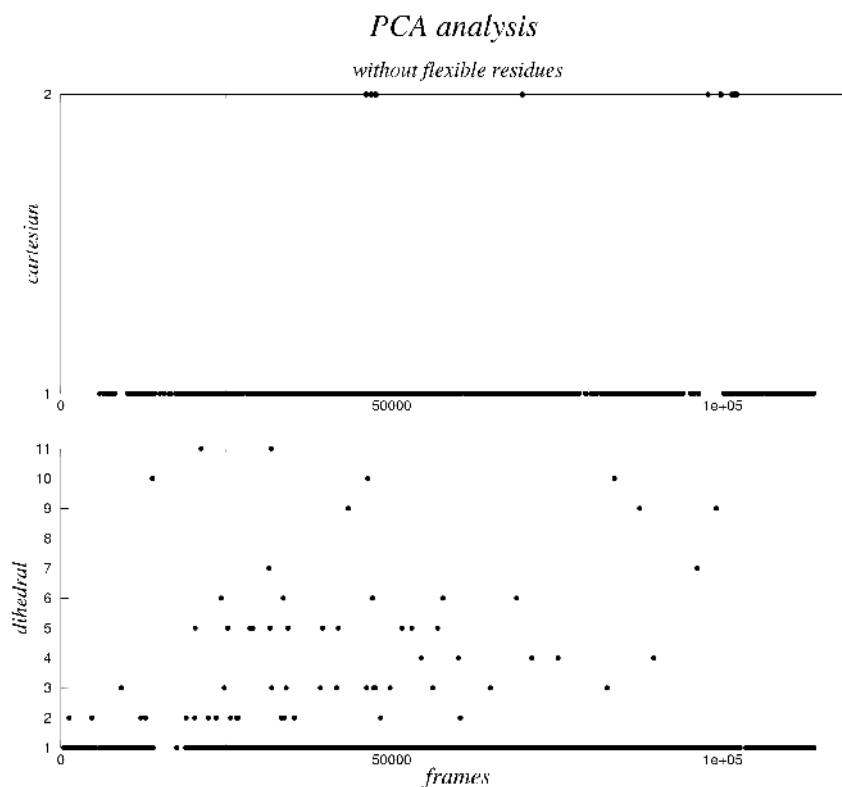


1-2

1-3

2-3

**Εικόνα 4.20** Τα γραφήματα ελεύθερης ενέργειας από dPCA δείχνουν να υπάρχει 1 κυρίαρχο cluster χωρίς sufficient sampling. (-2,60 μέχρι +2,60 με ενεργειακό ελάχιστο 2.11 kcal/mol )

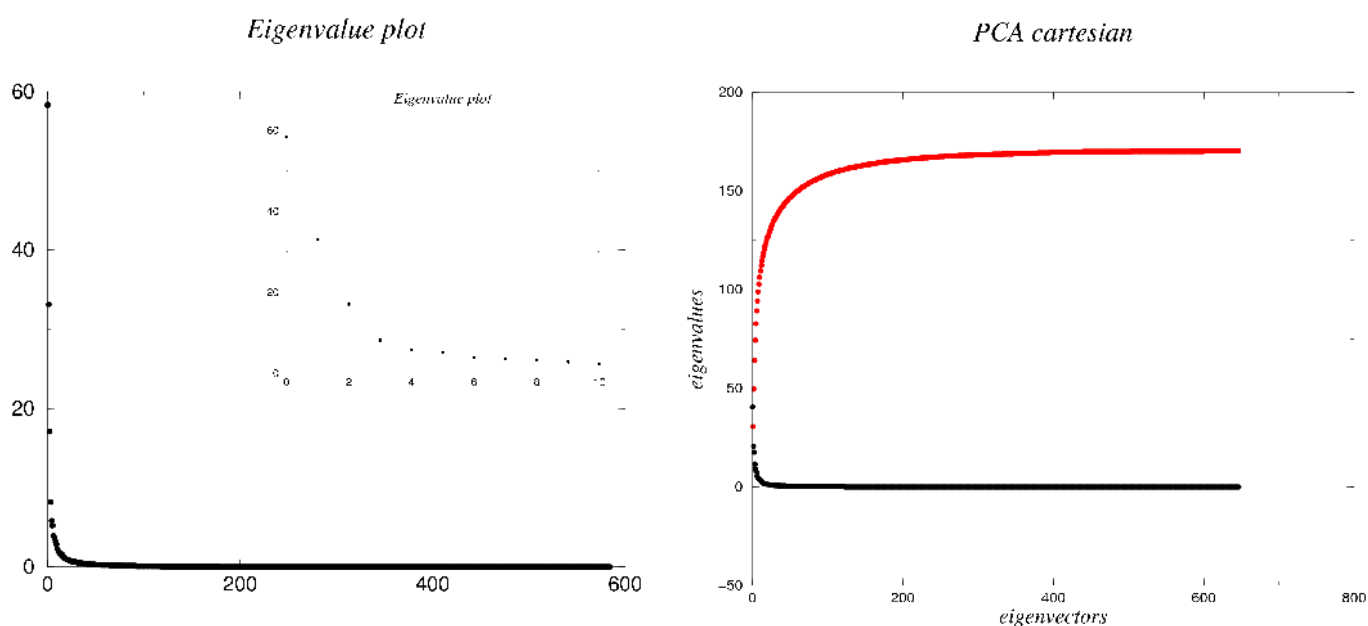


Εικόνα 4.21 Απεικόνιση των *frames* που αντιστοιχούν σε κάθε *cluster*, 2 για Cartesian και 11 για το *dihedral* PCA.

Συγκεντρωτικά, λοιπόν, παρατηρούμε ότι μετά την αφαίρεση των 20 υπερκινητικών καταλοίπων και οι δύο μέθοδοι έδειξαν την ύπαρξη ενός μόνο κυρίαρχου *cluster*, δηλαδή υπάρχει συμφωνία και με τα αποτελέσματα από την εικόνα 4.11, το cartesian-PCA έδειξε 2 *cluster* με ένα κυρίαρχο, ενώ το dPCA 11 *cluster* όπου ως κυρίαρχο *cluster* θεωρείται το πρώτο το οποίο φαίνεται να περιέχει μεγάλο αριθμό από ίδια *frames* με το κυρίαρχο cartesian *cluster*. Από εδώ και στο εξής όταν θα αναφερόμαστε στο κυρίαρχο *cluster* θα εννοούμε το κυρίαρχο *cluster* από το cartesian-PCA είτε συμπεριλαμβάνονται τα υπερκινητικά κατάλοιπα είτε όχι ανάλογα με την ανάλυση.

- **Eigenvectors και Eigenvalues**

Συνεχίζοντας την ανάλυση των κύριων συνιστωσών αφού υπολογίσαμε τον αριθμό των cluster και καθορίσαμε το κυρίαρχο, υπολογίζουμε τις χαρακτηριστικές τιμές (eigenvalues) και τα αντίστοιχα χαρακτηριστικά ανύσματα (eigenvectors) των κύριων συνιστωσών, χρησιμοποιώντας όλα τα κατάλοιπα ακόμα και τα υπερκίνητικα. Όπως βλέπουμε και στην εικόνα 4.22, οι πρώτες κύριες συνιστώσες εξηγούν το μεγαλύτερο κομμάτι της διακύμανσης των eigenvectors οι οποίοι μας πληροφορούν για το σύνολο των κινήσεων της δομής του δεματίου κατά τη διάρκεια του trajectory (Amadei et al., 2003).



*Εικόνα 4.22 Αριστερά: Απεικόνιση των eigenvectors και σε μεγέθυνση φαίνονται οι 10 πρώτοι αυτών.*

*Δεξιά: Απεικόνιση των eigenvectors (μαύρη γραμμή) και το ολοκλήρωμα των κύριων συνιστωσών (κόκκινη γραμμή)*

Πιο συγκεκριμένα οι μεγάλες κινήσεις του μορίου φαίνονται στους πρώτους eigenvectors, ενώ προχωρώντας οι υπόλοιποι ασχολούνται με τις μικρότερες κινήσεις του. Στο διάγραμμα 4.22 είναι φανερό για την δομή μας ότι από τους πρώτους 5 eigenvectors περιγράφεται το 54.7% της συνολικής κίνησης του μορίου υπολογίζοντας όλα τα άτομα.

Ακολούθως, θα ελέγξουμε εάν υπάρχει sufficient sampling, γι' αυτό θα υπολογίσουμε τις διακυμάνσεις των 5 πρώτων eigenvector οι οποίοι όπως φαίνεται και στο γράφημα (εικόνα 4.23) είναι αντιπροσωπευτικοί των κινήσεων του μορίου. Ο έλεγχος θα γίνει με βάση την κατανομή των 5 πρώτων eigenvector στο επίπεδο, σε δύο διαστάσεις και σε τρεις διαστάσεις. Στην περίπτωση όπου υπάρχει sufficient sampling τα ιστογράμματα που θα προκύψουν θα έχουν κανονική κατανομή γύρω από το μηδέν, όπως

επίσης και οι διακυμάνσεις των eigenvectors για δυο και τρεις διαστάσεις.

Παρατηρούμε στην εικόνα 4.23α τα ιστογράμματα των 5 πρώτων eigenvector από τους οποίους ο 1<sup>ος</sup> έχει σχεδόν κανονική κατανομή με μια πολύ μικρή κορυφή στα δεξιά του, το μικρό cluster που αντιστοιχεί στο στάδιο του heating. Φαίνεται πως ο χρόνος ήταν σχεδόν αρκετός για να έχουμε ικανοποιητικό δείγμα της κίνησης του μορίου υπάρχει δηλαδή αρκετά ικανοποιητικό sufficient sampling, ενώ όσο προχωράμε στους επόμενους eigenvectors, αυτοί εμφανίζουν σαφή αποτελέσματα με κανονική κατανομή και ξεκάθαρο sufficient sampling ο καθένας. Οπότε αρκεί μονάχα η μελέτη των 3 πρώτων eigenvector για να πληροφορηθούμε με αξιοπιστία για την κίνηση του μορίου.

Όσο αφορά τις διακυμάνσεις των 3 πρώτων eigenvector σε 2 διαστάσεις (εικόνα 4.23β) όπου απεικονίζονται ανά ζεύγη οι eigenvectors, το ζεύγος 1-2 δηλώνει, ότι υπάρχει ένα κυρίαρχο cluster αλλά, δεν είναι καλοσχηματισμένο, άρα το sufficient sampling είναι σχεδόν ικανοποιητικό και επίσης δημιουργείται και ένα δεύτερο μικρότερο (το heating). Τέλος, η απεικόνιση των διακυμάνσεων και των τριών eigenvector σε τρεις διαστάσεις σε στέρεο εικόνα (εικόνα 4.23γ) επιβεβαιώνουν τις προηγούμενες παρατηρήσεις.

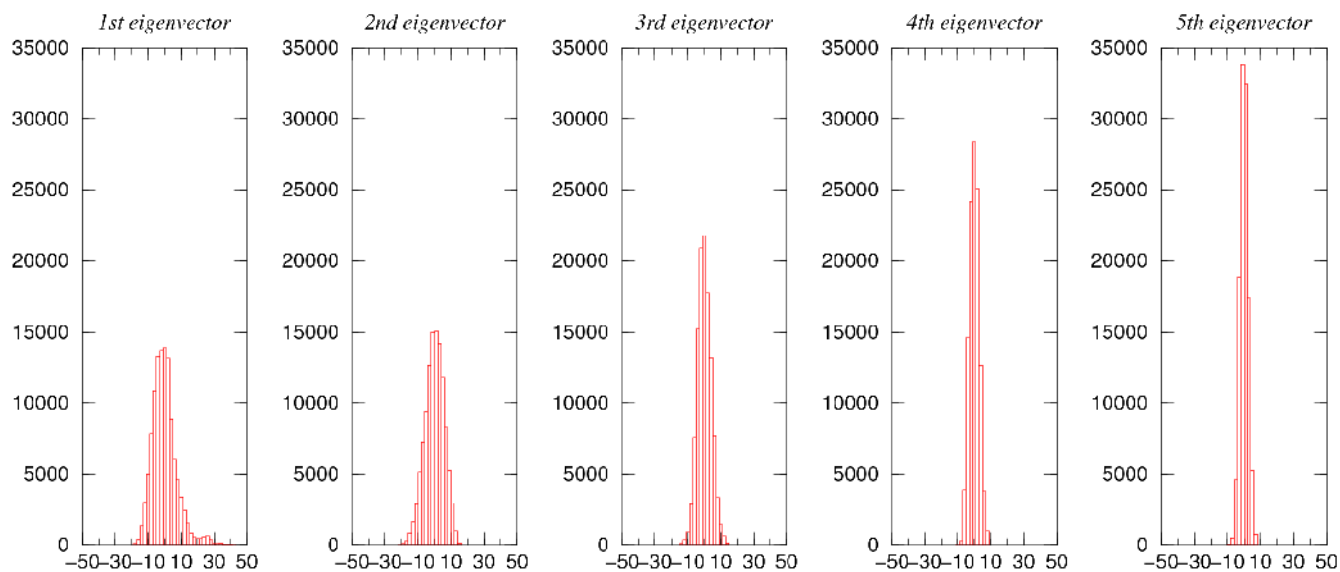
Από την αρχή των προσομοιώσεων υπήρχε το ερώτημα εάν είναι επαρκείς ο χρόνος στον οποίο εκτελούνται, ώστε να είναι αρκετά τα δεδομένα για την εξαγωγή εύλογων συμπερασμάτων. Οι αναλύσεις PCA παρέχουν πληροφορίες για τις επιμέρους κινήσεις των μορίων και επίσης, είναι ένα καλό μέτρο για να συμπεράνουμε εάν ο χρόνος της προσομοίωσης ήταν αρκετός ώστε να είναι ικανοποιητικό το δείγμα που μελετάμε. Κρίναμε σκόπιμο να συγκεντρώσουμε τα ΔG γραφήματα των PCA αναλύσεων (εικόνα 4.24) ώστε να δούμε συνολικά σε ποια σημεία είχαμε sufficient sampling.

Μελετάμε διαδοχικά τα γραφήματα των PCA αναλύσεων καθώς περιορίζουμε και συγκεκριμενοποιούμε το επίπεδο των υπολογισμών μας με την αφαίρεση των υπερκινητικών καταλοίπων και τον υπολογισμό PCA που αφορά τα κυρίαρχα Cluster. Με αυτόν τον τρόπο, διακρίνουμε πιο ξεκάθαρα στοιχεία για την ύπαρξη sufficient sampling, ενώ στον αρχικό υπολογισμό ολόκληρου του trajectory τα δεδομένα ήταν θολά.

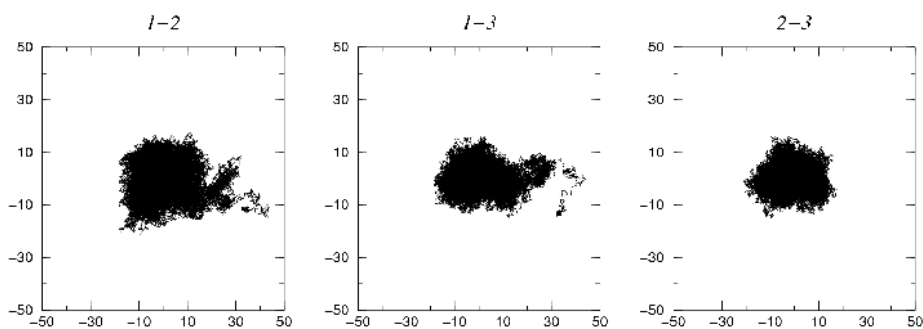


**Εικόνα 4.23**

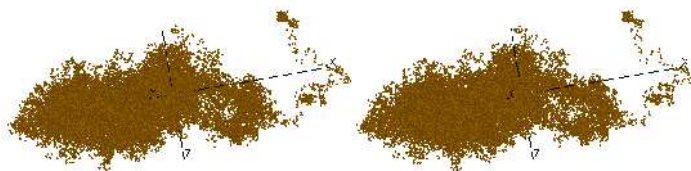
A) Προβολή των διακομάνσεων των 5 πρώτων eigenvectors στη μια διάσταση

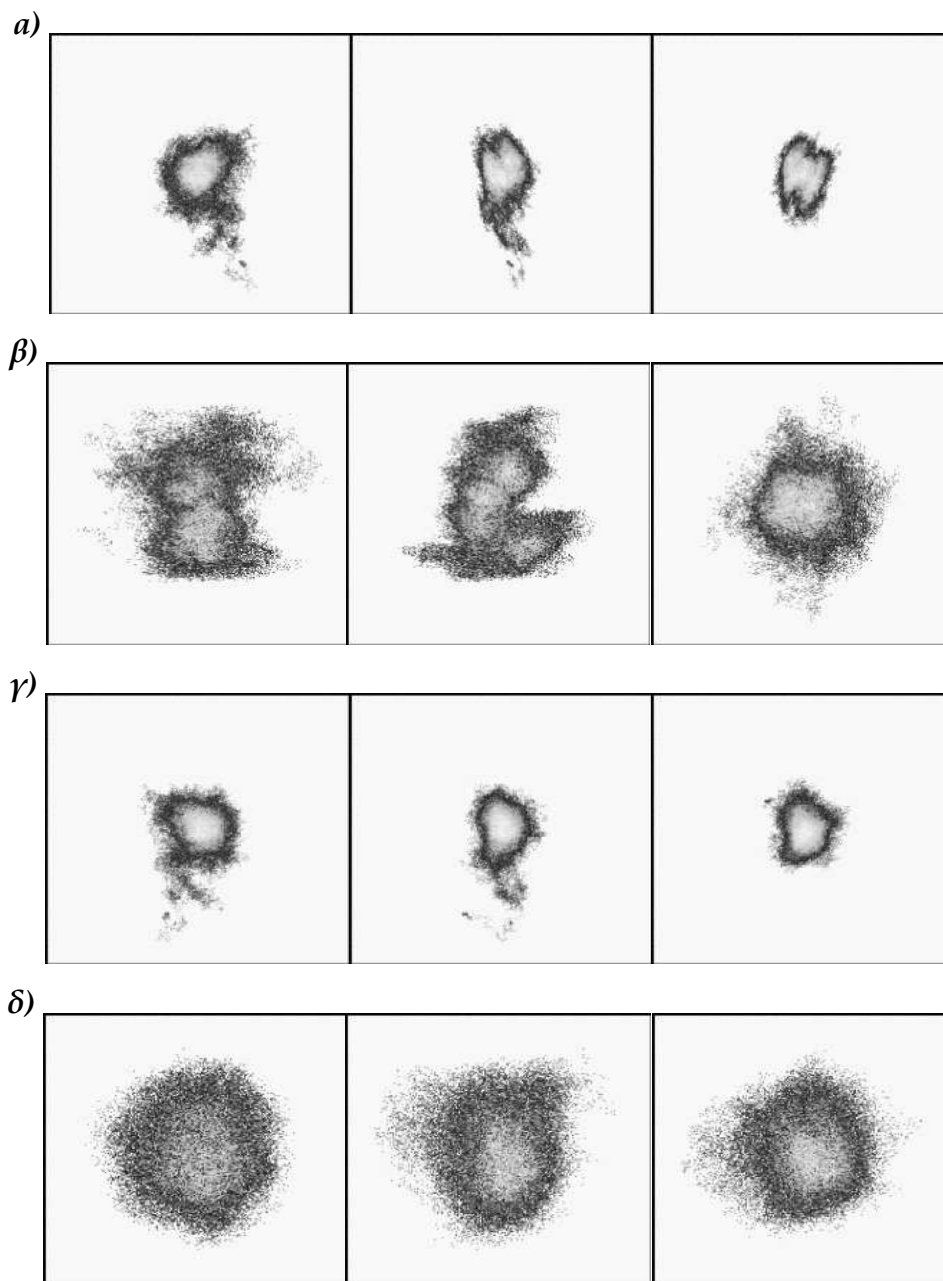


B) Προβολή των διακομάνσεων των eigenvectors ανά ζεύγη



Γ) Προβολή των διακομάνσεων και των τριών eigenvector μαζί (stereo)

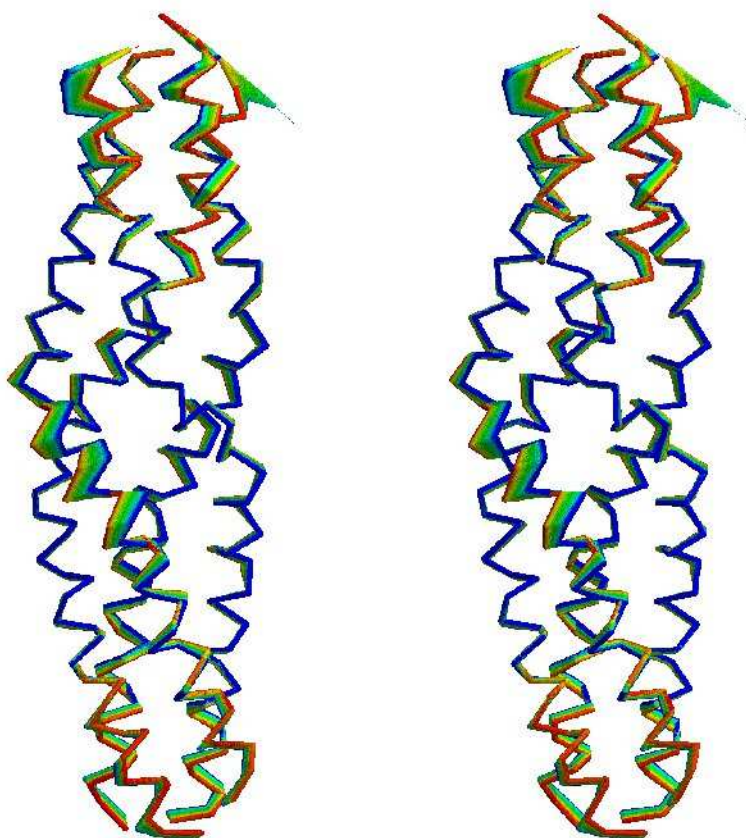




*Εικόνα 4.24* ΔG γραφήματα που προκύπτουν από PCA ανάλυση α) όλου του trajectory και β) του αντίστοιχου κυρίαρχου cluster, γ) όλου του trajectory αφαιρώντας τα 20 υπερκινητικά κατάλοιπα και του αντίστοιχου κυρίαρχου cluster του.

- Προσδιορισμός της συνολικής κίνησης του μορίου

Στην συνέχεια θα δείξουμε την κίνηση που πραγματοποιεί το μόριο μας με βάση τον πρώτο eigenvector όμως επειδή δεν έχει επιτευχθεί sufficient sampling για ολόκληρο το trajectory η μελέτη θα γίνει σε ένα τμήμα αυτού. Έγινε εκ νέου, λοιπόν, pca υπολογισμός για το κυρίαρχο cluster, το οποίο έχει sufficient sampling, χρησιμοποιώντας όλα τα άτομα (εικόνα 4.26) και έπειτα χωρίς τα υπερκινητικά κατάλοιπα (εικόνα 4.27), ώστε να φανεί η κίνηση του μορίου χωρίς να επηρεάζεται από αυτά.

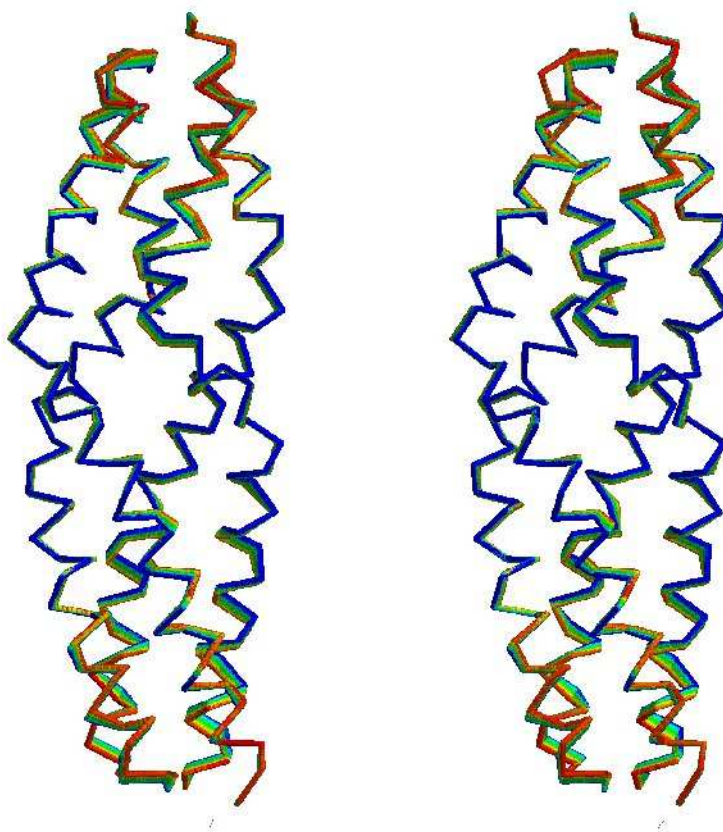


*Εικόνα 4.26 Αναπαράσταση της κίνησης του μορίου συμφώνα με τον πρώτο eigenvector σε stereo μορφή. Η μεγαλύτερη κίνηση παρατηρείται στα άκρα των ελίκων καθώς επίσης και στο κεντρικό τμήμα της C αλυσίδας το οποίο έχει αναλυθεί σε προηγούμενες μελέτες.*

Παρατηρούμε ότι σε όλο το μήκος τους οι αλυσίδες εμφανίζουν μια ελαφριά κινητικότητα με ιδιαίτερη έμφαση στα κατάλοιπα των άκρων καθώς επίσης και στα κατάλοιπα στο εσωτερικό τμήμα κυρίως της αλυσίδας C, όπως παρατηρήσαμε και στις προηγούμενες αναλύσεις μας. Διακρίνουμε το κεντρικό τμήμα του δεματίου να είναι αρκετά σταθερό (μπλε χρώμα) ενώ τα άκρα του δεματίου να εμφανίζουν μεγάλο εύρος στην κίνηση των ατόμων τους. Τα αποτελέσματα αυτά έρχονται σε συμφωνία με τα αποτελέσματα του

πίνακα variance-covariance (εικόνα 4.10) ο οποίος επιπλέον έδειξε ότι η υψηλή κίνηση των άκρων είναι αντίθετη από την κίνηση του μεσαίου τμήματος των αλυσίδων. Το εύρος της κίνησης δηλώνεται από την χρωματική κλίμακα μπλε σε κόκκινο (μεγάλη κίνηση).

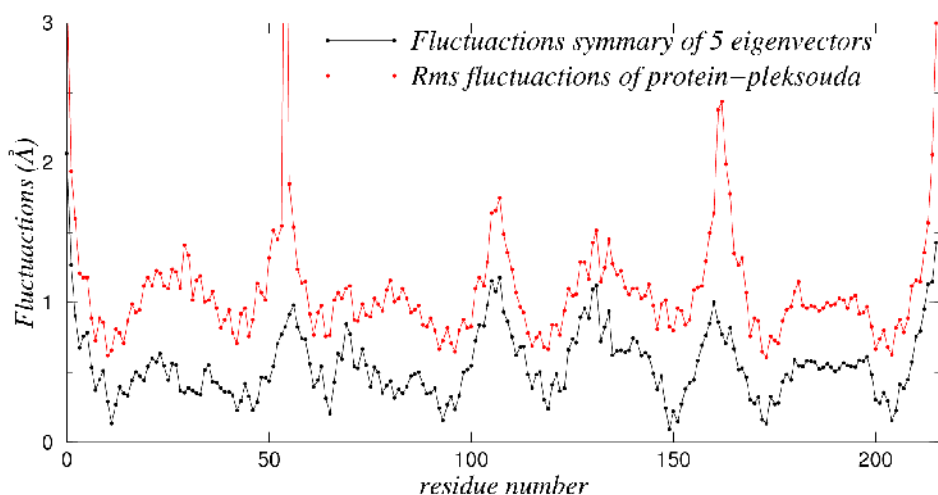
Προφανώς η υψηλή κινητικότητα των καταλοίπων αυτών να επηρεάζει και να καλύπτει την κινητικότητα των υπολοίπων. Με την αφαίρεση τους, μας δίνεται η δυνατότητα να εξετάσουμε την κίνηση του υπόλοιπου μορίου. Η εικόνα 4.27 δείχνει ότι μετά την αφαίρεση των υπερκινητικών καταλοίπων στα άκρα, η κίνηση των γειτονικών τους καταλοίπων είναι πιο ευδιάκριτη και οι αλυσίδες στο κεντρικό τους τμήμα εμφανίζουν μειωμένη κινητικότητα. Ωστόσο, φαίνεται να έχουν απομακρυνθεί η μια από την άλλη διότι το δεμάτιο εμφανίζει μια πάχυνση.



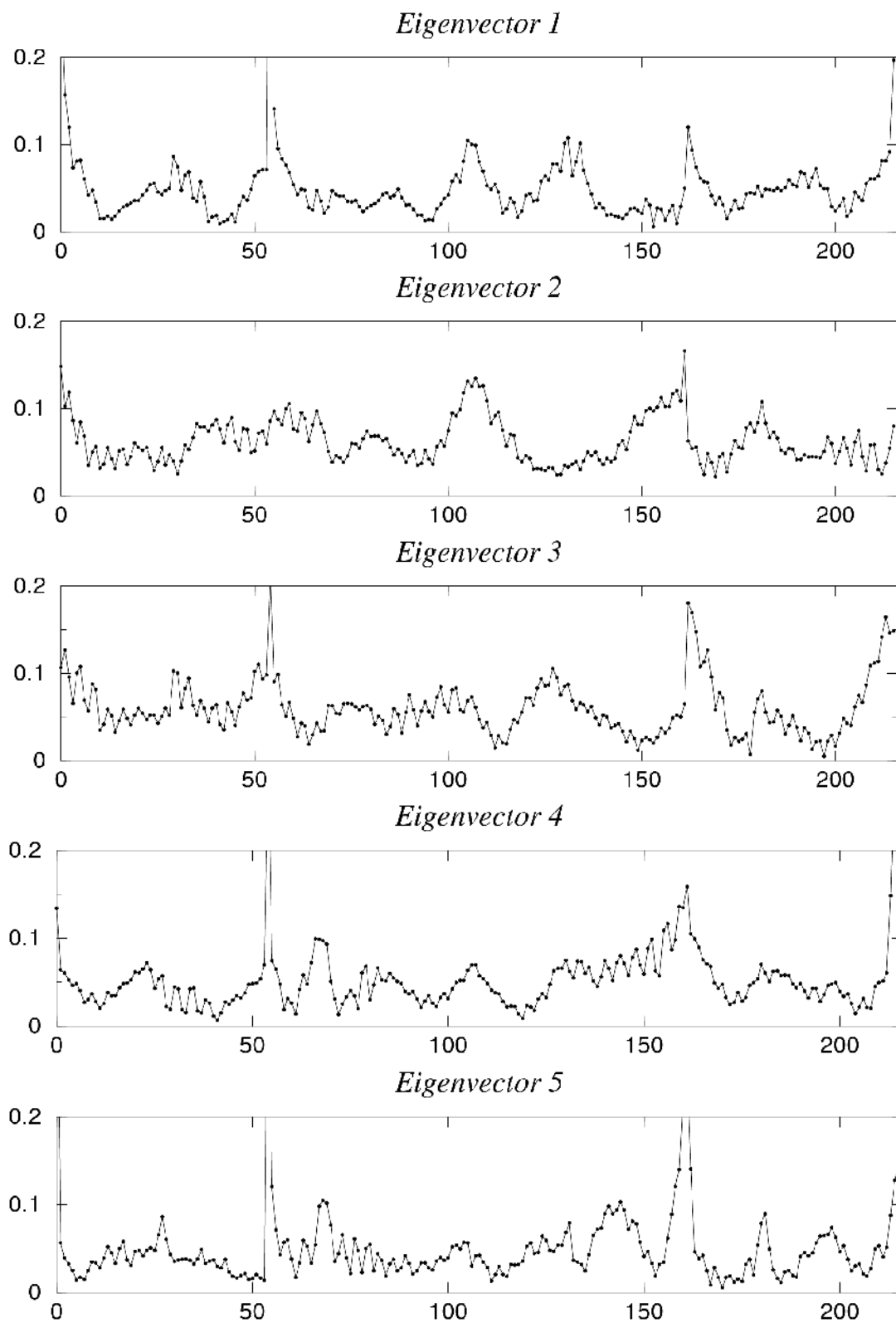
*Εικόνα 4.27 Απεικόνιση της κίνησης του μορίου στο κορίαρχο Cluster σε stereo χωρίς τα υπερκινητικά κατάλοιπα.*

Τέλος, μπορούμε να αναπαραστήσουμε την κατανομή των διακυμάνσεων του κάθε eigenvector ανά κατάλοιπο σε ένα γράφημα για τον καθένα ξεχωριστά (εικόνα 4.29) βλέποντας ποια κατάλοιπα συμμετέχουν σε κάθε κίνηση αλλά και στο σύνολό τους (εικόνα 4.28) παρατηρώντας αθροιστικά την διακύμανση των eigenvectors για κάθε κατάλοιπο. Ταυτόχρονα, χρησιμοποιούμε και τις ατομικές κινητικότητες-rmsf των καταλοίπων, ώστε να δούμε εάν συμφωνούν. Παρατηρούμε ότι οι διακυμάνσεις των eigenvector συμφωνούν

με τις ατομικές κινητικότητες του τετραμερούς σε μεγάλο βαθμό. Συγκρίνοντας τις δυο καμπύλες ο βαθμός γραμμικής συσχέτισης τους είναι  $0.7\text{\AA}$ , ενώ όταν απαλείψουμε τα κατάλοιπα που έχουν rmsf μεγαλύτερο από  $2\text{\AA}$  (cutoff= $2\text{\AA}$ , 6 κατάλοιπα) ο βαθμός συσχέτισης τους γίνεται  $0.78\text{\AA}$  (μεγαλώνει) ο οποίος είναι αρκετά ισχυρός (αφού η τιμές του είναι από -1 μέχρι +1) όμως για cutoff τα  $1.5\text{\AA}$  γίνεται  $0.73$  (20 κατάλοιπα). Όπως παρατηρήσαμε ήδη η μεγαλύτερη κινητικότητα εμφανίζεται στα άκρα κάθε μονομερούς αλλά και στο κεντρικό τμήμα των αλυσίδων. Η κατανομή των διακυμάνσεων ανά κατάλοιπο για τον κάθε eigenvector αναπαρίσταται στην εικόνα 4.29. Για την κατασκευή των γραφημάτων που αφορούν τις διακυμάνσεις χρησιμοποιήθηκαν τα script στο παράρτημα 4.2 και 4.3.



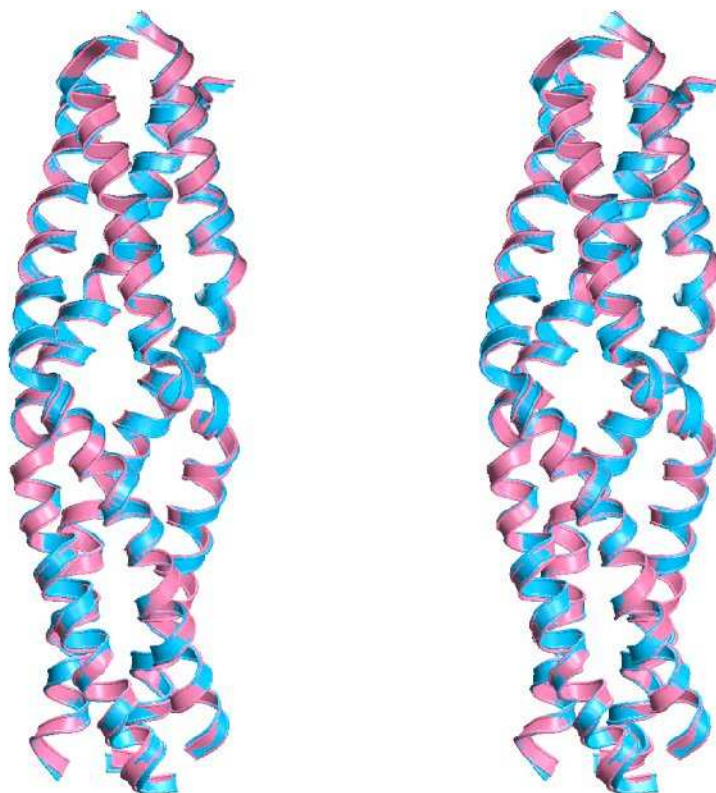
**Εικόνα 4.28** Γραφική αναπαράσταση των διακυμάνσεων των 5 πρώτων eigenvector αθροιστικά ανά κατάλοιπο (μαύρη γραμμή) και τα rmsf κόκκινη γραμμή. Για να είναι τα δεδομένα μας συγκρίσιμα και να μην χάνεται πληροφορία από τις διακυμάνσεις λόγω κλίμακας, μετατοπίστηκε η καμπύλη των διακυμάνσεων στον άξονα y αφού πολλαπλασιαστική η τιμή της με την σταθερά 4. Τα κατάλοιπα με rmsf μεγαλύτερο του 3 τα οποία δεν φαίνονται στο γράφημα είναι το Glu 1 και Glu 55.



*Εικόνα 4.29 Αναπαράσταση των διακρομήσεων των 5 πρώτων eigenvector ανά κατάλοιπο*

#### 4.1.12 Σύγκριση των μέσων δομών από τα κυρίαρχα cluster (Cartesian και Dihedral)

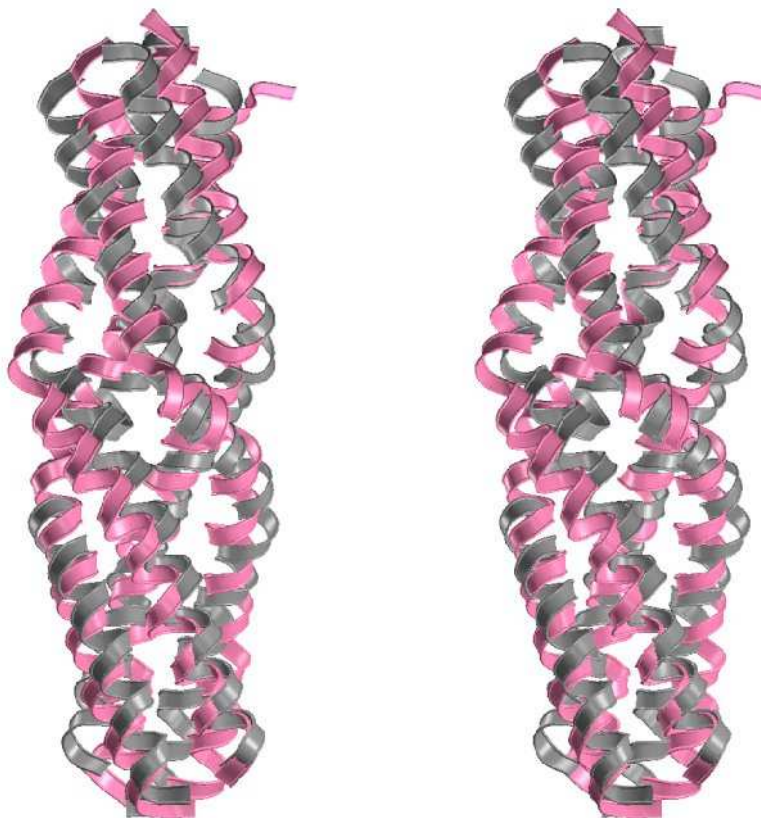
Στην προηγούμενη ενότητα προσδιορίστηκε ο αριθμός των cluster, εδώ θα βρούμε την μέση δομή για το καθένα και θα τις συγκρίνουμε μεταξύ τους αλλά και με την αρχική δομή. Συγκρίνουμε αρχικά της μέσες δομές των κυρίαρχων cluster από το cartesian (χρώμα ροζ) και από το dihedral (μπλε χρώμα) PCA, εξαιρώντας τα υπερκινητικά κατάλοιπα.



*Εικόνα 4.30* Στερεοδιάγραμμα της μέσης δομής του κυρίαρχου cluster από το Cartesian (ροζ χρώμα) και από το dihedral (μπλε χρώμα) σε υπέρθεση. Έχει γίνει στοίχιση των δομών και στα 1-216 κατάλοιπα με απόκλιση Rmsd να είναι 0.48 Å.

Στην εικόνα 4.30 βλέπουμε ότι δεν εμφανίζεται καμία σημαντική διαφορά ανάμεσα στις δύο δομές, σχεδόν επικαλύπτονται σε όλο το μήκος των αλυσίδων. Οι δομές έχουν στοιχηθεί για όλο το μήκος του σκελετού των αλυσίδων σε όλα τα κατάλοιπα της πρωτεΐνης (1-216) και εμφανίζουν τυπική απόκλιση Rmsd μόλις 0.48Å, κάτι το οποίο περιμέναμε αφού τα δύο κυρίαρχα cluster ομοιάζουν σύμφωνα με την εικόνα 4.21 και οι δομές θα είναι αρκετά όμοιες.

Η επόμενη σύγκριση είναι ανάμεσα στην μέση δομή από το κυρίαρχο cluster που έδωσε το Cartesian (ροζ χρώμα) και στην αρχική δομή (γκρι χρώμα) πριν την εξέλιξη των προσομοιώσεων. Στην εικόνα 4.31 είναι εμφανείς οι διαφορές μεταξύ των δύο δομών. Έχουν μετατοπιστεί οι τέσσερις αλυσίδες αρκετά κατά τη διάρκεια του trajectory από τις αρχικές τους θέσεις. Στο κεντρικό τμήμα της αλυσίδας A φαίνεται τα κατάλοιπα να δημιουργούν μια μικρότερη  $\alpha$ -έλικα που φεύγει από την ευθεία της αλυσίδας. Στην C αλυσίδα τα κατάλοιπα στο κέντρο έχουν μετατοπιστεί αρκετά σε σχέση με την αρχική δομή του δεματίου.

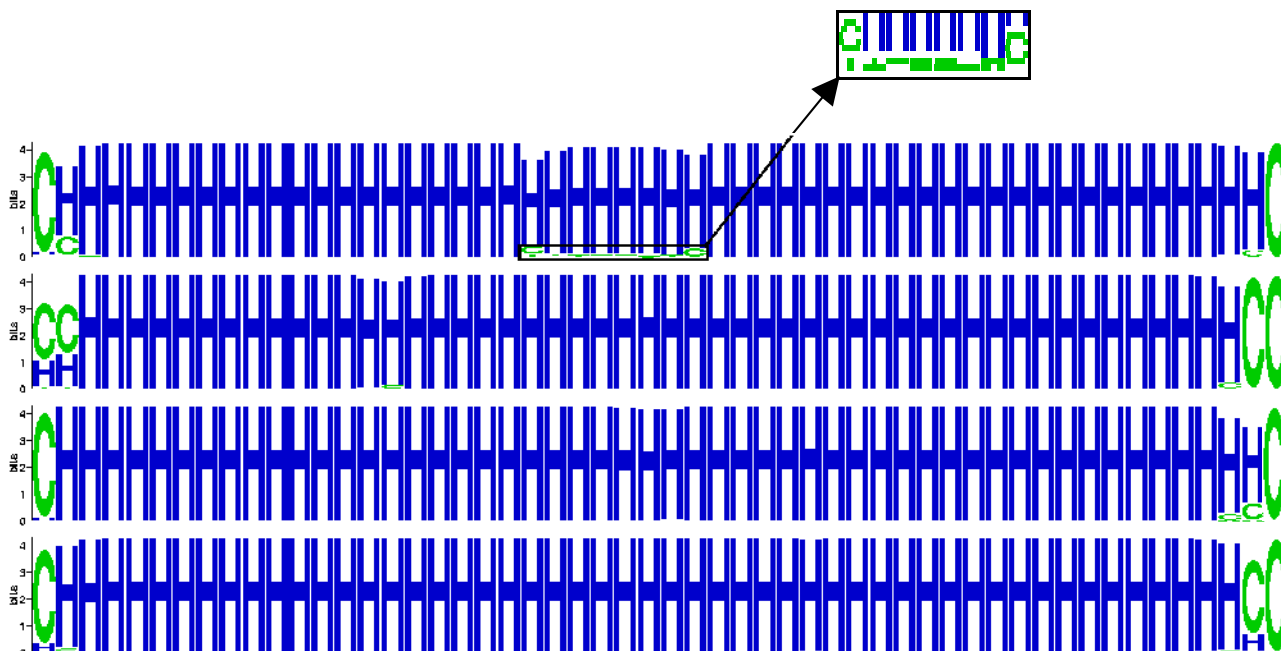


**Εικόνα 4.31** Στερεοδιάγραμμα της μέσης δομής του κυρίαρχου cluster από το Cartesian (ροζ χρώμα) και η αρχική δομή (γκρι χρώμα) σε υπέρθεση. Έχει γίνει στοίχιση των δομών στα κατάλοιπα 1- 216 με απόκλιση *Rmsd* να είναι αρκετά μεγάλη 3.58 Å.



### 4.1.13 Συντήρηση δευτεροταγούς δομής

Με αυτήν την ανάλυση θα ερευνήσουμε την συντήρηση της δευτεροταγούς δομής κατά τη διάρκεια της προσομοίωσης, δηλαδή κατά πόσο και αν διατηρούνται οι α-έλικες ή αλλοιώνονται σε ορισμένα σημεία.



Εικόνα 4.32 Συντήρηση δευτεροταγούς δομής για κάθε θέση της αλληλουχίας

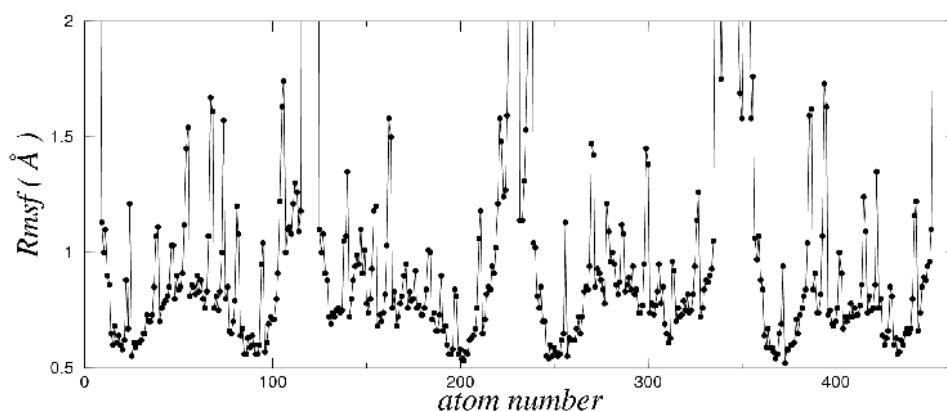
Στο γράφημα (εικόνα 4.32) το οποίο κατασκευάστηκε με το πρόγραμμα weblogo (Shaner et al., 2003) παρουσιάζονται οι έλικες ως H (helix), οι στροφές ως T (turn), ενώ το C (coils) αναφέρεται σε θηλιές. Στον κάθετο άξονα βρίσκονται με τη σειρά τα μονομερή A, B, C, D και σε κάθε σειρά βρίσκονται 54 θέσεις της αλληλουχία τους. Σε κάθε σειρά βρίσκονται 54 θέσεις αλληλουχίας των 4 μονομερών με την σειρά A, B, C, D. Οι α-έλικες διατηρούνται καθ' όλη τη διάρκεια της προσομοίωσης και είναι αρκετά σταθερές. Ωστόσο οι περιοχές στις οποίες σχηματίζονται θηλιές και όχι σταθερή δομή είναι στην A αλυσίδα το κατάλοιπο 1 (Glu) στο N τελικό της άκρο, το κατάλοιπο 54 (Phe) στο C άκρο ενώ στην περιοχή της στροφής παρατηρείται μια μικρή αλλοίωση των ελίκων στα κατάλοιπα 22-29 της A αλυσίδας (εικόνα σε μεγέθυνση). Στη B αλυσίδα χάνεται η ελικοειδής μορφή στα κατάλοιπα 1, 2 του N-τελικού άκρου, 53, 54 του C-τελικού. Όσο αφορά τις αλυσίδες C και D η δευτεροταγής δομής τους διατηρείται σε πολύ μεγάλο βαθμό παρά μόνο στα άκρα τα κατάλοιπα 1 (Glu) 53 και 54 της D η α-έλικα μετατρέπεται σε θηλιά.

#### 4.1.14. Ανάλυση διεδρων γωνιών ( $\chi$ -value)

Προκειμένου η μελέτη και ανάλυση της δομής μας να είναι πιο ολοκληρωμένη θα μελετήσουμε την σταθερότητα και την διαμόρφωση του υδρόφοβου πυρήνα του δεματίου, αναλύοντας την κινητικότητα των πλευρικών ομάδων των καταλοίπων του υδρόφοβου πυρήνα και έπειτα αναλύοντας την κατανομή των rotamers των καταλοίπων αυτών.

Η κινητικότητα των καταλοίπων του υδρόφοβου πυρήνα δίνεται από την παράμετρο rmsf των αντίστοιχων καταλοίπων, η οποία φαίνεται στην εικόνα 4.33. Συγκρίνοντας τις ατομικές κινητικότητες των ατόμων όλων των καταλοίπων (εικόνα 4.5) με αυτών των υδρόφοβων καταλοίπων, παρατηρούμε ότι η κινητικότητα των υδρόφοβων καταλοίπων είναι αρκετά υψηλή. Ιδιαίτερα ευκίνητα φαίνονται τα κατάλοιπα των ατόμων των άκρων, του υδρόφοβου πυρήνα του δεματίου σε σχέση με τα αυτά όλου του σκελετού με μέση τιμή 1.17Å. Όσο αφορά τα κατάλοιπα του κεντρικού τμήματος των μονομερών έχουν επίσης υψηλές τιμές για άτομα που ανήκουν σε πακεταρισμένα κατάλοιπα.

*Rmsf of core residues (heavy atoms)*



Εικόνα 4.33 Παρουσιάζεται στο γράφημα η κινητικότητα των πλευρικών ομάδων των καταλοίπων του υδρόφοβου πυρήνα.

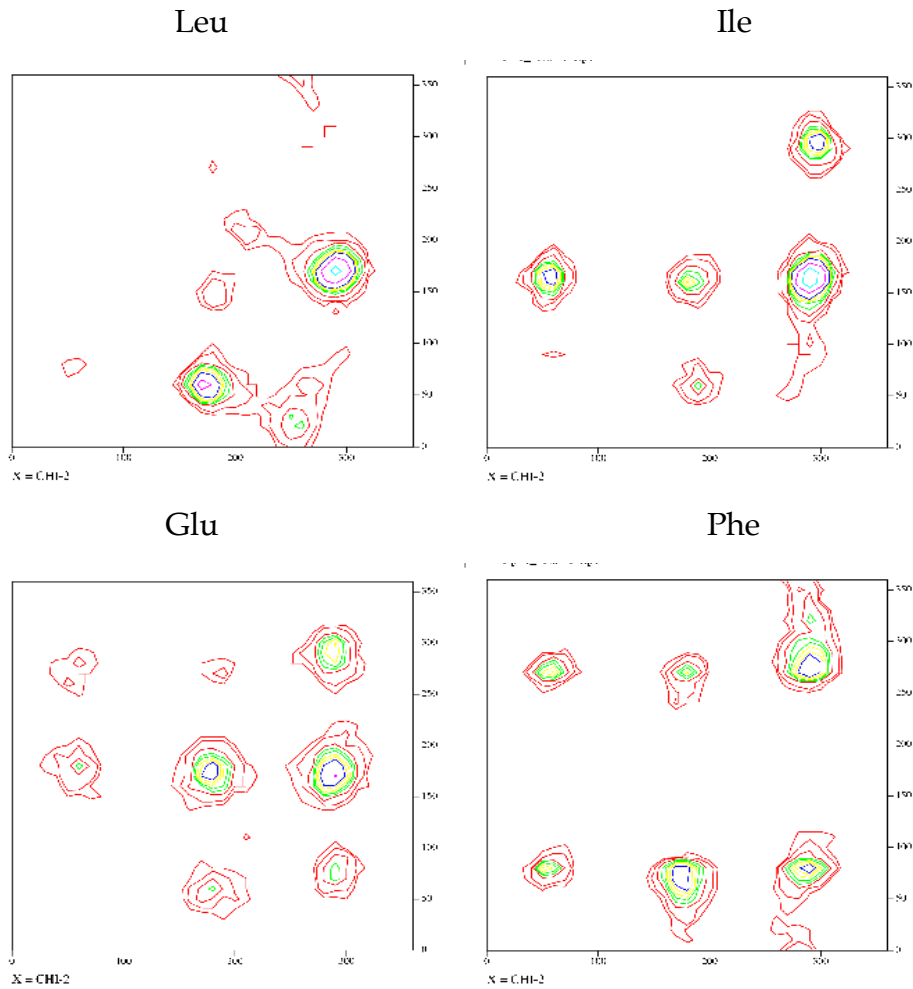
Για την μελέτη της κατανομής των rotamers των καταλοίπων του υδρόφοβου πυρήνα θα χρειαστεί ο υπολογισμός των διεδρων γωνιών των πλευρικών ομάδων. Οι διεδρες γωνίες  $\chi_i$ -διεδρες γωνίες ( $i: 1, 2, 3, 4,$ ) περιγράφουν τις περιστροφές γύρω από τους απλούς δεσμούς των πλευρικών ομάδων. Όμως κατά την περιστροφή δημιουργούνται στερικές παρεμποδίσεις που δεν επιτρέπουν ορισμένες διαμορφώσεις της πλευρικής ομάδας. Οι  $\chi$  γωνίες αντιστοιχούν στα ελάχιστα δυναμικής ενέργειας των αδρανοποιημένων αμινοξέων ενώ η έννοια του rotamer ορίστηκε από τους Ponder &

Richards ως πυκνή συνάθροιση των  $\chi$ -γωνιών (Ponder & Richards, 1987). Θα ασχοληθούμε με την ανάλυση μόνο των  $\chi_1$ ,  $\chi_2$  γωνιών καθώς δεν συνηθίζονται στον υδρόφοβο πυρήνα κατάλοιπα με γωνίες  $\chi_3$ ,  $\chi_4$  στις πλευρικές τους και επιπλέον προσδιορίζονται με μικρή ακρίβεια (άρα μεγάλη αβεβαιότητα) οι θέσεις των ατόμων τους, ενώ η περιστροφή γύρω από τους δεσμούς που αντιστοιχούν στις γωνίες  $\chi_3$ ,  $\chi_4$  παρεμποδίζεται λιγότερο σε σχέση με τις  $\chi_1$   $\chi_2$ . Θα υπολογίσουμε λοιπόν τις διεδρες  $\chi_1$ ,  $\chi_2$  γωνίες για κάθε κατάλοιπο κατά τη διάρκεια του trajectory ώστε να προσδιορίσουμε τα rotamers από τα οποία περνάει το κάθε κατάλοιπο. Η γωνία  $\chi_1$  καθορίζεται από τα άτομα N-C $\alpha$ -C $\beta$ -C $\gamma$  και η  $\chi_2$  καθορίζεται από τα άτομα C $\alpha$ -C $\beta$ -C $\gamma$ -C $\delta$ 1. Ο υπολογισμός των γωνιών θα πραγματοποιηθεί με το πρόγραμμα αναλύσεων μοριακής δυναμικής euclb (I. Γ. Τσούλος και Αθανάσιος Σταβρακούδης), ενώ θα χρειαστούμε και ένα Script γραμμένο σε γλώσσα προγραμματισμού C το οποίο θα μετατρέπει τις γωνίες από  $\pm 180^\circ$  σε  $0-360^\circ$  (πάρτημα 4.4).

Το μόριο μας αποτελείται από 216 κατάλοιπα εκ των οποίων τα 64 ανήκουν στα 16 επίπεδα που σχηματίζουν τον υδρόφοβο πυρήνα και είναι αλανίνες (Ala), φαινυλαλανίνες (Phe), κυστεΐνες (Cys), γλουταμικά οξέα (Glu), λευκίνες (Leu), ισολευκίνες (Ile) και θρεονίνες (Thr). Διεδρες γωνίες  $\chi_1$  έχουν και τα έξι είδη καταλοίπων εκτός από τις αλανίνες των οποίων η πλευρική ομάδα είναι μια μεθυλομάδα, ενώ  $\chi_2$  γωνίες έχουν μόνο οι Leu, Ile, Phe και Glu επίσης λόγω της πλευρικής τους ομάδας. Γι' αυτά τα κατάλοιπα θα μελετήσουμε τα rotamers που παίρνουν κατά τη διάρκεια της προσομοίωσης και θα συγκρίνουμε τις τιμές τους με τις τιμές των rotamers που είναι κρυσταλλογραφικά προσδιορισμένα (εικόνα 4.34).

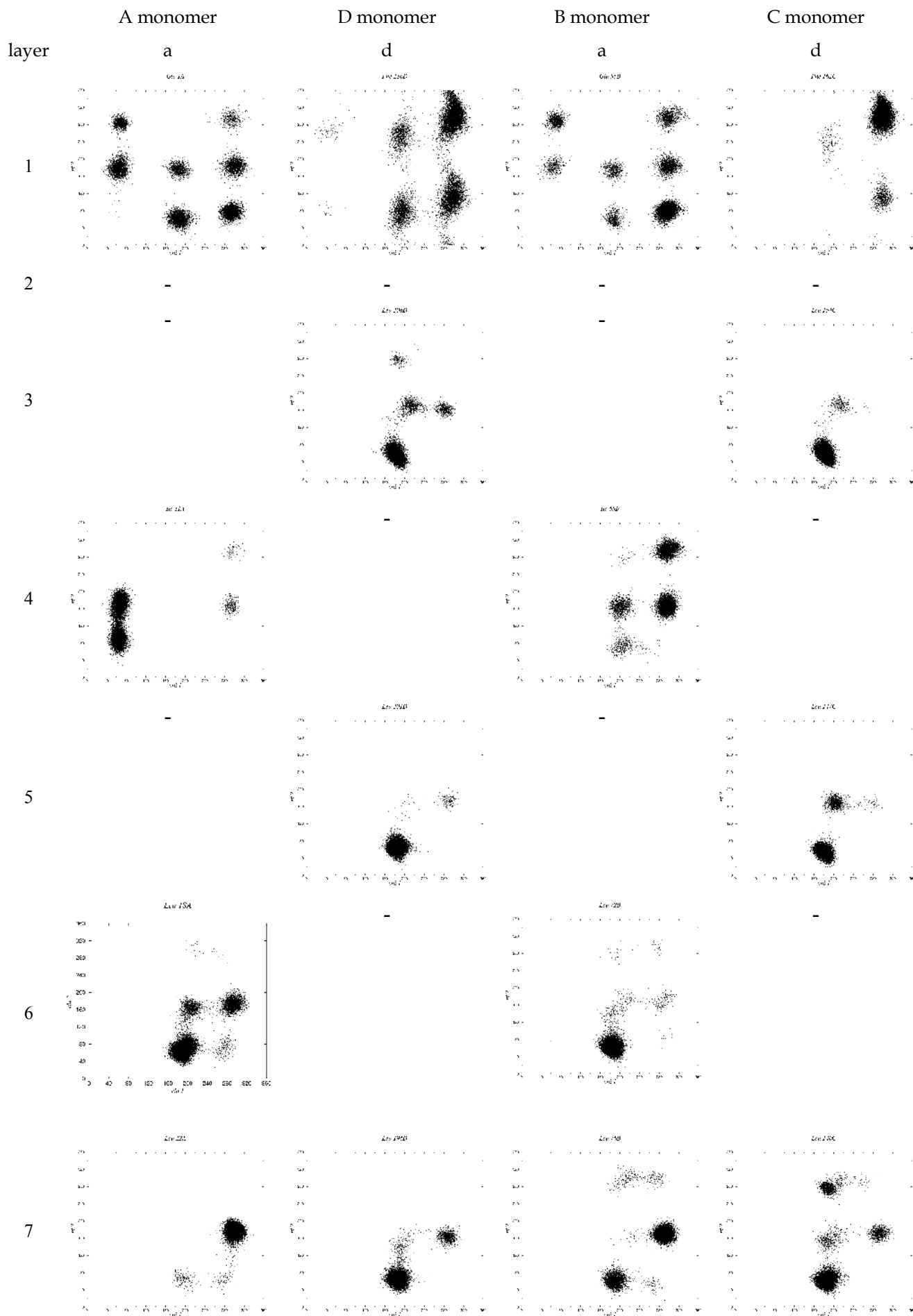
Τα γραφήματα τα οποία προκύπτουν έχουν στον κάθετο άξονα τους τις τιμές των  $\chi_2$  γωνιών και στον οριζόντιο άξονα τις τιμές των  $\chi_1$  γωνιών (εικόνα 4.35). Οι κουκίδες που εμφανίζονται αναπαριστούν την κατανομή των rotamers για τις γωνίες  $\chi_1$ ,  $\chi_2$ , ενώ η μέγιστη συγκέντρωση των τιμών φανερώνει τις προτιμήσεις τους. Έπειτα από σύγκριση των rotamers που προκύπτουν από τις προσομοιώσεις (εικόνα 4.35) με τα rotamers που προσδιορίστηκαν κρυσταλλογραφικά (εικόνα 4.34) για τα κατάλοιπα αυτά, παρατηρούμε ότι κάποια συμφωνούν ενώ, κάποια δεν εμφανίζουν όλα τα προσδιορισμένα rotamers. Συγκεκριμένα, τα κατάλοιπα Glu1, Glu55 συμφωνούν, ενώ το Glu109 έχει κάποια από τα παρατηρούμενα rotamers και ένα επιπλέον που δεν ανήκει σε αυτά, όπως συμβαίνει και στο Glu 163. Οι φαινυλαλανίνες καταλαμβάνουν από 1 έως και 4 rotamers τα οποία συμφωνούν με τα κρυσταλλογραφικά δεδομένα. Το ίδιο συμβαίνει και με τις ισολευκίνες

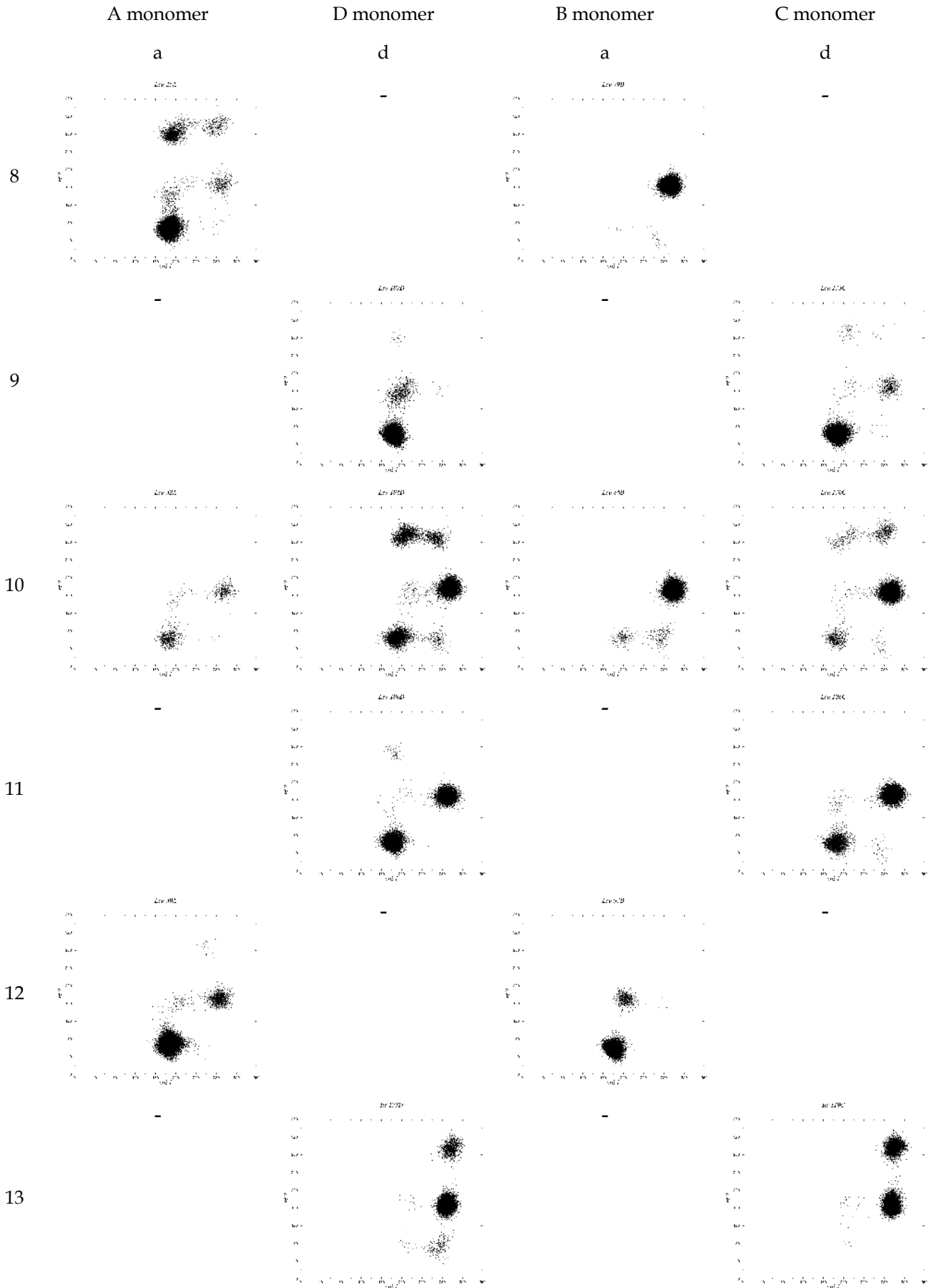
οι οποίες εμφανίζουν από 2 έως 3 rotamers. Τέλος, η κατανομή των rotamers για τα κατάλοιπα λευκίνης είναι πιο περίπλοκη, καθώς άλλα συμφωνούν περισσότερο και άλλα λιγότερο και άλλα παίρνουν τις πιο σπάνιες μορφές των rotamers όπως συμβαίνει με τα κατάλοιπα Leu 76B, Leu 140C, Leu 25A, Leu 184D, Leu46A, Leu 100B.

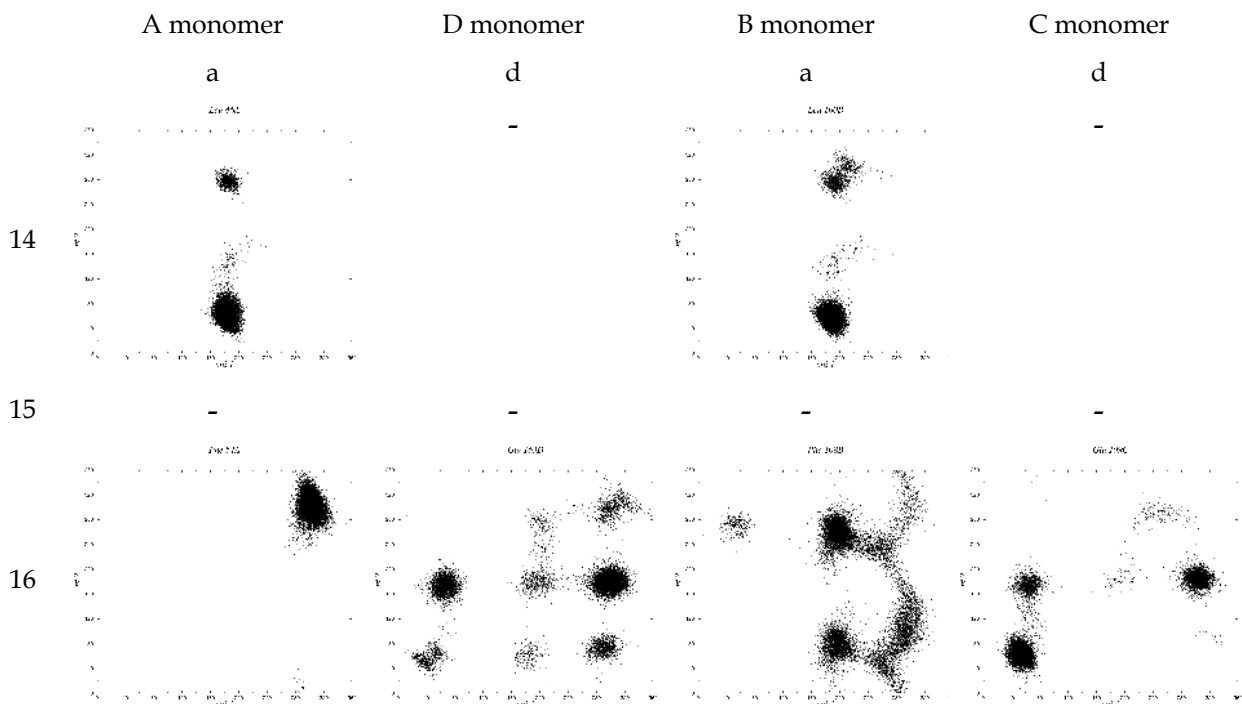


Εικόνα 4.34 Κατανομή λευκίνης, ισολευκίνης, γλουταμικού οξέος και φαινολανίνης των  $\chi_2$  διεδρων γωνιών

Plots of chi1-chi2 dihedral angles

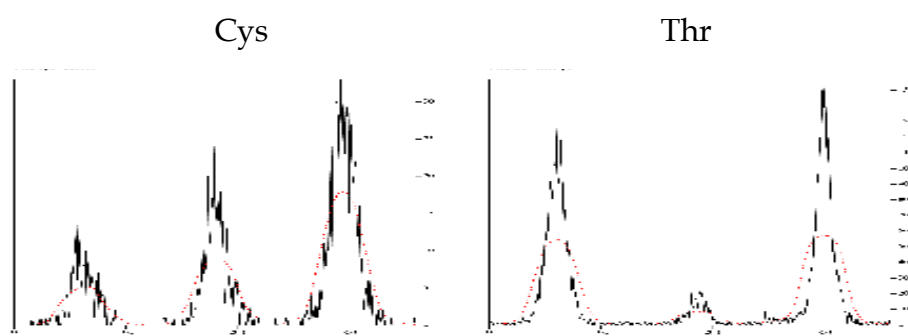




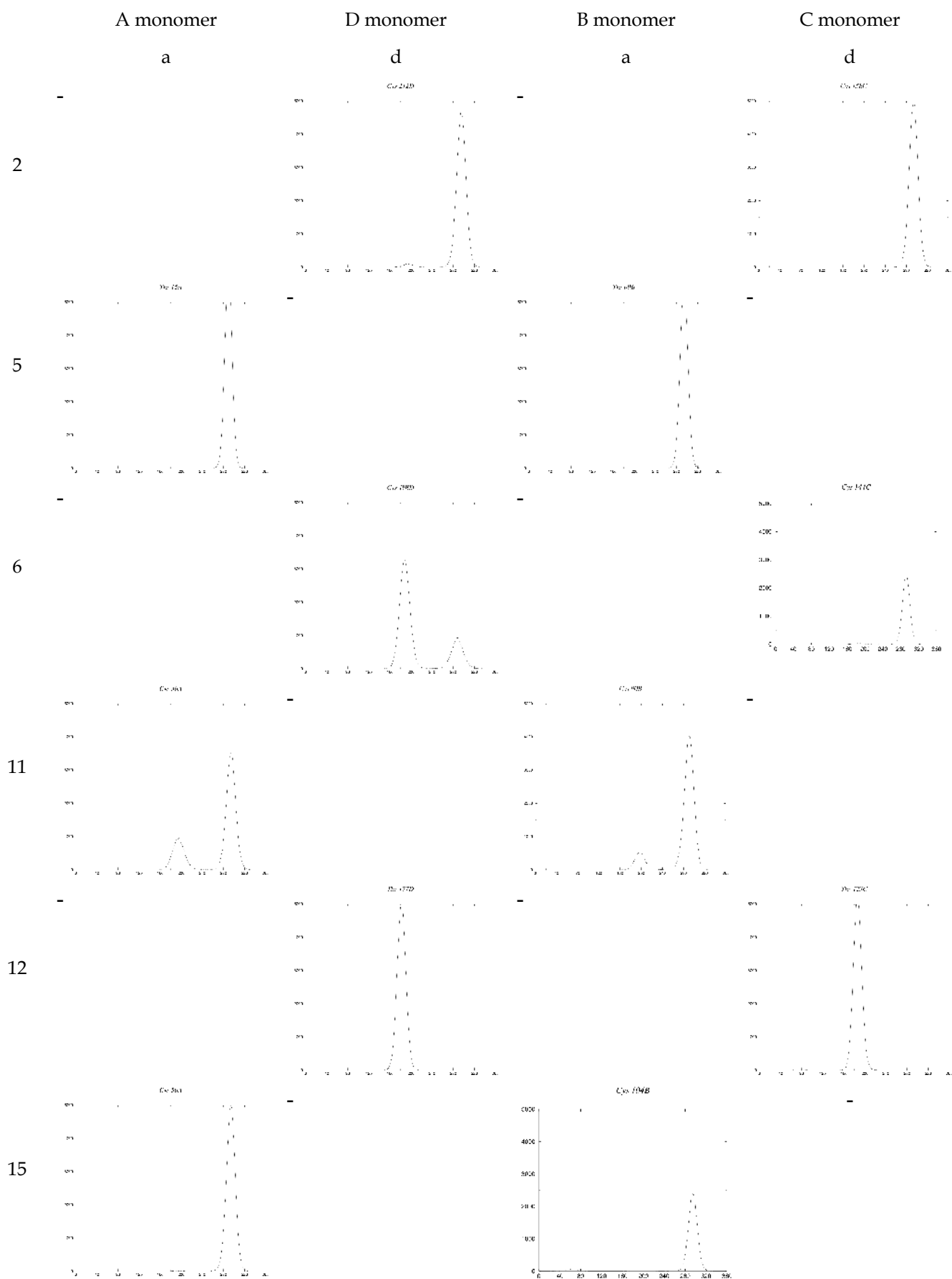


**Εικόνα 4.35** Κατανομή των διεδρων γωνιών  $\chi_1$ ,  $\chi_2$  μόνο των κατάλοιπων λευκίνης, ισολευκίνης, γλουταμικού οξέος και φαινυλαλανίνης του υδρόφοβου πυρήνα καθενός επιπέδου (layer), η οποία διαμορφώνεται κατά τη διάρκεια του trajectory. Τα κατάλοιπα είναι τοποθετημένα σύμφωνα τις θέσεις a και d που καταλαμβάνουν στον υδρόφοβο πυρήνα. Τα layer 2 και 15 αποτελούνται από κατάλοιπα Cys και Ala τα οποία δεν έχουν  $\chi_2$  γωνίες.

Όσο αφορά τα κατάλοιπα θρεονίνης και κυστεΐνης τα οποία δεν έχουν  $\chi_2$  γωνίες κατασκευάσαμε τα γραφήματα για τις  $\chi_1$  γωνίες τους (εικόνα 4,37). Στην εικόνα 4.36 φαίνονται τα γραφήματα των γωνιών για τα κατάλοιπα αυτά όπως έχουν προσδιοριστεί κρυσταλλογραφικά.



**Εικόνα 4.36** Κατανομή των  $\chi_1$  γωνιών για τις κυστεΐνες και τις θρεονίνες όπως έχουν προσδιοριστεί κρυσταλλογραφικά. (<http://xray.bmc.uu.se/gerard/rama/chi.html>)



**Εικόνα 4.37** Κατανομή των τιμών των διεδρων γωνιών  $\chi_1$  των καταλοίπων θρεονίνης και κυστεΐνης σε μία διάσταση κατά τη διάρκεια της προσομοίωσης για τις θέσεις όπου όπου βρίσκονται.



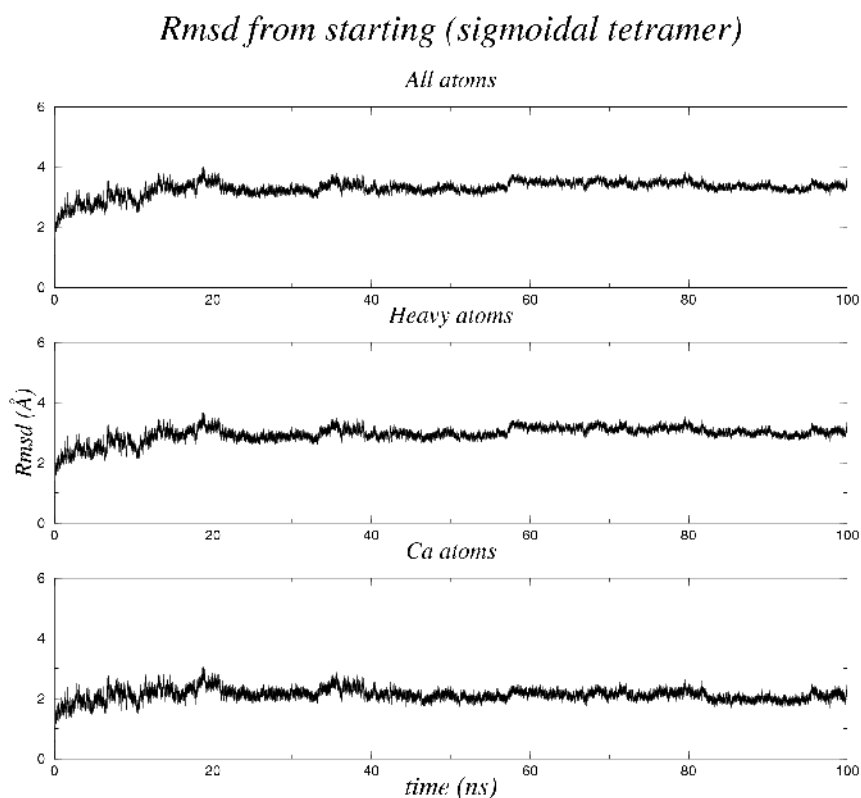
Έπειτα από σύγκριση των γραφημάτων συμπεραίνουμε ότι οι Thr15 και 69 συμφωνούν με τις πειραματικά προσδιορισμένες τιμές ενώ οι Thr 123 και 177 πήραν την πιο σπάνια διαμόρφωση από τις 3 και εμφανίζουν υψηλή τιμή λόγω του πλήθους των frames. Τέλος, τα γραφήματα των Cys επίσης συμφωνούν αλλά δεν εμφανίζουν όλες τις μορφές και στην ίδια τιμή.

## 4.2 Ανάλυση προσομοίωσης μοντέλου sigmoidal

Η ενότητα που ακολουθεί, αφορά στην ανάλυση των προσομοιώσεων του δεύτερου βελτιωμένου μοντέλου το οποίο προτείνουμε για τη δομή της 2ααQL, το sigmoidal. Αυτή η δομή κατασκευάστηκε χρησιμοποιώντας το πρόγραμμα Modeller όπως έχει αναφερθεί στο κεφάλαιο 3.

### 4.2.1 Σταθερότητα αρχικής δομής

Ως πρώτο στάδιο στην ανάλυση του μοντέλου sigmoidal θα χρησιμοποιήσουμε την παράμετρο rmsd (root-mean-square-deviation) από την αρχική δομή, η οποία μελετά τη σταθερότητα και το βαθμό απόκλισης της δομής του μοντέλου, από την αρχική του διαμόρφωση κατά τη διάρκεια της προσομοίωσης.



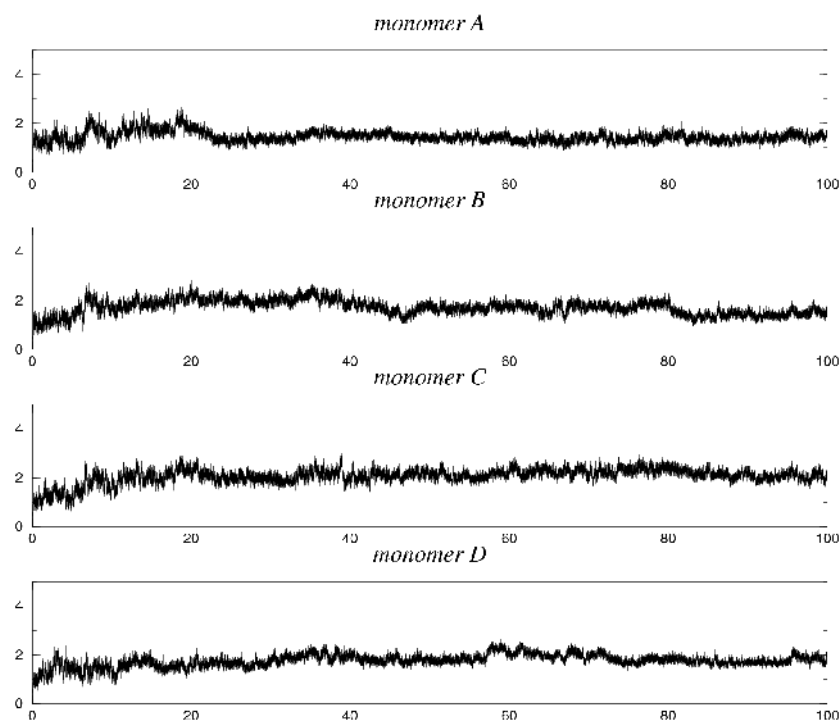
**Εικόνα 4.38** Γράφημα του rmsd από την αρχική δομή συναρτήσει του χρόνου (ns) για τη δομή του sigmoidal: για όλα τα άτομα, για όλα τα άτομα εκτός τα πρωτόνια (Heavy atoms), μόνο για τα Ca άτομα του σκελετού. Στον άξονα των y (rmsd τιμές) κρατήσαμε ίδια κλίμακα με εικόνα 4.1 ώστε να είναι άμεσα συγκρίσιμες.

Στην εικόνα 4.38 απεικονίζεται η rmsd απόκλιση της δομής χρησιμοποιώντας αρχικά όλα τα άτομα της πρωτεΐνης (all atoms), έπειτα αφαιρώντας τα πρωτόνια (Heavy

atoms) και τέλος υπολογίζοντας μόνο τα Ca άτομα του πρωτεϊνικού σκελετού. Παρατηρούμε ότι στην αρχή το rmsd εμφανίζει μια απότομη αύξηση μέχρι τα 20ns περίπου με μεγάλη διακύμανση η οποία μικραίνει όσο περνάει ο χρόνος και το rmsd σταθεροποιείται. Το rmsd που εμφανίζεται στην διάρκεια της προσομοίωσης είναι 2.65Å, με μικρή διακύμανση 0.19Å όταν υπολογίζουμε όλα τα άτομα, ενώ μειώνεται στην τιμή 2.12Å όταν χρησιμοποιούμε για τον υπολογισμό του, μόνο τα Ca άτομα του σκελετού με διακύμανση 0.20Å. Υπολογίσουμε ξανά το rmsd χωρίς τα πρώτα 20ns ή 50000 frames (18ns 26000 frames) τότε βλέπουμε το rmsd να γίνεται 3.35Å με διακύμανση 0.14Å. Επίσης παρατηρούμε ότι μετά την αύξηση του rmsd μέχρι τα 20ns, αυτό σταθεροποιείται αλλά εμφανίζονται κάποιες μικρές αλλαγές στη σταθερότητα του οι οποίες συμβαίνουν, κατά προσέγγιση, στα 33ns και 58ns όπου στο διάγραμμα διακρίνονται σαν μικρά σκαλοπάτια. Οι αλλαγές αυτές φαίνονται και στα διαγράμματα με τα βαριά άτομα και με τα Ca άτομα, σε μικρότερη κλίμακα πράγμα αναμενόμενο αφού μειώθηκε ο αριθμός των ατόμων. Αυτό σημαίνει ότι δεν είχαμε μικρές αλλαγές μόνο στη διαμόρφωση κάποιων πλευρικών ομάδων αλλά και στο σκελετό του μορίου. Με τη μείωση του rmsd συμπεραίνουμε ότι μετά τα 50000 frames η δομή σταθεροποιείται σε μια άλλη διαμόρφωση την οποία διατηρεί σε μεγάλο βαθμό για το υπόλοιπο της προσομοίωσης με κάποιες μικρές αλλαγές στους χρόνους που αναφέρθηκαν.

Μελετώντας εκτενέστερα την δομή θα αναφερθούμε στην συμπεριφορά του εκάστοτε μονομερούς υπολογίζοντας το rmsd από την αρχική δομή για το καθένα (εικόνα 4.39). Μέχρι τα 20ns οι διακυμάνσεις είναι ιδιαίτερα έντονες, το rmsd δεν εμφανίζει ιδιαίτερα απότομη αύξηση στα μονομερή A και D, όπως συμβαίνει στα μονομερή B, C και σε ολόκληρη την πρωτεΐνη, ωστόσο η τιμή του είναι υψηλή στην αρχή και έπειτα σταθεροποιείται. Συγκεκριμένα, το μονομερές A εμφανίζει rmsd 1.45Å με διακύμανση 0.21Å, το οποίο μειώνεται ελάχιστα στο 1.41Å όταν αφαιρούμε από τους υπολογισμούς μας τα πρώτα frames των έντονων διακυμάνσεων. Στο μονομερές B το rmsd είναι 1.70Å με αρκετά υψηλή διακύμανση 1.30Å. Στην αρχή εμφανίζεται απότομη αύξηση της τιμής του μέχρι τα 20ns και έπειτα σταθεροποιείται με κάποιες αλλαγές όπως αναφέρεται και στην εικόνα 4.38 ότι εμφανίζει το τετραμερές. Το μονομερές C στην αρχή, εμφανίζει την χαρακτηριστική απότομη αύξηση, έπειτα όμως παρατηρείται ότι έχει αρκετά σταθερό rmsd, όμως με μεγάλη τιμή 2.05 Å και μεγάλη διακύμανση 1.31Å και τις λιγότερο έντονες αλλαγές. Τέλος το μονομερές D εκτός από τις έντονες διακυμάνσεις στην αρχή εμφανίζεται αρκετά σταθερό με μικρό rmsd 1.75Å και μικρή διακύμανση 0.25Å.

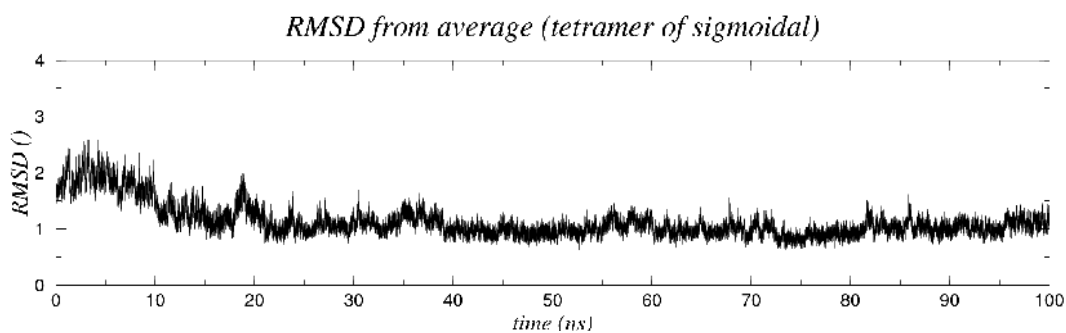
### Rmsd from starting (only Ca atoms)



Εικόνα 4.39 Rmsd από την αρχική δομή των τεσσάρων μονομερών, για τα Ca άτομα του πρωτεϊνικού σκελετού.

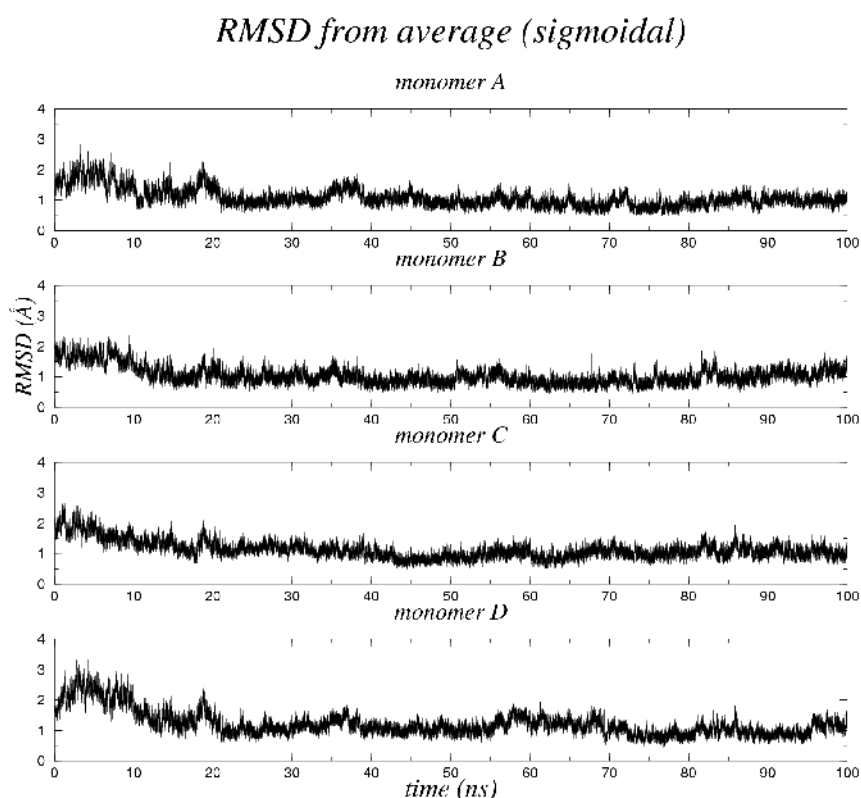
### 4.2.2 Σταθερότητα δομών από το trajectory και υπολογισμός μέσης δομής

Μια σημαντική παράμετρος για την μελέτη της σταθερότητας της δομής μας είναι ο υπολογισμός της τυπικής απόκλισης rmsd από τη μέση δομή. Πληροφορούμαστε για το πόσο σταθερό είναι το μόριο και για το αν σταθεροποιείται σε μια δομή ή αν συγκλίνει σε πολλές, όπως έχουμε αναφέρει στην παράγραφο 4.1.2. Στην εικόνα 4.40 απεικονίζεται το rmsd από τη μέση δομή για τα Ca άτομα του σκελετού του μορίου της πλεξούδας. Παρατηρούμε ότι σταθεροποιείται σε μια μέση δομή μετά από τα 20ns με rmsd 1.12Å και διακύμανση 0.29Å.



Εικόνα 4.40 Τοπική απόκλιση rmsd από τη μέση δομή για τα Ca άτομα του πρωτεϊνικού σκελετού της πρωτεΐνης.

Όσο αφορά τα rmsd από την μέση δομή για κάθε μονομερές της πρωτεΐνης (εικόνα 4.41), παρατηρούμε ότι το rmsd για το μονομερές A είναι αρκετά σταθερό με μικρή τιμή 1.07Å και μικρή διακύμανση 0.30Å. Το μονομερές B, έχει το πιο σταθερό rmsd με μικρή τιμή 1.03Å και διακύμανση 0.28Å, που σημαίνει ότι βρήκε την μέση δομή και σταθεροποιήθηκε σε αυτήν. Το μονομερές C φαίνεται αρκετά σταθερό με rmsd 1.12Å όμως με αρκετά υψηλή διακύμανση 0.28Å και μια εμφανή αλλαγή στα 60ns, ενώ το μονομερές D έχει επίσης μικρό rmsd 1.21Å με διακύμανση 0.41Å όπως παρουσιάζονται στον πίνακα 4.3.



Εικόνα 4.41 Το Rmsd από την μέση δομή για τα Ca άτομα του πρωτεϊνικού σκελετού των μονομερών της πρωτεΐνης

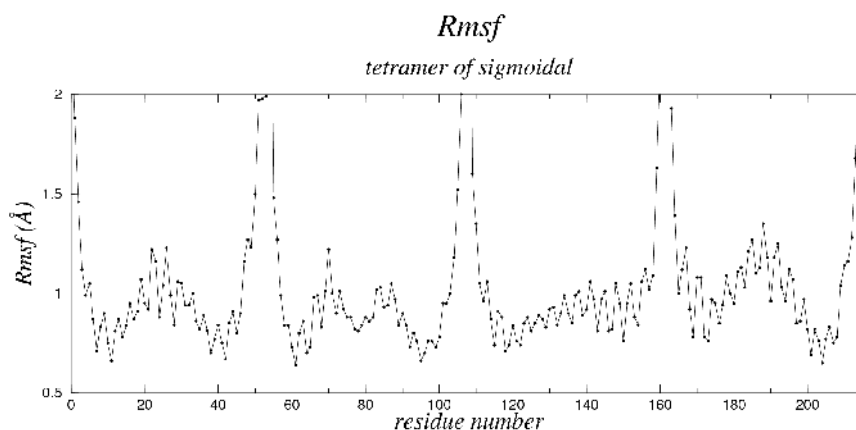
Πίνακας 4.3 Συγκεντρωτικός πίνακας των μέσων τιμών και αντίστοιχων αποκλίσεων των Rmsd από τη μέση και την αρχική δομή

	RMSD FROM STARTING		RMSD FROM AVERAGE	
	Μέση τιμή	Διακύμανση	Μέση τιμή	Διακύμανση
<b>A monomer</b>	1.45	0.21	1.07	0.30
<b>B monomer</b>	1.70	0.30	1.03	0.28
<b>C monomer</b>	2.05	0.31	1.12	0.28
<b>D monomer</b>	1.75	0.25	1.21	0.41
<b>All atoms</b>	2.65	0.19	1.12	0.29
<b>Heavy atoms</b>	2.95	0.25		
<b>Ca atoms</b>	2.12	0.20		

### 4.2.3 Ανάλυση της κινητικότητας των ατόμων (Rmsf)

Θα προχωρήσουμε στην μελέτη της κινητικότητας των Ca ατόμων του κάθε καταλοίπου από το μοντέλο “sigmoidal” χρησιμοποιώντας την παράμετρο rmsf, η οποία όπως αναφέραμε δίνει πληροφορίες για τις ατομικές διακυμάνσεις των ατόμων σε σχέση με την θέση που είχαν στη μέση δομή και τονίζει τα ευκίνητα κατάλοιπα.

Παρατηρώντας την εικόνα 4.42, βλέπουμε όπως ήταν αναμενόμενο να εμφανίζουν μεγαλύτερη κινητικότητα, συνεπώς και μεγαλύτερες διακυμάνσεις, τα κατάλοιπα στα άκρα των τεσσάρων ελίκων, ενώ υπάρχει και μια αξιοσημείωτη αύξηση της κινητικότητας στο κεντρικό τους τμήμα, το οποίο παρατηρείται και στα rmsf των καταλοίπων της πρότυπης δομής της RM6 (Glykos et al., 2003) όπως επίσης και στο προηγούμενο μοντέλο, την πλεξούδα.



**Η εικόνα 4.42** Απεικόνιση της κινητικότητας των Ca ατόμων της πρωτεΐνης σύμφωνα με την παράμετρο Rmsf. Για λόγους ευκρίνειας έχουν αφαιρεθεί από την εικόνα τα κατάλοιπα εκείνα τα οποία είχαν υψηλό rmsf (πάνω από 2 Å) δηλαδή τα: Glu(E) 1 της A αλυσίδας, Glu(E) 55, Arg(R) 107, Phe(F) 108 της B αλυσίδας, Glu(E) 109, Arg(R) 161 και Phe(F) 162 της C αλυσίδας, Glu(E) 163, Arg(R) 215 και Phe(F) 216 της αλυσίδας D.

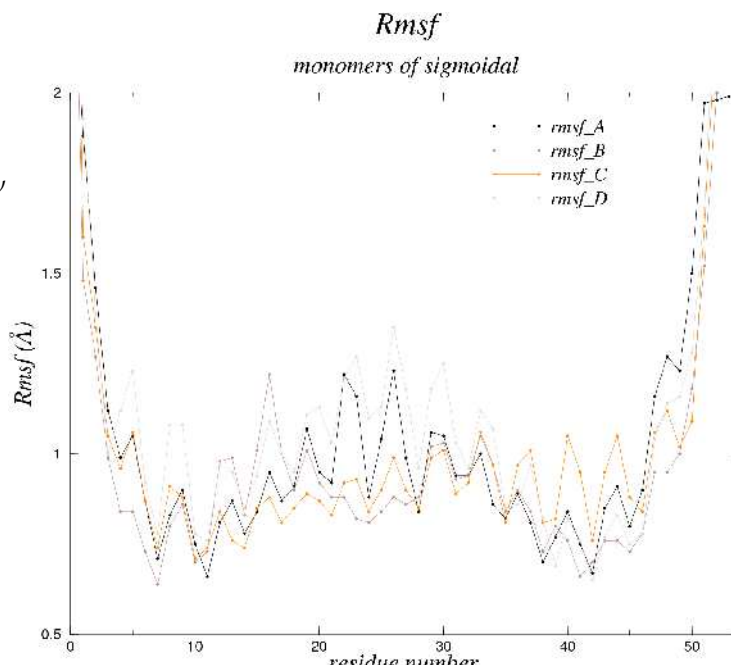
Στο γράφημα απεικονίζονται διαδοχικά τα κατάλοιπα των τεσσάρων αλυσίδων όπου εμφανίζουν μεγαλύτερη κινητικότητα τα αυτά των αμινοτελικών άκρων από αυτά των καρβοξυτελικών. Επιπλέον, η διακύμανση που παρατηρείται στα ενδιάμεσα κατάλοιπα της κάθε αλυσίδας είναι μικρότερη από αυτήν των άκρων αλλά είναι αρκετά υψηλή για άτομα που ανήκουν στον υδρόφοβο πυρήνα. Υψηλότερη κινητικότητα εμφανίζουν τα άτομα των A και D αλυσίδων, κάποια μείωση φαίνεται να έχουν τα ενδιάμεσα άτομα της B αλυσίδας και πιο σταθερά μοιάζουν να είναι τα κατάλοιπα της C.

Στην εικόνα 4.43 μελετώνται όλα τα μονομερή καθώς υπολογίζεται ο συντελεστής συσχέτισης τους. Περιμέναμε οι αλυσίδες που έχουν μικρή μέση απόσταση, δηλαδή οι A-D και B-C, να έχουν μεγάλο συντελεστή συσχέτισης (πίνακας 4.4), κάτι το οποίο δεν φαίνεται να ισχύει, ωστόσο οι συντελεστές συσχέτισης όλων των συνδυασμών των αλυσίδων έχουν αρκετά υψηλές τιμές.

**Πίνακας 4.4**

Γραμμικός συντελεστής συσχέτισης μεταξύ των μονομερών.

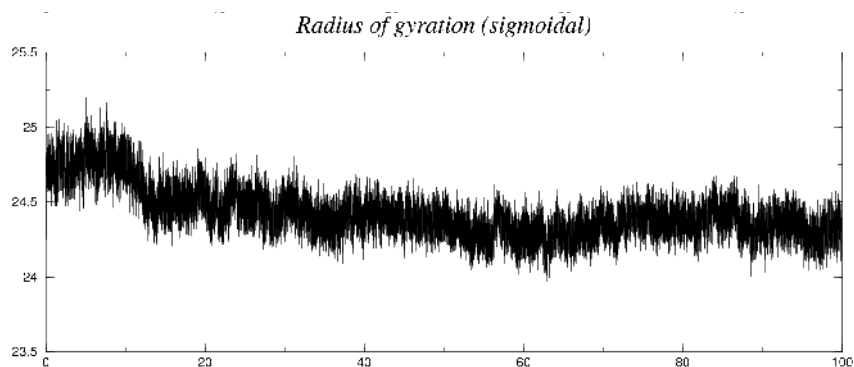
-	A	B	C'	D'
A	-	0.89	0.77	0.84
B		-	0.84	0.90
C			-	0.92
D				-



Εικόνα 4.43 Rmsf των καταλοίπων των μονομερών της πρωτεΐνης.

#### 4.2.4 Radius of gyration (Rg)

Μια άλλη παράμετρος με την οποία ελέγχουμε τη σταθερότητα του μορίου είναι το Rg μέσω της οποίας πληροφορούμαστε για το πόσο συμπαγές είναι το μόριο και κατά πόσο διατηρείται η δομή του.



Εικόνα 4.44 Εξέλιξη του Rg για το μόριο της πλεξούδας.

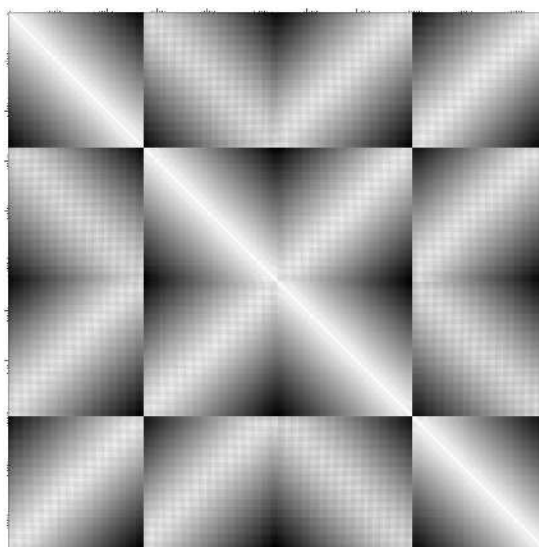
Συγκεκριμένα, στην εικόνα 4.44 διακρίνεται η τιμή του Rg να κυμαίνεται γύρω από μια μέση τιμή το 24.5Å. Το γεγονός ότι μετά τα πρώτα 18ns σταθεροποιείται η τιμή του στα 24.5Å σημαίνει ότι το μόριο είναι αρκετά συμπαγές και η δομή του διατηρείται μέχρι το τέλος της προσομοίωσης.

#### 4.2.5. Χάρτης αποστάσεων Ca-Ca ατόμων και ανάλυση των διακυμάνσεων του.

Στρέφοντας την προσοχή μας στην δυναμική του sigmoidal θα εξετάσουμε την διακύμανση των αποστάσεων των Ca-Ca ατόμων των δομών κατά τη διάρκεια της προσομοίωσης. Επειδή η προσομοίωση μας είχε 250.000 frames θα χρησιμοποιήσουμε ένα πιο μικρό αντιπροσωπευτικό trajectory ανά 100 frames το οποίο θα περιλαμβάνει μονάχα τα Ca άτομα και θα έχει διαστάσεις 2.500 x 2.500, ώστε να είναι μικρότερες οι υπολογιστικές απαιτήσεις.

- Χάρτης Ca-Ca distance map

Όπως έχουμε αναφέρει στην παράγραφο 4.1.4 για την κατασκευή του χάρτη αυτού χρησιμοποιούμε τις μέσες αποστάσεις μεταξύ των ατόμων Ca από όπου προκύπτει ένας συμμετρικός πίνακας με άξονα συμμετρίας την διαγώνιο. Ισχύει ο ίδιος χρωματικός κώδικας με την εικόνα 4.8, όπου το λευκό χρώμα αντιστοιχεί σε μικρότερες αποστάσεις ατόμων ενώ το μαύρο σε μεγαλύτερες.



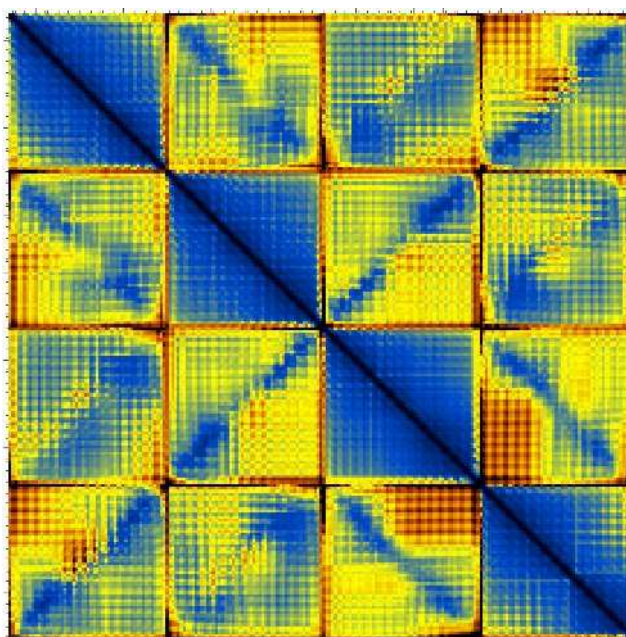
Εικόνα 4.45 Αριστερά Ca-Ca distance map

Η διαγώνιος αντιπροσωπεύει την απόσταση του κάθε καταλοίπου από τον εαυτό του γι' αυτό είναι λευκή. Η μέγιστη παρατηρούμενη απόσταση ανάμεσα στα άτομα είναι 78.3Å, ενώ επιβεβαιώνεται ότι οι αλυσίδες των C και D είναι αντιπαράλληλες με τις A και B.



- **Rmsd από τις μέσες αποστάσεις**

Ο χάρτης Rmsd από τις μέσες αποστάσεις (εικόνα 4.46) δείχνει τις αποκλίσεις των αποστάσεων των ατόμων από τις μέσες τιμές του προηγούμενου πίνακα και μας πληροφορεί για τη διακύμανση τους. Ο πίνακας είναι συμμετρικός με άξονα συμμετρίας την διαγώνιο και ισχύει ο ίδιο χρωματικός κώδικας με της εικόνας 4.9.



*Εικόνα 4.46 Rmsd του χάρτη Ca-Ca distance map.*

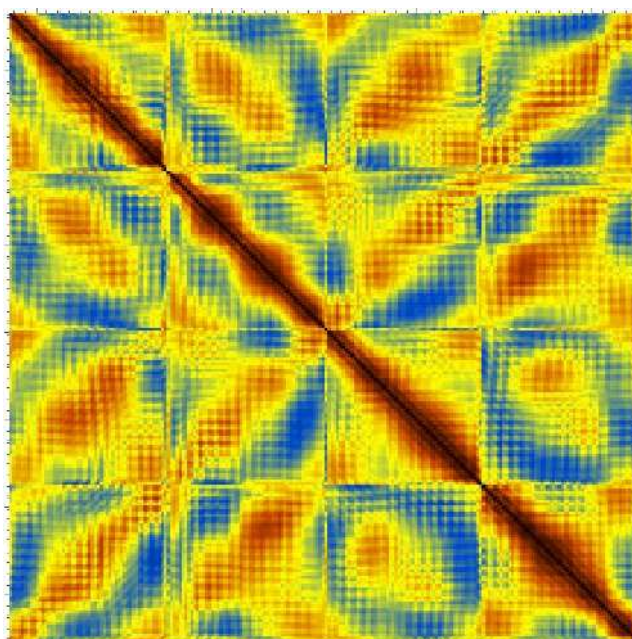
Παρατηρούμε ότι στα τετράγωνα της διαγωνίου, που αντιστοιχούν στο κάθε μονομερές, σχεδόν όλη η επιφάνεια του τετραγώνου καλύπτεται από μπλε χρώμα, αυτό δείχνει ότι είναι αρκετά σταθερή η δομή τους που σημαίνει ότι οι παραμετροί που είχαν οριστεί κατά τη μοντελοποίηση ήταν κατάλληλοι. Υπάρχει έντονη κινητικότητα μεταξύ του καρβοξυτελικού τμήματος της C με το αμινοτελικό τμήμα της D, λόγω της θέσης τους. Επίσης, τα αμινοτελικά άκρα των μονομερών A και D εμφανίζουν υψηλή κινητικότητα διότι βρίσκονται σε αντίθετες πλευρές του δεματίου αφού οι έλικες είναι αντιπαράλληλες.

Η μέγιστη απόκλιση που βρέθηκε ήταν 4.79Å, όμως για να έχουμε πιο ξεκάθαρο και σαφές χρωματικό διάγραμμα χρησιμοποιούμε ως μέγιστη τιμή 1.5Å, ώστε να πετύχουμε και σύγκριση με τον αντίστοιχο πίνακα της πλεξούδας. Επιπλέον, σχηματίζεται μια σκακιέρα από μαύρες κάθετες και οριζόντιες γραμμές οι οποίες αντιστοιχούν στα υπερκηνιτικά άτομα των καταλοίπων των άκρων. Συγκεκριμένα, η πρώτη μαύρη γραμμή που εμφανίζεται είναι το Glu1 στο N τελικό άκρο του A μονομερούς, οι επόμενες αντιστοιχούν στα κατάλοιπα Glu55 (N-τελικό) και Phe 108 (C-τελικό) της αλυσίδας B,

ακολουθούν τα Glu 109 (N-τελικό) και Phe 162 (C-τελικό) της αλυσίδας C, και τέλος υψηλή κινητικότητα έχουν τα τελικά κατάλοιπα Glu 163 (N-τελικό) και Phe 216 (C-τελικό) του μονομερούς D. Τέλος, τα δεδομένα του πίνακα βρίσκονται σε συμφωνία με τα αποτελέσματα από τον υπολογισμό rmsf όπου τα πιο κινητικά κατάλοιπα ήταν τα Glu 55, Glu109, Glu163 και Phe 216.

#### 4.2.6 Υπολογισμός cross-correlation

Προκειμένου να ερμηνεύσουμε την κίνηση ομάδων ατόμων του μορίου, υπολογίζουμε τον πίνακα variance-covariance ο οποίος έπειτα από την κανονικοποίηση των τιμών του, δίνει το γράφημα cross-correlation, εικόνα 4.47. Το γράφημα αυτό αποκαλύπτει την συσχέτιση των διαφόρων κινήσεων που πραγματοποιεί κάθε τμήμα του μορίου βάση του χρωματικού κώδικα που αναφέρεται στην παράγραφο 4.1.6.



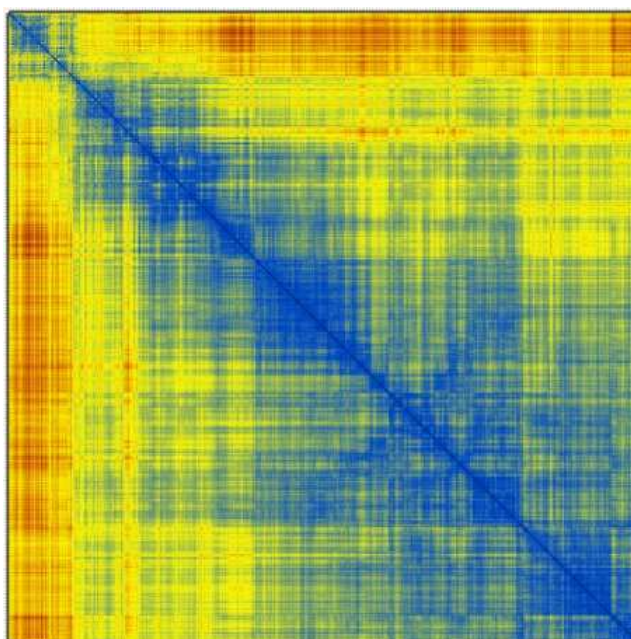
Εικόνα 4.47 Γράφημα του cross correlation

Ο πίνακας αυτός μας δείχνει της κίνηση του κάθε καταλοίπου σε σχέση με τα υπόλοιπα (εδώ υπολογίζουμε μόνο Ca άτομα). Πάνω στη διαγώνιο παρουσιάζεται η κίνηση των ατόμων με τους εαυτούς τους οπότε προφανώς κινούνται μαζί (κόκκινο χρώμα). Παρατηρούμε ότι το κάθε μονομερές χαρακτηρίζεται από διαφορετική κινητικότητα σε σχέση με τα υπόλοιπα μονομερή, αλλά και τα άτομα του κάθε μονομερούς εμφανίζουν διαφορετική κινητικότητα. Στο μονομερές A το αμινοτελικό και το καρβοξυτελικό του

άκρο έχουν θετική συσχέτιση (κόκκινο χρώμα), οπότε κινούνται προς την ίδια κατεύθυνση η οποία είναι αντίθετη από το κεντρικό τμήμα, όπου εμφανίζουν αρνητική συσχέτιση (μπλε χρώμα). Επίσης το αμινοτελικό άκρο της A δεν έχει πολύ μεγάλη συσχέτιση με το αμινοτελικό της B (κίτρινο χρώμα) αλλά έχει μεγαλύτερη συσχέτιση (κόκκινο χρώμα) με το αμινοτελικό της C και το καρβοξυτελικό της D. Επίσης το αμινοτελικό του C προς την ίδια κατεύθυνση με το καρβοξυτελικό του B. Στο B μονομερές διακρίνουμε 4 πιο μικρά κομμάτια όπου το πρώτο με το δεύτερο κινούνται αντίθετα (μπλε), το τρίτο αντίθετα από το 2 (μπλε) αλλά μοιάζει να κινείται προς την ίδια κατεύθυνση με το 1ο (κόκκινο-κίτρινο). Το 4 κομμάτι κινείται αντίθετα από το 3, αλλά στην ίδια κατεύθυνση με το 1ο και το 2ο. Τέλος, το D εμφανίζει το μοτίβο κινητικότητας που παρατηρείται στα μονομερή της πλεξούδας και στην RM6, δηλαδή τα άκρα κινούνται μαζί και σε αντίθετη κατεύθυνση από το κεντρικό τμήμα του μονομερούς.

#### 4.2.7 Rmsd μεταξύ όλων των δομών του trajectory

Ακολουθώντας υπολογίζουμε απόκλιση rmsd μεταξύ δομών του τροχιακού από κάθε frame και δημιουργούμε τον χάρτη frame to frame rmsd (crossDCD.matrix). Αποκτούμε έτσι, μια άποψη της κινητικότητας του μορίου και των διακυμάνσεων των δομών που υιοθετεί, το οποίο μας επιτρέπει μια πρώτη ομαδοποίηση των δομών του trajectory. Ο πίνακας που σχηματίζεται είναι δυσδιάστατος και περιέχει τις δομές ανά 100 στιγμιότυπα, ενώ ο χρωματικός κώδικας είναι ως εξής: το μπλε χρώμα δηλώνει μικρές τιμές rmsd μεταξύ των δομών άρα μοιάζουν πολύ, με κόκκινο, μεγάλες και με κίτρινο, ενδιάμεσες.

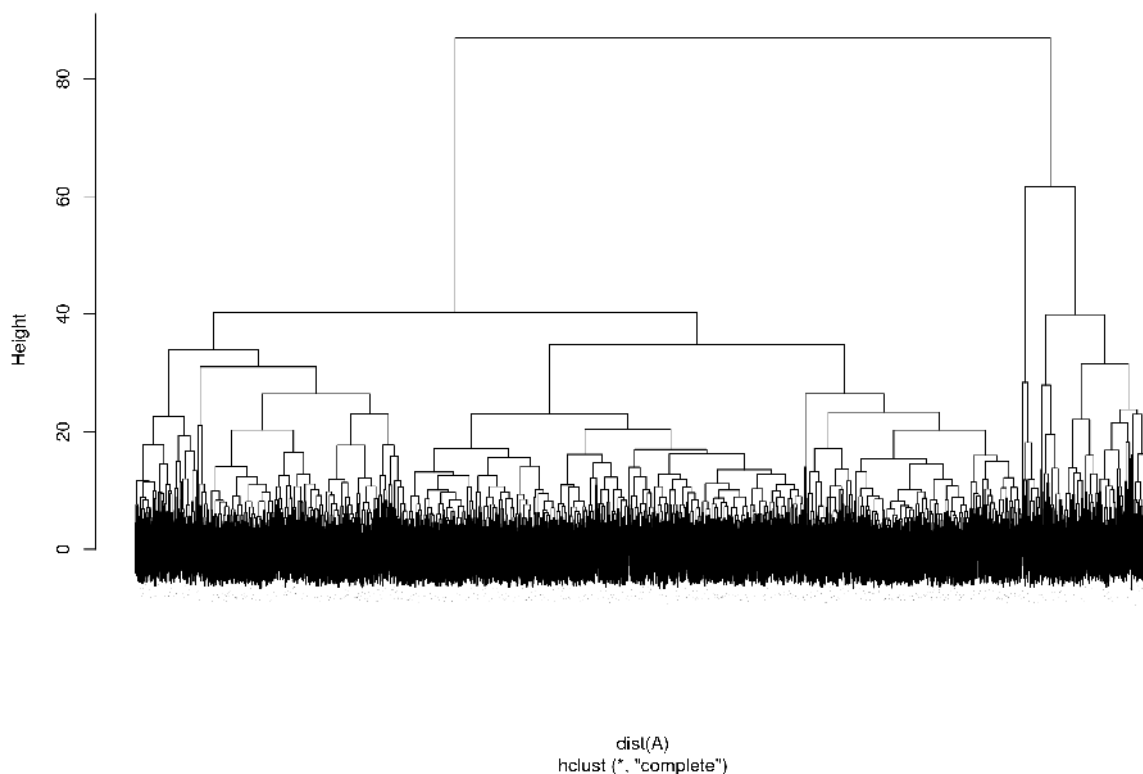


Εικόνα 4.48 Frame to frame rmsd (crossDCD.matrix)

Στην εικόνα 4.48 βλέπουμε ότι το μόριο του sigmoidal ξεκινάει από μια δομή στην οποία μένει περίπου για 26000 frames και έπειτα περνάει σε μια άλλη ομάδα δομών διαφορετική από την αρχική, όπου φαίνεται να δημιουργούνται μικρότερες ομάδες με πολλά όμοια στοιχεία μεταξύ τους (αφού αλληλεπικαλύπτονται). Οπότε συμπεραίνουμε ότι υπάρχουν 2 cluster με το ένα πολύ μεγάλο. Τα αποτελέσματα του crossDCD συμφωνούν με το γράφημα rmsd από την αρχική δομή διότι και στην εικόνα 4.38 φαίνεται ότι υπάρχουν 2 ομάδες δομών, η μια στην διάρκεια της απότομης αύξησης και η άλλη στο σταθερό κομμάτι του rmsd, το οποίο φέρει πολλές αλλαγές αλλά όχι έντονες ώστε να θεωρηθούν σημαντικές και να υποθέσουμε την δημιουργία ξεχωριστής ομάδας. Για να διευκρινίσουμε την ομαδοποίηση των δομών θα προχωρήσουμε σε cluster analysis με το πρόγραμμα R.

## 4.2.8 Ομαδοποίηση των δομών του trajectory (cluster analysis)

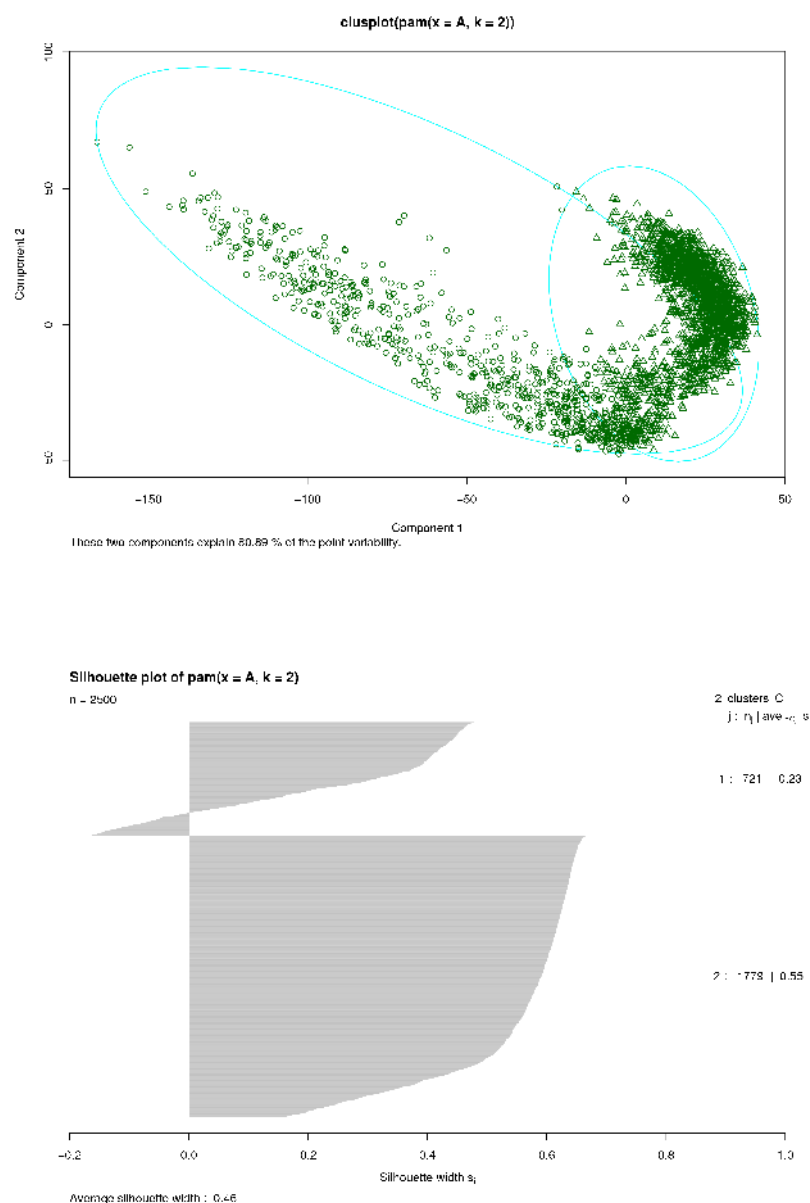
Ένας πολύ κατανοητός τρόπος ομαδοποίησης των δομών του τροχιακού μας σχηματικά, είναι το cluster analysis το οποίο όπως αναφέραμε στην παράγραφο 4.1.8 ομαδοποιούνται οι δομές σύμφωνα με τον βαθμό ομοιότητας τους και κατασκευάζεται ένα δενδρόγραμμα βάση του πίνακα αποστάσεων rmsd (εικόνα 4.48). Πραγματοποιείται το cluster analysis μέσω του προγράμματος R, το οποίο έδωσε το διάγραμμα της εικόνας 4.49. Όπως βλέπουμε στο δενδρόγραμμα δημιουργούνται ένα μεγάλο cluster και ένα πιο μικρό το



**Εικόνα 4.49** Δενδρόγραμμα των δομών του sigmoidal που παρατηρούνται κατά την διάρκεια της προσομοίωσης χρησιμοποιώντας τον πίνακα αποστάσεων rmsd.

Συγκρίνοντας τα δεδομένα αυτά με τον πίνακα crossDCD συμπεραίνουμε ότι το μικρό cluster (στα δεξιά) είναι το ίδιο με το αρχικό cluster του πίνακα crossDCD το οποίο αντιστοιχεί στα 26000 πρώτα frames του trajectory. Όμως προκειμένου να βεβαιωθούμε για το εάν το μεγάλο cluster είναι ένα ή δημιουργούνται και άλλα πιο μικρά θα υπολογίσουμε ξανά τον πίνακα crossDCD και θα σχηματίσουμε το αντίστοιχο

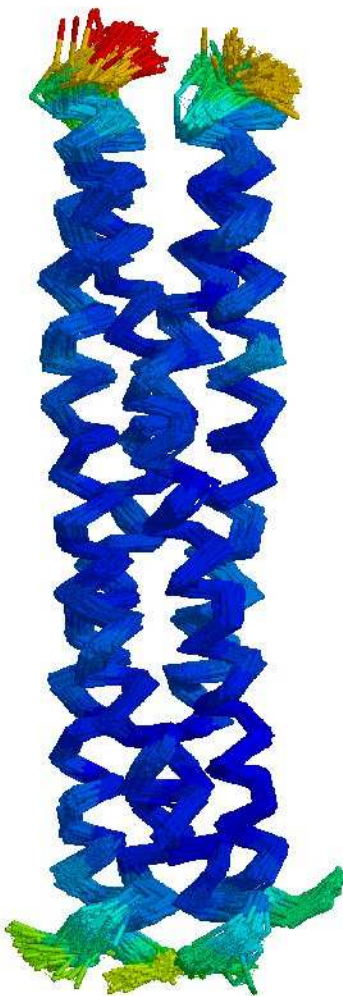
δενδρόγραμμα με τις καινούργιες αποστάσεις (με τα καινούργια δεδομένα). Σε ένα καινούργιο dcd αρχείο αποκόβουμε τα 26000 πρώτα frames. Στην περίπτωση αυτή βλέπουμε ότι crossDCD σχηματίζει ένα μπλε κουτί σχεδόν ομοιόμορφο που σημαίνει ότι υπάρχει ένα κυρίαρχο cluster το οποίο επιβεβαιώνεται και από το δενδρόγραμμα που προέκυψε από το καινούργιο crossDCD. Όταν χρησιμοποιούσαμε τους αλγόριθμους pam και fanny, ώστε να πάρουμε ποσοτικοποιημένα δεδομένα, στους υπολογισμούς για περισσότερο των 2 cluster έδιναν στο γράφημα silhouette plot αρκετές αρνητικές τιμές που σημαίνει ότι είναι λάθος οι υποθέσεις μας. Οπότε έχουμε 2 cluster εκ των οποίων το 1 μικρό και το άλλο το κυρίαρχο.



**Εικόνα 4.50** Τα γραφήματα όπου δείχνουν την κατανομή σε cluster (επάνω) και στο διάγραμμα silhouette το ποσοστό του καθενός.

#### 4.2.9 Υπέρθεση των δομών με τα αντίστοιχα rmsf fluctuations του κυρίαρχου cluster τους.

Συγκεντρώνοντας τα αποτελέσματα των τριών μεθόδων για την ομαδοποίηση των δομών του trajectory (στατιστικού προγράμματος R, ανάλυση των κύριων συνιστωσών (PCA) παράγραφος 4.2.11, πίνακας crossDCD), πληροφορούμαστε για την ύπαρξη ενός μοναδικού κυρίαρχου cluster.



**Εικόνα 4.51** Υπέρθεση των δομών του κυρίαρχου cluster κάθε 1000 frame, χρησιμοποιώντας τα rmsf της μέσης δομής στη θέση των ατομικών θερμικών παραγόντων (*B factors*). Αναπαράσταση της κινητικότητας των κατάλοιπων σύμφωνα με τη χρωματική κλίμακα από το μπλε (μικρή κινητικότητα) στο κόκκινο (μεγάλη κινητικότητα). Τα rmsf τα οποία έχουν χρησιμοποιηθεί στη θέση των ατομικών θερμικών παραγόντων (*B factors*) για την απόδοση της χρωματικής κλίμακας έχουν τροποποιηθεί με κοινή κλίμακα από μπλε ( $0.47\text{\AA}$ ) - κόκκινο ( $2.82\text{\AA}$ ) για λόγους σύγκρισης με το πρώτο μοντέλο την πλεξούδα.

Για το κυρίαρχο cluster, λοιπόν, θα κάνουμε υπέρθεση των δομών του trajectory με βήμα 1000 frames χρησιμοποιώντας στη θέση των ατομικών θερμικών παραγόντων (*B*

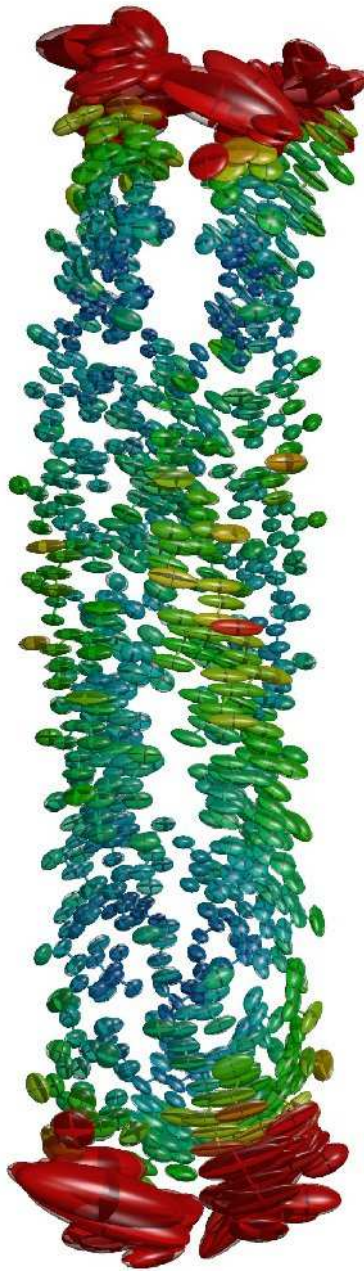
factors), τα rmsf από την μέση δομή. Η εικόνα 4.51 αναπαριστά την κινητικότητα των καταλοίπων (του κυρίαρχου cluster) κατά την διάρκεια του trajectory σύμφωνα με μια χρωματική κλίμακα από το μπλε (μικρή κινητικότητα) στο κόκκινο (μεγάλη κινητικότητα).

Παρατηρούμε ότι μεγάλη κινητικότητα εμφανίζουν τα κατάλοιπα στα άκρα του δεματίου και στις τέσσερις αλυσίδες και ιδιαίτερα στο επάνω άκρο της δομής που αντιστοιχεί στο N τελικό άκρο των A και B αλυσίδων και στο C τελικό των C και D με ιδιαίτερα ευκίνητα κατάλοιπα τα Glu 1 της A και Glu 55 της B (κίτρινο χρώμα), Arg 215 και Phe 216 της D (κόκκινο) με rmsf μεγαλύτερο από 2Å. Επιπλέον, με πράσινο χρώμα απεικονίζονται τα κατάλοιπα της A αλυσίδας Lys 2, της B αλυσίδας Lys 56 και της D αλυσίδας Ala 214 με rmsf μεγαλύτερο από 1.5 Å. Στο άκρο με τα καρβοξυτελικά κατάλοιπα των A και B και τα αμινοτελικά των C και D υπερκινητικό ρόλο έχουν η Phe 108 της B, το Glu 109 της C (κίτρινο), Phe 54 της A, Lys 110 της C, Glu 163 της D (πράσινο), Ala 52 της A, Leu 105 της B, Thr 111 της C και Thr 165 της D (γαλάζιο). Τέλος, παρατηρούμε μια ελαφριά αύξηση της κινητικότητας (γαλάζιο) των καταλοίπων και στο ενδιάμεσο τμήμα των αλυσίδων A και B. Συγκεκριμένα τα κατάλοιπα Thr 17 και Glu 14 του A μονομερούς, ενώ από την B είναι τα κατάλοιπα Thr 71, Glu85, Asp 88.

#### 4.2.10 Anisotropic fluctuations

Όπως αναφέραμε οι πληροφορίες οι οποίες φέρουν οι χάρτες των μέσων αποστάσεων (Ca-Ca rmsd) είναι περιορισμένες και αγνοούν παντελώς την δυναμική του trajectory. Γι' αυτό τον λόγο χρησιμοποιούμε μια γραφική αναπαράσταση που βασίζεται στον χάρτη αυτόν και αναπαριστά την μέση δομή της πλεξούδας χρησιμοποιώντας τους θερμικούς ανισότροπους παράγοντες ως ελλειψοειδή. Τα ελλειψοειδή υπολογίζονται με βάση την κίνηση του σκελετού και των C $\beta$  ατόμων κατά τη διάρκεια της προσομοίωσης. Η κατανομή των ελλειψοειδών αποκαλύπτει την δομικά σημαντική κινητικότητα των καταλοίπων.





**Εικόνα 4.52** Απεικόνιση της μέσης δομής της πλεξούδας χρησιμοποιώντας τους θερμικούς ανισότροπους ελλειψοειδείς παράγοντες (*thermal anisotropic ellipsoids*) υπολογιζόμενοι από την κίνηση του σκελετού και των C $\beta$  ατόμων κατά τη διάρκεια της προσομοίωσης. Τα *rmsf* τα οποία έχουν χρησιμοποιηθεί στη θέση των ατομικών θερμικών παραγόντων (*B factors*) για την απόδοση της χρωματικής κλίμακας έχουν τροποποιηθεί με κοινή κλίμακα από μπλε (0.47 $\text{\AA}$ ) - κόκκινο (2.82 $\text{\AA}$ ) όπως και στην εικόνα 4.48. Αυτή η εικόνα ετοιμάστηκε με το Raster3D (Merritt et al., 1997)

Συγκεκριμένα στην εικόνα 4.52 παρατηρούμε εκτεταμένη κίνηση στα άκρα του δεματίου και κυρίως στο πάνω άκρο, στο N τελικό άκρο των A, B αλυσίδων και στο C τελικό των C, D που αντιστοιχούν στα κατάλοιπα Glu 1 της A και Glu 55 της B, Phe 162

της C και Phe 216 της D ως μεγαλύτερα ελλειψοειδή. Με μικρότερη κινητικότητα, σχηματίζονται ως μικρότερα ελλειψοειδή, τα κατάλοιπα Lys 2 και Thr 3 της A αλυσίδας, της B αλυσίδας Lys 56 και Thr 57, της C αλυσίδας Phe 162, της D αλυσίδας Arg215 και Ala 214. Τέλος, εμφανίζονται και αρκετά ευέλικτα άτομα στο ενδιάμεσο τμήμα των αλυσίδων με την μεγαλύτερη κινητικότητα να εμφανίζονται στην B αλυσίδα το Thr 71 όπως επίσης και τα κατάλοιπα Thr 17 και Glu 14 στην αλυσίδα A.

Συμπεραίνουμε, λοιπόν, από την εικόνα των δομών σε υπέρθεση, από τη μέση δομή με βάση τους ανισότροπους παράγοντες αλλά και από τα rmsf ότι υπάρχουν ιδιαίτερα ευκίνητα κατάλοιπα (ουρές) τα οποία μπορούν να επηρεάσουν σημαντικά την αξιοπιστία των αποτελεσμάτων κατά την εξέλιξη των αναλύσεων της προσομοίωσης εξαιτίας της τυχαίας και ακανόνιστης κίνησης τους μιας και δεν συμμετέχουν σε κάποια σταθερή δομή. Συνήθως τα κατάλοιπα αυτά δεν φέρουν κάποια σημαντική πληροφορία όταν μελετάμε τις διαμορφώσεις της μέσης δομής διότι μας ενδιαφέρει ο σκελετός της δομής του μορίου. Οπότε συνηθίζεται όταν οι αναλύσεις απαιτούν λεπτομερείς πληροφορίες, οι υπολογισμοί να γίνονται χωρίς να λαμβάνονται υπόψη τα συγκεκριμένα κατάλοιπα.

#### 4.2.11 Principal component analysis.

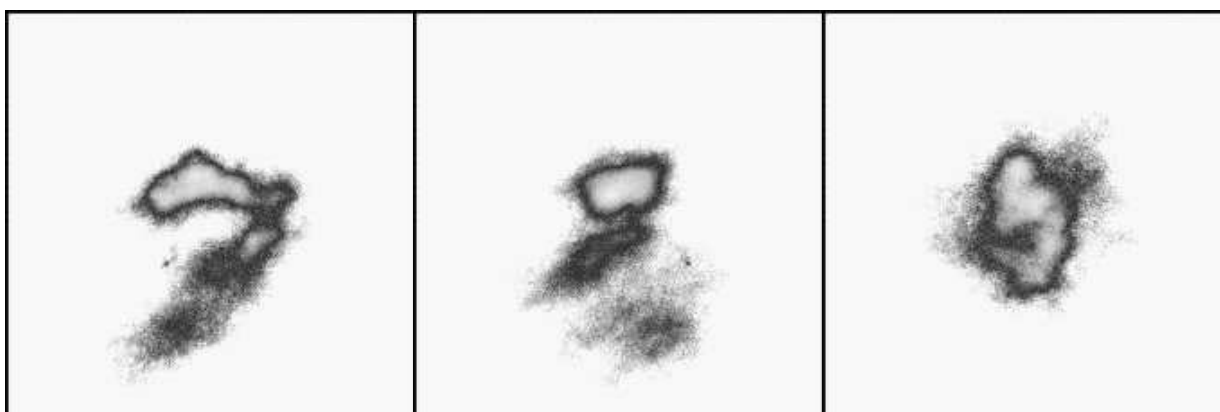
Προχωρώντας στις αναλύσεις του sigmoidal, θα χρησιμοποιήσουμε τη στατιστική μέθοδο των κύριων συνιστωσών, δηλαδή το Principal Component analysis (PCA), ένα χρήσιμο εργαλείο για να αναλύσουμε τις μεγάλες πρωτεϊνικές διακυμάνσεις, να διαχωρίσουμε τις κινήσεις του μορίου κατά τη διάρκεια του trajectory, καθώς επίσης να μελετήσουμε και να ομαδοποιήσουμε τα δεδομένα μας. Οι τιμές των eigenvalues και eigenvectors που προκύπτουν είναι κατάλληλες σε επόμενο επίπεδο για τον χαρακτηρισμό της κίνησης του μορίου και επίσης για να καθοριστεί εάν ήταν αρκετή η διάρκεια της προσομοίωσης (sufficient sampling).

Είναι δύσκολο να αποφασίσουμε για το εάν ο χρόνος της προσομοίωσης ήταν αρκετός ώστε να παρατηρηθούν ολοκληρωμένα οι κινήσεις των μορίων. Εκτός από τα γραφήματα ελεύθερης ενέργειας που προκύπτουν από το carma, ένα πιθανό ποσοτικό κριτήριο για τον έλεγχο του δείγματος είναι η ομοιότητα που εμφανίζει με την τυχαία διάχυση. Ένα μέτρο για την ομοιότητα αυτή είναι η περιεκτικότητα σε συνημίτονο

(παράμετρος cosine content) των κύριων συνιστωσών. Η παράμετρος αυτή μπορεί να πάρει τιμές που κυμαίνονται ανάμεσα στο 0 και το 1. Όταν η πρώτη κύρια συνιστώσα (PC) έχει τιμή κοντά στο 0 το sufficient sampling είναι ικανοποιητικό, ενώ όταν υπάρχει υψηλό cosine content κοντά στο 1, είναι πολύ πιθανό ότι δεν υπάρχει sufficient sampling και ότι το μόριο πραγματοποιεί τυχαία κίνηση. Οπότε το cosine content μπορεί να χαρακτηριστεί ως ένας αρνητικός δείκτης. Ο υπολογισμός του cosine content γίνεται εύκολα με την βοήθεια του πρόγραμμα `cos_content.pl` (παράρτημα 4.5) γραμμένο σε γλώσσα προγραμματισμού perl.

- **Cartesian PCA**

Με το Cartesian-PCA κάνουμε ανάλυση των κύριων συνιστωσών στο καρτεσιανό σύστημα μέσω του προγράμματος `carma`. Το πρόγραμμα υπολόγισε 5 clusters, από τα οποία το πρώτο είναι το κυρίαρχο, το οποίο αποτελείται από 75245 frames και καταλαμβάνει το 97.47% των frames που ανήκουν σε cluster. Η λεπτή δομή που εμφανίζεται στα γραφήματα  $\Delta G$  energy landscapes (εικόνα 4.53) δείχνει ότι μάλλον δεν ήταν αρκετός ο χρόνος της προσομοίωσης για να δούμε ικανοποιητικά τις χαρακτηριστικές κινήσεις του μορίου. Άρα δεν έχουμε ικανοποιητικό sufficient sampling το οποίο επιβεβαιώνεται από την τιμή του cosine content για την πρώτη κύρια συνιστώσα η οποία είναι 0.5749.



1-2

1-3

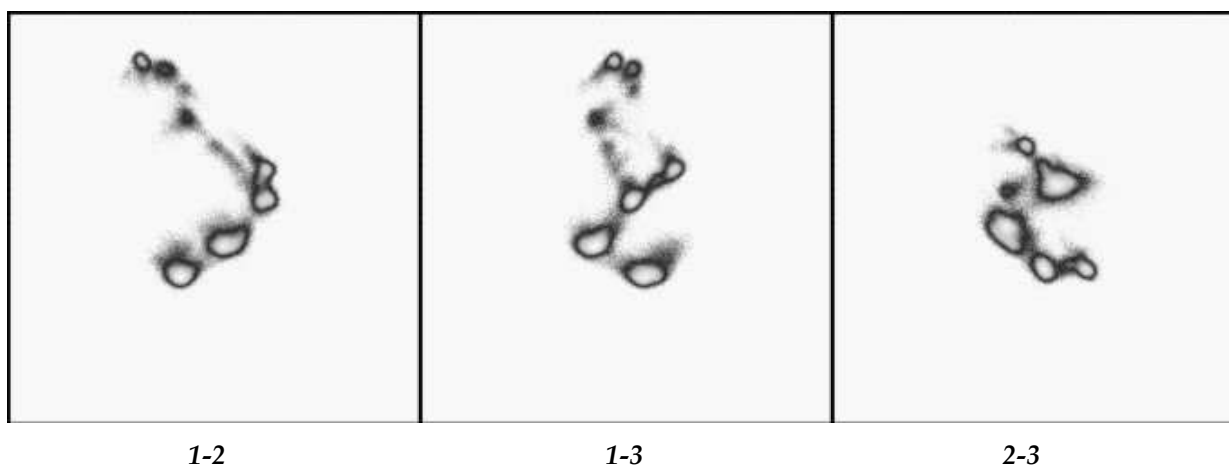
2-3

**Εικόνα 4.53** Γραφήματα ελεύθερης ενέργειας για τις κύριες συνιστώσες 1, 2, 3, οι οποίες καθορίζονται από τα χαρακτηριστικά ανόσματα (eigenvectors) σε ζευγάρια, με κλίμακα -39.39 μέχρι +39.39 και με ενεργειακό ελάχιστο 2.77 kcal/mol.

- **Dihedral PCA**

Στην περίπτωση όπου χρησιμοποιούμε τις διεδρες γωνίες ( $\phi, \psi$ ) της πρωτεΐνης για την ανάλυση των κύριων συνιστωσών, πραγματοποιείται δηλαδή dihedral-PCA, το cluster analysis έδειξε την ύπαρξη 5 cluster τα οποία αποτελούνται από:

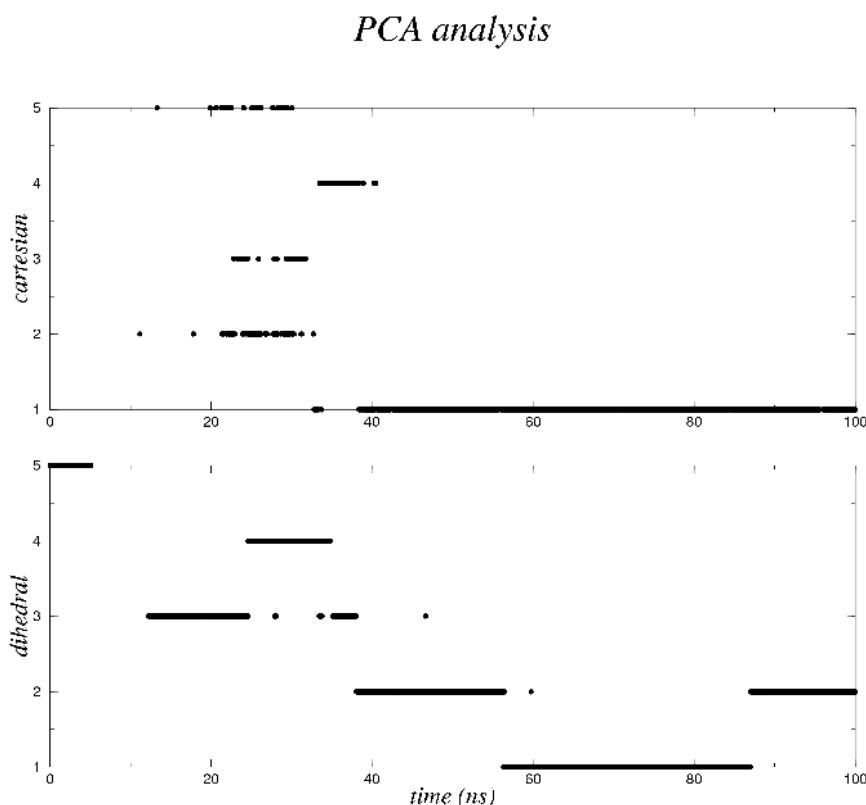
4.1o cluster -->	44611 frames με ποσοστό	occupancy	33,8%
5.2o cluster -->	45963 frames με ποσοστό	“	34,8%
6.3o cluster -->	23744 frames με ποσοστό	“	17,9%
7.4o cluster -->	11738 frames με ποσοστό	“	8,9%
8.5o cluster -->	6056 frames με ποσοστό	“	4,6%



**Εικόνα 4.54** Γραφήματα της ελεύθερης ενέργειας για τις κύριες συνιστώσες 1, 2, 3, οι οποίες καθορίζονται από τα χαρακτηριστικά ανόσματα (eigenvectors) σε ζευγάρια με κλίμακα -4,96 μέχρι +4,96 και ενεργειακό ελάχιστο 3.60 value kcal/mol.

Ως κυρίαρχο cluster θεωρούμε το cluster με τα περισσότερα frames μόνο που εδώ έχουμε 2 cluster με σχεδόν παρόμοιο occupancy. Ερευνώντας το περιεχόμενο του κάθε cluster που προέκυψε από τους δύο διαφορετικούς υπολογισμούς, στην εικόνα 4.54, συμπεραίνουμε ότι το cluster που θεωρούσε το cartesian ως κυρίαρχο, το dihedral το αναγνωρίζει ως δύο, χωρίζοντας το δεύτερο σε δύο μικρότερα κομμάτια, τα οποία όμως συνολικά αποτελούνται περίπου από τα ίδια frames. Το cosine content με την υψηλή τιμή του 0.6340, έδειξε ότι δεν έχουμε ικανοποιητικό sufficient sampling, οπότε θα πρέπει να μελετήσουμε τι συμβαίνει στην περίπτωση που αφαιρεθούν τα υπερκινητικά κατάλοιπα. Σε αυτά τα ευκίνητα κατάλοιπα ίσως να ευθύνεται και ο διαχωρισμός των cluster κατά αυτόν τον τρόπο. Γι' αυτό θα υπολογίσουμε PCA ξανά με περισσότερη έμφαση στα κύρια

τμήματα της δομής του μορίου χωρίς τα υπερκινητικά κατάλοιπα, ώστε να δούμε τις διαμορφώσεις των δομών κατά τη διάρκεια του trajectory.

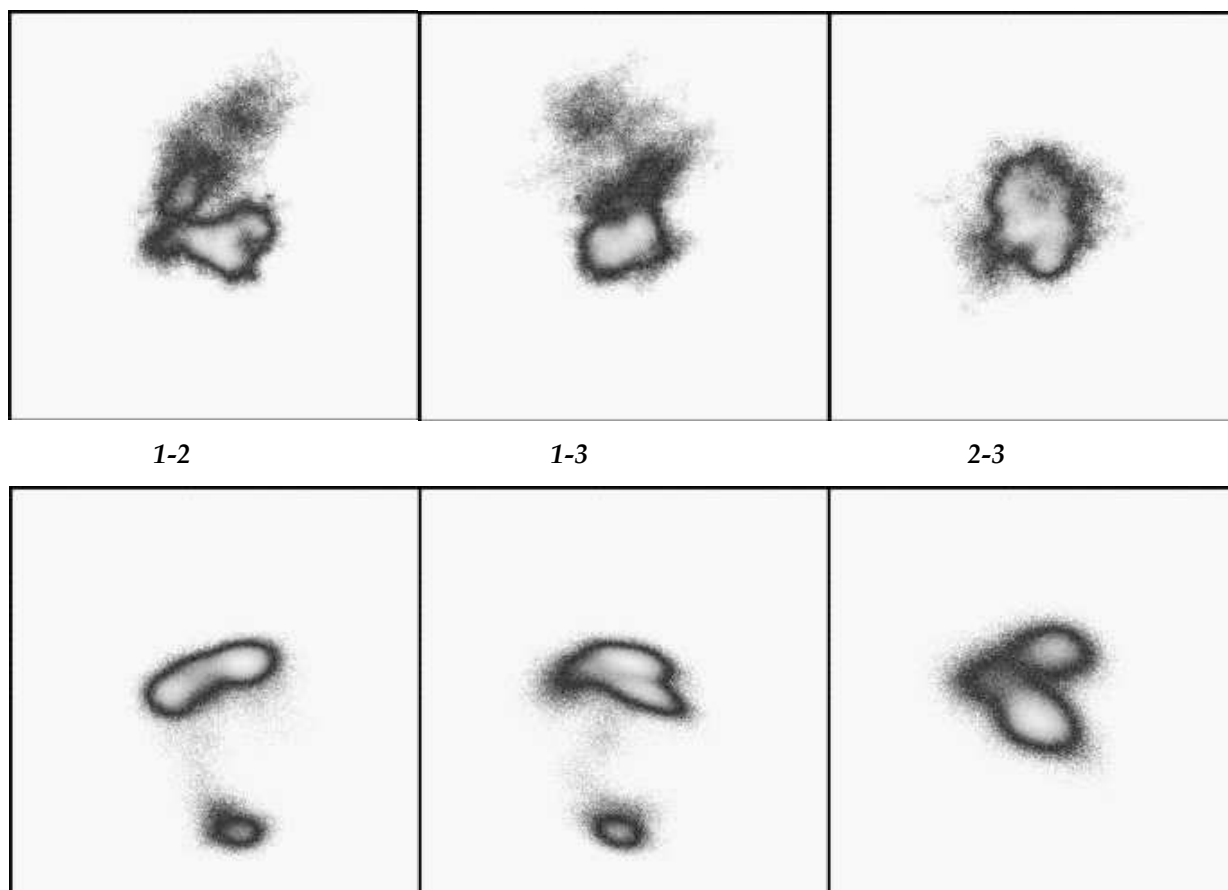


Εικόνα 4.55 Βλέπουμε την κατανομή των frames του κάθε cluster μέσω cartesian και dihedral PCA ανάλυση.

- **Cartesian και dihedral PCA ( χωρίς υπερκινητικά κατάλοιπα)**

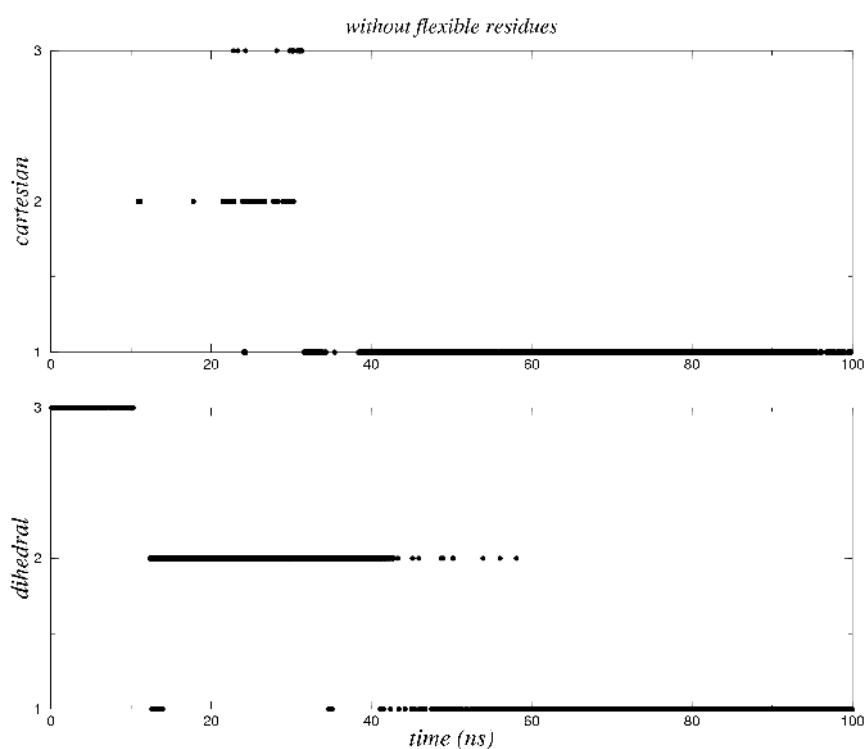
Αποκόβουμε, λοιπόν, τα κατάλοιπα εκείνα που έχουν rmsf μεγαλύτερο από 1,5 Å (20 κατάλοιπα), προκειμένου να μην αλλοιώνονται τα αποτελέσματα μας. Το PCA όπως και το dPCA, έπειτα από την αφαίρεση αυτών των καταλοίπων, έδειξαν την ύπαρξη 3 cluster. Το PCA υπολόγισε 3 cluster από τα οποία το κυρίαρχο είχε 71635 frames και κάλυπτε το 99,77% (occupancy) των συνολικών frames που ανήκουν σε cluster, ενώ τα άλλα δύο είχαν μόλις 0.2% και 0.03%. Στα διαγράμματα ΔG φαίνεται σχετικά ευδιάκριτα ότι υπάρχει ένα κυρίαρχο cluster με σχεδόν ικανοποιητικό sufficient sampling ενώ η τιμή του cosine content είναι 0.4652 και επιβεβαιώνει την υποψία μας.

Το Dihedral PCA έδειξε επίσης ότι υπάρχουν 3 cluster από τα οποία το κυρίαρχο είναι το πρώτο και αυτό με το οποίο θα ασχοληθούμε, αποτελείται από 56834 στιγμιότυπα και έχει occupancy 70.03%, ενώ τα άλλα δύο καλύπτουν το 29.53% και το 0.44% όσο αφορά το cosine content έχοντας την τιμή 0.4506 δείχνει επίσης αρκετά ικανοποιητικό sufficient sampling.



**Εικόνα 4.56** Γραφήματα της ελεύθερης ενέργειας για τις κύριες συνιστώσες 1, 2, 3, οι οποίες καθορίζονται από τα χαρακτηριστικά ανόσματα (*eigenvectors*) σε ζευγάρια 1-2, 1-3, 2-3. Στην πάνω τριάδα απεικονίζονται τα  $\Delta G$  γραφήματα όπως προκύπτουν από το *cartesian PCA* (με κλίμακα -37.47 έως +37.47 με ενεργειακό ελάχιστο  $\Delta G$  2.82 kcal/mol) και στη κάτω από το *dihedral*. ( με κλίμακα +2.97 -2.97 με ενεργειακό ελάχιστο  $\Delta G$  2.97 kcal/mol.)

### PCA analysis

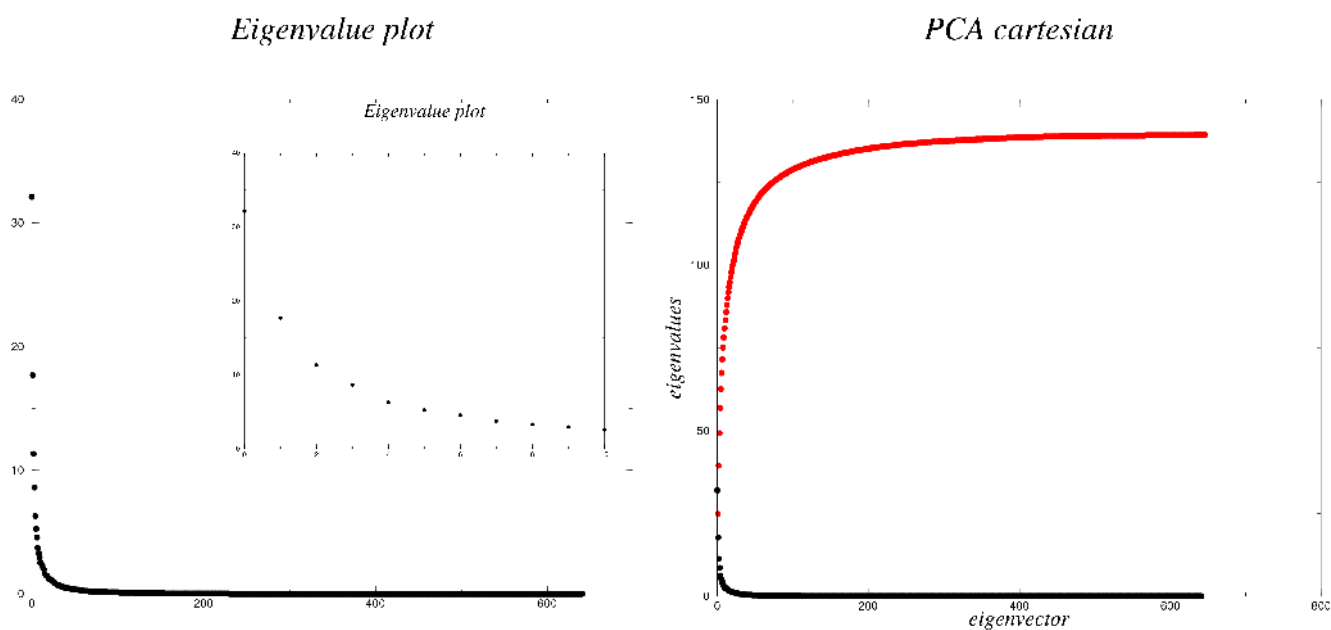


**Εικόνα 4.57** Απεικόνιση των *frames* που αντιστοιχούν σε κάθε *cluster*.

Συγκεντρώνοντας τα αποτελέσματα των υπολογισμών έπειτα από την αφαίρεση των 20 υπερκινητικών καταλοίπων και οι δύο μέθοδοι έδειξαν την ύπαρξη ενός μόνο κυρίαρχου cluster αποτελούμενο από περίπου ίδιο αριθμό στιγμιότυπων και έτσι συμφωνούν τα αποτελέσματα και με την εικόνα 4.48. Από εδώ και στο εξής οι υπολογισμοί μας θα επικεντρωθούν στην μελέτη της συμπεριφοράς του μορίου για το κυρίαρχο cluster από το cartesian-PCA είτε συμπεριλαμβάνονται τα υπερκινητικά κατάλοιπα είτε όχι ανάλογα με την ανάλυση.

- **Eigenvectors και eigenvalues**

Συνεχίζοντας την ανάλυση των κύριων συνιστωσών θα υπολογίσουμε τις χαρακτηριστικές τιμές (eigenvalues) και τα αντίστοιχα χαρακτηριστικά ανόσματα (eigenvectors) τους, χρησιμοποιώντας όλα τα κατάλοιπα ακόμα και τα υπερκινητικά. Οι τιμές των eigenvalues και eigenvectors είναι χρήσιμες για να μελετήσουμε το σύνολο των κινήσεων του μορίου διότι οι πρώτες κύριες συνιστώσες αρκούν ώστε να μας εξηγήσουν το μεγαλύτερο τμήμα της διακύμανσης όλων των eigenvectors όπως επισημαίνεται και στο διάγραμμα 4.58. Είναι φανερό για την δομή μας ότι οι 5 πρώτοι eigenvectors είναι ικανοί να περιγράψουν το 51.7% της συνολικής κίνησης του μορίου.

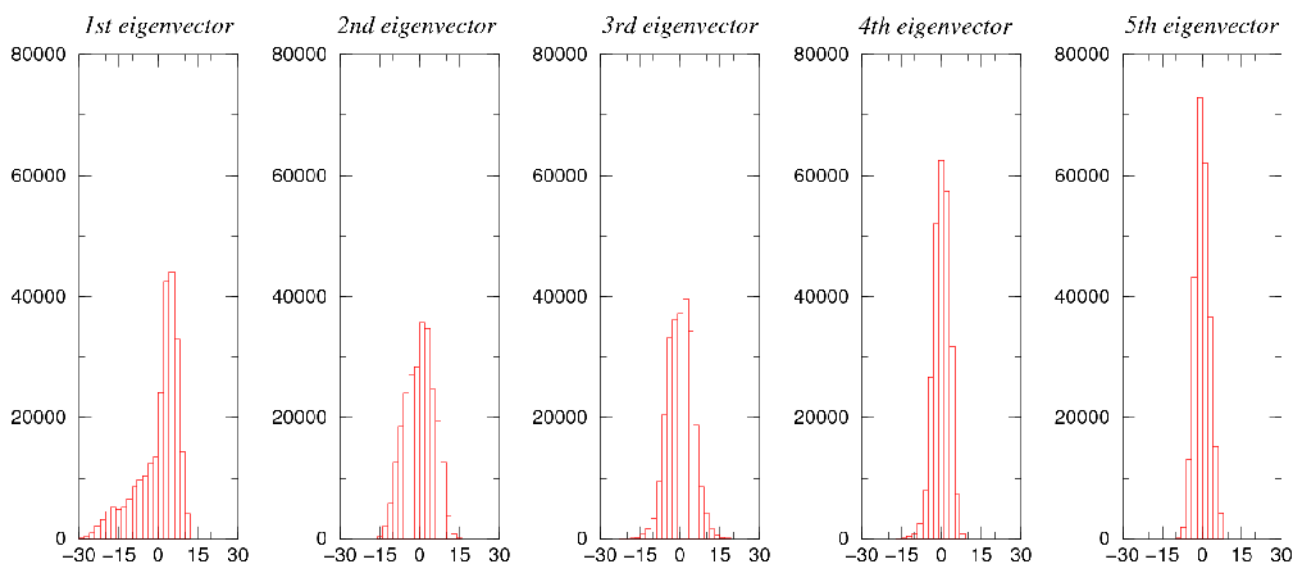


**Εικόνα 4.58 Αριστερά:** Απεικονίζονται οι eigenvectors και σε μεγέθυνση οι φαίνονται οι 10 πρώτοι.

**Δεξιά:** Οι eigenvectors και το ολοκλήρωμα των Eigenvalues

Θα επικεντρωθούμε στο κυρίαρχο cluster χωρίς τα υπερκινητικά κατάλοιπα. Μία ακόμη ένδειξη για την ύπαρξη ή όχι sufficient sampling είναι μέσω των διακυμάνσεων των 5 πρώτων eigenvector βάση της κατανομής τους στο επίπεδο. Στην περίπτωση όπου υπάρχει sufficient sampling τα ιστογράμματα που θα προκύψουν θα έχουν κανονική κατανομή γύρω από το μηδέν.

Παρατηρούμε στην εικόνα 4.59 τα ιστογράμματα των 5 πρώτων eigenvector από τους οποίους ο 1<sup>ος</sup> δεν εμφανίζει κανονική κατανομή αλλά εμφανίζει πολλές αρνητικές τιμές οι οποίες ανήκουν στο αρχικό μικρό cluster των 26000 frame. Οπότε ο χρόνος της προσομοίωσης δεν ήταν αρκετός για να έχουμε ικανοποιητικό δείγμα της κίνησης του μορίου σύμφωνα με τον 1ο eigenvector. Ωστόσο, όσο προχωράμε στη μελέτη των επόμενων eigenvectors τα ιστογράμματα τους εμφανίζουν κανονική κατανομή οπότε έχουν sufficient sampling.



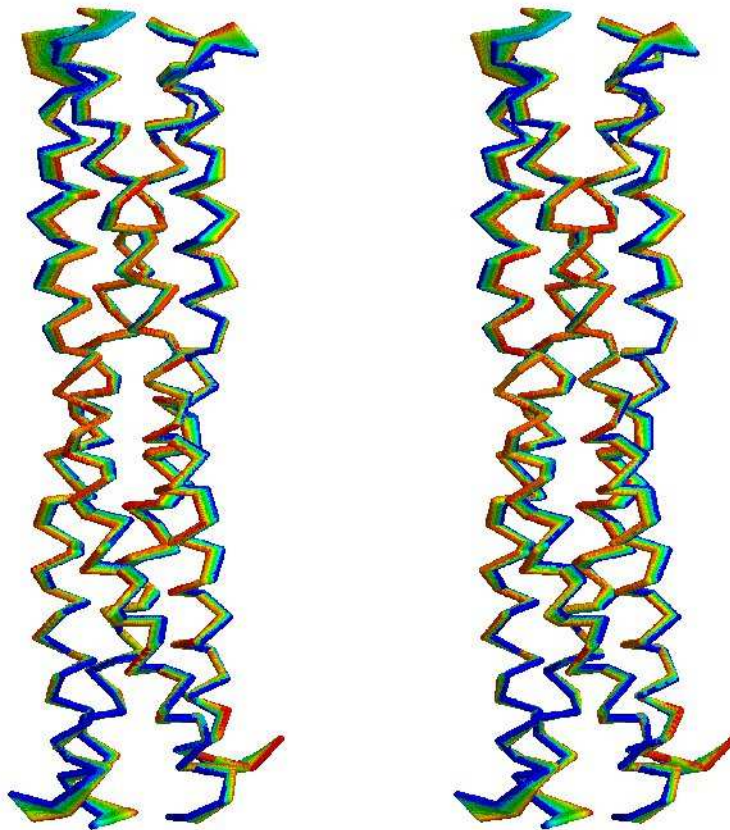
Εικόνα 4.59 Προβολή των διακυμάνσεων των 5 eigenvectors στη μία διάσταση

Όσο αφορά τις διακυμάνσεις των 3 πρώτων eigenvector σε 2 διαστάσεις (εικόνα 4.56) όπου απεικονίζονται ανά ζεύγη οι eigenvectors, θα μελετήσουμε τα  $\Delta G$  γραφήματα ελεύθερης ενέργειας που προέρχονται από την ανάλυση PCA. Στο γράφημα του ζεύγους 1-2 διακρίνονται 3 cluster χωρίς ξεκάθαρο sufficient sampling.



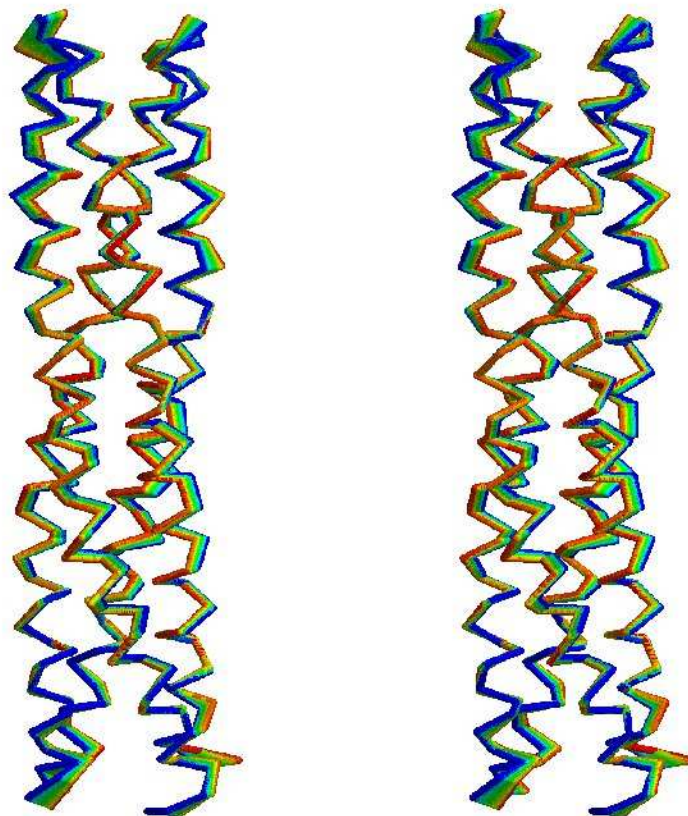
- **Προσδιορισμός της συνολικής κίνησης του μορίου**

Στην συνέχεια θα δείξουμε την κίνηση που πραγματοποιεί το μόριο μας με βάση τον πρώτο eigenvector, αφού έχει ικανοποιητικό sufficient sampling, κατασκευάζοντας ένα καινούργιο trajectory. Ο υπολογισμός έχει γίνει για το κυρίαρχο cluster χρησιμοποιώντας όλα τα άτομα (εικόνα 4.60) και έπειτα αφαιρώντας τα κατάλοιπα τα οποία γνωρίζουμε ότι κουνιούνται ιδιαίτερα πολύ (εικόνα 4.61) ώστε να φανεί η κίνηση του μορίου χωρίς να επηρεάζεται από τα υπερκινητικά αυτά κατάλοιπα.



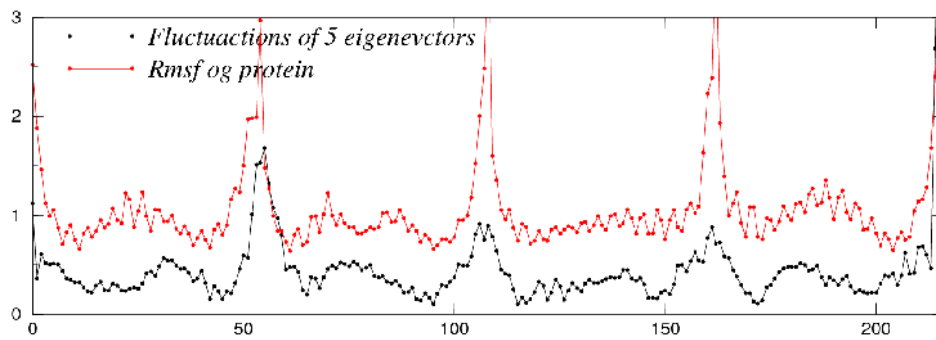
*Εικόνα 4.60 Αναπαράσταση της κίνησης του μορίου σύμφωνα με τον πρώτο eigenvector σε stereo μορφή. Η μεγαλύτερη κίνηση παρατηρείται στα άκρα των ελίκων.*

Παρατηρούμε ότι η δομή αυτή σύμφωνα με τον πρώτο eigenvector εμφανίζει σε όλο το ήκος της μια μικρή κινητικότητα με ιδιαίτερη αύξηση στα κατάλοιπα των άκρων, ενώ δεν φαίνεται να υπάρχει κάποια ιδιαίτερη αύξηση στο εύρος της κίνησης κάποιου καταλοίπου στο εσωτερικό. Ωστόσο, με την αφαίρεση των υπερκινητικών καταλοίπων των άκρων (εικόνα 4.61) διακρίνεται η κίνηση του υπόλοιπου μορίου χωρίς την επιρροή των άκρων. Παρατηρούμε, λοιπόν, ότι η κίνηση για το υπόλοιπο μόριο εξακολουθεί να είναι στον ίδιο βαθμό.

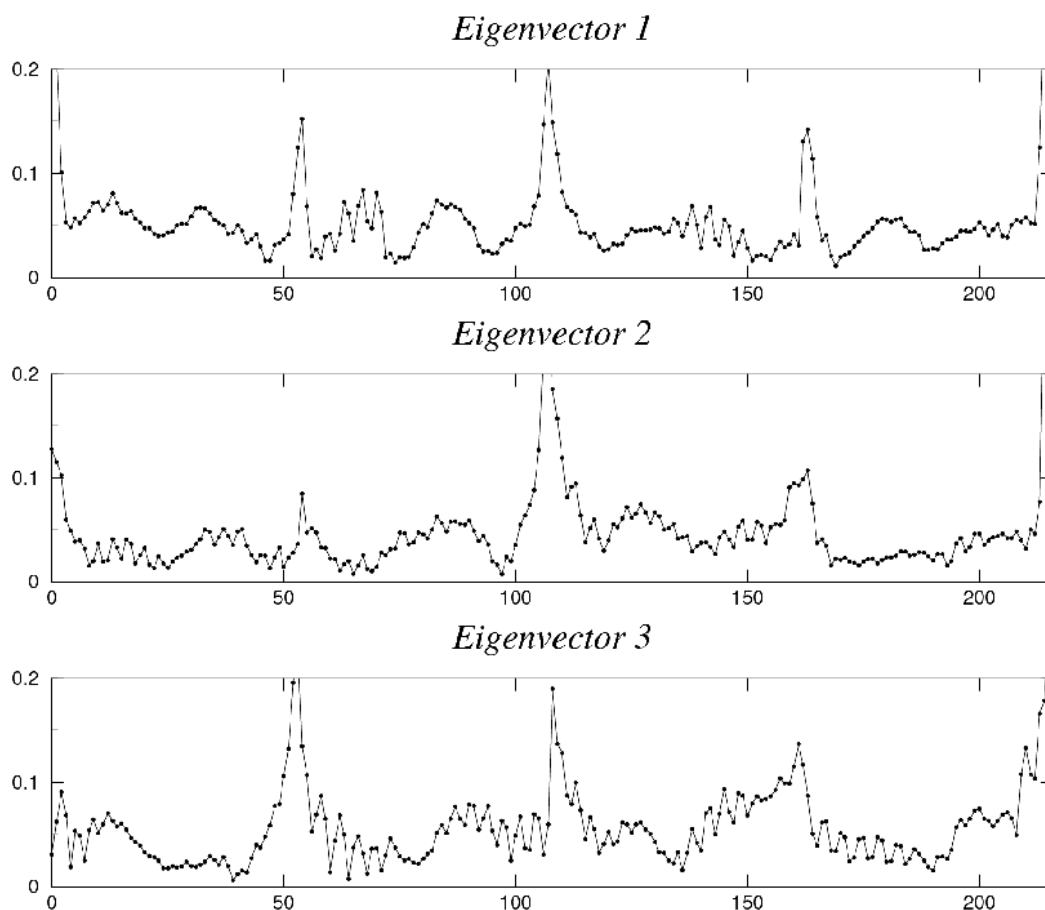


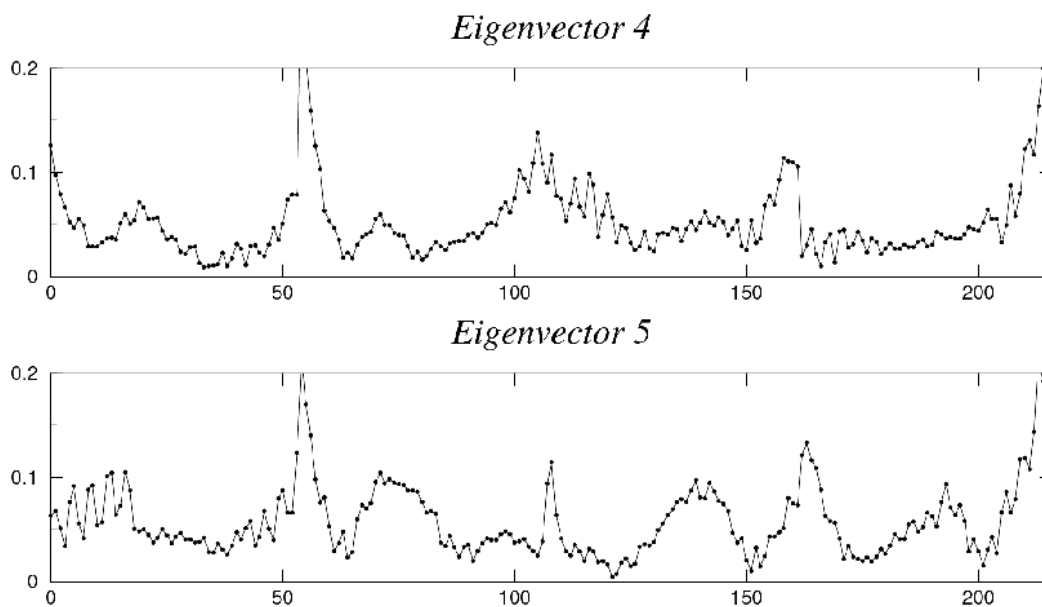
*Εικόνα 4.61 Αναπαράσταση της κίνησης του μορίου στο κυρίαρχο cluster σε Stereo μορφή, χωρίς τα υπερκίνητικα κατάλοιπα των άκρων.*

Τέλος, έχουμε την δυνατότητα να αναπαραστήσουμε την κατανομή των διακυμάνσεων του κάθε eigenvector ανά κατάλοιπο σε ένα γράφημα για τον καθένα ξεχωριστά (εικόνα 4.63) βλέποντας ποια κατάλοιπα συμμετέχουν σε κάθε κίνηση αλλά και στο σύνολό τους (εικόνα 4.62) παρατηρώντας αθροιστικά την διακύμανση των eigenvectors για κάθε κατάλοιπο. Τις διακυμάνσεις μπορούμε επίσης να τις παρατηρήσουμε και με την παράμετρο  $rmsf$  η οποία είναι ισοδύναμη με την τετραγωνική ρίζα της αθροιστικής διακύμανσης των eigenvector (Hess 2002). Παρατηρούμε ότι οι διακυμάνσεις των eigenvector συμφωνούν με τις ατομικές κινητικότητες του τετραμερούς σε μεγάλο βαθμό. Συγκρίνοντας τις δυο καμπύλες ο βαθμός γραμμικής συσχέτισης τους είναι 0,64. Όπως παρατηρήσαμε ήδη η μεγαλύτερη κινητικότητα εμφανίζεται στα άκρα κάθε μονομερούς αλλά και στο κεντρικό τμήμα των αλυσίδων. Η κατανομή των διακυμάνσεων ανά κατάλοιπο για τον κάθε eigenvector αναπαρίσταται στην εικόνα 4.63



**Εικόνα 4.62** Γραφική αναπαράσταση των διακομάνσεων των 5 πρώτων *eigenvector* αθροιστικά ανά κατάλοιπο (μαύρη γραμμή) και τα *rmsf* κόκκινη γραμμή. Για να είναι τα δεδομένα μας συγκρίσιμα και να μην χάνεται πληροφορία από τις διακομάνσεις λόγω κλίμακας, μετατοπίστηκε η καμπύλη των διακομάνσεων στον άξονα  $\psi$  αφού πολλαπλασιαστική η τιμή της με την σταθερά 4. Τα κατάλοιπα με *rmsf* μεγαλύτερο του 3 τα οποία δεν φαίνονται στο γράφημα είναι το Glu 109 (C μονομερές), Glu 163 και Phe 216 (D μονομερές)



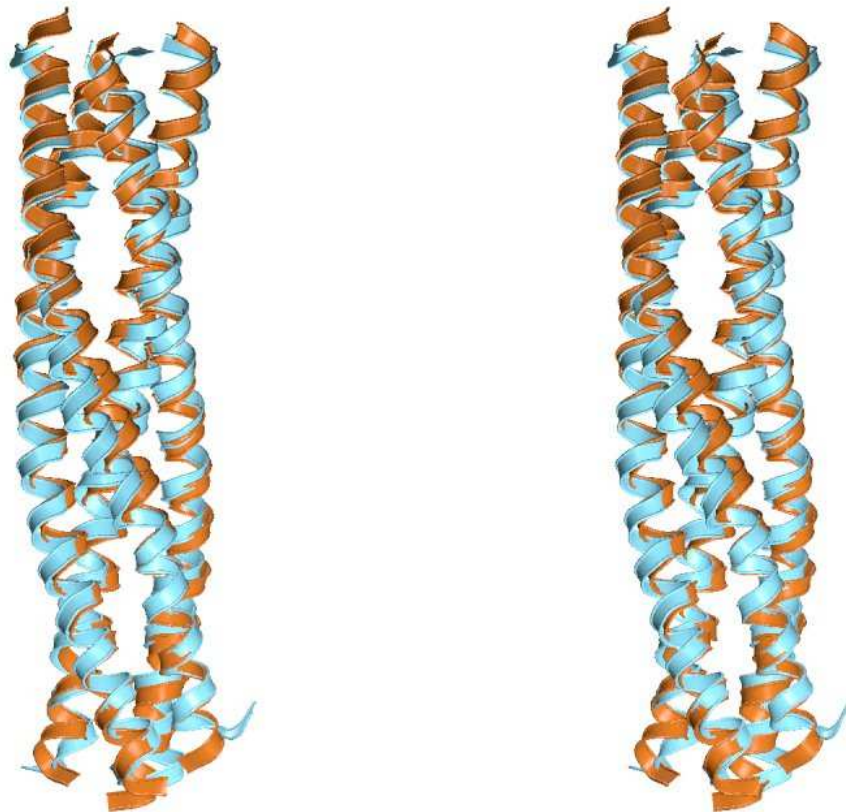


*Εικόνα 4.63 Αναπαράσταση των διακομάνσεων των 5 πρώτων eigenvector ανά κατάλοιπο*

#### 4.2.12 Σύγκριση των μέσων δομών από τα κυρίαρχα cluster (Cartesian και Dihedral)

Στην προηγούμενη ενότητα προσδιορίστηκε ο αριθμός των cluster, εδώ θα βρούμε την μέση δομή για το καθένα και θα τις συγκρίνουμε μεταξύ τους αλλά και με την αρχική δομή. Η σύγκριση των δομών των κυρίαρχων cluster από το cartesian και από το dihedral PCA δεν έχει ιδιαίτερο νόημα να γίνει σχηματικά καθώς γνωρίζουμε από την εικόνα 4.56 ότι τα cluster αποτελούνται από τα ίδια frames συνεπώς δεν θα εμφανίζεται καμία σημαντική διαφορά ανάμεσα στις δύο δομές.

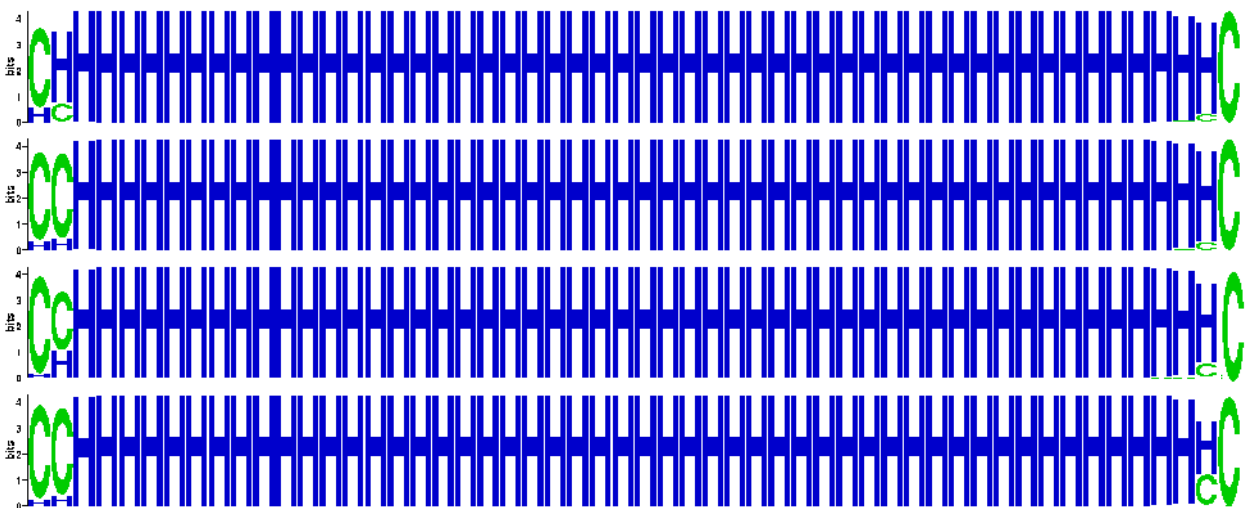
Με την σύγκριση μεταξύ της μέσης δομής από το κυρίαρχο cluster που έδωσε το Cartesian (γαλάζιο χρώμα) και της αρχικής δομής ( πορτοκαλί χρώμα) πριν την εξέλιξη των προσομοιώσεων, παρατηρούμε (εικόνα 4.64) ότι εμφανίζονται κάποιες δομικές διαφορές. Οι αλυσίδες είναι μετατοπισμένες σε όλο το μήκος τους χωρίς να έχουν αποκλίνει από την αρχική τους ευθεία και τα ακραία κατάλοιπα έχουν διαφορετικούς προσανατολισμούς. Η τυπική απόκλιση των δύο δομών είναι 2,25Å.



**Εικόνα 4.64** Στερεοδιάγραμμα της μέσης δομής του κοριαρχου cluster από το Cartesian (γαλάζιο χρώμα) και η αρχική δομή του Sigmoidal (πορτοκαλί χρώμα) σε υπέρθεση. Έχει γίνει στοίχιση των δομών στα κατάλοιπα 1- 216 με απόκλιση Rmsd να είναι αρκετά μεγάλη 2.25 Å.

#### 4.2.13 Συντήρηση δευτεροταγούς δομής

Επιθυμώντας να ερευνήσουμε το βαθμό στον οποίο συντηρήθηκε η δευτεροταγή δομή κατά τη διάρκεια της προσομοίωσης χρησιμοποιήσαμε το πρόγραμμα weblogo το οποίο δηλώνει την δευτεροταγή δομή για κάθε κατάλοιπο της έλικας σχέδια: Η για έλικες και C για θηλιές.



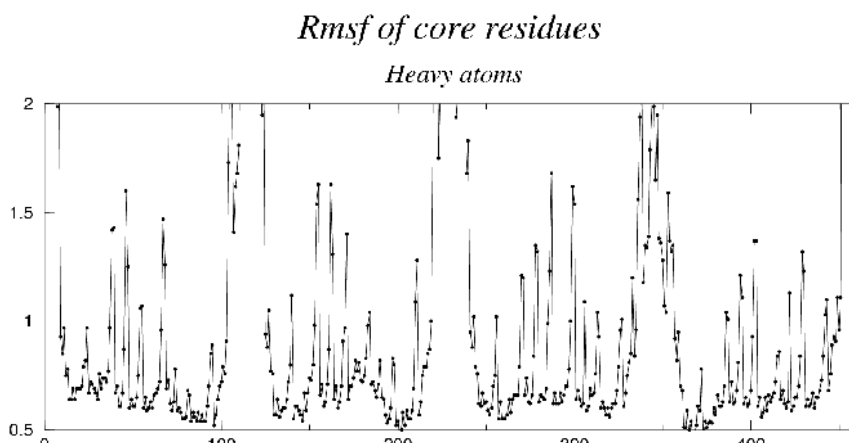
**Εικόνα 4.65** Συντήρηση δευτεροταγούς δομής για κάθε θέση της αλληλουχίας του κάθε μονομερούς.

Απεικονίζονται στους άξονες τα μονομερή με τα αντίστοιχα 54 κατάλοιπα της αλληλουχίας τους. Οι α-έλικες διατηρούνται καθ' όλη τη διάρκεια της προσομοίωσης και είναι αρκετά σταθερές. Ωστόσο οι περιοχές στις οποίες σχηματίζονται θηλιές και όχι σταθερή δομή είναι στα άκρα των 4 μονομερών και συγκεκριμένα χάνεται η ελικοειδή μορφή των καταλοίπων 1 (Glu) στο N τελικό άκρο της A και του καταλοίπου 54 (Phe) στο C άκρο της. Στη B αλυσίδα χάνεται η ελικοειδής μορφή στα κατάλοιπα 1, 2 του N-τελικού άκρου και 54 του C-τελικού, όπως το ίδιο συμβαίνει και στις αλυσίδες C και D.

#### 4.2.14. Ανάλυση διεδρων γωνιών ( $\chi$ -value)

Ολοκληρώντας την μελέτη της δομής αυτής, θα ασχοληθούμε με την σταθερότητα και τη διαμόρφωση του υδρόφοβου πυρήνα. Για το σκοπό αυτό θα αναλύσουμε την κινητικότητα των πλευρικών ομάδων των καταλοίπων του υδρόφοβου πυρήνα και την κατανομή των rotamers των καταλοίπων αυτών στην εκάστοτε δομή.

Για την κινητικότητα των καταλοίπων του, θα χρησιμοποιήσουμε την παράμετρο rmsf των αντίστοιχων καταλοίπων η οποία σχηματικά αναπαρίσταται στην εικόνα 4.66. Συγκρίνοντας τις ατομικές κινητικότητες των ατόμων όλων των καταλοίπων (Ca άτομα) (εικόνα 4.42) με αυτών των υδρόφοβων καταλοίπων (heavy άτομα) παρατηρούμε ότι η κινητικότητα των υδρόφοβων καταλοίπων είναι αρκετά υψηλή. Ιδιαίτερα ευκίνητα φαίνεται να είναι τα άτομα των άκρων του υδρόφοβου πυρήνα του δεματίου, σε σχέση με τα κατάλοιπα όλου του σκελετού με μέση τιμή 1,07 Å



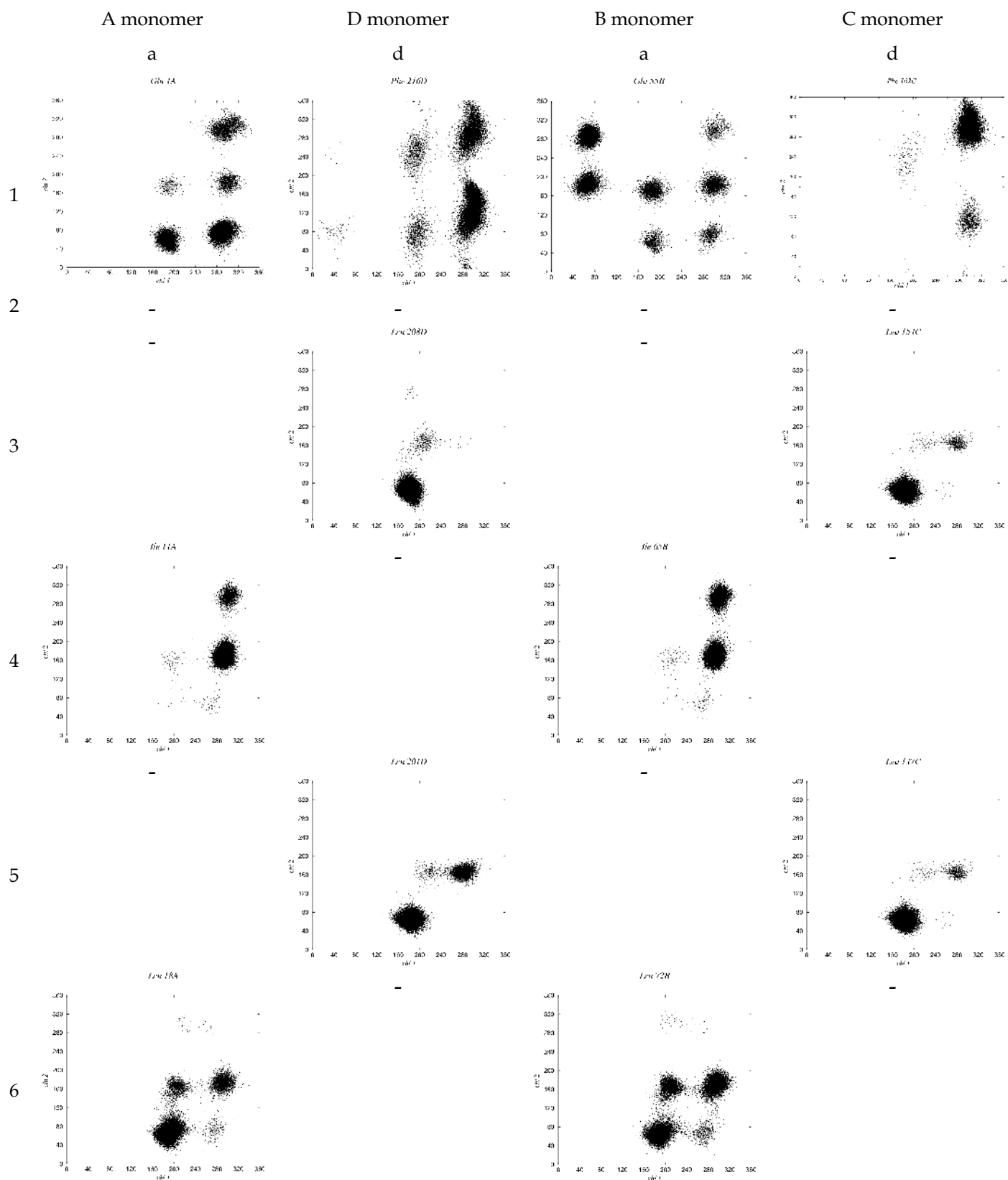
Εικόνα 4.66 Παρουσιάζεται στο γράφημα η κινητικότητα των πλευρικών ομάδων των καταλοίπων του υδρόφοβου πυρήνα.

Για την μελέτη της κατανομής των rotamers των καταλοίπων του υδρόφοβου πυρήνα θα χρειαστεί ο υπολογισμός των διεδρών γωνιών των πλευρικών ομάδων  $\chi_1$  και  $\chi_2$  οι οποίες όπως αναφέραμε στην παράγραφο 4.1.14 περιγράφουν τις περιστροφές γύρω από τους απλούς δεσμούς των πλευρικών ομάδων οπότε πληροφορούν για το ποιες είναι οι επιτρεπόμενες διαμορφώσεις που μπορούν να πάρουν και ποιές όχι, λόγω στερικών παρεμποδίσεων. Θα υπολογίσουμε λοιπόν τις διεδρες  $\chi_1$ ,  $\chi_2$  γωνίες για κάθε κατάλοιπο κατά τη διάρκεια του trajectory ώστε να προσδιορίσουμε τα rotamers από τα οποία περνάει το κάθε κατάλοιπο. Η γωνία  $\chi_1$  καθορίζεται από τα άτομα N-C $\alpha$ -C $\beta$ -C $\gamma$  και η  $\chi_2$  καθορίζεται από τα άτομα C $\alpha$ -C $\beta$ -C $\gamma$ -C $\delta_1$  για Leu και Phe, ενώ για Ile C $\alpha$ -C $\beta$ -C $\gamma_1$ -C $\delta$ .

Στο μόριο μας 64 κατάλοιπα από τα 216 ανήκουν στα 16 επίπεδα που συγκροτούν τον υδρόφοβο πυρήνα και είναι αλανίνες (Ala), φαινυλαλανίνες (Phe), κυστεΐνες (Cys), γλουταμικά οξέα (Glu), λευκίνες (Leu), ισολευκίνες (Ile) και θρεονίνες (Thr). Από τα κατάλοιπα αυτά θα υπολογίσουμε  $\chi_1$  διεδρες γωνίες για όλα εκτός από τις αλανίνες και  $\chi_2$  γωνίες μόνο για τα κατάλοιπα Leu, Ile, Phe και Glu.

Έπειτα από σύγκριση των rotamers που προκύπτουν από τις προσομοιώσεις (εικόνα 4.67) με τα rotamers που προσδιορίστηκαν κρυσταλλογραφικά (εικόνα 4.34, σελίδα 98) για τα κατάλοιπα αυτά, παρατηρούμε ότι κάποια συμφωνούν, ενώ κάποια δεν εμφανίζουν όλα τα προσδιορισμένα rotamers. Πιο συγκεκριμένα οι περισσότερες λευκίνες εμφανίζουν κάποια ή όλα τα προσδιορισμένα rotamers, όμως υπάρχουν κατάλοιπα λευκίνης που εμφανίζουν πιο σπάνιες διαμορφώσεις, όπως τα Leu 25A, Leu 130C και Leu 100B. Από τα γλουταμικά οξέα μόνο το Glu 163D αποκλείει διότι έχει το μοναδικό του rotamer στο λιγότερο συνηθισμένη θέση. Τέλος, όλα τα κατάλοιπα ισολευκίνης και φαινυλαλανίνης συμφωνούν με τα κρυσταλλογραφικά δεδομένα.

Plots of chi1-chi2 dihedral angles



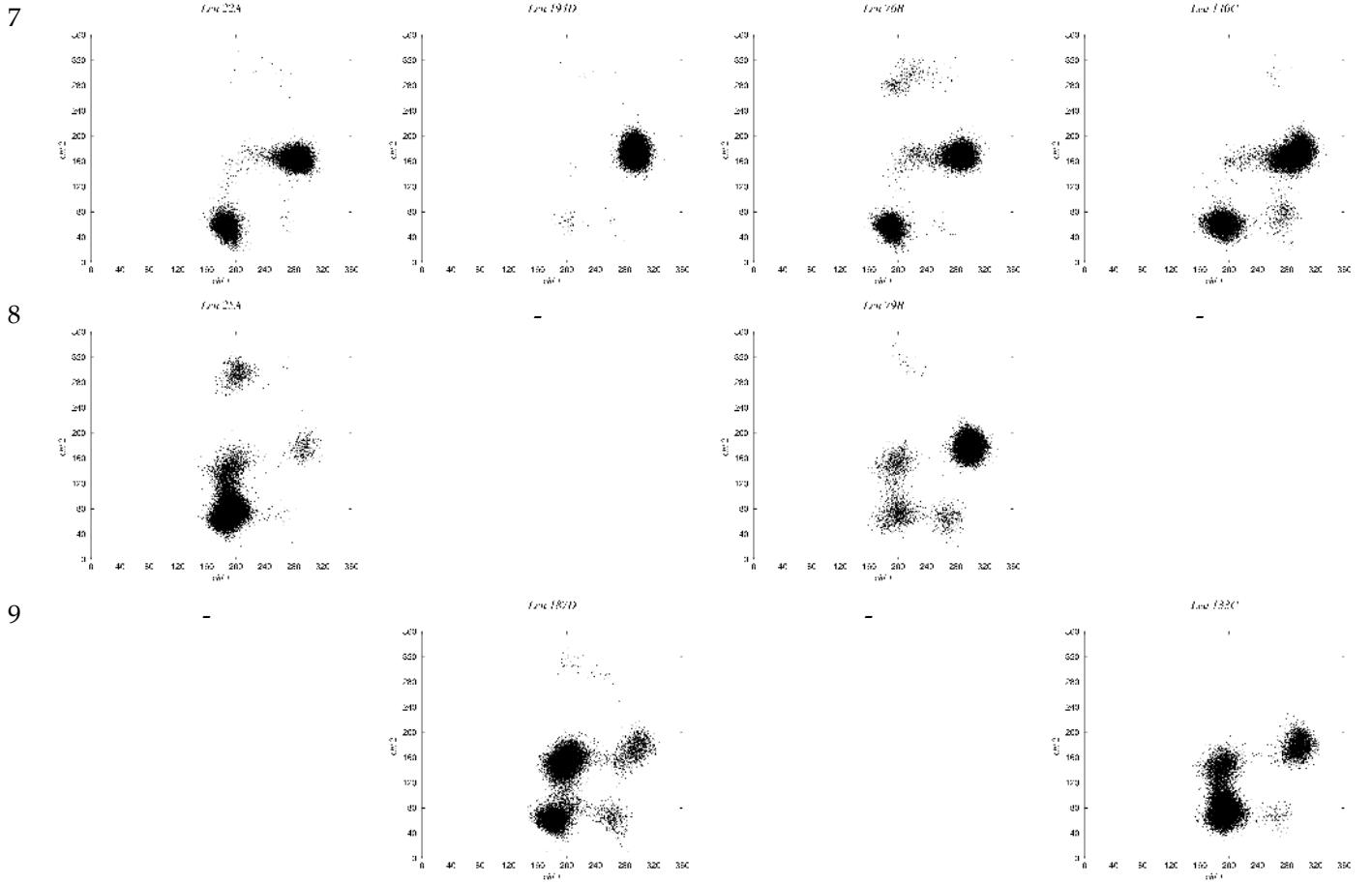


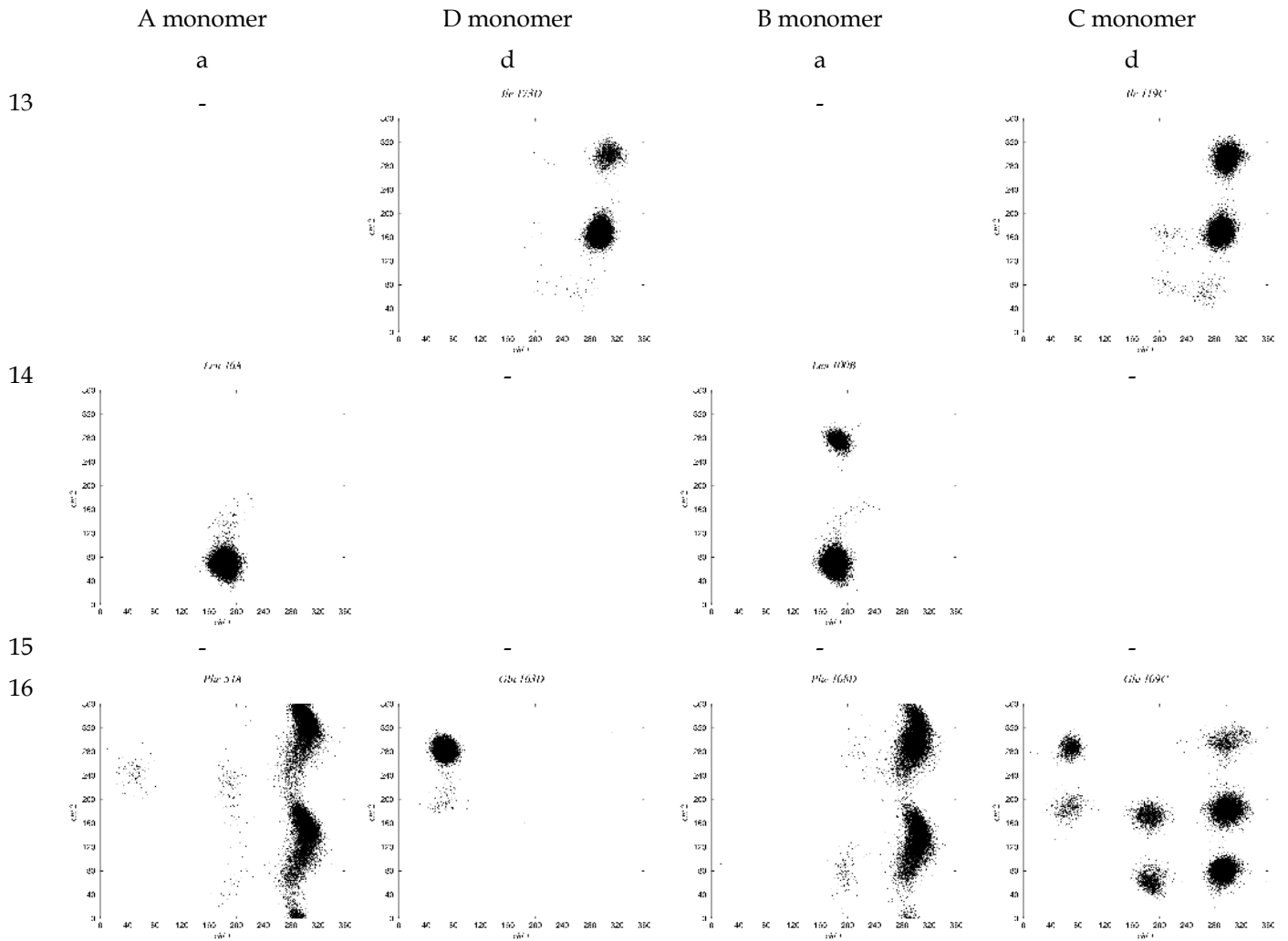
A monomer

D monomer

B monomer

C monomer

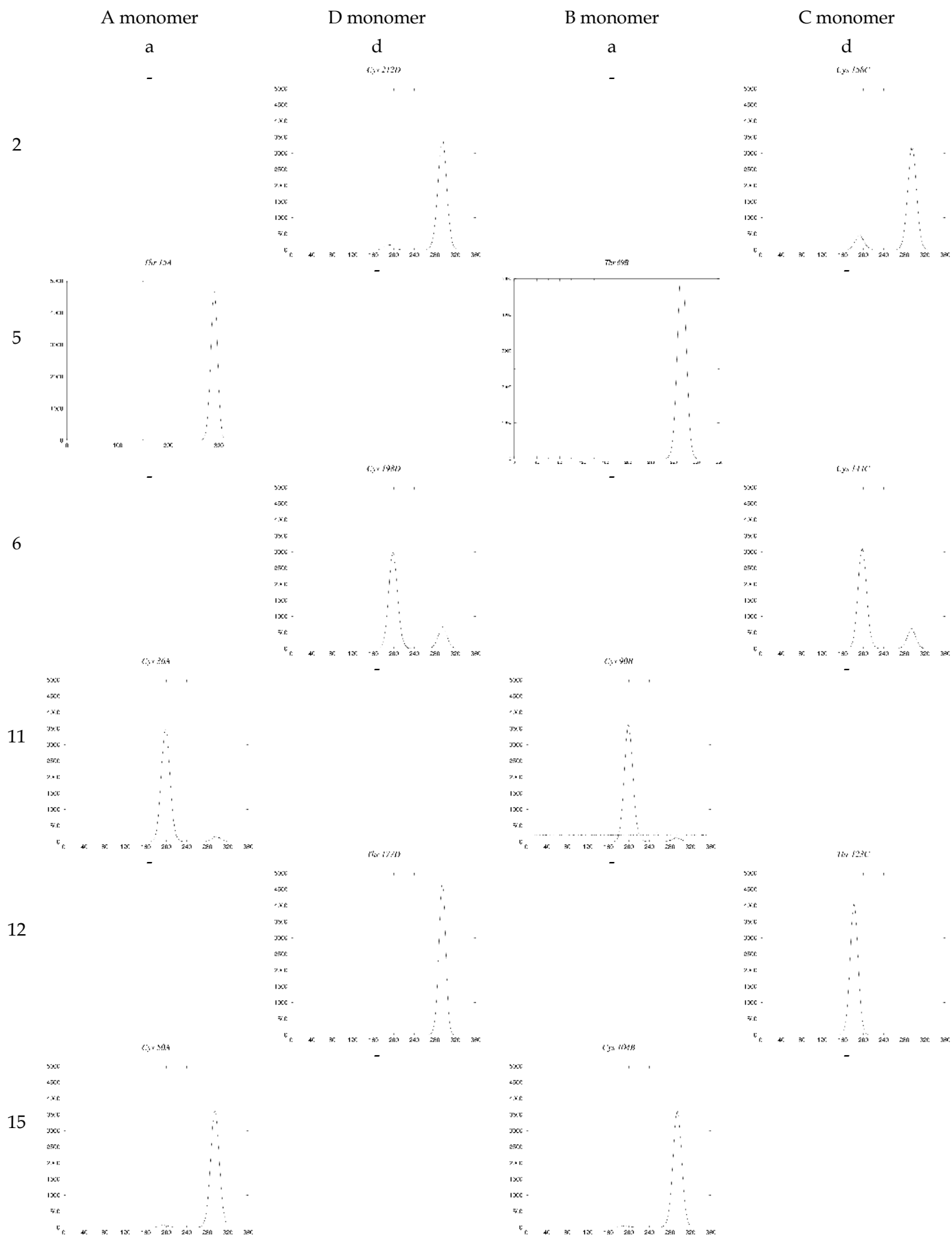




**Εικόνα 4.67** Κατανομή των διεδρων γωνιών  $\chi_1$ ,  $\chi_2$  μόνο των κατάλοιπων λευκίνης, ισολευκίνης, γλουταμικού οξέος και φαινυλαλανίνης του υδρόφοβου πυρήνα καθενός επιπέδου (layer), η οποία διαμορφώνεται κατά τη διάρκεια του trajectory. Τα κατάλοιπα είναι τοποθετημένα σύμφωνα τις θέσεις *a* και *d* που καταλαμβάνουν στον υδρόφοβο πυρήνα. Τα layer 2 και 15 αποτελούνται από κατάλοιπα Cys και Ala τα οποία δεν έχουν  $\chi_2$  γωνίες.

Τα κατάλοιπα θρεονίνης και κυστεΐνης δεν έχουν  $\chi_2$  γωνίες όμως έχουν  $\chi_1$  για τις οποίες κατασκευάστηκαν τα αντίστοιχα γραφήματα (εικόνα 4.68). Μετά από σύγκριση με τα κρυσταλλογραφικά προσδιορισμένα γραφήματα των γωνιών (εικόνα 4.36), συμπεραίνουμε ότι μόνο το κατάλοιπο Thr 123C εμφανίζει σπάνιο rotamer, ενώ τα υπόλοιπα συμφωνούν με τα κρυσταλλογραφικά δεδομένα.

Plots of chi1 dihedral angles



Εικόνα 4.68 Κατανομή των τιμών των διεδρων γωνιών  $\chi_1$  των καταλοίπων θρεονίνης και κυστεΐνης σε μία διάσταση κατά τη διάρκεια της προσομοίωσης για τις θέσεις όπου όπου βρίσκονται.

Επειδή οι πρωτεΐνες κατά τη διάρκεια της προσομοίωσης τείνουν να μεταπηδούν σε διαφορετικές διαμορφώσεις είναι αδύνατον να πάρουμε μια ολοκληρωμένη εικόνα του διαθέσιμου χώρου με τους εν δυνάμει υπολογιστές μας. Ακόμα και όταν η πρωτεΐνη μένει σε μια διαμόρφωση κατά τη διάρκεια της προσομοίωσης, μια μεταπήδηση μπορεί να συμβεί εάν η προσομοίωση είχε επεκταθεί. Ο μόνος τρόπος για να πετύχουμε sufficient sampling και αξιοπιστα αποτελέσματα σε μια μικρής διάρκειας προσομοίωση είναι πραγματοποιώντας μια μεγαλύτερη.

## Κεφάλαιο 5

### Σύγκριση αποτελεσμάτων ανάλυσης μεταξύ των δύο μοντέλων πλεξούδα και sigmoidal

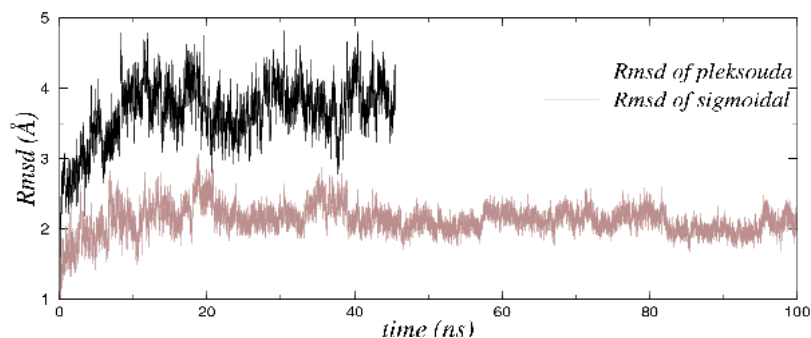
Μέχρι τώρα έχουμε αναφερθεί αναλυτικά στην μελέτη της δομής και της δυναμικής των δύο μοντέλων, που σχεδιάστηκαν αρχικά ως πιθανές διαμορφώσεις για την πρωτεΐνη 2aaQL, με απώτερο σκοπό τον καθορισμό της πιο σταθερής δομής. Με την σύγκριση των αποτελεσμάτων των δύο μοντέλων που ακολουθεί θα μελετήσουμε τη σταθερότητα και τις κινητικές τους ιδιότητες.

Ξεκινώντας τη σύγκριση απαιτείται να επισημάνουμε ότι ο χρόνος προσομοίωσης δεν ήταν ο ίδιος και για τα δύο μοντέλα. Για την πλεξούδα ήταν 45.7ns και για το sigmoidal 100ns. Επίσης για να γίνουν οι αναλύσεις έχουν αφαιρεθεί όλες οι μεταθέσεις και οι μετατοπίσεις που συνέβησαν κατά τη διάρκεια της προσομοίωσης με τη βοήθεια του προγράμματος carma.

Αρχικά, με την ανάλυση Rmsd από την αρχική δομή για όλα τα Ca άτομα του σκελετού (εικόνα 4.69) μπορούμε να συμπεράνουμε ότι το sigmoidal είναι πιο σταθερό μόριο από ότι η πλεξούδα. Το rmsd της πλεξούδας είναι αρκετά υψηλό, γύρω στο 3.66Å με έντονες διακυμάνσεις 0.44Å, ενώ φαίνεται πως όλες οι δομές ομαδοποιούνται σε ένα cluster συμπεριλαμβανομένου και των αρχικών 210 frames του heating. Από την άλλη, το μοντέλο sigmoidal φαίνεται να είναι πιο σταθερό με rmsd πολύ μικρότερο 2.12Å και διακυμάνσεις επίσης μικρές 0.20Å. Συγκεκριμένα η τιμή του αυξάνεται για τα πρώτα 18ns, έπειτα όμως για τον υπόλοιπο χρόνο της προσομοίωσης διατηρείται σταθερό.

Διακρίνεται, λοιπόν, μια μικρή διαφοροποίηση των δομών σε 2 cluster, οι δομές στην αρχή της προσομοίωσης, όπου αυξάνεται η τυπική απόκλιση και έπειτα το υπόλοιπο πολύ σταθερό τμήμα.

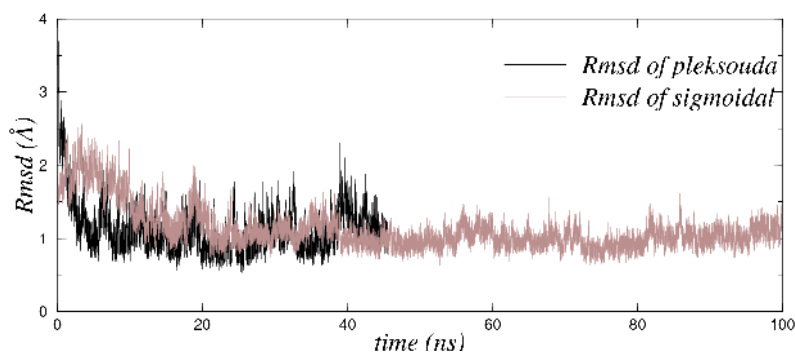
### *Rmsd from starting*



**Εικόνα 4.69** Καμπύλη της τοπικής απόκλισης *Rmsd* από την αρχική δομή συναρτήσει του χρόνου για τα άτομα του σκελετού (Ca άτομα) των μοντέλων πλεξούδας (μαύρο) και sigmoidal (καφέ).

Στη συνέχεια θα μελετήσουμε κατά πόσο μεταβάλλεται το rmsd από τη μέση δομή κατά τη διάρκεια του trajectory, ώστε να διαπιστώσουμε τη σταθερότητα της δομής. Στην εικόνα 4.70 παρατηρούμε ότι και τα δύο μοντέλα βρίσκουν την μέση δομή σχετικά γρήγορα με rmsd 1.12 Å, μόνο που η πλεξούδα έχει διακύμανση 0,34, ενώ το sigmoidal 0,29Å. Συμπεραίνουμε, λοιπόν, ότι δεν αποκλίνουν ιδιαίτερα πολύ από τη μέση δομή οπότε είναι αρκετά σταθερές.

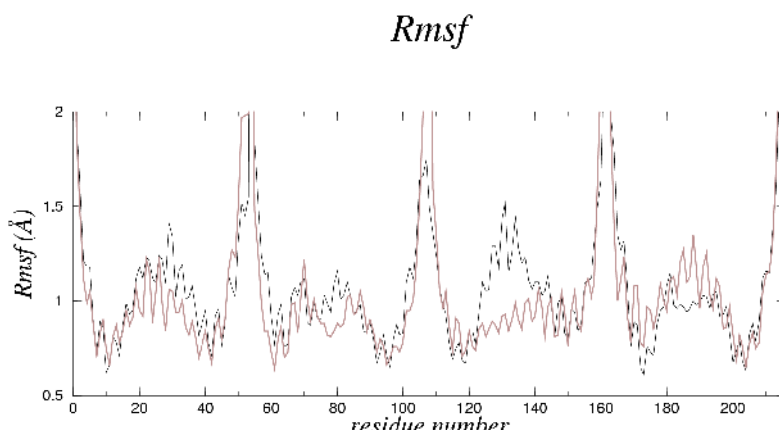
### *Rmsd from average*



**Εικόνα 4.70** Καμπύλη τοπικής απόκλισης *Rmsd* από την μέση δομή των μοντέλων πλεξούδας και sigmoidal.

Ακολούθως, θα ασχοληθούμε με τα στοιχεία που μας δίνει η παράμετρος rmsf για την κινητικότητα των ατόμων του εκάστοτε καταλοίπου των δυο μοντέλων. Όπως φαίνεται στην εικόνα 4.71 οι δύο δομές εμφανίζουν παρόμοιες κινητικότητες με υψηλές

τιμές στα άκρα των μονομερών όπως ήταν φυσικό, αλλά και με μία ασυνήθιστη αύξηση στο εσωτερικό τμήμα του κάθε μονομερούς, η οποία οφείλεται στην εισαγωγή των επτά καταλοίπων κατά την μοντελοποίηση. Ωστόσο, πρέπει να σημειώσουμε ότι οι τιμές της πλεξούδας είναι πιο υψηλές από ότι του sigmoidal, εκτός από το μονομερές D και τα τελικά κατάλοιπα του καρβοξυτελικού άκρου της B και του αμινοτελικού της C του sigmoidal τα οποία εμφανίζουν μεγαλύτερες τιμές (παράγραφος 4.1.4).



**Εικόνα 4.71** Rmsf των Ca ατόμων του σκελετού των μοντέλων σε σχέση με τη θέση τους στη μέση δομή. Απεικόνιση της κινητικότητας των Ca ατόμων της πρωτεΐνης σύμφωνα με την παράμετρο Rmsf. Χρησιμοποιούνται τα ίδια χρώματα με τις προηγούμενες εικόνες δηλαδή μαύρο το μόριο της πλεξούδας και καφέ του sigmoidal.

Συνεχίζοντας την σύγκριση των δύο μορίων χρήσιμο θα ήταν να ελέγξουμε τη συμπεριφορά των καταλοίπων του υδρόφοβου πυρήνα μελετώντας την κινητικότητα τους, καθώς επίσης να μελετήσουμε και τις διαμορφώσεις που λαμβάνουν οι πλευρικές ομάδες τους (εικόνες 4.34, 4.36, 4.67 και 4.68). Μελετώντας την κινητικότητα των πλευρικών ομάδων των καταλοίπων (rmsf καταλοίπων υδρόφοβου πυρήνα) που ανήκουν στον πυρήνα, συμπεραίνουμε ότι η κινητικότητα των υδρόφοβων καταλοίπων του μοντέλου πλεξούδας (εικόνα 4.32), είναι αρκετά υψηλή, rmsf 1.17Å και διακύμανση 1.00Å, με ιδιαίτερα ευκίνητα να είναι τα άτομα των τελικών καταλοίπων των μονομερών καθώς βρίσκονται στην άκρη του δεματίου και δεν συμμετέχουν στον σχηματισμό σταθερής δομής. Από την εικόνα 4.66 συμπεραίνουμε ότι παρόμοια γεγονότα συμβαίνουν και στον πυρήνα του Sigmoidal με μικρότερη Rmsf τιμή γύρω στα 1.07Å και διακύμανση 0,87Å.

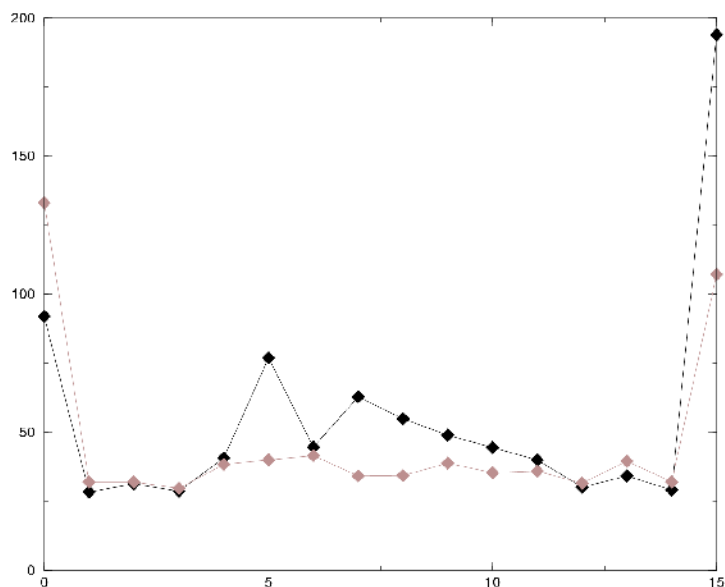
Από την κατανομή των rotamers των αντίστοιχων καταλοίπων ανάλογα με τις διαμορφώσεις που παίρνουν αλλά και το χρονικό διάστημα παραμονής τους στο εκάστοτε

πιθανό rotamer συμπεραίνουμε τον βαθμό σταθερότητας του πυρήνα. Σύμφωνα με τα γραφήματα των διεδρων γωνιών (4.35, 4.37, 4.67 και 4.68) μπορούμε να ενημερωθούμε για το αν εμφανίζεται μικρός αριθμός από διαφορετικά πιθανά rotamers, τότε πιθανόν οι πλευρικές ομάδες δεν έχουν το περιθώριο να περιστρέφονται ελεύθερες, οπότε αναπτύσσονται αλληλεπιδράσεις μεταξύ γειτονικών καταλοίπων, με αποτέλεσμα το ισχυρό πακετάρισμα τους και έτσι πετυχαίνεται σταθερότητα. Από την άλλη όμως, εάν εμφανίζονται πολλές διαφορετικές διαμορφώσεις σημαίνει ότι έχουν δημιουργηθεί κενά στην δομή, cavities, οπότε ο πυρήνας εμφανίζεται ασταθής. Συγκεκριμένα στον πυρήνα της πλεξούδας παρατηρούμε τις πλευρικές ομάδες των εκάστοτε καταλοίπων να αποκτούν τα πιο συνήθη rotamers, ενώ εννέα από αυτά φέρουν σπάνιες διαμορφώσεις. Όσο αφορά το μοντέλο της sigmoidal, σχεδόν όλες οι πλευρικές ομάδες του πυρήνα εμφανίζουν της κρυσταλλογραφικά προσδιορισμένες διαμορφώσεις εκτός από 5 κατάλοιπα (αναλυτικά στις αντίστοιχες παραγράφους). Οπότε συμπεραίνουμε ότι ο πυρήνας του sigmoidal είναι πιο καλά πακεταρισμένος.

Ιδιαίτερο ενδιαφέρον όμως, εμφανίζει το πακετάρισμα των 16 επιπέδων (layers) που απαρτίζουν τον πυρήνα. Όσο καλύτερα πακεταρισμένα είναι τα επίπεδα τόσο πιο σταθερός ο υδρόφοβος πυρήνας. Προς εκπλήρωση του σκοπού αυτού θα υπολογίσουμε το εμβαδό που καταλαμβάνουν τα layers που βρίσκονται στον πυρήνα με τη βοήθεια ενός προγράμματος γραμμένο σε γλώσσα προγραμματισμού perl (παράρτημα 5). Το πρόγραμμα αυτό υπολογίζει το εμβαδό του τετράπλευρου που σχηματίζεται από τα κατάλοιπα των θέσεων "a" και "d" του υδρόφοβου πυρήνα και συγκεκριμένα των ατόμων Cβ. Μπορούμε να υπολογίσουμε το εμβαδό του τετράπλευρου ως άθροισμα των εμβαδών των 2 τριγώνων τα οποία αποτελούνται από μια κοινή πλευρά την υποτείνουσα, δηλαδή την απόσταση των ατόμων στις "a" θέσεις. Από το μέγεθος του εμβαδού τους διακρίνονται οι αποστάσεις μεταξύ των πλευρικών αλυσίδων και σχηματίζουμε μια γενική άποψη για το βαθμό της μεταξύ τους αλληλεπίδρασης, ένδειξη σταθερότητας. Στην εικόνα 4.72 παρατηρούμε ότι το εμβαδό των επιπέδων την πλεξούδας έχουν μεγαλύτερη τιμή από τα επίπεδα του sigmoidal που σημαίνει ότι ο πυρήνας του δεύτερου μορίου είναι σαφώς πιο σταθερός. Οι υψηλές τιμές εμβαδού αντιστοιχούν στα ακραία επίπεδα του δεματίου.



### Area of layers



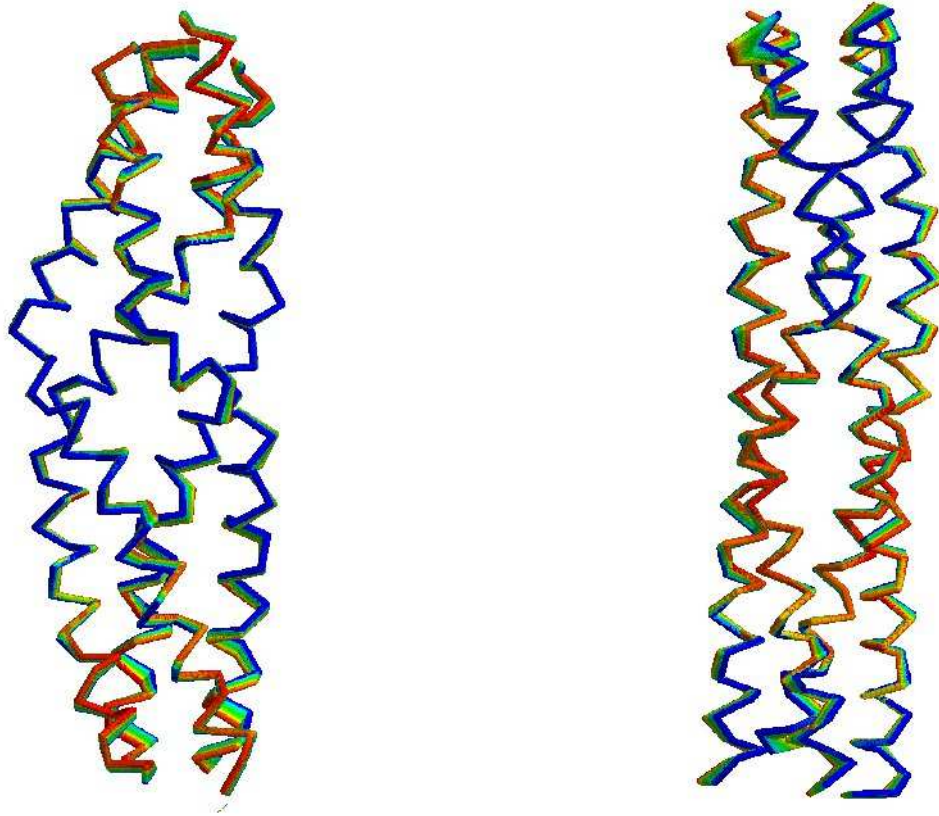
Εικόνα 4.72 Εμβαδό των επιπέδων των καταλοίπων που σχηματίζουν τον υδρόφοβο πυρήνα. Η μαύρη γραμμή αντιστοιχεί στον πυρήνα της πλεξούδας και η καφέ στο sigmoidal.

Συγκρίνοντας την αθροιστική διακύμανση των 5 πρώτων eigenvector ανά κατάλοιπο (εικόνες 4.28 και 4.62), βλέπουμε ότι η συνεισφορά τους στην συνολική κίνηση των μορίων είναι αρκετά σημαντική. Ειδικότερα, στο μόριο της πλεξούδας παρατηρούμε πολύ υψηλές διακυμάνσεις τόσο στα αμινοτελικά-καρβοξυτελικά άκρα των μονομερών όσο και στο εσωτερικό τους τμήμα και σε πολύ υψηλότερες τιμές από τα αντίστοιχα κατάλοιπα του μορίου sigmoidal.

Με την μέχρι τώρα σύγκριση μπορούμε να ισχυριστούμε ότι η δομή του Sigmoidal εμφανίζει μεγαλύτερη σταθερότητα από την πλεξούδα. Μελετώντας συγκριτικά τις δομές των δύο μοντέλων εμφανίζονται κάποιες βασικές διαφορές. Η πλεξούδα δεν εμφανίζει τα βασικά δομικά χαρακτηριστικά ενός 4-α-ελικοειδούς δεματίου και χαρακτηρίζεται από μία κάμψη στο κεντρικό τμήμα και των τεσσάρων ελίκων προς την εξωτερική πλευρά του δεματίου. Με την εισαγωγή των επτά καταλοίπων δεν υπήρχε χώρος ώστε να διευθετηθούν οι πλευρικές ομάδες κατάλληλα και έτσι δημιουργήθηκε αυτή η κάμψη.

Τέλος, θα συγκρίνουμε τα αποτελέσματα των αναλύσεων των κινήσεων βάση του 1 eigenvector (εικόνα 4.73) για τα δύο μοντέλα τα οποία συμφωνούν και με τον πίνακα variance-covariance. Στην πλεξούδα είναι εμφανής ότι τα άκρα του δεματίου κινούνται περισσότερο από ότι το εσωτερικό της τμήμα όμως υπάρχουν κατάλοιπα στο εσωτερικό τα

οποία έχουν κινητικότητα μεγαλύτερη από ότι συνηθίζεται. Από την άλλη τα άκρα του sigmoidal χαρακτηρίζονται επίσης από υψηλή κινητικότητα η οποία δεν εμφανίζεται στο εσωτερικό της τμήμα.



Εικόνα 4.73 Η χαρακτηριστική κίνηση των μορίων στο κυρίαρχο cluster, απαλείφοντας τα υπεκινητικά κατάλοιπα, βάση τον 1<sup>ο</sup> eigenvector. Αριστερά το μοντέλο της πλεξούδας και δεξιά του Sigmoidal.

## Συμπεράσματα

Από την ανάλυση των δύο μοντέλων συμπεραίνουμε ότι μεγαλύτερη σταθερότητα στη δομή εμφανίζεται στο μόριο του sigmoidal από ότι στο μόριο της πλεξούδας, χωρίς όμως πολύ σημαντικές διαφορές. Ωστόσο δεν θα μπορούσε να θεωρηθεί το μοντέλο πλεξούδα ως η δομή της Rop2aL(QL), καθώς αποκλίνει από τα βασικά χαρακτηριστικά των ελικοειδών δεματίων, ενώ η δομή του μοντέλου sigmoidal είναι πιο αληθοφανής. Στην πρότυπη δομή τους, την RM6, υπήρχε μια χαρακτηριστική κίνηση του καρβοξυτελικού και αμινοτελικού άκρου του κάθε μονομερούς προς την ίδια κατεύθυνση, αλλά αντίθετη από το κεντρικό του τμήμα. Αυτό το μοτίβο εμφανίζεται στα μονομερή της πλεξούδας, όμως μόνο σε ένα από τα 4 μονομερή του sigmoidal. Από δομικής πλευράς, το sigmoidal μοιάζει περισσότερο με την πρότυπη δομή σε σχέση με την πλεξούδα, διότι η πλεξούδα εμφανίζει μια έντονη κάμψη στην περιοχή εισαγωγής των επτά καταλοίπων η οποία απουσιάζει από το sigmoidal.

Τέλος, δεν μπορούμε να ισχυριστούμε ότι η δομή του sigmoidal είναι η δομή της παραλλαγής Rop2aL(QL) αφού ο χρόνος των προσομοιώσεων δεν ήταν αρκετός και δεν είχαμε ξεκάραθα αποτελέσματα. Η επίλυση της δομής του μεταλλάγματος αυτού με κρυσταλλογραφία ακτινών Χ, θα δώσει σαφή αποτελέσματα. Έτσι μας δίνεται η δυνατότητα να πραγματοποιήσουμε προσομοιώσεις και στην κρυσταλλογραφικά προσδιορισμένη δομή ώστε να συγκριθεί με την δομή που κατασκευάσαμε υπολογιστικά και να εξάγουμε συμπεράσματα για την αξιοπιστία του.

# Βιβλιογραφία

1. Allen, M.P. (2004) Introduction to molecular dynamics simulation. NIC Series 23, 1-28.
2. Amadei, A., Lissen, A. & Berendsen, H. J.C. (1993) Essential dynamics of proteins. *Proteins* 17, 412-415.
3. Arthanari, H., McConnell, K. J., Beger, R., Young, M. A., Beveridge, D.L. & Bolton, P. H. (2003) Assessment of the molecular dynamics structure of DNA in solution based on calculated and observed NMR NOESY volumes and dihedral angles from scalar coupling constraints. *Biopolymers* 68, 3-15.
4. Aszodi, A., & Taylor, W. R., (1996) Homology modeling by distance geometry. *Folding and Design* 1 325.
5. Banner, D. W., Kokkinidis, M. & Tsernoglou, D (1987) Structure of the ColE1Rop protein at 1.7 Å resolution. *J. Mol. Biol.* 196, 657-675.
6. Bernstein, H. J. (1999) Rasmol 2.7.1. Molecular Graphics Visualisation Tool (VMD).
7. Blundell, T. L., Sibanda, B. L. & Pearl, I., (1985). Three-dimensional Structure, Specificity and Catalytic mechanism of Renin. *Nature* (London). 304, 273-275.
8. Castagnoli, L., Scarpa, M., Kokkinidis, M., Banner, D. W., Tsernoglou, D. & Cesareni, G. (1989). Genetic and structural analysis of the ColE1 Rop (Rom) protein. *EMBO J.* 8, 621-629.
9. Castagnoli, L., Vetriani, C., and Cesareni, G. (1994) Linking an easily detectable phenotype to the folding of a common structural motif. Selection of rare turn mutations that prevent the folding of Rop, *J. Mol. Biol.* 237, 378-387.
10. Cavalli, A., Ferrara, P. & Calfisch, A. (2002) Weak temperature dependence of the free energy surface and folding pathways of structured peptides. *Proteins Struct. Funct. Genet.* 47, 305-314. 2002

11. Cheatham III, T. E. & Brooks, B.R. (1998) Recent advances in molecular dynamics simulation towards the realistic representation of biomolecules in solution. *Theor. Chem. Acc.* 99, 279-288.
12. Chothia, C. (1992) One thousand families for the molecular biologist. *Nature* 357, 543-544.
13. Chothia, C., and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*5, 823-826
14. Chou, K.-C., Maggiora, G. M., Ne'methy, G., and Scheraga, H. A. (1988) Energetics of the structure of the four-R-helix bundle in proteins, *Proc. Natl. Acad. Sci. U.S.A.* 85, 4295-4299.
15. Cohen, C., and Parry, D. A. D. (1990) R-Helical Coiled Coils and Bundles: How to Design an R-Helical Protein, *Proteins* 7, 1-15.
16. Cregut, D., Drin, G., Liautard, J.-P. & Chiche, L. (1998). Hinge-bending motions in annexins: Molecular dynamics and essential dynamics of apo-annexin V and of calcium bound annen V and I. *Protein Eng.* 11, 891-900.
17. Crick, F. H. C (1953). The packing of a-helices: Simple coiled-ciols. *Acta Crystallographica* 6, 689-697
18. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald. An  $N \log(N)$  method for Ewald sums in large systems, *J. Chem. Phys.* 98, 10089-10092.
19. Diemand,A., Guex,N. and Peitsch,M.C. (2000) Protein structure computing in the genomic era. *Res. Microbiol.*, 151, 107-112.
20. Eberle, W., Pastore, A., Sander, C., and Rosch, P. (1991) The structure of ColE1 rop in solution, *J. Biomol. NMR* 1, 71-82.
21. [EMBnet molecular dynamics simulation Charmm](http://www.ch.embnet.org/MD_tutorial/)  
[http://www.ch.embnet.org/MD\\_tutorial/](http://www.ch.embnet.org/MD_tutorial/)
22. Ercolessi , F. (1997) A molecular dynamics primer. International School for Advanced Studies (Sissa-isas). Italy.
23. Eswar N., Webb B.,Marti-Renom, M.A., Madhusudhan M.S., Eramian D., Shen M.Y., Pieper U., Sali, A., (2006) Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics.* 5.6.1-5.6.30.

24. Eucb is a C++ program written by Ioannis G. Tsoulos and Athanassios Stavrakoudis.
25. Faduloglou, V., Glykos, N. M. & Kokkinidis, M. (2001) Side-chain conformations in 4-a-helical bundles. *Protein Eng.* 14, 321-328.
26. Fiser A., Do R.K.G., Sali A., (2000) Modeling of loops in protein structures. *Protein Sci* 9, 1753-1773.
27. Fiser, A., and Sali, A. (2003) Comparative Protein Structure Modeling. *Protein Structure*. Chasman, D., editor. Marcel Dekker, Inc.; New York: p. 167-206.
28. Garcia A.E (1992) Large-amplitude nonlinear motions in proteins *Phys. Rev. Lett.* 68, 2696.
29. Garemyr, R. & Elofsson A. (1999). Study of the Electrostatics Treatment in molecular dynamics simulations. *PROTEINS: Structure, function and genetics* 37, 417-428.
30. Gibas, C & Jambeck, P. (2001) Development bioinformatics computer skills. Published by O'Reilly & Associates, Inc.
31. Glykos, N. M. (2006) Carma: A molecular dynamics analysis program, *J. Comput. Chem.* (in press).
32. Glykos, N. M., and Kokkinidis, M. (1999) Meaningful refinement of poly-alanine models using rigid-body simulated annealing: Application to the structure determination of the A31P Rop mutant, *Acta Crystallogr.* D55, 1301-1308.
33. Glykos, N. M., and Kokkinidis, M. (2004) Structural Polymorphism of a Marginally Stable 4-R-Helical Bundle. Images of a Trapped Molten Globule? *Proteins* 56, 420-425.
34. Glykos, N. M., Y. Papanikolaou, et al. 2006 . Loopless Rop: structure and dynamics of an engineered homotetrameric variant of the repressor of primer protein, *Biochemistry* 45, 10905-10919.
35. Gobel, U., C., S., Scheider, R., & Valencia, A., (1994) Correlated mutations and residue constraints in proteins. *Proteins* 18 309.
36. Greer. J. (1985a). Molecular Structure of the Tnflammatory Protein C5a. *Science*, 228. 1055-1060.

37. Grossfield, A., Feller, S. E. & Pitman, M. C. (2007). Convergence of molecular dynamics simulations of membrane proteins. *proteins*, 67, 31-40.
38. Gruia, A.D. Fisher, S. & Smith, J.C. (2003) Molecular dynamics simulation reveals a surface salt bridge forming a kinetic trap in Unfolding of truncated staphylococcal nuclease. *Proteins* 50, 507-515.
39. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18, 2714-2723.
40. Guex, N., Diemand, A. And Peitsch, M.C. (1999) Protein modelling for all. *Trends Biochem. Sci.*, 24, 364-367.
41. Haile, J. M. (1997). *Molecular dynamics simulation*. Jhn Wiley & Sons.
42. Havel, T. & Snow, M. (1991), A new method for building protein conformations from sequence alignments with homologue of known structure. *J. Mol. Biol.* 217, 1-7
43. Hess, B., Convergence of sampling in protein simulations. *Phys. Rev.* E65,
44. Huber, R. & Bennett, Jr., W. S. (1983). Functional significance of flexibility in proteins. *Biopolymers*, 22, 261-279.
45. Hummer, G., Schotte, F. & Anfinrud, P. A. (2004). Unveiling Functional protein morions with picosecond x-ray crystallography and molecular dynamics simulations. *PNAS* 101, 15330-15334.
46. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics, *J. Mol. Graphics* 14, 33-38.
47. Hunenberger, P.H., Mark, A.E. & van Gunsteren W. F. (1995), Fluctuation and Cross-correlation Analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* 252, 492-503.
48. Ichiye, T., and Karplus, M. (1991) Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations, *Proteins* 11, 205-217.
49. Izaquire, J. A., Reich, S. & Skeel, R. D (1999). Longer time steps for molecular dynamics. *J. Chem. Phys.* 110, 9853-9864.

50. Johnson, M., Srinivasan, N., Sowdhamini, R. and Blundell, T. (1994), Knowledge-based protein modeling, *CRC Crit. Rev. Biochem. Mol. Biol.* 29, 1-68
51. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79, 926- 935.
52. Kale, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics, *J. Comput. Phys.* 151, 283-312.
53. Karplus, M. (2003). Molecular dynamics of biological macromolecules: a brief history and perspective. *Biopolymers* 68, 350-358.
54. Karplus, M. & McCammon, J. A. (2002) Molecular dynamics simulation of biomolecules. *Nat. Struct. Biol.* 9, 646-652.
55. Kleywegt, G.J. (1995) [O/X-PLOR Dictionaries] Dictionaries for Heteros, *ESF/CCP4 Newsletter* 32, 45-50.
56. Kleywegt, G.J.& Jones, T.A. (1998). Database in protein crystallography. *Acta Cryst.* D54, 1119-1132. (<http://xray.bmc.uu.se/gerard/rama/chi.html>)
57. Kyriazidis, I. (2006) The role of turn and heptad pattern in folding and stability of Rop protein, Master Thesis.
58. Lange O, Grubmüller H (2006) Can principal components yield a dimension reduced description of protein dynamics on long time scales? *J Phys Chem B* 110:22842-22852.
59. Larson, S. M., Snow, C. D., Shirts, M. & Pande, V. S. (2003) Folding@Home and Genome@Home: Using distributed computing to tackle previously intractable problems in computational biology. *Computational Genomics*.
60. Lassalle, M. W., and Hinz, H. J. (1998) Unfolding of the tetrameric loop deletion mutant of ROP protein is a second-order reaction, *Biochemistry* 37, 8465-8472.
61. Lassalle, M. W., Hinz, H.-J., Wenzel, H., Vlasi, M., Kokkinidis, M., and Cesareni, G. (1998) Dimer-to-Tetramer Transformation: Loop Excision Dramatically Alters Structure and Stability of the ROP Four R-helix Bundle Protein, *J. Mol. Biol.* 279, 987-1000.



62. Levitt, M (1992), Accurate modeling of protein conformation by automatic segment matching, *J. Mol. Biol.* 226, 506-533
63. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis, *J. Mol. Model.* 7, 306-317
64. MacKerell, A. D., Bashfold, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fisher, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith. C., Stote, R Straub. J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. & Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586-3616.
65. Madhusudhan M.S., Martin-Remon M.A., Eswar N., John B., Pieper U., Karchin R., Min-yi. Shen, Sali A. (2005) Comparative Protein Structure Modeling. In: *Proteomics Protocols Handbook*. Ed: J.M. Walker, pp. 831-860, Humana Press Inc., Totowa, NJ.
66. Magliery, T.J. & Regan, L. (2004). A cell-based screen for function of the four-helix bindle protein Rop: a new tool for combinatorial experiments in biophysics. *Protein Engineering, Design & Selection* 17, 77-83.
67. Martin-Remon M.A., Stuart A.C., Fiser A., Sanchez R., Melo F., Sali A., (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29, 291-325.
68. McRee, D. E. (1992) A visual protein crystallographic software system for X11/XView, *J. Mol. Graphics* 10, 44-46.
69. Merritt, E. A. (1999). Expanding the Model: Anisotropic Displacement Parameters in Protein Structure Refinement. *Acta Cryst. D* 55, 1109-1117.
70. Merritt, E. A., & Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics, *Methods Enzymol.* 277, 505-524.
71. Munson, M., R. O'Brein, et al., (1994). Redesigning the hydrophobic core of a four-helix-bindle protein." *Protein Sci* 3: 2015-2022

72. Munson, M., S. Balasubramaian, S., Fleming, K. G., Nagi, A. D., O'Brien R., Sturtevant J. M. & Regan, L. (1996). What makes a protein protein? Hydrophobic core designs that specify stability and structural properties. *Prot. Sci* 5, 1584-1593.
73. Nagi, A. D. & Regan L. (1997). An inverse correlation between loop length and stability in a four-helix-bundle protein. *Fold Des* 2, 67-75
74. Nagi, A. D., Andersson, K.S. et al., & Regan, L. (1999). Using loop length variants to dissect the folding pathway of a four-helix-bundle protein. *J. Mol Biol* 286, 257-265.
75. Nguyen, P.H., Mu, Y., Emil Mittag., Chin-Kun Hu., and Mai Suan Li & Stock, G., (2005) Free Energy Landscape and Folding Mechanism of a Hairpin in Explicit Water: A Replica Exchange Molecular Dynamics Study *Proteins* 58,45
76. Paliakasis, C.D. and Kokkinidis, M. (1992). Relationships between sequence and structure for the four-alpha-helix bundle tertiary motif in proteins. *Protein Eng* 5, 739-748
77. Palmer, K. A., Scheraga: H. A., Riordan, J. F. & Vallee, B. L. (1986). A Preliminary Three-dimensional Structure of angiogenin. *Proc. Nat. Acad. Sci., U.S.A.* 83, 1965-1969.
78. Peitsch, M.C. (1995) Protein modelling by E-Mail. *BioTechnology*, 13, 658-660.
79. Peitsch, M.C. (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.*, 24, 274-279.
80. Peitsch, M.C., & Jongeneel, C.V. (1993) A 3-model for the CD40 ligand predicts that it is a compact trimer similar to the tumor necrosis factors. *Intern. Immunol.*, 5, 233-238.
81. Polisky, B. (1988) ColE1 replication control circuitry: Sense from antisense, *Cell* 55, 929-932.
82. Ponder, J.W. & Richards, F.M. (1987) Tertiary templates for proteins -- use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* 193, 775-791.
83. Predki, P.F., Nayak, L. M., Gottlieb, M.B.C., & Regan, L. (1995) Dissecting RNA-Protein Interactions: RNA-RNA Recognition by Rop, *Cell* 80, 41-50.

84. Presnell, S. R. & Cohen, F. E. (1989) Topological distribution of four- $\alpha$ -helix-bundles. PNAS, 86, 6592-6596.
85. R Development Core Team (2004). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna Austria. ISBN 3-900051-07-0. URL <http://www.R-project.org>.
86. Ranson, N.A. et al., (2001), ATP-bound states of GrpEL captured by cryo-electron microscopy. Cell 107, 869-879.2001
87. Rossi V, Gaboriaud C, Lacroix M, et al. (1995) Structure of the catalytic region of human complement protease Cls: study by chemical cross-linking and three-dimensional homology modeling. Biochemistry ; 34:7311-7321.
88. Rost, B., Protein prediction in 1D, 2D and 3D. Heidelberg, Germany.
89. Sali A., (1995) Modeling mutations and homologous proteins. Curr Opin Biotechnol 6, 437-451.
90. Sali A., Blundell T.L., (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815.
91. Sali, A., & Overington, J.P. (1994) Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci 3, 1582-1596.
92. Sanchez, R., & Sali A.,(2000) Comparative protein structure modeling. Introduction and practical examples with modeller. Methods Mol Biol 143, 97-129, 2000.
93. Sanchez, R., & SaliA., (1997) Advances in comparative protein-structure modelling JMolStruct (Theochem) 1997;398-399:489-496
94. Schwede,T., Kopp, J., Guex,N., Peitsch,M.C. (2003) , SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research., 13, 3381-3385
95. Shaner, M.C., Blair, I.M. & Svhneider, T.D. (Mudge, T.N., Milutinovic, V. & Hunter, L. eds) (2003) Sequence logos: A poewrful, yet simple tool, Proceeding of the Twenty-Sixth Annual Hawaii International Conference on System Science, Volume 1: Architechure and Biotechnology Computing", 813-821", IEEE Computer Society Press", Los Alamitos, CA", 1993 From the "Alternative Approachs to Sequence Representation" minitrack of the Biotechnology computing Track, Hawaii International conference on system science-26, Kauai, Hawaii-January 5-8, 1993.

96. Sigler, P.B., Ma, J.P., Xu, Z.H. & Karplus, M. (2000) A dynamic model for the allosteric mechanism of GroEL. *J. Mol. Biol.* 302, 303-313.
97. Steif, C., Hinz, H.-J., and Cesareni, G. (1995) Effects of Cavity- Creating Mutations on Conformational Stability and Structure of the Dimeric 4-R-Helical Protein ROP: Thermal Unfolding Studies, *Proteins* 23, 83-96.
98. Sutcliffe, M. J., Dobson, C. M., & Oswald, R. (1992) E., Solution structure of neuronal bungarotoxin determined by two-dimensional NMR spectroscopy: calculation of tertiary structure using systematic homologous model building, dynamical simulated annealing, and restrained molecular dynamics. *Biochemistry* 31 2962.
99. Jolliffe, T., (2002) *Principal component analysis* (Springer, New York,)
100. Tai, K., Shen, T., Borjesson, U., Philippopoulos, M. & McCammon, J.A. (2001) Analysis of a ten-nanosecond molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J.* 81, 715-724.
101. Taylor, W. R., (1993) Protein fold refinement: building models from idealised folds using motif constraints and multiple sequence data. *Protein Eng.* 6 593
102. Tooze & Branden (1999). *Introduction to Protein Structure, Second Edition*, Garland Publishing.
103. Tramontano A, Leplae R. And Mora, V. (2001) Analysis and assessment of comparative modeling predictions in CASP4. *Proteins*, 45 (Suppl. 5), 22-38
104. Turner, P.J. (1991-1995) & ACE/gr Development Team (1996-1998), Xmgr v4.1.2.
105. Van Gunsteren, W. F., Burgi, R., Peter, C. & Daura, X (2001). The key to solving the protein-folding problem lies in an accurate description of the denatured state. *Angew. Chem. Int. Ed.* 40, 352-355.
106. Vlassi, M., et al. (1994) Restored heptad pattern continuity does not alter the folding of a four-R-helix bundle, *Nat. Struct. Biol.* 1, 706-716.
107. Xiang, Z., (2006), *Advances in Homology Protein Structure Modeling*, *Curr Protein Sci.* 7, 217-227.
108. York D.M., Darden T.A., Pedersen L.G. (1993), The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Eward and truncated list methods. *J. Chem. Phys.* 99, 8345-8348.

109. Young, M.A., Gonfloni, S., Superti-Furga, G., Roux, B & Kuriyan, J. (2001), Dynamic coupling between the SH2 and SH3 domains of c-Src and hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* 105, 115
110. Zeilkiewicz, J. (2005). Structural properties of water: Comparison of the SPC, SPCE, TIP4P and TIP5P models of water. *J. Chem. Phys.* 123, 104501.

# Παράρτημα

## Ομόλογη μοντελοποίηση

**Script A** Το script βάση του οποίου γίνεται η μοντελοποίηση της αλληλουχίας 2aaQLa με βάση τη γνωστή δομή 1QX8a γραμμένο σε γλώσσα προγραμματισμού python.

```
# Homology modeling by the automodel class
from modeller import *          # Load standard Modeller classes
from modeller.automodel import *  # Load the automodel class

log.verbose()      # request verbose output
env = environ()   # create a new MODELLER environment to build this model in

# directories for input RM6-tests files
env.io.RM6_test_directory = ['.', 'RM6_test/']

a = automodel(env,
alnfile = 'alignment_aQL.ali',      # alignment filename
knowns = '1QX8a',                  # codes of the templates
sequence = '2aaQLa')                # code of the target

a.starting_model= 1                  # index of the first model
a.ending_model = 1                  # index of the last model
# (determines how many models to calculate)
a.make()                             # do the actual homology modeling
```

**Script B.** Δηλώνει ποια κατάλοιπα θα γίνουν α-έλικα και ποια θα κρατηθούν σταθερά.

```
# Homology modeling by the automodel class
#
# Demonstrates how to refine only a part of the model.
#
# You may want to use the more exhaustive "loop" modeling routines instead.

from modeller import *          # Load standard Modeller classes
from modeller.automodel import *  # Load the automodel class

log.verbose()

env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']
# selected atoms do not feel the neighborhood
env.edat.nonbonded_sel_atoms = 2

class MyModel(automodel):
def special_restraints(self, aln):
    rsr = self.restraints
    at = self.atoms
    #     Add some restraints from a file:
    #     rsr.append(file='my_rsrsl.rsr')

    #     Residues 28 through 34 should be an alpha helix:
    rsr.add(secondary_structure.alpha(self.residue_range('28:', '34:')))

    #     Restrain the specified CA-CA distance to 10 angstroms (st. dev.=0.1)
    #     Use a harmonic potential and X-Y distance group.
    rsr.add(forms.gaussian(group=physical.xy_distance,
        feature=features.distance(at['CA:35'],
        at['CA:40']),
        mean=10.0, stdev=0.1))

mdl = model(env, file='1QX8a')

# Keep residues 5-27 in chain A rigid:
r = rigid_body(mdl.residue_range('5:A', '27:A'))
mdl.restraints.rigid_bodies.append(r)

# Keep residues 35-54 in chain A rigid:
r = rigid_body(mdl.residue_range('35:A', '51:A'))
mdl.restraints.rigid_bodies.append(r)

a = MyModel(env,
alnfile = 'alignment_aQL.ali',      # alignment filename
knowns = '1QX8a',                  # codes of the templates
sequence = '2aaQLa')               # code of the target
a.starting_model = 3                # index of the first model
a.ending_model = 3                  # index of the last model
# (determines how many models to calculate)
a.make()                             # do homology modeling
```

**Script C** To script με το οποίο, μέσω του υποσυνόλου *MyModel*, προσδιορίζει την τετραμερή μορφή της πρωτεΐνης με συμμετρία A C και B D.

```
# Homology modeling by the automodel class
#
# Demonstrates how to build multi-chain models, and symmetry restraints
#
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()

# Override the 'special_restraints' and 'user_after_single_model' methods:
class MyModel(automodel):
def special_restraints(self, aln):
# Constrain the A and B chains to be identical (but only restrain
# the C-alpha atoms, to reduce the number of interatomic distances
# that need to be calculated):
s1 = selection(self.chains['A']).only_atom_types('CA')
s2 = selection(self.chains['B']).only_atom_types('CA')
s3 = selection(self.chains['C']).only_atom_types('CA')
s4 = selection(self.chains['D']).only_atom_types('CA')
self.restraints.symmetry.append(symmetry(s1, s3, 1.0))
self.restraints.symmetry.append(symmetry(s2, s4, 1.0))

env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Be sure to use 'MyModel' rather than 'automodel' here!
a = MyModel(env,
alnfile = 'fourchain.ali' ,          # alignment filename
knowns = '1QX81',                    # codes of the templates
sequence = '2aaQL')                 # code of the target

a.starting_model= 1                   # index of the first model
a.ending_model = 1                   # index of the last model
# (determines how many models to calculate)
a.make()                              # do homology modeling
```



*Script D* το script το οποίο κάνει την δομή τετραμερές και ταυτόχρονα τα κατάλοιπα εισαγωγής της κάθε αλυσίδας α έλικα

```
# Homology modeling by the automodel class
#
# Demonstrates how to build multi-chain models, and symmetry restraints
#
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()

# Override the 'special_restraints' and 'user_after_single_model' methods:
class MyModel(automodel):
def special_restraints(self, aln):
# Constrain the A and B chains to be identical (but only restrain
# the C-alpha atoms, to reduce the number of interatomic distances
# that need to be calculated):
s1 = selection(self.chains['A']).only_atom_types('CA')
s2 = selection(self.chains['B']).only_atom_types('CA')
s3 = selection(self.chains['C']).only_atom_types('CA')
s4 = selection(self.chains['D']).only_atom_types('CA')
self.restraints.symmetry.append(symmetry(s1, s3, 1.0))
self.restraints.symmetry.append(symmetry(s2, s4, 1.0))

#residues should be an alpha helix:
rsr = self.restraints
at = self.atoms
rsr.add(secondary_structure.alpha(self.residue_range('28:A', '34:A')))
rsr.add(secondary_structure.alpha(self.residue_range('82:B', '88:B')))
rsr.add(secondary_structure.alpha(self.residue_range('136:C', '142:C')))
rsr.add(secondary_structure.alpha(self.residue_range('190:D', '196:D')))
def user_after_single_model(self):
# Report on symmetry violations greater than 1A after building
# each model:
self.restraints.symmetry.report(1.0)

env = environ()
# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Be sure to use 'MyModel' rather than 'automodel' here!
a = MyModel(env,
alnfile = 'fourchain.ali' ,      # alignment filename
knowns = '1QX81',                # codes of the templates
sequence = '2aaQL')              # code of the target

a.starting_model= 1                # index of the first model
a.ending_model = 1                 # index of the last model
# (determines how many models to calculate)
a.make()                            # do homology modeling
```

**Script E.** Μοντελοποίηση της δομής, βάση του αρχείου χωρικών περιορισμών (rsr) το οποίο θέτει ο χρήστης.

```
# Modeling using a provided restraints file (csrfile)
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

a = automodel(env,
alnfile = 'fourchain.ali',      # alignment filename
knowns = '1QX81',              # codes of the templates
sequence = '2aaQL',            # code of the target
csrfile = 'my.pleksouda.rsr')   # use 'my' restraints file
a.starting_model= 1              # index of the first model
a.ending_model = 1              # index of the last model
# (determines how many models to calculate)
a.make()                         # do homology modeling
```

**Αρχείο rsr:** Ενδεικτικό παράδειγμα του αρχείου rsr

```
MODELLERS5 VERSION: MODELLER FORMAT
R   3   1   1   1   2   2   1   3   2           1.5380   0.0364
R   3   1   1   1   2   2   1   4   3           1.5300   0.0364
R   3   1   1   1   2   2   1   5   4           1.5220   0.0384
R   3   1   1   1   2   2   1   6   5           1.2600   0.0237
R   3   1   1   1   2   2   1   7   5           1.2600   0.0237
R   3   1   1   1   2   2   1   1   2           1.4800   0.0384
R   3   1   1   1   2   2   1   9   8           1.2300   0.0218
R   3   1   1   1   2   2   1   8   2           1.4900   0.0344
R   3   1   1   1   2   2   1   8   10          1.3450   0.0282
R   3   1   1   1   2   2   1  12  11          1.5380   0.0364
R   3   1   1   1   2   2   1  13  12          1.5300   0.0364
.
R   3   1   3   4   4   2   1  111  110  113  112           0.0000   0.0810
R   3   1   3   4   4   2   1  108  107  114  109           0.6155   0.0732
R   3   1   3   4   4   2   1  121  117  123  122           0.0000   0.0496
R   3   1   3   4   4   2   1  117  116  121  118           0.6155   0.0732
R   3   1   3   4   4   2   1  118  119  120  117           0.6155   0.0732
R   3   1   3   4   4   2   1  129  124  131  130           0.0000   0.0496
R   3   1   3   4   4   2   1  124  123  129  125           0.6155   0.0732
R   3   1   3   4   4   2   1  126  128  127  125           0.6155   0.0732
R   3   1   3   4   4   2   1  136  132  138  137           0.0000   0.0496
R   3   1   3   4   4   2   1  132  131  136  133           0.6155   0.0732
R   3   1   3   4   4   2   1  133  134  135  132           0.6155   0.0732
R   3   1   3   4   4   2   1  144  139  146  145           0.0000   0.0496
R   3   1   3   4   4   2   1  139  138  144  140           0.6155   0.0732
```

**Πίνακας 2.1** Λίστα τύπων «φυσικών» περιορισμών

Python object	Index	Group
physical.bond	1	Bond length potential
physical.angle	2	Bond angle potential
physical.dihedral	3	Stereochemical cosine dihedral potential
physical.improper	4	Stereochemical improper dihedral potential
physical.soft_sphere	5	soft-sphere overlap restraints
physical.lennard_jones	6	Lennard-Jones 6-12 potential
physical.coulomb	7	Coulomb point-point electrostatic potential
physical.h_bond	8	H-bonding potential
physical.ca_distance	9	Distance restraints 1 ( $C_{\alpha} - C_{\alpha}$ )
physical.n_o_distance	10	Distance restraints 2 (N-O)
physical.phi_dihedral	11	Mainchain $\Phi$ dihedral restraints
physical.psi_dihedral	12	Mainchain $\Psi$ dihedral restraints
physical.omega_dihedral	13	Mainchain $\omega$ dihedral restraints
physical.chi1_dihedral	14	Sidechain $\chi_1$ dihedral restraints
physical.chi2_dihedral	15	Sidechain $\chi_2$ dihedral restraints
physical.chi3_dihedral	16	Sidechain $\chi_3$ dihedral restraints
physical.chi4_dihedral	17	Sidechain $\chi_4$ dihedral restraints
physical.disulfide_distance	18	Disulfide distance restraints
physical.disulfide_angle	19	Disulfide angle restraints
physical.disulfide_dihedral	20	Disulfide dihedral angle restraints
physical.lower_distance	21	X lower bound distance restraints
physical.upper_distance	22	X upper bound distance restraints
physical.sd_mn_distance	23	Distance restraints 3 (SDCH-MNCH)
physical.chi5_dihedral	24	Sidechain $\chi_5$ dihedral restraints
physical.phi_psi_dihedral	25	( $\Phi, \Psi$ ) binomial dihedral restraints
physical.sd_sd_distance	26	Distance restraints 4 (SDCH-SDCH)
physical.xy_distance	27	Distance restraints 5 (X-Y)
physical.nmr_distance	28	NMR distance restraints 6 (X-Y)
physical.nmr_distance2	29	NMR distance restraints 7 (X-Y)
physical.min_distance	30	Minimal distance restraints
physical.nonbond_spline	31	Non-bonded spline restraints
physical.accessibility	32	Atomic accessibility restraints
physical.density	33	Atom density restraints
physical.absposition	34	Absolute position restraints
physical.dihedral_diff	35	Dihedral angle difference restraints
physical.gbsa	36	GBSA implicit solvent potential
physical.em_density	37	EM density fitting potential

physical.saxs	38	SAXS restraints
physical.symmetry	39	Symmetry restraints

## Προσομοιώσεις μοριακής δυναμικής

### A. Minimization and heating up script (*heat.namd*)

#### *pleksouda system*

```

# Input files
#
structure           ionized.psf
coordinates         ionized.pdb
parameters         par_all27_prot_na.inp
paraTypeCharmm     on

#
# Output files & writing frequency for DCD
# and restart files
#
outputname         output/heat_out
binaryoutput       off
restartname        output/restart
restartfreq        1000
binaryrestart       yes
dcdFile            output/heat_out.dcd
dcdFreq            200

#
# Frequencies for logs and the xst file
#
outputEnergies     40
outputTiming       200
xstFreq            200

#
# Timestep & friends
#
timestep           2.0
stepsPerCycle      20
nonBondedFreq      2
fullElectFrequency 4

#
# Simulation space partitioning
#
switching          on
switchDist         10
cutoff             12
pairlistdist       13.5

#
# Basic dynamics
#
temperature        0
COMmotion          no
dielectric          1.0
exclude            scaled1-4
1-4scaling         1.0
rigidbonds         all

```

```

#
# Particle Mesh Ewald parameters.
#
Pme                on
PmeGridsizeX      128          # <===== CHANGE ME
PmeGridsizeY      64           # <===== CHANGE ME
PmeGridsizeZ      64           # <===== CHANGE ME

#
# Periodic boundary things
#
wrapWater          on
wrapNearest        on

cellBasisVector1   109.97  00.00  00.00  # <===== CHANGE ME
cellBasisVector2   00.00   60.86  00.00  # <===== CHANGE ME
cellBasisVector3   00.00   00.00  58.65  # <===== CHANGE ME
cellOrigin         0.61   -0.13  -0.88  # <===== CHANGE ME

#
# Fixed atoms for initial heating-up steps
#
fixedAtoms         on
fixedAtomsForces   on
fixedAtomsFile     fix_backbone.pdb
fixedAtomsCol      B

#
# Restrained atoms for initial heating-up steps
#
constraints        on
consRef            restrain_ca.pdb
consKFile          restrain_ca.pdb
consKCol           B

#
# Langevin dynamics parameters
#
langevin           on
langevinDamping    10
langevinTemp       320          # <===== Check me
langevinHydrogen   on

langevinPiston     on
langevinPistonTarget 1.01325
langevinPistonPeriod 200
langevinPistonDecay 100
langevinPistonTemp  320          # <===== Check me

useGroupPressure   yes

#####
# The actual minimisation and heating-up #
# protocol follows. The number of steps #
# shown below are too small for a real run#
#####

#
# run one step to get into scripting mode
#
minimize           0

#
# turn off pressure control until later
#
langevinPiston     off

```

```

#
# minimize nonbackbone atoms
#
minimize          2000          ;# <===== CHANGE ME
output            output/min_fix

#
# min all atoms
#
fixedAtoms        off
minimize          2000          ;# <===== CHANGE ME
output            output/min_all

#
# heat with CAs restrained
#
set temp 20;
while { $temp < 321 } {      ;# <===== Check me
  langevinTemp      $temp
  run                2000      ;# <===== CHANGE ME
  output            output/heat_ca
  set temp [expr $temp + 20]
}

#
# equilibrate volume with CAs restrained
#
langevinPiston    on
run                2000          ;# <===== CHANGE ME
output            output/equil_ca

#
# equilibrate volume without restraints
#
constraintScaling  0
run                4000          ;# <===== CHANGE ME

```

## ***B. Equilibration script (equi.namd)***

### ***pleksouda system***

```
# Input files
#
structure          ionized.psf
coordinates        heat_out.coor
velocities         heat_out.vel
extendedSystem    heat_out.xsc
parameters        par_all27_prot_na.inp
paraTypeCharmm    on

#
# Output files & writing frequency for DCD
# and restart files
#
outputname         output/equi_out
binaryoutput      off
restartname       output/restart
restartfreq       10000
binaryrestart     yes
dcdFile          output/equi_out.dcd
dcdFreq          200
DCDunitcell      yes

#
# Frequencies for logs and the xst file
#
outputEnergies    40
outputTiming      200
xstFreq           200

#
# Timestep & friends
#
timestep          2.0
stepsPerCycle     20
nonBondedFreq    2
fullElectFrequency 4

#
# Simulation space partitioning
#
switching         on
switchDist        10
cutoff            12
pairlistdist     14.5

#
# Basic dynamics
#
COMmotion        no
dielectric        1.0
exclude          scaled1-4
1-4scaling       1.0
rigidbonds       all

#
# Particle Mesh Ewald parameters.
#
Pme              on
PmeGridsizeX    112
PmeGridsizeY    60
PmeGridsizeZ    60
# <===== CHANGE ME
# <===== CHANGE ME
# <===== CHANGE ME

PMEprocessors    10
```

```

#
# Periodic boundary things
#
wrapWater          on
wrapNearest        on
wrapAll            on

#
# Langevin dynamics parameters
#
langevin           on
langevinDamping    1
langevinTemp       320          # <===== Check me
langevinHydrogen   on

langevinPiston     on
langevinPistonTarget 1.01325
langevinPistonPeriod 200
langevinPistonDecay 100
langevinPistonTemp 320          # <===== Check me

useGroupPressure   yes

firsttimestep      0           # <===== CHANGE ME
run                50000000    ;# <===== CHANGE ME

```



## *C. Minimization and heating up script (heat.namd)*

### *sigmoidal system*

```
#
# Input files
#
structure           ionized.psf
coordinates         ionized.pdb
parameters          par_all27_prot_na.inp
paraTypeCharmm     on
margin 2.5

#
# Output files & writing frequency for DCD
# and restart files
#
outputname          output/heat_out
binaryoutput        off
restartname          output/restart
restartfreq         10000
binaryrestart        yes
dcdFile             output/heat_out.dcd
dcdFreq             200

#
# Frequencies for logs and the xst file
#
outputEnergies      40
outputTiming        200
xstFreq             200

#
# Timestep & friends
#
timestep            2.0
stepsPerCycle       20
nonBondedFreq       2
fullElectFrequency  4

#
# Simulation space partitioning
#
switching           on
switchDist          8
cutoff              9
pairlistdist        10

#
# Basic dynamics
#
temperature         0
COMmotion           no
dielectric           1.0
exclude             scaled1-4
1-4scaling          1.0
rigidbonds          all

#
# Particle Mesh Ewald parameters.
#
Pme                 on
PmeGridsizeX        128
PmeGridsizeY        54
PmeGridsizeZ        54
# <===== CHANGE ME
# <===== CHANGE ME
# <===== CHANGE ME
```

```

#
# Periodic boundary things
#
wrapWater          on
wrapNearest        on
wrapAll            off

cellBasisVector1   119.24  00.00  00.00  # <===== CHANGE ME
cellBasisVector2   00.00  42.85  24.74  # <===== CHANGE ME
cellBasisVector3   00.00  00.00  49.92  # <===== CHANGE ME
cellOrigin         0.00    0.00    0.00  # <===== CHANGE ME

#
# Fixed atoms for initial heating-up steps
#
fixedAtoms         on
fixedAtomsForces   on
fixedAtomsFile     fix_backbone.pdb
fixedAtomsCol      B

#
# Restrained atoms for initial heating-up steps
#
constraints        on
consRef            restrain_ca.pdb
consKFile          restrain_ca.pdb
consKCol           B

#
# Langevin dynamics parameters
#
langevin           on
langevinDamping    10
langevinTemp       320                # <===== Check me
langevinHydrogen   on

langevinPiston     on
langevinPistonTarget 1.01325
langevinPistonPeriod 200
langevinPistonDecay 100
langevinPistonTemp 320                # <===== Check me

useGroupPressure   yes

#####
# The actual minimisation and heating-up #
# protocol follows. The number of steps #
# shown below are too small for a real run#
#####

```

---

```

#
# run one step to get into scripting mode
#
minimize           0

#
# turn off pressure control until later
#
langevinPiston     off

#
# minimize nonbackbone atoms
#
minimize           2000                ;# <===== CHANGE ME
output             output/min_fix

#

```

```

# min all atoms
#
fixedAtoms          off
minimize            2000                ;# <===== CHANGE ME
output              output/min_all

#
# heat with CAs restrained
#
set temp 20;
while { $temp < 321 } {                ;# <===== Check me
langevinTemp        $temp
run                 2000                ;# <===== CHANGE ME
output              output/heat_ca
set temp [expr $temp + 20]
}

#
# equilibrate volume with CAs restrained
#
langevinPiston      on
run                 2000                ;# <===== CHANGE ME
output              output/equil_ca

#
# equilibrate volume without restraints
#
constraintScaling    0
run                 20000               ;# <===== CHANGE ME

```

## *D. Equilibration script (equi.namd)*

### *sigmoidal system*

```
# Input files
#
structure          ionized.psf
coordinates        heat_out.coor
velocities         heat_out.vel
extendedSystem     heat_out.xsc
parameters         par_all27_prot_na.inp
paraTypeCharmm    on

#
# Output files & writing frequency for DCD
# and restart files
#
outputname         output/equi_out
binaryoutput      off
restartname        output/restart
restartfreq        10000
binaryrestart      yes
dcdFile           output/equi_out.dcd
dcdFreq           200
DCDunitcell       yes

#
# Frequencies for logs and the xst file
#
outputEnergies    40
outputTiming      200
xstFreq           200

#
# Timestep & friends
#
timestep          2.0
stepsPerCycle     20
nonBondedFreq     2
fullElectFrequency 4

#
# Simulation space partitioning
#
switching         on
switchDist        10
cutoff            12
pairlistdist      14.5

#
# Basic dynamics
#
COMmotion         no
dielectric         1.0
exclude           scaled1-4
1-4scaling        1.0
rigidbonds        all

#
# Particle Mesh Ewald parameters.
#
Pme               on
PmeGridsizeX      112          # <===== CHANGE ME
PmeGridsizeY      60          # <===== CHANGE ME
PmeGridsizeZ      60          # <===== CHANGE ME

PMEprocessors     10
```

```

#
# Periodic boundary things
#
wrapWater          on
wrapNearest        on
wrapAll            on

#
# Langevin dynamics parameters
#
langevin           on
langevinDamping    1
langevinTemp       320          # <===== Check me
langevinHydrogen   on

langevinPiston     on
langevinPistonTarget 1.01325
langevinPistonPeriod 200
langevinPistonDecay 100
langevinPistonTemp 320          # <===== Check me

useGroupPressure   yes

firsttimestep      0           # <===== CHANGE ME
run                50000000    ;# <===== CHANGE ME

```

# Αναλύσεις προσομοιώσεων μοριακής δυναμικής

## 4.1 Script για την αλλαγή φοράς των μονομερών C και D

```
#include <stdio.h>
#include <math.h>

main()
{
float x[100];
int N;
int i;

N=1;
while( scanf("%f", &x[N]) == 1 )
{
N++;
}

for ( i=N-1 ; i > 0 ; i-- )
{
printf(" %f\n", x[i] );
}
}
```

## 4.2 Οι διακομάνσεις ανά κατάλοιπο (mkfluc.c)

```
#include <stdio.h>
#include <math.h>

main()
{
float x, y, z;
int c;

while( ( c = scanf("%f %f %f", &x, &y, &z) ) == 3 )
printf("%10.8f\n", sqrt( x*x + y*y + z*z));
if ( c > 0 )
fprintf(stderr, "ERROR. Not triplets ???\n");
}
```

## 4.3 Αθροιστική διακόμανση (Sumfluc.c)

```
#include <stdio.h>
#include <math.h>

main()
{
float x, y, z, v, w;
int c;

while( ( c = scanf("%f %f %f %f %f", &x, &y, &z, &v, &w) ) == 5 )
printf("%10.8f\n", ( x + y + z + v + w));
if ( c > 0 )
fprintf(stderr, "ERROR. ???\n");
}
```

#### 4.4 Υπολογισμός του *cosine contente* για κύριες συνιστώσες σε γλώσσα Perl.

```
#!/usr/bin/perl -w

use constant PI => 4 * atan2(1, 1);

(@ARGV==2) or die "Usage: cos_content.pl filename component\n";
open ( FILE, $ARGV[0] ) or die "Can not open $ARGV[0]\n";

$i = $ARGV[1];

$T = 0;
while ( $line = <FILE> )
{
    $T++;
}

close ( FILE );

open ( FILE, $ARGV[0] ) or die "Can not open $ARGV[0]\n";

$t = 0;
$sum_1 = 0.0;
$sum_2 = 0.0;
while ( $line = <FILE> )
{
    @q = split(' ', $line);
    $sum_1 += cos( $i * PI * $t / $T ) * $q[$i];
    $sum_2 += $q[$i] * $q[$i];
    $t++;
}

printf ("Cosine content for component $i is %6.4f\n", (2.0 * $sum_1 * $sum_1) / ( $T * $sum_2 ) );

close ( FILE );
```

#### 4.5 Μετατροπή των διεδρων γωνιών σε κλίμακα 0-360° (*angles180.c*)

```
#include <stdio.h>
#include <math.h>

main ()
{
    float val1;
    float val2;

    while ( scanf ( "%f %f", &val1, &val2)== 2 )

    {
        if ( val2 > 0.0 )
            printf( "%15.6f %15.6f\n", val1, val2);

        else
        {
            val2+=360.0;
            printf( "%15.6f %15.6f\n", val1, val2);
        }
    }
}
```

## Σύγκριση προσομοιώσεων των δύο μοντέλων

### 5.1 Υπολογισμός εμβαδού των επιπέδων του υδρόφοβου πυρήνα. (AreaLayers.pl)

```
#!/usr/bin/perl -w

while (1) {

$layer1 = <STDIN> or exit;
@pos1 = split(' ', $layer1);
$layer2 = <STDIN>;
@pos2 = split(' ', $layer2);
$layer3 = <STDIN>;
@pos3 = split(' ', $layer3);
$layer4 = <STDIN>;
@pos4 = split(' ', $layer4);

$dista = Distance( "$pos1[0]", "$pos1[1]", "$pos1[2]", "$pos2[0]", "$pos2[1]", "$pos2[2]");
$distB = Distance( "$pos1[0]", "$pos1[1]", "$pos1[2]", "$pos3[0]", "$pos3[1]", "$pos3[2]");
$distC = Distance( "$pos2[0]", "$pos2[1]", "$pos2[2]", "$pos3[0]", "$pos3[1]", "$pos3[2]");
$distD = Distance( "$pos1[0]", "$pos1[1]", "$pos1[2]", "$pos4[0]", "$pos4[1]", "$pos4[2]");
$distE = Distance( "$pos2[0]", "$pos2[1]", "$pos2[2]", "$pos4[0]", "$pos4[1]", "$pos4[2]");

$layerArea = Area( "$dista", "$distB", "$distC", "$distD", "$distE");

print "$layerArea\n";

}

sub Area {

$a = $_[0];
$b = $_[1];
$c = $_[2];
$d = $_[3];
$e = $_[4];

$T1 = ($a+$b+$c) * ($a+$b-$c) * ($b+$c-$a) * ($a+$c-$b);
$T1 = sqrt($T1)/4;

$T2 = ($a+$d+$e) * ($a+$d-$e) * ($d+$e-$a) * ($a+$e-$d);
$T2 = sqrt($T2)/4;

$area = $T1+$T2;

return ($area);
}

sub Distance {

$x1 = $_[0];
$y1 = $_[1];
$z1 = $_[2];

$x2 = $_[3];
$y2 = $_[4];
$z2 = $_[5];

$sumX = ($x1-$x2) * ($x1-$x2);
$sumY = ($y1-$y2) * ($y1-$y2);
$sumZ = ($z1-$z2) * ($z1-$z2);
$sum = $sumX + $sumY + $sumZ;
$dist = sqrt($sum);

return ($dist);
```