

Δημοκρίτειο Πανεπιστήμιο Θράκης
Τμήμα Μοριακής Βιολογίας και Γενετικής

Διπλωματική εργασία

Προσομοιώσεις αναδίπλωσης της χολινο-προσδενόμενης επανάληψης από την πρωτεΐνη LytA του *Streptococcus pneumoniae* και σύγκριση με τις πειραματικά προσδιορισμένες δομές.

Πατμανίδης Ηλίας

AEM -796-

Επιβλέπων καθηγητής:

Νικόλαος Μ. Γλυκός

Περίληψη

Το πρόβλημα της αναδίπλωσης των πρωτεϊνών αποτελεί ένα από τα άλυτα προβλήματα της βιολογίας. Για την επίλυση του προβλήματος της αναδίπλωσης έχουν χρησιμοποιηθεί διάφορες προσεγγίσεις και έχουν διατυπωθεί αρκετές θεωρίες. Στην εργασία αυτή, θα μελετήσουμε υπολογιστικά τη συμπεριφορά του 14-πεπτιδίου DKFEKIDGTWYYFD. Το πεπτίδιο αυτό είναι μετάλλαγμα του LytA197-210[wt], το οποίο αποτελεί τμήμα της κύριας αυτολυσίνης του βακτηρίου *Streptococcus pneumoniae*. Πειράματα έδειξαν πως τόσο το πεπτίδιο άγριου τύπου όσο και το μετάλλαγμα του αναδιπλώνονται σε υδατικά διαλύματα σε δομές παρόμοιες με τη φυσική κατάσταση, οι οποίες έχουν τη δυνατότητα να προσδέουν χολίνη και ανάλογά της. Στόχος της εργασίας είναι η επίλυση της αναδίπλωσης του πεπτιδίου LytA197-210[ND] με τη χρήση physics-based μεθόδων και η σύγκριση των αποτελεσμάτων των προσομοιώσεων με τα πειραματικά δεδομένα, καθώς και η σύγκριση των αποτελεσμάτων μεταξύ διαφορετικών δυναμικών πεδίων. Σύμφωνα με τα αποτελέσματα των αναλύσεων, το πεπτίδιο για το μεγαλύτερο διάστημα των προσομοιώσεων υιοθετεί ημισταθερές διαμορφώσεις, οι οποίες διαφέρουν από τη φυσική κατάσταση. Παράλληλα εντοπίστηκε ένα μοναδικό γεγονός αναδίπλωσης με απόλυτη συμφωνία με την κρυσταλλογραφική δομή. Τα δεδομένα αυτά υποδηλώνουν την ύπαρξη πολλών τοπικών ελαχίστων στο ενεργειακό πεδίο της αναδίπλωσης του πεπτιδίου, που λειτουργούν ως κινητικές παγίδες και καθυστερούν την αναδίπλωση του πεπτιδίου.

Abstract

Protein folding is considered to be one of the unsolved problems of biology. Different approaches have been used and many theories have been proposed, in order to solve this problem. In this work we will study the behaviour of the 14-peptide DKFEKIDGTWYYFD, a mutant of LytA197-210[wt]. It is experimentally known that both the wild type peptide and the mutant have the ability to fold in aqueous solution and recognize choline and other molecules such as cholin. The aim of this work is studying the folding of LytA197-210[ND] by using physics-based methods and the comparison of the results with the experimental findings, as well as the comparison of the results obtained from the different force fields. According to the results of our analysis, this peptide spent most of the simulation time adopting semi-stable conformation, that differ from the native state. We characterized one unique folding event with excellent agreement with the crystallographic structure. The results indicate the presence of local minima in the energy landscape (kinetic frustration), which act like kinetic traps and slow down the peptide's folding.

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή	σελ.
1.1 Πρωτεϊνική αναδίπλωση	1
1.2 Το πρόβλημα της αναδίπλωσης των πρωτεϊνών	3
1.3 Μοντέλα αναδίπλωσης πρωτεϊνών	5
1.4 Υπολογιστικές προσεγγίσεις για την επίλυση του προβλήματος της αναδίπλωσης των πρωτεϊνών	8
1.5 Πεπτίδια ως συστήματα μοντέλα για την αναδίπλωση των πρωτεϊνών	10
1.6 Η πρωτεΐνη LytA και το πεπτίδιο LytA197–210[ND]	10
Κεφάλαιο 2: Εισαγωγή στις προσομοιώσεις μοριακής δυναμικής	
2.1 Εισαγωγή	14
2.2 Κινητική θεωρία και στατιστική μηχανική	16
2.3 Προσομοιώσεις Μοριακής Δυναμικής	16
2.4 Δυναμικά Πεδία	20
2.5 AMBER	23
2.6 Μοντέλα νερού	23
Κεφάλαιο 3: Προσομοιώσεις	
3.1 Τεχνικά χαρακτηριστικά υπολογιστών	25
3.2 Έναρξη προσομοιώσεων μοριακής δυναμικής με NAMD	27
3.3 Προετοιμασία συστημάτων	30
Κεφάλαιο 4: Ανάλυση προσομοιώσεων	
4.1 Εισαγωγή	33
4.2 Eigenvector overlap	36
4.3 RMSD από την φυσική δομή	37
4.4 AMBER-99SB	40
4.5 ILDN	47
4.6 ILDN-TIP4	56
4.7 ILDN-340K	63

4.8	ILDN-360K	70
4.9	Σύγκριση αποτελεσμάτων	78
4.10	RMSD από πειραματικά προσδιορισμένες δομές	80
4.11	Πρώτα συμπεράσματα	84

Κεφάλαιο 5: Συμπεράσματα και συζήτηση

5.1	Συμπεράσματα και συζήτηση	88
-----	---------------------------	----

Κεφάλαιο 1

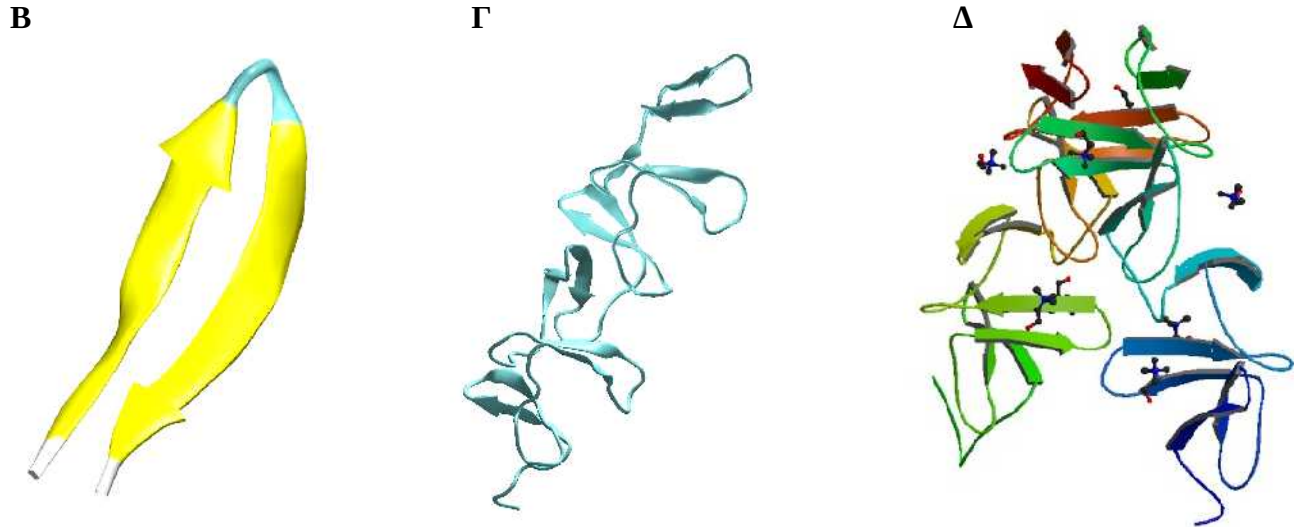
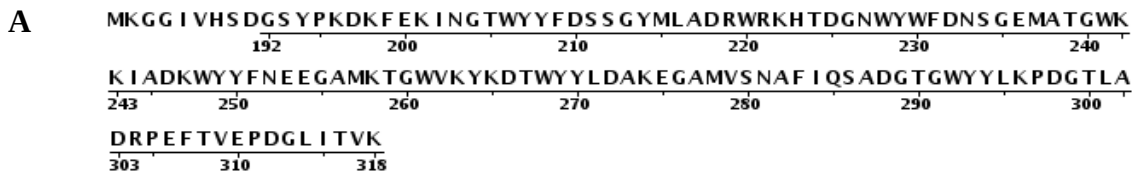
Εισαγωγή

Only entropy comes easy

Anton Chekhov

1.1 Πρωτεϊνική αναδίπλωση

Οι πρωτεΐνες αποτελούν ένα από τα θεμελιώδη μόρια των οργανισμών, τα οποία λειτουργούν είτε ως δομικά στοιχεία ή ως λειτουργικά μόρια, καταλύτες, σε πολλές διεργασίες. Ο δομικός λίθος των πρωτεϊνών είναι τα αμινοξέα, τα οποία ενώνονται μεταξύ τους με πεπτιδικούς δεσμούς για να σχηματίσουν πολυπεπτιδικές αλυσίδες. Για να είναι λειτουργική μια πρωτεΐνη, θα πρέπει η πολυπεπτιδική της αλυσίδα ή το σύνολο αυτών να έχουν τη σωστή τριτοταγή δομή. Η διαδικασία με την οποία οι πολυπεπτιδικές αλυσίδες αναδιπλώνονται για να σχηματίσουν πρωτεΐνες στη φυσική τους διαμόρφωση (native state) ονομάζεται πρωτεϊνική αναδίπλωση.



Εικόνα 1.1 . Α) Πρωτοταγής δομή επικράτειας πρόσδεσης της χολίνης της πρωτεΐνης LytA. Β) Δευτεροταγής δομή του πεπτιδίου LytA[197-210] σε cartoon, με γαλάζιο χρώμα έχουν χρωματιστεί τα κατάλοιπα της στροφής, με κίτρινο όσα συμμετέχουν στο σχηματισμό της β-φουρκέτας και με άσπρο τα κατάλοιπα των άκρων. Γ) Τριτοταγής δομή του μονομερούς της LytA. Δ) Τεταρτοταγής δομή του ομοδιμερούς της LytA, στην εικόνα παρουσιάζονται και μόρια χολίνης σε ball and stick(οι εικόνες αναπαράγονται άνευ αδειάς)

Η ενεργειακή διαφορά που παρουσιάζει η αποδιαταγμένη κατάσταση από τη διπλωμένη, επηρεάζεται από δυο παράγοντες, την εντροπία και την ενθαλπία. Η εντροπία προκύπτει σαν μια ποσοτική διατύπωση του δευτέρου νόμου της θερμοδυναμικής. Από τη σκοπιά της στατιστικής μηχανικής είναι το μέτρο της αταξίας ενός συστήματος και μπορεί να χρησιμοποιηθεί για να εκφράσει τον αριθμό των διαμορφώσεων που θα μπορούσε να αποκτήσει ένα σύστημα. Οι πρωτεΐνες στη φυσική τους κατάσταση αποκτούν μία συμπαγή δομή με υψηλό βαθμό οργάνωσης, ενώ στην αποδιαταγμένη κατάσταση είναι ξεδιπλωμένες και μπορούν να υιοθετήσουν πολλές πιθανές διαμορφώσεις. Η ενθαλπία είναι το μέτρο της της ενέργειας που προέρχεται από μη ομοιοπολικές αλληλεπιδράσεις των ατόμων των πρωτεϊνών όπως Van der Waals, δεσμούς υδρογόνου κ.α. Η ενεργειακή συνεισφορά των μη ομοιοπολικών αλληλεπιδράσεων είναι πολύ σημαντική, διότι υπάρχει μεγάλος βαθμός διαφορετικότητας μεταξύ αποδιαταγμένων δομών και της φυσικής κατάστασης και οι αλληλεπιδράσεις αυτές συνεισφέρουν με τον καλύτερα δυνατό τρόπο ώστε να δημιουργηθεί ένα συμπαγές μόριο. Η ολική διαφορά ενέργειας μιας πρωτεΐνης μεταξύ της αποδιαταγμένης και της φυσικής κατάστασης αποτελείται από το άθροισμα των

διαφορών της εντροπίας και της ενθαλπίας και ονομάζεται διαφορά ελεύθερης ενέργειας.

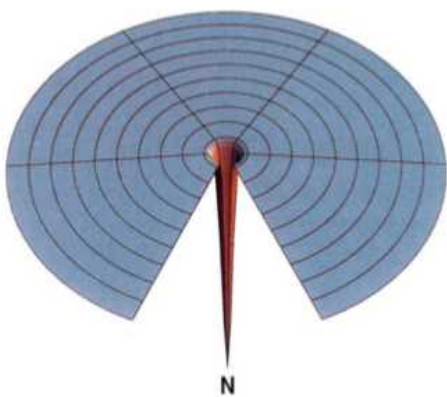
Το γεγονός ότι η διαφορά ελεύθερης ενέργειας είναι της τάξης των 5-15 kcal/mol υποδηλώνει ότι στους υπολογισμούς θα πρέπει να λαμβάνονται υπόψη όλες οι αλληλεπιδράσεις μεταξύ των ατόμων μιας πρωτεΐνης καθώς και των ατόμων του διαλύτη. Από τις παρατηρήσεις προκύπτει ότι η αναδίπλωση δεν είναι πιθανό να κυριαρχείται από ηλεκτροστατικές αλληλεπιδράσεις μεταξύ των φορτισμένων καταλοίπων, λόγω της χαμηλής συχνότητας τους και της κύριας συγκέντρωσής τους σε περιοχές με μεγάλες διηλεκτρικές σταθερές. Οι δεσμοί υδρογόνου φαίνεται να παίζουν σημαντικό ρόλο στη σταθερότητα των μορίων στη φυσική τους κατάσταση και αποτελούν συστατικά κλειδί όλων των στοιχείων δευτεροταγούς δομής. Αν και τα σημερινά δεδομένα υποδεικνύουν τη σημασία των δεσμών υδρογόνου, των δυνάμεων Van der Waals, των υδρόφοβων αλληλεπιδράσεων, των πλευρικών ομάδων των καταλοίπων και διάφορων άλλων στοιχείων, δεν υπάρχει κάποιο γενικό μοντέλο που να μπορεί να προβλέψει με λεπτομέρεια τη συμπεριφορά μιας πρωτεΐνης από την αποδιαταγμένη μέχρι τη φυσική κατάσταση. Ο μηχανισμός με τον οποίο μια πρωτεΐνη αναδιπλώνεται στη φυσική της κατάσταση και ο τρόπος με τον οποίο η αλληλουχία των αμινοξέων επηρεάζει την αναδίπλωση αποτελούν δύο εκφράσεις του προβλήματος που προκύπτει. [1, 2, 3]

1.2 Το πρόβλημα της αναδίπλωσης των πρωτεϊνών

Το πρόβλημα της αναδίπλωσης των πρωτεϊνών αποτελεί ένα από τα άλυτα προβλήματα της βιολογίας. Η πρώτη αντίληψη του προβλήματος προέκυψε κάπου στη δεκαετία του '60 μετά τον προσδιορισμό των πρώτων πρωτεϊνικών δομών με ατομική λεπτομέρεια και περιστρεφόμενα γύρω από το ερώτημα, πώς η αλληλουχία των αμινοξέων μιας πρωτεΐνης καθορίζει την τρισδιάστατη δομή της. Σημαντική συνεισφορά στην κατανόηση και την επίλυση του προβλήματος ήταν τα συμπεράσματα που προέκυψαν από τα πειράματα του Anfinsen [4, 5]. Ο Anfinsen και οι συνεργάτες του κατέληξαν στο ότι οι φυσικές δομές των πρωτεϊνών είναι οι θερμοδυναμικά σταθερές και οι διαμορφώσεις αυτές βρίσκονται σε ολικά ελάχιστα της ελεύθερης ενέργειας. Ένα άλλο συμπέρασμα ήταν ότι η φυσική δομή των πρωτεϊνών είναι ανεξάρτητη από την κινητική πορεία της αναδίπλωσης και εξαρτάται μόνο από την αλληλουχία των αμινοξέων και τις συνθήκες του διαλύματος. Αν υποθέσουμε ότι έχουμε μια αλληλουχία 100 αμινοξέων και ότι υπάρχουν δύο πιθανές διαμορφώσεις για κάθε αμινοξύ, τότε υπάρχουν συνολικά 10^{30} διαμορφώσεις. Αν το πεπτίδιο μπορεί να αλλάζει από τη μια διαμόρφωση στην άλλη σε διάστημα 10^{-11} s, ο συνολικός χρόνος που χρειάζεται για την τυχαία αναζήτηση της κατάλληλης διαμόρφωσης ανάμεσα σε όλες

τις υπόλοιπες διαμορφώσεις είναι περίπου 10^{19} s. Ο χρόνος αυτός ξεπερνά την ηλικία του σύμπαντος και φυσικά δεν μπορεί να ανταποκρίνεται στην πραγματικότητα. Στο παραπάνω παράδειγμα συνοψίζεται ο προβληματισμός που έθεσε ο Levinthal στα τέλη της δεκαετίας του '60, που αργότερα ονομάστηκε το παράδοξο του Levinthal (Levinthal's paradox).

Μια σχηματική αναπαράσταση του προβλήματος παρουσιάζεται στην Εικόνα 1.2, όπου όπως φαίνεται όλες οι διαμορφώσεις της πολυπεπτιδικής αλυσίδας είναι εξίσου πιθανές και το πεπτίδιο περιπλανιέται μεταξύ των πιθανών διαμορφώσεων μέχρι να καταλήξει στην φυσική δομή. Αυτό το σχεδιάγραμμα της εσωτερικής ενέργειας ενός πεπτιδίου αναφέρεται συχνά ως “γήπεδο του γκολφ” (golf-course). Το συμπέρασμα του Levinthal ήταν ότι είναι απίθανο η αναζήτηση της φυσικής δομής να γίνεται τυχαία και ότι υπάρχουν συγκεκριμένα και σαφώς προσδιορισμένα μονοπάτια (pathways).[6] Κατά συνέπεια, η πρωτεϊνική αναδίπλωση βρίσκεται κάτω από κινητικό έλεγχο, δηλαδή η αναδίπλωση είναι μία πολύ σύντομη διαδικασία αφού ακολουθεί κάποιο συγκεκριμένο μονοπάτι και αφού ακολουθεί κάποιο συγκεκριμένο μονοπάτι μπορεί να περνά μόνο από τοπικές βέλτιστες διαμορφώσεις. Τα πειράματα που ακολούθησαν κινούνταν γύρω από την ιδέα ότι η κινητική μπορεί να μας διδάξει πολλά για τον κώδικα της αναδίπλωσης και γι' αυτό δόθηκε ιδιαίτερη προσοχή στον εντοπισμό ενδιάμεσων διαμορφώσεων πάνω στην τροχιά των μονοπατιών. Από τότε έχουν αναπτυχθεί αρκετά μοντέλα με στόχο να μειωθεί ο υπολογιζόμενος χρόνος αναδίπλωσης ώστε να πλησιάζει τα πειραματικά δεδομένα και να ελαχιστοποιηθεί ο όγκος του προβλήματος περιορίζοντας τις πιθανές διαμορφώσεις που πρέπει να διερευνηθούν. [1, 3, 7, 8]



Εικόνα 1.2 Ενεργειακό πεδίο τύπου “γήπεδο γκολφ”. Η αλυσίδα ερευνά τυχαία τις πιθανές διαμορφώσεις ώσπου να καταλήξει στη φυσική κατάσταση

1.3 Μοντέλα αναδίπλωσης πρωτεϊνών

Ένα πιθανό μοντέλο είναι δημιουργία “υδρόφοβων πυρήνων” (hydrophobic core) [9, 10]. Οι πολυπεπτιδικές αλυσίδες πολλές φορές, στα πρώτα στάδια της αναδίπλωσης, θάβουν τις υδρόφοβες πλευρικές ομάδες σχηματίζοντας συμπαγείς δομές. Το αποτέλεσμα είναι να μειώνεται σημαντικά ο αριθμός των πιθανών διαμορφώσεων, ενώ ταυτόχρονα υπάρχει η τάση σχηματισμού στοιχείων δευτεροταγούς δομής, α -ελικών και β -φύλλων. Αυτό συμβαίνει γιατί οι πολικές ομάδες της κύριας αλυσίδας των μερικώς θαμμένων υδρόφοβων αμινοξέων είναι εκτεθειμένες και η κάλυψη τους δεν είναι ενεργειακά ευνοϊκή εκτός και αν σχηματίσουν δεσμούς υδρογόνου μεταξύ τους. Τα στοιχεία δευτεροταγούς δομής είναι ο συνηθέστερος και απλούστατος τρόπος. Η δημιουργία των στοιχείων δευτεροταγούς δομής στα αρχικά στάδια της αναδίπλωσης είναι πιθανό να οφείλεται στο θάψιμο των υδρόφοβων πλευρικών ομάδων. [1]

Η άμεση μετάπτωση από την αποδιαταγμένη πολυπεπτιδική αλυσίδα σε μια μερικώς οργανωμένη κατάσταση όπου έχει σχηματιστεί το μεγαλύτερο τμήμα της δευτεροταγούς δομής, οδηγεί στο σχηματισμό μίας δομής που ονομάζεται “εύπλαστη σφαίρα” (molten globule) [11, 12]. Στη φάση αυτή, οι εσωτερικές πλευρικές ομάδες παρουσιάζουν μεγάλη κινητικότητα και τα στοιχεία που βρίσκονται εκτεθειμένα στην επιφάνεια του μορίου, όπως βρόγχοι, υιοθετούν πολλές διαμορφώσεις. Στην επόμενη φάση, η οποία μπορεί να διαρκέσει μέχρι και ένα δευτερόλεπτο, το πολυπεπτίδιο αρχίζει σιγά σιγά να πακετάρεται ενώ κάνουν την εμφάνιση τους στοιχεία τριτοταγούς δομής με τη μορφή υπο-επικρατειών. Στην τελική φάση, η αναδίπλωση ολοκληρώνεται με τη σταθεροποίηση εσωτερικών και εξωτερικών στοιχείων και με το σχηματισμό αλληλεπιδράσεων σε ολόκληρη την πρωτεΐνη. [1]

Μικρές πρωτεΐνες της τάξης των 100 ή λιγότερων αμινοξέων αναδιπλώνονται συνήθως με κινητική δύο σταδίων (two-state kinetics) [13]. Η διάρκεια της αναδίπλωσης για τέτοια μόρια ξεκινά από μερικά μ s. Τα τελευταία χρόνια, έχει προταθεί η ύπαρξη ενδιάμεσων μερικώς αποδιαταγμένων διαμορφώσεων, οι οποίες είναι αρκετά δύσκολο να μετρηθούν με πειράματα κινητικής. Αν και πρόσφατες μελέτες προτείνουν νέους τρόπους για τον εντοπισμό των δομών αυτών και τη σημασία τους στην αναδίπλωση, οι διαμορφώσεις είναι συνήθως δύσκολο να εντοπιστούν είτε γιατί είναι σπάνιες στον πληθυσμό ή διαρκούν μικρά χρονικά διαστήματα.

Σύμφωνα με το μοντέλο diffusion-collision [16, 17] στα πρώτα στάδια της αναδίπλωσης σχηματίζονται μικροεπικρατείες, οι οποίες μπορεί να είναι τμήματα στοιχείων δευτεροταγούς

δομής, όπως α-έλικες και β-φύλλα) ή υδρόφοβων πυρήνων. Οι μικροεπικράτειες αυτές κινούνται και συγκρούονται μεταξύ τους για να σχηματίσουν μεγαλύτερες δομές. Η πληροφορία φαίνεται να βρίσκεται στη δημιουργία των μικροεπικρατειών και όχι σε κάθε αμινοξύ ξεχωριστά. Με τον τρόπο αυτό μειώνονται σε μεγάλο βαθμό οι πιθανές διαμορφώσεις και περιορίζεται το πρόβλημα της εύρεσης της κατάλληλης διαμόρφωσης. Η αναδίπλωση μπορεί να ακολουθεί ένα συγκεκριμένο και σαφώς προσδιορισμένο μονοπάτι, είτε πολλαπλά μονοπάτια από παράλληλα γεγονότα αναδίπλωσης.[3]

Ένα άλλο μοντέλο είναι ο μηχανισμός “ενπυρήνωσης-συμπήκνωσης” (nucleation-condensation) [18, 19], ο οποίος προτείνει τον αρχικό σχηματισμό ενός πυρήνα. Ο πυρήνας αυτός παρουσιάζει μεγάλη σταθερότητα λόγω των στοιχείων δευτεροταγούς δομής, τα οποία διακρίνονται για τις τεταρτοταγείς επαφές που σχηματίζουν και είναι ίδιες με αυτές της φυσικής κατάστασης. Στη συνέχεια, γύρω από τον πυρήνα σχηματίζονται περαιτέρω δομές αφού ο ίδιος δρα σαν βάση για τη σωστή διεύθυνση των υπόλοιπων στοιχείων στο χώρο. Η βασική διαφορά του μοντέλου αυτού είναι η συνύπαρξη στοιχείων δευτεροταγούς και τεταρτοταγούς δομής.

Οι πιο πρόσφατες απόψεις διαφέρουν από τα προηγούμενα φαινομενολογικά μοντέλα και δίνουν ιδιαίτερη έμφαση στα χαρακτηριστικά των ενεργειακών πεδίων της πολυπεπτιδικής αλυσίδας. Τα νέα αυτά μοντέλα αναφέρονται ως “χωριά αναδίπλωσης” (folding funnels) ή “ενεργειακά πεδία” (energy landscapes). Το ιδιαίτερο χαρακτηριστικό τους είναι ότι χρησιμοποιούν μεθόδους στατιστικής μηχανικής με απλοποιημένες αναπαραστάσεις των δομών και των αλληλεπιδράσεων, οι οποίες έχουν τη δυνατότητα να υπολογίζουν τις μικροσκοπικές αλληλεπιδράσεις που αναπτύσσονται μεταξύ των πρωτεϊνών. Σε αντίθεση με τα προηγούμενα μοντέλα, εισάγεται η έννοια του συνόλου των διαμορφώσεων για την περιγραφή των καταστάσεων. Η ιδέα των πιθανών μονοπατιών αναδίπλωσης με διαδοχικά γεγονότα αντικαθίσταται από πολλαπλά μονοπάτια με παράλληλα γεγονότα που μοιάζουν με διάχυση μέχρι όλα να καταλήξουν στην κανονική δομή. Αυτή η αντιμετώπιση ελαχιστοποιεί το παράδοξο της χρονοβόρας αναζήτησης. Η σχηματική αναπαράσταση των ενεργειακών πεδίων αποτελείται από δισδιάστατα και τρισδιάστατα διαγράμματα που παρουσιάζουν τη διακύμανση της ελεύθερης ενέργειας των διαμορφώσεων ως συνάρτηση των βαθμών ελευθερίας, Εικόνες 1.2-5 [7] (οι εικόνες αναπαράγονται άνευ αδειάς). Ο οριζόντιος άξονας αναπαριστά την εντροπία, η οποία μειώνεται όσο πλησιάζει η πρωτεΐνη την κανονική της δομή και ο κάθετος άξονας την “εσωτερική ελεύθερη ενέργεια”, δηλαδή την ενέργεια των δεσμών υδρογόνου, των ιοντικών αλληλεπιδράσεων και

πολλών άλλων αλληλεπιδράσεων της πολυπεπτιδικής αλυσίδας, εκτός της εντροπίας. Η δυσκολία στην εύρεση των κατάλληλων διαμορφώσεων είναι η ουσία του παραδόξου του Levinthal. Η ελεύθερη ενέργεια προκύπτει από το αλγεβρικό άθροισμα της ενθαλπίας και της εντροπίας:

$$\Delta G_T = \Delta H_T - T\Delta S_T$$

Η ενθαλπία υπολογίζεται από τη σχέση:

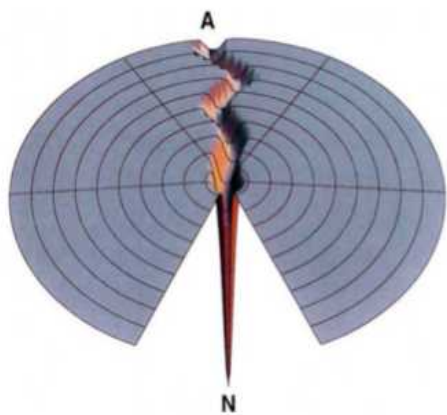
$$H = E + PV$$

όπου E είναι η εσωτερική ενέργεια, P είναι η πίεση και V ο όγκος του συστήματος.

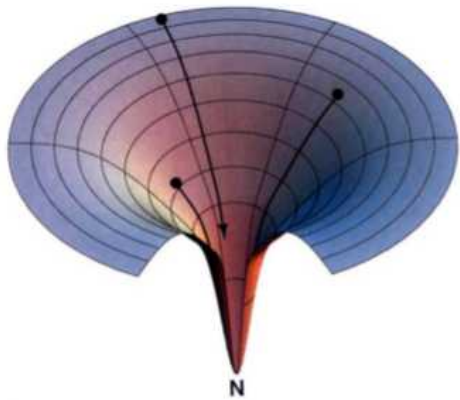
Η εσωτερική ενέργεια E είναι το άθροισμα της συνολικής κινητικής & δυναμικής ενέργειας των ατόμων του συστήματος:

$$E = \sum E_{\text{κινητική}} + \sum E_{\text{δυναμική}}$$

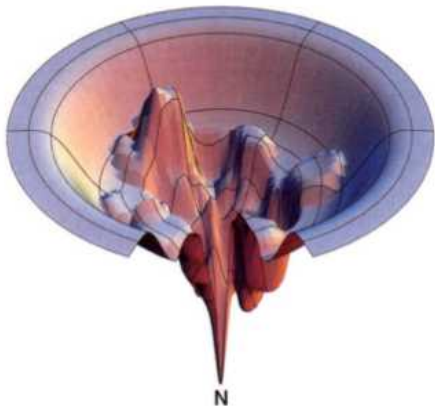
Η δυναμική ενέργεια οφείλεται στις αλληλεπιδράσεις και τους δεσμούς μεταξύ ατόμων των ατόμων του συστήματος και μειώνεται κατά την πορεία προς την κανονική δομή, ευνοώντας την αναδίπλωση, παράλληλα με τη δυσμενή συνεισφορά της εντροπίας. Η αλγεβρική σύνδεση της ελεύθερης ενέργειας και της εντροπίας οδηγεί στον σχηματισμό φραγμάτων ελεύθερης ενέργειας και οφείλεται για τη μορφή των διαγραμμάτων που μοιάζουν με χωνί. [7, 20, 43]



Εικόνα 1.3. Αναπαράσταση ενός μονοπατιού σε ένα ενεργειακό πεδίο. Η αναζήτηση περιορίζεται σε ένα συγκεκριμένο μονοπάτι διαμορφώσεων, μειώνοντας τις πιθανές διαμορφώσεις και το χρόνο της αναδίπλωσης



Εικόνα 1.4. Εξιδανικευμένη αναπαράσταση ενεργειακού πεδίου σε σχήμα χωνιού. Η ελεύθερη ενέργεια μειώνεται, όσο αναπτύσσονται εσωτερικές αλληλεπιδράσεις που οδηγούν στην φυσική κατάσταση



Εικόνα 1.5. Μια πιο ρεαλιστική απεικόνιση ενός ενεργειακού πεδίου. Φαίνονται τα ενεργειακά εμπόδια, καθώς και τα τοπικά ελάχιστα ενέργειας, ενώ σκιαγράφονται πιθανά μονοπάτια

1.4 Υπολογιστικές προσεγγίσεις για την επίλυση του προβλήματος της αναδίπλωσης των πρωτεϊνών

Για την κατανόηση της αναδίπλωσης των πρωτεϊνών έχουν χρησιμοποιηθεί θεωρητικές, πειραματικές καθώς και υπολογιστικές προσεγγίσεις. Συνοπτικά, οι πειραματικές προσεγγίσεις που έχουν χρησιμοποιηθεί για να ρίξουν φως στο πρόβλημα της αναδίπλωσης είναι α) η κρυσταλλογραφία ακτίνων X [28], η οποία είναι η σημαντικότερη μέθοδος καθορισμού της δομής πεπτιδίων και πρωτεϊνών, β) ο πυρηνικός μαγνητικός συντονισμός (NMR) [29], ο οποίος μπορεί να δώσει αποτελέσματα υψηλής ακρίβειας σε ατομικό επίπεδο, όσον αφορά την συμπεριφορά των ατόμων σε ένα μόριο και να υπολογίσει την κατανομή των δομών σε ένα σύνολο διαμορφώσεων, γ) ο κυκλικός διχροϊσμός (circular dichroism) [21], για τον καθορισμό διαμορφώσεων και στοιχείων δευτεροταγούς δομής κατά τη διάρκεια της αναδίπλωσης πολλών μορίων, δ) μέθοδοι πρωτεϊνικής μηχανικής, για τη σύγκριση των ιδιοτήτων μεταλλαγμάτων σε σχέση με τις πρωτεΐνες αγρίου τύπου. Πρόσφατες εξελίξεις στις πειραματικές μεθόδους έδωσαν τη δυνατότητα για

λεπτομερέστερη ανάλυση των διαμορφώσεων σε ατομικό επίπεδο κατά τη διάρκεια της αναδίπλωσης. Τέτοιες μέθοδοι είναι mass spectrometry, high resolution-hydrogen exchange, fluorescence resonance energy transfer (FRET), atomic force microscopy (AFM). [7, 23, 24]

Το πρόβλημα με τις πειραματικές προσεγγίσεις είναι ότι δεν μπορούν να αποτυπώσουν με ατομική ακρίβεια τις πολλαπλές διαμορφώσεις μεταξύ της αποδιαταγμένης και της αναδιπλωμένης δομής. Οι υπολογιστικές προσεγγίσεις και τα μοντέλα είναι ο συνδετικός κρίκος μεταξύ των θεωρητικών και πειραματικών αποτελεσμάτων για την κατανόηση των μηχανισμών της αναδίπλωσης καθώς και την ερμηνεία των κινητικών και θερμοδυναμικών στοιχείων της. Υπάρχουν διάφοροι τύποι υπολογιστικών προσεγγίσεων όπως, α) *threading* [32] και β) *ab initio* [33]. Στην πρώτη προσέγγιση χρησιμοποιείται η αλληλουχία των αμινοξέων μιας πρωτεΐνης και υπολογίζεται το πόσο καλά ταιριάζει με κάποιες από τις γνωστές τριτοταγείς δομές πρωτεϊνών. Οι *ab initio* προσεγγίσεις χρησιμοποιούν αλγόριθμους για να υπολογίσουν τις συντεταγμένες των ατόμων μιας πρωτεΐνης καθώς και ενεργειακούς παράγοντες, με στόχο την εύρεση των διαμορφώσεων με τη χαμηλότερη δυναμική ενέργεια. Η μελέτη της αναδίπλωσης μπορεί να γίνει είτε μελετώντας την πορεία από την αποδιαταγμένη πρωτεΐνη προς την κανονική δομή ή το αντίστροφο(πειράματα *fold*ing και *un*fold*ing* αντίστοιχα). Η βασική ιδέα στα πειράματα *un*fold*ing* είναι ότι από τα αρχικά στάδια της αποδιάταξης μπορούμε να βγάλουμε συμπεράσματα για τα τελευταία στάδια της αναδίπλωσης.

Υπάρχουν διάφορα μοντέλα που έχουν χρησιμοποιηθεί σε προσομοιώσεις μοριακής δυναμικής. Μερικά από αυτά είναι μινιμαλιστικά, δηλαδή έχουν χαμηλή ατομική λεπτομέρεια, και στόχος τους ο εντοπισμός σημαντικών στοιχείων που επηρεάζουν την αναδίπλωση καθώς και τη βελτιστοποίηση των αλγορίθμων. Τέτοια μοντέλα είναι τα μοντέλα *lattice*, τα μοντέλα τύπου-Go κ.α, τα οποία έχουν γενικό χαρακτήρα και δεν μπορούν να δώσουν απαντήσεις όσον αφορά ερωτήματα που χρειάζονται την βαθύτερη κατανόηση των αλληλεπιδράσεων σε ατομικό επίπεδο. Μοντέλα τα οποία λαμβάνουν υπόψη ατομικές λεπτομέρειες έχουν φανεί ιδιαίτερα χρήσιμα. Οι προσομοιώσεις μοριακής δυναμικής είναι ένα πολύτιμο εργαλείο για την απόκτηση πληροφοριών σε ατομικό επίπεδο. Το πλεονεκτήματά τους είναι οι λεπτομερείς πληροφορίες για τη φύση και τη σχέση των διαμορφώσεων κατά τη διάρκεια της αναδίπλωσης καθώς και η δυνατότητα εντοπισμού ενδιάμεσων διαμορφώσεων, το οποίο είναι σημαντικό στοιχείο για την κατανόηση των μηχανισμών αναδίπλωσης. Εκτός από τις προσομοιώσεις μοριακής δυναμικής υπάρχουν και άλλες προσεγγίσεις για την πρόβλεψη της συμπεριφοράς πεπτιδίων και πρωτεϊνών, όπως οι

προσομοιώσεις Monte Carlo [45] που χρησιμοποιούν στατιστική και πιθανότητες για τη μελέτη μοριακών συστημάτων. [25, 26, 30, 44]

1.5 Πεπτίδια ως συστήματα μοντέλα για την αναδίπλωση των πρωτεϊνών

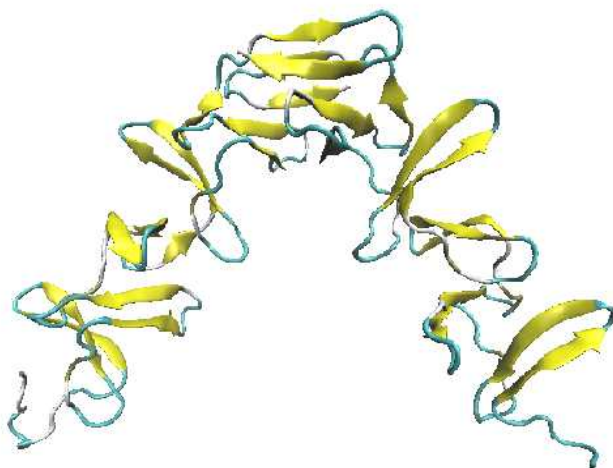
Οι προσομοιώσεις ολόκληρων μοριακών συστημάτων, όπως οι πρωτεΐνες, είναι περίπλοκες και χρονοβόρες, ενώ το υπολογιστικό κόστος τους είναι τις περισσότερες φορές απαγορευτικό. Ιδιαίτερα χρήσιμη στη μελέτη της αναδίπλωσης των πρωτεϊνών έχει αποδειχθεί η χρήση πεπτιδίων. Πεπτίδια οριοθετούνται πολυπεπτιδικές αλυσίδες με μήκος έως και 50 κατάλοιπα. Πειράματα έχουν δείξει ότι πεπτίδια σε διαλύματα έχουν την ικανότητα να σχηματίζουν φυσικές δευτεροταγείς δομές όπως α-έλικες [34] και β-φύλλα [35, 36]. Τέτοιου είδους αποτελέσματα ενισχύουν την υπόθεση, ότι σημαντικό κομμάτι της πληροφορίας για τη σωστή αναδίπλωση των πρωτεϊνών βρίσκεται τοπικά στην πολυπεπτιδική αλυσίδα και στρέφουν ένα κομμάτι της έρευνας στη μελέτη της αναδίπλωσης πεπτιδίων.

Οι προσομοιώσεις πεπτιδίων έχουν ως στόχο τη διαμόρφωση συμπερασμάτων για την καλύτερη κατανόηση των μηχανισμών της αναδίπλωσης καθώς και τη βελτιστοποίηση των αλγόριθμων που χρησιμοποιούνται στις σημερινές προσομοιώσεις ώστε να αναπτυχθούν καλύτερες physics-based μέθοδοι για την πρόβλεψη πρωτεϊνικών δομών. Αν και οι σημερινές physics-based μέθοδοι φαίνεται να παρουσιάζουν μεγάλη ακρίβεια, θα πρέπει να τονιστεί ότι τα αποτελέσματά τους θα πρέπει να αναλύονται από μια σκεπτικιστική σκοπιά. Το γεγονός ότι η πρωτεϊνική αναδίπλωση αποτελεί μια ιδιαίτερα πολύπλοκη διαδικασία που παρουσιάζει μεγάλη συνεργατικότητα, είναι οι λόγοι για τους οποίους αυτή η τακτική “διαίρει και βασίλευε” μπορεί να οδηγήσει σε παρερμηνείες. [27, 31]

1.6 Η πρωτεΐνη LytA και το πεπτίδιο LytA197–210[ND]

Οι αυτολυσίνες προσδένονται στα κατάλοιπα χολίνης των τειχικών οξέων του κυτταρικού τοιχώματος και στη συνέχεια υδρολύουν τους πεπτιδογλυκανικούς δεσμούς του. Η χολίνη είναι ένα σημαντικό μόριο πολλών βακτηρίων που αποτελεί δομικό στοιχείο των τειχικών οξέων τους και συμμετέχει στη φυσιολογική ανάπτυξη τους. Η ικανότητα των αυτολυσινών να υδρολύουν το πεπτιδογλυκανικό δίκτυο των βακτηρίων που τις εκφράζουν, υποδηλώνει την σημασία τους σε πολλές διαδικασίες του κυττάρου όπως η κυτταρική ανάπτυξη, ο πολλαπλασιασμός, η κυτταρική διαφοροποίηση κ.α. Η πρωτεΐνη LytA αποτελεί την κυριότερη αυτολυσίνη του βακτηρίου *Streptococcus pneumoniae*. Το μοριακό βάρος της LytA είναι 36.5-kDa και αποτελείται από 318

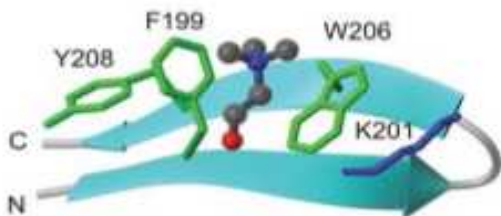
αμινοξικά κατάλοιπα. Η αναγνώριση και η πρόσδεση της με τα κατάλοιπα χολίνης της βακτηριακής επιφάνειας γίνεται μέσω συγκεκριμένων δομών στο C-τελικό άκρο της, οι οποίες ονομάζονται πρωτεΐνες που προσδέουν λυσίνη (choline-binding proteins, CBP). Οι CBP έχουν μελετηθεί εκτενώς και αποτελούν μερικές από τις καλύτερα μελετημένες πρωτεΐνες της κυτταρικής επιφάνειας. Πειράματα έχουν δείξει ότι η περιοχή πρόσδεσης της χολίνης αποτελείται από πολλαπλές επαναλήψεις καλώς συντηρημένων αλληλουχιών με μήκος 20 αμινοξικά κατάλοιπα, οι οποίες ονομάζονται επαναλήψεις πρόσδεσης χολίνης (choline-binding repeats, CBRs). Το πλήθος των επαναλήψεων μπορεί να κυμαίνεται από 3 έως 18, ενώ οι αλληλουχίες τους έχουν μεγάλη περιεκτικότητα σε αρωματικά κατάλοιπα. Οι επαναλήψεις αυτές αλληλεπιδρούν για να σχηματίσουν λειτουργικά ομοδιμερή, **Εικόνα 1.6**. Οι δομές αυτές φαίνεται να είναι υπεύθυνες για την ικανότητα της LytA να αναγνωρίζει τη κατάλοιπα χολίνης της βακτηριακής επιφάνειας, καθώς και την αγκυροβόλησή της. [37, 38, 39, 42]



Εικόνα 1.6. Η ομοδιμερής μορφή της πρωτεΐνης LytA. Η τεταρτοταγής δομή των μονομερών αποτελείται από έξι ανεξάρτητες β-φουρκέτες. Οι δομές με κίτρινο χρώμα αντιπροσωπεύουν τις β-φουρκέτες, ενώ οι γαλάζιες τις στροφές.

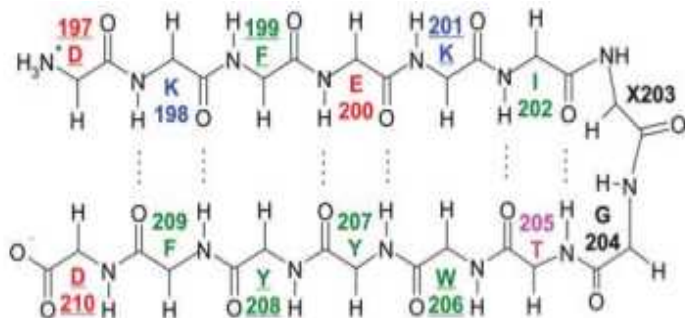
Οι πρώτες κρυσταλλογραφικές αναλύσεις της επικράτειας πρόσδεσης χολίνης της LytA έδειξαν μία β-σωληνοειδή διαμόρφωση, η οποία αποτελούνταν από επαναλήψεις διαμορφώσεων του τύπου στροφή-β-φουρκέτα που αντιστοιχούσαν στις CBRs [40]. Οι τέσσερις θέσεις πρόσδεσης χολίνης προκύπτουν από τη συνεισφορά των έξι β-φουρκετών, οι οποίες προσφέρουν δύο αρωματικά κατάλοιπα από τη μία αλυσίδα και ένα ακόμα από την επόμενη για τη δημιουργία τεσσάρων θέσεων πρόσδεσης της χολίνης, ενώ σημαντική είναι και η συνεισφορά μίας επιπλέον υδρόφοβης πλευρικής αλυσίδας, **Εικόνα 1.7** [38] (οι εικόνες αναπαράγονται άνευ αδειάς). Οι

στροφές παίζουν το ρόλο του συνδέσμου μεταξύ των φουρκετών. Η πρόσδεση και η σταθεροποίηση της χολίνης οφείλονται κυρίως σε υδρόφοβες αλληλεπιδράσεις. Αν και υπάρχουν μερικές διαφορές, αυτή η διαμόρφωση διατηρείται σε όλες τις CBRs που έχουν μελετηθεί μέχρι σήμερα. [38]



Εικόνα 1.7. Οι πλευρικές ομάδες των καταλοίπων που επηρεάζουν την ικανότητα πρόσδεσης της χολίνης

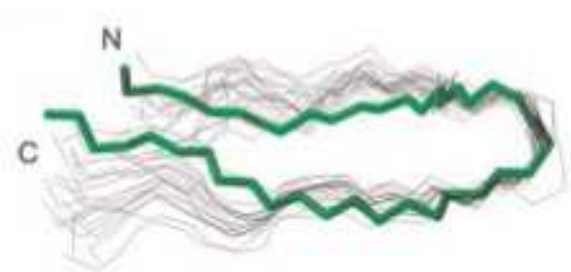
Στην εργασία αυτή, θα μελετήσουμε υπολογιστικά τη συμπεριφορά του 14-πεπτιδίου DKFEKIDGTWYYFD [38]. Το συγκεκριμένο πεπτίδιο είναι μετάλλαγμα ενός πεπτιδίου της *LytA*, συγκεκριμένα του *LytA*197–210[wt] **Εικόνα 1.8** (οι εικόνες αναπαράγονται άνευ αδείας), το οποίο αποτελεί μία από τις έξι β-φουρκέτες της αυτολυσίνης *LytA*. Από το αρχικό 20-πεπτίδιο έγινε η επιλογή να μη χρησιμοποιηθούν τα έξι κατάλοιπα των στρωφών, τα οποία πιθανώς θα ήταν ξεδιπλωμένα κατά τη διάρκεια της αναδίπλωσης λόγω της έλλειψης σταθερών αλληλεπιδράσεων με την υπόλοιπη πρωτεΐνη. Επιπλέον, πειράματα έδειξαν πως απομονωμένοι πυρήνες CBR αναδιπλώνονται σε υδατικά διαλύματα σε δομές παρόμοιες με τη φυσική κατάσταση, οι οποίες έχουν τη δυνατότητα να προσδέσουν χολίνη και ανάλογά της. [38]



Εικόνα 1.8 Σχηματική αναπαράσταση της β-φουρκέτας των *LytA*197–210[wt] και *LytA*197–210[ND]. Στο σχήμα φαίνονται και οι υδρογονοδεσμοί μεταξύ των καταλοίπων της κύριας αλυσίδας

Κρυσταλλογραφικά πειράματα του 14-πεπτιδίου *LytA*197–210[wt], DKFEKIDGTWYYFD, έδειξαν ότι σχηματίζει μια 2:2 β-φουρκέτα, καθώς και μία στροφή τύπου I, της οποίας η κύρια αλληλουχία NG φαίνεται να ευνοεί το σχηματισμό τέτοιου τύπου στρωφών [41]. Παρόλα αυτά, ο πλήρης δομικός χαρακτηρισμός του πεπτιδίου *LytA*197–210[wt] δεν ήταν εφικτός λόγω

προβλημάτων στα φάσματα NMR, τα οποία πιθανώς να οφείλονται στην απαμίνωση της ασπαραγίνης της στροφής που οδηγεί στην ύπαρξη δύο πεπτιδίων, από τα οποία το ένα φέρει το κατάλοιπο ασπαραγίνη ενώ το άλλο β-ασπαραγίνη. Το μετάλλαγμα με το οποίο θα ασχοληθούμε, το οποίο σχεδιάστηκε από τον Sanz και τους συνεργάτες του [38], διαφέρει ως προς το έβδομο κατάλοιπο της αλληλουχίας του LytA197–210[wt], όπου η ασπαραγίνη έχει αντικατασταθεί με ασπαρτικό οξύ, δηλαδή LytA197–210[ND]. Εκτός από το γεγονός ότι το ασπαρτικό οξύ δεν μπορεί να απαμινωθεί αφού δε φέρει αμινομάδα, άρα εμποδίζεται η παρουσία πολλαπλών δομών, το πεπτίδιο LytA197–210[ND] παρουσιάζει καλύτερη διαλυτότητα στο νερό και μικρότερη αυτοένωση λόγω της ύπαρξης περισσότερων καταλοίπων με αρνητικά φορτισμένες πλευρικές ομάδες. Επιπλέον γνωρίζουμε ότι πειράματα NMR έχουν δείξει ότι το LytA197–210[ND] μπορεί να αναδιπλώνεται σε δομές που μοιάζουν πολύ με αυτές της LytA και ότι η κεντρική αλληλουχία της στροφής DG είναι κατάλληλη για το σχηματισμό στροφών τύπου I. Ιδιαίτερα ενδιαφέρον είναι το γεγονός ότι η β-φουρκέτα του LytA197–210[ND] παρουσιάζει μεγάλη ομοιότητα με τη στροφή της κανονικής δομής της C-LytA, **Εικόνα 1.9**. Αν και υπάρχουν αρκετές διαφορές όσον αφορά τις πλευρικές ομάδες του πεπτιδίου και της κανονικής δομής, μερικές από αυτές παρουσιάζουν τον ίδιο προσανατολισμό. Για παράδειγμα η Tyr 207 παρουσιάζει χ_1 γωνίες παρόμοιες με τη LytA. Επιπλέον, οι πλευρικές ομάδες Phe 199 και Trp 206 των οποίων οι πλευρικές ομάδες συμμετέχουν στην πρόσδεση της χολίνης παρουσιάζουν μεγάλη ομοιότητα με τις αντίστοιχες της κανονικής δομής για τις περισσότερες από τις υπολογισμένες δομές του μεταλλάγματος. [38, 40]



Εικόνα 1.9. Υπέρθηση της κύριας αλυσίδας των υπολογισμένων δομών σε σύγκριση με τη δομή της φουρκέτας της LytA. Η κύρια αλυσίδα της πρωτεΐνης άγριου τύπου παρουσιάζεται με πράσινο χρώμα.

Στόχος της εργασίας είναι η μελέτη της αναδίπλωσης του 14-πεπτιδίου DKFEKIDGTWYYFD με τη χρήση physics-based μεθόδων και η σύγκριση των αποτελεσμάτων των προσομοιώσεων με άλλα πειραματικά δεδομένα, όπως πειραματικά προσδιορισμένες δομές από κρυσταλλογραφία ακτίνων-X και NMR, καθώς και η σύγκριση των αποτελεσμάτων μεταξύ διαφορετικών δυναμικών πεδίων.

Κεφάλαιο 2ο

Εισαγωγή στις προσομοιώσεις μοριακής δυναμικής

Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective positions of the beings which compose it, if moreover this intelligence were vast enough to submit these data to analysis, it would embrace in the same formula both the movements of the largest bodies in the universe and those of the lightest atom; to it nothing would be uncertain, and the future as the past would be present to its eyes.

Pierre Simon De Laplace, 1814

2.1 Εισαγωγή

Στην προσπάθεια μας να κατανοήσουμε τον κόσμο, λόγω της πολυπλοκότητας του προκύπτει πολλές φορές η ανάγκη να μελετάμε μικρότερα τμήματά του, έτσι ώστε να εξάγουμε ρεαλιστικά

συμπεράσματα που έχουν γενική ισχύ, μέχρι τουλάχιστον να αποδειχθεί το αντίθετο. Το τμήμα του φυσικού κόσμου που επιλέγουμε να μελετήσουμε κάθε φορά ονομάζεται σύστημα. Ένα σύστημα μπορεί να αποτελείται από όμοια ή διαφορετικά τμήματα. Τα άτομα και οι ιδιότητες τους περιγράφουν την κατάσταση του συστήματος, η οποία μπορεί να είναι αέρια, υγρή ή στερεή. Για να μελετήσουμε και να περιγράψουμε τη συμπεριφορά ενός συστήματος χρειαζόμαστε αριθμητικές τιμές οι οποίες περιγράφουν την κατάσταση και τις λειτουργίες του. Τέτοιες τιμές ονομάζονται παρατηρήσιμες. Για παράδειγμα έχουμε ένα σύστημα το οποίο περιέχει 10^{10} μόρια νερού. Η κατάσταση του περιγράφεται από τη θέση και την ορμή των μορίων, από τις οποίες προκύπτουν και οι παρατηρήσιμες τιμές, θερμοκρασία και πίεση. Για να αναλύσουμε και να ερμηνεύσουμε στη συνέχεια τις παρατηρήσεις μας φτιάχνουμε θεωρίες, οι οποίες πρέπει να επιβεβαιώνονται από πειράματα. Σε περιπτώσεις όπου δεν μπορούν να διεξαχθούν πειράματα που θα μας δώσουν τα απαιτούμενα αποτελέσματα χρησιμοποιούμε προσομοιώσεις με στόχο να αναπαράγουμε λεπτομερώς το σύνολο των αλληλεπιδράσεων στο σύστημα που μελετάμε. [49, 52]

Η ιστορία των προσομοιώσεων μοριακών συστημάτων ξεκινά στις αρχές της δεκαετίας του '50, όταν οι Metropolis, Rosenbluth, Rosenbluth, Teller and Teller εφάρμοσαν για πρώτη φορά τη μέθοδο Monte Carlo (MC) για τη μελέτη ενός συστήματος από μαλακές σφαίρες [45]. Η πρώτη προσομοίωση μοριακής δυναμικής σε ηλεκτρονικό υπολογιστή έγινε από τους Alder and Wainwright [46], οι οποίοι υπολόγισαν ιδιότητες ισορροπίας και ανισορροπίας σε σκληρά σφαιρίδια. Μερικά χρόνια αργότερα, οι Rahman και Stillinger [47] μελέτησαν την ιδιότητες του νερού σε υγρή μορφή χρησιμοποιώντας τεχνικές μοριακής δυναμικής, ενώ η πρώτη εφαρμογή μοριακής δυναμικής σε πρωτεΐνες έγινε από τους McCammon, Gelin, και Karplus [48]. Με το πέρασμα του χρόνου και την αύξηση της πρόσβασης των επιστημόνων στους ηλεκτρονικούς υπολογιστές καθώς και την αύξηση της υπολογιστικής τους δύναμης, άρχισε να διευρύνεται η εφαρμογή των προσομοιώσεων για τη μελέτη μεγαλύτερων και πολυπλοκότερων συστημάτων. Σήμερα είναι δυνατή η προσομοίωση λιπιδικών συστημάτων και συμπλεγμάτων DNA και πρωτεϊνών. Εδώ θα πρέπει να αναφερθεί ότι στο μεσοδιάστημα υπήρχε μια ισχυρή διαμάχη για το εάν οι προσομοιώσεις σε ηλεκτρονικούς υπολογιστές είναι θεωρίες ή πειράματα. Η μία τάση υποστήριζε ότι είναι καθαρή θεωρία αφού δεν διεξάγονται πειράματα σε πραγματικά συστήματα, ενώ στον αντίποδα στέκονταν η αντίληψη ότι οι προσομοιώσεις μπορούν να χρησιμοποιούνται όπως τα πειράματα για να επαληθεύονται οι θεωρίες. Στην πραγματικότητα, οι προσομοιώσεις σε υπολογιστές χρησιμοποιούνται παράλληλα με τις υπόλοιπες πειραματικές μεθόδους και συμπληρώνουν την πειραματική ανάλυση δίνοντας μας πληροφορίες οι οποίες δεν θα μπορούσαν

να αποκτηθούν διαφορετικά. Θα μπορούσαμε να πούμε ότι οι προσομοιώσεις αποτελούν το συνδυαστικό κρίκο μεταξύ της θεωρίας και των πειραμάτων. Αν και έχει γίνει μεγάλη πρόοδος στη μεθοδολογία και έχουν βελτιωθεί οι τεχνικές των προσομοιώσεων, οι αλγόριθμοι που χρησιμοποιούνται έχουν υποστεί πολύ μικρές αλλαγές, ενώ η βασική ιδέα παραμένει η ίδια. [49, 52, 54]

2.2 Κινητική θεωρία και στατιστική μηχανική

Για να αναλύσουμε και να εξηγήσουμε τη συμπεριφορά των παρατηρήσιμων τιμών φτιάχνουμε θεωρίες. Το επόμενο βήμα είναι να φτιάξουμε πιο περίπλοκες θεωρίες που θα συνδέουν τις παρατηρήσεις με το υπόβαθρο από το οποίο προκύπτουν, όπως η κινητική θεωρία που συνδέει την θερμοκρασία και την πίεση με την άτακτη κίνηση και τις αλληλεπιδράσεις των ατόμων ενός συστήματος. Μικροσκοπικά δεδομένα όπως η θέση και η ορμή των ατόμων μπορούν σε ορισμένες περιπτώσεις να χρησιμοποιηθούν για την περιγραφή της κατάστασης ενός μοριακού συστήματος, αν και είναι συνήθως προτιμότερο η περιγραφή και η μελέτη να γίνεται χρησιμοποιώντας μετρήσιμες μακροσκοπικές μεταβλητές όπως η πίεση, ο όγκος και η θερμοκρασία. Τα δεδομένα όμως που εξάγονται θεωρητικά μπορεί να διαφέρουν από τα δεδομένα που προκύπτουν από τις πειραματικές διαδικασίες. Η φύση της επαφής μας με τον κόσμο είναι τέτοια ώστε να μην μπορούμε να αποκτήσουμε την “πλήρη” γνώση και περιοριζόμαστε στο βαθμό που μας επιτρέπει η θεωρία. Θα μπορούσαμε να πούμε ότι η μελέτη του κόσμου μέσω λεπτομερώς προσδιορισμένων συστημάτων είναι μάλλον μια αφηρημένη διαδικασία. Η στατιστική μηχανική είναι ένα σημαντικό εργαλείο για τη ελαχιστοποίηση της αφαίρεσης και ενισχύει τη σύγκλιση της θεωρίας και των πειραματικών αποτελεσμάτων. Συγκεκριμένα, μελετάται η κατάσταση ενός συστήματος υπολογίζοντας τις παρατηρήσιμες τιμές ως συνάρτηση της πιθανότητας του συστήματος να βρίσκεται στη συγκεκριμένη κατάσταση, παρά ως συνάρτηση του υποβάθρου τους. [49, 50, 51]

2.3 Προσομοιώσεις Μοριακής Δυναμικής

Η μοριακή δυναμική (MD) εφαρμόζεται σε ένα ευρύ φάσμα επιστημονικών πεδίων που ξεκινά από εφαρμοσμένα μαθηματικά, φυσική, βιοχημεία μέχρι και επιστήμες υλικών. Οι προσομοιώσεις μοριακής δυναμικής (Molecular Dynamics Simulations) αποτελούν ένα χρήσιμο εργαλείο για τη μελέτη μοριακών συστημάτων σε ατομικό επίπεδο. Οι προσομοιώσεις μοριακής δυναμικής χρησιμοποιούνται για τον υπολογισμό της κίνησης των μορίων σε συστήματα στερεών, υγρών και αερίων. Η κίνηση είναι το συστατικό κλειδί της μελέτης, αφού περιγράφει το πως μεταβάλλονται οι θέσεις και οι ταχύτητες ως συνάρτηση του χρόνου. Μέσω των προσομοιώσεων είναι δυνατή η

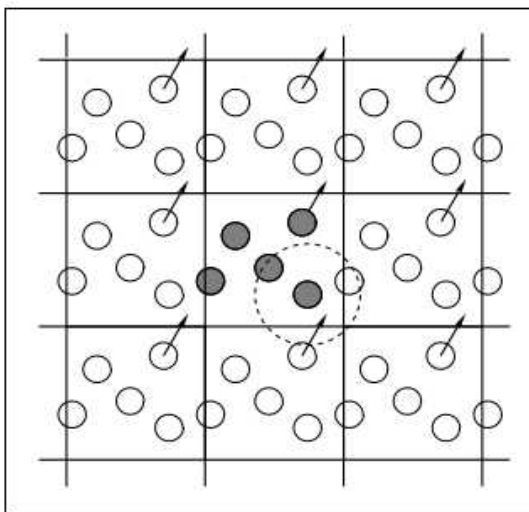
αναπαράσταση πολύπλοκων συστημάτων και η κατανόηση της συμπεριφοράς τους. Όπως έγινε ξεκάθαρο στο προηγούμενο κεφάλαιο η εργασία αυτή θα ασχοληθεί με την εφαρμογή προσομοιώσεων μοριακής δυναμικής στη μελέτη της αναδίπλωσης πρωτεϊνών *in silico*. Στόχος των προσομοιώσεων είναι η απόκτηση δεδομένων για τη συμπεριφορά ενός συστήματος σε μακροσκοπικό επίπεδο ως συνάρτηση των μικροσκοπικών αλληλεπιδράσεων. Η μοριακή δυναμική είναι στενά συνδεδεμένη με την κινητική θεωρία και λιγότερο με τη στατιστική μηχανική, οι οποίες δρουν ως συνδετικός κρίκος μεταξύ των μικροσκοπικών και μακροσκοπικών φαινομένων. Με τη χρήση προσομοιώσεων γίνεται εφικτή η λεπτομερής μελέτη και η πρόβλεψη των γεγονότων που λαμβάνουν χώρα σε ένα μοριακό σύστημα καθώς και ο υπολογισμός της δομής και της σταθερότητας του. Επιπλέον είναι δυνατή, η προσομοίωση συστημάτων που εξελίσσονται σε ακραίες τιμές μεταβλητών όπως η θερμοκρασία και η πίεση, τα οποία δεν είναι δυνατό να μελετηθούν πειραματικά. [49, 55, 56]

Τα στάδια για την έναρξη μιας προσομοίωσης είναι τα εξής:

1. Θέτουμε τις παραμέτρους που ορίζουν την κατάσταση του συστήματος, όπως αρχική θερμοκρασία, αριθμό σωματιδίων κ.α.
2. Αρχικοποιούμε το σύστημα ορίζοντας αρχικές θέσεις και ταχύτητες των σωματιδίων
3. Υπολογίζουμε τις δυνάμεις σε κάθε σωματίδιο, καθώς επίσης θέσεις και ταχύτητες σε όλη τη διάρκεια του τροχιακού
4. Συλλέγουμε και αναλύουμε τις τιμές που προκύπτουν από τα τροχιακά

Οι προσομοιώσεις μοριακής δυναμικής γίνονται συνήθως με μόρια που αποτελούνται από εκατοντάδες ή χιλιάδες άτομα. Σε τέτοιου είδους συστήματα κυριαρχούν δυνάμεις που οφείλονται στις αλληλεπιδράσεις των ατόμων με τα τοιχώματα του περιβάλλοντος στο οποίο λαμβάνει χώρα η προσομοίωση. Επειδή συνήθως μας ενδιαφέρει ποια θα ήταν η συμπεριφορά του υπό μελέτη μορίου σε ένα μεγάλο όγκο νερού, οι πιθανές αλληλεπιδράσεις του μορίου με τη στερεή επιφάνεια των τοιχωμάτων είναι αναγκαίο να αφαιρεθούν. Μία μέθοδος είναι η χρήση συνθηκών περιοδικών συνόρων (periodic boundary conditions). Σε αυτή τη μέθοδο, το υπό μελέτη μόριο τοποθετείται σε ένα συγκεκριμένο όγκο, ο οποίος αποτελεί το πρωταρχικό κελί. Για να προσομοιωθεί ο τεράστιος όγκος νερού θεωρείται ότι το σύστημα αποτελείται από αντίγραφα του πρωταρχικού κελιού προς όλες τις κατευθύνσεις. Ουσιαστικά, το σύστημα αποτελείται από περιοδικές επαναλήψεις του πρωταρχικού κελιού που ως στόχο έχουν τη δημιουργία ενός μακροσκοπικού δείγματος και την περιοδική έκφραση των θέσεων και των ορμών των ατόμων. Τα κελιά χωρίζονται μεταξύ τους από

νοητά σύνορα από τα οποία μπορούν ελεύθερα να περάσουν τα άτομα σε ένα διπλανό κελί. Όταν ένα άτομο ή ένα μέρος του μορίου περάσουν τα σύνορα από το πρωταρχικό κελί, άμεσα εισέρχεται το αντίστοιχο τμήμα από το αντιπαράλληλο κελί, **Εικόνα 2.1**.



Εικόνα 2.1. Periodic Boundary Conditions, το πρωταρχικό κελί βρίσκεται στο κέντρο και εφάπτεται με τα αντίγραφα του

Το πρόβλημα της μεθόδου αυτής είναι η τεχνητή περιοδικότητα που εισάγεται στο σύστημα και μπορεί να επηρεάσει τα αποτελέσματα της προσομοίωσης. Για να είναι ακριβείς οι προσομοιώσεις βιομορίων σε διαλύματα είναι ανάγκη να υπάρχει όσο το δυνατόν περισσότερη πληροφορία για τις εσωτερικές ιδιότητες των σωματιδίων του υπό μελέτη συστήματος. Γι' αυτό το λόγο προσομοιώσεις όλων των σωματιδίων με καλά προσδιορισμένες ιδιότητες θεωρείται ότι παρουσιάζουν την πιο ρεαλιστική προσέγγιση.

Στη μοριακή δυναμική οι θέσεις, οι ταχύτητες και οι επιταχύνσεις των σωματιδίων υπολογίζονται από τη λύση των Νευτώνειων διαφορικών εξισώσεων της κίνησης. Σύμφωνα με το γνωστό σε όλους μας, δεύτερο νόμο του Νεύτωνα:

$$F_i = m_i a_i$$

όπου F είναι η δύναμη που ασκείται σε ένα σωματίδιο, m η μάζα του και a η επιτάχυνσή του.

Για τον υπολογισμό λοιπόν ενός τροχιακού χρειάζονται η αρχικές θέσεις των ατόμων, μια αρχική κατανομή των ταχυτήτων και η επιτάχυνση. Οι αρχικές θέσεις μπορούν να αποκτηθούν από πειράματα, όπως κρυσταλλογραφία ακτίνων X ή φασματοσκοπία NMR . Η αρχική κατανομή των ταχυτήτων προκύπτει συνήθως από την κατανομή Maxwell-Boltzmann, από την οποία προκύπτει

η πιθανότητα ενός ατόμου i με μάζα m_i να έχει μια συγκεκριμένη ταχύτητα v_x προς την κατεύθυνση x σε μια συγκεκριμένη θερμοκρασία T .

$$p(v_{ix}) = (m_i / 2\pi k_B T)^{1/2} \exp(-m_i v_{ix}^2 / 2k_B T)$$

όπου k_B η σταθερά Boltzmann.

Γνωρίζοντας τη δύναμη που ασκείται σε κάθε άτομο μπορούμε να υπολογίσουμε την επιτάχυνση του. Για να υπολογίσουμε τη δύναμη χρησιμοποιούμε δυναμικά πεδία τα οποία θα αναλυθούν επόμενη παράγραφο. [49, 60]

Η επαναλαμβανόμενη χρονο-εξαρτώμενη ολοκλήρωση των κλασικών εξισώσεων της κίνησης παρέχει ουσιαστικά ατομικά τροχιακά από τα οποία μπορούν να υπολογιστούν στη συνέχεια μακροσκοπικά δεδομένα του συστήματος. Το αποτέλεσμα της χρήσης τέτοιων ολοκληρώσεων είναι ότι οι ατομικές θέσεις που προκύπτουν κατά τη διάρκεια του τροχιακού σχετίζονται χρονικά μεταξύ τους. Αυτή η ανάλυση ονομάζεται ντετερμινιστική. Σε ένα ντετερμινιστικό σύστημα δεν υπάρχει τυχαιότητα στην εξέλιξη του, το οποίο σημαίνει ότι για μια συγκεκριμένη αρχική θέση το τελικό αποτέλεσμα θα είναι πάντα ίδιο. Υπάρχουν όμως και μέθοδοι, όπως η Monte Carlo, στις οποίες οι θέσεις προκύπτουν στοχαστικά, δηλαδή το σύνολο των θέσεων των σωματιδίων εξαρτάται μόνο από το προηγούμενο στιγμιότυπο. Όταν σε μία αλληλουχία τυχαίων γεγονότων, η τρέχουσα κατάσταση επηρεάζεται μόνο από την αμέσως προηγούμενη και από καμία άλλη κατάσταση που προηγούνταν, η αλληλουχία αυτή ονομάζεται αλυσίδα Markov. Επιπλέον υπάρχουν προσομοιώσεις που αποτελούνται από το συνδυασμό ντετερμινιστικών και στοχαστικών διαδικασιών, όπως General Langevin Dynamics, Brownian Dynamics κ.α. Σε προσομοιώσεις μοριακής δυναμικής περίπλοκων και χρονοβόρων συστημάτων όπως οι πρωτεΐνες, συνίσταται η χρήση υβριδικών προσομοιώσεων όπως οι μέθοδοι Langevin και Brownian, κυρίως λόγω του περίπλοκων εξισώσεων κίνησης. Τέτοιες μέθοδοι μειώνουν σημαντικά το υψηλό υπολογιστικό κόστος που απαιτείται για τη διεξαγωγή περίπλοκων προσομοιώσεων. [49, 53]

Η ολοκλήρωση των κλασικών εξισώσεων της κίνησης βασίζεται σε απλές διαφορικές μεθόδους, συνήθως χρησιμοποιώντας διαφορετικές εκδόσεις του αλγορίθμου Verlet, ο οποίος κατά βάση αποτελεί συνδυασμό δύο επεκτάσεων του Taylor. Οι σειρές Taylor χρησιμοποιούνται για να μετατρέψουν μια εξίσωση με άπειρους όρους σε μια εξίσωση με πεπερασμένο αριθμό όρων

ώστε να είναι δυνατή η επίλυσή της. Τα βασικά χαρακτηριστικά του αλγορίθμου Verlet είναι η απλότητα και η καλή σταθερότητα για σχετικά μεγάλα χρονικά διαστήματα. Αρχικά γράφουμε τις σειρές Taylor για τη θέση από τη χρονική στιγμή t έως την $t + \Delta t$:

$$x(t + \Delta t) = x(t) + dx(t) \Delta t / dt + 1/2 d^2x(t) \Delta t^2 / dt^2 + 1/3! d^3x(t) \Delta t^3 / dt^3 + O(\Delta t^4)$$

Στη συνέχεια γράφουμε τις σειρές Taylor για τη θέση από τη χρονική στιγμή t έως την $t - \Delta t$:

$$x(t - \Delta t) = x(t) - dx(t) \Delta t / dt + 1/2 d^2x(t) \Delta t^2 / dt^2 - 1/3! d^3x(t) \Delta t^3 / dt^3 + O(\Delta t^4)$$

Αν προσθέσουμε αυτές τις δύο επεκτάσεις προκύπτει ο αλγόριθμος Verlet των θέσεων:

$$x(t + \Delta t) = 2x(t) - x(t - \Delta t) + d^2x(t) \Delta t^2 / dt^2 + O(\Delta t^4)$$

Ο αλγόριθμος αυτός παρουσιάζει τοπικά σφάλματα αποκοπής που επηρεάζονται από το (Δt^4) . Οι αλγόριθμοι ολοκλήρωσης που υπολογίζουν τις θέσεις, την ταχύτητα και την επιτάχυνση μπορούν να υπολογιστούν με επεκτάσεις των σειρών Taylor:

$$r(t + \Delta t) = r(t) + v(t) \Delta t + 1/2 a(t) \Delta t^2 + \dots$$

$$v(t + \Delta t) = v(t) + a(t) \Delta t + 1/2 b(t) \Delta t^2 + \dots$$

$$a(t + \Delta t) = a(t) + b(t)\Delta t + \dots$$

Υπάρχουν αρκετοί αλγόριθμοι που χρησιμοποιούνται για την ολοκλήρωση των εξισώσεων της κίνησης:

- Verlet algorithm
- Leap-frog algorithm
- Velocity Verlet
- Beeman's algorithm

Οι εξισώσεις της κίνησης ακόμα και για τα πιο απλά συστήματα είναι αρκετά περίπλοκες ώστε η ολοκλήρωση να πρέπει να γίνεται σε ένα μεγάλο αριθμό διακριτών μικρών χρονικών διαστημάτων. Αυτή η χρήση του χρόνου προϋποθέτει ότι σε κάθε χρονικό διάστημα οι ατομικές συντεταγμένες θα είναι λεπτομερώς προσδιορισμένες και θα χρησιμοποιούνται για τη μέτρηση της δυναμικής ενέργειας και κατά συνέπεια των δυνάμεων χρησιμοποιώντας το κατάλληλο δυναμικό πεδίο. [49, 56, 60]

2.4 Δυναμικά Πεδία

Οι προσομοιώσεις μοριακής δυναμικής βιολογικών συστημάτων απαιτούν υπολογισμούς της δυναμικής ενέργειας του υπό μελέτη συστήματος. Τα δυναμικά πεδία αποτελούν μαθηματικά μοντέλα που χρησιμοποιούνται στις προσομοιώσεις μοριακής δυναμικής για την υπολογισμό των δυνάμεων που λειτουργούν σε ατομικό επίπεδο καθώς επίσης και την περιγραφή της δυναμικής ενέργειας των υπό μελέτη συστημάτων ως συνάρτηση των αλληλεπιδράσεων και της θέσης των ατόμων στο χώρο. Το κάθε άτομο μπορεί να αλληλεπιδρά με άτομα με τα οποία σχηματίζει άμεσους δεσμούς, οι αλληλεπιδράσεις αυτές ονομάζονται δεσμικές, και με άλλα με τα οποία δεν σχηματίζει άμεσες επαφές. Οι αλληλεπιδράσεις μεταξύ των δεσμικών ατόμων αφορούν το μήκος των δεσμών, τη γωνία που σχηματίζουν καθώς επίσης και τις δίεδρες γωνίες. Οι αλληλεπιδράσεις των μη συνδεδεμένων ατόμων χωρίζονται σε δύο κατηγορίες, τις ηλεκτροστατικές και αυτές που προκύπτουν από τις δυνάμεις van der Waals. Για τις ηλεκτροστατικές χρησιμοποιούνται εξισώσεις που βασίζονται σε δυναμικά όπως προβλέπει ο νόμος του Coulomb, ενώ για τις δυνάμεις van der Waals χρησιμοποιούνται οι εξισώσεις Lennard-Jones (LJ). Οι πιο συνηθισμένες εκφράσεις του δυναμικού LJ είναι:

$$V_{LJ} = 4\epsilon [(\sigma/r)^{12} - (\sigma/r)^6] = \epsilon [(r_m/r)^{12} - 2(r_m/r)^6]$$

όπου ϵ είναι το βάθος του δυναμικού πηγαδιού, δηλαδή της περιοχής που περιβάλλει ένα τοπικό ελάχιστο δυναμικής ενέργειας, σ είναι η πεπερασμένη απόσταση όπου το δυναμικό μεταξύ των σωματιδίων είναι 0, r είναι η απόσταση των σωματιδίων και r_m είναι η απόσταση όπου το δυναμικό φτάνει στο ελάχιστο. Η σχέση των αποστάσεων είναι $r_m = 2^{1/6}\sigma$.

Για να περιγραφεί σωστά η συμπεριφορά ενός μοριακού συστήματος θα πρέπει το δυναμικό πεδίο που χρησιμοποιείται να είναι ακριβές, έτσι ώστε τα αποτελέσματα των προσομοιώσεων να έχουν κάποιο βιολογικό περιεχόμενο και να συγκλίνουν με τα πειραματικά δεδομένα. Η χρήση εξισώσεων κβαντικής μηχανικής για τη μελέτη της συμπεριφοράς του συστήματος θα μπορούσαν να εξασφαλίσουν τη μεγαλύτερη ακρίβεια. Δυστυχώς, η μέθοδος αυτή λόγω της πολυπλοκότητας τέτοιων συστημάτων είναι χρονοβόρα και έχει υψηλό υπολογιστικό κόστος. Γι' αυτό το λόγο, οι περισσότερες προσομοιώσεις χρησιμοποιούν εξισώσεις κλασικής μοριακής δυναμικής, οι οποίες είναι πιο απλές και οι εμπειρικοί τους παράμετροι μπορούν να υποστούν αρκετές τροποποιήσεις. Σήμερα, υπάρχουν πολλά δυναμικά πεδία των οποίων οι μαθηματικοί τύποι είναι παρόμοιοι. Οι περισσότερες διαφορές βρίσκονται στις παραμέτρους που περιγράφουν τα ενεργειακά χαρακτηρι-

στικά και στις μεθόδους από τις οποίες αυτά έχουν προκύψει. Υπολογισμοί κβαντικού επιπέδου και πειράματα φασματοσκοπίας και κρυσταλλογραφίας είναι συνήθως οι μέθοδοι από τις οποίες προκύπτουν οι παραπάνω παράμετροι. Τα δυναμικά πεδία υπόκεινται συνεχώς βελτιώσεις που σκοπό έχουν την αύξηση της ακρίβειας και της σύγκλισης με τα πειραματικά δεδομένα.

Τα δυναμικά πεδία που χρησιμοποιούνται κυρίως σε προσομοιώσεις μοριακής δυναμικής για τη μελέτη της αναδίπλωσης των μακρομορίων και πρωτεϊνών είναι το AMBER [59] (Assisted Model Building and Energy Refinement), CHARMM [77] (Chemistry HARvard Molecular Mechanics), το OPLS [78] (Optimized Potential for Liquid Simulations) και το GROMOS [79] (Groningen Molecular Simulation). Η μορφή της εξίσωσης του AMBER είναι:

$$V(rN) = \sum_{\text{bonds}} k_b (l - l_0)^2 + \sum_{\text{angles}} k_a (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} 1/2 V_n [1 + \cos(n\omega - \gamma)] + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \{ \epsilon_{i,j} [(r_{0ij}/r_{ij})^{12} - 2(r_{0ij}/r_{ij})^6] + q_i q_j / 4\epsilon_0 r_{ij} \}$$

όπου ο πρώτος όρος αντιπροσωπεύει την ενέργεια της έκτασης μεταξύ των δεσμικών ατόμων, ο δεύτερος την ενέργεια που προκύπτει από τη γεωμετρία των γωνιών μεταξύ των δεσμικών ατόμων, ο τρίτος την ενέργεια από την περιστροφή των ατόμων γύρω από τους δεσμούς και ο τέταρτος τη μη δεσμική ενέργεια όλων των ατόμων. Οι παράγοντες k είναι σταθερές και το n αντιπροσωπεύει την πολλαπλότητα της εξίσωσης.

Η μορφή της εξίσωσης του δυναμικού πεδίου CHARMM22 (μια από τις εκδόσεις του CHARMM) είναι:

$$V = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 + \sum_{\text{nonbonded}} \epsilon [(r_{0ij}/r_{ij})^{12} - 2(r_{0ij}/r_{ij})^6] + q_i q_j / \epsilon r_{ij}$$

όπου ο πρώτος όρος αντιπροσωπεύει την ενέργεια της έκτασης μεταξύ των δεσμικών ατόμων, ο δεύτερος την ενέργεια των γωνιών, ο τρίτος τις διεδρες γωνίες και ο τέταρτος τις “improper” οι οποίες χρησιμοποιούνται για τη συγκράτηση των χειρόμορφων και των επίπεδων κέντρων των αμινοξικών δεσμών. Ο όρος Urey-Bradley αφορά τις αλληλεπιδράσεις οι οποίες βασίζονται στην απόσταση ατόμων που χωρίζονται από δύο δεσμούς (αλληλεπιδράσεις 1,3) και το u αντιπροσωπεύει την απόσταση μεταξύ των ατόμων 1, 3.

Και τα δύο αυτά δυναμικά πεδία χρησιμοποιούν αρμονικές προσεγγίσεις για τις αλληλεπιδράσεις των συνδεδεμένων ατόμων. Επιπλέον, μοιάζει ο τρόπος με τον οποίο αντιμετωπίζονται οι αλληλεπιδράσεις των συνδεδεμένων ατόμων, ενώ υπάρχουν ουσιαστικές διαφορές όσον αφορά τα μη συνδεδεμένα άτομα και συγκεκριμένα στον υπολογισμό των φορτίων των ατόμων. Και τα δύο δυναμικά πεδία συναντούν δυσκολία στον υπολογισμό των μεταβολών στους εξωτερικούς ηλεκτρο-

νιακούς φλοιούς των ατόμων. Τα τελευταία χρόνια, τα AMBER, CHARMM και οι επεκτάσεις τους, έχουν υποστεί πολλές τροποποιήσεις σε διάφορες παραμέτρους τους. Μερικές από αυτές αφορούν τη παραμέτρους των δίδρων γωνιών της κύριας αλυσίδας ώστε να επιτευχθεί καλύτερη σύγκλιση με τα αποτελέσματα της κβαντικής μηχανικής, την αντιμετώπιση των δεσμών, των στοιχείων δευτεροταγούς δομής καθώς και των αλληλεπιδράσεων με το νερό. [52, 57, 58, 59]

2.5 AMBER

Το δυναμικό πεδίο AMBER [59] έχει υποστεί πολλές τροποποιήσεις από την πρώτη έκδοση του, στα μέσα της δεκαετίας του 80 μέχρι σήμερα. Το AMBER ff94 ήταν το πλέον χρησιμοποιούμενο τις τελευταίες δύο δεκαετίες για σε προσομοιώσεις πρωτεϊνών με όλα τα άτομα. Μερικά από τα χαρακτηριστικά του ff94 είναι καθορισμένα μερικά φορτία στο κέντρο των ατόμων, έλλειψη συγκεκριμένου τύπου για τους δεσμούς υδρογόνου και παραμετροποίηση των δίδρων γωνιών με βάση υπολογισμών κβαντικής μηχανικής σε μικρά μόρια. Οι επεκτάσεις ff96 και ff99 αφορούν αλλαγές στις παραμέτρους για τον υπολογισμό των δίδρων. Το δυναμικό πεδίο Amber ff99SB είναι το αποτέλεσμα τροποποιήσεων του ff99 όσον αφορά τα δυναμικά της κύριας αλυσίδας, όπως και το ff99SB*. Επιπλέον τροποποιήσεις στη συστροφή των πλευρικών ομάδων των αμινοξέων ισολευκίνη, λευκίνη, ασπαρτικό οξύ και ασπαραγίνη δημιούργησαν το δυναμικό πεδίο Amber ff99SB-ILDN. Αν και ορισμένες έρευνες έχουν δείξει μία υπερεκτίμηση ορισμένων μελών αυτής της οικογένειας των δυναμικών πεδίων για συγκεκριμένα στοιχεία δευτεροταγούς δομής [59], προσομοιώσεις μοριακής δυναμικής χρησιμοποιώντας το AMBER έχουν αναπαράγει με μεγάλη ακρίβεια δομικές και δυναμικές ιδιότητες νουκλεϊκών οξέων, πρωτεϊνών και DNA. [58, 59, 61]

2.6 Μοντέλα νερού

Το νερό είναι ο πιο συνηθισμένος και σημαντικός διαλύτης σε βιολογικά συστήματα. Αν και οι μακροσκοπικές ιδιότητες του νερού έχουν μελετηθεί εκτενώς και είναι καλά προσδιορισμένες, οι μικροσκοπικές του ιδιότητες δεν είναι επαρκώς διευκρινισμένες. Οι ιδιαίτερες ιδιότητες του νερού οφείλονται στην ικανότητα του να αλληλεπιδρά με τα γειτονικά μόρια νερού και να σχηματίζει δεσμούς υδρογόνου. Υπάρχουν διάφορες θεωρίες που προσπαθούν να εξηγήσουν τη συμπεριφορά και τις ιδιότητες του νερού. Στις προσομοιώσεις μοριακής δυναμικής χρησιμοποιούνται συγκεκριμένα μοντέλα νερού με στόχο την αναπαραγωγή των ιδιοτήτων του νερού στο υπό μελέτη σύστημα. Τα μοντέλα νερού, TIP3P (transferable intermolecular potential 3P), TIP4P, SPC (simple point charge) και SPC/E (extended simple point charge) αντιμετωπίζουν τα άτομα των μορίων νερού σαν άκαμπτες σφαίρες, χρησιμοποιούν εξισώσεις Lennard-Jones (LJ) για τον

υπολογισμό αφόρτιστων μορίων ή ατόμων και την εξίσωση του Coulomb για τον υπολογισμό ηλεκτροστατικών αλληλεπιδράσεων. Τα πιο συνηθισμένα μοντέλα που χρησιμοποιούνται σε προσομοιώσεις θεωρούν ότι τα μόρια νερού έχουν τρεις θέσεις αλληλεπίδρασης, οι οποίες βρίσκονται στα ατομικά κέντρα. Τέτοια μοντέλα είναι τα TIP3P, SPC και SPC/E. Το μοντέλο TIP4P υποθέτει ότι τα μόρια νερού έχουν τέσσερις θέσεις με τις οποίες μπορούν να αλληλεπιδράσουν με τα γειτονικά άτομα, στα τρία άτομα και μία τέταρτη πάνω στον άξονα συμμετρίας μεταξύ του οξυγόνου και των δύο υδρογόνων. [61]

Κεφάλαιο 3

Προσομοιώσεις

Computers are useless. They can only give you answers.

Pablo Picasso

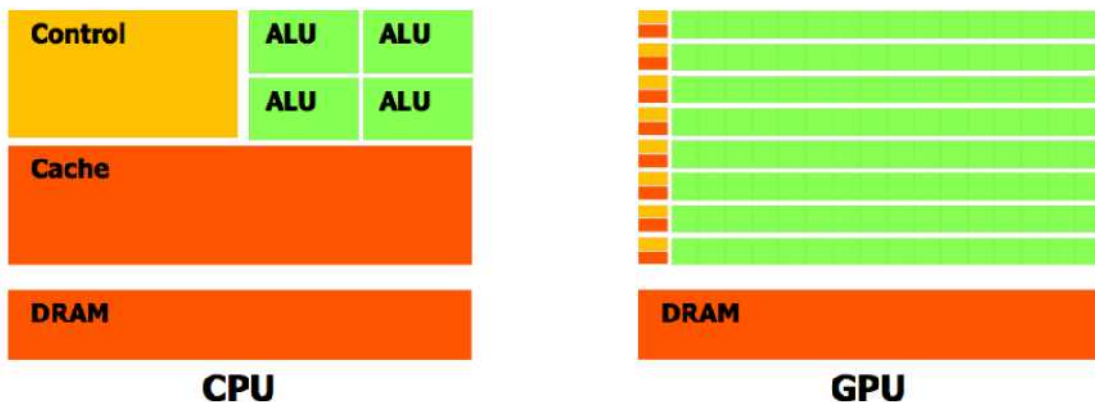
3.1 Τεχνικά χαρακτηριστικά υπολογιστών

Όπως περιγράφηκε στο προηγούμενο κεφάλαιο, οι προσομοιώσεις μοριακής δυναμικής υπολογίζουν τα ατομικά τροχιακά μοριακών συστημάτων λύνοντας τις εξισώσεις της κίνησης με τη χρήση εμπειρικών δυναμικών πεδίων. Λόγω του μεγάλου υπολογιστικού κόστους των διεργασιών αυτών, είναι προτιμότερο να χρησιμοποιούνται παράλληλα συνδεδεμένοι υπολογιστές, οι οποίοι προσφέρουν ισχυρή υπολογιστική δύναμη με σχετικά μικρό κόστος. Η προσέγγιση αυτή αποδίδει όταν μελετώνται προσομοιώσεις μεγάλων συστημάτων με τουλάχιστον δεκάδες χιλιάδες άτομα, ενώ προσομοιώσεις μικρών συστημάτων είναι πιο συμφέρον να γίνονται σε μεμονωμένους υπολογιστές. Τα υπολογιστικά συγκροτήματα (computer clusters) τύπου Beowulf αποτελούνται

από υπολογιστές παράλληλα συνδεδεμένους σε ένα τοπικό δίκτυο. Οι υπολογιστές αυτοί περιέχουν βιβλιοθήκες και προγράμματα που επιτρέπουν τον καταμερισμό των διεργασιών ανάμεσά τους. Ένα τέτοιο υπολογιστικό συγκρότημα είναι η Norma. [62,63]

Η Norma περιλαμβάνει 40 κεντρικές μονάδες επεξεργασίας (central processing unit, CPU), 46 Gbytes φυσικής μνήμης και 6 GPGPUs (General-purpose computing on graphics processing units) καταμεμημένα σε 10 κόμβους. Οι 10 κόμβοι αποτελούνται από 8 Intel's Q6600 Kentsfield 2.4GHz quad processors, τέσσερις από τους οποίους φέρουν nvidia GTX-460 GPU, έναν i7-extreme-GTX295 με κάρτα CUDA-capable GTX-295 (Compute Unified Device Architecture) και τον κεντρικό κόμβο με 1.5 TB RAID-5 array τεσσάρων δίσκων και μια nvidia GTX-260 συνδεδεμένους με μεταγωγέα HP ProCurve 1800-24G Gigabit ethernet. [63, 64]

Ρόλος των CPUs είναι η εκτέλεση πολλαπλών εντολών, οι οποίες αποτελούν τα διάφορα προγράμματα. Τα κύρια στοιχεία των CPUs είναι η αριθμητική λογική μονάδα (arithmetic logic unit, ALU), στην οποία γίνονται οι αριθμητικές και λογικές λειτουργίες, η μονάδα ελέγχου (control unit), η οποία λαμβάνει και εκτελεί εντολές από τη μνήμη και η κρυφή μνήμη (cache), στην οποία αποθηκεύονται δεδομένα ώστε να μπορούν να ανακτηθούν γρηγορότερα αν προκύψει ανάγκη. Οι μονάδες επεξεργασίας γραφικών (graphics processing unit, GPU) είναι κατασκευασμένες για να διαχειρίζονται τα γραφικά του υπολογιστή, να κατασκευάζουν και να παρουσιάζουν εικόνες. Στην **Εικόνα 3.1** παρουσιάζεται μία σχηματική αναπαράσταση της έκτασης των κύριων στοιχείων στις CPU και στις GPU.



Εικόνα 3.1 Σχηματική αναπαράσταση των κύριων στοιχείων στις CPU και στις GPU. Στην εικόνα παρουσιάζονται η μονάδα ελέγχου, η αριθμητική λογική μονάδα, η κρυφή μνήμη και η μνήμη DRAM, στην οποία αποθηκεύονται δεδομένα για γρηγορότερη ανάκτηση.

Όπως φαίνεται στην **Εικόνα 3.1**, η έκταση του υπολογιστικού τμήματος στις GPU είναι συγκριτικά μεγαλύτερη από την έκταση του στις CPU. Επιπλέον, οι CPUs έχουν έναν εγγενή περιορισμό, ο οποίος είναι ότι δεν μπορούν να διαχειριστούν παράλληλα πολλές εντολές και ότι πρέπει να ολοκληρωθεί μια εντολή πριν αρχίσει η εκτέλεση της επόμενης. Για να ξεπεραστεί ο περιορισμός αυτός χρησιμοποιούνται παράλληλοι υπολογισμοί είτε σε επίπεδο διαχείρισης των δεδομένων ή σε επίπεδο υπολογισμών. Σε αντίθεση με τις CPUs, οι GPUs έχουν παράλληλη αρχιτεκτονική απόδοση (parallel throughput architecture) και μπορούν να επεξεργάζονται ταυτόχρονα πολλά δεδομένα και σύνολα εντολών, μειώνοντας το χρόνο που χρειάζεται για τη διεξαγωγή ανεξάρτητων εργασιών. CUDA (Computed Unified Device Architecture) ονομάζεται η υπολογιστική μηχανή και το προγραμματιστικό μοντέλο που χρησιμοποιείται από τις Nvidia GPUs. Η CUDA επιτρέπει την πρόσβαση στο εικονικό σύνολο εντολών (virtual instruction set) και στη μνήμη των παράλληλων υπολογιστικών στοιχείων των CUDA GPUs, επιτρέποντας τη δημιουργία GPGPU. Μέσω των GPGPUs είναι εφικτή η διενέργεια εργασιών που συνήθως γίνονται στις CPUs σε GPUs (graphics processing unit). [64]

Όλες οι προσομοιώσεις που θα αναλυθούν στην παρούσα εργασία (και πολλές άλλες) έγιναν στη Norma. Η απόδοση του κάθε κόμβου για τα συστήματα που μελετήσαμε ήταν περίπου 31 ns την ημέρα και κατ' επέκταση για κάθε 1μs προσομοιώσεων χρειάστηκαν περίπου 32 μέρες. Συνολικά χρειάστηκαν 13 μήνες φυσικού χρόνου για όλες τις προσομοιώσεις.

3.2 Έναρξη προσομοιώσεων μοριακής δυναμικής με NAMD

Το NAMD [65] είναι ένα πρόγραμμα σχεδιασμένο για τη διεξαγωγή προσομοιώσεων μοριακής δυναμικής μεγάλης ακρίβειας για μεγάλα βιολογικά συστήματα. Για τη διεξαγωγή μιας προσομοίωσης το NAMD χρειάζεται τα παρακάτω αρχεία:

α) ένα αρχείο pdb (Protein Data Bank), το οποίο περιέχει τις συντεταγμένες των ατόμων ενός συστήματος. Το NAMD ενδιαφέρεται μόνο για τις καταχωρήσεις ATOM, οι οποίες αφορούν τις συντεταγμένες των ατόμων της πρωτεΐνης, μορίων νερού και επιπρόσθετων ιόντων, καθώς και τις συντεταγμένες ετερογενών ατόμων στους κρυστάλλους αντίστοιχα. Αρχεία pdb μπορούν να κατασκευαστούν χειροκίνητα ή να ληφθούν έτοιμα από την ιστοσελίδα <http://www.pdb.org>. Αρχεία pdb, όπως τα χρειάζεται το NAMD, μπορεί να έχουν την παρακάτω μορφή:

ATOM	1	N	ASP	1	-5.907	-4.289	-0.539	1.00	0.00
ATOM	2	H1	ASP	1	-5.072	-4.848	-0.644	1.00	0.00
ATOM	3	H2	ASP	1	-6.669	-4.948	-0.611	1.00	0.00
ATOM	4	H3	ASP	1	-5.942	-3.887	0.387	1.00	0.00
ATOM	5	CA	ASP	1	-6.048	-3.350	-1.632	1.00	0.00
ATOM	6	HA	ASP	1	-5.257	-2.644	-1.382	1.00	0.00
ATOM	7	CB	ASP	1	-5.761	-3.900	-3.051	1.00	0.00
ATOM	8	HB2	ASP	1	-6.089	-3.150	-3.771	1.00	0.00
ATOM	9	HB3	ASP	1	-4.679	-3.975	-3.160	1.00	0.00
ATOM	10	CG	ASP	1	-6.499	-5.212	-3.413	1.00	0.00

Εικόνα 3.2 Οι στήλες από τα αριστερά προς τα δεξιά περιέχουν τον τύπο καταχώρησης, τον αριθμό του ατόμου, το όνομα του ατόμου, το όνομα του καταλοίπου, τον αριθμό του καταλοίπου, συντεταγμένες x, y και z, την κατοχή και τον παράγοντα θερμοκρασίας.

β) ένα psf (Protein Structure File), το οποίο περιέχει τις μοριακές πληροφορίες για τα άτομα της πρωτεΐνης που χρειάζονται για την εφαρμογή ενός δυναμικού πεδίου σε ένα μοριακό σύστημα. Πολλές από τις πληροφορίες που υπάρχουν στα psf εντοπίζονται και στα αρχεία pdb, με αποτέλεσμα συχνά οι συντεταγμένες των ατόμων να είναι το μοναδικό ξεχωριστό στοιχείο που θα χρησιμοποιηθεί σε προσομοιώσεις με το NAMD. Επιπλέον τα psfs περιέχουν πληροφορίες για τους ομοιοπολικούς δεσμούς μεταξύ των ατόμων, τις δίεδρες γωνίες, τις improper κ.α. Στις προσομοιώσεις που θα αναλύσουμε δε χρησιμοποιήθηκαν psfs, διότι το δυναμικό πεδίο που εφαρμόσαμε ήταν το AMBER και επεκτάσεις του, αλλά παραθέτουμε τις πληροφορίες αυτές για λόγους πληρότητας. Psfs χρειάζονται μόνο όταν οι προσομοιώσεις γίνονται με το δυναμικό πεδίο CHARMM. Η καταχώρηση των πληροφοριών γίνεται με την παρακάτω μορφή:

244 !NATOM

1 A	1	ASP	N	N3	0.078200	14.0100	0
2 A	1	ASP	H1	H	0.220000	1.0080	0
3 A	1	ASP	H2	H	0.220000	1.0080	0
4 A	1	ASP	H3	H	0.220000	1.0080	0
5 A	1	ASP	CA	CT	0.029200	12.0100	0
6 A	1	ASP	HA	HP	0.114100	1.0080	0

7 A	1	ASP	CB	CT	-0.023500	12.0100	0
8 A	1	ASP	HB2	HC	-0.016900	1.0080	0
9 A	1	ASP	HB3	HC	-0.016900	1.0080	0
10 A	1	ASP	CG	C6	0.819400	12.0100	0

Εικόνα 3.3 Αρχικά αναφέρεται ο αριθμός των ατόμων που περιγράφονται στο παρόν αρχείο. Οι στήλες από τα αριστερά προς τα δεξιά περιέχουν τον αριθμό του ατόμου, το όνομα του τμήματος, τον αριθμό του καταλοίπου, το όνομα του ατόμου, τον τύπο του, το φορτίο του, τη μάζα του και ένα μηδέν στην τελευταία στήλη κάθε σειράς

249 !NBOND: bonds

1	2	1	3	1	4	1	5
5	6	5	7	5	13	7	8
7	9	7	10	10	11	10	12
13	14	13	15	15	16	15	17
17	18	17	19	17	35	19	20

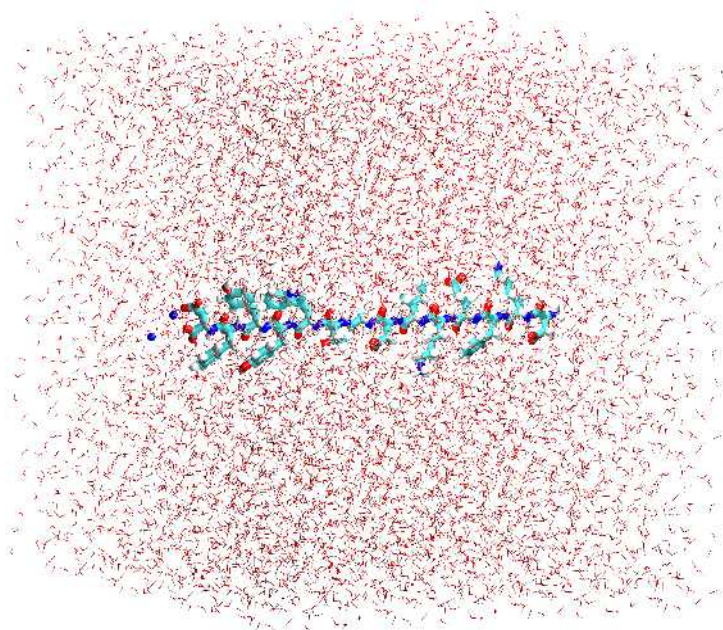
Εικόνα 3.4 Στην εικόνα φαίνονται οι ομοιοπολικοί δεσμοί ανά τέσσερα ζευγάρια σε κάθε σειρά. Με αντίστοιχο τρόπο καταχωρούνται οι δίεδρες, οι improper και τα υπόλοιπα δεδομένα.

γ) ένα αρχείο παραμετροποίησης του δυναμικού πεδίου(force-field parameters file), που περιέχει όλες τις αριθμητικές σταθερές που χρειάζονται για τον υπολογισμό των δυνάμεων που ασκούνται στα άτομα καθώς και της ενέργειας του μοριακού συστήματος. Επιπλέον καθορίζει διάφορους παράγοντες όπως η δύναμη των δεσμών. Το NAMD έχει την ικανότητα να χρησιμοποιεί αρχεία παραμετροποίησης από διαφορετικά δυναμικά πεδία όπως AMBER και CHARMM.

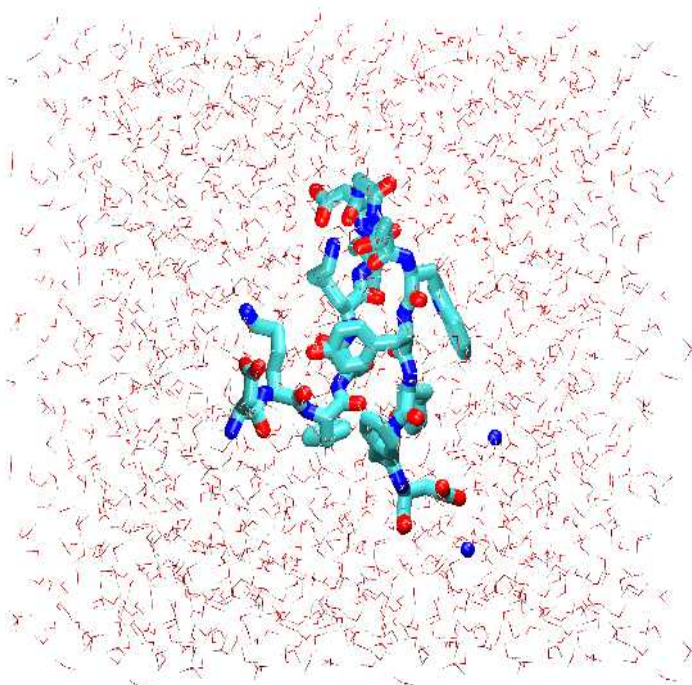
δ) ένα αρχείο διαμόρφωσης (configuration file), στο οποίο ο χρήστης μπορεί να ορίσει όλες τις ρυθμίσεις τις οποίες το NAMD θα λάβει υπόψη κατά τη διάρκεια μιας προσομοίωσης. Τέτοιες ρυθμίσεις μπορεί να είναι ο καθορισμός της εισόδου και της εξόδου του προγράμματος, ο καθορισμός της αρχικής θερμοκρασίας του συστήματος, το χρονικό βήμα της προσομοίωσης, τη συχνότητα καταγραφής των συντεταγμένων στα τροχιακά κ.α. [66]

3.3 Προετοιμασία συστημάτων

Όλες οι προσομοιώσεις ξεκίνησαν από την αποδιαταγμένη δομή του 14-πεπτιδίου DKFEKIDGTWYYFD όπως προήλθε από το πρόγραμμα RIBOSOME (<http://www.roselab.jhu.edu/~raj/Manuals/ribosome.html>). Στο αρχικό πεπτίδιο προστέθηκαν τα υδρογόνα που έλειπαν. Επιπλέον, προστέθηκαν ιόντα ώστε το σύστημα να μοιάζει περισσότερο με ένα τυπικό βιολογικό περιβάλλον. Η προσθήκη των υδρογόνων και των ιόντων έγινε με το πρόγραμμα LEAP από AMBER tools distribution [67]. Τα ιόντα που προστέθηκαν συνέβαλαν στην εξουδετέρωση του φορτίου του συστήματος και βοήθησαν το σύστημα να γίνει ουδέτερο, πράγμα απαραίτητο για την εφαρμογή της μεθόδου Particle Mesh Ewald, PME [64]. Η μέθοδος αυτή χρησιμοποιείται για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων σε μεγάλες αποστάσεις, όταν χρησιμοποιούνται συνθήκες περιοδικών συνόρων. Το αποτέλεσμα είναι η δημιουργία ενός τρισδιάστατου πλέγματος μέσα στο οποίο κατανέμεται το φορτίο του συστήματος. Η αρχική δομή του πεπτιδίου τοποθετήθηκε σε ένα σύστημα με περίπου 9.000 μόρια νερού. Η προσομοίωση αυτή έγινε ώστε το πεπτίδιο να αποκτήσει μια πιο σφαιρική συμπαγή δομή, συγχρόνως πιο ρεαλιστική αφού είναι δύσκολο το πεπτίδιο να βρίσκεται στην πλήρως αποδιαταγμένη του μορφή σε ένα τυχαίο περιβάλλον. Στη συνέχεια, απομακρύνθηκε ένας μεγάλος αριθμός μορίων νερού, ώστε να μειωθεί το υπολογιστικό κόστος για τον υπολογισμό των αλληλεπιδράσεων των ατόμων του πεπτιδίου με το νερό. Η μορφή του συστήματος πριν και μετά την εξισορρόπηση παρουσιάζεται στις παρακάτω εικόνες:



Εικόνα 3.5 Στην εικόνα παρουσιάζεται το αρχικό σύστημα, του οποίου οι διαστάσεις είναι $65\text{\AA} \times 65\text{\AA} \times 65\text{\AA}$ και περιέχει 27237 άτομα.



Εικόνα 4. Στην εικόνα παρουσιάζεται το σύστημα μετά την μείωση των ατόμων νερού, οι διαστάσεις του είναι 38Å x 38Å x 38Å και περιέχει 5493 άτομα.

Όλες οι προσομοιώσεις έγιναν με το πρόγραμμα NAMD και ο συνολικός χρόνος των προσομοιώσεων ήταν 13.69 μ s. Χρησιμοποιήθηκαν τα δυναμικά πεδία AMBER99SB και AMBER99SB-ILDN, με διαφορετικούς συνδυασμούς θερμοκρασιών και μοντέλων νερού. Για όλες τις προσομοιώσεις χρησιμοποιήθηκαν συνθήκες περιοδικών συνόρων με διαστάσεις κυψελίδας 38Å x 38Å x 38Å. Στον **Πίνακα 3.1** παρουσιάζονται οι διαφορές της κάθε προσομοίωσης:

Προσομοιώσεις	Συνολικά άτομα συστήματος	Αριθμός ατόμων πεπτιδίου	Αριθμός ατόμων νερού	Αριθμός ιόντων	Διάρκεια προσομοίωσης (μ s)	Θερμοκρασία	Μοντέλο νερού
99SB	5493	244	5247	2 Na ⁺	3.76	320K	TIP3P
ILDN					4.48		
ILDN-340K					1.40	340K	
ILDN-360K					1.42	360K	
ILDN-TIP4P	7762		7516		2.60	320K	TIP4P

Πίνακας 3.1 Συγκεντρωτικός πίνακας των προσομοιώσεων που πραγματοποιήθηκαν.

Πριν την έναρξη κάθε προσομοίωσης είναι απαραίτητη η ελαχιστοποίηση της ενέργειας. Ο ρόλος της είναι να βρεθεί ένα τοπικό ενεργειακό ελάχιστο και να επιδιορθωθούν πιθανά μικρά σφάλματα στη διευθέτηση και τη συγκρότηση των ατόμων, όπως παραμορφωμένες γωνίες και μήκη δεσμών, της αρχικής δομής. Μετά την ελαχιστοποίηση της ενέργειας ακολουθεί η φάση της εξισορρόπησης. Κατά τη διάρκεια της εξισορρόπησης υπολογίζονται μεταβλητές όπως η πίεση, η θερμοκρασία και η ενέργεια. Στόχος της εξισορρόπησης είναι η δημιουργία μιας σύντομης προσομοίωσης κατά την οποία οι μεταβλητές θα γίνουν σταθερές με σχέση με το χρόνο. Στο κάθε σύστημα αρχικά έγινε ελαχιστοποίηση ενέργειας ως προς τα άτομα που δεν ανήκουν στην κύρια αλυσίδα με βήμα 2000 και στη συνέχεια ακολούθησε θέρμανση του συστήματος από τους 0 K μέχρι την επιθυμητή θερμοκρασία σε κάθε προσομοίωσης (320 K, 340 K, 360 K) με βήμα 20 K. Για την εξισορρόπηση των συστημάτων χρησιμοποιήθηκαν συνθήκες NpT χωρίς κάποιον περιορισμό μέχρις ότου να εξισορροπηθεί ο όγκος. Η ρύθμιση της θερμοκρασίας και της πίεσης έγινε χρησιμοποιώντας το δυναμικό Nosè-Hoover Langevin dynamics και μεθόδους Langevin piston barostat control. Θα μπορούσαμε να πούμε ότι κατά τη διάρκεια της εξισορρόπησης, το σύστημα περνά ένα χρονικό διάστημα, αρκετό ώστε το σύστημα να “ξεχάσει” πώς έχει κατασκευαστεί και να μοιάζει περισσότερο με ένα βιολογικό περιβάλλον. Για την παραγωγική φάση χρησιμοποιήθηκε ο αλγόριθμος Verlet-I και το βήμα που χρησιμοποιήθηκε για τον υπολογισμό των αλληλεπιδράσεων ήταν 2 fs. Ο υπολογισμός των ηλεκτροστατικών αλληλεπιδράσεων έγινε με τη χρήση της μεθόδου Particle Mesh Ewald. Το όριο για τον υπολογισμό των δυνάμεων van der Waals ορίστηκε στα 8 Å χρησιμοποιώντας τη λειτουργία switch και ο αλγόριθμος SHAKE χρησιμοποιήθηκε για τον περιορισμό όλων των δεσμών μεταξύ υδρογόνων και άλλων ατόμων. Τέλος, η αποθήκευση των ατομικών συντεταγμένων όλων των ατόμων του συστήματος γινόταν κάθε 800 fs. [49, 63, 68, 69]

Κεφάλαιο 4

Ανάλυση προσομοιώσεων

Nothing will work unless you do

Maya Angelou

4.1 Εισαγωγή

Όπως αναφέρθηκε και σε προηγούμενα κεφάλαια στόχος αυτής της εργασίας είναι η μελέτη της αναδίπλωσης του 14-πεπτιδίου DKFEKIDGTWYYFD με τη χρήση physics-based μεθόδων και η σύγκριση των αποτελεσμάτων των προσομοιώσεων με άλλα πειραματικά δεδομένα καθώς και η σύγκριση των αποτελεσμάτων μεταξύ διαφορετικών δυναμικών πεδίων. Παράλληλα, θα προσπαθήσουμε να εντοπίσουμε σημαντικά στοιχεία στοιχεία που συμβάλλουν στην αναδίπλωση του και να δώσουμε μια ερμηνεία στη συμπεριφορά του πεπτιδίου. Για την ανάλυση των προσομοιώσεων χρησιμοποιήθηκαν διάφορες μέθοδοι για τη σύγκριση, την ομαδοποίηση και τον προσδιορισμό

χαρακτηριστικών των διαφορετικών διαμορφώσεων του πεπτιδίου σε κάθε τροχιακό που σκοπό έχουν τον προσδιορισμό κινητικών και δυναμικών ιδιοτήτων των εκάστοτε μορίων. Για την καλύτερη κατανόηση και ερμηνεία των αποτελεσμάτων θα προσπαθήσουμε να παρουσιάσουμε την κεντρική ιδέα γύρω από κάθε μέθοδο που χρησιμοποιήθηκε.

Μία από τις πιο συνηθισμένες μεθόδους ανάλυσης προσομοιώσεων μοριακής δυναμικής είναι η ρίζα της μέσης τετραγωνικής απόκλισης (root-mean-square deviation, RMSD). Το RMSD χρησιμοποιείται για τον υπολογισμό της απόκλισης των διαμορφώσεων ενός μορίου από ένα μόριο αναφοράς. Βασίζεται στον υπολογισμό των ευκλείδειων αποστάσεων των ατόμων και η μαθηματική του έκφραση δίνεται από την εξίσωση:

$$\text{RMSD} = \sqrt{\sum(x_i - x_{\text{ref}})^2/N}$$

όπου N είναι ο αριθμός των ατόμων, x_i είναι οι συντεταγμένες των ατόμων του εκάστοτε μορίου και x_{ref} είναι οι συντεταγμένες των ατόμων του μορίου αναφοράς.

Το RMSD είναι ένα χρήσιμο εργαλείο για τη σύγκριση δομών, το οποίο παρέχει μεγάλη ακρίβεια. Στη συνέχεια του κεφαλαίου, ο υπολογισμός μέσω δομών από τους πίνακες RMSD γίνεται επιλέγοντας τη δομή με χαμηλότερο RMSD από τη μέση δομή για κάθε σύνολο.

Η διακύμανση του RMSD γύρω από μία μέση δομή αποτελεί το root-mean-square fluctuation, RMSF, το οποίο μπορεί να χρησιμοποιηθεί ως μέτρο για τον υπολογισμό της κινητικότητας ατόμων και περιοχών πάνω σε ένα συγκεκριμένο μόριο. Το RMSF χρησιμοποιείται συχνά για τον εντοπισμό περιοχών με μεγάλη διακύμανση και τη σύγκρισή τους με πειραματικά δεδομένα.

Σημαντικό κομμάτι της ανάλυσης είναι ο προσδιορισμός στοιχείων δευτεροταγούς δομής για το χαρακτηρισμό των τρισδιάστατων δομών σε όλη τη διάρκεια του τροχιακού. Για το σκοπό αυτό, υπάρχουν διάφορες μέθοδοι που προσεγγίζουν το ίδιο πρόβλημα από διαφορετική σκοπιά, όπως εντοπισμός μοτίβων μεταξύ των αποστάσεων των $C\alpha$, ανάλυση των γωνιών και του μήκους των δεσμών μεταξύ των $C\alpha$, εντοπισμός μοτίβων μεταξύ των δεσμών υδρογόνου κ.α. Ο αλγόριθμος DSSP χρησιμοποιείται συχνά για τον καθορισμό στοιχείων δευτεροταγούς δομής με βάση τα επαναλαμβανόμενα μοτίβα δεσμών υδρογόνου που εμφανίζονται. Η ενέργεια των δεσμών υδρογόνου υπολογίζεται μέσω μιας εμπειρικής συνάρτησης της ενέργειας χρησιμοποιώντας την απόσταση μεταξύ

του δότη και του δέκτη καθώς και την γωνία που σχηματίζεται ανάμεσά τους. Ένας άλλος αλγόριθμος για τον προσδιορισμό στοιχείων δευτεροταγούς δομής είναι STRIDE, ο οποίος λαμβάνει υπόψη τα επαναλαμβανόμενα μοτίβα δεσμών υδρογόνου και τη γεωμετρία της κύριας αλυσίδας, πχ δίεδρες γωνίες, σε αντίθεση με τον DSSP. [70, 71]

Η ανάλυση κύριων συνιστωσών (Principal Component Analysis, PCA) είναι μια μέθοδος που χρησιμοποιείται για τη μείωση της πολυδιαστατικότητας σε ένα περίπλοκο σύστημα [72]. Η PCA επιτρέπει τον εντοπισμό συγκεκριμένων μοτίβων σε ένα σύνολο δεδομένων και την έκφραση των δεδομένων του συνόλου με τέτοιο τρόπο ώστε να επισημαίνονται οι ομοιότητες και οι διαφορές τους. Η μέθοδος στηρίζεται σε πίνακες συνδιακύμανσης συγκεκριμένων μεταβλητών του συστήματος. Η συνδιακύμανση χρησιμοποιείται ως μέτρο υπολογισμού της από κοινού μεταβλητότητας μεταξύ δύο μεταβλητών και ως μέτρο συσχέτισης τους. Η μαθηματική εξίσωση της συνδιακύμανσης είναι εξής:

$$\text{cov}(X,Y)= 1/n \sum_{i=1}^n (x_i - x_m) (y_i - y_m)$$

όπου x_m και y_m οι μέσες τιμές των μεταβλητών.

Η μείωση της πολυδιαστατικότητας επιτυγχάνεται μέσω της μετατροπής των σχετιζόμενων μεταβλητών σε τιμές γραμμικώς μη σχετιζόμενων μεταβλητών που ονομάζονται κύριες συνιστώσες (Principal Components, PCs). Ουσιαστικά, τα δεδομένα δεν εκφράζονται πια ως συνάρτηση των αξόνων x , y , z , αλλά ως συνάρτηση των ιδιοδιανυσμάτων (eigenvectors) που προκύπτουν από τους πίνακες της συνδιακύμανσης. Η μετατροπή έχει ως στόχο την έκφραση των μεταβλητών με βάση μοτίβα που περιγράφουν καλύτερα τις κινήσεις και τις σχέσεις μεταξύ τους. Αν και όπως είναι φυσικό, ένα ποσοστό της πληροφορίας χάνεται κατά τη μετατροπή, οι κύριες συνιστώσες με τις μεγαλύτερες ιδιοτιμές (eigenvalues) φαίνεται να μπορούν να περιγράφουν με μεγάλη ακρίβεια τις κινήσεις και τις σχέσεις μεταξύ των μεταβλητών. Η αρίθμηση των ιδιοδιανυσμάτων γίνεται με βάση το πόσο μεγάλη είναι η ιδιοτιμή που τα χαρακτηρίζει. Βασικός στόχος για την ανάλυση είναι ο διαχωρισμός των μοτίβων της κίνησης με βάση τη βαρύτητά τους και πιο συγκεκριμένα ο διαχωρισμός της εσωτερικής κίνησης του υπό μελέτη αντικειμένου από την ολική κίνηση μέσα στο σύστημα. Η Cartesian Principal Component Analysis (CPCA) χρησιμοποιείται για τον υπολογισμό των κυρίων συνιστωσών με βάση τις καρτεσιανές συντεταγμένες των ατόμων του συστήματος. Αν και η CPCA είναι ευρέως χρησιμοποιημένη, ευρήματα δείχνουν την CPCA να αδυνατεί να διαχωρίσει την εσωτερική από την ολική κίνηση. Παράλληλα με την CPCA, έχει προταθεί η χρήση της dihedral Principal Component

Analysis (dPCA) [73, 74], η οποία λαμβάνει υπόψη τις διεδρες γωνίες (ϕ , ψ), για τον καλύτερο προσδιορισμό της εσωτερικής κίνησης. Οι υπολογισμοί για τις αναλύσεις έγιναν με το πρόγραμμα CARMA [75].

4.2 Eigenvector overlap

Η επικάλυψη (overlap) διακυμάνσεων των ιδιοδιανυσμάτων μπορεί να χρησιμοποιηθεί ως μέτρο σύγκλισης (convergence) για διαφορετικά τροχιακά. Η μαθηματική εξίσωση της επικάλυψης είναι η εξής:

$$\text{overlap}(v, w) = 1/n \sum_{i=1}^n (v_i w_i)^2$$

όπου v και w είναι δύο σύνολα ιδιοδιανυσμάτων, που το καθένα περιέχει n ιδιοδιανύσματα.

Η επικάλυψη των ιδιοδιανυσμάτων είναι ένα μέσο για την ποσοτική σύγκριση των ιδιοδιανυσμάτων των διαφορετικών τροχιακών και μας επιτρέπει να εντοπίσουμε το αν τα αποτελέσματα των τροχιακών συμφωνούν μεταξύ τους. Τιμές κοντά στο ένα δηλώνουν μεγάλη ομοιότητα, ενώ ενώ όσο οι τιμές πλησιάζουν στο μηδέν τα τροχιακά φαίνεται να είναι όλο και πιο αταίριαστα. Τα αποτελέσματα από τον υπολογισμό της επικάλυψης των ιδιοδιανυσμάτων για όλα τα τροχιακά που θα αναλύσουμε παρουσιάζονται στον **Πίνακα 4.1**.

AMBER-99SB	ILDN	ILDN-TIP4	ILDN-340K	ILDN-360K
Eig. 1	0.2364	0.0001	0.1618	0.3312
Eig. 2	0.2577	0.2127	0.1598	0.3031
Eig 3	0.2056	0.3190	0.2458	0.3070
Eig 4	0.3734	0.3044	0.2674	0.2521
Eig 5	0.4174	0.3755	0.3247	0.3108

ILDN	ILDN-TIP4	ILDN-340K	ILDN-360K
Eig. 1	0.4205	0.2353	0.3404
Eig. 2	0.3976	0.2612	0.2067
Eig 3	0.2878	0.3304	0.2758
Eig 4	0.3479	0.5032	0.2855
Eig 5	0.3475	0.5012	0.4058

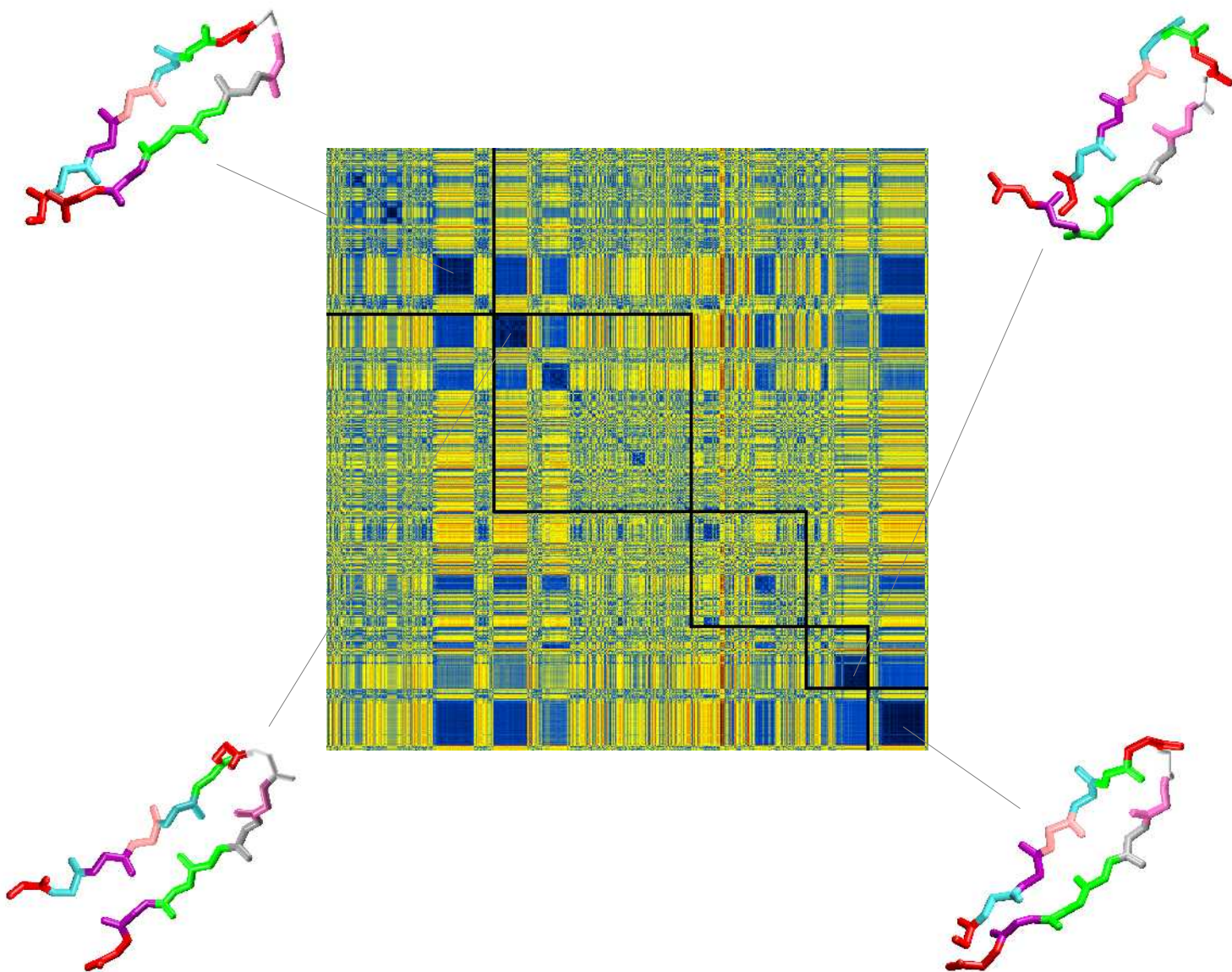
ILDN-TIP4	ILDN-340K	ILDN-360K	ILDN-360K	ILDN-340K
Eig. 1	0.1440	0.0442	Eig. 1	0.2015
Eig. 2	0.0997	0.1456	Eig. 2	0.2875
Eig 3	0.3064	0.2309	Eig 3	0.3991
Eig 4	0.1798	0.2606	Eig 4	0.3197
Eig 5	0.3396	0.2865	Eig 5	0.3252

Πίνακας 4.1: Οι τιμές επικάλυψης μεταξύ όλων των τροχιακών. Ο υπολογισμός γίνεται μεταξύ των πέντε πρώτων ιδιοδιανυσμάτων για κάθε τροχιακό, τα οποία προήλθαν από dPCA. Για τις dPCA χρησιμοποιήθηκαν όλα τα κατάλοιπα εκτός από αυτά των άκρων και κάθε ανάλυση έγινε στην θερμοκρασία του εκάστοτε τροχιακού.

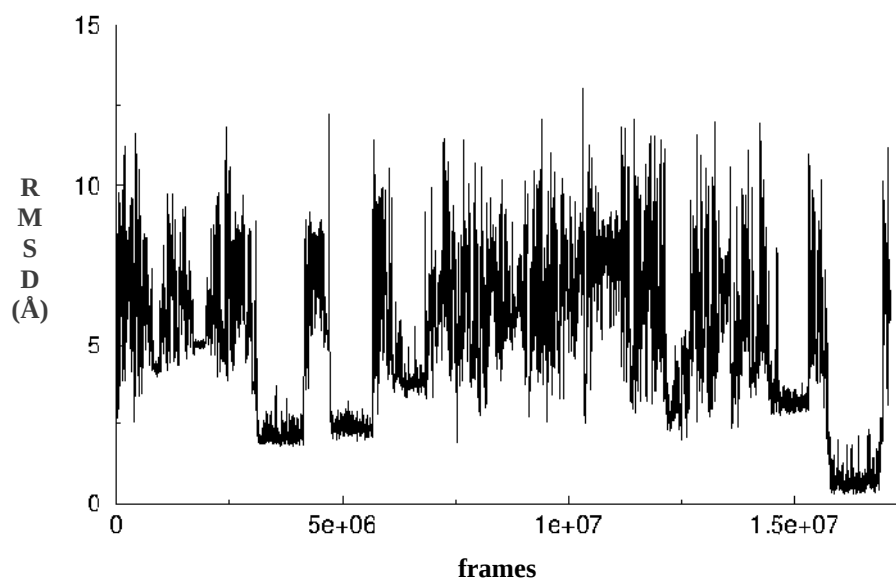
Όπως φαίνεται στους παραπάνω πίνακες, οι τιμές για την επικάλυψη των ιδιοδιανυσμάτων κυμαίνονται μεταξύ των 0.1-0.4. Η μεγαλύτερη σύγκλιση υπάρχει μεταξύ των τροχιακών ILDN και ILDN-340K για τα οποία τα ιδιοδιανύσματα 4 και 5 παρουσιάζουν τιμές κοντά στο 0.5, ενώ τιμές που πλησιάζουν στο 0 φαίνεται να υπάρχουν ανάμεσα στα αρκετά τροχιακά. Τα παραπάνω δεδομένα δηλώνουν ότι τα τροχιακά δεν μοιάζουν σε μεγάλο βαθμό μεταξύ τους.

4.3 RMSD από την φυσική δομή

Επόμενος στόχος, μετά τον υπολογισμό της σύγκλισης των ιδιοδιανυσμάτων, ήταν να εντοπίσουμε αν υπάρχουν γεγονότα αναδίπλωσης στο σύνολο των προσομοιώσεων, τη διάρκεια τους και την ομοιότητα τους με τη φυσική δομή. Αρχικά κατασκευάσαμε ένα καινούριο τροχιακό μέσα από την ένωση όλων των τροχιακών. Η σειρά με την οποία τοποθετήθηκαν τα τροχιακά ήταν AMBER-99SB, ILDN, ILDN-TIP4, ILDN-340K και ILDN-360K. Στη συνέχεια, κατασκευάσαμε πίνακες που παρουσιάζουν το RMSD μεταξύ όλων των ζευγών των δομών που εντοπίστηκαν στο τροχιακό. Η **Εικόνα 4.1** αποτελεί μια χρωματική αναπαράσταση των τιμών του RMSD που προέκυψαν, οι μπλε περιοχές αντιστοιχούν σε χαμηλές τιμές RMSD και υποδεικνύουν την ύπαρξη σταθερών δομών. Η έκταση των περιοχών δηλώνει το χρόνο που πέρασε το πεπτιδίο στις δομές αυτές. Οι κίτρινες περιοχές αντιστοιχούν σε μεσαίες τιμές RMSD και οι κόκκινες περιοχές σε μεγάλες τιμές. Η διακύμανση από κίτρινες σε κόκκινες περιοχές υποδεικνύει όλο και μεγαλύτερες δομικές διαφορές καθώς και την έλλειψη σταθερών διαμορφώσεων. Η **Εικόνα 4.2** αποτελεί μια γραφική αναπαράσταση των τιμών RMSD από τη φυσική δομή, κάθετος άξονας, για κάθε 1000 frames του ολικού τροχιακού, οριζόντιος άξονας.



Εικόνα 4.1: Πίνακας RMSD, κάθε σημείο της διαγωνίου που ξεκινά από πάνω δεξιά αντιστοιχεί σε μία διαμόρφωση από το ολικό τροχιακό (17.114.276 διαμορφώσεις) συνολικής διάρκειας 13.7 μ s. Η προβολή της κάθε διαμόρφωσης, στον οριζόντιο πάνω οριζόντιο και αριστερό κάθετο, χρωματίζεται με βάση την τιμή του RMSD. Οι περιοχές με σκούρες αποχρώσεις του μπλε αντιστοιχούν σε τιμές κοντά στο 0, ενώ η διαβάθμιση τους σε πιο ανοιχτές αποχρώσεις αυξάνεται ανάλογα με τις τιμές του RMSD. Οι κίτρινες αποχρώσεις αντιστοιχούν σε μεσαίες τιμές και οι κόκκινες αντιστοιχούν σε μεγάλες. Το RMSD υπολογίστηκε για τα Cas όλων των καταλοίπων του πεπτιδίου με βήμα 1000 διαμορφώσεις και το εύρος των τιμών είναι 0.3 με 12.8 \AA . Οι μπλε περιοχές που δεν ανήκουν στη διαγώνιο δηλώνουν ομοιότητα μεταξύ των δομών στις μπλε περιοχές της διαγωνίου. Τα μαύρα πλαίσια οριοθετούν το κάθε ξεχωριστό τροχιακό. Οι δομές αντιστοιχούν στις κύριες αλυσίδες των μέσων δομών κάθε συνόλου, με βάση το RMSD, και ο χρωματισμός έγινε με βάση το όνομα καταλοίπου όπως ορίζονται από το VMD, βλ. **Παράρτημα**. Ο υπολογισμός του Πίνακα RMSD έγινε με τη χρήση του προγράμματος CARMA και οι εικόνες των δομών προετοιμάστηκαν με το πρόγραμμα VMD.



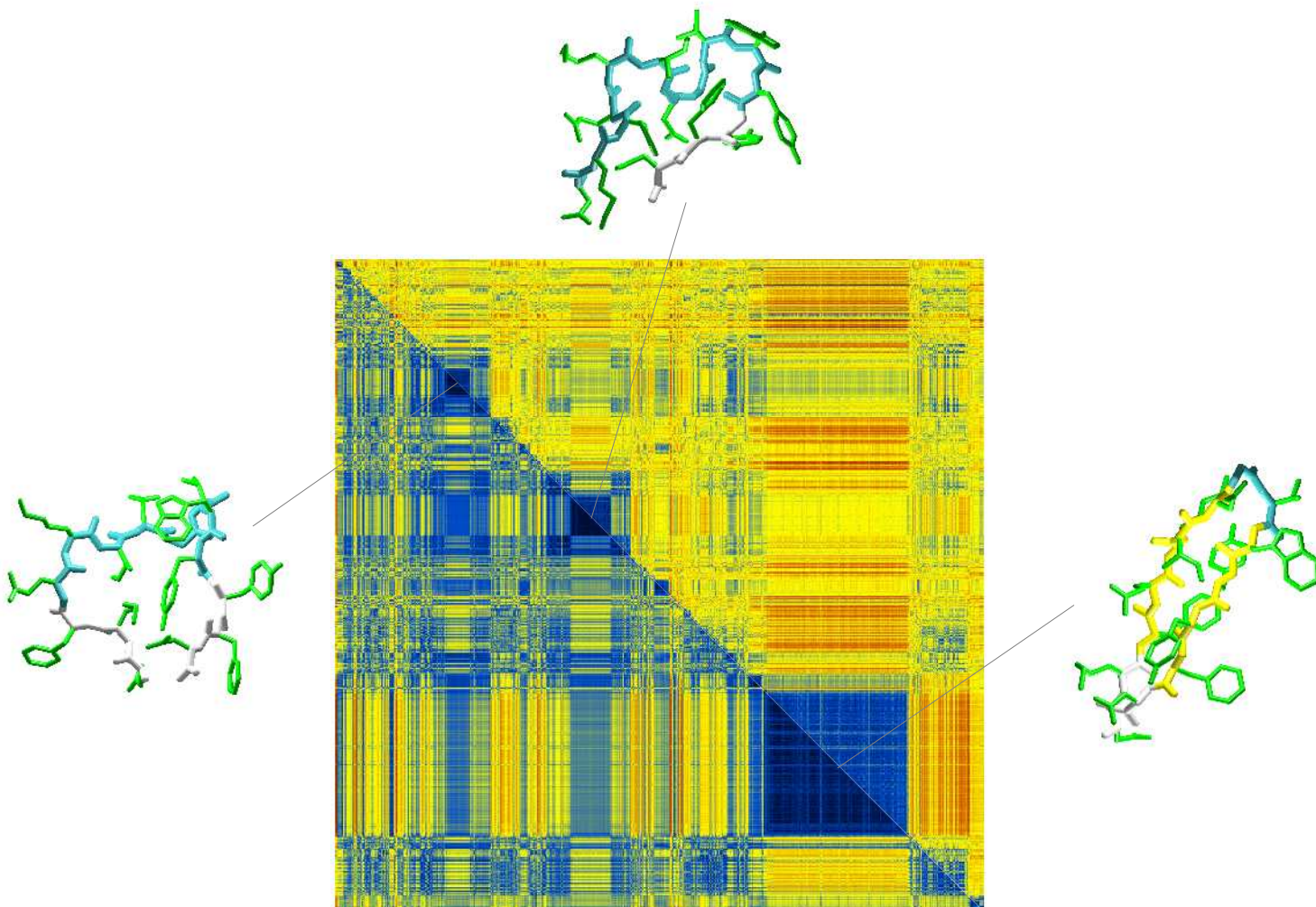
Εικόνα 4.2 Γραφική αναπαράσταση του RMSD για κάθε διαμόρφωση του ολικού τροχιακού ως προς φυσική δομή. Το RMSD υπολογίστηκε με βάση τα Cas όλων των καταλοίπων του πεπτιδίου με βήμα 1000 διαμορφώσεις. Ο υπολογισμός των RMSDs έγινε με τη χρήση του προγράμματος CARMA και η γραφική παράσταση με το πρόγραμμα Xmgr.

Όπως φαίνεται στις παραπάνω εικόνες, το πεπτίδιο υιοθετεί πολλές διαφορετικές διαμορφώσεις των οποίων η διάρκεια κυμαίνεται από 0.5 έως 0.9 μ s. Το μοναδικό γεγονός αναδίπλωσης εμφανίζεται στο τελευταίο τροχιακό και φαίνεται να έχει τη μεγαλύτερη διάρκεια (0.9 μ s) καθώς και το μικρότερο εύρος τιμών RMSD από τη φυσική δομή (0.3 έως 1.5 Å). Στην **Εικόνα 4.1** παρουσιάζονται οι μέσες δομές για κάθε ομάδα (cluster) που ξεπερνά τα 0.5 μ s και φαίνεται να διατηρεί μία σταθερή διαμόρφωση. Εκ πρώτης όψεως, στις ομάδες που δεν υπάρχει αναδίπλωση φαίνεται να μην έχει σχηματιστεί σωστά η στροφή και να μην υπάρχει σωστή διεύθυνση των καταλοίπων της β -φουρκέτας που σχηματίζουν υδρογονοδεσμούς. Επιπλέον, στη δομή που απομονώθηκε από το τροχιακό ILDN-340K φαίνεται να υπάρχει μία ολίσθηση όσον αφορά το σχηματισμό της β -φουρκέτας. Τα αποτελέσματα αυτά πρέπει να αναλυθούν από μία σκεπτικιστική σκοπιά, διότι οι δομές που απεικονίζονται αφορούν τις μέσες δομές των ομάδων, οι οποίες είναι πιθανό να μην είναι αντιπροσωπευτικές για το σύνολο της ομάδας.

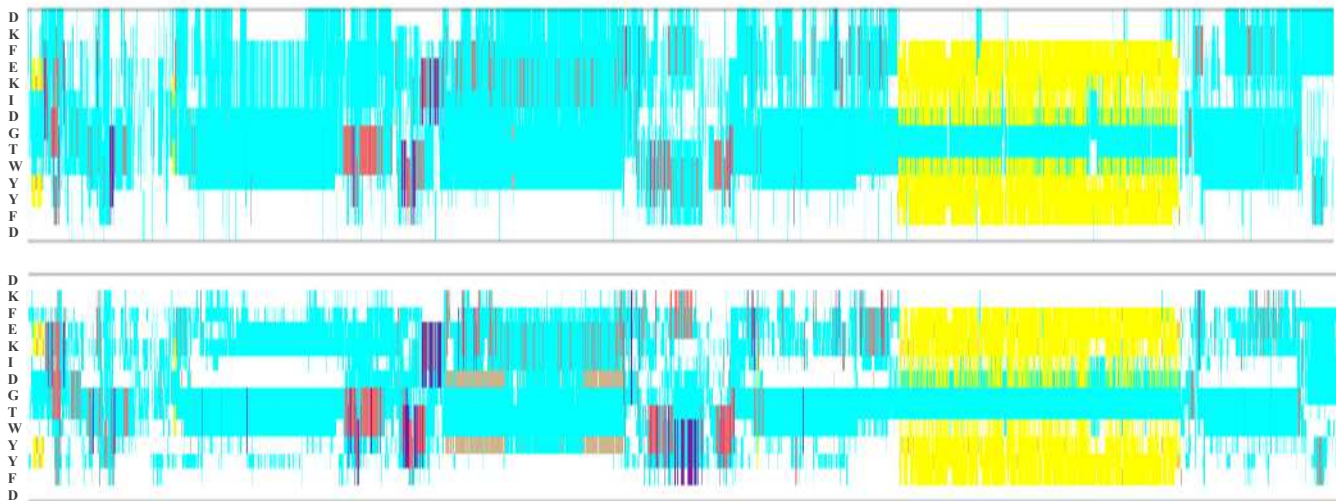
Στη συνέχεια θα αναλύσουμε κάθε τροχιακό ξεχωριστά χρησιμοποιώντας διαφορετικές μεθόδους, έτσι ώστε να αποκτήσουμε μια πιο σφαιρική άποψη για το τι πραγματικά συμβαίνει στα τροχιακά, να εντοπίσουμε σταθερές διαμορφώσεις και να αναδείξουμε τις διαφορές και τις ομοιότητες ανάμεσα τους.

4.4 AMBER-99SB

Το πρώτο κομμάτι της ανάλυσης της προσομοίωσης AMBER-99SB περιλαμβάνει τον υπολογισμό του πίνακα RMSD για το τροχιακό καθώς και τον καθορισμό των στοιχείων δευτεροταγούς δομής. Για το αριστερό κάτω μέρος του πίνακα RMSD χρησιμοποιήθηκαν τα άτομα Ca όλων των καταλοίπων του πεπτιδίου, ενώ για το δεξί πάνω μέρος χρησιμοποιήθηκαν όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα, **Εικόνα 4.3**. Για τον καθορισμό των στοιχείων δευτεροταγούς δομής του τροχιακού εφαρμόσαμε τους αλγόριθμους DSSP και STRIDE, **Εικόνα 4.4**.



Εικόνα 4.3 Πίνακας RMSD. Στο αριστερό κάτω μέρος χρησιμοποιήθηκαν τα άτομα Ca, ενώ στο δεξί πάνω μέρος όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Ο υπολογισμός του RMSD έγινε με βήμα 1000. Στην εικόνα παρουσιάζονται και οι μέσες δομές για κάθε διακριτή ομάδα. Ο χρωματισμός της κύριας αλυσίδας έγινε με βάση τη δευτεροταγή τους δομή όπως προσδιορίστηκε από το VMD, με τον οποίο κατασκευάστηκαν εικόνες των δομών. Το γαλάζιο αντιστοιχεί σε στροφές, το κίτρινο σε β-επιφάνειες και το άσπρο σε τυχαία σπειράματα.



Εικόνα 4.4 Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην πρώτη εικόνα έγινε με τη χρήση του αλγορίθμου DSSP και η δεύτερη με τον STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου DKFEKIDGTWYYFD. Οι γαλάζιες περιοχές αντιπροσωπεύουν στροφές, οι κίτρινες β-επιφάνειες και οι λευκές τυχαία σπειράματα. Οι υπόλοιποι χρωματισμοί και οι αντίστοιχες δομές υπάρχουν στο **Παράρτημα**. Οι υπολογισμοί έγιναν με βήμα 500 διαμορφώσεις.

Οι μέσες δομές των δύο πρώτων ομάδων φαίνεται να απέχουν πολύ από τη φυσική δομή, αφού δεν έχουν σχηματίσει το βασικό χαρακτηριστικό της φυσικής δομής, δηλαδή τη β-φουρκέτα, **Εικόνα 4.3**. Επιπλέον, τα αποτελέσματα των DSSP και STRIDE δείχνουν ότι οι ομάδες αυτές, αποτελούνται από τέτοιες εκτεταμένες δομές που μοιάζουν με στροφές ή περιοχές που δεν αντιστοιχούν σε κάποια συγκεκριμένη διαμόρφωση (οι περιοχές αυτές θα αναφέρονται αδιευκρίνιστες) για το μεγαλύτερο μέρος της ομάδας, **Εικόνα 4.4**. Η τρίτη ομάδα φαίνεται να διατηρεί δομή β-φουρκέτας για το μεγαλύτερο μέρος της. Τα διαφορετικά αμινοξέα μεταξύ των οποίων σχηματίζεται η β-φουρκέτα καθώς και η μετατόπιση της στροφής από τα κατάλοιπα D-G στα κατάλοιπα G-T είναι οι κύριες διαφορές της φουρκέτας από τη φυσική δομή, ενώ παράλληλα υπάρχουν διαφορές στη διεύθυνση των πλευρικών αλυσίδων για τα περισσότερα κατάλοιπα. Τα άκρα του πεπτιδίου για το μεγαλύτερο κομμάτι της προσομοίωσης δεν σχηματίζουν κάποια συγκεκριμένη δομή. Περιστασιακά σε όλη τη διάρκεια της προσομοίωσης εντοπίζονται έλικες και άλλες δομές. Αν και οι τρεις ομάδες φαίνεται να μοιάζουν μεταξύ τους, αριστερό τμήμα πίνακα RMSD, όταν στους υπολογισμούς εισήχθησαν όλα τα άτομα, όπως ήταν αναμενόμενο οι πλευρικές ομάδες εκτόξευσαν τις τιμές του RMSD, γεγονός το οποίο φαίνεται στο δεξί τμήμα του πίνακα RMSD, και υποδεικνύει τη διαφορετικότητα των δομών.

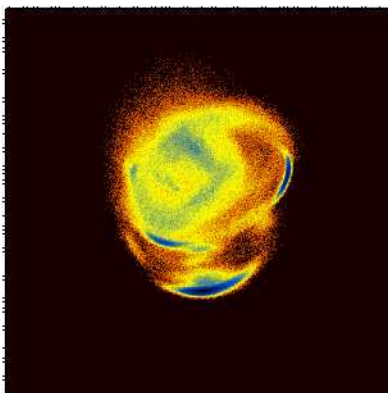
Το επόμενο βήμα της ανάλυσης ήταν ο εντοπισμός ομάδων (clusters), όπως υπολογίζονται από cPCA και dPCA, η ομαδοποίηση τους και η σύγκριση των αποτελεσμάτων από τις διαφορετικές μεθόδους. Στους υπολογισμούς που θα ακολουθήσουν δε θα ληφθούν υπόψη τα κατάλοιπα των άκρων, δηλαδή ασπαρτικό οξύ και λυσίνη για το N-τελικό άκρο και φαινυλαλανίνη και ασπαρτικό οξύ για το C-τελικό άκρο, λόγω της μεγάλης κινητικότητας των πλευρικών ομάδων τους.

Η ανάλυση κύριων συνιστωσών επιτρέπει τον εντοπισμό συγκεκριμένων μοτίβων κίνησης σε ένα τροχιακό. Η ομαδοποίηση των δεδομένων γίνεται με βάση τις κύριες κινήσεις του τροχιακού, οι οποίες αποτυπώνονται στις κύριες συνιστώσες (PC) με τα μεγαλύτερα ιδιοδιανύσματα. Αρχικά εφαρμόσαμε Cartesian Principal Component Analysis (CPCA) για τον εντοπισμό των κυρίαρχων ομάδων χρησιμοποιώντας τις τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν παράλληλα και τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.5**. Πριν την έναρξη των υπολογισμών προηγήθηκε το ταίριασμα (fitting) του τροχιακού με βάση την κύρια αλυσίδα των καταλοίπων 3-12, δηλαδή η απομάκρυνση των ολικών περιστροφών και μεταθέσεων. Οι κύριες ομάδες (clusters) που εντοπίστηκαν ήταν εξής:

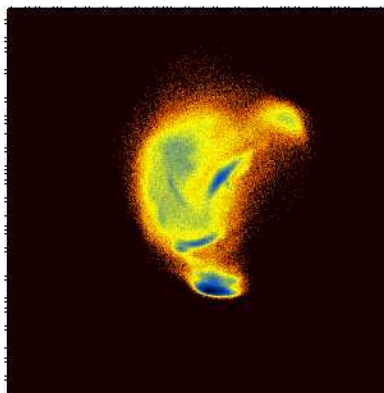
Cluster 1 : 617990 από τα 4703300 frames

Cluster 2 : 212196 από τα 4703300 frames

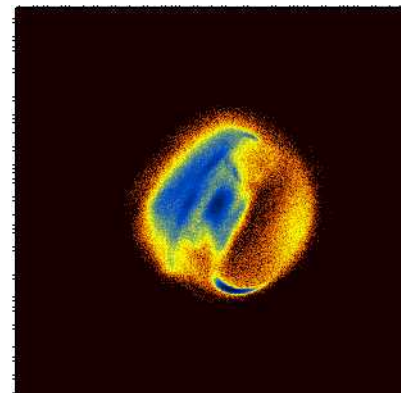
Cluster 3 : 130910 από τα 4703300 frames



PC_01_02



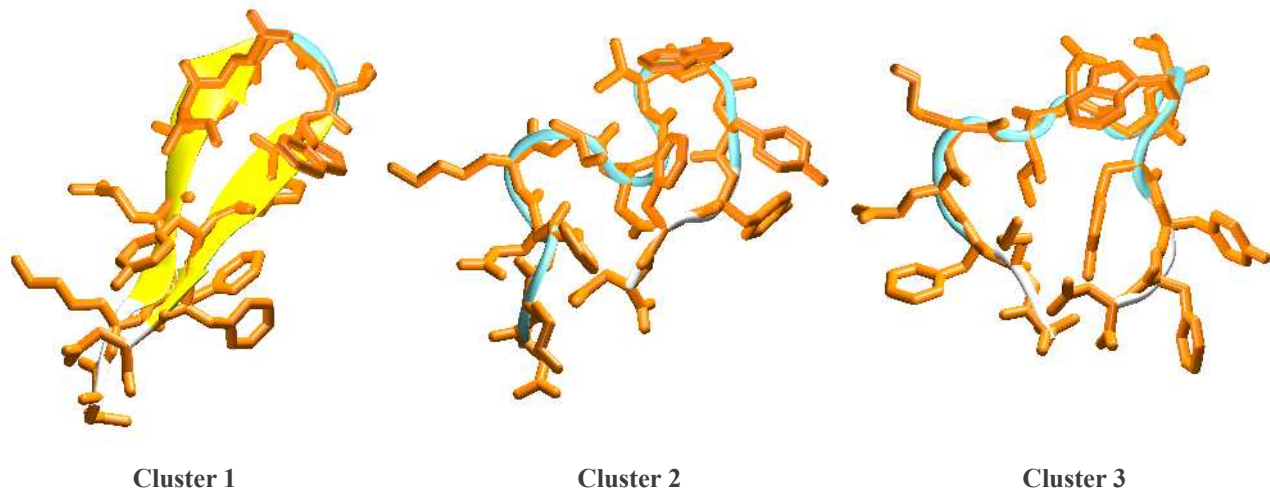
PC_01_03



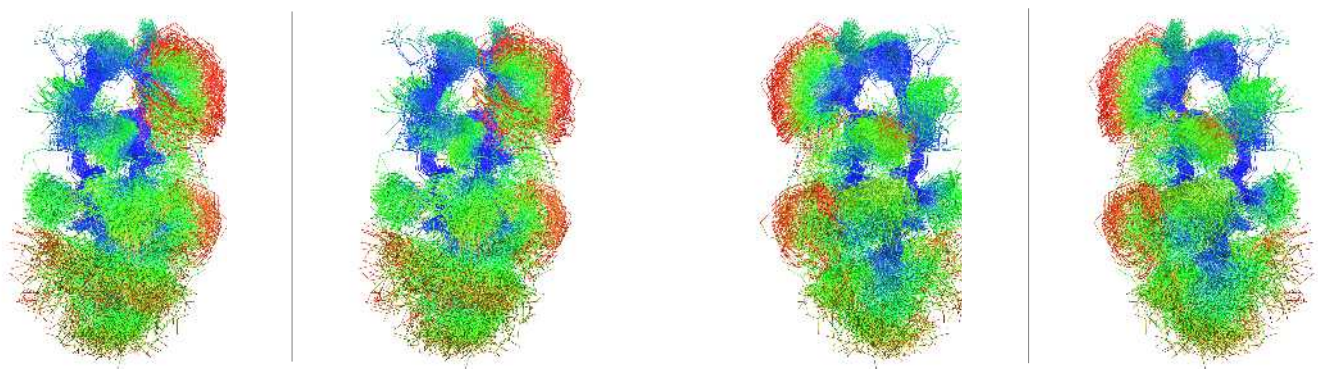
PC_02_03

Εικόνα 4.5 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη. Οι μπλε περιοχές υποδεικνύουν την ύπαρξη τοπικών ελαχίστων ελεύθερης ενέργειας για συγκεκριμένες ομάδες δομών. Όσο πιο σκούρα είναι η απόχρωση του μπλε τόσο μικρότερη είναι η τιμή της ελεύθερης ενέργειας, ενώ το μέγεθος των περιοχών δηλώνει το μέγεθος της ομάδας.

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 13% του συνολικού τροχιακού, η δεύτερη στο 4% και η τρίτη περίπου στο 3%. Στα διαγράμματα ελεύθερης ενέργειας διαγράφονται ξεκάθαρα οι κύριες ομάδες που εντοπίστηκαν από την CPCA. Η έκταση και η θέση των τοπικών ελαχίστων στα δύο πρώτα διαγράμματα διευκολύνει την αντιστοίχιση τους με τις κύριες ομάδες, ενώ στο τρίτο διάγραμμα είναι ασαφής ο διαχωρισμός των ομάδων. Για παράδειγμα, στο πρώτο διάγραμμα η μπλε περιοχή στο κάτω μέρος του διαγράμματος αντιστοιχεί στην πρώτη ομάδα, η δεξιά περιοχή αντιστοιχεί στη δεύτερη ομάδα και η περιοχή στο κεντρικό και αριστερό τμήμα στην τρίτη, **Εικόνα 4.5**. Στη συνέχεια απομονώθηκε η αντιπροσωπευτική δομή για κάθε ομάδα, **Εικόνα 4.6**.



Εικόνα 4.6 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την CPCA.



Εικόνα 4.7 Υπέρθωση δομών σε όψη stereo από την κυρίαρχη ομάδα. Ο χρωματισμός του πεπτιδίου έγινε με βάση την κινητικότητα του κάθε ατόμου. Στο αριστερό τμήμα, το N-τελικό άκρο βρίσκεται στα αριστερό και κάτω μέρος του πεπτιδίου, ενώ στο δεξί τμήμα έχει γίνει περιστροφή κατά 180°.

Οι αντιπροσωπευτικές δομές των ομάδων ταιριάζουν σε μεγάλο βαθμό με τις μέσες δομές που απομονώθηκαν με βάση τα RMSDs, **Πίνακας 4.2**. Εκτός από την υψηλή κινητικότητα των ακραίων καταλοίπων, η υπέρθεση των δομών της πρώτης ομάδας φανερώνει την υψηλή κινητικότητα των πλευρικών ομάδων των καταλοίπων 10-13, δηλαδή της τρυπτοφάνης, της τυροσίνης και της φαινυλαλανίνης, **Εικόνα 4.7**.

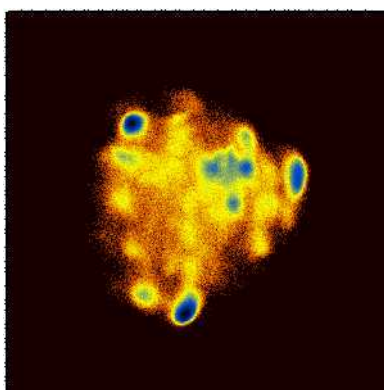
Για την πληρότητα της ανάλυσης και τη σύγκριση των αποτελεσμάτων μεταξύ τους, εφαρμόσαμε dihedral Principal Component Analysis (dPCA), με και χωρίς τις γωνίες χ_1 , στο τροχιακό. Στη dPCA χρησιμοποιούνται οι δίεδρες γωνίες του πεπτιδίου και συνεπώς δεν είναι απαραίτητη η απομάκρυνση των ολικών περιστροφών και μεταθέσεων. Για τους υπολογισμούς χρησιμοποιήθηκαν οι τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν και τα διαγράμματα ελεύθερης ενέργειας. Οι κύριες ομάδες που εντοπίστηκαν στους υπολογισμούς χωρίς τις γωνίες χ_1 ήταν οι εξής:

Cluster 1 : 358346 από τα 4703300 frames

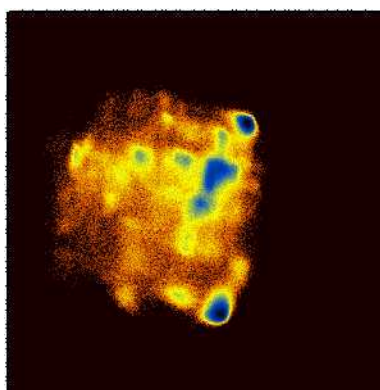
Cluster 2 : 477048 από τα 4703300 frames

Cluster 3 : 66879 από τα 4703300 frames

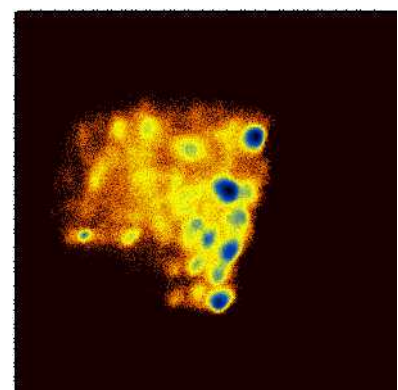
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί περίπου στο 7% του συνολικού τροχιακού, η δεύτερη στο 10% και η τρίτη περίπου στο 1%. Η κυρίαρχη ομάδα είναι η δεύτερη, αφού έχει τον μεγαλύτερο αριθμό στιγμιότυπων συγκριτικά με τις υπόλοιπες. Όπως φαίνεται στα διαγράμματα ελεύθερης ενέργειας, υπάρχουν αρκετοί μικροί πληθυσμοί από διαμορφώσεις, οι οποίοι δεν πέρασαν το κατώφλι και δεν εντοπίστηκαν ως διακριτές ομάδες, **Εικόνα 4.8**.



PC_01_02



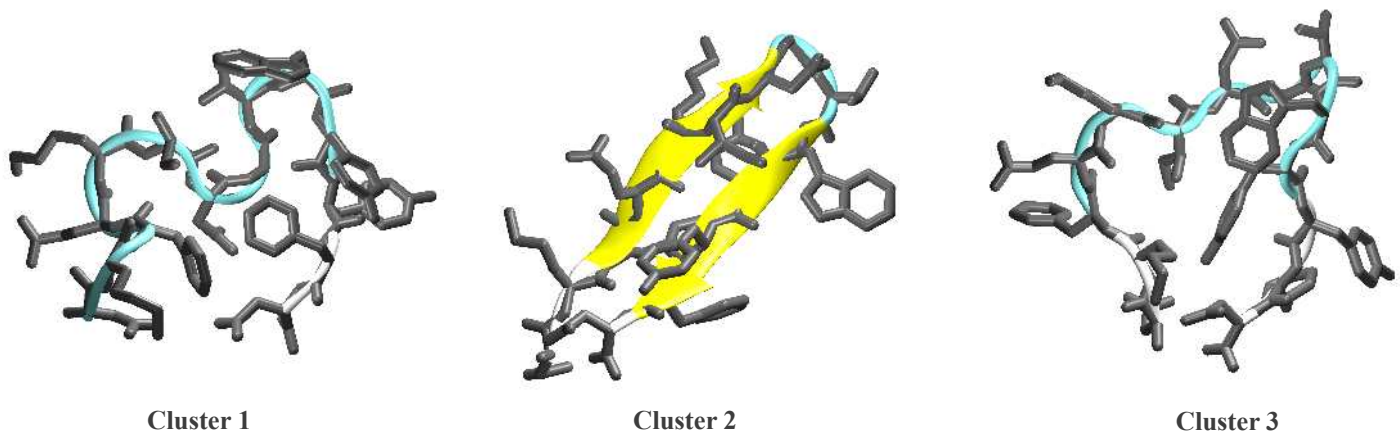
PC_01_03



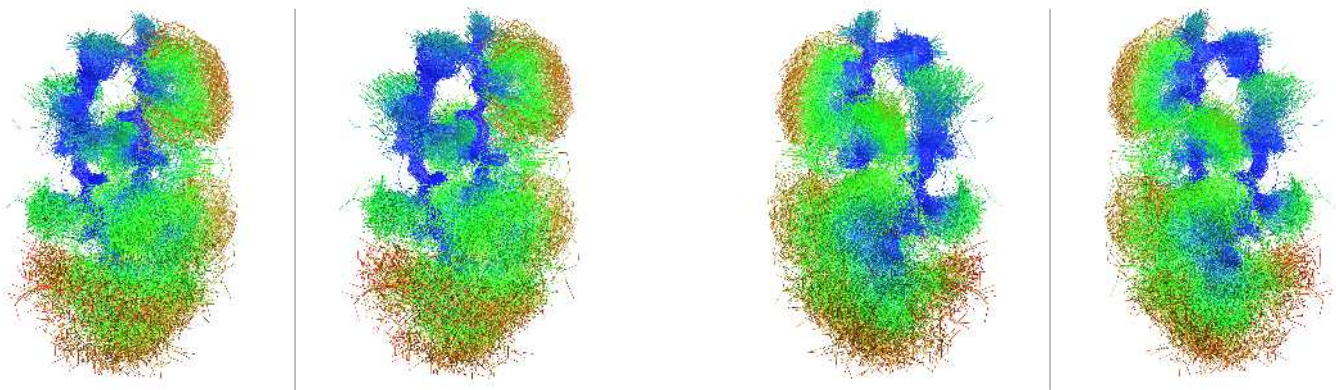
PC_02_03

Εικόνα 4.8 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

Αν και υπάρχουν αρκετοί πληθυσμοί, οι κύριες ομάδες διακρίνονται εύκολα από τα παραπάνω διαγράμματα. Για κάθε κύρια ομάδα απομονώσαμε τις αντιπροσωπευτικές δομές, **Εικόνα 4.9**.



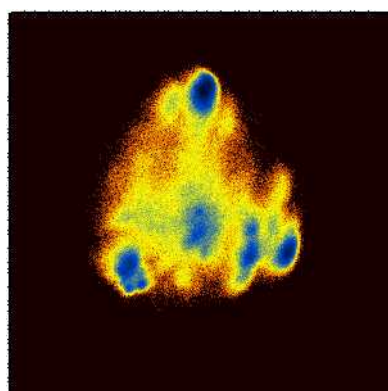
Εικόνα 4.9 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA χωρίς γωνίες χ_1 .



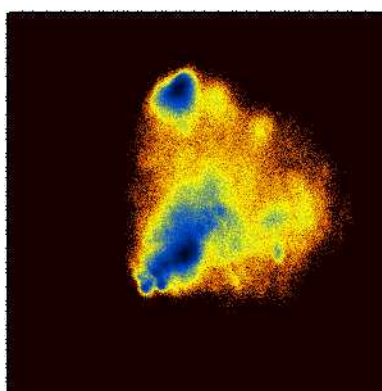
Εικόνα 4.10 Υπέρθωση δομών σε όψη stereo από την κυρίαρχη ομάδα.

Τα αποτελέσματα της ανάλυσης χρησιμοποιώντας τις γωνίες χ_1 παρουσιάζονται στη συνέχεια:

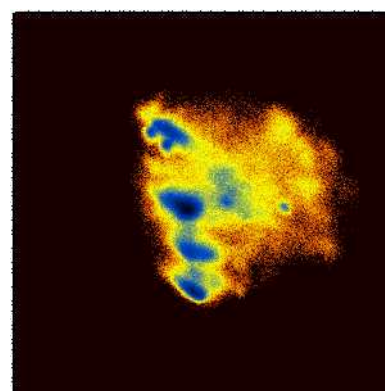
- Cluster 1 : 393959 από τα 4703300 frames
- Cluster 2 : 183064 από τα 4703300 frames
- Cluster 3 : 33402 από τα 4703300 frames
- Cluster 4 : 146747 από τα 4703300 frames
- Cluster 5 : 43098 από τα 4703300 frames



PC_01_02

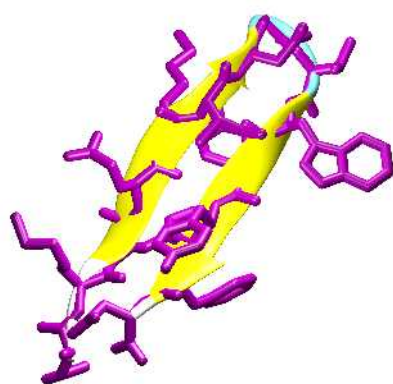


PC_01_03

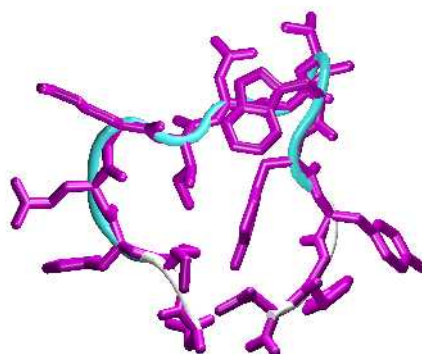


PC_02_03

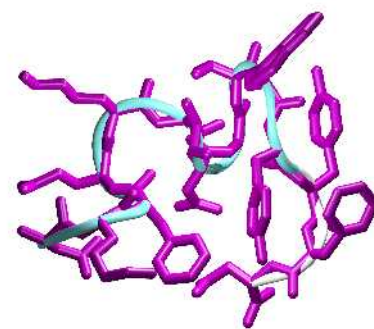
Εικόνα 4.11 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



Cluster 1

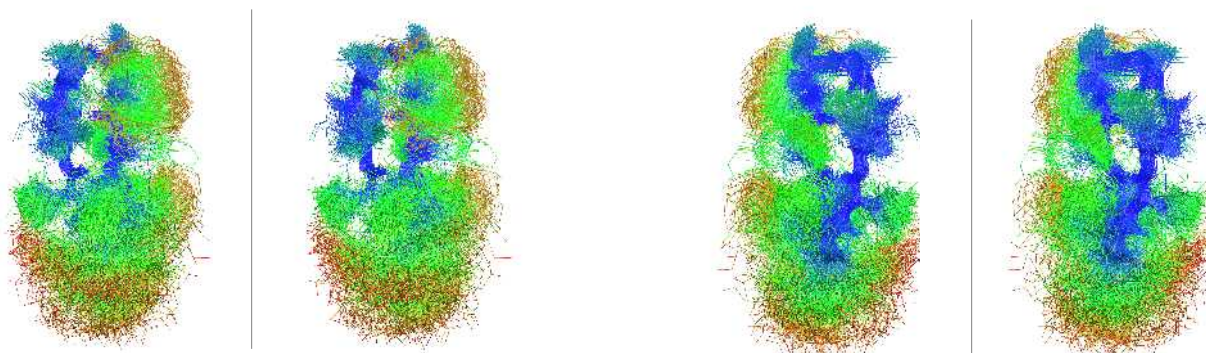


Cluster 2



Cluster 4

Εικόνα 4.12 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA χρησιμοποιώντας τις γωνίες χ_1 .



Εικόνα 4.13 Υπέρθυση δομών σε όψη stereo από την κυρίαρχη ομάδα.

Παρόλες τις διαφορές στην ομαδοποίηση των δεδομένων η κάθε μία PCA φαίνεται να κατάφερε να εντοπίσει τις κύριες κινήσεις του τροχιακού. Τα αποτελέσματα των PCA συγκλίνουν μεταξύ τους και συμφωνούν με τα αποτελέσματα από τον πίνακα RMSD, όπως φαίνεται στον **Πίνακα 4.2**, και τις μεθόδους DSSP και STRIDE. Η απόκλιση των δομών από την φυσική δομή παρουσιάζεται στον **Πίνακα 4.8**, στο τέλος του κεφαλαίου, όπου παρατίθενται οι τιμές RMSD των κύριων διαμορφώσεων που εντοπίστηκαν για κάθε προσομοίωση ξεχωριστά από τη φυσική δομή.

	RMSD_1	CPCA_3	DPCA_3	DPCA_CHI1_2
RMSD_1	-	0.72	0.62	0.62
CPCA_3	0.72	-	0.64	0.64
DPCA_3	0.62	0.64	-	0.00
DPCA_CHI1_2	0.62	0.64	0.00	-

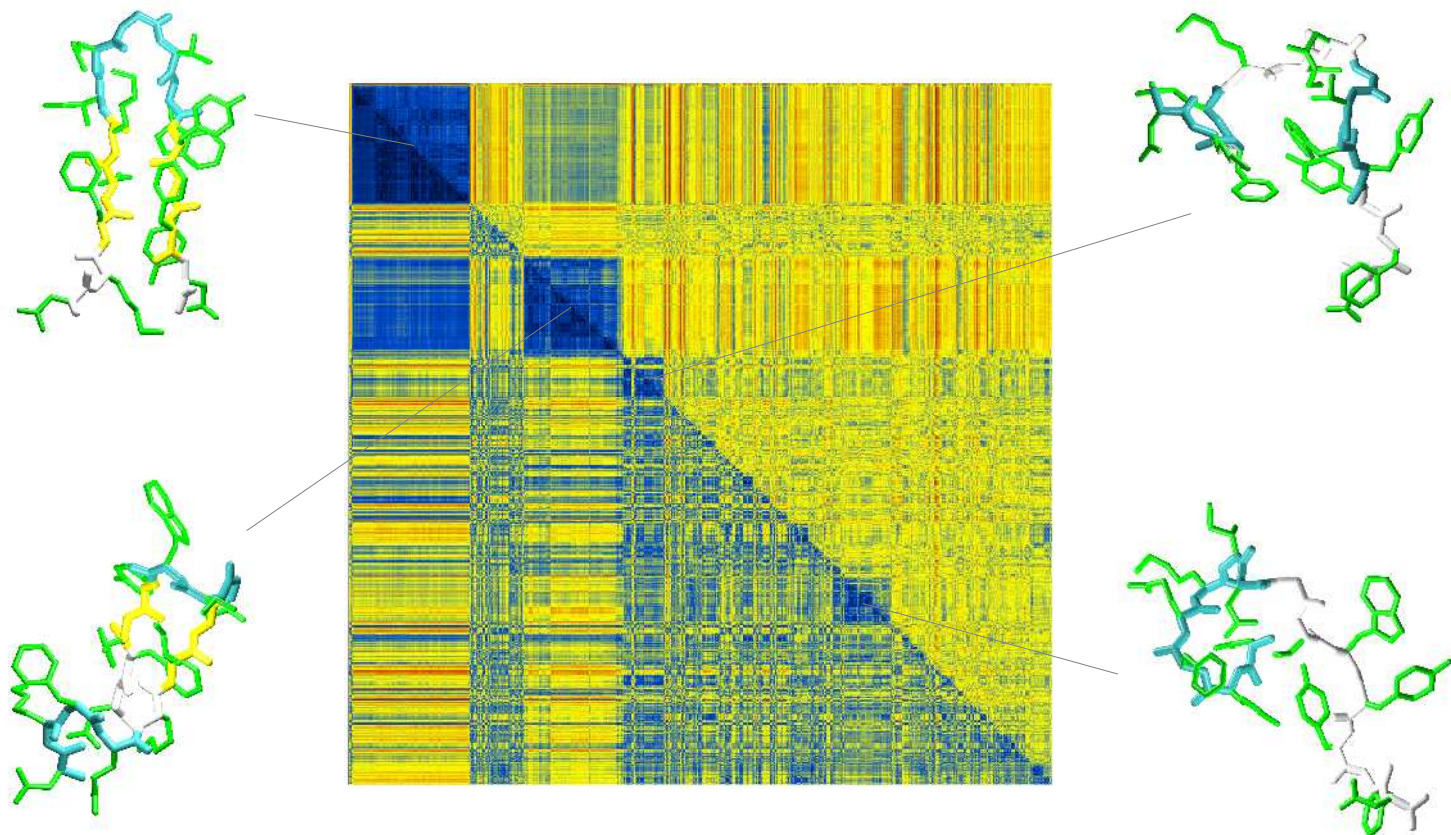
	RMSD_2	CPCA_2	DPCA_1	DPCA_CHI1_4
RMSD_2	-	0.35	1.52	1.98
CPCA_2	0.35	-	2.47	1.98
DPCA_1	1.52	1.47	-	1.38
DPCA_CHI1_4	1.98	1.98	1.38	-

	RMSD_3	CPCA_1	DPCA_2	DPCA_CHI1_1
RMSD_3	-	0.38	0.52	0.52
CPCA_1	0.38	-	0.45	0.45
DPCA_2	0.52	0.45	-	0.00
DPCA_CHI1_1	0.52	0.45	0.00	-

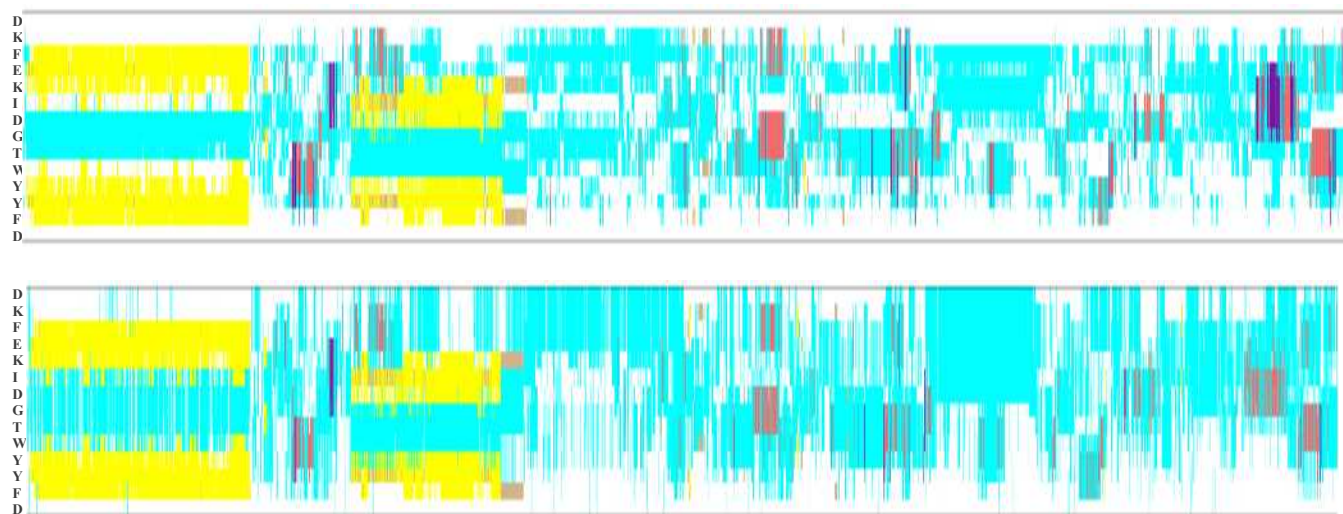
Πίνακες 4.2 RMSD μεταξύ των αντιπροσωπευτικών δομών που απομονώθηκαν για κάθε σημαντική ομάδα. Για τους υπολογισμούς χρησιμοποιήθηκαν όλα τα άτομα της κύριας αλυσίδας. Οι τίτλοι των επικεφαλίδων αναφέρονται στην μέθοδο και την ομάδα για την οποία έγιναν οι υπολογισμοί. Οι τιμές RMSD αναγράφονται σε Angstrom.

4.5 ILDN

Αρχικά υπολογίσαμε τον πίνακα RMSD, **Εικόνα 4.14**. Στο κάτω αριστερό μέρος του πίνακα χρησιμοποιήθηκαν τα άτομα Ca όλων των καταλοίπων του πεπτιδίου. ενώ για το δεξί πάνω μέρος χρησιμοποιήθηκαν όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Για τον καθορισμό των στοιχείων δευτεροταγούς δομής του τροχιακού εφαρμόσαμε τους αλγόριθμους DSSP και STRIDE, **Εικόνα 4.15**.



Εικόνα 4.14 Πίνακας RMSD. Στο αριστερό κάτω μέρος χρησιμοποιήθηκαν τα άτομα Ca, ενώ στο δεξί πάνω μέρος όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Ο υπολογισμός του RMSD έγινε με βήμα 1000. Στην εικόνα παρουσιάζονται και οι μέσες δομές για κάθε διακριτή ομάδα.



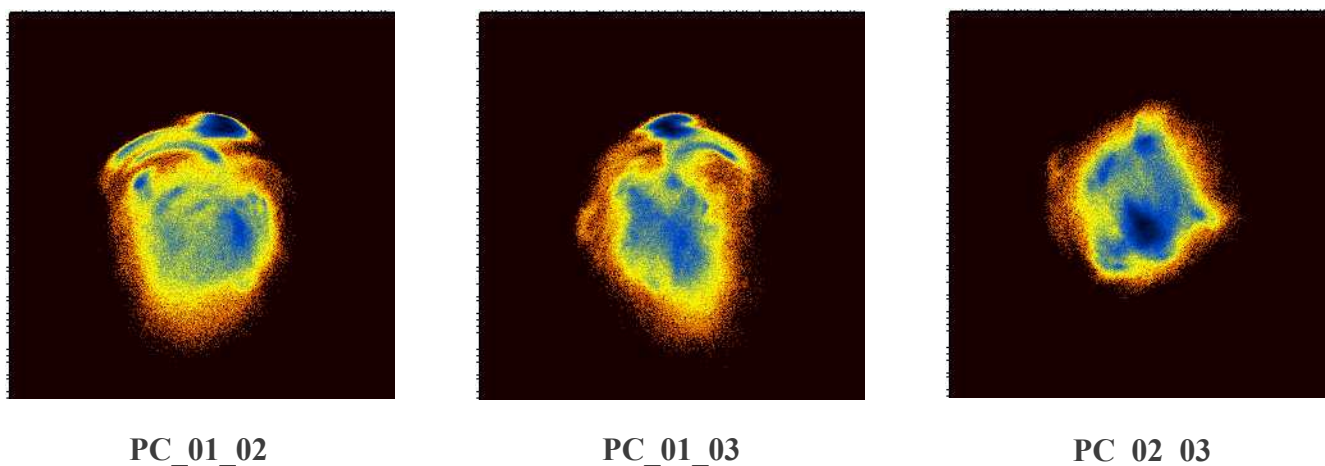
Εικόνα 4.15 Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην πρώτη εικόνα έγινε με τη χρήση του αλγορίθμου DSSP και η δεύτερη με τον STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου DKFEKIDGTWYYFD. Οι υπολογισμοί έγιναν με βήμα 500 διαμορφώσεις.

Η μέση δομή της πρώτης ομάδας φαίνεται να διατηρεί για αρκετό χρόνο τη δομή β-φουρκέτας, η οποία όμως παρουσιάζει διαφορές από την φυσική δομή. Αυτές οι διαφορές αφορούν τα κατάλοιπα που συμμετέχουν στο σχηματισμό της β-φουρκέτας καθώς και το σχηματισμό της χαρακτηριστικής στροφής τύπου I. Παρά τις διαφορές, οι περισσότερες πλευρικές ομάδες φαίνεται να έχουν πακεταριστεί σωστά στην πρώτη δομή. Τα αποτελέσματα από τις μεθόδους DSSP και STRIDE συμφωνούν σε μεγάλο βαθμό, διαφέρουν όμως στην ανάθεση στοιχείων δευτεροταγούς δομής για την πρώτη ομάδα, κυρίως όσον αφορά τα κατάλοιπα που συμμετέχουν στη στροφή και τη β-φουρκέτα, και για τους μικρούς πληθυσμούς των οποίων απομονώθηκαν οι μέσες δομές. Η δεύτερη ομάδα παρουσιάζει μεγαλύτερες αποκλίσεις από τη φυσική δομή σε σχέση με την πρώτη, αφού η β-φουρκέτα και η στροφή φαίνεται να έχουν μετατοπιστεί προς το C-τελικό άκρο. Μικροί πληθυσμοί από στροφές και αδιευκρίνιστες περιοχές καταλαμβάνουν το μεγαλύτερο μέρος του τροχιακού, ενώ όπως και στο προηγούμενο τροχιακό παρουσιάζονται έλικες και άλλες δομές για μικρά χρονικά διαστήματα που προφανώς δεν είναι αρκετά σταθερές για να δημιουργήσουν ξεχωριστούς πληθυσμούς.

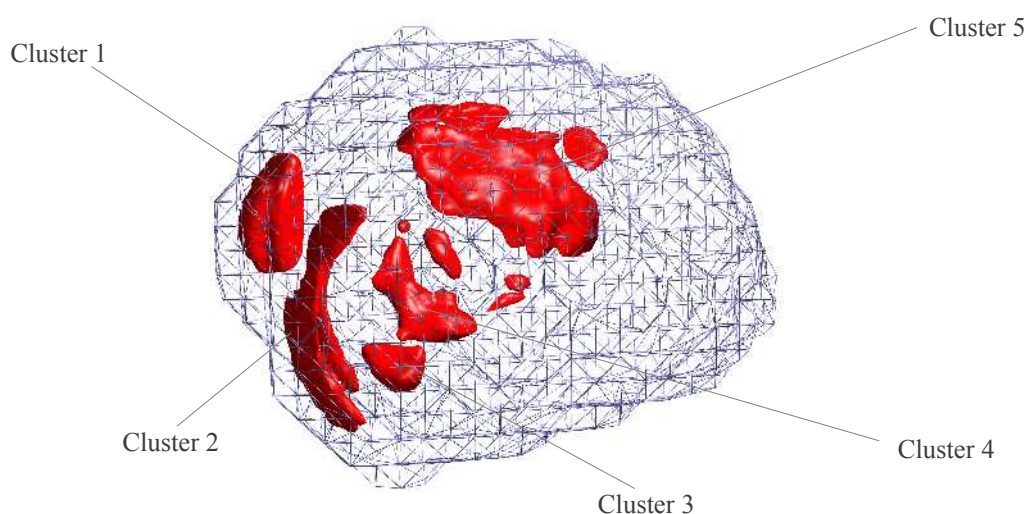
Στη συνέχεια, εφαρμόσαμε Cartesian Principal Component Analysis (CPCA) για τον εντοπισμό των κύριων ομάδων χρησιμοποιώντας τις τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν παράλληλα και τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.16**. Πριν την έναρξη των υπολογισμών προηγήθηκε το ταίριασμα (fitting) του τροχιακού με βάση την κύρια αλυσίδα των καταλοίπων 3-12. Οι κύριες ομάδες (clusters) που εντοπίστηκαν ήταν εξής:

- Cluster 1 : 851625 από τα 5610675 frames
- Cluster 2 : 325581 από τα 5610675 frames
- Cluster 3 : 99399 από τα 5610675 frames
- Cluster 4 : 107229 από τα 5610675 frames
- Cluster 5 : 692455 από τα 5610675 frames
- Cluster 6 : 50993 από τα 5610675 frames
- Cluster 7 : 17418 από τα 5610675 frames

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 15% του συνολικού τροχιακού, η πέμπτη στο 12%, η δεύτερη περίπου στο 6% και η τέταρτη περίπου στο 2%. Οι υπόλοιπες ομάδες είναι πολύ μικρές συγκριτικά με το μέγεθος του τροχιακού. Λόγω πολλών ομάδων, στα διαγράμματα ελεύθερης ενέργειας δε διαγράφονται ξεκάθαρα οι κύριες ομάδες εκτός από την πρώτη. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στην **Εικόνα 4.17**.



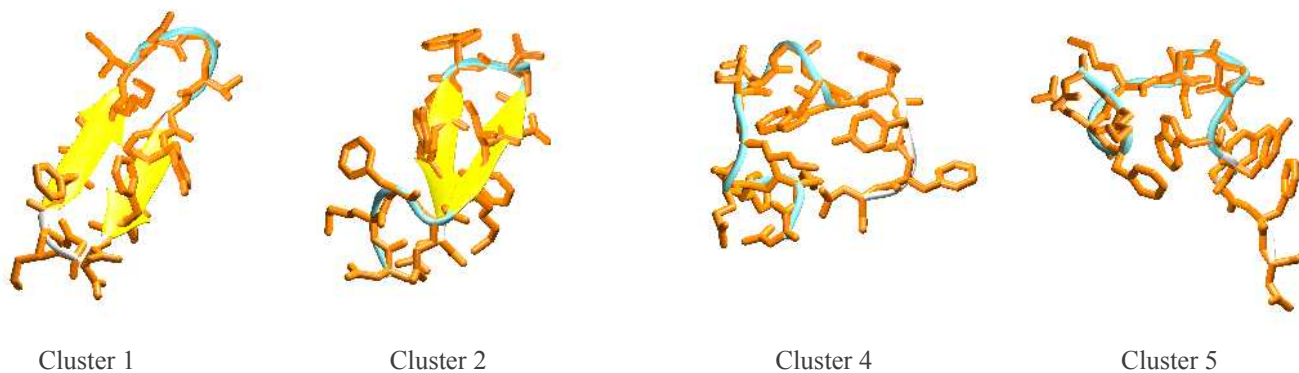
Εικόνα 4.16 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



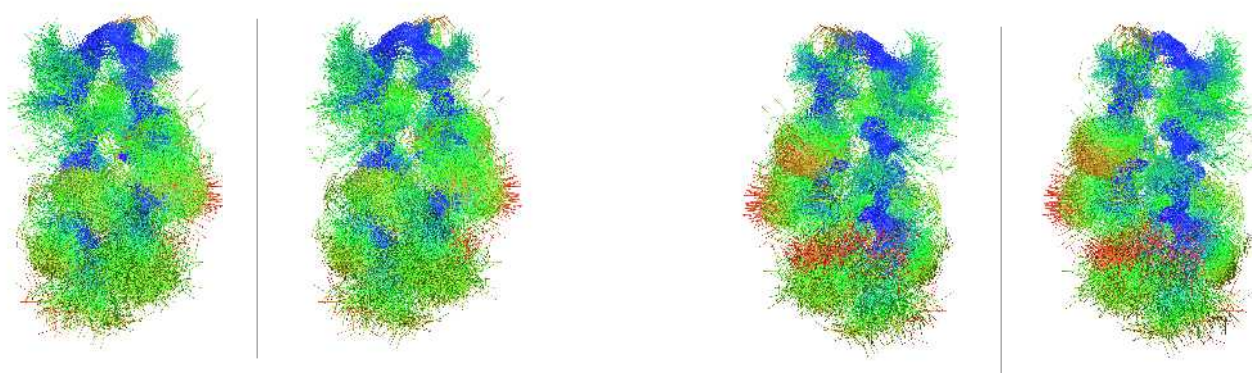
Εικόνα 4.17 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC. Η εικόνα προετοιμάστηκε με το πρόγραμμα VMD

Στη συνέχεια απομονώθηκε η αντιπροσωπευτική δομή για κάθε μία από τις μεγαλύτερες ομάδες, **Εικόνα 4.18**, και κατασκευάστηκαν εικόνες από την υπέρθεση των διαμορφώσεων της κυρίαρχης ομάδας, **Εικόνα 4.19**.

Η αντιπροσωπευτική δομή της δεύτερης ομάδας ταιριάζει με την αντίστοιχη που απομονώθηκε από τους υπολογισμούς με βάση το RMSD, **Πίνακας 4.3**. Τα αποτελέσματα δεν είναι ίδια, για τις υπόλοιπες δομές που παρουσιάζουν σημαντικές διαφορές σε επίπεδο κύριας αλυσίδας και διευσθέτησης των πλευρικών ομάδων. Οι υπολογισμοί των αντιπροσωπευτικών δομών από τους πίνακες RMSD έγιναν με γραμμικό τρόπο, δηλαδή επιλέγοντας συνεχόμενες διαμορφώσεις από τις περιοχές με χα-



Εικόνα 4.18 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την CPCA.



Εικόνα 4.19 Υπέρθυση δομών σε όψη stereo από την κυρίαρχη ομάδα.

μηλό RMSD και υπολογίζοντας τη μέση διαμόρφωση αυτών, ενώ στην PCA η ομαδοποίηση των δομών δεν γίνεται γραμμικά σε μία περιοχή αλλά σε όλη την έκταση του τροχιακού. Η υπέρθεση των δομών της πρώτης ομάδας παρουσιάζει την υψηλή κινητικότητα των άκρων και των πλευρικών ομάδων των καταλοίπων 10-13, **Εικόνα 4.19**.

Τα αποτελέσματα από dihedral Principal Component Analysis (dPCA), με και χωρίς τις γωνίες χ_1 , παρουσιάζονται στις επόμενες σελίδες. Για την ανάλυση χρησιμοποιήθηκαν όλα τα άτομα των καταλοίπων 3-12 και για τους υπολογισμούς οι τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν και τα διαγράμματα ελεύθερης ενέργειας. Οι κύριες ομάδες που εντοπίστηκαν στους υπολογισμούς χωρίς τις γωνίες χ_1 ήταν οι εξής:

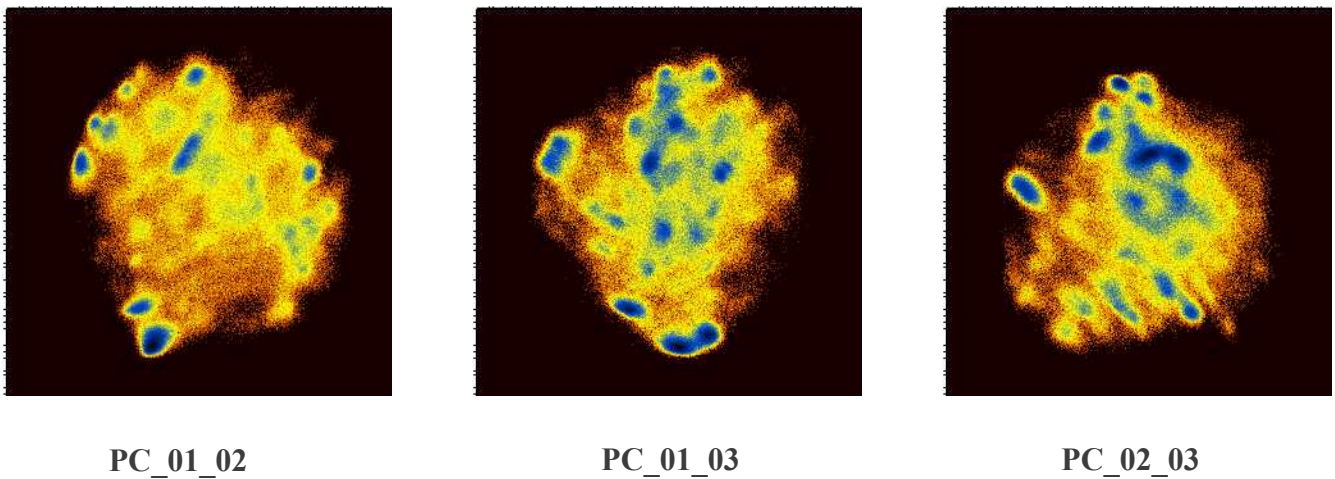
- Cluster 1 : 321976 από τα 5610675 frames
- Cluster 2 : 71109 από τα 5610675 frames
- Cluster 3 : 74569 από τα 5610675 frames
- Cluster 4 : 28238 από τα 5610675 frames

Cluster 5 : 22276 από τα 5610675 frames

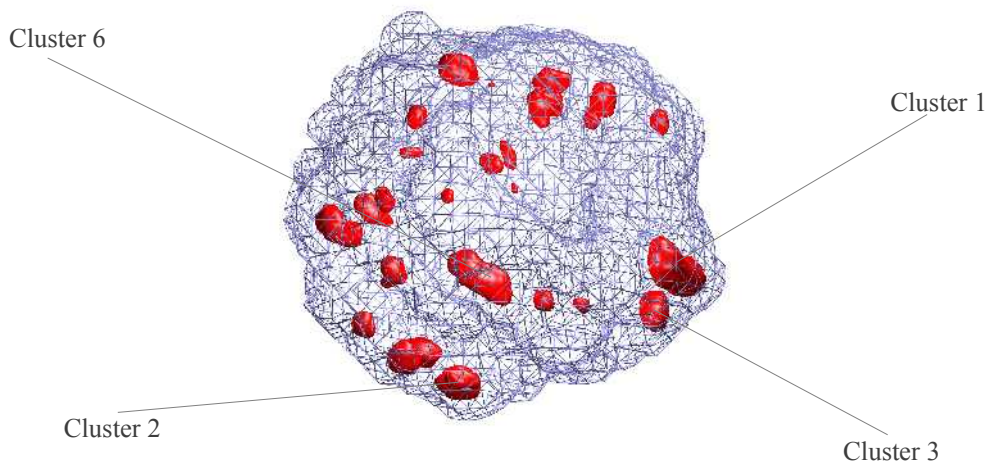
Cluster 6 : 53890 από τα 5610675 frames

Cluster 7 : 3427 από τα 5610675 frames

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί περίπου στο 5% του συνολικού τροχιακού, ενώ οι υπόλοιπες μεγάλες ομάδες κυμαίνονται γύρω από το 1%. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στις **Εικόνες 4.20** και **4.21**.

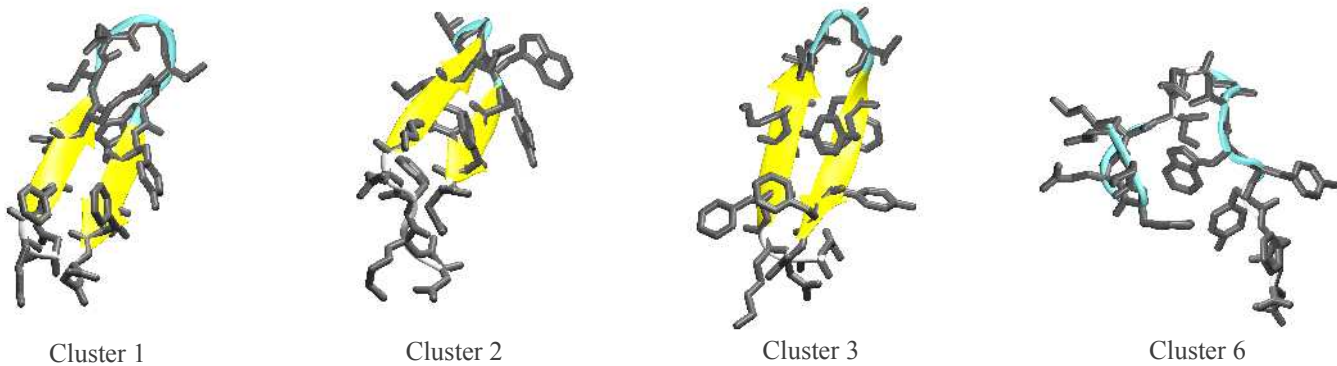


Εικόνα 4.20 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

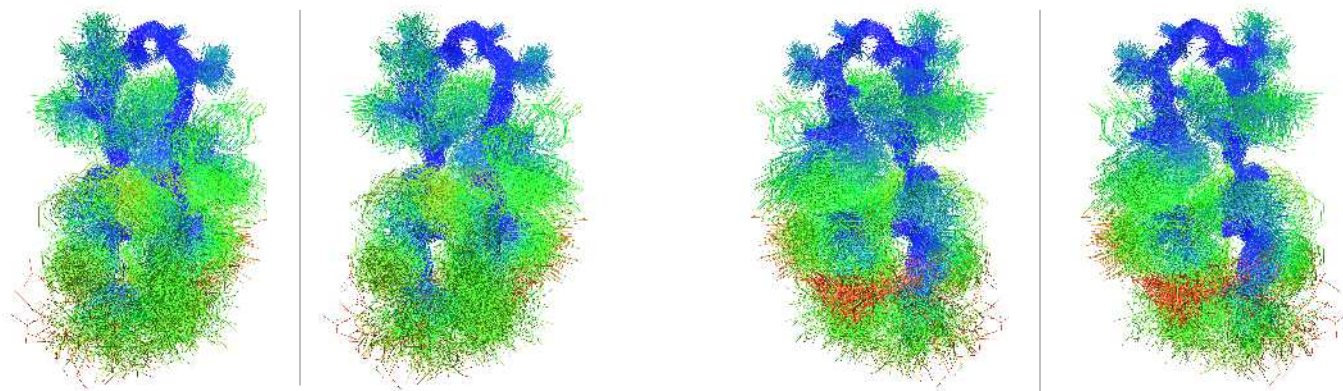


Εικόνα 4.21 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC.

Μετά την ομαδοποίηση των σημαντικότερων ομάδων απομονώθηκαν οι αντιπροσωπευτικές δομές, **Εικόνα 4.22**, και προετοιμάστηκαν εικόνες από την υπέρθεση των διαμορφώσεων της κυρίαρχης ομάδας, **Εικόνα 4.23**.



Εικόνα 4.22 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.



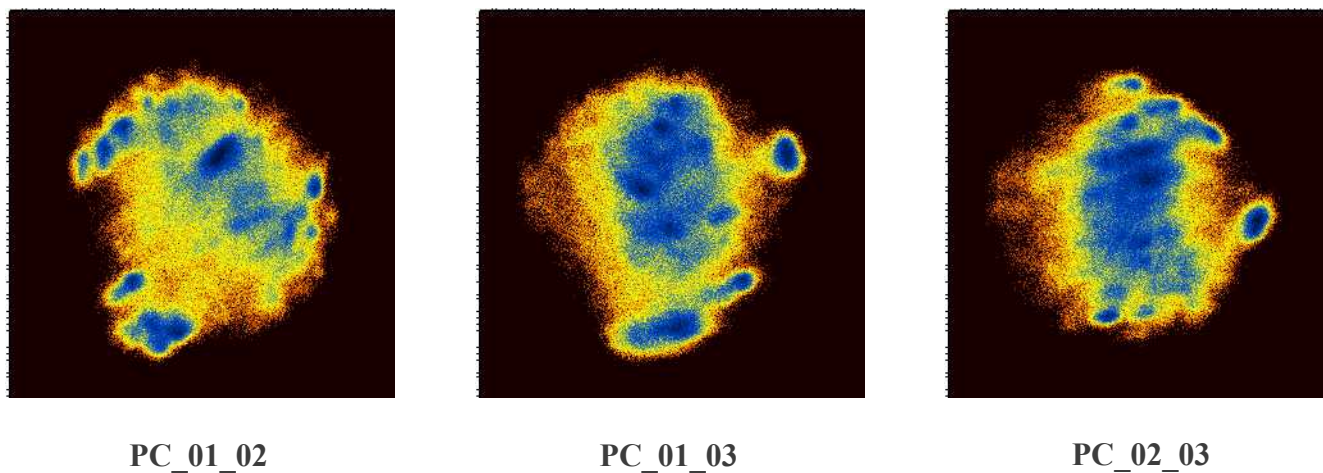
Εικόνα 4.23 Υπέρθωση δομών σε όψη stereo από την κυρίαρχη ομάδα.

Τα αποτελέσματα της ανάλυσης χρησιμοποιώντας τις γωνίες χ_1 παρουσιάζονται στη συνέχεια:

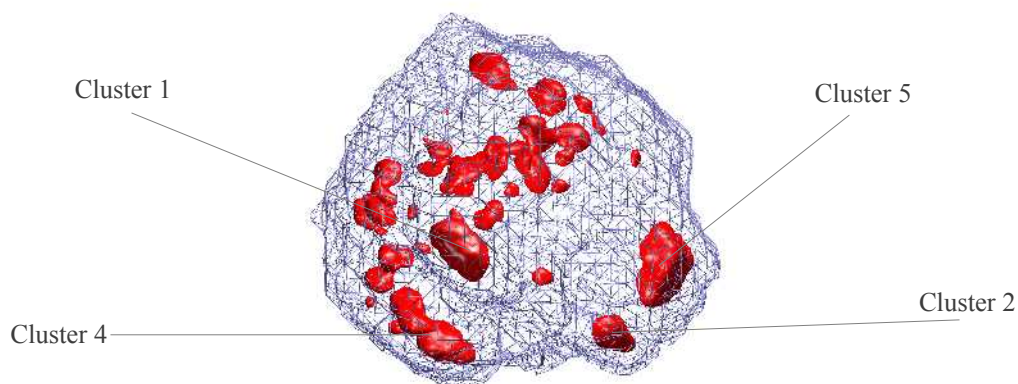
- Cluster 1 : 227277 από τα 5610675 frames
- Cluster 2 : 136782 από τα 5610675 frames
- Cluster 3 : 132657 από τα 5610675 frames
- Cluster 4 : 305277 από τα 5610675 frames
- Cluster 5 : 446726 από τα 5610675 frames
- Cluster 6 : 36471 από τα 5610675 frames
- Cluster 7 : 30025 από τα 5610675 frames

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί περίπου στο 4% του συνολικού τροχιακού, η δεύτερη στο 2%, η τέταρτη στο 5% και η πέμπτη περίπου στο 8%. Η κυρίαρχη

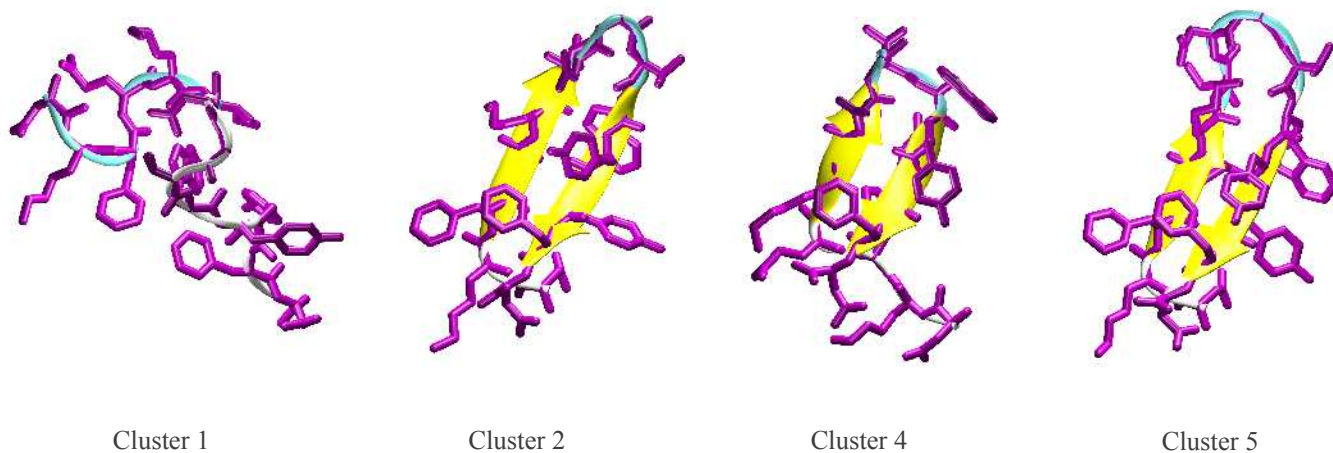
ομάδα είναι η πέμπτη αφού καταλαμβάνει το μεγαλύτερο κομμάτι του τροχιακού. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στην **Εικόνα 4.24** και **Εικόνα 4.25**.



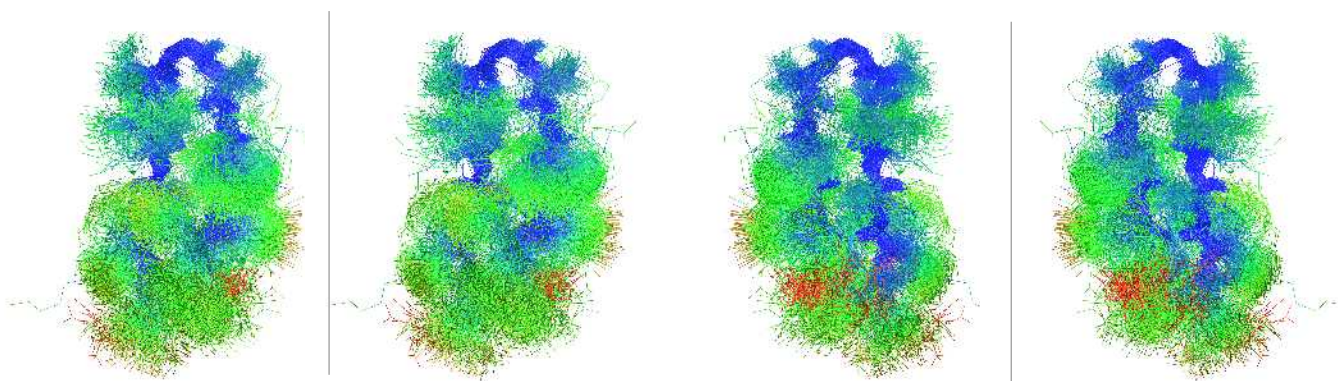
Εικόνα 4.24 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



Εικόνα 4.25 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC.



Εικόνα 4.26 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.



Εικόνα 4.27 Υπέρθυση δομών σε όψη stereo από την κυρίαρχη ομάδα.

Τα αποτελέσματα των dPCAs, με και χωρίς τις χ_1 γωνίες, διαφέρουν από αυτά της CPCA κυρίως ως προς τον αριθμό των διακριτών ομάδων που εντοπίστηκαν από κάθε μέθοδο. Για λόγους απλότητας, στα αποτελέσματα της ομαδοποίησης δεν παρουσιάζονται οι ομάδες που εντοπίστηκαν από τις μεθόδους με ποσοστό πολύ μικρότερο της μονάδας, γεγονός που φαίνεται από τους μικρούς πληθυσμούς στα διαγράμματα ελεύθερης ενέργειας. Παρά τις διαφορές στην ομαδοποίηση των διαμορφώσεων, τα αποτελέσματα των διαφορετικών PCA φαίνεται να συγκλίνουν, ιδιαίτερα για τις κυρίαρχες ομάδες κάθε ανάλυσης, αφού οι αντιπροσωπευτικές δομές που απομονώθηκαν από κάθε κύρια ομάδα μοιάζουν σε κάποιο βαθμό μεταξύ τους, **Πίνακες 4.3**. Στις περιπτώσεις όπου είχαμε διακριτές περιοχές με μεγάλη έκταση και χαμηλές τιμές RMSD στην **Εικόνα 4.14**, τα αποτελέσματα συμφωνούν με ορισμένες ομάδες που απομονώθηκαν από τις PCA, ενώ για το μεγαλύτερο κομμάτι της προσομοίωσης όπου το πεπτιδίο δεν παρουσιάζει κάποια σταθερότητα τα αποτελέσματα προφανώς διαφέρουν. Ενδιαφέρον παρουσιάζει η Ομάδα 3 της dPCA χωρίς τις γωνίες χ_1 και η αντίστοιχη Ομάδα 2 της dPCA με τις γωνίες χ_1 , οι οποίες δεν εντοπίστηκαν στον πίνακα RMSD, ούτε από την CPCA. Στις αντιπροσωπευτικές δομές των ομάδων αυτών φαίνεται να έχει σχηματιστεί στο μεγαλύτερο τμήμα του πεπτιδίου μια δομή β-φουρκέτας, η οποία όμως είναι μετατοπισμένη κατά ένα κατάλοιπο προς το C-τελικό άκρο.

ILDN

	RMSD_1	CPCA_1	DPCA_1	DPCA_CHI1_5
RMSD_1	-	2.21	2.06	2.09
CPCA_1	2.21	-	0.59	0.49
DPCA_1	2.06	0.59	-	0.40
DPCA_CHI1_5	2.09	0.49	0.40	-

	RMSD_2	CPCA_2	DPCA_3	DPCA_CHI1_4
RMSD_2	-	0.48	2.87	2.87
CPCA_2	0.48	-	2.76	2.76
DPCA_3	2.87	2.76	-	0.00
DPCA_CHI1_4	2.87	2.76	0.00	-

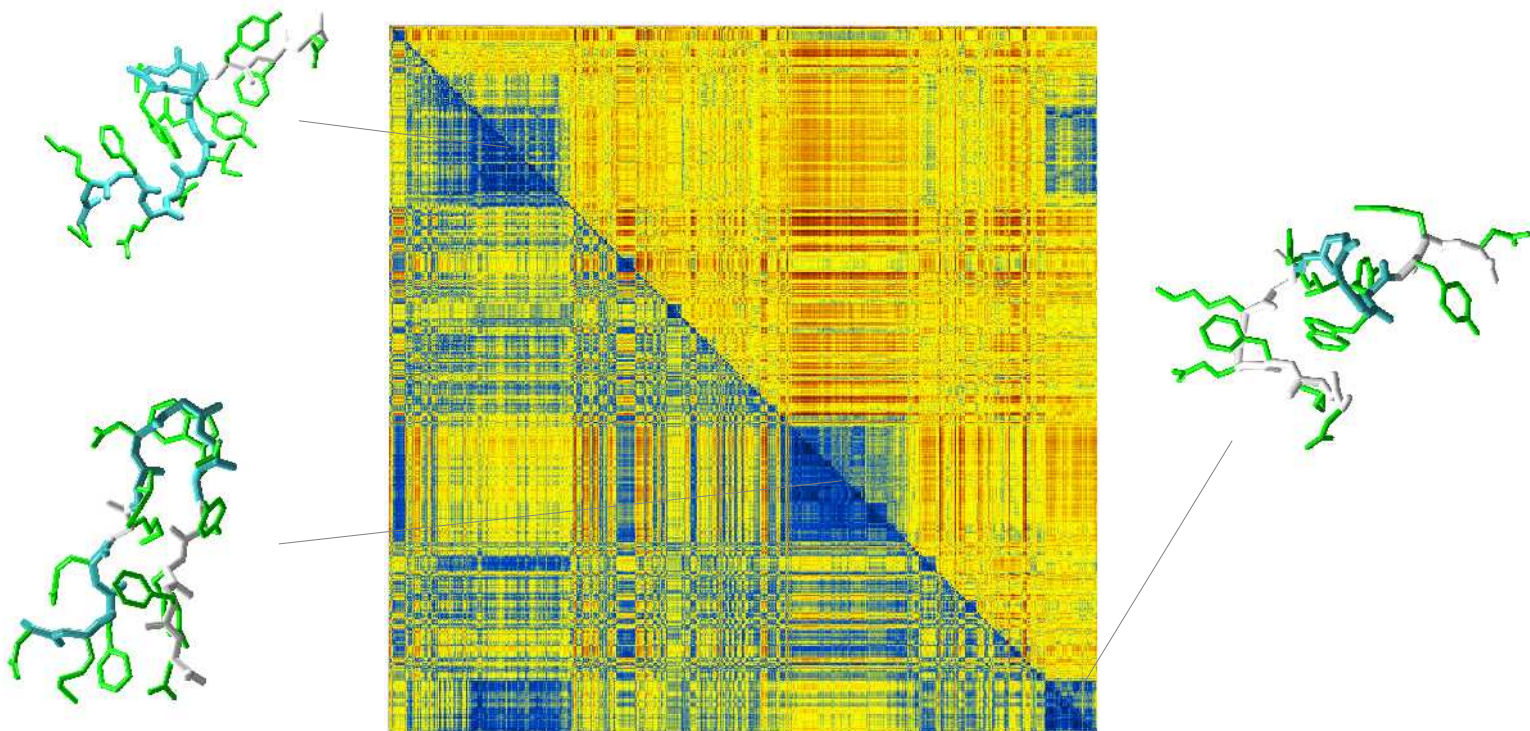
	RMSD_4	CPCA_5	DPCA_6	DPCA_CHI1_1
RMSD_4	-	3.62	3.75	3.96
CPCA_5	3.62	-	1.69	1.45
DPCA_6	3.75	1.69	-	1.67
DPCA_CHI1_1	3.96	1.45	1.67	-

Πίνακες 4.3 RMSD μεταξύ των αντιπροσωπευτικών δομών που απομονώθηκαν για κάθε σημαντική ομάδα.

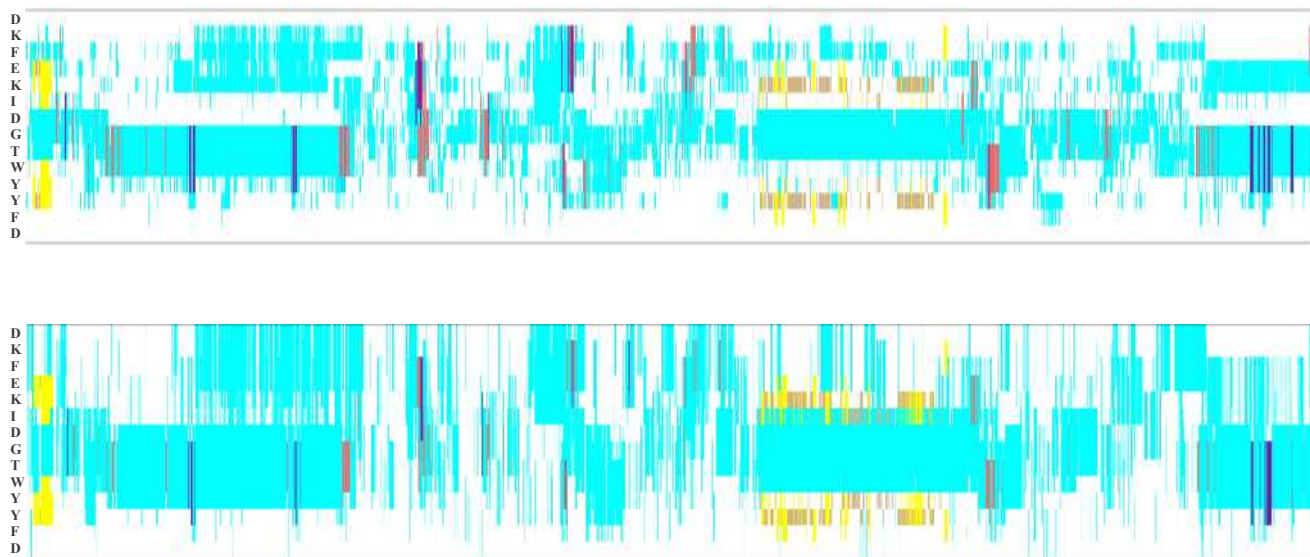
4.6 ILDN-TIP4

Η ανάλυση του τροχιακού ξεκίνησε υπολογίζοντας τον πίνακα RMSD, **Εικόνα 4.28**. Αρχικά υπολογίστηκε το RMSD χρησιμοποιώντας μόνο τα άτομα Ca, ενώ σε διαφορετικό υπολογισμό χρησιμοποιώντας όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Έπειτα έγινε ο καθορισμός των στοιχείων δευτεροταγούς δομής του τροχιακού εφαρμόζοντας τους αλγόριθμους DSSP και STRIDE, **Εικόνα 4.29**.

Τα αποτελέσματα από τον πίνακα RMSD δείχνουν ότι οι μέσες δομές των σταθερότερων ομάδων απέχουν σε μεγάλο βαθμό από τη φυσική δομή. Όπως φαίνεται στην **Εικόνα 4.28**, σε καμία από τις αντιπροσωπευτικές δομές δεν έχει σχηματιστεί το χαρακτηριστικό στοιχείο δευτεροταγούς δομής της φυσικής δομής, δηλαδή η β-φουρκέτα. Τα αποτελέσματα από τις μεθόδους DSSP και STRIDE παρουσιάζουν ελάχιστες διαφορές. Το πεπτίδιο φαίνεται για το μεγαλύτερο διάστημα του τροχιακού να υιοθετεί εκτεταμένες δομές που αποτελούνται κυρίως από στροφές και μεγάλες αδιευκρίνιστες περιοχές, **Εικόνα 4.29**. Περιστασιακά εμφανίζονται έλικες και άλλες διαμορφώσεις, οι οποίες βέβαια δεν είναι στατιστικά σημαντικές.



Εικόνα 4.28 Πίνακας RMSD. Στο αριστερό κάτω μέρος χρησιμοποιήθηκαν τα άτομα Cα, ενώ στο δεξιό πάνω μέρος όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Ο υπολογισμός του RMSD έγινε με βήμα 1000. Στην εικόνα παρουσιάζονται και οι μέσες δομές για κάθε διακριτή ομάδα.



Εικόνα 4.29 Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην πρώτη εικόνα έγινε με τη χρήση του αλγορίθμου DSSP και η δεύτερη με τον STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου DKFEKIDGTWYYFD. Οι υπολογισμοί έγιναν με βήμα 500 διαμορφώσεις.

Στη συνέχεια, εφαρμόσαμε Cartesian Principal Component Analysis (CPCA) για τον εντοπισμό των κύριων ομάδων χρησιμοποιώντας τις τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν παράλληλα και τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.30**. Πριν την έναρξη των υπολογισμών προηγήθηκε το ταίριασμα (fitting) του τροχιακού με βάση την κύρια αλυσίδα των καταλοίπων 3-12. Οι κύριες ομάδες (clusters) που εντοπίστηκαν ήταν εξής:

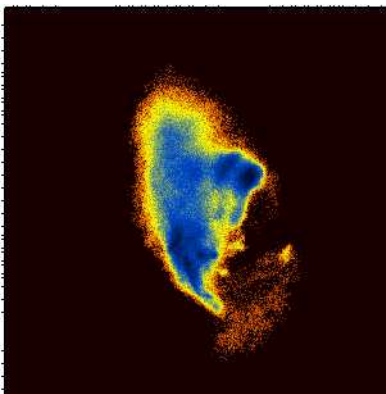
Cluster 1 : 341588 από τα 3256125 frames

Cluster 2 : 592104 από τα 3256125 frames

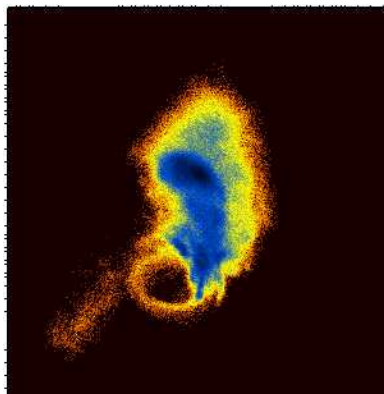
Cluster 3 : 15755 από τα 3256125 frames

Cluster 4 : 3056 από τα 3256125 frames

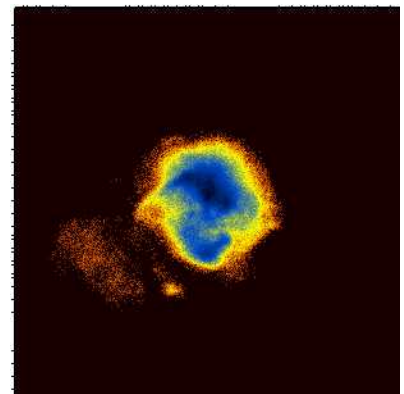
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 10% του συνολικού τροχιακού, η δεύτερη στο 18%, ενώ το ποσοστό της τρίτης και της τέταρτης είναι πολύ μικρότερο της μονάδας. Η κυρίαρχη ομάδα είναι η δεύτερη ομάδα, αφού καταλαμβάνει το μεγαλύτερο κομμάτι του τροχιακού. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στην **Εικόνα 4.31**.



PC_01_02



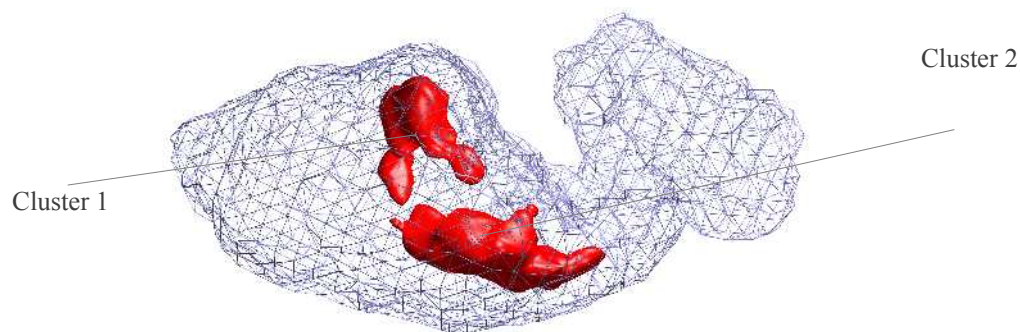
PC_01_03



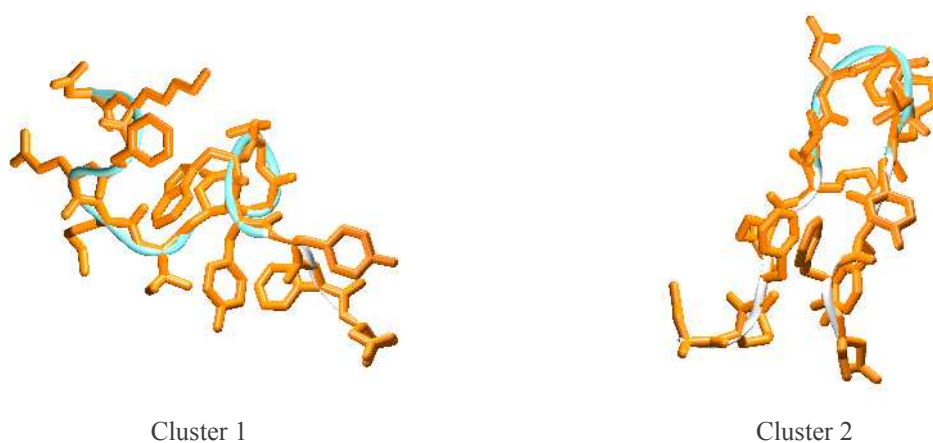
PC_02_03

Εικόνα 4.30 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

Στη συνέχεια παρουσιάζονται οι αντιπροσωπευτικές δομές των δύο μεγαλύτερων ομάδων, **Εικόνα 4.32**. Από τα διαγράμματα συμπεραίνουμε ότι και οι δύο ομάδες παρουσιάζουν μια σχετική ετερογένεια. Μία υπέρθεση των δομών των ομάδων θα περιείχε πολύ θόρυβο και δε θα μπορούσε να διακριθεί η κινητικότητα των ατόμων.



Εικόνα 4.31 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC.



Εικόνα 4.32 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την CPCA.

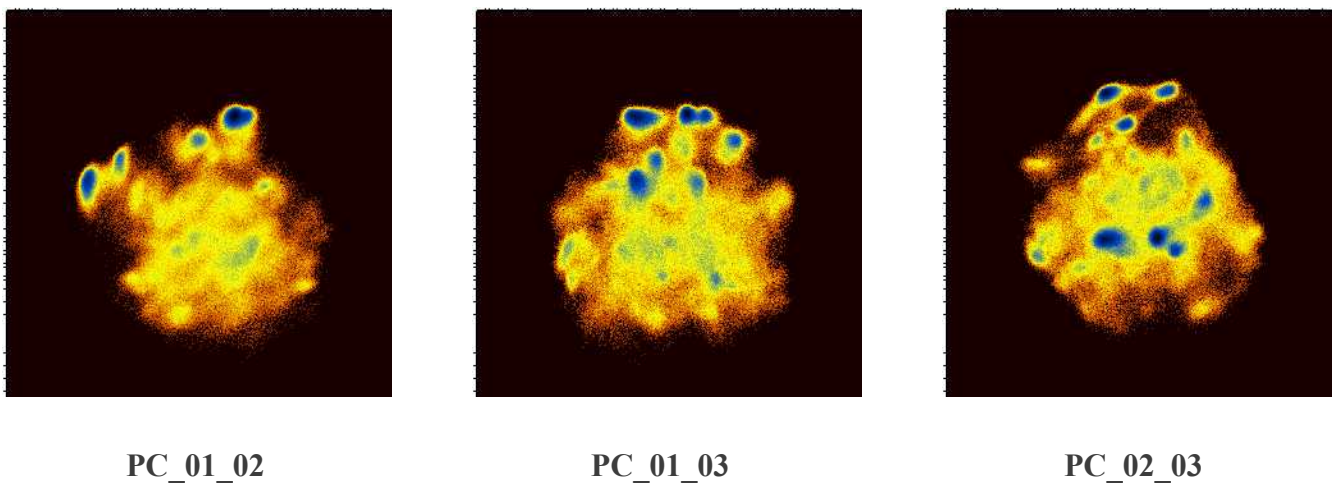
Οι αντιπροσωπευτικές δομές για τις δύο πρώτες ομάδες παρουσιάζουν ομοιότητα με τις δομές που απομονώθηκαν από τους υπολογισμούς με βάση το RMSD, **Πίνακας 4.4**.

Στη συνέχεια παρουσιάζονται τα αποτελέσματα από dihedral Principal Component Analysis (dPCA), με και χωρίς τις γωνίες χ_1 . Για την ανάλυση χρησιμοποιήθηκαν όλα τα άτομα των καταλοίπων 3-12 και για τους υπολογισμούς οι τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν και τα διαγράμματα ελεύθερης ενέργειας. Οι κύριες ομάδες που εντοπίστηκαν στους υπολογισμούς χωρίς τις γωνίες χ_1 ήταν οι εξής:

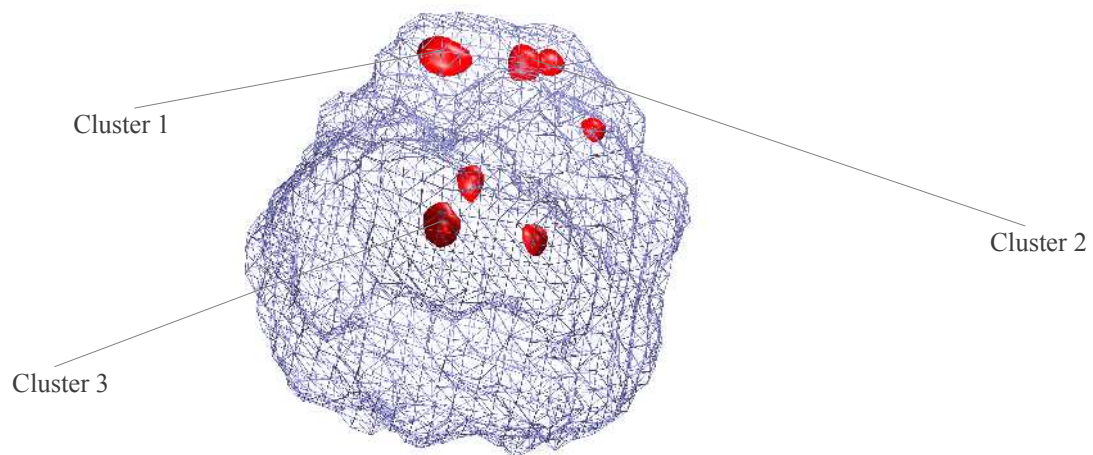
- Cluster 1 : 215096 από τα 3256125 frames
- Cluster 2 : 243146 από τα 3256125 frames
- Cluster 3 : 146208 από τα 3256125 frames
- Cluster 4 : 54347 από τα 3256125 frames

Cluster 5 : 41600 από τα 3256125 frames
Cluster 6 : 35418 από τα 3256125 frames
Cluster 7 : 2812 από τα 3256125 frames

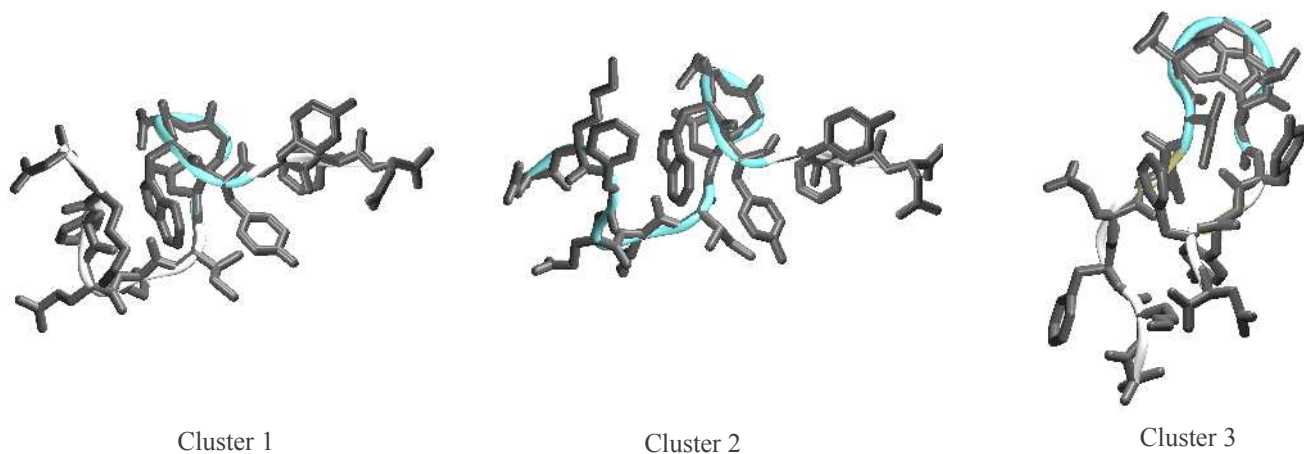
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 6% του συνολικού τροχιακού, η δεύτερη στο 7% και η τρίτη στο 4%. Οι υπόλοιπες ομάδες κυμαίνονται σε ποσοστά γύρω στο 1%. Κυρίαρχη ομάδα είναι η δεύτερη ομάδα. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στην **Εικόνα 4.33** και **Εικόνα 4.34** και οι αντιπροσωπευτικές δομές των τριών μεγαλύτερων ομάδων παρουσιάζονται στην **Εικόνα 4.35**.



Εικόνα 4.33 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



Εικόνα 4.34 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC.



Εικόνα 4.35 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.

Τα αποτελέσματα της ανάλυσης χρησιμοποιώντας τις γωνίες χ_1 παρουσιάζονται στη συνέχεια:

Cluster 1 : 286835 από τα 3256125 frames

Cluster 2 : 239903 από τα 3256125 frames

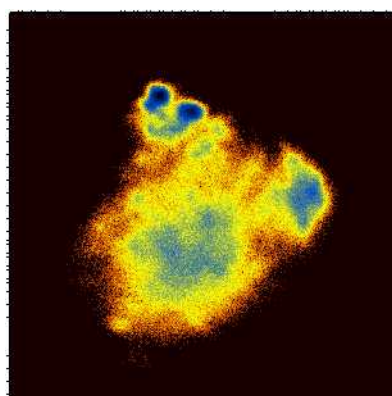
Cluster 3 : 307831 από τα 3256125 frames

Cluster 4 : 8205 από τα 3256125 frames

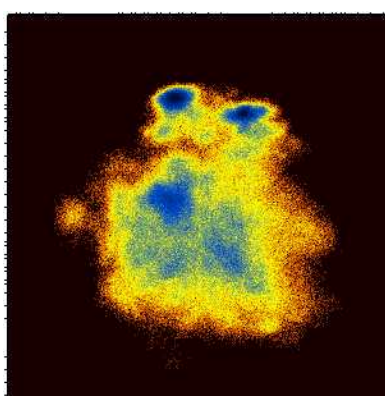
Cluster 5 : 19077 από τα 3256125 frames

Cluster 6 : 23718 από τα 3256125 frames

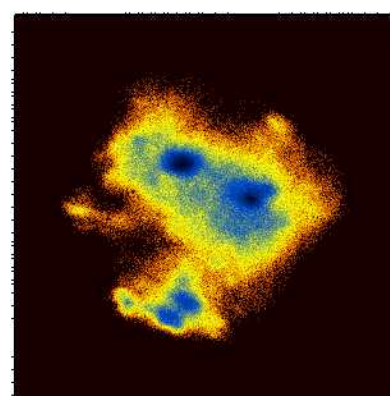
Η πρώτη ομάδα αντιστοιχεί στο 8% του συνολικού τροχιακού, η δεύτερη στο 7% και η τρίτη στο 9%. Κυρίαρχη ομάδα είναι η τρίτη ομάδα. Η κατανομή των ομάδων πάνω στο διάγραμμα της ελεύθερης ενέργειας φαίνεται στην **Εικόνα 4.36** και **Εικόνα 4.37** και οι αντιπροσωπευτικές δομές των τριών μεγαλύτερων ομάδων παρουσιάζονται στην **Εικόνα 4.38**.



PC_01_02

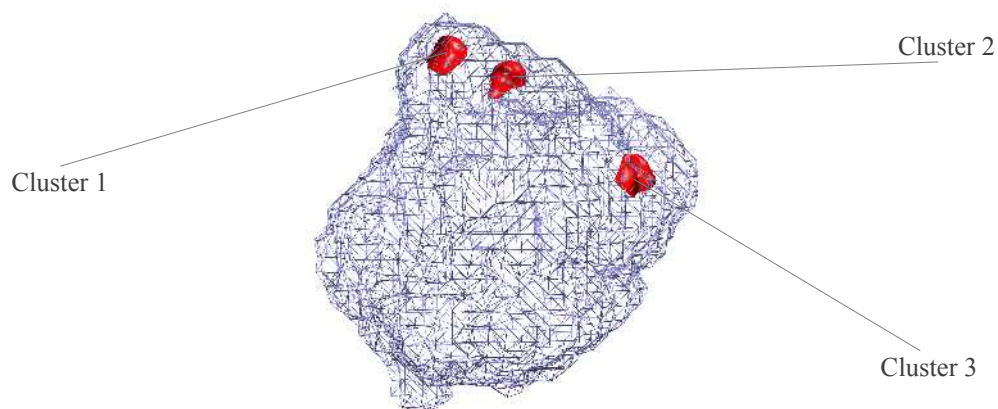


PC_01_03

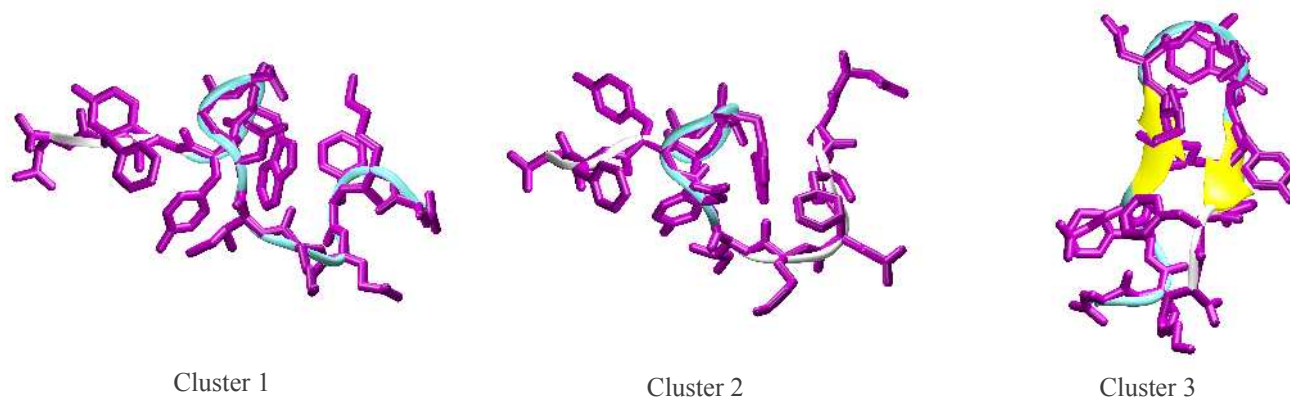


PC_02_03

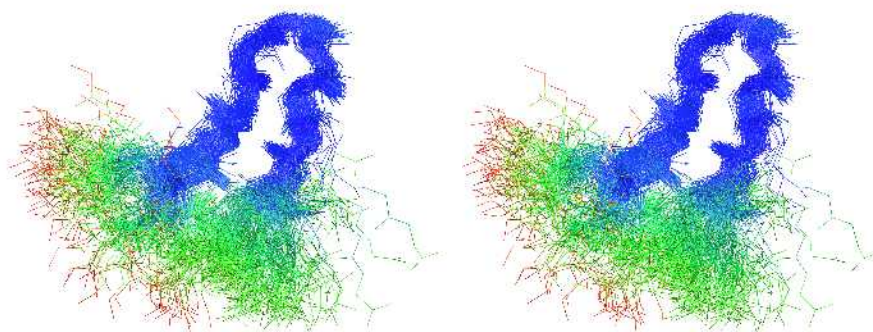
Εικόνα 4.36 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



Εικόνα 4.37 Γραφική παράσταση της κατανομής των τριών μεγαλύτερων PC.



Εικόνα 4.38 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.



Εικόνα 4.39 Υπέρθυση δομών σε όψη stereo από την κύρια αλυσίδα της κυρίαρχης ομάδας.

Σύμφωνα, με τα αποτελέσματα των αναλύσεων στο τροχιακό κυριαρχούν εκτεταμένες δομές στις οποίες δε φαίνεται να έχει σχηματιστεί κάποιο κύριο χαρακτηριστικό της φυσικής δομής. Παράλληλα εμφανίζονται διαμορφώσεις στις οποίες έχει σχηματιστεί μια στροφή στα σωστά κατάλοιπα, η οποία

όμως διαφέρει από τη στροφή τύπου I της φυσικής δομής. Επιπλέον, τα υπόλοιπα κατάλοιπα του πεπτιδίου για τις διαμορφώσεις αυτές δε φαίνεται να έχουν στοιχηθεί σωστά. Ενδιαφέρον παρουσιάζει η αντιπροσωπευτική δομή της κυρίαρχης ομάδας στην οποία φαίνεται να έχει σχηματιστεί σωστά η στροφή και το κομμάτι της β-φουρκέτας δίπλα στη στροφή, **Εικόνα 4.38** και **Εικόνα 4.39**. Σύμφωνα με τα αποτελέσματα, η ομαδοποίηση των διαμορφώσεων διαφέρει από μέθοδο σε μέθοδο, ενώ τα αποτελέσματα των διαφορετικών PCA δε παρουσιάζουν μεγάλες ομοιότητες όσον αφορά τις κυρίαρχες κινήσεις του πεπτιδίου και τις αντιπροσωπευτικές δομές, **Πίνακας 4.4**. Για λόγους απλότητας, στα αποτελέσματα της ομαδοποίησης δεν παρουσιάζονται οι ομάδες που εντοπίστηκαν από τις μεθόδους με ποσοστό πολύ μικρότερο της μονάδας.

ILDN-TIP4

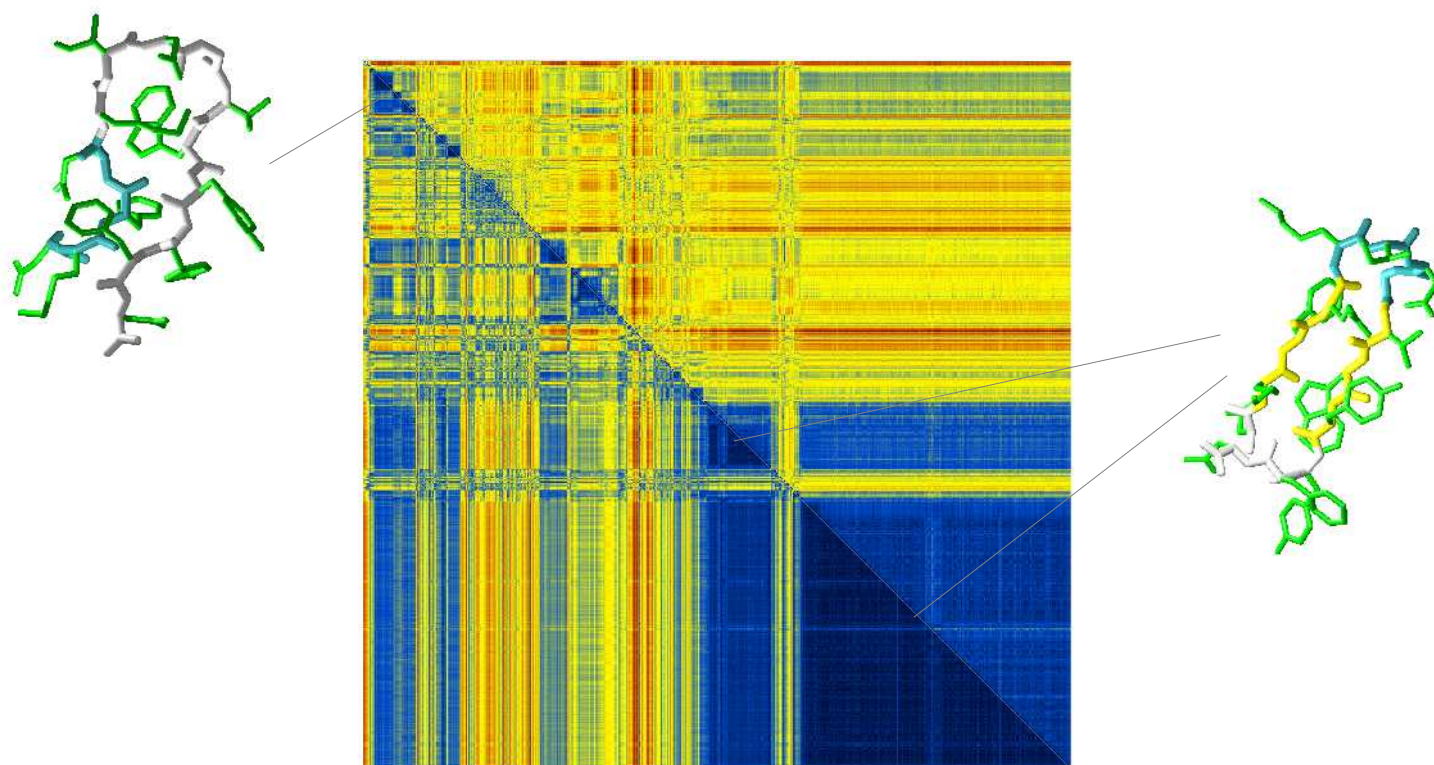
	RMSD_1	CPCA_1	DPCA_1	DPCA_CHI1_1
RMSD_1	-	2.24	3.65	0.77
CPCA_1	2.24	-	2.67	1.89
DPCA_1	3.65	2.67	-	3.67
DPCA_CHI1_1	0.77	1.89	3.67	-

	RMSD_2	CPCA_2	DPCA_2	DPCA_CHI1_2
RMSD_2	-	1.54	1.92	1.57
CPCA_2	1.54	-	2.02	1.85
DPCA_2	1.92	2.02	-	1.61
DPCA_CHI1_2	1.57	1.85	1.61	-

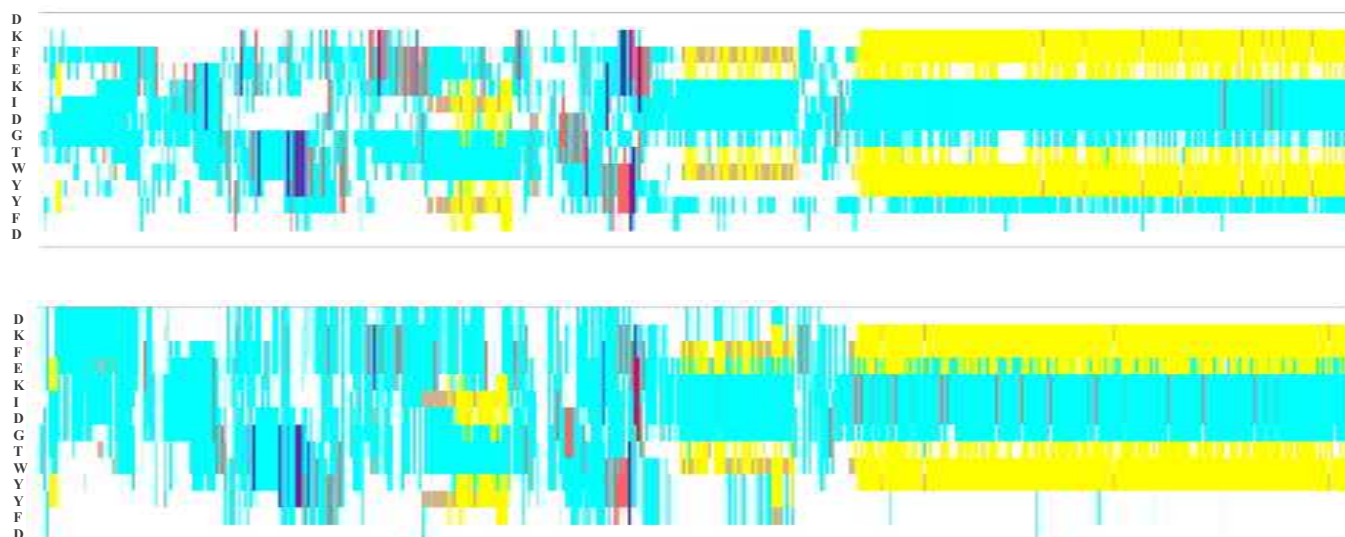
Πίνακας 4.4 RMSD μεταξύ των αντιπροσωπευτικών δομών που απομονώθηκαν για κάθε σημαντική ομάδα.

4.7 ILDN-340K

Αρχικά υπολογίσαμε τον πίνακα RMSD, **Εικόνα 4.40**. Στο κάτω αριστερό μέρος του πίνακα χρησιμοποιήθηκαν τα άτομα Ca όλων των καταλοίπων του πεπτιδίου, ενώ για το δεξί πάνω μέρος χρησιμοποιήθηκαν όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Για τον καθορισμό των στοιχείων δευτεροταγούς δομής του τροχιακού εφαρμόσαμε τους αλγόριθμους DSSP και STRIDE, **Εικόνα 4.41**.



Εικόνα 4.40 Πίνακας RMSD. Στο αριστερό κάτω μέρος χρησιμοποιήθηκαν τα άτομα Ca, ενώ στο δεξί πάνω μέρος όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Ο υπολογισμός του RMSD έγινε με βήμα 1000. Στην εικόνα παρουσιάζονται και οι μέσες δομές για κάθε διακριτή ομάδα.



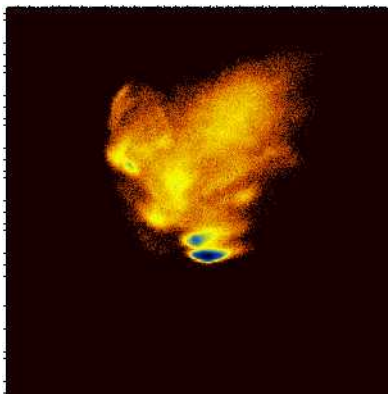
Εικόνα 4.41 Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην πρώτη εικόνα έγινε με τη χρήση του αλγορίθμου DSSP και η δεύτερη με τον STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου DKFEKIDGTWYYFD. Οι υπολογισμοί έγιναν με βήμα 500 διαμορφώσεις.

Η πρώτη ομάδα φαίνεται να αποτελείται από εκτεταμένες δομές χωρίς κάποια συγκεκριμένη διαμόρφωση. Αντιθέτως, οι πληθυσμοί των επόμενων δύο ομάδων παρουσιάζουν σταθερότητα για το μεγαλύτερο τμήμα του τροχιακού. Στην αντιπροσωπευτική δομή τους φαίνεται να έχει σχηματιστεί μία β-φουρκέτα, η οποία διαφέρει από αυτή της φυσικής δομής λόγω της μετατόπισης της προς το Ν-τελικό άκρο, **Εικόνα 4.40**. Τα αποτελέσματα από τις μεθόδους DSSP και STRIDE επιβεβαιώνουν την αρχική παρατήρηση ενώ συμφωνούν μεταξύ τους σε μεγάλο βαθμό, **Εικόνα 4.41**. Κύρια διαφορά τους είναι το πλήθος των δομών από έλικες που εντοπίζεται στα μέσα του τροχιακού.

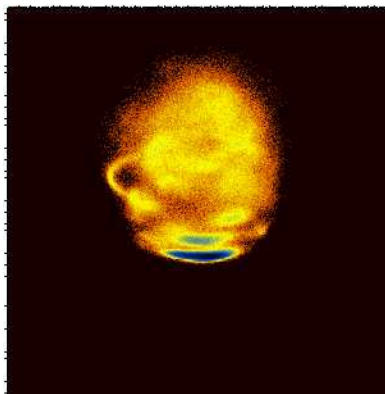
Στη συνέχεια, εφαρμόσαμε Cartesian Principal Component Analysis (CPCA) για τον εντοπισμό των κύριων ομάδων χρησιμοποιώντας τις τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν παράλληλα και τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.42**. Πριν την έναρξη των υπολογισμών προηγήθηκε το ταίριασμα (fitting) του τροχιακού με βάση την κύρια αλυσίδα των καταλοίπων 3-12. Οι κύριες ομάδες (clusters) που εντοπίστηκαν ήταν εξής:

Cluster 1 : 709482 από τα 1756850 frames

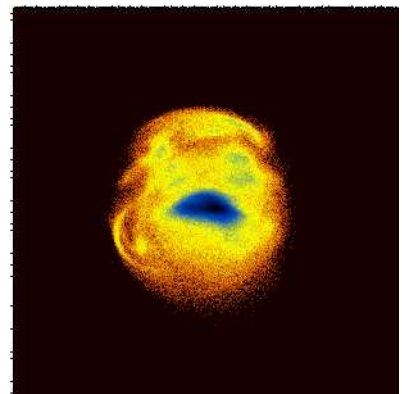
Cluster 2 : 4251 από τα 1756850 frames



PC_01_02



PC_01_03



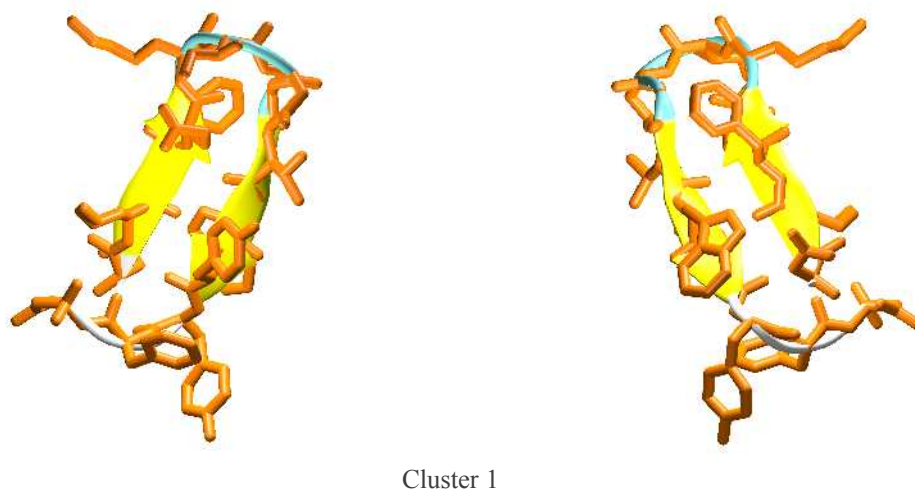
PC_02_03

Εικόνα 4.42 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

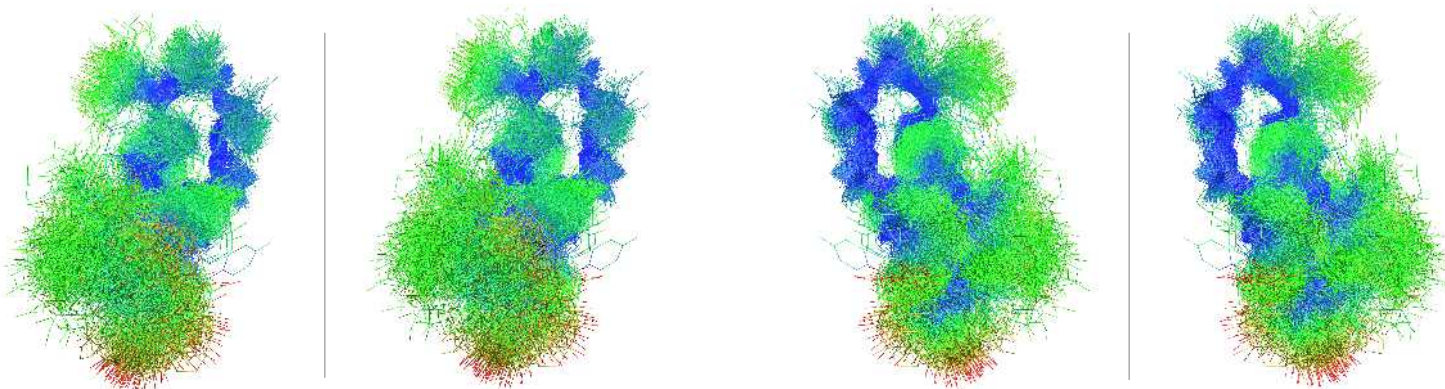
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 40% του συνολικού τροχιακού, ενώ η δεύτερη αντιστοιχεί σε ποσοστό μικρότερο της μονάδας. Προφανώς κυρίαρχη ομάδα είναι η πρώτη, αφού καταλαμβάνει το μεγαλύτερο τμήμα του τροχιακού. Η κατανομή της διαγράφεται ξεκάθαρα στα διαγράμματα ελεύθερης ενέργειας. Στη συνέχεια απομονώθηκε η αντιπροσωπευτική δομή της κυρίαρχης ομάδας, **Εικόνα 4.43**, και προετοιμάστηκαν εικόνες από την υπέρθεση των δομών

της, **Εικόνα 4.44**.

Όπως φαίνεται στις εικόνες, η αντιπροσωπευτική δομή της κύριας ομάδας ταιριάζει με αυτήν που απομονώθηκε από τον πίνακα RMSD, **Πίνακες 4.5**. Η υψηλή κινητικότητα των άκρων φαίνεται στην **Εικόνα 4.44**.



Εικόνα 4.43 Αντιπροσωπευτικές δομές από την κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την CPCA.



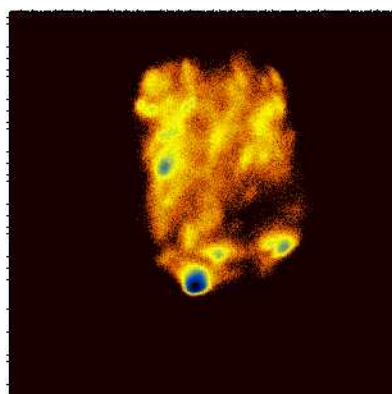
Εικόνα 4.44 Υπέρθυση δομών της κυρίαρχης ομάδας σε όψη stereo.

Έπειτα εφαρμόσαμε dihedral Principal Component Analysis (dPCA), με και χωρίς τις γωνίες χ_1 , στο τροχιακό. Για τους υπολογισμούς χρησιμοποιήθηκαν οι τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν και τα διαγράμματα ελεύθερης ενέργειας. Οι κύριες ομάδες που εντοπίστηκαν στους υπολογισμούς χωρίς τις γωνίες χ_1 ήταν οι εξής:

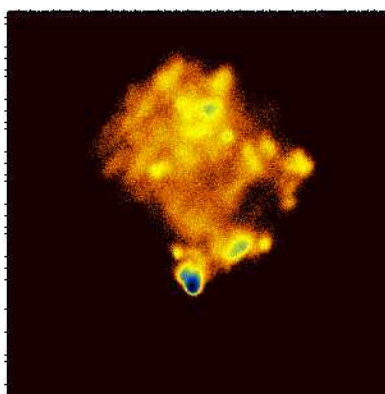
Cluster 1 : 603553 από τα 1756850 frames

Cluster 2 : 22826 από τα 1756850 frames

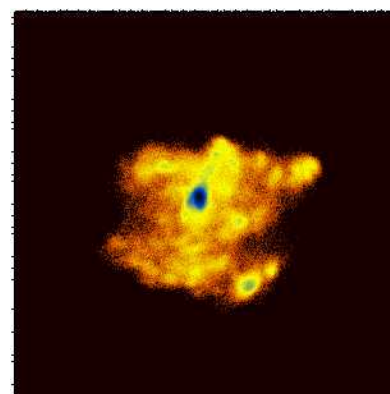
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 34% του συνολικού τροχιακού και η δεύτερη στο 1%. Κυρίαρχη ομάδα είναι η πρώτη, της οποίας η κατανομή φαίνεται εύκολα από τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.45**. Στη συνέχεια απομονώθηκε η αντιπροσωπευτική δομή της κυρίαρχης ομάδας, **Εικόνα 4.46**, και προετοιμάστηκαν εικόνες από την υπέρθεση των δομών της, **Εικόνα 4.47**.



PC_01_02

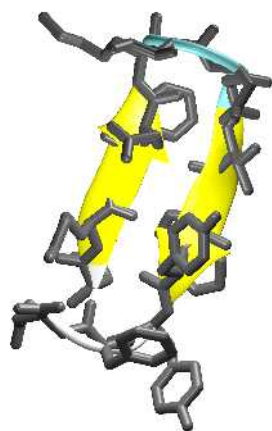


PC_01_03

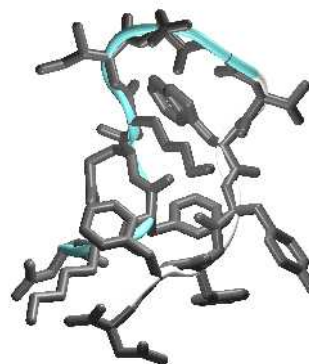


PC_02_03

Εικόνα 4.45 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

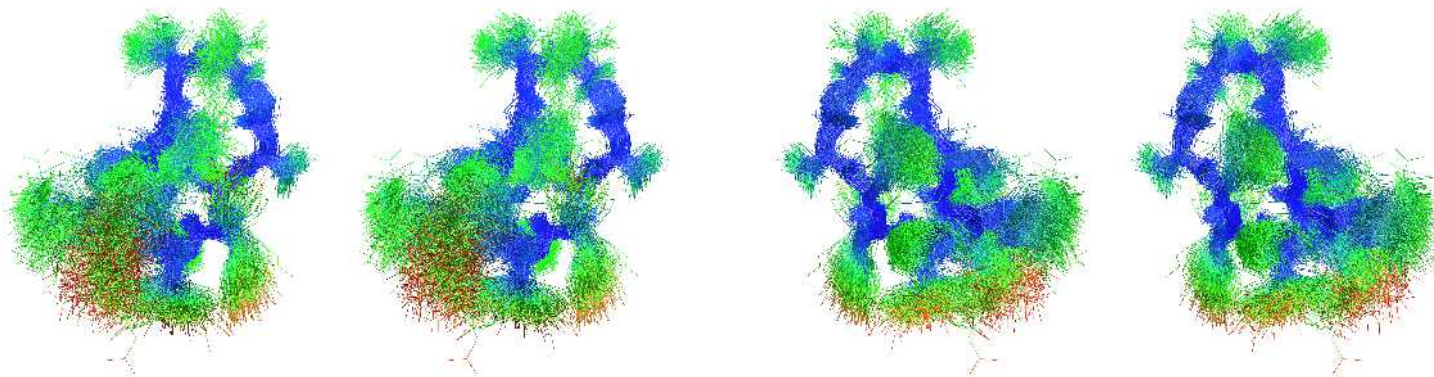


Cluster 1



Cluster 2

Εικόνα 4.46 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.



Εικόνα 4.47 Υπέρθωση δομών της κυρίαρχης ομάδας σε όψη stereo.

Τα αποτελέσματα της ανάλυσης χρησιμοποιώντας τις γωνίες χ_1 παρουσιάζονται στη συνέχεια:

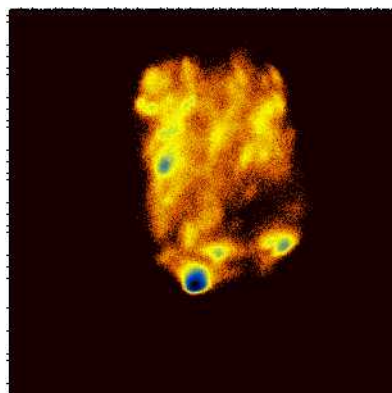
Cluster 1 : 514977 από τα 1756850 frames

Cluster 2 : 18970 από τα 1756850 frames

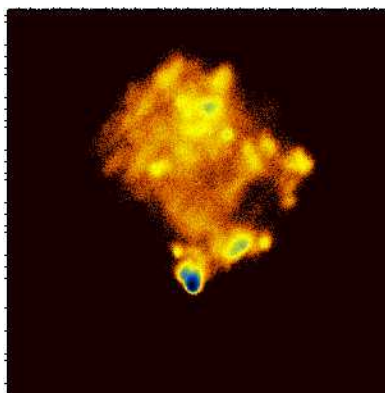
Cluster 3 : 2457 από τα 1756850 frames

Cluster 4 : 880 από τα 1756850 frames

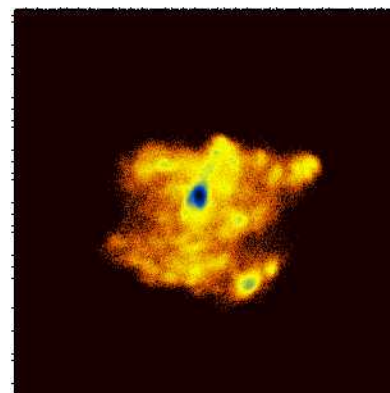
Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 29% του συνολικού τροχιακού, η δεύτερη στο 1% και οι υπόλοιπες σε ποσοστό μικρότερο της μονάδας. Κυρίαρχη ομάδα είναι η πρώτη. Η κατανομή της διαγράφεται ξεκάθαρα στα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.48**. Στη συνέχεια απομονώθηκε η αντιπροσωπευτική δομή της κυρίαρχης ομάδας, **Εικόνα 4.49**, και προετοιμάστηκαν εικόνες από την υπέρθεση των δομών της, **Εικόνα 4.50**.



PC_01_02

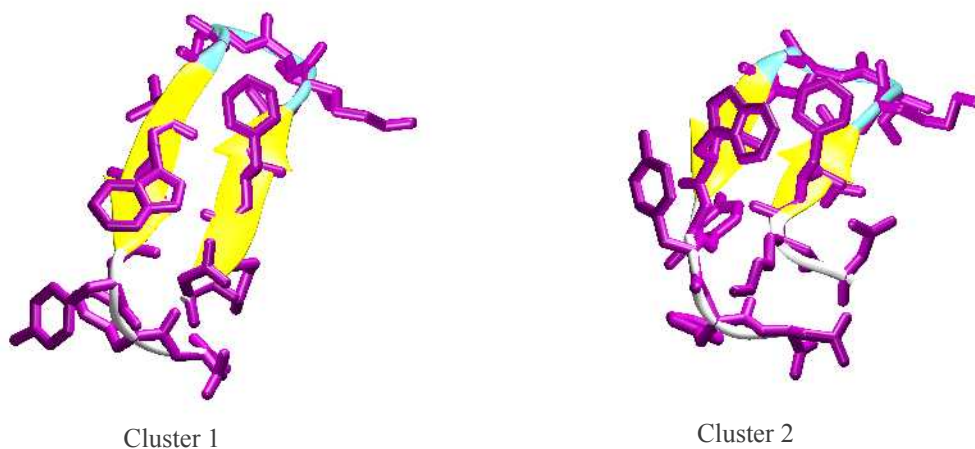


PC_01_03

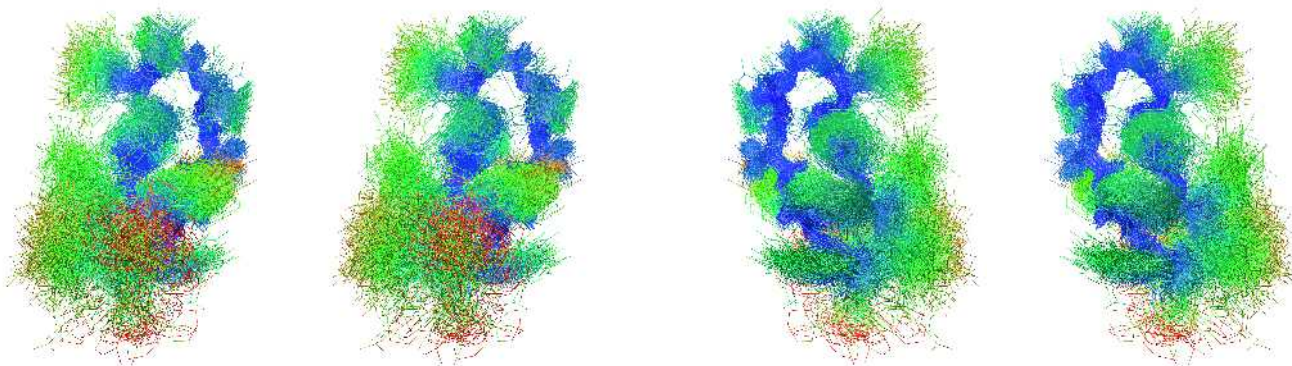


PC_02_03

Εικόνα 4.48 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.



Εικόνα 4.49 Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.



Εικόνα 4.50 Υπέρθωση δομών της κυρίαρχης ομάδας σε όψη stereo.

Οι κύριες κινήσεις του τροχιακού φαίνεται να αποτυπώνονται σε κάθε μέθοδο που χρησιμοποιήθηκε. Το πεπτίδιο για το μεγαλύτερο διάστημα της προσομοίωσης υιοθετεί τη χαρακτηριστική δομή που απομονώθηκε από κάθε μέθοδο, δηλαδή μία μετατοπισμένη, ως προς τη φυσική δομή, β-φουρκέτα. Οι dPCAs εντόπισαν και άλλους μικρότερους πληθυσμούς από αδιευκρίνιστες περιοχές, ενώ όταν χρησιμοποιήθηκαν οι γωνίες χ_1 , εντοπίστηκε ένας πληθυσμός με ανεστραμμένες τις πλευρικές αλυσίδες της τρυπτοφάνης και της τυροσίνης και λιγότερα κατάλοιπα να συμμετέχουν στο σχηματισμό της β-φουρκέτας. Παρά τις μικρές διαφορές, τα αποτελέσματα από τις αναλύσεις συμφωνούν σε μεγάλο βαθμό, **Πίνακες 4.5**.

ILDN-340K

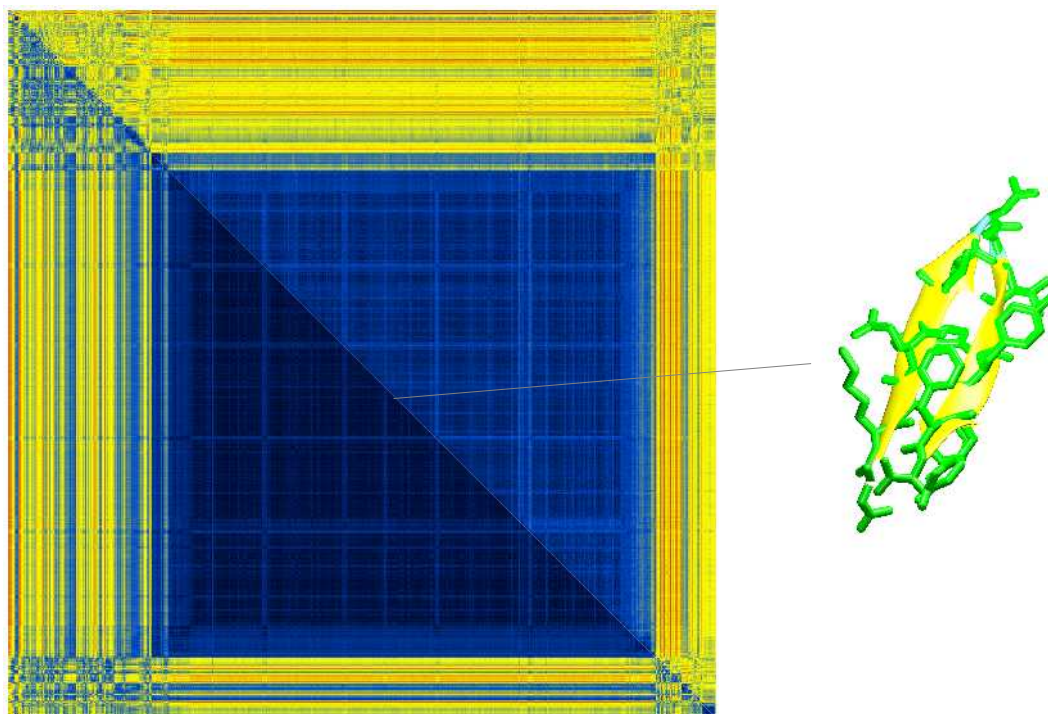
	RMSD_1	DPCA_2	DPCA_CHI1_2
RMSD_1	-	1.26	2.92
DPCA_2	1.26	-	2.90
DPCA_CHI1_2	2.92	2.90	-

	RMSD_2	CPCA_1	DPCA_1	DPCA_CHI1_1
RMSD_2	-	0.77	0.64	0.64
CPCA_1	0.77	-	0.40	0.40
DPCA_1	0.64	0.40	-	0.00
DPCA_CHI1_1	0.64	0.40	0.00	-

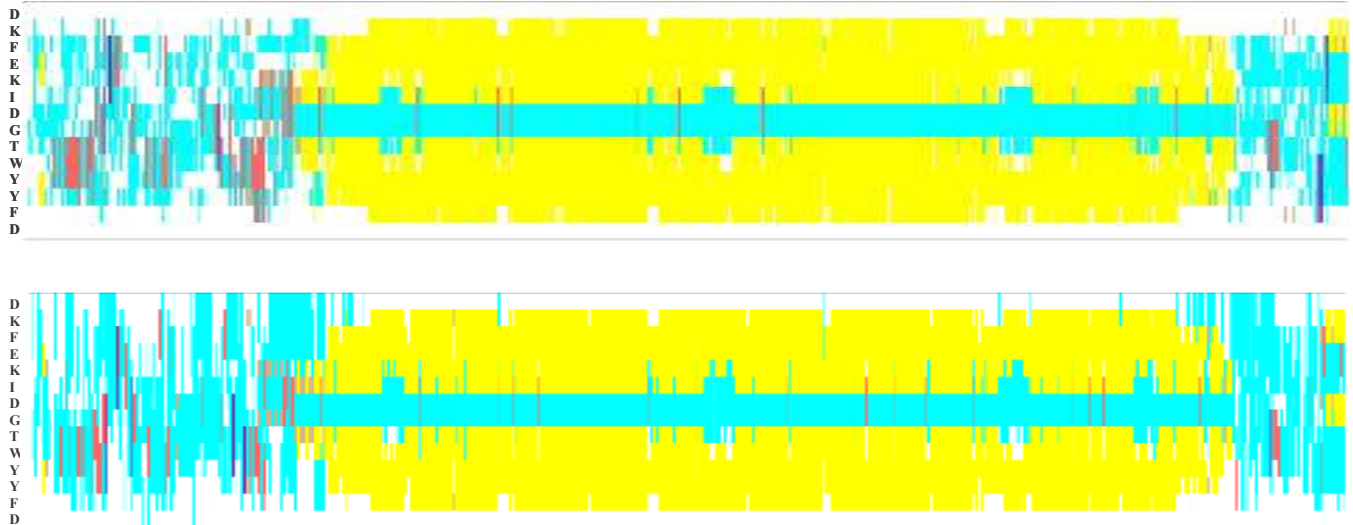
Πίνακες 4.5 RMSD μεταξύ των αντιπροσωπευτικών δομών που απομονώθηκαν για κάθε σημαντική ομάδα.

4.8 ILDN-360K

Αρχικά υπολογίσαμε τον πίνακα RMSD, **Εικόνα 4.51**. Στο κάτω αριστερό μέρος του πίνακα χρησιμοποιήθηκαν τα άτομα Ca όλων των καταλοίπων του πεπτιδίου, ενώ για το δεξί πάνω μέρος χρησιμοποιήθηκαν όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Για τον καθορισμό των στοιχείων δευτεροταγούς δομής του τροχιακού εφαρμόσαμε τους αλγόριθμους DSSP και STRIDE, **Εικόνα 4.52**.



Εικόνα 4.51 Πίνακας RMSD. Στο αριστερό κάτω μέρος χρησιμοποιήθηκαν τα άτομα Ca, ενώ στο δεξί πάνω μέρος όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα. Ο υπολογισμός του RMSD έγινε με βήμα 1000. Στην εικόνα παρουσιάζονται και οι μέσες δομές για κάθε διακριτή ομάδα.



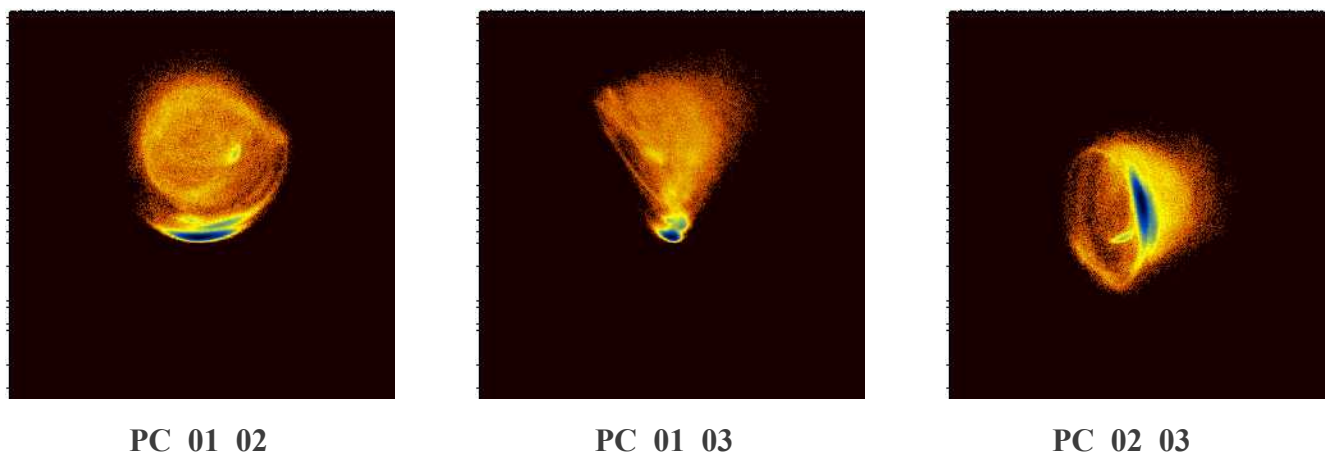
Εικόνα 4.52 Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην πρώτη εικόνα έγινε με τη χρήση του αλγορίθμου DSSP και η δεύτερη με τον STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου DKFEKIDGTWYYFD. Οι υπολογισμοί έγιναν με βήμα 500 διαμορφώσεις.

Ο πίνακας RMSD δείχνει την ύπαρξη μιας μεγάλης ομάδας, η οποία καταλαμβάνει το μεγαλύτερο τμήμα του τροχιακού. Όπως φαίνεται στην **Εικόνα 4.51**, η αντιπροσωπευτική δομή της ομάδας αυτής έχει σχηματίσει σωστά την στροφή τύπου I και τα κατάλοιπα του πεπτιδίου έχουν στοιχηθεί σωστά ώστε να σχηματίζεται η β-φουρκέτα της φυσικής δομής. Οι παραπάνω παρατηρήσεις επιβεβαιώνονται από τον **Πίνακα 4.8**, καθώς επίσης και από τα αποτελέσματα της εφαρμογής των μεθόδων DSSP και STRIDE στο τροχιακό, **Εικόνα 4.52**.

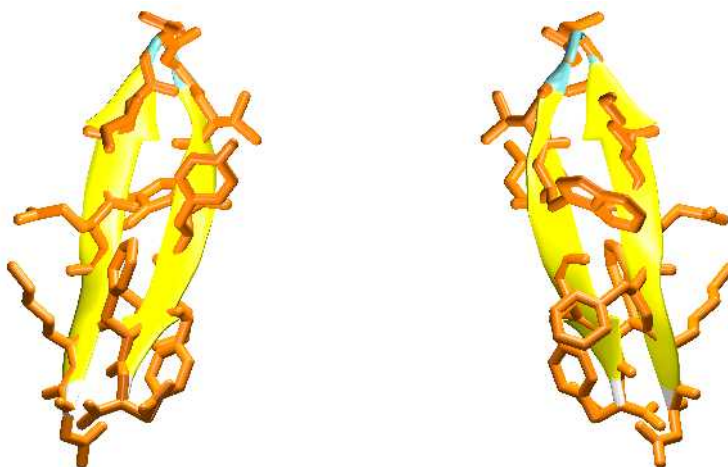
Στη συνέχεια, εφαρμόσαμε Cartesian Principal Component Analysis (CPCA) για τον εντοπισμό των κύριων ομάδων χρησιμοποιώντας τις τρεις πρώτες κύριες συνιστώσες για τις οποίες υπολογίστηκαν παράλληλα και τα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.53**. Πριν την έναρξη των υπολογισμών προηγήθηκε το ταίριασμα (fitting) του τροχιακού με βάση την κύρια αλυσίδα των καταλοίπων 3-12. Οι κύριες ομάδες (clusters) που εντοπίστηκαν ήταν εξής:

- Cluster 1 : 1107688 από τα 1787325 frames
- Cluster 2 : 4437 από τα 1787325 frames
- Cluster 3 : 749 από τα 1787325 frames

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 62% του συνολικού τροχιακού, ενώ η δεύτερη αντιστοιχεί σε ποσοστό μικρότερο της μονάδας. Όπως ήταν αναμενόμενο εμφανίζεται μία κυρίαρχη ομάδα της οποίας η κατανομή διαγράφεται ξεκάθαρα στα διαγράμματα ελεύθερης ενέργειας και η αντιπροσωπευτική δομή παρουσιάζεται στην **Εικόνα 4.54**. Η κινητικότητα των ατόμων του πεπτιδίου φαίνεται στην υπέρθεση των δομών της κυρίαρχης ομάδας, **Εικόνα 4.55**.



Εικόνα 4.53 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

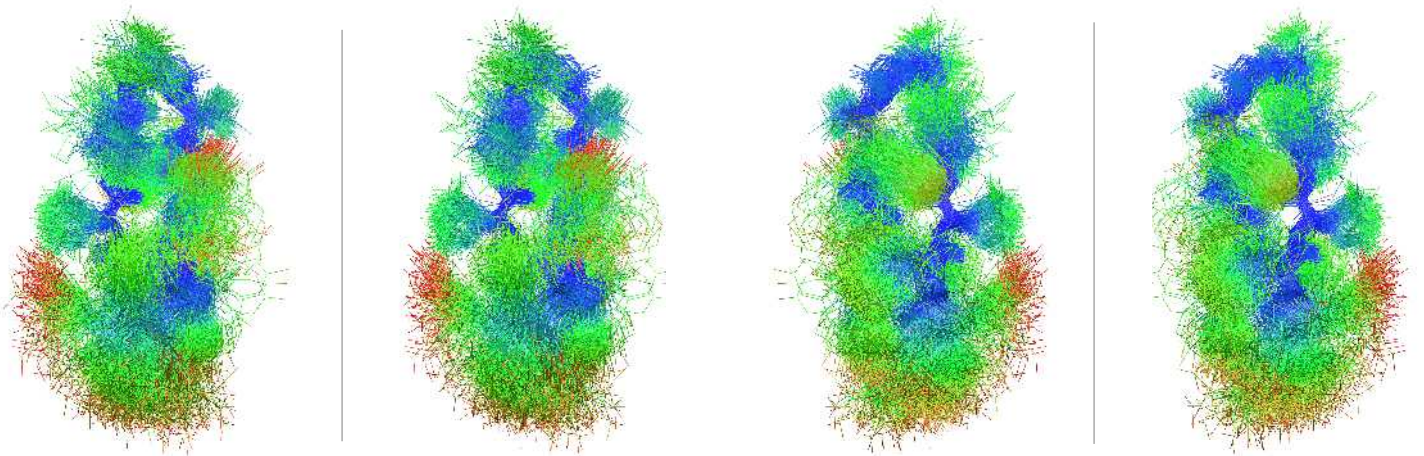


Cluster 1

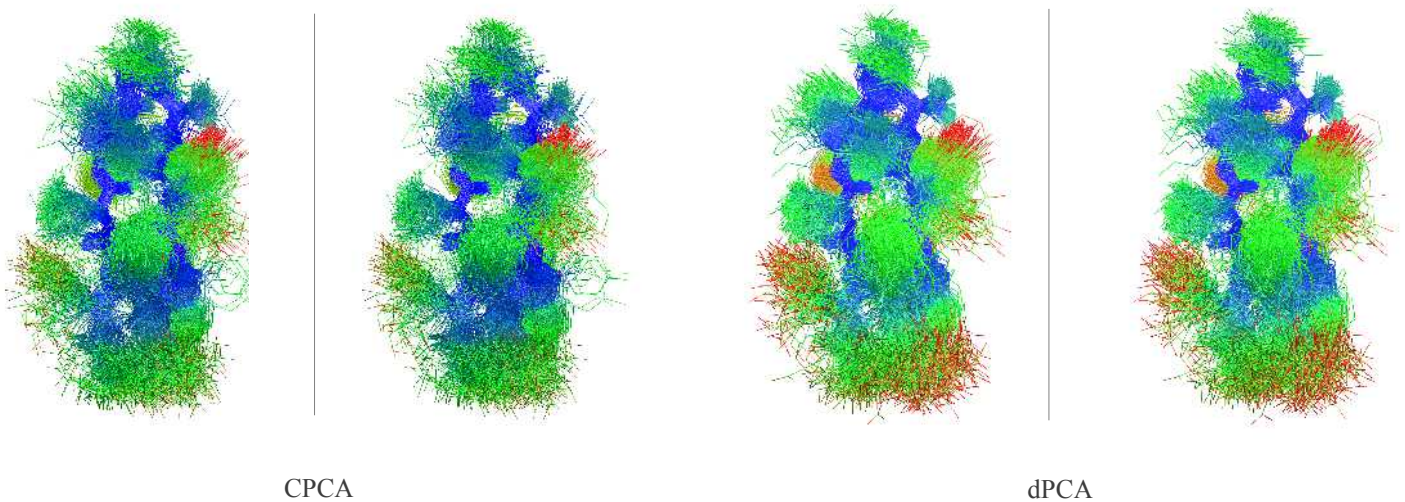
Εικόνα 4.54 Αντιπροσωπευτική δομή της κύριας ομάδας του τροχιακού όπως υπολογίστηκε από την CPCA.

Τα κατάλοιπα των άκρων και η πλευρική ομάδα της τυροσίνης παρουσιάζουν τη μεγαλύτερη κινητικότητα. Για να εντοπίσουμε τις κύριες διαφορές μέσα στην κυρίαρχη ομάδα, δημιουργήσαμε ένα καινούργιο τροχιακό που αποτελούνταν αποκλειστικά από τις δομές της κυρίαρχης ομάδας και εφαρμόσαμε στο νέο τροχιακό CPCA και dPCA. Σε κάθε ανάλυση εντοπίστηκε μόνο μία κύρια ομάδα

με ποσοστό 60-70%. Η υπέρθεση των δομών της κύριας ομάδας από κάθε ανάλυση παρουσιάζεται στην **Εικόνα 4.56**.



Εικόνα 4.55 Υπέρθεση δομών της κυρίαρχης ομάδας σε όψη stereo.



Εικόνα 4.56 Υπέρθεση δομών της κυρίαρχης ομάδας σε όψη stereo.

Η ομάδα που απομονώθηκε από την dPCA φαίνεται να έχει καλύτερα προσδιορισμένη την κύρια αλυσίδα κουβαλώντας όμως το κόστος των υπερκινητικών άκρων και πλευρικών ομάδων. Το όμοιο χαρακτηριστικό των δύο ομάδων είναι η υπερκινητική πλευρική ομάδα της τυροσίνης.

Έπειτα εφαρμόσαμε dihedral Principal Component Analysis (dPCA), με και χωρίς τις γωνίες χ_1 , στο τροχιακό. Για τους υπολογισμούς χρησιμοποιήθηκαν οι τρεις πρώτες κύριες συνιστώσες για τις οποίες

υπολογίστηκαν και τα διαγράμματα ελεύθερης ενέργειας. Οι κύριες ομάδες που εντοπίστηκαν στους υπολογισμούς χωρίς τις γωνίες χ_1 ήταν οι εξής:

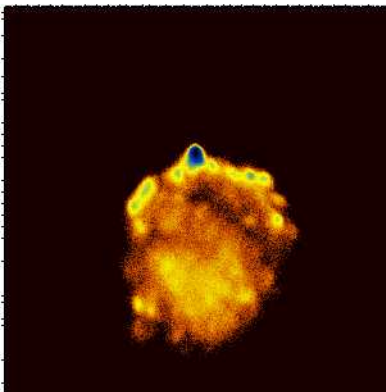
Cluster 1 : 824606 από τα 1787325 frames

Cluster 2 : 51954 από τα 1787325 frames

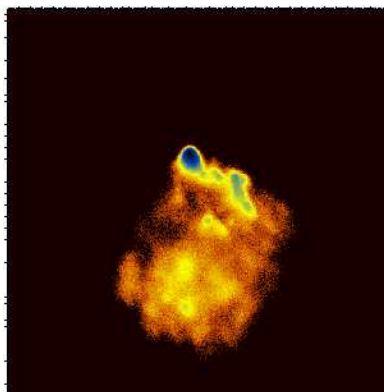
Cluster 3 : 59388 από τα 1787325 frames

Cluster 4 : 9595 από τα 1787325 frames

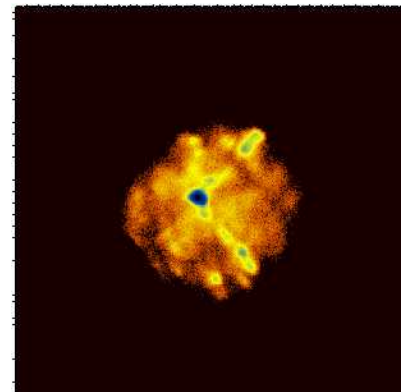
Cluster 5 : 10220 από τα 1787325 frames



PC_01_02



PC_01_03

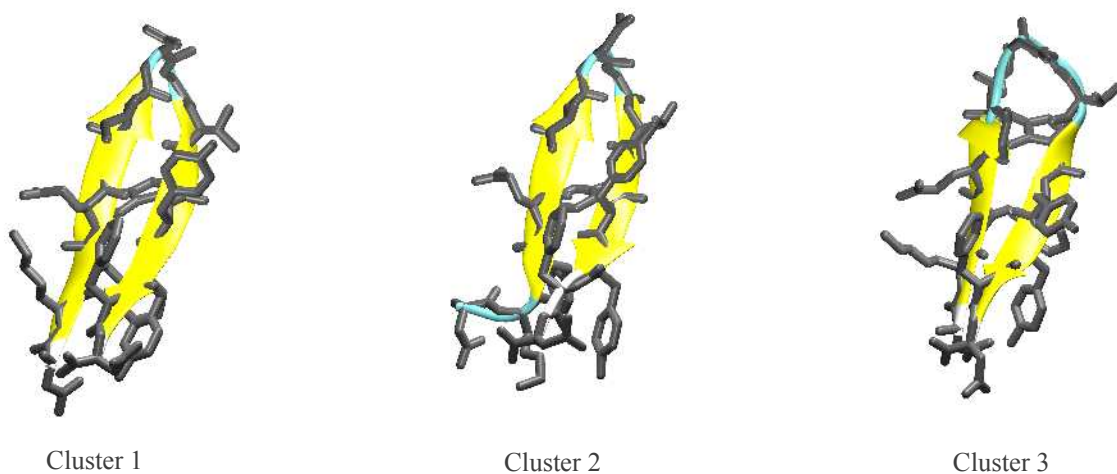


PC_02_03

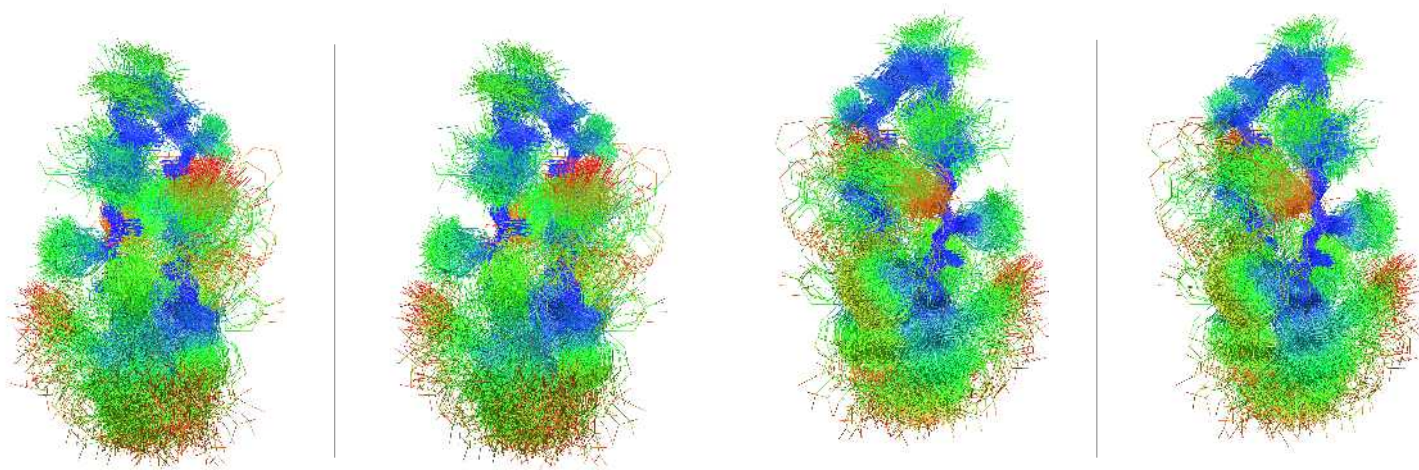
Εικόνα 4.57 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 46% του συνολικού τροχιακού, ενώ η δεύτερη και τρίτη αντιστοιχούν περίπου στο 3% η κάθε μια ξεχωριστά. Η κυρίαρχη ομάδα φαίνεται εύκολα στα διαγράμματα ελεύθερης ενέργειας, **Εικόνα 4.57**, και η αντιπροσωπευτική δομή της παρουσιάζεται στην **Εικόνα 4.58** μαζί με τις αντιπροσωπευτικές δομές των μικρότερων ομάδων. Η κινητικότητα των ατόμων του πεπτιδίου φαίνεται στην υπέρθεση των δομών της κυρίαρχης ομάδας, **Εικόνα 4.59**.

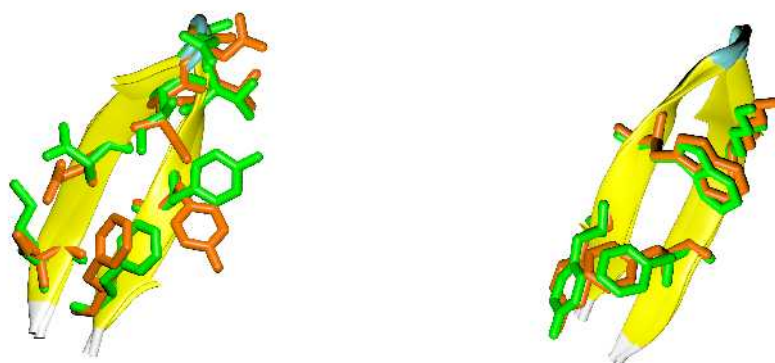
Όπως φαίνεται από την υπέρθεση των δομών της πρώτης ομάδας, **Εικόνα 4.59**, τη μεγαλύτερη κινητικότητα παρουσιάζουν τα άκρα του πεπτιδίου, καθώς επίσης και οι πλευρικές ομάδες των καταλοίπων της τρυπτοφάνης και της τυροσίνης. Η εφαρμογή cPCA στην κύρια ομάδα αποκάλυψε την ύπαρξη δύο μικρότερων πληθυσμών, των οποίων οι αντιπροσωπευτικές δομές παρουσιάζονται στην **Εικόνα 4.60** και οι υπερθέσεις των πληθυσμών στην **Εικόνα 4.61**. Η μεγαλύτερη ομάδα καταλαμβάνει το 50% της κυρίαρχης και η δεύτερη το 10%. Η κύρια διαφορά τους είναι η διεύθυνση της πλευρική ομάδας της τυροσίνης.



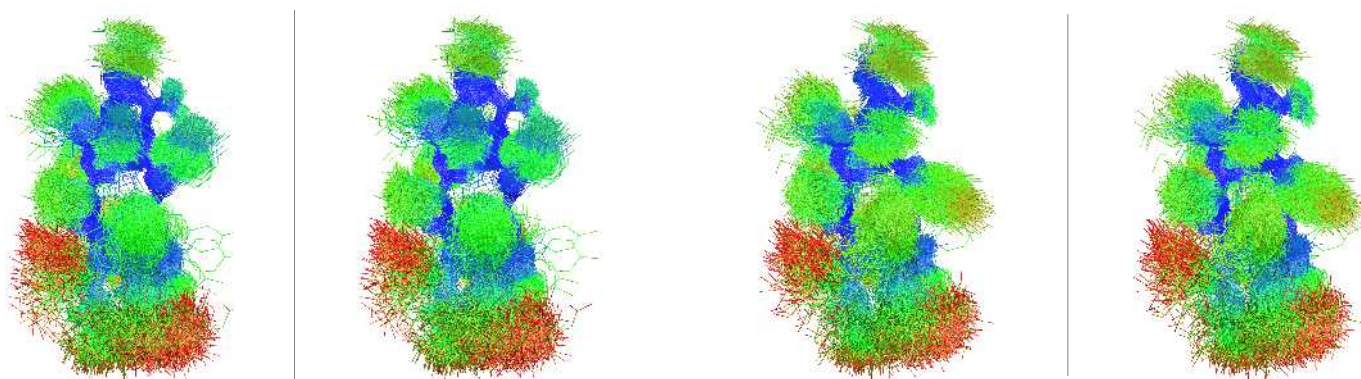
Εικόνα 4.58 Αντιπροσωπευτικές δομές από τις κύριες ομάδες του τροχιακού όπως υπολογίστηκαν από την dPCA.



Εικόνα 4.59 Υπέρθωση δομών της κυρίαρχης ομάδας σε όψη stereo.



Εικόνα 4.60 Αντιπροσωπευτικές δομές από τις κύριες ομάδες του τροχιακού όπως υπολογίστηκαν από την CPCA. Με πράσινο χρώμα είναι χρωματισμένα τα κατάλοιπα της μεγαλύτερης ομάδας και με πορτοκαλί τα κατάλοιπα της μικρότερης.

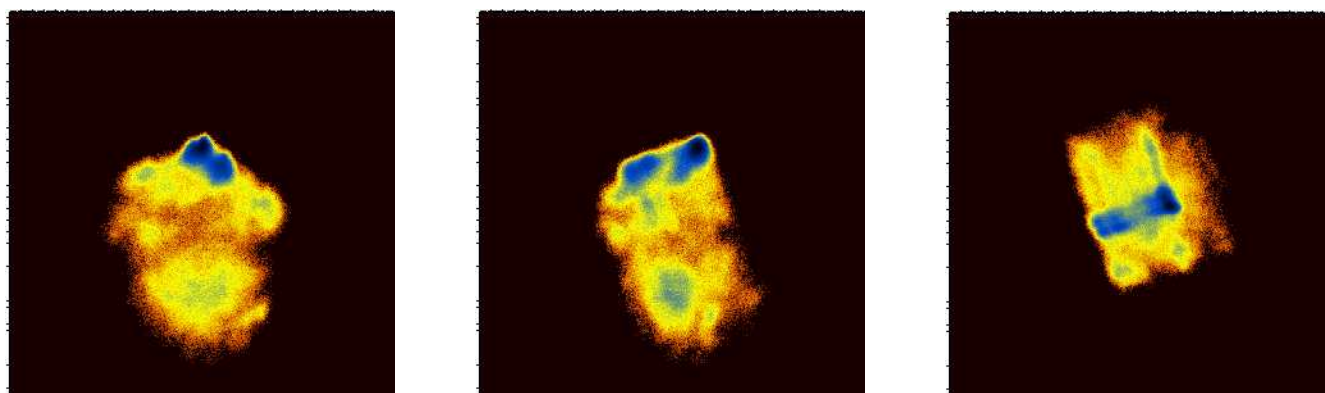


Εικόνα 4.61 Υπέρθυση δομών της κυρίαρχης ομάδας σε όψη stereo.

Τα αποτελέσματα της ανάλυσης χρησιμοποιώντας τις γωνίες χ_1 παρουσιάζονται στη συνέχεια:

Cluster 1 : 745533 από τα 1787325 frames

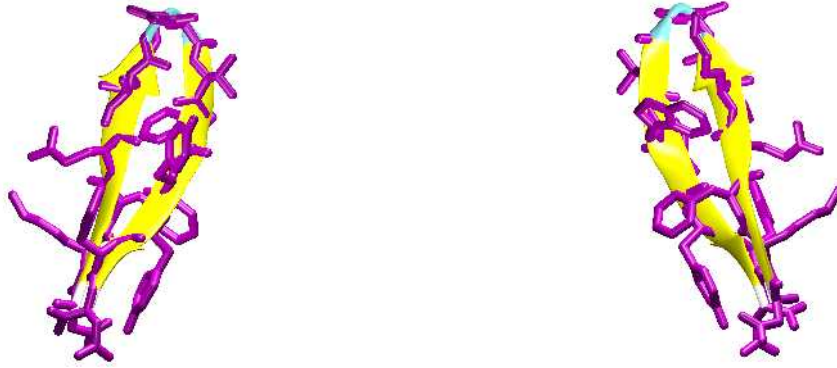
Cluster 2 : 3221 από τα 1787325 frames



Εικόνα 4.62 Γραφικές αναπαραστάσεις της ελεύθερης ενέργειας ως προς τους τρεις πρώτους PC ανά ζεύγη.

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η πρώτη ομάδα αντιστοιχεί στο 41% του συνολικού τροχιακού, ενώ η δεύτερη αντιστοιχεί σε ποσοστό μικρότερο της μονάδας. Η αντιπροσωπευτική δομή της κυρίαρχης ομάδας παρουσιάζεται στην **Εικόνα 4.63**. Η κινητικότητα των ατόμων του πεπτιδίου φαίνεται στην υπέρθεση των δομών της κυρίαρχης ομάδας, **Εικόνα 4.64**.

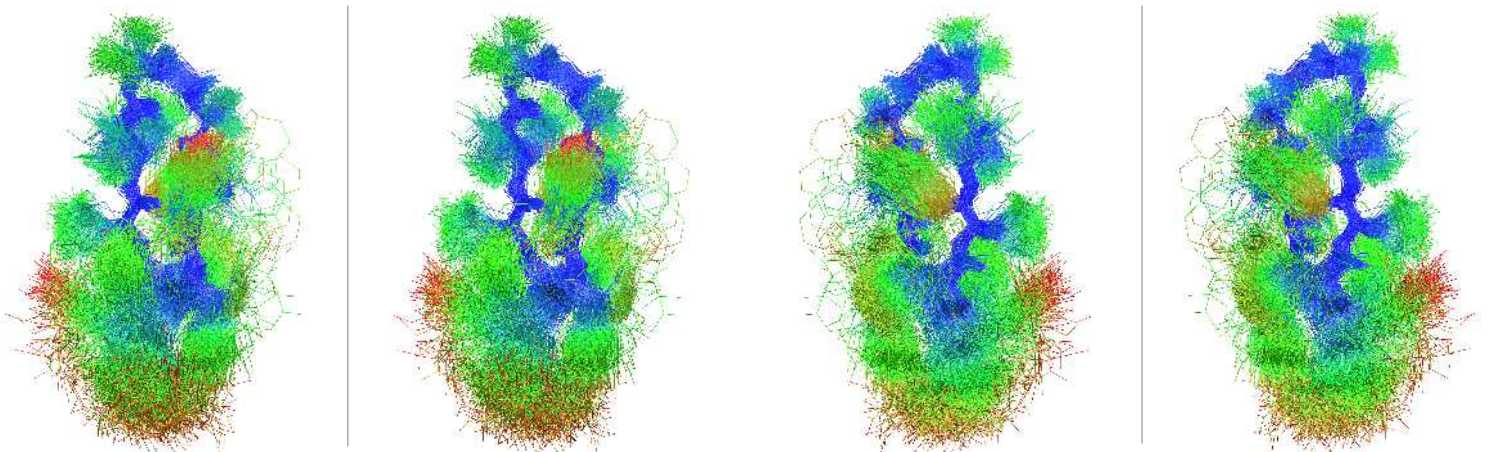
Στη συνέχεια εφαρμόσαμε CPCA και dPCA στην κυρίαρχη ομάδα για να εντοπίσουμε τις διαφορές μεταξύ των διαμορφώσεων που ανήκουν στην ομάδα αυτή. Σε κάθε ανάλυση εντοπίστηκε μόνο μια



Cluster 1

Εικόνα 4.63 Αντιπροσωπευτικές δομές από την κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA.

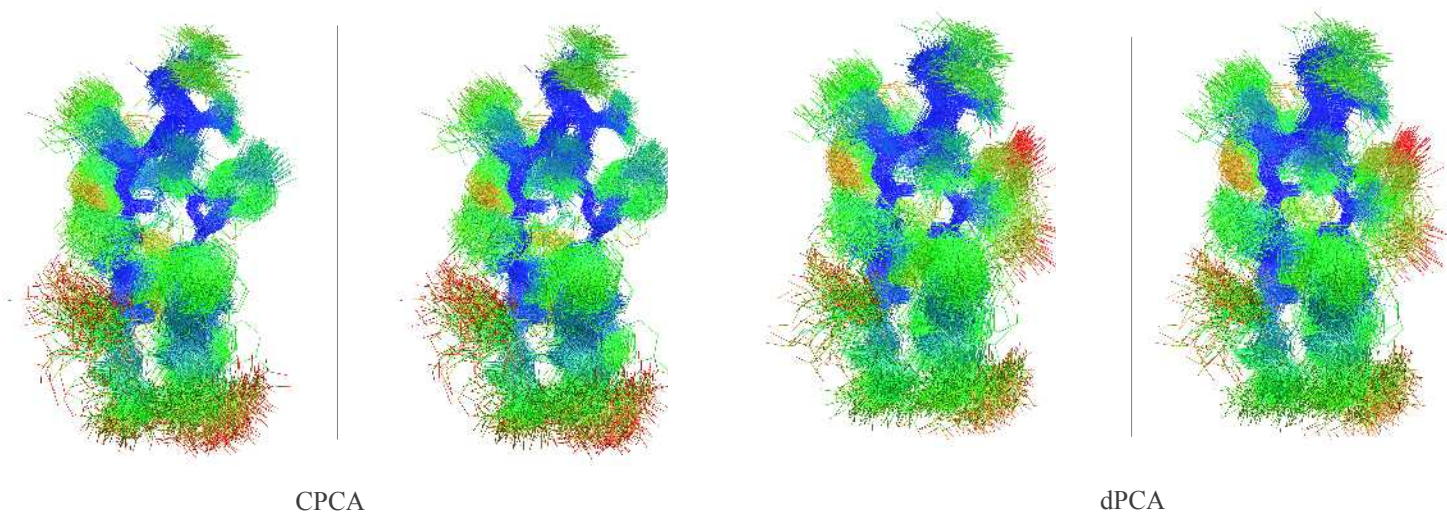
κύρια ομάδα η οποία κυμαίνονταν μεταξύ 40-50% της αρχικής ομάδας. Η υπέρθεση των δομών της κύριας ομάδας από κάθε ανάλυση παρουσιάζεται στην **Εικόνα 4.65**.



Εικόνα 4.64 Υπέρθεση δομών της κυρίαρχης ομάδας σε όψη stereo.

Στο τροχιακό αυτό έχουμε το μοναδικό γεγονός αναδίπλωσης, το οποίο καλύπτει τη μεγαλύτερη έκταση του τροχιακού. Παρά τις διαφορές που παρουσιάζουν οι αναλύσεις στην ομαδοποίηση των δεδομένων, τα αποτελέσματα των αναλύσεων συγκλίνουν και οι κύριες κινήσεις του πεπτιδίου αποτυπώνονται με ακρίβεια στις μεγαλύτερες ομάδες, **Πίνακας 4.6**. Οι αντιπροσωπευτικές δομές που απομονώθηκαν από κάθε ανάλυση παρουσιάζουν μεγάλη ομοιότητα με τη φυσική δομή, **Πίνακας 4.8**. Το πεπτίδιο φαίνεται να υιοθετεί τη φυσική δομή για το μεγαλύτερο κομμάτι του τροχιακού, αφού έχει σωστά διαμορφωμένες τη στροφή τύπου I και β-φουρκέτα. Η κύρια αλυσίδα φαίνεται να διατηρείται σταθερή σε κάθε κύρια ομάδα, ενώ η βασικές διαφορές μεταξύ των διαμορφώσεων της κάθε ομάδας

είναι η διευσθέτηση των πλευρικών ομάδων κυρίως της τυροσίνης και λιγότερο της τρυπτοφάνης.



Εικόνα 4.65 Υπέρθθεση δομών της κυρίαρχης ομάδας σε όψη stereo.

ILDN-360K

	RMSD_1	CPCA_1	DPCA_1	DPCA_CHI1_1
RMSD_1	-	0.37	0.37	0.39
CPCA_1	0.37	-	0.00	0.41
DPCA_1	0.37	0.00	-	0.41
DPCA_CHI1_1	0.39	0.41	0.41	-

Πίνακες 4.6 RMSD μεταξύ των αντιπροσωπευτικών δομών που απομονώθηκαν για κάθε σημαντική ομάδα.

4.9 Σύγκριση αποτελεσμάτων

Για να δούμε την ομοιότητα μεταξύ των αποτελεσμάτων κάθε ανάλυσης, υπολογίσαμε το RMSD για κάθε σημαντικά στατιστική δομή που απομονώθηκε. Στις περιπτώσεις όπου οι αντιπροσωπευτικές δομές από κάθε ανάλυση έμοιαζαν σε μεγάλο βαθμό μεταξύ τους, **Πίνακες 4.2-4.6**, υπολογίστηκαν οι μέσες δομές των δομών αυτών για το συγκεκριμένο τροχιακό. Για τους υπολογισμούς χρησιμοποιήθηκαν όλα τα άτομα της κύριας αλυσίδας. Τα αποτελέσματα παρουσιάζονται στον **Πίνακα 4.7**.

Σύμφωνα με τα αποτελέσματα του υπολογισμού του RMSD, οι αντιπροσωπευτικές δομές από κάθε τροχιακό παρουσιάζουν ομοιότητες μεταξύ τους. Γνωρίζουμε από τις αρχικές παρατηρήσεις της ανάλυσης ότι σε κάθε τροχιακό εντοπίστηκαν πληθυσμοί στους οποίους είχαν σχηματιστεί τμηματικά

	99SB	ILDN_ RMSD_1	ILDN'	TIP4'	TIP4_ DPCA_1	TIP4_ DPCA_ CHI1_3	340K	360K
99SB	-	2.51	2.14	2.29	8.01	2.23	4.06	1.71
ILDN_ RMSD_1	2.51	-	2.08	1.50	7.21	2.50	4.48	2.61
ILDN'	2.14	2.08	-	2.48	7.99	2.73	3.78	2.10
TIP4'	2.29	1.50	2.48	-	6.86	1.85	4.80	2.72
TIP4_ DPCA_1	8.01	7.21	7.99	6.86	-	6.75	7.76	8.15
TIP4_ DPCA_ CHI1_3	2.23	2.50	2.73	1.85	6.75	-	4.40	2.54
340K	4.06	4.48	3.78	4.80	7.76	4.40	-	2.29
360K	1.71	2.61	2.10	2.72	8.15	2.54	2.29	-

Πίνακας 4.7 Τιμές RMSD μεταξύ των αντιπροσωπευτικών δομών για κάθε σημαντικά στατιστική ομάδα που εντοπίστηκε σε κάθε τροχιακό. Στις περιπτώσεις όπου οι αντιπροσωπευτικές δομές από κάθε ανάλυση έμοιαζαν σε μεγάλο βαθμό μεταξύ τους, υπολογίστηκαν οι μέσες δομές των δομών αυτών για το συγκεκριμένο τροχιακό. Η δομή 99SB αντιστοιχεί στη μέση δομή των αντιπροσωπευτικών δομών όλων των κυρίαρχων ομάδων του τροχιακού, RMSD_3, CPCA_1, DPCA_2 και DPCA_CHI1_1, η δομή ILDN αντιστοιχεί στις ομάδες CPCA_1, DPCA_1 και DPCA_CH1_5, η TIP4' αντιστοιχεί στις ομάδες RMSD_2 και CPCA_2 και τέλος οι 340K και 360K αντιστοιχούν στη μέση δομή όλων των κυρίαρχων ομάδων κάθε ανάλυσης για κάθε τροχιακό αντίστοιχα. Για τους υπολογισμούς χρησιμοποιήθηκαν όλα τα άτομα της κύριας αλυσίδας και οι τιμές RMSD αναγράφονται σε Angstrom.

μία β-φουρκέτα και μία στροφή, οι οποίες όμως διέφεραν από αυτές της φυσικής δομής. Οι διαφορές εντοπίζονταν κυρίως στα κατάλοιπα που σχημάτιζαν τη β-φουρκέτα και στα κατάλοιπα που σχημάτιζαν τη στροφή, καθώς και τον τύπο της, όσον αφορά τη διευθέτηση των καταλοίπων και τον τύπο της στροφής αντίστοιχα.

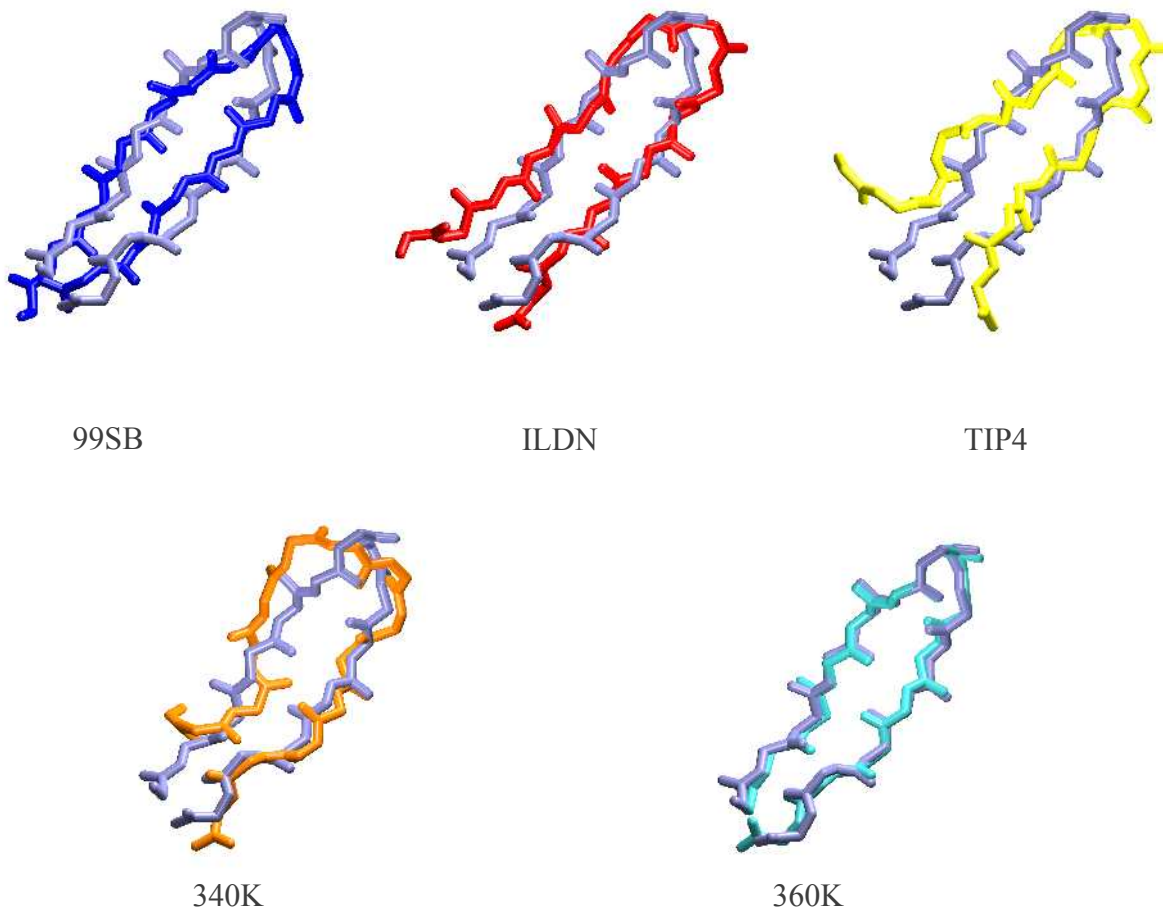
4.10 RMSD από πειραματικά προσδιορισμένες δομές

Αφού αποκτήσαμε μια καλύτερη ιδέα για το “κινείται” στις προσομοιώσεις, για κάθε σημαντική ομάδα που εντοπίστηκε από τις αναλύσεις υπολογίστηκε το RMSD από την προσδιορισμένη δομή των πειραμάτων κρυσταλλογραφίας ακτίνων-X, **Πίνακας 4.7**. Οι υπολογισμοί έγιναν χρησιμοποιώντας διαφορετικούς συνδυασμούς ατόμων και καταλοίπων. Στην **Εικόνα 4.66** παρουσιάζεται η υπέρθεση των κυρίων αλυσίδων για τις αντιπροσωπευτικές δομές με χαμηλότερο RMSD από τη δομή των πειραμάτων κρυσταλλογραφίας.

RMSD	HEAVY			Ca		
	1-14	3-12	6-9	1-14	3-12	6-9
99SB_1	6.34	6.79	3.06	4.39	4.58	1.63
99SB_2	6.98	7.28	2.07	5.03	4.72	0.92
99SB_3	4.85	5.01	2.13	1.88	1.90	1.16
ILDN_1	5.01	4.74	1.93	2.80	2.53	0.77
ILDN_2	5.95	6.01	2.49	3.67	3.35	0.99
ILDN_4	6.97	5.79	2.61	6.49	4.26	0.99
TIP4_1	8.61	6.77	3.00	8.29	5.05	1.56
TIP4_2	4.25	3.91	1.86	2.92	2.24	1.08
340K_1	3.64	3.49	2.45	2.31	1.86	1.36
340K_2	4.76	4.80	2.18	4.14	2.89	1.87
360K_1	1.12	1.09	0.84	0.32	0.26	0.14

Πίνακας 4.7 Τιμές RMSD των αντιπροσωπευτικών ομάδων που απομονώθηκαν από τις αναλύσεις σε σχέση με την πειραματικά προσδιορισμένη δομή. Στην πρώτη στήλη αναγράφονται οι αντιπροσωπευτικές δομές των ομάδων. Η αρίθμηση έγινε με βάση τη θέση τους στους πίνακες RMSD. Οι πρώτοι υπολογισμοί έγιναν χρησιμοποιώντας όλα τα άτομα του πεπτιδίου εκτός από τα υδρογόνα, ενώ οι δεύτεροι χρησιμοποιώντας μόνο τα άτομα Ca. Επιπλέον έγιναν τρεις διαφορετικοί υπολογισμοί για τις τιμές RMSD όσον αφορά όλα τα κατάλοιπα του πεπτιδίου, χωρίς κατάλοιπα των άκρων και χρησιμοποιώντας τα κατάλοιπα γύρω από τη στροφή. Οι τιμές RMSD αναγράφονται σε Angstrom.

Από τα παραπάνω αποτελέσματα γίνεται εύκολα αντιληπτό πως οι περισσότερες δομές διαφέρουν σε μεγάλο βαθμό από τη φυσική δομή. Ο **Πίνακας 4.8** αποτελεί μια ποσοτική ανάλυση της ομοιότητας, ενώ η **Εικόνα 4.66** μας δίνει μια ποιοτική αίσθηση για τη διαφορετικότητα των ομάδων που εντοπίστηκαν στις αναλύσεις.



Εικόνα 4.66 Υπέρθυση των κυρίων αλυσίδων των αντιπροσωπευτικών δομών των ομάδων της ανάλυσης με τη δομή των πειραμάτων κρυσταλλογραφίας. Η δομή των πειραμάτων κρυσταλλογραφίας έχει γκρι απόχρωση.

Στο μοναδικό γεγονός αναδίπλωσης υπάρχει μεγάλη συμφωνία με τη δομή των πειραμάτων κρυσταλλογραφίας σε επίπεδο κύριας αλυσίδας, όπου βλέπουμε να έχει σχηματιστεί σωστά η στροφή και να έχουν διευθετηθεί σωστά τα κατάλοιπα της β-φουρκέτας. Εκτός από την κύρια αλυσίδα σωστά φαίνεται να έχουν διευθετηθεί και οι πλευρικές ομάδες των αρωματικών καταλοίπων στα οποία προσδέεται η χολίνη, **Εικόνα 4.67**. Στις υπόλοιπες ομάδες δεν έχει σχηματιστεί σωστά κάποιο στοιχείο δευτεροταγούς δομής της φυσικής δομής, αλλά το πεπτίδιο φαίνεται να υιοθετεί μη φυσικές δομές, εκτεταμένες ή πιο συμπαγείς, για μεγάλα χρονικά διαστήματα.

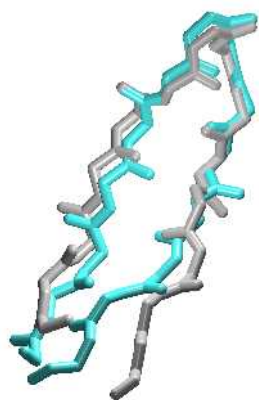
Στη συνέχεια υπολογίσαμε το RMSD της αντιπροσωπευτικής δομής που απομονώθηκε από το μοναδικό γεγονός αναδίπλωσης των προσομοιώσεων με τις προσδιορισμένες δομές των πειραμάτων NMR, **Πίνακας 4.8**. Στην **Εικόνα 4.68** παρουσιάζεται η υπέρθεση των κυρίων αλυσίδων για τις δομές με χαμηλότερο RMSD ως προς τα Ca άτομα, ενώ στην **Εικόνα 4.69** παρουσιάζεται η υπέρθεση της



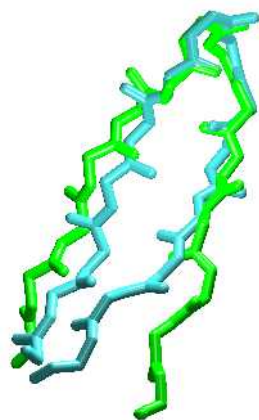
Εικόνα 4.67 Υπέρθυση της δομής της προσομοίωσης και της δομής της κρυσταλλογραφίας. Οι πλευρικές ομάδες της δομής της κρυσταλλογραφίας έχουν γκρι απόχρωση, ενώ οι κυανές ανήκουν στη δομή της προσομοίωσης.

RMSD	HEAVY			C α		
	1-14	3-12	6-9	1-14	3-12	6-9
NMR_1	2.74	2.53	0.64	2.25	1.66	0.33
NMR_2	2.74	2.53	0.71	1.88	1.12	0.35
NMR_3	2.42	1.97	0.64	1.16	0.89	0.30
NMR_4	2.92	2.77	0.78	2.29	1.24	0.34
NMR_5	2.12	1.92	0.74	1.51	1.03	0.31
NMR_6	2.73	2.71	0.67	1.63	0.93	0.33
NMR_7	1.92	1.97	0.71	1.46	0.90	0.35
NMR_8	3.10	3.22	0.70	2.09	1.51	0.31
NMR_9	2.89	2.78	0.66	2.04	1.32	0.35
NMR_10	2.00	1.85	0.94	1.35	0.73	0.32
NMR_11	2.14	2.23	1.17	1.12	0.73	0.33
NMR_12	2.95	2.63	0.87	2.43	1.43	0.33
NMR_13	3.12	2.32	0.94	2.56	1.89	0.37
NMR_14	2.58	1.98	0.64	1.57	0.61	0.35
NMR_15	2.01	2.03	0.81	0.99	0.70	0.32
NMR_16	2.46	2.01	0.61	1.52	0.99	0.31
NMR_17	2.36	1.82	0.70	1.88	0.90	0.34
NMR_18	2.49	2.23	1.30	1.81	1.09	0.63
NMR_19	2.15	2.18	0.76	1.41	1.16	0.35
NMR_20	2.16	2.73	0.62	1.92	1.30	0.32

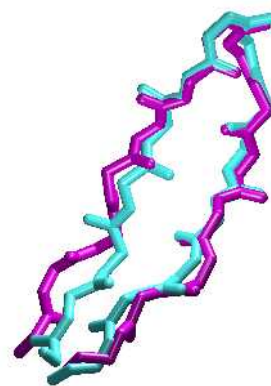
Πίνακας 4.8 Τιμές RMSD της αντιπροσωπευτικής δομής που απομονώθηκε από το γεγονός αναδίπλωσης με τις προσδιορισμένες δομές από τα πειράματα NMR του Sanz J . Οι τιμές RMSD αναγράφονται σε Angstrom.



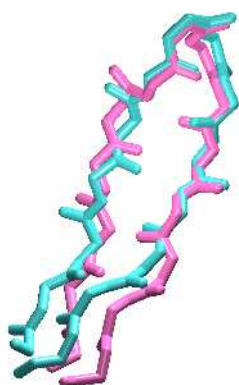
360K-NMR_3



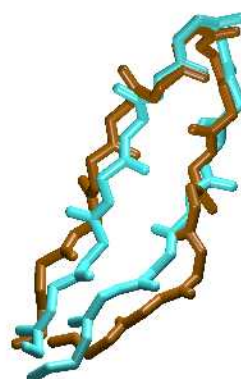
360K-NMR_9



360K-NMR_10

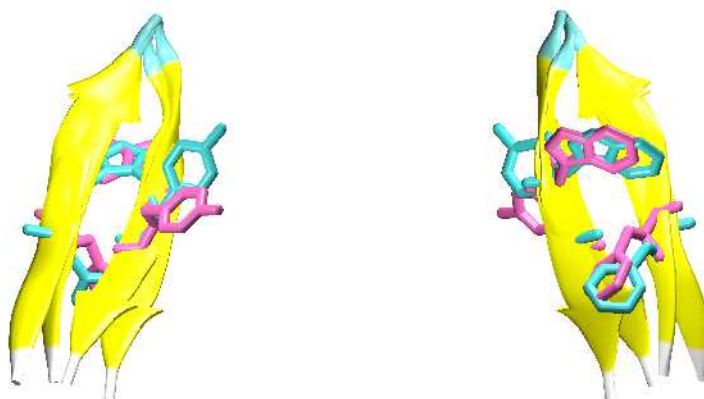


360K-NMR_15



360K-NMR_19

Εικόνα 4.68 Υπέρθηση των κυρίων αλυσίδων για τις δομές με χαμηλότερο RMSD μεταξύ της αντιπροσωπευτικής δομής που απομονώθηκε από το γεγονός αναδίπλωσης με τις προσδιορισμένες δομές των πειραμάτων NMR. Με κυανό χρώμα παρουσιάζεται η αναδιπλωμένη δομή που απομονώθηκε από τις προσομοιώσεις.

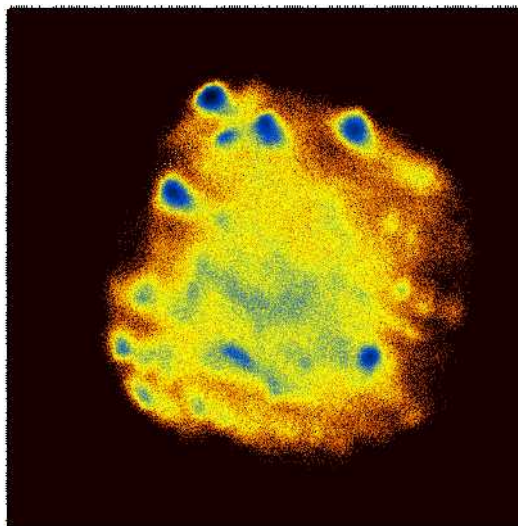


Εικόνα 4.69 Υπέρθηση της δομής της προσομοίωσης και της δομής των πειραμάτων NMR. Οι πλευρικές ομάδες της δομής των πειραμάτων NMR έχουν ροζ χρώμα, ενώ οι κυανές ανήκουν στη δομή της προσομοίωσης.

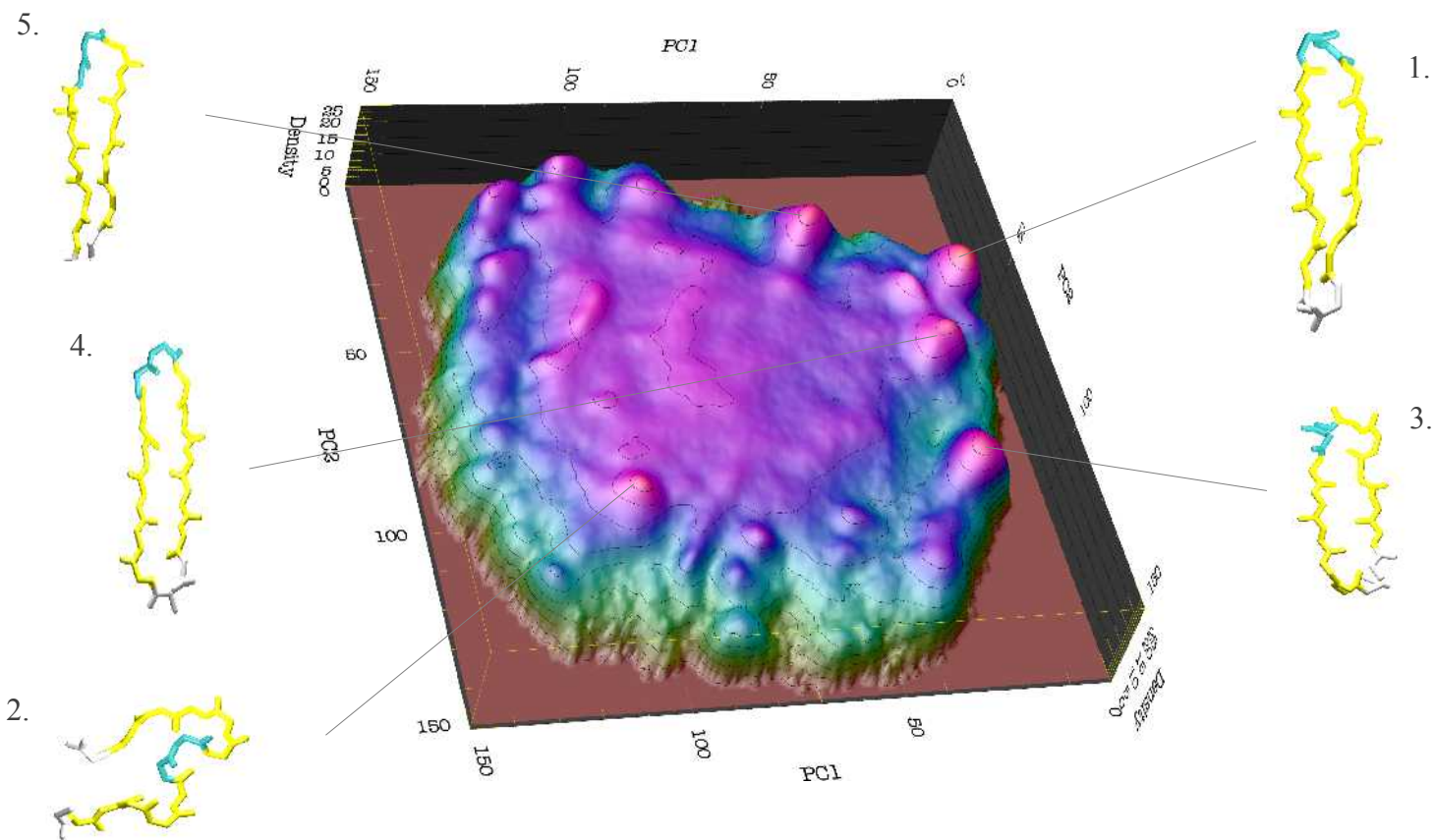
δομής της προσομοίωσης και της δομής των πειραμάτων NMR. Από τα παραπάνω αποτελέσματα προκύπτει ότι η αντιπροσωπευτική δομή του γεγονότος αναδίπλωσης που εντοπίστηκε παρουσιάζει ομοιότητες με μερικές από τις δομές των πειραμάτων NMR σε επίπεδο κύριας αλυσίδας και πλευρικών ομάδων. Παρόλα αυτά, η σύγκλιση μεταξύ της δομής της προσομοίωσης και της δομής της κρυσταλλογραφίας είναι πολύ μεγαλύτερη σε σχέση με οποιαδήποτε δομή των πειραμάτων NMR.

4.11 Πρώτα συμπεράσματα

Για να ανακεφαλαιώσουμε, στην εργασία αυτή μελετήσαμε πέντε διαφορετικές προσομοιώσεις του πεπτιδίου LytA197-210[ND] με διάφορους συνδυασμούς θερμοκρασίας και μοντέλων νερού. Η συνολική διάρκεια των προσομοιώσεων ήταν 13.7 μ s. Στο διάστημα αυτό, παρατηρήθηκε μόνο ένα γεγονός αναδίπλωσης με διάρκεια σχεδόν 0.9 μ s, η αντιπροσωπευτική δομή του οποίου ταυτίζεται με τη δομή των πειραμάτων κρυσταλλογραφίας. Το υπόλοιπο διάστημα των προσομοιώσεων το πεπτίδιο υιοθετεί διάφορες μη φυσικές δομές διάρκειας 0.3 μ s με 0.6 μ s, οι οποίες διαφέρουν αρκετά από τη φυσική δομή, συνήθως λόγω της λάθος σχηματισμένης β -φουρκέτας. Στην προσπάθειά μας να ερμηνεύσουμε τα αποτελέσματα αυτά, κάναμε μια σειρά από υπολογισμούς. Αρχικά κατασκευάσαμε ένα καινούργιο τροχιακό, το οποίο προήλθε από την ένωση των πέντε αρχικών τροχιακών, για να αποκτήσουμε μια γενική ιδέα για τη συνολική συμπεριφορά του πεπτιδίου. Γνωρίζουμε εξ' αρχής ότι το καινούργιο τροχιακό αποτελείται από τροχιακά με διαφορετικές θερμοκρασίες και παραμέτρους και κατά συνέπεια τα διαγράμματα που κατασκευάστηκαν δεν αποτελούν φυσική περιγραφή ενός ενεργειακού τοπίου. Για να εντοπίσουμε τις συνολικές κινήσεις εφαρμόσαμε dPCA στο καινούργιο τροχιακό. Στη συνέχεια παρουσιάζεται το διάγραμμα ελεύθερης ενέργειας που προέκυψε, **Εικόνα 4.70-71**.



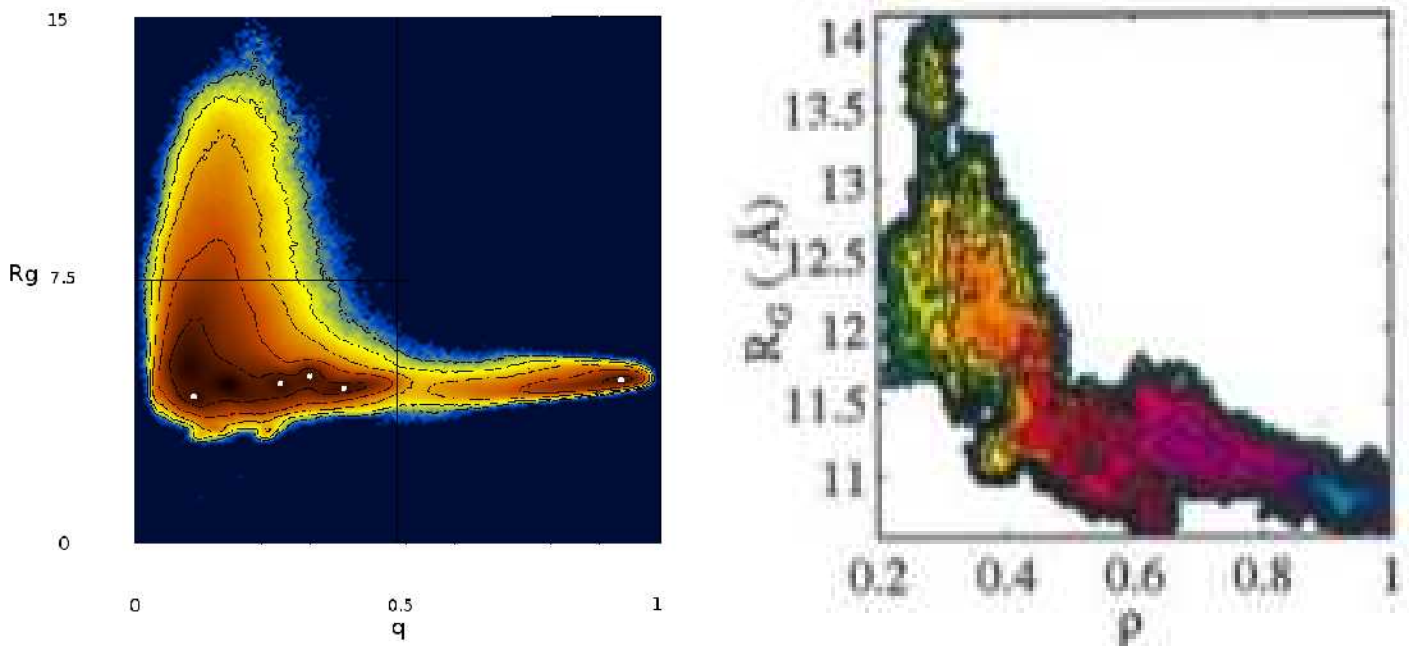
Εικόνα 4.70 Γραφική αναπαράσταση της ελεύθερης ενέργειας ως προς τους δύο πρώτους PCs.



Εικόνα 4.71 Γραφική αναπαράσταση της ελεύθερης ενέργειας ως προς τους δύο πρώτους PCs. Στην εικόνα παρουσιάζονται και οι κύριες αλυσίδες των σημαντικότερων ομάδων που εντοπίστηκαν, ενώ πάνω δεξιά βρίσκεται η αναδιπλωμένη δομή. Η αρίθμηση υποδηλώνει την ομάδα όπως υπολογίστηκε από την dPCA. Η εικόνα προετοιμάστηκε με το πρόγραμμα orpndx.

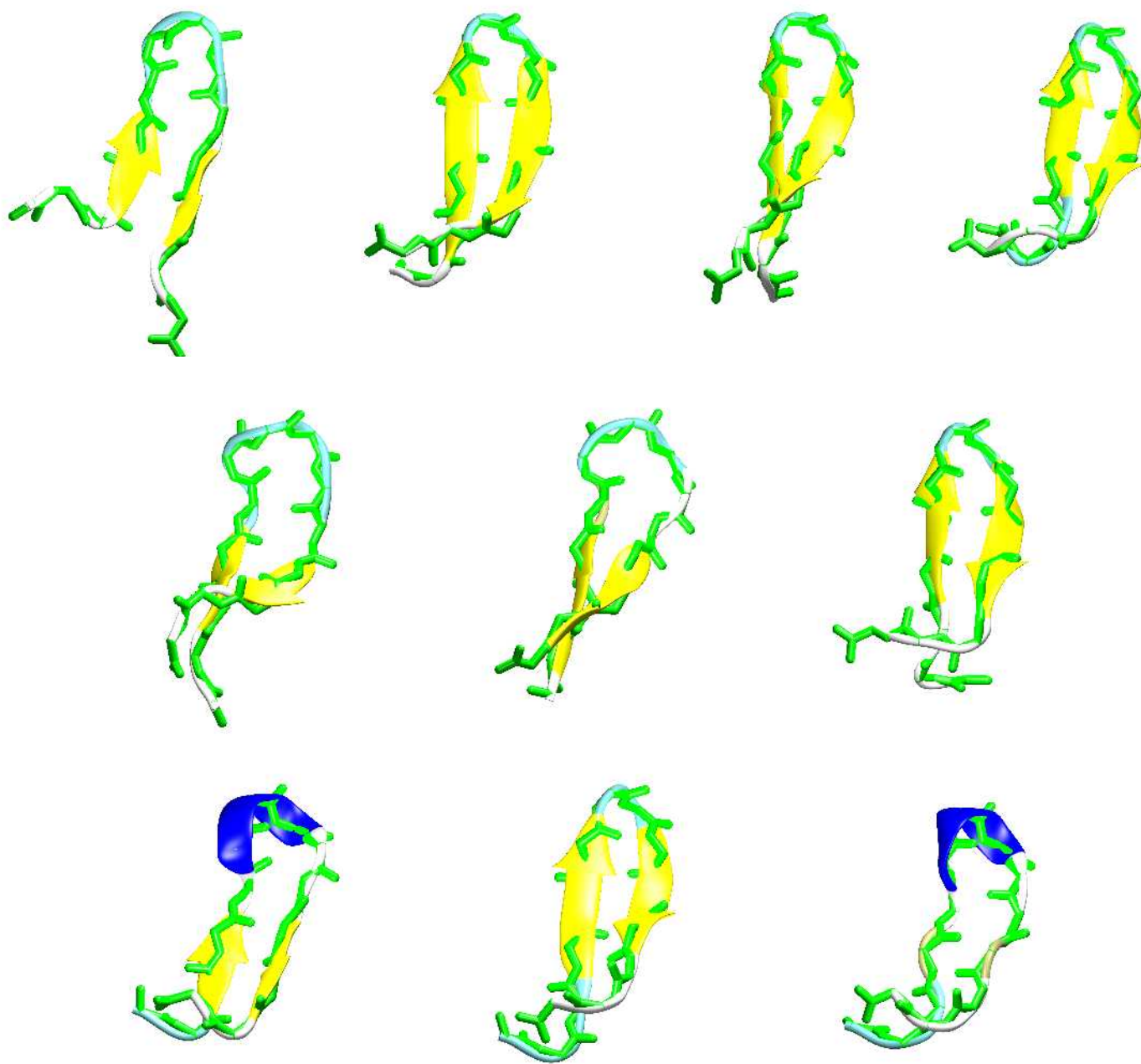
Στην **Εικόνα 4.71** φαίνεται ότι το πεπτίδιο υιοθετεί αρκετές μη φυσικές διαμορφώσεις των οποίων η πυκνότητα συγκρίνεται με αυτή των αναδιπλωμένων διαμορφώσεων. Στις περισσότερες μη φυσικές δομές παρατηρείται λάθος σχηματισμός της β-φουρκέτας. Η **Εικόνα 4.72** αποτελεί μια διαφορετική αναπαράσταση της κατανομής των διαμορφώσεων του τροχιακού με βάση την ομοιότητα τους με τη φυσική δομή, παράμετρος q , και την παράμετρο Radius of Gyration (R_g), η οποία είναι ενδεικτική του μεγέθους του μορίου. Η γραφική αναπαράσταση αυτή, μοιάζει με το ενεργειακό τοπίο της αναδίπλωσης πρωτεϊνών [76], διαφέρει όμως στο γεγονός ότι λείπει η χαρακτηριστική κύρτωση που παρουσιάζουν τα πεπτίδια των οποίων τα ενεργειακά πεδία μοιάζουν με ένα ιδανικό χωνί. Στο διάγραμμα φαίνεται η ύπαρξη πολλών διαμορφώσεων με χαμηλό R_g που διαφέρουν όμως σημαντικά από φυσική δομή, χαμηλές τιμές q . Η λεπτή λωρίδα των διαμορφώσεων με τιμές q περίπου 0.5 αντιπροσωπεύει διαμορφώσεις που πιθανώς ανήκουν σε κάποια μεταβατική κατάσταση (transition

state) με υψηλή ενέργεια η οποία οδηγεί στην αναδίπλωση στη φυσική δομή. Στην **Εικόνα 4.73** παρουσιάζονται δομές που απομονώθηκαν από τον πληθυσμό της πιθανής μεταβατικής κατάστασης. Η απομόνωση των δομών έγινε εντοπίζοντας τις διαμορφώσεις από το ολικό τροχιακό των οποίων οι τιμές q κυμαίνονταν από 0.525 έως 0.55 και οι τιμές R_g κυμαίνονταν από 4.2 έως 4.8, από τις οποίες επιλέχθηκαν τυχαία ορισμένες από αυτές.



Εικόνα 4.72 Γραφική αναπαράσταση των διαμορφώσεων του τροχιακού με βάση τις παραμέτρους R_g και q (ρ). Η αριστερή εικόνα αντιστοιχεί στο ολικό τροχιακό. Η κορυφή με τιμές q κοντά στο 1 αντιστοιχεί στις αναδιπλωμένες διαμορφώσεις. Οι άσπρες κουκκίδες αντιστοιχούν στις δομές 2, 5, 3, 4, 1 ξεκινώντας από αριστερά προς δεξιά. Στη δεξιά εικόνα παρουσιάζεται το ενεργειακό πεδίο της πρωτεύουσας GB1. (Η εικόνα αναπαράγεται άνευ αδείας)

Οι δομές που απομονώθηκαν παρουσιάζουν σημαντική ετερογένεια όσον αφορά τα κατάλοιπα που συμμετέχουν στο σχηματισμό των στοιχείων δευτεροταγούς δομής. Επιπλέον στον πληθυσμό αυτό εντοπίζονται δομές, οι οποίες υιοθετούν έλικες στα κατάλοιπα της στροφής. Παρά τις διαφορές τους, οι περισσότερες δομές παρουσιάζουν ένα κοινό χαρακτηριστικό, το οποίο είναι η σωστή διεύθυνση των καταλοίπων της στροφής και ο σχηματισμός της στροφής τύπου I που εντοπίζεται και στη φυσική δομή. Λόγω του πεπερασμένου χρόνου και των πλαισίων της διπλωματικής εργασίας, η διενέργεια υπολογισμών για την εύρεση πιθανών μεταβατικών καταστάσεων δεν ήταν δυνατή.



Εικόνα 4.73 Δομές που απομονώθηκαν από τον πληθυσμό της πιθανής μεταβατικής κατάστασης.

Κεφάλαιο 5

Συμπεράσματα και συζήτηση

*We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.*

T.S. Eliot

5.1 Συμπεράσματα και συζήτηση

Η πρωτεΐνη LytA αποτελεί την κυριότερη αυτολυσίνη του βακτηρίου *Streptococcus pneumoniae*. Στην εργασία αυτή, μελετήθηκε υπολογιστικά η συμπεριφορά του 14-πεπτιδίου DKFEKIDGTWYYFD, το οποίο αποτελεί μετάλλαγμα του πεπτιδίου LytA197–210[wt]. Το πεπτίδιο άγριου τύπου συμμετέχει στο σχηματισμό δομών που προσδένουν χολίνη, CBRs, στη βακτηριακή επιφάνεια, οι οποίες αποτελούνται από πολλαπλές επαναλήψεις καλώς συντηρημένων αλληλουχιών που αποτελούνται από

β-φουρκέτες. Ο Sanz και οι συνεργάτες του [38] έδειξαν ότι το πεπτιδίο άγριου τύπου, LytA197–210[wt], και το μετάλλαγμα του LytA197 – 210[ND] έχουν την ικανότητα να αναδιπλώνονται σε υδατικά διαλύματα σε δομές παρόμοιες με τη φυσική κατάσταση, οι οποίες έχουν τη δυνατότητα να προσδέσουν χολίνη και ανάλογά της.

Συνολικά διεξήχθησαν 13.7 μs προσομοιώσεων του LytA197 – 210[ND], με διαφορετικούς συνδυασμούς παραμέτρων, εφαρμόζοντας το δυναμικό πεδίο AMBER και ορισμένες επεκτάσεις του. Στόχος τους ήταν η μελέτη της αναδίπλωσης του πεπτιδίου και η σύγκριση των αποτελεσμάτων μεταξύ τους, καθώς και με πειραματικά δεδομένα, όπως κρυσταλλογραφία ακτίνων-Χ. Στη συνολική διάρκεια των προσομοιώσεων που διεξήχθησαν, εντοπίστηκε ένα μοναδικό γεγονός αναδίπλωσης του οποίου η αντιπροσωπευτική δομή ταυτίζεται με τη δομή των πειραμάτων κρυσταλλογραφίας, σε επίπεδο κύριας αλυσίδας και πλευρικών ομάδων. Για το μεγαλύτερο διάστημα των προσομοιώσεων το πεπτιδίο υιοθετεί ημισταθερές διαμορφώσεις στις οποίες δεν υπάρχει αναδίπλωση. Στις διαμορφώσεις αυτές έχουν σχηματιστεί κάποια στοιχεία δευτεροταγούς δομής, τα οποία όμως διαφέρουν από τη φυσική δομή. Οι κύριες διαφορές εντοπίζονται μεταξύ των καταλοίπων που σχηματίζουν τις β-φουρκέτες και τις στροφές στις ημισταθερές αυτές δομές. Εκτός από τις διαφορές με τη φυσική δομή, διαφέρουν και για κάθε τροχιακό ξεχωριστά αφού δεν υπάρχει κάποιο συντηρημένο χαρακτηριστικό μεταξύ των κυρίαρχων δομών που απομονώθηκαν από κάθε προσομοίωση. Περιστασιακά ορισμένα κατάλοιπα του πεπτιδίου φαίνεται να υιοθετούν ελικοειδείς δομές, οι οποίες δεν είναι στατιστικά σημαντικές.

Στην εργασία αυτή υπάρχουν δύο κύρια σημεία τα οποία θα πρέπει να επισημανθούν. Πρώτον, η ικανότητα του δυναμικού πεδίου AMBER99SB-ILDN να αναπαράγει δομές, οι οποίες συγκλίνουν τόσο με τα πειραματικά δεδομένα ώστε να ταυτίζονται με τις δομές των πειραμάτων κρυσταλλογραφίας ακτίνων-Χ, γεγονός που υποστηρίζει τη χρήση του δυναμικού πεδίου για περαιτέρω προσομοιώσεις και πιθανές βελτιώσεις. Δεύτερον, η ύπαρξη ενός μοναδικού γεγονότος αναδίπλωσης και αρκετών ημισταθερών δομών με μεγάλη έκταση υποστηρίζουν την ύπαρξη κινητικών εμποδίων στο ενεργειακό πεδίο της αναδίπλωσης του πεπτιδίου (kinetic frustration). Το ενεργειακό πεδίο αυτό, πιθανώς δε θα έχει σχήμα χωνιού αλλά θα μοιάζει περισσότερο με ένα εκτεταμένο πεδίο με αρκετά ενεργειακά ελάχιστα όπου θα παγιδεύεται το πεπτιδίο και ένα ενεργειακό εμπόδιο στο οποίο θα εντοπίζεται κάποια μεταβατική κατάσταση που θα οδηγεί στο γεγονός της αναδίπλωσης.

Παράρτημα

Χρωματισμός καταλοίπων σύμφωνα με το VMD

Αμινοξύ	Χρώμα
Ασπαραγίνη	Κόκκινο
Λυσίνη	Γαλάζιο
Φαινυλαλανίνη	Μωβ
Γλουταμινικό οξύ	Ροζ
Ισολευκίνη	Ανοιχτό πράσινο
Γλυκίνη	Λευκό
Θρεονίνη	Βυσσινί
Τρυπτοφάνη	Γκρι
Τυροσίνη	Πράσινο

Χρωματισμός καταλοίπων σύμφωνα με τις μεθόδους DSSP και STRIDE

Στοιχείο Δευτεροταγούς Δομής	Χρώμα
Στροφή	Γαλάζιο
B-φύλλα	Κίτρινο
Απομονωμένη γέφυρα	Γκρι
α έλικα	Κόκκινο
3-10 έλικα	Μωβ
π έλικα	Μπλε
Σπείρα	Λευκό

Βιβλιογραφία

1. Branden C and Tooze J. 1999. Introduction to protein structure
2. Οικονόμου Ε. Ν. 2004. Η φυσική σήμερα
3. Dill K. A, Ozkan S. B, Shell M. S, and Weikl T. R. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:289–316
4. Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181:223– 30
5. Haber E, Anfinsen CB. 1962. Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* 237:1839–44
6. Levinthal C. 1968. Are there pathways for protein folding? *J. Chim. Phys.* 65, 44-45.
7. Dill K, Chan H.S. 1997. From Levinthal to pathways to funnels *Nat. Struct. Biol.* 4 , 10–19
8. Karplus M. 1997. The Levinthal paradox: yesterday and today. *Elsevier* 2:69-75
9. Agashe VR, Shastry MC, Udgaonkar JB. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature* 377: 754-757.
10. Brylinskia M, Leszek Koniecznyc, Irena Rotermana. 2006. Hydrophobiccollapse in (in silico) protein folding : 255–267
11. Ohgushi M, Wada A. 1983. Molten-globule state': a compact form of globular proteins with mobile side-chains. *FEBS Lett.* 164: 21– 24.
12. Christopher M. Dobson. 1992. Unfolded proteins, compact states and molten globules: *Current Opinion in Structural Biology*, 2:6–12
13. Jackson SE. 1998. How do small single-domain proteins fold? 4: 81- 91.
14. Religa TL, Markson JS, Mayor U, Freund SM, Fersht AR. 2005. Solution structure of a protein denatured state and folding intermediate. *Nature*; 437:1053–1056.
15. Feng HQ, Vu ND, Bai YW. 2005. Detection of a hidden folding intermediate of the third domain of PDZ. *Journal of Molecular Biology*;346:345–353.
16. Karplus M, Weaver DL. 1979. Diffusion–collision model for protein folding. *Biopolymers.* 18: 1421- 37.
17. Karplus M, Weaver DL. 1994. The Protein Society Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Science* 3: 650- 68.
18. Fersht A R. 1997. Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* 7, 3–9
19. Nolting B, Agard D. A. 2008. How General Is The Nucleation–Condensation Mechanism?. *Proteins* 73(3): 754-64
20. Karplus M. 2011. Behind the folding funnel diagram. *Nature Chemical Biology* 7, 401–404

21. Kelly S. M and Price N. C. 2000. The Use of Circular Dichroism in the Investigation of Protein Structure and Function. *Current Protein and Peptide Science*, 1, 349-384
22. Greenfield N. J. 2007. Using circular dichroism spectra to estimate protein secondary structure *Nature Protocols* 1, 2876 – 2890
23. Yiwen Chen, Feng Ding, Huifen Nie, Serohijos A, Shantanu Sharma, Kyle C. Wilcox, Shuangye Yin, and Dokholyan N. Y. 2008. Protein Folding: Then and Now . *Arch Biochem Biophys.* 469(1): 4–19.
24. Dinner A.R, Sali A, Smith L. J, Dobson C. B, and Karplus M. 2002 Understanding protein folding via free-energy surfaces from theory and experiment . *TIBS* 25: 331-339
25. Freddolino P L., Harrison C. B., Yanxin Liu, and Schulten K. 2010. Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat Phys.* 6(10): 751–758.
26. Protein and peptide folding explored with molecular simulations. 2002. *Accounts of Chemical Research* 35 (6), 447-54.
27. Ho B. K, Dill K.A. 2006. Folding Very Short Peptides Using Molecular Dynamics. *PLoS Comput Biol* 2(4): e27
28. Bernal J., and D. Crowfoot. 1934. X-ray photographs of crystalline pepsin. *Nature.* 133:794-795.
29. Wutrich K. 1986. *NMR of Proteins and Nucleic Acids.* J. Wiley, New York.
30. Nicosia G, Stracquadanio G. 2008. Generalized Pattern Search Algorithm for Peptide Structure Prediction. *Biophysical Journal* Volume 95 Issue 10 4988–4999
31. Voelz VA, Shell MS, Dill K. A. 2009. Predicting Peptide Structures in Native Proteins from Physical Simulations of Fragments. *PLoS Comput Biol* 5(2): e1000281. doi:10.1371/journal.pcbi.1000281
32. Forster M J. 2001. Molecular modelling in structural biology. *Micron* 33: 365-384
33. Carloni P, Rothlisberger U, and Parrinello M. 2002. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* 35:455–464.
34. Marqusee S, Robbins VH, Baldwin R. L. 1989. Unusually stable helix formation in short alanine-based peptides. *Proc Natl Acad Sci U S A* 86: 5286–5290.
35. Blanco FJ, Rivas G, Serrano L .1994. A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat Struct Biol* 1: 584–590
36. Searle MS, Williams DH, Packman LC. 1995. A short linear peptide derived from the N-

- terminal sequence of ubiquitin folds into a water-stable non- native beta-hairpin. *Nat Struct Biol* 2: 999–1006
37. Patricia Romero Rubens López and Ernesto García. 2007 .Key role of amino acid residues in the dimerization and catalytic activation of the autolysin LytA, an important virulence factor in *Streptococcus pneumoniae*. *J Biol Chem* 282(24):17729-37
 38. Maestro B, Santiveri C. M, Jimenez A and.Sanz J. M. 2010. Structural autonomy of a b-hairpin peptide derived from the pneumococcal choline-binding protein LytA. *Protein Engineering, Design & Selection* vol. 24 no. 1 –2 pp. 113– 122
 39. Tomasz A. .1967. Choline in the Cell Wall of a Bacterium: Novel Type of Polymer-Linked Choline in *Pneumococcus*. *Science* 157, 694 – 697
 40. Fernandez-Tornero C., Lopez R, Garcia E, Gimenez-Gallego G. and Romero A. 2001. A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. *Nat. Struct. Biol.*, 8, 1020–1024.
 41. Hutchinson E.G. and Thornton J.M. 1994. A revised set of potentials for α -turn formation in protein. *Protein Sci.*, 3, 2207– 2216.
 42. Fernandez-Tornero C, Garcia E, Lopez R, Gimenez-Gallego G, Romero A. 2002. Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer, *J.Mol.Biol.* 321:163
 43. http://users.auth.gr/paloura/nucleation_text.pdf
 44. Hiinenberger P. H, Mark A.E, and van Gunsteren W. F. 1995. Computational Approaches to Study Protein Unfolding: Hen Egg White Lysozyme as a Case Study. *Proteins*. Mar 21, 3:196– 213.
 45. Metropolis N, Rosenbluth A. W, Rosenbluth M. N, Teller A. H and Teller E. 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21 1087-1092.
 46. Alder B. J and Wainwright T. E. Phase transition for a hard sphere system. 1957. *J. Chem. Phys.*, 27 1208-1209
 47. Rahman A, Stillinger F. H. 1971. Molecular dynamics study of liquid water. *J. Chem. Phys.* 55:3336–59
 48. McCammon J. A, Gelin B. R, Karplus M. 1977. Dynamics of folded proteins. *Nature* 267:585–90

49. J. M. Haile. 1992. *Molecular Dynamics Simulation: Elementary Methods* Wiley-Interscience
50. Tolman R. C. 1938. *The Principles of Statistical Mechanics* Courier Dover Publications
51. Frenkel D and Smit B. 1996. *Understanding Molecular Simulation, From Algorithms to Applications*
52. Haile j.M. 1992. *Introduction to Molecular Dynamics Simulation*
53. Jabbarzadeh A and Tanner R. I. 1998. *Molecular Dynamics Simulation and its Application to Nano-Rheology*. *Journal of Non-Newtonian Fluid Mechanics*, 77, 53-78
54. Scheraga HA, Khalili M, Liwo A. 2007. *Protein Folding Dynamics: Overview of Molecular Dynamics Techniques*. *Annu Rev Phys Chem*. 58:57-83.
55. W. F van Gunsteren. 1994 *Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry*. *Annu. Rev. Biophys. Biomol. Struct.* .23. 847 63
56. Freddolino P. L, Harrison C. B, Yanxin Liu and Schulten K. 2010. *Challenges in protein folding simulations: Timescale, representation, and analysis*. *Nat Phys* 6(10): 751–758
57. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood M. P, Dror R. O, et al. 2012. *Systematic Validation of Protein Force Fields against Experimental Data*. *PLoS ONE* 7(2)
58. Perez A, Marchan I, Svozil D, Sponer J, Cheatham III T. E, Laughton C. A and Orozco M. 2007. *Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of a/g Conformers*. *Biophys J*. 92(11): 3817–3829.
59. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C. 2006. *Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters*. *PROTEINS: Structure, Function, and Bioinformatics* 65:712–725
60. http://www.ch.embnet.org/MD_tutorial/pages/MD.Part1.html
61. Mark K. P and Nilsson L. 2001. *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298*. *J. Phys. Chem.*105, 9954-9960
62. Radajewski J and Eadline D. 1998. *Beowulf HOWTO* Jacek Radajewski and Douglas Eadline v1.1.1
63. <http://utopia.duth.gr/~glykos/norma.html>
64. www.wikipedia.org
65. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K and Schulten K. 1999. *J. Comput. Phys.* 151, 283-312.
66. <http://www.ks.uiuc.edu/Research/namd/>

67. Case D. A, Cheatham 3rd T. E, Darden T, Gohlke H, LuoR, Merz Jr K. M, Onufrie A, Simmerling C, Wang B and Woods R. J. 2005. *J. Comput. Chem.* 26, 1668-1688.
68. <http://compbio.biosci.uq.edu.au/education/mdcourse/MDpract.html>
69. Berendensen H. J. V. 2007. Simulating the physical world, Hierarchical modeling from quantum mechanics to fluid dynamics
70. Frishman D and Argos P. 1995. Knowledge-Based Protein Secondary Structure Assignment. *Proteins: Structure, Function and Genetics* 23:556-579
71. Heinig M and Frishman D. 2004, STRIDE: a web server for secondary structure assignment from atomic coordinates. *Nucleic Acids Research*. Vol.32. Web server Issue
72. Smith L. I. 2002 A tutorial on Principal Component Analysis
73. Altis A, Otten M, Nguyen P. H, Hegger R and Stock G. 2007. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Phys. Chem* 126, 244111
74. Altis A, Otten M, Nguyen P. H, Hegger R and Stock G. 2008. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J. Phys. Chem* 128, 245102
75. Glykos N. M. 2006. *J. Comp. Chem.* 27, 1765-1768
76. Brooks III C. L, Gruebele M, Onuchic J. N, Wolynes P. G. 1998. Chemical physics of protein folding. *Proc. Natl. Acad. Sci. USA* Vol. 95, pp. 11037–11038
77. MacKerell Jr., A. D., Bashford D, ., Karplus M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
78. Jorgensen W, Maxwell D. S, Tirado-Rives J. 1996. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* 118: 11225– 11236
79. van Gunsteren W. F, Berendsen H. J. C. 1987. *Groningen Molecular Simulation (GROMOS) Library Manual*; BIOMOS bv, University of Groningen, Groningen.