



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ

ΤΜΗΜΑ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΓΕΝΕΤΙΚΗΣ

Διπλωματική Εργασία

**“Θεώρηση τριών δυναμικών πεδίων AMBER:
διεξαγωγή προσομοιώσεων μοριακής δυναμικής
στη μίνι-πρωτεΐνη CLN025”**

Σεραφείμ Αθανασία – Παναγιώτα
1015

Επιβλέπων Καθηγητής:
Νικόλαος Μ. Γλυκός

Αλεξανδρούπολη, 2015

Ευχαριστίες

Πάνω απ' όλους θα ήθελα να ευχαριστήσω ειλικρινά τον επιβλέποντα καθηγητή μου Νικόλαο Μ. Γλυκό για την καθοδήγησή του κατά την εκπόνηση αυτής της εργασίας. Μου μεταλαμπάδευσε γνώσεις και ιδέες, μου αφιέρωσε τον πολύτιμο χρόνο του και έδειξε μεγάλη υπομονή.

Επίσης θα ήθελα να ευχαριστήσω την οικογένειά μου, τους φίλους μου, τους συναδέλφους και καθηγητές μου αλλά και κάθε άλλο άτομο που συνάντησα κατά τις σπουδές μου. Με ποικίλους τρόπους συνέβαλαν στην δημιουργία αναμνήσεων και στην ωρίμανσή μου ως άτομο.

“Motivation? What more do you need than pride?”

-Yamaguchi Tadashi

Περιεχόμενα

Περίληψη	4
Abstract	4
Κεφάλαιο 1: Εισαγωγή	5
1.1 Το πρόβλημα της αναδίπλωσης των πρωτεϊνών	6
1.2 Αναδίπλωση πρωτεϊνών και πεπτιδίων	7
1.3 Εφαρμογή προσομοιώσεων μοριακής δυναμικής για τη μελέτη της αναδίπλωσης	10
1.4 Το πεπτίδιο CLN025	11
1.5 Στόχος της παρούσας μελέτης	14
Κεφάλαιο 2: Προσομοιώσεις μοριακής δυναμικής	15
2.1 Εισαγωγή	16
2.2 Μοριακές αλληλεπιδράσεις και δυναμική ενέργεια	16
2.3 Το δυναμικό πεδίο Amber	22
Κεφάλαιο 3: Μέθοδοι	25
3.1 Τεχνικά χαρακτηριστικά του υπολογιστικού συστήματος	26
3.2 Το πρωτόκολλο των προσομοιώσεων	27
Κεφάλαιο 4: Αποτελέσματα	33
4.1 Ανάλυση των προσομοιώσεων με βάση το RMSD	34
4.2 Ανάλυση των προσομοιώσεων με βάση τη μέθοδο της Ανάλυσης Κύριων Συνιστωσών	39
4.2.1 Η μέθοδος	39
4.2.2 Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό ILDN	40
4.2.3 Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό NMR	47
4.2.4 Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό STAR	52
4.3 Ανάλυση των προσομοιώσεων με βάση τις “native contacts”	59
4.4 Σύγκριση με τα πειραματικά δεδομένα	65
4.4.1 Εισαγωγή	65
4.4.2 Υπολογισμός και σύγκριση των NOEs	66
Κεφάλαιο 5: Συμπεράσματα	72
Βιβλιογραφία	74

Περίληψη

Μικρά πεπτίδια με καλά καθορισμένες δομές και χρόνους αναδίπλωσης της τάξεως των μικρο-δευτερολέπτων μπορούν να παρέχουν ένα εξαιρετικό σύστημα για τη δοκιμή και τη σύγκριση δυναμικών πεδίων. Σε αυτή τη μελέτη προσομοιώσαμε την αναδίπλωση μιας μίνι-πρωτεΐνης, του CLN025, με τη χρήση τριών δυναμικών πεδίων Amber, των ff99sb-ildn, ff99sb-ildn-nmr και ff99sb*-ildn. Το CLN025 είναι ένα συνθετικό μόριο που αποτελείται από δέκα κατάλοιπα και υιοθετεί τη δομή μιας β-φουρκέτας όπως αποδείχθηκε από πειράματα κρυσταλλογραφίας ακτίνων-Χ και φασματοσκοπίας NMR. Η ανάλυση των τροχιακών που παρήχθησαν φανέρωσε τις διαφορές που παρήγαγε η εφαρμογή του κάθε δυναμικού πεδίου. Περαιτέρω σύγκριση με τα πειραματικά δεδομένα και συγκεκριμένα με τις αποστάσεις λόγω NOE κατέδειξε ότι η παραλλαγή ff99sb*-ildn είναι αυτή που αποδίδει καλύτερα από τις άλλες δύο.

Abstract

Small peptides with well-defined 3D structures and folding times on the microsecond time scale are excellent systems to test and compare force fields. In this study we simulate the folding of a mini-protein, CLN025, with three Amber force fields: ff99sb-ildn, ff99sb-ildn-nmr and ff99sb*-ildn. CLN025 is a synthetic molecule which consists of ten residues and forms a beta hairpin as proven through X-ray crystallography and NMR spectroscopy. Analysis of the trajectories illuminates the differences that the application of each force field produced. Further comparison against the raw experimental data, specifically the NOE distances, indicates that the ff99sb*-ildn variant outperforms the other two.

Κεφάλαιο 1

Εισαγωγή

1.1 | Το πρόβλημα της αναδίπλωσης των πρωτεϊνών

Με ποιο τρόπο αποκτούν οι πρωτεΐνες τη φυσική τους δομή; Πώς η αμινοξική αλληλουχία καθορίζει την τρισδιάστατη δομή της πρωτεΐνης; Πώς επιτυγχάνεται τόσο γρήγορα η αναδίπλωση; Είναι δυνατόν να προβλεφθεί η δομή μιας νέας πρωτεΐνης γνωρίζοντας μονάχα την αλληλουχία της; Το πρόβλημα της αναδίπλωσης των πρωτεϊνών υφίσταται για τουλάχιστον 50 χρόνια. Αν και αρχικά φάνταζε δυσεπίλυτο, τα τελευταία χρόνια έχει σημειωθεί σημαντική πρόοδος στην προσπάθεια επίλυσής του.

Το πρώτο ορόσημο τέθηκε από τον Christian Anfinsen το 1961 όταν με τα πειράματά του στη ριβονουκλεάση προέκυψε η “θερμοδυναμική υπόθεση” σύμφωνα με την οποία η φυσική δομή μιας πρωτεΐνης είναι εκείνη η διαμόρφωση στην οποία το μόριο φέρει το ολικό ελάχιστο ελεύθερης ενέργειας. Επομένως σε ένα δεδομένο περιβάλλον όλες οι απαραίτητες πληροφορίες για να αναδιπλωθεί μια πρωτεΐνη στη φυσική τρισδιάστατη δομή της βρίσκονται στην αλληλουχία των αμινοξέων της (*Anfinsen et al. 1961*). Επόμενο ορόσημο αποτέλεσε η παρατήρηση του Levinthal ότι θα ήταν αδύνατο μια πρωτεΐνη να αναδιπλωθεί στη φυσική της δομή μεταβαίνοντας από τυχαίες διαμορφώσεις έως ότου καταλήξει στην φυσική καθώς αυτό θα απαιτούσε χρόνο μεγαλύτερο από αυτόν της ύπαρξης του σύμπαντος. Οι πρωτεΐνες όμως αναδιπλώνονται εντός δευτερολέπτων ή λιγότερο, πράγμα που σημαίνει ότι θα πρέπει να υπάρχουν ειδικά μονοπάτια αναδίπλωσης (*Levinthal, 1968*). Η ερμηνεία σε αυτό το παράδοξο δόθηκε όταν αποδείχθηκε πως ακόμη και η μικρότερη ενεργειακή “προτίμηση” στο δυναμικό αναδίπλωσης επιταχύνει σε βιολογικώς παρατηρήσιμες τιμές τον χρόνο εύρεση μιας σταθερής διαμόρφωσης (*Zwanzig et al, 1992*).

Ως απάντηση στην υπόθεση ύπαρξης μονοπατιών αναδίπλωσης διενεργήθηκαν αρκετές μελέτες και διατυπώθηκαν δύο κύριες προσεγγίσεις επί του θέματος. Στο ένα άκρο κυριάρχησε η άποψη ότι επρόκειτο για μια σταδιακή αναδίπλωση όπου πρώτα σχηματίζονται ταχέως και σε τοπικό επίπεδο δευτεροταγείς διαμορφώσεις οι οποίες λειτουργούν ως ικρίωμα για την απόκτηση τεταρτοταγούς δομής. Στο μοντέλο αυτό δόθηκε η ονομασία «framework model» και περιελάμβανε δύο μηχανισμούς. Ο πρώτος μηχανισμός «diffusion-collision» προέβλεπε ότι ο σχηματισμός των δευτεροταγών δομών ακολουθείται από την διάχυση, σύγκρουση και τελικώς συνένωση τους με στόχο να σχηματίσουν την τεταρτοταγή δομή (*Karplus, 1976*). Ο δεύτερος μηχανισμός «nucleation» υποστήριζε ότι πρώτα σχηματίζεται ο πυρήνας της δομής και ακολουθεί η γρήγορη διάδοσή της (*Wetlaufer,*

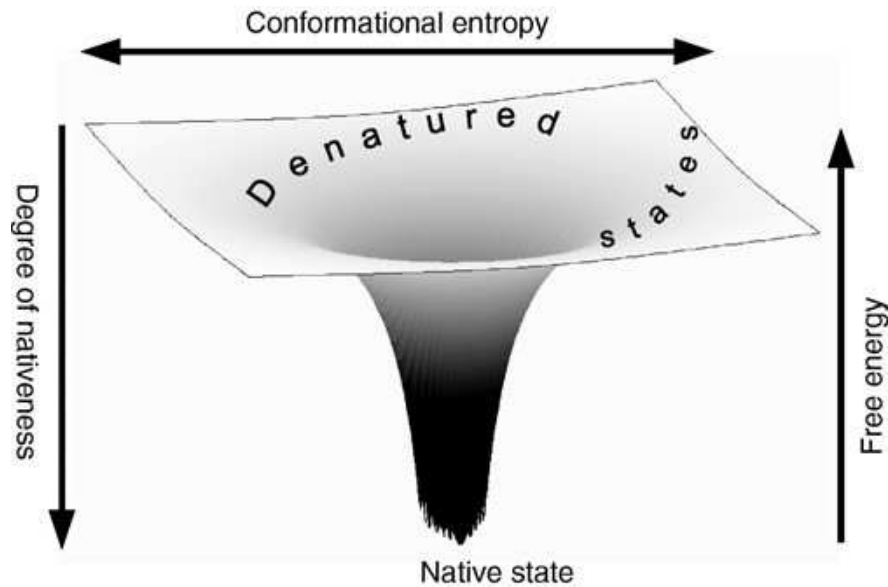
1973). Στο άλλο άκρο βρισκόταν ο μηχανισμός «hydrophobic collapse» σύμφωνα με τον οποίο η κινητήριος δύναμη της αναδίπλωσης είναι το υδροφοβικό φαινόμενο. Στην πρωτεΐνη κατά τα πρώτα στάδια της αναδίπλωσης σχηματίζονται ταχέως δομές λόγω των υδροφοβων αλληλεπιδράσεων και έπειτα ακολουθούν οι δευτεροταγείς διαμορφώσεις επιταχύνοντας έτσι την αναδίπλωση (Ptitsyn, 1987).

1.2 | Αναδίπλωση πρωτεϊνών και πεπτιδίων

Σε μία πρωτεΐνη ο εσωτερικός πυρήνας σχηματίζεται από υδροφοβα κατάλοιπα τα οποία συγκρατούνται μεταξύ τους χάρις στις δυνάμεις van der Waals ενώ η επιφάνειά της φέρει φορτισμένα και πολικά κατάλοιπα. Η φυσική δομή μιας πρωτεΐνης είναι ενεργειακά σταθερή γεγονός που επιτυγχάνεται από τέσσερις κύριους συντελεστές: το υδροφοβικό φαινόμενο, την ενέργεια των δεσμών υδρογόνου, την ενέργεια των ηλεκτροστατικών αλληλεπιδράσεων και την εντροπία που οφείλεται στην περιορισμένη κίνηση της κύριας και των πλευρικών αλυσίδων.

Ποια είναι όμως η δύναμη που κατευθύνει την αναδίπλωση; Κατ' αρχάς οι ηλεκτροστατικές αλληλεπιδράσεις μεταξύ φορτισμένων πλευρικών ομάδων είναι απίθανο να κατευθύνουν την αναδίπλωση καθώς τα φορτισμένα κατάλοιπα είναι ελάχιστα στις πρωτεΐνες. Επιπλέον το pH και η συγκέντρωση αλάτων δεν επηρεάζουν τη σταθερότητα της δομής όπως επίσης και οι σημειακές μεταλλάξεις οι οποίες έχουν συνήθως μικρό αντίκτυπο στη διαμόρφωση. Διαπιστώθηκε επίσης ότι οι δεσμοί υδρογόνου διαδραματίζουν σημαντικό ρόλο λόγω του ότι αποτελούν την κύρια συνιστώσα όλων των δευτεροταγών δομών (Dill, 1990). Καθώς όμως οι διαφορές μεταξύ δύο πρωτεϊνών έγκειται στις πλευρικές τους ομάδες και όχι τόσο στον πεπτιδικό σκελετό, η πληροφορία για την αναδίπλωση θα πρέπει να κωδικοποιείται εκεί. Πράγματι μελέτες απέδειξαν ότι οι υδροφοβες αλληλεπιδράσεις παίζουν κύριο ρόλο στην πρωτεϊνική αναδίπλωση (Dill, 1999).

Τα τελευταία χρόνια η κατανόηση του τρόπου με τον οποίο λαμβάνει χώρα η σύνθετη διαδικασία της αναδίπλωσης έχει αυξηθεί. Έχει διατυπωθεί η αρχή του “minimal frustration” κατά την οποία οι περισσότερες αλληλεπιδράσεις που αναπτύσσονται μεταξύ των αλυσίδων αλληλοϋποστηρίζονται και δρουν συνεργατικά ώστε να οδηγήσουν σε μια χαμηλής ενέργειας δομή (Bryngelson, 1987). Αυτή η αρχή οδήγησε στη διαπίστωση ότι οι πρωτεΐνες θα πρέπει να έχουν ενεργειακά τοπία με το σχήμα χωνιού (Εικόνα 1.2.1) (Dill, 1997).

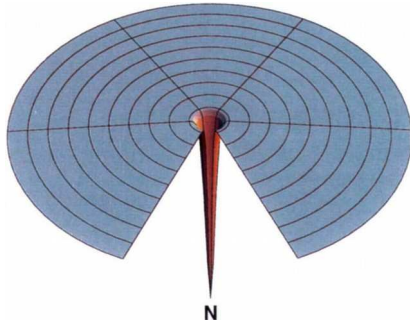


Εικόνα 1.2.1: Σχηματική απεικόνιση ενός ενεργειακού τοπίου με σχήμα χωνιού.

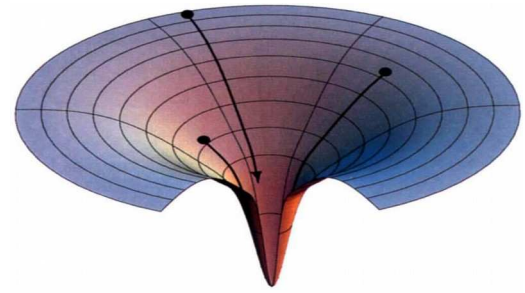
Ένα ενεργειακό τοπίο αναπαριστά την ελεύθερη ενέργεια κάθε διαμόρφωσης σε συνάρτηση με τους βαθμούς ελευθερίας όπως για παράδειγμα τις διέδρες γωνίες του πεπτιδικού σκελετού. Ο κάθετος άξονας του χωνιού αναπαριστά την εσωτερική ελεύθερη ενέργεια κάθε δεδομένης διαμόρφωσης ενώ οι οριζόντιοι άξονες αναπαριστούν τις συντεταγμένες των διαμορφώσεων. Κάθε διαμόρφωση αντιπροσωπεύεται με μία τελεία στο πολυδιάστατο ενεργειακό τοπίο. Δομές με διαμορφώσεις που ομοιάζουν εντοπίζονται κοντά η μία στην άλλη. Τα ενεργειακά τοπία μπορεί να εμφανίζουν διάφορα χαρακτηριστικά όπως λόφους όπου συναντώνται δομές με υψηλή ενέργεια και πεδιάδες που περιέχουν ενεργειακά ευνοϊκές διαμορφώσεις (**Εικόνα 1.2.2**).

Με την έναρξη της αναδίπλωσης η πρωτεΐνη τείνει να αλλάζει την διαμόρφωση της με τρόπο τέτοιο ώστε να μειώσει την ενέργειά της. Ένα βασικό μοντέλο προτείνει ότι ξεκινώντας από μια αποδιατεταγμένη διαμόρφωση μια πρωτεΐνη τυπικά οδηγείται απευθείας σε μια διαμόρφωση όπου το ένα τέταρτο περίπου των επαφών μεταξύ των ατόμων είναι παρόμοιο με τη φυσική δομή. Αυτή η κατάσταση αντιστοιχεί σε ένα ευρύ ελάχιστο ελεύθερης ενέργειας όπου οι διαμορφώσεις έχουν τη δομή μιας άναρχα διατεταγμένης σφαίρας. Έπειτα ακολουθεί μια τυχαία αναζήτηση της μεταβατικής δομής που θα οδηγήσει ταχέως στην φυσική δομή (*Dinner et al., 2000*). Αυτό το μοντέλο καταδεικνύει ότι είναι δυνατόν να βρεθεί η φυσική δομή ύστερα από διέλευση μέσω ενός μικρού μονάχα τμήματος του συνόλου των διαμορφώσεων.

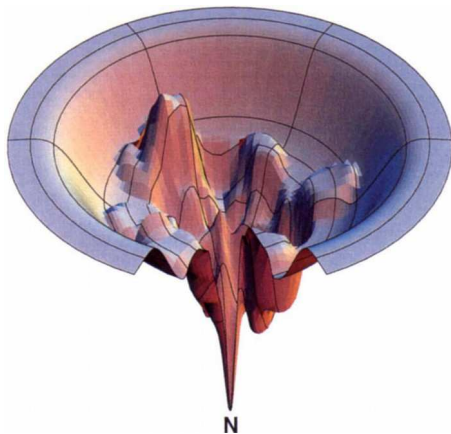
α.



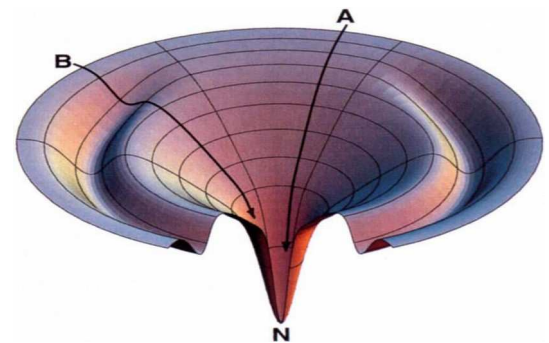
β.



γ.



δ.



Εικόνα 1.2.2: Χαρακτηριστικά ενεργειακά τοπία πρωτεϊνών. α) ενεργειακό τοπίο όπου η αναδίπλωση κυριαρχείται από την αναζήτηση διαμορφώσεων μέσω διάχυσης, β) λείο ενεργειακό πεδίο ενός ταχέως αναδιπλούμενου μορίου, γ) τραχύ ενεργειακό πεδίο γεμάτο κινητικές παγίδες και δ) ένα πεδίο στο οποίο η αναδίπλωση διέρχεται από ένα ενδιάμεσο.

Παρ' όλη την πρόοδο που έχει λάβει χώρα η γνώση γύρω από την αναδίπλωση πιο σύνθετων πρωτεϊνών παραμένει ελλιπής. Τα πεπτίδια αποτελούν το καταλληλότερο εργαλείο για την μελέτη των μηχανισμών που λαμβάνουν χώρα στα αρχικά στάδια της αναδίπλωσης των πρωτεϊνών. Αναδιπλώνονται γρήγορα και παρέχουν επίσης ένα πιο εύκολα ελεγχόμενο σύστημα σε σχέση με τις πρωτεΐνες. Με αυτό τον τρόπο τα πεπτίδια παρέχουν ένα μέσο για τον καθορισμό των δυνάμεων που επενεργούν κατά την αναδίπλωση και την σταθεροποίηση των πρωτεϊνών. Την τάση για μελέτη της πεπτιδικής αναδίπλωσης ενίσχυσαν τρεις σημαντικές επιστημονικές εξελίξεις. Καταρχάς ο σχεδιασμός πεπτιδίων που μιμούνται την πολυπλοκότητα των πρωτεϊνών αλλά είναι επαρκώς μικρά ώστε να μελετηθούν εκτενώς μέσω προσομοιώσεων. Κατά δεύτερον η ανάπτυξη ταχέων

φασματοσκοπικών μεθόδων για την μελέτη της δυναμική της πεπτιδικής αναδίπλωσης στην ίδια χρονική κλίμακα με τις υπολογιστικές προσομοιώσεις. Τέλος η δημιουργία αλγορίθμων προσομοίωσης που ξεπερνούν προβλήματα μη επαρκούς δείγματος. Όλα τα παραπάνω καθιστούν πολύ σημαντικό το να δοθεί βάση στη επικύρωση τεχνικών προσομοίωσης και δυναμικών πεδίων και στην επαλήθευση των αποτελεσμάτων των προσομοιώσεων (Gnanakaran et al., 2003).

1.3 | Εφαρμογή προσομοιώσεων μοριακής δυναμικής για τη μελέτη της αναδίπλωσης

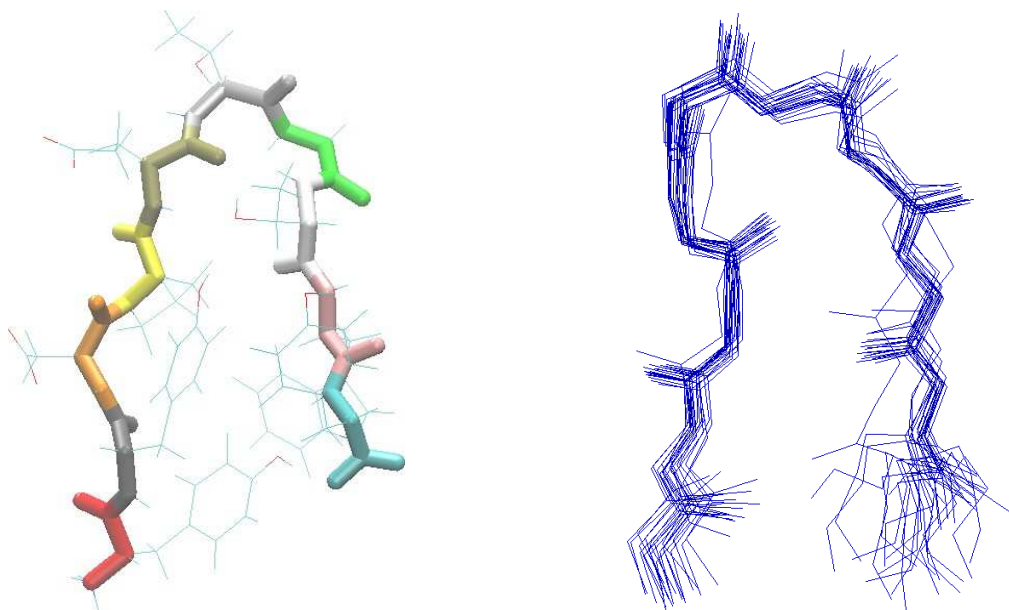
Οι προσομοιώσεις μοριακής δυναμικής έχουν μετατραπεί σε ένα εξαιρετικά χρήσιμο εργαλείο για την πρόβλεψη δομών και για την μελέτη της διαδικασίας της αναδίπλωσης. Η επιτυχημένη χρήση τους δεν κρίνεται μονάχα από την πρόβλεψη της φυσικής δομής αλλά και από την παροχή βαθύτερης κατανόησης των δυνάμεων που διέπουν τις ιδιότητες μιας πρωτεΐνης. Η χρήση τους μπορεί να μας φέρει σε θέση να διεξάγουμε ποικιλία μελετών που έχουν να κάνουν με την πρόβλεψη αλλαγών στη διαμόρφωση, έναν κύριο άξονα του σχεδιασμού φαρμάκων. Επίσης μελέτες κατανόησης της συμπεριφοράς των πρωτεϊνών σε ποικιλία συνθηκών όπως υπό την επίδραση διαλυτών, αλλαγών στο pH, αποδιατακτικών παραγόντων κλπ. Τέλος, όπως έχει αναφερθεί ήδη, μελέτες που έχουν να κάνουν με την κίνηση της πρωτεΐνης σε καταστάσεις όπως η αναδίπλωση και η ενζυματική κατάλυση.

Υπολογιστικά μοντέλα και προσομοιώσεις έχουν βελτιώσει σημαντικά την κατανόησή μας γύρω από την αναδίπλωση. Η συνεργεία μεταξύ πειράματος και θεωρίας είναι δεδομένη καθώς οι πληροφορίες που παρέχουν οι υπολογιστικές μελέτες έρχονται να συμπληρώσουν τις πειραματικές μελέτες. Από την πειραματική σκοπιά τεχνικές όπως η φασματοσκοπία μαγνητικού συντονισμού παρήγαγαν πληθώρα πληροφορίας για τη μοριακή δομή, τη δυναμική και τη λειτουργία των πρωτεϊνών. Από τη σκοπιά της θεωρίας οι προσομοιώσεις μοριακής δυναμικής με τη χρήση δυναμικών πεδίων είναι σε θέση να καταγράψουν τη δομή και την κίνηση των βιομορίων με μικροσκοπική λεπτομέρεια. Οι προσομοιώσεις συγκεκριμένα βοηθούν στον εντοπισμό και την πρόβλεψη ενδιάμεσων του μονοπατιού αναδίπλωσης, παρέχουν τη δυνατότητα πρόβλεψης του ρυθμού αναδίπλωσης και σε ορισμένες περιπτώσεις είναι σε θέση να προβλέψουν την τελική αναδιπλωμένη δομή.

1.4 | Το πεπτίδιο CLN025

Οι Honda et al. κατασκεύασαν ένα συνθετικό μόριο το οποίο αποτελείται από 10 φυσικώς απαντώμενα αμινοξέα και το ονόμασαν CLN025. Το πεπτίδιο αυτό σχεδιάστηκε με βάση την αλληλουχία της chignolin, μιας πρωτεΐνης 10 αμινοξέων που είχαν σχεδιάσει οι ίδιοι προηγουμένως. Τα κατάλοιπα Gly1 και Gly10 της chignolin αντικαταστάθηκαν από διάφορα άλλα αμινοξέα και η παραλλαγή που επέδειξε την μεγαλύτερη σταθερότητα ήταν αυτή του CLN025 με αλληλουχία YYDPETGTWY (Honda et al., 2008).

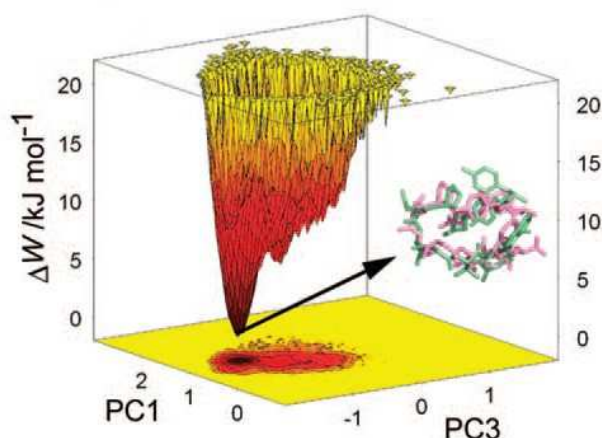
Η κρυσταλλική δομή του μορίου προσδιορίστηκε σε διακριτικότητα ίση με 1,11Å και αποκαλύπτει ότι το CLN025 σχηματίζει μία β-φουρκέτα. Η δομή σταθεροποιείται χάρη σε 6 ενδομοριακούς δεσμούς υδρογόνου και μία γέφυρα άλατος. Η εφαρμογή της φασματοσκοπίας NMR παρήγαγε 20 μοντέλα για τα οποία το bb-RMSD από την κρυσταλλική δομή ήταν 1.75Å. Επομένως οι δομές στο διάλυμα στους 298K διατηρούν παρόμοια διαμόρφωση (Εικόνα 1.4.1). Οι διαφορές που παρατηρούνται είναι η απουσία της γέφυρας άλατος και του δεσμού υδρογόνου μεταξύ των καταλοίπων Tyr1 και Tyr10.



Εικόνα 1.4.1: Η κρυσταλλική δομή του CLN025 και τα 20 διαθέσιμα μοντέλα NMR.

Επίσης πραγματοποιήθηκαν προσομοιώσεις μοριακής δυναμικής οι οποίες δείχνουν ότι το μόριο αναδιπλώνεται αρκετές φορές *in silico* και τελικά φτάνει σε μία κατάσταση ισορροπίας παράγοντας μια επιφάνεια ελεύθερης ενέργειας σε σχήμα χωνιού. Συμφωνεί

δηλαδή με την αρχή του minimum frustration. Η δομή που εντοπίζεται στη βάση του χωνιού συμφωνεί με τα πειραματικά δεδομένα συμπεριλαμβανομένης της διευσθέτησης των πλευρικών αλυσίδων (Εικόνα 1.4.2).



Εικόνα 1.4.2: Η επιφάνεια ελεύθερης ενέργειας του CLN025

Το CLN025 ως πεπτίδιο διαθέτει ορισμένα χαρακτηριστικά τα οποία το κάνουν να ξεχωρίζει. Η μελέτη των Honda et al. κατέδειξε την καλά καθορισμένη τρισδιάστατη δομή του πεπτιδίου. Η δομή αυτή διαθέτει δυναμική συμπεριφορά που περιλαμβάνει την αναδιάταξη των πλευρικών αλυσίδων και την αναστρέψιμη αναδίπλωση. Το CLN025 είναι το μικρότερο ευθύγραμμο πολυπεπτίδιο το οποίο αποκτά μια μοναδική δομή σε υδατικά διαλύματα και μπορεί να κρυσταλλωθεί. Όλα τα παραπάνω το καθιστούν ικανό να θεωρηθεί πρωτεΐνη. Μάλιστα το χαρακτηρίζουν ως “ιδανική πρωτεΐνη” παρόλο που είναι ένα μόριο το οποίο δεν απαντάται στη φύση.

Το μόριο έπειτα από τη σύνθεσή του μελετήθηκε εκτενώς και παρουσιάζονται εδώ τα σημαντικότερα ευρήματα που παρήχθησαν. Το 2010 έπειτα από την εφαρμογή ECD, VCD και προσομοιώσεων μοριακής δυναμικής αποδείχθηκε ότι το μόριο υιοθετεί μια σταθερή δομή β-φουρκέτας σε περιβάλλοντα όπως νερό, TFE, MeOH και DMSO (Hatfield et al., 2010). Η ίδια ομάδα πραγματοποίησε επίσης προσομοιώσεις μοριακής δυναμικής σε διάφορες θερμοκρασίες και το CLN025 κατέληξε και πάλι σε σταθερή διαμόρφωση β-φουρκέτας. (Hatfield et al., 2010).

Έπειτα, το 2011, πραγματοποιήθηκε μια υπολογιστική μελέτη στην οποία το πεπτίδιο μελετήθηκε με τρία διαφορετικά πρωτόκολλα. Το πρώτο περιλάμβανε την αναλυτική παρουσία διαλύτη και χρησιμοποιήθηκε στην προσομοίωση το δυναμικό πεδίο parm99SB. Τα

υπόλοιπα είχαν μη αναλυτική παρουσία διαλύτη και χρησιμοποιήθηκαν τα δυναμικά πεδία parm94, parm96 και parm99SB. Παρατηρήθηκαν διαφορές στα αποτελέσματα από τα διαφορετικά περιβάλλοντα γεγονός που ανέδειξε την ισορροπία μεταξύ ενδο- και διαμοριακών δυνάμεων ως την καθοριστική δύναμη πίσω από την υιοθέτηση μιας διαμόρφωσης σε διαλύτη (*Rodriguez et al., 2011*).

Την ίδια χρονιά πραγματοποιήθηκε μια μελέτη η οποία απέδειξε ότι ακόμη και υψηλές συγκεντρώσεις από τους αποδιατακτικούς παράγοντες ουρία και GdmCl δεν κατάφεραν να επιδράσουν και να διαταράξουν ικανοποιητικά τη δομή β-φουρκέτας. Αυτός ήταν ένας από τους λόγους που έκαναν αυτή την ομάδα επιστημόνων να αμφισβητήσει τον ισχυρισμό των Honda et al. ότι το CLN025 είναι μια μίνι-πρωτεΐνη. Επίσης το γεγονός ότι από το μόριο λείπει μια καλά καθορισμένη ταταρτοταγής δομή και ότι δεν διαθέτει αυξημένη θερμοχωρητικότητα κοντά στην θερμοκρασία τήξης συνέβαλλαν επιπλέον σε αυτήν την πρόταση (*Hatfield et al. 2011*).

Έπειτα πραγματοποιήθηκε μια μελέτη στην οποία έγιναν προσομοιώσεις μοριακής δυναμικής που ερευνούσαν την αναδίπλωση του μορίου. Παρατηρήθηκε ότι η διαδικασία αναδίπλωσης διαφέρει από αυτή των υπολοίπων πεπτιδίων. Το φαινόμενο αυτό κατέδειξε ότι οι ηλεκτροστατικές αλληλεπιδράσεις μεταξύ των φορτισμένων άκρων παίζουν ένα πολύ σημαντικό ρόλο στη σταθερότητα της β-φουρκέτας (*Zhao et al., 2012*).

Η διαδικασία με την οποία το μόριο αναδιπλώνεται μελετήθηκε εκ νέου και φάνηκε ότι το μόριο αναδιπλώνεται εξαιρετικά γρήγορα καθώς διαθέτει ένα μικρό φράγμα ελεύθερης ενέργειας. Επίσης η αναδίπλωση του δεν περιλαμβάνει τη συνήθη διαδικασία των δύο καταστάσεων αλλά είναι μια πιο ετερογενής διαδικασία (*Davis et al., 2012*). Το 2013 η ομάδα επέκτεινε τη μελέτη της πραγματοποιώντας ένα πείραμα στο οποίο εισήγαγε το CLN025 σε κάθε βρόχο του FBP28 WW domain και παρατήρησε ότι η δυναμική αναδίπλωσης παραμένει η ίδια στα πλαίσια μια μεγαλύτερης πρωτεΐνης. Άρα τέτοιου είδους subdomains μπορούν να χρησιμοποιηθούν ώστε να επιταχύνουν την αναδίπλωση μεγαλύτερων πρωτεϊνών (*Davis et al., 2013*)

Τέλος στην πιο πρόσφατη μελέτη πραγματοποιήθηκαν προσομοιώσεις μοριακής δυναμικής με παράγωγα των δυναμικών πεδίων Amber στα οποία οι μάζες των ατόμων είχαν μειωθεί κατά δέκα φορές. Το μόριο αναδιπλώθηκε στην φυσική του δομή γεγονός που δεν παρατηρήθηκε όταν οι προσομοιώσεις επαναλήφθηκαν με τα πρωτότυπα amberff12sb και ff14sb (*Pang, 2014*).

1.5 | Στόχος της παρούσας μελέτης

Πραγματοποιήθηκαν προσομοιώσεις μοριακής δυναμικής της αναδίπλωσης του πεπτιδίου CLN025 με τη χρήση τριών δυναμικών πεδίων της οικογένειας Amber: ff99sb-ildn, ff99sb-ildn-nmr και ff99sb*-ildn. Στόχος είναι να συγκριθούν τα αποτελέσματα των προσομοιώσεων με τα διαθέσιμα πειραματικά δεδομένα και να κριθεί η αποτελεσματικότητα των τριών δυναμικών πεδίων όπως επίσης και να αναδειχθούν οι διαφορές που προκύπτουν από την εφαρμογή του καθενός.

Κεφάλαιο 2

Προσομοιώσεις μοριακής δυναμικής

2.1 | Εισαγωγή

Οι προσομοιώσεις μοριακής δυναμικής είναι ένα βασικό εργαλείο της θεωρητικής μελέτης των βιομορίων και αποτελούν μια μέθοδο για τον υπολογισμό της συμπεριφοράς ενός μοριακού συστήματος στο χρόνο. Η εφαρμογή τους στοχεύει στην κατανόηση των ιδιοτήτων μοριακών διατάξεων όσον αφορά τη δομή αλλά και τις μεταξύ τους αλληλεπιδράσεις σε μικροσκοπικό επίπεδο. Με αυτό τον τρόπο συμπληρώνουν τις κλασσικές πειραματικές μεθόδους παρέχοντας επιπλέον πληροφορία που δεν θα μπορούσε διαφορετικά να αποκτηθεί. Έτσι δίνεται η δυνατότητα να ελεγχθεί η θεωρία με τη διεξαγωγή της προσομοίωσης στο ίδιο σύστημα αλλά και να ελεγχθεί ένα μοντέλο συγκρίνοντάς το με τα πειραματικά δεδομένα. Έτσι λοιπόν οι προσομοιώσεις μοριακής δυναμικής χρησιμοποιούνται πλέον τακτικά για τη μελέτη της δομής και της θερμοδυναμικής των βιομορίων όπως επίσης και για τον καθορισμό της δομής από πειράματα κρυσταλλογραφίας ακτίνων X και φασματοσκοπίας NMR.

Η μέθοδος είχε αναπτυχθεί ήδη από την δεκαετία του 50 από τους Alder και Wainwright (Alder, 1957) ενώ η πρώτη προσομοίωση πρωτεϊνών πραγματοποιήθηκε το 1977 με την προσομοίωση του αναστολέα της παγκρεατικής θρυψίνης βοοειδών (BPTI) (McCammon et al., 1977). Έκτοτε οι προσομοιώσεις μοριακής δυναμικής έχουν χρησιμοποιηθεί σε πληθώρα μελετών όπως στην μελέτη διαλυτοποιημένων πρωτεϊνών, συμπλόκων πρωτεϊνών-DNA και λιπιδικών συστημάτων.

2.2 | Μοριακές αλληλεπιδράσεις και δυναμική ενέργεια

Μια προσομοίωση μοριακής δυναμικής ουσιαστικά αποτελεί μία βήμα-προς-βήμα επίλυση της κλασσικής εξίσωσης της κίνησης. Για ένα απλό σύστημα η εξίσωση αυτή έχει ως εξής:

$$F_i = m_i \cdot a_i$$

όπου F_i είναι η δύναμη που ασκείται σε ένα σωματίδιο i , m_i η μάζα του και a_i η επιτάχυνση του σωματιδίου. Γνωρίζοντας την δύναμη F_i είναι δυνατόν να υπολογιστεί η επιτάχυνση κάθε ατόμου στο σύστημα. Έπειτα ολοκληρώνοντας τις εξισώσεις της κίνησης προκύπτει ένα τροχιακό το οποίο περιγράφει τις θέσεις, τις ταχύτητες και τις επιταχύνσεις των σωματιδίων καθώς μεταβάλλονται με το χρόνο. Για την εξαγωγή ενός τροχιακού απαιτείται η γνώση των

αρχικών συντεταγμένων των ατόμων, μία αρχική κατανομή των ταχυτήτων και η επιτάχυνση. Για το σκοπό αυτό θα πρέπει να υπολογιστούν οι δυνάμεις F_i που δρουν στα άτομα και αυτό επιτυγχάνεται μέσω της δυναμικής ενέργειας. Η δύναμη F_i μπορεί να εκφραστεί ως βαθμίδωση της δυναμικής ενέργειας:

$$F_i = m_i (dv_i/dt) = m(d^2r_i/dt^2) = - \nabla_i E$$

Για τον υπολογισμό της δυναμικής ενέργειας ενός συστήματος ατόμων έχουν αναπτυχθεί τα δυναμικά πεδία (force fields) τα οποία αποτελούνται από μια συνάρτηση και ορισμένες παραμέτρους. Είναι σημαντικό να σημειωθεί ότι τα δυναμικά πεδία είναι εμπειρικά, πράγμα που σημαίνει ότι δεν υπάρχει ένα ιδανικό σετ συναρτήσεων και παραμέτρων. Η βασική συνάρτηση αποτελείται από δύο σκέλη. Το πρώτο σκέλος αφορά στις αλληλεπιδράσεις μεταξύ ατόμων που συνδέονται μέσω ομοιοπολικών δεσμών και έχει να κάνει με την παραμόρφωση δεσμών, τις αλλαγές στην γεωμετρία των γωνιών και την περιστροφή γύρω από δίεδρες γωνίες. Από την άλλη μεριά το δεύτερο σκέλος σχετίζεται με ηλεκτροστατικές αλληλεπιδράσεις και αλληλεπιδράσεις van der Waals. Η δυναμική ενέργεια του συστήματος επομένως αναπαρίσταται ως (MacKeller et al., 1998):

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

Ο όρος E_{bonded} υπολογίζεται ως το άθροισμα των παρακάτω επιμέρους όρων:

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

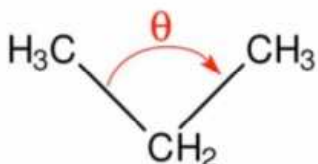
Ο πρώτος όρος είναι ένα αρμονικό δυναμικό το οποίο αναπαριστά την αλληλεπίδραση μεταξύ ζευγών ατόμων στα οποία τα άτομα χωρίζονται από έναν ομοιοπολικό δεσμό.



$$E_{\text{bond-stretch}} = K_b (r - r_0)^2$$

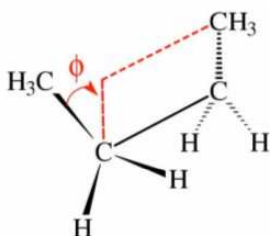
Ο όρος αυτός δίνει την ενέργεια ενός δεσμού ως την συνάρτηση της μετατόπισης από το ιδανικό μήκος r_0 το οποίο αποκτάται όταν όλοι οι υπόλοιποι όροι της συνάρτησης είναι μηδέν. Η σταθερά K_b σχετίζεται με το σθένος του δεσμού και είναι γενικά υψηλή

υποδεικνύοντας ότι απαιτείται ένα μεγάλο ποσό ενέργειας για να μεταβληθεί το μήκος ενός χημικού δεσμού. Οι τιμές τόσο του r_0 όσο και του K_b είναι ειδικές για κάθε ζεύγος ατόμων. Ο δεύτερος όρος αναπαρίσταται επίσης από ένα αρμονικό δυναμικό και αφορά στη μεταβολή των γωνιών θ ως προς την ιδανική τιμή θ_0 . Σε αυτή την περίπτωση επίσης οι τιμές των θ_0 και K_θ εξαρτώνται από τον χημικό τύπο του ζεύγους των ατόμων. Η τιμή της σταθεράς K_θ είναι σαφώς μικρότερη λόγω του ότι απαιτείται λιγότερη ενέργεια για να αποκλίνει η γωνία του δεσμού από την τιμή αναφοράς.



$$E_{\text{angle-bend}} = K_\theta (\theta - \theta_0)^2$$

Ο τελευταίος όρος αναπαριστά τη δυναμική ενέργεια του συστήματος ως συνάρτηση της περιστροφής γύρω από κάθε δίεδρη γωνία. Η ενέργεια που απαιτείται σε αυτή την περίπτωση είναι σαφώς μικρότερη απ' ότι στις δύο προηγούμενες περιπτώσεις. Το δυναμικό είναι περιοδικό, αφορά άτομα που διαχωρίζονται από τρεις δεσμούς και αναπαρίσταται ως συνημιτονοειδής συνάρτηση.



$$E_{\text{rotate-along-bond}} = K_\phi [1 + \cos(n\phi - \delta)]$$

Αξίζει να αναφερθεί ότι ο διαχωρισμός στους παραπάνω επιμέρους όρους δεν έχει κάποια βάση στη κβαντική φύση των ενδο- και διαμοριακών αλληλεπιδράσεων. Αποτελεί όμως μια χρήσιμη προσέγγιση που παρέχει πληροφορίες ως προς τους καθοριστικούς παράγοντες της δομής και της δυναμικής του μορίου.

Ο όρος $E_{\text{non-bonded}}$ αποτελείται από δύο επιμέρους όρους και αναφέρεται σε ζεύγη ατόμων που διαχωρίζονται από τρεις ή περισσότερους δεσμούς ή μεταξύ ατόμων σε διαφορετικά άτομα.

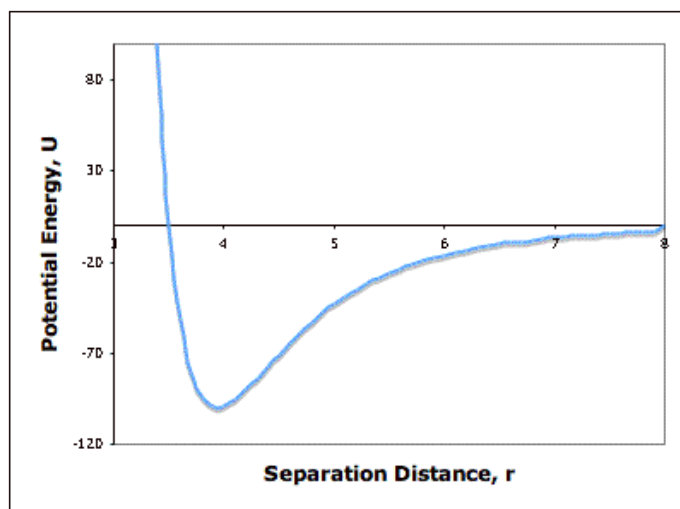
$$E_{\text{non-bonded}} = E_{\text{van-der-Waals}} + E_{\text{electrostatic}}$$

Ο πρώτος όρος, η ενέργεια που προκύπτει από τις αλληλεπιδράσεις van der Waals ονομάζεται

αλλιώς και δυναμικό Lennard-Jones και δίνεται από τη σχέση:

$$E_{\text{van-der-Waals}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6\right]$$

Στην **Εικόνα 2.2.1** δίνεται η γραφική αναπαράσταση της παραπάνω συνάρτησης. Η αλληλεπίδραση van der Waals μεταξύ δύο ατόμων προκύπτει από μια ισορροπία ελκτικών και απωστικών δυνάμεων. Καθώς η απόσταση r μεταξύ δύο ατόμων μικραίνει η πιθανότητα να αλληλεπιδράσουν αυξάνεται. Τα άτομα πλησιάζουν έως ότου φτάσουν σε ένα εύρος απόστασης στο οποίο είναι πλέον δεσμευμένα, τότε η δυναμική ενέργεια μειώνεται από το μηδέν σε αρνητικές τιμές. Η μεταξύ τους απόσταση συνεχίζει να μειώνεται και φτάνουν σε ισορροπία όταν η απόσταση αντιστοιχεί στο ελάχιστο της δυναμικής ενέργειας. Αν τα άτομα πλησιάσουν περισσότερο από την απόσταση ισορροπίας τότε αρχίζουν να αναπτύσσουν απωστικές δυνάμεις και η δυναμική ενέργεια αυξάνεται ταχέως.



Εικόνα 2.2.1: Γραφική αναπαράσταση του δυναμικού Lennard-Jones

Ο δεύτερος όρος αφορά στις ηλεκτροστατικές αλληλεπιδράσεις και αναπαρίσταται με το δυναμικό Coulomb,

$$E_{\text{electrostatic}} = q_i q_j / 4\pi\epsilon_0 \epsilon r_{ij}$$

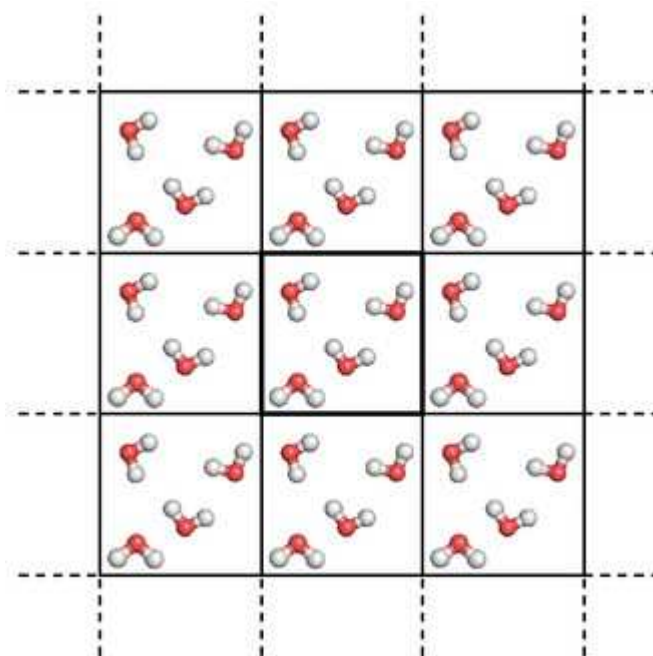
όπου ϵ είναι η διηλεκτρική σταθερά του μέσου και r η απόσταση μεταξύ των δύο ατόμων με φορτία q_i και q_j .

Ο υπολογισμός των ηλεκτροστατικών αλληλεπιδράσεων είναι ένα κομβικό σημείο των προσομοιώσεων μοριακής δυναμικής. Οι ηλεκτροστατικές αλληλεπιδράσεις είναι μεγάλης εμβέλειας και έχουν αναπτυχθεί αρκετές προσεγγίσεις για τον υπολογισμό τέτοιου είδους αλληλεπιδράσεων. Στην απλούστερη περίπτωση οι αλληλεπιδράσεις περικόπτονται έπειτα από μια συγκεκριμένη απόσταση, συνήθως 0,8 με 1,4nm. Η εισαγωγή όμως ενός τέτοιου ορίου οδηγεί στη γένεση τεχνουργημάτων (*Brooks et al., 1985*). Αυτό μπορεί να μειωθεί με τη χρήση “shift” και “switch” συναρτήσεων οι οποίες επιδρούν στο δυναμικό με τέτοιο τρόπο ώστε κοντά στην οριακή τιμή οι δυνάμεις να γίνονται μηδέν παρά να αποκτούν κάποια πεπερασμένη τιμή. Η προσέγγιση αυτή όμως έχει πάψει να χρησιμοποιείται. Με σκοπό να αποφευχθούν τα τεχνουργήματα γίνεται η υπόθεση ότι ένα κεντρικό άτομο σε μεγάλες αποστάσεις περιβάλλεται από ένα ομογενές περιβάλλον το οποίο περιγράφεται αναλυτικά. Γίνεται λοιπόν η υπόθεση ότι από ένα σημείο και μετά ένα φορτισμένο άτομο δεν βλέπει μεμονωμένα φορτία αλλά τη διηλεκτρική σταθερά του μέσου (*Tironi, 1995*). Αναπτύχθηκε επίσης μία τρίτη προσέγγιση με βάση την οποία το κελί της προσομοίωσης αναπαράγεται επ' αόριστον και έπειτα γίνεται χρήση των εξισώσεων Fourier για να επιλυθεί η εξίσωση Poisson για το νέο σύστημα. Συνήθως χρησιμοποιείται ο αλγόριθμος PME (Particle Mesh Ewald). Στην περίπτωση αυτή όλες οι αλληλεπιδράσεις του συστήματος αθροίζονται και συνεπώς περιλαμβάνουν τις μεγάλης-εμβέλειας ηλεκτροστατικές αλληλεπιδράσεις (*Essman et al., 1995*).

Ένα ακόμα σημαντικό κομμάτι μιας προσομοίωσης είναι ο διαλύτης, ο οποίος συνηθέστερα είναι το νερό. Το νερό ασκεί σημαντική επίδραση στη δομή και τη θερμοδυναμική των βιομορίων και κυρίως στις ηλεκτροστατικές αλληλεπιδράσεις. Έτσι λοιπόν είναι σημαντικό να συνυπολογίζεται η επίδραση που έχει ο διαλύτης και αυτό μπορεί να γίνει με δύο τρόπους. Στην πρώτη εκδοχή, αυτή του implicit solvent, τα μόρια νερού δεν περιλαμβάνονται στην προσομοίωση αλλά χρησιμοποιείται μία διηλεκτρική σταθερά η οποία εξαρτάται από την απόσταση. Στην δεύτερη εκδοχή ο διαλύτης είναι explicit, δηλαδή τα μόρια νερού είναι παρόντα στην προσομοίωση. Σε αυτή την περίπτωση απαιτείται να χρησιμοποιηθούν οριακές συνθήκες (boundary conditions) γεγονός που θα αποτρέψει την διάχυση του διαλύτη μακριά από το μόριο που προσομοιώνεται αλλά και θα δώσει τη δυνατότητα υπολογισμού μακροσκοπικών ιδιοτήτων με τη χρήση λιγοστών μόνο διαλυμένων μορίων.

Υπάρχουν πολλοί τρόποι να διαχειριστούν οι οριακές συνθήκες καθέναν από τους οποίους εφαρμόζεται ανάλογα με τη φύση της προσομοίωσης. Όσον αφορά την

προσομοίωση πρωτεϊνών χρησιμοποιούνται συνήθως οι συνθήκες περιοδικών ορίων (periodic boundary conditions) (**Εικόνα 2.2.2**). Σε αυτή την περίπτωση παρακολουθείται η κίνηση των N σωματιδίων που περιέχονται εντός ενός “υπερ-κελιού” το οποίο περιβάλλεται από απείρως επαναλαμβανόμενα αντίγραφα του εαυτού του. Σε αυτό ένα σωματίδιο μπορεί να αλληλεπιδράσει τόσο με άλλα σωματίδια του ίδιου “υπερ-κελιού” όσο και με σωματίδια προσκείμενων κελιών. Στα “υπερ-κελιά” μπορούν να δοθούν ποικίλα σχήματα όμως επικρατεί αυτό ενός απλού παραλληλεπίπεδου. Αυτό που επιτυγχάνεται με τις συνθήκες περιοδικών ορίων είναι ότι καθώς ένα σωματίδιο μετακινείται εκτός του κεντρικού κελιού που προσομοιώνεται ένα άλλο εισέρχεται και το αντικαθιστά. Για τον υπολογισμό του διατομικού δυναμικού ορίζεται μια μέγιστη ακτίνα πέρα από την οποία οι αλληλεπιδράσεις αγνοούνται.



Εικόνα 2.2.2: Αναπαράσταση των συνθηκών περιοδικών ορίων.

Έχοντας λοιπόν ορίσει την συνάρτηση της δυναμικής ενέργειας, ακολουθεί ο υπολογισμός των δυνάμεων που ασκούνται σε κάθε μόριο. Λόγω όμως της πολυπλοκότητας της συνάρτησης ο υπολογισμός γίνεται κατά προσέγγιση με τη χρήση αλγορίθμων. Οι αλγόριθμοι εξυπηρετούν στο ότι αν χρησιμοποιηθούν καταλλήλως καταφέρνουν να είναι υπολογιστικά αποδοτικοί. Οι γνωστότεροι αλγόριθμοι που χρησιμοποιούνται για αυτό το σκοπό είναι οι: Verlet, Leap frog, Velocity Verlet και Beeman's. Όλοι αυτοί οι αλγόριθμοι ολοκλήρωσης βασίζονται στις επεκτάσεις του Taylor, που παρουσιάζονται παρακάτω, για να

πραγματοποιηθεί ο υπολογισμός των θέσεων, των ταχυτήτων και των επιταχύνσεων των ατόμων.

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

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

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

2.3 | Το δυναμικό πεδίο Amber

Οι προσομοιώσεις μοριακής δυναμικής και τα δυναμικά πεδία έχουν χρησιμοποιηθεί σε πληθώρα μελετών και έχουν γνωρίσει σημαντική εξέλιξη. Πλέον είναι δυνατή η εφαρμογή τους σε πολύπλοκα βιολογικά συστήματα με χιλιάδες άτομα σε ένα χρονικό πλαίσιο της τάξης των μικρο-δευτερολέπτων. Παρ' όλα αυτά γίνεται μια συνεχής προσπάθεια για την περαιτέρω βελτίωσή τους με σκοπό να αυξηθεί ακόμη περισσότερο η ακρίβειά που επιτυγχάνουν.

Για την βελτίωση των δυναμικών πεδίων απαιτείται η επικύρωσή τους η οποία ξεκάθαρα βασίζεται στην λεπτομερή σύγκριση με τα πειραματικά δεδομένα. Ένα δυναμικό πεδίο δομείται με μία συγκεκριμένη παραμετροποίηση. Αυτό που το καθιστά χρήσιμο είναι η δυνατότητα να αναπαράγει δεδομένα που μπορούν να μετρηθούν με ακρίβεια αλλά δεν έχουν χρησιμοποιηθεί στην παραμετροποίηση. Σε γενικές γραμμές χρήσιμο θεωρείται καθετί που έχει υπολογιστεί πειραματικά και μπορεί να συγκριθεί με δεδομένα που προέκυψαν από την προσομοίωση. Παρ' όλα αυτά υπάρχουν κάποιοι περιορισμοί καθώς είναι δυνατόν να έχουν προκύψει λανθασμένα πειραματικά δεδομένα και να έχουν υπάρξει τεχνικά προβλήματα με την προσομοίωση. Η εξέλιξη που γνωρίζουν όμως και οι δύο τομείς επιτρέπει την ταχεία και ακριβή εξαγωγή ασφαλών αποτελεσμάτων στα οποία μπορεί να βασιστεί η βελτιστοποίηση των δυναμικών πεδίων.

Τα πιο ευρέως χρησιμοποιούμενα πρωτεϊνικά δυναμικά πεδία είναι τα Amber (*Cornell et al., 1995*), CHARMM (*MacKerell et al., 1998*), GROMOS (*Oostenbrink et al., 2004*) και OPLS-AA (*Jorgensen et al., 1996*) μιας και έχουν χρησιμοποιηθεί στην πλειοψηφία των δημοσιευμένων μελετών που αφορούν την προσομοίωση πρωτεϊνών. Καθένα από αυτά έχει αναπτυχθεί από πανεπιστημιακές ερευνητικές ομάδες και έχει ελεγχθεί μέσω μελετών από ακαδημαϊκούς συναδέλφους τους.

Ακολουθεί μια σύντομη αναφορά στα σημεία που διαφέρουν τα δυναμικά πεδία αυτά. Δύο κύριες διαφορές εντοπίζονται στο δεσμικό κομμάτι της εξίσωσης. Η πρώτη έχει να κάνει με τη χρήση των “improper” διέδρων γωνιών οι οποίες χρησιμοποιούνται για να διατηρηθεί η χειρικότητα και η επιπεδότητα ενός κεντρικού ατόμου το οποίο δεσμεύεται με τρία άλλα άτομα. Στην περίπτωση των Amber και OPLS-AA οι “ακατάλληλες” διέδρες γωνίες περιλαμβάνονται στον όρο που αφορά στις διέδρες γωνίες. Από την άλλη στην περίπτωση των CHARMM και GROMOS υπάρχει ένας ξεχωριστός όρος γι' αυτές ο οποίος έχει μια δευτέρου βαθμού εξάρτηση από την τιμή της γωνίας. Η δεύτερη διαφορά είναι ότι το δυναμικό CHARMM διαθέτει τον επιπλέον όρο Urey-Bradly ο οποίος σχετίζεται με την κάμψη γωνιών και λαμβάνει υπόψιν την απόσταση των δύο τελικών ατόμων που απαρτίζουν την γωνία.

Όπως και με τους δεσμικούς όρους δύο είναι οι διαφορές που εντοπίζονται στη μεταχείριση των μη δεσμικών όρων. Κατά πρώτον, διαφέρουν οι κανόνες συνδυασμού που χρησιμοποιούνται για τον υπολογισμό της ενέργειας αλληλεπίδρασης μεταξύ δύο ανόμοιων ατόμων. Ο υπολογισμός αυτός αφορά το δυναμικό Lennard Jones και συγκεκριμένα τις παραμέτρους ϵ και σ . Τα δυναμικά OPLS-AA και GROMOS χρησιμοποιούν γεωμετρικούς κανόνες για τον συνδυασμό και των δύο όρων ενώ στην περίπτωση των CHARMM και Amber για τον υπολογισμό του σ χρησιμοποιείται η αριθμητική μέση τιμή. Κατά δεύτερον, κάθε δυναμικό πεδίο διαχειρίζεται με διαφορετικό τρόπο τις 1,4-μη-δεσμικές αλληλεπιδράσεις μεταξύ των δύο ακραίων ατόμων μιας διέδρης γωνίας. Συγκεκριμένα κάθε δυναμικό πεδίο εφαρμόζει διαφορετικές σταθερές κλιμάκωσης για τις αλληλεπιδράσεις Lennard-Jones και Coulomb μεταξύ αυτού του είδος ζευγών ατόμων. Το δυναμικό πεδίο Amber κλιμακώνει τις 1,4-αλληλεπιδράσεις LJ κατά 1/2 και τις Coulomb κατά 1/1.2. Το δυναμικό OPLS-AA κλιμακώνει και τις δύο τύπου αλληλεπιδράσεις κατά 1/2. Το GROMOS από την άλλη χρησιμοποιεί διαφορετικές παραμέτρους για τις αλληλεπιδράσεις Lennard-Jones ανάλογα με τον τύπο του ατόμου. Τέλος, το δυναμικό CHARMM προσεγγίζει το ζήτημα με τον απλούστερο τρόπο επιλέγοντας να μην κλιμακώνει τις 1,4-μη δεσμικές αλληλεπιδράσεις παρά σε ορισμένες μόνο περιπτώσεις.

Πέρα από τα δημοφιλή δυναμικά πεδία που αναφέρθηκαν παραπάνω υπάρχουν αρκετά ακόμη καθένα από τα οποία διαθέτει ένα εύρος παραλλαγών. Τα τελευταία χρόνια πληθαίνουν οι μελέτες που στοχεύουν στο να αναδείξουν ποια είναι τα πλεονεκτήματα και οι αδυναμίες του καθενός (*Beauchamp et al., 2012, Lindorff-Larsen et al., 2012*). Στα πλαίσια αυτής της διπλωματικής εργασίας επιλέχθηκε οι προσομοιώσεις του πεπτιδίου CLN025 να

γίνουν με το δυναμικό πεδίο Amber και συγκεκριμένα τρεις παραλλαγές αυτού, το Amber ff99sb-ildn, το ff99sb-ildn-nmr και το ff99sb*-ildn.

Το δυναμικό πεδίο AMBER (Assisted Model Building with Energy Refinement) αναπτύχθηκε από την ομάδα του Peter Kollman στο Πανεπιστήμιο της Καλιφόρνια που εδράζεται στο Σαν Φρανσίσκο (*Weiner et al., 1984*). Η βασική εξίσωση του δυναμικού πεδίου έχει ως εξής:

$$\begin{aligned}
 V(r^N) = & \sum_{\text{bonds}} k_b(l - l_0)^2 + \sum_{\text{angles}} k_a(\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} \sum_n \frac{1}{2} V_n [1 + \cos(n\omega - \gamma)] \\
 & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}
 \end{aligned}$$

Το Amber έκτοτε συνεχώς βελτιώνεται. Το πρωτότυπο δυναμικό πεδίο Amber ff99 τροποποιήθηκε με την προσθήκη νέων παραμέτρων για τις δίεδρες γωνίες του σκελετού της πολυπεπτιδικής αλυσίδας. Η παραλλαγή αυτή ονομάστηκε ff99sb και επιτύγχανε μεγαλύτερη ισορροπία στα στοιχεία δευτεροταγούς δομής (*Simmerling et al., 2006*). Έπειτα το ff99sb υπέστη περαιτέρω βελτίωση μέσω αλλαγών στο δυναμικό συστροφής των πλευρικών ομάδων. Συγκεκριμένα οι τροποποιήσεις αφορούσαν 4 “προβληματικά” κατάλοιπα, την ισολευκίνη, την λευκίνη, το ασπαραγινικό και την ασπαραγίνη. Έτσι προέκυψε το ff99sb-ildn (*Lindorff-Larsen et al., 2010*). Το δυναμικό πεδίο ff99sb-ildn-nmr είναι παράγωγο του συνδυασμού των βελτιστοποιήσεων στις πλευρικές ομάδες του ff99sb-ildn με τις βελτιστοποιήσεις για τη συστροφή του σκελετού της αλυσίδας που είχε το ff99sb-nmr (*Li et al., 2010*). Τέλος οι βελτιστοποιήσεις που αφορούσαν το ff99sb-ildn συνδυάστηκαν με το ff99sb* δίνοντας γένεση στο ff99sb*-ildn. Το δυναμικό πεδίο αυτό περιλαμβάνει διορθώσεις που αφορούν τους όρους συστροφής ϕ και ψ (*Piana et al., 2011*).

Κεφάλαιο 3

Μέθοδοι

3.1 | Τεχνικά χαρακτηριστικά υπολογιστικού συστήματος

Όλες οι προσομοιώσεις πραγματοποιήθηκαν χάρις στο υπολογιστικό σύμπλεγμα Norma το οποίο εδράζεται στο Τμήμα Μοριακή Βιολογίας και Γενετικής του Δημοκρίτειου Πανεπιστημίου Θράκης. Πρόκειται για ένα υπολογιστικό cluster της κλάσης Beowulf το οποίο αποτελείται από 12 κόμβους. Η παράλληλη σύνδεση των υπολογιστών σε ένα τέτοιο σύστημα επιτυγχάνει τον καταμερισμό των εργασιών στους επιμέρους κόμβους. Με αυτό τον τρόπο απαιτητικές διεργασίες όπως μια προσομοίωση μοριακής δυναμικής δύναται να πραγματοποιηθούν με μειωμένο υπολογιστικό κόστος και αυξημένη απόδοση. Η Norma αποτελείται από 96 κεντρικές μονάδες επεξεργασίας, 114 Gb φυσικής μνήμη και 6 GPGPUs που κατανέμονται σε 12 κόμβους. Ο κεντρικός κόμβος διαθέτει 4 πυρήνες, 8 Gb φυσικής μνήμης, 1,5 Tb αποθηκευτικό χώρο στη μορφή μιας RAID-5 συστοιχίας 4 δίσκων και μια μονάδα επεξεργασίας γραφικών nvidia GTX-260. Επιπλέον υπάρχουν οκτώ κόμβοι με τετραπύρηνους επεξεργαστές Intel Q6600 Kentsfield των 2,4 Ghz καθένας από τους οποίους προσφέρει 4Gb φυσικής μνήμης. Τέσσερις εξ' αυτών διαθέτουν μια μονάδα επεξεργασίας γραφικών nvidia GTX-460. Η Norma ακόμη διαθέτει ένα κόμβο με επεξεργαστή Intel i7 965 extreme που προσφέρει 6 Gb φυσικής μνήμης και μία GPU GTX-295. Επίσης υπάρχει ένας κόμβος με οκταπύρηνο επεξεργαστή AMD FX-8150 με 4 Gb φυσικής μνήμης και μία nvidia GTX-570. Τέλος, ο κόμβος IBM X3755 περιλαμβάνει 48 AMD πυρήνες και προσφέρει 64 Gb φυσικής μνήμης και 1,8 Tb αποθηκευτικό χώρο. Όλοι οι κόμβοι συνδέονται μεταξύ τους με μεταγωγή HP ProCurve 1800-24G Gigabit Ethernet.



Εικόνα 3.1.1: Το υπολογιστικό cluster Norma.

3.2 | Το πρωτόκολλο των προσομοιώσεων

Στα πλαίσια αυτής της εργασίας πραγματοποιήθηκαν τρεις προσομοιώσεις του πεπτιδίου CLN025 με τρία δυναμικά πεδία της οικογένειας Amber, τα ff99sb-ildn, ff99sb-ildn-nmr και ff99sb*-ildn. Για να πραγματοποιηθούν οι προσομοιώσεις χρησιμοποιήθηκε το πρόγραμμα NAMD (*Phillips et al., 2005*). Πρόκειται για έναν κώδικα μοριακής δυναμικής σχεδιασμένο για υψηλών επιδόσεων προσομοιώσεις βιομοριακών συστημάτων. Το NAMD απαιτεί για την εφαρμογή του τα παρακάτω αρχεία:

- ένα αρχείο PDB με τις αρχικές συντεταγμένες του μοριακού συστήματος

Για το πεπτίδιο CLN025 που σχεδιάστηκε από τους Honda et al. υπάρχει διαθέσιμη τόσο η κρυσταλλική διαμόρφωση όπως προέκυψε από την κρυσταλλογραφία ακτίνων X αλλά και είκοσι μοντέλα NMR από πειράματα φασματοσκοπίας. Στο NAMD δόθηκε ένα αρχείο .pdb το οποίο περιελάμβανε τις συντεταγμένες του μορίου σε εκτεταμένη διαμόρφωση, για λόγους οι οποίοι περιγράφονται παρακάτω. Σε κάθε στήλη παρέχεται και ένα είδος πληροφορίας, από αριστερά προς τα δεξιά: ο τύπος καταχώρησης, ο αριθμός ατόμου, το άτομο, το κατάλοιπο στο οποίο ανήκει το άτομο, η πεπτιδική αλυσίδα, ο αριθμός του καταλοίπου, οι συντεταγμένες x, y και z του ατόμου, η πληρότητα, ο θερμοκρασιακός παράγοντας και τέλος ο τύπος του ατόμου.

ATOM	68	N	GLY	A	7	16.216	19.385	6.751	1.00	6.30	N
ATOM	69	CA	GLY	A	7	17.256	18.676	6.039	1.00	5.88	C
ATOM	70	C	GLY	A	7	18.259	19.556	5.361	1.00	6.51	C
ATOM	71	O	GLY	A	7	19.015	19.091	4.506	1.00	9.78	O
ATOM	72	N	THR	A	8	18.362	20.804	5.695	1.00	5.98	N
ATOM	73	CA	THR	A	8	19.242	21.760	5.125	1.00	5.66	C
ATOM	74	C	THR	A	8	20.419	22.026	6.074	1.00	5.15	C
ATOM	75	O	THR	A	8	20.317	21.872	7.313	1.00	6.36	O
ATOM	76	CB	THR	A	8	18.553	23.073	4.762	1.00	9.09	C
ATOM	77	CG2	THR	A	8	17.498	22.871	3.686	1.00	11.85	C
ATOM	78	OG1	THR	A	8	17.905	23.602	5.922	1.00	10.05	O
ATOM	79	N	TRP	A	9	21.517	22.418	5.517	1.00	5.75	N
ATOM	80	CA	TRP	A	9	22.762	22.680	6.241	1.00	4.78	C
ATOM	81	C	TRP	A	9	22.687	24.060	6.869	1.00	4.96	C
ATOM	82	O	TRP	A	9	22.321	25.031	6.185	1.00	7.83	O
ATOM	83	CB	TRP	A	9	23.924	22.604	5.307	1.00	6.90	C
ATOM	84	CG	TRP	A	9	25.267	22.745	5.969	1.00	6.58	C
ATOM	85	CD1	TRP	A	9	26.024	23.847	6.060	1.00	9.48	C
ATOM	86	CD2	TRP	A	9	25.975	21.701	6.641	1.00	6.96	C
ATOM	87	CE2	TRP	A	9	27.178	22.245	7.100	1.00	7.80	C
ATOM	88	CE3	TRP	A	9	25.747	20.347	6.894	1.00	8.84	C

- ένα αρχείο διαμόρφωσης (configuration file)

Το αρχείο αυτό δίνεται στο NAMD και καθορίζει τα πάντα γύρω από την προσομοίωση που πρόκειται να συμβεί. Στην **Εικόνα 3.2.1** δίνεται το αρχείο που χρησιμοποιήθηκε.

```

#
# Input files
#
amber                on
readexclusions       yes
parmfile             c1n025.prmtop
coordinates          heat_out.coor
bincoordinates       restart.coor
binvelocities        restart.vel
extendedSystem       restart.xsc

#
# Adaptive ...
#
adaptTempMD          on
adaptTempInFile      restart.tempering
adaptTempRestartFile output/restart.tempering
adaptTempRestartFreq 10000
adaptTempLangevin    on
adaptTempRescaling   off
adaptTempOutFreq     400

#
# Output files & writing frequency for DCD
# and restart files
#
outputname           output/equil_out
binaryoutput         off
restartname          output/restart
restartfreq          10000
binaryrestart        yes
dcdFile              output/equil_out.dcd
dcdFreq             400
DCDunitcell         yes

#
# Frequencies for logs and the xst file
#
outputEnergies       400
outputTiming         1600
xstFreq              400

#
# Timestep & friends
#
timestep             2.0
stepsPerCycle        20
nonBondedFreq        1
fullElectFrequency   2

#
# Simulation space partitioning
#
switching            on
switchDist           7
cutoff               8
pairlistdist         10

twoAwayX             yes

#
# Basic dynamics
#
COMmotion            no
dielectric            1.0
exclude              scaled1-4
1-4scaling           0.833333
rigidbonds           all

#
# Particle Mesh Ewald parameters:
#
Pme                  on
PmeGridsizeX         32
PmeGridsizeY         32
PmeGridsizeZ         32

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

#
# Langevin dynamics parameters
#
langevin              on
langevinDamping       1
langevinTemp          370
langevinHydrogen     off

langevinPiston        on
langevinPistonTarget 1.01325
langevinPistonPeriod 400
langevinPistonDecay  200
langevinPistonTemp    370

useGroupPressure     yes

firsttimestep        10000
run                   500000000

```

Εικόνα 3.2.1: Το αρχείο διαμόρφωσης NAMD που εκτέλεσε το πρόγραμμα για την διεξαγωγή των προσομοιώσεων.

➤ ένα αρχείο PRMTOP για την παραμετροποίηση του δυναμικού πεδίου Amber
 Αυτό το αρχείο περιέχει μία περιγραφή της τοπολογίας του μορίου αλλά και όλες τις απαραίτητες παραμέτρους του δυναμικού πεδίου. Το αρχείο παράγεται από το πρόγραμμα LEaP. Παρακάτω δίνεται το αρχικό τμήμα του .prmtop αρχείου που χρησιμοποιήθηκε.

```
%VERSION VERSION_STAMP = V0001.000 DATE = 12/26/12 20:31:37
%FLAG TITLE
%FORMAT(20a4)
default_name
%FLAG POINTERS
%FORMAT(10I8)
    3377      16      3280      98      163      136      349      299      0      0
    5188     1083      98      136      299      33      65      50      25      1
      0         0         0         0         0         0         0         1      24      0
      0
%FLAG ATOM_NAME
%FORMAT(20a4)
N H1 H2 H3 CA HA CB HB2 HB3 CG CD1 HD1 CE1 HE1 CZ OH HH CE2 HE2 CD2
HD2 C O N H CA HA CB HB2 HB3 CG CD1 HD1 CE1 HE1 CZ OH HH CE2 HE2
CD2 HD2 C O N H CA HA CB HB2 HB3 CG OD1 OD2 C O N CD HD2 HD3
CG HG2 HG3 CB HB2 HB3 CA HA C O N H CA HA CB HB2 HB3 CG HG2 HG3
CD OE1 OE2 C O N H CA HA CB HB CG2 HG21HG22HG23OG1 HG1 C O N
H CA HA2 HA3 C O N H CA HA CB HB CG2 HG21HG22HG23OG1 HG1 C O
N H CA HA CB HB2 HB3 CG CD1 HD1 NE1 HE1 CE2 CZ2 HZ2 CH2 HM2 CZ3 HZ3 CE3
HE3 CD2 C O N H CA HA CB HB2 HB3 CG CD1 HD1 CE1 HE1 CZ OH HH CE2
HE2 CD2 HD2 C O OXT Na+ CL- Na+ Na+ O H1 H2 O H1 H2 O H1 H2 O H1
H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2
O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1
H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O
H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2
O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1 H2 O H1
```

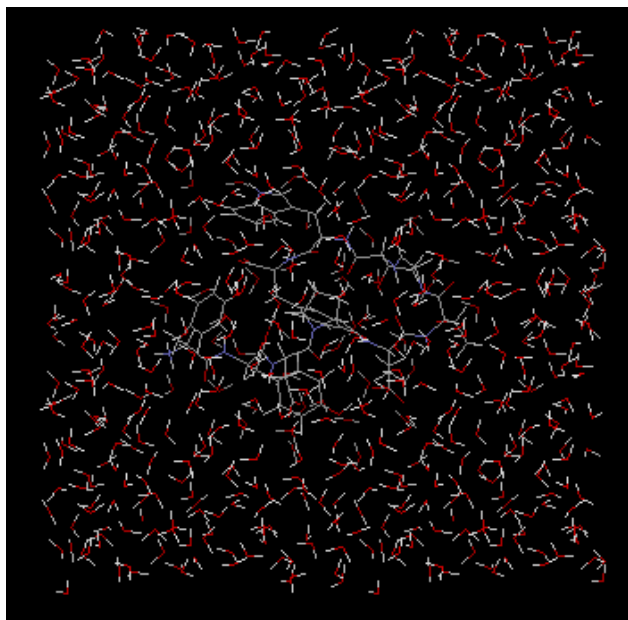
Κάθε προσομοίωση διέπεται από μια σειρά συγκεκριμένων βημάτων όπως παρουσιάζονται στο οργανόγραμμα της **Εικόνας 3.2.2**.



Εικόνα 3.2.2: Γενικό οργανόγραμμα του πρωτοκόλλου μιας προσομοίωσης μοριακής

δυναμικής.

Το εναρκτήριο σημείο μιας προσομοίωσης, ο χρόνος $t=0$, είναι εξαιρετικά σημαντικό. Το τελικό τροχιακό που παράγεται μπορεί να είναι εξαρτώμενο σε μεγάλο βαθμό από την αρχική διαμόρφωση απ' όπου ξεκινά η προσομοίωση. Αυτό συμβαίνει γιατί ένα κακώς προετοιμασμένο σύστημα το οποίο περιέχει ατομικές συγκρούσεις θα ξεκινήσει την προσομοίωση με δυνάμεις που πιθανόν να επέμβουν στην δομή του μορίου που εξετάζεται. Για το λόγο αυτό χρησιμοποιήθηκε το πρόγραμμα RIBOSOME (*Shrinivasan*) χάρις στο οποίο δημιουργήθηκαν οι αρχικές ατομικές συντεταγμένες της πεπτιδικής αλυσίδας σε πλήρως εκτεταμένη διαμόρφωση. Έπειτα με το πρόγραμμα MOLEMAN (*Kleywegt et al., 2001*) ευθυγραμμίστηκαν οι άξονες του πεπτιδίου με το ορθοκανονικό σύστημα και μετατοπίστηκε το κέντρο βάρους του ώστε να συμπίπτει με την αρχή των αξόνων. Στη συνέχεια με τη χρήση του προγράμματος LEaP, το οποίο παρέχεται από τους δημιουργούς του δυναμικού πεδίου Amber, έγινε η προσθήκη των ατόμων υδρογόνου και των ιόντων στη δομή. Συγκεκριμένα τα ιόντα που προστέθηκαν ήταν τρία άτομα Na^+ και ένα άτομο Cl^- . Ακολούθως πραγματοποιήθηκε η διαλυτοποίηση του πεπτιδίου εντός του επιλεγμένου διαλύτη. Η παρουσία του διαλύτη που είναι το νερό ήταν αναλυτική και έγινε με τη χρήση του μοντέλου νερού TIP3P (*Jorgensen et al., 1983*). Συγκεκριμένα δημιουργήθηκε ένα ρυθμιστικό διάλυμα από TIP3P νερό 6\AA γύρω από το πεπτίδιο προς κάθε κατεύθυνση. Συνολικά στο τέλος το κουτί προσομοίωσης περιλάμβανε 3377 άτομα.



Εικόνα 3.2.3: Κυψελίδα που περιλαμβάνει το πεπτίδιο CLN025 πλαισιωμένο από τα μόρια

του διαλύτη H₂O.

Πριν ξεκινήσει η προσομοίωση λαμβάνει χώρα μια ελαχιστοποίηση της ενέργειας του συστήματος. Με τον τρόπο αυτό αφαιρούνται τυχόν ισχυρές αλληλεπιδράσεις van der Waals οι οποίες μπορούν να επιδράσουν τοπικά στη δομή και να αποσταθεροποιήσουν την προσομοίωση. Η ενέργεια αυτή πραγματοποιήθηκε σε 1000 βήματα. Ακολούθησε μια σύντομη φάση θέρμανσης και αμέσως μετά το σύστημα εξισορροπήθηκε για 1000 ακόμη βήματα.

Τελικό βήμα είναι η κύρια φάση της προσομοίωσης η οποία αλλιώς ονομάζεται παραγωγική φάση. Η προσομοίωση διενεργήθηκε υπό συνθήκες adaptive tempering, πράγμα που σημαίνει ότι η προσομοίωση δεν έλαβε χώρα σε συγκεκριμένη θερμοκρασία αλλά εντός ενός εύρους θερμοκρασιών και συγκεκριμένα 300-500K. Η θερμοκρασία μεταβάλλεται με βάση τη δυναμική ενέργεια του συστήματος. Όταν η δυναμική ενέργεια έχει χαμηλότερη τιμή από την τρέχουσα μέση τιμή τότε η θερμοκρασία μειώνεται, ενώ όταν η δυναμική ενέργεια έχει μεγαλύτερη τιμή από την τρέχουσα μέση τιμή τότε η θερμοκρασία αυξάνεται. Αυτό επιτυγχάνεται μέσω ενός θερμοστάτη, του Langevin thermostat.

Για τον υπολογισμό των θέσεων και των ταχυτήτων των ατόμων στο χρόνο χρησιμοποιήθηκε ο αλγόριθμος ολοκλήρωσης Verlet-I. Για να μειωθεί το κόστος υπολογισμού των ηλεκτροστατικών αλληλεπιδράσεων χρησιμοποιήθηκε πολλαπλό χρονικό βήμα της τάξεως των 2fs. Οι τοπικές αλληλεπιδράσεις (δεσμικές, van der Waals και οι ηλεκτροστατικές μέχρι μια συγκεκριμένη τιμή) υπολογίζονται σε κάθε βήμα ενώ οι μεγαλύτερης εμβέλειας ηλεκτροστατικές αλληλεπιδράσεις κάθε δύο βήματα.

Για την εξαγωγή του τροχιακού είχε οριστεί να αποθηκεύονται οι ατομικές συντεταγμένες κάθε 400 βήματα. Η τιμή αυτή αντιστοιχεί σε 0,8ps φυσικού χρόνου. Καθεμιά από τις προσομοιώσεις που διενεργήθηκαν είχε διαφορετική διάρκεια. Στον Πίνακα 1 παρουσιάζονται αναλυτικά δεδομένα που αφορούν την διάρκεια κάθε προσομοίωσης και το μήκος του αντίστοιχου τροχιακού.

Πίνακας 1

	Στιγμιότυπα (frames)	Φυσικός χρόνος (μs)
ff99sb-ildn	6310325	5,04
ff99sb-ildn-nmr	6273375	5,01
ff99sb*-ildn	6677500	5,34

Αφότου η προσομοιώσεις έλαβαν τέλος ξεκίνησε η ανάλυση των αποτελεσμάτων και η σύγκρισή τους με τα πειραματικά δεδομένα. Σημειώνεται εδώ ότι από το επόμενο κεφάλαιο τα δυναμικά πεδία Amber ff99sb-ildn, ff99sb-ildn-nmr και ff99sb*-ildn και τα τροχιακά τους θα αναφέρονται συχνά για λόγους συντομίας ως ILDN, NMR και STAR αντίστοιχα.

Κεφάλαιο 4

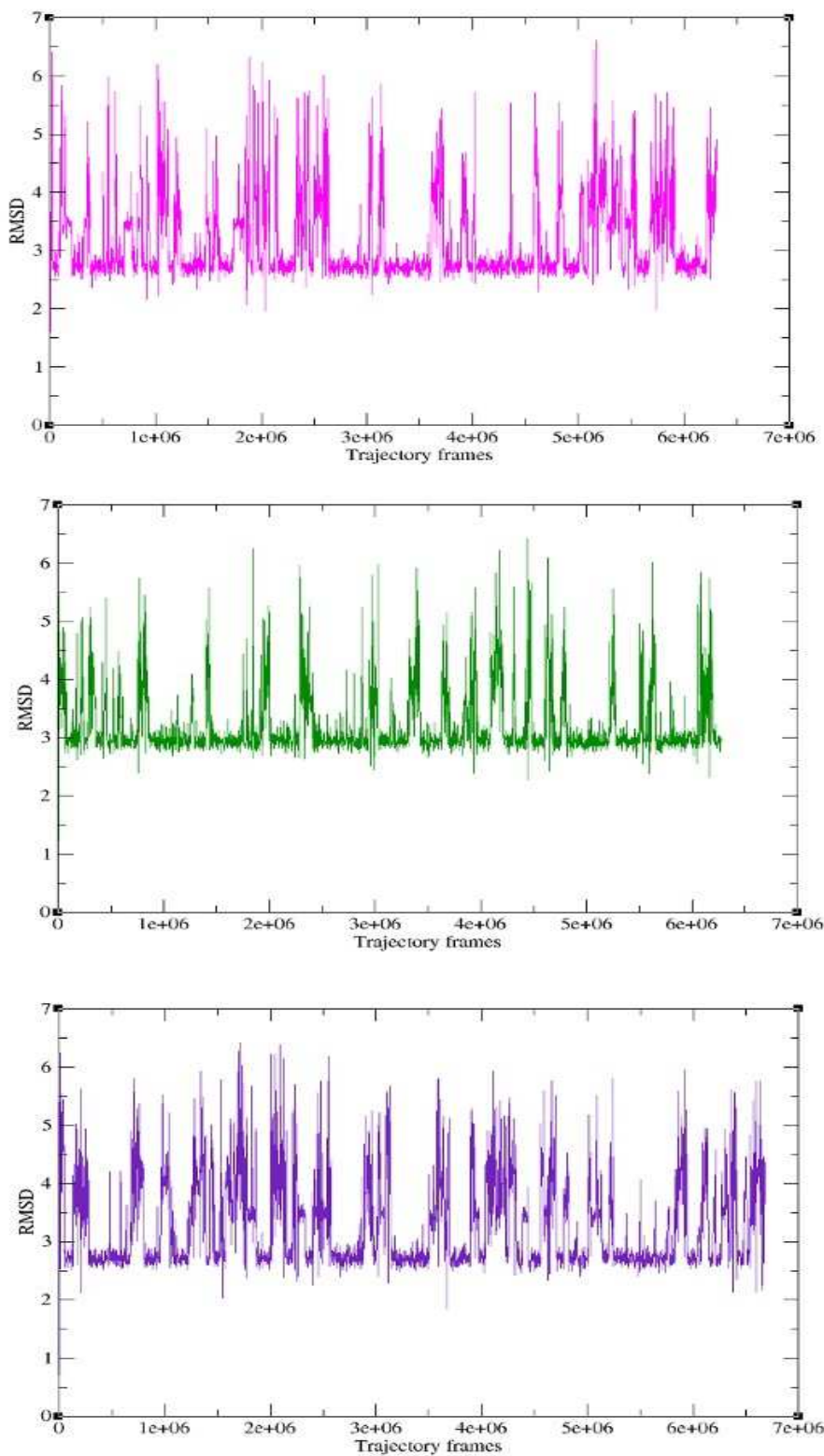
Αποτελέσματα

4.1 | Ανάλυση των προσομοιώσεων με βάση το RMSD

Η ανάλυση των προσομοιώσεων μοριακής δυναμικής πραγματοποιήθηκε με τη χρήση του προγράμματος carma (Glykos, 2006) και ειδικότερα του grcarma (Koukos, 2013), της αυτοματοποιημένης διεπαφής που σχεδιάστηκε για την απλούστευση της εφαρμογής του. Το πρόγραμμα τροφοδοτείται με δύο αρχεία, ένα DCD και ένα PSF. Το πρώτο περιλαμβάνει το τροχιακό που εξήγαγε η προσομοίωση, δηλαδή τις συντεταγμένες των ατόμων καθ' όλη τη διάρκεια της προσομοίωσης του μορίου. Το δεύτερο περιέχει δομική πληροφορία για το υπό μελέτη μόριο όπως άτομα, δεσμούς και γωνίες.

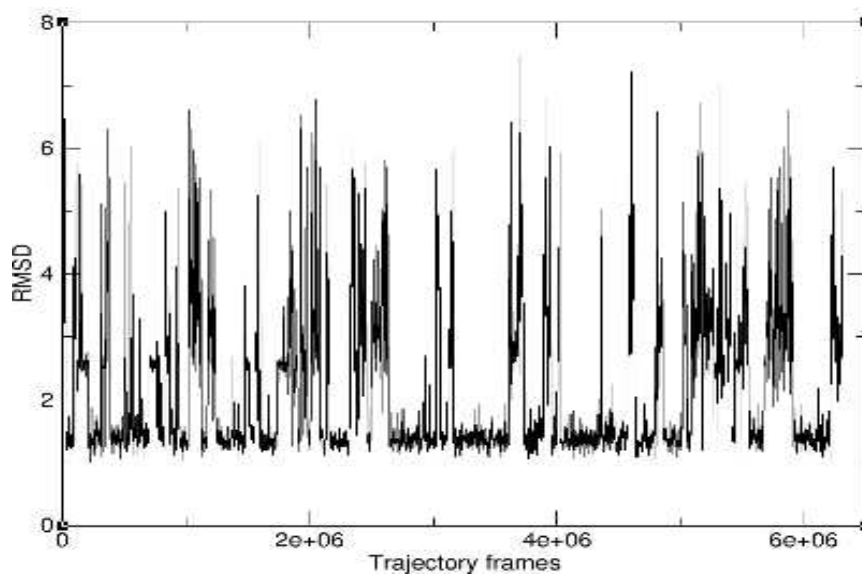
Πρώτο βήμα στη διαδικασία της ανάλυσης ήταν η λήψη μιας αρχικής εικόνας για την κινητικότητα του πεπτιδίου κατά την προσομοίωση. Για το λόγο αυτό έγινε υπολογισμός της τιμής του RMSD της κάθε διαμόρφωσης που περιέχεται στο τροχιακό από την πρώτη κατά σειρά διαμόρφωση. Το RMSD (root-mean-square deviation) αποτελεί το μέτρο της μέσης απόστασης μεταξύ των ατόμων σε δύο πρωτεϊνικά μόρια το οποία βρίσκονται σε υπέρθεση. Ο τύπος για τον υπολογισμό του είναι ο εξής: $RMSD = \sqrt{[1/n(x_i - x_{ref})^2]}$, όπου x_i είναι οι συντεταγμένες των ατόμων μια δεδομένη στιγμή, x_{ref} οι συντεταγμένες των ατόμων στη δομή που αποτελεί την αναφορά και n ο αριθμός των ατόμων. Στην εν προκειμένω μέτρηση χρησιμοποιήθηκαν μόνο τα Ca άτομα του σκελετού. Η γραφική παράσταση (**Εικόνα 4.1.1**) αναπαριστά στον άξονα Y την τιμή του RMSD και στον άξονα X τις διαμορφώσεις του τροχιακού. Για να αποκτήσει ευκρίνεια η εικόνα και να μειωθεί το μέγεθος της πληροφορίας που αναπαρίσταται, η μέτρηση έγινε με βήμα 2000. Τα αποτελέσματα καταδεικνύουν την ακεραιότητα της προσομοίωσης και στις τρεις περιπτώσεις. Το πεπτίδιο φαίνεται να είναι σταθερό σε εκτενή διαστήματα καθ' όλη τη διάρκεια. Η πλειοψηφία των δομών έχει RMSD από την αρχική δομή της τάξεως των $\sim 2,8 \text{ \AA}$ ενώ η δεύτερη σε μέγεθος ομάδα διαμορφώσεων περιλαμβάνει αυτές με RMSD $\sim 3,5 \text{ \AA}$.

Είναι όμως σημαντικό να εκτιμηθεί αν και πόσο κοντά στην πειραματικά προσδιορισμένη δομή φτάνει το πεπτίδιο στην διάρκεια της κάθε προσομοίωσης. Για το σκοπό αυτό έγινε υπολογισμός της τιμής του RMSD από τα είκοσι διαθέσιμα μοντέλα NMR για το CLN025. Για να καταστεί δυνατός αυτός ο υπολογισμός κατασκευάστηκε ένα νέο τροχιακό το οποίο διαθέτει πέρα από τις διαμορφώσεις του προϋπάρχοντος τροχιακού και τα μοντέλα NMR ως τις πρώτες κατά σειρά διαμορφώσεις. Ενδεικτικά παρατίθεται το γράφημα για τις τιμές του RMSD που έχουν οι δομές που προέκυψαν από την προσομοίωση

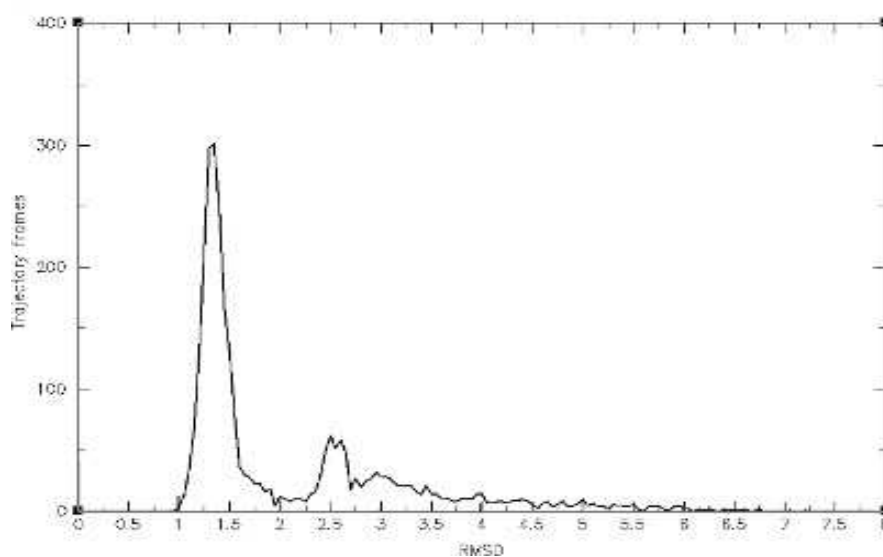


Εικόνα 4.1.1: Γραφική αναπαράσταση των τιμών του RMSD κάθε διαμόρφωσης του τροχιακού ως προς την πρώτη κατά σειρά διαμόρφωση. Τα δυναμικά πεδία έχουν ως εξής: ILDN (magenta), NMR (green) και STAR (indigo)

A.



B.



Εικόνα 4.1.2: Α. Γραφική αναπαράσταση των τιμών του RMSD των διαμορφώσεων του τροχιακού ILDN ως προς το πρώτο μοντέλο NMR και Β. Ιστόγραμμα όπου παρουσιάζεται το πλήθος των διαμορφώσεων που αντιστοιχεί σε κάθε τιμή RMSD

με το ILDN σε σχέση με το πρώτο μοντέλο NMR (Εικόνα 4.1.2). Η πλειονότητα των διαμορφώσεων, όπως φαίνεται στο ιστόγραμμα, έχει τιμή RMSD από την πειραματικά

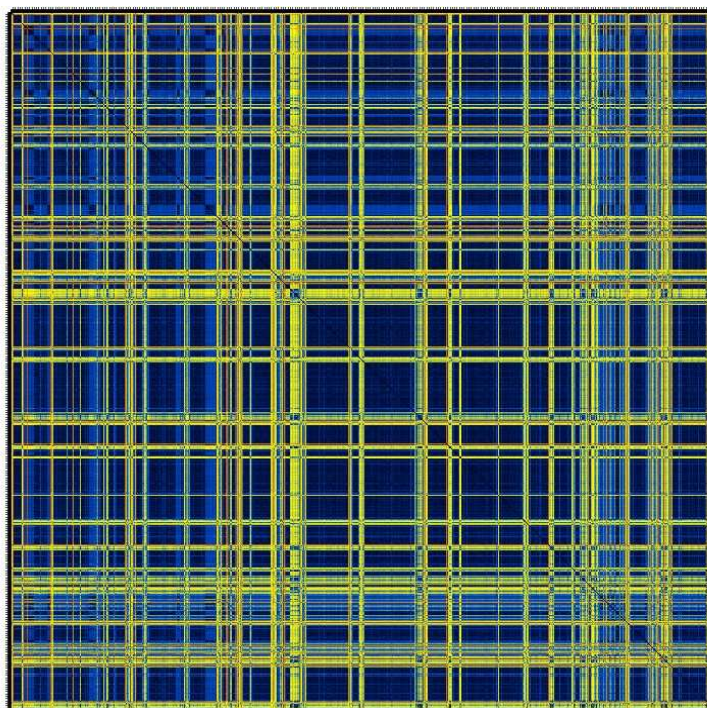
προσδιορισμένη δομή που κυμαίνεται στα 1,3Å. Οι δομές αυτές βρίσκονται αρκετά κοντά στην πειραματική δομή του μοντέλου και μπορούν να χαρακτηριστούν ως αναδιπλωμένες. Αξίζει να σημειωθεί ότι υπάρχει και μια δεύτερη μικρότερου αλλά σημαντικού μεγέθους ομάδα δομών για την οποία η τιμή του RMSD είναι περίπου 2,6Å.

Η ίδια διαδικασία για την κατασκευή νέων τροχιακών και τη σύγκριση τους με τις διαθέσιμες πειραματικές δομές έλαβε χώρα και για τα τροχιακά που αφορούν και τα άλλα δύο δυναμικά πεδία. Εξήχθησαν παρόμοια αποτελέσματα τα οποία επίσης δείχνουν ότι η προσομοίωση καταλήγει σε αναδιπλωμένες διαμορφώσεις σε εκτενή διαστήματα καθ' όλη τη διάρκεια της προσομοίωσης.

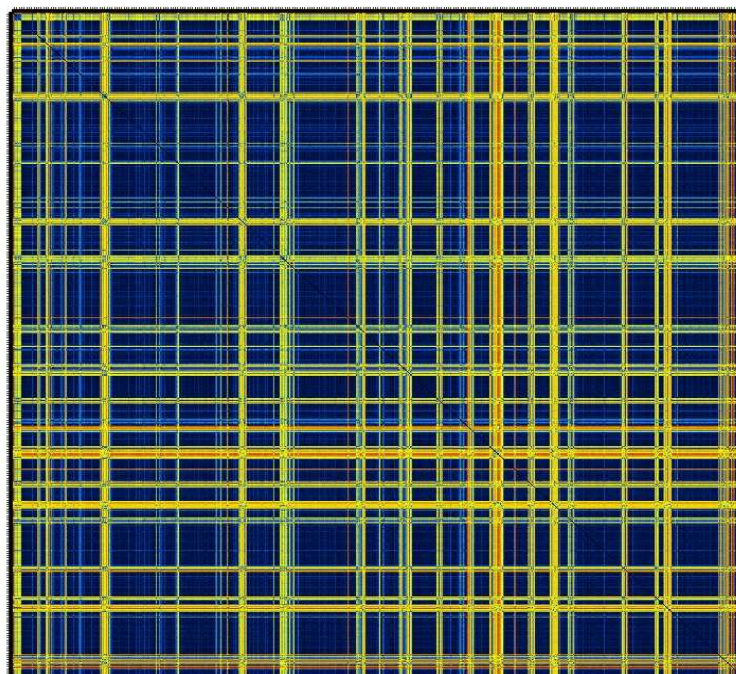
Εν συνεχεία ακολούθησε η κατασκευή διδιάστατων πινάκων RMSD. Ένας πίνακας RMSD αντιπροσωπεύει το λεγόμενο frame-to-frame RMSD, δηλαδή την τιμή του RMSD κάθε διαμόρφωσης σε σχέση με καθεμιά από τις υπόλοιπες και όχι ως προς μια μοναδική δομή αναφοράς. Ο κάθε άξονας του πίνακα περιέχει όλες τις διαμορφώσεις σε χρονική σειρά και η διαγώνιος περιλαμβάνει την τιμή του RMSD της κάθε διαμόρφωσης ως προς τον εαυτό της. Ο χρωματικός κώδικας που χρησιμοποιείται έχει ως εξής. Υψηλές τιμές RMSD αντιπροσωπεύονται με κόκκινο χρώμα και ενδιάμεσες τιμές με κίτρινο. Χαμηλές τιμές RMSD αναπαριστώνται από προοδευτικά σκούρο μπλε χρώμα. Η κατασκευή των πινάκων έγινε και πάλι χάριν μείωσης τους μεγέθους με βήμα 1000.

Αρχικά παρατίθεται ο πίνακας RMSD που αφορά το τροχιακό του ILDN (**Εικόνα 4.1.3**). Ένας πίνακας RMSD αναδεικνύει τις περιοδικές αλλαγές και τη σταθερότητα του μορίου στην προσομοίωση. Εδώ φαίνεται ότι στο μέσο της προσομοίωσης κυριαρχούν περιοχές με σκούρο μπλε χρώμα πράγμα που σημαίνει ότι σε αυτά τα διαστήματα το πεπτίδιο κατέληξε και διατήρησε σταθερές διαμορφώσεις. Παρόμοια κινητικότητα είχε και το πεπτίδιο στην προσομοίωση με το NMR (**Εικόνα 4.1.4**). Σε αυτή την περίπτωση όμως τα διαστήματα όπου η τιμή του RMSD είναι χαμηλή κυριαρχούν σε όλη τη διάρκεια της προσομοίωσης. Τέλος, παρουσιάζεται ο πίνακας RMSD για την προσομοίωση που έγινε με το δυναμικό πεδίο STAR (**Εικόνα 4.1.5**). Εδώ οι κίτρινες περιοχές έχουν πιο έντονη παρουσία σε σχέση με τους προηγούμενους πίνακες. Το πεπτίδιο φαίνεται να διέρχεται από συνεχείς αλλαγές στη δομή του όμως αξιολογικά είναι και τα διαστήματα στα οποία παραμένει σταθερό.

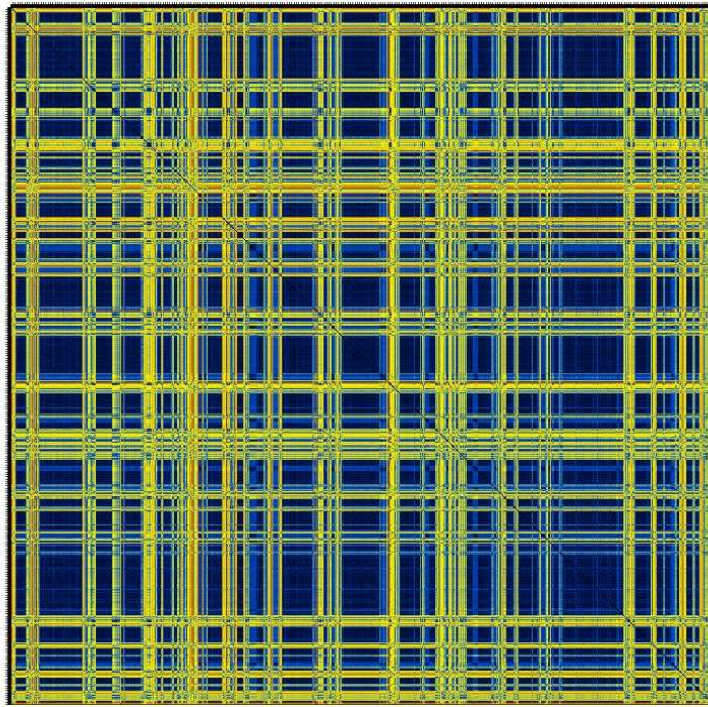
Συνοψίζοντας ως εδώ σημειώνονται τα εξής. Όλες οι προσομοιώσεις που πραγματοποιήθηκαν παρήγαγαν τροχιακά των οποίων οι δομές εμφανίζουν σταθερότητα. Επιπλέον προσομοιάζουν αρκετά τις διαθέσιμες πειραματικές δομές και θεωρείται ότι καταλήγουν σε αναδιπλωμένες διαμορφώσεις.



Εικόνα 4.1.3: Πίνακας RMSD για το τροχιακό ILDN



Εικόνα 4.1.4: Πίνακας RMSD για το τροχιακό NMR



Εικόνα 4.1.5: Πίνακας RMSD για το τροχιακό STAR

4.2 | Ανάλυση των προσομοιώσεων με βάση τη μέθοδο της Ανάλυσης Κύριων Συνιστωσών

4.2.1. Η μέθοδος

Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA) είναι μια στατιστική τεχνική που εφαρμόζεται με σκοπό την ανεύρεση προτύπων σε ένα σύνολο δεδομένων και τελικώς στην παρουσίαση των δεδομένων με τρόπο που θα αναδεικνύει τις ομοιότητες και τις διαφορές που έχουν μεταξύ τους. Ουσιαστικά η τεχνική αποσκοπεί στη “συμπύεση” διανυσμάτων σε μικρότερο αριθμό διαστάσεων. Όσο περισσότερο συσχετισμένες είναι οι μεταβλητές των αρχικών διανυσμάτων μεταξύ τους, τόσο μεγαλύτερη συμπύεση επιτυγχάνεται. Οι νέες μεταβλητές αποκαλούνται κύριες συνιστώσες (principal components, PC) και κατασκευάζονται έτσι ώστε ένας μικρός αριθμός από κύριες συνιστώσες να καλύπτει ένα μεγάλο ποσοστό της μεταβλητότητας του δείγματος. Το μέτρο του πόσο συσχετισμένες

είναι οι μεταβλητές μεταξύ τους είναι ο πίνακας συνδιακύμανσης (covariance matrix). Οι τιμές του πίνακα αποτελούν μέτρο του κατά πόσο μια μεταβλητή μεταβάλλεται μαζί με μία άλλη.

Η Ανάλυση Κύριων Συνιστωσών που εφαρμόζεται στη μελέτη της κινητικότητας ενός πρωτεϊνικού μορίου περιλαμβάνει την κλασική Cartesian PCA (cPCA) που λαμβάνει υπόψιν τις καρτεσιανές συντεταγμένες των ατόμων του μορίου και την Dihedral PCA (dPCA) η οποία προσανατολίζεται γύρω από τις δίεδρες γωνίες (φ , ψ) του πεπτιδικού σκελετού. Η dPCA προτάθηκε ως αποτελεσματικότερη σε κινήσεις μεγάλου πλάτους όπως αυτές που λαμβάνουν χώρα κατά την αναδίπλωση ενός πεπτιδίου. Αντίθετα η cPCA είναι επιτυχής σε περιπτώσεις μικρών διακυμάνσεων (Altis et al., 2007).

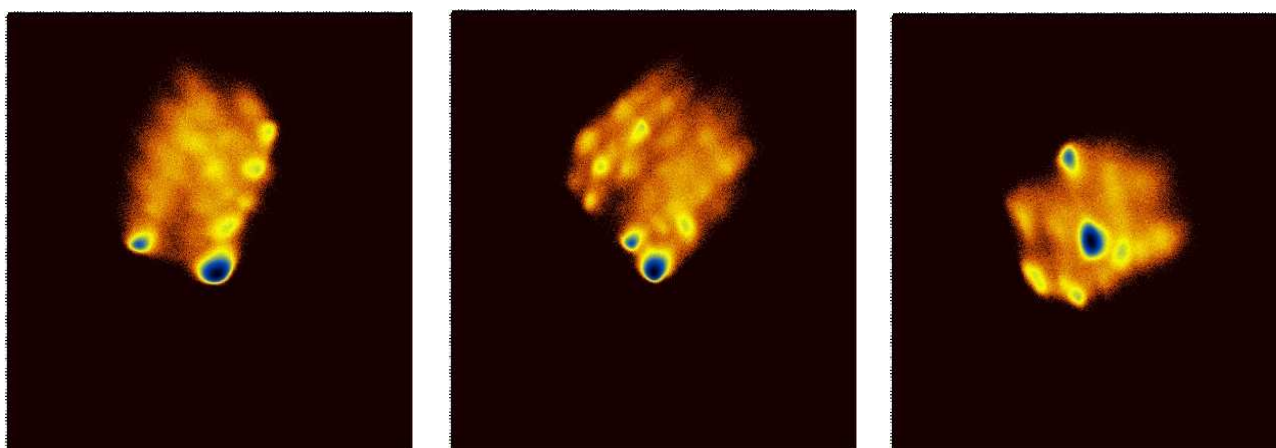
4.2.2. Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό ILDN

Ξεκινώντας από τη μελέτη του τροχιακού που προέκυψε από την προσομοίωση με το δυναμικό πεδίο ILDN και χρησιμοποιώντας το πρόγραμμα gcrarma εφαρμόστηκε dPCA στα άτομα C, Ca, N και O του πεπτιδικού σκελετού. Τα αποτελέσματα δείχνουν ότι στο τροχιακό περιέχονται δύο κύρια clusters.

1o cluster: 3336245 από τις 6310325 διαμορφώσεις

2o cluster: 245478 από τις 6310325 διαμορφώσεις

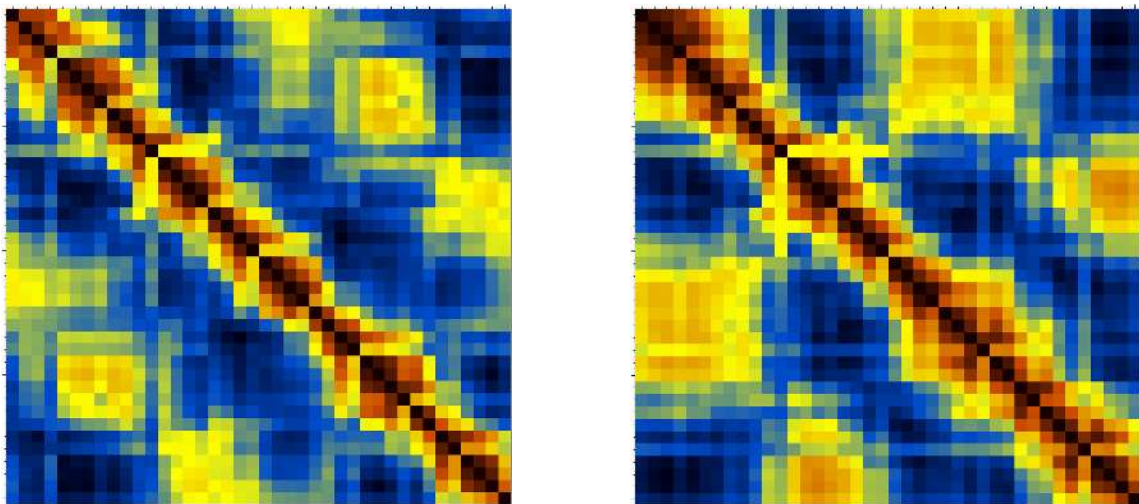
Το πρώτο και κυρίαρχο cluster αντιστοιχεί στο 52,8% του τροχιακού και το δεύτερο στο 3,8%. Στα διαγράμματα ελεύθερης ενέργειας (Εικόνα 4.2.2.1) το πρώτο cluster καταλαμβάνει την μεγαλύτερη σε μέγεθος συμπαγή περιοχή με το σκούρο μπλε χρώμα.



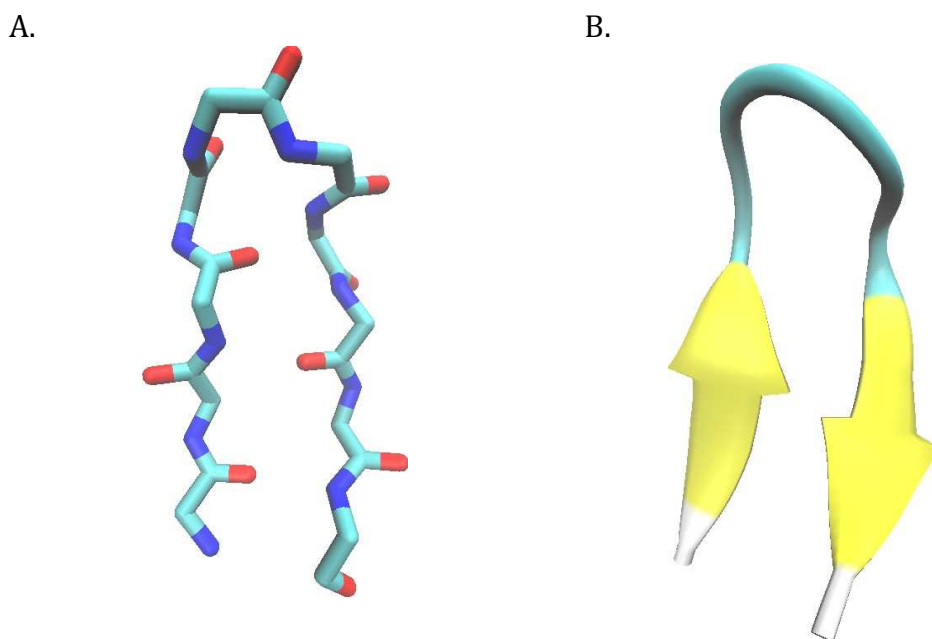
Εικόνα 4.2.2.1: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή dPCA στο ILDN. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε

σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα

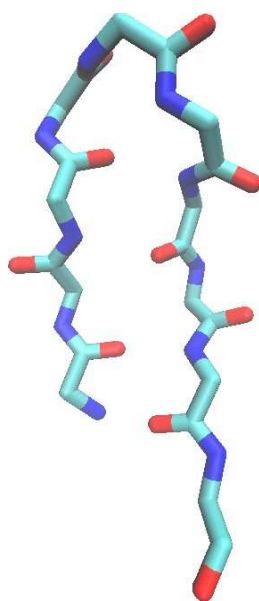
Όπως έχει αναφερθεί προηγουμένως, ο πίνακας συνδιακύμανσης αποτελεί μέτρο του πόσο συσχετισμένες είναι μεταξύ τους οι μεταβλητές. Στην **Εικόνα 4.2.2.2** παρατίθενται οι πίνακες συνδιακύμανσης για το πρώτο και δεύτερο cluster. Η αντιπροσωπευτική δομή του πρώτου cluster είναι ξεκάθαρα μία β-φουρκέτα και αποτελείται από δύο αντιπαράλληλους β κλώνους οι οποίοι συνδέονται μέσω ενός βρόχου (**Εικόνα 4.2.2.3**). Όσον αφορά την αντιπροσωπευτική δομή του δεύτερου cluster πρόκειται για μια offset διαμόρφωση στην οποία το ν-τελικό άκρο προεξέχει κατά ένα κατάλοιπο κάνοντας τη δομή λιγότερο σταθερή σε σύγκριση με αυτή του πρώτου cluster (**Εικόνα 4.2.2.4**).



Εικόνα 4.2.2.2: Πίνακες συνδιακύμανσης κατά σειρά για το πρώτο και δεύτερο cluster που προέκυψαν από την εφαρμογή dPCA στο ILDN. Τμήματα του μορίου που κινούνται μαζί έχουν πορτοκαλί και κόκκινο χρώμα



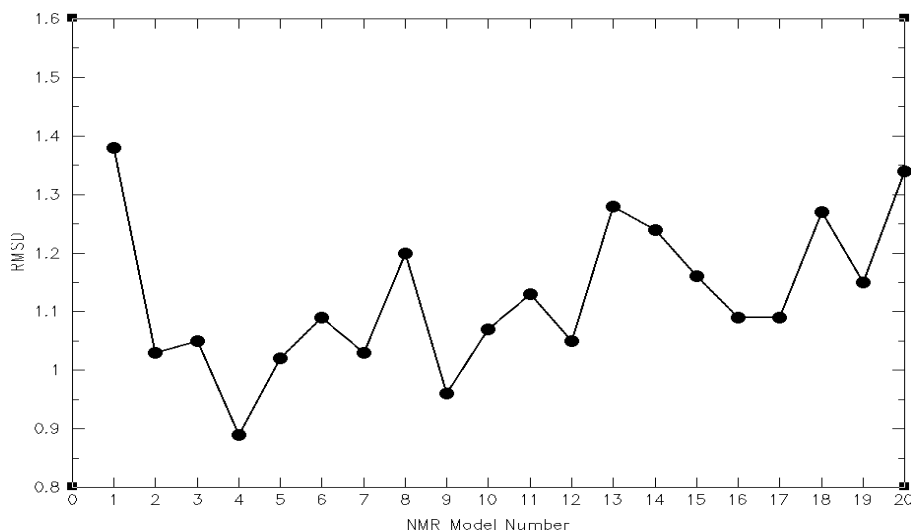
Εικόνα 4.2.2.3: Η αντιπροσωπευτική διαμόρφωση του πρώτου cluster όπως προέκυψε από την dPCA για το τροχιακό ILDN σε licorice (A) και cartoon (B) αναπαράσταση



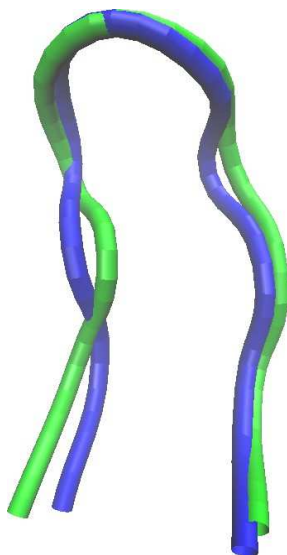
Εικόνα 4.2.2.4: Η αντιπροσωπευτική διαμόρφωση του δεύτερου cluster όπως προέκυψε από την dPCA για το τροχιακό ILDN

Εν συνεχεία προκειμένου να ελεγχθεί το πόσο ακριβώς προσεγγίζει την πειραματικά προσδιορισμένη δομή η αντιπροσωπευτική δομή του επικρατούς πρώτου cluster, έγινε υπέρθεση του με τα μοντέλα NMR. Στο διάγραμμα της **Εικόνας 4.2.2.5** φαίνεται ότι το μοντέλο NMR με το μικρότερο RMSD από την αντιπροσωπευτική δομή του πρώτου cluster

είναι το τέταρτο. Συγκεκριμένα η τιμή του RMSD ήταν στα 0,89Å (Εικόνα 4.2.2.6). Στην υπέρθεση η κύρια διαφορά μεταξύ των δομών αφορά τον β-κλώνο με τα τρία πρώτα κατάλοιπα. Σημειώνεται επίσης ότι το RMSD για όλα τα μοντέλα ήταν μικρότερο του 1,4Å και κατά συνέπεια ικανοποιητικά κοντά.



Εικόνα 4.2.2.5: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο ILDN



Εικόνα 4.2.2.6: Υπέρθεση του τέταρτου μοντέλου NMR (μπλε αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο ILDN (πράσινη αλυσίδα)

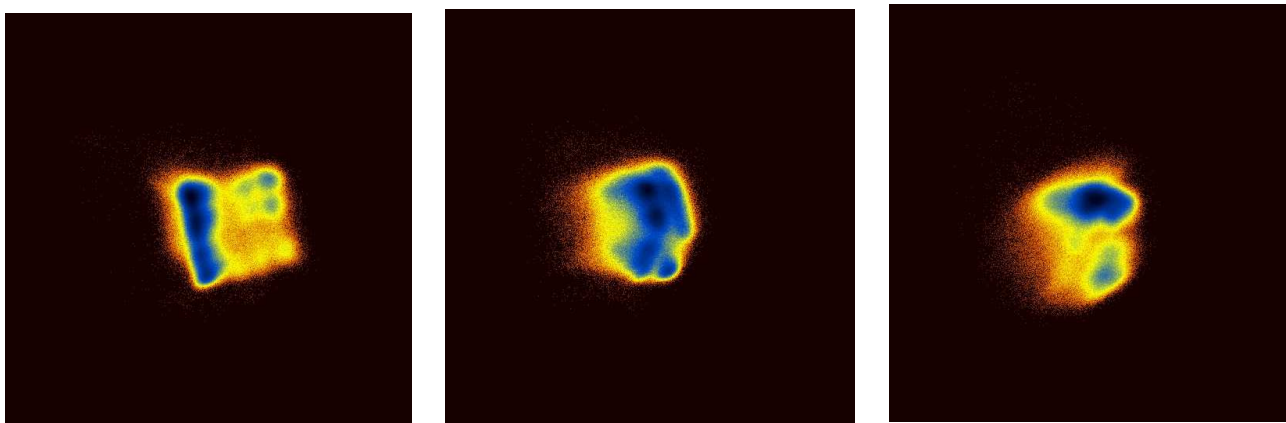
Έπειτα με σκοπό την μελέτη της συμπεριφοράς των πλευρικών ομάδων έγινε χρήση της cartesian PCA (cPCA) στο πρώτο και κυρίαρχο τροχιακό. Η εφαρμογή της μεθόδου προϋποθέτει την αφαίρεση των ολικών περιστροφών και μεταθέσεων με σκοπό να μην ληφθούν υπόψιν αλλαγές στις ατομικές συντεταγμένες λόγω της κινητικότητας του σκελετού του πεπτιδίου. Για το σκοπό αυτό έγινε fitting των διαμορφώσεων χρησιμοποιώντας ως αναφορά μία εξ αυτών. Παράχθηκε έτσι ένα νέο τροχιακό στο οποίο εφαρμόστηκε η μέθοδος ενώ χρησιμοποιήθηκαν μόνο τα heavy άτομα. Τα αποτελέσματα έδειξαν ότι εντοπίστηκαν δύο clusters:

1o cluster: 1654717 από τις 3336245 διαμορφώσεις

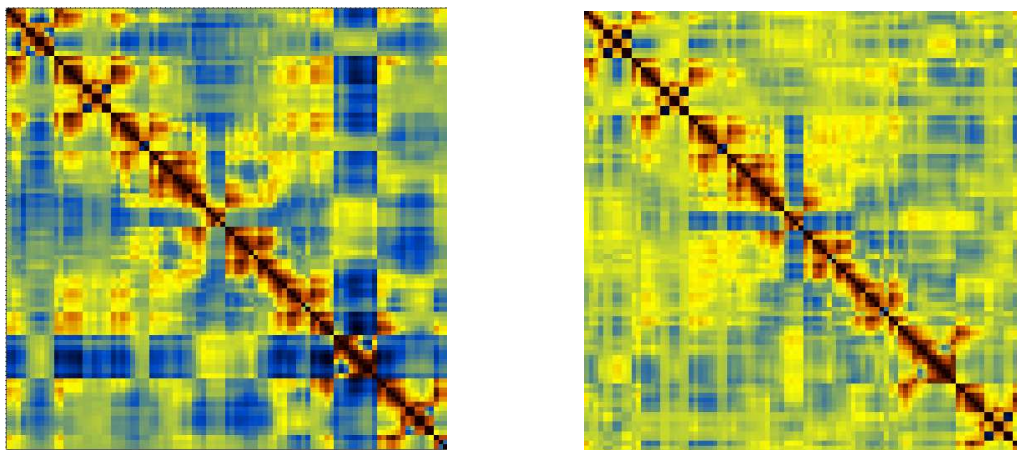
2o cluster: 16809 από τις 3336245 διαμορφώσεις

Το πρώτο cluster αντιστοιχεί στο 49,5% του τροχιακού και το δεύτερο μόλις στο 0,5%.

Στην **Εικόνα 4.2.2.7** παρουσιάζονται τα διαγράμματα ελεύθερης ενέργειας για όλα τα ζεύγη κύριων συνιστωσών και στην **Εικόνα 4.2.2.8** παρουσιάζονται και πάλι οι πίνακες συνδιακύμανσης, όπου αυτή τη φορά επικρατεί το κίτρινο χρώμα γεγονός που προϊδεάζει την συμπεριφορά των πλευρικών ομάδων.

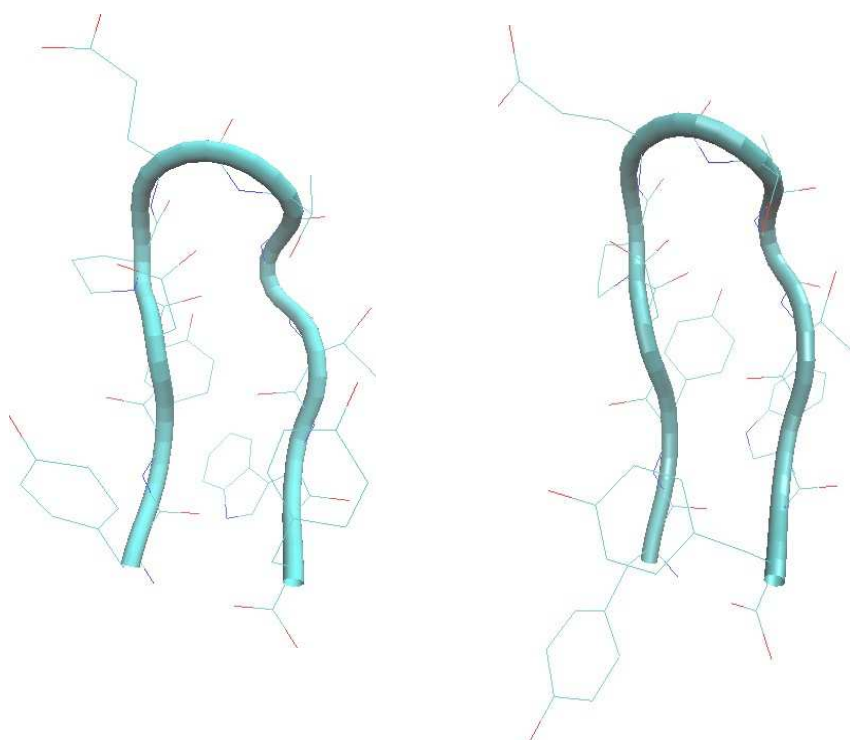


Εικόνα 4.2.2.7: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή cPCA στο ILDN. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα.



Εικόνα 4.2.2.8: Πίνακες συνδιακύμανσης κατά σειρά για το πρώτο και δεύτερο cluster που προέκυψαν από την εφαρμογή cPCA στο ILDN

Οι αντιπροσωπευτικές δομές του κάθε cluster φαίνονται στην **Εικόνα 4.2.2.9**.

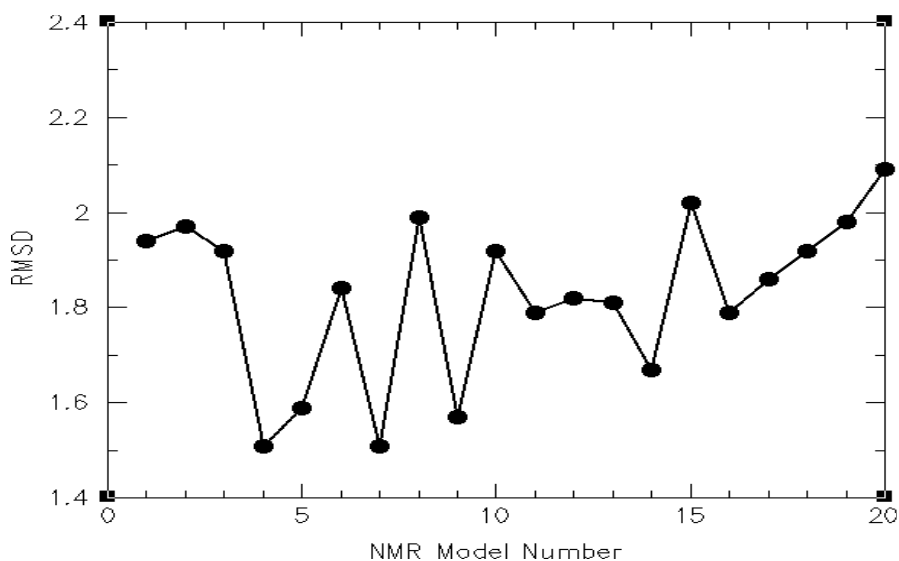


Εικόνα 4.2.2.9: Η αντιπροσωπευτική διαμόρφωση του πρώτου και δεύτερου cluster αντίστοιχα όπως προέκυψε από την cPCA για το τροχιακό ILDN

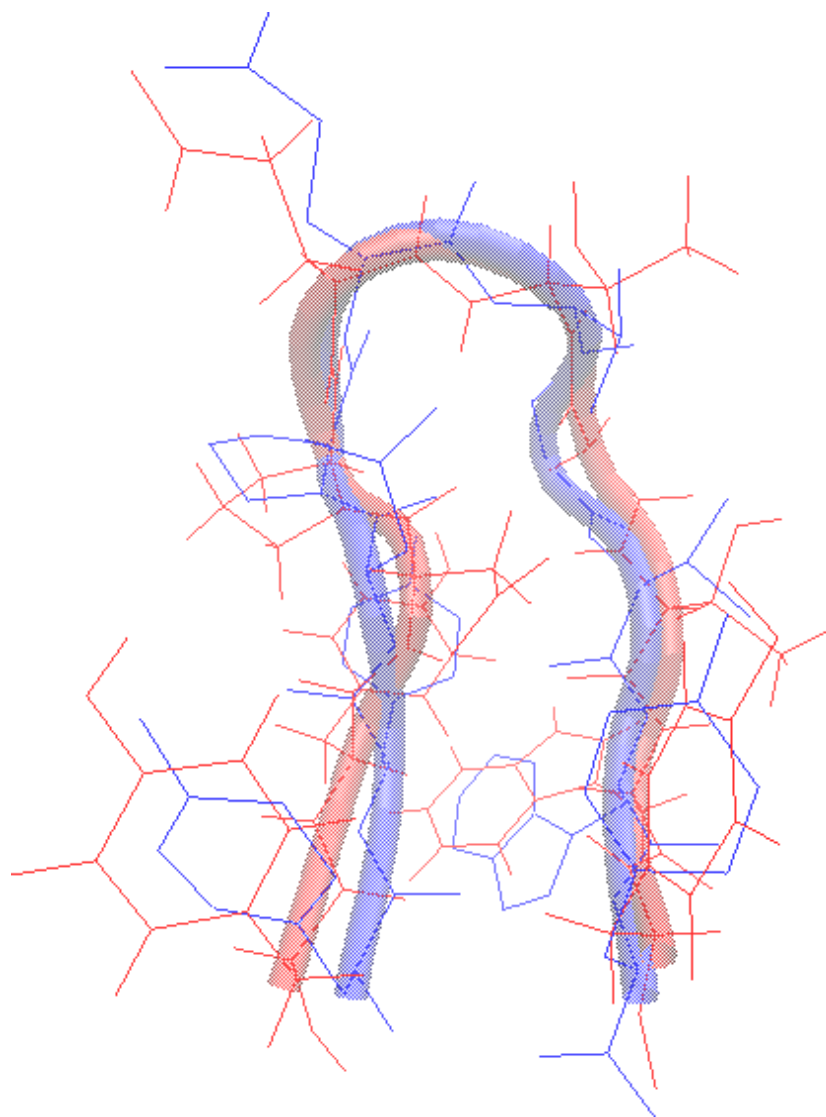
Από την dPCA προέκυψε η επικρατέστερη διαμόρφωση όσον αφορά τον πεπτιδικό σκελετό και έπειτα με την εφαρμογή cPCA αναδείχθηκε η επικρατέστερη διαμόρφωση σε συνδυασμό με τον προσανατολισμό των πλευρικών ομάδων. Με σκοπό να υπολογιστεί πόσο

κοντά στις πειραματικές δομές είναι η διαμόρφωση αυτή πραγματοποιήθηκε υπέρθεση των μοντέλων NMR με την αντιπροσωπευτική δομή του πρώτου cluster (**Εικόνα 4.2.2.10**). Η προσομοίωση φτάνει πιο κοντά στα μοντέλα 4 και 7 για τα οποία η τιμή του RMSD είναι 1.51Å. Στην **Εικόνα 4.2.2.11** παρατίθεται η υπέρθεση μεταξύ του τέταρτου μοντέλου και της αντιπροσωπευτικής δομής του πρώτου cluster. Οι διαφορές εντοπίζονται κυρίως στον προσανατολισμό των καταλοίπων των άκρων.

Η εφαρμογή της Ανάλυσης Κύριων Συνιστωσών στο τροχιακό που εξήγαγε η προσομοίωση με το ILDN φανέρωσε τελικώς τα εξής. Το πεπτίδιο διέρχεται από δύο διαμορφώσεις, τη δομή β-φουρκέτας που έχουν και οι πειραματικές δομές NMR και μια διαμόρφωση που αποτελεί έναν τύπο β-φουρκέτας στην οποία οι δύο κλώνοι έχουν offset ενός πεπτιδίου. Ο πρώτος τύπος διαμορφώσεων καταλαμβάνει συντριπτικά μεγαλύτερο ποσοστό της προσομοίωσης σε σχέση με τον δεύτερο. Μάλιστα συγκρίνοντας την ομάδα διαμορφώσεων με δομή β-φουρκέτας με τα μοντέλα NMR προέκυψε ότι έχουν διαφορά της τάξεως των μόλις 1,51Å. Επομένως προσεγγίζουν τις πειραματικές δομές ικανοποιητικά κοντά.



Εικόνα 4.2.2.10: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή cPCA στο ILDN



Εικόνα 4.2.2.11: Υπέρθυση του τέταρτου μοντέλου NMR (μπλε αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster (κόκκινη αλυσίδα) που προέκυψε από την εφαρμογή dPCA στο ILDN

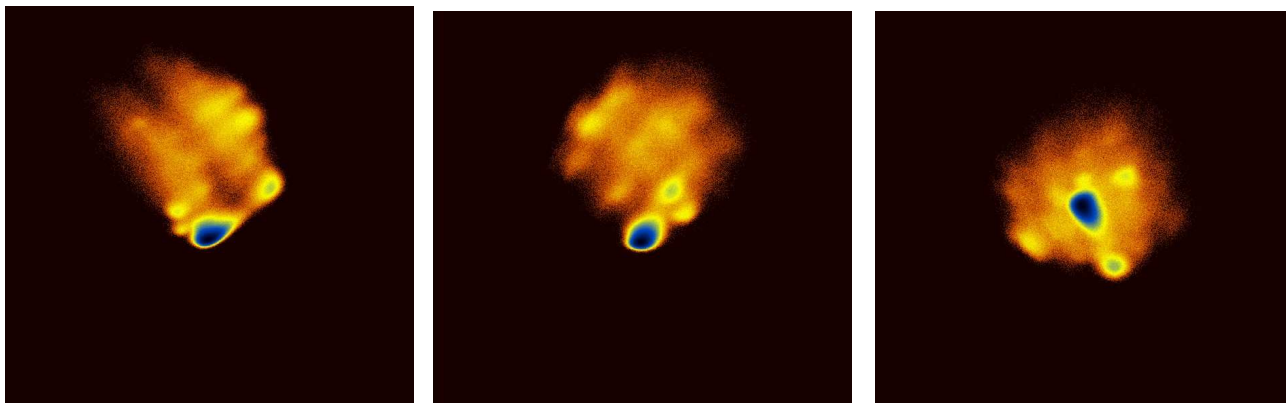
4.2.3. Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό NMR

Η ίδια διαδικασία ακολουθήθηκε και στην εφαρμογή της Ανάλυσης Κύριων Συνιστωσών στο τροχιακό που προέκυψε από την προσομοίωση με το δυναμικό πεδίο NMR. Τα αποτελέσματα παρουσιάζονται ακολούθως.

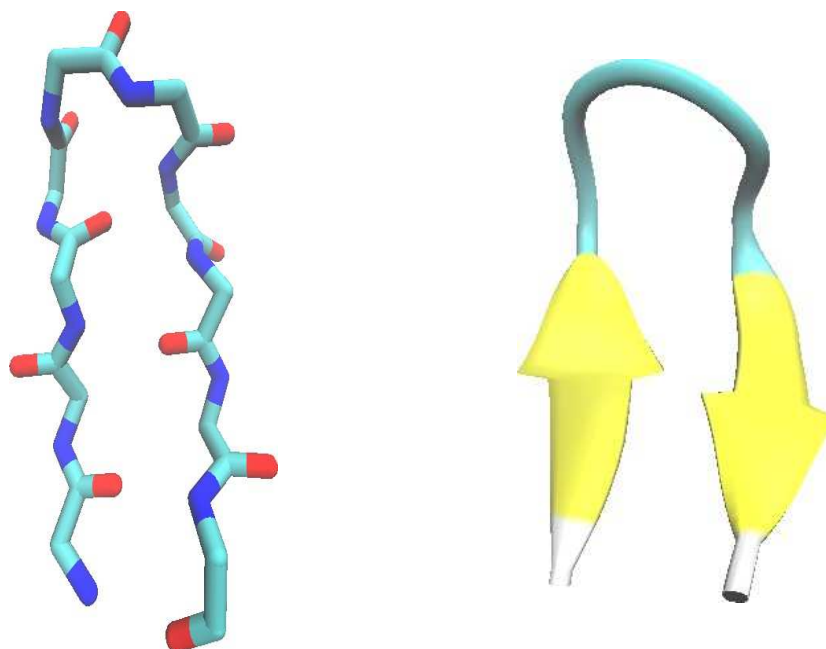
Εφαρμόζοντας dPCA εξήχθη ένα cluster το οποίο καταλαμβάνει το 60,41% του τροχιακού:

1o cluster: 3790133 από τις 6273375 διαμορφώσεις

Παρουσιάζονται τα διαγράμματα ελεύθερης ενέργειας (Εικόνα 4.2.3.1) και η αντιπροσωπευτική διαμόρφωση του cluster (Εικόνα 4.2.3.2). Η διαμόρφωση αυτή έχει την δομή β-φουρκέτας.

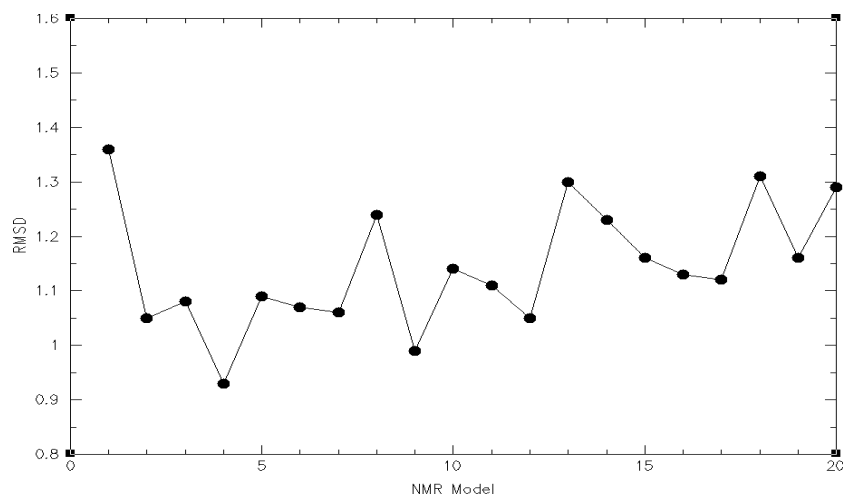


Εικόνα 4.2.3.1: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή dPCA στο NMR. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα

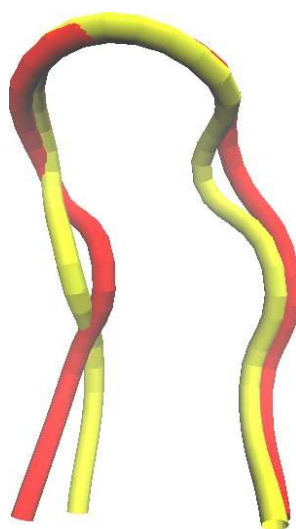


Εικόνα 4.2.3.2: Η αντιπροσωπευτική διαμόρφωση του πρώτου cluster όπως προέκυψε από την dPCA για το τροχιακό NMR σε (A) licorice και (B) cartoon αναπαράσταση

Ακολούθησε η σύγκριση της αντιπροσωπευτικής δομής με τα διαθέσιμα μοντέλα NMR για να υπολογιστεί πόσο κοντά προσεγγίζει τις πειραματικές δομές ο αντιπροσωπευτικός σκελετός του πεπτιδίου της προσομοίωσης. Τα αποτελέσματα (Εικόνα 4.2.3.3) δείχνουν ότι η τιμή RMSD για τα μοντέλα κινείται μεταξύ 0.93Å και 1.36Å. Το μοντέλο με την μικρότερη τιμή RMSD είναι το τέταρτο και παρατίθεται η υπέρθεση των δομών (Εικόνα 4.2.3.4). Η πιο έντονη διαφορά των δύο δομών βρίσκεται στους διαφορετικούς προσανατολισμούς των β-κλώνων στο c-τελικό άκρο.



Εικόνα 4.2.3.3: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο NMR



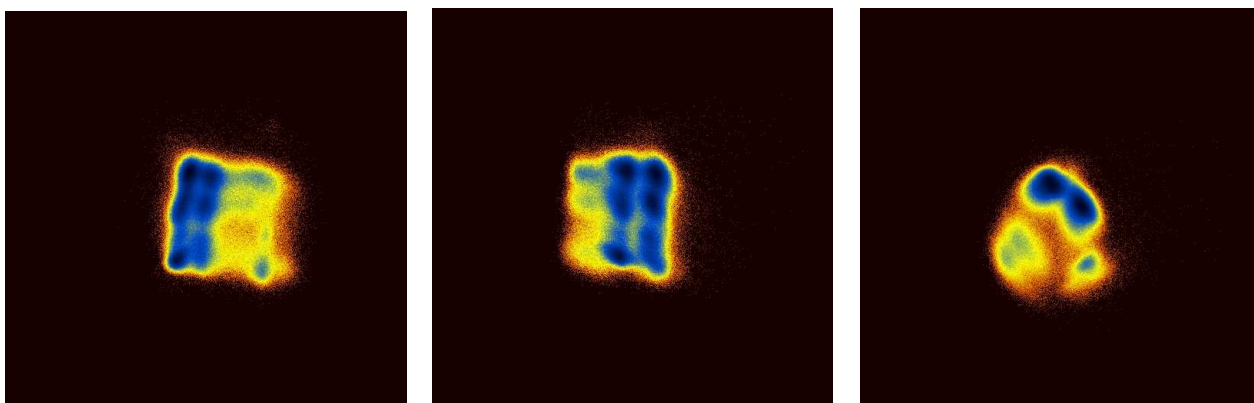
Εικόνα 4.2.3.4: Υπέρθεση του τέταρτου μοντέλου NMR (κόκκινη αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο ILDN (κίτρινη αλυσίδα)

Έπειτα, όπως και πριν, εφαρμόστηκε cPCA για τα heavy άτομα του μορίου στις διαμορφώσεις που περιλαμβάνει το cluster που προέκυψε προηγουμένως. Εξήχθησαν δύο clusters:

1o cluster: 2224040 από τις 3790133 διαμορφώσεις

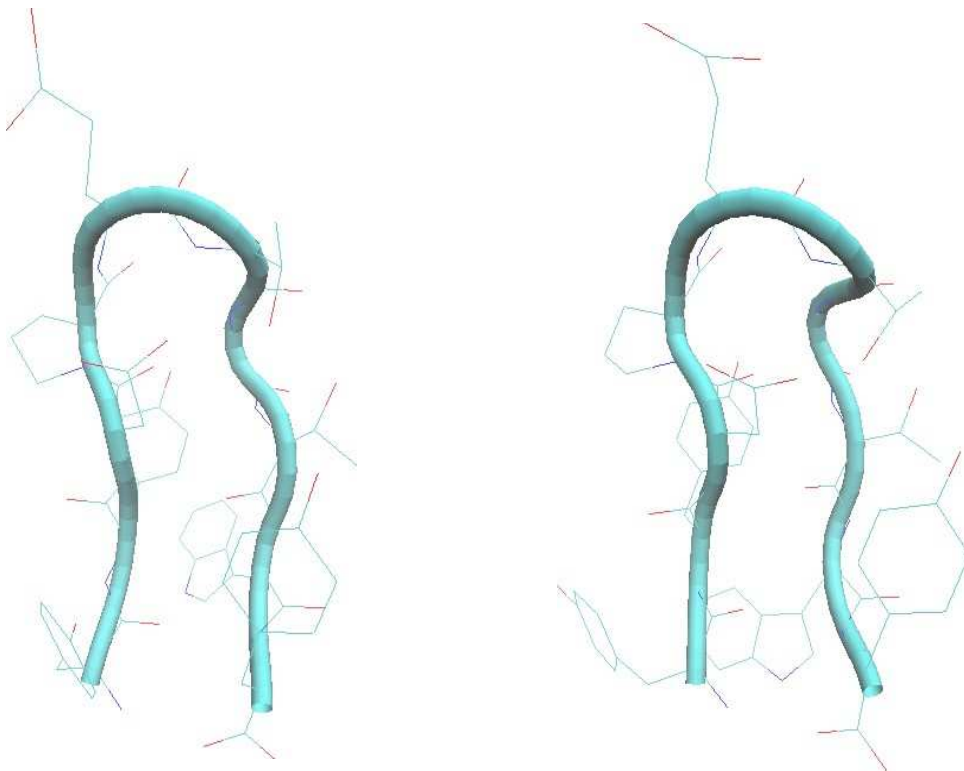
2o cluster: 19812 από τις 3790133 διαμορφώσεις

Το πρώτο αντιστοιχεί στο 58,6% του τροχιακού που εξήγαγε η dPCA και το δεύτερο στο 0.52%. Παρατίθενται τα διαγράμματα ελεύθερης ενέργειας (**Εικόνα 4.2.3.5**) και οι αντιπροσωπευτικές διαμορφώσεις (**Εικόνα 4.2.3.6**).

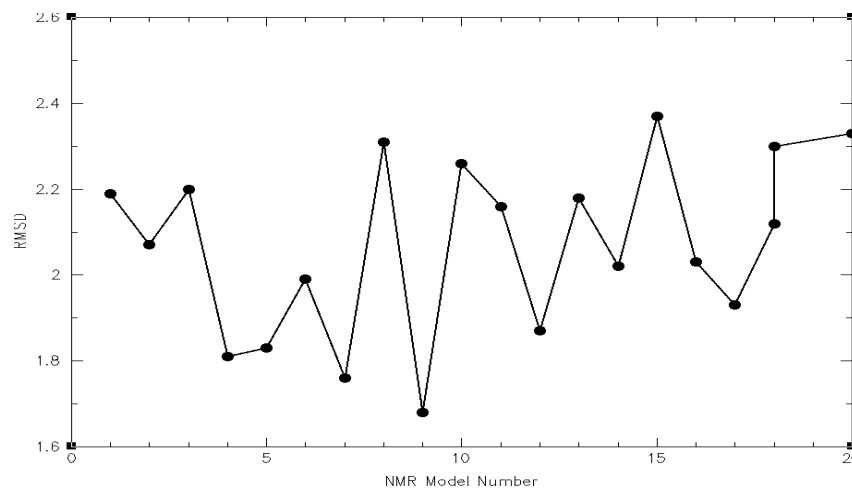


Εικόνα 4.2.3.5: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή cPCA στο NMR. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα.

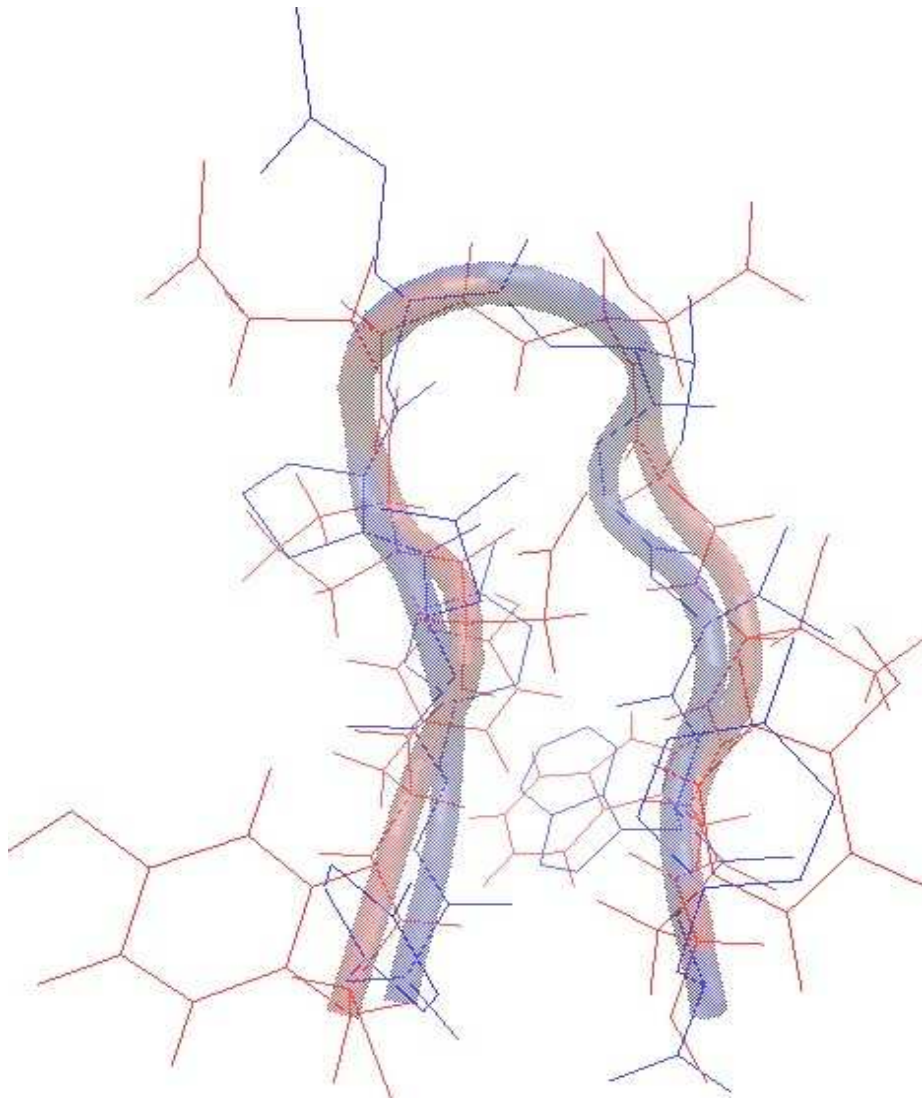
Παίρνοντας την κυρίαρχη διαμόρφωση του πρώτου cluster συγκρίθηκε το πόσο ομοιάζει με τις πειραματικά προσδιορισμένες δομές. Στο διάγραμμα της **Εικόνας 4.2.3.7** φαίνεται ότι οι τιμές του RMSD έχουν ελάχιστο στα 1,68Å και μέγιστο στα 2,37Å. Η υπέρθεση του ένατου μοντέλου, δηλαδή αυτού με το μικρότερο RMSD και της αντιπροσωπευτικής δομής του πρώτου cluster που εξήγαγε η cPCA φανερώνει ότι οι μεγαλύτερες διαφορές αφορούν στα κατάλοιπα Tyr1 και Trp9 (**Εικόνα 4.2.3.8**).



Εικόνα 4.2.3.6: Η αντιπροσωπευτική διαμόρφωση του πρώτου και δεύτερου cluster αντίστοιχα όπως προέκυψε από την cPCA για το τροχιακό NMR



Εικόνα 4.2.3.7: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή cPCA στο NMR



Εικόνα 4.2.3.8: Υπέρθυση του ένατου μοντέλου NMR (μπλε αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster (κόκκινη αλυσίδα) που προέκυψε από την εφαρμογή dPCA στο NMR

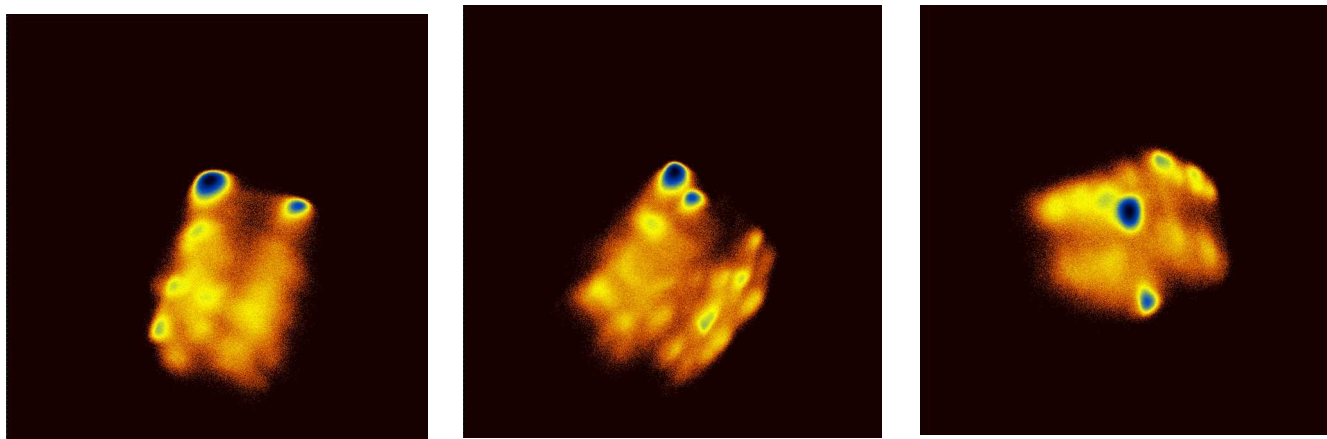
4.2.4. Εφαρμογή Ανάλυσης Κύριων Συνιστωσών στο τροχιακό STAR

Όπως και πριν η ανάλυση ξεκίνησε με εφαρμογή της dPCA στα άτομα του σκελετού του πεπτιδίου, δηλαδή τα C, C α , N, O. Η μέθοδος εντόπισε δύο cluster, το πρώτο καταλαμβάνει το 44,7% του ολικού τροχιακού και το δεύτερο είναι πολύ μικρότερο και καταλαμβάνει μόνο το 3,9%.

1o cluster: 2989452 από τις 6677500 διαμορφώσεις

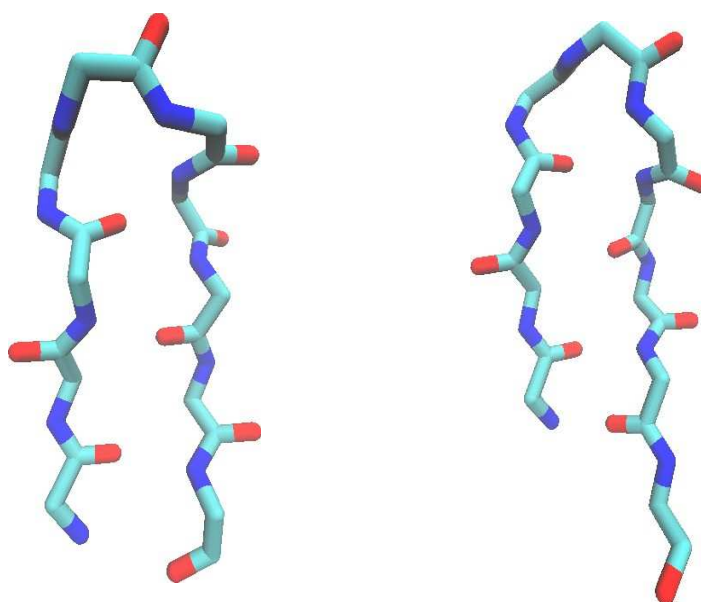
2o cluster: 265822 από τις 6677500 διαμορφώσεις

Παρουσιάζονται τα διαγράμματα ενέργειας ως προς τα ζεύγη των τριών κύριων συνιστωσών (Εικόνα 4.2.4.1).



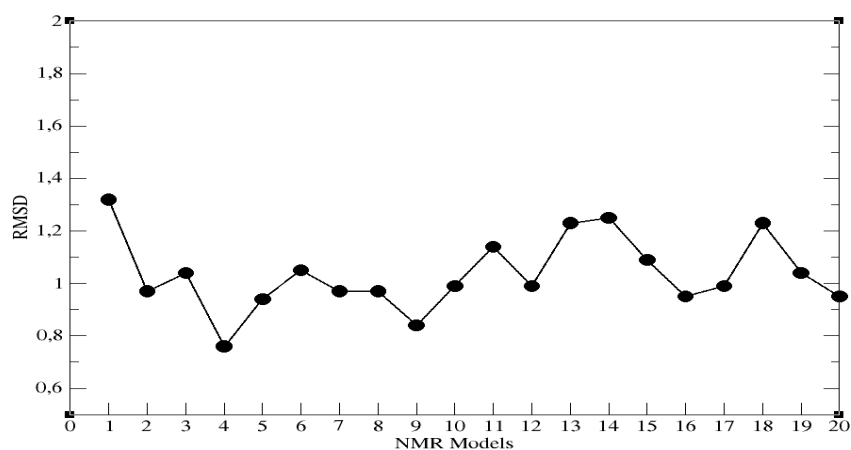
Εικόνα 4.2.4.1: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή dPCA στο STAR. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα

Οι διαμορφώσεις που ανήκουν στο πρώτο cluster έχουν τη δομή β-φουρκέτας που παρατηρήθηκε έως τώρα. Το δεύτερο cluster και σε αυτή την περίπτωση εμφανίζει τη δομή ενός τύπου β-φουρκέτας με όφσεντ ενός πεπτιδίου (Εικόνα 4.2.4.2).

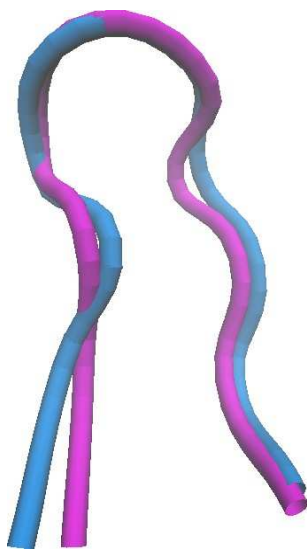


Εικόνα 4.2.4.2: Η αντιπροσωπευτική διαμόρφωση του πρώτου (A) και δεύτερου (B) cluster όπως προέκυψε από την dPCA για το τροχιακό STAR

Ακολούθησε η σύγκριση με την πειραματικά προσδιορισμένη δομή και για το λόγο αυτό έγινε υπέρθεση κάθε μοντέλου NMR με την αντιπροσωπευτική δομή του κυρίαρχου cluster. Το μοντέλο που προσεγγίζεται στον μεγαλύτερο βαθμό είναι το τέταρτο. Το RMSD που του αντιστοιχεί είναι 0,76Å (Εικόνα 4.2.4.3). Παρουσιάζεται επίσης η υπέρθεση των δύο αυτών δομών (Εικόνα 4.2.4.4).



Εικόνα 4.2.4.3: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο STAR



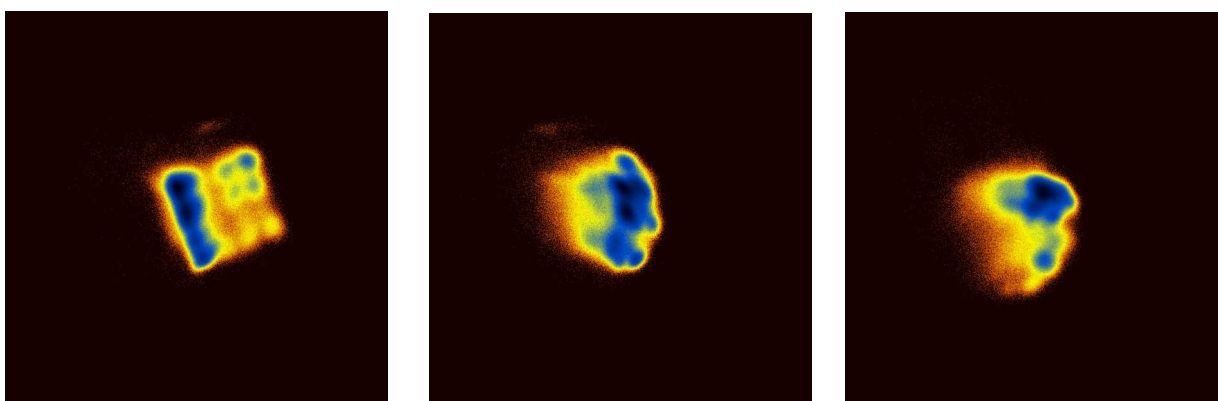
Εικόνα 4.2.4.4: Υπέρθεση του τέταρτου μοντέλου NMR (μπλε αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή dPCA στο STAR (μωβ αλυσίδα)

Από την εφαρμογή της cPCA στο τροχιακό με τις διαμορφώσεις του πρώτου cluster προέκυψαν τα εξής.

1o cluster: 1497255 από τις 2989452 διαμορφώσεις

2o cluster: 38755 από τις 2989452 διαμορφώσεις

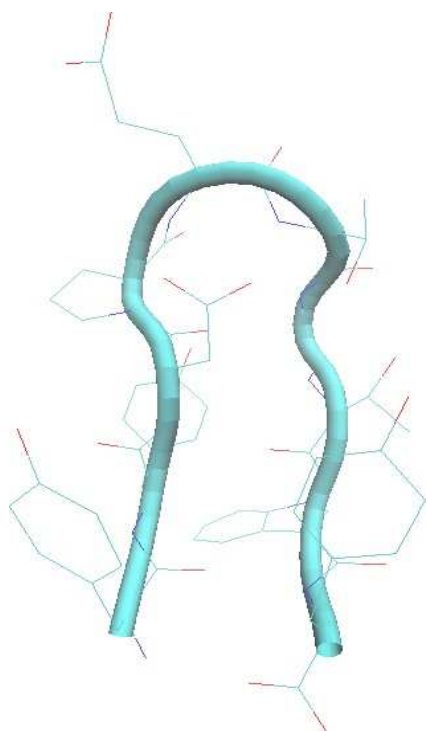
με το πρώτο να περιλαμβάνει το 50% του τροχιακού και το δεύτερο μόλις το 1,2%. Παρατίθενται τα διαγράμματα ελεύθερης ενέργειας (**Εικόνα 4.2.4.5**) και οι αντιπροσωπευτικές διαμορφώσεις (**Εικόνα 4.2.4.6**).



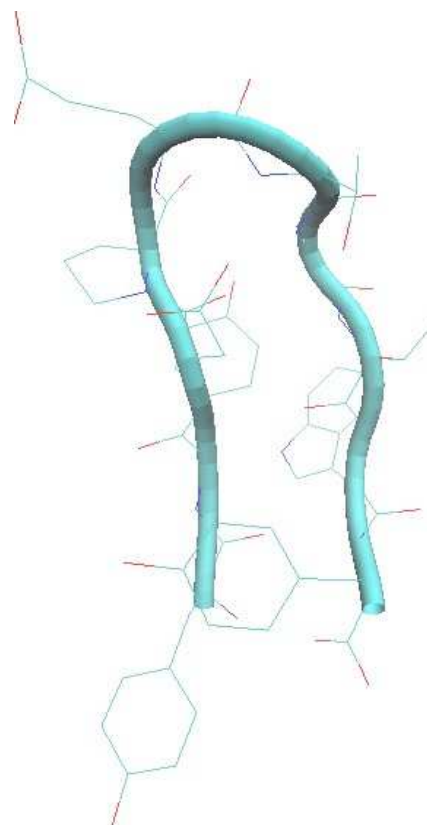
Εικόνα 4.2.4.5: Διαγράμματα ελεύθερης ενέργειας για τα clusters που προέκυψαν από την εφαρμογή cPCA στο STAR. Από τα αριστερά προς τα δεξιά: η πρώτη κύρια συνιστώσα σε σχέση με την δεύτερη, η πρώτη σε σχέση με την τρίτη και τέλος η δεύτερη σε σχέση με την τρίτη κύρια συνιστώσα.

Από τη σύγκριση με τις πειραματικά προσδιορισμένες δομές προκύπτουν τιμές RMSD οι οποίες κυμαίνονται από έως (**Εικόνα 4.2.4.7**). Η μικρότερη τιμή αντιστοιχεί στο τέταρτο μοντέλο. Στην **Εικόνα 4.2.4.8** παρατίθεται η υπέρθεση των δύο δομών. Οι πιο έντονες διαφορές εντοπίζονται στα κατάλοιπα Tyr1 και Tyr10.

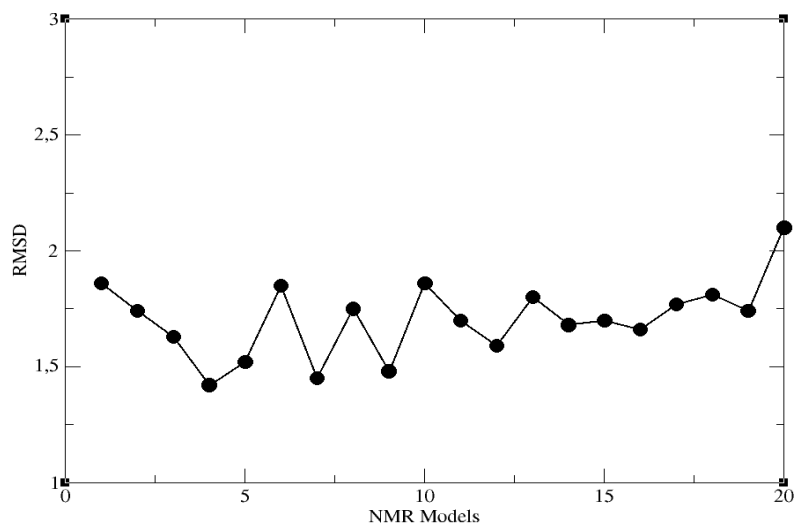
(A)



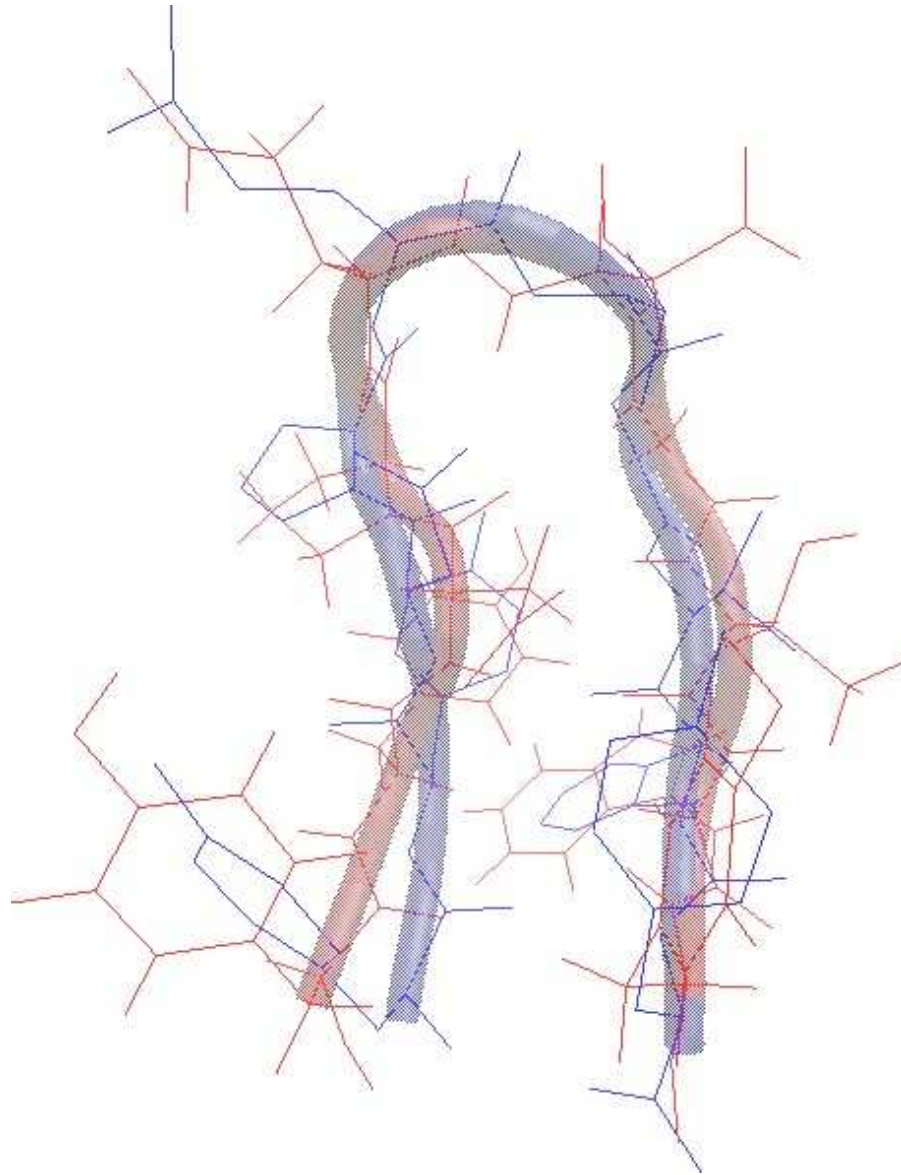
(B)



Εικόνα 4.2.4.6: Η αντιπροσωπευτική διαμόρφωση του πρώτου και δεύτερου cluster αντίστοιχα όπως προέκυψε από την cPCA για το τροχιακό STAR



Εικόνα 4.2.4.7: Διαγραμματική απεικόνιση του RMSD για κάθε μοντέλο NMR από την αντιπροσωπευτική δομή του πρώτου cluster που προέκυψε από την εφαρμογή cPCA στο STAR



Εικόνα 4.2.4.8: Υπέρθηση του τέταρτου μοντέλου NMR (μπλε αλυσίδα) με την αντιπροσωπευτική δομή του πρώτου cluster (κόκκινη αλυσίδα) που προέκυψε από την εφαρμογή dPCA στο STAR

Συνοψίζοντας ως εδώ παρουσιάζονται τα αποτελέσματα στον **Πίνακα 4.1**. Η εφαρμογή της μεθόδου και στις τρεις περιπτώσεις έδειξε ότι η προσομοίωση διέρχεται επανειλημμένα από διαμορφώσεις οι οποίες έχουν τη δομή της β-φουρκέτας. Το τροχιακό NMR είναι αυτό που περιέχει το μεγαλύτερο cluster διαμορφώσεων με αυτή τη δομή. Η δεύτερη ομάδα διαμορφώσεων που εντοπίστηκε αφορούσε ένα τύπο β-φουρκέτας ο οποίος έχει όψετ ενός πεπτιδίου. Αυτός ο τύπος διαμορφώσεων εμφανίζεται συχνότερα στο τροχιακό STAR. Η δομή όμως αυτή καταλαμβάνει ένα πολύ μικρό ποσοστό του τροχιακού και

μάλιστα στην περίπτωση του NMR δεν εντοπίστηκε καν cluster. Συγκρίνοντας τις αντιπροσωπευτικές διαμορφώσεις από κάθε cluster με τα πειραματικά δεδομένα αποδεικνύεται ότι όλες οι προσομοιώσεις πλησιάζουν περισσότερο τη δομή του τέταρτου μοντέλου NMR. Σημειώνεται ότι το τροχιακό με τις μικρότερες τιμές RMSD είναι το STAR.

Πίνακας 4.1

	ILDN		NMR		STAR	
	1o cluster	2o cluster	1o cluster	2o cluster	1o cluster	2o cluster
dPCA	52,8%	3,8%	60,4%	-	44,7%	3,9%
cPCA	49,5%	0,5%	58,6%	0,5%	50,0%	1,2%
RMSD από τη φυσική δομή (Å)						
Backbone	0.89		0.93		0.76	
Heavy	1.51		1.68		1.42	

4.3 | Ανάλυση των προσομοιώσεων με βάση τις “native contacts”

Με σκοπό να αναδειχθούν οι διαφορές μεταξύ των προσομοιώσεων έγινε αναπαράσταση της κατανομής των διαμορφώσεων του κάθε τροχιακού ως προς την θερμοκρασία και την παράμετρο q , ενός μέτρου της ομοιότητας μεταξύ δύο δομών.

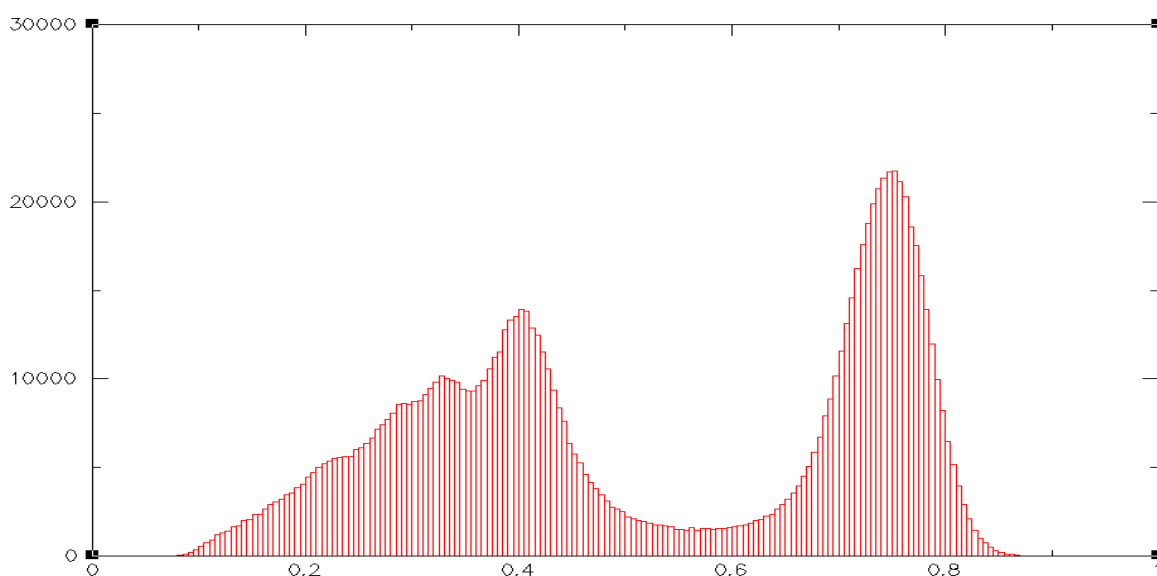
Οι προσομοιώσεις πραγματοποιήθηκαν σε συνθήκες “adaptive tempering”, πράγμα που σημαίνει ότι δεν υπήρχε μία σταθερή θερμοκρασία την οποία διατήρησε το σύστημα κατά την προσομοίωση. Αντιθέτως η θερμοκρασία μεταβαλλόταν μέσω ενός θερμοστάτη ο οποίος λειτουργεί ως εξής. Όταν η δυναμική ενέργεια είναι χαμηλότερη από την τρέχουσα μέση τιμή τότε η θερμοκρασία μειώνεται ενώ όταν η δυναμική τιμή αποκτήσει μεγαλύτερη τιμή από την τρέχουσα μέση τιμή, η θερμοκρασία αυξάνεται. Αυτό που επιτυγχάνεται είναι να υπερβαίνονται τα ενεργειακά φράγματα και η προσομοίωση να επισκέπτεται ενεργειακά ελάχιστα ταχύτερα, εξοικονομώντας έτσι υπολογιστικό χρόνο (*Zhang et al., 2010*). Στον **Πίνακα 4.2** παρουσιάζεται το θερμοκρασιακό εύρος στο οποίο κινήθηκε η κάθε προσομοίωση.

Πίνακας 4.2

	T_{min} (K)	T_{max} (K)
ILDN	276.8544	536.1313
NMR	279.2036	535.0327
STAR	276.9265	530.9874

Με τη χρήση του προγράμματος grcarma έγινε υπολογισμός του κλάσματος των “native contacts”. Μία επαφή μεταξύ δύο ατόμων συμβαίνει όταν η μεταξύ τους απόσταση είναι μικρότερη από μια δεδομένη ακτίνα. Για να αποκτήσει αυτή η επαφή τον χαρακτηρισμό native θα πρέπει να υπάρχει επίσης και στα ισοδύναμα άτομα της δομής-αναφοράς, εν προκειμένω της φυσικής δομής. Το κλάσμα των “native contacts” υπολογίζεται ως ο συνολικός αριθμός native επαφών μια συγκεκριμένη χρονική στιγμή διαιρούμενος με τον συνολικό αριθμό native επαφών στη δομή-αναφορά. Οι τιμές του q κυμαίνονται από 0, όταν πρόκειται για πλήρως ανόμοιες δομές, έως 1 στην περίπτωση που οι δύο δομές ταυτίζονται. Στους υπολογισμούς που πραγματοποιήθηκαν εδώ χρησιμοποιήθηκαν τα άτομα Ca, C, N, O του πεπτιδικού σκελετού, το όριο στην απόσταση μεταξύ δύο ατόμων τέθηκε στα 8Å και ορίστηκε ότι θα πρέπει να διαχωρίζονται από τουλάχιστον 2 κατάλοιπα.

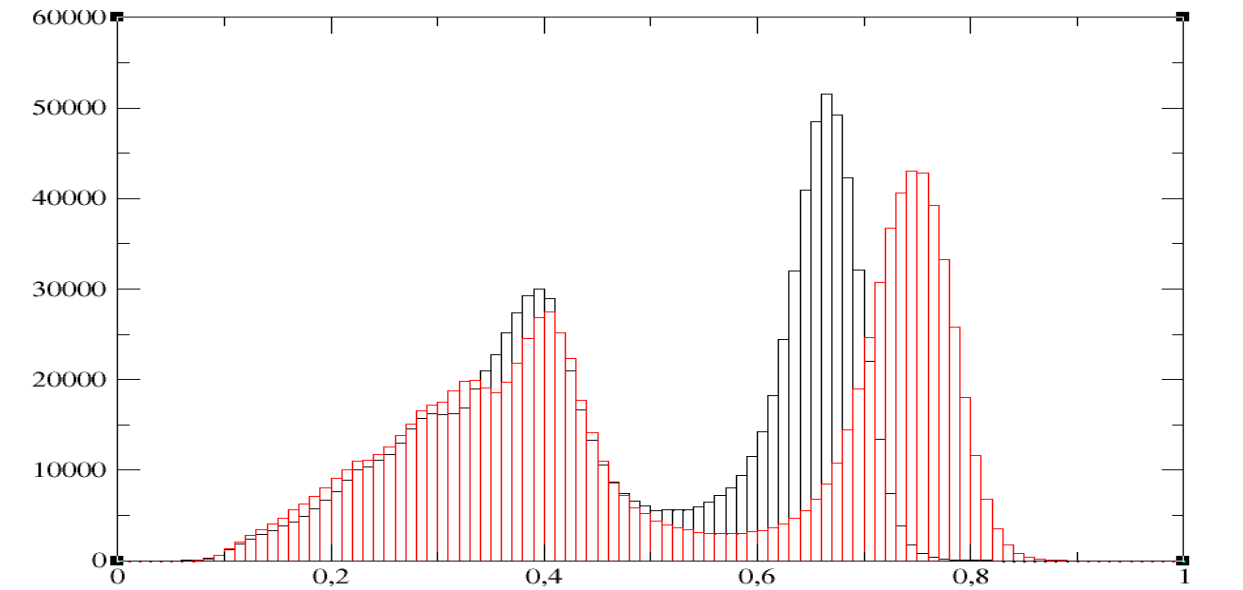
Πραγματοποιήθηκε λοιπόν υπολογισμός του κλάσματος των native επαφών και ως δομή αναφοράς χρησιμοποιήθηκε κάθε φορά ένα από τα είκοσι μοντέλα NMR. Τα αποτελέσματα επεξεργάστηκαν με τον εξής τρόπο. Επιλέχτηκαν 1.000.000 διαμορφώσεις για το κάθε τροχιακό, γεγονός που αντιστοιχεί περίπου στο 16% του συνόλου και δημιουργήθηκαν ιστογράμματα για να εξακριβωθεί σε ποιες τιμές q κυμαίνονται οι διαμορφώσεις. Παρατηρήθηκε ότι τα ιστογράμματα είχαν την μορφή που παρουσιάζεται στην **Εικόνα 4.3.1**. Σε κάθε περίπτωση διακρίνονται δύο κύριες ομάδες διαμορφώσεων με την πρώτη να έχει τιμή $q \approx 0.4$ και να αντιστοιχεί σε offset διαμορφώσεις και τη δεύτερη να έχει τιμή $q \approx 0.75$ και να αντιστοιχεί σε native διαμορφώσεις.



Εικόνα 4.3.1: Ιστόγραμμα όπου παρουσιάζεται το πλήθος των διαμορφώσεων που αντιστοιχεί σε κάθε τιμή q , αφορά το τροχιακό ILDN και δομή-αναφορά είναι το 4ο μοντέλο NMR.

Συγκρίνοντας τα αποτελέσματα που αφορούν στο κάθε μοντέλο προκύπτει ότι το 4ο μοντέλο είναι αυτό που διαθέτει μεγαλύτερο πλήθος διαμορφώσεων σε υψηλότερες τιμές q . Ενδεικτικά στην **Εικόνα 4.3.2** παρουσιάζεται η σύγκριση των μοντέλων 4 και 10.

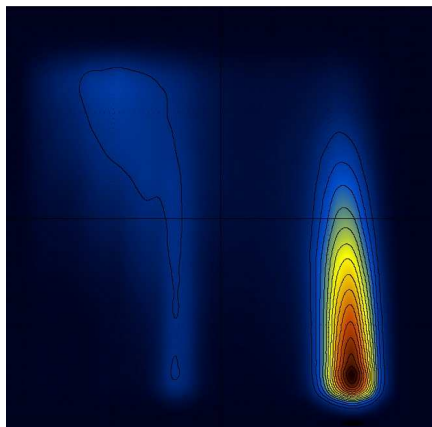
Η ίδια διαδικασία επαναλήφθηκε και για τα τροχιακά NMR και STAR και τα αποτελέσματα ήταν τα ίδια. Με λίγα λόγια το 4ο μοντέλο αποδείχθηκε ότι διαθέτει τον μεγαλύτερο αριθμό διαμορφώσεων σε τιμές q που αντιστοιχούν σε native διαμορφώσεις.



Εικόνα 4.3.2: Ιστόγραμμα του πλήθους των διαμορφώσεων που αντιστοιχούν σε κάθε τιμή q για τα μοντέλα 4 (κόκκινο) και 10 (μαύρο) όσον αφορά το τροχιακό ILDN.

Εν συνεχεία δημιουργήθηκαν πίνακες της θερμοκρασίας T ως προς το κλάσμα των native contacts q . Το κλάσμα των native contacts υπολογίστηκε με δομή-αναφορά την αντιπροσωπευτική δομή που προέκυψε από την εφαρμογή της dPCA σε κάθε cluster. Στην **Εικόνα 4.3.3** παρουσιάζονται δύο πίνακες αυτής της μορφής με σκοπό να περιγραφούν οι ιδιότητές τους. Ο πρώτος πίνακας περιλαμβάνει πυκνότητες και ο δεύτερος τον λογάριθμο αυτών των πυκνοτήτων. Με σκούρο κόκκινο χρώμα αντικατοπτρίζεται η υψηλή πυκνότητα διαμορφώσεων ενώ με πορτοκαλί και κίτρινο η ενδιάμεση και χαμηλή. Στις μπλε περιοχές

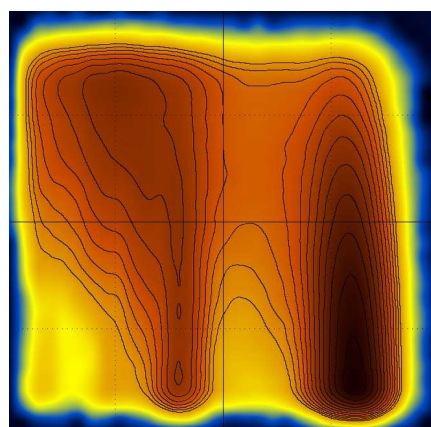
A)



0

q

B)



T

0

q

Εικόνα 4.3.3: A) Πίνακας πυκνοτήτων q - T , B) πίνακας των λογαρίθμων των πυκνοτήτων q - T

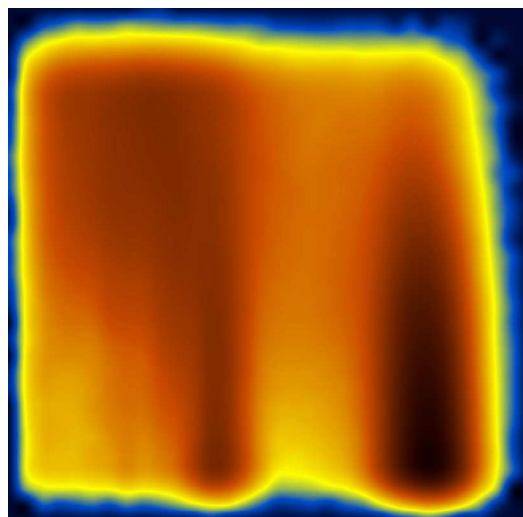
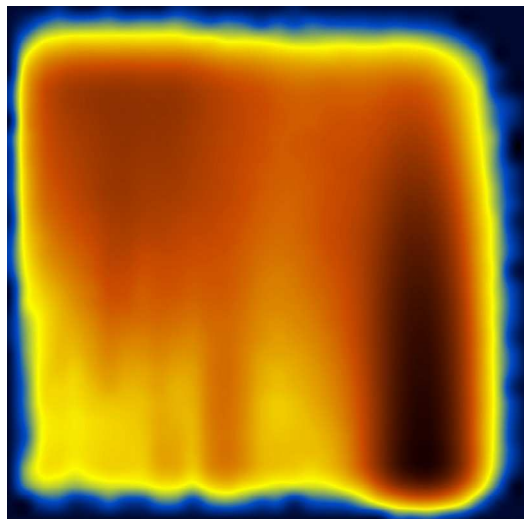
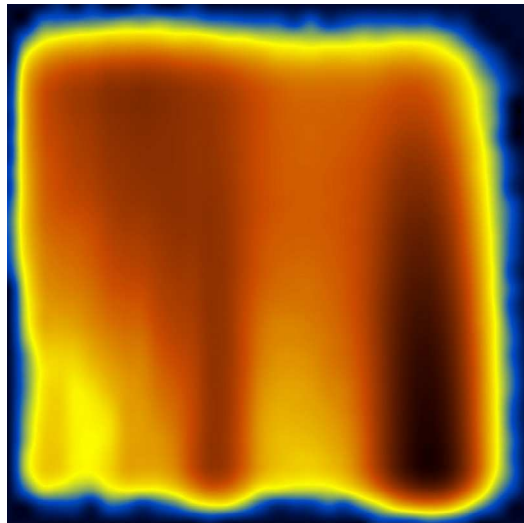
εντοπίζονται ελάχιστες διαμορφώσεις.

Αυτοί οι πίνακες κατασκευάστηκαν για κάθε τροχιακό και στον καθένα αναδεικνύονται σημαντικές λεπτομέρειες (**Εικόνα 4.3.4**). Το κάτω δεξιά τεταρτημόριο των πινάκων είναι η περιοχή με την εντονότερη παρουσία διαμορφώσεων. Οι διαμορφώσεις αυτές αντιστοιχούν σε υψηλές τιμές q και σε χαμηλή έως μέτρια θερμοκρασία. Αμέσως επόμενη περιοχή με σημαντική παρουσία διαμορφώσεων είναι αυτή στο κάτω αριστερά τεταρτημόριο που αντιστοιχεί σε ενδιάμεσες τιμές τόσο q όσο και T . Τέλος αξιοσημείωτη είναι η περιοχή στο άνω αριστερά τεταρτημόριο όπου εντοπίζονται οι μη αναδιπλωμένες διαμορφώσεις από τις οποίες διέρχεται η προσομοίωση.

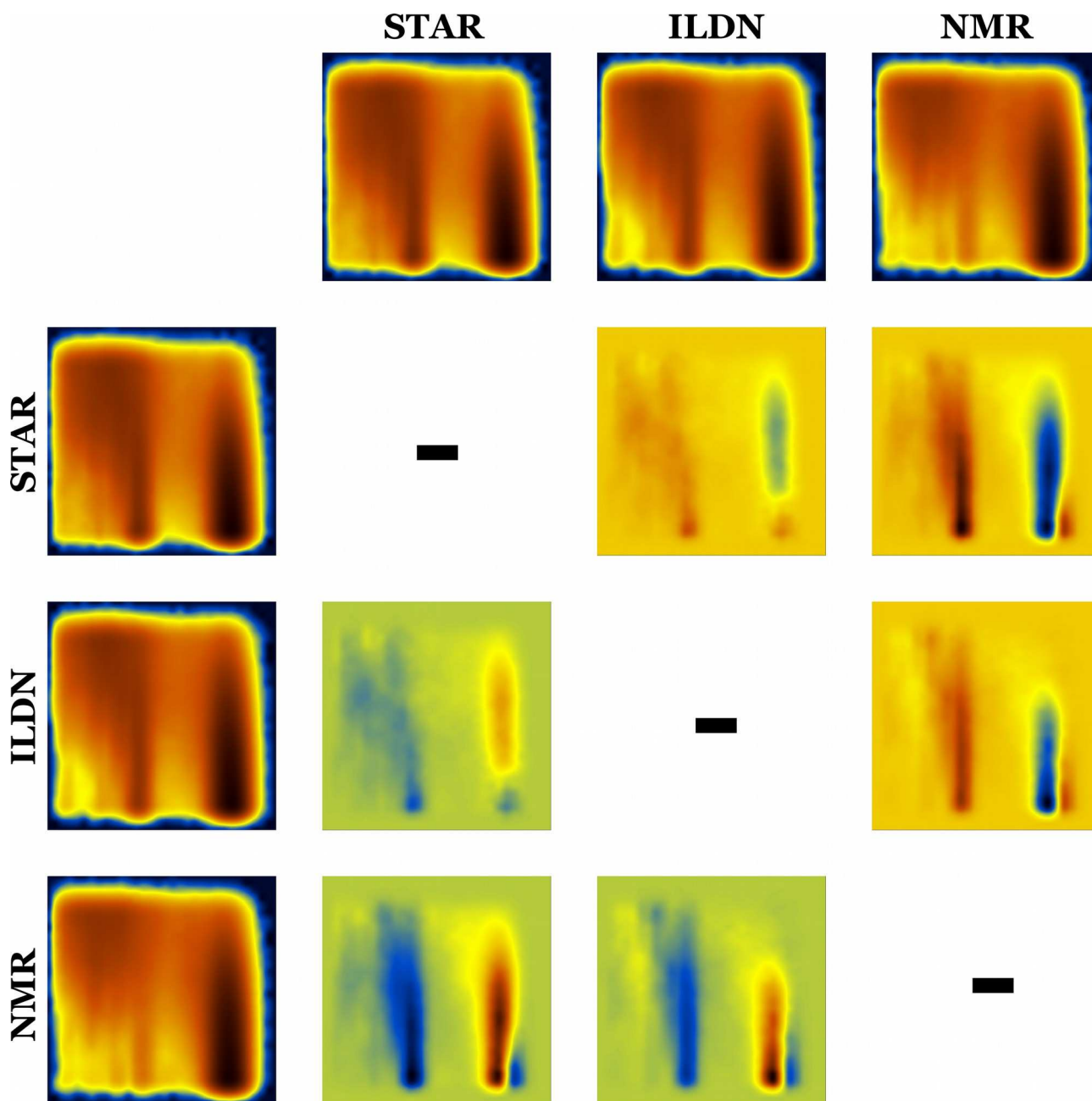
Με σκοπό να αναδειχθούν οι διαφορές μεταξύ των πινάκων η σύγκριση έγινε με τον ακόλουθο τρόπο. Αρχικά άλλαξαν τα μεγέθη των πινάκων έτσι ώστε να περιέχουν τον ίδιο αριθμό διαμορφώσεων μιας και τα τροχιακά έχουν διαφορετικά μεγέθη. Έπειτα έγινε αφαίρεση των πινάκων και και προέκυψαν 6 νέοι πίνακες οι οποίοι φαίνονται στην **Εικόνα 4.3.5**.

Κατ' αρχάς προκύπτει ότι από όλα τα ζεύγη τροχιακών αυτά που είναι μεταξύ τους αρκετά όμοια είναι τα ILDN και STAR. Η μόνη διαφορά έγκειται στο γεγονός ότι στην περίπτωση του STAR η θερμοκρασία τήξης για τις φυσικές διαμορφώσεις είναι σημαντικά μικρότερη. Το τροχιακό NMR είναι αυτό το οποίο διαθέτει τις περισσότερες διαφορές από τα άλλα δύο. Αυτό που ξεχωρίζει είναι ότι περιέχει λιγότερο αριθμό διαμορφώσεων με δομές offset. Αυτό εξηγεί γιατί ο μικρός πληθυσμός αυτών των διαμορφώσεων δεν εντοπίστηκε από την ανάλυση κύριων συνιστωσών. Σημειώνεται ότι το τροχιακό STAR είναι αυτό που διαθέτει την πολυπληθέστερη ομάδα με offset δομές. Η δεύτερη διαφορά που εντοπίζεται στο τροχιακό NMR είναι ότι του λείπει η περιοχή των native δομών που αντιστοιχεί σε υψηλές τιμές q .

Τα παραπάνω αποτελέσματα οδηγούν στο συμπέρασμα ότι οι μικρές αλλαγές στην παραμετροποίηση μεταξύ των τριών δυναμικών πεδίων παρήγαγαν σαφείς διαφορές στα παραγόμενα τροχιακά. Για το λόγο αυτό η προσπάθεια για σύγκριση και επικύρωση των δυναμικών πεδίων θα συνεχιστεί παραβάλλοντας τις προσομοιώσεις με τα πειραματικά δεδομένα.



Εικόνα 4.3.4: Πίνακες q-T για τα τροχιακά ILDN, NMR και STAR κατά σειρά.



Εικόνα 4.3.5: Σύγκριση των πινάκων q-T για τα τρία τροχιακά.

4.4 | Σύγκριση με τα πειραματικά δεδομένα

4.4.1 Εισαγωγή

Η φασματοσκοπία NMR (Nuclear Magnetic Resonance) είναι μία από της κυρίαρχες πειραματικές μεθόδους για τη μελέτη της δομής και της δυναμικής των βιομορίων. Παρατηρήσιμα μεγέθη που προκύπτουν από το NMR, όπως οι χημικές μετατοπίσεις (chemical shifts), οι συχνότητες J-coupling και οι εντάσεις NOE (Nuclear Overhauser Enhancement), μπορούν επίσης να υπολογιστούν και από το τροχιακό μιας προσομοίωσης και έπειτα να συγκριθούν. Οι εντάσεις NOE είναι ιδιαίτερα σημαντικές στο πλαίσιο αυτό γιατί εξαρτώνται αποκλειστικά από τις διατομικές αποστάσεις των υπό μελέτη ατόμων. Έτσι λοιπόν η σύγκριση των παρατηρούμενων από την φασματοσκοπία NMR αποστάσεων που προκύπτουν λόγω του Πυρηνικού Φαινομένου Overhauser με τα δεδομένα που προέκυψαν από τις προσομοιώσεις αποτελεί έναν τρόπο αξιολόγησης της ποιότητας ενός δυναμικού πεδίου. Ειδικότερα στόχος είναι να υπολογιστεί πόσο αποκλίνουν οι αποστάσεις στη δομή του πειραματικού μοντέλου από τις αποστάσεις που προβλέπονται από την προσομοίωση.

Η μέθοδος αυτή στηρίζεται στο φαινόμενο Overhauser. Το φαινόμενο Overhauser είναι η μεταβολή στην ένταση απορρόφησης ενός πρωτονίου όταν γειτονικό ή πλησίον ευρισκόμενο πρωτόνιο ακτινοβολείται. Πιο συγκεκριμένα σε ένα σύστημα δύο πυρήνων X και Y η αλλαγή στην ένταση μιας NMR απορρόφησης του πυρήνα X επέρχεται όταν διαταράσσονται οι πληθυσμοί των ενεργειακών καταστάσεων spin του πυρήνα Y. Αυτό είναι δυνατόν να συμβεί με το λεγόμενο κορεσμό της αντίστοιχης απορρόφησης δηλαδή την ακτινοβολία της, έτσι ώστε να μην υπάρχει περίσσεια πυρήνων στη βασική κατάσταση και κατά συνέπεια να υπάρχουν οι ίδιοι πληθυσμοί στη βασική και διεγερμένη κατάσταση. Σε αυτή την περίπτωση, το όλο σύστημα των πυρήνων X και Y επιχειρεί να επανέλθει σε μία κατάσταση θερμικής ισορροπίας και αυτό ακριβώς είναι και το Πυρηνικό Φαινόμενο Overhauser. Το φαινόμενο αυτό μπορεί επίσης να εκφραστεί με διαφορετικό τρόπο ως η προσπάθεια του συστήματος των πυρήνων X και Y να αντισταθμίσει μία τεχνητή μεταβολή που του προκλήθηκε κατά την ακτινοβολία ενός μέρους του συστήματος. Μία μεταβολή πληθυσμού ενός άλλου μέρους του συστήματος επέρχεται έτσι ώστε τελικά να επικρατήσει πάλι θερμική ισορροπία στο σύστημα. Μία έκφραση του φαινομένου δίνεται από τη σχέση:

$$\text{NOE}(n_{X(Y)}) = I - I_0 / I_0$$

όπου $n_{X(Y)}$ είναι το NOE στην απορρόφηση των πυρήνων X όταν η απορρόφηση των πυρήνων

Υ ακτινοβολείται, I_0 είναι η ένταση του σήματος των πυρήνων X σε κατάσταση θερμικής ισορροπίας και I η ένταση του σήματος των πυρήνων X σε κατάσταση θερμικής ισορροπίας έπειτα από την ακτινοβολήση των πυρήνων Υ.

Η σχέση μεταξύ NOE και της απόστασης r μεταξύ δύο πυρήνων X και Y δίνεται από τη σχέση $NOE = 1/r^6 \cdot f(\tau_c)$, όπου r η απόσταση μεταξύ των δύο πυρήνων και τ_c ο χρόνος που απαιτείται για την περιστροφή του μορίου κατά μία μονάδα ακτίνας 1 (1 radian). Από αυτή τη σχέση προκύπτει ότι για σταθερό τ_c δηλαδή για συγκεκριμένο μόριο και διαλύτη, το NOE αποτέλεσμα μεταξύ δύο πυρήνων είναι αντιστρόφως ανάλογο της έκτης δύναμης της μεταξύ τους απόστασης. Σημειώνεται ότι προκειμένου να υπάρχει NOE η απόσταση των δύο πυρήνων δεν θα πρέπει να ξεπερνά τα 5,5Å. Επομένως το NOE αποτελεί ένα εξαιρετικά σημαντικό εργαλείο στον υπολογισμό ενδομοριακών αποστάσεων μεταξύ διαφόρων ατόμων ενός μορίου.

Για να γίνει ο υπολογισμός των NOEs που προκύπτουν από την προσομοίωση μιας δομής ακολουθείται η εξής διαδικασία. Αρχικά εντοπίζονται τα ζεύγη πυρήνων που εμφανίζουν NOE τόσο στο πείραμα όσο και στην προσομοίωση. Έπειτα καταγράφεται η μεταβολή της δυναμικής τους στο χρόνο. Τελευταίο βήμα είναι η χρήση είτε r^{-3} είτε r^{-6} μέσου όρου για τον υπολογισμό των μέσων αποστάσεων οι οποίες εν συνεχεία θα συγκριθούν με τις πειραματικές αποστάσεις (Daura et al., 1999).

4.4.2 Υπολογισμός και σύγκριση των NOEs

Οι Honda et al., σχεδιαστές του CLN025, έχουν καταθέσει τις αποστάσεις NOE στη βάση δεδομένων BioMagResBank με αριθμό καταχώρησης 20009. Συνολικά το CLN025 διαθέτει 80 NOE constraints και στην **Εικόνα 4.4.2.1** παρουσιάζεται ένα τμήμα αυτών. Με τη σειρά δίνονται ο αριθμός και το όνομα του καταλοίπου στο οποίο ανήκει το πρώτο άτομο, το πρώτο άτομο, ο αριθμός και το όνομα του καταλοίπου στο οποίο ανήκει το δεύτερο άτομο, το δεύτερο άτομο και στην τελευταία στήλη το όριο της απόστασης σε Å.

Για τον υπολογισμό των NOEs που προβλέπουν οι προσομοιώσεις ακολουθήθηκε η παρακάτω διαδικασία. Αρχικά με την χρήση ενός script, του prep_proton.pl, δημιουργήθηκε μια λίστα με όλα τα πιθανά ζεύγη με βάση το αρχείο .psf του πεπτιδίου. Έπειτα η λίστα αυτή συγκρίθηκε με την καταχώρηση της BMRB και αφαιρέθηκαν όσα ζεύγη δεν διέθεταν πειραματικά παρατηρούμενες αποστάσεις. Στη συνέχεια η νέα λίστα, που περιελάμβανε τελικώς 80 ζεύγη, δόθηκε στο πρόγραμμα carma το οποίο πραγματοποίησε τον υπολογισμό

των αποστάσεων για το κάθε τροχιακό.

9 TRP H	9 TRP HB2	3.75
9 TRP HB2	10 TYR H	4.66
8 THR H	8 THR HB	3.61
9 TRP HB3	9 TRP HD1	3.82
9 TRP HB2	9 TRP HD1	3.82
8 THR H	8 THR QG2	4.51
6 THR QG2	7 GLY H	4.75
8 THR QG2	10 TYR H	4.82
2 TYR HB3	3 ASP H	4.82

Εικόνα 4.4.2.1: Εννέα εκ των 80 NOE constraints του CLN025

Λόγω του ότι οι προσομοιώσεις πραγματοποιήθηκαν υπό συνθήκες adaptive tempering τα αρχεία των αποστάσεων που παρήχθησαν επεξεργάστηκαν και δημιουργήθηκαν νέα τα οποία αντιστοιχούν σε καθορισμένες θερμοκρασίες. Συγκεκριμένα δημιουργήθηκαν αρχεία αποστάσεων για τις θερμοκρασίες των 298, 308, 318, 328 και 338 βαθμών Kelvin. Σημειώνεται ότι η φασματοσκοπία NMR πραγματοποιήθηκε στους 298K. Έπειτα πραγματοποιήθηκε ο υπολογισμός των r^{-3} και r^{-6} μέσω αποστάσεων χάρις στο πρόγραμμα noe_averaging.c ως εξής:

$$R = (\langle R_{ij}^{-6} \rangle)^{-1/6} \quad \text{και} \quad R = (\langle R_{ij}^{-3} \rangle)^{-1/3}$$

όπου το R_{ij} είναι λαμβάνει υπόψιν όλους τους συνδυασμούς αποστάσεων μεταξύ των ατόμων i και j . Τελικώς, για κάθε τροχιακό και για όλες τις παραπάνω θερμοκρασίες δημιουργούνται αρχεία που έχουν την μορφή αυτού της **Εικόνας 4.4.2.2**. Αρχικά δίνεται το ζεύγος των δύο ατόμων όπως περιγράφηκε παραπάνω και στις 4 τελευταίες στήλες δίνεται η πειραματική τιμή NOE, ο αριθμός του ζεύγους, η r^{-3} μέση τιμή και η r^{-6} μέση τιμή.

122	126	TRP	9 H	-	TRP	9 HB2	3.75	1	2.536623	2.498245
126	146	TRP	9 HB2	-	TYR	10 H	4.66	2	4.012692	3.881803
108	112	THR	8 H	-	THR	8 HB	3.61	3	2.746508	2.692972
127	130	TRP	9 HB3	-	TRP	9 HD1	3.82	4	2.890638	2.794805
126	130	TRP	9 HB2	-	TRP	9 HD1	3.82	5	3.189525	3.048767
108	114	THR	8 H	-	THR	8 HG21	4.51	6	4.343904	4.059005
93	101	THR	6 HG21	-	GLY	7 H	4.75	7	5.102448	5.036262
114	146	THR	8 HG21	-	TYR	10 H	4.82	8	5.584498	5.237354
30	46	TYR	2 HB3	-	ASP	3 H	4.82	9	4.032014	3.949336
29	46	TYR	2 HB2	-	ASP	3 H	4.82	10	4.061779	3.986985

Εικόνα 4.4.2.2: Τμήμα του αρχείου των NOE αποστάσεων για το τροχιακό NMR στη

θερμοκρασία των 328K

Έχοντας λοιπόν διαθέσιμες τις αποστάσεις τόσο των προσομοιώσεων όσο και του πειράματος ξεκίνησε η διαδικασία σύγκρισης τους. Για το σκοπό αυτό σε πρώτη φάση χρησιμοποιήθηκε ο Συντελεστής Γραμμικής Συσχέτισης (Linear Correlation Coefficient). Η τιμή του συντελεστή κυμαίνεται μεταξύ -1 και +1 και εκφράζει το σθένος της γραμμικής σχέσης μεταξύ δύο μεταβλητών. Ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών X και Y ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων και συμβολίζεται με r ως:

$$r = [\sum xy - (\sum x)(\sum y)] / [\sqrt{(n\sum x^2 - (\sum x)^2)} \sqrt{(n\sum y^2 - (\sum y)^2)}]$$
$$-1 \leq r \leq +1$$

Όταν η τιμή του συντελεστή είναι κοντά στο +1 τότε η συσχέτιση των δύο μεταβλητών είναι θετική ενώ όταν η τιμή είναι κοντά στο -1 τότε η συσχέτιση είναι αρνητική. Στην περίπτωση που η τιμή του συντελεστή είναι κοντά στο 0 τότε δεν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών.

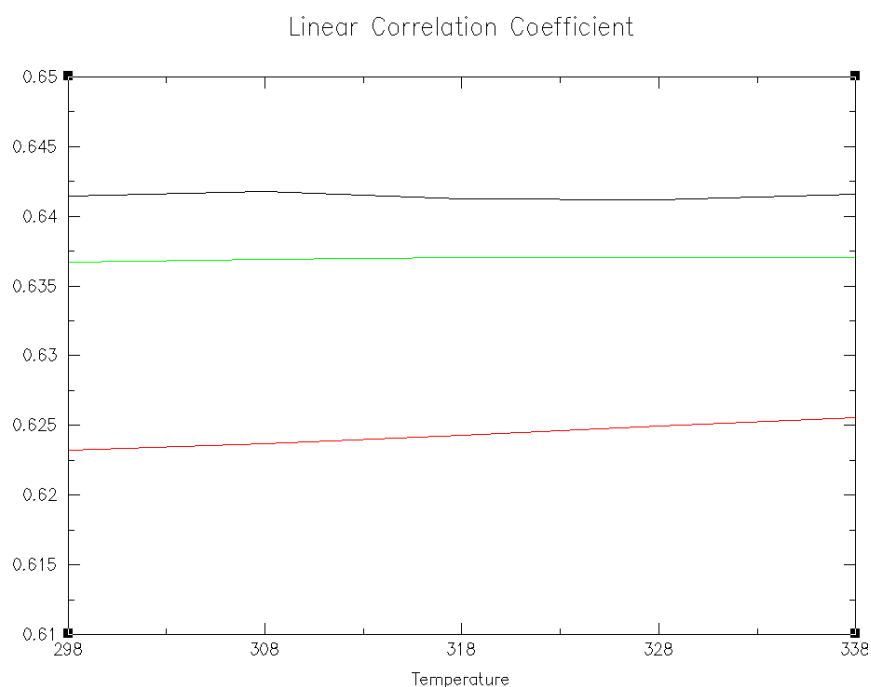
Πραγματοποιήθηκε υπολογισμός του Συντελεστή Γραμμικής Συσχέτισης για να αναδειχθεί το είδος της συσχέτισης δύο μεταβλητών, των πειραματικών αποστάσεων NOE από τη μια και των μέσων NOE αποστάσεων που προέβλεψε η κάθε προσομοίωση από την άλλη. Όπως φαίνεται στον **Πίνακα 4.3** οι τιμές του συντελεστή για τις r⁻³ μέσες αποστάσεις κυμαίνεται μεταξύ 0,62 και 0,64 ενώ για τις r⁻⁶ μέσες αποστάσεις μεταξύ 0,57 και 0,59. Τόσο η r⁻³ όσο και η r⁻⁶ προσέγγιση στον υπολογισμό των μέσων αποστάσεων εμφανίζουν μέση γραμμική συσχέτιση με τις πειραματικές τιμές NOE. Παρ' όλα αυτά η συσχέτιση είναι ισχυρότερη όσον αφορά την r⁻³ προσέγγιση. Επομένως κατά την περαιτέρω σύγκριση χρησιμοποιήθηκε αυτό το σύνολο τιμών.

Πίνακας 4.3

	r ⁻³	r ⁻⁶
298K		
ILDN	0,641469	0,591922
NMR	0,623242	0,573772
STAR	0,636734	0,588794
308K		
ILDN	0,641800	0,592021

NMR	0,623671	0,574703
STAR	0,636916	0,589039
318K		
ILDN	0,641215	0,591370
NMR	0,624294	0,575268
STAR	0,637021	0,588807
328K		
ILDN	0,641212	0,590890
NMR	0,624955	0,575852
STAR	0,637065	0,588699
338K		
ILDN	0,641611	0,590761
NMR	0,625567	0,576270
STAR	0,637068	0,589421

Αντιπαραβάλλοντας τη διακύμανση στην τιμή του συντελεστή γραμμικής συσχέτισης για το κάθε τροχιακό προκύπτουν τα αποτελέσματα που φαίνονται στην Εικόνα 4.4.2.3. Σε όλες τις θερμοκρασιακές τιμές το ILDN είναι το τροχιακό εκείνο που έχει την υψηλότερη γραμμική συσχέτιση NOE αποστάσεων με τις πειραματικές αποστάσεις.



Εικόνα 4.4.2.3: Οι τιμές του Συντελεστή Γραμμικής Συσχέτισης ως προς την θερμοκρασία για το τροχιακό ILDN (μαύρο), NMR (κόκκινο) και STAR (πράσινο)

Επόμενο βήμα ήταν να πραγματοποιηθεί η σύγκριση με τη χρήση της στατιστικής δοκιμής χ^2 . Η μέθοδος αυτή χρησιμοποιήθηκε για να εκτιμηθεί η καλή προσαρμογή (“goodness of fit”) των πειραματικά προσδιορισμένων τιμών με τις τιμές που προέκυψαν από την προσομοίωση. Για τον υπολογισμό του χ^2 ο βασικός τύπος είναι ο ακόλουθος:

$$\chi^2 = \sum_N \left\{ \frac{1}{\sigma_i^2} [y_i - f(x_i)]^2 \right\}$$

όπου N ο αριθμός των δεδομένων, σ_i^2 η διακύμανση, y η ανεξάρτητη και x η εξαρτημένη μεταβλητή. Στην προκειμένη περίπτωση χρησιμοποιήθηκε μια παραλλαγή του παραπάνω τύπου, που ονομάζεται reduced chi-square, στην οποία το χ^2 διαιρείται με τον αριθμό των βαθμών ελευθερίας. Το αποτέλεσμα κρίνεται ως εξής:

- Αν $\chi^2 > 1$ σημαίνει ότι η προσαρμογή δεν έχει συλλάβει πλήρως τα δεδομένα.
- Αν $\chi^2 = 1$ τότε η αντιστοιχία μεταξύ των δύο σετ τιμών είναι σε συμφωνία με τη διακύμανση σφάλματος.
- Τέλος, αν $\chi^2 < 1$ τότε πρόκειται για “υπέρ-προσαρμογή” στα δεδομένα.

Σε αυτό το στάδιο εφαρμόστηκε η δοκιμή χ^2 έτσι ώστε να συγκριθούν οι πειραματικά προσδιορισμένες αποστάσεις NOEs με τις r^{-3} μέσες αποστάσεις NOEs των προσομοιώσεων. Οι παρατηρούμενες NOE αποστάσεις αφορούσαν την θερμοκρασία των 318K. Ο υπολογισμός της διακύμανσης σ_i^2 έγινε με το script noe_averaging_sigma. Στον **Πίνακα 4.4** δίνονται τα αποτελέσματα που παρήχθησαν.

Πίνακας 4.4

	reduced χ^2
ILDN	2,77
NMR	3,28
STAR	2,40

Οι παραπάνω τιμές χ^2 είναι σαφώς μεγαλύτερες του 1 και αντιστοιχούν σε κακή προσαρμογή. Παρ' όλα αυτά όμως το τροχιακό STAR κατάφερε να επιδείξει την καλύτερη προσαρμογή από τα τρία. Σημειώνεται επίσης η εξής παρατήρηση. Κατά τον υπολογισμό του χ^2 υπήρξαν συγκεκριμένα ζεύγη τα οποία και στις τρεις περιπτώσεις είχαν την μεγαλύτερη συμβολή στον υπολογισμό του συνόλου Σ. Πρόκειται για το ζεύγος 12 H-HB3 στο αμινοξύ

ASP 3, για το ζεύγος 35 H-HA μεταξύ ASP 3 και TYR 2 και τέλος για το ζεύγος 65 HB2-HD2 μεταξύ TYR 2 και PRO 4. Επομένως η μεγαλύτερη ασυμφωνία παρατηρείται στα κατάλοιπα 2-4 του πεπτιδίου.

Η σύγκριση με βάση τον Συντελεστή Γραμμικής Συσχέτισης φανερώνει ότι υψηλότερη συσχέτιση με τις πειραματικές αποστάσεις διαθέτει το τροχιακό ILDN ενώ σύμφωνα με τη σύγκριση με βάση το reduced χ^2 καλύτερη προσαρμογή επιδεικνύει το STAR. Η ασυνέπεια μεταξύ των δύο μεθόδων σύγκρισης μπορεί να οφείλεται στο εξής. Η εξίσωση για τον υπολογισμό του χ^2 περιλαμβάνει την τυπική απόκλιση, κάτι που δεν συμβαίνει στην περίπτωση του Συντελεστή Γραμμικής Συσχέτισης. Καθώς το STAR, όπως έχει ήδη αναφερθεί, διαθέτει μεγαλύτερο αριθμό διαμορφώσεων με offset δομή οι τιμές της τυπικής απόκλισης είναι μεγαλύτερες και κατά συνέπεια το χ^2 μικρότερο σε σύγκριση με το ILDN.

Κεφάλαιο 5

Συμπεράσματα

Η παρούσα εργασία είχε ως στόχο να συγκρίνει και να επικυρώσει την αποτελεσματικότητα τριών διαφορετικών δυναμικών πεδίων της οικογένειας Amber, των ff99sb-ildn, ff99sb-ildn-nmr και ff99sb*-ildn. Για να γίνει αυτό διεξήχθησαν προσομοιώσεις μοριακής δυναμικής που μελετούσαν την αναδίπλωση ενός πεπτιδίου δέκα καταλοίπων, του CLN025. Το πεπτίδιο αυτό λογίζεται ως μίνι-πρωτεΐνη και διαθέτει μια καλά καθορισμένη δομή β-φουρκέτας.

Έχοντας πραγματοποιήσει τις προσομοιώσεις η ανάλυση των τροχιακών που προέκυψαν έγινε με τους εξής τρόπους. Αρχικά η σύγκριση έγινε με βάση το RMSD και εξετάστηκε η σταθερότητα του μορίου κατά την προσομοίωση. Οι δομές που περιλαμβάνονται και στα τρία τροχιακά είναι σταθερές σε εκτενή διαστήματα καθ' όλη τη διάρκεια της προσομοίωσης και προσομοιάζουν αρκετά σε αναδιπλωμένες διαμορφώσεις. Εν συνεχεία εφαρμόστηκε η τεχνική της Ανάλυσης Κύριων Συνιστωσών με την οποία αναδείχθηκε ότι η προσομοίωση διέρχεται επανειλημμένα από δύο τύπους διαμορφώσεων. Επικρατής δομή είναι αυτή της β-φουρκέτας ενώ η δεύτερη και κατά πολύ μικρότερη ομάδα διαμορφώσεων έχει τη δομή β-φουρκέτας με όφσεντ ενός καταλοίπου. Από τα 20 διαθέσιμα μοντέλα NMR για το CLN025 οι προσομοιώσεις προσομοιάζουν περισσότερο τη δομή του τέταρτου και μάλιστα το τροχιακό STAR επιτυγχάνει να έχει RMSD της τάξεως των 1,41Å από αυτό. Έπειτα δημιουργήθηκαν πίνακες των native contacts ως προς την θερμοκρασία και έτσι φανερώθηκαν οι λεπτές διαφορές ανάμεσα στα τρία τροχιακά. Το NMR διαφέρει περισσότερο από τα άλλα δύο. Περιέχει ελάχιστες διαμορφώσεις με την όφσεντ δομή και του λείπει επίσης ένα σημαντικό τμήμα διαμορφώσεων με υψηλές τιμές q , δηλαδή φυσικές διαμορφώσεις. Τα τροχιακά STAR και ILDN είναι αρκετά όμοια μεταξύ τους όμως το STAR διαθέτει χαμηλότερο σημείο τήξης και επομένως λιγότερη υπερσταθεροποίηση της φυσικής δομής του πεπτιδίου.

Επόμενο στάδιο ήταν να γίνει η άμεση σύγκριση με τα δεδομένα από τα πειράματα NMR του πεπτιδίου. Συγκρίθηκε λοιπόν πόσο αποκλίνουν οι πειραματικές αποστάσεις NOE από τις αποστάσεις που υπολογίστηκαν στην προσομοίωση. Με τη χρήση της δοκιμασίας χ^2 αντιπαραβλήθηκαν τα δύο σετ δεδομένων και προέκυψε ότι το STAR είναι αυτό που περιγράφει καλύτερα τα πειραματικά δεδομένα. Επομένως μπορούμε να αποφανθούμε ότι η προσομοίωση που διεξήχθη με το τροχιακό STAR επιτυγχάνει να αναπαράγει τα πειραματικά δεδομένα σε καλύτερο βαθμό απ' ό,τι οι άλλες δύο παραλλαγές του δυναμικού πεδίου Amber.

Βιβλιογραφία

- Alder BJ, Wainwright TE, Phase Transition for a Hard Sphere System, 1957, J. Chem. Phys. 27, 1208
- Altis A, Nguyen PH, Hegger R, Stocka G, Dihedral angle principal component analysis of molecular dynamics simulations, 2007, J. Chem. Phys. 126, 244111
- Anfinsen CB, Haber E, Sela M, Whitw FH Jr, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, Proc Natl Acad Sci U S A, 1961 Sep 15, 47:1309-14
- Beauchamp KA, Yu-Shan Lin, Rhiju Das, and Vijay S. Pande, Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements, J. Chem. Theory Comput., 2012, 8 (4), pp 1409–1414
- Brooks CL, Pettitt BM, Karplus M, Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids, 1985, J Chem Phys 83:5897–5908
- Bryngelson JD, Wolynes PG, Spin glasses and the statistical mechanics of protein folding, Proc Natl Acad Sci U S A. 1987 Nov;84(21):7524-8
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA , A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, 1995, J Am Chem Soc 117:5179–5197
- Daura X, Antes I, van Gunsteren WF, Thiel W, Mark AE, The effect of motional averaging on the calculation of NMR-derived structural properties, Proteins. 1999 Sep 1;36(4):542-55
- Davis CM, Xiao S, Raleigh DP, Dyer RB, Raising the speed limit for β -hairpin formation, J Am Chem Soc. 2012 Sep 5;134(35):14476-82
- Davis CM, Dyer RB, Dynamics of an ultrafast folding subdomain in the context of a larger protein fold, J Am Chem Soc. 2013 Dec 26;135(51):19260-7
- Dill KA, Dominant forces in protein folding, Biochemistry. 1990 Aug 7;29(31):7133-55
- Dill KA, Chan HS, From Levinthal to pathways to funnels, Nat Struct Biol. 1997 Jan, 4(1):10-9
- Dill KA, Polymer principles and protein folding, Protein Sci. 1999 Jun;8(6):1166-80
- Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M, Understanding protein folding via free-energy surfaces from theory and experiment, Trends Biochem Sci. 2000

Jul;25(7):331-9

- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG, A smooth particle mesh Ewald potential, 1995, J Chem Phys 103:8577–8592
- Glykos NM, Software news and updates carma: a molecular dynamics analysis program, 2006, J. Comput. Chem. 27, 1765-1768
- Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, García AE, Peptide folding simulations, Curr Opin Struct Biol. 2003 Apr;13(2):168-74
- Hatfield MP, Murphy RF, Lovas S, Molecular dynamics analysis of the conformations of a beta-hairpin miniprotein, J Phys Chem B. 2010 Mar 4;114(8):3028-37
- Hatfield MP, Murphy RF, Lovas S, VCD spectroscopic properties of the beta-hairpin forming miniprotein CLN025 in various solvents, Biopolymers. 2010 May;93(5):442-50
- Hatfield MP, Murphy RF, Lovas S, The CLN025 decapeptide retains a β -hairpin conformation in urea and guanidinium chloride, J Phys Chem B. 2011 May 5;115(17):4971-81
- Honda Shinya, Toshihiko Akiba, Yusuke S. Kato†, Yoshito Sawada, Masakazu Sekijima, Miyuki Ishimura, Ayako Ooishi, Hideki Watanabe, Takayuki Odahara and Kazuaki Harata, Crystal Structure of a Ten-Amino Acid Protein, J. Am. Chem. Soc., 2008, 130 (46), pp 15327–15331
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al., Comparison of multiple Amber force fields and development of improved protein backbone parameters, 2006, Proteins 65: 712–725
- Jorgensen WL., Chandrasekhar J., Madura J.D., Impey R. W. & Klein M. L., Comparison of simple potential functions for simulating liquid water, 1983, J. Chem. Phys. 79, 926-935
- Jorgensen WL, Maxwell DS, TiradoRives J, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, 1996, J Am Chem Soc 118:11225–11236
- Karplus M, Weaver DL, Protein-folding dynamics, Nature, 1976 Apr 1, 260(5550):404-6
- Kleywegt G.J., Zou J.Y., Kjeldgaard M. & Jones T.A. (2001). Around O. In: "International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules"

(Rossmann, M.G. & Arnold, E., Editors). Chapter 17.1, pp. 353-356, 366-367. Dordrecht: Kluwer Academic Publishers, The Netherlands

- Koukos PI, Glykos NM, grcarma: A Fully Automated Task-Oriented Interface for the Analysis of Molecular Dynamics Trajectories, 2013, *J. Comput. Chem.* 34: 2310-2312
- Levinthal C, Are there pathways for protein folding?, *Extrait du Journal de Chimie Physique*, 1968, 65 n° 1, p. 44
- Li D, Bruschweiler R. *Angew. Chem.* 2010;122:6930–6932
- Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field, 2010, *Proteins* 78: 1950–1958
- Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE, Systematic validation of protein force fields against experimental data, *PLoS One.* 2012; 7(2):e32131
- MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, and Karplus, M., 1998, All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102, 3586–3616
- McCammon JA, Gelin BR, Karplus M, Dynamics of folded proteins, *Nature.* 1977 Jun 16;267(5612):585-90
- Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, 2004, *J Comput Chem* 25:1656–1676
- Pang YP, Low-mass molecular dynamics simulation: a simple and generic technique to enhance configurational sampling, *Biochem Biophys Res Commun.* 2014 Sep 26;452(3):588-92
- Phillips JC, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781-1802,

2005

- Piana S, Lindorff-Larsen K, Shaw DE, How robust are protein folding simulations with respect to force field parameterization?, 2011, *Biophys J* 100: L47–L49
- Ptitsyn, OB, Protein folding: hypotheses and experiments, *J. Prot. Chem.* 1987, 6, 273-293
- Rodriguez A, Mokoema P, Corcho F, Bisetty K, Perez JJ, Computational study of the free energy landscape of the miniprotein CLN025 in explicit and implicit solvent, *J Phys Chem B.* 2011 Feb 17;115(6):1440-9
- Shrinivasan R, RIBOSOME, <http://www.roselab.jhu.edu/~raj/Manuals/ribosome.html>
- Tironi IG, Sperb R, Smith PE, Vangunsteren WF, A generalized reaction field method for molecular-dynamics simulations, 1995, *J Chem Phys* 102:5451–5459
- Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P, A new force field for molecular mechanical simulation of nucleic acids and proteins, 1984, *J Am Chem Soc* 106:765–784
- Wetlaufer DB, Nucleation, rapid folding, and globular intrachain regions in proteins, *Proc Natl Acad Sci U S A*, 1973 Mar, 70(3):697-701
- Zhang C. & Ma J. Enhanced sampling and applications in protein folding in explicit solvent, 2010, *J. Chem. Phys.* 132, 244101
- Zhao GJ, Cheng CL, Molecular dynamics simulation exploration of unfolding and refolding of a ten-amino acid miniprotein, *Amino Acids.* 2012 Aug;43(2):557-6
- Zwanzig R, Szabo A, Bagchi B, Levinthal's paradox, *Proc Natl Acad Sci U S A*, 1992 Jan 1, 89(1):20-2