

Δημοκρίτειο Πανεπιστήμιο Θράκης  
Τμήμα Μοριακής Βιολογίας και Γενετικής

## Διπλωματική εργασία

Προσομοιώσεις μοριακής δυναμικής ενός FMRF-amide-like  
πεπτιδίου του *C. elegans*.

*Κουμουνδούρου Ευσταθία Άρτεμις*  
*AEM: -1109-*

*Επιβλέπων Καθηγητής:*  
*Νικόλαος Μ. Γλυκός*

## Ευχαριστίες

Ξεκινώντας τη συγγραφή της παρούσας πτυχιακής εργασίας θα ήθελα να ευχαριστήσω και να εκφράσω την βαθειά ευγνωμοσύνη μου στα άτομα που στάθηκαν δίπλα μου στη δύσκολη αυτή πορεία. Αρχικά νιώθω την ανάγκη να ευχαριστήσω τα άτομα που για 22 ολόκληρα χρόνια με στήριξαν τόσο οικονομικά, αλλά κυρίως ψυχολογικά σε όποια απόφαση πήρα στη ζωή μου, σωστή ή λανθασμένη και με βοήθησαν να φτάσω σήμερα εδώ που είμαι και να εκπληρώσω ένα μέρος των ονείρων μου. Δε θα μπορούσα να μιλώ παρά για τους γονείς μου Ντόρα και Παναγιώτη και την αδερφή μου Μυρτώ. Οι άνθρωποι αυτοί είναι που κυρίως συνέβαλαν στη διαμόρφωση του χαρακτήρα και της προσωπικότητας μου και σίγουρα ένα μέρος των επιτυχιών μου είναι δικό τους έργο. Τα λόγια δεν αρκούν για να εκφράσω όλα όσα μου έχουν προσφέρει και το πόσο τους εκτιμώ γι' αυτό.

Στη συνέχεια θα ήθελα να ευχαριστήσω τις καλές μου φίλες Ταμάρα και Μαρία που υπήρξαν στο πλευρό μου σε δύσκολες στιγμές και μαζί δημιουργήσαμε αξέχαστες αναμνήσεις, κάνοντας τη ζωή μου και τη διαμονή μου στην πόλη αυτή μια όμορφη εμπειρία. Όμως δε θα μπορούσα να παραλείψω και τους υπόλοιπους φίλους μου που με στήριξαν με το δικό τους τρόπο Θάνο, Τζο, Σταυρούλα, Διονυσία και Μαρία.

Όσον αφορά την πραγματοποίηση της πτυχιακής μου εργασίας πρέπει σίγουρα να ευχαριστήσω τον επιβλέποντα μου Νικόλαο Γλυκό, ο οποίος με δέχτηκε στο εργαστήριο του, με ανέχτηκε και με βοήθησε, ώστε να κατανοήσω όσο το δυνατόν περισσότερα πράγματα. Την αμέριστη όμως ευγνωμοσύνη μου θέλω να την εκφράσω στον συμφοιτητή μου και συνάδελφο Θανάση Μπαλτζή, ο οποίος ήταν πάντα εκεί όταν τον χρειάστηκα και δεν αρνήθηκε ποτέ να μου προσφέρει τη βοήθεια του.

Σε όλα αυτά τα άτομα θέλω, λοιπόν, να υποσχεθώ ότι θα κάνω ότι καλύτερο μπορώ από εδώ και πέρα με όπλο τα εφόδια που μου προσέφεραν, για να τους γεμίσω περηφάνια.

Σας ευχαριστώ όλους,

Ευσταθία Αρτεμης Κουμουνδούρου

"As we express our gratitude, we must never forget that the highest appreciation is not to utter words, but to live by them."

John F. Kennedy

## Περίληψη

Ένα από τα πλέον σύγχρονα και άλυτα προβλήματα της βιολογίας είναι το πρόβλημα της αναδίπλωσης των πρωτεϊνών, δηλαδή η κατανόηση των μηχανισμών σχηματισμού της φυσικής διαμόρφωσης μιας πρωτεΐνης μέσα σε λίγα χιλιοστά του δευτερολέπτου. Για την επίλυση αυτού του προβλήματος έχουν αναπτυχθεί διάφορες θεωρίες και έχουν δοκιμασθεί κάποιες προσεγγίσεις. Στην παρούσα εργασία θα διερευνήσουμε με υπολογιστικές μεθόδους ένα από τα πεπτίδια FLP-18, με αλληλουχία EMPGVLRN-NH<sub>2</sub>. Το πεπτίδιο αυτό ανήκει στη μεγαλύτερη ομάδα νευρορυθμιστών του *C. elegans* και φαίνεται να ρυθμίζει μαζί με τα υπόλοιπα FLP-18 πεπτίδια, την τροφική συμπεριφορά στον *C. elegans*. Στόχος της εργασίας είναι η επίλυση της αναδίπλωσης του FLP-18 πεπτιδίου 1, με τη χρήση physics-based μεθόδων και η σύγκριση των αποτελεσμάτων των προσομοιώσεων με τα πειραματικά δεδομένα. Σύμφωνα με τα αποτελέσματα των αναλύσεων, το πεπτίδιο δε λαμβάνει κάποια συγκεκριμένη διαμόρφωση για μεγάλο χρονικό διάστημα και θα μπορούσε να χαρακτηριστεί ως μη δομημένο, γεγονός που συμβαδίζει με τα ευρήματα των πειραμάτων NMR.

## Abstract

One of the modern and unsolved problems in biology is the problem of protein folding, namely the understanding of the formation mechanisms of the native conformation of a protein in a few milliseconds. To solve this problem there have been developed various theories and some approaches have been tested. In this paper we will explore with computational methods one of the FLP-18 peptides that has the following sequence EMPGVLRN-NH<sub>2</sub>. This peptide belongs to the largest group of neuromodulation of *C. elegans* and it has been shown to be one of the regulators, among the FLP-18 peptides, of the food behavior in *C. elegans*. The aim of this work is to study the folding of FLP-18 peptide 1 by using physics-based methods and the comparison of the results with the experimental data. According to the results of the analysis, the peptide does not take a specific shape for a long time and could be described as unstructured, which is consistent with the findings of the NMR experiments.

# Περιεχόμενα

## Κεφάλαιο 1<sup>ο</sup>: Εισαγωγή

1.1 Αναδίπλωση των πρωτεϊνών .....	7
1.2 Το πρόβλημα της πρωτεϊνικής αναδίπλωσης .....	8
1.3 Μοντέλα πρωτεϊνικής αναδίπλωσης .....	9
1.4 Υπολογιστικές προσεγγίσεις για την επίλυση του προβλήματος της αναδίπλωσης των πρωτεϊνών .....	12
1.5 Ο ρόλος των πεπτιδίων στη μελέτη της αναδίπλωσης των πρωτεϊνών .....	15
1.6 Το νευροπεπτίδιο FMRFamide-Like-Peptide 18 .....	15
1.7 Στόχος της εργασίας .....	17

## Κεφάλαιο 2<sup>ο</sup>: Εισαγωγή στις προσομοιώσεις μοριακής δυναμικής

2.1 Εισαγωγή .....	19
2.2 Μοριακή Μηχανική .....	20
2.3 Προσομοιώσεις Μοριακής Δυναμικής .....	21
2.4 Δυναμικά Πεδία .....	23
2.5 Amber .....	24
2.6 Μοντέλα Νερού .....	26

## Κεφάλαιο 3<sup>ο</sup>: Προσομοιώσεις

3.1 Τεχνικά Χαρακτηριστικά Υπολογιστών .....	28
3.2 Έναρξη προσομοιώσεων μοριακής δυναμικής με NAMD .....	29
3.3 Προετοιμασία Συστήματος .....	31

## **Κεφάλαιο 4<sup>ο</sup>: Ανάλυση Προσομοιώσεων**

4.1 Εισαγωγή .....	36
4.2 RMSD based Ανάλυση των Τροχιακών .....	36
4.3 Ανάθεση Δευτεροταγούς Δομής .....	41
4.4 Cluster Analysis .....	43
4.5 Chemical Shifts .....	49

## **Κεφάλαιο 5<sup>ο</sup>: Συμπεράσματα και Συζήτηση**

5.1 Κεφάλαιο 5ο: Συμπεράσματα και Συζήτηση .....	54
--	----

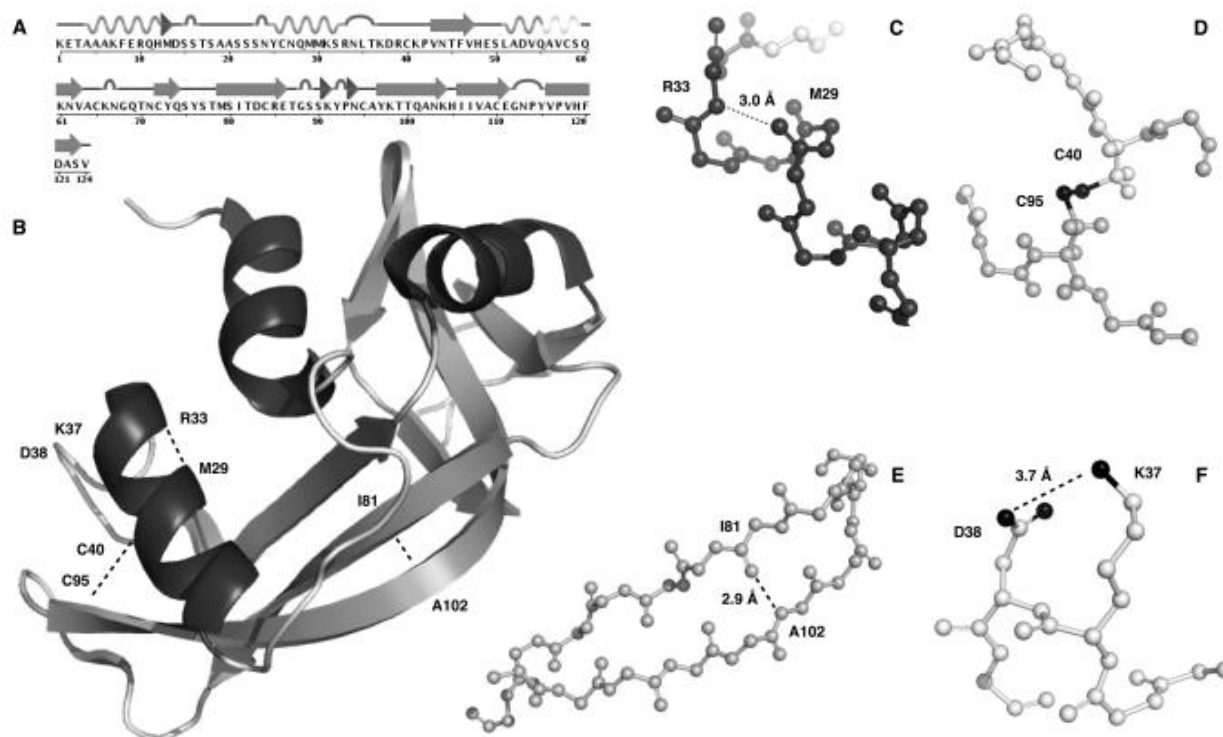
## **Βιβλιογραφία**

Βιβλιογραφία .....	56
--------------------	----

# 1. Εισαγωγή

## 1.1 Αναδίπλωση των Πρωτεϊνών

Οι πρωτεΐνες αποτελούν θεμελιώδη μόρια των οργανισμών και έχουν είτε δομικό είτε λειτουργικό ρόλο στα κύτταρα. Μια πρωτεΐνη αποτελείται από αμινοξέα, τα οποία ενώνονται μεταξύ τους με πεπτιδικούς δεσμούς σχηματίζοντας μια γραμμική αλυσίδα. Όμως για τη δημιουργία μιας λειτουργικής πρωτεΐνης απαιτείται εκτός από τη σωστή αλληλουχία των αμινοξέων και η σωστή τριτοταγής δομή, δηλαδή η σωστή αναδίπλωση στο χώρο (Εικόνα 1.1). Η διαδικασία με την οποία μια πολυπεπτιδική αλυσίδα αποκτά τρισδιάστατη δομή στο χώρο, ονομάζεται πρωτεϊνική αναδίπλωση. (Branden C and Tooze J. 1999)



**Εικόνα 1.1.** (αναπαράγεται άνευ αδείας) Διαφορετικές αλληλεπιδράσεις (δεσμοί υδρογόνου, γέφυρες άλατος, και δισουλφιδικοί δεσμοί μεταξύ κυστεϊνών) που καθορίζουν τη δευτεροταγή δομή των πρωτεϊνών. (A) Αλληλουχία πρωτεΐνης στην οποία οι συμβολισμοί δίνουν μια μονοδιάστατη απεικόνιση των περιοχών με δομή  $\alpha$ -έλικας (κυματομορφή),  $\beta$ -φύλλου (βέλος), και τυχαίας σπείρας (ευθεία γραμμή). Η συνολική 3D διαμόρφωση της ίδιας πρωτεΐνης φαίνεται στον πίνακα B. Παραδείγματα δεσμών υδρογόνου που σταθεροποιούν  $\alpha$ -έλικες και  $\beta$ -φύλλα αναφέρονται στα C και E, αντίστοιχα. Παραδείγματα δισουλφιδικού δεσμού και γέφυρας άλατος δείχνεται στα D και F, αντίστοιχα. (Compriani and Capriotti, 2013)

Υπάρχουν δύο κύριοι παράγοντες που συνεισφέρουν στην ενεργειακή διαφορά μεταξύ της διπλωμένης και της αποδιαταγμένης κατάστασης: η Ενθαλπία και η Εντροπία. Η Ενθαλπία προέρχεται από την ενέργεια των μη-ομοιοπολικών αλληλεπιδράσεων μεταξύ των ατόμων της πολυπεπτιδικής αλυσίδας, όπως αλληλεπιδράσεις van der Waals, δεσμούς-H, και ιοντικές αλληλεπιδράσεις. Είναι ισχυρότερες στην φυσική απ' ότι στην αποδιαταγμένη κατάσταση, καθώς αυτές συνεισφέρουν με τον καλύτερα δυνατό τρόπο ώστε να δημιουργηθεί ένα συμπαγές μόριο. Η Εντροπία αποτελεί έκφραση του 2ου νόμο της θερμοδυναμικής, δηλαδή την απαίτηση ενέργειας για να τη δημιουργία τάξης.. Η αποδιαταγμένη διαμόρφωση είναι ευνοϊκότερη εντροπικά, καθώς μπορεί να υιοθετήσει πολλές πιθανές διαμορφώσεις, ενώ οι αναδιπλωμένες πρωτεΐνες αποκτούν μία συμπαγή δομή με υψηλό βαθμό οργάνωσης. Όταν η πρωτεΐνη αναδιπλώνεται, το κόστος της μετάβασης από τη μη διπλωμένη στην αναδιπλωμένη μορφή είναι να μειωθεί η εντροπία (να γίνει λιγότερο αρνητική) (Branden C and Tooze J. 1999, Οικονόμου E. N. 2004).

Η ολική διαφορά ενέργειας ανάμεσα στην αποδιαταγμένη και στην αναδιπλωμένη μορφή της πρωτεΐνης, που αντιστοιχεί στην διαφορά μεταξύ των ενθαλπικών και εντροπικών διαφορών μεταξύ φυσικής και αποδιαταγμένης κατάστασης, ονομάζεται ελεύθερη ενέργεια. Η ελεύθερη ενέργεια προκύπτει από το αλγεβρικό άθροισμα της ενθαλπίας και της εντροπίας:

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

Η διεργασία της αναδίπλωσης μπορεί να λάβει χώρα όταν ο συνδυασμός της εντροπίας και η μεταβολή της ενθαλπίας λόγω των δεσμών υδρογόνου και των αλληλεπιδράσεων Van der Waals καταστήσουν τη συνολική ελεύθερη ενέργεια αρνητική. (Branden and Tooze 1999, Οικονόμου 2004).

Όμως η διαδικασία της αναδίπλωσης των πρωτεϊνών, παρότι είναι ζωτική και θεμελιώδης ουσιαστικά για όλη τη βιολογία, παραμένει από πολλές πλευρές ένα μυστήριο.

## 1.2 Το Πρόβλημα της πρωτεϊνικής αναδίπλωσης

Το πρόβλημα της αναδίπλωσης των πρωτεϊνών αποτελεί ένα από τα άλυτα προβλήματα της Μοριακής Βιολογίας. Το πρόβλημα της αναδίπλωσης των πρωτεϊνών είναι ουσιαστικά το ερώτημα του πώς η αλληλουχία των αμινοξέων μιας πρωτεΐνης οδηγεί στην τρισδιάστατη δομή της. Ένα σημαντικό ορόσημο στην πρωτεϊνική επιστήμη ήταν η υπόθεση που διατύπωσε ο Christian Anfinsen και οι συνεργάτες του (Dill et. al. 2008). Ο Anfinsen συμπέρανε ότι η φυσική δομή της πρωτεΐνης είναι η



θερμοδυναμικά σταθερή δομή, η οποία εξαρτάται μόνο από την αλληλουχία των αμινοξέων και τις συνθήκες του διαλύματος, και είναι ανεξάρτητη από την κινητική πορεία της αναδίπλωσης. Η αρχή του Anfinsen οδηγεί στο συμπέρασμα ότι ενώ η εξέλιξη μπορεί να ενεργεί για να αλλάξει μια αμινοξική αλληλουχία, η αναδίπλωση και η κινητική μιας δεδομένης αλληλουχίας ακολουθούν τους νόμους της φυσικοχημείας (Dill et. al. 2008, Anfinsen 1973).

Το 1969, ο Cyrus Levinthal επισήμανε ότι, λόγω του πολύ μεγάλου αριθμού των βαθμών ελευθερίας στη μη αναδιπλωμένη πολυπεπτιδική αλυσίδα, το μόριο έχει ένα τεράστιο αριθμό πιθανών διαμορφώσεων. Για παράδειγμα, ένα πολυπεπτίδιο 100 υπολειμμάτων θα έχουν 99 πεπτιδικούς δεσμούς, και ως εκ τούτου 198 διαφορετικές φ και ψ γωνίες. Εάν καθεμία από αυτές τις γωνίες δεσμού μπορεί να είναι σε μία από τις τρεις σταθερές διαμορφώσεις, η πρωτεΐνη μπορεί να πάρει  $3^{198}$  διαφορετικές διαμορφώσεις. Ως εκ τούτου, εάν μία πρωτεΐνη ήταν να λάβει τη σωστή διαμόρφωση του με διαδοχική λήψη όλων των πιθανών διαμορφώσεων, θα απαιτούσε χρόνο για να φτάσει στη σωστή φυσική διαμόρφωση περισσότερο από την ηλικία του σύμπαντος. Το «παράδοξο» είναι ότι οι περισσότερες μικρές πρωτεΐνες αναδιπλώνονται αυθόρμητα σε ένα χιλιοστό του δευτερολέπτου ή ακόμη και μικροδευτερολέπτα. Αυτή η παρατήρηση ονομάστηκε «το παράδοξο του Levinthal» (Branden C and Tooze J. 1999, Levinthal. 1968).

Ο Levinthal γνώριζε ότι οι πρωτεΐνες αναδιπλώνονται αυθόρμητα και σε σύντομο χρονικό διάστημα. Υπέδειξε ότι το παράδοξο μπορεί να επιλυθεί εάν η αναδίπλωση των πρωτεϊνών επιταχύνεται και καθοδηγείται από τον ταχύ σχηματισμό των τοπικών αλληλεπιδράσεων που στη συνέχεια καθορίζουν την περαιτέρω αναδίπλωση του πεπτιδίου. Η πρωτεϊνική αναδίπλωση βρίσκεται κάτω από κινητικό έλεγχο και ακολουθεί κάποιο συγκεκριμένο μονοπάτι, περνώντας μόνο από τοπικές βέλτιστες διαμορφώσεις (Zwanzig et al. 1992).

### **1.3 Μοντέλα πρωτεϊνικής αναδίπλωσης**

Έχουν προταθεί αρκετά μοντέλα αναδίπλωσης των πρωτεϊνών, χωρίς όμως κάποιο από αυτά να μπορεί να εξηγήσει επαρκώς την αναδίπλωση όλων των πρωτεϊνών. Τα κυριότερα μοντέλα είναι τα εξής:

## **Hydrophobic collapse hypothesis**

Η υπόθεση αυτή αναφέρεται σε ένα γεγονός που συμβαίνει κατά τη διάρκεια της διαδικασίας αναδίπλωσης των σφαιρικών πρωτεϊνών. Βασίστηκε στην παρατήρηση ότι στη φυσική τους κατάσταση οι πρωτεΐνες συχνά περιέχουν έναν υδρόφοβο πυρήνα από μη πολικές πλευρικές αλυσίδες αμινοξέων στο εσωτερικό της πρωτεΐνης, αφήνοντας το μεγαλύτερο μέρος των πολικών ή φορτισμένων καταλοίπων στην επιφάνεια της πρωτεΐνης. Η σταθεροποίηση της πρωτεΐνης με την παγίδευση των υδρόφοβων πλευρικών αλυσίδων από το περιβάλλον νερό, πιστεύεται ότι σταθεροποιεί τις ενδιάμεσες διαμορφώσεις. Η υπόθεση θεωρεί το “hydrophobic collapse” ως ένα σχετικά πρώιμο συμβάν στο μονοπάτι της αναδίπλωσης, που εμφανίζονται πριν από το σχηματισμό των πολλών δευτερογενών δομών και των φυσικών επαφών που υπάρχουν στο πλήρως αναδιπλωμένη τριτοταγή δομή. Η ενδιάμεση διαμόρφωση αναφέρεται επίσης ως εύπλαστη σφαίρα και αντιστοιχεί σε μια μερικώς αναδιπλωμένη κατάσταση της οποίας η ενέργεια είναι χαμηλότερη από εκείνη της αποδιαταγμένης κατάστασης, αλλά υψηλότερη από εκείνη της φυσικής κατάστασης. (Lapidus et al. 2007, Sadqi et.al. 2003, Brylinskia et al 2006)

## **Diffusion-collision model**

Σύμφωνα με το μοντέλο αυτό, η πρωτεΐνη αποτελείται από αρκετά τμήματα (elementary microdomains) , καθένα από τα οποία είναι αρκετά μικρό ώστε να λάβει γρήγορα όλες τις πιθανές διαμορφώσεις σε σύγκριση με τη χρονική κλίμακα ολόκληρης της διαδικασίας αναδίπλωσης. Έτσι, η ύπαρξη των microdomains κατά την αναδίπλωση της πολυπεπτιδικής αλυσίδας και η σημασία τους στην κινητική της αναδίπλωσης επιτρέπει στην πρωτεΐνη να αποφύγει την εξέταση όλων των εναλλακτικών διαμορφώσεων. Οι μικροεπικράτειες αυτές κινούνται και συγκρούονται μεταξύ τους για να σχηματίσουν μεγαλύτερες δομές. Η πληροφορία φαίνεται να βρίσκεται στη δημιουργία των μικροεπικρατειών και όχι σε κάθε αμινοξύ ξεχωριστά. (Karplus and Weaver. 1979, Karplus and Weaver. 1994)

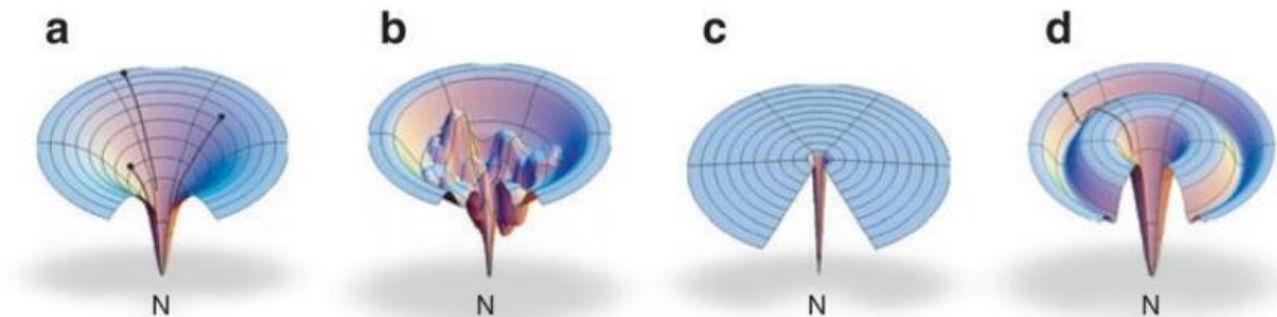
## **Nucleation-condensation mechanism**

Το μοντέλο αυτό προτείνει αρχικά το σχηματισμό ενός πυρήνα, δηλαδή, διάχυτων και εκτεταμένων περιοχών που παρατηρούνται σε μεταβατική κατάσταση και όχι ως όχι μια καλοσχηματισμένη δομή. Ο πυρήνας αυτός παρουσιάζει μεγάλη σταθερότητα λόγω των στοιχείων

δευτεροταγούς δομής, τα οποία διακρίνονται για τις τεταρτοταγείς επαφές που σχηματίζουν, οι οποίες είναι ίδιες με αυτές της φυσικής κατάστασης. Στη συνέχεια, ο πυρήνας δρα ως βάση για περαιτέρω δομές, σχηματίζοντας, έτσι, γύρω του τη δομή της πρωτεΐνης. Στο μοντέλου αυτό παρατηρούμε τη συνύπαρξη στοιχείων δευτεροταγούς και τεταρτοταγούς δομής. (Fersht 1997, Nolting and Agard 2008)

### Funnel Shaped- Energy Landscapes

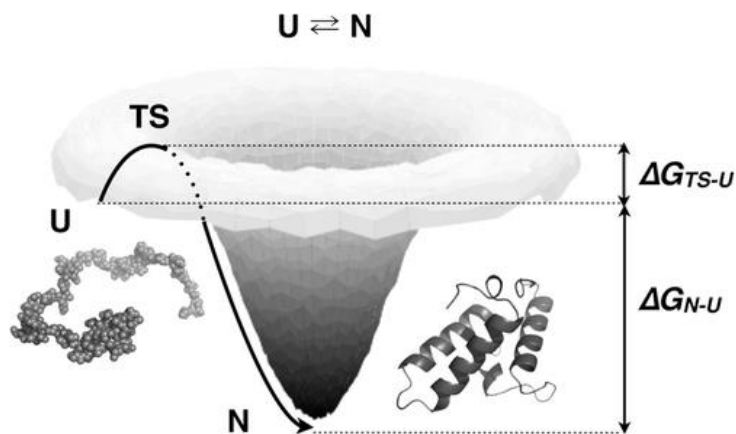
Τα μοντέλα των “ενεργειακών τοπίων” ή των “χωνιών αναδίπλωσης” είναι τα πιο πρόσφατα μοντέλα και προσφέρουν απλοποιημένες αναπαραστάσεις των δομών και των αλληλεπιδράσεων μέσω μεθόδων στατιστικής μηχανικής και υπολογίζουν τις μικροσκοπικές αλληλεπιδράσεις που αναπτύσσονται μεταξύ των πρωτεϊνών. Η σχηματική αναπαράσταση των ενεργειακών τοπίων αποτελείται από δισδιάστατα (Εικόνα 1.3) και τρισδιάστατα διαγράμματα που παρουσιάζουν τη διακύμανση της ελεύθερης ενέργειας των διαμορφώσεων ως συνάρτηση των βαθμών ελευθερίας. (Dill 1997)



**Εικόνα 1.2.** (αναπαράγεται άνευ αδείας) (a) Smooth energy landscape: για ένα ταχέως αναδιπλούμενο μόριο, (b) Rugged energy landscape: περιέχει ενεργειακές “παγίδες”, δηλαδή πολλούς πυθμένες που αντιστοιχούν σε ενδιάμεσα στάδια, (c) Golf course energy landscape: κατά την αναζήτηση αργή και τυχαία αναζήτηση σε όλη την επιφάνειά του, (d) Moat energy landscape: η αναζήτηση της φυσικής δομής πρέπει να περάσει από ένα υποχρεωτικό ενδιάμεσο στάδιο. (Dill et al, 2008)

Το ενεργειακό τοπίο μπορεί να απεικονιστεί ως ένα χωνί (Εικόνα 1.2). Το βάθος του χωνιού αντιπροσωπεύει την ενεργητική σταθεροποίηση της φυσικής κατάστασης έναντι της αποδιαταγμένης

κατάστασης, δηλαδή την εσωτερική ελεύθερη ενέργεια που προέρχεται από δεσμούς υδρογόνου, ιοντικές αλληλεπιδράσεις κλπ. Το πλάτος του χωνιού αντιπροσωπεύει την διαμορφωτική εντροπία του συστήματος. Η επιφάνεια έξω από το φρεάτιο παρουσιάζεται ως σχετικά επίπεδη για να δείχνει την ετερογένεια της τυχαίας κατάστασης. Έτσι η εντροπία μειώνεται όσο πλησιάζει η πρωτεΐνη την κανονική της δομή. (Karplus 2011) Η δυσκολία στην εύρεση των κατάλληλων διαμορφώσεων είναι ουσιαστικά το παραδόξο του Levinthal. Η αλγεβρική σύνδεση της ελεύθερης ενέργειας και της εντροπίας οδηγεί στον σχηματισμό φραγμάτων ελεύθερης ενέργειας και οφείλεται για τη μορφή των διαγραμμάτων που μοιάζουν με χωνί. (Dill 1997)



**Εικόνα 1.3.** (αναπαράγεται άνευ αδείας) Τυπική αναπαράσταση funnel-like μοντέλων ελεύθερης ενέργειας για τη διαδικασία αναδίπλωσης δύο σταδίων των πρωτεϊνών από μη διπλωμένη (U) στην μητρική (N) διαμόρφωση (Compiani and Capriotti, 2013)

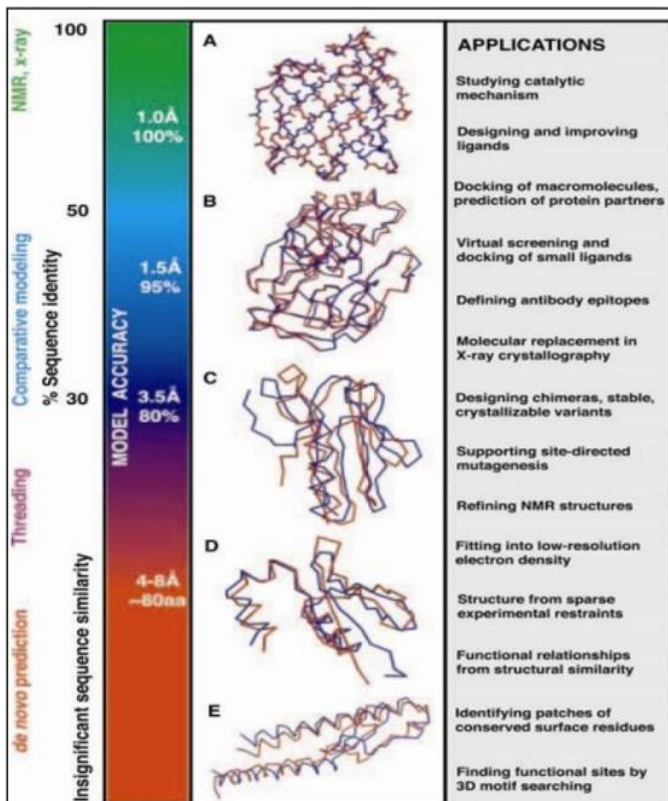
#### 1.4 Υπολογιστικές προσεγγίσεις για την επίλυση του προβλήματος της αναδίπλωσης των πρωτεϊνών

Αφού έχει καθοριστεί η αλληλουχία μιας πρωτεΐνης, η αποσαφήνιση της μοναδικής της τρισδιάστατης φυσικής δομής είναι ένα δύσκολο έργο. Έτσι έχουν χρησιμοποιηθεί όπως είδαμε θεωρητικές και πειραματικές προσεγγίσεις. Από μεθοδολογικής άποψης, οι παραδοσιακές προσομοιώσεις έχουν συνδεθεί με επεξεργασία μεγάλου όγκου δεδομένων με πειραματικές τεχνικές, όπως η κρυσταλλογραφία ακτίνων X, ο πυρηνικός μαγνητικός συντονισμός (NMR) (Compiani and Capriotti 2013) και τεχνικές απεικόνισης όπως τα mass spectrometry, high resolution-hydrogen exchange, fluorescence resonance energy transfer (FRET) και atomic force microscopy (AFM) (Chen et al 2008, Dinner et al 2002).

Ο προσδιορισμός της δομής μιας πρωτεΐνης πειραματικά είναι και ακριβός και το χρονοβόρος (Guntert, 2004). Αυτό έχει επιφέρει τη μαζική ανάπτυξη υπολογιστικών μεθόδων. Κατά τη διάρκεια των τελευταίων 10 ετών αρκετές υπολογιστικές μέθοδοι, συστήματα και αλγόριθμοι έχουν προταθεί ως λύση για το πρόβλημα της πρόβλεψης της τρισδιάστατης δομής των πρωτεϊνών. Αυτές οι μέθοδοι χωρίζονται σε τέσσερις κατηγορίες, (Floudas et al., 2006):

- (1) Ab initio methods (Osguthorpe, 2000)
- (2) First Principle Methods with Database Information (Röhl et al, 2004)
- (3) Fold Recognition and Threading methods (Forster 2001)
- (4) Comparative Modelling Methods (Carloni 2002)

Η πρώτη ομάδα μεθόδων στοχεύει στην πρόβλεψη νέων δομών μόνο μέσω (υπολογιστικών) προσομοιώσεων των φυσικοχημικών ιδιοτήτων της διαδικασίας αναδίπλωσης των πρωτεϊνών. Οι άλλες ομάδες αντιπροσωπεύουν μεθόδους που είναι σε θέση να εκτελέσουν γρήγορη και αποτελεσματική πρόβλεψη πρωτεϊνικών δομών όταν είναι διαθέσιμες γνωστές δομές πρότυπα σε βιβλιοθήκες (Kolinski, 2004).



**Εικόνα 1.4.** (αναπαράγεται άνευ αδείας)  
Συσχέτιση της μεθόδου πρόβλεψης της δομής μιας πρωτεΐνης με την ακρίβεια των μοντέλων που προκύπτουν και τις εφαρμογές για κάθε μοντέλο

## 1.5 Ο ρόλος των πεπτιδίων στη μελέτη της αναδίπλωσης των πρωτεϊνών

Η πολυπλοκότητα και η διάρκεια προσομοίωσης μεγάλων μοριακών συστημάτων, όπως οι πρωτεΐνες, είναι αυξημένη, με αποτέλεσμα και το υπολογιστικό κόστος τους να είναι πολλές φορές ιδιαίτερα υψηλό. Έτσι έχει προτιμηθεί η χρήση πεπτιδίων για τη μελέτη της αναδίπλωσης των πρωτεϊνών. (Gnanakaran et al., 2003) Η καταλληλότητα των πεπτιδίων σε τέτοιες μελέτες οφείλεται στην ικανότητα τους να σχηματίζουν σε διαλύματα φυσικές δευτεροταγείς δομές, πχ α-έλικες και β-φύλλα. (Marqusee 1989. Blanco 1994)

Μέσω των προσομοιώσεων μοριακής δυναμικής σε πεπτίδια μπορούν να χαρακτηριστούν η δυναμική και οι μοριακοί μηχανισμοί των αρχικών γεγονότων της πρωτεϊνικής αναδίπλωσης. Η κατανόηση του μηχανισμού της αναδίπλωσης των πεπτιδίων επιταχύνθηκε με το σχεδιασμό μικρών πεπτιδίων που μιμούνται την πολυπλοκότητα των πρωτεϊνών, αλλά επιτρέπουν τη λεπτομερή μελέτη τους, την ανάπτυξη γρήγορης time-resolved spectroscopy για την πειραματική μελέτη της δυναμικής της αναδίπλωσης των πεπτιδίων στην ίδια χρονική κλίμακα με τις προσομοιώσεις και την ανάπτυξη και εφαρμογή των αλγορίθμων για τις προσομοιώσεις (Neidigh et al., 2002, Sporlein et al., 2002, Shirts et al., 2001)

Καθώς οι προσομοιώσεις πεπτιδίων στοχεύουν στην καλύτερη κατανόηση των μηχανισμών της αναδίπλωσης και τη βελτιστοποίηση των αλγορίθμων των υπαρχουσών προσομοιώσεων, είναι ιδιαίτερα σημαντική η περαιτέρω ανάπτυξη τους για τη δημιουργία καλύτερων physics-based μεθόδων πρόβλεψης πρωτεϊνικών δομών. (Ho and Dill 2006)

## 1.6 Το νευροπεπτίδιο FMRFamide-Like-Peptide 18

Τα νευροπεπτίδια είναι μικρές αλληλουχίες αμινοξέων, που συμβάλουν με διάφορους τρόπους στη λειτουργία και στην ανάπτυξη του νευρικού συστήματος, λειτουργώντας είτε άμεσα είτε έμμεσα στη ρύθμιση της συναπτικής ενεργότητας, αλλά και ως νευροδιαβιβαστές. Στον *C. elegans* έχουν προβλεφθεί πάνω από 100 νευροπεπτίδια τα οποία χωρίζονται σε δύο μεγάλες κατηγορίες: στα Insulin-Like-Peptides και στα FMRFamide-Like-Peptides (FLPs), που είναι και η μεγαλύτερη ομάδα νευρορρυθμιστών του (Li and Kim 2008). Γονίδια *flp* έχουν βρεθεί και σε άλλους οργανισμούς όπως η *D.melanogaster*, ο *Ascaris suum*, αλλά και θηλαστικά. Πάνω από 50% των νευρώνων του *C. elegans* φέρουν FLPs.(Papaioannou et al 2005) Τα νευροπεπτίδια αυτά, όπως και τα περισσότερα, συντίθενται ως τμήμα μεγαλύτερων πρόδρομων πρωτεϊνών (Dossey et al 2006). Τα γονίδια *flp* εδράζονται στα 5

από τα 6 χρωμοσώματα του *C.elegans* και τα περισσότερα απ' αυτά εκφράζονται καθόλη τη ζωή του (Li et al 1999). Αν και κάθε *flp* γονίδιο εκφράζεται σε ένα ξεχωριστό σύνολο των νευρώνων, υπάρχει σημαντική επικάλυψη στο πρότυπο έκφρασης των διαφόρων γονιδίων FLP και ένας νευρώνας μπορεί να δείξει μια μεγάλη ποικιλία στην έκφραση FLPs (Li et al 1999).

Με 31 γονίδια να κωδικοποιούν FLPs στον *C.elegans* (Εικόνα 1.5) είναι αναμενόμενες οι πολλές λειτουργίες τους, όπως η συμβολή στον μυϊκό έλεγχο, στη ρύθμιση της καρδιάς, στην αίσθηση του πόνου, στη μάθηση, στην εναπόθεση των αυγών και τη συμπεριφορά κατά την πρόσληψη τροφής. (Li et al 1999, Papaioannou 2005, Dossey et al 2006, Li and Kim 2008, Li and Kim 2014.). Δεδομένου ότι πολλά από τα πεπτίδια που κωδικοποιούνται από ένα γονίδιο είναι πολύ παρόμοια, έχει προταθεί ότι τα πεπτίδια αυτά έχουν ελαφρώς διαφορετικές συγγένειες πρόσδεσης για τους υποδοχείς για να επιτραπούν μικρές διαφορές στη λειτουργία ενός κυκλώματος (Li et al 1999).

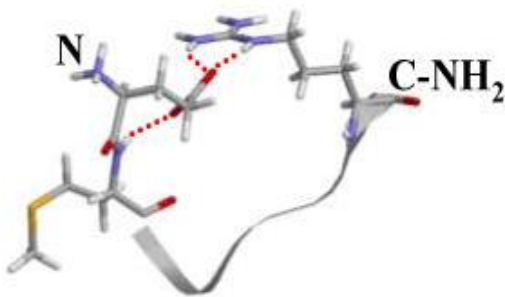
<b><i>flp-1</i> IV</b> *SADPNFLRFG *SQPNFLRFG *ASGDPNFLRFG *SDPNFLRFG *AAADPNFLRFG *(K)PNFLRFG AGSDPNFLRFG (K)PNFMRFG	<b><i>flp-7</i> X</b> x3 *SPMQRSSMVRFG x2 *TPMQRSSMVRFG SPMERSAMVRFG SPMDRSKMVRFG SSIDRASMVRLG	<b><i>flp-15</i> III</b> *GGPQGPLRFG *RGPSGPLRFG	<b><i>flp-24</i> III</b> *VPSAGDMMVRFG
<b><i>flp-2</i> X</b> *SPREPIRFG LRGEPIRFG	<b><i>flp-8</i> X</b> x3 *KNEFIRFG	<b><i>flp-16</i> II</b> x2 *AQTFVRFG *QQTFVRFG	<b><i>flp-25</i> III</b> DYDFVRFG *ASYDYIRFG
<b><i>flp-3</i> X</b> SPLGTMRFG *TPLGTMRFG *EAEELGTMRFG NPLGTMRFG *ASEDALFGTMRFG EDGNAPFGTMKFG *SAEPFGTMRFG *SADDSAPFGTMRFG *NPENDTPFGTMRFG	<b><i>flp-9</i> IV</b> x2 *KPSFVRFG	<b><i>flp-17</i> IV</b> x2 KSAFVRFG KSQYIRFG	<b><i>flp-26</i> X</b> *(E)FNADDLTRFG *GGAGEPLAFSPDMLSLRFG NYYESKPY
<b><i>flp-4</i> II</b> PTFIRFG *ASPSFIRFG	<b><i>flp-10</i> IV</b> QPKARSGYIRFG	<b><i>flp-18</i> X</b> *DFDGAMPVGLRFG *EMPGVLRFG x3 *(K)ISVPGVLRFG *EIPGVLRFG *SEVPGVLRFG *DVPVGLRFG	<b><i>flp-27</i> II</b> *(EASAFGDIIGELGK)- GLGGRMRFG *pQIPEERPIFME
<b><i>flp-5</i> X</b> *GAKFIRFG *AGAKFIRFG *APKPKFIRFG	<b><i>flp-11</i> X</b> *AMRNALVRFG *ASGGMRNALVRFG *NGAPQPFVRFG NGAPQPFG	<b><i>flp-19</i> X</b> *WANQVRFG *ASWASSVRFG	<b><i>flp-28</i> X</b> *APNRVLMRFG
<b><i>flp-6</i> V</b> x6 *KSAYMRFG	<b><i>flp-12</i> X</b> *RNKFEFIRFG	<b><i>flp-20</i> X</b> x2 AMMRFG AVFRMG	<b><i>flp-32</i> X</b> *AMRNLSLVRFG
	<b><i>flp-13</i> IV</b> *AMDSPLIRFG x2 *AADGAPLIRFG x2 *APEASPFIRFG *ASPSAPLIRFG *SPSAVPLIRFG *SAAAPLIRFG *ASSAPLIRFG	<b><i>flp-21</i> V</b> GLGPRPLRFG	<b><i>flp-33</i> I</b> *APLEGFEDMSGFLR- TIDGIQKPRFG
	<b><i>flp-14</i> III</b> x4 *KHEYLRFG	<b><i>flp-22</i> I</b> x3 *SPSAKWMRFG	<b><i>flp-34</i> I</b> ALNRDSVASLNNAERLRFG ADISTFASAINNAGRLRYG
		<b><i>flp-23</i> III</b> *WVGQQDFLRFG NDFLRFG EGHPEFMSQIAPP?	†Additional FLPs: APNRVLMRLVRF-NH <sub>2</sub> HFYNFSSSRKPNFLRF-NH <sub>2</sub> KXPXFFIRF-NH <sub>2</sub>

**Εικόνα 1.5.** (αναπαράγεται άνευ αδείας) Τα προβλεπόμενα FLP νευροπεπτίδια στον *C. elegans*. Η χρωμοσωμική θέση αναγράφεται μετά το όνομά του γονιδίου. Με μπλε υποδεικνύονται τα κοινά αμινοξέα μεταξύ των πεπτιδίων που κωδικοποιούνται από το ίδιο γονίδιο. (Li and Kim. 2014)

Όλα τα πεπτίδια FLPs περιλαμβάνουν την αλληλουχία Arg-Phe-NH<sub>2</sub> στο καρβοξυτελικό άκρο τους. Τα FLPs ποικίλουν σε μέγεθος από 5 ως 24 αμινοξέα (Li and Kim 2014) και μπορούν να διαφέρουν σημαντικά στο N-τελικό άκρο τους. Η ενεργότητά τους εξαρτάται από την διάταξη των αμινοξέων (Li et al 1999). Χαρακτηρίζονται από μειούμενη συντήρηση της αλληλουχίας από το C-τελικό προς το N-τελικό άκρο. Έχουν βρεθεί δύο είδη υποδοχέων για τα FLPs: οι GPCRs (G-Protein Coupled Receptors) και διάλυτοι Na (Dossey 2006).

Στην παρούσα εργασία θα ασχοληθούμε με ένα από τα 6 FLP-18 πεπτίδια (Li and Kim 2008), στο οποίο θα αναφερόμαστε ως πεπτίδιο 1, που έχει την αλληλουχία EMPGVLRN-NH<sub>2</sub>. Όλα τα πεπτίδια FLP-18 προκύπτουν από το ίδιο πρόδρομο μόριο και πιθανώς επεξεργάζονται και απελευθερώνονται ταυτόχρονα. Ο NPR-1, ένας GPCR που ρυθμίζει την τροφική συμπεριφορά στον *C.elegans*, ενεργοποιείται από δύο υποοικογένειες των FLPs, η μία εκ των οποίων είναι η FLP-18.[1] Η ενεργότητα του υποδοχέα NPR-1 φαίνεται να επηρεάζεται από τη δομή των πεπτιδίων. Το πεπτίδιο με το οποίο θα ασχοληθούμε φαίνεται να έχει αυξημένη ενεργότητα στον υποδοχέα NPR-1 σε σχέση με άλλα μακρύτερα. (Dossey et al. 2006)

Η υπόθεση ότι διαφορές στην τοπική δομή του μεταβλητού N-τελικού άκρου των ελεύθερων FLPs μπορεί να εξηγήσει εν μέρει τις διαφορές στην επίδραση στους υποδοχείς. Μέτρηση των chemical shifts του υπό μελέτη πεπτιδίου έδειξε ότι η συντηρημένη αλληλουχία PGVLRN-NH<sub>2</sub> είναι αδόμενη σε διάλυμα (Εικόνα 1.6). (Dossey et al. 2006)



**Εικόνα 1.6.** (αναπαράγεται άνευ αδείας) Οι πιο σημαντικοί δεσμοί H και ιοντικές αλληλεπιδράσεις για τα οποία έχουν υπάρξει αποδείξεις. Η αδόμενη περιοχή φαίνεται με μορφή ribbon. (Dossey et al. 2006)



## 1.7 Στόχος της εργασίας

Στόχος της εργασίας είναι η μελέτη της αναδίπλωσης του FLP-18 πεπτιδίου 1 με τη χρήση physics-based μεθόδων και η σύγκριση των αποτελεσμάτων των προσομοιώσεων με την πειραματικά προσδιορισμένη από NMR δομή.

## 2. Εισαγωγή στις προσομοιώσεις μοριακής δυναμικής

## 2.1 Εισαγωγή

Πειράματα με τη χρήση φασματοσκοπίας πυρηνικού μαγνητικού συντονισμού (NMR) ή κρυσταλλογραφίας ακτίνων-Χ μπορούν να παρέχουν δομικές πληροφορίες για την φυσική κατάσταση ή μερικές φορές για μετασταθείς καταστάσεις. Αλλά, όπως ένα σύστημα χαλαρώνει γρήγορα σε μια κατάσταση χαμηλότερης ενέργειας, η δυναμική της διαδικασίας αναδίπλωσης είναι δύσκολο να εκτιμηθεί από το πείραμα. Επιπλέον, τα παραδοσιακά πειράματα παρέχουν μόνο μέσο όρο των ποσοτήτων, όπως μέσες δομές, και όχι κατανομές και παραλλαγές. (Berman et al 2000)

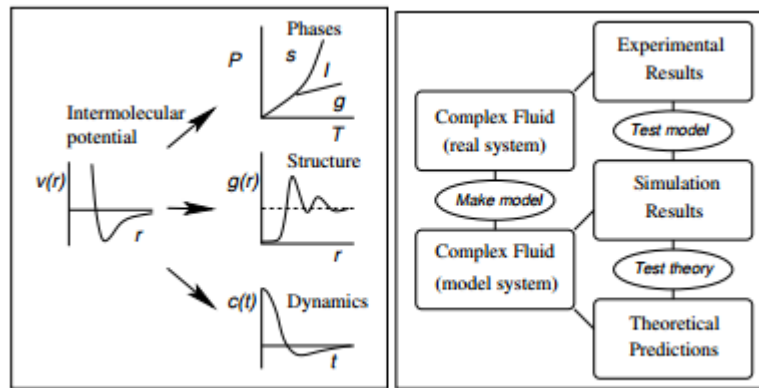
Οι προσομοιώσεις Μοριακής Δυναμικής σε υπολογιστή χρησιμοποιούνται για την απόκτηση μιας βαθύτερης κατανόησης της δυναμικής και των μηχανισμών που εμπλέκονται στην πρωτεϊνική αναδίπλωση και πραγματοποιούνται με την ελπίδα κατανόησης των ιδιοτήτων των μορίων, όσον αφορά τη δομή τους και τις μικροσκοπικές αλληλεπιδράσεις μεταξύ τους. Αυτό χρησιμεύει ως συμπλήρωμα των συμβατικών πειραμάτων, δίνοντάς μας τη δυνατότητα να μάθουμε κάτι καινούργιο, κάτι που δεν μπορεί να βρεθεί με άλλους τρόπους. (van Gunsteren et al 2007)

Οι προσομοιώσεις μοριακής δυναμικής αποτελούν τη ρεαλιστική αναπαράσταση του περιβάλλοντος των βιολογικών μακρομορίων σε ατομικό επίπεδο και αφορούν την περιγραφή των ατομικών και μοριακών αλληλεπιδράσεων που διέπουν τις μικροσκοπικές και μακροσκοπικές συμπεριφορές των φυσικών συστημάτων και περιλαμβάνουν όλες τις θεωρητικές μεθόδους και τις υπολογιστικές τεχνικές που χρησιμοποιούνται στη μοντελοποίηση ή στη μίμηση της συμπεριφοράς των μορίων. Λειτουργούν σαν μια γέφυρα μεταξύ του μικροσκοπικού και μακροσκοπικού κόσμου του εργαστηρίου (Εικόνα 2.1), δηλαδή ξεκινάμε με μια πρόβλεψη για τις αλληλεπιδράσεις μεταξύ των μορίων και καταλήγουμε σε “ακριβείς” προβλέψεις για την πρωτεΐνη (Allen, 2004).

Οι περισσότερες μελέτες μοριακής μοντελοποίησης συμπεριλαμβάνουν τρία στάδια μοριακών μοντέλων. Στο πρώτο στάδιο γίνεται η επιλογή μιας μεθόδου η οποία περιγράφει τη δομή και τις ιδιότητες του συστήματος. Οι δύο περισσότερο χρησιμοποιούμενες μέθοδοι είναι η Μοριακή Μηχανική και η Κβαντική Μηχανική. Οι μέθοδοι αυτές συμπεριλαμβάνουν μία συνάρτηση που περιγράφει πως μεταβάλλεται η ενέργεια του συστήματος καθώς μεταβάλλονται οι συντεταγμένες των πυρήνων των ατόμων που έχουμε ορίσει ως τις μικρότερες μονάδες που απαρτίζουν το σύστημα μας. (Allen, 2004, Κυρίκου et al. 2001)

Κατά το δεύτερο στάδιο πραγματοποιούνται οι υπολογισμοί, όπως είναι η ελαχιστοποίηση της ενέργειας, η μοριακή δυναμική (Molecular Dynamics), η προσομοίωση Monte Carlo και η συστηματική ανάλυση (Grid Scan). Ενώ στο τρίτο στάδιο αναλύονται οι δομές για να υπολογιστούν

κάποιες υπό μελέτη ιδιότητες (π.χ. ηλεκτρονική πυκνότητα, όγκος υπέρθεσης σε μία οδηγό φαρμακευτική ένωση κ.λπ.) οι οποίες συνεισφέρουν για να εξεταστεί η ορθότητα των υπολογισμών.



**Εικόνα 2.1.** (αναπαράγεται άνευ αδείας) Οι προσομοιώσεις ως γέφυρα μεταξύ (α) μικροκοσμου και μακροκοσμου (β) θεωρίας και πειράματος (Allen 2004)

## 2.2 Μοριακή Μηχανική

Πολλά από τα προβλήματα που θα θέλαμε να αντιμετωπίσουμε στη μοριακή μοντελοποίηση είναι, δυστυχώς, πολύ μεγάλα για να εξεταστούν από την κβαντομηχανική. Η μοριακή μηχανική αγνοεί τις ηλεκτρονικές κινήσεις και υπολογίζει την ενέργεια του συστήματος ως μια συνάρτηση μόνο των πυρηνικών θέσεων. Η Μοριακή μηχανική χρησιμοποιείται για να εκτελέσει υπολογισμούς σε συστήματα που περιέχουν σημαντικό αριθμό ατόμων. Σε ορισμένες περιπτώσεις τα δυναμικά πεδία μπορεί να δώσουν απαντήσεις που είναι τόσο ακριβείς, όπως ακόμη και υψηλού επιπέδου υπολογισμοί κβαντικής μηχανικής. Ωστόσο, η μοριακή μηχανική δεν μπορεί να παρέχει φυσικά ιδιότητες που εξαρτώνται από την ηλεκτρονική κατανομή σε ένα μόριο. (Κυρίκου et al. 2001)

Η Μοριακή Μηχανική, λοιπόν, χρησιμοποιείται ευρέως σε προσομοιώσεις μοριακής δυναμικής καθώς περιλαμβάνει τη χρήση της κλασσικής μηχανικής για την περιγραφή της φυσικής πίσω από τα μοριακά μοντέλα. Χρησιμοποιεί κλασσικού τύπου μοντέλα για την πρόβλεψη της ενέργειας ενός μορίου, ως συνάρτηση της διαμόρφωσης του. Αυτό επιτρέπει την πρόβλεψη μεταβατικών καταστάσεων. (Allinger, Burkert 1982)

Στην μοριακή μηχανική το μόριο αποτελείται από άτομα, που συνδέονται μεταξύ τους μέσω δεσμών, οι οποίοι λειτουργούν ως ελατήρια και βασίζεται σε ένα μάλλον απλό μοντέλο των αλληλεπιδράσεων μέσα σε ένα σύστημα με συνεισφορές από διαδικασίες όπως το τέντωμα δεσμών, το άνοιγμα και το κλείσιμο των γωνιών και των περιστροφών των απλών δεσμών. Οι εσωτερικές δυνάμεις περιγράφονται, συνήθως, χρησιμοποιώντας απλές συναρτήσεις, για παράδειγμα οι δεσμικές αλληλεπιδράσεις περιγράφονται από το νόμο του Hooke, ενώ οι μη δεσμικές αλληλεπιδράσεις μπορεί να περιγράφονται από το δυναμικό Lennard- Jones. (Κυρίκου et al. 2001)

## 2.3 Προσομοιώσεις Μοριακής Δυναμικής

Οι προσομοιώσεις μοριακής δυναμικής εισήχθησαν για πρώτη φορά από τους Alder και Wainwright στα τέλη της δεκαετίας του 1950, για να μελετήσουν τις αλληλεπιδράσεις σκληρών σφαιρών. (Alder and Wainwright, 1957,1959) Η πρώτη προσομοίωση μοριακής δυναμικής ενός ρεαλιστικού συστήματος έγινε από τον Rahman και Stillinger στην προσομοίωση νερού υγρής μορφής το 1974 (Stillinger and Rahman, 1974). Οι πρώτες προσομοιώσεις πρωτεΐνης εμφανίστηκαν το 1977 με την προσομοίωση του βοείου αναστολέα παγκρεατικής θρυψίνης (BPTI) (McCammon *et al*, 1977).

Πλέον έχουν γίνει προσομοιώσεις μοριακής δυναμικής επιδιαλυτωμένων πρωτεϊνών, συμπλόκων πρωτεΐνης-DNA, καθώς και συστημάτων λιπιδίων, που αντιμετωπίζουν μια ποικιλία θεμάτων, όπως τη θερμοδυναμική της δέσμευσης του προσδέτη και την αναδίπλωση μικρών πρωτεϊνών. Οι τεχνικές προσομοίωσης μοριακής δυναμικής χρησιμοποιούνται ευρέως σε πειραματικές διαδικασίες όπως κρυσταλλογραφία ακτίνων-X και τον προσδιορισμό δομής με NMR.(Van Gunsteren , Berendsen 1990)

Χρησιμοποιώντας μοντέλα μοριακής μηχανικής και τις εξισώσεις κίνησης του Νεύτωνα, οι προσομοιώσεις μοριακής δυναμικής μπορούν να παράγουν τη μεταβολή μιας δομής σε συνάρτηση με το χρόνο, δηλαδή τη θέση, την ταχύτητα και την επιτάχυνση όλων των ατόμων του μορίου κατά τη διάρκεια της προσομοίωσης (τροχιακό) (Haile 1992,1997). Επιπλέον είναι εφικτή η προσομοίωση συστημάτων που εξελίσσονται σε ακραίες τιμές μεταβλητών, όπως η θερμοκρασία και η πίεση, οι οποίες δεν είναι δυνατό να μελετηθούν πειραματικά. Οι προσομοιώσεις μοριακής μηχανικής εφαρμόζονται σήμερα ως επί το πλείστον στη φυσικοχημεία, στην επιστήμη υλικών και την μοντελοποίηση των βιομορίων. (Van Gunsteren , Berendsen 1990, Freddolino 2010)

Η βασική ιδέα είναι να επιτραπεί στο σύστημα να εξελιχθεί στο χρόνο, έτσι ώστε το σύστημα τελικά να περάσει από όλες τις πιθανές καταστάσεις. Υπάρχουν συγκεκριμένα βήματα που πρέπει να ακολουθηθούν για την πραγματοποίηση μιας προσομοίωσης. Αρχικά απαιτείται η προετοιμασία του συστήματος, δηλαδή ο καθορισμός της κατάστασης στην οποία θα συμβεί η προσομοίωση, όπως θερμοκρασία, αριθμός σωματιδίων κτλ.. Επίσης απαραίτητη είναι και η αρχικοποίηση του συστήματος, στην οποία δηλώνουμε τις αρχικές θέσεις και ταχύτητες των σωματιδίων καθώς και η ελαχιστοποίηση της ενέργειας. Ακολουθεί η παραγωγική φάση της προσομοίωσης κατά την οποία υπολογίζονται οι δυνάμεις σε κάθε σωματίδιο, καθώς επίσης θέσεις και ταχύτητες σε όλη τη διάρκεια του τροχιακού. Με το τέλος της προσομοίωσης συλλέγονται και αναλύονται τα δεδομένα.

Η προσομοιώσεις μοριακής δυναμικής βασίζονται στο νόμο του Νεύτωνα :

$$F = ma$$

Με τη γνώση της δύναμης που ενεργεί σε κάθε άτομο είναι δυνατό να προσδιοριστεί η επιτάχυνση του κάθε ατόμου. Η επίλυση της εξίσωσης δίνει ένα τροχιακό που περιγράφει τη θέση, την ταχύτητα και την επιτάχυνση του κάθε ατόμου, όπως αυτές μεταβάλλονται με το χρόνο. Μόλις είναι γνωστές η θέση και η ταχύτητα του κάθε ατόμου, η κατάσταση του συστήματος μπορεί να προβλεφθεί σε οποιαδήποτε στιγμή. Οι αρχικές θέσεις μπορούν να αποκτηθούν από πειράματα, πχ κρυσταλλογραφίας ακτίνων X ή φασματοσκοπίας NMR. Η αρχική κατανομή των ταχυτήτων προκύπτει συνήθων από την κατανομή Maxwell-Boltzmann:

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

από την οποία προκύπτει η πιθανότητα ενός ατόμου  $i$  με μάζα  $m_i$  να έχει μια συγκεκριμένη ταχύτητα  $v_x$  προς την κατεύθυνση  $x$  σε μια συγκεκριμένη θερμοκρασία  $T$ . Όπου  $k_B$  η σταθερά του Boltzmann. (Allen, 2004, Haile 1997)

Το δυναμικό ενέργειας είναι συνάρτηση των ατομικών θέσεων ( $3N$ ) όλων των ατόμων στο σύστημα. Λόγω της περίπλοκης φύσης αυτής της λειτουργίας, δεν υπάρχει αναλυτική λύση στις εξισώσεις κίνησης, πρέπει να επιλυθούν αριθμητικά. Πολυάριθμοι αριθμητικοί αλγόριθμοι έχουν αναπτυχθεί για την ενσωμάτωση των εξισώσεων κίνησης, που προσεγγίζουν τις θέσεις, ταχύτητες και επιταχύνσεις με την επέκταση του αλγόριθμου Taylor, όπως:

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

Κοινή είναι η χρήση του αλγόριθμου Verlet.

## 2.4 Δυναμικά πεδία

Στο πλαίσιο της μοριακής μοντελοποίησης, ένα δυναμικό πεδίο είναι μια συλλογή των εξισώσεων και των συναφών σταθερών που έχει σχεδιαστεί για να αναπαράγει τη μοριακή γεωμετρία και επιλεγμένες ιδιότητες ελεγμένων δομών και αναφέρεται στη λειτουργική μορφή και στο σύνολο των παραμέτρων που χρησιμοποιούνται για τον υπολογισμό της δυναμικής ενέργειας του συστήματος των ατόμων στη μοριακή μηχανική και στις προσομοιώσεις μοριακής δυναμικής. Η δυναμική ενεργεία στη μοριακή μηχανική αποτελείται από δεσμικές αλληλεπιδράσεις, δηλαδή τις αλληλεπιδράσεις των ατόμων που συνδέονται με ομοιοπολικούς δεσμούς και αφορούν το μήκος των δεσμών, τη γωνία που σχηματίζουν καθώς επίσης και τις διέδρες γωνίες, και μη δεσμικές (ονομάζεται επίσης μη ομοιοπολικές), που περιγράφουν την ηλεκτροστατική δύναμη μακράς εμβέλειας και τις δυνάμεις van der Waals. Η συγκεκριμένη ανάλυση των όρων εξαρτάται από το δυναμικό πεδίο, αλλά μια γενική μορφή για το σύνολο της ενέργειας σε ένα δυναμικό πεδίο μπορεί να γραφτεί ως:

$$V = E_{\text{bonded}} + E_{\text{nonbonded}}$$

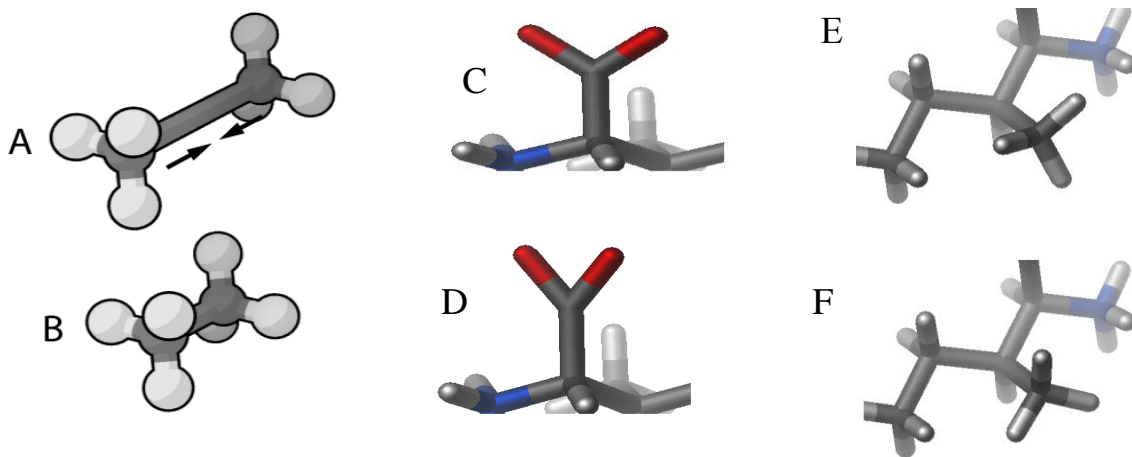
όπου το σύνολο των ομοιοπολικών και μη ομοιοπολικών συνεισφορές δίνεται από τα ακόλουθα αθροίσματα:

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

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

Δηλαδή η συνολική δυναμική ενέργεια αποτελείται από το άθροισμα της δυναμικής ενέργειας που προέρχεται από τις δεσμικές αλληλεπιδράσεις και της δυναμικής ενέργειας των μη δεσμικών αλληλεπιδράσεων. Στη δυναμική ενέργεια των δεσμικών αλληλεπιδράσεων συνεισφέρει η ενέργεια

που παράγεται από την έκταση των ομοιοπολικών δεσμών, την κάμψη των γωνιών και την περιστροφή των δεσμών (Εικόνα 2.2), ενώ στη δυναμική ενέργεια των μη δεσμικών αλληλεπιδράσεων συνεισφέρουν οι δυνάμεις van-der-Waals και οι ηλεκτροστατικές αλληλεπιδράσεις. Οι αλληλεπιδράσεις van der Waals συνήθως υπολογίζονται με τις εξισώσεις Lennard-Jones και οι ηλεκτροστατικές αλληλεπιδράσεις με το νόμο του Coulomb. (Ponder, Case 2003)



**Εικόνα 2.2.** (αναπαράγεται άνευ αδείας) A και B) Φαίνεται η έκταση του ομοιοπολικού δεσμού. ([http://en.wikipedia.org/wiki/Force\\_field\\_%28chemistry%29](http://en.wikipedia.org/wiki/Force_field_%28chemistry%29)) C και D) Φαίνεται η κάμψη της γωνίας που σχηματίζουν τα 3 άτομα. E και F) Φαίνεται η περιστροφή 4 ατόμων γύρω από τον κεντρικό δεσμό. Στο σύνολό τους αποτελούν τις πηγές της δυναμικής ενέργειας που προκύπτει από τις δεσμικές αλληλεπιδράσεις. ([http://www.ch.embnet.org/MD\\_tutorial/](http://www.ch.embnet.org/MD_tutorial/))

## 2.5 Amber

Τα πιο κοινώς χρησιμοποιούμενα, σε προσομοιώσεις μοριακής δυναμικής, δυναμικά πεδία για τη μελέτη της αναδίπλωσης των μακρομορίων και πρωτεϊνών είναι το AMBER (Assisted Model Building with Energy Refinement), το CHARMM (Chemistry HARvard Molecular Mechanics), το OPLS (Optimized Potential for Liquid Simulations ) και το GROMOS (Groningen Molecular Simulation). (Guvench and MacKerell, 2008, Ponder and Case 2003)

Ο όρος Amber αναφέρεται σε κάτι περισσότερο από ένα πακέτο μοριακής δυναμικής (MD). Περιλαμβάνει τη συλλογή πολυάριθμων προγραμμάτων που συνεργάζονται για την εγκατάσταση, την εκτέλεση, και την ανάλυση προσομοιώσεων MD, από την προετοιμασία των απαραίτητων αρχείων



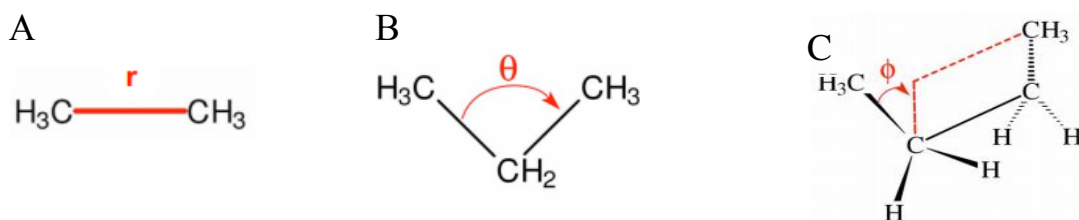
εισόδου, μέχρι την ανάλυση των αποτελεσμάτων. Το όνομα Amber αναφέρεται επίσης σε μια σειρά από δυναμικά πεδία κλασσικής μοριακής μηχανική, που έχουν σχεδιαστεί κυρίως για την προσομοίωση των βιομορίων. Το Amber είναι το αποτέλεσμα της συνεργασίας πάνω από 40 ερευνητών και εξωτερικών συνεργατών, που εργάζονται για την προώθηση των μεθόδων MD και τη μελέτη πολλών σημαντικών προβλημάτων της βιοχημείας. (Salomon-Ferrer et al 2012)

Οι πιο διαδεδομένες εκδόσεις του δυναμικού πεδίου Amber είναι οι ff94, ff99SB, ff03 και GAFF. Η ανάπτυξη δυναμικού πεδίου είναι μια σύνθετη διαδικασία στην οποία συνήθως απαιτούνται πολλές προσεγγίσεις. Μετά από μια δεκαετία εκτεταμένης χρήσης και δοκιμών του ff94 έχουν εμφανιστεί διάφοροι περιορισμοί στο δυναμικό πεδίο. Αυτό οδήγησε σε πολλές προσπάθειες για τη βελτίωση αυτών των παραμέτρων, με αποτέλεσμα τη δημιουργία μιας ποικιλίας δυναμικών πεδίων 'Amber' και τη δυσκολία στον προσδιορισμό της σωστής χρήσης τους για μια συγκεκριμένη εφαρμογή. Μελέτες έχουν δείξει ότι το σύνολο των παραμέτρων ff99SB παρέχουν ένα λογικό σύνολο παραμέτρων για τις προσομοιώσεις πρωτεϊνών, στο πλαίσιο ενός αποτελεσματικού μοντέλου μοριακής μηχανικής. (Hornak et al 2006)

Η μορφή της εξίσωσης του AMBER είναι:

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

Ο πρώτος όρος αντιπροσωπεύει την ενέργεια μεταξύ ομοιοπολικά συνδεδεμένων ατόμων κατά την έκταση των δεσμών. Ο δεύτερος όρος ορίζει την ενέργεια λόγω της κάμψης των γωνιών των δεσμικών ατόμων και ο τρίτος όρος την ενέργεια λόγω της περιστροφής των ατόμων γύρω από τους δεσμούς. Ο τέταρτος όρος αντιπροσωπεύει την ενέργεια των μη δεσμικών ατόμων, και μπορεί να αναλυθεί στην ενέργεια που οφείλεται σε δυνάμεις van der Waals (πρώτος όρος του αθροίσματος) και στις αλληλεπιδράσεις ηλεκτροστατικής φύσης (δεύτερος όρος αθροίσματος). (Cornell et al 1995)



**Εικόνα 2.3.** Στις εικόνες φαίνονται οι μεταβλητές των δεσμικών αλληλεπιδράσεων που χρησιμοποιούνται για τον υπολογισμό της δυναμικής ενέργειας. Α) Η απόσταση του δεσμού που βρίσκεται σε έκταση. Β) Το άνοιγμα της γωνίας που κάμπτεται .C) Η γωνία που σχηματίζεται κατά την περιστροφή των ατόμων γύρω από το δεσμό.

## 2.6 Μοντέλα Νερού

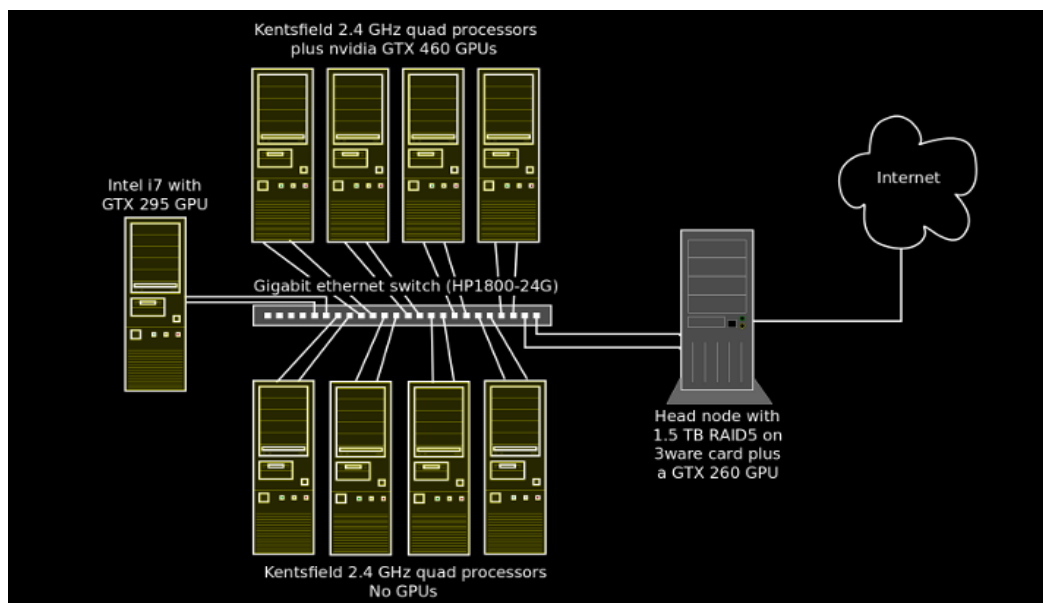
Το νερό, ο πιο σημαντικός διαλύτης στη φύση, έχει πολλές ειδικές και ασυνήθιστες ιδιότητες. Πολλές από αυτές τις ιδιότητες οφείλονται στην ικανότητα των μορίων του νερού να σχηματίζουν δεσμούς υδρογόνου με άλλα μόρια νερού σε τρισδιάστατα δίκτυα. Οι μακροσκοπικές ιδιότητες του νερού έχουν μελετηθεί σε βάθος και είναι τώρα καλά γνωστές, αλλά οι μικροσκοπικές δυνάμεις που καθορίζουν τη δομή του νερού δεν είναι πλήρως κατανοητές. (Mark and Nilsson 2001) Στις προσομοιώσεις μοριακής δυναμικής χρησιμοποιούνται συγκεκριμένα μοντέλα νερού. Αυτά είναι τα SPC (Simple Point Charge), SPC/E (Extended Simple Point Charge), TIP3P και TIP4P (Three Point και Four Point Transferable Intermolecular Potentials) (Van der Spoel et al 1998). Αυτά τα μοντέλα νερού έχει τρεις θέσεις αλληλεπίδρασης και είναι παρόμοιας φύσης (εκτός από το TIP4P που θεωρεί ότι υπάρχουν 4 θέσεις αλληλεπίδρασης), αλλά οι όροι Lennard-Jones (LJ) και Κουλόμπ διαφέρουν και δίνουν σημαντικές διαφορές στους υπολογισμούς των ιδιοτήτων του νερού (Mark and Nilsson 2001).

## 3. Προσομοιώσεις

### 3.1 Τεχνικά χαρακτηριστικά υπολογιστών

Καθώς το υπολογιστικό κόστος των προσομοιώσεων μοριακής δυναμικής είναι ιδιαίτερα αυξημένο, προτιμάται η χρήση παράλληλα συνδεδεμένων υπολογιστών για την αύξηση της υπολογιστικής ισχύος χωρίς την παράλληλη αύξηση του οικονομικού κόστους. Τα υπολογιστικά συγκροτήματα Beowulf αποτελούνται συνήθως από πανομοιότυπους υπολογιστές, παράλληλα συνδεδεμένους και δικτυομένους σε ένα μικρό τοπικό δίκτυο, με τις βιβλιοθήκες και τα εγκατεστημένα προγράμματα να οδηγούν στη μοιρασμένη μεταξύ του επεξεργασία των δεδομένων. (Becker et al 1995) Ένα τέτοιο υπολογιστικό συγκρότημα είναι η Norma.

Η Norma περιλαμβάνει συνολικά 40 CPU ( central processing unit) πυρήνες, συνολική φυσική μνήμη 46 Gbytes και 6 GPGPUs (General-purpose computing on graphics processing units) καταναμημένα σε 10 κόμβους. Οι οκτώ κόμβοι (πλην του σκληρού δίσκου) προσφέρουν, έκαστος, τέσσερις πυρήνες, 4 Gbytes φυσικής μνήμης και δύο(gigabit) διασυνδέσεις δικτύου. Ο ένατος κόμβος (Intel i7 965 extreme) προσφέρει 6 Gbytes φυσικής μνήμης και μία κάρτα γραφικών με δυνατότητα αύξησης της επεξεργαστικής ισχύος μέσω CUDA (Compute Unified Device Architecture) GTX-295. Οι κόμβοι συνδέονται μέσω ενός ειδικού ελεγκτή δικτύου HP ProCurve 1800-24G Gigabit Ethernet switch. Ο κύριος κόμβος παρέχει τέσσερις πυρήνες 8 Gbytes φυσικής μνήμης, 1,5 Tbytes αποθηκευτικής μνήμης, τρεις (gigabit) διασυνδέσεις δικτύου, και μια NVIDIAGTX-260 GPU (graphics processing unit).



Εικόνα 3.1. Σχηματική απεικόνιση του υπολογιστικού συστήματος Norma. (<http://norma.mbg.duth.gr/index.php?id=about:intro>)

### 3.2 Έναρξη προσομοιώσεων μοριακής δυναμικής με NAMD

Το πρόγραμμα που χρησιμοποιήθηκε προκειμένου να εκτελεστούν οι προσομοιώσεις στη Norma είναι το NAMD. Το NAMD είναι ένα πρόγραμμα μοριακής δυναμικής σχεδιασμένο για προσομοιώσεις υψηλής απόδοσης μεγάλων βιομοριακών συστήματα σε παράλληλους υπολογιστές. Οι τυπικές προσομοιώσεις NAMD περιλαμβάνουν μοντέλα με όλα τα άτομα των πρωτεϊνών, των λιπιδίων, και / ή των νουκλεϊκών οξέων, καθώς και το διαλύτη (νερό και ιόντα) και κυμαίνονται σε μέγεθος από 10.000 έως 10.000.000 άτομα. (Nelson et al 1996) Το NAMD λειτουργεί με τις λειτουργίες δυναμικού, τις παραμέτρους και τις μορφές αρχείων των AMBER και CHARMM. (Phillips et al 2005)

Οι προσομοιώσεις πραγματοποιήθηκαν με το πρόγραμμα NAMD χρησιμοποιώντας το δυναμικό πεδίο AMBER. Παράχθηκαν 11.323.925 διαμορφώσεις (frames). Για την έναρξη της προσομοίωσης το NAMD χρειάζονται συνήθως τουλάχιστον τέσσερις τύποι αρχείων:

**α) Αρχείο PDB:** είναι μια μορφή αρχείου κειμένου που περιγράφει τις τρισδιάστατες δομές των μορίων που προέρχεται από πειράματα περίθλασης ακτίνων X και μελέτες NMR. Περιέχει περιγραφή και σχολιασμό των δομών πρωτεϊνών και νουκλεϊκών οξέων, συμπεριλαμβανομένων των ατομικών συντεταγμένων, παρατηρούμενα ροταμερή της πλευρικής αλυσίδας, αναθέσεις δευτερεύουσας δομής, καθώς και συνδεσης μεταξύ των ατόμων. Τα αρχεία PDB μπορούν να δημιουργηθούν με το χέρι ή να παρθούν μέσω του Διαδικτύου από την ιστοσελίδα <http://www.pdb.org>. Το NAMD ενδιαφέρεται μόνο για τις καταχωρήσεις ATOM (Εικόνα 3.2), οι οποίες αφορούν τις συντεταγμένες των ατόμων της πρωτεΐνης, μορίων νερού και επιπρόσθετων ιόντων, καθώς και τις συντεταγμένες ετερογενών ατόμων στους κρυστάλλους αντίστοιχα. (Phillips 2012)

ATOM	1	N	GLH	A	1	0.943	-7.152	6.178	1.00	0.00
ATOM	2	H	GLH	A	1	1.823	-6.976	6.641	1.00	0.00
ATOM	3	CA	GLH	A	1	0.391	-5.904	5.575	1.00	0.00
ATOM	4	HA	GLH	A	1	-0.661	-6.110	5.379	1.00	0.00
ATOM	5	CB	GLH	A	1	0.660	-4.824	6.591	1.00	0.00
ATOM	6	HB2	GLH	A	1	0.517	-5.074	7.643	1.00	0.00
ATOM	7	HB3	GLH	A	1	1.704	-4.512	6.620	1.00	0.00
ATOM	8	CG	GLH	A	1	-0.226	-3.585	6.390	1.00	0.00
ATOM	9	HG2	GLH	A	1	0.011	-3.188	5.403	1.00	0.00
ATOM	10	HG3	GLH	A	1	-1.256	-3.942	6.388	1.00	0.00
ATOM	11	CD	GLH	A	1	-0.030	-2.519	7.464	1.00	0.00
ATOM	12	OE1	GLH	A	1	1.001	-2.442	8.109	1.00	0.00

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

**β) Αρχείο prmtop :** περιέχει όλες τις ειδικές πληροφορίες σχετικά με τη μοριακή τοπολογία, τις παραμέτρους δυναμικού πεδίου, τα άτομα και τα ονόματα των καταλοίπων. Ορίζονται η μάζα, οι δεσμοί, οι γωνίες και διέδρες γωνίες.

```

MASS
CU 65.36

BOND
NB-CU 70.000 2.05000 #kludge by JRS
CU-S 70.000 2.10000 #kludge by JRS
CU-SH 70.000 2.90000 #for pcy
CT-SH 222.000 1.81000 #met(aa)

ANGLE
CU-NB-CV 50.000 126.700 #JRS estimate
CU-NB-CR 50.000 126.700 #JRS estimate
CU-NB-CP 50.000 126.700 #JRS estimate
CU-NB-CC 50.000 126.700 #JRS estimate
CU-SH-CT 50.000 120.000 #JRS estimate
CU-S -CT 50.000 120.000 #JRS estimate
CU-S -C2 50.000 120.000 #JRS estimate
CU-S -C3 50.000 120.000 #JRS estimate
NB-CU-NB 10.000 110.000 #dac estimate
NB-CU-SH 10.000 110.000 #dac estimate
NB-CU-S 10.000 110.000 #dac estimate
SH-CU-S 10.000 110.000 #dac estimate
CU-SH-CT 50.000 120.000 #JRS estimate
CT-CT-SH 50.000 114.700 #met(aa)
HC-CT-SH 35.000 109.500
H1-CT-SH 35.000 109.500
CT-SH-CT 62.000 98.900 #MET(OL)

```

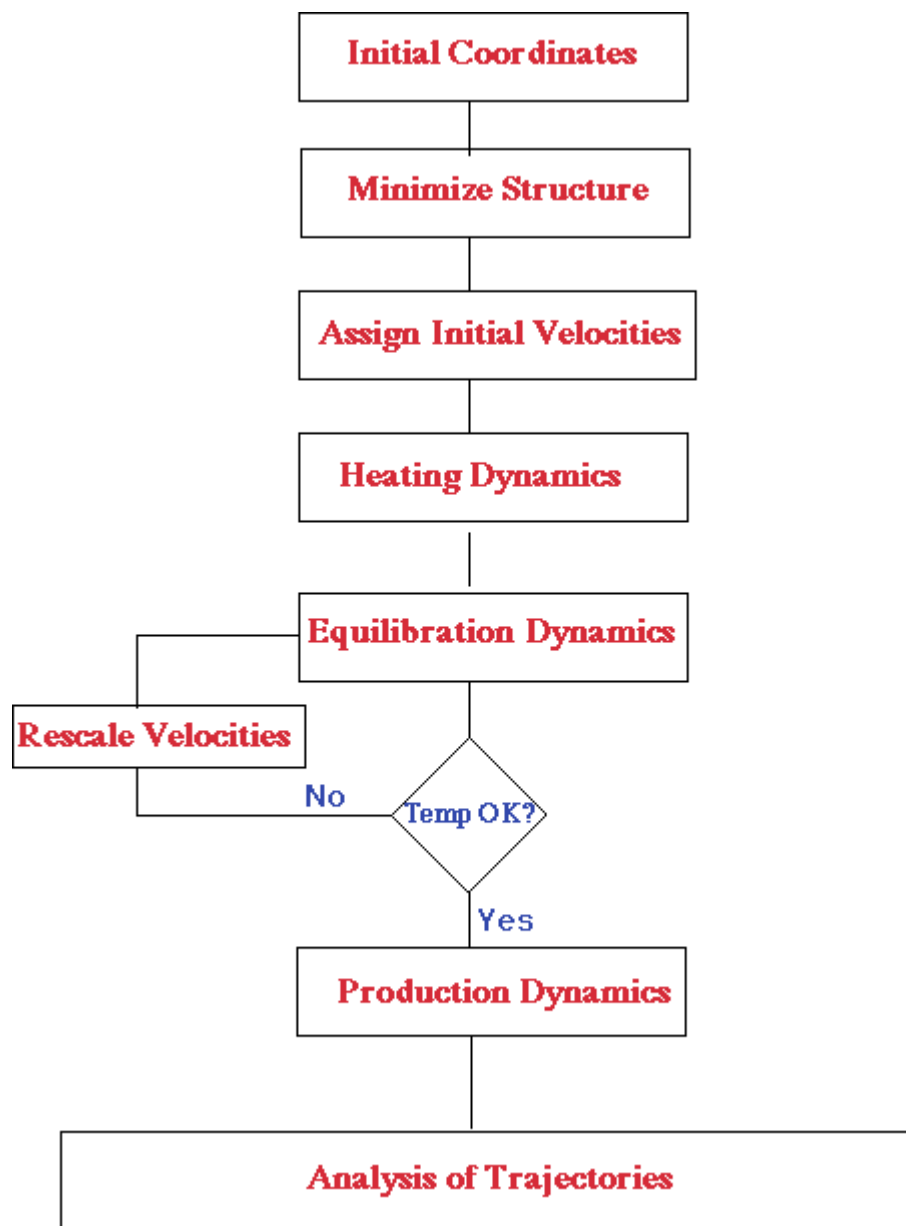
**Εικόνα 3.3:** Η μορφή ενός prmtop αρχείου

**γ) Αρχείο παραμετροποίησης δυναμικού πεδίου:** περιέχει το σύνολο των αριθμητικών σταθερών που απαιτούνται για την εκτίμηση των δυνάμεων και των ενεργειών, εφόσον δίνεται ένα αρχείο PSF και ατομικές συντεταγμένες και ορίζει την αντοχή του δεσμού, το μήκος ισορροπίας κλπ. Τα CHARMM, X-PLOR, AMBER, και GROMACS είναι τέσσερις τύποι δυναμικών πεδίων και το NAMD είναι σε θέση να τα χρησιμοποιεί όλα. (Phillips 2012)

**δ) Αρχείο ρυθμίσεων:** καθορίζει όλες τις επιλογές που το NAMD θα πρέπει να ακολουθήσει κατά τη λειτουργία της προσομοίωσης. Το αρχείο ρυθμίσεων λέει στο NAMD πώς πρέπει να τρέξει η προσομοίωση. Τέτοιες ρυθμίσεις μπορεί να είναι ο καθορισμός της εισόδου και της εξόδου του προγράμματος, ο καθορισμός της αρχικής θερμοκρασίας του συστήματος, το χρονικό βήμα της προσομοίωσης, τη συχνότητα καταγραφής των συντεταγμένων στα τροχιακά κ.α (Phillips 2012)

### 3.3 Προετοιμασία Συστήματος

Σε μια προσομοίωση μοριακής δυναμικής, η χρονοεξαρτώμενη συμπεριφορά του μοριακού συστήματος επιτυγχάνεται με την ενσωμάτωση των εξισώσεων του Νεύτωνα. Το αποτέλεσμα της προσομοίωσης είναι μια χρονική αλληλουχία από διαμορφώσεις. Αυτό ονομάζεται τροχιακό. Αυτή η αλληλουχία των δεδομένων, δηλαδή τα τροχιακά MD, καθορίζουν πλήρως την ιστορία της ατομικής κίνησης. (Shao et al 2007)



Εικόνα 3.4. (αναπαράγεται άνευ αδείας) Διάγραμμα ροής ποσομοιώσεων μοριακής δυναμικής. ([http://www.ch.embnet.org/MD\\_tutorial/](http://www.ch.embnet.org/MD_tutorial/))

Όπως φαίνεται και στην εικόνα 3.4 υπάρχουν συγκεκριμένα βήματα που πρέπει να ακολουθηθούν προκειμένου να λειτουργήσει σωστά μια προσομοίωση. Τα βήματα αυτά είναι η αρχικοποίηση των συντεταγμένων, η ελαχιστοποίηση της ενέργειας, η ανάθεση αρχικών ταχυτήτων, η θέρμανση, η εξισορρόπηση, τον έλεγχο της θερμοκρασίας, την παραγωγική φάση και τέλος την ανάλυση των τροχιακών.

### **Αρχικοποίηση συντεταγμένων**

Για να ξεκινήσει μια προσομοίωση μοριακής δυναμικής, θα πρέπει πρώτα να επιλεγθεί μια αρχική διαμόρφωση του συστήματος, ένα σημείο εκκίνησης, ή  $t = 0$ . Τις περισσότερες φορές, σε προσομοιώσεις βιομορίων, μια κρυσταλλική δομή με ακτίνες X ή μια δομή NMR λαμβάνεται από την Brookhaven Protein Databank (<http://www.rcsb.org/pdb/>) και χρησιμοποιείται ως η αρχική δομή. Είναι επίσης δυνατόν να χρησιμοποιηθεί μια θεωρητική δομή που έχει αναπτυχθεί από μοντελοποίηση ομολογίας. Η επιλογή της αρχικής διαμόρφωσης θα πρέπει να γίνει προσεκτικά, καθώς αυτό μπορεί να επηρεάσει την ποιότητα της προσομοίωσης. (Frenkel 2001) Το υπό εξέταση πεπτίδιο είχε μελετηθεί μέσω πειραμάτων NMR (Dossey 2006).

### **Ελαχιστοποίηση ενέργειας**

Πριν από την έναρξη μιας μοριακής προσομοίωσης δυναμική, είναι σκόπιμο να κάνουμε μια ελαχιστοποίηση της ενέργειας. Αυτό αφαιρεί οποιεσδήποτε ισχυρές αλληλεπιδράσεις van der Waals που μπορεί να υπάρχουν, οι οποίες διαφορετικά θα μπορούσαν να οδηγήσουν σε τοπική στρέβλωση της δομής και να οδηγήσουν σε μια ασταθή προσομοίωση (Frenkel 2001, Leach 2001).

### **Θέρμανση**

Κατά τη διάρκεια της φάσης θέρμανσης, οι αρχικές ταχύτητες αναθέτονται σε χαμηλή θερμοκρασία και η προσομοίωση αρχίζει. Σε τακτά χρονικά διαστήματα, νέες ταχύτητες αναθέτονται σε μια ελαφρά υψηλότερη θερμοκρασία και η προσομοίωση συνεχίζεται. Αυτό επαναλαμβάνεται έως ότου επιτευχθεί η επιθυμητή θερμοκρασία (Schlick, Tamar 2002).



## Εξισορρόπηση

Μόλις επιτευχθεί η επιθυμητή θερμοκρασία, η προσομοίωση του συστήματος πρωτεΐνης / νερού συνεχίζει και κατά τη διάρκεια αυτής της φάσης οι διάφορες ιδιότητες παρακολουθούνται, πχ η δομή, η πίεση, η θερμοκρασία και η ενέργεια. Κατά τη φάσης εξισορρόπησης εκτελείται η προσομοίωση μέχρι αυτές οι ιδιότητες να είναι πλέον σταθερές σε σχέση με το χρόνο. Εάν η θερμοκρασία αυξάνεται ή μειώνεται σημαντικά, οι ταχύτητες μπορεί να κλιμακωθούν ώστε η θερμοκρασία να επιστρέψει στην επιθυμητή τιμή της (Schlick, Tamar 2002).

Σε αυτά τα βήματα, λοιπόν, χρειαζόμαστε το αρχείο διαμόρφωσης `heat.namd`, το οποίο ορίζει πως ακριβώς θα γίνει η διαδικασία της ελαχιστοποίησης και της θέρμανσης του συστήματος. Τα αρχεία που χρησιμοποιούνται ως είσοδος για τη δημιουργία του `heat.namd` είναι το `.prmtop` και το `.pdb` που δημιουργήθηκαν στην προηγούμενη φάση της προετοιμασίας. Η ελαχιστοποίηση πραγματοποιείται αρχικά για τα άτομα που δεν ανήκουν στην κύρια αλυσίδα με βήμα 2000 (ένα βήμα αντιστοιχεί σε 2 fs). Στη συνέχεια η θερμοκρασία του συστήματος αυξάνεται σταδιακά με βήμα 20 μέχρι να φτάσει την επιθυμητή θερμοκρασία των 320K. Καθ' όλη τη διάρκεια της θέρμανσης διατηρούνται σταθερά τα άτομα Ca.

Οι περισσότερες προσομοιώσεις μοριακής δυναμικής εκτελούνται υπό συνθήκες σταθερής N, V, E, αλλά πιο καινούργιες μέθοδοι εκτελούν προσομοιώσεις με σταθερές N, T και P για να μιμούνται καλύτερα τις πειραματικές συνθήκες. Αυτό επιτεύχθηκε με τη χρήση της μεθόδου Langevin dynamics, που είναι μια προσέγγιση για τη μαθηματική μοντελοποίηση της δυναμικής των μοριακών συστημάτων που χαρακτηρίζεται από τη χρήση απλοποιημένων μοντέλων ενώ αντικαθιστά παραλειπόμενους βαθμούς ελευθερίας με τη χρήση των στοχαστικών διαφορικών εξισώσεων. (Schlick, Tamar, 2002) Ο υπολογισμός των ηλεκτροστατικών αλληλεπιδράσεων έγινε βάσει της μεθόδου Particle Mesh Ewald (PME), η οποία περιγράφει ηλεκτροστατικές αλληλεπιδράσεις σε ένα τρισδιάστατο πλέγμα. Ο αλγόριθμος SHAKE χρησιμοποιήθηκε για τον περιορισμό όλων των δεσμών μεταξύ υδρογόνων και άλλων ατόμων.

## Παραγωγική Φάση

Το τελικό βήμα της προσομοίωσης είναι να εκτελεστεί η προσομοίωση σε «παραγωγική» φάση για την επιθυμητή χρονική διάρκεια. Αυτό μπορεί να είναι από αρκετές εκατοντάδες ps μέχρι ns ή και περισσότερο. Κατά τη παραγωγική φάση οι θερμοδυναμικές παράμετροι μπορούν να υπολογιστούν,

ώστε η προσομοίωση να συμφωνεί με μία από τα σύνολα που περιγράφεται παραπάνω. (Schlick, Tamar 2002) Για να ξεκινήσει αυτή η φάση είναι απαραίτητο ένα αρχείο διαμόρφωσης της παραγωγικής φάσης, το `equi.namd` file, στο οποίο καθορίζεται από τον χρήστη ο τρόπος που θα γίνει η προσομοίωση. Οι συντεταγμένες, οι ταχύτητες και η ενέργεια του συστήματος που καταγράφηκαν στο τελευταίο βήμα της προηγούμενης φάσης χρησιμοποιούνται ως είσοδο για την έναρξη της παραγωγικής φάσης. Για την παραγωγική φάση χρησιμοποιήθηκε ο αλγόριθμος Verlet-I.

## 4. Ανάλυση Προσομοιώσεων

## 4.1 Εισαγωγή

Στόχος της παρούσας εργασίας, όπως έχει αναφερθεί και παραπάνω, είναι η μελέτη της διαδικασίας αναδίπλωσης του πεπτιδίου EMPGVLRN-NH<sub>2</sub> με τη χρήση physic-based μεθόδων καθώς και η σύγκριση των αποτελεσμάτων της προσομοιώσεως με πειραματικά δεδομένα. Θα προσπαθήσουμε δηλαδή, μέσα από τις πληροφορίες που λάβαμε από τις προσομοιώσεις να εξηγήσουμε τη συμπεριφορά του πεπτιδίου. Για το σκοπό αυτό χρησιμοποιήθηκαν διάφορες τεχνικές οι οποίες θα αναλυθούν στη συνέχεια για την καλύτερη κατανόηση των ευρημάτων. Στις αναλύσεις συμπεριλαμβάνονται οι εξής: 1) Ανάλυση χαρτών και γραφημάτων βασισμένων στη μέση απόκλιση (RMSD based matrices), 2) Ανάλυση δευτεροταγούς δομής (STRIDE και DSSP), 3) Ανάλυση βασισμένη σε ομαδοποίηση (cluster analysis), και 4) Ανάλυση κυρίων συνιστωσών (Principal Component Analysis). Στο τέλος θα γίνει σύγκριση των δομών που προέκυψαν από τα πειράματά μας με την πειραματικά προσδιορισμένη NMR δομή.

## 4.2 RMSD based Ανάλυση των Τροχιακών

Μια ανάλυση που χρησιμοποιείται συχνά στις αναλύσεις προσομοιώσεων είναι η RMSD (Root-Mean-Square Deviation). Συνήθως η RMSD χρησιμοποιείται για να κάνει μια ποσοτική σύγκριση μεταξύ της δομής μιας μερικώς διπλωμένης πρωτεΐνης και τη δομής της, στη φυσική κατάσταση. Η RMSD είναι το μέτρο της μέσης απόστασης μεταξύ των ατόμων (συνήθως τα άτομα του σκελετού) μιας πρωτεΐνης. Στη μελέτη των σφαιρικών διαμορφώσεων μιας πρωτεΐνης μετράται η ομοιότητα στην τριτοταγή δομή συνήθως με τη RMSD των C $\alpha$  ατομικών συντεταγμένων μετά από τη βέλτιστη υπέρθεση του σώματος. (Theobald 2005)

Όσο μικρότερη είναι η τιμή RMSD τόσο πιο χωροταξικά ισοδύναμες είναι οι δύο δομές. Για να θεωρηθούν δύο δομές ταυτόσημες θα πρέπει το RMSD να έχει τιμή 0.0 Å. Κατά σύμβαση δομές που έχουν RMSD μικρότερο του 2.0 Å μπορούν να θεωρηθούν όμοιες. Βασίζεται στον υπολογισμό των ευκλείδειων αποστάσεων των ατόμων και η μαθηματική του έκφραση δίνεται από την εξίσωση:

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

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

Όταν ένα δυναμικό σύστημα κυμαίνεται γύρω από κάποια καλά καθορισμένη μέση θέση, η RMSD μπορεί να αναφέρεται ως RMSF ή root mean square fluctuation. Μας βοηθά να κατανοήσουμε τη σταθερότητα της δομής σε σχέση με το χρόνο, καθώς αυτή τη φορά η μέση δομή είναι η δομή αναφοράς με την οποία συγκρίνεται ολόκληρο το τροχιακό.

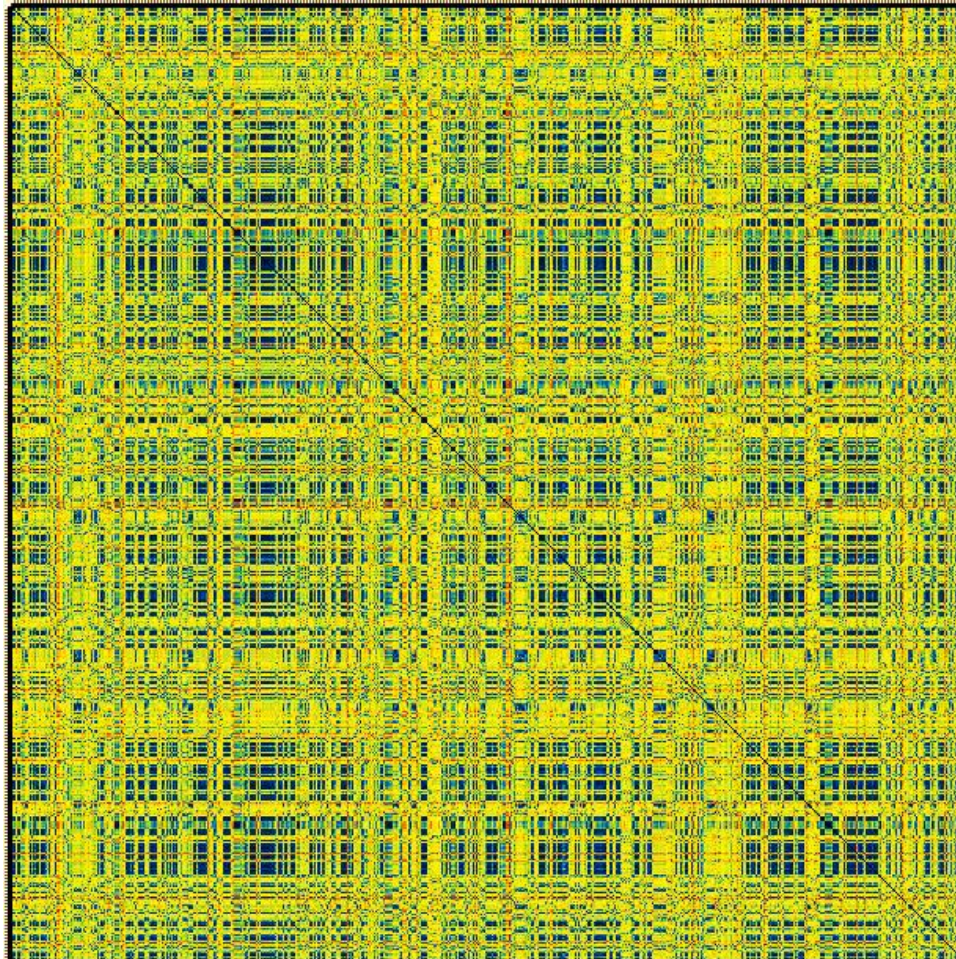
$$RMSF = \sqrt{[\sum(x - \langle x_i \rangle)^2 / N]}$$

Για να φέρουμε την παραπάνω μέτρηση πιο κοντά στην πραγματικότητα μπορούμε να βρούμε ποια δομή από το τροχιακό έχει το χαμηλότερο RMSD σε σχέση με τη μέση δομή και να χρησιμοποιήσουμε αυτή ως αντιπροσωπευτική δομή του τροχιακού. Έτσι οι υπολογισμοί θα αντιστοιχούν σε πραγματικές συντεταγμένες που υιοθετήθηκαν κατά τη διάρκεια της προσομοίωσης και θα μπορούμε να ελέγξουμε κατά πόσο, και για πόσο χρονικό διάστημα, αποκλίνουν ή συγκλίνουν οι δομές που παρήχθησαν με την αντιπροσωπευτική δομή.

Ο υπολογισμός μέσων δομών από τους πίνακες RMSD γίνεται επιλέγοντας τη δομή με το χαμηλότερο RMSD για κάθε σύνολο. Τα αποτελέσματα από μία ανάλυση RMSD είναι ένας πίνακας. Με τους πίνακες RMSD έχουμε τη δυνατότητα να συγκρίνουμε την απόσταση οποιουδήποτε ατόμου (ατόμων Ca, Backbone, Not hydrogen κ.λ.π) μεταξύ όλων των δομών που υιοθετούνται κατά τη διάρκεια της προσομοίωσης. Έτσι μπορούμε να έχουμε μια πιο ξεκάθαρη εικόνα της κίνησης κατά τη διάρκεια της προσομοίωσης. Ο πίνακας RMSD είναι ένα τετράγωνο ανάλογο του μήκους της εκάστοτε προσομοίωσης. Στους δύο άξονες τοποθετούνται οι δομές που θέλουμε να συγκρίνουμε με αύξουσα σειρά, ξεκινώντας από την πρώτη δομή και τελειώνοντας με την τελευταία δομή του τροχιακού. Σε περίπτωση που το τροχιακό συμπεριλαμβάνει πολύ μεγάλο αριθμό δομών υπάρχει η δυνατότητα χρήσης βήματος στη δημιουργία του πίνακα, προκειμένου να μικρύνει το μέγεθός του.

Η βαθμολόγηση του RMSD μεταξύ των δομών γίνεται σε χρωματική κλίμακα. Η διαγώνιος του πίνακα έχει πάντα το χρώμα που αντιστοιχεί στο ελάχιστο RMSD. Καθώς αυξάνει το RMSD των ατόμων αλλάζει και η απόχρωση στον πίνακα. Οι μπλε περιοχές αντιστοιχούν σε χαμηλές τιμές RMSD και υποδεικνύουν την ύπαρξη σταθερών δομών. Η έκταση των περιοχών δηλώνει το χρόνο που πέρασε το πεπτίδιο στις δομές αυτές. Οι κίτρινες περιοχές αντιστοιχούν σε μεσαίες τιμές RMSD και οι κόκκινες περιοχές σε μεγάλες τιμές. Η διακύμανση από κίτρινες σε κόκκινες περιοχές υποδεικνύει όλο

και μεγαλύτερες δομικές διαφορές καθώς και την έλλειψη σταθερών διαμορφώσεων. Η Εικόνα 4.1 αποτελεί μια χρωματική αναπαράσταση των τιμών του RMSD που προέκυψαν. Ο πίνακας κατασκευάστηκε με τη χρήση του προγράμματος GRCARMA (Koukos, Glykos 2013).

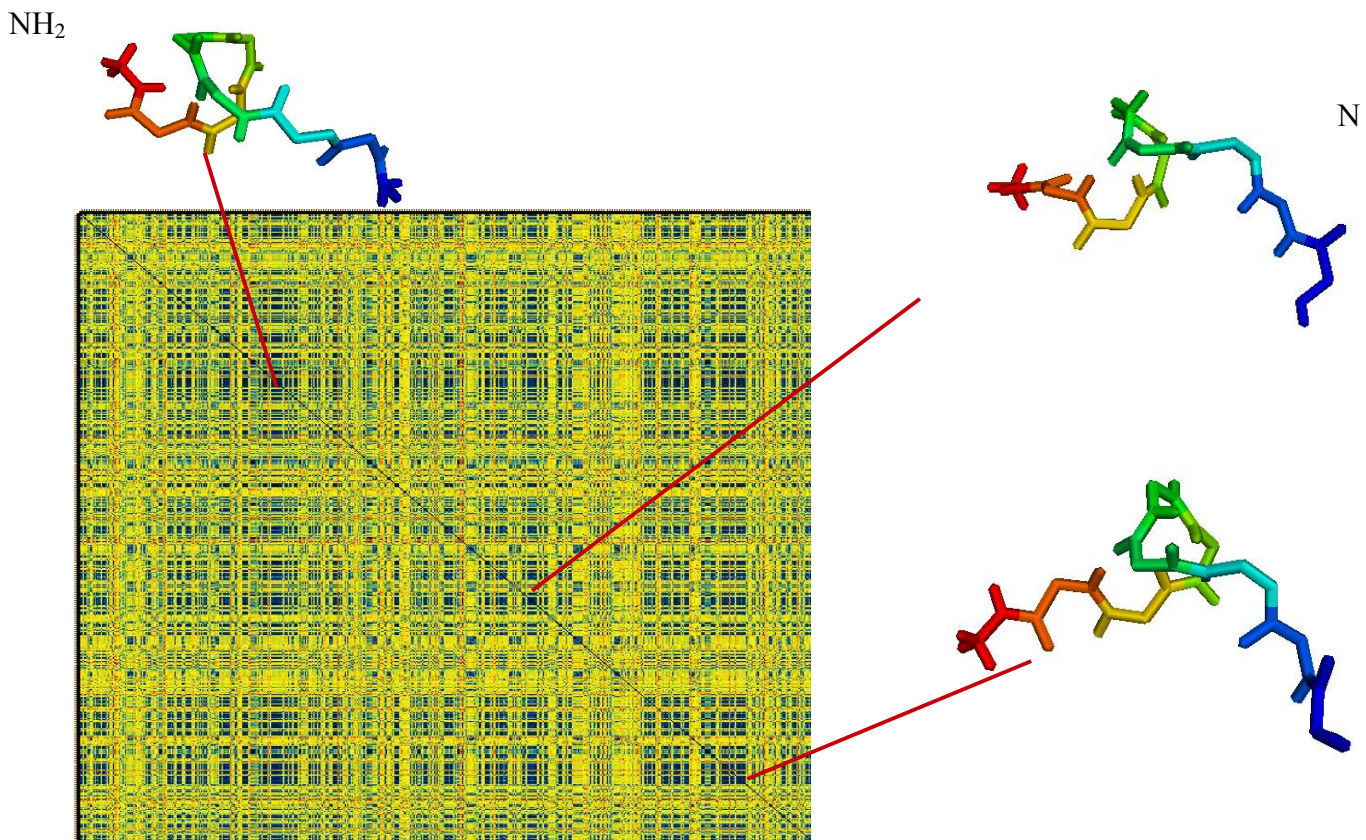


**Εικόνα 4.1:** Στην απεικόνιση RMSD κάθε σημείο της διαγωνίου που ξεκινά από πάνω αριστερά αντιστοιχεί σε μία διαμόρφωση από το ολικό τροχιακό (11.323.925 διαμορφώσεις). Βλέπουμε τη χρωματική αναπαράσταση των τιμών του RMSD που προέκυψαν, όπου οι περιοχές με σκούρες αποχρώσεις του μπλε αντιστοιχούν σε τιμές κοντά στο 0 και οι πιο ανοιχτές αποχρώσεις αντιστοιχούν σε μεγαλύτερες τιμές του RMSD. Οι κίτρινες αποχρώσεις αντιστοιχούν σε μεσαίες τιμές και οι κόκκινες αντιστοιχούν σε μεγάλες. Το RMSD υπολογίστηκε για τα Cas όλων των καταλοίπων του πεπτιδίου. Οι μπλε περιοχές που δεν ανήκουν στη διαγώνιο δηλώνουν ομοιότητα μεταξύ των δομών στις μπλε περιοχές της διαγωνίου. Ο υπολογισμός του Πίνακα RMSD έγινε με τη χρήση του προγράμματος GRCARMA.

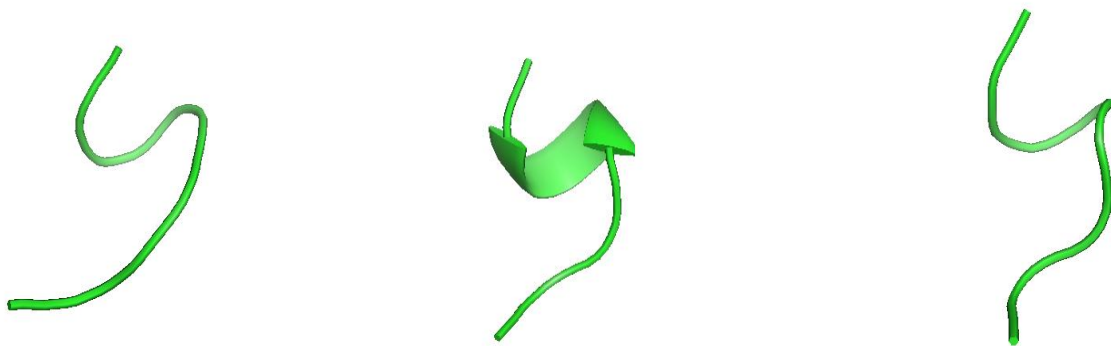
Αυτό που μας ενδιαφέρει σε έναν πίνακα RMSD είναι η παρατήρηση των χαμηλότερων τιμών κατά τη διάρκεια της προσομοίωσης, έτσι ώστε να δούμε τη σταθερότητα των δομών στο χρόνο.

Πρακτικά αυτό φαίνεται από το σχηματισμό μπλε τετραγώνων στη διαγώνιο γραμμή. Όσο μεγαλύτερα είναι τα τετράγωνα τόσο περισσότερο παραμένει σταθερή μια δομή. Επιπλέον με τον πίνακα RMSD μπορούμε να ελεγχεί η ομοιότητα των δομών μεταξύ τους, με την παρατήρηση των προβολών τους στα αντίστοιχα τετράγωνα.

Δε φαίνεται να υπάρχει κάποια ιδιαίτερα σταθερή δομή που να υπερισχύει των άλλων. Βλέπουμε ότι οι μπλε περιοχές περιορίζονται σε έκταση, πράγμα που σημαίνει ότι το πεπτίδιο δε προτιμά και δε λαμβάνει κάποια συγκεκριμένη δομή. Όμως συγκρίνοντας τις μέσες δομές από 3 επιλεγμένες περιοχές (Εικόνα 4.2), που φαίνεται να διατηρεί μία σταθερή διαμόρφωση, βλέπουμε ότι το πεπτίδιο έχει λάβει σε όλες παρόμοια διαμόρφωση. Από τις διαμορφώσεις του πεπτιδίου συμπεράνουμε πως είναι αδιάταχτο αφού δεν παρατηρείται κανένα στοιχείο δευτεροταγούς δομής. Η προβολή των διαμορφώσεων έγινε με το πρόγραμμα PyMOL Molecular Graphics System.

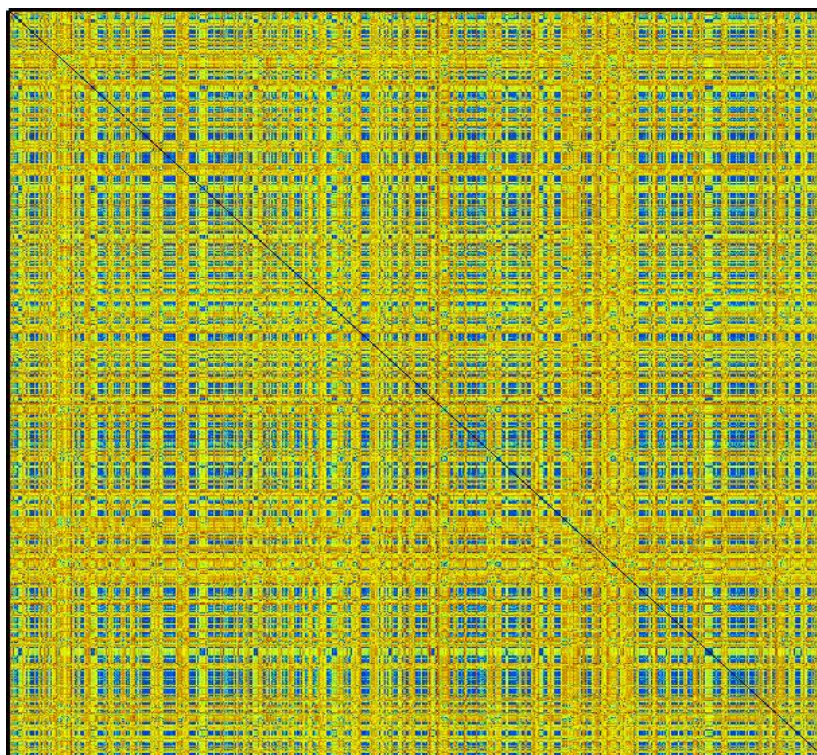


**Εικόνα 4.2.** Παρουσιάζονται οι δομές του τροχιακού που φαίνεται να παραμένουν σταθερές για κάποιο χρονικό διάστημα. Κάθε δομή ανήκει σε μια διαφορετική μπλε περιοχή και είναι συνδεδεμένη με αυτή με βέλος. Ο χρωματισμός έγινε με βάση το όνομα καταλοίπου και οι εικόνες των δομών προετοιμάστηκαν με το πρόγραμμα PyMOL.



**Εικόνα 4.3:** Οι μέσες δομές από τις 3 επιλεγμένες περιοχές της Εικόνα 4.2 σε προβολή cartoon.

Έπειτα με τη βοήθεια του προγράμματος CARMA επιλέχθηκαν μόνο τα frames που αντιστοιχούσαν σε χαμηλές θερμοκρασίες, δηλαδή κάτω των 300K, καθώς τα πειράματα NMR έγιναν σε θερμοκρασία 288K (Dossey 2006). Τα frames που πήραμε ήταν 2.744.964. Με αυτόν τον τρόπο μπορούμε να διαπιστώσουμε αν η θερμοκρασία επηρέασε τη σταθερότητα του μορίου. Ακολούθησε καινούργια ανάλυση RMSD η οποία φαίνεται στην Εικόνα 4.4.



**Εικόνα 4.4:** Πίνακας RMSD για τα frames που αντιστοιχούν σε θερμοκρασία κάτω των 300K, δηλαδή για 2.744.964 διαμορφώσεις.



Όπως φαίνεται και από την σύγκριση των δύο πινάκων RMSD δε φαίνεται να υπάρχει ιδιαίτερη διαφορά στη σταθερότητα των διαμορφώσεων που λαμβάνει το πεπτίδιο, κάτι που επιβεβαιώνεται και από την προβολή των πιο σταθερών διαμορφώσεων του πίνακα RMSD των χαμηλών θερμοκρασιών, οι οποίες ομοιάζουν πολύ με αυτές του πίνακα που περιλαμβάνει όλα τα frames.

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

### 4.3 Ανάθεση Δευτεροταγούς Δομής

Οι αναλύσεις που θα ακολουθήσουν σε αυτή την ενότητα αφορούν την δευτεροταγή δομή που υιοθετούν τα κατάλοιπα του πεπτιδίου κατά τη διάρκεια της προσομοίωσης. Η πιο διαδεδομένη μέθοδος, το DSSP, ορίζει τα στοιχεία της δευτεροταγούς δομής ως επαναλαμβανόμενα μοτίβα δεσμών υδρογόνου. Στην παρούσα εργασία χρησιμοποιήθηκε η μέθοδος STRIDE (STRuctural IDentification).

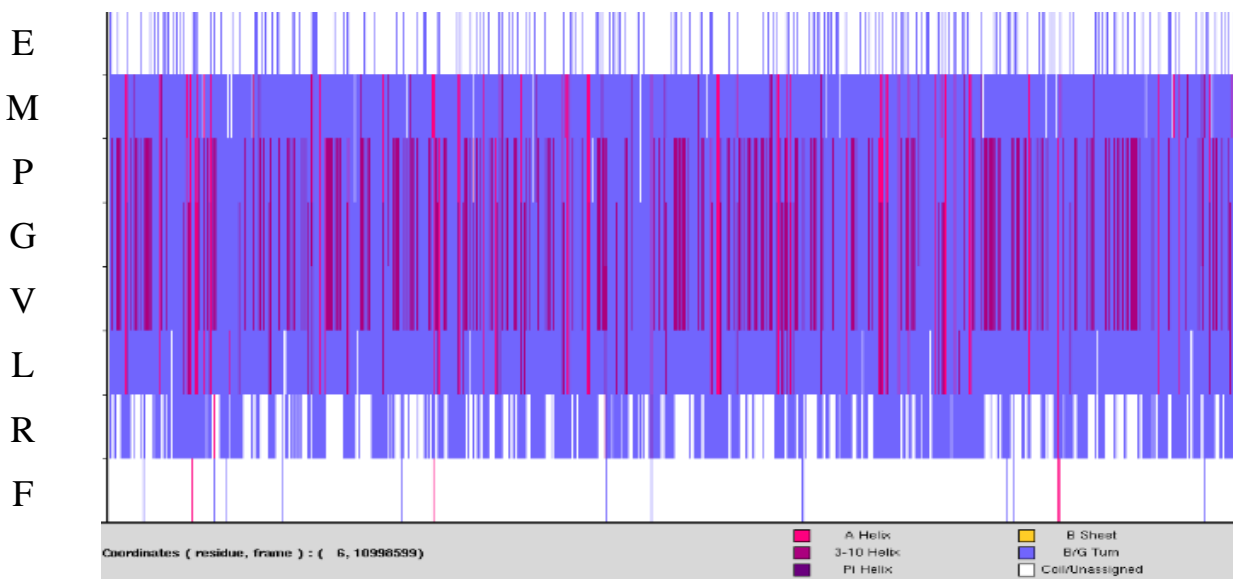
Το STRIDE είναι ένα εργαλείο για την ανάθεση δευτεροταγούς δομής από ατομικές αναλύσεις δομών πρωτεϊνών. Εφαρμόζει έναν αλγόριθμο χρησιμοποιεί συνδυαστική την ενέργεια των δεσμών υδρογόνου και τις στατιστικά προερχόμενες πληροφορίες των διεδρων γωνιών του κύριου σκελετού. (Heinig M, Frishman D 2004)

Το αποτέλεσμα από τον αλγόριθμο STRIDE είναι ένα γράμμα που αντιστοιχεί στη διαμόρφωση κάθε καταλοίπου της αλληλουχίας καθ' όλη την έκταση του τροχιακού. Η αντιστοίχιση γράμματος και δευτεροταγούς δομής είναι η παρακάτω :

- H: α-έλικα
- G: 3-10 έλικα
- I: π-έλικα
- E: Εκτεταμένη β-διαμόρφωση
- B: β-μεμονωμένη γέφυρα

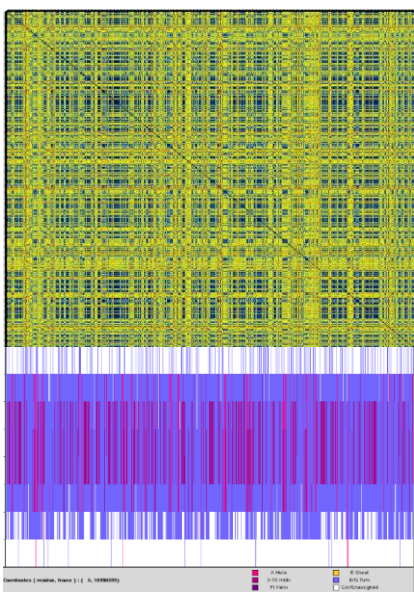
- T: Στροφή
- C: Σπείραμα (καμία από τις παραπάνω διαμορφώσεις)

Προκειμένου να γίνει πιο ευανάγνωστο το αποτέλεσμα, τα γράμματα αντικαταστάθηκαν με χρώμα (Εικόνα 4.5).



**Εικόνα 4.5:** Χρωματική αναπαράσταση των στοιχείων δευτεροταγούς δομής. Η ανάθεση των στοιχείων στην εικόνα έγινε με τη χρήση του αλγορίθμου STRIDE. Σε κάθε σημείο του οριζόντιου άξονα αντιστοιχούν διαμορφώσεις από το τροχιακό, ενώ στον κάθετο άξονα η αλληλουχία των αμινοξέων του πεπτιδίου EMPGVLR F. Οι μπλε περιοχές αντιστοιχούν σε στροφή, οι ροζ σε  $\alpha$ -έλικες και οι ανοιχτό μωβ σε 3-10 έλικες.

Το πεπτίδιο ταλαντεύεται ανάμεσα σε διαμορφώσεις όπως στροφή και έλικες με επικρατέστερη τη στροφή, χωρίς όμως να υιοθετεί κάποια συγκεκριμένη διαμόρφωση για μεγάλο χρονικό διάστημα, πράγμα που συμφωνεί και με την απεικόνιση RMSD (Εικόνα 4.6) Βλέπουμε ότι τη διαμόρφωση 3-10 έλικας την υιοθετούν μόνο τα κατάλοιπα P-G-V. Τα άκρα του πεπτιδίου για το μεγαλύτερο κομμάτι της προσομοίωσης δεν υιοθετούν κάποια συγκεκριμένη δομή, πράγμα που συμφωνεί με την ένδειξη των πειραμάτων NMR, ότι σε αντίθεση με άλλα FLP-18 πεπτίδια που σχηματίζουν ένα βρόγχο στο N-τελικό άκρο, το συγκεκριμένο πεπτίδιο δεν τον σχηματίζει. Επίσης το γεγονός ότι το πεπτίδιο δε λαμβάνει συγκεκριμένη διαμόρφωση για μεγάλο χρονικό διάστημα στηρίζει την υπόθεση ότι το συντηρημένο τμήμα PGVLR F-NH<sub>2</sub> μάλλον είναι αδόμητο.



**Εικόνα 4.6:** Πίνακες RMSD και ανάθεση δομής. Τα αποτελέσματα από τις τρεις μεθόδους φαίνεται να συμφωνούν για κάθε προσομοίωση.

## 4.4 Cluster Analysis

Οι μέθοδοι προσομοίωσης μοριακής δυναμικής παράγουν τροχιακά των ατομικών θέσεων (και προαιρετικά ταχύτητες και ενέργειες) ως συνάρτηση του χρόνου και παρέχουν μια αναπαράσταση του ενεργειακά προσβάσιμου διαμορφωτικού συνόλου ενός συγκεκριμένου μορίου. Με βελτιώσεις στην ταχύτητα των υπολογιστών και της αποδοτικότητας των αλγορίθμων, οι προσομοιώσεις MD εξετάζουν περισσότερες μοριακές και βιομοριακές διαμορφώσεις. Η δυνατότητα ποιοτικού και ποσοτικού διαχωρισμού αυτών των διαμορφώσεων σε ουσιώδεις ομάδες είναι ένα δύσκολο και σημαντικό έργο. (Wolf A, Kirschner 2013) Τεχνικές όπως το clustering παρέχουν ένα μέσο για την ομάδα και δίνουν νόημα στις πληροφορίες του τροχιακού. (Shao et al 2007)

Το Cluster Analysis είναι μια διαδικασία ομαδοποίησης, με χρήση αλγορίθμου, μιας σειράς από αντικείμενα με τέτοιο τρόπο, ώστε τα αντικείμενα στην ίδια ομάδα (που ονομάζεται cluster) να μοιάζουν περισσότερο (κατά κάποιο τρόπο) μεταξύ τους σε σχέση με αυτά των άλλων ομάδων (clusters) (Bailey, Ken. 1994). Η ανάλυση κύριων συνιστωσών επιτρέπει τον εντοπισμό συγκεκριμένων μοτίβων κίνησης σε ένα τροχιακό. Η ομαδοποίηση των δεδομένων γίνεται με βάση τις κύριες κινήσεις του τροχιακού, οι οποίες αποτυπώνονται στις κύριες συνιστώσες (PC) με τα μεγαλύτερα ιδιοδιανύσματα.

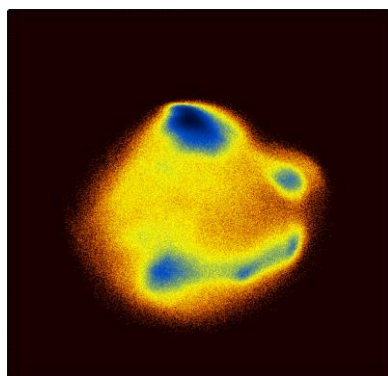
Στην παρούσα πτυχιακή εργασία για την ομαδοποίηση των διαμορφώσεων των τροχιακών χρησιμοποιήθηκε η Ανάλυση των Κύριων Συνιστωσών στο καρτεσιανό σύστημα (Cartesian principal component analysis) και η Ανάλυση Κύριων Συνιστωσών των διεδρων γωνιών  $\varphi/\psi$  (dihedral principal component analysis) αφού πρώτα εφαρμόστηκε προσαρμογή του πεπτιδίου με αφαίρεση του 9<sup>ο</sup> καταλοίπου, δηλαδή την απομάκρυνση του NME (N-methyl amide capping group) καταλοίπου.

Η Ανάλυση Κύριων Συνιστωσών (PCA, Principal Component Analysis), που ονομάζεται επίσης ανάλυση quasiharmonic ή μέθοδος essential dynamics, είναι μία από τις πιο δημοφιλείς μεθόδους για τη συστηματική μείωση της διάστασης ενός πολύπλοκου συστήματος. Η προσέγγιση βασίζεται σε έναν πίνακα συνδιακύμανσης (covariance matrix), ο οποίος παρέχει πληροφορίες σχετικά με τις συσχετίσεις δύο σημείων του συστήματος. Η PCA αντιπροσωπεύει ένα γραμμικό μετασχηματισμό που διαγωνιοποιεί τον πίνακα συνδιακύμανσης και έτσι αφαιρεί τις στιγμιαίες γραμμικές συσχετίσεις μεταξύ των μεταβλητών. (Ichiye Karplus 1991, Amadei et al 1993) Έτσι δημιουργείται ένα νέο σύνολο μεταβλητών, οι κύριες συνιστώσες (principal components, PCs), οι οποίες είναι μη συσχετιζόμενες και ταξινομούνται έτσι ώστε ορισμένες εκ των πρώτων να διατηρούν το μεγαλύτερο μέρος της διακύμανσης που παρουσιάζεται σε όλες τις αρχικές μεταβλητές (I.T. Jolliffe, 2002).

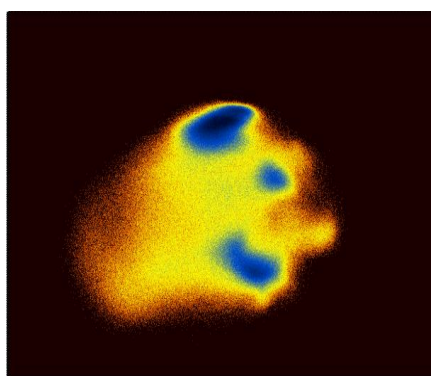
### **Cartesian Principal Component Analysis**

Ο τύπος PCA που χρησιμοποιεί καρτεσιανές συντεταγμένες για τις αναλύσεις ονομάζεται Cartesian Principal Component Analysis (cPCA). Οι αναλύσεις έγιναν με το πρόγραμμα GRCARMA. Στην ανάλυση λήφθηκαν υπόψιν όλα τα βαριά άτομα, έτσι ώστε να εξετάσουμε την κίνηση του πεπτιδίου συμπεριλαμβανομένων και των πλευρικών αλυσίδων. Στην Εικόνα 4.7 παρουσιάζεται το ενεργειακό τοπίο που προκύπτει από την προβολή των τροχιακών στα επίπεδα που ορίζεται από τους τρεις πρώτους eigenvectors με τα μεγαλύτερα eigenvalues, ανά ζεύγη.

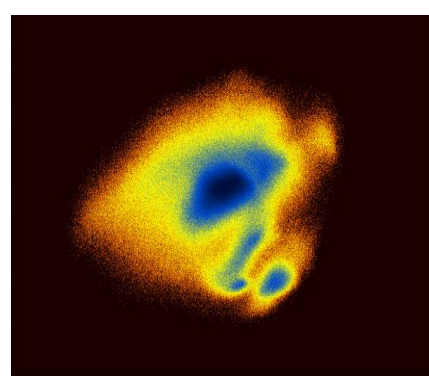
Τα πολυπληθέστερα cluster εμφανίζονται με μπλε χρώμα. Ιδανικά θα υπήρχε ένα μόνο συμπαγές cluster που θα αντιπροσωπεύει μία κυρίαρχη διαμόρφωση. Η έκταση και η θέση των τοπικών ελαχίστων στα διαγράμματα διευκολύνει την αντιστοίχιση τους με τις κύριες ομάδες.



PC\_01\_02



PC\_01\_03



PC\_02\_03

**Εικόνα 4.6:** Παρουσίαση των ενεργειακών τοπίων ( $\Delta G$ , energy landscapes) στο επίπεδο των τριών κύριων συνιστωσών (Pcs) ανά ζεύγη. Οι μπλε περιοχές υποδεικνύουν την ύπαρξη τοπικών ελαχίστων ελεύθερης ενέργειας για συγκεκριμένες ομάδες δομών. Όσο πιο σκούρα είναι η απόχρωση του μπλε τόσο μικρότερη είναι η τιμή της ελεύθερης ενέργειας, ενώ το μέγεθος των περιοχών δηλώνει το μέγεθος της ομάδας.

Στη συνέχεια απομονώθηκε η μέση δομή για κάθε cluster (Εικόνα 4.8)



cluster 1



cluster 2



cluster 3

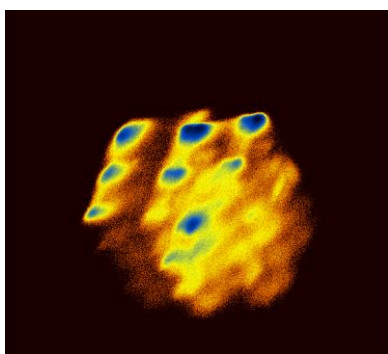
**Εικόνα 4.8 :** Μέση δομή για κάθε cluster του τροχιακού όπως υπολογίστηκαν από την cPCA .

Οι αντιπροσωπευτικές δομές των ομάδων ταιριάζουν σε μεγάλο βαθμό με τις μέσες δομές που απομονώθηκαν από τον πίνακα RMSDs (Εικόνα 4.1).

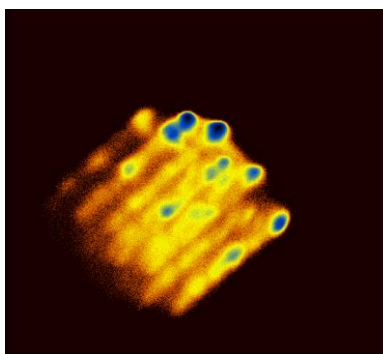
## Dihedral Angle Principal Component Analysis

Πειράματα έδειξαν ότι PCA αναλύσεις που χρησιμοποιούσαν καρτεσιανές συντεταγμένες δεν απέδιδαν το σωστό τοπίο ελεύθερης ενέργειας λόγω ύπαρξης artifacts από την ανάμιξη της εσωτερικής και της συνολικής κίνησης (Mu et al 2005). Πρόσφατα, προτάθηκε η χρήση εσωτερικών αντί για καρτεσιανών συντεταγμένων σε μια PCA. Σε βιομόρια, ιδίως, η εξέταση των διέδρων γωνιών φαίνεται ελκυστική, διότι άλλες εσωτερικές συντεταγμένες, όπως μήκη δεσμών και γωνίες δεσμών συνήθως δεν υπόκεινται σε αλλαγές μεγάλου μεγέθους. (Altis et al 2007) Έτσι το ενεργειακό τοπίο που προκύπτει από τις αναλύσεις dPCA (dihedral PCA) εμφανίζει πολλά ενεργειακά ελάχιστα τα οποία αντιστοιχούν σε καλά καθορισμένες δομές. Συνεπώς είναι πιο τραχύ από το (λείο) ενεργειακό τοπίο που παράγουν οι αναλύσεις cPCA.

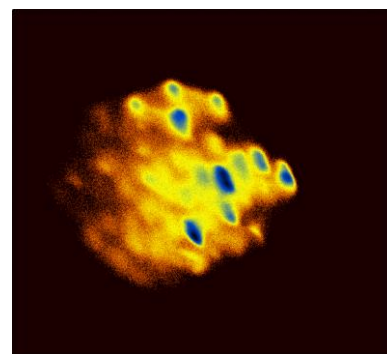
Η dPCA είναι πλέον μια καθιερωμένη μέθοδος για τη μείωση των διαστάσεων ενός υψηλών διαστάσεων συνόλου δεδομένων. Βασίζεται στις διέδρες γωνίες  $\phi_n$  και  $\psi_n$  του πεπτιδίου. Στην παρούσα πτυχιακή εργασία πραγματοποιήθηκε αναλύση dPCA, που αφορά τόσο τις διέδρες γωνίες  $\phi$  και  $\psi$ , όπου αγνοούνται οι πλευρικές ομάδες (Εικόνα 4.9), όσο και των γωνιών  $\phi$ ,  $\psi$  συμπεριλαμβανομένης και της  $\chi_1$  γωνίας (Εικόνα 4.11). Οι αναλύσεις dPCA περιορίζονται στην κίνηση του πεπτιδικού σκελετού. Οι αναλύσεις έγιναν με χρήση του προγράμματος GRCARMA.



PC\_01\_02



PC\_01\_03

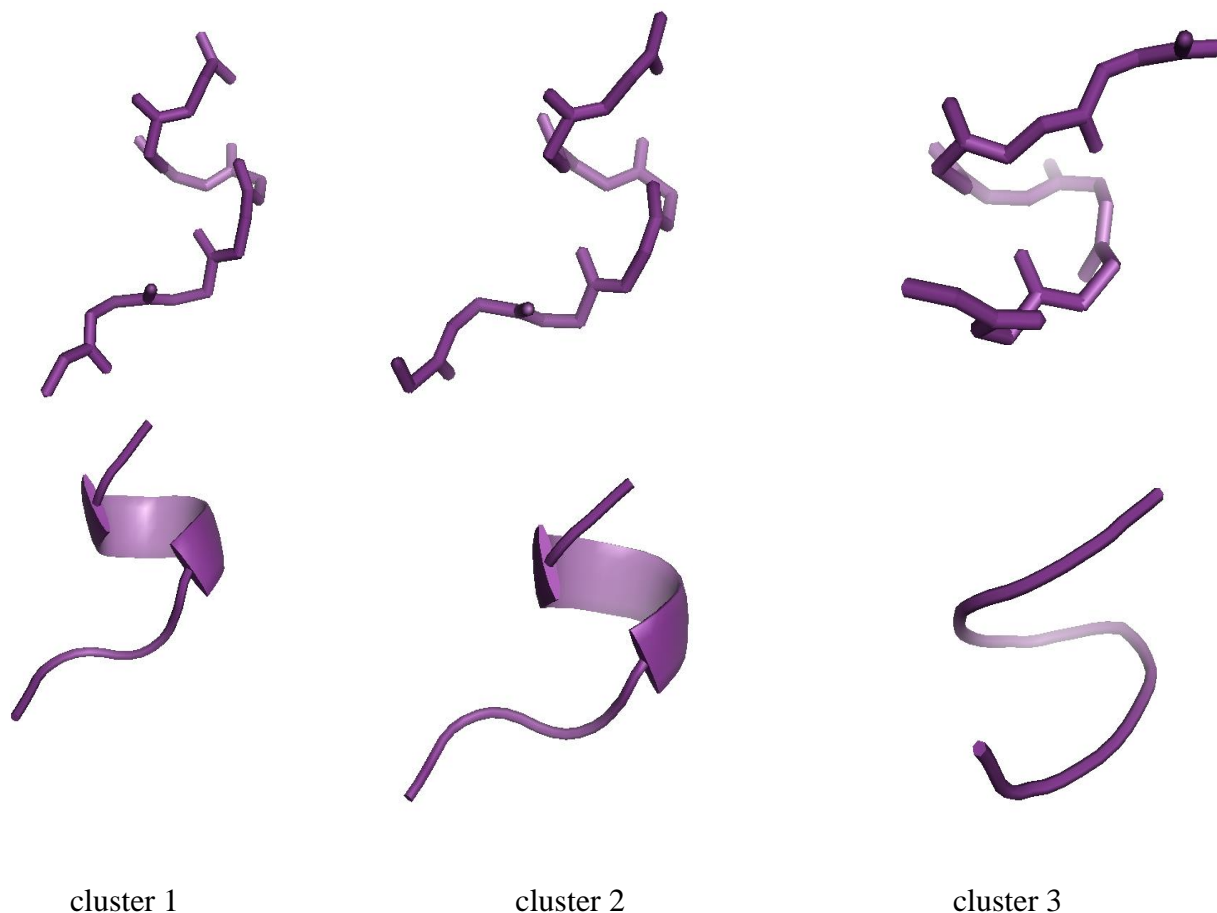


PC\_02\_03

**Εικόνα 4.09:** Παρουσίαση των ενεργειακών τοπίων ( $\Delta G$ , energy landscapes) στο επίπεδο των τριών κυρίαρχων συνιστωσών (Pcs) χωρίς τον υπολογισμό της  $\chi_1$  γωνίας.

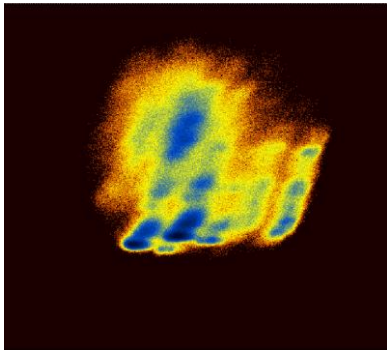
Στο πρώτο διάγραμμα ο διαχωρισμός των ομάδων είναι ασαφής, ενώ στο δεύτερο και στο τρίτο οι κύριες ομάδες διακρίνονται εύκολα.

Για κάθε κύρια ομάδα απομονώσαμε τις αντιπροσωπευτικές δομές (Εικόνα 4.10)

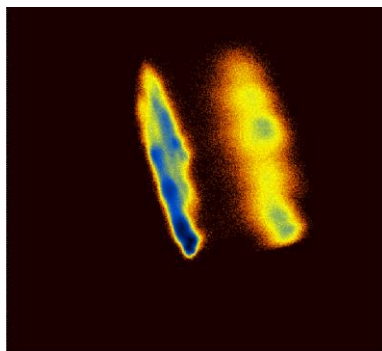


**Εικόνα 4.10:** Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA χωρίς τις γωνίες  $\chi_1$  σε προβολή sticks (πάνω) και cartoon (κάτω).

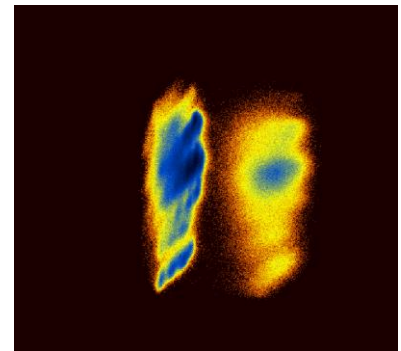
Τα αποτελέσματα των dPCAs, με τις  $\chi_1$  γωνίες (Εικόνα 4.11), διαφέρουν από αυτά της cPCA κυρίως ως προς τον αριθμό των διακριτών ομάδων που εντοπίστηκαν από κάθε μέθοδο. Για λόγους απλότητας, στα αποτελέσματα της ομαδοποίησης δεν παρουσιάζονται οι ομάδες που εντοπίστηκαν από τις μεθόδους με ποσοστό πολύ μικρότερο της μονάδας, γεγονός που φαίνεται από τους μικρούς πληθυσμούς στα διαγράμματα ελεύθερης ενέργειας.



PC\_01\_02



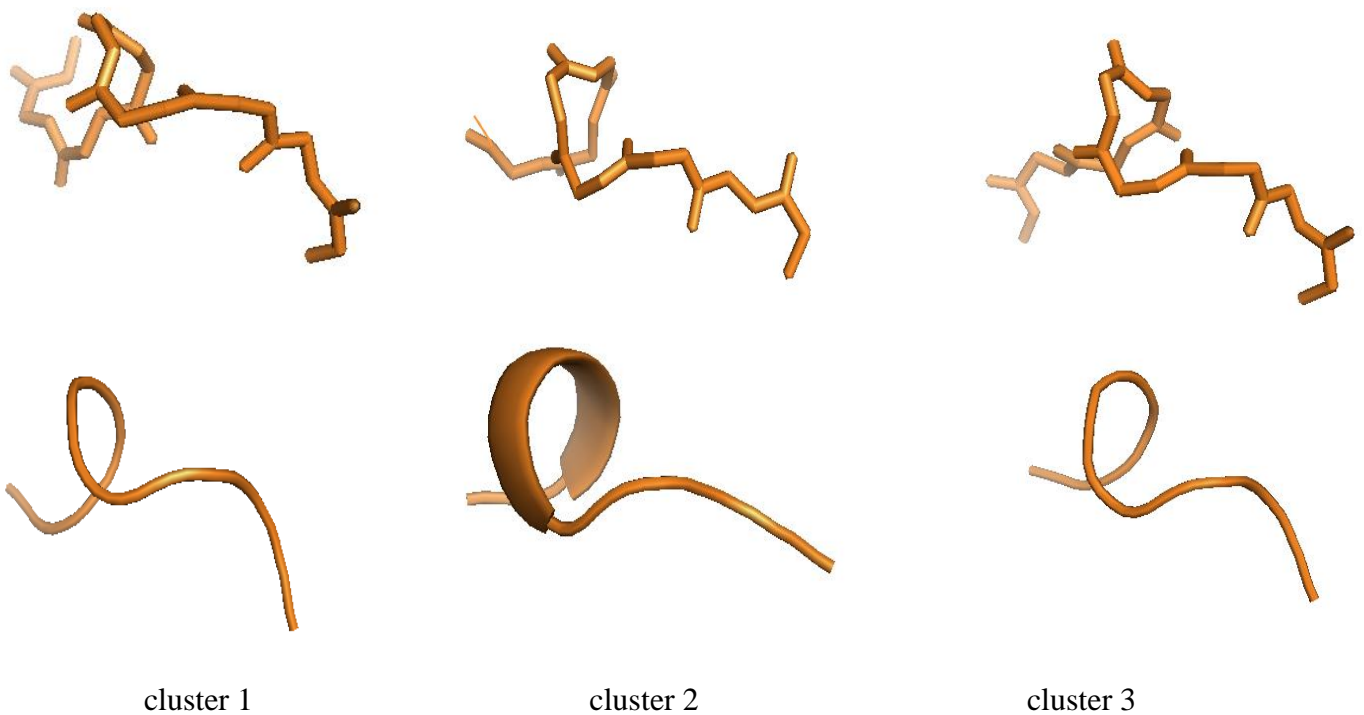
PC\_01\_03



PC\_02\_03

**Εικόνα 4.11:** Παρουσίαση των ενεργειακών τοπίων ( $\Delta G$ , energy landscapes) στο επίπεδο των τριών κυρίαρχων συνιστωσών (Pcs) με τον υπολογισμό της  $\chi_1$  γωνίας.

Για κάθε κύρια ομάδα απομονώσαμε τις αντιπροσωπευτικές δομές (Εικόνα 4.12)



**Εικόνα 4.10:** Αντιπροσωπευτικές δομές για κάθε κύρια ομάδα του τροχιακού όπως υπολογίστηκαν από την dPCA χρησιμοποιώντας τις γωνίες  $\chi_1$  σε προβολή sticks (πάνω) και cartoon (κάτω).



Παρόλο που δε συμπεριλαμβάνονται οι πλευρικές ομάδες στην ανάλυση dPCA, τα ενεργειακά τοπία που προέκυψαν φαίνεται να έχουν περισσότερα clusters σε σχέση με τα ενεργειακά τοπία των αναλύσεων Cartesian PCA, όπως αναμένεται από αυτή την ανάλυση. Επιβεβαιώνεται συνεπώς ο κανόνας των τραχέων ενεργειακών πεδίων που προκύπτουν από την dPCA. Με την ανάλυση dPCA μπορούμε να δούμε περισσότερες ομαδοποιημένες δομές που αντιστοιχούν σε χαμηλά ενεργειακά επίπεδα και οι οποίες δε διαφαίνονται στην ανάλυση cPCA.

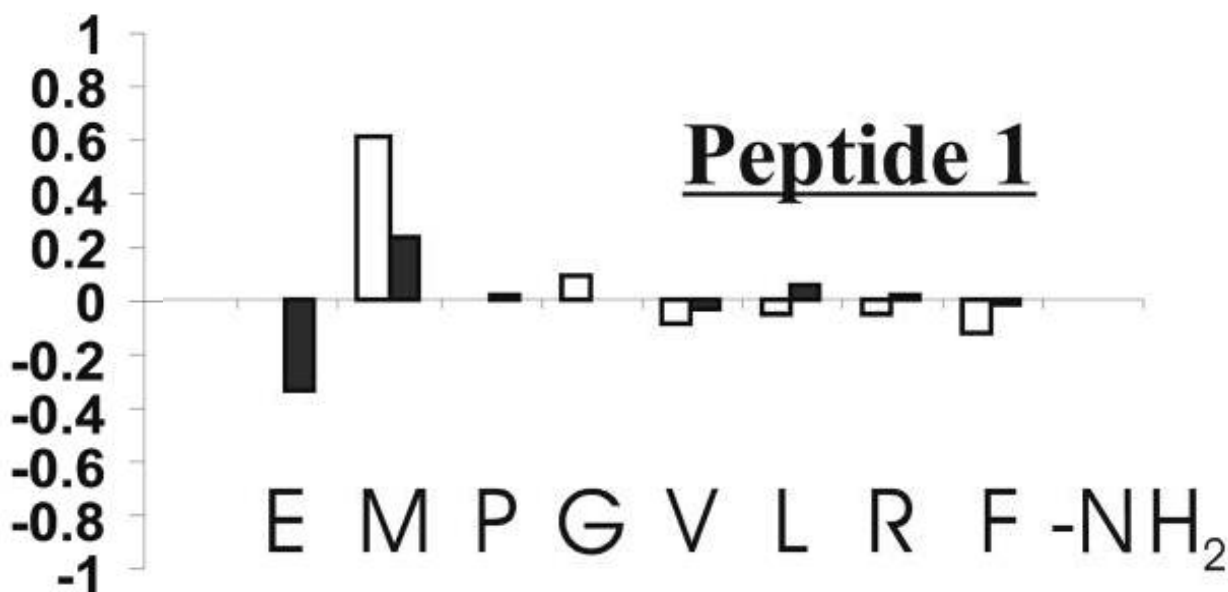
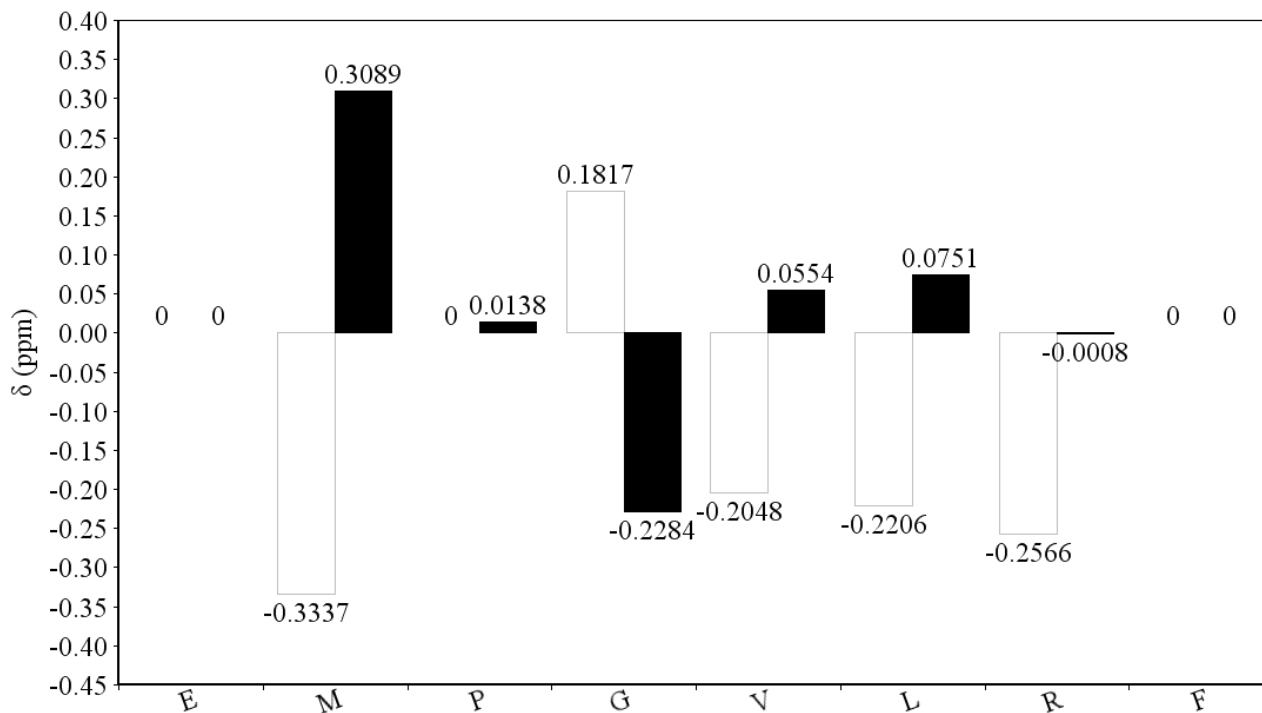
## 4.5 Chemical Shifts

Στον πυρηνικό μαγνητικό συντονισμό (NMR), τα chemical shifts είναι η συχνότητα συντονισμού του πυρήνα σε σχέση με ένα πρότυπο. Συχνά, η θέση και ο αριθμός των chemical shifts είναι διαγνωστικά της δομής ενός μορίου. Στην παρούσα εργασία έγινε υπολογισμός των chemical shifts του πεπτιδίου με το πρόγραμμα SPARTA+ (Shen, Bax 2010) για τα 11.323.925 frames με βήμα 10. Από τα αποτελέσματα του προγράμματος υπολογίστηκε το μέσο όρο των chemical shifts. Εμείς θα ασχοληθούμε μόνο με τα chemical shifts του Η. Τα αποτελέσματα παρουσιάζονται στον πίνακα 4.1.

Αριθμός καταλοίπου	Όνομα καταλοίπου	Όνομα ατόμου	Chemical shift	Τυπική απόκλιση
1	E	HA	0.0000	0.0000
2	M	HA	0.3089	0.1606
2	M	HN	-0.3337	0.2614
3	P	HA	0.0138	0.1463
4	G	HA	-0.2284	0.1394
4	G	HN	0.1817	0.2705
5	V	HA	0.0554	0.2184
5	V	HN	-0.2048	0.2895
6	L	HA	0.0751	0.2292
6	L	HN	-0.2206	0.3249
7	R	HA	-0.0008	0.2034
7	R	HN	-0.2566	0.3706
8	F	HA	0.0000	0.0000
8	F	HN	0.0000	0.0000

**Πίνακας 4.1:** Το μέσο όρο των chemical shifts όπως υπολογίστηκε από το πρόγραμμα SPARTA+.

Στα διαγράμματα που ακολουθούν παρουσιάζονται τα chemical shifts όπως υπολογίστηκαν από το πρόγραμμα SPARTA+ και τα πειράματα. Οι υπολογισμοί των chemical shifts έγιναν στις ίδιες συνθήκες με τις πειραματικές μετρήσεις (pH 2.3)



**Διάγραμμα 4.1:** Τα chemical shifts (ppm) για κάθε κατάλοιπο, όπως υπολογίστηκαν από το πρόγραμμα SPARTA+ (πάνω) και όπως υπολογίστηκαν από τα πειράματα (κάτω) (Dossey, 2006). Οι μαύρες μπάρες αντιπροσωπεύουν τα άλφα πρωτόνια και οι λευκές μπάρες πρωτόνια αμιδίου.

Οι διαφορές ανάμεσα στα chemical shifts που υπολογίστηκαν με το πρόγραμμα με αυτά που υπολογίστηκαν πειραματικά είναι εμφανείς. Οι μοναδικές μετρήσεις που ταιριάζουν αφορούν τα άλφα πρωτόνια των καταλοίπων P, L και R.

Ένας τρόπος με τον οποίο ένα μέτρο της ποιότητας προσαρμογής μπορεί να κατασκευαστεί, στην περίπτωση που είναι γνωστή η διακύμανση του σφάλματος μέτρησης, είναι να κατασκευαστεί ένα σταθμισμένο άθροισμα των τετραγώνων των σφαλμάτων:

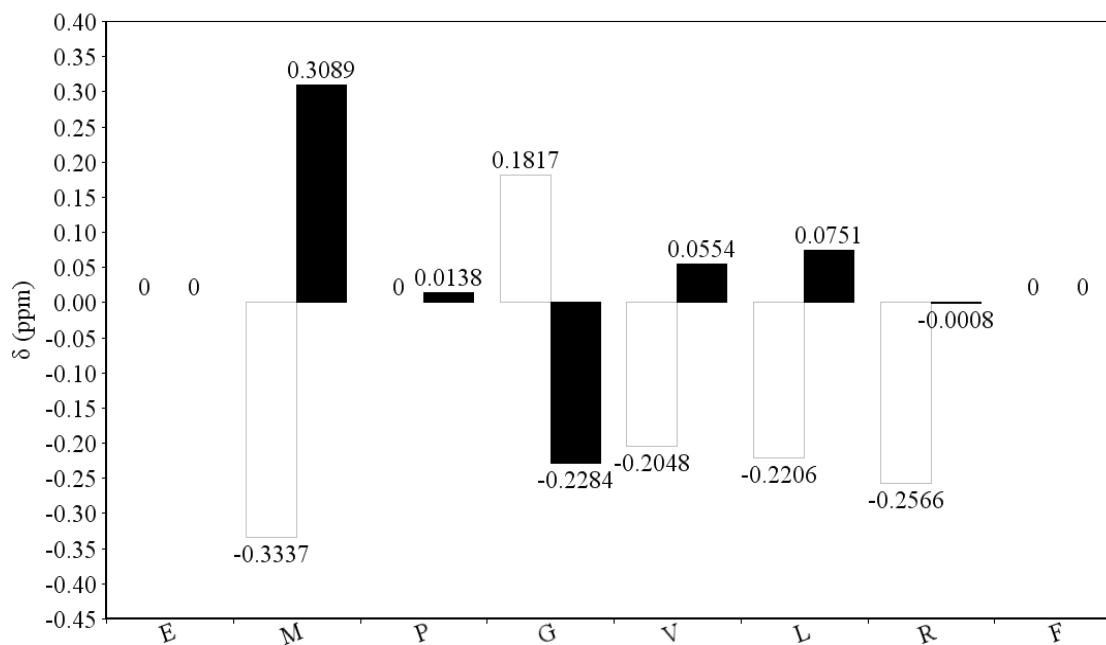
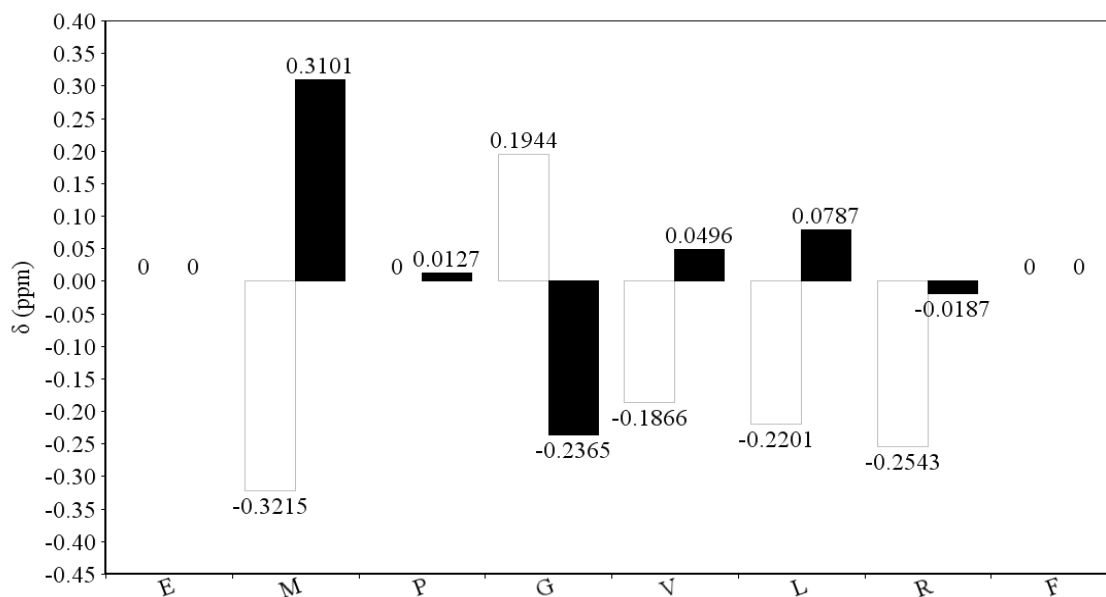
$$x_{red}^2 = \frac{1}{\nu} \sum \frac{(O-E)^2}{\sigma^2}$$

όπου  $\sigma$  είναι η γνωστή διακύμανση των παρατηρήσεων,  $O$  είναι η τιμή της παρατήρησης,  $E$  είναι τα θεωρητικά δεδομένα,  $\nu$  είναι ο αριθμός των βαθμών ελευθερίας, συνήθως δίδεται από τον τύπο  $N-n-1$ , όπου  $N$  είναι ο αριθμός των παρατηρήσεων, και  $n$  είναι ο αριθμός των fitted παραμέτρων.

Ως γενικός κανόνας (μόνο όταν η διακύμανση του σφάλματος μέτρησης είναι γνωστή εκ των προτέρων και δεν εκτιμάται από τα δεδομένα), αν  $x_{red}^2 \gg 1$  τότε έχουμε μια κακή προσαρμογή του μοντέλου. Αν  $x_{red}^2 > 1$ , μιλάμε για ένα μέτριο μοντέλο (ή ότι η διακύμανση του σφάλματος έχει υποτιμηθεί). Όταν  $x_{red}^2 = 1$  υποδηλώνεται ότι η έκταση του ταιριάσματος μεταξύ των παρατηρήσεων και των θεωρητικών δεδομένων είναι σε συμφωνία με την διακύμανση σφάλματος. Αν  $x_{red}^2 < 1$ , είτε το μοντέλο διαχειρίζεται λάθος το θόρυβο, ή η διακύμανση σφάλματος έχει υπερεκτιμηθεί.

Με το  $x_{red}^2 = 2.22$  μπορούμε να πούμε ότι μιλάμε για ένα μέτριο μοντέλο.

Όπως ανέφερα παραπάνω επιλέχθηκαν μόνο frames που αντιστοιχούν σε χαμηλές θερμοκρασίες, δηλαδή κάτω των 300K. Για αυτά τα frames υπολογίστηκαν ξεχωριστά τα chemical shifts. Στο Διάγραμμα 4.2 φαίνονται αυτά τα chemical shifts σε σύγκριση με τα chemical shifts όλων των frames.



**Διάγραμμα 4.2:** Στο πάνω διάγραμμα φαίνονται τα chemical shifts για τα επιλεγμένα frames που αντιστοιχούν σε θερμοκρασίες 300K, ενώ στο κάτω διάγραμμα φαίνονται τα chemical shifts όλων των frames, όπως υπολογίστηκαν από το πρόγραμμα SPARTA+.

Όπως φαίνεται από τα διαγράμματα δεν υπάρχουν ιδιαίτερες διαφορές ανάμεσα στα chemical shifts των frames που αφορούν χαμηλές θερμοκρασίες με τα υπόλοιπα frames, και έτσι επιβεβαιώνεται το συμπέρασμα ότι η θερμοκρασία δεν επηρέασε τη σταθερότητα και τη διαμόρφωση του πεπτιδίου.

## 5. Συμπεράσματα και Συζήτηση

## 5.1 Συμπεράσματα και Συζήτηση

Το νευροπεπτίδιο της οικογένειας FLP-18, με αλληλουχία EMPGVLRN-NH<sub>2</sub>, αποτελεί έναν από τους προσδέτες του NPR-1, ενός GPCR που ρυθμίζει την τροφική συμπεριφορά στον *C.elegans*. Η συγγένεια των προσδετών φαίνεται να εξαρτάται από τη δομή τους. Στην εργασία αυτή, μελετήθηκε υπολογιστικά η συμπεριφορά του πεπτιδίου αυτού. Ο Dossey και οι συνεργάτες του έδειξαν ότι, το συγκεκριμένο πεπτίδιο, σε αντίθεση με τα υπόλοιπα FLP-18 πεπτίδια, δε σχηματίζει βρόχο στο N-τελικό άκρο και φέρει αδόμητη τη συντηρημένη περιοχή PGVLRN-NH<sub>2</sub>.

Οι προσομοιώσεις του FLP-18 πεπτιδίου 1 εφάρμοζαν το δυναμικό πεδίο AMBER. Στόχος τους ήταν η μελέτη της αναδίπλωσης του πεπτιδίου και η σύγκριση των αποτελεσμάτων με πειραματικά δεδομένα από φασματοσκοπία NMR. Στη συνολική διάρκεια των προσομοιώσεων που διεξήχθησαν, δε φαίνεται να πραγματοποιείται κάποιο μοναδικό γεγονός αναδίπλωσης. Το πεπτίδιο φαίνεται να μη λαμβάνει κάποια συγκεκριμένη δομή, πράγμα που συμφωνεί με τα πειραματικά ευρήματα που θέλουν το πεπτίδιο αδόμητο. Το πεπτίδιο φαίνεται να υιοθετεί κατά διαστήματα κάποια από τα στοιχεία δευτεροταγούς δομής όπως στροφή και έλικες, χωρίς όμως αυτό να συμβαίνει για αρκετά σημαντικό χρονικό διάστημα. Βασικές διαφορές, όμως, εντοπίστηκαν στον υπολογισμό των secondary chemical shifts, τα οποία απέκλιναν αρκετά από τα πειραματικώς υπολογισμένα.

Τα σημεία στα οποία πρέπει να σταθούμε στην παρούσα εργασία είναι το γεγονός ότι το δυναμικό πεδίο AMBER έχει την ικανότητα να αναπαράγει δομές, που φαίνεται να συμφωνούν με τις πειραματικά προσδιορισμένες δομές, πράγμα που φανερώνει την χρησιμότητα των δυναμικών πεδίων στις προσομοιώσεις και την αναγκαιότητα για μελλοντικές βελτιώσεις.

Κανείς δεν μπορεί να περιγράψει τα δομικά χαρακτηριστικά αυτών των εγγενώς μη δομημένων πεπτιδίων με μια ενιαία διαμόρφωση. Ωστόσο, οι αντιπροσωπευτικές δομές, περιέχουν πληροφορίες σχετικά με το σύνολο των διαμορφώσεων που λαμβάνονται από τις προσομοιώσεις. Ως εκ τούτου, μπορεί κανείς να υποστηρίξει ότι μπορεί να παρέχουν χρήσιμες πληροφορίες σχετικά με τα δομικά χαρακτηριστικά των εγγενώς μη δομημένων πρωτεϊνών.

Εάν τα σύνολα των διαμορφώσεων για τις μη δομημένες πρωτεΐνες αποτελούνται από περισσότερα από ένα cluster, η ενιαία αντιπροσωπευτική δομή δεν μπορεί να είναι η φυσικά σχετική δομή. Σε αυτή την περίπτωση, μπορεί κανείς να ορίσει ξεχωριστές αντιπροσωπευτικές δομές για τα διαφορετικά cluster. Λαμβάνοντας υπόψη πειραματικές δυσκολίες όσον αφορά το διαρθρωτικό προσδιορισμό των μη δομημένων πρωτεϊνών, οι υπολογιστικές προσεγγίσεις που βασίζονται στην ιδέα

της αντιπροσωπευτικής δομής θα παίξουν σημαντικό ρόλο στην έρευνα σημαντικών διαδικασιών όπως η συσσωμάτωση των αδόμητων πρωτεϊνών. (Yoon et al 2008)

# Βιβλιογραφία

- Alder B. J., Wainwright T. E., 1957. Phase Transition for a Hard Sphere System. *J. Chem. Phys.* 27 (5): 1208
- Alder B. J., Wainwright T. E. 1959. Studies in molecular dynamics. I. General method. *J. Chem. Phys.* 31: 459
- Allen M. P. 2004. Introduction to Molecular Dynamics Simulation. NIC Series, 23: 1-28
- Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science* 181:223– 30
- Allinger NL, Burkert U. 1982. *Molecular Mechanics*. An American Chemical Society Publication
- Altis A, Nguyen PH, Hegger R, Stock G. 2007. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* 126
- Amadei A., Linssen A.B.M., Berendsen, H.J.C. 1993. Essential dynamics of proteins. *Proteins* 17: 412–425,
- Bailey, Ken. 1994. Numerical Taxonomy and Cluster Analysis. *Typologies and Taxonomies*: 34
- Becker, Donald J , Sterling, Thomas, Savarese, Daniel, Dorband, John E, Ranawak, Udaya A, Packer, Charles V. 1995. BEOWULF: A parallel workstation for scientific computation. in *Proceedings, International Conference on Parallel Processing* 95
- Berman, H., Westbrook, J., Feng, Z., Gilliland G., Bhat T., Weissig H., Shindyalov I., Bourne P., 2000. The Protein Data Bank. *Nucleic Acids Research* 28(1):235–242
- Blanco FJ, Rivas G, Serrano L .1994. A short linear peptide that folds into a native stable beta- hairpin in aqueous solution. *Nat Struct Biol* 1: 584–590
- Branden C and Tooze J. 1999. Introduction to protein structure
- Brooks CL, Chen J, Im W. 2006. Balancing solvation and intramolecular interactions: toward a consistent generalized born force field (CMAP opt. for GBSW). *J Am Chem Soc* 128 (11): 3728–3736.
- Brylinska M, Konieczny L, Rotermana I. 2006. Hydrophobic collapse in (in silico) protein folding. *Comp Biol and Chem* 30: 255–267
- Carloni P, Rothlisberger U, and Parrinello M. 2002. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Acc. Chem. Res.* 35:455–464.
- Chen Y, Ding F, Nie H, Serohijos A, Sharma S, Wilcox KS, Yin S, Dokholyan N. Y. 2008. Protein Folding: Then and Now. *Arch Biochem Biophys.* 469(1): 4–19.



- Compiani M, Capriotti E. 2013. Computational and Theoretical Methods for Protein Folding. *Biochemistry* 52(48): 8601–8624
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 117: 5179–5197
- Dill K, Chan H.S. 1997. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* 4: 10–19
- Dill K. A, Ozkan S. B, Shell M. S, and Weikl T. R. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:289–316
- Dinner A.R, Sali A, Smith L. J, Dobson C. B, and Karplus M. 2002. Understanding protein folding via free-energy surfaces from theory and experiment. *TIBS* 25: 331-339
- Dossey AT, Reale V., Chatwin H., Zachariah C., deBono M., Evans PD., Edison AS. 2006. NMR Analysis of *C. elegans* FLP-18 Neuropeptides: Implications for NPR-1 Activation. *Biochemistry.* 45(24): 7586-7597
- Fersht A.R. 1997. Nucleation mechanisms in protein folding. *Curr opin struct biol.* 7:3–9.
- Floudas, C., Fung, H., McAllister, S., Moennigmann, M., Rajgaria, R. 2006. Advances in protein structure prediction and de novo protein design: A review. *Chem Eng. Sci.* 61 (3), 966
- Forster M J. 2001. Molecular modelling in structural biology. *Micron* 33: 365-384
- Freddolino P. L, Harrison C. B, Yanxin Liu and Schulten K. 2010. Challenges in protein folding simulations: Timescale, representation, and analysis. *Nat Phys* 6(10): 751–758
- Frenkel D. 2001. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science).*
- Garcia AE. 1992. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* 68, 2696
- Gnanakaran, S., Nymeyer H., Portman J., Sanbonmatsu K.Y. , Garcia A.E. 2003. Peptide Folding Simulations. *Current opinion in structural biology* 13(2): 168-174
- Guntert, P., 2004. Automated nmr structure calculation with cyana. *Methods Mol. Biol.* 278, 353.
- Guvench O., MacKerell AD Jr., 2008. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol Biol.* 443: 63-88
- Haile J. M. 1992. *Introduction to Molecular Dynamics Simulation.*
- Haile J. M.. 1997. *Molecular Dynamics Simulation: Elementary Methods.* Wiley-Interscience
- Heinig M, Frishman D. 2004. STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl Acids Res* 32 : 500-502.
- Ho B. K, Dill K.A. 2006. Folding Very Short Peptides Using Molecular Dynamics. *PLoS Comput Biol* 2(4): 27

- Hornak V, Abel R, Oku A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *PROTEINS: Structure, Function, and Bioinformatics* 65: 712–725
- Ichiye T., Karplus M. 1991. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 11: 205–217
- Jolliffe IT. 2002. *Principal component analysis*. Springer.
- Κυρίκου Ι., Κάπου Α., Μαυρομούστακος Θ., Πούλος Κ. 2001. Μοριακά Μοντέλα: Ένα Ανεκτίμητο Εργαλείο στον Ορθολογιστικό Σχεδιασμό Φαρμάκων. *ΦΑΡΜΑΚΕΥΤΙΚΗ* 14 (3): 109-123
- Karplus M, Weaver DL. 1979. Diffusion–collision model for protein folding. *Biopolymers*. 18:1421-37.
- Karplus M, Weaver DL. 1994. The Protein Society Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Science* 3:650- 68.
- Karplus M. 2011. Behind the folding funnel diagram. *Nature Chemical Biology* 7: 401–404
- Kolinski, A., 2004. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* 51 (2), 349–371.
- Koukos P I, Glykos NM. 2013. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *J Comp Chem.* 34 (26): 2310-2312
- Lapidus L., Yao S., McGarrity K.S., Hertzog D.E., Tubman E., Bakajin O. 2007. Protein Hydrophobic Collapse and Early Folding Steps Observed in a Microfluidic Mixer. *Biophysical Journal* 93: 218–224
- Leach A. R. 2001. *Molecular Modelling: Principles and Applications*
- Levinthal C. 1968. Are there pathways for protein folding?. *J. Chim. Phys.* 65:44-45.
- Li C, Nelson LS, Kim K, Nathoo A, Hart AC. 1999. Neuropeptide gene families in the nematode *Caenorhabditis elegans*. *Ann N Y Acad Sci*, 897:239-52.
- Li C, Kim K. 2008. Neuropeptides. *WormBook*. 1-36
- Li, Kim K. 2014. Family of FLP peptides in *Caenorhabditis elegans* and related nematodes. *Front. Endocrinol.*,
- MacKerell, Jr. AD, Bashford D., Bellott, 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, Wirkiewicz-Kuczera J, Yin D, Karplus M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102 (18): 3586–3616.
- Mark P, Nilsson L. 2001. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298

- K. J. Phys. Chem. A 105, 9954-9960
- Marqusee S, Robbins VH, Baldwin R. L. 1989. Unusually stable helix formation in short alanine-based peptides. Proc Natl Acad Sci U S A 86: 5286–5290.
- McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. Nature 267 (5612): 585–59
- Mourad Sadqi, Lisa J. Lapidus, Victor Muñoz. 2003. How fast is protein hydrophobic collapse?. PNAS 100(21): 12117–12122
- Mu Y, Nguyen PH, Stock G. 2005. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. Proteins. 58( 1): 45-52.
- Neidigh J.W., Fesinmeyer R.M., Andersen N.H. 2002. Designing a 20-residue protein. Nat struct biol 9: 425-430
- Nelson M, Humphrey W, Gursoy A, Dalke A, Kale L, Skeel RD, Schulten K. 1996. NAMD {a Parallel, Object-Oriented Molecular Dynamics Program.
- Nolting B, Agard D. A. 2008. How General Is The Nucleation–Condensation Mechanism?. Proteins 73(3): 754-64
- Οικονόμου Ε. Ν. 2004. Η φυσική σήμερα.
- Osguthorpe D., 2000. Ab initio protein folding. Curr. Opin. Struct. Biol. 10 (2), 146.
- Papaioannou S., Marsden D., Franks CJ., Walker RJ., Holden-Dye L. 2005. Role of a FMRFamide-like Family of Neuropeptides in the Pharyngeal Nervous System of *Caenorhabditis elegans*. J Neurobiol. 65(3):304-19
- Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005. Scalable Molecular Dynamics with NAMD. J Comput Chem. 26(16): 1781–1802.
- Phillips J. 2012. NAMD Tutorial. Computational Biophysics Workshop
- Ponder JW, Case DA. 2003. *Force fields for protein simulations. Ad. Prot. Chem, 6: 27-85*
- Rahman A. 1964. Correlations in the Motion of Atoms in Liquid Argon. *Physical Review* 136 (2A): A405–A411
- Rohl, C., Strauss, C., Misura, K., Baker, D., 2004. Protein structure prediction using rosetta. Methods Enzymol. 383 (2): 66
- Salomon-Ferrer R, Case DA, Walker RC. 2012. An overview of the Amber biomolecular simulation package. WIREs Comput Mol Sci 0: 1-13
- Schlick, Tamar. 2002. Molecular Modeling and Simulation: An Interdisciplinary Guide.
- Shao J, Tanner SW, Thompson N, Cheatham TE. 2007. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. J. Chem. Theory Comput.

3: 2312-2334

- Shen Y, Bax A. 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR*, 48: 13-22
- Shirts M.R., Pande V.S. 2001. Mathematical analysis of coupled parallel simulations. *Phys rev lett* 86: 4983-4987
- Sporlein S., Carstens H., Satzger H., Renner C., Behrendt R., Moroder L., Tavan P., Zinth W., Wachtveitl J., 2002. Ultrafast spectroscopy reveals subnanosecond peptide conformation dynamics and validates molecular dynamics simulation. *Proc natl acad sci usa* 99:7998-8002
- Stillinger F. H., Rahman A. 1974. Improved Simulation of Liquid Water by Molecular Dynamics. *J. Chem. Phys.* 60: 1545-1557
- Theobald D.L. 2005. Rapid calculation of RMSDs using a quaternionbased characteristic polynomial. *Acta Cryst.* 61: 478–480
- Van der Spoel D. Van Maaren PJ, Berendsenb H JC. 1998. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *J. Chem. Phys.*, 108 (24): 10220-10230
- Van Gunsteren W. F., Berendsen H. J. C. 1990. *Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry*. *Angewandte Chemie International Edition in English*, 29: 992-1023
- van Gunsteren WF., Bakowies D., Baron R., Chandrasekhar I., Christen M., Daura X., Gee P., Geerke DP., Glattli A., Hunenberger P. H., Kastenholtz M. A., Oostenbrink C., Schenk M., Trzesniak D., van der Vegt NFA., Yu HB. 2007. Biomolecular modelling: goals, problems, perspectives. *Angew. Chem. Int. Ed.* 45:4064–4092
- Wolf A, Kirschner KN. 2013. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11·23S subdomain. *J Mol Model.* 19(2): 539-549.
- Yoon J, Park J, Jang S, Lee K, Shin S. 2008. Conformational Characteristics of Unstructured Peptides:  $\alpha$ -Synuclein. *J Biomol Struct Dyn.* 25(5) :505-515.
- Zwanzig R, Szabo A, Bagchi B. 1992. Levinthal's paradox. *Proc Natl Acad Sci.* 89(1): 20–22.