



**ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ**  
**Τμήμα Μοριακής Βιολογίας και Γενετικής**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Προσομοιώσεις Μοριακής Δυναμικής  
του πεπτιδίου αLa101-111 της α-λακταλβουμίνης**

Καλλιόπη Κ. Παταπάτη  
ΑΕΜ 772

Επίβλεψη  
Νικόλαος Μ. Γλυκός

Αλεξανδρούπολη  
Σεπτέμβριος 2010

The need for accuracy must be weighed against  
*the need for finality...*

*Σχόλιο του Ανώτατου Δικαστηρίου των ΗΠΑ, πριν την ανακοίνωση της τελικής του απόφασης  
περί του αποτελέσματος της εκλογικής αναμέτρησης ανάμεσα στους Bush και Gore,  
ένα μήνα μετά τις Προεδρικές εκλογές του 2000.*

## Περίληψη

Μπορούν τα εμπειρικά δυναμικά πεδία να χρησιμοποιηθούν ως μέσο για την επίλυση του προβλήματος της αναδίπλωσης των πεπτιδίων; Μπορεί ειδικότερα το CHARMM forcefield με την διόρθωση CMAP να αναπαράγει την πειραματικά προσδιορισθείσα δομή της  $3_{10}$ -έλικας του ενδεκαπεπτιδίου INYWLAHAKAG; Σε προσπάθεια απάντησης στα ερωτήματα αυτά, πραγματοποιήσαμε δύο προσομοιώσεις του ενδεκαπεπτιδίου. Η μία αναφέρεται ως '**fold**ing', πραγματοποιήθηκε στους 320K και διήρκεσε 1.3  $\mu$ s ενώ η άλλη ως '**native**', έγινε στους 283K και η διάρκειά της ήταν 0.4  $\mu$ s. Η αρχική δομή της προσομοίωσης '**fold**ing' είναι η πλήρως εκτεταμένη, ενώ της '**native**' είναι το μοντέλο 1 της πειραματικά προσδιορισμένης με NMR δομής (καταχώρηση 2DX2 της PDB, δομή  $3_{10}$ -έλικας). Και οι δύο προσομοιώσεις πραγματοποιήθηκαν με χρήση του προγράμματος NAMD και την βοήθεια του VMD, σε NpT ensemble, χρησιμοποιώντας περιοδικές οριακές συνθήκες, διαλύτη νερό, συγκέντρωση ιόντων ισοδύναμη με 100mM και forcefield το CHARMM με τη διόρθωση CMAP. Η ανάλυση των τροχιακών επιτεύχθηκε με την βοήθεια των προγραμμάτων CARMA και STRIDE. Και οι δύο προσομοιώσεις έδειξαν την αδυναμία του συγκεκριμένου δυναμικού πεδίου για αναπαραγωγή της δομής της  $3_{10}$ -έλικας και την σαφή προτίμησή τους για  $\alpha$ -ελικοειδή δομή. Ωστόσο, οι προσομοιώσεις συνέκλιναν μεταξύ τους και βρέθηκαν σε συμφωνία με τα πειραματικά δεδομένα σε ότι αφορά την σταθερότητα διαμόρφωσης των καταλοίπων 2-6 και την κινητικότητα του C-τελικού άκρου του ενδεκαπεπτιδίου. Από τα δεδομένα μας συμπεραίνουμε ότι το χρησιμοποιηθέν δυναμικό πεδίο με την συγκεκριμένη παραμετροποίηση δεν έχει μεγάλες πιθανότητες να συντελέσει στην επίλυση του προβλήματος της αναδίπλωσης των πεπτιδίων.

## Abstract

Is it possible that the use of empirical forcefield can be the answer to the peptide folding problem? Specifically: is the CHARMM forcefield with CMAP correction capable of reproducing the experimentally determined structure of  $3_{10}$ -helix of the undecamer peptide INYWLAKAKAG? In an attempt to answer these questions, we have performed two simulations, namely '**folding**' and '**native**'. The first one has been performed at 320K for a duration of 1.3  $\mu$ s and the last one, at 283K with a total simulation time of 0.4  $\mu$ s. The initial conformation of the '**folding**' simulation is the fully extended conformation of the peptide, whilst the '**native**' one has started from the model 1 of the experimentally determined by NMR structure (PDB entry: 2DX2, conformation of  $3_{10}$ -helix). Both simulations have been performed using the NAMD program, the assistance of VMD, under NpT conditions, using periodic boundary conditions, ion concentration of 100mM and forcefield the CHARMM with CMAP correction. The trajectory analysis has been accomplished using CARMA and STRIDE. Both simulations have shown the weakness of the specified forcefield for the reproduction of the  $3_{10}$ -helix and its bias towards  $\alpha$ -helices. Nevertheless, our simulations have converged on  $\alpha$ -helices and they have come to an agreement with the experimental data on the stability of the residues 2-6 and on the C-terminal mobility. We conclude that the forcefield under discussion, with the current parameterization will not contribute substantially to the solution of the peptide folding problem.

## Περιεχόμενα

<b>Πρόλογος</b>		v
<b>Κεφάλαιο 1</b>	<b>Περί πρωτεϊνικής αναδίπλωσης</b>	1
1.1	Μοντέλα αναδίπλωσης πρωτεϊνών	2
1.2	Μέθοδοι μελέτης αναδίπλωσης πρωτεϊνών	6
1.3	Πεπτίδιο ή πρωτεΐνη;	7
1.4	Χρήση πεπτιδίων στο protein folding	10
1.5	Περί $3_{10}$ - έλικας	12
1.6	Προέλευση του ενδεκαπεπτιδίου της παρούσας εργασίας	13
<b>Κεφάλαιο 2</b>	<b>Μοριακός μοντελισμός &amp; προσομοιώσεις μοριακής δυναμικής</b>	21
2.1	Υπολογιστικές προσομοιώσεις (computer simulations)	23
2.2	Μοριακή Μηχανική & προσομοιώσεις Μοριακής Δυναμικής	25
2.3	Στατιστική μηχανική (statistical mechanics)	26
2.4	Δυναμικά πεδία (force fields)	29
<b>Κεφάλαιο 3</b>	<b>Προσομοιώσεις των δύο συστημάτων</b>	31
3.1	Απαιτούμενα αρχεία για προσομοίωση μοριακής δυναμικής	32
3.2	Τα στάδια για την προετοιμασία του συστήματος	34
3.3	Προσομοίωση της 'folding'	36
3.3.1	Προετοιμασία του συστήματος της 'folding'	36
3.3.2	Ελαχιστοποίηση της ενέργειας και σταδιακή θέρμανση της 'folding'	37
3.3.3	Παραγωγή δεδομένων της 'folding'	38
3.3.4	Μειώσεις νερών και στοχαστικά φαινόμενα	38
3.3.5	Το τροχιακό της 'folding'	39
3.4	Προσομοίωση της 'native'	39
3.4.1	Προετοιμασία του συστήματος της 'native'	39
3.4.2	Ελαχιστοποίηση της ενέργειας και σταδιακή θέρμανση της 'native'	40
3.4.3	Παραγωγή δεδομένων της 'native'	40
3.4.4	Το τροχιακό της 'native'	40
<b>Κεφάλαιο 4</b>	<b>Ανάλυση προσομοιώσεων</b>	42
4.1	Στιγμιότυπα διαδοχικών δομών των δύο τροχιακών	45
4.2	Ανάλυση του τροχιακού της 'native'	47
4.2.1	Ανάλυση σε κύριες συνιστώσες με χρήση εσωτερικών συντεταγμένων	47
4.2.2	Ανάθεση δευτεροταγούς δομής – κατανομή των cluster	50
4.2.3	Πίνακες RMSD και ανάλυσή τους	51
4.2.4	RMSD από την πειραματικά προσδιορισμένη δομή	53
4.3	Ανάλυση του τροχιακού της 'folding'	56
4.3.1	Ανάλυση σε κύριες συνιστώσες με χρήση εσωτερικών συντεταγμένων	56
4.3.2	Ανάθεση δευτεροταγούς δομής	58
4.3.3	Πίνακες RMSD και ανάλυσή τους	59
4.3.4	RMSD από το πλησιέστερο στη μέση δομή frame της 'folding'	62
4.4	Σύγκριση των δύο προσομοιώσεων	65
4.4.1	RMSD της 'native' από τη 'μέση δομή' της 'folding'	65
4.4.2	RMSD της 'folding' από την πειραματικά προσδιορισμένη δομή	67
4.4.3	Υδροδυναμική ακτίνα του πεπτιδίου στις 'folding' και 'native'	69
4.5	Αλληλεπιδράσεις πλευρικών ομάδων	70
<b>Κεφάλαιο 5</b>	<b>Συζήτηση</b>	74
<b>Βιβλιογραφία</b>		77

## Πρόλογος

Μια ενδιαφέρουσα και σχετικά σπάνια δευτεροταγής δομή ενός πεπτιδίου, δομή  $3_{10}$ -έλικας, ήταν η αφορμή για την παρούσα εργασία. Η δομή αυτή ήταν η Μούσα, που μας ενέπνευσε για να ελέγξουμε τα εμπειρικά δυναμικά πεδία, όπως η μούσα Καλλιόπη ενέπνεε τους ραψωδούς στο να εκφράσουν και να καταφέρουν να μεταφέρουν στο κοινό τους την ουσία και το συναίσθημα των ασμάτων τους. Οι Πυθαγόρειοι φιλόσοφοι θεωρούσαν ότι οι Μούσες δίνουν την πρόσβαση στην γνώση για τον κόσμο και το σύμπαν που μας περιβάλλει. Εμείς – αιώνες μετά- θεωρούμε ότι η μελέτη της δομής αυτής, της δικής μας Μούσας, έχει πιθανότητες να προσφέρει έναν κόκκο γνώσης στο ευρύτερο πρόβλημα της αναδίπλωσης των πρωτεϊνών. Η δική μας αναζήτηση για δομή, δεν έγινε σε κάποιον Όλυμπο, αλλά στην protein data bank (PDB – εφόσον στον τρέχοντα αιώνα, υπάρχουν βάσεις δεδομένων, οι οποίες έχουν καταχωρημένες πειραματικά προσδιορισθείσες δομές μακρομορίων).

Ο αυτοσκοπός ήταν να κάνουμε την διπλωματική μας εργασία. Ως μέσο για την επίτευξη αυτού του σκοπού χρησιμοποιήσαμε την προσέγγιση της προσομοίωσης μοριακής δυναμικής ενός μικρού σε μέγεθος και ενδιαφέροντος συστήματος, με χρήση εμπειρικών δυναμικών πεδίων. Αναζητώντας στην PDB ενδιαφέρουσα για εμάς καταχώρηση και ταυτόχρονα δομή με μικρό αριθμό καταλοίπων (για επίτευξη ικανοποιητικού χρόνου προσομοίωσης στο χρονικό διάστημα ολοκλήρωσης μιας διπλωματικής εργασίας), επιλέξαμε από τρεις finalist, την 2DX2. Αυτή αντιστοιχεί στο ενδεκαπεπτίδιο INYWLAHAKAG, είναι τροποποιημένο τμήμα των καταλοίπων 101-111 της α-λακταλβουμίνης, έχει δε δομή  $3_{10}$ -έλικας. Αυτό το πεπτίδιο χρησιμοποιήσαμε για να ελέγξουμε τα δυναμικά πεδία της επιλογής μας και να απαντήσουμε σε ερωτήματα, για την αποκάλυψη των οποίων θα πρέπει, ο έχων την θέληση, την υπομονή και το κουράγιο, να ανατρέξει στα επόμενα κεφάλαια...

Στην Θεογονία του Ησιόδου, οι Μούσες συχνά παραποιούσαν την αλήθεια. Εμείς όμως, βασιστήκαμε στους (αιτιοκρατικούς) φυσικούς νόμους και όχι σε φιλοσοφικές ιδεολογίες. Και εκ του εκβάντος κρίνοντας, θεωρούμε ότι η Μούσα μας, η 2DX2, συνάδει με την άποψη περί Μουσών, των Πυθαγορίων φιλοσόφων.

# Κεφάλαιο 1<sup>ο</sup>

## Περί πρωτεϊνικής αναδίπλωσης.

Hofstadter's law:  
*It always takes longer than you expect, even when you take into account Hofstadter's law.*

Πρωτεϊνική αναδίπλωση είναι η διεργασία εκείνη κατά την διάρκεια της οποίας μία πολυπεπτιδική αλυσίδα αναδιπλώνεται στην χαρακτηριστική τρισδιάστατη φυσική, βιολογικά ενεργή δομή της. Τα νεοσυντεθέντα στο ριβόσωμα πολυπεπίδια, στους ζωντανούς οργανισμούς, αναδιπλώνονται τελικά στην φυσική μορφή τους, με ή χωρίς βοήθεια άλλων πρωτεϊνών. Κατά την *in vitro* αναδίπλωση η πρωτεΐνη λαμβάνει την λειτουργική της δομή, όταν οι συνθήκες το επιτρέπουν, από ένα τυχαίο σπείραμα (random coil). Το τελευταίο μπορεί να οριστεί ως στατιστική κατανομή όλων των πιθανών διαμορφώσεων του πληθυσμού του μακρομορίου [1].

Από το 1961, ο Christian Anfinsen [2,3] με τα πειράματά του πάνω στην αναδίπλωση της ριβονουκλεάσης έδειξε ότι, τουλάχιστον για μικρές σφαιρικές πρωτεΐνες, η πρωτοταγής δομή είναι αυτή που καθορίζει την τριτοταγή (σε θερμοκρασίες και καταστάσεις του διαλύτη κοντά στο φυσιολογικό για την πρωτεΐνη περιβάλλον) και ότι η φυσική (native) διαμόρφωση της πρωτεΐνης θα πρέπει να αντιστοιχεί σε στερεοδιάταξη ελάχιστης ελεύθερης ενέργειας.

Πρώτος ο Levinthal [4,5] πρότεινε ότι θα πρέπει να υπάρχουν σαφώς καθορισμένα μονοπάτια κατά την διάρκεια της αναδίπλωσης, εφόσον δεν μπορεί να δικαιολογηθεί χρονικά η αναδίπλωση μιας πρωτεΐνης με την αναζήτηση όλων των πιθανών στερεοδιατάξεων μέχρι να βρεθεί το ολικό ενεργειακό ελάχιστό της. Η διαδικασία αναδίπλωσης θα πρέπει να είναι εφικτή τόσο θερμοδυναμικά, όσο και κινητικά (σε κλίμακα τάξης του βιολογικού χρόνου της αναδίπλωσης).

### **1.1. Μοντέλα αναδίπλωσης πρωτεϊνών.**

Αρκετά μοντέλα έχουν προταθεί [6,7,8] -μετά την διατύπωση του ερωτήματος του Levinthal- σχετικά με τον τρόπο αναδίπλωσης των πρωτεϊνών. Κανένα όμως από αυτά δεν μπορεί να εξηγήσει την αναδίπλωση όλων ανεξαιρέτως των πρωτεϊνών. Κάποια μοντέλα εξετάζουν το θέμα από την θερμοδυναμική του πλευρά, προσπαθώντας να προβλέψουν την φυσική δομή μιας πρωτεΐνης από την αλληλουχία της. Άλλα μοντέλα μελετούν την δυναμική της αναδίπλωσης, εξετάζοντας το μηχανισμό αναδίπλωσης αποδιαταγμένων πρωτεϊνών *in vitro* [9,10].

Πολλές μικρές μονομερείς πρωτεΐνες αναδιπλώνονται με απλή κινητική δύο σταδίων (two-state kinetics), με μεγάλη ωστόσο χρονική διακύμανση των ρυθμών αναδίπλωσης (από εκατομμυριοστά του δευτερολέπτου έως και δευτερόλεπτα). Αυτό δείχνει ότι η ύπαρξη ενδιάμεσων (intermediates) δεν είναι προαπαιτούμενη για την γρήγορη και αποτελεσματική αναδίπλωση και ότι είναι πιθανόν να υπάρχουν κινητικές παγίδες, οι οποίες επιβραδύνουν την διαδικασία [11,12].

Ο τρόπος αναδίπλωσης μερικών πρωτεϊνών φαίνεται να ακολουθεί το μοντέλο 'diffusion-collision' [13,14], το οποίο μελετά την αναδίπλωση από την πλευρά της

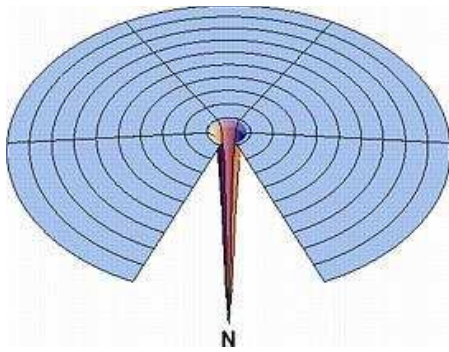


κινητικής. Σύμφωνα με το μοντέλο αυτό, οι πρωτεΐνες θεωρούνται ως αποτελούμενες από πολλές στοιχειώδεις μικροεπικράτειες, κάθε μία εκ των οποίων είναι αρκετά μικρή, ώστε να μπορεί να λάβει σε σύντομο χρονικό διάστημα (σε σύγκριση με την κλίμακα χρόνου που χρειάζεται ολόκληρη η πρωτεΐνη) όλες τις πιθανές διαμορφώσεις [7].

Το μοντέλο ‘hydrophobic collapse’ περιγράφει το αρχικό, πολύ γρήγορο στάδιο της αναδίπλωσης σφαιρικών πρωτεϊνών [15,16], οφείλεται σε υδρόφοβες αλληλεπιδράσεις και έχει ως αποτέλεσμα τον σχηματισμό της εύπλαστης σφαίρας (molten globule state) [17]. Η εύπλαστη σφαίρα μπορεί να θεωρηθεί ως σύνολο παραπλήσιων δομών, οι οποίες μετατρέπονται ταχύτατα η μία στην άλλη, έχει δε σχηματιστεί το μεγαλύτερο μέρος της δευτεροταγούς δομής, χωρίς όμως να έχουν εκδηλωθεί όλες οι αλληλεπιδράσεις της φυσικής κατάστασης.

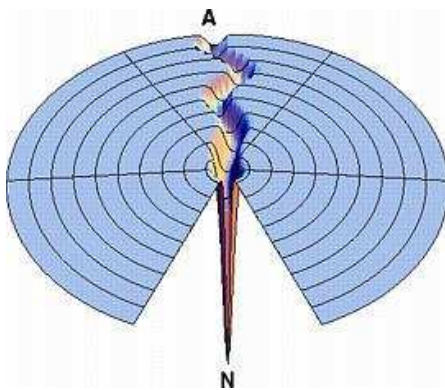
Σύμφωνα με ένα άλλο μοντέλο, αυτό του μηχανισμού εμπυρήνωσης [18] (nucleation-like pathway) αρχικά, κατά την διαδικασία αναδίπλωσης, δημιουργείται ένα σύμπλεγμα (cluster), το οποίο περιλαμβάνει κατάλοιπα σε φυσική διαμόρφωση. Το σύμπλεγμα αυτό στην συνέχεια αυξάνεται, καθώς προστίθενται και τα κατάλοιπα από τα μη αναδιπλωμένα τμήματα της πρωτεΐνης.

Μία νέα θεωρία για την κινητική της πρωτεϊνικής αναδίπλωσης έχει αντικαταστήσει τα τελευταία χρόνια την παλιά ιδέα των μονοπατιών αναδίπλωσης, εισάγοντας τους όρους των τοπίων ενέργειας (energy landscapes) και χωνιών αναδίπλωσης (folding funnels). Η νέα αυτή άποψη δημιουργήθηκε με την βοήθεια της στατιστικής μηχανικής, δεδομένης της προβάσιμης σε ατομικό επίπεδο πληροφορίας περί της δομής της πρωτεΐνης. Χρησιμοποιώντας την μεταφορά των τοπίων ενέργειας, η πρωτεϊνική αναδίπλωση παρομοιάζεται με την ροή νερού από την κορυφή ενός βουνού με ανώμαλες πλαγιές και όχι με την ροή του νερού σε ένα μόνο αυλάκι. Η θεωρία αυτή, δίνει έμφαση στην ομαδοποίηση (ensemble) των διαμορφώσεων των ενδιάμεσων της αναδίπλωσης με την ταυτόχρονη παρουσία πολλαπλών μονοπατιών αναδίπλωσης [19]. Όλα τα μοντέλα που είχαν προταθεί στο παρελθόν μπορούν να σχηματοποιηθούν βάσει της τρέχουσας άποψης των τοπίων ενέργειας. Επιπλέον είναι δυνατή και η απεικόνιση του μοντέλου αναδίπλωσης μεγάλων πρωτεϊνών, οι οποίες δεν υπάκουαν σε κάποιο από τα προηγούμενα μοντέλα. Οι εικόνες 1.1-1.6 [20,21] απεικονίζουν τις διαφορετικές όψεις της θεωρίας του τοπίων ενέργειας.



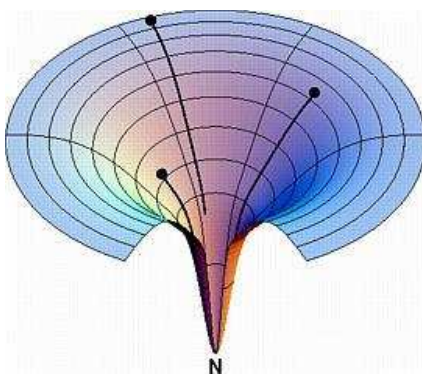
**Εικόνα 1.1:** Golf-course

Το τοπίο ‘γήπεδο του golf’ του Levinthal, όπου **N** η φυσική (native) διαμόρφωση. Η πολυπεπτιδική αλυσίδα αναζητεί τυχαία την φυσική κατάσταση σε ολόκληρη την επιφάνεια του γηπέδου. Δύσκολη η πρόβλεψη της διάρκειας της αναδίπλωσης.



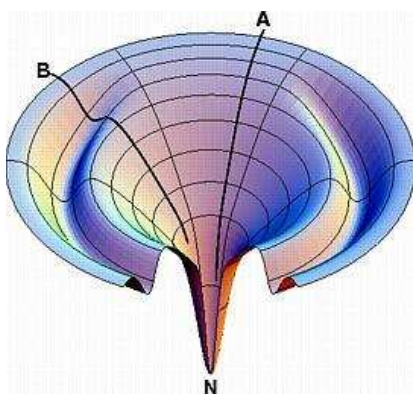
**Εικόνα 1.2:** Grooved golf course

Απεικόνιση προτεινόμενης λύσης του προβλήματος της Εικόνας 1. Αυτή της πιθανής ύπαρξης μονοπατιού στην τυχαία αναζήτηση της φυσικής κατάστασης. Το μονοπάτι οδηγεί την δομή από μια τυχαία μη φυσική διαμόρφωση στην φυσική, με συνέπεια τον περιορισμό της αναζήτησης και αποτέλεσμα την συντομότερη αναδίπλωση.



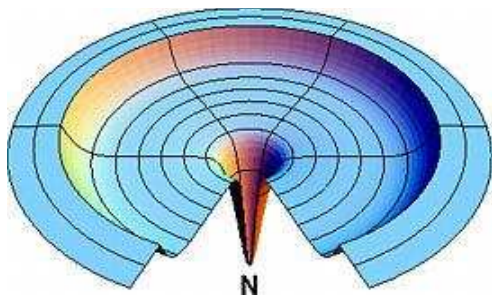
**Εικόνα 1.3:** Smooth funnel

Εικόνα ενός ιδανικού τοπίου χωνιού. Η ομαλή μείωση της ελεύθερης ενέργειας του μορίου μειώνει τον αριθμό των πιθανών διαμορφώσεων.



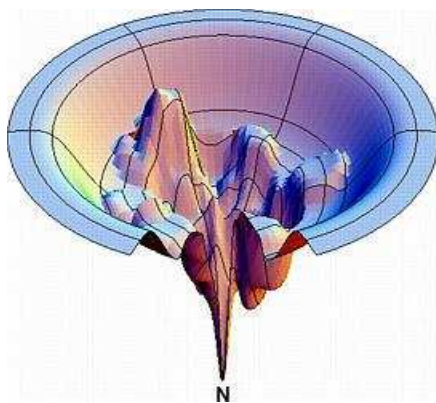
**Εικόνα 1.4:** Moat Landscape

Απεικόνιση του τρόπου με τον οποίο μία πρωτεΐνη μπορεί να αναδιπλώνεται γρήγορα (fast-folding) μέσω του μονοπατιού **A** και μιας πιο αργής (slow-folding) διεργασίας **B**, η οποία περιλαμβάνει μία κινητική παγίδα-τάφρο.



**Εικόνα 1.5:** Champagne Glass Landscape

Στην εικόνα αυτή φαίνεται πώς η εντροπία της διαμόρφωσης μπορεί να δημιουργεί φράγμα στην περαιτέρω μείωση της ελεύθερης ενέργειας. Με τον τρόπο αυτό η αλυσίδα παγιδεύεται στο επίπεδο χωρίς να μπορεί να φτάσει γρήγορα στην φυσική της μορφή.



**Εικόνα 1.6:** Bumpy bowl

Ένα ανώμαλο τοπίο ενέργειας με πολλές κινητικές παγίδες, ενεργειακούς φραγμούς και μερικά στενά μονοπάτια προς την διαδρομή για την φυσική N κατάσταση. Η αναδίπλωση μπορεί να ακολουθεί πολλαπλά μονοπάτια.

Αξίζει να σημειωθεί ότι, ακόμα και με την θεωρία περί τοπίων ενέργειας και παρά όλες τις θεωρητικές και πειραματικές προσπάθειες, δεν μπορεί να δοθεί λεπτομερής περιγραφή του μηχανισμού με τον οποίο αναδιπλώνονται όλες οι πρωτεΐνες. Θα μπορούσε κάποιος να πει ότι το πρόβλημα είναι διττό και ότι αφορά, τόσο την πρόβλεψη της τρισδιάστατης δομής μιας πρωτεΐνης δεδομένης της αλληλουχίας της, όσο και την κινητική και την δυναμική της διαδικασίας αναδίπλωσης. Το πρώτο τμήμα του προβλήματος λύνεται κατά κάποιον τρόπο με έμμεσες μεθόδους, όπως το threading και το homology modeling, η επίλυση όμως του δεύτερου σκέλους είναι αυτή που θα έλυνε οριστικά το πρόβλημα [19].

## 1.2. Μέθοδοι μελέτης αναδίπλωσης πρωτεϊνών.

Στην προσπάθεια για την κατανόηση της διαδικασίας αναδίπλωσης των πρωτεϊνών έχει μελετηθεί μεγάλος αριθμός πρωτεϊνών με την χρήση πολλών διαφορετικών μεθόδων (μετά την πρώτη χρονολογικά πειραματική μέθοδο του Anfinsen). Αναφορικά, κάποιες από τις μεθόδους αυτές είναι: Η ηλεκτροφόρηση (urea-gradient-gel-electrophoresis), η οποία χρησιμοποιείται για την μελέτη των μεταβολών της διαμόρφωσης των πρωτεϊνών σε συνθήκες αποδιάταξης [22,23,24], η χρήση της μεθόδου του κυκλικού διχρωϊσμού (Circular Dichroism) για την διερεύνηση της δευτεροταγούς δομής πρωτεϊνών [25], η διαδικασία του equilibrium unfolding με την οποία καθορίζεται η σταθερότητα διαμόρφωσης του μορίου με αλλαγή των συνθηκών του περιβάλλοντός του [26], η μέθοδος differential scanning calorimetry-DSC [27], η οποία μελετά τις μεταβολές των θερμοδυναμικών παραμέτρων της πρωτεΐνης κατά τη μετουσίωσή της λόγω θερμότητας [28].

Σημαντική λεπτομερειακή γνώση της τρισδιάστατης δομής των βιολογικών μακρομορίων σε ατομικό επίπεδο έχει ληφθεί με τα δεδομένα περίθλασης ακτίνων Χ από κρυστάλλους πρωτεϊνών καθώς και με την μέθοδο του πυρηνικού μαγνητικού συντονισμού (nuclear magnetic resonance, NMR), η οποία εφαρμόζεται σε διαλύματα πρωτεϊνών. Η μέθοδος NMR (εκμεταλλευόμενη την ιδιότητα του spin των πυρήνων των ατόμων), δεν παράγει είδωλο της δομής της πρωτεΐνης, αλλά δίνει άφθονη δομική πληροφορία (όπως περιορισμούς αποστάσεων ανάμεσα στα άτομα) από την οποία τελικά καθορίζεται η τρισδιάστατη δομή. Περιληπτικά τα βήματα καθορισμού δομής με NMR είναι: προετοιμασία του διαλύματος της πρωτεΐνης, μετρήσεις NMR, ανάθεση των σημάτων NMR σε μεμονωμένα άτομα, συλλογή των περιορισμών διαμόρφωσης (conformational constraints), όπως τις αποστάσεις μεταξύ ατόμων υδρογόνου και τέλος, καθορισμός της τρισδιάστατης δομής βάσει των πειραματικά προσδιοριζόμενων περιορισμών διαμόρφωσης. Η μέθοδος αυτή χρησιμοποιείται ευρύτατα τα τελευταία 25 χρόνια για προσδιορισμούς δομών βιολογικών μακρομορίων [29]. Ωστόσο είναι δύσκολο να προσδιοριστεί δομή με μοριακό βάρος μεγαλύτερο των 30-35 kDa, γεγονός που οφείλεται σε δύο λόγους. Ο ένας λόγος είναι ότι στα μεγαλύτερα μόρια το σήμα δεν υποχωρεί απότομα και έχει μικρότερο ενδιάμεσο χρόνο ηρεμίας, με αποτέλεσμα τα μεγάλου πλάτους σήματα, που δημιουργούνται να αλληλεπικαλύπτονται και να

δημιουργείται έτσι πρόβλημα στην ανάθεση του κάθε σήματος στο σωστό άτομο. Ο δεύτερος οφείλεται στο ότι κάθε ενεργός πυρήνας NMR δίνει εξατομικευμένο σήμα (resonance line), το οποίο πρέπει στην συνέχεια να αναλυθεί με πολυδιάστατες τεχνικές NMR., πράγμα που γίνεται πιο δύσκολο για μεγαλύτερες μοριακές δομές. Οι δύο αυτοί τεχνικοί φραγμοί θέτουν, πρακτικά, όριο στο μέγεθος των μορίων, που μπορούν να μελετηθούν με τη μέθοδο αυτή [30]. Σήμερα, ωστόσο, το όριο των 35 kDa για τον προσδιορισμό δομών πρωτεϊνών έχει αυξηθεί σημαντικά [31,32].

Τέλος, με την χρήση υπολογιστικών μεθόδων προσπαθούμε να απαντήσουμε σε μεγάλο εύρος σύνθετων ερωτημάτων ανάλογων της πολυπλοκότητας των ίδιων των βιολογικών συστημάτων. Οι υπολογιστικές προσεγγίσεις χρειάζονται για να γεμίσουν τα κενά των πειραματικά προσδιοριζόμενων με τους ανωτέρω αναφερόμενους τρόπους δομών και για να οικοδομήσουν καλύτερα μοντέλα και θεωρίες, οι οποίες να μπορούν να κάνουν ελέγξιμη πρόβλεψη της αναδίπλωσης μακρομορίων. Ο δικαιολογημένος ενθουσιασμός για την βιολογία *in silico* οφείλεται τόσο στη βελτίωση των πειραματικών τεχνικών οι οποίες επιτρέπουν να αναλυθούν υπολογιστικά πολυσύνθετες δομές, όσο και στην ταυτόχρονη ανάπτυξη νέων μοντέλων και αλγορίθμων για προσομοιώσεις μορίων. Επιπλέον η αυξανόμενη ταχύτητα των υπολογιστών, οι παράλληλοι επεξεργαστές, αλλά και η συνεργασία μεταξύ προγραμματιστών και μηχανικών υπολογιστών με βιολόγους, φυσικούς και μαθηματικούς καθιστούν την συνεισφορά των υπολογιστικών μεθόδων ιδιαίτερα ελπιδοφόρο προσέγγιση στην προσπάθεια επίλυσης των άλυτων μέχρι σήμερα προβλημάτων της βιολογίας. Σε αυτό συνεισφέρει και η ελεύθερη διακίνηση της πληροφορίας μέσω διαδικτύου και η ύπαρξη πολλών ελεύθερα προσβάσιμων βάσεων δεδομένων και αυτοματοποιημένων εργαλείων ανάλυσης [33].

### **1.3. Πεπτίδιο ή πρωτεΐνη;**

Όσο μικρότερη είναι μια πρωτεΐνη τόσο μικρότερη είναι η πολυπλοκότητά της και τόσο λιγότερο δύσκολη η μελέτη της. Μία πρωτεΐνη αποτελούμενη από μικρότερο αριθμό αμινοξικών καταλοίπων έχει (σχεδόν πάντα) μικρότερο αριθμό ομοιοπολικών δεσμών και μικρότερο αριθμό ατόμων στο μόριό της και συνεπώς μικρότερο αριθμό

φυσιολογικών τρόπων ταλάντωσης (normal modes, βασικών μοριακών κινήσεων). Ο αριθμός των normal modes για πρωτεΐνη N ατόμων είναι  $3N-6$ : 3 βαθμοί ελευθερίας ανά άτομο μείον 6 (3 translational και 3 rotational) βαθμούς ελευθερίας του μορίου συνολικά. Η ταλάντευση ομοιοπολικά συνδεδεμένων ατόμων εξαρτάται από το είδος και τον αριθμό των ατόμων που συμμετέχουν και μπορεί να είναι μορφής έκτασης (stretching), κάμψης (bending) ή στρέψης (torsional) αλλά και συνδυασμός των τριών. Είναι λοιπόν προφανές ότι μικρότερο μόριο θα έχει μικρότερο αριθμό τέτοιων δονήσεων. Αλλά μία μικρότερη πρωτεΐνη θα έχει και μικρότερο αριθμό ατόμων που θα μπορούν να αλληλεπιδρούν μη ομοιοπολικά υπό κατάλληλες συνθήκες με αλληλεπιδράσεις Van der Waals, ηλεκτροστατικές, ακόμα και με δεσμούς υδρογόνου. Συμπερασματικά είναι απλούστερη η εμπειριστατωμένη μελέτη μικρών πεπτιδίων από αυτή των μεγαλύτερων πρωτεϊνών.

Πότε όμως μπορούμε να χρησιμοποιούμε τον όρο πεπτίδιο και πότε αυτόν της πρωτεΐνης; Σε μία απλή αναζήτηση στο διαδίκτυο βρίσκουμε:

- ‘Πεπτίδιο: Μόριο που αποτελείται από δύο ή περισσότερα αμινοξέα. Τα πεπτίδια είναι μικρότερα από τις πρωτεΐνες, που είναι επίσης αλυσίδες αμινοξέων. Μόρια αρκετά μικρά που μπορούν να συντεθούν χημικά από τα αμινοξέα τους ονομάζονται κατά σύμβαση πεπτίδια και όχι πρωτεΐνες. Η διαχωριστική γραμμή βρίσκεται περίπου στα 50 αμινοξέα’ [34].
- ‘Πεπτίδια με αλυσίδες αποτελούμενες από περισσότερες από λίγες δωδεκάδες (!) αμινοξέων ονομάζονται πρωτεΐνες’ [35].
- ‘Πολυπεπτίδιο: ένα πεπτίδιο που περιέχει 10 έως περισσότερα από 100 αμινοξέα. Τα πεπτίδια είναι μικρού μήκους αλυσίδες από ενωμένα μεταξύ τους αμινοξέα. Αν υπάρχουν μόνο δύο αμινοξέα το πεπτίδιο ονομάζεται διπεπτίδιο. Ομοίως υπάρχουν τριπεπτίδια, τετραπεπτίδια κλπ. Αν ο αριθμός των αμινοξέων της αλυσίδας φτάνει τον αριθμό δέκα ή κάπου εκεί (!) η αλυσίδα θεωρείται πολυπεπτίδιο, ενώ τα μεγαλύτερα πολυπεπτίδια ονομάζονται πρωτεΐνες. Δεν υπάρχει καθορισμένο μέγεθος στο οποίο ένα μεγάλο πολυπεπτίδιο γίνεται μία μικρή πρωτεΐνη, αλλά γενικά τα πολυπεπτίδια έχουν μοριακό βάρος λίγων χιλιάδων ενώ οι πρωτεΐνες δεκάδων χιλιάδων. Ανάλογα με το είδος του

εμπλεκόμενου αμινοξέος, επτά με δέκα κατάλοιπα προσθέτουν περίπου 1000 στο μοριακό βάρος' [36].

- 'Ένα πεπτίδιο είναι μία μικρού μήκους πολυμερής αλυσίδα, η οποία αποτελείται από ενωμένα μεταξύ τους μόρια αμινοξέων' [37].
- 'Οι πρωτεΐνες είναι πολυπεπτιδικά μόρια, τα οποία αποτελούνται από πολλαπλές πολυπεπτιδικές υπομονάδες καθεμία από τις οποίες περιέχει ειδική ακολουθία από τα 22 αμινοξέα, που παράγουν πρωτεΐνες (proteinogenic amino acids) [38], τα οποία κωδικοποιούνται από τον γενετικό κώδικα. Ο διαχωρισμός είναι ότι τα πεπτίδια είναι μικρού μήκους, ενώ τα πολυπεπτίδια μεγάλου. Υπάρχουν κάποιες συμβάσεις για τον διαχωρισμό των δύο όρων, που όμως περιέχουν πολλές ασάφειες. Μία σύμβαση είναι ότι αυτά, των οποίων η πολυπεπτιδική αλυσίδα είναι αρκετά μικρή ώστε να μπορεί να συντεθεί χημικά από τα αμινοξέα της, ονομάζονται πεπτίδια και όχι πρωτεΐνες. Ωστόσο με την ανάπτυξη της τεχνολογίας σήμερα μπορούν να συντεθούν πεπτίδια αποτελούμενα από εκατοντάδες αμινοξέα, ακόμη και ολόκληρες πρωτεΐνες όπως η ουβικιτίνη. Με την χρήση δε της τεχνικής για δημιουργία μεγάλων πεπτιδίων από δύο ή περισσότερα πεπτίδια με μη προστατευμένα άκρα (native chemical ligation), δεν υπάρχει θεωρητικά περιορισμός στο μέγεθος του συντιθέμενου πολυπεπτιδίου, οπότε η σύμβαση αυτή δεν μπορεί να έχει πλέον εφαρμογή. Μία άλλη σύμβαση θέτει ανεπίσημα το όριο των περίπου 50 καταλοίπων, ενώ κατά άλλους ο αριθμός αυτός είναι ακόμα μικρότερος. Ο διαχωρισμός αυτός είναι κατά κάποιον τρόπο αυθαίρετος. Μεγάλα πεπτίδια όπως τα β-αμυλοειδή μπορούν να θεωρηθούν πρωτεΐνες και μικρές πρωτεΐνες όπως η ινσουλίνη μπορούν να θεωρηθούν πεπτίδια' [39].

Μετά λοιπόν από αυτό το συνονθύλευμα ασαφών ορισμών κάναμε μια αναζήτηση στην PDB [40] για να δούμε εκεί πως χρησιμοποιούν τους όρους 'πεπτίδιο' και 'πρωτεΐνη'. Στην απλή αναζήτηση των δομών της PDB με τον όρο 'peptide' δίνει 64229 δομές, με τον όρο 'polypeptide' 62071 και με τον όρο 'protein' 62024 από τις 64229 καταχωρημένες δομές [41], ενώ στην αναζήτηση για 'not peptide' παίρνουμε 2087 συνολικά καταχωρήσεις, που περιέχουν κυρίως νουκλεϊκά οξέα αλλά και πολυσακχαρίτες. Στην απλή επίσης αναζήτηση για 'peptide and not polypeptide' η PDB μας δίνει 71 δομές με τα αμινοξέα, που περιλαμβάνονται σε αυτές να είναι

τροποποιημένα και όχι στην μορφή που μεταφράζονται σύμφωνα με τον γενετικό κώδικα. Στην advanced αναζήτηση για ‘protein and not peptide’ δίνει 1251 δομές, ενώ για ‘peptide and not protein’ 1369 και για ‘protein and not polypeptide’ 1305 δομές που περιλαμβάνουν κυρίως γλυκοπεπτίδια. Τέλος ακόμα και στην ασαφή αναζήτηση για ‘protein and polypeptide and not peptide’ δίνει τελικά 1412 πρωτεϊνικές δομές σε σύμπλοκα κυρίως με νουκλεϊκά οξέα.

Φαίνεται λοιπόν ότι ούτε στην PDB υπάρχει σαφής διαχωρισμός ανάμεσα στους όρους πεπτίδιο, πολυπεπτίδιο και πρωτεΐνη. Συμπεραίνουμε ότι οι όροι αυτοί είναι απλά keywords και υπάρχουν στις περισσότερες πρωτεϊνικής φύσεως καταχωρήσεις. Φαίνεται επίσης, ότι δεν έχει σε τελική ανάλυση σημασία πως ακριβώς αποκαλεί κανείς το μόριο, αρκεί να αναφέρει τον αριθμό των καταλοίπων, όταν αυτό απαιτείται. Ο όρος πρωτεΐνη χρησιμοποιείται κύρια για μεγάλου μήκους αλυσίδες και σχεδόν πάντα για τις φυσικά υπάρχουσες πρωτεΐνες (για την τιτίνη των 34350 καταλοίπων δεν θα χρησιμοποιήσει κανείς τον όρο πεπτίδιο, αλλά για την βαζοπρεσίνη των 9 είναι δόκιμος και αυτός ο όρος). Όταν κάποιος ασχολείται με τα επίπεδα μιας φυσικά ευρισκόμενης στο οργανισμό πρωτεΐνης ως μέσου διάγνωσης ασθενειών (ενδοκρινολογία - μικροβιολογία) χρησιμοποιεί τον όρο πρωτεΐνη (εναλλακτικά πεπτιδική ορμόνη). Όταν όμως δίδεται ιδιαίτερη έμφαση στην πρωτεΐνη αυτή καθαυτή (πχ για ανάλυση δομής) και μέχρι κάποιου ‘λογικού’ αριθμού καταλοίπων προτιμάται ο όρος πεπτίδιο ή πολυπεπτίδιο. Τον όρο πεπτίδιο χρησιμοποιούμε και στην παρούσα εργασία για το ενδεκαπεπτίδιο μας αλληλουχίας INYWLAHAKAG.

#### **1.4 Χρήση πεπτιδίων στο protein folding.**

Από την στιγμή, που αναγνωρίστηκε ότι οι πρωτεΐνες στην φυσική τους διαμόρφωση αποτελούνται κύρια από στοιχεία δευτεροταγούς δομής, οι θεωρίες για την αναδίπλωση των πρωτεϊνών άρχισαν να χρησιμοποιούν αυτά τα στοιχεία ως σημαντικές δομικές μονάδες προς μελέτη. Τα πεπτίδια έχουν διαδραματίσει κεντρικό ρόλο στην διασαφήνιση των παραγόντων, οι οποίοι σταθεροποιούν τα μεμονωμένα στοιχεία δευτεροταγούς δομής και πλέον χρησιμοποιούνται για την μελέτη υψηλότερων βαθμών οργάνωσης. Ο έλεγχος της διαμόρφωσης των πεπτιδίων έχει γίνει σημαντικό πεδίο



μελέτης μετά την αναγνώριση του κεντρικού ρόλου της λανθασμένης αναδίπλωσης των πρωτεϊνών (misfolding) σε πολλές ασθένειες [42].

Ο λόγος για τον οποίο προτιμώνται τα πεπτίδια –έναντι των μεγαλύτερων πρωτεϊνών- στο ζήτημα της αναδίπλωσης είναι προφανής (επιπλέον έχει αναφερθεί στην προηγούμενη παράγραφο). Η προέλευση των πεπτιδίων και ο τρόπος με τον οποίο χρησιμοποιούνται αυτά ποικίλει. Όσο αφορά την προέλευση, χρησιμοποιούνται είτε μοντέλα φυσικών πεπτιδίων [43,44], είτε πεπτίδια τα οποία είναι μικρά τμήματα φυσικά υπαρχόντων πρωτεϊνών (αυτούσια [45] ή τροποποιημένα [46,47]), είτε συνθετικά πεπτίδια [48]. Η δε προσέγγιση με την οποία αυτά εξετάζονται μπορεί να είναι είτε θεωρητική [49,50], είτε πειραματική [46,47,51,52,53,54], είτε υπολογιστική [48,55,56,57,58,59]. Και η πλευρά από την οποία εξετάζεται η κάθε παράμετρος, που πιθανώς συνεισφέρει στον τρόπο αναδίπλωσης, (όπως φαίνεται από το σύνολο των αναφορών της παραγράφου αυτής), ποικίλει.

Η κατανόηση της συμπεριφοράς των πεπτιδίων - ως θραυσμάτων πρωτεϊνών -σε υδατικά διαλύματα έχει χρησιμοποιηθεί για την μελέτη των αρχικών σταδίων της πρωτεϊνικής αναδίπλωσης [60]. Εκτεταμένες θεωρητικές και πειραματικές εργασίες έχουν δημοσιευτεί για τον τρόπο δημιουργίας της δευτεροταγούς δομής σε πεπτίδια και νέες τεχνικές προέκυψαν από τις μελέτες αυτές [44]. Έχει δειχθεί τρόπος αναδίπλωσης διμερών πρωτεϊνών με την χρήση πεπτιδίων [61]. Μεγάλη βαρύτητα έχει δοθεί και στο θέμα της σταθεροποίησης ενδιάμεσων δομών και του ρόλου των υδρόφοβων αλληλεπιδράσεων στον αναδίπλωση των πρωτεϊνών, χρησιμοποιώντας πάντα πεπτίδια [62,63]. Εμπειριστατωμένες μελέτες έχουν γίνει με σκοπό την πρόβλεψη της επίδρασης μεμονωμένων καταλοίπων στην δυναμική αναδίπλωσης [60], τον έλεγχο της συνεισφοράς των γεφυρών άλατος στην σταθερότητα των πρωτεϊνών [64], την παρατήρηση της επίδρασης του διαλύτη και του pH στην διαδικασία αναδίπλωσης [62]. Τα πεπτίδια χρησιμοποιούνται ευρύτατα και για υπολογιστικές προσομοιώσεις μοριακής δυναμικής για την μελέτη ‘προτιμήσεων’ διαμόρφωσης τους [60], για την πρόβλεψη της δομής των πρωτεϊνών με υπολογιστικές μεθόδους [65] και για τον καθορισμό της δομής των ενδιάμεσων διαμορφώσεων της πορείας αναδίπλωσης [66]. Ακόμα και για τον έλεγχο και την βελτίωση των δυναμικών πεδίων, τα οποία χρησιμοποιούνται στις προσομοιώσεις μοριακής δυναμικής, τα πεπτίδια έχουν πρωταγωνιστικό ρόλο [67].

## 1.5. Περί $3_{10}$ έλικας.

Το τρία με τέσσερα τοις εκατό των καταλοίπων σε κρυσταλλικές δομές πρωτεϊνών έχουν δομή  $3_{10}$ -έλικας [68,69]. Σύμφωνα με το ποσοστό αυτό οι  $3_{10}$ -έλικες καταλαμβάνουν την τέταρτη θέση των πλέον συνηθισμένων διαμορφώσεων δευτεροταγούς δομής στις πρωτεΐνες (μετά την  $\alpha$ -έλικα, την  $\beta$ -πτυχωτή δομή και τις στροφές (reverse turns)). Οι περισσότερες  $3_{10}$ -έλικες είναι μικρού μήκους (3-4 καταλοίπων) σε σχέση με τον μέσο όρο των 10 καταλοίπων των  $\alpha$ -ελίκων [68], συναντώνται δε συνήθως στα N και C-τελικά άκρα των  $\alpha$ -ελίκων [70,68]. Το 32% των ελίκων στις κρυσταλλικές δομές έχει βρεθεί να έχουν τύπου  $3_{10}$  υδρογονοδεσμούς στο N-τελικό άκρο τους και 34% στο C-τελικό [70]. Ο μικρός αριθμός των καταλοίπων της  $3_{10}$ -έλικας μπορεί να δικαιολογηθεί μετά την παρατήρηση κατά την οποία ο αριθμός των καταλοίπων ανά στροφή των  $3_{10}$ -ελίκων μειώνεται - αυξανόμενου του μήκους της έλικας - οπότε αυτή γίνεται λεπτότερη και μακρύτερη. Αυτό έχει ως αποτέλεσμα ο ορισμός της έλικας αυτής (όσο αφορά τον αριθμό των καταλοίπων ανά στροφή και τις διέδρες) να μην ισχύει στην νέα αυτή λεπτότερη μορφή της έλικας [71].

Το όνομα της  $3_{10}$ -έλικας οφείλεται στην ύπαρξη τριών καταλοίπων ανά στροφή και των δέκα ατόμων που περικλείονται στον δακτύλιο που δημιουργείται από κάθε δεσμό υδρογόνου (με την ίδια ονοματολογία η  $\alpha$ -έλικα ονομάζεται  $3.6_{13}$ ). Οι  $3_{10}$ -έλικες είναι δεξιόστροφες και το κάθε κατάλοιπο αντιστοιχεί σε στροφή  $120^\circ$  της έλικας. Η ακτίνα της είναι  $1.9 \text{ \AA}$  και η αξονική απόσταση ανά Ca (translation)  $2.0 \text{ \AA}$ . Με τον αριθμό των 3 καταλοίπων ανά στροφή, το βήμα της έλικας είναι  $6 \text{ \AA}$  (Η αντίστοιχη γεωμετρία της  $\alpha$  έλικας είναι: ακτίνα  $2.3 \text{ \AA}$ , μετατόπιση  $1.5 \text{ \AA}$  και βήμα  $5.4 \text{ \AA}$ , εφόσον στην περίπτωση αυτή έχουμε 3.6 κατάλοιπα ανά στροφή). Δεσμοί υδρογόνου δημιουργούνται ανάμεσα στο καρβονυλικό οξυγόνο ενός καταλοίπου και το υδρογόνο της ιμινομάδας του καταλοίπου που βρίσκεται τρία κατάλοιπα μετά, στην πρωτοταγή δομή (στην περίπτωση της  $\alpha$  έλικας έχουμε  $i$  με  $i+4$ ). Για τα περισσότερα κατάλοιπα που συμμετέχουν στην δημιουργία της  $3_{10}$ -έλικας οι κυρίαρχες διέδρες  $\phi$  και  $\psi$  είναι αντίστοιχα  $-49$  και  $-26$  (ενώ της  $\alpha$ -έλικας  $-58$  και  $-47$  αντίστοιχα). Επειδή οι αζιμουθιακές θέσεις των πλευρικών αλυσίδων συμπίπτουν και οι Ca βρίσκονται στις κορυφές του ίδιου τριγώνου (σε προβολή) δημιουργείται διάταξη που δεν είναι ενεργειακά σταθερή, λόγω στερεοχημικών παρεμποδίσεων των ατόμων των πλευρικών

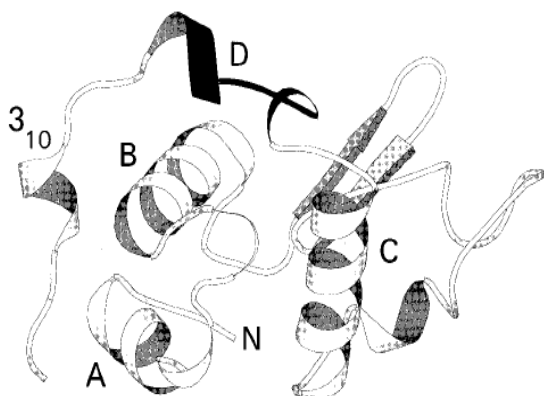
αλυσίδων. Επιπλέον οι δεσμοί υδρογόνου, που δημιουργούνται ανάμεσα στα κατάλοιπα  $i$  και  $i+3$ , δημιουργούν διάταξη, η οποία δεν αντιστοιχεί σε ενεργειακό ελάχιστο, λόγω της μη ευθυγράμμισης των διπόλων των υδρογονοδεσμών (στις  $\alpha$  έλικες η μη ύπαρξη ακεραίων καταλοίπων ελαχιστοποιεί την στερεοχημική παρεμπόδιση και οι δεσμοί υδρογόνου είναι ευθυγραμμισμένοι).

Παρατηρήθηκαν διαφορές στο είδος των καταλοίπων που συμμετέχουν στην δημιουργία της  $3_{10}$ -έλικας (το ασπαρτικό, για παράδειγμα, συναντάται με διπλάσια συχνότητα στο N-τελικό άκρο της  $3_{10}$  από ότι στο ίδιο άκρο της  $\alpha$ ). Εμφανής είναι και η διαφορετική κατανομή των υδρόφοβων καταλοίπων ανάμεσα στις δύο έλικες με αποτέλεσμα να συναντάμε συχνά αμφιπαθείς  $3_{10}$ . Οι τοπικές αλληλεπιδράσεις που σταθεροποιούν την  $3_{10}$ -έλικα οφείλονται τόσο στο είδος των καταλοίπων που απαντώνται συνήθως σε αυτήν, όσο και στην ύπαρξη μικρο-επικρατειών, όπου οι τριτοταγείς αλληλεπιδράσεις αντικαθιστούν τις τοπικές. Έχει επίσης παρατηρηθεί ότι η αναδίπλωση προς  $\alpha$ -έλικα μπορεί να γίνεται μέσω ενδιάμεσων στροφών (reverse-turn) και  $3_{10}$ -ελίκων [72]. Και σε απομονωμένα πεπτίδια η  $3_{10}$  είναι πιθανό ενδιάμεσο της δημιουργίας της  $\alpha$ -έλικας [73], εφόσον δεν υπάρχουν μη επιτρεπόμενες περιοχές ανάμεσα στις δύο αυτές έλικες στο διάγραμμα Ramachadran και ένα πεπτίδιο μπορεί να μετατρέπεται από την μία δομή στην άλλη [74].

## 1.6. Προέλευση του ενδεκαπεπτιδίου της παρούσας εργασίας.

Το πεπτίδιο των ένδεκα καταλοίπων, το οποίο χρησιμοποιήσαμε στις προσομοιώσεις μας, αλληλουχίας INYWLAHAKAG, αντιστοιχεί στο τροποποιημένο τμήμα των καταλοίπων 101-111 της  $\alpha$ -λακταλβουμίνης ( $\alpha$ -LA). Η  $\alpha$ -LA συντίθεται μόνο στους γαλακτοφόρους αδένες των θηλαστικών και μαζί με την  $\beta$ -1,4 γλυκοζυλοτρανφεράση, σχηματίζει την συνθετάση της λακτόζης. Η  $\alpha$ -LA προάγει την σύνδεση της γλυκόζης στο σύμπλοκο, διευκολύνοντας με αυτόν τον τρόπο την βιοσύνθεση της λακτόζης. Το γονίδιο της  $\alpha$ -λακταλβουμίνης είναι υψηλά συντηρημένο στα θηλαστικά. Αυτό της ανθρώπινης  $\alpha$ -λακταλβουμίνης έχει μέγεθος 2575 βάσεων και περιλαμβάνει 5 ιντρόνια και 4 εξόνια και κωδικοποιεί για πρωτεΐνη 123 καταλοίπων [75]. Δομικά, η πρωτεΐνη, στην φυσική της διαμόρφωση αποτελείται από

μία επικράτεια κυρίως ελικοειδή, η οποία δημιουργείται από τα κατάλοιπα 1-39 και 81-123 και μία β-επικράτεια (sheet/coil), η οποία περιλαμβάνει τα υπόλοιπα κατάλοιπα του μορίου [47], όπως φαίνεται στην Εικόνα 1.7.



**Εικόνα 1.7:** Διάγραμμα με MolScript της κρυσταλλικής μορφής της ανθρώπινης α-LA. Η α-επικράτεια είναι λευκή, η β-επικράτεια σκιασμένη και τα κατάλοιπα 101-111 της περιοχής D είναι μαύρα και σχηματίζουν α-έλικα. Μία μικρή  $3_{10}$ -έλικα υπάρχει στο C-τελικό άκρο.

Αναπαράγεται άνευ αδείας από τους Demarest, Hua & Raleigh (1999).

Η περιοχή 101-111 του μορίου έχει αλληλουχία IDYWLAHKALC και προέρχεται από δύο διαφορετικά εξόνια. Από το τρίτο εξόνιο του γονιδίου της α-λακταλβουμίνης προέρχονται τα κατάλοιπα IDYW και από το τέταρτο τα LAHKALC [75]. Το τμήμα αυτό λαμβάνει πολύ διαφορετικές διαμορφώσεις στις κρυσταλλικές δομές υψηλού και χαμηλού pH. Στην κρυσταλλική δομή των 37° C, σε pH 6.5 των Harata και Muraki (που δεν είναι καταχωρημένη στην PDB) τα κατάλοιπα A106 έως C111 λαμβάνουν διαμόρφωση έλικας με την H107 να είναι θαμμένη κάτω από την επιφάνεια του μορίου και να αλληλεπιδρά με την καρβοξυλομάδα του E25. Η δομή αυτή του pH 6.5 των 37°C (310K) θεωρείται η φυσική. Στην κρυσταλλική δομή των 286K σε pH 4.2 (δομή των Harata και Muraki, καταχωρημένη στην PDB με κωδικό 1B9O [76]), τα κατάλοιπα A106 έως C111 δημιουργούν loop, αυτή δε η περιοχή εκτίθεται περισσότερο στον διαλύτη από την αντίστοιχη περιοχή της δομής της υψηλής θερμοκρασίας. Με τις δύο αυτές δομές (της θεωρούμενης φυσικής δομής της ανθρώπινης α-λακταλβουμίνης της υψηλής θερμοκρασίας και της δομής του ίδιου μορίου της χαμηλής), οι Harata και Muraki [77] έδειξαν την μεγάλη διακύμανση της δομής του μορίου, ιδιαίτερα της περιοχής W104-C111 και την μετάπτωση έλικας-βρόχου, λόγω τοπικών αλληλεπιδράσεων, στην α-La.

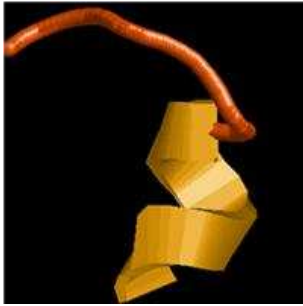
Σε αναζήτηση δομικών μοντέλων α-λακταλβουμίνης άλλων θηλαστικών [78] βρίσκουμε στην PDB, τις κρυσταλλικές δομές της α-La με καταχωρήσεις 1HFY, 1HFZ και 1HFZ.

Η 1HFY δομή των 123 καταλοίπων της α-λακταλβουμίνης (Εικόνα 1.8) προέρχεται από το θηλαστικό *Cavia porcellus* (guinea pig) και έχει προσδιοριστεί στους 298K και pH 4.6. Τα κατάλοιπα 101-105 έχουν διαμόρφωση α-έλικας και τα υπόλοιπα 106-111 σχηματίζουν στροφή (turn-like). Η αλληλουχία της στα κατάλοιπα 101-111 είναι παρόμοια με την ανθρώπινη με την διαφορά μίας προλίνης αντί αλανίνης στην θέση 109 (αλληλουχία IDYWLAHKPLC).

Η δομή 1HFY (Εικόνα 1.9) με 120 κατάλοιπα προέρχεται από την α-λακταλβουμίνη της κατσίκας (*Capra hircus*), έχει προσδιοριστεί στους 298K σε pH 4.5. Η αλληλουχία των καταλοίπων 101-111 είναι IDYWLAHKALC, ολόγεια με αυτήν της ανθρώπινης. Βλέπουμε τα κατάλοιπα 101-103 και 108-111 με μορφή έλικας.

Στην 1HFZ κρυσταλλική δομή της α-La της αγελάδας (*Bos Taurus*, Εικόνα 1.10), με αλληλουχία στα κατάλοιπα 101-111 την INYWLAHKALC, που προσδιορίστηκε στους 298K σε pH 8, βλέπουμε ότι τα μόνα κατάλοιπα που δεν συμμετέχουν σε έλικα είναι τα 104 και 111.

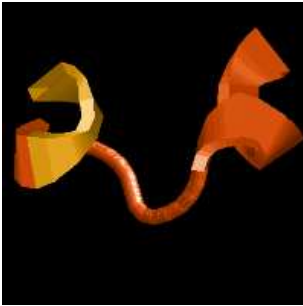
Στις τρεις επόμενες εικόνες απεικονίζεται, με την βοήθεια του rasmol [79] το τμήμα των καταλοίπων 101-111 των κρυσταλλικών δομών της α-λακταλβουμίνης των διαφορετικών θηλαστικών, που αναφέρθηκαν ανωτέρω (καταχωρήσεις 1HFY, 1HFZ και 1B9O).



**Εικόνα 1.8:**

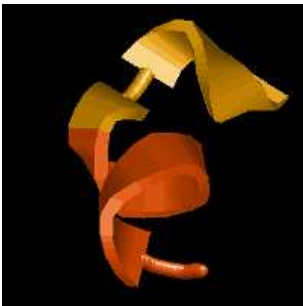
Τμήμα των ένδεκα καταλοίπων 101-111 της καταχωρημένης με κωδικό 1HFY δομής της PDB. Το C-τελικό άκρο απεικονίζεται στο πάνω τμήμα της εικόνας.

Δεξιά από την δομή βλέπουμε το guinea pig, του οποίου ευγενική προσφορά είναι η α-La της ανωτέρω καταχώρησης.



**Εικόνα 1.9:**

Στην καταχωρημένη με κωδικό 1HFY δομή της PDB της α-La της κασίικας βλέπουμε το τμήμα των καταλοίπων 101-111 με το C-τελικό άκρο δεξιά. Ακόμα πιο δεξιά βλέπουμε την κασίικα, της οποίας η α-La δεν απεικονίζεται.



**Εικόνα 1.10:**

Αριστερά: το τμήμα των καταλοίπων 101-111 της καταχώρισης 1HFZ της α-La της αγελάδας. Το C -τελικό άκρο είναι κάτω.



Κέντρο: επάνω η αγελάδα, που γελά: κάτω μία πιο σοβαρή αγελάδα.



**Εικόνα 1.11:**

Η καταχώρηση 1B9O της ανθρώπινης α-La, με αλληλουχία IDYWLAHKALC.

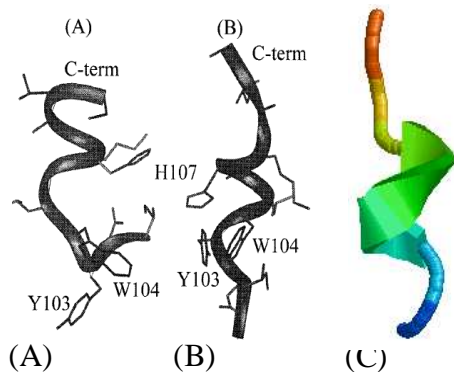
Αριστερά: Το τμήμα των καταλοίπων 101-111 με την C-τελική κυστεΐνη επάνω.

Κέντρο: Η δομή, όπως προσδιορίστηκε κρυσταλλογραφικά στους 286K και pH 4.2

Δεξιά: Τυχαίος εκπρόσωπος του είδους Homo sapiens sapiens.

Από τις παραπάνω εικόνες, παρατηρούμε την ποικιλία των διαμορφώσεων του τμήματος των καταλοίπων 101 έως 111 της προερχόμενης από διαφορετικά θηλαστικά α-λακταλβουμίνης με σαφή μεν ελικοειδή διαμόρφωση, αλλά δημιουργούμενη από διαφορετικά κατάλοιπα. Συγκρίνοντας την δομή της αλληλουχίας των καταλοίπων 101-111 της α-La της κατσίκας και του ανθρώπου των Εικόνων 1.9 και 1.11 (οι οποίες, στο τμήμα αυτό των καταλοίπων έχουν ίδια ακριβώς αλληλουχία και έχουν προσδιοριστεί και οι δύο σε χαμηλό pH) παρατηρούμε σημαντικές διαφορές, ενώ σημαντική ομοιότητα παρουσιάζουν οι αντίστοιχες δομές του guinea pig και του ανθρώπου (Εικόνες 8 και 11) αν και διαφέρουν σε μία προλίνη.

Το τμήμα 101-111 αποτελεί ιδιαίτερα ενδιαφέρουσα περιοχή του μορίου, εξαιτίας του σημαντικού ρόλου, που διαδραματίζει στην σταθεροποίηση της εύπλαστης σφαίρας. Το αυτό τμήμα αυτό της α-λακταλβουμίνης συντέθηκε χημικά, για διερεύνηση των τάσεων διαμόρφωσης του συγκεκριμένου τμήματος. Αντικαταστάθηκε η κυστεΐνη με αλανίνη και προστέθηκαν προστατευμένα άκρα (N-τελική ακετυλομάδα και C-τελική αμινομάδα). Συντέθηκε με αυτόν τον τρόπο το ενδεκαπεπτίδιο Ac-IDYWLAHKALA-NH<sub>2</sub> (καταχώρηση PDB: 1CB3). Ερευνήθηκε η δομή του πεπτιδίου τόσο σε υδατικό διάλυμα, όσο και σε διάλυμα σε 30% TFE με χρήση CD και MNR. Το τελικό αποτέλεσμα ήταν να ληφθούν δύο διαφορετικές μη φυσικές δομές. Η δομή στο υδατικό διάλυμα (αυτή, που μας ενδιαφέρει) λαμβάνει διαμόρφωση, που μοιάζει με 3<sub>10</sub> έλικα. Παρατηρήθηκαν δεσμοί υδρογόνου ανάμεσα στο καρβονυλικό οξυγόνο της Y103 και στο αμιδικό πρωτόνιο της A106, όπως επίσης μεταξύ W104 και H107. Η δομή φάνηκε να σταθεροποιείται από μη φυσικές αλληλεπιδράσεις των πλευρικών αλυσίδων και συγκεκριμένα από την αλληλεπίδραση των καταλοίπων I101, Y103 και W104 και αυτήν ανάμεσα στα κατάλοιπα Y103 και H107. Η δομή αυτή των καταλοίπων 101-107 είναι ιδιαίτερα μη φυσική, αλλά καλά διαμορφωμένη, αποτέλεσμα των τοπικών αλληλεπιδράσεων, διαφέρει δε σημαντικά από την φυσική διαμόρφωση του τμήματος 101-107 της α-LA [47](Εικόνα 1.12).



**Εικόνα 1.12:**

(A) Διάγραμμα κορδέλας των καταλοίπων 101-111 της κρυσταλλικής δομής υψηλού pH της ανθρώπινης α-La της φυσικής διαμόρφωσης.

(B) Διάγραμμα κορδέλας του ενδεκαπεπτιδίου, που καταχωρήθηκε τελικά ως 1CB3. Φαίνονται οι μη φυσικές αλληλεπιδράσεις των πλευρικών ομάδων, που αναφέρθηκαν στο κείμενο.

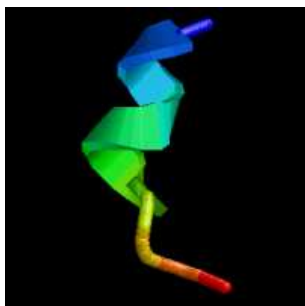
(C) Το μοντέλο 1 της προσδιορισθείσης με NMR δομής, της 1CB3 με το rasmol. Στην έλικα συμμετέχουν τα κατάλοιπα W4, L5, A6 και H7.

Η τοπική δομή του πεπτιδίου φαίνεται να μιμείται τμήμα της δομής, που δημιουργείται σε αυτήν την περιοχή των καταλοίπων του ολόκληρου μορίου της α-LA στην κατάσταση της εύπλαστης σφαίρας. Το αποτέλεσμα της μελέτης διαμόρφωσης αυτού του απομονωμένου πεπτιδίου είναι ένα καλό παράδειγμα, το οποίο συνηγορεί στην υπόθεση ότι η δομή της εύπλαστης σφαίρας επάγεται από τοπικές αλληλεπιδράσεις.

Το ενδεκαπεπτίδιο αυτό επιλέχθηκε από τους Araki και Tamura [46], για να μελετηθεί η επίδραση μη τοπικών αλληλεπιδράσεων (που θα προέκυπταν με μία προσθήκη επτά επιπλέον καταλοίπων στο C-τελικό άκρο) υποθέτοντας ότι οι μη τοπικές αλληλεπιδράσεις, που θα προέκυπταν από την προσθήκη των επτά καταλοίπων θα αποσταθεροποιούσαν την δομή του TP και θα μετέτρεπαν την δομή σε β-hairpin. Για την μελέτη αυτή το αρχικό ενδεκαπεπτίδιο της καταχώρησης 1CB3 της PDB τροποποιήθηκε. Το ασπαρτικό της θέσης 2 αντικαταστάθηκε από ασπαραγίνη, για να παρεμποδιστεί η πιθανή ευαισθησία του πεπτιδίου σε αλλαγές pH. Επιπλέον τα κατάλοιπα 8-11 (KALA) του C-τελικού άκρου αντικαταστάθηκαν με το την αλληλουχία AKAG, η οποία έχει την ικανότητα να παίρνει σταθερή διαμόρφωση στροφής σε μερικά μικρά πεπτίδια, με την λογική ότι η αλληλουχία στροφής διαδραματίζει σημαντικό ρόλο στην αναδίπλωση για β-hairpin. Σχεδιάστηκε με αυτόν τον τρόπο το ενδεκαπεπτίδιο INYWLAHAKAG, που ονομάστηκε TP (Target Peptide). Αυτό συντέθηκε χημικά, ερευνήθηκε η δευτεροταγής δομή του με CD, αποτιμήθηκε η μονομερής του οντότητα με NMR, υπολογίστηκε η δομή του με NMR και καταχωρήθηκε στην PDB ως 2DX2 (Εικόνα 1.13). Η δομή του είναι ελικοειδής και



μάλιστα 3<sub>10</sub> έλικα, που όπως και του αρχικού πεπτιδίου, αυτού της καταχώρησης 1CB3, η δε έλικά του σχηματίζεται από τα κατάλοιπα 2-7.



**Εικόνα 1.13:**

Το μοντέλο 1 της υπολογισθείσης NMR δομής 2DX2. Η θερμοκρασία του πειράματος ήταν 283K, το pH 4.5 και η ιονική ισχύς του διαλύματος 6.5mM. Δεσμοί υδρογόνου δημιουργούνται μεταξύ καταλοίπων N2-L5, Y3-A6, W4-H7. Οι πλευρικές ομάδες των Y3, W4 και H7 αλληλεπιδρούν, σταθεροποιώντας την διαμόρφωση της 3<sub>10</sub> έλικας. Το C-τελικό άκρο είναι κάτω.

Σχεδιάστηκαν 5 επιπλέον πεπτίδια (DPs –Design Peptides) με προσθήκη 7 καταλοίπων παρόμοιας αλληλουχίας στο C-τελικό άκρο του TP, υποθέτοντας ότι οι αλληλεπιδράσεις των ατόμων των επιπλέον επτά καταλοίπων με τα άτομα της TP, θα ήταν ικανές να αποσταθεροποιήσουν την ελικοειδή δομή. Οι δομές τους υπολογίστηκαν με NMR σε διάλυμα ιονικής ισχύος 6.5 mM, στους 283K και σε pH 4.5. Τα τρία από αυτά τα DPs διαμόρφωσαν πληθυσμούς τόσο με α, όσο και με β διαμόρφωση, ενώ τα υπόλοιπα δύο αποτελούνταν από πληθυσμό αποκλειστικά ελικοειδούς δομής. Το DP5 είχε τον μεγαλύτερο αριθμό NOEs, που συμφωνούσε με την ύπαρξη δύο μονομερών, εξίσου σταθερών πληθυσμών δομών: ενός πληθυσμού σταθερής ελικοειδούς δομής και ενός πληθυσμού σταθερής β-δομής. Η έλικα της DP5 είναι α-έλικα, που δείχνει ότι η ελικοειδής δομή του TP τελικά σταθεροποιήθηκε περισσότερο με την προσθήκη των επτά επιπλέον καταλοίπων. Οι υπολογισθείσες με NMR δομές των δύο πληθυσμών καταχωρήθηκαν στην PDB με κωδικούς 2DX3 και 2DX4. Η πρώτη αντιστοιχεί στην ελικοειδή διαμόρφωση (DP5\_conformation 1) και η δεύτερη στην τύπου β-δομή (DP5\_conformation 2). Τα μοντέλα 1 των δύο δομών απεικονίζονται στην Εικόνα 1.14, ως cartoons με το rasmol.



**Εικόνα 1.14:** Το δεκαοκταπεπτίδιο DP5. Αριστερά η 2DX3 και δεξιά η 2DX4.

Το ενδεκαπεπτιδίο INYWLAHAKAG (το TP της Εικόνας 1.13) επιλέξαμε να μελετήσουμε επιτελώντας δύο ανεξάρτητες προσομοιώσεις μοριακής δυναμικής χρησιμοποιώντας το πρόγραμμα NAMD [80], τη βοήθεια του VMD [81] και δυναμικά πεδία τα CHARMM [82] για να ερευνήσουμε αφενός την δυνατότητα των σύγχρονων δυναμικών πεδίων να διατηρήσουν την όχι ιδιαίτερα σταθερή δομή της  $3_{10}$ -έλικας, αφετέρου την ικανότητά τους να δημιουργήσουν την σπάνια αυτή έλικα με αρχική δομή τελείως εκτεταμένη διαμόρφωση και τέλος για να αναλύσουμε τα δομικά μοντέλα των προσομοιώσεών μας και να συγκρίνουμε τα δεδομένα, τα οποία θα λάβουμε, με τα πειραματικά δεδομένα του NMR και με τις καταχωρημένες στην PDB δομές.

# Κεφάλαιο 2<sup>ο</sup>

## Μοριακός μοντελισμός και προσομοιώσεις μοριακής δυναμικής.

The purpose of models is not to fit the data but to sharpen the questions.  
Samuel Karlin.

Μοριακός μοντελισμός (molecular modeling) είναι η επιστήμη και τέχνη της μελέτης της μοριακής δομής και λειτουργίας μέσω κατασκευής μοντέλων και χρήσης μεθόδων βασιζόμενων σε υπολογιστές. Η κατασκευή μοντέλων μπορεί να είναι απλή και να πραγματοποιείται με πλαστικά πολύχρωμα σφαιρικά κομμάτια, τα οποία δηλώνουν τα άτομα και με (ή χωρίς) ραβδόμορφα τμήματα, τα οποία δηλώνουν τους δεσμούς μεταξύ των ατόμων. Μπορεί να είναι σκελετικά μοντέλα, όπως αυτά του DNA των Crick και Watson και της μυοσφαιρίνης του Kendrew. Μπορεί ακόμα να

πρόκειται για πολυεδρικά (συνηθέστερο για ανόργανα στερεά) ή για πολύπλοκα μοντέλα (όπως τα μοντέλα μακρομορίων του Nicholson). Για την κατασκευή μοντέλων εκμεταλλευόμαστε σήμερα και τα νέες τεχνολογίας, ελεγχόμενα από υπολογιστές, μηχανήματα, τα οποία δημιουργούν με αυτοματοποιημένο τρόπο συμπαγή μοντέλα πολύπλοκων μορίων (μέθοδος rapid prototyping –δημιουργία μοντέλων αποτελούμενων από πολλά στρώματα με τεχνολογία additive manufacturing). Επιπλέον, με χρήση laser για εγχάραξη, μπορούμε να δημιουργούμε μοντέλα μορίων μέσα σε γυαλί (μέθοδος subsurface laser engraving). Τέλος υπάρχει και ο μοριακός μοντελισμός για την δημιουργία μοντέλων τριών διαστάσεων με την βοήθεια υπολογιστών και οπτικοποίηση των μοντέλων αυτών με χρήση μοριακών γραφικών.

Ο μοριακός μοντελισμός με την βοήθεια υπολογιστών περιλαμβάνει μεθόδους κβαντικής μηχανικής (ab initio και semi-empirical), μοριακής (εμπειρικής) μηχανικής και πολλές άλλες διαδικασίες (Μοριακής Δυναμικής, Monte Carlo, free energy και solvation methods, SAR (structure/activity relationship), χρήση χημικής και βιοχημικής πληροφορίας και πληροφορίας αποθηκευμένης σε βάσεις δεδομένων). Ακόμα και το αποτέλεσμα της επεξεργασίας της πληροφορίας που προέρχεται από NMR και τα είδωλα της κρυσταλλογραφία ακτίνων X θα μπορούσαν να θεωρηθούν είδος μοριακού μοντελισμού [33].

Οι ρίζες του μοριακού μοντελισμού δημιουργήθηκαν με την ιδέα ότι η γεωμετρία του μορίου, η ενέργεια και πολλές από τις ιδιότητες του μορίου μπορούν να υπολογιστούν από μηχανικά μοντέλα, που υπόκεινται σε φυσικές δυνάμεις. Ένα μόριο αναπαριστάται ως ένα μηχανικό σύστημα, του οποίου τα σωματίδια –τα άτομα- είναι συνδεδεμένα μεταξύ τους με ελατήρια –τους δεσμούς. Το μόριο στην συνέχεια περιστρέφεται, δονείται και μετατοπίζεται, ως συνολική αντίδραση στις ενδομοριακές και διαμοριακές δυνάμεις, που ενεργούν σε αυτό. Οι δυνάμεις αυτές εκφράζονται ως άθροισμα των αρμονικών εκφράσεων (θεωρώντας ότι ισχύει ο νόμος του Hooke) για τις διακυμάνσεις του μήκους και της γωνίας των δεσμών, ως αρμονικές ή τριγωνομετρικές εκφράσεις για την ροπή που υπολογίζεται για την περιστροφή τμημάτων του μορίου γύρω από δεσμούς, που ενώνουν τα τμήματα αυτά και τέλος εκφράζονται και για τις μη δεσμικές (δυνάμεις van der Waals με την μορφή δυναμικού 12-6 Lennard-Jones και ηλεκτροστατικές). Περισσότερα για την χρήση του μηχανικού μοντέλου της Μοριακής

Μηχανικής στις υπολογιστικές προσομοιώσεις ακολουθεί παρακάτω. Πρώτα, όμως, θα αναφέρουμε ιστορικά στοιχεία για τις υπολογιστικές προσομοιώσεις.

## **2.1 Υπολογιστικές προσομοιώσεις (computer simulations).**

Η ιστορία των υπολογιστικών προσομοιώσεων έχει ως εξής: Κατά την διάρκεια του δεύτερου παγκόσμιου πολέμου και αμέσως μετά, είχε αναπτυχθεί η τεχνολογία για δημιουργία υπολογιστικών μηχανών για να μπορούν αυτές να εκτελούν πολύπλοκους υπολογισμούς, με τελικό σκοπό την ανάπτυξη πυρηνικών όπλων καθώς και την χρήση τους για αποκρυπτογράφηση (code breaking). Στις αρχές της δεκαετίας του '50 αυτοί οι υπολογιστές έγιναν διαθέσιμοι και για μη στρατιωτικούς σκοπούς. Τότε άρχισε η χρήση τους για την πραγματοποίηση προσομοιώσεων λιγότερο ή περισσότερο πολύπλοκων συστημάτων. Και ο λόγος που χρησιμοποιήθηκαν γι' αυτόν τον σκοπό ήταν απλά για να εκτιμηθεί η 'λογική' και τα όρια των ικανοτήτων των υπολογιστών [83].

Πριν από την εποχή των προσομοιώσεων, η προσπάθεια για την πρόβλεψη των ιδιοτήτων των πολύπλοκων συστημάτων βασιζόταν στις προσεγγιστικές θεωρίες (approximate theories). Δεδομένης ικανοποιητικής (όχι όμως πλήρους) πληροφορίας για τις ενδομοριακές αλληλεπιδράσεις, οι θεωρίες αυτές δίνουν μια γενική εκτίμηση των ιδιοτήτων του υπό εξέταση υλικού. Το γεγονός της περιορισμένης γνώσης όλων των ενδομοριακών αλληλεπιδράσεων, ακόμα και για σχετικά απλά μόρια, αποτελεί πρόβλημα όταν θέλουμε να ελέγξουμε την εγκυρότητα μιας συγκεκριμένης θεωρίας με πείραμα. Αν η θεωρία και το πείραμα δεν είναι σύμφωνα τότε είναι πιθανό είτε η θεωρία να μην είναι σωστή είτε να μην έχουμε υπολογίσει σωστά τις ενδομοριακές αλληλεπιδράσεις στο πείραμα είτε να συμβαίνουν και τα δύο αυτά μαζί. Με τις υπολογιστικές προσομοιώσεις μπορούμε να συγκρίνουμε τις ιδιότητες ενός υπολογιστικού μοριακού συστήματος απευθείας με αυτές του πειραματικού συστήματος. Αν διαφωνούν, τότε το υπολογιστικό μοντέλο είναι ανεπαρκές και χρειάζεται βελτίωση. Ακόμα μπορούμε να συγκρίνουμε το αποτέλεσμα της προσομοίωσης του δεδομένου μοριακού συστήματος με τις προβλέψεις για το μοντέλο αυτό των αναλυτικών προσεγγιστικών θεωριών. Αν σε αυτήν την περίπτωση υπάρχει διαφωνία θεωρίας και προσομοίωσης τότε η θεωρία έχει το πρόβλημα. Σε αυτήν την

περίπτωση η υπολογιστική προσομοίωση είναι το πείραμα το οποίο ελέγχει την θεωρία (computer experiment). Με αυτόν τον τρόπο μπορούμε να βελτιώσουμε υπάρχουσες θεωρίες και να αλλάξουμε τον τρόπο με τον οποίο κατασκευάζουμε τις νέες. Σήμερα σπάνια μια θεωρία εφαρμόζεται στον αληθινό κόσμο αν πριν δεν έχει ελεγχθεί με υπολογιστικές προσομοιώσεις. Θα πρέπει βέβαια, να λαμβάνεται υπόψη ότι τα δεδομένα με τα οποία τροφοδοτούμε τους υπολογιστές για την διεξαγωγή των υπολογιστικών πειραμάτων μπορεί να έχουν στατιστικά λάθη, όπως τέτοια λάθη υπάρχουν και στα αληθινά πειράματα. Και επιπλέον, όπως και στα αληθινά πειράματα, θα πρέπει να εξάγουμε τελικά τη σημαντική πληροφορία [83].

Όταν οι υπολογιστές έγιναν τελικά διαθέσιμοι για μη πολεμικούς σκοπούς, χρησιμοποιήθηκαν αρχικά για την προσομοίωση των dense liquids. Η πρώτη προσομοίωση ενός υγρού έγινε από τους Metropolis, Rosenbluth, Rosenbluth, Teller and Teller στους υπολογιστές MANIAC στο Los Alamos εισάγοντας την μέθοδο Monte Carlo [84]. Η ονομασία 'προσομοιώσεις Monte Carlo' είχε επινοηθεί νωρίτερα από τους Metropolis και Ulam [85], επειδή στην μέθοδο αυτή χρησιμοποιούνται κατά κόρον τυχαίοι (παραγόμενοι από υπολογιστές) αριθμοί. Το ίδιο περίπου χρονικό διάστημα οι Fermi, Pasta και Ulam [86] παρουσίασαν την μελέτη της δυναμικής μη αρμονικού (anharmonic) μονοδιάστατου κρυστάλλου. Η πρώτη αξιοπρεπής προσομοίωση μοριακής δυναμικής (Molecular Dynamic simulation, MD simulation) αναφέρεται το 1956 από τους Alder και Wainwright [87], οι οποίοι μελέτησαν την δυναμική της συναρμολόγησης των σκληρών σφαιρών (hard spheres). Η πρώτη προσομοίωση με μοντέλο πραγματικό υλικό δημοσιεύτηκε το 1960 από την ομάδα του Vineyard [88], η οποία μελέτησε την δυναμική της καταστροφής από ραδιενέργεια, ενώ τα πρωτεία στην προσομοίωση πραγματικού υγρού έχουν οι Rahman at Argonne [89]. Η πρώτη προσομοίωση μοριακής δυναμικής σε πρωτεΐνη αναφέρεται το 1977 από τους McCammon, Gelin και Karplus και αφορά τον αναστολέα πρωτεασών σερίνης BPTI [90] (Bovine Pancreatic Trypsine Inhibitor, πρωτεΐνη 500 ατόμων, με διάρκεια προσομοίωσης μόλις 9.2 ps). Η μεθοδολογία των υπολογιστικών προσομοιώσεων έχει αναπτυχθεί από τότε κύρια λόγω της βελτίωσης της τεχνολογίας των υπολογιστών, ενώ οι βασικοί αλγόριθμοι των MC και MD δεν έχουν αλλάξει σημαντικά από το 1950.

Η πλέον κοινή εφαρμογή των υπολογιστικών προσομοιώσεων είναι η πρόβλεψη των ιδιοτήτων και της συμπεριφοράς των υλικών κάτω από συγκεκριμένες συνθήκες. Η

χρησιμότητα τέτοιων προσομοιώσεων μπορεί να μην είναι εκ πρώτης όψεως προφανής. Για παράδειγμα είναι πολύ πιο εύκολο να μετρήσει κανείς το σημείο πήξης του νερού από το να το εξάγει υπολογιστικά με προσομοιώσεις. Το ζήτημα είναι ότι είναι εύκολο να μετρηθεί το σημείο πήξης του νερού σε πίεση μίας ατμόσφαιρας, αλλά πολύ δύσκολο έως αδύνατο και σίγουρα δαπανηρό να μετρηθούν οι ιδιότητες ενός πραγματικού υλικού σε πολύ υψηλές πιέσεις ή θερμοκρασίες. Με τους υπολογιστές μπορούμε να προσομοιώσουμε ένα σύστημα σε οποιαδήποτε (σχεδόν) θερμοκρασία ή πίεση. Επιπλέον με υπολογιστικές προσομοιώσεις μπορούμε να προβλέψουμε ακόμα και ιδιότητες υλικών που ακόμα δεν έχουν ακόμα κατασκευαστεί. Τέλος, οι υπολογιστικές προσομοιώσεις μπορούν να χρησιμοποιηθούν απλά και ως εξερευνητικό εργαλείο που μπορεί να προσφέρει αξιόλογη πληροφορία για την κατανόηση των βιολογικών διεργασιών [83].

## **2.2. Μοριακή Μηχανική & προσομοιώσεις μοριακής Δυναμικής.**

Η Μοριακή Μηχανική (μέθοδος των δυναμικών πεδίων ή της δυναμικής ενέργειας) βασίζεται στην αρχή κατά την οποία το άθροισμα των φυσικών δυνάμεων, οι οποίες επενεργούν σε ένα μόριο, μπορεί να χρησιμοποιηθεί για να περιγράψει την γεωμετρία και την ενέργεια του μορίου αυτού, η τελική δε διαμόρφωσή του μορίου αυτού στον χώρο είναι αυτή της ελάχιστης ολικής εσωτερικής ενέργειας. Ένα σύστημα μορίων, ως κλασσικό σύστημα της μοριακής μηχανικής, έχει αναφερθεί σε προηγούμενη παράγραφο. Η μοριακή μηχανική απαιτεί την κατασκευή μίας συνάρτησης, η οποία θα περιγράφει την δυναμική ενέργεια του συστήματος αυτού ως προς τις θέσεις (και συνεπώς και τις αποστάσεις) των ατόμων, που το απαρτίζουν. Οι ατομικές θέσεις λαμβάνονται από πειραματικά δεδομένα.

Οι αρχές στις οποίες βασίζεται η μοριακή δυναμική είναι τρεις. Η πρώτη είναι η βασική θερμοδυναμική υπόθεση, κατά την οποία ένα σύστημα οδηγείται αυθόρμητα σε κατάσταση χαμηλότερης ελεύθερης ενέργειας. Η προσθετικότητα των δυναμικών είναι η δεύτερη αρχή. Οι δυνάμεις, οι οποίες επενεργούν στα άτομα εκφράζονται ως άθροισμα των δεσμικών και των μη δεσμικών δυνάμεων. Η τρίτη αρχή είναι αυτή της δυνατότητας μεταβίβασης των υπολογισθέντων από απλές αντιπροσωπευτικές δομές

δυναμικών σε άλλα πολυπλοκότερα μόρια, τα οποία όμως αποτελούνται από τις ίδιες χημικές υπομονάδες, εισάγοντας με αυτό τον τρόπο την έννοια του τύπου ατόμου (atom type). Ο τύπος ατόμου απεικονίζει το περιβάλλον του κάθε ατόμου [33].

Η Μοριακή Δυναμική επιτρέπει την αλληλεπίδραση των ατόμων ενός μορίου για μία χρονική περίοδο και προβλέπει την μοριακή κίνηση σε ατομικό επίπεδο κατά την διάρκεια αυτής της περιόδου. Με την υπολογιστική μέθοδο των προσομοιώσεων μοριακής δυναμικής υπολογίζεται η συμπεριφορά ενός μοριακού συστήματος συναρτήσει του χρόνου παρέχοντας λεπτομερή πληροφορία για τις διακυμάνσεις και τις αλλαγές διαμόρφωσης του υπό προσομοίωση μορίου.

Η μέθοδος των προσομοιώσεων μοριακής δυναμικής βασίζεται στον δεύτερο νόμο της κίνησης του Νεύτωνα (κλασική μηχανική). Σύμφωνα με αυτόν μπορούμε να προβλέψουμε την επιτάχυνση ενός αντικειμένου, δεδομένης της μάζας του και της συνολικής δύναμης η οποία εφαρμόζεται σε αυτό. Ο νόμος αυτός εκφράζεται με την εξίσωση  $\mathbf{F} = m \cdot \mathbf{a}$ , όπου  $\mathbf{F}$  είναι η συνολική δύναμη (δηλ το άθροισμα των ανυσμάτων των εφαρμοζόμενων στο σώμα δυνάμεων),  $m$  είναι η μάζα του σώματος και  $\mathbf{a}$  η επιτάχυνσή του. Το ολοκλήρωμα της ανωτέρω εξίσωσης της κίνησης παράγει ένα τροχιακό (trajectory), το οποίο περιγράφει τις θέσεις, τις ταχύτητες και τις επιταχύνσεις των σωματιδίων του συστήματος σε όλη την διάρκεια του χρόνου προσομοίωσης. Από αυτό το τροχιακό μπορούν να καθοριστούν οι μέσες τιμές πολλών ιδιοτήτων του συστήματος. Στην κλασική μηχανική ισχύει η αιτιοκρατία (determinism). Εφόσον οι θέσεις και οι ταχύτητες όλων των ατόμων είναι γνωστές, η κατάσταση του απορτιζόμενου από αυτά τα άτομα συστήματος μπορεί να προβλεφθεί επακριβώς σε οποιαδήποτε μελλοντική χρονική στιγμή. Και η χρονική στιγμή μπορεί να καθοριστεί, εφόσον ο χρόνος στην κλασική μηχανική είναι απόλυτος.

### **2.3. Στατιστική μηχανική (statistical mechanics).**

Όπως αναφέραμε προηγουμένως, οι υπολογιστικές προσομοιώσεις μας επιτρέπουν να μελετήσουμε τις ιδιότητες ενός συστήματος πολλών σωματιδίων. Ωστόσο αυτές, δεν μπορούν να μετρήσουν άμεσα όλες τις ιδιότητες του συστήματος.



Ουσιαστικά, οι περισσότερες από τις ποσότητες που μπορούν να μετρηθούν με την υπολογιστική προσομοίωση δεν ανταποκρίνονται σε αυτές που μετρούνται στα πραγματικά πειράματα. Για παράδειγμα, με μια προσομοίωση του υγρού νερού μπορούμε να μετρήσουμε τις στιγμιαίες θέσεις και τις ταχύτητες όλων των ατόμων του υγρού. Αυτή όμως η πληροφορία δεν μπορεί να συγκριθεί με πειραματικά δεδομένα εφόσον κανένα πείραμα δεν μπορεί να μας δώσει τέτοια στοιχεία. Σε ένα πείραμα μετράμε την μέση τιμή μιας ιδιότητας του συνόλου των σωματιδίων, τα οποία απαρτίζουν το υπό εξέταση υλικό και για το σύνολο της χρονικής διάρκειας του πειράματος. Αν θέλουμε να χρησιμοποιήσουμε τα δεδομένα των υπολογιστικών προσομοιώσεων ως συμπληρωματικά του πειράματος θα πρέπει να μετατρέψουμε την μικροσκοπική πληροφορία, την οποία αυτά παράγουν (ατομικές θέσεις και ταχύτητες) σε μακροσκοπικά παρατηρούμενα με το πείραμα δεδομένα (πίεση, ενέργεια κλπ). Αυτό το επιτυγχάνουμε με την στατιστική μηχανική. Η στατιστική μηχανική παρέχει ακριβείς μαθηματικές συναρτήσεις, οι οποίες συσχετίζουν τις μακροσκοπικές ιδιότητες με την κατανομή και την κίνηση των ατόμων (και μορίων) ενός συστήματος  $N$ -σωματιδίων, ενώ οι προσομοιώσεις μοριακής δυναμικής παρέχουν το μέσο για την επίλυση της εξίσωσης της κίνησης των σωματιδίων και την εκτίμηση αυτών των μαθηματικών τύπων. Με τις προσομοιώσεις μοριακής δυναμικής μπορούν να μελετηθούν τόσο οι θερμοδυναμικές ιδιότητες όσο και εξαρτώμενα από τον χρόνο κινητικά φαινόμενα [91,92].

Η θερμοδυναμική κατάσταση ενός συστήματος καθορίζεται από μικρό αριθμό παραμέτρων, όπως η θερμοκρασία, η πίεση και ο αριθμός των σωματιδίων (ατόμων) του συστήματος. Άλλες θερμοδυναμικές ιδιότητες μπορούν να παραχθούν από τις καταστατικές και από άλλες βασικές θερμοδυναμικές εξισώσεις. Οι θερμοδυναμικές αυτές μεταβλητές (state variables) όπως είναι η ελεύθερη ενέργεια Gibbs, η ενθαλπία, η εντροπία και άλλες, περιγράφουν την στιγμιαία κατάσταση ενός συστήματος και είναι ανεξάρτητες από το μονοπάτι που ακολούθησε το σύστημα για να φτάσει στην κατάσταση αυτή. Βάσει της θερμοδυναμικής υπόθεσης, ένα φυσικό σύστημα μεταβάλλεται με τον χρόνο κατευθυνόμενο σε κατάσταση μικρότερης ελεύθερης ενέργειας. Η θερμοδυναμική περιγράφει την κινητήρια δύναμη για τις χημικές διεργασίες, ενώ η κινητική τον μηχανισμό με τον οποίο γίνονται αυτές.

Η μηχανική ή μικροσκοπική κατάσταση ενός συστήματος καθορίζεται από την θέση  $q$  και την ορμή  $p$  του κάθε ατόμου που απαρτίζει το σύστημα. Τα  $p$  και  $q$  μπορούν να θεωρηθούν ως συντεταγμένες ενός πολυδιάστατου χώρου, του phase space. Αυτός είναι ο χώρος μέσα στον οποίο αντιπροσωπεύονται όλες οι πιθανές καταστάσεις ενός συστήματος, με κάθε πιθανή κατάσταση (μικροκατάσταση - microstate) να αντιπροσωπεύεται σε ένα μοναδικό σημείο του phase space. Όλες οι πιθανές αυτές καταστάσεις των  $p$  και  $q$  σε συνάρτηση με τον χρόνο, δίνουν τα διαγράμματα φάσεων (phase diagrams). Και τα τελευταία (σε μακροσκοπικό επίπεδο) μας δείχνουν τις περιοχές των θερμοδυναμικών φάσεων (πίεση, θερμοκρασία). Για ένα σύστημα  $N$  σωματιδίων ο χώρος αυτός έχει  $6N$  διαστάσεις. Ένα μοναδικό σημείο αυτού του χώρου περιγράφει την κατάσταση όλου του συστήματος. Το σύνολο των σημείων του phase space που ικανοποιεί τις συνθήκες μίας συγκεκριμένης θερμοδυναμικής κατάστασης ονομάζεται ensemble. Ένα ensemble είναι το σύνολο όλων των μικροκαταστάσεων, οι οποίες μπορούν να έχουν διαφορετικές μικροσκοπικές καταστάσεις, έχουν όμως ολόδια μακροσκοπική ή θερμοδυναμική κατάσταση. Η έννοια του ensemble είναι ευρεστική (εύλογη και διαφωτιστική, όχι όμως αυστηρά θεμελιωμένη) και η ιδέα είναι να μπορούμε να υπολογίζουμε τα μη εξαρτώμενα από τον χρόνο εξέλξης του συστήματος μακροσκοπικά παρατηρούμενα με τον υπολογισμό των μέσων τιμών. Υπάρχουν διάφορα ensemble με διαφορετικά χαρακτηριστικά: Το ισοβαρές – ισόθερμο ensemble ( $NpT$ ) χαρακτηρίζεται από συγκεκριμένο αριθμό ατόμων  $N$ , σταθερή πίεση  $P$  και σταθερή θερμοκρασία  $T$ , το microcanonical ensemble ( $NVE$ ) από συγκεκριμένο αριθμό ατόμων  $N$ , σταθερό όγκο  $V$  και σταθερή ενέργεια  $E$ . Το canonical ensemble ( $NVT$ ) χαρακτηρίζεται επίσης από σταθερό αριθμό ατόμων  $N$  και σταθερό όγκο  $V$ , αλλά και από σταθερή θερμοκρασία  $T$ . Τέλος υπάρχει και το Grand canonical ensemble ( $mVT$ ) του οποίου η θερμοδυναμική κατάσταση έχει ως χαρακτηριστικά συγκεκριμένο χημικό δυναμικό  $m$ , σταθερό όγκο  $V$  και σταθερή θερμοκρασία  $T$ . Ιδανικά, μία προσομοίωση μοριακής δυναμικής δημιουργεί μία ακολουθία σημείων στο phase space σε συνάρτηση με τον χρόνο με τα σημεία αυτά να ανήκουν στο ίδιο ensemble και να αντιστοιχούν σε διαφορετικές διαμορφώσεις του υπό προσομοίωση συστήματος.

## 2.4. Δυναμικά πεδία (force fields).

Οι μέθοδοι των δυναμικών πεδίων (μέθοδοι μοριακής μηχανικής) αγνοούν τις κινήσεις των ηλεκτρονίων και υπολογίζουν την ενέργεια του υπό εξέταση συστήματος, ως συνάρτηση των θέσεων των πυρήνων μόνο. Όταν μιλάμε για δυναμικά πεδία (forcefields), εννοούμε την μορφή της εξίσωσης και το σύνολο των παραμέτρων, που χρησιμοποιούνται για να περιγράψουν την δυναμική ενέργεια ενός συστήματος σωματιδίων, συνήθως ατόμων. Όπως αναφέραμε σε προηγούμενο κεφάλαιο, η δυναμική ενέργεια ενός συστήματος ατόμων αποτελείται από το άθροισμα της δυναμικής ενέργειας των δεσμικών και των μη δεσμικών αλληλεπιδράσεων. Κάθε forcefield χρησιμοποιεί παρόμοια εξίσωση για την δυναμική ενέργεια. Η μορφή της εξίσωσης της δυναμικής ενέργειας, του CHARMM22 αποδίδεται με την εξίσωση, η οποία ακολουθεί:

$$\begin{aligned} V = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \\ & + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 \\ & + \sum_{\text{nonbonded}} \epsilon \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}} \end{aligned}$$

Ο πρώτος όρος της εξίσωσης της ενέργειας αφορά την διάταση των δεσμών ανάμεσα σε δύο άτομα, ο δεύτερος εκτιμά τις γωνίες των δεσμών, ο τρίτος όρος τις διέδρες, ο τέταρτος τις impropers για την επιπεδότητα. Ο πέμπτος όρος της εξίσωσης (όρος Urey-Bradley) αφορά την κάμψη των γωνιών. Οι δύο τελευταίοι όροι αφορούν τις μη δεσμικές αλληλεπιδράσεις: Η ενέργεια van der Waals υπολογίζεται με το δυναμικό 12-6 Lennard-Jones και η ηλεκτροστατική ενέργεια με το δυναμικό Coulomb [93].

Για την σωστή παραμετροποίηση ενός forcefield και εξαιτίας του μεγάλου αριθμού των παραμέτρων, απαιτείται μεγάλη προσπάθεια. Ακόμα και η προσθήκη λίγων επιπλέον παραμέτρων σε ένα υπάρχον forcefield με σκοπό την μοντελοποίηση νέας τάξης μορίων, είναι δύσκολη και χρονοβόρα διαδικασία. Μόνο, όμως, με την σωστή παραμετροποίηση θα μπορέσουν να γίνουν αξιόπιστες προβλέψεις για την δομή.

Όλα τα δυναμικά πεδία βασίζονται σε πολυάριθμες προσεγγίσεις που προέρχονται από διάφορης μορφής πειραματικά δεδομένα. Για τον λόγο αυτόν ονομάζονται 'εμπειρικά'. Μελέτες πρωτεϊνών και άλλων βιολογικών μακρομορίων κυριαρχούνται από προσεγγίσεις βασιζόμενες σε εμπειρικά δυναμικά πεδία. Σε συνέπεια με τον αυξανόμενο αριθμό των μελετών είναι και ο μεγάλος αριθμός των εμπειρικών δυναμικών πεδίων για πρωτεΐνες. Υπάρχουν πολλά εμπειρικά δυναμικά πεδία, τα οποία σχεδιάστηκαν για διαφορετικές χρήσεις. Δεν υπάρχει ένα οικουμενικό forcefield με δυνατότητα εφαρμογής σε οποιοδήποτε μόριο. Ονομαστικά, κάποια από τα ευρέως χρησιμοποιούμενα δυναμικά πεδία είναι: το AMBER (Assisted Model Building and Energy Refinement), το οποίο χρησιμοποιείται για πρωτεΐνες [94] και DNA [95], το OPLS [96,97] (Optimized Potential for Liquid Simulations), το GROMOS [98,99,100], το ECEPP [101] και το TINKER [102,103] και το CHARMM (Chemistry HARvard Molecular Mechanics) [104], το οποίο χρησιμοποιείται τόσο για μικρά μόρια, όσο και για μακρομόρια.

Αν και τα ανωτέρω δυναμικά πεδία έχουν εφαρμοστεί επιτυχώς σε μεγάλο αριθμό πρωτεϊνών, υπάρχουν ενδογενείς περιορισμοί, όσο αφορά την ακρίβειά τους. Για παράδειγμα σε προσομοιώσεις μοριακής δυναμικής με την χρήση του CHARMM είχε γίνει εμφανής η ύπαρξη π-έλικας σε πεπτίδια σε διάλυμα, η οποία όμως οφειλόταν στην τάση για π-έλικες του συγκεκριμένου δυναμικού πεδίου. Σε μικρά μόρια έχει φανεί η αδυναμία αυτού του δυναμικού πεδίου και για την αναπαραγωγή των ενεργειακών δεδομένων της κβαντικής μηχανικής σε διπεπτίδια και τετραπεπτίδια αλανίνης, κάνοντας προφανή την ανάγκη για τη βελτίωση της συμπεριφοράς του σκελετού των πεπτιδίων, όπως αυτή είχε αποδοθεί από το δυναμικό πεδίο.

Η βελτίωση των δυναμικών πεδίων είναι ακόμα μια συνεχιζόμενη προσπάθεια, η οποία καθίσταται αναγκαία εξαιτίας του συνεχώς αυξανόμενου αριθμού των πειραματικών δεδομένων και αυτών της κβαντικής μηχανικής, δεδομένων τα οποία αφορούν τα συστήματα για τα οποία έχουν σχεδιαστεί τα εκάστοτε χρησιμοποιούμενα δυναμικά πεδία [105].

# Κεφάλαιο 3<sup>ο</sup>

## Προσομοιώσεις.

Time is defined so that motion looks simple.  
John Archibald Wheeler.

Οι προσομοιώσεις, τα δεδομένα των οποίων χρησιμοποιήθηκαν για την επίτευξη του στόχου της παρούσας εργασίας, είναι δύο. Η αναφερόμενη ως '**fold**ing', της οποίας η αρχική δομή ήταν η εκτεταμένη μορφή του πεπτιδίου INYWLAHAKAG και πραγματοποιήθηκε στους 320K και η αναφερόμενη ως '**native**' με αρχική δομή 3<sub>10</sub>-έλικας του πεπτιδίου της ίδιας αλληλουχίας και τη θερμοκρασία κατά την προσομοίωση στους 283K. Και οι δύο προσομοιώσεις πραγματοποιήθηκαν με χρήση του

προγράμματος NAMD [106], σε NpT ensemble, με χρήση περιοδικών οριακών συνθηκών, διαλύτη νερό, συγκέντρωση ιόντων ισοδύναμη με 100mM και forcefield το CHARMM. Τα παραχθέντα τροχιακά των δύο προσομοιώσεων χρησιμοποιούνται για σύγκριση, πιθανή σύγκλιση και περαιτέρω ανάλυση.

### **3.1. Απαιτούμενα αρχεία για προσομοίωση μοριακής δυναμικής.**

Οι πληροφορίες, οι οποίες αναφέρονται στο κεφάλαιο αυτό αφορούν τις προσομοιώσεις μοριακής δυναμικής, οι οποίες επιτεύχθηκαν με χρήση του προγράμματος NAMD, την βοήθεια του VMD [107] και την χρήση του CHARMM forcefield [104] με τη διόρθωση CMAP [105]. Αυτό δεν σημαίνει ότι για τις προσομοιώσεις με χρήση άλλων προγραμμάτων και άλλων δυναμικών πεδίων (forcefield) η διαδικασία είναι τελείως διαφορετική, αλλά ότι θα αναφέρουμε αποκλειστικά τη διαδικασία, που ακολουθήσαμε στην παρούσα εργασία.

Για την αρχική προετοιμασία χρειάζεται ένα αρχείο της βάσης δεδομένων δομών βιολογικών μακρομορίων protein data bank (αρχείο .pdb). Στο αρχείο αυτό βρίσκονται αποθηκευμένες, εκτός άλλων, οι συντεταγμένες των ατόμων του πεπτιδίου, πληροφορία η οποία μαζί με το όνομα του καταλοίπου και το όνομα του ατόμου, απαιτείται για την προετοιμασία του συστήματος. Χρειάζεται ακόμα και ένα αρχείο με τον χάρτη τοπολογίας του CHARMM, το οποίο καθορίζει τον τύπο, την μάζα και το φορτίο κάθε ατόμου του κάθε καταλοίπου και παρέχει πληροφορία για την ομοιοπολική σύνδεση ανάμεσα στα άτομα. Το CHARMM forcefield topology file, το οποίο χρησιμοποιήσαμε (έμμεσα με το VMD) είναι το top\_all22\_prot.inp. Το αρχείο αυτό μαζί με το .pdb χρειάζονται για την δημιουργία του αρχείου protein structure file (.psf.).

Στο παραγόμενο αρχείο (.psf) βρίσκεται αποθηκευμένη πληροφορία, η οποία σχετίζεται με την δομή του συστήματος. Χωρίζεται σε έξι ενότητες. Στην πρώτη αναφέρονται τα άτομα του συστήματος, στην δεύτερη οι ομοιοπολικοί δεσμοί μεταξύ των ατόμων (τέσσερα ζεύγη ανά γραμμή), στην τρίτη οι γωνίες, τις οποίες σχηματίζουν τρία συνδεδεμένα άτομα (τρεις τριάδες ατόμων ανά γραμμή). Στην τέταρτη, την πέμπτη και την τελευταία από τις έξι ενότητες, η καταχώρηση γίνεται με δύο τετράδες ανά

γραμμή. Η τέταρτη αφορά τις δίεδρες γωνίες, η πέμπτη τις (τύπου δίεδρης) improper γωνίες (για να εμποδίζεται η παραμόρφωση μεταξύ ενός ατόμου και άλλων τριών ομοιοπολικά συνδεδεμένων) και η έκτη ενότητα δίνει τις παρεμβαλλόμενες τιμές για τον χάρτη διόρθωσης της ενέργειας του backbone (CMAP). Εμείς δημιουργήσαμε ένα X-PLOR formatted CHARMM .psf αρχείο με το VMD (χρήση του PSFGEN [108]), το οποίο χρησιμοποιεί το NAMD μαζί με το CHARMM parameter file.

Το αρχείο παραμετροποίησης του forcefield (parameter file - εμείς χρησιμοποιήσαμε το par\_all22\_prot.inp, τύπου CHARMM με CMAP) περιέχει τις απαραίτητες παραμέτρους για τον υπολογισμό της ενέργειας του μορίου με δεδομένα τα αρχεία δομής (.psf) και συντεταγμένων (.pdb). Το αρχείο αυτό των παραμέτρων είναι στενά συνδεδεμένο με το αντίστοιχο της τοπολογίας και διανέμονται μαζί, τα δε ονόματά τους μοιάζουν. Το CHARMM parameter file αποτιμά δυνάμεις και τις ενέργειες, τόσο τις ομοιοπολικές, οι οποίες αφορούν δεσμούς, γωνίες, δίεδρες, impropers, όσο και τις μη δεσμικές. Το μερικό φορτίο του κάθε ατόμου, που είναι διακριτό για κάθε τύπο ατόμου (atom type), ορίζεται στο αρχείο τοπολογίας (top\_all22\_prot.inp) και στο αρχείο δομής (.psf). Στο αρχείο της παραμετροποίησης ωστόσο, τροποποιούνται οι ηλεκτροστατικές δυνάμεις ανάμεσα σε άτομα, που διαχωρίζονται από τρεις δεσμούς (χρησιμοποιούμε την τακτική scale1-4 exclusion). Το αρχείο παραμετροποίησης, το οποίο χρησιμοποιήσαμε περιέχει και την διόρθωση με το όνομα CMAP. Το CMAP (correction energy map) βελτιώνει τη συμπεριφορά του σκελετού και δίνει πιο σωστά την δυναμική συμπεριφορά του πεπτιδίου.

Τέλος χρειάζεται το αρχείο που λέει στο NAMD πως ακριβώς να τρέξει η προσομοίωση. Αυτό το αρχείο διαμόρφωσης (configuration file) είναι το .namd file. Το .namd file καθορίζει τα αρχεία που περιέχουν την μοριακή δομή και τις αρχικές συντεταγμένες (structure-coordinates), την θερμοκρασία του συστήματος, το βήμα για την προσομοίωση (by default 1 fs), τα ονόματα των παραγόμενων αρχείων και το βήμα με το οποίο γράφονται τα αρχεία αυτά. Επιπλέον καθορίζει το αρχείο παραμετροποίησης, το οποίο θα χρησιμοποιηθεί, καθώς και τον τρόπο χειρισμού των αλληλεπιδράσεων των ατόμων. Στις προσομοιώσεις, που έγιναν, χρησιμοποιήθηκαν περιοδικές οριακές συνθήκες (periodic boundary conditions) και η μέθοδος particle-mesh Ewald για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλου βεληνεκούς στο περιοδικό αυτό σύστημα (PME full electrostatics). Οι προσομοιώσεις

έγιναν σε σταθερή πίεση. Οι προσομοιώσεις σε σταθερή πίεση απαιτούν την ύπαρξη περιοδικών συνθηκών. Η πίεση ρυθμίζεται με δυναμική προσαρμογή του μεγέθους της στοιχειώδους κυψελίδας (unit cell) με το να αλλάζει την κλίμακα (rescaling) όλων (σχεδόν) των ατομικών συντεταγμένων κατά την διάρκεια της προσομοίωσης. Οι προσομοιώσεις σε σταθερή πίεση με το NAMD γίνονται με χρήση της τροποποιημένης μεθόδου Nosé-Hoover, με την οποία χρησιμοποιείται η δυναμική Langevin για τον έλεγχο των διακυμάνσεων στο βαρόμετρο. Η δυναμική Langevin (στοχαστική δυναμική, εναλλακτική της Νευτώνειας) επιδιώκει να προσομοιώσει την ύπαρξη διαμοριακών συγκρούσεων και χρησιμοποιείται για τον έλεγχο της θερμοκρασίας και για την επιτάχυνση της δυναμικής του συστήματος. Για την προσομοίωση σε NpT ensemble με το NAMD, η μέθοδος του πιστονιού των Nosé-Hoover συνδυάζεται πάντα με την μέθοδο ελέγχου της θερμοκρασίας της δυναμικής Langevin.

### **3.2. Τα στάδια για την προετοιμασία του συστήματος.**

Το πρώτο στάδιο, το οποίο πρέπει να ολοκληρωθεί πριν ουσιαστικά αρχίσει η προσομοίωση είναι αυτό το οποίο ονομάζουμε προετοιμασία του συστήματος (system preparation). Κατά την διάρκεια αυτού του σταδίου δημιουργούμε τα αρχεία .pdb και .psf, τα οποία θα χρησιμοποιήσουμε στο αμέσως επόμενο στάδιο. Φορτώνουμε το αρχικό .pdb αρχείο στο πρόγραμμα VMD και χρησιμοποιούμε τον αυτόματο κατασκευαστή psf (μέσω extensions-modeling). Μέσω του VMD χρησιμοποιείται (ως default) το αρχείο τοπολογίας του CHARMM (top\_all22\_prot.inp). Το αποτέλεσμα είναι να δημιουργηθούν τρία αρχεία με καταλήξεις autopsf.pdb, autopsf.psf και autopsf.log. Το αρχείο .log έχει δεδομένα για την πορεία της διαδικασίας, ενώ χρησιμοποιώντας τα άλλα δύο αρχεία (τα οποία περιέχουν την πληροφορία για τις συντεταγμένες των ατόμων και τον τρόπο σύνδεσης των ατόμων μεταξύ τους) και με την βοήθεια του VMD, προσθέτουμε τα νερά (δημιουργούμε το solvation box δίνοντας τις επιθυμητές διαστάσεις για την στοιχειώδη κυψελίδα). Από αυτό το βήμα παράγονται τρία επίσης αρχεία, τα solvate.pdb, solvate.psf και solvate.log. Για μία ακόμα φορά φορτώνουμε στο VMD τα δύο πρώτα solvate αρχεία για να προσθέσουμε ιόντα νατρίου και χλωρίου για τελική ιοντική συγκέντρωση 100mM και εξουδετέρωση των φορτίων. Τα εξαγόμενα αρχεία από αυτή την διαδικασία, τα ionized.pdb και ionized.psf είναι τα



απαιτούμενα δεδομένα για το επόμενο στάδιο, ενώ παράγεται και ένα αρχείο .log (ionized.log). Ελέγχοντας τα .log μπορούμε να βλέπουμε αν η δημιουργία των επιθυμητών αρχείων έχει γίνει σωστά.

Τα μόρια νερού που χρησιμοποιήσαμε για solvation και στα δύο συστήματα είναι τροποποιημένα μοντέλα νερού τύπου TIP3 (TIP3P). Σε αυτά το μόριο του νερού είναι άκαμπτο και λαμβάνονται υπόψη μόνο οι μη δεσμικές αλληλεπιδράσεις των ατόμων του. Οι ηλεκτροστατικές αλληλεπιδράσεις μοντελοποιούνται με τον νόμο του Coulomb και οι δυνάμεις διασποράς και απώθησης με την χρήση του δυναμικού Lennard-Jones. Σε κάθε άτομο του μοντέλου αυτού των τριών σημείων (με τρία σημεία αλληλεπίδρασης, όσα και τα άτομα του νερού) έχει ανατεθεί ένα σημειακό φορτίο, το δε άτομο του οξυγόνου έχει επιπλέον τις παραμέτρους Lennard-Jones. Η γεωμετρία του μορίου του τύπου TIP3 είναι η παρατηρούμενη και στο μόριο του νερού (γωνία H-O-H 104.5°).

Το επόμενο στάδιο είναι το πρώτο πραγματικό βήμα της προετοιμασίας του συστήματος για την προσομοίωση μοριακής δυναμικής και περιλαμβάνει την ελαχιστοποίηση της ενέργειάς του (energy minimization) και την σταδιακή θέρμανση (heating). Για την ολοκλήρωση αυτού του σταδίου χρειάζονται τα αρχεία ionized.pdb, ionized.psf, τα οποία δημιουργήσαμε με την διαδικασία της προετοιμασίας του συστήματος. Χρειάζεται επίσης το par\_all22\_prot.inp (το αρχείο με τις παραμέτρους του CHARMM force field), καθώς και configuration file (το heat.namd), το οποίο τροποποιούμε κατάλληλα για τα δεδομένα της προσομοίωσης. Το ίδιο αρχείο διαμόρφωσης περιέχει την πληροφορία και για την θέρμανση του συστήματος μέχρι την επιθυμητή θερμοκρασία. Το NAMD διαβάζει αυτό το αρχείο (heat.namd) χρησιμοποιώντας τα αρχεία ionized.pdb και ionized.psf. Οι παράμετροι του forcefield (αλληλεπιδράσεις ατόμων-ατόμων) διαβάζονται επίσης από το heat.namd. Τα παραγόμενα αρχεία (heat\_out και restart) καταγράφονται σε ένα κατάλογο, όπως δηλώνεται επίσης στο script του heat.namd. Από την διαδικασία του heating παράγεται επίσης ένα τροχιακό (.dcd), το οποίο μπορούμε να οπτικοποιήσουμε με την χρήση κατάλληλου προγράμματος (πχ με CARMA [108]).

Το τρίτο και τελευταίο στάδιο είναι αυτό της συλλογής των δεδομένων. Είναι το πλέον χρονοβόρο στάδιο της όλης διαδικασίας. Τα αρχεία που χρειάζεται το NAMD

είναι τα `ionized.pdb` και `ionized.psf` και τα παραγόμενα από την διαδικασία του heating (`heat_out.coor`, `heat_out.vel` `heat_out.xsc`, τα οποία δίνουν τις συντεταγμένες, τις ταχύτητες των ατόμων και τις διαστάσεις της στοιχειώδους κυψελίδας, αντίστοιχα). Χρειάζεται ακόμη το αρχείο παραμετροποίησης (`par_all22_prot.inp`) και το script `equi.namd`, στο οποίο τοποθετούμε τα ανωτέρω δεδομένα για να διαβαστούν από το NAMD. Μετά το τέλος του σταδίου αυτού παίρνουμε το συνολικό τροχιακό (`trajectory`), το οποίο δείχνει την τροχιά που διέγραψε κάθε άτομο του συστήματος που προσομοιώσαμε και σε όλη την διάρκεια του χρόνου προσομοίωσης, με την μορφή ενός (μεγάλου) αρχείου `.dcd`. Αυτό το δυαδικό αρχείο με το αντίστοιχό του αρχείο `.psf`, είναι αυτά τα οποία χρησιμοποιούμε στις αναλύσεις μας. Μπορούμε επίσης να μελετήσουμε το παραγόμενο αρχείο LOG για λήψη επιπλέον πληροφορίας όσο αφορά τον τρόπο με τον οποίο έγινε η διαδικασία της συλλογής των δεδομένων.

### **3.3. Προσομοίωση της 'folding'.**

#### **3.3.1. Προετοιμασία του συστήματος της 'folding'.**

Για την δημιουργία του συστήματος της προσομοίωσης '**folding**' αρχικά δημιουργήσαμε το ενδεκαπεπτιδίο αλληλουχίας INYWLAHAKAG με την χρήση του προγράμματος RIBOSOME [109] το οποίο παράγει εικονικά πεπτίδια και δημιουργήσαμε ένα αρχείο με το ενδεκαπεπτιδίο σε εκτεταμένη μορφή χωρίς τα υδρογόνα (`extended.rib`). Στην συνέχεια χρησιμοποιήσαμε το πρόγραμμα RIBOSOME και πήραμε το αντίστοιχο με το `.rib` αρχείο, το `extended.pdb`.

Στο αρχείο `extended.pdb` που λάβαμε, αντικαταστήσαμε την ιστιδίνη HSE (την default ιστιδίνη του RIBOSOME των 17 ατόμων) με HSP, επειδή η προσομοίωση γίνεται σε χαμηλό pH, όπου η ιστιδίνη είναι πλήρως πρωτονιωμένη. Από το τροποποιημένο αρχείο `extended.pdb` βλέπουμε ότι το μόριο αποτελείται από τα κατάλοιπα INYWLAHAKAG, ο συνολικός αριθμός των ατόμων του είναι 107 και η ιστιδίνη είναι στην μορφή HSP.

Στην συνέχεια χρησιμοποιήσαμε τον automatic psf builder του VMD (με δεδομένο το τροποποιημένο αρχείο `.pdb`) και πήραμε τρία νέα αρχεία (`.psf`, `.pdb` και

.log). Από τα νέα αρχεία βλέπουμε ότι τώρα έχουμε 177 άτομα (και όλα το υδρογόνα) στο πεπτιδίο των ένδεκα καταλοίπων, του οποίου η αλυσίδα ονομάστηκε P1.

Κατόπιν χρησιμοποιήσαμε και πάλι το VMD για την ενυδάτωση (solvation) του συστήματος. Με την εισαγωγή των νέων αρχείων .psf και .pdb, δημιουργήσαμε τα δύο αρχεία solvate.psf και solvate.pdb (extensions- modeling- add solvation box και μετά από rotate to minimize volume, boundary 1.8Å και διαστάσεις στοιχειώδους κυψελίδας 60\*60\*60 Å<sup>3</sup>). Από τα δύο solvate αρχεία βλέπουμε ότι έχουμε δύο segment: P1 του πεπτιδίου με 11 κατάλοιπα και 177 άτομα και WT1 του νερού με 20322 άτομα, με ένα σύνολο ατόμων 20499.

Ακολουθεί, πάλι με το VMD, η προσθήκη ιόντων για συγκέντρωση 100mM για την εξουδετέρωση του συνολικού φορτίου του συστήματος. Πήραμε το ionized.pdb και το ionized.psf. Τώρα στο σύστημά μας έχουμε συνολικά 20475 άτομα σε τρία segments. Το P1 του πεπτιδίου των 11 καταλοίπων με 177 άτομα, το WT1 του νερού με 20286 άτομα και το ION με 12 άτομα (189-177) από τα οποία τα 5 είναι νάτρια (SOD) και τα 7 χλώρια (CLA).

Ακολουθώντας την σειρά αυτή που αναφέρθηκε τελείωσε η προετοιμασία του συστήματος για την προσομοίωση της '**fold**ing'. Το προοριζόμενο για την προσομοίωση σύστημα (όπως δείχνουν τα αρχεία ionized.pdb και ionized.psf, των οποίων η δημιουργία ήταν και ο σκοπός της προετοιμασίας) με διαστάσεις της κυψελίδας είναι 60\*60\*60 Å<sup>3</sup> είναι έτοιμο για το επόμενο στάδιο.

### 3.3.2. Ελαχιστοποίηση της ενέργειας και σταδιακή θέρμανση της '**fold**ing'.

Για την πραγματοποίηση του σταδίου αυτού (energy minimization και heating με το πρόγραμμα NAMD) χρειάζονται τα αρχεία ionized.pdb, ionized.psf (τα οποία δημιουργήσαμε με την προετοιμασία του συστήματος), το par\_all22\_prot.inp, (το αρχείο με τις παραμέτρους του CHARMM forcefield) και ένα κατάλογο, όπου θα γράφονται τα παραγόμενα αρχεία, όπως καθορίζεται στο αρχείο διαμόρφωσης το heat.namd. Στο αρχείο αυτό αλλάζουμε κάποιες παραμέτρους (διαστάσεις κυψελίδας 59.5\*59.5\*59.5 Å<sup>3</sup>, θερμοκρασία: 320K). Στην συνέχεια βάζουμε το NAMD να τρέξει και από τα παραχθέντα αρχεία χρησιμοποιούμε τα heat\_out.coor, heat\_out.vel και heat\_out.xsc για το επόμενο στάδιο. Αυτά δίνουν τις συντεταγμένες των ατόμων, τις

ταχύτητες των ατόμων και τις διαστάσεις της κυψελίδας αντίστοιχα και καλούνται από το script του NAMD της φάσης παραγωγής (το equi.namd). Τα 135 frames του .dcd, που παρήχθησαν, μπορούμε να τα παρακολουθήσουμε με κάποιο πρόγραμμα οπτικοποίησης (carma, VMD).

### 3.3.3. Παραγωγή δεδομένων της 'folding'.

Τροποποιούμε το script equi.namd. Προσθέτουμε σε αυτό τα αρχεία, που δημιουργήθηκαν από το προηγούμενο βήμα του heating του συστήματος (heat\_out.coor, heat\_out.vel και heat\_out.xsc), ρυθμίζουμε τις συχνότητες εγγραφής των αρχείων που θα παράγονται, δηλώνουμε το βήμα προσομοίωσης (2fs) και βάζουμε την θερμοκρασία στους 320K. Χρειαζόμαστε τα ionized.pdb και ionized.psf, το αρχείο με τις παραμέτρους, το par\_all22\_prot.inp του CHARMM και έναν κενό κατάλογο για την εγγραφή των παραγόμενων αρχείων. Βάζουμε το πρόγραμμα NAMD να τρέξει σε 4 πυρήνες για 50000000 βήματα για την παραγωγή του τροχιακού. Κατά την διάρκεια της παραγωγής του .dcd βλέπουμε (με το carma) ότι το πεπτίδιο αναδιπλώθηκε αρκετά και ότι μπορούμε να μειώσουμε τα νερά του συστήματος για να προσομοιώνεται το σύστημα γρηγορότερα. Σταματάμε την προσομοίωση για να μειώσουμε τα νερά. Έχουν παραχθεί συνολικά 3545 frames στο .dcd αρχείο.

### 3.3.4. Μειώσεις νερών και στοχαστικά φαινόμενα.

Χρησιμοποιήσαμε το frame 3478 του equi\_out.dcd και δημιουργήσαμε το αντίστοιχο .pdb αρχείο με το πρόγραμμα carma. Επαναλάβαμε την διαδικασία της προετοιμασίας του συστήματος με αρχικό το 3478\_frame.pdb και χρησιμοποιήσαμε διαστάσεις κυψελίδας  $50*50*50 \text{ \AA}^3$ . Με το τέλος της προετοιμασίας πήραμε το νέο ionized.pdb και το αντίστοιχο του ionized.psf με 11678 άτομα, εκ των οποίων τα 177 του πεπτιδίου, 11493 νερά και 8 ιόντα (3 SOD και 5 CLA). Συνεχίσαμε με το heating και το production run έως ότου μία απρόβλεπτη διακοπή ρεύματος από την ΔΕΗ διέκοψε την ροή της προσομοίωσης. Είχαν ήδη παραχθεί 130604 frames.

Επειδή 'ουδέν κακόν αμιγές καλού', η απρόβλεπτη αυτή διακοπή ελήφθη σαν ευκαιρία για μια περαιτέρω μείωση των μορίων νερού του συστήματος, εφόσον το πεπτίδιο είχε ήδη αναδιπλωθεί πολύ. Μετατρέποντας το τελευταίο καταγραφέν frame του dcd σε .pdb και με χρήση πάλι του VMD, ετοιμάσαμε ένα νέο σύστημα.

Χρησιμοποιήσαμε διαστάσεις κουτιού  $40*40*40 \text{ \AA}^3$  και τα τελικά `ionized.pdb` και `ionized.psf`, που πήραμε αποτελούντο από 6352 άτομα: 177 του πεπτιδίου, 6171 του νερού και τέσσερα ιόντα (1SOD και 3CLA). Επαναλάβουμε την διαδικασία του heating και βάλουμε πάλι το NAMD να τρέξει. Αυτή ήταν η δεύτερη και τελευταία μείωση νερών του συστήματος της **'folding'**.

Κατά την διάρκεια των μηνών της προσομοίωσης ακολούθησαν τέσσερις ακόμα διακοπές ρεύματος με ισάριθμες διακοπές και των προσομοιώσεών μας. Στην τελευταία, αποφασίσαμε να σταματήσουμε την προσομοίωση για πάντα, εφόσον είχαμε συνολικά προσομοιώσει το σύστημα για 1.3 περίπου μs.

### **3.3.5. Το τροχιακό της 'folding'.**

Συνολικά πήραμε 7 `dcd` αρχεία από τις φάσεις παραγωγής: 3546 frames από την αρχική προσομοίωση των 20475 ατόμων, 13604 frames από αυτή των 11678 ατόμων και 445018, 24428, 189327, 375331 και 236563 frames από τις προσομοιώσεις των 6352 ατόμων. Για κάθε ένα από τα ανωτέρω `.dcd` αρχεία και με το αντίστοιχο `.psf` τους, χρησιμοποιώντας το πρόγραμμα CARMA παίρνουμε τα επτά `.dcd` αρχεία (με τα τροχιακά του πεπτιδίου μόνο) και τα αντίστοιχά τους `.psf` (και όχι μόνο). Αφού τα μετονομάσουμε, τα μεταφέρουμε σε έναν κατάλογο και τα συνενώνουμε με το πρόγραμμα `catdcd`. Το αρχείο `.dcd` των 1621038 frames που παράγεται από την συνένωση, μετά από ένα τελικό fitting και μία μετονομασία (σε `folding.dcd`) είναι το αρχείο του συνολικού τροχιακού του αρχικά εκτεταμένου ενδεκαπεπτιδίου, που θα χρησιμοποιηθεί για ανάλυση και αντιστοιχεί σε χρόνο προσομοίωσης 1.3 μs.

## **3.4. Προσομοίωση της 'native'.**

### **3.4.1. Προετοιμασία του συστήματος της 'native'.**

Το πρώτο βήμα για την προετοιμασία του συστήματος για την προσομοίωση της **'native'** είναι διαφορετικό. Στην περίπτωση αυτή χρησιμοποιήθηκε το μοντέλο 1 της καταχώρησης 2DX2 της PDB ως αρχικό αρχείο `.pdb`. Τα βήματα, αμέσως μετά, είναι τα ίδια: Χρησιμοποιώντας το VMD, τέθηκαν περιοδικές οριακές συνθήκες και

προστέθηκαν νερά και ιόντα στην ίδια τελική συγκέντρωση των 100mM. Τελικά παρήχθησαν τα αρχεία `ionized.pdb` και `ionized.psf` που θα χρησιμοποιηθούν για την προσομοίωση που ονομάζουμε '**native**'. Το `ionized.pdb` αποτελείται από 5890 άτομα. Από αυτά τα 177 ανήκουν στο πεπτίδιο (segid P1 με αλληλουχία των 11 καταλοίπων (INYWLAHAKAG), υπάρχει ένα ιόν νατρίου και τρία ιόντα χλωρίου (segment ION) και τα υπόλοιπα 5709 είναι νερά (segment WT1).

#### 3.4.2. Ελαχιστοποίηση της ενέργειας και σταδιακή θέρμανση της '**native**'.

Πήραμε τα αρχεία `ionized.pdb` και `ionized.psf` που δημιουργήσαμε καθώς και το αρχείο με τις παραμέτρους του CHARMM force field (`par_all22_prot.inp`). Δημιουργήσαμε και ένα κατάλογο για την εγγραφή των δεδομένων και μετατρέψαμε το αρχείο διαμόρφωσης `heat.namd`, θέτοντας θερμοκρασία-στόχο 283K και διαστάσεις κυψελίδας  $40*40*40 \text{ \AA}^3$ . Στην συνέχεια χρησιμοποιήσαμε το NAMD βάζοντας το πρόγραμμα να τρέξει σε 4 πυρήνες. Μετά το τέλος του προγράμματος πήραμε το αρχείο του τροχιακού (`.dcd`) με 130 frames, τα αρχεία `heat_out.coor`, `heat_out.vel` και `heat_out.xsc`, που θα χρησιμοποιηθούν στο επόμενο στάδιο καθώς και τα αρχεία `restart`.

#### 3.4.3. Παραγωγή δεδομένων της '**native**'.

Χρειαζόμαστε τα αρχεία `ionized.pdb`, `ionized.psf`, `heat_out.coor`, `heat_out.vel`, `heat_out.xsc`, `par_all22_prot.inp`, `equi.namd`. Δημιουργούμε και τον κατάλογο για την εγγραφή των παραγόμενων αρχείων. Χρησιμοποιούμε το NAMD για την παραγωγή του τροχιακού. Στην περίπτωση της παραγωγής τροχιακού της '**native**' μόνο μία διακοπή της προσομοίωσης συνέβη εξαιτίας διακοπής ρεύματος. Δεν έγινε καμία μείωση νερών. Η πρώτη προσομοίωση παρήγαγε αρχείο `.dcd` με 245492 frames και η δεύτερη με 260837 frames.

#### 3.4.4. Το τροχιακό της '**native**'.

Όπως και στην περίπτωση της '**foldng**' συγχωνεύσαμε τα `.dcd` για να πάρουμε το συνολικό τροχιακό της προσομοίωσης. Στην περίπτωση, όμως, της '**native**' χρησιμοποιήσαμε και το αρχικό τροχιακό, που προήλθε από το heating. Έτσι είχαμε το `heat_out.dcd` των 130 frames, το `equi_out.dcd` της πρώτης production run των 245492 frames και το `equi_out.dcd` της δεύτερης των 260837 frames. Αφού αφαιρέσαμε τα

νερά και τα ιόντα από το καθένα .dcd με το carma, τα μετονομάσαμε, τα μεταφέραμε στο ίδιο directory και τα συνενώσαμε με catdcd. Παρήχθη το αρχείο .dcd με 506452 frames, που μετά από ένα fitting το ονομάσαμε native.dcd. Αυτό αντιστοιχεί σε χρόνο προσομοίωσης 0.4 μs. Το σχετικό .psf των 177 ατόμων της '**native**' το μετονομάσαμε σε rept.psf.

# Κεφάλαιο 4<sup>ο</sup>

## Ανάλυση προσομοιώσεων.

*It is a pollster's maxim that the truth lies not in any one poll  
but at the center of gravity of several polls.  
Michael R. Kagay, New York Times.*

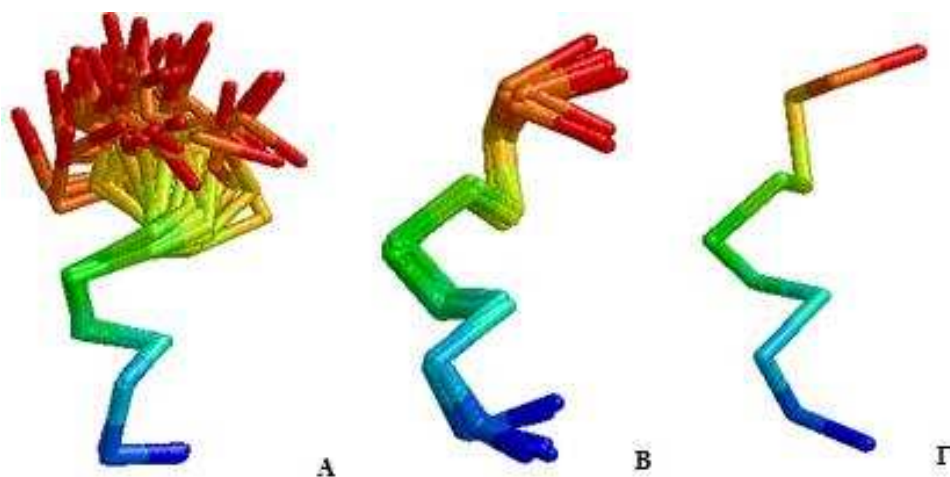
Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, η εργασία αυτή επικεντρώνεται στην ανάλυση δύο προσομοιώσεων μοριακής δυναμικής του ενδεκαπεπτιδίου INYWLAHAKAG. Η αναφερόμενη ως '**native**' προσομοίωση των 283K με αρχική δομή το μοντέλο 1 της καταχώρησης της PDB με κωδικό 2DX2 παρήγαγε 506452 frames και διήρκεσε 0.4  $\mu$ s. Η προσομοίωση '**fold**' των 320K με αρχική δομή την εκτεταμένη μορφή του πεπτιδίου, η οποία δημιουργήθηκε με την



βοήθεια του προγράμματος RIBOSOME, παρήγαγε 1621038 frames και η διάρκειά της ήταν 1.3 μs.

Η παρουσίαση της ανάλυσης των τροχιακών στο παρόν κεφάλαιο γίνεται με την εξής σειρά:

- Απεικόνιση στιγμιότυπων των δύο τροχιακών, σε τακτά χρονικά διαστήματα.
- Ανάλυση της προσομοίωσης 'native'.
- Ανάλυση της προσομοίωσης '*fold*ing'.
- Σύγκριση των δύο προσομοιώσεων ('native' και '*fold*ing') αρχικά μεταξύ τους και στην συνέχεια με τα πειραματικά προσδιορισθέντα με NMR δομικά μοντέλα, τόσο της 2DX2, όσο και της στενά συσχετιζόμενης δομής, της 1CB3 (αλληλουχία IDYWLAHKALA). Και οι δύο αναφερόμενες καταχωρήσεις της PDB (Εικόνα 4.1) έχουν δομή 3<sub>10</sub>-έλικας.



**Εικόνα 4.1:** Απεικόνιση με *rasmol* των δομών των δύο καταχωρήσεων της PDB. Το N-τελικό άκρο απεικονίζεται στο κάτω τμήμα της εικόνας. A: Τα 30 μοντέλα της 1CB3. B: Τα 10 μοντέλα της 2DX2. Γ: Το αρχικό μοντέλο της 'native' προσομοίωσης (μοντέλο 1 της 2DX2).

Από τα αποτελέσματα που πήραμε από τις αναλύσεις των τροχιακών και των δύο προσομοιώσεων παρατηρήθηκε ότι:

- Το τμήμα των καταλοίπων 2-7 παρουσιάζει αξιοσημείωτη σταθερότητα σε αντίθεση με το C-τελικό άκρο του ενδεκαπεπτιδίου, το οποίο φαίνεται ιδιαίτερα

υπερκινητικό. Η παρατήρηση αυτή συνάδει με τα πειραματικά δεδομένα που προέκυψαν κατά τον προσδιορισμό των δομών των 2DX2 και 1CB3.

- Το σταθερό τμήμα των επτά πρώτων καταλοίπων του πεπτιδίου των προσομοιώσεων έχει μεν ελικοειδή δομή, αλλά η δομή αυτή είναι α-έλικα και όχι  $3_{10}$ , όπως αυτή των πειραματικών μοντέλων.
- Στα τροχιακά παρατηρούμε μικρής διάρκειας δομές  $3_{10}$ -έλικας σε διαφορετικά χρονικά διαστήματα των προσομοιώσεων. Οι έλικες όμως αυτές διαλύονται γρήγορα και αντικαθιστώνται από την σταθερή (σε ότι αφορά το forcefield) δομή της α-έλικας. Το γεγονός αυτό, συμφωνεί με δημοσιευμένες μελέτες περί της  $3_{10}$  ως ενδιάμεσου της α-έλικας [73,74]. Η σταθερότητα, όμως, της α-έλικας των προσομοιώσεων σε ποσοστό μεγαλύτερο του 50%, αποκλίνει σημαντικά από το ποσοστό του περίπου 20% της α-έλικας, το οποίο βρέθηκε ως δομικός πληθυσμός στο πεπτίδιο, κατά την διάρκεια της πειραματικού καθορισμού της δομής του με NMR [46].

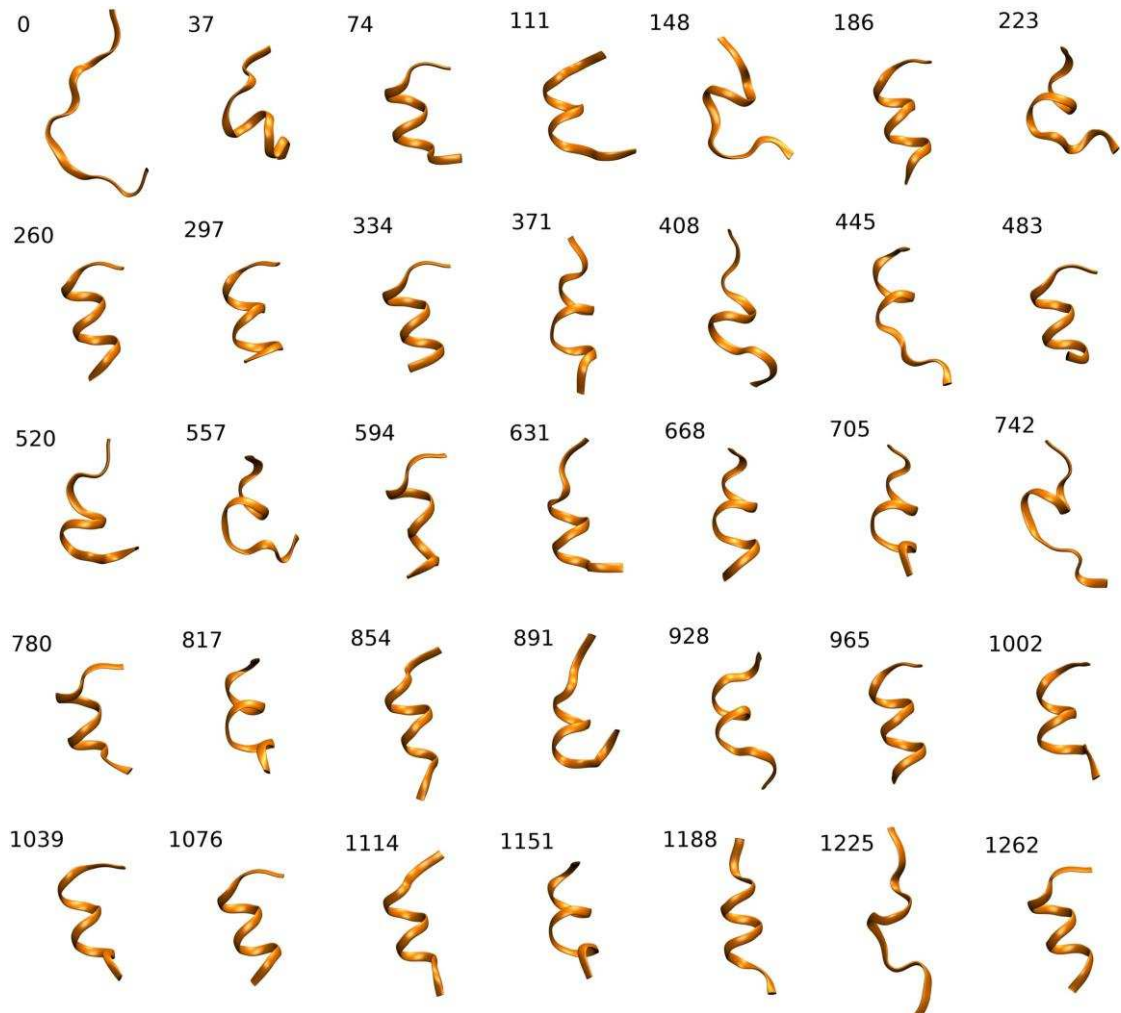
Συμπεράσματα, τα οποία εξήχθησαν είναι:

- Δεν μπόρεσε το CHARMM forcefield να αναπαράγει την  $3_{10}$ -έλικα, ξεκινώντας με αρχική δομή την εκτεταμένη μορφή του πεπτιδίου.
- Δεν κατάφερε το CHARMM forcefield να διατηρήσει την δομή της  $3_{10}$ -έλικας, χρησιμοποιώντας ως αρχική δομή την  $3_{10}$ -έλικα της καταχώρησης 2DX2.
- Το CHARMM forcefield έδειξε για μία ακόμα φορά την τάση του για δημιουργία α-έλικας και την (αφύσικη) σταθεροποίησή της.
- Τόσο η κινητικότητα του C-τελικού άκρου όσο και η σταθερότητα της διαμόρφωσης του N-τελικού, έχουν αναπαραχθεί από το ανωτέρω δυναμικό πεδίο, σε συμφωνία με τα πειραματικά δεδομένα.
- Η αλληλεπίδραση των πλευρικών ομάδων των καταλοίπων Tyr3, Trp4 και His7 είναι εμφανής σε σημαντικό ποσοστό του κυρίαρχου cluster του τροχιακού και σε συμφωνία με τα πειραματικά δεδομένα της 1CB3 καταχώρησης.
- Υπήρξε σημαντική σύγκλιση των δύο προσομοιώσεών μας.

Αλλά καλύτερα να πάρουμε την ιστορία από την αρχή...

#### 4.1. Στιγμιότυπα διαδοχικών δομών των δύο τροχιακών.

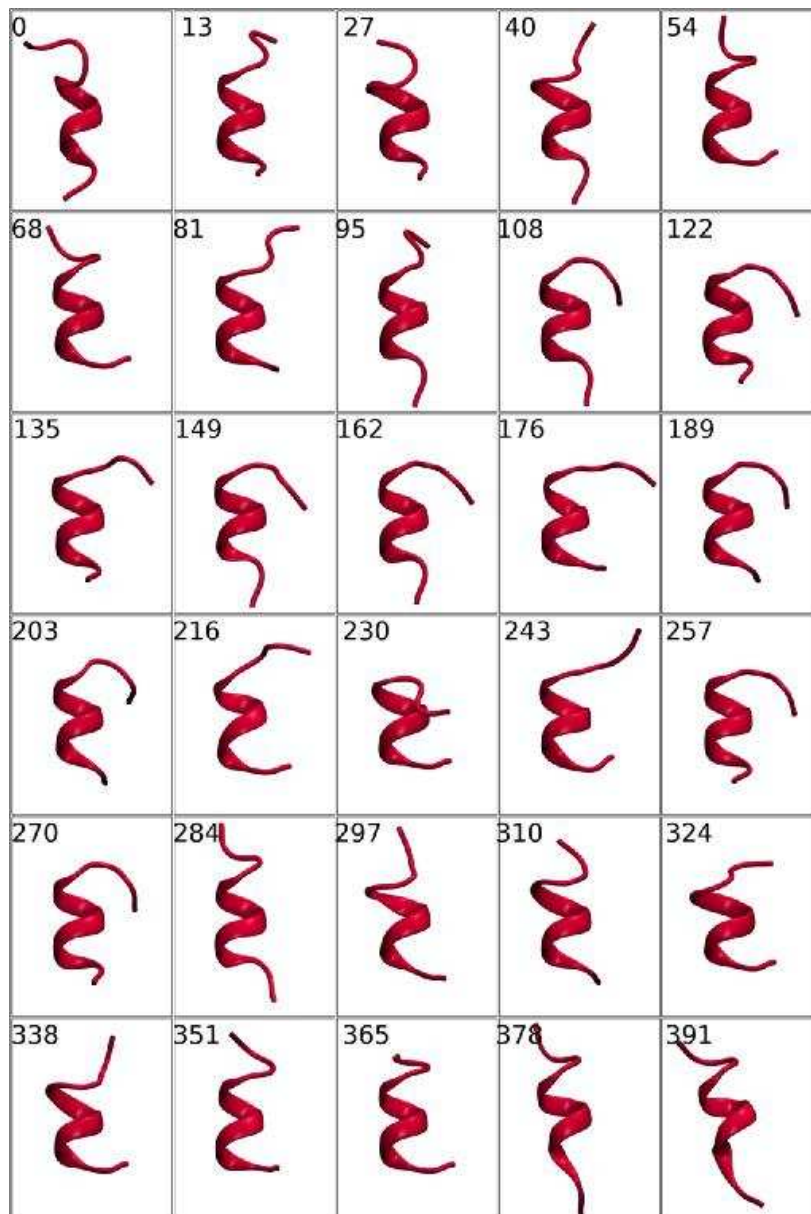
Ως πρώτη πληροφορία περί της ποικιλίας (ή όχι) των δομών από τις οποίες έχει περάσει το πεπτίδιο κατά την διάρκεια του χρόνου των προσομοιώσεων, παίρνουμε τα στιγμιότυπα των διαδοχικών δομών των Εικόνων 4.2 και 4.3.



**Εικόνα 4.2:** Στιγμιότυπα από την προσομοίωση *'folding'*. Οι διαδοχικές δομές απέχουν μεταξύ τους 37.1 ns. Το N-τελικό άκρο του πεπτιδίου βρίσκεται επάνω. Το πρώτο frame (t=0) είναι το τελευταίο του heating.

Η Εικόνα 4.2 απεικονίζει 35 χρονικά ίσης απόστασης στιγμιότυπα (καταγεγραμμένα απευθείας από το τροχιακό μοριακής δυναμικής) της προσομοίωσης της *'folding'*. Όπως είναι εμφανές από αυτήν την εικόνα, κατά τη διάρκεια των περίπου ~1.3 μs της προσομοίωσης, το πεπτίδιο περνάει από διαδοχικούς κύκλους αναδίπλωσης και αποδιάταξης. Για παράδειγμα, από την αρχική εκτεταμένη μορφή, το πεπτίδιο

περνάει σε μια σχεδόν αποκλειστικά ελικοειδή δομή μέσα σε μόλις δυο στιγμιότυπα (στη δομή που αντιστοιχεί στα 74ns). Αυτή η ελικοειδής δομή αποδιατάσσεται στα 148 ns (με βάση πάντα τα στιγμιότυπα) και επανεμφανίζεται στα 186 ns, ακολουθεί δε αντιστοίχους κύκλους αποδιάταξης και αναδιάταξης στα 408, 742 και 1225 ns. Δεδομένου ότι πρόκειται για ένα ενδεκακαπεπτίδιο και ότι αυτή η προσομοίωση πραγματοποιήθηκε στους 320K (47 βαθμούς Κελσίου), το πεπτιδίο φαίνεται μάλλον αξιοσημείωτα σταθερό.



**Εικόνα 4.3:** Στιγμιότυπα από την προσομοίωση 'native'. Οι διαδοχικές δομές απέχουν μεταξύ τους 13.5 ns. Το C-τελικό άκρο του πεπτιδίου βρίσκεται επάνω.

Στην Εικόνα 4.3 βλέπουμε 30 χρονικά ισαπέχοντα κατά 13.5 ns στιγμιότυπα, τα οποία έχουν καταγραφεί από το τροχιακό μοριακής δυναμικής της προσομοίωσης **'native'**. Στην περίπτωση αυτή γνωρίζουμε ότι το πρώτο στιγμιότυπο απεικονίζει την αρχική 3<sub>10</sub>-έλικα. Παρατηρούμε ότι η ελικοειδής διαμόρφωση των καταλοίπων 2 έως 7 παραμένει καθόλη τη διάρκεια της προσομοίωσης, αλλά φαίνεται να χάνεται η διαμόρφωση της 3<sub>10</sub>-έλικας αμέσως μετά το πρώτο στιγμιότυπο. Εμφανής είναι και η κινητικότητα του C-τελικού άκρου του πεπτιδίου.

Από τα στιγμιότυπα και των δύο προσομοιώσεων, γίνεται εμφανής η σταθερότητα της δομής του ενδεκαπεπτιδίου στο τμήμα των καταλοίπων 2-7 και η κινητικότητα του C-τελικού τμήματος των τεσσάρων καταλοίπων. Όσον αφορά το κατάλοιπο 1, αυτό δείχνει να έχει σταθερότερη διαμόρφωση στα στιγμιότυπα της **'folding'**, λαμβάνοντας υπόψη τόσο την διαφορά θερμοκρασίας, όσο και την διάρκεια της προσομοίωσης.

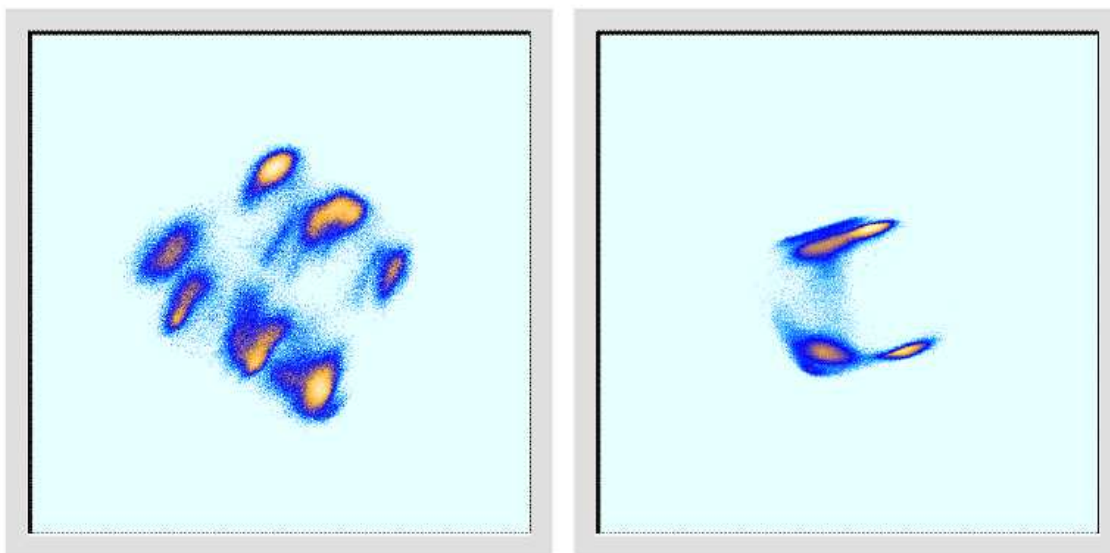
Με τις εικόνες από τα στιγμιότυπα παίρνουμε μία γενική ιδέα για τις δομές, που έχει λάβει το πεπτίδιο, αλλά κανένα ασφαλές συμπέρασμα δεν μπορεί να εξαχθεί μόνο από αυτές. Εμείς έχουμε ως δεδομένα τα αρχεία των δύο τροχιακών (της **'native'** και της **'folding'**). Κάθε ένα από αυτά περιέχει την πληροφορία για τις συντεταγμένες του καθενός ατόμου από τα 177 σε 506452 χρονικές στιγμές για την **'native'** και σε 1621038 χρονικές στιγμές για την **'folding'**. Για την συλλογή ουσιαστικής πληροφορίας και άντληση δεδομένων, τα οποία να μπορούμε να ερμηνεύσουμε, απαιτείται ομαδοποίηση των δεδομένων και σειρά αναλύσεων, η οποία ακολουθεί.

## 4.2. Ανάλυση του τροχιακού της **'native'**.

### 4.2.1. Ανάλυση σε κύριες συνιστώσες με χρήση εσωτερικών συντεταγμένων.

Αρχικά ομαδοποιούμε τα δεδομένα του τροχιακού της **'native'** εκτελώντας ανάλυση σε κύριες συνιστώσες και στον χώρο των διεδρων γωνιών (dPCA), με την χρήση του προγράμματος CARMA. Πραγματοποιούμε δύο τέτοιες αναλύσεις. Η μία ανάλυση αφορά τις διέδρες του συνόλου των καταλοίπων του ενδεκαπεπτιδίου. Η δεύτερη έχει γίνει με χρήση των διεδρων των καταλοίπων 1 έως και 7, για τα οποία

αφενός έχουμε παρατηρήσει εκ πρώτης όψεως (από την Εικόνα 4.3) μια σταθερότητα διαμόρφωσης και αφετέρου η σταθερότητα αυτή έχει φανεί και πειραματικά [46,47]. Από την ανάλυση αυτή σε ομάδες (cluster analysis) παίρνουμε την Εικόνα 4.4, που ακολουθεί:



**Εικόνα 4.4:** Εικόνες από την dihedral PCA. Αριστερά: Εννέα cluster της '**native**' με χρήση όλων των διέδρων. Δεξιά: Τέσσερα cluster της '**native**' όταν χρησιμοποιηθούν τα κατάλοιπα 1-7.

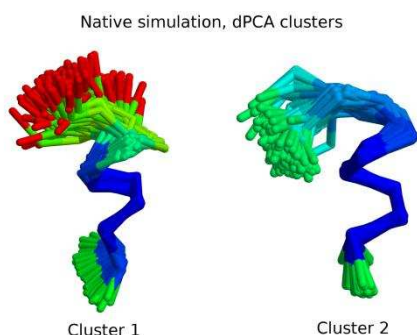
Από τα δεδομένα της ανάλυσης και από την Εικόνα 4.4, βλέπουμε ότι με την ομαδοποίηση που έκανε το CARMA, δημιουργήθηκαν εννέα cluster με χρήση όλων των διέδρων και τέσσερα cluster με χρήση των διέδρων των καταλοίπων 1-7.

Από τα 135767 frames, τα οποία το carma απέδωσε σε cluster (ποσοστό 27% του συνολικού τροχιακού), χρησιμοποιώντας όλες τις διέδρες, το 31.4% (42633 frames) ανήκουν στο κυρίαρχο cluster και 28.6% (38797 frames) στο αμέσως επόμενο.

Από τα 105434 frames των cluster, με χρήση των διέδρων των καταλοίπων 1-7 (21% του τροχιακού), το 82% (86562 frames) ανήκουν στο κυρίαρχο cluster.

Η ανωτέρω ομαδοποίηση είναι καλή ένδειξη για την σταθερότητα του σκελετού του πεπτιδίου όσο αφορά τα πρώτα επτά κατάλοιπων, ενώ φαίνεται η ευμεταβλητότητα των επόμενων τεσσάρων καταλοίπων.

Για να οπτικοποιήσουμε τα δομικά μοντέλα, τα οποία ομαδοποιήθηκαν στα δύο μεγαλύτερα cluster στην περίπτωση χρήσης όλων των διεδρων, κάνουμε υπέρθεση των δομών των δύο μεγαλύτερων cluster και με χρήση σταθερής χρωματικής κλίμακας (μπλε για μικρά rmsf και κόκκινο για μεγάλα), παίρνουμε την Εικόνα 4.5.



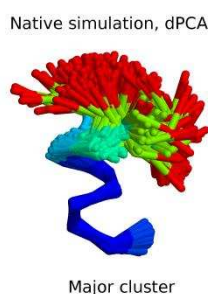
**Εικόνα 4.5:**

Υπερθέσεις των δομών των δύο cluster, που αποδόθηκαν από το carma με την ανάλυση σε κύριες συνιστώσες και με χρήση όλων των διεδρων γωνιών.

Αριστερά: Το κυρίαρχο cluster της '**native**'.  
Δεξιά: Το δεύτερο cluster της '**native**'.

Βλέπουμε το κυρίαρχο cluster με ελικοειδή την διαμόρφωση των καταλοίπων 2-7, ενώ στο δεύτερο cluster παρατηρούμε το αυξημένο μήκος της έλικας, το οποίο επεκτείνεται κατά ένα κατάλοιπο προς την C-τελική περιοχή. Παρατηρώντας και πάλι τα στιγμιότυπα της Εικόνας 4.3, γίνεται αντιληπτή αυτή η μεγαλύτερου μήκους έλικα σε κάποια από αυτά. Καμία από τις ανωτέρω έλικες δεν φαίνεται να έχει χαρακτηριστικά  $3_{10}$ -έλικας.

Όσο για τις ομαδοποιημένες δομές του κυρίαρχου cluster στην περίπτωση χρήσης των διεδρων των καταλοίπων 1-7, αυτές απεικονίζονται στην Εικόνα 4.6.



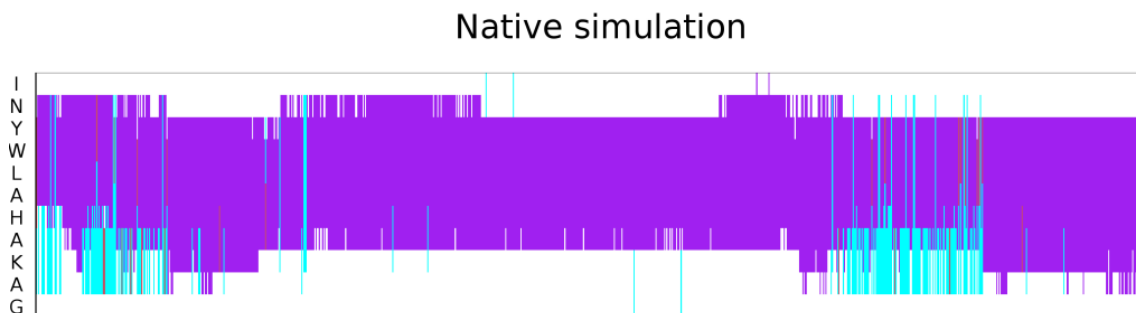
**Εικόνα 4.6:**

Υπερθέσεις των δομών του κυρίαρχου (80%) cluster της '**native**', μετά την ανάλυση σε κύριες συνιστώσες με την χρήση του προγράμματος carma και με χρήση μόνο των διεδρων των καταλοίπων 1 έως και 7 του πεπτιδίου.

Το κυρίαρχο cluster είναι ελικοειδές στο N-τελικό άκρο του, περιλαμβάνοντας στην έλικα τα κατάλοιπα 2-7. Και αυτή η έλικα δεν μοιάζει για  $3_{10}$ .

#### 4.2.2. Ανάθεση δευτεροταγούς δομής – κατανομή των cluster.

Χρησιμοποιούμε το πρόγραμμα ανάθεσης δευτεροταγούς δομής STRIDE [110] για να δούμε τι είδους δευτεροταγή δομή αναθέτει αυτό, στα κατάλοιπα του πεπτιδίου κατά την διάρκεια της προσομοίωσης 'native'. Στην επόμενη Εικόνα 4.7, οπτικοποιείται χρωματικά η ανάθεση αυτή.



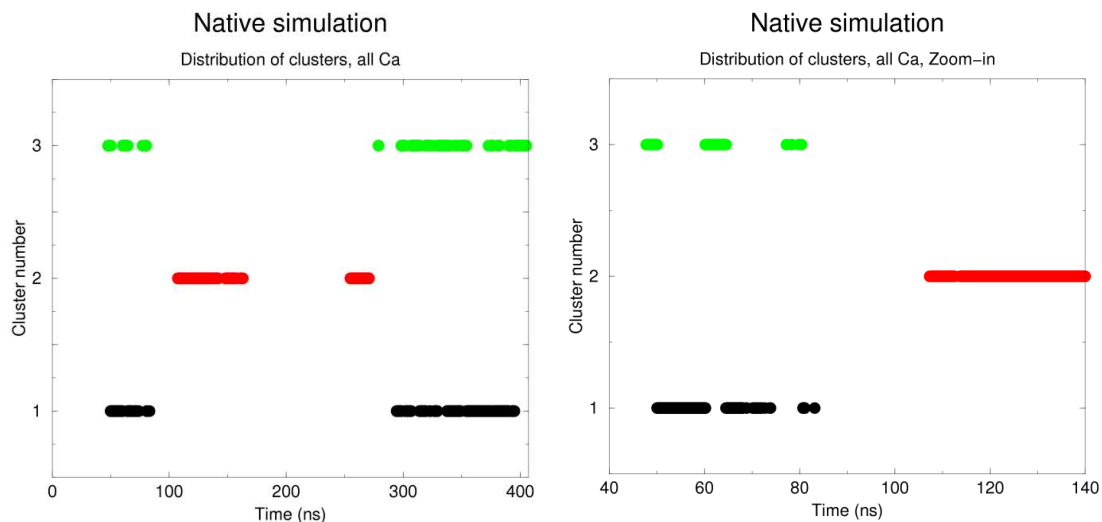
**Εικόνα 4.7:** Χρωματική απεικόνιση της δευτεροταγούς δομής, την οποία αναθέτει το πρόγραμμα STRIDE, με βάση το τροχιακό της προσομοίωσης 'native'.

Από την ανωτέρω εικόνα βλέπουμε ότι οι καταγεγραμμένες δομές της προσομοίωσης της 'native' κυριαρχούνται από α-έλικια (μωβ) στη N-τελική περιοχή σε όλο σχεδόν το χρόνο της προσομοίωσης. Το C-τελικό άκρο σπάζει την μονοτονία της σταθερότητας της N-τελικής περιοχής, ευρισκόμενο για μεγάλο χρονικό διάστημα, είτε σε κατάσταση random coil (λευκό), είτε σε διαμόρφωση στροφής (σιέλ), λαμβάνει δε διαμόρφωση α-έλικας (μωβ) για μικρά χρονικά διαστήματα, τα οποία, όμως, δεν είναι ικανοποιητικού χρόνου (ή ικανοποιητικού αριθμού frames) για να ενταχθούν σε cluster από το carma, όπως είδαμε στην περίπτωση της dPCA με χρήση όλων των διέδρων γωνιών (cluster 2, Εικόνα 4.5). Σποραδικά, η περιοχή των πρώτων 6-7 καταλοίπων παίρνει πολύ μικρής διάρκειας διαμόρφωση  $3_{10}$ -έλικας (κόκκινο – δεν είναι αρκούντως ορατό). Φαίνεται ότι η προσομοίωση δεν διατήρησε την αρχική δομή της  $3_{10}$ -έλικας..

Στην Εικόνα 4.8 βλέπουμε την κατανομή των τριών πρώτων cluster της 'native', με χρήση όλων των διέδρων γωνιών. Η κατανομή των cluster δίνει πληροφορία για τις συχνότητες επανεμφάνισης των ομαδοποιημένων δομών καθόλη την χρονική διάρκεια της προσομοίωσης. Στην συγκεκριμένη περίπτωση μπορεί να μας δώσει επιπλέον την πληροφορία για την παρουσία ή όχι της δομής της  $3_{10}$ -έλικας, με δεδομένο ότι το αρχικό μοντέλο της προσομοίωσης αυτής ήταν ακριβώς αυτός ο τύπος έλικας. Στην



αριστερή πλευρά της Εικόνας 4.8, απεικονίζεται γραφικά η παρουσία του κάθε cluster στην χρονική διάρκεια των 0.4 μs της προσομοίωσης και στην δεξιά πλευρά μία μεγέθυνση ενός μικρού τμήματος της δεξιάς εικόνας, η οποία αντιστοιχεί στο διάστημα των 40 έως 140 ns του χρόνου προσομοίωσης. Η μεγέθυνση αυτή δείχνει ότι δεν υπάρχει αλληλοεπικάλυψη των cluster, όπως εκ πρώτης όψης μπορεί να υποθέσει κανείς, αλλά γρήγορη μεταφορά της δομής από την μία ομάδα δομών στην άλλη.



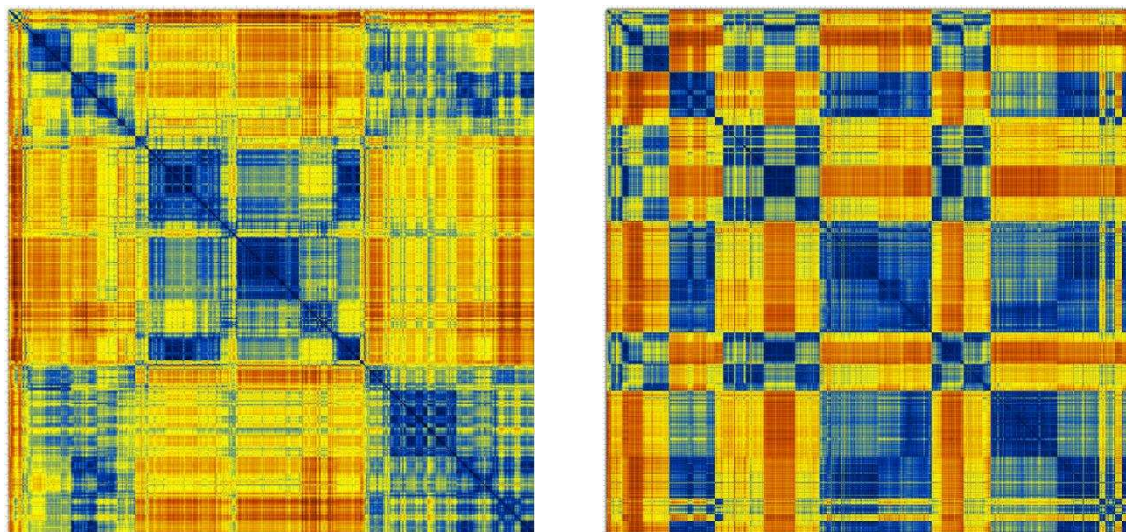
**Εικόνα 4.8:** Αριστερά: η παρουσία του καθενός από τα τρία πρώτα cluster της **'native'** σε σχέση με τον συνολικό χρόνο προσομοίωσης. Δεξιά: μία μεγέθυνση της περιοχής των 40-140 ns.

Παρατηρούμε ότι η περιοχή, η οποία ανήκει στον αρχικό χρόνο της προσομοίωσης δεν ανήκει σε κάποιο από αυτά τα μεγάλα cluster. Δεδομένης της αρχικής δομής 3<sub>10</sub>-έλικας, που χρησιμοποιήσαμε, έχουμε πλέον σαφείς ενδείξεις για την αποτυχία της προσομοίωσης της **'native'** στην αναπαραγωγή της σπάνιας αυτής έλικας.

#### 4.2.3. Πίνακες RMSD και ανάλυσή τους.

Η Εικόνα 4.9 είναι μια έγχρωμη αναπαράσταση των τιμών των RMSD μεταξύ όλων των πιθανών συνδυασμών των δομών του πεπτιδίου και με χρήση μόνο των Ca ατόμων, όπως έχουν καταγραφεί κατά τη διάρκεια της προσομοίωσης **'native'**. Το εύρος των παρατηρούμενων τιμών απεικονίζεται χρωματικά, με τις μικρές τιμές RMSD (σε Å) να αντιστοιχούν σε χρώμα σκούρου μπλε, τις μεγαλύτερες σε χρώμα σκούρου κόκκινου και τις ενδιάμεσες (από μικρότερες σε μεγαλύτερες τιμές) σε πράσινο, κίτρινο και πορτοκαλί. Οι τιμές της κύριας διαγωνίου είναι μηδενικές, εφόσον αντιστοιχούν σε

RMSD ανάμεσα σε κάθε δομή με τον εαυτό της, το δε χρώμα της διαγωνίου αυτής απεικονίζεται ως μαύρο.



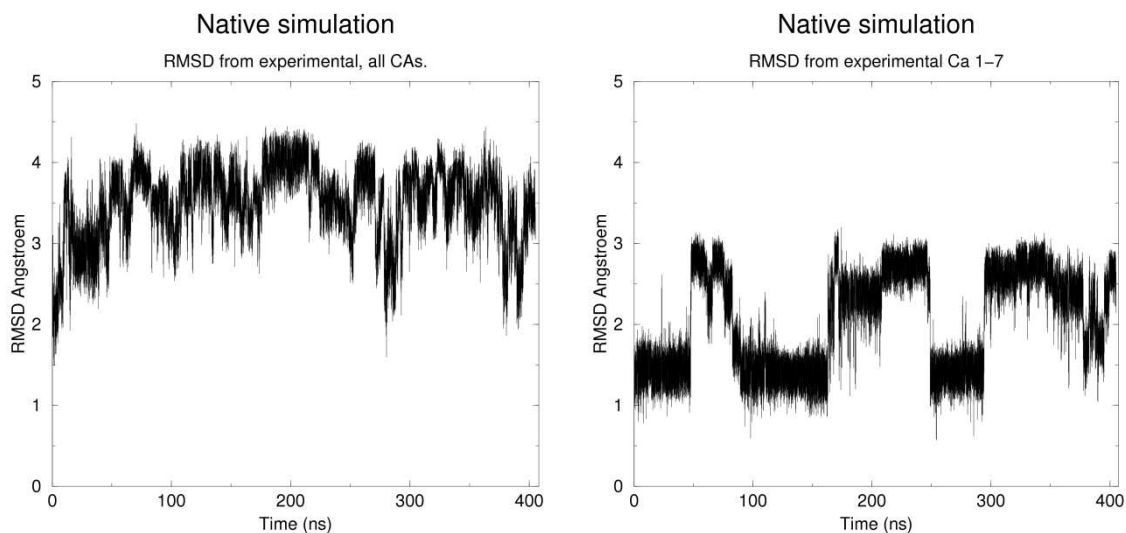
**Εικόνα 4.9:** Πίνακες RMSD με χρήση των συντεταγμένων των Ca ατόμων της 'native'. Αριστερά: Πίνακας με χρήση όλων των Ca ατόμων και μέγιστη τιμή 5.32 Å. Δεξιά: Πίνακας με χρήση των Ca των καταλοίπων 1-8. Το εύρος των RMSD κυμαίνεται από 0.00 Å έως 2.93 Å.

Στον αριστερό πίνακα παρατηρούμε κάποια επαναλαμβανόμενα σκούρα μπλε τμήματα, που αντιστοιχούν, όπως έχουμε αναφέρει, σε μικρά RMSD, λίγα εκ των οποίων συνδέονται τόσο οριζόντια όσο και κάθετα με μπλε τμήματα. Αυτό δείχνει ότι το πεπτίδιο σχηματίζει ποικιλία δομών σε διαφορετικές φάσεις της προσομοίωσης. Χρησιμοποιώντας την βοήθεια της Εικόνας 4.8, βλέπουμε την κατανομή του κυρίαρχου cluster στον χρόνο από 50 έως 100 ns και από 300ns έως τέλους χρόνου και του δεύτερου cluster στην κεντρική περιοχή των 100-300ns.

Λόγω της ευκινησίας του C-τελικού άκρου, η οποία έχει παρατηρηθεί (από τις Εικόνες 4.3, 4.5, 4.6 και από την ύπαρξη του ενός κυρίαρχου (80%) cluster της dPCA ανάλυσης με τις διέδρες των καταλοίπων 1-7), υπολογίσαμε τον ίδιο πίνακα RMSD, χρησιμοποιώντας αυτή τη φορά τις συντεταγμένες του σκελετού μόνο των οκτώ N-τελικών καταλοίπων. Στον πίνακα αυτόν, ο οποίος απεικονίζεται στην δεξιά πλευρά της Εικόνας 4.9 φαίνεται ότι, μετά την εξαίρεση των C-τελικών καταλοίπων, ο πίνακας κυριαρχείται από μικρές τιμές RMSD (σκούρο μπλε και με μέγιστο RMSD τα 2.93 Å.

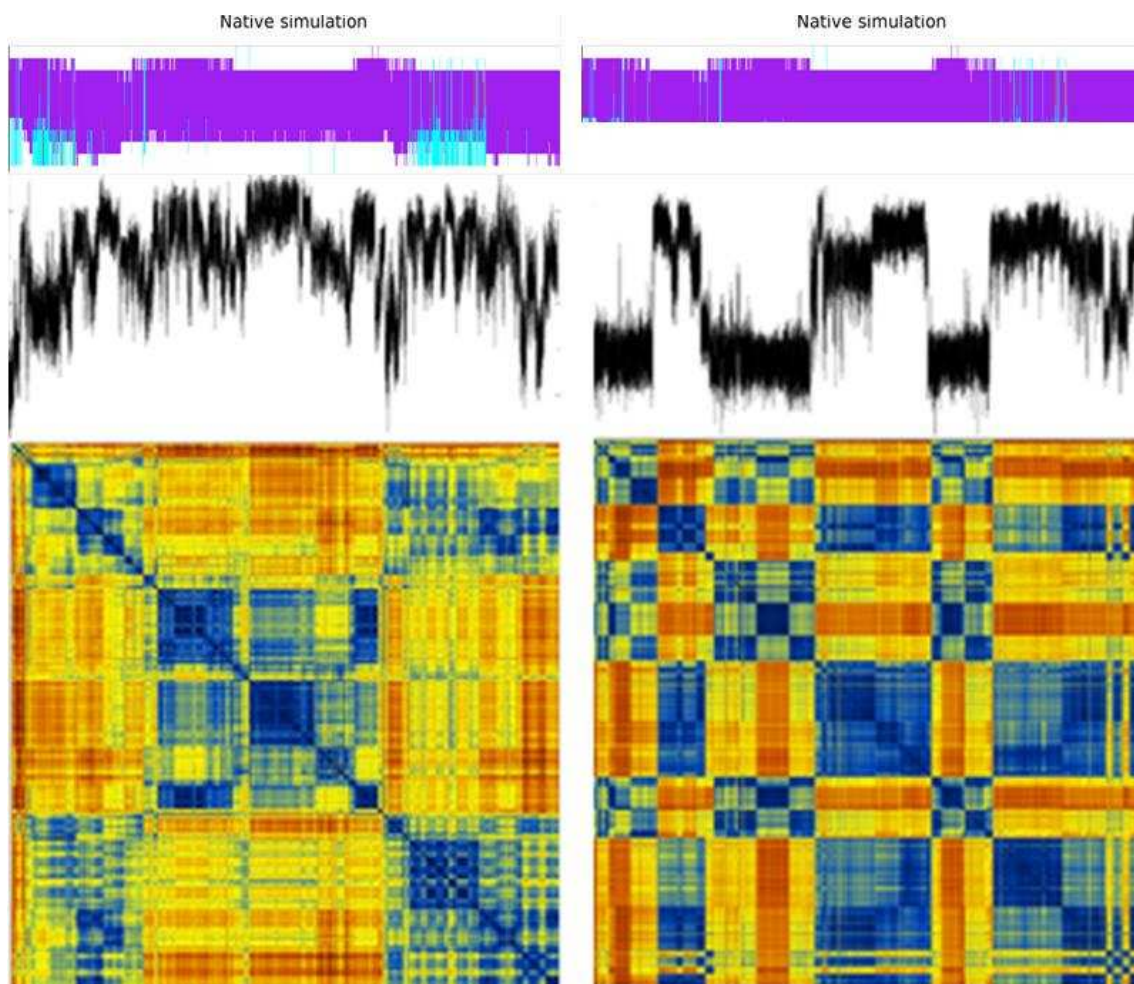
#### 4.2.4. RMSD από την πειραματικά προσδιορισμένη δομή.

Στην Εικόνα 4.10 βλέπουμε την γραφική παράσταση των RMSD των Ca ατόμων ως προς την πειραματικά προσδιορισμένη με NMR δομή (μοντέλο 1 της 2DX2), η οποία τυχαίνει να είναι και το πρώτο frame της παρούσας προσομοίωσης.



**Εικόνα 4.10:** Γραφική παράσταση των RMSD (Å) από την πειραματικά προσδιορισμένη δομή ως προς τον συνολικό χρόνο προσομοίωσης (ns). Αριστερά: Με όλα τα Ca άτομα. Δεξιά: Με τα Ca των καταλοίπων 1-7.

Παρατηρούμε από την εικόνα αριστερά, ότι από την αρχή ακόμα της προσομοίωσης και σχεδόν σε όλη την διάρκειά της τα RMSD από την αρχική δομή είναι μεγάλα, κυμαινόμενα στο διάστημα των 2 με 4 Å, όταν χρησιμοποιούμε όλα τα Ca άτομα. Αυτό δείχνει ότι η αρχική δομή της  $3_{10}$ -έλικας, πολύ γρήγορα μετά την αρχή της προσομοίωσης, αλλάζει δομή και δεν επανέρχεται σε αυτή σε όλο το επόμενο χρονικό διάστημα. Έχοντας, πλέον, ως δεδομένη την κυρίαρχη δομή της  $\alpha$ -έλικας της 'native' (Εικόνα 4.7) συμπεραίνουμε (και πάλι) ότι το δυναμικό πεδίο δεν κατάφερε να διατηρήσει την αρχική  $3_{10}$ -έλικα. Μετά την εξαίρεση των τελευταίων τεσσάρων καταλοίπων, όπως φαίνεται δεξιά στην Εικόνα 4.10, τα RMSD μειώνονται σε τιμές οι οποίες κυμαίνονται από 1 έως 3 Å. Η μείωση αυτή δείχνει την κινητικότητα της C-τελικής περιοχής, η οποία όταν εξαιρεθεί, έχει ως αποτέλεσμα την μείωση της ποικιλίας των διαμορφώσεων του πεπτιδίου. Με σύνθεση των Εικόνων 4.7, 4.9 και 4.10, έχουμε ακόμα καλύτερη απεικόνιση της κατάστασης (Εικόνα 4.11).



**Εικόνα 4.11:** Σύνθεση εικόνων 4.7, 4.9 και 4.10. Αριστερά: Στήλη διαγραμμάτων που αφορούν όλα τα κατάλοιπα (από επάνω προς κάτω συμμετέχουν: όλα τα άτομα, τα C $\alpha$ , τα άτομα του σκελετού). Δεξιά: Αντίστοιχη στήλη που αφορά τα κατάλοιπα 1-7.

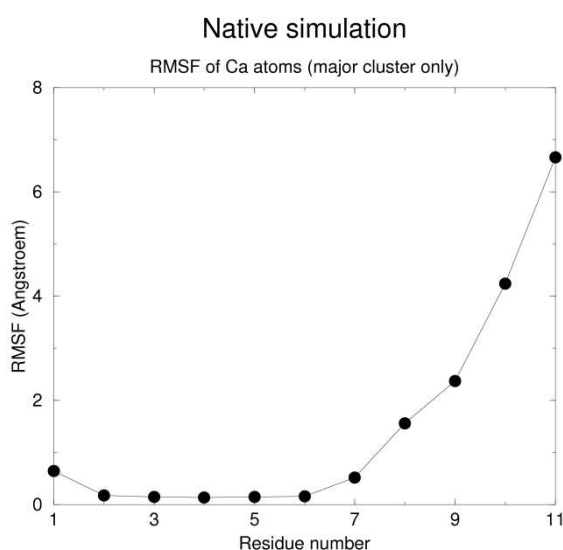
Ανεξάρτητα από τις απόλυτες τιμές των RMSD, στην ανωτέρω εικόνα φαίνεται ότι στην περίπτωση που αφορά και τα ένδεκα κατάλοιπα, υπάρχει μεγάλη ποικιλία διαμορφώσεων του πεπτιδίου συνολικά ενώ, μετά την εξαίρεση του C-τελικού άκρου, παρατηρείται μία σταθερότητα όσο αφορά τις επαναλαμβανόμενες δομές, τις οποίες επισκέπτονται τα πρώτα 7 κατάλοιπα καθώς και το χρονικό διάστημα για το οποίο παραμένουν οι δομές αυτές.

Σε σχέση με τη δευτεροταγή δομή της περίπτωσης αριστερά, οι μικρότερες τιμές RMSD λαμβάνονται όταν η δομή παίρνει διαμόρφωση  $\alpha$ -έλικας και στο C-τελικό άκρο (στα 300 και στα 380 ns) και οι αμέσως μεγαλύτερες όταν το C-τελικό έχει διαμόρφωση στροφής. Με την επιπλέον βοήθεια της Εικόνας 4.8 συμπεραίνουμε ότι η κεντρική μπλε περιοχή (που αντιστοιχεί σε χρόνο 100 έως 250 ns) αποδίδεται στο

δεύτερο cluster της μεγαλύτερης α-έλικας, συμπέρασμα το οποίο συμφωνεί με την ανάθεση δευτεροταγούς δομής από το STRIDE. Επιπλέον (από την ίδια Εικόνα 4.8), συμπεραίνουμε ότι το κύριο cluster εμφανίζεται μετά τα πρώτα 50 ns και επανεμφανίζεται κάποιες φορές μέχρι τα 100 ns, στη συνέχεια εξαφανίζεται για τα επόμενα 200 ns και επανεμφανιζόμενο γύρω στα 300 ns, υφίσταται τακτικά μέχρι του τέλους της προσομοίωσης. Αξίζει να σημειωθεί ότι το cluster 1 εναλλάσσεται με το cluster 3, σε όλη την διάρκεια της παρουσίας και των δύο. Το γεγονός αυτό δεν έρχεται σε αντίθεση με τον πίνακα RMSD της εικόνας αριστερά, όπου είναι εμφανής η παρουσία δύο κυρίως ομάδων δομών.

Όσο αφορά την δευτεροταγή δομή της περίπτωσης χωρίς τα τελευταία κατάλοιπα (Εικόνα 4.11, δεξιά), το ένα τέταρτο των frames του τροχιακού έχει αποδοθεί σε cluster, εκ των οποίων το 80% ανήκει στο κυρίαρχο. Φαίνεται ότι το κυρίαρχο cluster είναι αυτό, που αντιστοιχεί σε μικρά RMSD, η δε δομή του να είναι αυτή κατά την οποία και το δεύτερο κατάλοιπο ανήκει στην έλικα.

Είναι εμφανής η κινητικότητα του C-τελικού άκρου, όσο αφορά τα Ca άτομα. Για να την ποσοτικοποιήσουμε δημιουργούμε την γραφική παράσταση της Εικόνας 4.12.



**Εικόνα 4.12:**

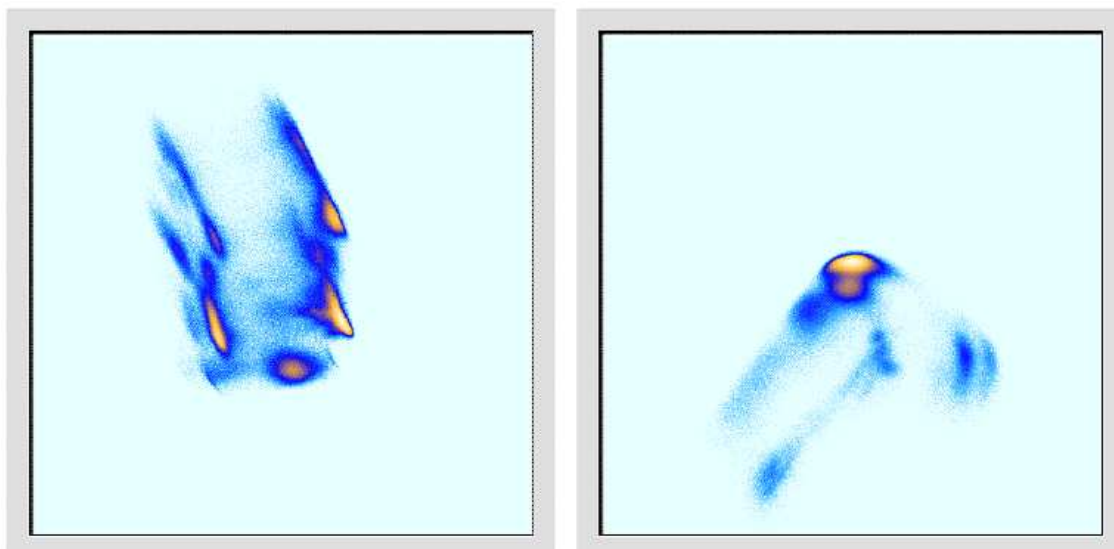
Γραφική παράσταση του εύρους των διακυμάνσεων σε Å καθενός Ca ατόμου του ενδεκαπεπτιδίου καθόλη την διάρκεια του τροχιακού της 'native'.

Είναι εμφανής η σχεδόν αμελητέα κινητικότητα των καταλοίπων 2, 3, 4, 5 και 6, η πολύ μικρή διακύμανση των 1 και 7 και η αυξανόμενη διακύμανση των Ca των καταλοίπων 8-11 σε συνάρτηση με την απόστασή τους από το C-τελικό άκρο.

### 4.3. Ανάλυση του τροχιακού της ‘folding’.

#### 4.3.1. Ανάλυση σε κύριες συνιστώσες με χρήση εσωτερικών συντεταγμένων.

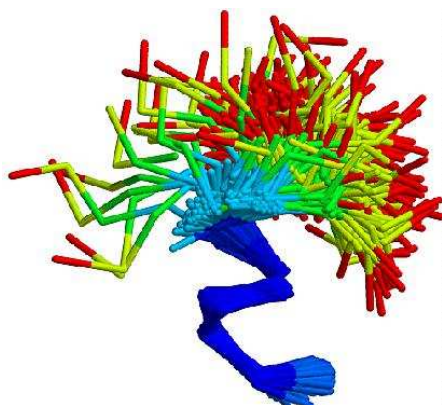
Αρχικά, όπως και στην περίπτωση της ‘native’, ομαδοποιούμε τα δεδομένα του τροχιακού της ‘folding’, εκτελώντας ανάλυση σε κύριες συνιστώσες και στον χώρο των διεδρων γωνιών (dPCA), με την χρήση του προγράμματος carma. Πραγματοποιούμε, επίσης, δύο τέτοιες αναλύσεις. Η μία ανάλυση γίνεται για το σύνολο των διεδρων του ενδεκαπεπτιδίου και η δεύτερη με χρήση των διεδρων των σταθερών καταλοίπων 1 έως και 7. Από την ανάλυση αυτή σε ομάδες (cluster analysis) παίρνουμε την Εικόνα 4.13, που ακολουθεί:



**Εικόνα 4.13:** Εικόνες από την dihedral PCA. Αριστερά: Τα τέσσερα cluster της ‘folding’ με χρήση όλων των διεδρων. Δεξιά: Ένα μοναδικό κυρίαρχο cluster, με χρήση των διεδρων των καταλοίπων 1-7.

Ο αριθμός των frames, τα οποία αποδόθηκαν από το CARMA στα τέσσερα cluster, με χρήση όλων των διεδρων είναι 371883 (από τα 1621038), ποσοστό 23%. Το 55% αυτών (204458 frames) αντιστοιχούν στο κυρίαρχο cluster, με τα επόμενα να αντιστοιχούν σε ποσοστά 26%, 10% και 9%.

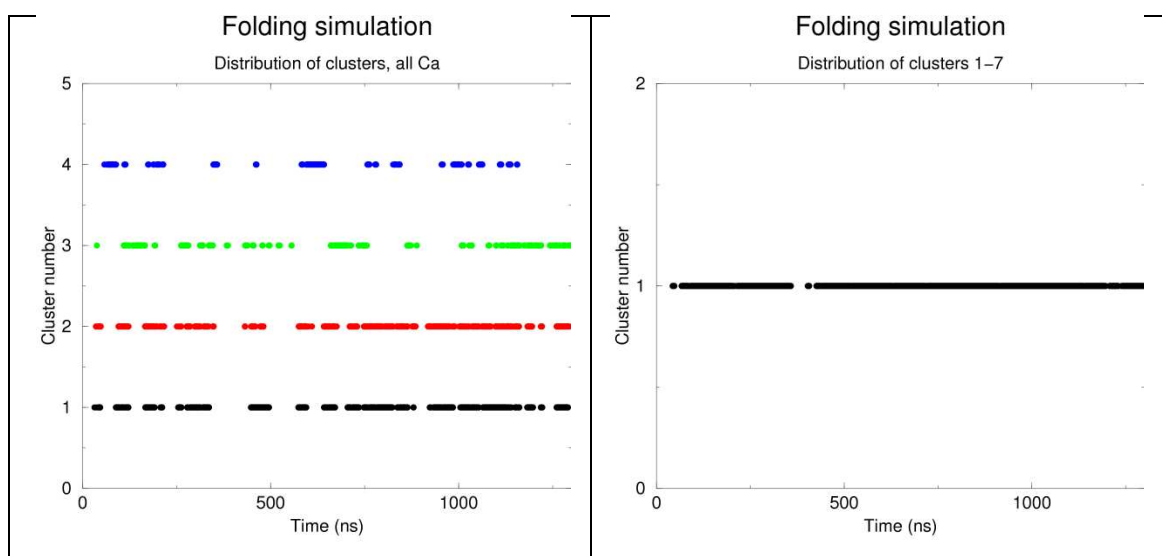
Με υπέρθεση των δομών του μοναδικού cluster με χρήση των διεδρων των καταλοίπων 1-7 και με την βοήθεια του rasmol, παίρνουμε την Εικόνα 4.14, που ακολουθεί:



**Εικόνα 4.14:**

Υπέρθωση 920 δομών του κυρίαρχου cluster με χρήση των διέδρων των καταλοίπων 1-7 και απεικόνιση μόνο των Ca ατόμων. Η χρωματική κλίμακα είναι ανάλογη με τις τιμές rmsf σε χρωματικό φάσμα εύρους, κυμαινόμενου από το χρώμα μπλε (για τις μικρότερες τιμές rmsf έως το κόκκινο (για τις μεγαλύτερες). Η υπέρθεση έγινε με την χρήση των Ca των καταλοίπων 2-6).

Είναι εμφανής η ελικοειδής διαμόρφωση των καταλοίπων 1-7 και η έντονη κινητικότητα της C-τελικής περιοχής των τεσσάρων καταλοίπων. Στην περίπτωση αυτή της *'folding'*, η έλικα δεν επεκτείνεται προς τα κατάλοιπα 8 και 9, όπως συμβαίνει στην αντίστοιχη περίπτωση της *'native'*. Δεδομένης της υψηλής θερμοκρασίας της προσομοίωσης *'folding'*, μια πιθανή ερμηνεία είναι ότι η μεγαλύτερη κινητικότητα των ατόμων στην συγκεκριμένη προσομοίωση, εμπόδισε το μόριο να εγκλωβιστεί σε διαμόρφωση μεγαλύτερου μήκους έλικας, όπως θα μπορούσε να έχει συμβεί στην περίπτωση της *'native'*. Για να ελέγξουμε την πιθανότητα αυτή, ετοιμάζουμε την Εικόνα 4.15, στην οποία φαίνεται η κατανομή των τεσσάρων cluster με την χρήση όλων των διέδρων και η κατανομή του μοναδικού cluster στην περίπτωση της χρήσης των διέδρων των 1-7.

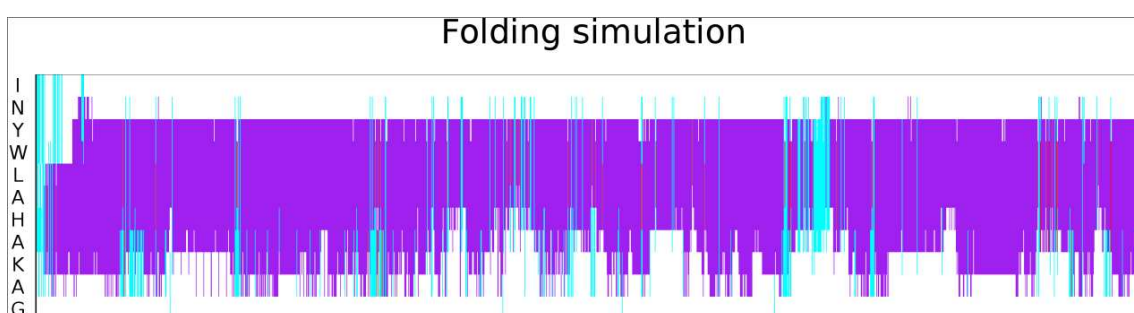


**Εικόνα 4.15:** Γραφική παράσταση της παρουσίας των cluster σε σχέση με τον συνολικό χρόνο προσομοίωσης των 1262 ns. Αριστερά: τα τέσσερα cluster της *'folding'* με χρήση όλων των διέδρων γωνιών. Δεξιά: το μοναδικό cluster με χρήση των διέδρων των 1-7 καταλοίπων.

Επιβεβαιώνεται η παρουσία του κυρίαρχου cluster σε ολόκληρη την χρονική διάρκεια της προσομοίωσης όσο αφορά τις διέδρες 1-7 και η γρήγορη εναλλαγή των τεσσάρων cluster της δομής στην περίπτωση όλων των διέδρων. Η υψηλή θερμοκρασία μπορεί να δικαιολογήσει την εναλλαγή των δομών, συνηγορεί δε σε αυτό και η αντίστοιχη Εικόνα 4.8 της **'native'**.

### 4.3.2. Ανάθεση δευτεροταγούς δομής.

Με το ίδιο πρόγραμμα, το STRIDE, το οποίο χρησιμοποιήσαμε και στην περίπτωση της **'native'**, παίρνουμε την χρωματική απεικόνιση της ανάθεσης δευτεροταγούς δομής στην Εικόνα 4.16, η οποία ακολουθεί:



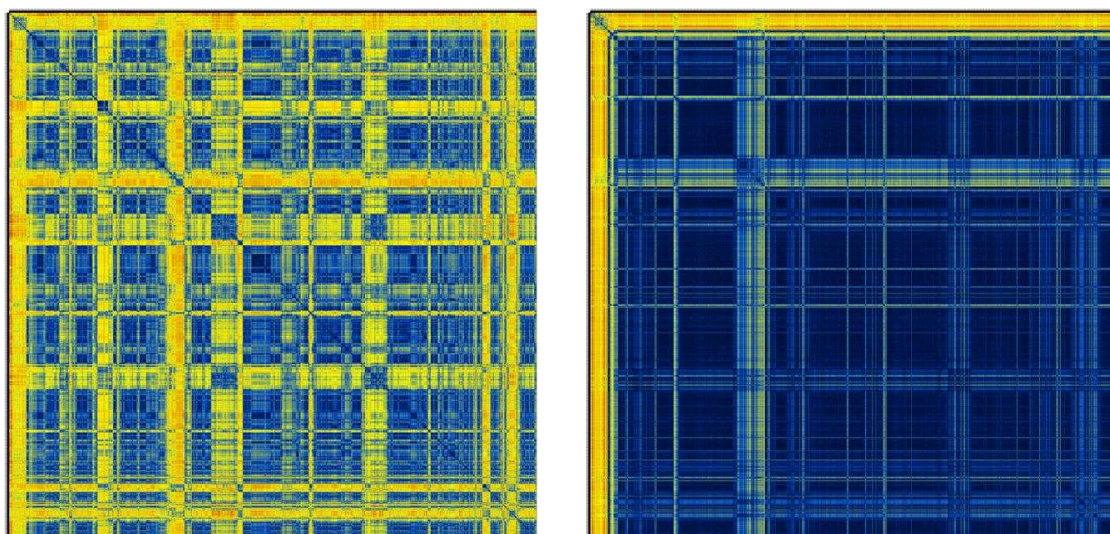
**Εικόνα 4.16:** Χρωματική απεικόνιση της δευτεροταγούς δομής, την οποία αναθέτει το πρόγραμμα STRIDE, με βάση το τροχιακό της προσομοίωσης **'folding'**.

Και στην περίπτωση της **'folding'**, παρατηρούμε ότι οι δομές κυριαρχούνται από  $\alpha$ -έλικα (μωβ) στη N-τελική περιοχή στο μεγαλύτερο χρονικό διάστημα της προσομοίωσης. Ωστόσο η έλικα αυτή εναλλάσσεται για μικρά χρονικά διαστήματα με διαμόρφωση στροφής, γεγονός το οποίο δεν παρατηρείται στην περίπτωση της **'native'**. Το C-τελικό άκρο λαμβάνει συχνότερα διαμόρφωση έλικας, ωστόσο αυτή δεν αποδίδεται στο κυρίαρχο cluster. Παρατηρείται και ένα επιπλέον χρονικό διάστημα (πλην του αρχικού), γύρω από το χρόνο των 2/3 της προσομοίωσης, όπου το πεπτίδιο διατηρεί σταθερή την διαμόρφωση στροφής. Σποραδικά, η περιοχή των καταλοίπων 4-7 παίρνει μία, πολύ μικρής διάρκειας, διαμόρφωση  $3_{10}$ -έλικας. Ούτε η προσομοίωση **'folding'** μπόρεσε να αναπαράγει δομή  $3_{10}$ -έλικας, γεγονός από το οποίο φαίνεται και πάλι, η τάση του CHARMM forcefield για  $\alpha$ -έλικες.



### 4.3.3. Πίνακες RMSD και ανάλυσή τους.

Στην επόμενη εικόνα (Εικόνα 4.17) βλέπουμε έγχρωμη αναπαράσταση των τιμών των RMSD μεταξύ όλων των πιθανών συνδυασμών των δομών του πεπτιδίου και με χρήση όλων των Ca ατόμων (αριστερά) και των Ca 1-7 (δεξιά), όπως έχουν καταγραφεί κατά τη διάρκεια της *'folding'* προσομοίωσης. Το εύρος των παρατηρούμενων τιμών απεικονίζεται χρωματικά, με τον ίδιο τρόπο, όπως και στην περίπτωση της *'native'*, με μικρές τιμές RMSD (σε Å) να αντιστοιχούν σε χρώμα σκούρου μπλε και τις μεγαλύτερες σε χρώμα σκούρου κόκκινου. Οι παρατηρούμενες τιμές έχουν εύρος που κυμαίνεται από 0.00 μέχρι 6.95Å στην εικόνα αριστερά, μέγιστου δε, RMSD 5.34Å στην εικόνα δεξιά.



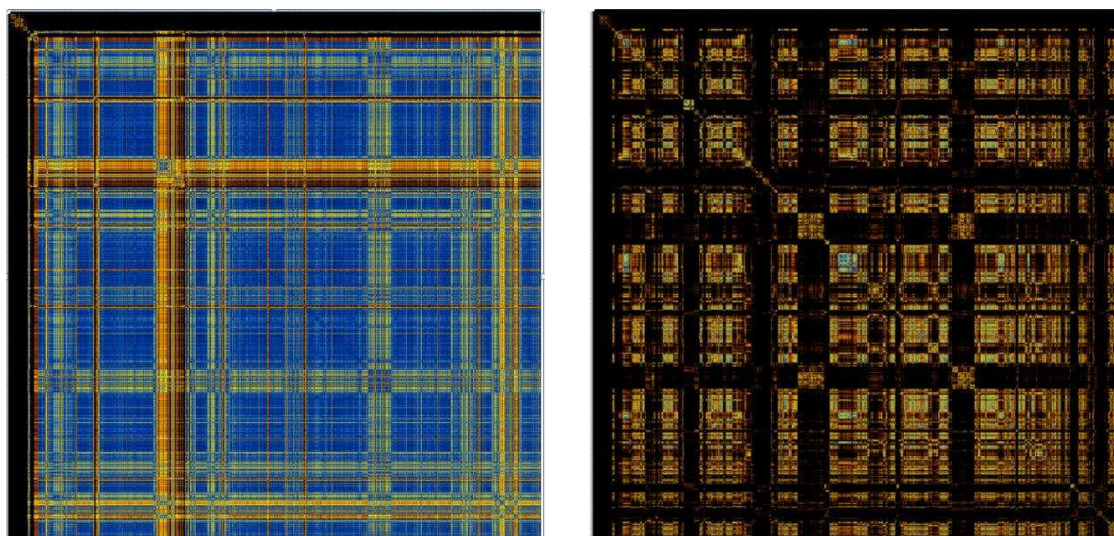
**Εικόνα 4.17:** Αριστερά: Πίνακας RMSD της *'folding'* προσομοίωσης με χρήση των συντεταγμένων όλων των Ca ατόμων και μέγιστο RMSD 6.95Å. Δεξιά: Ίδιου τύπου πίνακας, της ίδιας προσομοίωσης, με χρήση των συντεταγμένων των επτά N-τελικών Ca ατόμων και μέγιστο RMSD 5.34 Å.

Στον πίνακα αριστερά, παρατηρούμε πολλά επαναλαμβανόμενα σκούρα μπλε τμήματα, που αντιστοιχούν, όπως έχουμε αναφέρει, σε μικρά RMSD. Επιπλέον πολλά σκούρα μπλε τμήματα της κεντρικής διαγωνίου συνδέονται τόσο οριζόντια όσο και κάθετα με μπλε τμήματα. Αυτό δείχνει ότι το πεπτίδιο σχηματίζει παρόμοιες δομές σε διαφορετικές φάσεις της προσομοίωσης. Το γεγονός ότι αυτά τα μπλε τμήματα διαχωρίζονται από τις κίτρινες - κόκκινες περιοχές μεγαλύτερου RMSD δηλώνει ότι οι παρόμοιες αυτές δομές προκύπτουν από γεγονότα επαναλαμβανόμενης αποδιάταξης

και αναδίπλωσης του πεπτιδίου. Η ανάλυση αυτή συμφωνεί και με τα αρχικά συμπεράσματα από τα στιγμιότυπα της Εικόνας 4.2, όπου επίσης έχουμε συναντήσει παρόμοιες δομές σε διαφορετικά στιγμιότυπα της προσομοίωσης.

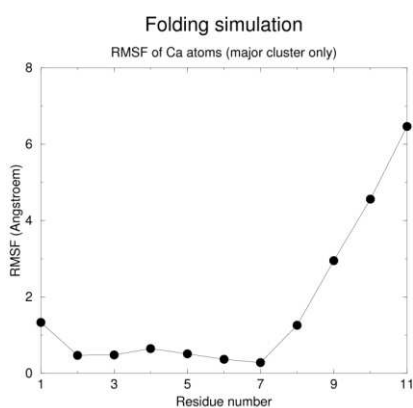
Λόγω της ευκινησίας του C-τελικού άκρου και της δημιουργίας του ερωτήματος κατά πόσο τα γεγονότα αναδίπλωσης – αποδιάταξης της εικόνας 4.17 επηρεάζονται από την παρουσία του υπερκινητικού C-τελικού άκρου, υπολογίσαμε τον πίνακα RMSD της Εικόνας 4.17 δεξιά, χρησιμοποιώντας μόνο τις συντεταγμένες των επτά N-τελικών Ca ατόμων. Αυτή τη φορά το μέγιστο παρατηρούμενο RMSD έχει τιμή 5.34 Å και φαίνεται ότι, εξαιρώντας τα C-τελικά κατάλοιπα, ο πίνακας καταλαμβάνεται σχεδόν ολοκληρωτικά από πολύ μικρές τιμές RMSD (μπλε σκούρο).

Επειδή ο ανωτέρω πίνακας κυριαρχείται από μικρές τιμές RMSD (όταν έχουμε ως μέγιστο τα 5.34 Å) και για να έχουμε σαφέστερη χρωματική απεικόνιση της διακύμανσης των RMSD, κατασκευάζουμε τον ίδιο πίνακα με μέγιστο τα 2.00 Å. Ο πίνακας αυτός απεικονίζεται στο αριστερό μέρος της Εικόνας 4.18. Για να μπορέσουμε να συγκρίνουμε χρωματικά (εφόσον το χρώμα αντιστοιχεί σε τιμές RMSD) τους δύο πίνακες της εικόνας 4.17, υπολογίζουμε και τον αριστερά ευρισκόμενο (αυτόν με χρήση όλων των Ca ατόμων), με μέγιστο RMSD τα 2.00 Å. Ο πίνακας αυτός απεικονίζεται στο δεξί μέρος της Εικόνας 4.18.



**Εικόνα 4.18:** Αριστερά: Πίνακας RMSD της 'folding' προσομοίωσης με τις συντεταγμένες των επτά N-τελικών Ca ατόμων και μέγιστο RMSD 2.00 Å. Δεξιά: Πίνακας RMSD της 'folding' προσομοίωσης με χρήση των συντεταγμένων όλων των Ca ατόμων και μέγιστο RMSD 2.00 Å.

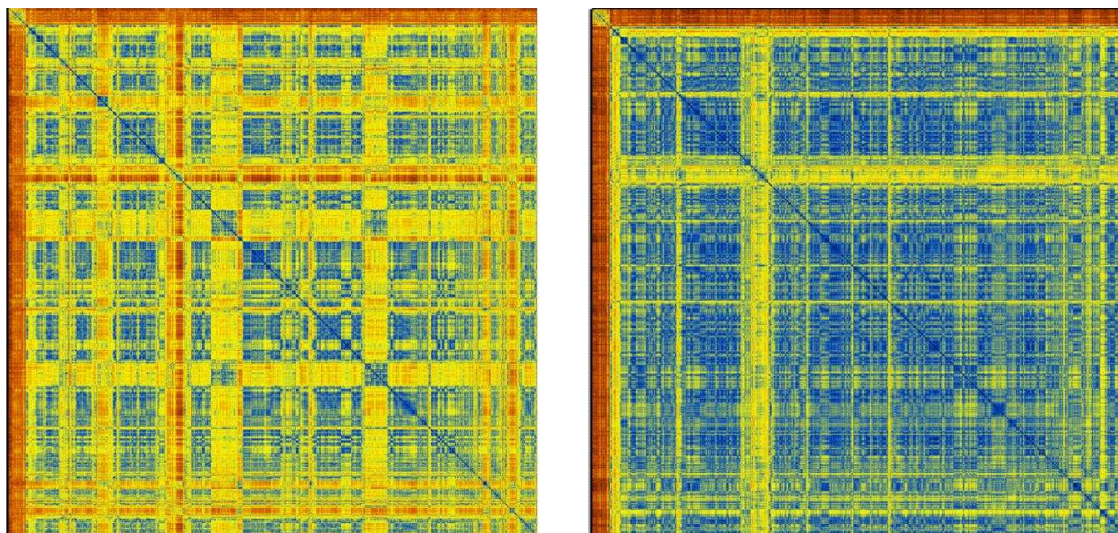
Παρατηρούμε, στο αριστερό μέρος του ανωτέρω πίνακα, ότι τα επτά N-τελικά κατάλοιπα συνεχίζουν να διατηρούν σταθερή την διαμόρφωσή τους σε σημαντικό ποσοστό των frames του τροχιακού. (όσο αφορά πάντα τα Ca άτομα). Όσο για τον πίνακα δεξιά, παρατηρούμε ότι αυτός κυριαρχείται πλέον από μεγάλα RMSD (μεγάλα για το εύρος 0.00 έως 2.00 Å). Φαίνεται ότι και στην περίπτωση της 'folding' ότι τα επιπλέον Ca άτομα του ευκίνητου C-τελικού άκρου (της AKAG), επηρεάζουν σημαντικά την σταθερότητα της δομής του ενδεκαπεπτιδίου κατά την διάρκεια της προσομοίωσης. Για να δούμε την διακύμανση του καθενός εκ των Ca ατόμων των ένδεκα καταλοίπων καθόλη την διάρκεια της προσομοίωσης, δημιουργήσαμε την γραφική παράσταση της Εικόνας 4.19.



**Εικόνα 4.19:**

Γραφική παράσταση της διακύμανσης γύρω από την μέση θέση του κάθε Ca ατόμου του κυρίαρχου cluster, κατά το χρονικό διάστημα της προσομοίωσης. Παρατηρείται η πολύ μικρή διακύμανση των Ca ατόμων των καταλοίπων 2-7, η μικρή διακύμανση των 1 και 8 και η μεγάλη διακύμανση των καταλοίπων του C-τελικού άκρου.

Μέχρι τώρα έχουμε χρησιμοποιήσει στους υπολογισμούς μας μόνο τα ένδεκα Ca άτομα από τα 177 συνολικά άτομα του πεπτιδίου. Για να δούμε την επίδραση των υπόλοιπων ατόμων του μορίου (πλην των υδρογόνων) στην σταθερότητα ή καλύτερα αποσταθεροποίησή του, υπολογίζουμε δύο καινούργιους πίνακες RMSD στους οποίους συνυπολογίζουμε και τις θέσεις των βαρέων ατόμων. Στον πίνακα, που απεικονίζεται αριστερά στην Εικόνα 4.20, έχουν συνυπολογιστεί οι συντεταγμένες όλων των βαρέων ατόμων. Στον πίνακα, που βρίσκεται δεξιά στην ίδια εικόνα χρησιμοποιήθηκαν τα βαρέα άτομα των καταλοίπων 1-7.

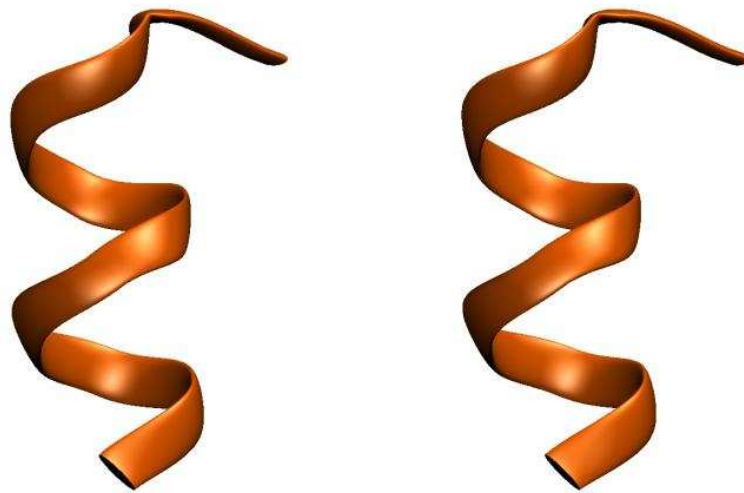


**Εικόνα 4.20:** Αριστερά: Πίνακας RMSD της **'folding'** προσομοίωσης με χρήση των συντεταγμένων όλων των βαρέων ατόμων με μέγιστο RMSD 7.89 Å. Δεξιά: Πίνακας RMSD της **'folding'** προσομοίωσης χρησιμοποιώντας τις συντεταγμένες όλων των βαρέων ατόμων των καταλοίπων 1-7 και μέγιστο RMSD 6.93.

Στον αριστερό πίνακα της εικόνας αυτής, παρατηρούμε την αυξημένη κινητικότητα του μορίου, που φαίνεται από τις μεγαλύτερες τιμές RMSD (κίτρινα και κόκκινα χρώματα ακόμα και με μέγιστο τα 7.8 Å). Σε γενικές γραμμές αυτό είναι αναμενόμενο και δείχνει το πώς η κινητικότητα των πλευρικών αυτών ομάδων μειώνει σημαντικά την σταθερότητα της διαμόρφωσης του μορίου στην διάρκεια των 1262 ns. Ωστόσο παρατηρούμε ακόμα τον ικανοποιητικό αριθμό επαναλήψεων παρόμοιας διαμόρφωσης κατά την διάρκεια του συνολικού χρόνου της προσομοίωσης. Στον πίνακα δεξιά, βλέπουμε την επίδραση των βαρέων ατόμων των τελευταίων υπερκινητικών C-τελικών καταλοίπων. Παρατηρούμε σημαντική μείωση των διαμορφώσεων της δομής κατά την προσομοίωση, όταν εξαιρεθούν τα βαρέα άτομα των τελευταίων τεσσάρων καταλοίπων.

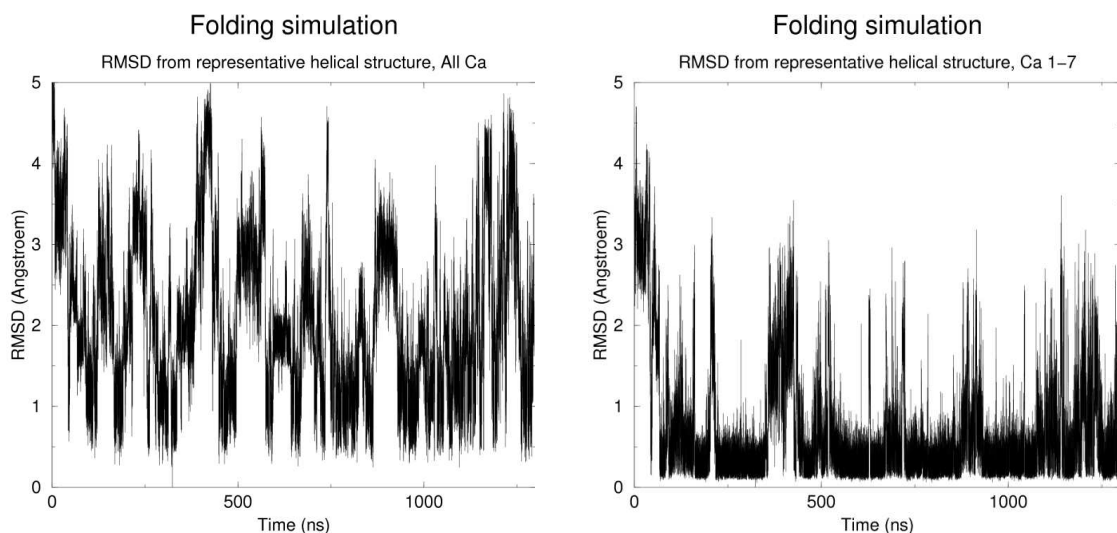
#### 4.3.4. RMSD από το πλησιέστερο στη μέση δομή frame της **'folding'**.

Στην επόμενη εικόνα 4.21 βλέπουμε την δομή, η οποία αντιστοιχεί στο frame 404945 της προσομοίωσης **'folding'** και το οποίο είναι το frame με το μικρότερο RMSD από την υπολογισθείσα μέση δομή αυτής της προσομοίωσης. Η μέση δομή υπολογίστηκε από το κυρίαρχο cluster της ανάλυσης dPCA.

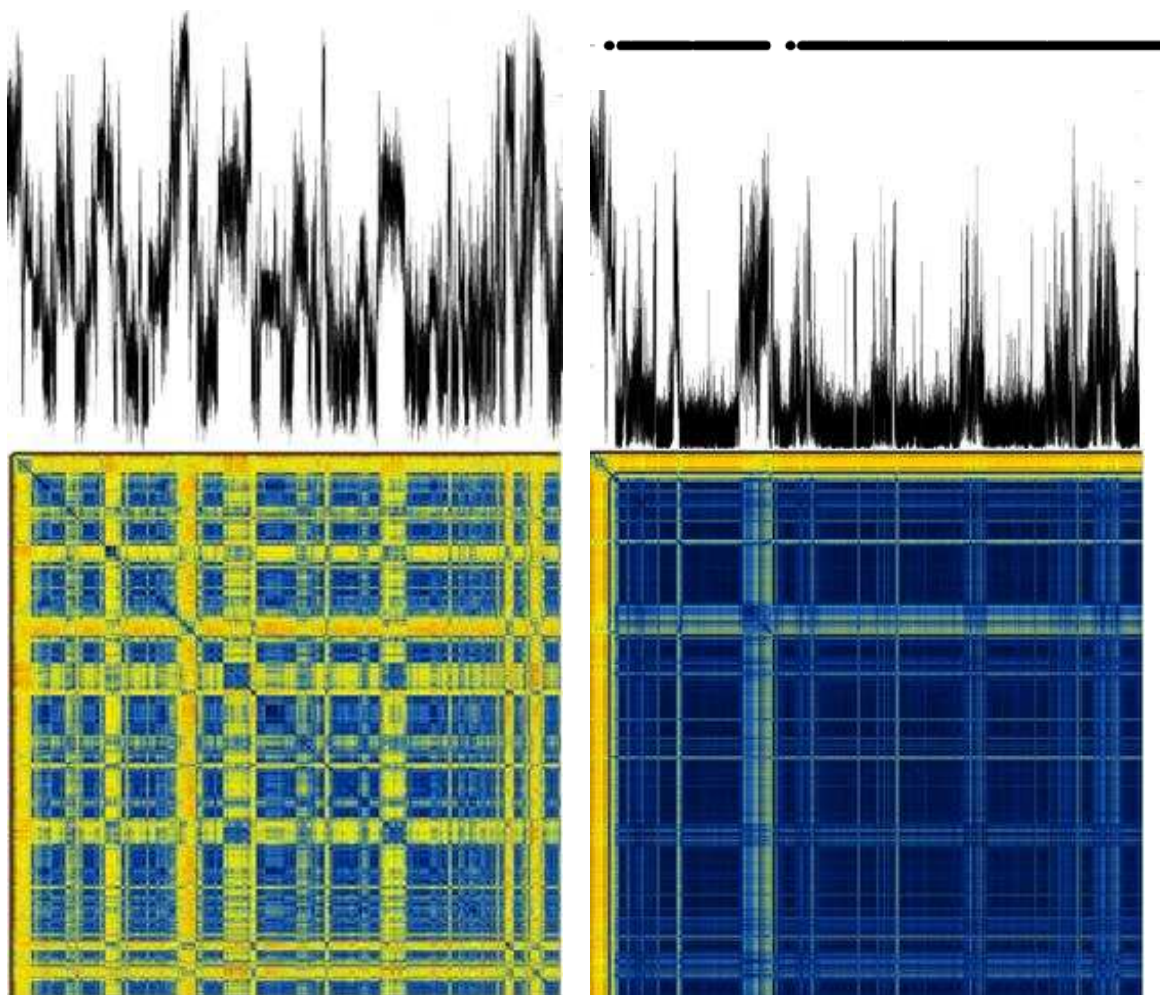


**Εικόνα 4.21:** Η πλέον αντιπροσωπευτική δομή της προσομοίωσης **'folding'** σε στερεοσκοπική απεικόνιση.

Στις δύο επόμενες εικόνες (Εικόνα 4.22 και 4.23), που ακολουθούν βλέπουμε την γραφική παράσταση της απόκλισης του κάθε frame της προσομοίωσης **'folding'** από την πλησιέστερη στην αντιπροσωπευτική δομή της Εικόνας 4.21, σε συνάρτηση με τον συνολικό χρόνο της προσομοίωσης, τόσο με την χρήση όλων των Ca, όσο και με χρήση των Ca των καταλοίπων 1-7 καθώς και τους αντίστοιχους πίνακες RMSD.



**Εικόνα 4.22:** RMSD όλων των δομών της προσομοίωσης από την αντιπροσωπευτική δομή της **'folding'**. Αριστερά: με όλα τα Ca. Δεξιά: με τα Ca 1-7.



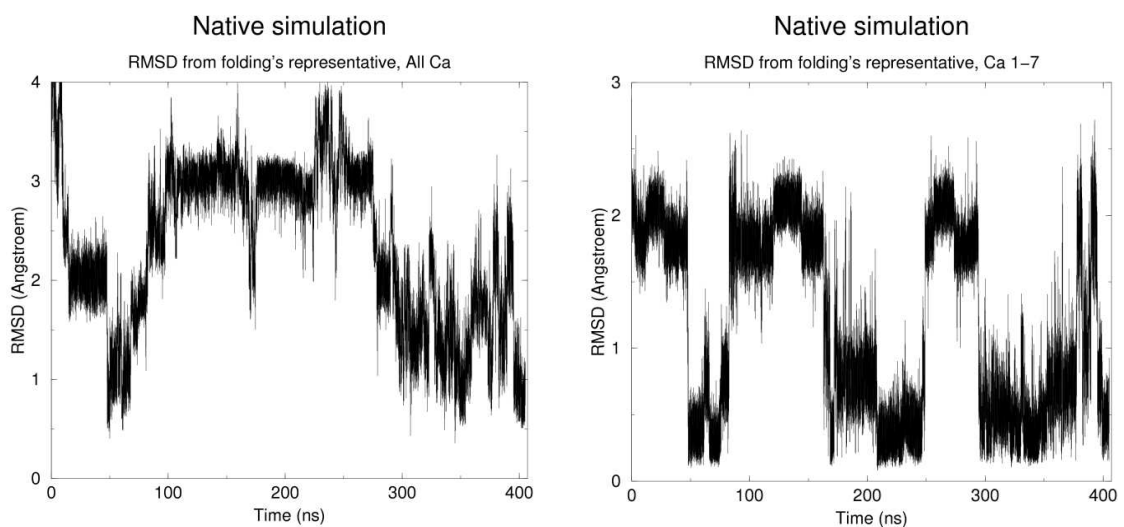
**Εικόνα 4.23:** Απεικόνιση των πινάκων RMSD της Εικόνας 4.17 σε σύνθεση με τα RMSD από την αντιπροσωπευτική δομή της *'folding'* της Εικόνας 4.21. Αριστερά: με όλα τα Ca. Δεξιά: με τα Ca των 1-7 και επάνω η κατανομή του κυρίαρχου cluster.

Στην γραφική παράσταση, αριστερά στην Εικόνα 4.23, φαίνεται η γρήγορη εναλλαγή των διαμορφώσεων του πεπτιδίου, με τα σχετικά μικρά RMSD να κυριαρχούν σε μεγάλο ποσοστό του συνολικού χρόνου. Παρατηρείται η ανάλογη αυξομείωση των RMSD του πίνακα και των RMSD από το αντιπροσωπευτικό frame της μέσης δομής. Στο δεξί τμήμα της Εικόνας 4.23, όπου χρησιμοποιήθηκαν τα κατάλοιπα 1-7, υπάρχει σχεδόν σε όλο το χρονικό διάστημα της προσομοίωσης μια σχεδόν απόλυτη κυριαρχία των μικρών RMSD. Αυτή, επίσης βρίσκεται σε αναλογία με το γράφημα των RMSD από το αντιπροσωπευτικό frame της μέσης δομής. Οι ελάχιστες χρονικές στιγμές, κατά τις οποίες οι τιμές RMSD είναι σχετικά μεγάλες, αντιστοιχούν σε αυτές, κατά τις οποίες οι διαμορφώσεις του μορίου δεν αντιστοιχούν στο κυρίαρχο (και μοναδικό) cluster της dPCA με χρήση των καταλοίπων 1-7.

#### 4.4. Σύγκριση των δύο προσομοιώσεων.

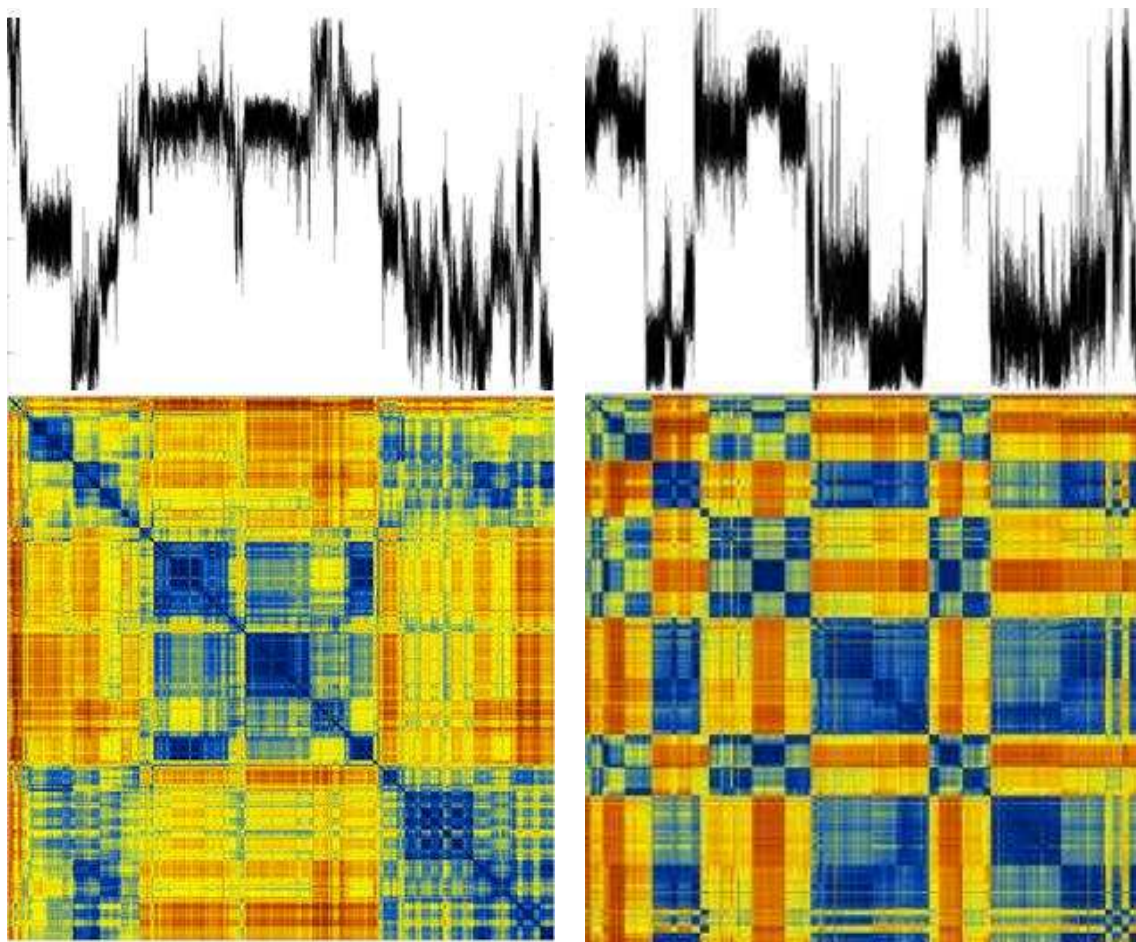
##### 4.4.1. RMSD της 'native' από τη 'μέση δομή' της 'folding'.

Χρησιμοποιώντας τον όρο 'μέση δομή' στην παρούσα παράγραφο, εννοούμε το frame, το οποίο έχει το μικρότερο RMSD από την πραγματική μέση δομή. Αυτό το frame 404945 της προσομοίωσης 'folding', το οποίο απεικονίζεται με το rasmol στην Εικόνα 4.21, χρησιμοποιήθηκε για τους υπολογισμούς των RMSD του κάθε frame του τροχιακού της 'native' από αυτό, της 'μέσης δομής' της 'folding', τόσο με όλα τα Ca, όσο και με τα Ca 1-7. Στην Εικόνα 4.24 απεικονίζονται οι δύο γραφικές παραστάσεις των RMSD όλων των δομών της προσομοίωσης 'native' από την αντιπροσωπευτική δομή της 'folding'.



**Εικόνα 4.24:** RMSD όλων των δομών της προσομοίωσης 'native' από την αντιπροσωπευτική δομή της 'folding'. Αριστερά: με όλα τα Ca. Δεξιά: με τα Ca 1-7.

Στην Εικόνα 4.25 η οποία ακολουθεί, απεικονίζεται η ανωτέρω ευρισκόμενη εικόνα και επιπλέον περιλαμβάνονται και οι έγχρωμοι πίνακες RMSD ανάμεσα σε όλους τους πιθανούς συνδυασμούς των δομών, τις οποίες έλαβε το πεπτίδιο κατά την διάρκεια της 'native'.



**Εικόνα 4.25:** Απεικόνιση του πίνακα RMSD της **'native'** και RMSD της **'native'** από την αντιπροσωπευτική δομή της **'folding'** κατά τον χρόνο της προσομοίωσης. Αριστερά με όλα τα Cα, δεξιά με τα Cα των 1-7.

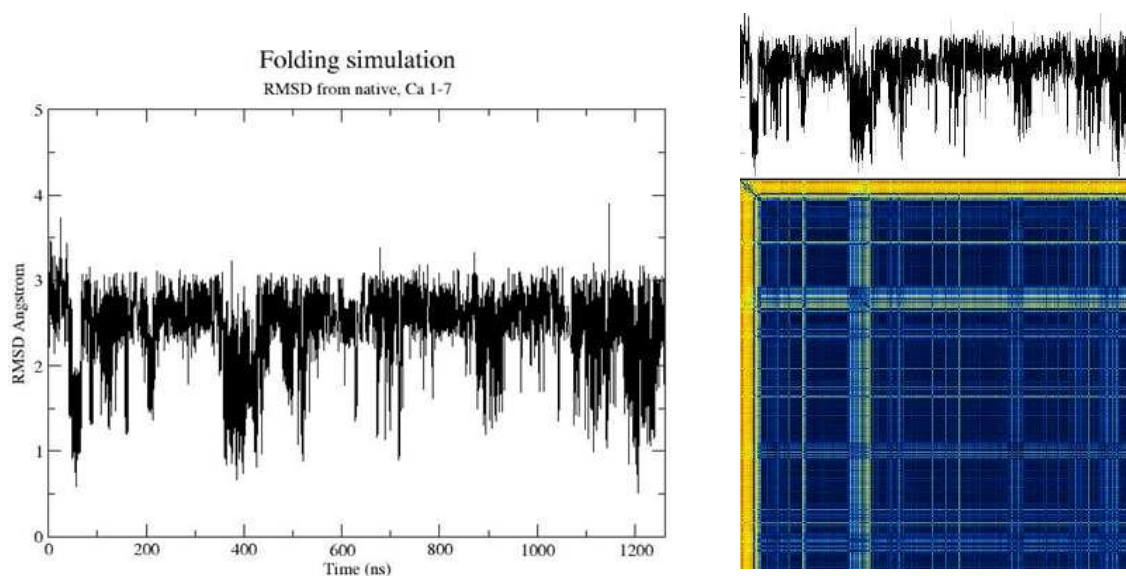
Η αρχική μεγάλη απόκλιση δικαιολογείται από την αρχική δομή  $3_{10}$ -έλικας της προσομοίωσης **'native'**. Το πρώτο cluster της **'native'** με όλα τα Cα, έχει τα μικρότερα RMSD, ενώ το δεύτερο cluster έχει από τα μεγαλύτερα RMSD ως προς την 'μέση δομή της **'folding'**. Το γεγονός αυτό μπορεί να αιτιολογηθεί δεδομένης της δομής α-έλικας στα κατάλοιπα 3-8 της 'μέσης δομής' και της διαμόρφωσης των τεσσάρων τελικών καταλοίπων της. Το cluster 1 είναι α-έλικα στα 2-7 με έντονη κινητικότητα του πρώτου N-τελικού καταλοίπου, μικρή κινητικότητα του όγδοου και παρόμοια διαμόρφωση με την μέση δομή της **'folding'** στα κατάλοιπα 9-11. Το cluster 2 έχει α-έλικα στα 2-8 και μικρή κινητικότητα στο κατάλοιπο 1. Η μικρή αυτή κινητικότητα των καταλοίπων 1 και 2 του cluster 2 μαζί με την μεγάλη διαφορά στην διαμόρφωση των τελευταίων τεσσάρων καταλοίπων από αυτή της 'μέσης δομής' της **'folding'**, μπορεί να δικαιολογήσει αυτά τα μεγάλα RMSD.



Φαίνεται επίσης ότι τα κατάλοιπα 1-7 επισκέπτονται την πλησιέστερη στην μέση δομή της *'folding'* στον μισό περίπου χρόνο της προσομοίωσης, ενώ το κυρίαρχο cluster της *'native'* με τα 1-7, έχει τα μεγαλύτερα RMSD από την 'μέση δομή' της *'folding'*. Όταν συμπεριλαμβάνονται όλα τα Ca, η προσομοίωση πλησιάζει την δομή κατά το ένα τρίτο του συνολικού χρόνου. Θα μπορούσαμε να πούμε ότι οι δύο προσομοιώσεις (*'native'* και *'folding'*) συγκλίνουν.

#### 4.4.2. RMSD της *'folding'* από την πειραματικά προσδιορισμένη δομή.

Στην εικόνα 4.26 (αριστερά), έχουμε δημιουργήσει γραφική παράσταση των RMSD των Ca, των καταλοίπων 1 – 7 του τροχιακού της προσομοίωσης *'folding'*, ως προς την πειραματικά προσδιορισμένη με NMR δομή (την οποία έχουμε προσθέσει ως πρώτο frame), σε συνάρτηση με τον χρόνο προσομοίωσης, τα 1262 ns. Παρατηρούμε ότι από την αρχή ακόμα της προσομοίωσης και σχεδόν σε όλη την διάρκειά της τα RMSD από την πειραματικά προσδιορισμένη δομή (μοντέλο 1 της 2DX2), είναι σημαντικά μεγάλα, κυμαινόμενα στο διάστημα των 3 με 4 Å, υπάρχει δε μείωσή τους - μικρής χρονικής διάρκειας - στην περιοχή των 50, 400, 520, 750, 850, 1100 και 1200 ns. Έχοντας επιπλέον τα δεδομένα από τον αντίστοιχο πίνακα RMSD του δεξιού τμήματος της Εικόνας 4.17 και για να αιτιολογήσουμε τα μεγάλα RMSD, τοποθετήσαμε τον πίνακα αυτό, κάτω ακριβώς από την γραφική παράσταση της αριστερής εικόνας. Το αποτέλεσμα απεικονίζεται δεξιά στη κάτωθι Εικόνα 4.26.

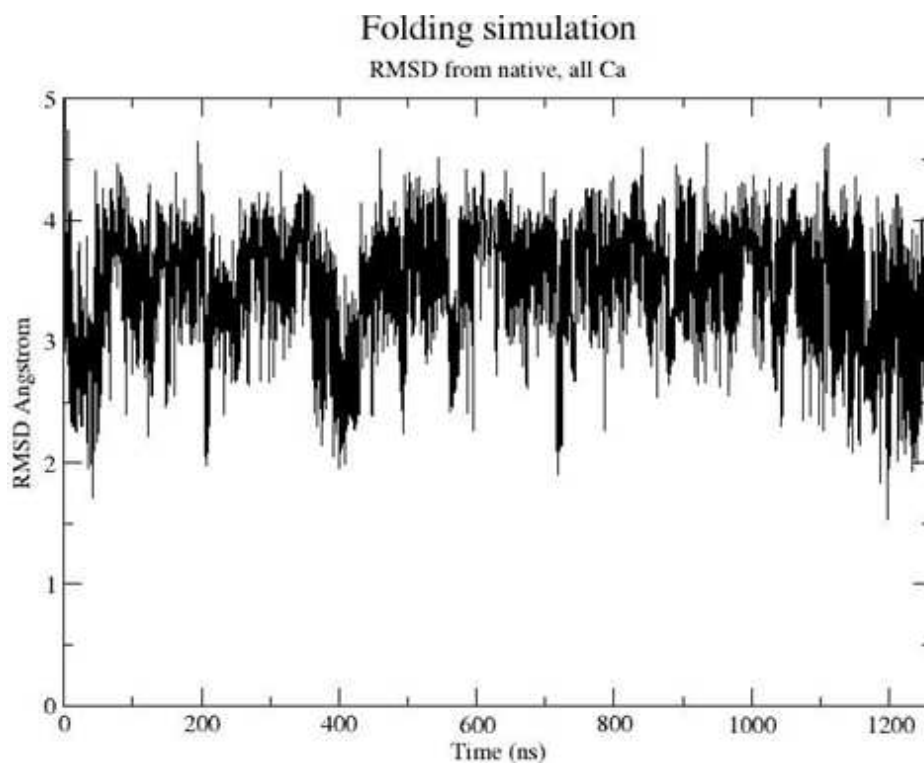


**Εικόνα 4.26:** Αριστερά: Γραφική παράσταση των RMSD (Å) από την πειραματικά προσδιορισμένη δομή ως προς τον συνολικό χρόνο προσομοίωσης (ns) με χρήση των Ca των καταλοίπων 1-7. Δεξιά: Η αναφερόμενη στο κείμενο σύνθεση εικόνων.

Παρατηρούμε ότι οι μικρές τιμές των RMSD αντιστοιχούν στα σημεία εκείνα του πίνακα, όπου η δομή παίρνει την λιγότερο σταθερή δομή της, ενώ οι μεγάλες τιμές RMSD στην πλέον σταθερή διαμόρφωση που παρατηρούμε επανειλημμένα κατά την διάρκεια σχεδόν ολόκληρης της προσομοίωσης.

Θεωρούμε αναμενόμενη την αρχική μεγάλη απόκλιση του γραφήματος, εφόσον έχουμε υπολογίσει το RMSD από το μοντέλο 1 της  $3_{10}$ -έλικας, ενώ το πρώτο frame της προσομοίωσης αντιστοιχεί στην (σχεδόν) πλήρως εκτεταμένη διαμόρφωση του πεπτιδίου. Μετά την αρχική μεγάλη τιμή RMSD θα περιμέναμε στην συνέχεια σημαντική μείωσή της καθώς το πεπτίδιο παίρνει την ελικοειδή σταθερή δομή του στα διάφορα στάδια της προσομοίωσης. Αλλά μία ελικοειδής δομή, όπως γνωρίζουμε, μπορεί να είναι είτε  $\alpha$ -έλικα είτε  $3_{10}$  είτε  $\pi$ -έλικα και κάθε μία από αυτές έχει διαφορετικές  $\phi$  και  $\psi$  γωνίες σε διαδοχικά κατάλοιπα, άλλο αριθμό καταλοίπων ανά στροφή, διαφέρουν δε και στην συχνότητα δημιουργίας δεσμών υδρογόνου. Δηλαδή, ακόμα και με την ίδια πρωτοταγή αλληλουχία διαφορετικοί τύποι έλικας έχουν άλλη διαμόρφωση. Η φυσική δομή της πειραματικά προσδιορισμένης δομής του πεπτιδίου είναι  $3_{10}$ -έλικα. Η σταθερή δομή, που παίρνουμε από την προσομοίωση είναι μεν ελικοειδής, αλλά είναι  $\alpha$ -έλικα. Και μόνο όταν αυτή αποσταθεροποιείται, πλησιάζει περισσότερο σε δομή που έχει παρόμοια χαρακτηριστικά με  $3_{10}$  έλικα. Έτσι μπορούμε να αιτιολογήσουμε τις μεγάλες τιμές RMSD, που παρατηρούμε σε όλη την διάρκεια της προσομοίωσης.

Αν χρησιμοποιήσουμε όλα τα Ca άτομα στην γραφική παράσταση παίρνουμε ακόμα μεγαλύτερες τιμές RMSD, ωστόσο τόσο οι μέγιστες όσο και οι ελάχιστες τιμές παρουσιάζουν συνέπεια στους αντίστοιχους χρόνους της προσομοίωσης (Εικόνα 4.27).

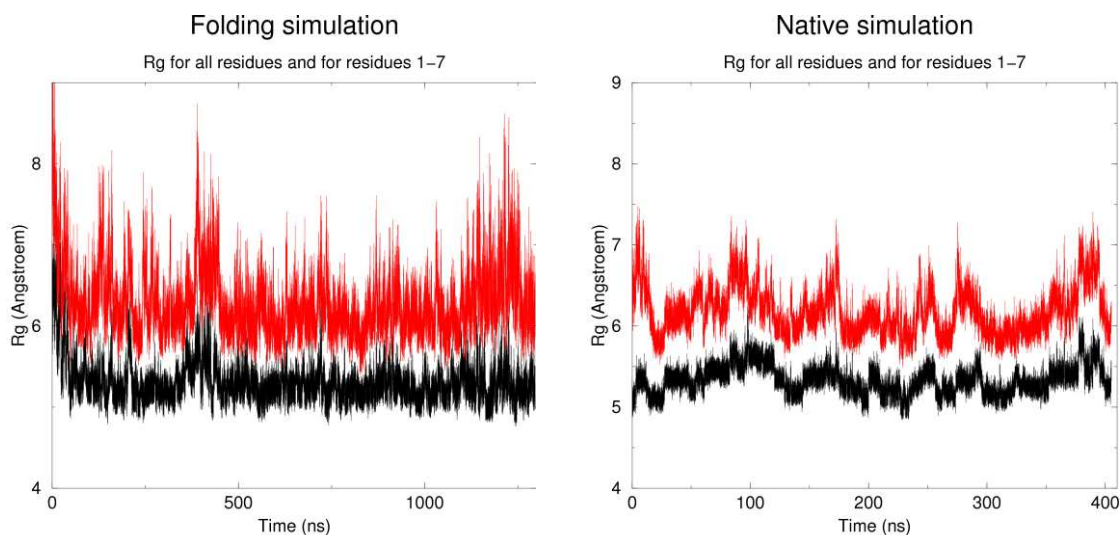


**Εικόνα 4.27:** Γραφική παράσταση των RMSD (Å) από την πειραματικά προσδιορισμένη δομή της προσομοίωσης *'folding'*, ως προς τον συνολικό χρόνο της προσομοίωσης (ns) με χρήση όλων των Ca.

#### 4.4.3. Υδροδυναμική ακτίνα του πεπτιδίου στις *'folding'* και *'native'*.

Ο υπολογισμός της υδροδυναμικής ακτίνας και στην *'folding'* και στην *'native'*, έγινε λαμβάνοντας υπόψη μόνο τα βαρέα άτομα και για ολόκληρη την χρονική διάρκεια των δύο προσομοιώσεων.

Για κάθε μία προσομοίωση η υδροδυναμική ακτίνα υπολογίστηκε εις διπλούν. Ο ένας υπολογισμός αφορούσε το σύνολο των βαρέων ατόμων του πεπτιδίου και ο δεύτερος τα βαρέα άτομα των καταλοίπων 1-7. Στις γραφικές παραστάσεις της επόμενης εικόνας (Εικόνα 4.28) απεικονίζεται χρωματικά η πρώτη περίπτωση με κόκκινο χρώμα και η δεύτερη με μαύρο.



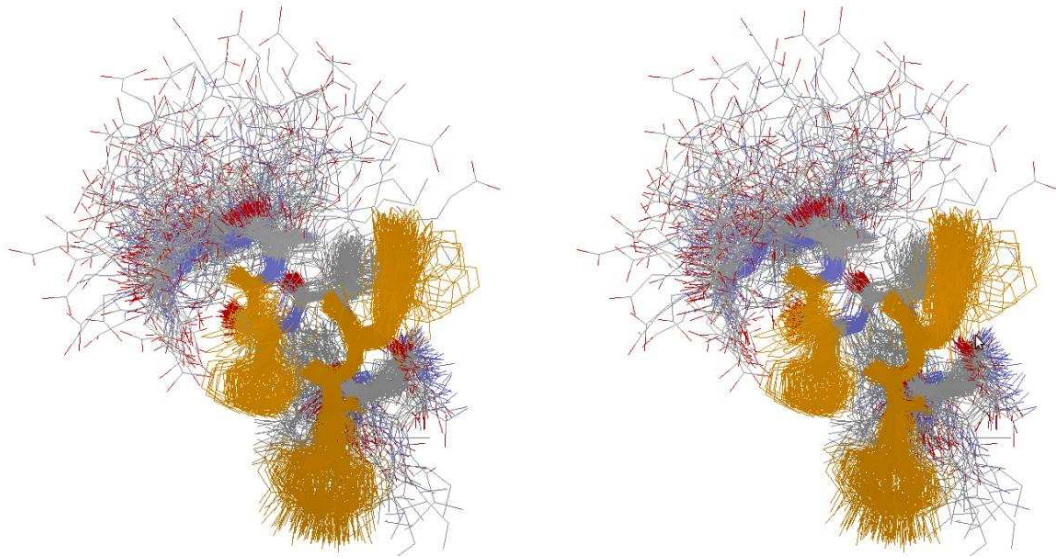
**Εικόνα 4.28:** Οι υδροδυναμικές ακτίνες του πεπτιδίου

Βλέπουμε ότι το μόριο στην περίπτωση της *'folding'* είναι λιγότερο συμπαγές από την περίπτωση της *'native'*. Αυτή η ανοικτότερη διαμόρφωση μπορεί να αιτιολογηθεί λόγω εντονότερης κινητικότητας του μορίου εξαιτίας της υψηλότερης θερμοκρασίας της *'folding'* προσομοίωσης. Σημαντικό ρόλο παίζει και η αρχική εκτεταμένη μορφή του πεπτιδίου και η οποία είναι εμφανής στον αρχικό χρόνο της προσομοίωσης. Αυτή, επαναλαμβάνεται και στον χρόνο των 400ns και βρίσκεται σε συμφωνία με τις μεγαλύτερες αποκλίσεις RMSD, οι οποίες παρατηρήθηκαν σε όλες τις αναλύσεις της *'folding'* προσομοίωσης.

#### 4.5. Αλληλεπιδράσεις πλευρικών ομάδων.

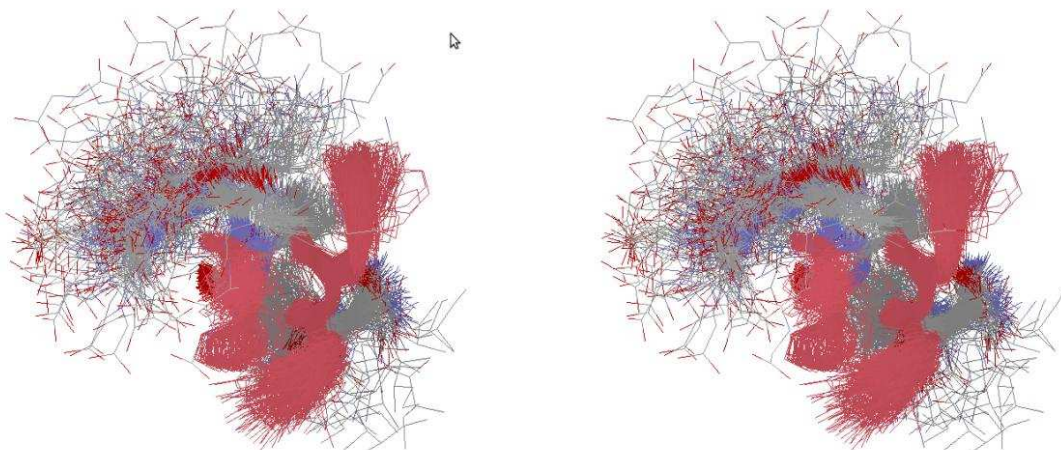
Στην καταχώρηση 1CB3 της PDB έχει παρατηρηθεί η αλληλεπίδραση ανάμεσα στις πλευρικές ομάδες των καταλοίπων της Y103, W104 και H107. Θέλοντας να ελέγξουμε, αν αυτή η αλληλεπίδραση, η οποία δεν υφίσταται στην φυσική μορφή της αλακταλβουμίνης, έχει αναπαραχθεί από το CHARMM forcefield με CMAP correction στο ενδεκαπεπτίδιό μας, επιτελούμε μία επιπλέον ανάλυση σε κύριες συνιστώσες, αυτή τη φορά στον χώρο των καρτεσιανών συντεταγμένων. Χρησιμοποιούμε το αρχείο του τροχιακού της προσομοίωσης *'folding'*, στο οποίο έχουν χρησιμοποιηθεί για υπέρθεση τα Ca άτομα 2-7, η δε ανάλυσή μας γίνεται με χρήση των βαρέων ατόμων των καταλοίπων 3,4 και 7 (ονομαστικά Tyr, Trp και His).

Από την ανάλυση αυτή, αποδίδονται από το CARMA 420982 frames σε 11 cluster, με το πρώτο να περιλαμβάνει 160674 frames και το δεύτερο 254921. Με υπέρθεση των δομών του πρώτου cluster και με την χρήση του rasmol, παίρνουμε την επόμενη εικόνα.



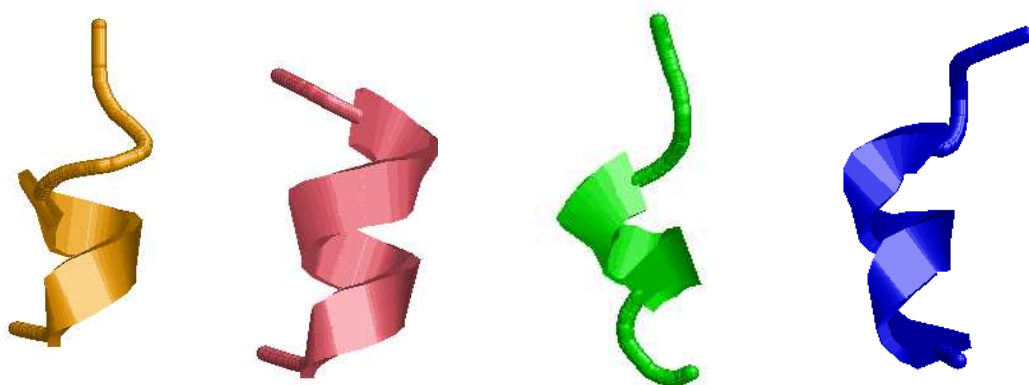
**Εικόνα 4.29:** Στερεοσκοπική απεικόνιση του πρώτου cluster της cPCA ανάλυσης της 'folding' με τα κατάλοιπα 3,4 και 7 να απεικονίζονται σε χρώμα πορτοκαλί.

Με υπέρθεση των δομών του δεύτερου cluster και πάλι με την βοήθεια του rasmol, παίρνουμε την εικόνα, που ακολουθεί:



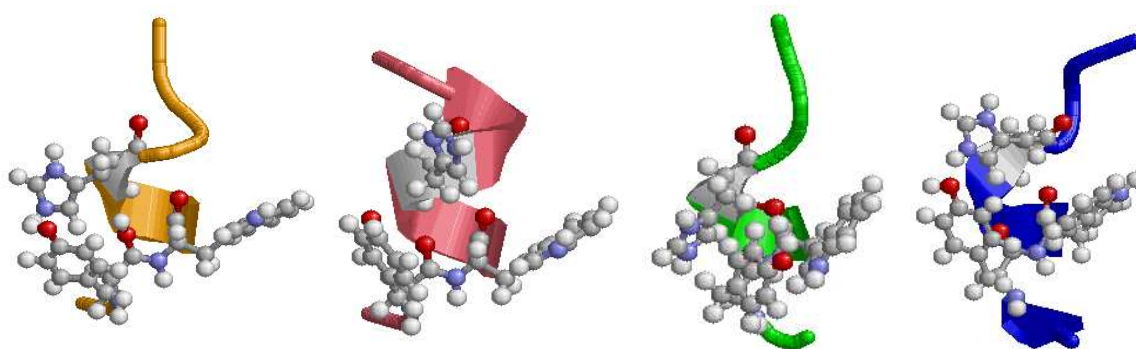
**Εικόνα 4.30:** Στερεοσκοπική απεικόνιση του δεύτερου cluster της cPCA ανάλυσης της 'folding' με τα κατάλοιπα 3,4 και 7 σε χρώμα ροζ.

Από τις ανωτέρω εικόνες μπορεί να παρατηρηθεί ότι η His7 λαμβάνει διαφορετικές διαμορφώσεις στα δύο cluster. Εξαιτίας του πολύ μεγαλύτερου αριθμού των frames του δεύτερου cluster, προτιμάται η χρήση αυτού, για μία επιπλέον ομαδοποίηση, με χρήση μόνο των βαρέων ατόμων της His7. Η νέα cPCA ανάλυση με την χρήση πάντα του CARMA, έχει αποτέλεσμα να αποδοθούν συνολικά σε cluster 85496 frames (από τα 254921), εκ των οποίων το πρώτο περιλαμβάνει 29932 frames και το δεύτερο 48057. Από το κάθε ένα cluster, επιλέγουμε το frame, το οποίο είναι το πλέον αντιπροσωπευτικό της μέσης δομής του καθενός (Εικόνα 4.31).



**Εικόνα 4.31:** Διαγράμματα cartoon με rasmol. Με πορτοκαλί και ροζ, τα πλησιέστερα στην μέση δομή frames του πρώτου και του δεύτερου cluster αντίστοιχα, της προσομοίωσης 'folding'. Πράσινο: Το μοντέλο 1 της 1CB3. Μπλε: Το μοντέλο 1 της 2DX2.

Στα ανωτέρω διαγράμματα αντικαθιστώντας τα κατάλοιπα 3,4 και 7 με ball&sticks, παίρνουμε την επόμενη εικόνα.



**Εικόνα 4.32:** Η Εικόνα 4.31 με αντικατάσταση των καταλοίπων Tyr3, Trp4 και His7 με ball&sticks. Το C-τελικό απεικονίζεται επάνω.

Οι εικόνες απεικονίζουν την Tyr4 στην δεξιά πλευρά. Η Tyr3 είναι σε κάτω αριστερά, ενώ η His7 εμφανίζεται σε διαφορετική θέση (σχετικά με τα W4 και Y3). Η σχετική θέση των καταλοίπων 3 και 4 είναι παρόμοια και στις τέσσερις δομές.

Στην πορτοκαλί δομή φαίνεται άτομα του δακτυλίου της His να βρίσκονται σε ακτίνα van der Waals με άτομα της Tyr. Στην ροζ δομή δεν φαίνεται κάποια αλληλεπίδραση ανάμεσα στις πλευρικές ομάδες των τριών αυτών καταλοίπων, όπως συμβαίνει και στη δομή της 2DX2 (μπλε), σε αντίθεση με τη δομή της 1CB3 (πράσινη), όπου η υπό συζήτηση αλληλεπίδραση είναι εμφανής.

Είναι αξιοσημείωτη η ομοιότητα των δομών του cluster 1 με την 1CB3, όσο αφορά τόσο το μήκος της έλικας (όχι όμως το είδος της τελευταίας), όσο και την πιθανή αλληλεπίδραση των τριών καταλοίπων 3,4 και 7, όπως αξιόλογη είναι και η ομοιότητα των δομών του cluster 2 και της 2DX2, για τους ίδιους ακριβώς λόγους.

# Κεφάλαιο 5<sup>ο</sup>

## Συζήτηση

Every sentence I utter must be understood not as an affirmation but as a question.  
Niels Bohr.

Με την ολοκλήρωση των αναλύσεων των δεδομένων των δύο προσομοιώσεων έχουμε δείξει την αδυναμία του δυναμικού πεδίου CHARMM με τη διόρθωση CMAP να αναπαράγει την δομή της  $3_{10}$ -έλικας από την αρχική εκτεταμένη δομή του πεπτιδίου INYWLAHAKAG στην περίπτωση της προσομοίωσης '*folding*', καθώς και την αδυναμία του να διατηρήσει την αρχική  $3_{10}$ -έλικα της προσομοίωσης '*native*'. Έγινε εμφανής η τάση δημιουργίας α-έλικας από το συγκεκριμένο forcefield, ωστόσο δεν παρατηρήθηκε δομή π-έλικας σε καμία από τις δύο προσομοιώσεις, γεγονός που πιθανά οφείλεται στη βελτιωμένη παραμετροποίηση η οποία συμπεριλαμβάνει τη διόρθωση



CMAP. Παρατηρήθηκε η σύγκλιση των δύο προσομοιώσεων σε ότι αφορά την κινητικότητα της C-τελικής ουράς και η μικρή κινητικότητα των καταλοίπων 2-7, γεγονός το οποίο βρίσκεται σε συμφωνία με τα πειραματικά δεδομένα. Οι αλληλεπιδράσεις των πλευρικών ομάδων, που έχουν παρατηρηθεί πειραματικά αποδόθηκε μερικά από το δυναμικό πεδίο, αποσπασματικά σε κάποια cluster των δύο προσομοιώσεων. Εμφανής ήταν η σταθερότητα της α-ελικοειδούς διαμόρφωσης και στις δύο προσομοιώσεις, η οποία όμως δεν βρίσκεται σε συμφωνία με τα πειραματικά δεδομένα των καταχωρημένων στην PDB δομών.

Αν θέλαμε να υπερασπιστούμε το δυναμικό πεδίο, που χρησιμοποιήσαμε θα μπορούσαμε να το δικαιολογήσουμε για την τάση του για δημιουργία α-έλικας [111] με την λογική της θερμοδυναμικά σταθερότερης δομής της έλικας αυτής από αυτή της  $3_{10}$ . Θα μπορούσαμε να πούμε ότι η παραμετροποίηση είναι τόσο σωστή, ώστε το δυναμικό πεδίο να δημιουργεί θερμοδυναμικά σταθερές δομές και όχι δομές αμφισβητούμενης σταθερότητας, όπως οι  $3_{10}$ -έλικες, για τις οποίες θα πρέπει να επισημάνουμε ότι ούτε η ίδια η φύση έχει την τάση να τις δημιουργεί.

Στη συνέχεια του ρόλου μας, ως δικηγόρου του δυναμικού πεδίου CHARMM με CMAP correction, μπορούμε να προσθέσουμε την ικανοποιητική σύγκλιση των δύο προσομοιώσεων, αν και αυτές έγιναν σε πολύ διαφορετικές θερμοκρασίες και με διαφορετικές αρχικές δομές. Αν η θερμοκρασία ήταν  $37^{\circ}\text{C}$  (310K) θα παίρναμε διαφορετικό είδος έλικας; Επιπλέον να προσθέσουμε ότι η μητρική πρωτεΐνη του ενδεκαπεπτιδίου που χρησιμοποιήσαμε, η α-λακταλβουμίνη, δεν έχει δομή  $3_{10}$ -έλικας, όχι τουλάχιστον στο τμήμα των καταλοίπων 101-111, όπου είναι σαφώς α-ελικοειδής, όπως και το πεπτίδιο των προσομοιώσεων και συνεπώς οι προσομοιώσεις συγκλίνουν με την φυσική δομή του τμήματος 101-111 της α-λακταλβουμίνης και όχι με την πειραματικά προσδιορισθείσα. Ποιόν θα πρέπει να πιστέψουμε: την φύση ή το δομικό μοντέλο της PDB;

Και συνεχίζοντας ως δικηγόροι: Το τμήμα των καταλοίπων 101-111 της α-λακταλβουμίνης είναι τόσο διαφορετικής φυσικής διαμόρφωσης σε κάθε θηλαστικό, που ακόμα και με την ίδια ακριβώς αλληλουχία (βλ. καταχωρήσεις 1HFY και 1B9O της PDB) λαμβάνει τόσο διαφορετικές δομές στα διαφορετικά θηλαστικά στην αναδιπλωμένη του διαμόρφωση. Γιατί λοιπόν να μη λάβουμε υπόψη ότι το CHARMM

προσομοιώνει τη φύση και σ' αυτό, εφόσον κατάφερε να μοντελοποιήσει τόσο μοναδικά την πολυμορφικότητα του τμήματος αυτού;

Μήπως φταίει η πρωτονιωμένη ιστιδίνη, που δεν επέτρεψε στο δυναμικό πεδίο να δημιουργήσει δομικά μοντέλα σύμφωνα με τα πειραματικά; Μήπως αν χρησιμοποιούσαμε ασπαρτικό στην θέση 2 και όχι ασπαραγίνη θα μας έδινε ακόμα και σε αυτές τις θερμοκρασίες  $3_{10}$ -έλικα; [72] Μήπως φταίει το ίδιο το πειραματικά καθορισμένο μοντέλο εφόσον υπάρχει η υποψία ότι δεν αποδόθηκαν σωστά οι περιορισμοί αποστάσεων του NMR στο μοντέλο της PDB; [46] και αν αποδόθηκε σωστά το μοντέλο μήπως η  $3_{10}$ -έλικα αυτή λειτουργεί απλά ως το ενδιάμεσο της σταθερής  $\alpha$ -έλικας; [72,73] Εξάλλου το 20% του πληθυσμού του ενδεκαπεπτιδίου έδειξε κατά την διάρκεια καθορισμού της δομής του με NMR, ότι έχει περιορισμούς αποστάσεων, οι οποίες συνάδουν με  $\alpha$ -έλικα [46].

Και ως δικηγόροι επί πτωμάτων πατώντας, προσθέτουμε: Ναι, το χρονικό διάστημα των προσομοιώσεων αποδείχθηκε ικανοποιητικό. Αυτό δεν το αμφισβητούμε. Μήπως όμως η ανάλυση των τροχιακών δεν έγινε σωστά; Μήπως έπρεπε να αναθέσουμε την δευτεροταγή δομή και να αναλύσουμε τα τροχιακά και με άλλα προγράμματα; Είναι στραβός ο γιαλός ή στραβά αρμενίζουμε;

Για να μη μακρηγορούμε σταματάμε εδώ. Θα μπορούσαμε να βρούμε και άλλα ασήμαντα ελαφρυντικά για την αδυναμία του συγκεκριμένου δυναμικού πεδίου σε ότι αφορά την ικανότητα αναπαραγωγής  $3_{10}$ -έλικας. Το πρόβλημα όμως δεν είναι το συγκεκριμένο δυναμικό πεδίο. Το πρόβλημα είναι η εύρεση του μέσου (υπολογιστικές μέθοδοι;) και του τρόπου (ab initio μέθοδοι, ίσως;) μέσω των οποίων θα καταστεί δυνατή η επίλυση του προβλήματος της αναδίπλωσης των πεπτιδίων.

Κλείνοντας θα υπενθυμίσουμε την ρήση του Bohr την οποία αναφέραμε στην αρχή του παρόντος κεφαλαίου: Every sentence I utter must be understood not as an affirmation but as a question και θα τολμήσουμε να προσθέσουμε την φράση: . .and a question as a question and not a rhetorical one...

## Βιβλιογραφία

- [1] Baldwin RL, Zimm BH (2000) Are denatured proteins ever random coils? PNAS 97: 12391- 12392.
- [2] Anfinsen CB, Haber E, Sela M, White FH (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A: 47: 1309– 1314.
- [3] Haber E, Anfinsen CB (1962) Side-chain Interactions Governing the Pairing of Half-cystine residues in Ribonuclease. J. Biol. Chem. 237: 1839- 1844.
- [4] Levinthal C (1969) Are there pathways for protein folding? J. Chim. Physique 65: 44- 45.
- [5] Zwanzig R, Szabo A, Bagchi B (1992) Levinthal's paradox. Proc. Nail. Acad. Sci. USA 89: 20- 22.
- [6] Bergasa-Caceres F, Ronneberg A, Rabitz H (1999). Sequential Collapse Model for Protein Folding Pathways. J. Phys. Chem. B 44: 9749– 9758
- [7] Beck X, Allen C (2001) Diffusion-collision model calculations of protein folding. Thesis (PhD), pp. 915.
- [8] Chan HS, Bromberg S, Dill K (1995) Models of cooperativity in protein folding. Biol Sci 1323: 61- 70.
- [9] Vasilkoski Z, Weaver D (2004) Diffusion–collision model algorithms for protein folding kinetics. Journal of Computational Chemistry 2: 1101-1107.
- [10] Leopold PE, Montal M, Onuchic JN (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship PNAS 18: 8721- 8725.
- [11] Jackson SE (1998) How do small single-domain proteins fold? 4: 81- 91.
- [12] Kim PS, Baldwin RL (1990) Intermediates in the Folding Reactions of Small Proteins. Annu. Rev. Biochem (59): 631-660.
- [13] Karplus M, Weaver D (1979) Diffusion–collision model for protein folding Biopolymers. 18: 1421- 1437.
- [14] Karplus M, Weaver D (1994) The Protein Society Protein folding dynamics: The diffusion-collision model and experimental data Protein Science 3: 650- 668.
- [15] Arai M, Kuwajima K. (1996). Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. Fold Des 1: 275-287.

- [16] Agashe VR, Shastry MC, Udgaonkar JB. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature* 377: 754-757.
- [17] Ohgushi M, Wada A (1983). "'Molten-globule state': a compact form of globular proteins with mobile side-chains." *FEBS Lett.* 164: 21– 24.
- [18] Djikaev YS, Ruckenstein E (2007) Model for the Nucleation Mechanism of Protein Folding *J. Phys. Chem. B.* 111:886–897.
- [19] Karlus M (1997) The Levinthal paradox: yesterday and today. *Elsevier* 2: 69- 75.
- [20] Dill K, Chan SH (1997) From Levinthal to pathways to funnels. *Nature Structural Biology* 4:1.
- [21] Dill K, Chan SH (1998) Protein Folding in the Landscape Perspective: Chevron Plots and Non-Arrhenius Kinetics, *Proteins: Structure, Function, and Genetics* 301: 2-33.
- [22] France RM, Grossman SH(1996) Denaturation and urea gradient gel electrophoresis of arginine kinase: evidence for a collapsed-state conformation. *Arch Biochem Biophys* 326: 93-99.
- [23] Goldenberg DP (2001) Transverse urea-gradient gel electrophoresis. *Curr Protoc Protein Sci* 7: 7.4.
- [24] Kalnine NN, Schachman HK (2002) Quantitative urea gradient gel electrophoresis for studies of dissociation and unfolding of oligomeric proteins. *Bioph. Chem* 101-102: 133- 144.
- [25] Greenfield NJ (2006). Using circular Dichroism spectra to estimate protein secondary structure. *Nature protocols* 6: 2876– 2890.
- [26] Pace CN, Hermans J (1975) "The Stability of Globular Proteins", *CRC Critical Reviews in Biochemistry.* 3:1-43.
- [27] Pungor E (1995) *A Practical Guide to Instrumental Analysis.* Florida: Boca Raton. pp. 181–191.
- [28] O'Neill MJ (1964) The Analysis of a Temperature-Controlled Scanning Calorimeter. *Anal. Chem.* 36: 1238– 1245.
- [29] Wider G (2000) Structure determination of biological macromolecules in solution using NMR spectroscopy. *BioTechniques* 29: 1278- 1294.
- [30] Hongtao Y (1999) Extending the size limit of protein nuclear magnetic resonance. *Proc. Natl. Acad. Sci.* 96: 332- 334.

- [31] Garrett DS, Seok YJ, Peterkofsky A, Gronenborn AM, Clore GM (1999) Solution structure of the 40000 M phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. Struct. Biol.* 6: 166- 173.
- [32] Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, Kay LE (2000) Global folds of proteins with low densities of NOEs using dipolar couplings: application to the 370-residue maltodextrin binding protein. *J.Mol.Biol* 300: 197- 112.
- [33] Schlick, T (2002) *Molecular Modeling and Simulation: An Interdisciplinary Guide.* Springer-Verlag, New York.
- [34] "Peptide Definition - Medical Dictionary Definitions of Popular Medical Terms Easily Defined on MedTerms." Web. 23 Mar. 2010.  
<<http://www.medterms.com/script/main/art.asp?articlekey=24643>>.
- [35] "Peptide: Definition from Answers.com." Answers.com: Wiki Q&A Combined with Free Online Dictionary, Thesaurus, and Encyclopedias. Web. 23 Mar. 2010.  
<<http://www.answers.com/topic/peptide>>.
- [36] "Peptide - Definition of Peptide by the Free Online Dictionary, Thesaurus and Encyclopedia." Dictionary, Encyclopedia and Thesaurus - The Free Dictionary. Web. 23 Mar. 2010. <<http://www.thefreedictionary.com/peptide>>.
- [37] "Peptide Definition - Definition of Peptide." Chemistry - Periodic Table, Chemistry Projects, and Chemistry Homework Help. Web. 23 Mar. 2010.  
<<http://chemistry.about.com/od/chemistryglossary/g/peptide-definition.htm>>.
- [38] "Proteinogenic Amino Acid." Wikipedia, the Free Encyclopedia. Web. 23 Mar. 2010. <[http://en.wikipedia.org/wiki/Proteinogenic\\_amino\\_acid](http://en.wikipedia.org/wiki/Proteinogenic_amino_acid)>.
- [39] "Peptide." Wikipedia, the Free Encyclopedia. Web. 23 Mar. 2010.  
<<http://en.wikipedia.org/wiki/Peptide>>.
- [40] Bernstein FC, Koetzle TF, Williams GJB, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112: 535-542.
- [41] "RCSB Protein Data Bank." Web. 23 Mar. 2010.  
<<http://www.pdb.org/pdb/home/home.do>>.
- [42] Osterhout JJ (2005) Understanding protein folding through peptide models. *Protein Pept Lett* 12: 159- 164.
- [43] Rizo J, Gierasch M (1992) Constrained Peptides: Models of Bioactive Peptides and Protein Substructures. *Annual Review of Biochemistry* 61: 387- 416.
- [44] Suenaga A, Narumi T, Futatsugi N, Yanai R, Ohno Y, Okimoto N, Taiji M (2007) Folding dynamics of 10-residue  $\beta$ -hairpin peptide chignolin. *Chemistry* 2: 591- 598.

- [45] Thomas A, Deshaves S, Decaffmever M, Van Eyck MH, Charloteaux B, Brasseur R (2006). Prediction of peptide structure: how far are we? *Proteins* 65: 889-897.
- [46] Araki M, Tamura A (2007) Transformation of an alpha-helix peptide into a beta-hairpin induced by addition of a fragment results in creation of a coexisting state. *Proteins* 66: 860-868.
- [47] Demarest SJ, Hua Y, Raleigh DP (1999) Local interactions drive the formation of nonnative structure in the denatured state of human alpha-lactalbumin: a high resolution structural characterization of a peptide model in aqueous solution. *Biochemistry* 38: 7380- 7387.
- [48] Gnanakaran S, Garcia AE (2005) Helix-coil transition of alanine peptides in water: Force field dependence on the folded and unfolded structures. *Proteins* 59: 773- 782.
- [49] Leopold P, Montal M, Onuchic N (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship. *Analytical theory of probability flow. Biophysics* (89): 8721-8725.
- [50] Onuchic HN (1997) Theory of protein folding: The Energy Landscape Perspective. *Physical Chemistry* 48: 545- 600.
- [51] Montalvo G, Waagele MM, Shandler S, Gai F, DeGrado WF (2010) Infrared signature and folding dynamics of a helical beta-peptide. *J Am Chem Soc* 132: 5616-5618.
- [52] Lin Z, Kornfeld J, Machler M, van Gunsteren WF (2010) Prediction of folding equilibria of differently substituted peptides using one-step perturbation. *J Am Chem Soc* 132: 7276- 7278.
- [53] Mohammed OF, Jas GS, Lin MM, Zewail AH (2009) Primary peptide folding dynamics observed with ultrafast temperature jump *Angew Chem Int Ed Engl* 48: 5628-5632.
- [54] Greenfield NJ (2006) Analysis of the kinetics of folding of proteins and peptides using circular dichroism *Nat Protoc* 1: 2891- 2899.
- [55] Jiang X, Chen C, Xiao Y (2010) Improvements of network approach for analysis of the folding free-energy surface of peptides and proteins. *J Comput Chem* 31: 2502-2509.
- [56] Khandogin J, Brooks CL 3rd (2007) Linking folding with aggregation in Alzheimer's beta-amyloid peptides. *Proc Natl Acad Sci U S A* 104: 16880- 16885.
- [57] Morra G, Meli M, Colombo G (2008) Molecular dynamics simulations of proteins and peptides: from folding to drug design. *Curr Protein Pept Sci.* 9: 181-196.
- [58] Thorpe IF, Zhou J, Voth GA (2008) Peptide folding using multiscale coarse-grained models. *J. Phys Chem B* 112: 13079-13090.

- [59] Khandogin J, Brooks CL (2007) Linking folding with aggregation in Alzheimer's betaamyloid peptides. *Proc Natl Acad Sci U S A* 104: 16880-1 6885.
- [60] Gooding E, Ramajo AP, Wang J, Palmer C, Fouts E, Volk M (2005). The effects of individual amino acids on the fast folding dynamics of helical peptides. *Chem.Commun* 5985- 5987.
- [61] Trinkl S, Glockshuber R, Jaenicke R (1994). Dimerization of beta B2-crystallin: the role of the linker peptide and the N- and C-terminal extensions. *Protein Sci.* 3: 1392–1400.
- [62] Ramajo AP, Petty S, Volk M (2006). Fast folding dynamics of  $\alpha$ -helical peptides – Effect of solvent additives and pH. *Chem Physics* 323: 11- 20.
- [63] Dyson HJ, Wright PE (1993) Peptide conformation and protein folding. *Current Opinion in Structural Biology* 3: 60- 65.
- [64] Wimley W, Gawrisch K, Creamers T, White S (1996). Direct measurement of salt-bridge solvation energies using a peptide model system: Implications for protein stability *Biochemistry* 93: 2985- 2990.
- [65] Ho BK, Dill KA (2006) Folding Very Short Peptides Using Molecular Dynamics. *PLoS Comput Biol* 2: 27.
- [66] Oas TG, Kim PS (1988). A peptide model of a protein folding intermediate. *Nature* 336: 42- 48.
- [67] Gnanakaran S, Nymeyer H, Portman J, Sanbonmatsu KY, Garcia AE (2003) Peptide folding simulations. *Curr Opin Struct Biol* 13: 168- 174.
- [68] Barlow, D.J. and Thornton, J.M. (1988) *J. Mol. Biol* 201: 601– 619
- [69] Karpen, M.E., DeHaseth, P.L. and Neet, K.E. (1992) *Protein Sci.* 1: 1333– 1342.
- [70] Baker, E.N. and Hubbard, R.E. (1984) *Prog. Biophys. Mol. Biol.* 44: 97– 179.
- [71] Enkhbayar P, Hikichi K, Osaki M, Kretsinger RH, Matsushima N (2006) 3(10)-helices in proteins are parahelices. *Proteins* 64: 691-699.
- [72] Karpen M, De Haseth PL, Neet KE (1992) Differences in the amino acid distributions of 3<sub>10</sub>-helices and  $\alpha$ -helices. *The Protein Society* 1:1333-1342.
- [73] Millhauser GL (1995) Views of Helical Peptides: A Proposal for the Position of 310-Helix along the Thermodynamic Folding Pathway. *Biochemistry* 34: 3873– 3877.
- [74] Toniolo C, Benedetti E (1991) The polypeptide 310-helix. *Trends Biochem Sci* 16: 350- 353.

- [75] Hall L, Emery DC, Davies MS, Parker D, Craig RK (1987) Organization and sequence of the human  $\alpha$ -lactalbumin gene. *Biochem. J.* 242: 735- 742.
- [76] Harata K, Abe Y, Muraki M (1999) Crystallographic evaluation of internal motion of human  $\alpha$ -lactalbumin refined by full-matrix least-squares method. *J.Mol.Biol.* 287: 347- 358.
- [77] Permyakov EA (2005)  $\alpha$ -lactalbumin. Nova Science Publishers, New York, pp 18-19.
- [78] Pike CWA, Brew K, Acharya KR (1996) Crystal structures of guinea-pig, goat and bovine  $\alpha$ -lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure* 4: 691- 703.
- [79] Sayle R, Milner-White EJ (1995) "RasMol: Biomolecular graphics for all". *Trends in Biochemical Sciences* 20: 374
- [80] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, and Schulten K (2005) Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 26: 1781- 1802
- [81] Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Molec Graphics* 14: 33- 38.
- [82] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comp Chem* 4: 187- 217.
- [83] Frenkel D, Smit B (2000) *Understanding Molecular Simulation*. Academic Press pp 1- 6.
- [84] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AN, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21: 1087-1092.
- [85] Metropolis N (1987) The beginning of the Monte Carlo method. *Los Alamos Science* 12: 125- 130.
- [86] Fermi E, Pasta JG, Ulam SM (1955) Studies of non-linear problems. LASL Report LA-1940.
- [87] Alder BJ, Wainwright TE (1958) Molecular dynamics by electronic computers. *Proc of the Int Symp on Statistical Mechanical Theory of Transport Processes*, Interscience Wiley, New York: pp. 97-131.
- [88] Gibson JB, Goland AN, Milgram M, Vineyard GH (1960) Dynamics of radiation damage. *Phys Rev* 120: 1229- 1253.



- [89] Rahman A (1964) Correlation in the motion of atom in liquid argon. *Phys Rev* 136: 405- 411.
- [90] McCammon JA, Gelin BR, Karplus M (1977) Molecular Dynamics Study of the Bovine Pancreatic Trypsin Inhibitor. *Nature* 267: 585-89.
- [91] Chandler D (1987) *Introduction to Modern Statistical Mechanics*. Oxford University Press. New York.
- [92] "Statistical Mechanics." Swiss EMBnet Node Server. Web. 12 Jul. 2010. <[http://www.ch.embnet.org/MD\\_tutorial/](http://www.ch.embnet.org/MD_tutorial/)>.
- [93] MacKerell et al, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem* 102:3586-3616.
- [94] Ponder JW, Case DA (2003) Force fields for protein simulations. *Adv. Prot. Chem.* 66: 27- 85.
- [95] Cheatham TE, Young MA (2001) Molecular dynamics simulation of nucleic acids: Successes, limitations and promise. *Biopolymers* 56: 232- 256.
- [96] Jorgensen W, Maxwell DS, Tirado-Rives J (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* 118: 11225– 11236.
- [97] Kaminski G, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *Phys. Chem. B* 105: 6474– 6487.
- [98] van Gunsteren WF, Berendsen, HJC (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*; BIOMOS bv, University of Groningen, Groningen.
- [99] van Gunsteren W F, Billeter SR, Eising AA.; Hunenberger PH, Kruger P, Mark AE, Scott WRP, Tironi IG (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; BIOMOS bv, Zurich.
- [100] Arnautova YA, Anna Jagielska A, Scheraga HA (2006) A New Force Field (ECEPP-05) for Peptides, Proteins, and Organic Molecules. *J. Phys Chem B* 110 (10), pp 5025–5044.
- [101] Nemethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HAJ (1992) Energy parameters in polypeptides. X. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides, *J. Phys. Chem.* 96: 6472-6484.
- [102] Ren P, Ponder JW (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J Comp Chem* Vol 23: 1497- 1506.

- [103] "Tinker: a tool for designing data-centric sensor networks" (2006) Elson J, Parker A. New York NY USA pp:350 – 357.
- [104] MacKerell AD, et al, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem* 102: 3586-3616.
- [105] Mackerell AD, Feig M, Brooks CL (2004) Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J Comp Chem*. Vol 25:11.
- [106] Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) NAMD2: Greater scalability for parallel molecular dynamics. *J Comput Phys* 151: 283-312.
- [107] Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Molec Graphics* 14: 33-38.
- [108] Glykos NM (2006) Carma: a molecular dynamics analysis program. *J Comp Chem* 27: 1765- 1768.
- [109] Srinivasan, Raj. "RIBOSOME - PROGRAM TO BUILD COORDINATES FOR PEPTIDES FROM SEQUENCE." RoseLab Home Page. Web. 28 May 2010. <<http://www.roselab.jhu.edu/~raj/Manuals/ribosome.html>>.
- [110] Heinig M, Frishman D (2004) STRIDE: a Web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl Acids Res* 32:W500-2.
- [111] Freddolino PL, Park S, Roux B, Schulten K (2009) Force Field Bias in Protein Folding Simulations. *Bioph J* 96: 3772-3780