



ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ

ΤΜΗΜΑ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ ΚΑΙ ΓΕΝΕΤΙΚΗΣ

Διπλωματική Εργασία

**Μελέτη αναδίπλωσης του
αντιμικροβιακού πεπτιδίου
VG16KRKP**

Εφραιμίδης Ευάγγελος Ορφέας
-1716-

Επιβλέπων καθηγητής:
Νικόλαος Μ. Γλυκός

Πίνακας περιεχομένων

<i>Περίληψη</i>	4
1. Εισαγωγή	5
1.1. Πρωτεΐνες & πρωτεϊνική αναδίπλωση.....	5
1.2. Το πρόβλημα της αναδίπλωσης πρωτεϊνών	6
1.3. Μοντέλα πρωτεϊνικής αναδίπλωσης	6
1.3.1. Diffusion-Collision Model	6
1.3.2. Hydrophobic Collapse Model	7
1.3.3. Nucleation Condensation Model.....	7
1.3.4. Energy Landscapes & Folding Funnels Model.....	7
1.4. Πειραματικές μέθοδοι ανάλυσης της πρωτεϊνικής δομής.....	9
2. Μοριακή Δυναμική	10
2.1. Προσομοιώσεις Μοριακής Δυναμικής (MD)	10
2.2. Δυναμικά Πεδία.....	12
2.2.1. Δεσμικές Μοριακές Αλληλεπιδράσεις.....	12
2.2.2. Μη-Δεσμικές Μοριακές Αλληλεπιδράσεις.....	13
2.3. Ο ρόλος του διαλύτη στις προσομοιώσεις μοριακής δυναμικής	14
3. Προσομοιώσεις Μοριακής Δυναμικής	16
3.1. Το αντιμικροβιακό πεπτίδιο VG16KRRP	16
3.2. Στόχος της παρούσας εργασίας.....	16
3.3. Υπολογιστικό σύστημα.....	17
3.3.1. Προσομοίωση με NAMD	17
3.3.2. Προετοιμασία συστήματος και πρωτόκολλο προσομοιώσεων	18

4.	<i>Αποτελέσματα</i>	20
4.1.	RMSD	20
4.2.	Ανάλυση δευτεροταγούς δομής	21
4.3.	Principal Component Analysis (PCA)	23
4.4.	Υπολογισμός NOE & σύγκριση με πειραματικά δεδομένα	25
5.	<i>Συμπεράσματα</i>	30
6.	<i>Βιβλιογραφία</i>	32

Περίληψη

Ο τομέας της Μοριακής Δυναμικής (Molecular Dynamics) αποτελεί πλέον ένα ευρέως διαδεδομένο εργαλείο ανάλυσης της δομής αλλά και της διαδικασίας αναδίπλωσης των βιομορίων. Στην παρούσα εργασία χρησιμοποιούνται δύο προσομοιώσεις Μοριακής Δυναμικής του πεπτιδίου VG16KRRP βασισμένες στα δυναμικά πεδία AMBER-14SB & AMBER99SB-STAR-ILDN και τα αποτελέσματα των πειραμάτων Πυρηνικού Μαγνητικού Συντονισμού (NMR) (Bhunia & Datta, 2016) προκειμένου να μελετηθεί η δομή και να πραγματοποιηθεί σύγκριση των τροχιακών με τη φυσική πραγματικότητα. Η αλληλουχία του πεπτιδίου είναι VARGWKRRKCPFLGKGG. Για την παραπάνω μελέτη, έγιναν αναλύσεις των τροχιακών των προσομοιώσεων εστιασμένες τόσο στη δευτεροταγή όσο και στην τριτοταγή διαμόρφωση του πεπτιδίου. Οι τεχνικές που εφαρμόστηκαν περιλαμβάνουν τον υπολογισμό RMSD, τη χρήση του αλγόριθμου STRIDE για τον εντοπισμό δευτεροταγών δομών, την ανάλυση PCA δίδρων γωνιών και τη σύγκριση των NOEs μεταξύ προσομοιώσεων και πειραμάτων NMR.

1. Εισαγωγή

1.1. Πρωτεΐνες & πρωτεϊνική αναδίπλωση

Οι πρωτεΐνες είναι βιομόρια, ποικίλου μοριακού βάρους, που δομούνται από μία ή περισσότερες αλυσίδες αμινοξικών καταλοίπων. Ο ρόλος των πρωτεϊνών έγκειται στην πραγματοποίηση μιας πληθώρας βιολογικών λειτουργιών όπως: ο πολλαπλασιασμός του γενετικού υλικού, οι ενζυμικές αντιδράσεις του μεταβολισμού, η διακυτταρική επικοινωνία, η συμμετοχή σε δομικές λειτουργίες στήριξης κ.α.

Τα αμινοξέα αποτελούνται από ένα κεντρικό άτομο άνθρακα (C_α), μία αμινομάδα (NH_2), μία καρβοξυλομάδα ($COOH$) και μία πλευρική αλυσίδα (R) χαρακτηριστική για κάθε αμινοξύ. Υπάρχουν 20 διαφορετικές πλευρικές αλυσίδες γεγονός που οδηγεί στην ύπαρξη 20 διαφορετικών αμινοξέων. Τα τελευταία συνδέονται μεταξύ τους με έναν πεπτιδικό δεσμό ο οποίος δημιουργείται ανάμεσα στην αμινομάδα ($-NH_2$) του ενός αμινοξέος και την καρβοξυλομάδα ($-COOH$) του επόμενου. Η γραμμική αλληλουχία των αμινοξικών καταλοίπων περιγράφεται από τον όρο «πρωτοταγής» δομή. Δεσμοί υδρογόνου μεταξύ του οξυγόνου $C=O$ και του υδρογόνου $N-H$ που συμμετέχουν σε διαφορετικό πεπτιδικό δεσμό οδηγούν στο σχηματισμό διαμορφώσεων «δευτεροταγούς» δομής όπως α -έλικες και β -πτυχωτά. Η διαδικασία της πρωτεϊνικής αναδίπλωσης αποτελεί το στάδιο κατά το οποίο η πρωτεΐνη αναδιπλώνεται κατάλληλα στον χώρο και καθίσταται βιολογικά ενεργή. Η τρισδιάστατη ή «τριτοταγής» διαμόρφωση που αποκτάται μετά την αναδίπλωση χαρακτηρίζεται ως «φυσική δομή» ή “native state”. Η διαδικασία της αναδίπλωσης μπορεί να συμβαίνει είτε αυθόρμητα είτε με τη συμβολή άλλων πρωτεϊνών^[10].

Η σωστή πρωτεϊνική λειτουργία εξαρτάται τόσο από την αλληλουχία καταλοίπων όσο και από την τριτοταγή δομή. Η διαφορά μεταξύ των πρωτεϊνών οφείλεται αρχικά στο γεγονός ότι από τον συνδυασμό των 20 αμινοξέων μπορεί να προκύψει ένας εξαιρετικά μεγάλος αριθμός πιθανών αλυσίδων και ακολούθως στο ότι μια αλληλουχία μπορεί να υιοθετήσει μια αντίστοιχα μεγάλη πληθώρα τρισδιάστατων διαμορφώσεων^[5].

1.2. Το πρόβλημα της αναδίπλωσης πρωτεϊνών

Βασικό και αναπάντητο ερώτημα της μελέτης των πρωτεϊνών αποτελεί ο τρόπος με τον οποίο επιλέγεται η φυσική δομή μεταξύ όλων των πιθανών διαμορφώσεων.

Η πρώτη απάντηση διατυπώθηκε μέσω της «θερμοδυναμικής υπόθεσης» του Christian Anfinsen το 1961. Η υπόθεσή του υποστηρίζει πως η φυσική δομή προκύπτει λόγω μέγιστης θερμοδυναμικής σταθερότητας, δηλαδή είναι η διαμόρφωση με την ελάχιστη ολική ελεύθερη ενέργεια. Συμπέρασμα του παραπάνω αποτελεί ότι η φυσική δομή καθορίζεται από το σύνολο των ενδοατομικών αλληλεπιδράσεων των αμινοξέων, επομένως η αλληλουχία καθ' αυτή περιέχει την πληροφορία που οδηγεί στην κατάλληλη αναδίπλωση της πρωτεΐνης^[3].

Στην παρουσίαση της μελέτης του το 1969, ο Cyrus Levinthal περιέγραψε ότι εφόσον οι πρωτεΐνες στη φύση αναδιπλώνονται σε ελάχιστα δευτερόλεπτα, είναι απίθανο αυτό να συμβαίνει περνώντας τυχαία από όλες τις πιθανές διαμορφώσεις καθώς για μια πρωτεΐνη 150 αμινοξέων υπολογίζεται να υπάρχουν 10^{300} πιθανές δομές. Προτάθηκε λοιπόν μια «κινητική θεωρία» που εισάγει τη δημιουργία μεταβατικών διατάξεων τοπικής θερμοδυναμικής σταθερότητας που οδηγούν την πρωτεΐνη σε καθορισμένα μονοπάτια αναδίπλωσης^[27].

1.3. Μοντέλα πρωτεϊνικής αναδίπλωσης

1.3.1. Diffusion-Collision Model

Το 1976 οι Karplus & Weaver πρότειναν ότι οι πρωτεΐνες αποτελούνται από πολυάριθμες μικροεπικράτειες χαμηλής σταθερότητας οι οποίες εντοπίζονται διάχυτες στον χώρο. Η κίνηση τους οδηγεί σε συγκρούσεις που ακολουθούνται από σύμπτυξη αυτών προς σχηματισμό μεγαλύτερων και σταθερότερων μικροδομών. Η συνεχής αυτή διαδικασία καταλήγει τελικά στην διαμόρφωση της ολικής δομής της πρωτεΐνης^[24].

1.3.2. Hydrophobic Collapse Model

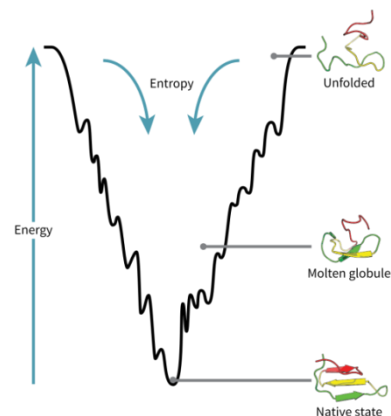
Βάσει του μοντέλου αυτού, νωρίς κατά τη διαδικασία της αναδίπλωσης σχηματίζονται αρχικές δευτεροταγείς δομές από πλευρικές αλυσίδες λόγω υδροφοβικών αλληλεπιδράσεων. Οι δομές αυτές λόγω της υδροφοβικής τους φύσης εισέρχονται στο εσωτερικό της δομής καθώς πολικά αμινοξέα τοποθετούνται στην επιφάνεια όπου εκτίθενται στο υδάτινο περιβάλλον. Η έμπνευση αυτού του μοντέλου προήλθε από την παρατήρηση ύπαρξης σφαιρικών πρωτεϊνών με υδρόφοβο πυρήνα^[41].

1.3.3. Nucleation Condensation Model

Βάσει του μοντέλου αυτού, πρώτο βήμα της αναδίπλωσης αποτελεί ο σχηματισμός ενός σταθερού πυρήνα ο οποίος θα αποτελέσει το θεμέλιο με βάση το οποίο θα σχηματιστούν όλες οι δομές που το περιβάλλουν^{[14][31]}.

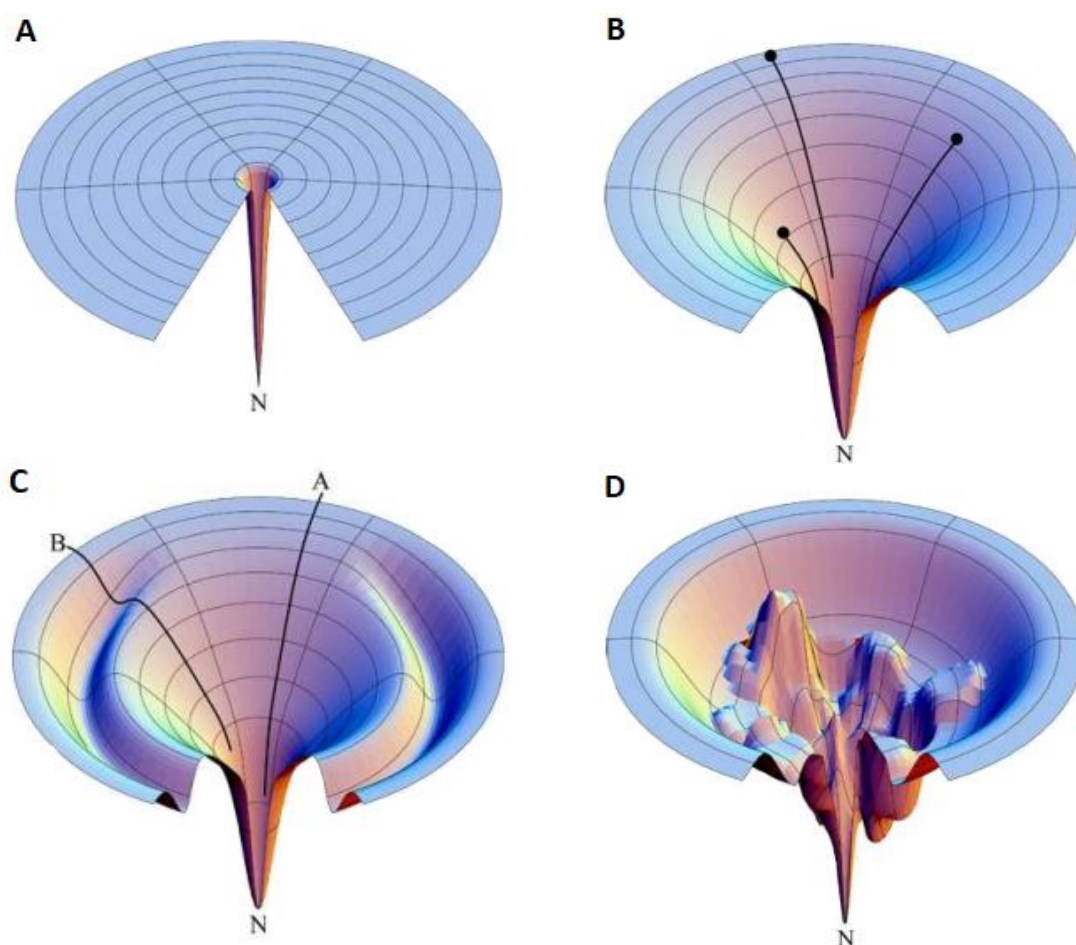
1.3.4. Energy Landscapes & Folding Funnels Model

Η πλέον διαδεδομένη θεωρία των energy landscapes καταφέρνει να συνδυάσει την θερμοδυναμική υπόθεση του Anfinsen και την κινητική υπόθεση του Levinthal θεωρώντας ενεργειακά τοπία σε σχήμα χωνιού. Για την κατανόηση της θεωρίας είναι σημαντικό να εισάγουμε τις έννοιες: της ενθαλπίας H (ως μέτρο θερμότητας), της εντροπίας S (ως μέτρο αταξίας) και της ελεύθερης ενέργειας Gibbs G . Η τελευταία δίνεται από τη σχέση $\Delta G = \Delta H - T\Delta S$. Η φυσική δομή της πρωτεΐνης είναι και η πιο θερμοδυναμικά σταθερή, άρα θα έχει και την ελάχιστη τιμή ελεύθερης ενέργειας. Τα ενεργειακά τοπία



Εικόνα 1: Απλή αναπαράσταση ενεργειακού τοπίου αναδίπλωσης. Παρατηρείται μείωση στην ελεύθερη ενέργεια Gibbs καθώς η πρωτεΐνη αναδιπλώνεται έως ότου τελικά φτάσει στην φυσική δομή της. (Αναπαραγωγή δίχως άδεια) ^[36]

αναπαρίστανται μέσω δισδιάστατων ή τρισδιάστατων διαγραμμάτων τα οποία παρουσιάζονται στην *Εικόνα 2* ως χωνιά. Το μοντέλο θεωρεί πως η πρωτεΐνη μπορεί να «κατρακυλήσει», μέσω πολυάριθμων μονοπατιών αναδίπλωσης, από την μη αναδιπλωμένη μορφή της στο πάνω μέρος του χωνιού (διαμορφώσεις με υψηλή ελεύθερη ενέργεια) προς τον πάτο αυτού όπου βρίσκεται η φυσική δομή (διαμόρφωση με το ελάχιστο ελεύθερης ενέργειας). Το βάθος του χωνιού αναπαριστά το ποσό της ελεύθερης ενέργειας ενώ το πλάτος την εντροπία^{[11][12][16][21]}.



Εικόνα 2: Χαρακτηριστικά ενεργειακά τοπία αναδίπλωσης πρωτεϊνών. Αναπαραγωγή δίχως άδεια από “Dill, K., Chan, H. (1997). From Levinthal to Pathways to Funnels. *Nature Structural & Molecular Biology*, 4(1), 10-19.”

1.4. Πειραματικές μέθοδοι ανάλυσης της πρωτεϊνικής δομής

Υπάρχουν αρκετές μέθοδοι που αναλύουν την τρισδιάστατη δομή μια πρωτεΐνης, οι τρεις σημαντικότερες και πιο διαδεδομένες αυτών είναι η Κρυσταλλογραφία Ακτίνων Χ, η Φασματοσκοπία Πυρηνικού Μαγνητικού Συντονισμού (NMR: Nuclear Magnetic Resonance) και η Κρυοηλεκτρονική Μικροσκοπία (Cryo-EM: Cryogenic Electron Microscopy).

Η κρυσταλλογραφία ακτίνων Χ βασίζεται στην απόκτηση δομικής πληροφορίας μέσω των προτύπων διάθλασης που προκύπτουν κατόπιν έκθεσης καθαρών δειγμάτων που έχουν υποστεί κρυστάλλωση σε δέσμη ακτίνων Χ. Η αποτελεσματικότητα της μεθόδου περιορίζεται σε πρωτεΐνες οι οποίες έχουν τη δυνατότητα να διαμορφώσουν σταθερές και ομοιόμορφες κρυσταλλικές δομές^[35].

Η φασματοσκοπία NMR βασίζεται στην ανάλυση της ακτινοβολίας που εκπέμπεται από τους πυρήνες της πρωτεΐνης όταν αυτοί εκτίθενται σε ένα ελεγχόμενο εξωτερικό μαγνητικό πεδίο.

Η Cryo-EM βασίζεται στην ανάλυση θολών εικόνων που προκύπτουν από την ακτινοβολία της πρωτεΐνης η οποία πρώτα έχει υποστεί γρήγορο πάγωμα μέσα σε λεπτή στρώση νερού. Η χρήση υπολογιστικών συστημάτων για την ταξινόμηση και επεξεργασία πολλών χιλιάδων τέτοιων εικόνων παρέχει τη δυνατότητα δημιουργίας τρισδιάστατων χαρτών υψηλής ανάλυσης^[4].

2. Μοριακή Δυναμική

2.1. Προσομοιώσεις Μοριακής Δυναμικής (MD)

Η Μοριακή Δυναμική είναι ο τομέας της επιστήμης που χρησιμοποιεί υπολογιστικές μεθόδους για να προσομοιώσει την συμπεριφορά των ατόμων και των μορίων συναρτήσει του χρόνου. Χρησιμεύει ιδιαίτερα στον καθορισμό τρισδιάστατων μοριακών δομών. Επίσης, παρέχει τη δυνατότητα επιβεβαίωσης πειραματικών δεδομένων ή ακόμα την πρωταρχική μελέτη συστημάτων πριν τεθούν στη χρονοβόρα και συνήθως ακριβή διαδικασία μελέτης με εργαστηριακές μεθόδους. Σε μια προσομοίωση καθορισμένης χρονικής διάρκειας, τα τροχιακά των σωματιδίων ενός συστήματος μπορούν να υπολογιστούν με τη βοήθεια του δεύτερου νόμου του Newton:

$$F = ma \quad (1)$$

όπου F είναι η δύναμη που ασκείται στο σωματίδιο (ή το άθροισμα των δυνάμεων αυτών), m η μάζα του σωματιδίου και a η επιτάχυνσή του.

Το τροχιακό που προκύπτει από την επίλυση του δεύτερου νόμου για όλα τα σωματίδια του συστήματος περιγράφει και τις τρεις συνιστώσες της εξίσωσης για κάθε σωματίδιο, την κάθε χρονική στιγμή της προσομοίωσης.

Η εξαρτώμενη της δυναμικής ενέργειας έκφραση της δύναμης δίνεται από τον τύπο:

$$F = -\frac{dV}{dr} \quad (2)$$

όπου V η δυναμική ενέργεια και r θέση του σωματιδίου.

Ο συνδυασμός των εξισώσεων (1) & (2) οδηγεί στις σχέσεις:

$$a = -\frac{1}{m} \frac{dV}{dr} \quad (3)$$

$$-\frac{dV}{dr} = m \frac{d^2r}{dt^2} \quad (4)$$

Εξαιτίας των παραπάνω σχέσεων, για τον υπολογισμό ενός τροχιακού αρκεί να είναι γνωστές οι αρχικές θέσεις των σωματιδίων, οι επιταχύνσεις από τις αρχικές θέσεις, η αρχική κατανομή των ταχυτήτων και η διαβάθμιση δυναμικής ενέργειας.

Η εύρεση των αρχικών θέσεων μπορεί να πραγματοποιηθεί με πειραματικές μεθόδους ανάλυσης της δομής όπως η κρυσταλλογραφία ακτινών Χ και η φασματοσκοπία NMR.

Η αρχική κατανομή των ταχυτήτων μπορεί να υπολογιστεί μέσω της κατανομής Maxwell-Boltzmann ή της κατανομής Gaussian.

$$p(V_x) = \left(\frac{m}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left(-\frac{mv_x^2}{2k_B T}\right) \quad (5)$$

όπου V_x η ταχύτητα ενός ατόμου προς τη διεύθυνση x , T η θερμοκρασία και k_B η σταθερά του Boltzmann^[20].

Η αρχική επιτάχυνση εξάγεται από τον υπολογισμό της δυναμικής ενέργειας ο οποίος ωστόσο αποτελεί ένα ιδιαίτερα απαιτητικό και εξαιρετικά χρονοβόρο έργο. Για τον λόγο αυτό έχουν σχεδιαστεί αρκετοί αλγόριθμοι όπως οι: Verlet, Leap-Frog, Velocity Verlet, Beeman's. Οι αλγόριθμοι αυτοί αποκαλούνται αλγόριθμοι ολοκλήρωσης (integration algorithms) και η στρατηγική τους επιδιώκει την ει δυνατόν μείωση των όρων μιας εξίσωσης θεωρώντας πως οι θέσεις, οι ταχύτητες και οι επιταχύνσεις μπορούν να υπολογιστούν προσεγγιστικά με τη χρήση της επέκτασης σειράς Taylor (Taylor series expansion).

$$r(t + dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 + \dots \quad (6)$$

$$v(t + dt) = v(t) + a(t)dt + \frac{1}{2}b(t)dt^2 + \dots \quad (7)$$

$$a(t + dt) = a(t) + b(t)dt + \dots \quad (8)$$

όπου r είναι η θέση, v η ταχύτητα και a η επιτάχυνση^[20].

Επειδή οι αλγόριθμοι ολοκλήρωσης χαρακτηρίζονται από το αρνητικό χαρακτηριστικό μη υψηλής ακρίβειας, για να παραχθούν τα βέλτιστα αποτελέσματα είναι απαραίτητο να επιλεγεί ο κατάλληλος αλγόριθμος βάσει διάφορων κριτηρίων.

2.2. Δυναμικά Πεδία

Όπως έχει γίνει σαφές, ο υπολογισμός της δυναμικής ενέργειας ενός συστήματος είναι απαραίτητος για την πραγματοποίηση μιας προσομοίωσης. Η εξίσωση του Schrödinger μπορεί να παρέχει αυτή την πληροφορία, ωστόσο η επίλυσή της αποτελεί μια εξαιρετικά χρονοβόρα διαδικασία ακόμα και για πολύ ισχυρά υπολογιστικά συστήματα. Η εναλλακτικές μέθοδοι υπολογισμού της δυναμικής ενέργειας που έχουν αναπτυχθεί είναι τα εμπειρικά δυναμικά πεδία (Force Fields)^{[20][28]}. Τα πιο διαδεδομένα δυναμικά πεδία είναι τα: AMBER (Assisted Model Building with Energy Refinement)^[37], CHARMM (Chemistry at Harvard Macromolecular Mechanics)^[19], GROMOS (Groningen Molecular Simulation)^[8], OPLS (Optimized Potential for Liquid Simulations)^[23]. Κατά συνθήκη, τα δυναμικά πεδία εκφράζουν την ολική δυναμική ενέργεια του συστήματος ως άθροισμα της ενέργειας των δεσμικών και αυτής των μη-δεσμικών αλληλεπιδράσεων.

$$V = E_{bonding} + E_{non-bonding} \quad (9)$$

2.2.1. Δεσμικές Μοριακές Αλληλεπιδράσεις

Ο όρος $E_{bonding}$ υπολογίζεται ως το άθροισμα:

$$E_{bonding} = E_{stretching} + E_{bending} + E_{rotation} \quad (10)$$

Ο όρος $E_{stretching}$ είναι ένα αρμονικό δυναμικό που αφορά την αλληλεπίδραση μεταξύ δυο ατόμων συνδεδεμένα με έναν ομοιοπολικό δεσμό. Συγκεκριμένα περιγράφει την απόκλιση του μήκους δεσμού από την ιδανική τιμή του b_0 . Στην σχέση (11) που ακολουθεί, το K_b είναι μια σταθερά που αντιπροσωπεύει το σθένος δεσμού^[20].

$$E_{stretching} = \sum_{1,2 \text{ pairs}} K_b (b - b_0)^2 \quad (11)$$

Ο όρος $E_{bending}$ είναι ένα αρμονικό δυναμικό που περιγράφει την απόκλιση της γωνίας του δεσμού από την ιδανική τιμή της θ_0 .

$$E_{bending} = \sum_{angles} K_{\theta}(\theta - \theta_0)^2 \quad (12)$$

Για τις $E_{stretching}$ και $E_{bending}$, οι τιμές b_0 και θ_0 αντίστοιχα εξαρτώνται από τον τύπο των ατόμων που συμμετέχουν στους δεσμούς που περιγράφονται.

Ο όρος $E_{rotation}$ υπολογίζει τη δυναμική ενέργεια που σχετίζεται με την περιστροφή των δίεδρων γωνιών μεταξύ 4 ατόμων συνδεδεμένων μεταξύ τους με 3 ομοιοπολικούς δεσμούς. Μέσω αυτού του όρου λαμβάνονται υπόψη οι πιθανοί στερεοχημικοί φραγμοί περιστροφής που υπάρχουν μεταξύ ατόμων^[20].

$$E_{rotation} = \sum_{1,4 \text{ pairs}} K_{\varphi}(1 - \cos(n\varphi)) \quad (13)$$

2.2.2. Μη-Δεσμικές Μοριακές Αλληλεπιδράσεις

Ο όρος $E_{non-bonding}$ υπολογίζεται ως το άθροισμα της ενέργειας των αλληλεπιδράσεων Van der Waals και της ενέργειας των ηλεκτροστατικών αλληλεπιδράσεων:

$$E_{non-bonding} = E_{Van-der-Waals} + E_{electrostatic} \quad (14)$$

Οι αλληλεπίδραση Van der Waals μεταξύ ατόμων προκύπτουν όταν οι ελκτικές και οι απωστικές δυνάμεις μεταξύ των ατόμων έρχονται σε ισορροπία και επιτελούν πολύ σημαντικό ρόλο στη σταθερότητα των μακρομορίων. Οι δυνάμεις άπωσης δημιουργούνται όταν τα ηλεκτρόνια βρίσκονται σε κοντινές αποστάσεις μεταξύ τους. Οι δυνάμεις έλξης προκύπτουν λόγω των παροδικών δίπολων που σχηματίζονται λόγω των μεταβολών φορτίου που παρατηρούνται στα ηλεκτρονικά νέφη. Τα δίπολα αυτά επάγουν με τη σειρά τους αντίστοιχα παροδικά δίπολα σε γειτονικά άτομα με αποτέλεσμα την γέννηση ελκτικής δύναμης. Οι αλληλεπιδράσεις αυτές εξαρτώνται από την απόσταση r μεταξύ των ατόμων. Μέγιστη σταθερότητα παρατηρείται σε συγκεκριμένη απόσταση που αποκαλείται απόσταση ισορροπίας και χαρακτηρίζεται από ελάχιστο ελεύθερης ενέργειας Gibbs^[20].

Για την έκφραση της ενέργειας των αλληλεπιδράσεων Van der Waals συνήθως χρησιμοποιείται το δυναμικό Lennard-Jones 6-12. Στο δυναμικό αυτό συμμετέχουν οι σταθερές A και C που εξαρτώνται από τον τύπο των ατόμων που αλληλεπιδρούν.

$$E_{Van-der-Waals} = \sum_{non-bonded\ pairs} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^6} \right) \quad (15)$$

Στην έκφραση της ενέργειας των ηλεκτροστατικών αλληλεπιδράσεων χρησιμεύει το δυναμικό του Coulomb.

$$E_{electrostatic} = \sum_{non-bonded\ pairs} \frac{q_i q_k}{D r_{ik}} \quad (16)$$

Όπου q_i και q_k είναι τα φορτία των δυο ατόμων, r η μεταξύ τους απόσταση και D η διηλεκτρική σταθερά του υλικού μέσου^[20].

2.3. Ο ρόλος του διαλύτη στις προσομοιώσεις μοριακής δυναμικής

Κρίσιμο ρόλο στον καθορισμό της δομής μιας πρωτεΐνης παίζει το περιβάλλον της και οι αλληλεπιδράσεις της με αυτό. Όπως στη φύση, έτσι και στις προσομοιώσεις μοριακής δυναμικής, το μακρομόριο περιβάλλεται από έναν διαλύτη ο οποίος επιδρά τόσο στις θερμοδυναμικές παραμέτρους του συστήματος όσο και στις ηλεκτροστατικές αλληλεπιδράσεις μεταξύ των μορίων. Ο πιο συνηθισμένος διαλύτης είναι το νερό. Τα μόριά του, έχουν την ιδιότητα να αλληλεπιδρούν με τα γειτονικά τους σχηματίζοντας δεσμούς υδρογόνου. Η εισαγωγή του διαλύτη σε μια προσομοίωση μπορεί γίνει με δύο τρόπους. Η πρώτη μέθοδος “implicit modeling” θεωρεί πως τα μόρια του διαλύτη μπορούν να αντικατασταθούν από μια καθολική διηλεκτρική σταθερά αντιπροσωπευτική των ιδιοτήτων του διαλύτη. Η παραδοχή αυτή οδηγεί σε προσεγγίσεις που επηρεάζουν την ακρίβεια των αποτελεσμάτων. Η δεύτερη μέθοδος “explicit modeling” περιλαμβάνει στους υπολογισμούς την ακριβή συμπεριφορά των μορίων του διαλύτη. Μια τέτοια διαδικασία απαιτεί φυσική οριοθέτηση προκειμένου να εξασφαλίζεται η μη διάχυση των μορίων του διαλύτη μακριά από το μακρομόριο και να παρέχεται η δυνατότητα χρήσης περιορισμένου αριθμού μορίων διαλύτη^[9].

Μερικά ευρέως διαδεδομένα μοντέλα νερού είναι τα TIP3P (Transferable Intermolecular Potential 3P)^[30], F3C (Flexible Three Center)^[39], SPC/E (Extended Simple Point Charge)^[30]. Τα παραπάνω μοντέλα θεωρούν ότι κάθε μόριο νερού διαθέτει τρεις θέσεις αλληλεπίδρασης ενώ τα άτομά του συμπεριφέρονται ως άκαμπτες σφαίρες που υπακούουν στα δυναμικά Lennard-Jones και Coulomb.

3. Προσομοιώσεις Μοριακής Δυναμικής

3.1. Το αντιμικροβιακό πεπτίδιο VG16KRKP

Το πεπτίδιο VG16KRKP είναι ένα τεχνητό ανάλογο της ικής πρωτεΐνης VG16. Έχει μήκος 16 αμινοξέα και μπορεί να σχεδιαστεί στο εργαστήριο μεταλλάσσοντας την αλληλουχία στα κατάλοιπα 2, 6, 7, 8 και 10^[15].

VG16 V D R G W G N G C G L F G K G G

VG16KRKP V A R G W K R K C P L F G K G G

Η σημασία της μελέτης του πεπτιδίου αυτού έγκειται στην αντιμικροβιακή του δράση έναντι αρκετών Gram(-) βακτηρίων.

3.2. Στόχος της παρούσας εργασίας

Στην παρούσα εργασία συγκρίνονται δύο προσομοιώσεις μοριακής δυναμικής με τα πειραματικά δεδομένα NMR (Bhunia & Datta, 2016). Οι προσομοιώσεις είναι βασισμένες στα δυναμικά πεδία AMBER-14SB^[29] και AMBER99SB-STAR-ILDN. Ο στόχος της μελέτης είναι να συγκριθεί η φυσική πραγματικότητα με τα μοντέλα των τροχιακών μοριακής δυναμικής.

Η ανάλυση των δεδομένων από τις προσομοιώσεις μοριακής δυναμικής υλοποιήθηκε με τη χρήση του προγράμματος *carma* και του αυτοματοποιημένου γραφικού περιβάλλοντος αυτού *grcarma*^[17] ^[26].

3.3. Υπολογιστικό σύστημα

Οι προσομοιώσεις της εν λόγω μελέτης εκτελέστηκαν μέσω της Norma. Η Norma, είναι ένα υπολογιστικό συγκρότημα κλάσης Beowulf. Βρίσκεται στο Τμήμα Μοριακής Βιολογίας & Γενετικής του Δημοκρίτειου Πανεπιστημίου Θράκης στην πόλη της Αλεξανδρούπολης και λειτουργεί από τις 29 Ιανουαρίου 2009.

Η Norma αποτελείται από 40 CPU πυρήνες, 46 GB φυσικής μνήμης και 6 GP GPUs μοιρασμένα σε 10 κόμβους. Όλοι οι κόμβοι συνδέονται μεταξύ τους με ειδικό Ethernet switch *HP ProCurve 1800-24G*. Οι 8 κόμβοι προσφέρουν στο σύστημα από έναν τετραπύρρηνο επεξεργαστή *Intel Q6600 Kentsfield 2.4GHz*, 4 GB φυσικής μνήμης και δύο διασυνδέσεις δικτύου. Τέσσερις από αυτούς προσφέρουν επίσης από μία κάρτα γραφικών *Nvidia GTX-460*. Ο ένατος κόμβος προσφέρει στο σύστημα έναν τετραπύρρηνο επεξεργαστή *Intel i7-965 Extreme*, 6 GB φυσικής μνήμης, μια κάρτα γραφικών *Nvidia GTX-295* και δύο διασυνδέσεις δικτύου. Ο κεντρικός κόμβος παρέχει επίσης τέσσερις πυρήνες επεξεργασίας, 8 GB φυσικής μνήμης, μια κάρτα γραφικών *Nvidia GTX-260*, μια συστοιχία τεσσάρων δίσκων συνολικής χωρητικότητας 1.5 TB (συνδεδεμένων βάσει της τεχνολογίας RAID-5) και τρεις διασυνδέσεις δικτύου. Η Norma χρησιμοποιείται σχεδόν αποκλειστικά για μελέτες υπολογιστικής βιολογίας και κρυσταλλογραφίας από την ομάδα Δομικής και Υπολογιστικής Βιολογίας του Δημοκρίτειου Πανεπιστημίου Θράκης^[38].

3.3.1. Προσομοίωση με NAMD

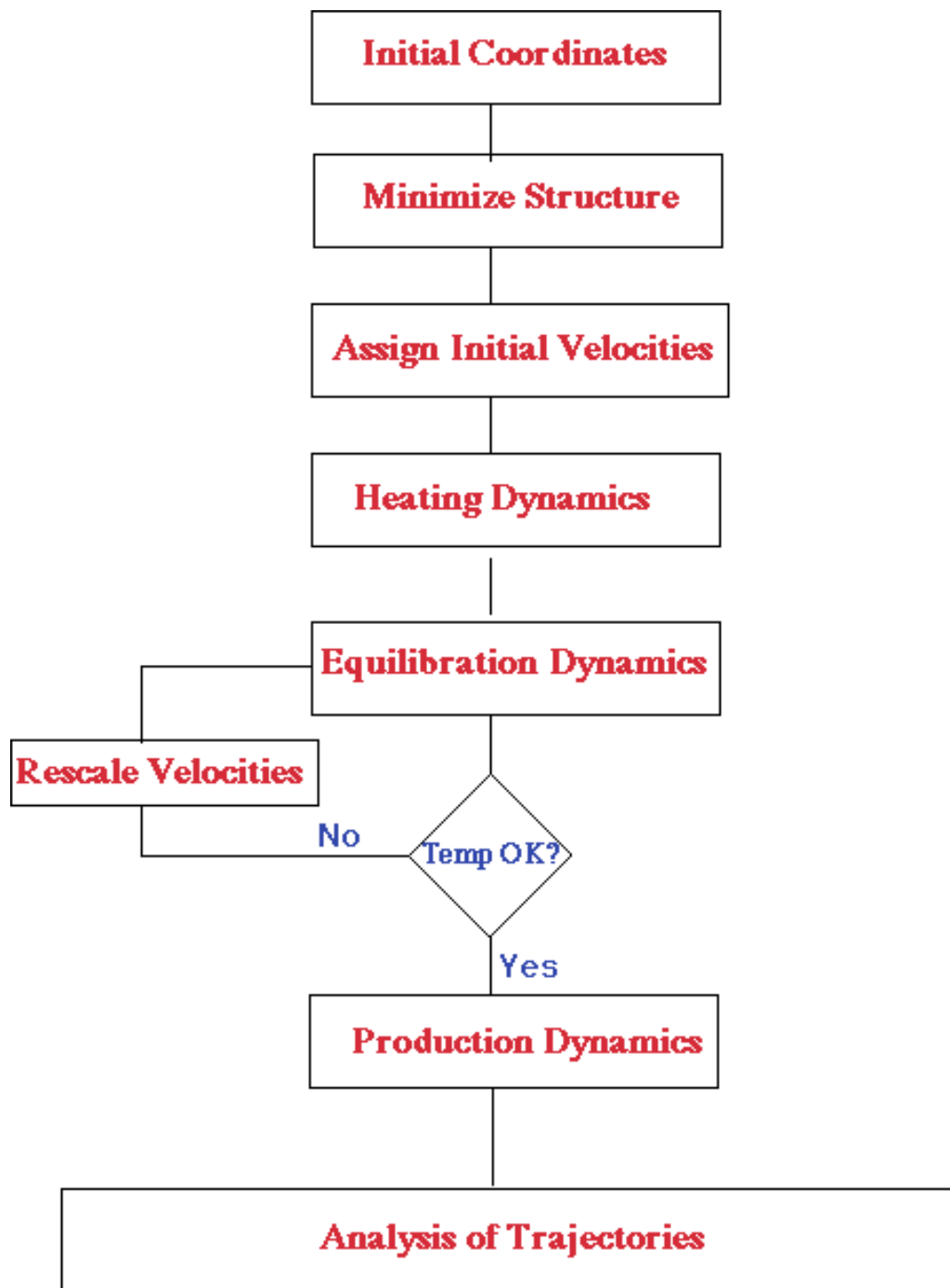
Οι προσομοιώσεις διεξήχθησαν μέσω του λογισμικού NAMD (Nanoscale Molecular Dynamics). Το NAMD αναπτύχθηκε από την Ομάδα Θεωρητικής και Υπολογιστικής Βιοφυσικής στο Ινστιτούτο Προηγμένης Επιστήμης και Τεχνολογίας του Beckman στο Πανεπιστήμιο του Illinois στην Urbana-Champaign^[22]. Προκειμένου να ξεκινήσει μια προσομοίωση NAMD απαιτούνται τρία αρχεία:

- Ένα PDB αρχείο (.pdb) συνήθως αντλημένο από τη βάση δεδομένων Protein Data Bank. Το αρχείο αυτό περιέχει όλες τις αρχικές συντεταγμένες του μοριακού συστήματος.

- Ένα αρχείο (.prmtop) που περιγράφει όλες τις παραμέτρους του δυναμικού πεδίου που θα χρησιμοποιηθεί (AMBER-14SB & AMBER99SB-STAR-ILDN) και είναι συμβατό με το λογισμικό NAMD.
- Ένα αρχείο διαμόρφωσης (configuration file) που περιλαμβάνει πληθώρα απαραίτητων πληροφοριών και επιλογών σχετιζόμενες με την προσομοίωση.

3.3.2. Προετοιμασία συστήματος και πρωτόκολλο προσομοιώσεων

Για την υλοποίηση μιας προσομοίωσης είναι απαραίτητο να ακολουθηθούν ορισμένα βήματα. Αρχικά είναι απαραίτητος ο καθορισμός των αρχικών συντεταγμένων του συστήματος όπως αυτές καθορίζονται από την πρωτοταγή, μη αναδιπλωμένη δομή. Ακολουθεί ελαχιστοποίηση της ενέργειας του συστήματος προκειμένου να αποφευχθούν πιθανές δομικές παραμορφώσεις και συνεπώς αστάθεια στην προσομοίωση από τυχών ισχυρούς δεσμούς Van der Waals. Στη συνέχεια προστίθενται τα μόρια του διαλύτη και με μια ελάχιστη αύξηση θερμοκρασίας δίνονται οι αρχικές ταχύτητες σε κάθε άτομο του συστήματος. Περιοδικά αυξάνεται η θερμοκρασία μέχρι να επιτευχθεί η επιθυμητή. Στην κατάσταση αυτή, ελέγχονται διάφοροι παράμετροι όπως η θερμοκρασία, η δομή κ.α. έως ότου σταθεροποιηθούν κοντά στις επιθυμητές τιμές. Όταν αυτό συμβεί, ξεκινάει η φάση «παραγωγής» κατά την οποία η προσομοίωση συνεχίζεται για το επιθυμητό χρονικό διάστημα και καταμετράται πληθώρα δεδομένων. Τέλος αναλύονται τα παραγόμενα τροχιακά για την εκάστοτε επιθυμητή μελέτη^[32].



Εικόνα 3: Μεθοδολογία προσομοιώσεων Μοριακής Δυναμικής. Αναπαραγωγή δίχως άδεια από τον ιστότοπο https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part3.html

4. Αποτελέσματα

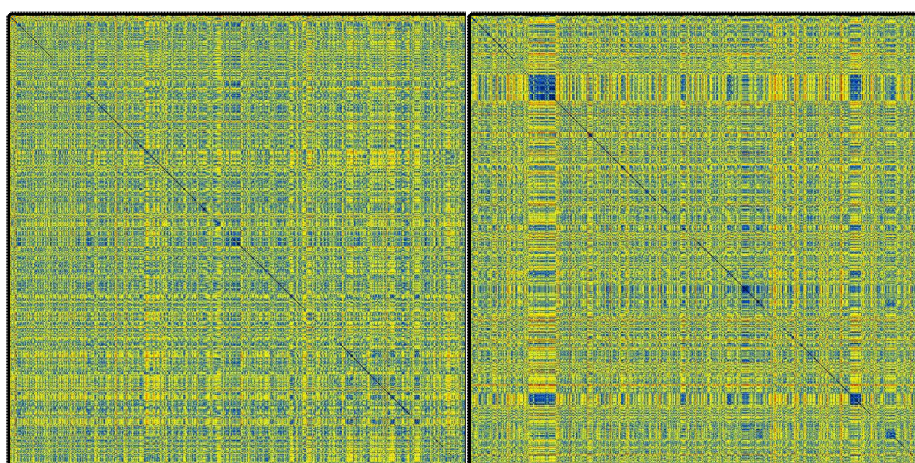
4.1. RMSD

Η ανάλυση RMSD (Root Mean Square Deviation) είναι μια μέθοδος με εφαρμογή σε αποτελέσματα προσομοιώσεων Μοριακής Δυναμικής και υπολογίζει τη μέση απόσταση χώρου μεταξύ των ατόμων δύο δομών μέσω της παρακάτω εξίσωσης:

$$RMSD = \sqrt{\frac{\sum(x_i - x_{ref})^2}{N}} \quad (17)$$

Όπου x_i οι συντεταγμένες των ατόμων της πρώτης δομής μια καθορισμένη χρονική στιγμή, x_{ref} οι συντεταγμένες των ατόμων της δομής αναφοράς την ίδια χρονική στιγμή και N ο αριθμός των ατόμων.

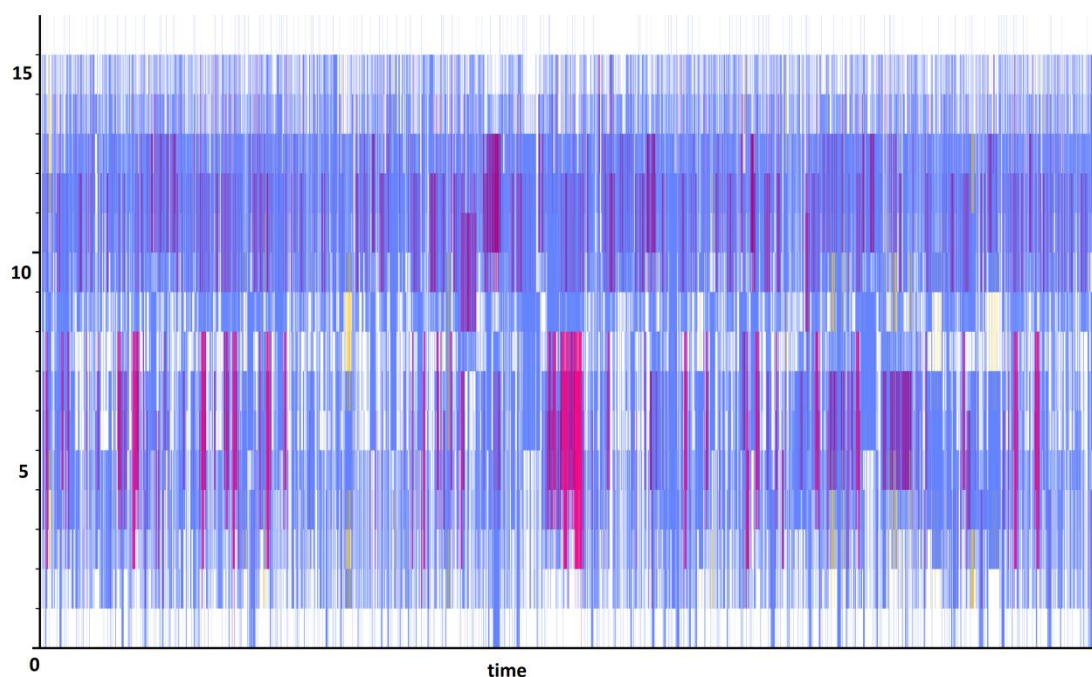
Όταν το υπολογιζόμενο RMSD είναι χαμηλό, οι δομές είναι ιδιαίτερα όμοιες μεταξύ τους. Τιμή RMSD ίση με 0Å σημαίνει πανομοιότυπες δομές. Η οπτικοποίηση των αποτελεσμάτων της μεθόδου πραγματοποιείται με τη χρήση πινάκων. Στους πίνακες αυτούς, τα χρώματα αντικατοπτρίζουν εύρη τιμών RMSD. Οι μπλε περιοχές αντιστοιχούν σε πολύ χαμηλές τιμές RMSD, οι κίτρινες περιοχές σε ενδιάμεσες τιμές ενώ οι κόκκινες περιοχές σε σχετικά υψηλές τιμές^[25].



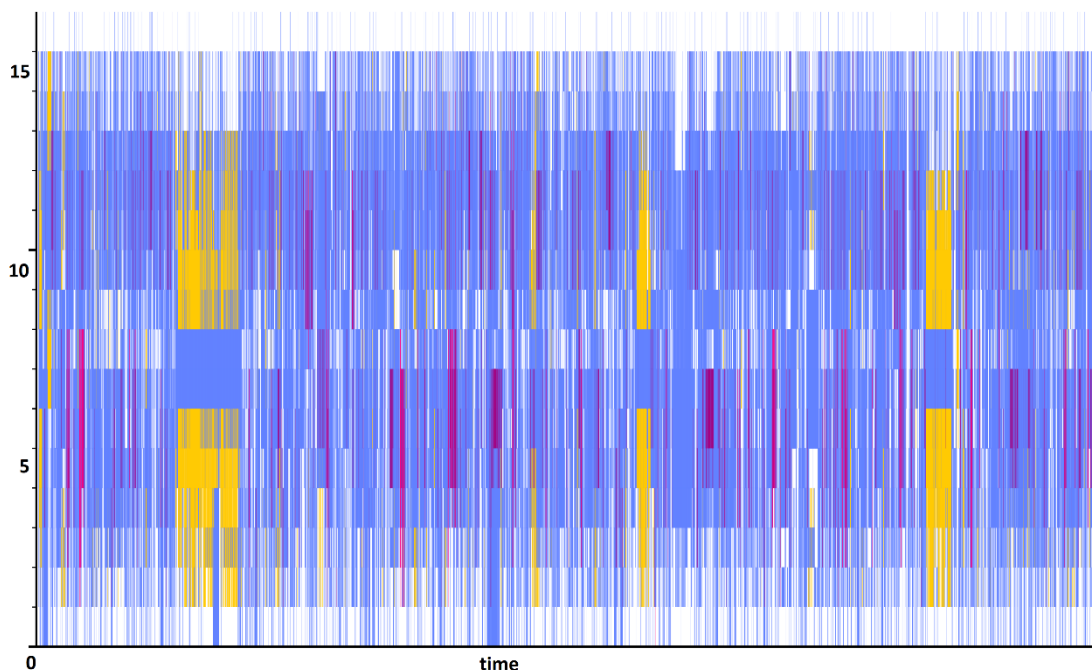
Εικόνα 4: Πίνακες RMSD. Ο αριστερός πίνακας αντιστοιχεί στο τροχιακό 14SB ενώ ο δεξιός στο STAR-ILDN. Οι περιοχές με χαμηλό RMSD (μπλε χρώμα) είναι ιδιαίτερα λίγες. Φαίνεται πως τα RMSD κυμαίνονται πιο κοντά σε ενδιάμεσες τιμές.

4.2. Ανάλυση δευτεροταγούς δομής

Η μέθοδος αυτή υπολογίζει την τάση των καταλοίπων μιας δομής να επιλέγουν συγκεκριμένες δευτεροταγείς δομές κατά τη διάρκεια της προσομοίωσης. Η ανάλυση αυτή έγινε με τη χρήση του αλγορίθμου STRIDE ο οποίος συνυπολογίζει την ενέργεια των δεσμών υδρογόνου και τις γωνίες στρέψης προκειμένου να εξάγει ένα αρχείο κειμένου με τις προβλεπόμενες δευτεροταγείς δομές^[18]. Στη συνέχεια το αρχείο αυτό δίνεται στο *gcrarna* το οποίο σχεδιάζει έναν πίνακα που περιγράφει τις δομές χρωματικά.



Εικόνα 5: Διάγραμμα δευτεροταγούς δομής από την προσομοίωση του δυναμικού πεδίου 14SB. Ο κάθετος άξονας υποδεικνύει τον αριθμό των αμινοξέων ενώ ο οριζόντιος τον χρόνο.



Εικόνα 6: Διάγραμμα δευτεροταγούς δομής από την προσομοίωση του δυναμικού πεδίου STAR-ILDN. Ο κάθετος άξονας υποδεικνύει τον αριθμό των αμινοξέων ενώ ο οριζόντιος τον χρόνο.

Στα διαγράμματα του αλγορίθμου STRIDE (Εικόνες 5 & 6) παρατηρείται πως τα δύο διαφορετικά δυναμικά πεδία προβλέπουν και διαφορετική δευτεροταγή διαμόρφωση του πεπτιδίου. Στην περίπτωση του 14SB παρατηρούνται κυρίως δομές α-έλικας ενώ σε αυτή τη STAR-ILDN εμφανίζονται έντονα δομές β-φύλλων. Συγκρίνοντας τον πίνακα RMSD από το δυναμικό πεδίο STAR-ILDN με τον αντίστοιχο πίνακα STRIDE μπορεί να παρατηρηθεί πως οι χαμηλές τιμές RMSD συμπίπτουν χρονικά με τις δομές β-φύλλων.

Ακολούθως, μέσω του προγράμματος WebLogo δημιουργήθηκε ένα διάγραμμα που παρουσιάζει την προτίμηση δευτεροταγούς δομής κάθε αμινοξέος κατά τη διάρκεια της προσομοίωσης.

Στον πίνακα που παράγεται από το WebLogo, κάθε στήλη αποτελείται από στοιβαγμένα γράμματα και αντιστοιχεί σε ένα αμινοξύ. Το ύψος των γραμμάτων εξαρτάται από τη συχνότητα της δομής στο εν λόγω αμινοξύ^{[13][34]}.

Κάθε γράμμα του πίνακα WebLogo αναπαριστά ένα τύπο δευτεροταγούς δομής. Η αντιστοιχία έχει ως εξής:

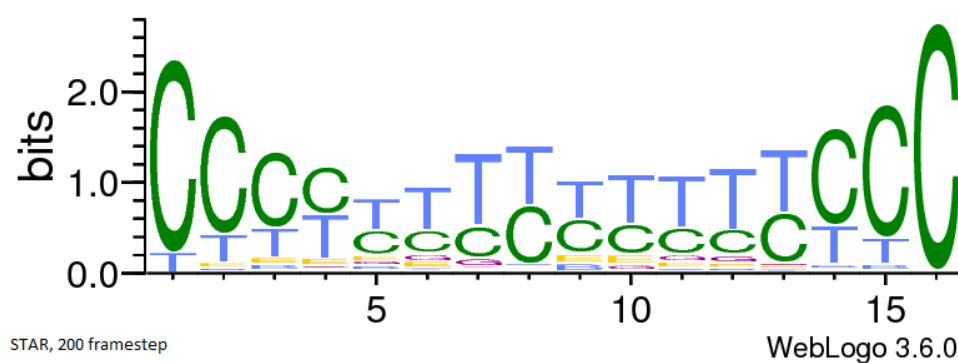
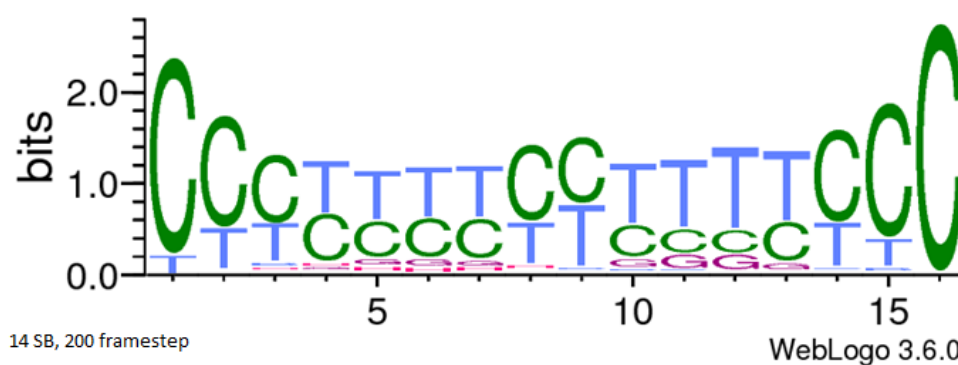
C : τυχαίο σπείραμα

T: στροφή

H: α-έλικα

E: β-φύλλο

G: 3_{10} έλικα



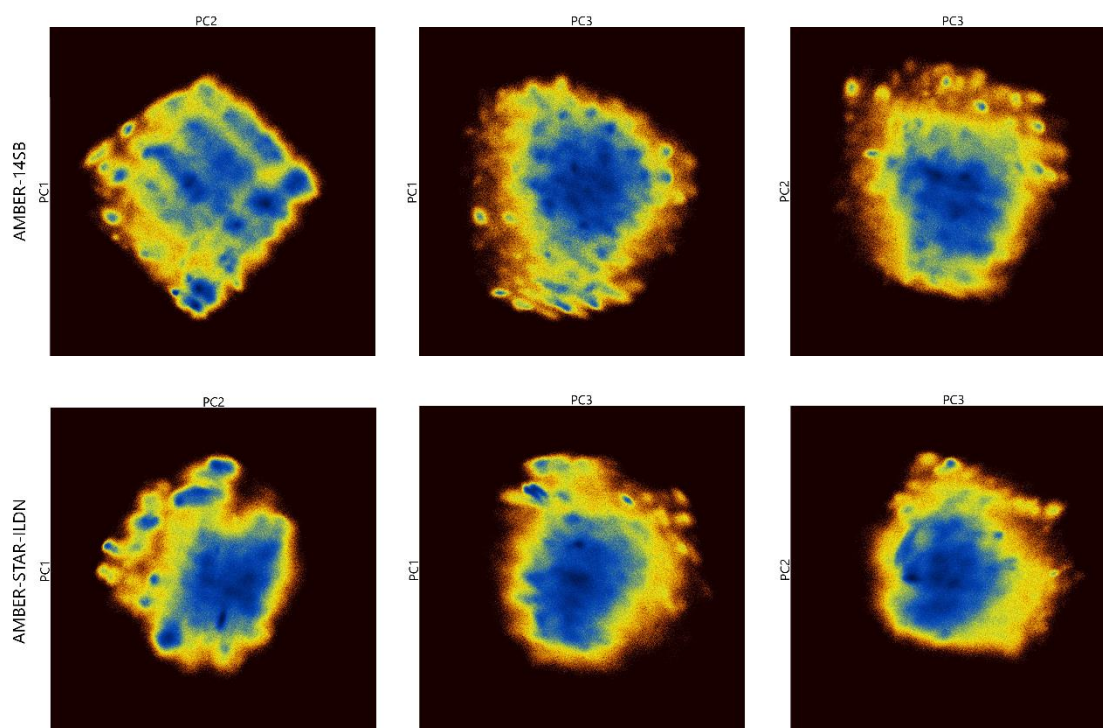
Εικόνα 7: Πίνακες WebLogo. Ο πρώτος αντιστοιχεί στο τροχιακό 14SB ενώ ο δεύτερος στο STAR-ILDN. Και στους δύο πίνακες παρατηρείται πως όλα σχεδόν τα αμινοξέα υιοθετούν πιο συχνά δομή τυχαίου σπειράματος ή στροφής ενώ πιο σπάνια φαίνεται πως τα αμινοξέα στη μέση του πεπτιδίου επιλέγουν δομές έλικας στο 14SB και β-φύλλων στο STAR-ILDN.

4.3. Principal Component Analysis (PCA)

Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis, PCA) είναι μια μαθηματική τεχνική ιδιαίτερα διαδεδομένη στη Μοριακή Δυναμική. Χρησιμοποιείται για τη μετατροπή ενός συνόλου πολυδιάστατων και συσχετιζόμενων δεδομένων σε ένα σύνολο γραμμικών συνιστωσών (κύριες συνιστώσες ή principal components). Το πόσο συσχετιζόμενα είναι τα αρχικά δεδομένα μεταξύ τους επηρεάζει το επίπεδο

συμπύεσης που μπορεί να επιτευχθεί. Η μετατροπή αυτή οδηγεί σε απλοποίηση των δεδομένων καθιστώντας δυνατή τη μελέτη και ερμηνεία τους.

Στο πεδίο της Μοριακής Δυναμικής, η PCA χρησιμεύει στη μελέτη της κινητικότητας πρωτεϊνικών μορίων. Δύο είναι τα είδη ανάλυσης κυρίων συνιστωσών που εφαρμόζονται σε τέτοιες μελέτες, η cPCA και η dPCA. Στην cPCA (Cartesian PCA) χρησιμοποιούνται οι καρτεσιανές συντεταγμένες κάθε ατόμου ενώ στην dPCA (Dihedral PCA) οι διέδρες γωνίες φ & ψ του πολυπεπτιδικού σκελετού. Η dPCA χαρακτηρίζεται συνήθως ως πιο αποτελεσματική της cPCA επειδή η υψηλή κινητικότητα των πρωτεϊνικών μορίων κατά την αναδίπλωση συχνά οδηγεί σε καταγραφή ασήμαντων συνολικών κινήσεων όταν λαμβάνεται υπόψη η θέση κάθε ατόμου. Η μελέτη βάσει των διέδρων γωνιών εξαιρεί τέτοιες κινήσεις και εστιάζει στις σημαντικές κινήσεις μεταξύ των πεπτιδικών δεσμών εντός του μοριακού σκελετού^{[1][2]}.



Εικόνα 8: Διαγράμματα ελεύθερης ενέργειας δύο διαστάσεων. Σχεδιάστηκαν από το πρόγραμμα *gcrarma* εφαρμόζοντας dPCA για όλες τις διέδρες γωνίες του πεπτιδίου. Τα πρώτα τρία προέρχονται από το τροχιακό AMBER-14SB ενώ τα τελευταία τρία από το τροχιακό AMBER-STAR-ILDN.

Στα διαγράμματα ελεύθερης ενέργειας, οι μπλε περιοχές αντιπροσωπεύουν ενεργειακό ελάχιστο, η κλιμάκωση του χρώματος προς το κόκκινο αντιστοιχεί σε ανάλογη κλιμάκωση της ελεύθερης ενέργειας. Και στα δύο τροχιακά, τα διαγράμματα που εξήχθησαν εμφανίζουν διευρυμένα ενεργειακά ελάχιστα.

4.4. Υπολογισμός NOE & σύγκριση με πειραματικά δεδομένα

Το NOE (Nuclear Overhauser Effect) είναι ένα φαινόμενο που παρατηρείται σε γειτονικούς πυρήνες και είναι ιδιαίτερα σημαντικό στην φασματοσκοπία NMR. Η μαγνητική διέγερση ενός πυρήνα μπορεί να επιφέρει διαταραχή στα spin-γειτονικών πυρήνων. Η μεταβολή αυτή είναι δυνατό να εντοπιστεί και να αποτελέσει μέτρο της απόστασης μεταξύ δύο ατόμων της τριτοταγούς δομής μιας πρωτεΐνης. Οι ατομικές αποστάσεις που υπολογίζονται από την ανάλυση των NOE είναι ιδιαίτερα χρήσιμες όταν αφορούν ζεύγη ατόμων που στην πρωτοταγή δομή απέχουν το ένα από το άλλο ενώ στον τρισδιάστατο χώρο βρίσκονται κοντά μεταξύ τους. Το σήμα του φάσματος NOE υπολογίζεται από την εξίσωση:

$$NOE = \frac{1}{r^6} f(t_c) \quad (18)$$

Όπου r η απόσταση μεταξύ των δύο πυρήνων και t_c ο χρόνος που απαιτείται για μια πλήρη περιστροφή 1 rad. Έχει γίνει σαφές ότι η τιμή του NOE είναι αντιστρόφως ανάλογη της 6^{ης} δύναμης της απόστασης των ατόμων. Η δύναμη αυτή είναι ιδανική για μετρήσεις σε μικρά πεπτίδια. Σε μεγαλύτερες πρωτεΐνες, προτιμάται η χρήση της 3^{ης} δύναμης αυτής της απόστασης^[40].

Τα σήματα NOE υπάγονται σε τρεις κατηγορίες:

Ισχυρό (1.8 – 2.7 Å)

Μέτριο (2.7 – 3.3 Å)

Ασθενές (3.3 – 5.0 Å)

Για σήματα με τιμή μεγαλύτερη από 5.0 Å θεωρείται πως δεν υπάρχει ατομική αλληλεπίδραση.

Τα ζεύγη ατόμων από τα αποτελέσματα NMR με $\text{NOE} \leq 5\text{\AA}$ επιλέχθηκαν και καταγράφηκαν σε μια λίστα (proton.list) ώστε να συγκριθούν με τα αντίστοιχα ζεύγη της προσομοίωσης. Η λίστα εξήχθη από τα αρχεία αποτελεσμάτων των πειραμάτων NMR που βρίσκονται καταχωρημένα στην βάση δεδομένων PDB με κωδικούς 2N9N & 2N9M^{[6][7]}.

Ο υπολογισμός των ατομικών αποστάσεων πραγματοποιήθηκε μέσω του προγράμματος *carma* το οποίο παρήγαγε ένα αρχείο ατομικών αποστάσεων (*carma.distances*). Το αρχείο αυτό δόθηκε στο πρόγραμμα *noe_averaging.c*^[42] το οποίο υπολογίζει τις αποστάσεις r^{-3} και r^{-6} .

Στο σημείο αυτό έχουμε τις πειραματικές αποστάσεις και τις αποστάσεις από τις προσομοιώσεις για όλα τα ζεύγη ατόμων που στα πειράματα NMR απέχουν $\leq 5\text{\AA}$. Επόμενο βήμα αποτελεί η εύρεση των ζευγών για τα οποία η απόσταση που προκύπτει από την προσομοίωση είναι μεγαλύτερη από αυτή του πειράματος και ο υπολογισμός του upper bound violation. Το τελευταίο υπολογίζεται βάσει της παρακάτω εξίσωσης.

$$v(i, j) = r^{-6} - nmr(i, j) \quad (19)$$

Το μέσο upper bound violation υπολογίζεται προσθέτοντας όλες τις τιμές των violations και διαιρώντας με τον αριθμό των ατομικών ζευγών που χρησιμοποιήθηκαν. Παρακάτω παρουσιάζονται οι πίνακες με τις αποστάσεις και τα upper bound violations όπως υπολογίστηκαν σε θερμοκρασία 300K μεταξύ των πειραματικών δεδομένων NMR (2N9N & 2N9M) και των προσομοιώσεων με δυναμικά πεδία AMBER-14SB & AMBER99SB-STAR-ILDN.

Proton pair	NMR upper bound	14SB r^6	Upper bound violation	STAR-ILDN r^6	Upper bound violation
1	3.5	2.188464	-	2.212407	-
2	3.5	2.629834	-	2.698162	-
3	3.5	2.281372	-	2.317162	-
4	3.8	3.139621	-	3.116689	-
5	3.5	2.356819	-	2.401036	-
6	3.8	2.618368	-	2.777990	-
7	3.8	2.552611	-	2.676450	-
8	3.5	2.354329	-	2.398280	-
9	3.5	2.442366	-	2.440156	-
10	3.5	2.348451	-	2.418306	-
11	3.5	2.479802	-	2.558903	-
12	3.8	2.547012	-	2.623958	-
13	3.8	2.590730	-	2.855960	-
14	5.3	3.094962	-	3.029227	-
15	5.3	3.480053	-	3.461198	-
16	3.5	2.553990	-	2.580355	-
17	4.0	2.881296	-	2.964483	-
18	3.8	2.645897	-	2.684511	-
19	3.8	2.962063	-	2.956308	-
20	4.5	2.683002	-	2.704000	-
21	4.5	2.781154	-	2.743723	-
22	5.0	5.011775	0.011775	4.998799	-
23	5.0	4.982050	-	4.988446	-
24	3.5	2.518011	-	2.482939	-
25	5.0	3.430498	-	3.266495	-
26	5.0	3.237265	-	3.120502	-
27	3.5	2.613750	-	2.583941	-
28	3.5	2.554854	-	2.511887	-
29	3.5	2.414882	-	2.440168	-
30	3.5	3.572445	0.072445	3.125924	-
31	3.5	2.875161	-	2.625890	-
32	3.5	3.085201	-	3.225026	-
33	3.5	2.871053	-	2.557056	-
34	3.5	2.809440	-	2.721026	-
35	3.5	3.020007	-	2.864021	-
36	3.5	2.707600	-	2.455308	-

37	3.5	2.776359	-	2.608721	-
38	3.5	2.595895	-	2.534542	-
39	3.5	3.100897	-	3.240891	-
40	3.5	2.741321	-	2.606856	-
41	3.5	3.215003	-	3.256329	-
42	3.5	6.699517	3.199517	5.765712	2.265712
43	3.8	6.146686	2.346686	6.479050	2.679050
44	4.5	5.143162	0.643162	5.527572	1.027572

14SB average violation	0.142581
-----------------------------------	-----------------

STAR-ILDN average violation	0.135735
--	-----------------

Στο τροχιακό 14SB παρατηρούνται 5 upper bound violations ενώ στο STAR-ILDN παρατηρούνται 3. Τα δύο πρώτα violation του τροχιακού 14SB (ζεύγη 22 & 30) έχουν τιμές 0.011775 & 0.072445 οι οποίες χαρακτηρίζονται ως χαμηλές και αποδεκτές αποκλίσεις. Ωστόσο, τα υπόλοιπα 3 violations είναι κοινά και για τα δύο τροχιακά (ζεύγη 42, 43, 44) και έχουν ιδιαίτερα μεγάλες τιμές. Αποτέλεσμα των αποκλίσεων στα τρία αυτά ζεύγη είναι ο υπολογισμός πολύ μεγάλων average violations. Το ζεύγος 42 αντιστοιχεί στα αμινοξέα GLY4 – PHE12 τα οποία στην πρωτοταγή δομή διαχωρίζονται από 7 αμινοξέα, στην τριτοταγή διαμόρφωση όμως, τα πειραματικά δεδομένα υπολογίζουν απόσταση 3.5 Å ενώ από τα δυναμικά πεδία υπολογίζεται r^{-6} με τιμή 6.699517. Τα ζεύγη 43, 44 αντιστοιχούν και τα δύο στα αμινοξέα ALA2 – TRP5 αλλά σε διαφορετικά ζεύγη πρωτονίων το καθένα.

Για την ολοκλήρωση της ανάλυσης και την επιβεβαίωση των αποτελεσμάτων, πραγματοποιήθηκε ένας ακόμη έλεγχος. Μελετήθηκαν όλες οι δομές των καταχωρήσεων 2N9N και 2N9M, καθεμία εκ των οποίων περιέχει 20 πιθανές διαμορφώσεις, και υπολογίστηκαν οι μέσες αποστάσεις για τα ζεύγη των ατόμων που εμφανίζουν μεγάλα violations στην προηγούμενη σύγκριση των NOEs.

Από τις δομές της καταχώρησης 2N9N εξήχθησαν οι αποστάσεις:

Ζεύγος 42: 5.1339 Å
 Ζεύγος 43: 6.6495 Å
 Ζεύγος 44: 6.6189 Å

Από τις δομές της καταχώρησης 2N9M εξήχθησαν οι αποστάσεις:

Ζεύγος 42: 10.9326 Å

Ζεύγος 43: 6.1365 Å

Ζεύγος 44: 5.5613 Å

Η σύγκριση αυτών των αποστάσεων με τα NOEs των αρχείων των πειραμάτων NMR οδηγούν στον υπολογισμό των παρακάτω violations:

	2N9N	2N9M
Ζεύγος 42:	1.6339 Å	7.4326 Å
Ζεύγος 43:	2.8495 Å	2.3365 Å
Ζεύγος 44:	2.1189 Å	1.9613 Å

Η ύπαρξη τόσο μεγάλων violations μεταξύ των αποτελεσμάτων NMR και των καταχωρημένων δομών δικαιολογεί τα violations που υπολογίστηκαν στη σύγκριση των τροχιακών με το NMR και παραπέμπει στη θεωρία ότι το πεπτίδιο είναι ιδιαίτερα κινητικό και άτακτο.

5. Συμπεράσματα

Η παρούσα εργασία ξεκίνησε με τις προσομοιώσεις μοριακής δυναμικής του πεπτιδίου VG16KRKP βάσει δύο διαφορετικών δυναμικών πεδίων AMBER, του AMBER-14SB και του AMBER99-STAR-ILDN. Στα αποτελέσματα των προσομοιώσεων εφαρμόστηκε μια σειρά από υπολογιστικές τεχνικές ανάλυσης προκειμένου να εξαχθούν τα ακόλουθα συμπεράσματα.

Η μελέτη ξεκίνησε με την ανάλυση του RMSD των δυο τροχιακών. Στους πίνακες που προέκυψαν παρατηρήθηκαν κυρίως κίτρινες περιοχές. Στον πίνακα του τροχιακού STAR-ILDN εμφανίζεται αυξημένη σταθερότητα στα αμινοξικά κατάλοιπα 3 και 14. Ωστόσο αυτός του τροχιακού 14SB χαρακτηρίζεται στο σύνολό του από χρώμα πλησιέστερο στο μπλε άρα και από μερικώς αυξημένη ολική σταθερότητα. Τα αρχικά αποτελέσματα αυτά προϋδεάζουν για ένα πεπτίδιο με αυξημένη κινητικότητα το οποίο δεν διατηρεί μια διαμόρφωση για μεγάλο χρονικό διάστημα.

Η ανάλυση δευτεροταγούς δομής που ακολούθησε, οδήγησε στην εξαγωγή συμπερασμάτων για τις προτιμήσεις του πεπτιδίου σε επίπεδο δευτεροταγών διαμορφώσεων. Στο τροχιακό 14SB παρατηρήθηκαν κυρίως δομές έλικας. Οι δομές αυτές προκύπτουν σχεδόν σε ολόκληρο το μήκος του πεπτιδίου αλλά εστιάζονται κυρίως στα αμινοξέα 3-8 και 10-13. Στο τροχιακό STAR-ILDN παρατηρούνται έντονα δομές β-φύλλων στα αμινοξέα 2-6 και 9-13. Λαμβάνοντας υπόψη τα δεδομένα της ανάλυσης RMSD προκύπτει το συμπέρασμα ότι η δομή β-φύλλου στο τροχιακό STAR-ILDN έχει αυξημένη σταθερότητα στην περιοχή κοντά στο αμινοξύ ARG3.

Οι χάρτες ελεύθερης ενέργειας που παρήγαγε η μέθοδος dPCA, όταν εφαρμόστηκε σε όλο το μήκος του πεπτιδίου, δεν εμφανίζουν ιδιαίτερα διακριτά μεταξύ τους ενεργειακά ελάχιστα. Τα παραπάνω εμφανίζονται συμπυγμένα και είναι σχεδόν αδύνατη η διάκρισή τους σε τοπικά ελάχιστα.

Τελική τεχνική που εφαρμόστηκε ήταν η σύγκριση των τιμών NOE μεταξύ των προσομοιώσεων και των πειραματικών δεδομένων NMR. Τα average upper bound violations που υπολογίστηκαν έχουν ως εξής: 0.142581 Å (14SB) και 0.135735 Å (STAR-ILDN). Οι τιμές αυτές είναι πολύ υψηλές και για τα δύο τροχιακά και μια αρχική

προσέγγιση θα ερμήνευε τα αποτελέσματα θεωρώντας πως οι προσομοιώσεις δε συμβαδίζουν με τα πειραματικά δεδομένα NMR. Ωστόσο, το μέγεθος των violations οφείλεται κυρίως σε τρία ζεύγη πρωτονίων τα οποία μάλιστα δεν ανήκουν σε γειτονικά αμινοξέα. Ακόμη, και τα τρία αυτά ζεύγη πρωτονίων έχουν ασθενές σήμα NOE μεγαλύτερο των 3.5 Å. Μια ερμηνεία θα μπορούσε να δικαιολογήσει τις τιμές αυτές θεωρώντας ότι το πεπτίδιο αποκτά μια προεξέχουσα δευτεροταγή δομή παρούσα και στα δύο Force Fields.

	Αμινοξύ	Τύπος πρωτονίου	NMR upper bound	Αμινοξύ	Τύπος πρωτονίου
Ζεύγος 42	GLY 4	H	3.5	PHE 12	HA
Ζεύγος 43	ALA 2	HA	3.8	TRP 5	H
Ζεύγος 44	ALA 2	MB (HB1)	4.5	TRP 5	HD1

Περαιτέρω ανάλυση των τριών αυτών ζευγών απέδειξε την ύπαρξη μεγάλων violations ανάμεσα στις καταχωρημένες δομές και τα αποτελέσματα NMR. Οι τιμές αυτών έχουν ως εξής:

	2N9N - NMR	2N9M - NMR
Ζεύγος 42	1.6339 Å	7.4326 Å
Ζεύγος 43	2.8495 Å	2.3365 Å
Ζεύγος 44	2.1189 Å	1.9613 Å

Ο εντοπισμός αυτών των violations επιβεβαιώνει την ιδέα ότι το πεπτίδιο είναι εξαιρετικά κινητικό και χαρακτηρίζεται από υψηλή αταξία.

Το τελικό συμπέρασμα που προκύπτει από τις αναλύσεις των δύο δυναμικών πεδίων και των αποτελεσμάτων NMR είναι ότι το πεπτίδιο VG16KRRKP εμφανίζει μια έντονη αταξία που δεν επιτρέπει τον σχεδιασμό δομών, ικανών να πληρούν όλους τους περιορισμούς και να διαθέτουν μεγάλη σταθερότητα.

6. Βιβλιογραφία

1. Altis, A., Otten, M., Nguyen, P., Hegger, R., Stock, G. (2008). Construction of the Free Energy Landscape of Biomolecules via Dihedral Angle Principal Component Analysis. *The Journal of Chemical Physics*, 128(24), 245102.
2. Amadei, A., Linssen, A., Berendsen, H. (1993). Essential Dynamics of Proteins. *Proteins: Structure, Function, And Genetics*, 17(4), 412-425.
3. Anfinsen, C. (1973). Principles that Govern the Folding of Protein Chains. *Science*, 181(4096), 223-230.
4. Baker, M. (2018). Cryo-electron Microscopy Shapes Up. *Nature*, 561(7724), 565-567.
5. Berg, J., Tymoczko, J., & Stryer, L. (2012). *Biochemistry* (7 ed.). New York: W. H. Freeman.
6. Bhunia, A., & Datta, A. (2016). Solution Structure of VG16KRKP in *C. neoformans* (conformation 1). doi: 10.2210/pdb2n9n/pdb
7. Bhunia, A., & Datta, A. (2016). VG16KRKP Solution Structure in *C. neoformans* live cells (conformation 2). doi: 10.2210/pdb2n9m/pdb
8. Biomolecular Simulation - The GROMOS Software. (2020). Retrieved 29 January 2020, from <http://www.gromos.net/>
9. Bizzarri, A., Cannistraro, S. (2002). Molecular Dynamics of Water at the Protein-Solvent interface. *Journal of Physical Chemistry B*, 106(26), 6617-6633.
10. Branden, C., & Tooze, J. (1999). *Introduction to Protein Structure* (2 ed.). New York: Garland Science.
11. Bryngelson, J., Onuchic, J., Socci, N., Wolynes, P. (1995). Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins: Structure, Function, and Genetics*, 21(3), 167-195.
12. Chan, H., Dill, K. (1998). Protein Folding in the Landscape Perspective: Chevron Plots and non-Arrhenius Kinetics. *Proteins: Structure, Function, and Genetics*, 30(1), 2-33.

13. Crooks, G. (2004). WebLogo: A Sequence Logo Generator. *Genome Research*, 14(6), 1188-1190.
14. Daggett, V., Fersht, A. (2003). Is there a Unifying Mechanism for Protein Folding? *Trends in Biochemical Sciences*, 28(1), 18-25.
15. Datta, A., Ghosh, A., Airoldi, C. et al. (2016). Antimicrobial Peptides: Insights into Membrane Permeabilization, Lipopolysaccharide Fragmentation and Application in Plant Disease Control. *Scientific Reports*, 5(11951).
16. Dill, K., Chan, H. (1997). From Levinthal to Pathways to Funnels. *Nature Structural & Molecular Biology*, 4(1), 10-19.
17. Glykos, N.M. (2006), "Carma: a molecular dynamics analysis program". *J. Comput. Chem.*, 27, 1765-1768.
18. Heinig, M., Frishman, D. (2004). STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Research*, 32(Web Server), W500-W502.
19. Home - CHARMM. (2020). Retrieved 29 January 2020, from <https://www.charmm.org/charmm/?CFID=03380c6f-266a-494d-bc3d-4741256b2f84&CFTOKEN=0>
20. Home MD. (2020). Retrieved 29 January 2020, from https://www.ch.embnet.org/MD_tutorial/
21. Honig, B., Yang, A. (1995). Free Energy Balance in Protein Folding. *Advances in Protein Chemistry*, 46, 27-58.
22. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26:1781-1802, 2005. abstract, journal
23. Jorgensen, W., Tirado, J. (1988). The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *Journal of the American Chemical Society*, 110(6), 1657-1666.

24. Karplus, M., Weaver, D. (1994). Protein Folding Dynamics: The Diffusion-Collision Model and Experimental Data. *Protein Science*, 3, 650-668.
25. Knapp, B., Frantal, S., Cibena, M., Schreiner, W., Bauer, P. (2011). Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible?. *Journal of Computational Biology*, 18(8), 997-1005.
26. Koukos, P.I. & Glykos, N.M. (2013), "grcarma: A Fully Automated Task-Oriented Interface for the Analysis of Molecular Dynamics Trajectories". *J. Comput. Chem.*, 34, 2310-2312.
27. Levinthal, C. (1969) How to Fold Graviously. Mössbauer Spectroscopy in Biological Systems Proceedings.
28. Mackerell, A. (2004). Empirical Force Fields for Biological Macromolecules. *Journal of Computational Chemistry*, 25(13), 1584-1604.
29. Maier, J., Kasavajhala, K., Wickstrom, L. Hauser, K., Simmerling C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696-3713.
30. Mark, P., Nilsson, L. (2001). Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105(43), 9957-9960.
31. Nölting, B., Agard, D. (2008). How General is the Nucleation-Condensation Mechanism?. *Proteins*, 73(3), 754-764.
32. Part 3. (2020). Retrieved 29 January 2020, from https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part3.html
33. research:howto:md_and_nmr_calculation_of_noes [Norma]. (2020). Retrieved 31 January 2020, from http://norma.mbg.duth.gr/index.php?id=research:howto:md_and_nmr_calculation_of_noes
34. Schneider, T., Stephens, R. (1990). Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Research*, 18(20), 6097-6100.

35. Smyth, M., Martin, J. (2000). *Journal of Clinical Pathology - Molecular Pathology*, 53(1), 8-14.
36. Splettstoesser, T. (2013). Folding funnel schematic. Retrieved 01 February 2020, from https://commons.wikimedia.org/wiki/File:Folding_funnel_schematic.svg
37. The Amber Molecular Dynamics Package. (2020). Retrieved 29 January 2020, from <http://ambermd.org/>
38. The Norma computing cluster. (2020). Retrieved 29 January 2020, from <https://norma.mbg.duth.gr/>
39. Wu, Y., Tepper, H., Voth, G. (2006). Flexible simple point-charge water model with improved liquid-state properties. *Journal of Chemical Physics*, 124(2), 024503.
40. Zagrovic, B., van Gunsteren, W. (2006). Comparing Atomistic Simulation Data With the NMR Experiment: How Much Can NOEs Actually Tell Us?. *Proteins: Structure, Function, And Bioinformatics*, 63(1), 210-218.
41. Zhou, R., Huang, X., Margulis, C., Berne, B. (2004). Hydrophobic Collapse in Multidomain Protein Folding. *Science*, 305(5690), 1605-1609.

42. noe_averaging.c

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

int main(argc,argv)
int  argc;
char*argv[];
{
    FILE          *in;
    char          c;
    char          string[100];
    int           columns;
    int           junk;
    int           i;
    long double r6, r3;
    float         val;
    int           nof_values;
    int           nof_values2;
    int           fraction;
    double        *data_r3;
    double        *data_r6;

    double        *sd_data_r3;
    double        *sd_data_r6;

    setlinebuf( stdout );
    in = fopen( argv[1], "r" );
    if ( in == NULL )
    {
        printf("Can't open file %s for reading. Abort.\n", argv[1] );
        exit(1);
    }

    columns = 0;
    while( fscanf(in, "%c", &c) == 1 )
    {
        if ( c == '\n' )
            break;
        if ( c == '.' )
            columns++;
    }

    fprintf(stderr, "Found %d columns with distances.\n", columns );
    fprintf(stderr, "Allocating memory ...");
    data_r3 = (double *)calloc( columns, sizeof(double));
    data_r6 = (double *)calloc( columns, sizeof(double));
    sd_data_r3 = (double *)calloc( columns, sizeof(double));
    sd_data_r6 = (double *)calloc( columns, sizeof(double));
    if ( data_r3 == NULL || data_r6 == NULL || sd_data_r3 == NULL || sd_data_r6 == NULL )
    {
        printf("Failed to allocate %ld bytes of memory. Abort.\n", (long int)(2*columns*sizeof(double)));
        exit(1);
    }
}
```

```

fprintf(stderr, " done.\nNow reading line : ");
rewind( in );

nof_values = 0;
fprintf(stderr, "%8d", nof_values );
while( 1 )
{
if ( fscanf(in, "%s", &string[0]) != 1)
break;

for ( i=0 ; i< columns ; i++ )
if ( fscanf(in, "%f", &val) != 1 )
{
printf("Ooops. Corrupt input file (or not from carma ?). Abort.\n");
exit(1);
}
else
{
data_r3[i] += 1.0l / (val*val*val);
data_r6[i] += 1.0l / (val*val*val*val*val*val);
}

nof_values++;
fprintf(stderr, "%8d", nof_values );

}

printf("\n\nSecond pass to calculate standard deviations ...\n");

rewind( in );

nof_values2 = 0;
fprintf(stderr, "%8d", nof_values2 );
while( 1 )
{
if ( fscanf(in, "%s", &string[0]) != 1) /* skip frame number */
break;

for ( i=0 ; i< columns ; i++ )
if ( fscanf(in, "%f", &val) != 1 )
{
printf("Ooops. Corrupt input file (or not from carma ?). Abort.\n");
exit(1);
}
else
{
sd_data_r3[i] += ( 1.0l / (val*val*val) - (data_r3[i]/nof_values) ) * ( 1.0l / (val*val*val) -
(data_r3[i]/nof_values) );
sd_data_r6[i] += (1.0l / (val*val*val*val*val*val) - (data_r6[i]/nof_values)) * (1.0l /
(val*val*val*val*val*val) - (data_r6[i]/nof_values));
}

nof_values2++;
fprintf(stderr, "%8d", nof_values2 );

}

```

```

if ( nof_values != nof_values2 )
{
printf("Different number of values during second pass ??? Panic.\n");
exit(1);
}

printf("\nColumn  r^-3  sd(r^-3)  r^-6  sd(r^-6)\n");
for ( i=0 ; i< columns ; i++ )
{
printf("% 5d % 10.6f% 10.6f% 10.6f% 10.6f \n", i+1, (double)(pow((data_r3[i]/nof_values),-
1.0/3.0)),
(double)(pow(sqrt(sd_data_r3[i]/nof_values2),-1.0/3.0)),
(double)(pow((data_r6[i]/nof_values),-1.0/6.0)),
(double)(pow(sqrt(sd_data_r6[i]/nof_values2),-1.0/6.0)) );
}
}

```