



DEMOCRITUS UNIVERSITY OF THRACE
SCHOOL OF HEALTH SCIENCES
DEPARTMENT OF MOLECULAR BIOLOGY & GENETICS

Μελέτη του διαμορφωτικού χώρου αναδίπλωσης πεπτιδίου άνευ παρουσίας δευτεροταγών δομών: EPI-X4

Διπλωματική εργασία

Νικόλαος Παπουτσήs
-616-

Επιβλέπων καθηγητής
Νικόλαος Μ. Γλυκός

Καρλόβασι Σάμου
Απρίλιος 2025

Περιεχόμενα

Abstract	3
Περίληψη	4
Εισαγωγή	5
1.1 Το πρόβλημα της πρωτεϊνικής αναδίπλωσης	5
1.2 EPI-X4: Ενδογενής ανταγωνιστής του CXCR4	6
1.3 Ηλεκτροστατικές Ιδιότητες και Ανάλυση Επιφανειακών Χαρακτηριστικών.....	7
1.4 Ο στόχος της συγκεκριμένης εργασίας	9
Προσομοίωση Μοριακής Δυναμικής	9
2.1 Χρήση και στόχοι της τεχνικής MD	9
2.2 Προετοιμασία Συστήματος	11
Μέθοδοι και Αποτελέσματα	11
3.1 Πίνακας RMSD και διάγραμμα δευτεροταγών δομών	11
3.2 Nuclear Overhauser Effects (NOE)	14
3.3 Υπολογισμός NOE και σύγκριση	14
3.5 Principal Component Analysis (PCA)	19
3.6 Παρουσίαση Αποτελεσμάτων PCA και Clustering	21
Συμπεράσματα	28
Βιβλιογραφία	29

Abstract

This study investigates the folding behavior of the 16-mer peptide EPI-X4, a naturally occurring antagonist of the chemokine receptor CXCR4. By applying molecular dynamics (MD) simulations in combination with the use of computational tools for statistical and structural analysis, we aim to assess the conformational space of the peptide and evaluate the reliability of computational (in silico) molecular structure determination methods in relation to experimental data. Structural information derived from the MD-defined trajectory was subjected to various analyses, including: NOE (nuclear Overhauser effect) distance calculations and PCA (dihedral and Cartesian) analyses.

Comparisons with experimental NOE data indicated general agreement between the computational structures and the experimental ones, while PCA and clustering analyses were able to identify dominant structures with relatively high accuracy. The overall picture highlighted the increasing reliability of computational methods in comparison to experimental ones.

Περίληψη

Η συγκεκριμένη εργασία ερευνά την συμπεριφορά αναδίπλωσης του 16-μερούς πεπτιδίου EPI-X4, ένας φυσικά παραγόμενος ανταγωνιστής του υποδοχέα χημοκινών CXCR4. Με την εφαρμογή προσομοίωσης μοριακής δυναμικής σε συνδυασμό με τη χρήση υπολογιστικών εργαλείων στατιστικής και δομικής ανάλυσης, στοχεύουμε στην εκτίμηση του διαμορφωτικού χώρου του πεπτιδίου και την αξιολόγηση της αξιοπιστίας των υπολογιστικών (*in silico*) μεθόδων προσδιορισμού μοριακής δομής σε σχέση με τα πειραματικά δεδομένα. Δομικές πληροφορίες προερχόμενες από το MD (molecular dynamics) προσδιορισμένο τροχιακό υποβλήθηκαν σε διάφορες αναλύσεις συμπεριλαμβανομένου: υπολογισμοί αποστάσεων φαινομένων NOE (nuclear overhauser effects) και αναλύσεις PCA (dihedral και Cartesian).

Οι συγκρίσεις με πειραματικά δεδομένα NOE, ενέδειξαν γενική συμφωνία των υπολογιστικών δομών με τις πειραματικές, ενώ οι αναλύσεις PCA και ομαδοποίησης ήταν σε θέση να εντοπίσουν κυρίαρχες δομές, με εξαιρετική σχετικά ακρίβεια. Η συνολική εικόνα κατέδειξε την αυξανόμενη αξιοπιστία των υπολογιστικών μεθόδων συγκριτικά με τις πειραματικές.

1. Εισαγωγή

1.1 Το πρόβλημα της πρωτεϊνικής αναδίπλωσης

Για πολλά πλέον χρόνια, το πρόβλημα της πρωτεϊνικής αναδίπλωσης αποτελεί κεντρική πρόκληση στον κλάδο της μοριακής βιολογίας, της βιοφυσικής και των υπολογιστικών επιστημών. Στον πυρήνα του, το πρόβλημα συμπεριλαμβάνει την κατανόηση του πως μια πολυπεπτιδική αλυσίδα, που απαρτίζεται από μια συγκεκριμένη αμινοξική αλληλουχία, αναδιπλώνεται σε μια καθορισμένη τριτοταγή δομή. Η τελευταία είναι αυτή που τελικά καθορίζει και τη λειτουργία της εκάστοτε πρωτεΐνης. Παρά της αξιόλογης προόδου στις πειραματικές και *in silico* μεθόδους, παραμένουν σημαντικά ερωτήματα όσον αφορά τους μηχανισμούς της αναδίπλωσης πρωτεϊνών.

Οι πρωτεΐνες αναδιπλώνονται τάχιστα, σε χρονικές κλίμακες της τάξεως των *millisecond*, με τρόπους που μπορούν να αναπαραχθούν *in vivo*, ωστόσο το ύψος της θεωρητικής πολυπλοκότητας των διαδικασιών αυτών είναι δυσθεώρητο. Ο διάσημος καθηγητής του MIT Cyrus Levinthal είχε διατυπώσει το πασίγνωστο, στην βιολογία, παράδοξο(5), το οποίο διατυπώνεται ως εξής: Αν μια πρωτεΐνη στην πρωτοταγή της διάταξη, ήταν απαραίτητο να περάσει από όλες τις δυνατές διαμορφώσεις για να καταλήξει στην λειτουργική της δομή, ο απαιτούμενος χρόνος θα ήταν αστρονομικών διαστάσεων (ποσότητα συγκρίσιμη με την ηλικία του σύμπαντος) . Παρ' όλα αυτά κάτι τέτοιο πρακτικά δεν συμβαίνει.

Το παράδοξο αυτό προδίδει την ύπαρξη κάποιου πιο πρακτικού μοντέλου που μπορεί να λειτουργεί σε λογικά, για τους βιολογικούς οργανισμούς, χρονικά πλαίσια. Μια υποψία η οποία οδήγησε στη σύλληψη της θεωρίας των τοπίων ελεύθερης ενέργειας Gibbs (*energy landscapes*), που διευκρινίζει ότι η εξερεύνηση των πιθανών διαμορφώσεων κατά την αναδίπλωση δεν είναι εντελώς τυχαία, αλλά κατά-κάποιο τρόπο καθοδηγούμενη από 'αυθυποβαλλόμενους' περιορισμούς (κάθε διαφορετική, δηλαδή, διαμόρφωση επιτρέπει τη μετεξέλιξη σε ορισμένες παράγωγες επακόλουθες διαμορφώσεις και όχι σε οποιαδήποτε άλλη τυχαία) καταλήγοντας στην ευνοϊκότερη ενεργειακά διαμόρφωση που αποτελεί και τη φυσική της κατάσταση. Στα *free-energy landscapes* αυτό συνήθως αναπαρίσταται με μια δομή χωνιού που καταλήγει στο ελάχιστο δυνατό ενεργειακό σημείο(3). Ενδιάμεσες καταστάσεις και μεταβατικές διαδρομές (*folding pathways*) επηρεάζονται από παράγοντες όπως δεσμοί υδρογόνου, δυνάμεις *van der Waals* και υδροφοβικότητα εξ αιτίας των οποίων σε κάθε στάδιο της αναδίπλωσης υπάρχουν και οι περιορισμοί στις δυνατές/πιθανές ακόλουθες διαμορφώσεις. Θα ήταν σωστό να επισημάνουμε σ' αυτό το σημείο ότι υφίστανται και περιπτώσεις με περισσότερες από μία τελικές διαμορφώσεις.

Πρόσφατες, σχετικά, πρόοδοι στο πεδίο των προσομοιώσεων μοριακής

δυναμικής έχουν ξεδιαλύνει το ζήτημα της αναδίπλωσης σε σημαντικό βαθμό. Οι εν λόγω προσομοιώσεις χρησιμοποιούνται στην εξερεύνηση διαδρομών αναδίπλωσης, ενδιάμεσων διαμορφώσεων, την επιρροή μεταλλάξεων καθώς και τη θερμοδυναμική διάσταση του φαινομένου (7,11). Επιτρέπουν επίσης στους επιστήμονες την αναπαράσταση σπάνιων μεταβάσεων και φαινομένων αναδίπλωσης που είναι πολύ δύσκολο να προσδιοριστούν πειραματικά. Για παράδειγμα, μελέτες αναδίπλωσης μικρών πρωτεϊνών όπως κεφαλές πρωτεϊνών villin-1 ή trp-cage μέχρι στιγμής επιβεβαιώνουν την ισχύ των προβλέψεων αυτής της τεχνικής.

Παρά το σημαντικό βαθμό προόδου, πάνω στο ζήτημα, παραμένουν αρκετές προκλήσεις στη βελτίωση της ακρίβειας προβλέψεων στους ρυθμούς αναδίπλωσης, την κατανόηση της επιρροής του κυτταρικού περιβάλλοντος και τη γεφύρωση ανάμεσα στα αναμενόμενα αποτελέσματα των προσομοιώσεων και των πειραματικών μεθόδων. Η έλευση τεχνολογιών όπως το machine-learning και άλλων ενισχυμένων υπολογιστικών εργαλείων, το ζήτημα της πρόβλεψης της πρωτεϊνικής αναδίπλωσης αντιμετωπίζεται με αυξανόμενη ακρίβεια, ενώ υπόσχεται βαθύτερη κατανόηση της θεμελιώδους αυτής βιολογικής διαδικασίας.

1.2 EPI-X4: Ενδογενής ανταγωνιστής του υποδοχέα CXCR4

Το EPI-X4 είναι μία 16μερής πεπτιδική αλυσίδα που έχει ταυτοποιηθεί ως ενδογενής ανταγωνιστής του υποδοχέα χημοκινών CXCR4 που συμμετέχει σε πληθώρα βιολογικών διεργασιών, όπως αιμοποίηση, κυκλοφορία κυττάρων του ανοσοποιητικού, οργανογένεση ενώ εμπλέκεται αξιοσημείωτα σε παθολογικές καταστάσεις όπως η καρκινική μετάσταση, η φλεγμονή και η μόλυνση από HIV-1. Η EPI-X4 αποτελεί τμήμα της αλβουμίνης του ανθρώπινου ορού, από την οποία και αποσπάται κάτω από συγκεκριμένες όξινες συνθήκες, προσδένεται ειδικά σε εξωκυττάριο βρόχο του CXCR4, εμποδίζοντας έτσι την αλληλεπίδραση με τη χημοκίνη CXCL12 (γνωστής και ως SDF-1a) το φυσιολογικό προσδέτη (ligand) αυτού του υποδοχέα. Η καταστολή αυτή εμποδίζει την CXCL12-εξαρτώμενη κυτταρική σηματοδότηση, μετανάστευση και εισβολή πράγμα που προσδίδει στην EPI-X4 υψηλό ενδιαφέρον για πιθανές θεραπευτικές εφαρμογές.

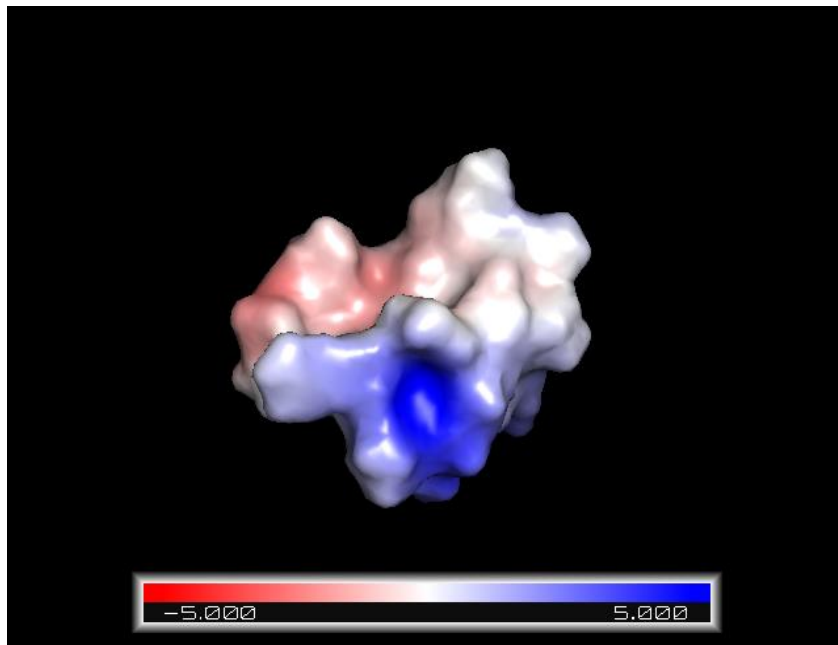
Λειτουργικά η EPI-X4 έχει παρατηρηθεί να φέρει την ικανότητα κινητοποίησης αιμοποιητικών βλαστοκυττάρων και πρόδρομων κυττάρων, μείωσης επιστράτευσης ανοσοποιητικών κυττάρων της φλεγμονώδους αντίδρασης και καταστολής της μεταστατικής καρκινικής επέκτασης σε προκλινικά μοντέλα. Επιπλέον έχει δείξει πιθανή αντι-ασθματική δράση, ρυθμίζοντας την ενεργότητα ανοσοποιητικών κυττάρων στους πνεύμονες. Εν αντιθέσει με συνθετικούς ανταγωνιστές CXCR4 (όπως π.χ. η AMD3100) η EPI-X4 επίσης καταστέλλει βασικού επιπέδου CXCR4 σηματοδότηση πράγμα που ενδέχεται να επιτρέψει

ευρύτερα οφέλη με ταυτόχρονη αποφυγή παρενεργειών συνδεδεμένων με τη χρήση συνθετικών φαρμάκων.

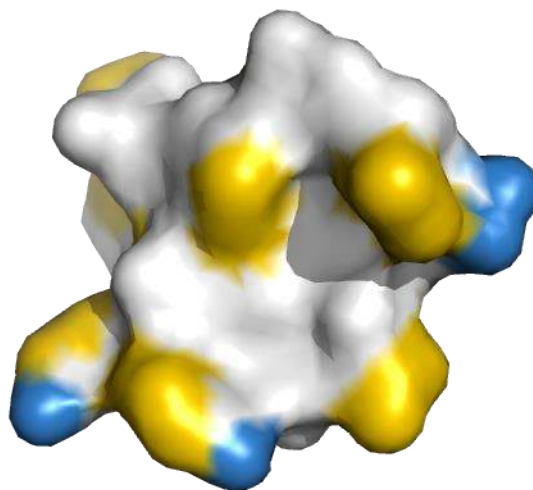
Δομικά, η EPI-X4 υιοθετεί μια διαμόρφωση βελτιστοποιημένη για την πρόσδεση σε CXCR4. Η παραγωγή του πεπτιδίου στηρίζεται σε ασπαρτικές πρωτεάσες (όπως οι καθεψίνες D και E) που τέμνουν την αλβουμίνη του ανθρώπινου ορού κάτω από συνθήκες τοπικής οξύνισης που συναντάται συνήθως στο μικροπεριβάλλον φλεγμονής ή όγκου. Πρόσφατες δομικές μελέτες έχουν προσδιορίσει συγκεκριμένες περιοχές της EPI-X4 εμπλεκόμενες σε αλληλεπιδράσεις με υποδοχείς, διαφωτίζοντας το μηχανισμό δράσης και επιτρέποντας έτσι το σχεδιασμό ενισχυμένων συνθετικών παράγωγων με βελτιωμένη σταθερότητα και αποτελεσματικότητα. Αυτά τα παράγωγα είναι πολλά υποσχόμενα για την εξέλιξη του EPI-X4 από απλή ανακάλυψη σε κλινικά βιώσιμο υποψήφιο στόχο φαρμακευτικού σχεδιασμού. Η πεπτιδική αλληλουχία του EPI-X4, όπως έχει προσδιοριστεί πειραματικά μέσω τεχνικών φασματοσκοπίας NMR (TOCSY, NOESY και τη βοήθεια του αλγορίθμου CYANA) (6), αποκαλύπτει μια ιδιαίτερα δυναμική και ευέλικτη δομή, η οποία χαρακτηρίζεται από την απουσία σταθερών δευτεροταγών στοιχείων όπως α-έλικες και β-πτυχωτά φύλλα. Αυτή η εγγενής αταξία επιτρέπει στο πεπτίδιο να προσαρμόζει τη διαμόρφωσή του ανάλογα με το περιβάλλον του, γεγονός που πιθανώς παίζει καθοριστικό ρόλο στην αναστολή της δραστηριότητας του CXCR4.

1.3 Ηλεκτροστατικές Ιδιότητες και Ανάλυση Επιφανειακών Χαρακτηριστικών

Η σύνδεση του EPI-X4 με τον CXCR4 καθορίζεται σε μεγάλο βαθμό από ηλεκτροστατικές αλληλεπιδράσεις και την κατανομή των φορτισμένων και υδρόφοβων περιοχών στην επιφάνειά του. Ο υπολογισμός του ηλεκτροστατικού δυναμικού με τη μέθοδο APBS μέσω PyMOL αποκαλύπτει σημαντικές περιοχές θετικού (μπλε) και αρνητικού (κόκκινου) φορτίου, οι οποίες πιθανώς αντιστοιχούν σε κρίσιμα σημεία σύνδεσης με τον υποδοχέα CXCR4 (Εικόνα 1 και 2).



Εικόνα 1: Ηλεκτροστατικό δυναμικό στην επιφάνεια του μορίου EPI-X4 (υπολογισμένο με το APBS plugin στο PyMol) Διακρίνουμε το σχηματισμό μιας θετικά φορτισμένης περιοχής που συμπίπτει με το σχηματισμό βρόχου (από τα κατάλοιπα Arg 3, Lys 6 και Lys 7) πράγμα που επηρεάζει και διαδικασία αναδίπλωσης και την ενεργότητα του πεπτιδίου (π.χ. αλληλεπίδραση με CXCR4).



Εικόνα 2: Αναπαράσταση ηλεκτροστατικού δυναμικού και υδροφοβικότητας με τη χρήση YRB plugin.

Κίτρινο: Υδροφοβικά άτομα C μη συνδεδεμένα απ' ευθείας με άτομα N ή O.

Μπλέ: Θετικά φορτισμένα άτομα N που συμπεριλαμβάνονται σε κατάλοιπα Arg και Lys

Λευκό: Τα υπόλοιπα άτομα του πεπτιδικού σκελετού

1.4 Ο στόχος της συγκεκριμένης εργασίας.

Η παρούσα μελέτη ερευνά την συμπεριφορά αναδίπλωσης του 16μερους πεπτιδίου EPI-X4, που αποσπάται κάτω από ορισμένες συνθήκες από την αλβουμίνη του ανθρώπινου ορού. Στόχος η ανάλυση δεδομένων προσομοίωσης μοριακής δυναμικής και η τελική σύγκριση της ακρίβειας και ποιότητας τους με τα διαθέσιμα πειραματικά δεδομένα που έχουν ήδη προσδιορίσει την τελική βιολογικά ενεργή στερεοδιαμόρφωση του EPI-X4. Οι σχετικές προσομοιώσεις προσφέρουν το πλεονέκτημα της απεικόνισής της διαδρομής προς την τελική διαμόρφωση εξ ολοκλήρου, συλλαμβάνοντας και τις ενδιάμεσες δομές, τα ενεργειακά όρια και τα συχνότερα εμφανιζόμενα δομικά μοτίβα προσδιορίζοντας έτσι το διαμορφωτικό χώρο του πεπτιδίου στο σύνολό του. Προηγούμενες μελέτες έχουν ασχοληθεί εκτενώς με την συμπεριφορά μικρών πρωτεϊνών-πεπτιδίων κατά την αναδίπλωση ενώ έχουν επίσης επιδείξει την ισχύ της εφαρμογής των MDs στην εξερεύνηση των συστημάτων αυτών όσον αφορά τις αλληλεπιδράσεις μοριακής δυναμικής, τα τοπία ελεύθερης ενέργειας και τις μεταβατικές δομές (7,11). Η σύγκριση των αποτελεσμάτων αυτών με τα πειραματικά δεδομένα είναι ένας κατηγορηματικός τρόπος αξιολόγησης τους που μπορεί ενδεχομένως να χρησιμεύσει στον εντοπισμό των αδυναμιών ή ελλείψεων των προσομοιώσεων και στην επακόλουθη μελλοντική βελτίωση των μοντέλων αυτών.

Στα πλαίσια αυτής της εργασίας χρησιμοποιήσαμε προγράμματα που μπορούν να διεκπεραιώσουν στατιστικές αναλύσεις (carma,grcarma) για τον εντοπισμό στιγμιότυπων των πιο ενεργειακά και θερμοδυναμικά σταθερών δομών σε δεδομένα αρχείων dcd-psf. Λόγω της ύπαρξης μη διαχειρίσιμου αριθμού μεταβλητών και βαθμών ελευθερίας πραγματοποιήσαμε ανάλυση κύριων συνιστωσών (PCA). Έτσι μειώνεται σημαντικά το σύνολο των μεταβλητών ενώ ταυτόχρονα διατηρείται το μεγαλύτερο ποσοστό χρήσιμης πληροφορίας.

2. Προσομοίωση Μοριακής Δυναμικής

2.1 Χρήση και στόχοι της τεχνικής των προσομοιώσεων μοριακής δυναμικής

Οι προσομοιώσεις μοριακής δυναμικής (molecular dynamics simulations – MDs) αποτελούν μια πανίσχυρη τεχνική για τη μελέτη της δυναμικής συμπεριφοράς βιολογικών μακρομορίων σε ατομικό επίπεδο. Επιλύοντας τις εξισώσεις κίνησης του

Νεύτωνα, οι MDs επιτρέπουν την εξερεύνηση δομικών διαμορφώσεων των βιομορίων, αλληλεπιδράσεων και διαδικασιών, εκτεινόμενων σε χρονικά πλαίσια της τάξεως από femtosecond σε millisecond. Το είδος και η ποσότητα των αποκτηθέντων πληροφοριών μέσω των εν λόγω προσομοιώσεων τις καθιστά φυσικά ανεκτίμητες στην κατανόηση πολύπλοκων βιολογικών δεδομένων εκ φύσεως αδύνατον να μελετηθούν με συμβατικές πειραματικές μεθόδους.

Οι MDs χρησιμοποιούνται ευρέως στη μελέτη πρωτεϊνικής αναδίπλωσης και αλλαγών στερεοδιαμόρφωσης, παρέχοντας λεπτομερείς υπολογισμούς πάνω στο concept των τοπίων ελεύθερης ενέργειας Gibbs, που διέπει τις διαδικασίες αυτές (11). Επίσης επιτρέπουν, στους ερευνητές, την μελέτη και παρατήρηση των ενδιάμεσων μεταβατικών δομών καθώς και προβλέψεις πιθανών διαδρομών αναδίπλωσης με εντυπωσιακή, πλέον, ακρίβεια. Παραδείγματος χάριν, τέτοιες προσομοιώσεις υπήρξαν καίριες στη διαλεύκανση των μηχανισμών αναδίπλωσης σε ολιγοπεπίδια και μικρές πρωτεΐνες όπως οι κεφαλές villin-1 κάτω από σχετικές πειραματικές συνθήκες (7)

Μία έτερη ενδιαφέρουσα εφαρμογή των MDs εναπόκειται στην κατανόηση αλληλεπιδράσεων πρωτεΐνης-συνδέτη ή πρωτεΐνης-πρωτεΐνης. Η αναπαράσταση των αλληλεπιδράσεων αυτών, σε ατομικό επίπεδο παρέχουν τη δυνατότητα παρατήρησης των μηχανισμών σχηματισμού χημικών δεσμών καθοδηγώντας έτσι και το σχετικό φαρμακευτικό σχεδιασμό (8). Πέρα από τα αμιγώς πρωτεϊνικά φαινόμενα δυναμικής, οι MDs έχουν αποδειχτεί ανεκτίμητες και στη μελέτη διαδικασιών συναρμολόγησης βιομορίων όπως σχηματισμός μεμβρανών, συσσωμάτων μυκκιλίων, και αναδίπλωση νουκλεϊκών οξέων.

Οι συνεχείς βελτιώσεις στα διάφορα υπολογιστικά μοντέλα force field και στη διαθέσιμη υπολογιστική ισχύ επεκτείνουν διαρκώς τις δυνατότητες των MDs ως προς την εφαρμογή τους για μεγαλύτερη χρονική διάρκεια και συστήματα αυξημένης πολυπλοκότητας. Ενισχυμένες τεχνικές δειγματοληψίας έχουν ήδη διευρύνει την έκταση των δυνατών πεδίων εφαρμογής των MDs επιτρέποντας την παρατήρηση και μελέτη σπάνιων φαινομένων (11)

Εν ολίγοις, οι προσομοιώσεις μοριακής δυναμικής αποτελούν ένα ευέλικτο εργαλείο στη διαλεύκανση λειτουργίας βιολογικών διαδικασιών, από το θεμελιώδες μοριακό επίπεδο έως το ιδιαίτερα πολύπλοκο κυτταρικό περιβάλλον. Η ισχύς των δυνατών προβλέψεων και η ικανότητα τους να παρέχουν συμπληρωματικά επιβεβαιωτικά δεδομένα σε σχέση με τα πειραματικά, τις καθιστούν αναντικατάστατο τμήμα της σύγχρονης έρευνας στον κλάδο της μοριακής βιολογίας.

2.2 Προετοιμασία συστήματος:

Για τη διερεύνηση της δομικής συμπεριφοράς του EPI-X4, κάτω από φυσιολογικές συνθήκες, επιστρατεύσαμε την τεχνική της προσομοίωσης μοριακής δυναμικής χρησιμοποιώντας το πρόγραμμα NAMD (nanoscale molecular dynamics). Πρόκειται για ευρέως διαδεδομένο λογισμικό, διεκπεραίωσης προσομοιώσεων, υψηλής απόδοσης, σχεδιασμένο για το χειρισμό μεγάλης γκάμας μεγεθών και πολυπλοκότητας, συστημάτων βιομορίων. Για την λειτουργία του εν λόγω λογισμικού είναι απαραίτητα 3 διαφορετικά αρχεία: Ένα αρχείο τοπολογίας που ορίζει την δομή και συνδεσμολογία του μορίου και τις παραμέτρους του πεδίου (υπολογιστικό μοντέλο force field)(.prmtop), ένα αρχείο συντεταγμένων με θέσεις ατόμων(.pdb), και ένα αρχείο προσαρμογής που διευκρινίζει τις συνθήκες και το πρωτόκολλο της προσομοίωσης (collapse.namd).

Στην παρούσα εργασία η αρχική δομή του EPI-X4, βρισκόταν σε πλήρως ανεπτυγμένη μορφή (unfolded state) για να διεξαχθεί η προσομοίωση αναδίπλωσης. Η τοπολογία του συστήματος και οι αντίστοιχες παράμετροι ορίστηκαν με τη χρήση AMBER force field ff99SB-ILDN, το οποίο συμπεριλαμβάνει βελτιωμένα στοιχεία όσον αφορά τον κύριο σκελετό του μορίου (backbone) και τον ορισμό των βαθμών ελευθερίας περιστροφής των πλευρικών αλυσίδων. Το μοντέλο νερού που χρησιμοποιήθηκε ήταν το TIP3P. Οι διαστάσεις του «δοχείου» της προσομοίωσης ήταν : κυβικό «δοχείο» με πλευρές μήκους 39.16 Å

3.Μέθοδοι και Αποτελέσματα

3.1 Πίνακας RMSD και ανάλυση διαγράμματος εμφάνισης δευτεροταγών δομών σε συνάρτηση με το χρόνο

Για την αξιολόγηση του εύρους της ποικιλίας διαμορφώσεων και δομικών μεταβάσεων κατά τη διάρκεια της MD προσομοίωσης, πραγματοποιήθηκε (μέσω carma) ανάλυση rmsd (μεταξύ όλων των πιθανών ζευγών frames του τροχιακού) και η παραγωγή σχετικού πίνακα παρουσίασης των αποτελεσμάτων. Ο υπολογισμός rmsd μεταξύ δυο δομών ορίζεται ως:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_i^{(a)} - \mathbf{r}_i^{(b)}\|^2}$$

όπου N ο αριθμός ατόμων που λαμβάνεται υπόψη (τυπικά άτομα του backbone του

μορίου ή 'α' άνθρακες (Cα)) ενώ $r_i^{(a)}$ και $r_i^{(b)}$ οι θέσεις του εκάστοτε ατόμου εντός των δομών a και b αντίστοιχα. Αυτός ο υπολογισμός ποσοτικοποιεί την μέση μετατόπιση των ατόμων ανάμεσα σε δυο διαφορετικά στιγμιότυπα του τροχιακού και χρησιμοποιείται ευρέως σαν μέτρο δομικής ομοιότητας. Ο πίνακας RMSD που ακολουθεί προκύπτει με χρωματική διαβάθμιση που αντιστοιχεί στα εξής:

Μπλε: ~0-4 Å

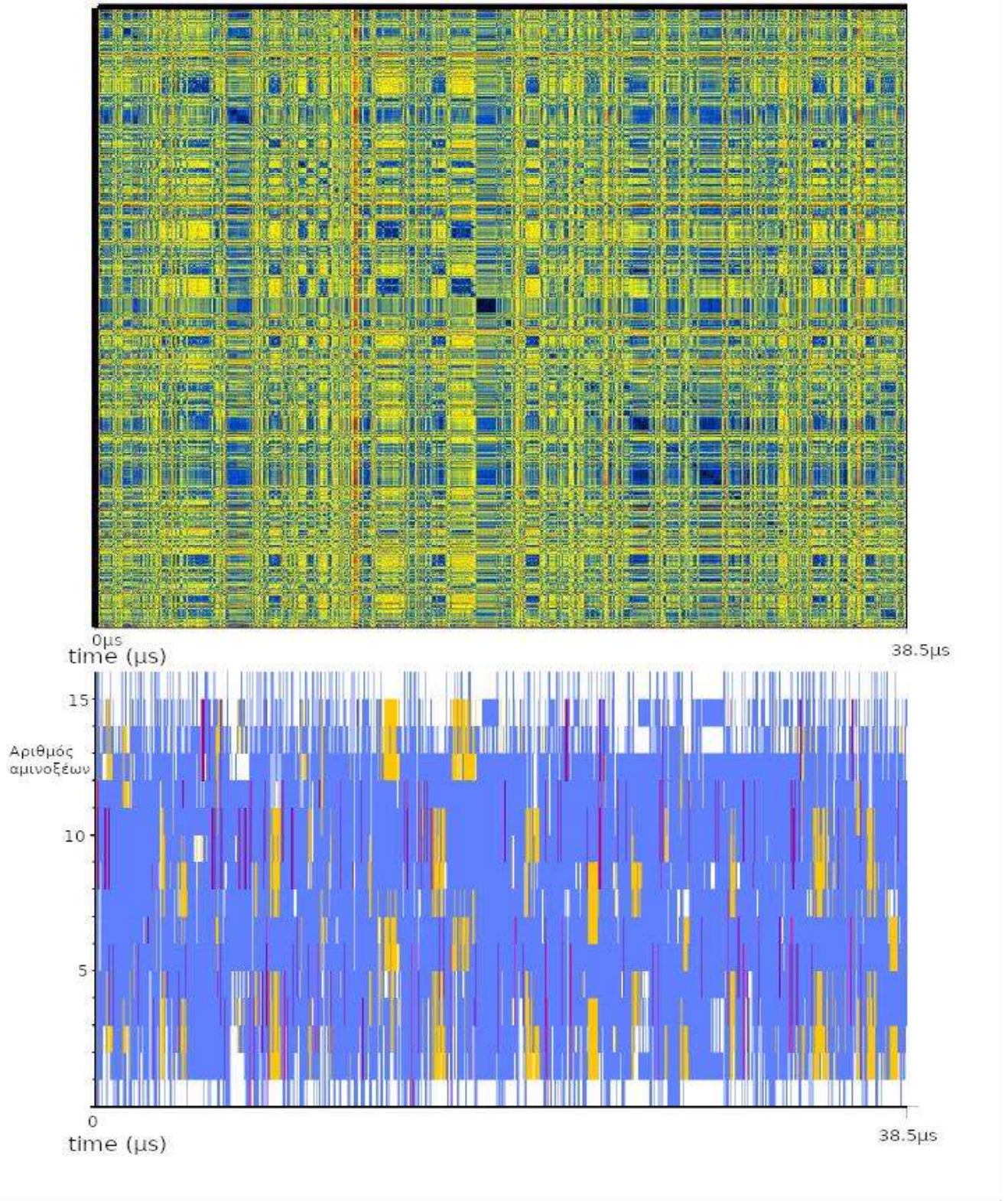
Κίτρινο: ~5-7.5 Å

Κόκκινο: > 8Å (με ανώτατες τιμές να πλησιάζουν ~12 Å)

Ο πίνακας RMSD αποτελεί μια συμπαγή οπτική αναπαράσταση της εξέλιξης των δομικών διαμορφώσεων του μορίου κατά τη διάρκεια της προσομοίωσης. Διακριτά μπλε τμήματα κοντά στη διαγώνιο συνήθως ανταποκρίνονται σε διαμορφώσεις με μεγαλύτερη σχετικά σταθερότητα.

Ακριβώς από κάτω από τον πίνακα ακολουθεί στοιχισμένο γράφημα εμφάνισης δευτεροταγών δομών σε συνάρτηση με το χρόνο. Ο υπολογισμός αυτός πραγματοποιήθηκε με εφαρμογή του γνωστού αλγορίθμου STRIDE (μέσω της εφαρμογής gcrarma) ο οποίος προσδιορίζει δευτεροταγή δομή βάση της γεωμετρίας του βασικού σκελετού του μορίου και του υπάρχοντος μοτίβου δεσμών υδρογόνου. Το παρόν γράφημα αναπαριστά την εξέλιξη της εμφάνισης, όπως αναφέραμε, δευτεροταγών δομών σε συνάρτηση με το χρόνο, με κάθε χρώμα να αντιστοιχεί σε έναν συγκεκριμένο τύπο δευτεροταγούς δομής (όπως παρουσιάζονται στο υπόμνημα). Ο άξονας y αντιστοιχεί στον αριθμό αμινοξέων και στο είδος δευτεροταγούς δομής (ή την απουσία αυτής) που εμφανίζεται σε κάθε χρονική στιγμή.

Σε συνδυασμό με τον πίνακα RMSD το διάγραμμα δευτεροταγούς δομής προσφέρονται για μια διορατική εξαγωγή συμπερασμάτων όσον αφορά τη δυναμική της πρωτεϊνικής αναδίπλωσης και δομικής σταθερότητας του πεπτιδίου κατά τη διάρκεια της προσομοίωσης. Διαπιστώνουμε δε τη χαρακτηριστική έλλειψη σταθερών δευτεροταγών δομών κατά τη διάρκεια εξερεύνησης του διαμορφωτικού χώρου.



Εικόνα 3: Πίνακας RMSD σε στοιχισή με διάγραμμα secondary structure

3.2 Nuclear overhauser effects (NOE)

Πρόκειται για το θεμελιώδες φυσικό φαινόμενο στο οποίο βασίζεται ολόκληρη η τάξη τεχνικών πυρηνικής μαγνητικής φασματοσκοπίας (NMR spectroscopy). Περιγράφεται από την αλλαγή έντασης του εκπεμπόμενου σήματος ενός πυρήνα σε ιδιοπεριστροφή σαν αποτέλεσμα της εγγύτητας σε κάποιον άλλο περιστρεφόμενο πυρήνα. Στις τεχνικές NMR που χρησιμοποιούνται στη μελέτη βιομορίων, το φαινόμενο αυτό επιτρέπει την εκτίμηση αποστάσεων μεταξύ ατόμων H(πρωτονίων) της τάξεως των $\sim 5 \text{ \AA}$. Η πλέον αξιοσημείωτη ιδιότητα των NOE είναι ότι η ένταση τους είναι αντιστρόφως ανάλογη της απόστασης μεταξύ αλληλεπιδρώντων πυρήνων στην 6^η δύναμη:

$$I \propto \frac{1}{r^6}$$

Η εμφανώς ισχυρή εξάρτηση από την απόσταση, καθιστά τους περιορισμούς στην ταυτοποίηση τοπολογίας των ατόμων ιδιαίτερος καθοριστικούς. Ακόμα και μικρές, σχετικά, αλλαγές στις εν λόγω αποστάσεις επιφέρουν σημαντικές αλλαγές στην εκτίμηση των αποστάσεων αυτών με βάση τα δεδομένα NOE, πράγμα που εξηγεί το γιατί το r^{-6} averaging (στις αναλύσεις MD) αντικατοπτρίζει πιο πιστά την εικόνα πειραματικά προσδιορισμένων NOE, που άλλωστε αποτελεί και το λόγο προτίμησης στην παρουσίαση της σύγκρισης πειραματικών και MD δεδομένων.

Είναι δε αξιοσημείωτο, ότι οι μετρήσεις NOE δεν παρέχουν ακριβείς αποστάσεις αλλά μάλλον τα «ανώτατα επιτρεπτά όρια» αυτών. Τις μέγιστες παρατηρούμενες αποστάσεις που παραμένουν συμβατές με το καταγεγραμμένο σήμα παλμού ραδιοσυχνότητας NOE. Στην σύγκριση συνεπώς, με MD δεδομένα παρατηρούμε τις r^{-6} averaged αποστάσεις αν υπερβαίνουν τα NMR προσδιορισμένα ανώτατα όρια (οι ενδεχόμενες υπερβάσεις αποτελούν χαρακτηριστικό δείκτη ασυμβατότητας ανάμεσα στα αποτελέσματα MD και πειραματικών δεδομένων)

3.3 Υπολογισμός NOE και σύγκριση με πειραματικά δεδομένα

Για την εκτίμηση της σύγκλισης των αποτελεσμάτων συναγόμενων εκ της προσομοίωσης MD και των πειραματικά προσδιορισμένων δεδομένων, πραγματοποιούμε συγκρίσεις των υπολογισμένων αποστάσεων μεταξύ πρωτονίων (όπως προκύπτουν από την προσομοίωση) με τις αντίστοιχες αποστάσεις από μετρήσεις NOE όπως αυτές παρουσιάζονται στο κατατεθειμένο αρχείο στην pdb (2n0x.mr). Επειδή, ωστόσο, η αρχική MD προσομοίωση πραγματοποιήθηκε με παραμέτρους μεταβλητής θερμοκρασίας, θεωρήθηκε σκόπιμο να παρουσιαστούν

συγκρίσεις με υπολογισμένες αποστάσεις από frames που ανταποκρίνονται σε διαφορετικά όρια θερμοκρασίας. Τέλος όλοι οι πίνακες συγκρίσεων ακολουθούνται από αντίστοιχους πίνακες υπολογισμών συσχετισμού Pearson για την εξέταση της γραμμικής συσχέτισης μεταξύ των υπολογιστικών και πειραματικών αποστάσεων του εκάστοτε πίνακα. Όλες οι συγκρίσεις γίνονται με τις υπολογισμένες αποστάσεις r^{-6} καθώς αντανακλούν ακριβέστερα τη φυσική εξάρτηση της έντασης των σημάτων NOE από την απόσταση μεταξύ πυρήνων, όπως ορίζεται από την αντίστοιχη εξίσωση της φασματοσκοπίας NMR. Η μέθοδος των υπολογισμών αποστάσεων από τα αποτελέσματα MD συμπεριέλαβε τα εξής βήματα:

1. Εξαγωγή λίστας ζευγών πρωτονίων από το αρχικό αρχείο .psf (πληροφορίες τοπολογίας και παραμέτρων συνδεσιμότητας του μορίου υπό μελέτη) με τη βοήθεια κατάλληλου perl script
2. Από την εν λόγω λίστα επιλέχθηκαν τα ζεύγη πρωτονίων που επίσης αναφέρονται ονομαστικά στο αρχικό paper. Στη διαλογή αυτή λήφθηκαν υπόψη οι ενδείξεις ονομασίας κατά IUPAC (20)
3. Το Carma με τη χρήση σχετικής εντολής (-dist) παρήγαγε κατά μέσο όρο υπολογισμένες πρωτονιακές αποστάσεις εξετάζοντας το σύνολο των ~22 εκατομμυρίων στιγμιότυπων του τροχιακού.
4. Το αρχείο υπολογισμών του Carma (Carma.distances) μαζί με την επεξεργασμένη λίστα πρωτονίων, δόθηκε προς επεξεργασία στο noe.averaging.c script για τον τελικό κατά μέσο όρο υπολογισμών r^{-3} και r^{-6} αποστάσεων

Πίνακας 1: Σύγκριση πειραματικά υπολογισμένων NOE αποστάσεων και υπολογισμών μέσω MD αποτελεσμάτων (όπως προκύπτουν από το σύνολο των frames του τροχιακού)

Proton 1	Proton 2	MD r^{-6} (Å)	Exp. NOE (Å)	Deviation
QD2 LEU 1	HB3 PRO 9	5.53	4.47	+1.06
QD2 LEU 1	HB2 PRO 9	5.27	4.47	+0.80
QD1 LEU 1	HB3 PRO 9	5.24	4.47	+0.77
QD2 LEU 1	HD3 PRO 9	6.79	5.50	+1.29
QD TYR 4	HD3 PRO 9	4.41	5.11	No violation
QD TYR 4	HG2 PRO 9	5.79	5.50	+0.29
QE TYR 4	HG2 PRO 9	5.60	5.14	+0.46
QD TYR 4	QB PRO 9	4.45	5.34	No violation
QD TYR 4	QG2 THR 13	4.16	5.20	No violation
QG2 THR 5	QG2 THR 13	4.50	3.05	+1.45
HA THR 5	HD3 PRO 14	2.25	3.46	No violation
HB THR 5	QD PRO 14	6.66	3.76	+2.90
QG2 THR 5	HD3 PRO 14	5.84	3.87	+1.97

Parameter	Value
Pearson correlation coefficient (r)	0.2776
r ²	0.07707
P-value	0.3585
Covariance	0.2712
Sample size (n)	13
Statistic	0.9584

Η σύγκριση μεταξύ των υπολογισμένων αποστάσεων NOE από την προσομοίωση μοριακής δυναμικής και των αντίστοιχων πειραματικών τιμών έδειξε μέτρια συνολική συμφωνία για το σύνολο των δομών του τροχιακού. Η τιμή του συσχετισμού Pearson για το σύνολο των frames είναι θετική αλλά όχι υψηλή, αντανακλώντας την ύπαρξη γενικής τάσης συμφωνίας, αλλά και σημαντικών αποκλίσεων σε επιμέρους περιπτώσεις. Οι μεγαλύτερες αποκλίσεις σχετίζονται κυρίως με περιοχές του πεπτιδίου που εμφανίζουν μεγαλύτερη ευελιξία ή λιγότερη δομική σταθερότητα, ενώ άλλες επαφές –όπως οι HA–HD3 και QD2–HB3–εμφανίζονται πιο κοντά στις πειραματικές τιμές.

Ακολουθούν συγκρίσεις υπολογισμών στους οποίους λήφθηκαν υπόψη frames του τροχιακού που ανταποκρίνονται σε διαφορετικές θερμοκρασίες.

Πίνακας 2: Θερμοκρασία < 320K

Proton 1	Proton 2	MD r ⁻⁶ (Å)	Exp. NOE (Å)	Deviation
QD2 LEU 1	HB3 PRO 9	5.44	4.47	+0.97
QD2 LEU 1	HB2 PRO 9	5.09	4.47	+0.62
QD1 LEU 1	HB3 PRO 9	5.10	4.47	+0.63
QD2 LEU 1	HD3 PRO 9	7.03	5.50	+1.53
QD TYR 4	HD3 PRO 9	4.38	5.11	No violation
QD TYR 4	HG2 PRO 9	5.62	5.50	+0.12
QE TYR 4	HG2 PRO 9	5.50	5.14	+0.36
QD TYR 4	QB PRO 9	4.32	5.34	No violation
QD TYR 4	QG2 THR 13	3.98	5.20	No violation
QG2 THR 5	QG2 THR 13	4.34	3.05	+1.29
HA THR 5	HD3 PRO 14	2.25	3.46	No violation
HB THR 5	QD PRO 14	6.83	3.76	+3.07
QG2 THR 5	HD3 PRO 14	5.76	3.87	+1.89

Parameter	Value
Pearson correlation coefficient (r)	0.2629
r^2	0.06911
P-value	0.3855
Covariance	0.2692
Sample size (n)	13
Statistic	0.9037

Πίνακας 3: Θερμοκρασία < 300K

Proton 1	Proton 2	MD r^{-6} (Å)	Exp. NOE (Å)	Deviation
QD2 LEU 1	HB3 PRO 9	5.39	4.47	+0.92
QD2 LEU 1	HB2 PRO 9	5.01	4.47	+0.54
QD1 LEU 1	HB3 PRO 9	5.02	4.47	+0.55
QD2 LEU 1	HD3 PRO 9	7.13	5.50	+1.63
QD TYR 4	HD3 PRO 9	4.37	5.11	No violation
QD TYR 4	HG2 PRO 9	5.58	5.50	+0.08
QE TYR 4	HG2 PRO 9	5.50	5.14	+0.36
QD TYR 4	QB PRO 9	4.25	5.34	No violation
QD TYR 4	QG2 THR 13	3.90	5.20	No violation
QG2 THR 5	QG2 THR 13	4.26	3.05	+1.21
HA THR 5	HD3 PRO 14	2.25	3.46	No violation
HB THR 5	QD PRO 14	7.03	3.76	+3.27
QG2 THR 5	HD3 PRO 14	5.76	3.87	+1.89

Parameter	Value
Pearson correlation coefficient (r)	0.2472
r^2	0.06111
P-value	0.4155
Covariance	0.263
Sample size (n)	13
Statistic	0.8461

Πίνακας 4: Θερμοκρασία < 280K

Proton 1	Proton 2	MD r^{-6} (Å)	Exp. NOE (Å)	Deviation
QD2 LEU 1	HB3 PRO 9	5.39	4.47	+0.92
QD2 LEU 1	HB2 PRO 9	5.00	4.47	+0.53
QD1 LEU 1	HB3 PRO 9	4.96	4.47	+0.49
QD2 LEU 1	HD3 PRO 9	7.28	5.50	+1.78
QD TYR 4	HD3 PRO 9	4.36	5.11	No violation
QD TYR 4	HG2 PRO 9	5.55	5.50	+0.05
QE TYR 4	HG2 PRO 9	5.49	5.14	+0.35
QD TYR 4	QB PRO 9	4.20	5.34	No violation
QD TYR 4	QG2 THR 13	3.82	5.20	No violation
QG2 THR 5	QG2 THR 13	4.20	3.05	+1.15
HA THR 5	HD3 PRO 14	2.24	3.46	No violation
HB THR 5	QD PRO 14	7.33	3.76	+3.57
QG2 THR 5	HD3 PRO 14	5.78	3.87	+1.91

Parameter	Value
Pearson correlation coefficient (r)	0.2239
r^2	0.05012
P-value	0.4622
Covariance	0.2517
Sample size (n)	13
Statistic	0.7618

Η εφαρμογή ορίων θερμοκρασίας στην επιλογή δομών του τροχιακού δείχνει σταδιακή βελτίωση της συμφωνίας με τα πειραματικά δεδομένα. Οι τιμές των αποκλίσεων μειώνονται σταδιακά με την επιλογή χαμηλότερων θερμοκρασιών (κάτω των 320K, 300K και 280K), γεγονός που υποστηρίζεται και από τις αντίστοιχες αυξήσεις στους συσχετισμούς Pearson. Η τάση αυτή είναι αναμενόμενη, καθώς οι δομές χαμηλότερης θερμοκρασίας τείνουν να αντανακλούν πιο σταθερές, «διπλωμένες» καταστάσεις, συγκρίσιμες με αυτές που παρατηρούνται πειραματικά. Τα αποτελέσματα αυτά επιβεβαιώνουν τη σημασία της θερμοκρασιακής επιλογής στην αξιολόγηση υπολογιστικών δεδομένων NOE, και ενισχύουν την αξιοπιστία της προσομοίωσης σε συνθήκες που προσεγγίζουν τη φυσιολογική σταθερότητα του πεπτιδίου.

3.5 Principal Component Analysis (PCA)

Πρόκειται για μια πανίσχυρη στατιστική μέθοδο που επιτυγχάνει το συνδυασμό διαθέσιμων μεταβλητών έπειτα από αξιολόγησή της επιρροής τους στην εικόνα του τελικού αποτελέσματος. Στόχος η σύμπτυξη του αριθμού των μεταβλητών (του υπό μελέτη συνόλου δεδομένων) σε έναν πιο διαχειρίσιμο αριθμό, ο οποίος ωστόσο εξακολουθεί να διατηρεί το μεγαλύτερο δυνατό ποσοστό της αρχικής πληροφορίας. Οι τελικές μεταβλητές που προκύπτουν από αυτή την επεξεργασία είναι οι πλέον σημαντικές για το εκάστοτε data set και είναι γνωστές ως κύριες συνιστώσες (principal components). Ευρέως, φυσικά, χρησιμοποιούμενη στην επεξεργασία τροχιακών που προκύπτουν από προσομοιώσεις MD. Εκ φύσεως τα βιομοριακά συστήματα μπορούν να συμπεριλαμβάνουν χιλιάδες ατόμων που υφίστανται συνεχείς διακυμάνσεις, στα πλαίσια μιας προσομοίωσης, στη διάρκεια εκατομμυρίων χρονικών βημάτων. Οι κύριες συνιστώσες στα πλαίσια της ανάλυσης ενός τροχιακού MD περιγράφουν επί συνόλου την κίνηση μορίων και ατόμων.

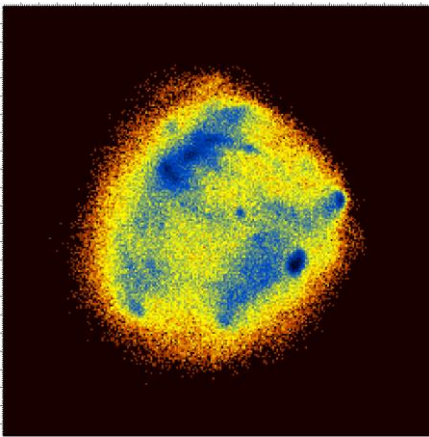
Επί της ουσίας, η PCA ταυτοποιεί τις κατευθύνσεις εντός του διαμορφωτικού χώρου προς τις οποίες, η κίνηση του συστήματος παρουσιάζει τις μεγαλύτερες διακυμάνσεις κατευθύνσεις αυτές προκύπτουν μέσω της κατασκευής ενός πίνακα συνδιακύμανσης των διακυμάνσεων των ατομικών θέσεων και στη συνέχεια με αποσύνθεση ιδιοτιμών. Τα ιδιοδιανύσματα που προκύπτουν αντιπροσωπεύουν ανεξάρτητους τρόπους κίνησης, ενώ οι αντίστοιχες ιδιοτιμές αποτυπώνουν το πλάτος της κίνησης σε κάθε άξονα. Οι πρώτες λίγες κύριες συνιστώσες περιλαμβάνουν συνήθως το μεγαλύτερο ποσοστό της συνολικής δυναμικής του συστήματος, επιτρέποντας έτσι την οπτικοποίηση της πορείας της προσομοίωσης σε δύο ή τρεις διαστάσεις, με ελάχιστη απώλεια κρίσιμης πληροφορίας.

Η χρησιμότητα της PCA, σ' αυτό το επίπεδο εναπόκειται στην ικανότητά της να εντοπίζει βαθύτερες δομικές τάσεις και μετασχηματισμούς, που συχνά δεν είναι ορατοί στη μελέτη ακατέργαστων δεδομένων. Ιδιαίτερα πολύτιμη στην εξέταση φαινομένων όπως η πρωτεϊνική αναδίπλωση. Μια επακόλουθη ανάλυση ομαδοποίησης δεδομένων (clustering), στις συντεταγμένες διαμορφωτικού χώρου παραγόμενες εκ της PCA, μπορεί να κατηγοριοποιήσει τα δεδομένα του τροχιακού περπατήριου, σε διακριτές μεταβατικές δομές.

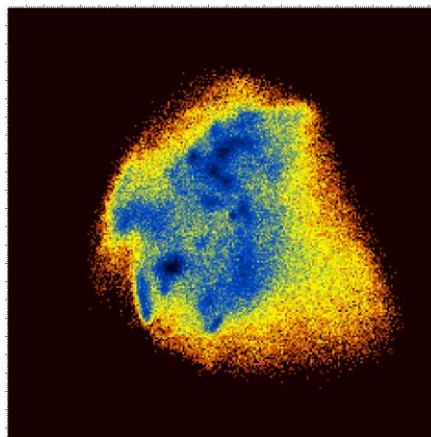
Στην παρούσα μελέτη, πραγματοποιήθηκαν αναλύσεις PCA με τη βοήθεια του CARMA(17). Το πρόγραμμα παρέχει τη δυνατότητα εφαρμογής cPCA και dPCA ανάλυσης. Στην καρτεσιανή PCA αναλύονται οι κινήσεις των ατόμων σε σύστημα καρτεσιανών συντεταγμένων στο σύνολο του διαμορφωτικού χώρου (καλύτερη για μια πιο μακροσκοπική σκοπιά των αλλαγών στο σύνολο του συστήματος). Στην «διεδρική» (dihedral) PCA η ανάλυση αφορά τις δίεδρες γωνίες (ϕ, ψ) των αμινοξέων και τις πιθανές διαμορφώσεις που προκύπτουν από την περιστροφή τους. Αν και

αμφότερες οι τεχνικές αυτές αποσκοπούν στην μείωση της πολυπλοκότητας στην οπτικοποίηση των δεδομένων του τροχιακού και στην αποκάλυψη των κυρίαρχων δομών και μοτίβων, η διαφορά στον τρόπο ορισμού και διαχείρισης των συντεταγμένων στο χώρο οδηγεί σε διαφορετικής φύσεως συμπεράσματα. Ο συνδυασμός των συμπληρωματικών αυτών τεχνικών δίνει, εν τέλει μια πιο εμπειριστατωμένη εικόνα για τον διαμορφωτικό χώρο του υπό μελέτη συστήματος.

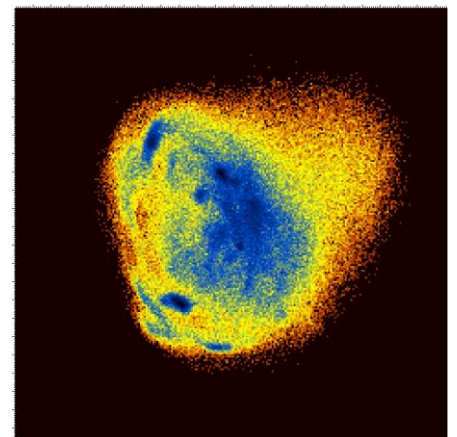
cPCA free energy landscapes



εικόνα 4: $PC1$ vs $PC2$

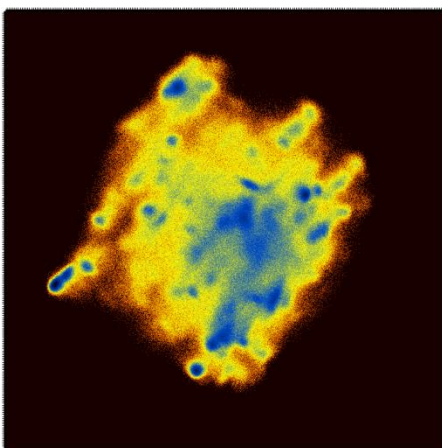


εικόνα 5: $PC1$ vs $PC3$

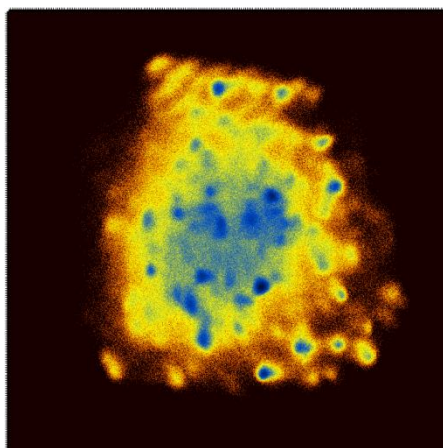


εικόνα 6: $PC2$ vs $PC3$

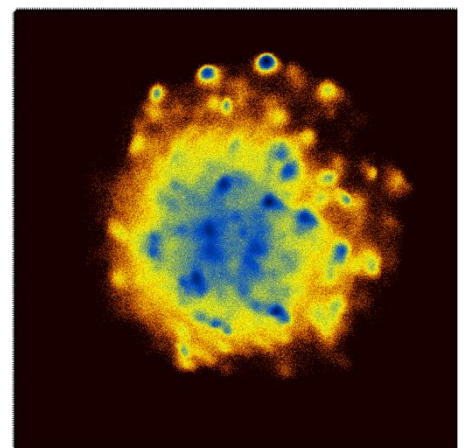
dPCA free energy landscapes



εικόνα 7: $PC1$ vs $PC2$



εικόνα 8: $PC1$ vs $PC3$



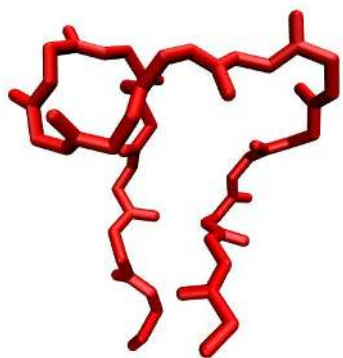
εικόνα 9: $PC2$ vs $PC3$

Κατά τη διεκπεραίωση των αναλύσεων PCA το CARMA παρήγαγε γραφικές αναπαραστάσεις των τοπίων ελεύθερης ενέργειας του συστήματος, σε δισδιάστατες προβολές, κάθε μια εκ των οποίων ανταποκρίνεται στο συνδυασμό δυο κυρίων συνιστωσών από τις τρεις πιο σημαντικές. Η χρωματική διαβάθμιση εκφράζει τα επίπεδα ελεύθερης ενέργειας στον διαμορφωτικό χώρο με τις μπλε περιοχές να αντιστοιχούν στις χαμηλότερες τιμές και συνεπώς πιο σταθερές δομές (global minima). Τα principal components στα οποία ανταποκρίνεται κάθε διάγραμμα σημειώνονται στον τίτλο της εκάστοτε εικόνας (εικόνες 3 έως 8). Οι πρώτες τρεις, κατά σειρά εμφάνισης προέκυψαν από την cPCA ενώ οι τελευταίες από την dPCA

3.6 cPCA και dPCA cluster analysis παρουσίαση αποτελεσμάτων

Οι εικόνες που ακολουθούν παρουσιάζουν δομικές συγκρίσεις ανάμεσα στο backbone της πειραματικής δομής του EPI-X4 (από επεξεργασία της PDB entry κωδικός:2n0x), και τις εξαχθείσες δομές εκ του τροχιακού όπως προέκυψαν από PCA και cluster analysis. Το backbone της πειραματικής δομής έχει διατηρηθεί στον ίδιο προσανατολισμό όπου αυτό ήταν δυνατό (με ορισμένες εξαιρέσεις για λόγους ευκρίνειας). Το backbone είναι σταθερά χρωματισμένο σε κόκκινο χρώμα ενώ οι υπόλοιπες δομές ανά κατάλοιπο. Σε κάθε εικόνα σημειώνονται ο αντίστοιχος τύπος ανάλυσης (cPCA, dPCA) αριθμός ομάδας, και RMSD σε Å όπως υπολογίστηκε μέσω των εργαλείων του VMD. Συμπεριλαμβάνονται συγκρίσεις με τις αντιπροσωπευτικές δομές κάθε cluster καθώς και οι δομές ανά cluster σε υπέρθεση (όπου αυτό ήταν ευκρινές).

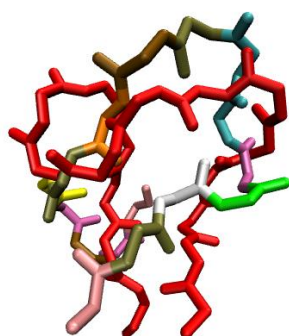
cPCA representatives



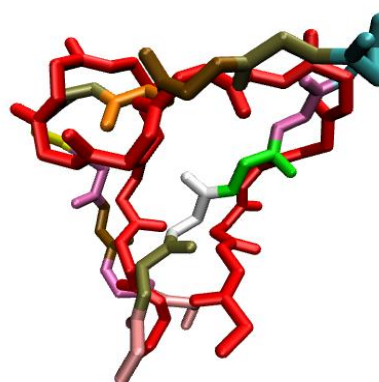
10:EPI-X4 backbone



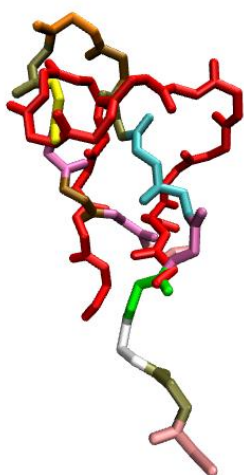
11:cPCA.rep.cluster 1 rmsd: 3,438 Å



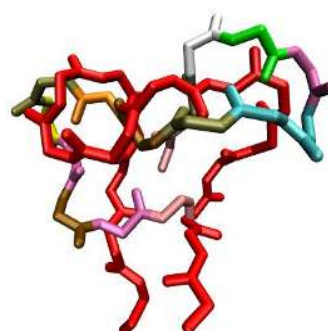
12:cPCA.rep.cluster 2 rmsd: 5,065



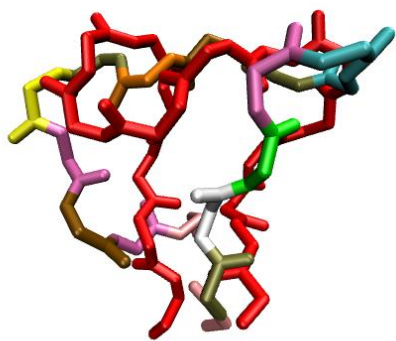
13:cPCA.rep.cluster 3 rmsd: 3,932



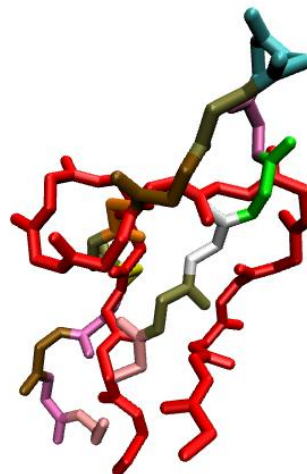
14:cPCA.rep.cluster 4 rmsd: 5,065 Å



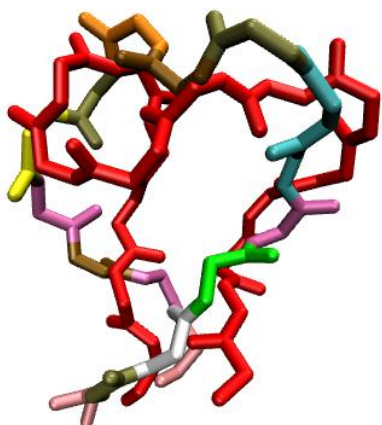
15:cPCA.rep.cluster 5 rmsd: 4,128 Å



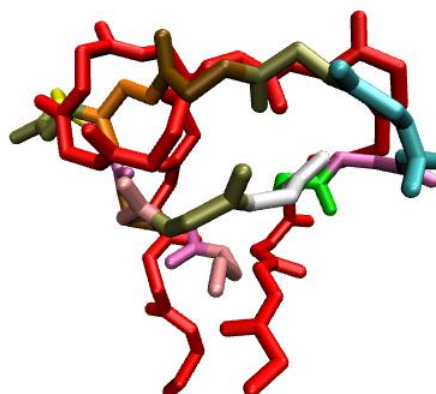
16:cPCA.rep.cluster 6 rmsd:4,687Å



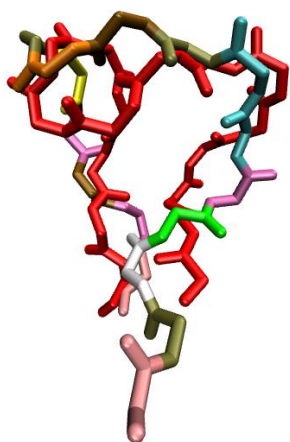
17:cPCA.rep.cluster 7 rmsd:5,232Å



18:cPCA.rep.cluster 8 rmsd:4,029Å

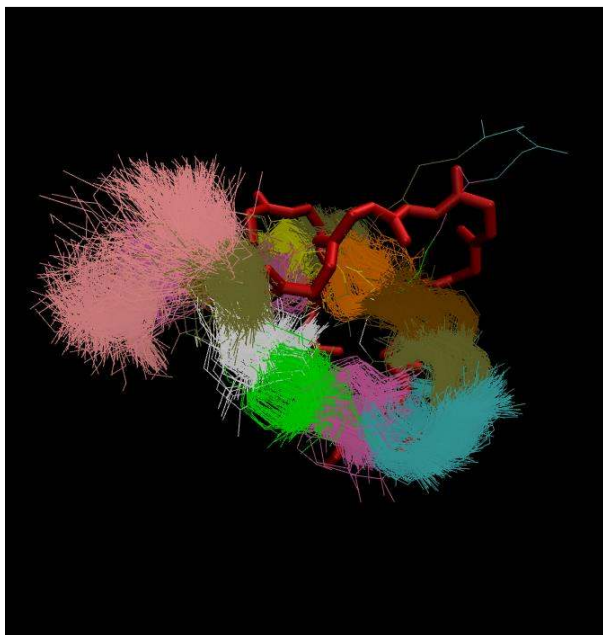


19:cPCA.rep.cluster 9 rmsd:4,377Å

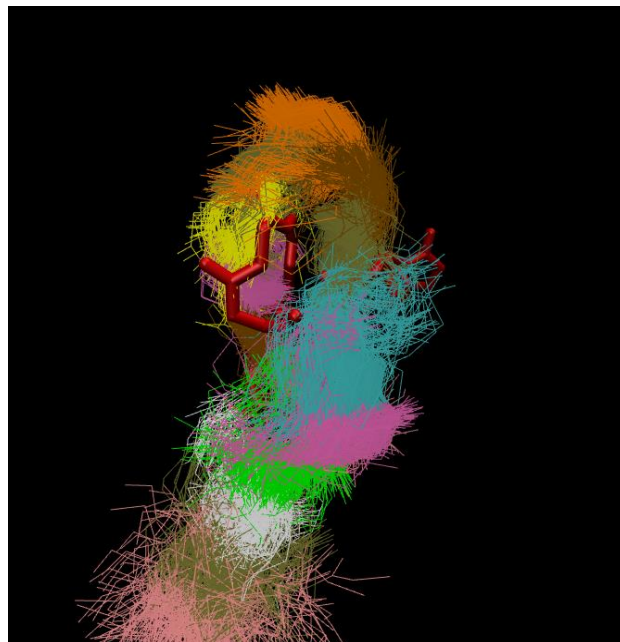


20:cPCA.rep.cluster 10 rmsd: 2,912Å

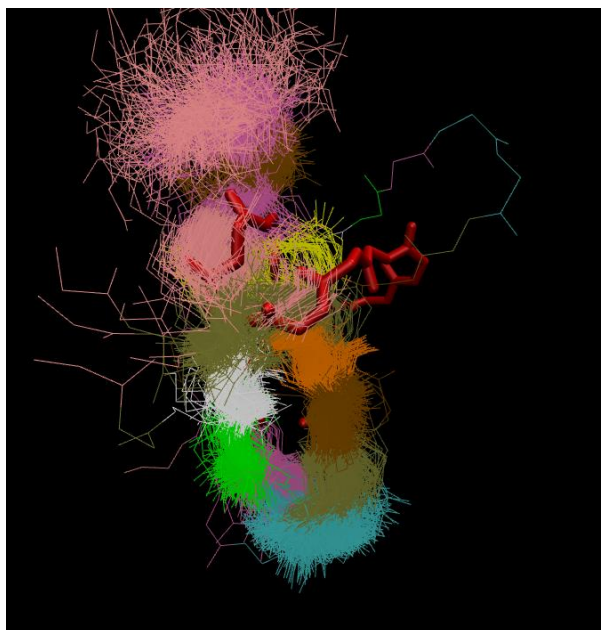
cPCA superimposed structures



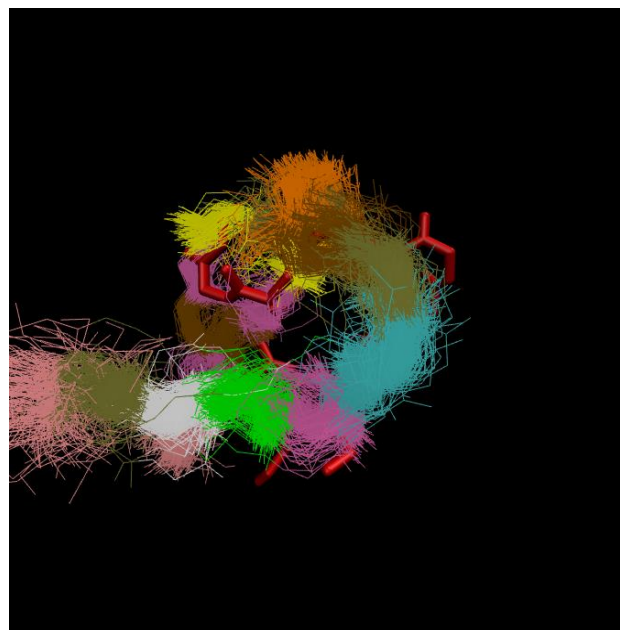
21:cPCA.superposition.cluster 3



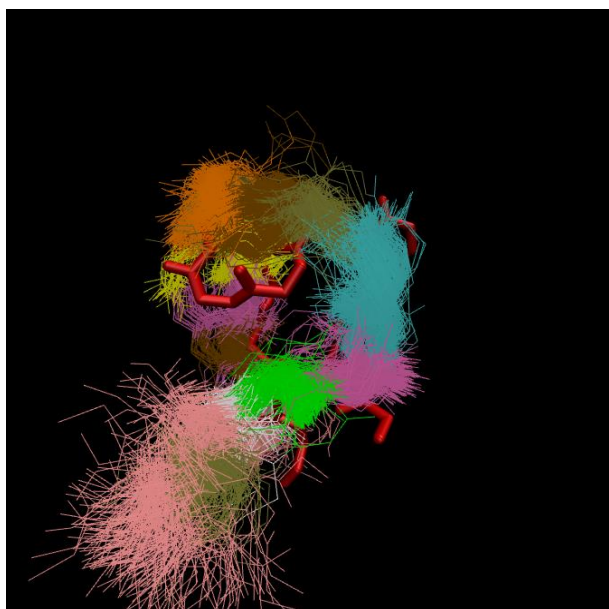
22:cPCA.superposition.cluster 4



23:cPCA.superposition.cluster 7

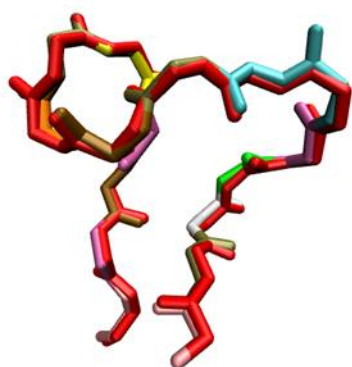


24:cPCA.superposition.cluster 8

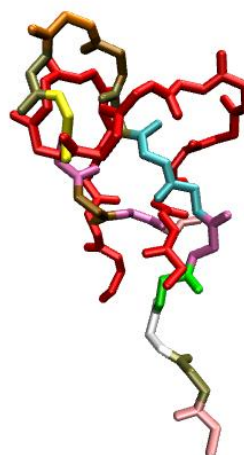


25:cpca.superposition.cluster 10

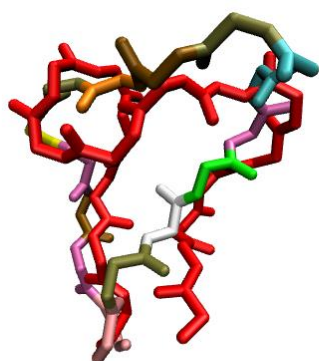
dPCA representatives



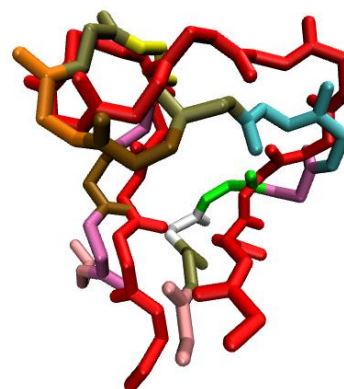
26:dpca.rep.cluster 1 rmsd: 0,333 Å



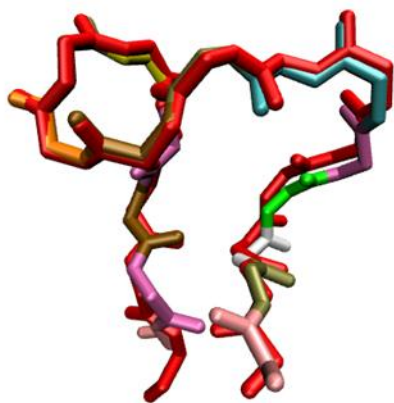
27:dpca.rep.cluster 2 rmsd: 5,095Å



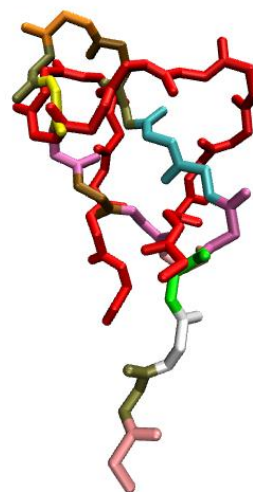
28:dpca.rep.cluster 3 rmsd:3,932Å



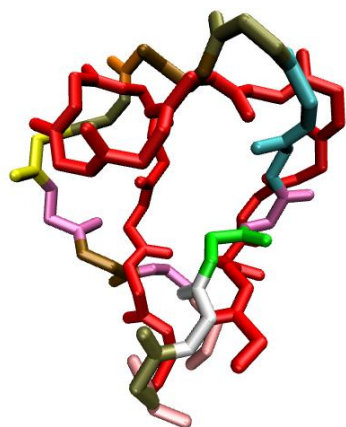
29:dpca.rep.cluster 4 rmsd:4,063Å



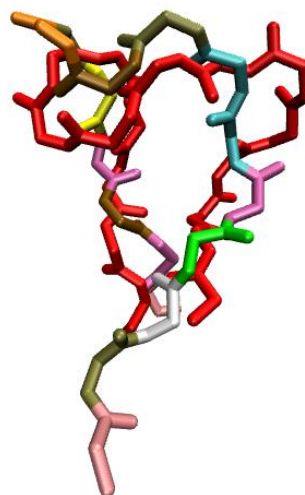
30: *d pca.rep.cluster 5 rmsd:0,972Å*



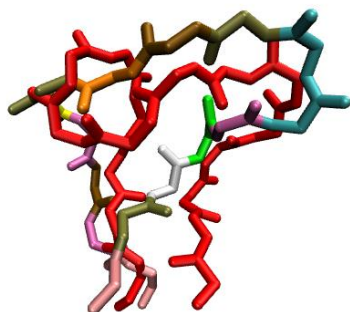
31: *d pca.rep.cluster 6 rmsd:5,012Å*



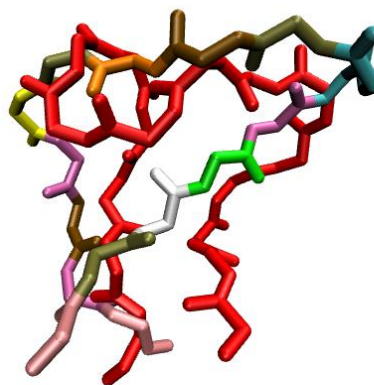
32: *d pca.rep.cluster 7 rmsd:4,293Å*



33: *d pca.rep.cluster 8 rmsd:2,952Å*

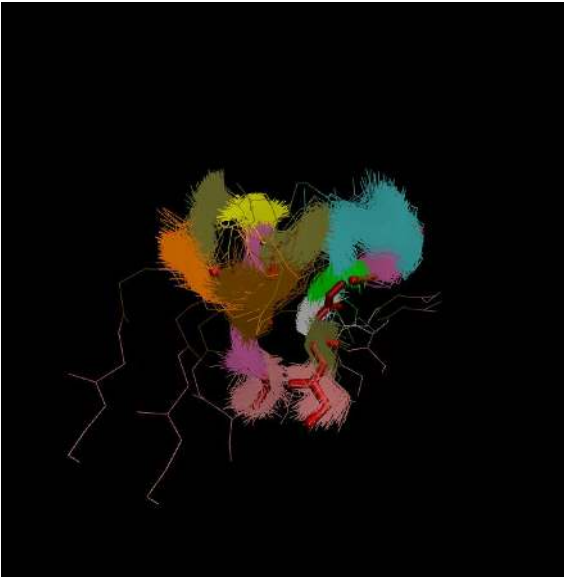


34: *d pca.rep.cluster 9 rmsd:3,750Å*

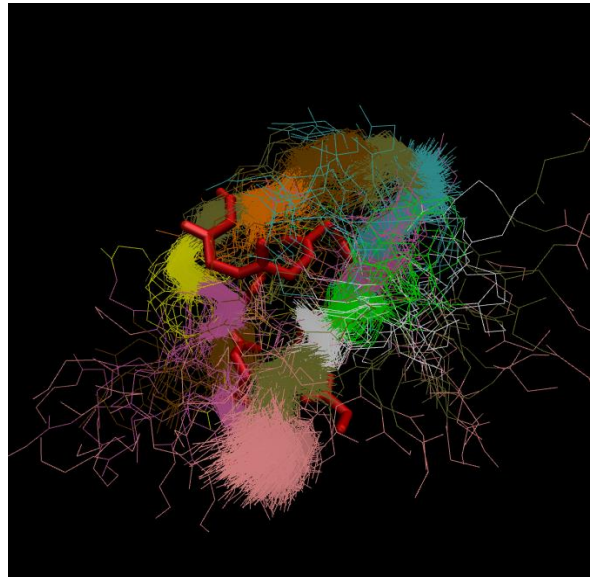


35: *d pca.rep.cluster 10 rmsd:3,756Å*

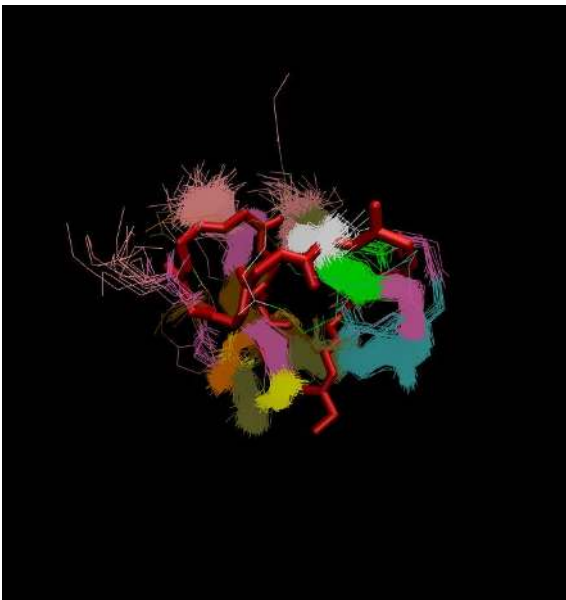
dPCA superpositions



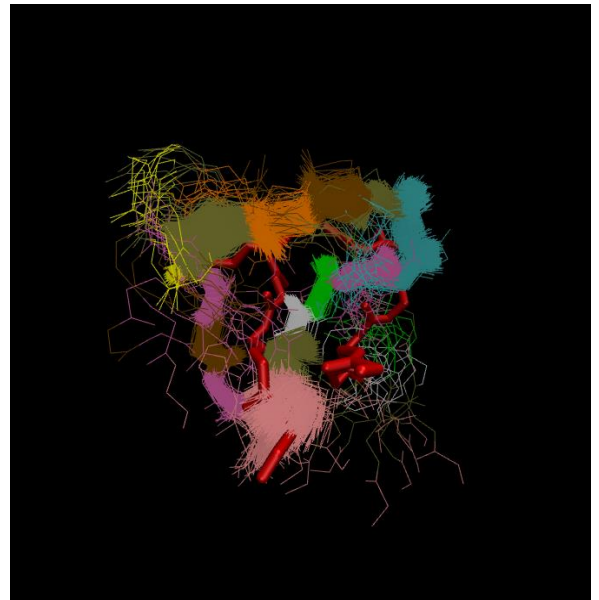
36:dPCA.superposition.cluster 1



37:dPCA.superposition.cluster 3



38:dPCA.superposition.cluster 5



39:dPCA.superposition.cluster 9

Συνολικά, μπορούμε να διακρίνουμε, ότι η ανάλυση dPCA (δεδομένου του πιο περιορισμένου διαμορφωτικού χώρου, που καλείται να εξερευνηθεί σε σχέση με την cPCA) παράγει συνεκτικότερα και συνεπέστερα αποτελέσματα που πλησιάζουν αισθητά περισσότερο το πειραματικά προσδιορισμένο native state. Χαρακτηριστικά να αναφέρουμε τις συγκρίσεις με τις αντιπροσωπευτικές δομές των dPCA clusters 1 και 5 (εικόνες 26 και 30) με rmsd ως προς το backbone $< 1 \text{ \AA}$. Συνοχή που παρατηρείται και στην απεικόνιση του dPCA cluster 1 σε υπέρθεση (εικόνα 36).

4. Συμπεράσματα

Κρίνοντας εκ των αποτελεσμάτων της παρούσας μελέτης, θεωρούμε ότι αναδείχθηκε επαρκώς η σημαντική δυναμική των υπολογιστικών μεθόδων για τον προσδιορισμό της φυσικής δομής ενός βιομορίου. Μέθοδοι που χρησιμοποιήσαμε όπως οι προσομοιώσεις μοριακής δυναμικής με στατιστική ανάλυση των αποτελεσμάτων (PCA), κατάφεραν με αξιοπιστία να εντοπίσουν την δομική συμπεριφορά ενός βιολογικά ενεργού πεπτιδίου όπως το EPI-X4. Οι συγκρίσεις των ανώτατων τιμών αποστάσεων ατόμων μέσω NOE φασματοσκοπίας (NOESY) με τις κατά μέσω όρο υπολογισμένες r^{-6} αποστάσεις μέσω του τροχιακού από την προσομοίωση MD έδειξε, όχι απόλυτο, παρ' όλα αυτά σημαντικό βαθμό συνοχής των πειραματικών και υπολογιστικών δομών. Πράγμα που υποστηρίζει την εγκυρότητα των παραμέτρων της προσομοίωσης.

Αμφότερες δε οι αναλύσεις PCA (dihedral, cartesian) παρείχαν συμπληρωματικά συμπεράσματα σε σχέση με τον διαμορφωτικό χώρο του συστήματος. Ήταν, παρ' όλα αυτά, εμφανής η αυξημένη συνέπεια των σχετικά πιο ρεαλιστικών αποτελεσμάτων που αποκτήθηκαν μέσω ανάλυσης dPCA. Η ανάλυση αυτή κατάφερε να εντοπίσει δομές με χαρακτηριστικά χαμηλά root mean square deviation (rmsd) σε σχέση με την πειραματικά προσδιορισμένη δομή (κάτω του 1 Å). Δημιουργήθηκαν επίσης απεικονίσεις υπερθέσεων ομαδοποιημένων δομών με εμφανή συνοχή.

Συνολικά, τα αποτελέσματα της μελέτης είναι ενδεικτικά, παρά τις προκλήσεις που ακόμα παραμένουν (έως ότου ξεπεραστούν), οι 'in silico' μέθοδοι προσφέρουν συνεχώς αυξανόμενη ακρίβεια και μπορούν πλέον να χαρακτηριστούν όχι απλώς ως συμπληρωματικές ως προς τις πειραματικές, αλλά πλησιάζουν δε, τον όρο 'ανταγωνιστικές'. Ενώ βεβαίως υπάρχει χώρος βελτίωσης στην πληρότητα των τεχνικών δειγματοληψίας και μοντέλων προσομοίωσης περιβαλλοντικών παραγόντων (force fields, water models, etc) η σημειούμενη πρόοδος είναι αναγνωρίσιμη και υποδηλώνει την μελλοντική ικανότητα της επιστημονικής έρευνας να στηριχθεί εξ ολοκλήρου σε υπολογιστικά μοντέλα για την μελέτη, εξερεύνηση και κατανόηση βιολογικά σημαντικών μοριακών συστημάτων, δεδομένου ότι και αυτά διατηρούν το πλεονέκτημα της ευκολίας και χαμηλότερου κόστους σε σχέση με τις πειραματικές μεθόδους.

Βιβλιογραφία

1. Monticelli, L., & Tieleman, D. P. (2013) 'Force fields for classical molecular dynamics', *Biomolecular Simulations: Methods and Protocols, Methods in Molecular Biology*, 924, pp. 197–243. DOI: 10.1007/978-1-62703-017-5_8.
2. Buchner, G. S., Murphy, R. D., Buchete, N.-V., & Kubelka, J. (2011) 'Dynamics of protein folding: probing the kinetic network of folding–unfolding transitions with experiment and theory', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(8), pp. 1001–1020. DOI: 10.1016/j.bbapap.2010.09.013.
3. Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G. (1997) 'Theory of protein folding: the energy landscape perspective', *Annual Review of Physical Chemistry*, 48, pp. 545–600. DOI: 10.1146/annurev.physchem.48.1.545.
4. Dinner, A. R., Šali, A., Smith, L. J., Dobson, C. M., & Karplus, M. (2000) 'Understanding protein folding via free-energy surfaces from theory and experiment', *Trends in Biochemical Sciences*, 25(7), pp. 331–339. DOI: 10.1016/S0968-0004(00)01610-8.
5. Karplus, M. (1997) 'The Levinthal paradox: yesterday and today', *Folding and Design*, 2, pp. S69–S75. DOI: 10.1016/S1359-0278(97)00067-9.
6. Zirafi, O., et al. (2015) 'Discovery and characterization of an endogenous CXCR4 antagonist', *Cell Reports*, 11(5), pp. 737–747. DOI: 10.1016/j.celrep.2015.03.061.
7. Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2011) 'How fast-folding proteins fold', *Science*, 334(6055), pp. 517–520. DOI: 10.1126/science.1208351.
8. Scheraga, H. A., Khalili, M., & Liwo, A. (2007) 'Protein-folding dynamics: overview of molecular simulation techniques', *Annual Review of Physical Chemistry*, 58, pp. 57–83. DOI: 10.1146/annurev.physchem.58.032806.104614.
9. Sagini, M. N., Zepp, M., Eyol, E., et al. (2024) 'EPI-X4, a CXCR4 antagonist inhibits tumor growth in pancreatic cancer and lymphoma models', *Peptides*, 175, 171111. DOI: 10.1016/j.peptides.2023.171111.
10. Buske, C., Kirchhoff, F., & Münch, J. (2015) 'EPI-X4, a novel endogenous antagonist of CXCR4', *Oncotarget*, 6(34), pp. 35137–35138. DOI:

10.18632/oncotarget.100710.

11. Georgoulia, P. S., & Glykos, N. M. (2019) ‘Molecular simulation of peptides coming of age: Accurate prediction of folding, dynamics and structures’, *Archives of Biochemistry and Biophysics*, 664, pp. 76–88. DOI: 10.1016/j.abb.2019.01.033.

12. VMD – Visual Molecular Dynamics. (2020) Available at: <https://www.ks.uiuc.edu/Research/vmd>.

13. Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G. (1997) ‘Theory of protein folding: the energy landscape perspective’, *Annual Review of Physical Chemistry*, 48, pp. 545–600. DOI: 10.1146/annurev.physchem.48.1.545.

14. Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology*, 4(1), 10–19. DOI: 10.1038/nsb0197-10.

15. Dominique Hagemans, Ianthe A. E. M. van Belzen, Tania Morán Luengo, and Stefan G. D. Rüdiger 2015 Oct 13 A script to highlight hydrophobicity and charge on protein surfaces Published online doi:10.3389/fmolb.2015.00056

16. Lindsay I. Smith February 26, 2002 A tutorial on Principal Components Analysis https://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf

17. Glykos, N. M. (2006). Software news and updates: Carma - A molecular dynamics analysis program. *Journal of Computational Chemistry*, 27(14), 1765-1768. Published online 17 August 2006 in Wiley InterScience. DOI: 10.1002/jcc.20482

18. Koukos, P. I. and Glykos, N. M. (2013). grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. *Journal of Computational Chemistry*, 34, 2310–2312. Published online 23 July 2013 in Wiley Online Library. DOI: 10.1002/jcc.23381

19. Thomas, J. (2011). YRB script for PyMOL: Visualizing hydrophobic and charged regions. PyMOLWiki. Available at: <https://pymolwiki.org/index.php/YRB> (Accessed: 5 May 2024).

20. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E., & Wüthrich, K. (1998). *Recommendations for the presentation of NMR structures of proteins and nucleic acids*. *Pure and Applied Chemistry*, 70(1), 117–142. <https://doi.org/10.1351/pac199870010117>

21. 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.
22. Marion, D. (2013) ‘An introduction to biological NMR spectroscopy’, *Molecular & Cellular Proteomics*, 12(11), pp. 3006–3025. DOI: 10.1074/mcp.O113.030239.
23. Koukos, P. I., & Glykos, N. M. (2014) ‘Folding Molecular Dynamics Simulations Accurately Predict the Effect of Mutations on the Stability and Structure of a Vammin-Derived Peptide’, *The Journal of Physical Chemistry B*, 118(37), pp. 10076–10084. DOI: 10.1021/jp5046113.
24. Stote, R. H. (n.d.) ‘Introduction to Molecular Dynamics Simulations’, *Institut de Chimie LC3-UMR 7177, Université Louis Pasteur, Strasbourg, France*. Available at: http://www.ch.embnet.org/MD_tutorial
25. Cuendet, M. (2008) ‘*Molecular Dynamics Simulation: A Short Introduction*’, EMBL. Available at: <https://www.embl.org>
26. Guvench, O., & MacKerell, A. D. (2008) ‘Comparison of protein force fields for molecular dynamics simulations’, *Molecular Modeling of Proteins*, *Methods in Molecular Biology*, 443, pp. 63–88. DOI: 10.1007/978-1-59745-177-2_4.
27. Schlick, T. (2010) *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 2nd edn. *Interdisciplinary Applied Mathematics, Volume 21*. New York: Springer. DOI: 10.1007/978-1-4419-6351-2