

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΧΗΜΙΚΩΝ ΜΗΧΑΝΙΚΩΝ  
ΤΟΜΕΑΣ (Ι) ΧΗΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ



---

## Υπολογισμοί διατάραξης ελεύθερης ενέργειας ως εργαλείο για τη βελτιστοποίηση ενώσεων-οδηγών

---

*Διπλωματική εργασία*

**Ντεκουμές Δημήτριος**

Επιβλέπουσα

Αναστασία Δέτση

Αναπληρώτρια Καθηγήτρια Ε.Μ.Π.

**ΑΘΗΝΑ 2019**



## Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στο Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών υπό την επίβλεψη της ερευνήτριας Γ' Δρ. Ζωής Κούρνια, κατά το ακαδημαϊκό έτος 2018-2019 και σηματοδοτεί το πέρας του πενταετούς κύκλου σπουδών μου στη Σχολή Χημικών Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θα ήθελα να ευχαριστήσω την Αναπληρώτρια Καθηγήτρια Ε.Μ.Π. Δρ. Αναστασία Δέτση, για την επίβλεψη της διπλωματικής εργασίας, την καθοδήγηση που μου προσέφερε σε όλο αυτό το διάστημα, καθώς και για τη δυνατότητα που μου έδωσε να πραγματοποιήσω την εργασία στο Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών σε συνεργασία με την Δρ. Ζωή Κούρνια.

Ταυτόχρονα οφείλω να ευχαριστήσω βαθύτατα τη Δρ. Ζωή Κούρνια για την εμπιστοσύνη που έδειξε στο πρόσωπό μου και για τη δυνατότητα που μου έδωσε να γνωρίσω τον τομέα της υπολογιστικής χημείας μέσα από την παρούσα διπλωματική εργασία. Θα ήθελα να εκφράσω την ευγνωμοσύνη μου για την πολύτιμη βοήθεια και τις συμβουλές της καθ' όλη τη διάρκεια της εκπόνησης της εργασίας.

Ιδιαίτερες ευχαριστίες θα ήθελα να αποδώσω στην τριμελή εξεταστική επιτροπή για τον χρόνο που διέθεσε για την κριτική ανάγνωση της εργασίας μου.

Στη συνέχεια, θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στον Dr. John Chodera και τον Andrea Rizzi, οι οποίοι μου έδωσαν την ευκαιρία να συμμετάσχω στον παγκόσμιο διαγωνισμό υπολογιστικού σχεδιασμού φαρμάκων SAMPL6 SAMPLing, καθώς και για όλη την καθοδήγηση και την πολύτιμη βοήθειά τους κατά την διάρκεια των προσομοιώσεων.

Επιπλέον, αισθάνομαι την ανάγκη να ευχαριστήσω τον Υποψήφιο Διδάκτορα Μιχάλη Παπαδουράκη, ο οποίος με εισήγαγε στους υπολογισμούς απόλυτης ελεύθερης ενέργειας, καθώς και με υποστήριξε καθ' όλη τη διάρκεια της πρόκλησης SAMPLing. Επίσης, ευχαριστώ θερμά τον Υποψήφιο Διδάκτορα Αλέξη Χατζηγούλα και τον μεταδιδακτορικό ερευνητή Στέφανο Ντερρ για την υποστήριξη που μου παρείχαν στα τεχνικά ζητήματα που προέκυπταν και με βοηθούσαν με την εμπειρία τους να ώστε να επιλυθούν. Επιπρόσθετα, θα ήθελα να ευχαριστήσω θερμά την Υποψήφια Διδάκτορα Δώρα Αργυροπούλου και τον μεταδιδακτορικό ερευνητή Χρήστο Λαμπράκη, οι οποίοι με συμβούλευαν στα καθημερινά προβλήματα που αντιμετώπιζα. Αισθάνομαι την ανάγκη να ευχαριστήσω την Αναστασία Θεοδωροπούλου για τη όλη την συμπαράσταση και την ανταλλαγή απόψεων, κατά τη διάρκεια εκπόνησης των διπλωματικών μας εργασιών στο εργαστήριο της Δρ. Ζωής Κούρνια.

Θα ήθελα να ευχαριστήσω ολόψυχα τους γονείς μου Μαρία και Γιάννη, καθώς και τον αδερφό μου Κωνσταντίνο, για την αγάπη και τη στήριξή τους καθ' όλη τη διάρκεια των σπουδών μου. Τέλος, δεν

θα μπορούσα να παραλείψω τους ανθρώπους στο κοντινό φιλικό μου περιβάλλον, για την υποστήριξη και την υπομονή τους όλα αυτά τα χρόνια.

Δημήτρης Ντεκουμές

Αθήνα, 14 Σεπτεμβρίου 2019

## ΠΕΡΙΛΗΨΗ

Μία από τις πιο απαιτητικές διαδικασίες της βελτιστοποίησης υποψηφίων φαρμάκων είναι να προβλεφθεί μεταξύ μιας σειράς ενώσεων-οδηγών ποιες από αυτές θα προσδεθούν ισχυρότερα στο φαρμακευτικό στόχο, ο οποίος είναι συνήθως μία πρωτεΐνη. Η πρόσδεση πραγματοποιείται με την εξισορρόπηση ευνοϊκών διαμοριακών αλληλεπιδράσεων και εντροπικών ποινών μεταξύ του υποψηφίου φαρμάκου (ένωση-οδηγός) και της πρωτεΐνης, οι οποίες συνεισφέρουν στην ελεύθερη ενέργεια πρόσδεσης. Η ελεύθερη ενέργεια πρόσδεσης καθορίζει το πόσο ισχυρά θα δεσμευθεί η ένωση-οδηγός στην συγκεκριμένη πρωτεΐνη. Μια από τις πιο ακριβείς προσεγγίσεις για τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης ή της σχετικής ελεύθερης ενέργειας πρόσδεσης μεταξύ δύο ενώσεων-οδηγών στην πρωτεΐνη είναι η Διατάραξη της Ελεύθερης Ενέργειας (FEP) εξαιτίας του υπόβαθρου στατιστικής μηχανικής της μεθόδου. Ωστόσο, το υψηλό υπολογιστικό κόστος, η περιορισμένη ακρίβεια των πεδίων δυνάμεων που χρησιμοποιούνται και οι τεχνικές προκλήσεις για την προετοιμασία, την εκτέλεση και την ανάλυση προσομοιώσεων ελεύθερης ενέργειας, έχουν επιβραδύνει την ενσωμάτωση των προσομοιώσεων Διατάραξης Ελεύθερης Ενέργειας στο σχεδιασμό νέων φαρμάκων. Προκειμένου να επιλυθούν αυτά τα ζητήματα, τις τελευταίες δεκαετίες η κοινότητα της υπολογιστικής χημείας κατέβαλε σημαντική προσπάθεια να επικυρώσει τη μέθοδο μέσω μελετών επικύρωσης μεγάλης κλίμακας, παγκόσμιων προκλήσεων και ανάπτυξης ποικίλων εφαρμογών για την αυτοματοποίηση των υπολογισμών.

Προς αυτή την κατεύθυνση, η παρούσα διπλωματική εργασία επικεντρώνεται σε δύο διαφορετικές περιπτώσεις, προκειμένου να συμβάλει στην προσπάθεια αυτή. Αρχικά, συμμετείχαμε στον παγκόσμιο διαγωνισμό SAMPL6 SAMPLing, ο οποίος εστιάζει στην εκτίμηση της σύγκλισης και αναπαραγωγιμότητας (μεταξύ διάφορων υπολογιστικών πακέτων) των προβλέψεων απόλυτης ελεύθερης ενέργειας πρόσδεσης. Για το σκοπό αυτό πραγματοποιήθηκαν υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης για τρία συστήματα προσδέτη-δέκτη χρησιμοποιώντας το λογισμικό NAMD, ώστε να εκτιμηθεί η σύγκλιση των υπολογισμών καθώς και να γίνει σύγκριση της ελεύθερης ενέργειας πρόσδεσης με πειραματικά δεδομένα και άλλα λογισμικά προσομοίωσης. Τα αποτελέσματά μας είναι σε συμφωνία με τους υπολογισμούς αναφοράς που πραγματοποιήθηκαν μέσω του λογισμικού YANK και άλλα λογισμικά προσομοίωσης όπως τα GROMACS, AMBER και SOMD, ωστόσο έχουν αποκλίσεις από τα πειραματικά αποτελέσματα.

Στη συνέχεια, μελετήθηκαν 15 ανάλογα μόρια της ένωσης CK-666, η οποία αποτελεί έναν υποψήφιο αναστολέα του πρωτεϊνικού συμπλόκου Arp2/3, και έγινε σύγκριση της προβλεπόμενης εκτιμήσεων σχετικής ελεύθερης ενέργειας πρόσδεσης για τις 15 ανάλογες ενώσεις μεταξύ δύο λογισμικών Μοριακής Δυναμικής (NAMD και GROMACS) καθώς και με τα αντίστοιχα πειραματικά δεδομένα. Τα αποτελέσματα των προσομοιώσεων φανερώνουν πως η εκτιμώμενη ελεύθερη ενέργεια πρόσδεσης σε σχέση με τα πειραματικά αποτελέσματα βρίσκεται εντός του σφάλματος της μεθόδου (1 – 2 kcal/mol). Επιπλέον, οι προσομοιώσεις με το πακέτο GROMACS σε συνδυασμό με το πεδίο δυνάμεων GAFF ανέδειξαν βελτιωμένη συσχέτιση από τις αντίστοιχες προσομοιώσεις του NAMD με το πεδίο OPLS.

Οι συνεχείς βελτιώσεις των τεχνικών του υπολογιστικού σχεδιασμού φαρμάκων θα επιτρέψουν την ενσωμάτωση των μοριακών προσομοιώσεων στη φαρμακοβιομηχανία, καθιστώντας έτσι το σχεδιασμό καινοτόμων φαρμακευτικών μορίων οικονομικά και χρονικά αποδοτικότερο.

**Λέξεις κλειδιά:** Μοριακή Δυναμική, Διατάραξη Ελεύθερης Ενέργειας, πρόκληση SAMPL6 SAMPLing, αναστολέας CK-666



## ABSTRACT

One of the most demanding tasks in the lead optimization phase of the drug design process is to predict among a series of lead candidates, which ones bind more strongly to the therapeutic target, which is most commonly a protein. This binding process occurs with a balance between the formation of favourable intermolecular interactions and entropic penalties between the lead compound and the protein, which contribute to the free energy of binding. The magnitude of the free energy of binding determines the affinity of the drug to the specific protein. One of the most accurate approaches to calculate the absolute free energy or the relative free energy of binding between two lead candidates is the alchemical Free Energy Perturbation (FEP) framework due to its rigorous statistical mechanics background. However, the high computational costs, the limited force field accuracy, and the technical challenges to setup, run and analyse free energy simulations, have hampered the success and throughput of free energy simulations in drug discovery. In order to tackle these issues, during the past few decades the computational chemistry community has made a significant effort to validate the method through validation studies, blind challenge predictions, and prospective applications.

In this direction, this thesis is focused on two different cases in order to contribute to this endeavour. First, we participate in the SAMPL6 SAMPLing challenge, which is focused on evaluating the convergence and reproducibility (across codes) of absolute free energy of binding predictions. To this end, we run absolute free energy calculations for three host – guest systems using the Molecular Dynamics (MD) software NAMD to assess the convergence of the calculation, and to also compare the binding affinity estimates against experimental data and other simulation platforms. Our results show that our FEP predictions are in agreement with reference calculations performed by YANK and other simulation platforms such as GROMACS, AMBER and SOMD, but show some deviation from the experimental results.

Second, we study a series of 15 analogs of CK-666, a micromolar Arp2/3 inhibitor, and compare the difference in the free energy of binding estimates for the analog series between two popular MD packages, NAMD and GROMACS, and against experimental data. The simulations reveal that the predicted binding affinities vs experimental results, for both packages, lie within the error of the method (1 – 2 kcal/mol) that other studies suggest. Moreover, GROMACS simulations using the GAFF force field reveal improved correlation than NAMD simulations using the OPLS force field.

Continuing improvements of computer-aided drug design techniques will enable molecular simulations to become an integral part of the pharmaceutical workflow, rendering the drug design process more cost- and time- efficient.

**Keywords:** Molecular Dynamics, Free Energy Perturbation, SAMPL6 SAMPLing challenge, CK-666 inhibitor





## Συντομογραφίες

**ADME** Absorption, Distribution, Metabolism, Excretion

**Ala** Alanine

**Arg** Arginine

**ARIS** Advanced research information system

**Arp2/3** Actin related protein 2/3

**Asp** Aspartic acid

**ATP** Adenosine triphosphate

**BAR** Bennet's acceptance ratio

**CADD** Computer Aided Drug Design

**CB8** Cucurbit[8]uril

**CB8-G3** quinine guest bound to cucurbit[8]uril host

**Cryo-EM** Cryogenic electron microscopy

**Cys** Cysteine

**DFT** Density function theory

**ECM** Extracellular matrix

**FEP** Free Energy Perturbation

**FP** Fluorescence polarization

**GAFF** General amber force field

**Glu** Glutamic acid

**Gly** Glycine

**GPCRs** G protein coupled receptors

**HPC** High performance computing

**HTS** High throughput screening

**Ile** Isoleucine

**ITC** Isothermal titration calorimetry

**Leu** Leucine

**LIE** Linear Interaction Energy

**Lys** Lysine

**MCSS** Maximum common substructure search

**MM/PBSA** Molecular Mechanics Generalized Born Surface Area  
**MM/PBSA** Molecular Mechanics Poisson-Boltzmann Surface Area  
**NMR** Nuclear Magnetic Resonance  
**NNRTIs** Non-nucleoside reverse transcriptase inhibitors  
**NPFs** Nucleation promoting factors  
**NPT** Ισόθερμη – ισοβαρής κατανομή  
**NVE** Μικρο-κανονική κατανομή  
**NVT** Κανονική κατανομή  
**OA** Octa-acid  
**OA-G3** 5-hexenoic acid guest bound to octa-acid host  
**OA-G6** 4-methylpentenoic acid guest bound to octa-acid host  
**OPLS-AA** All atoms optimized potentials for liquid simulations  
**PDB** Protein Data Bank  
**PES** Potential energy surface  
**Phe** Phenylalanine  
**PME** Particle Mesh Ewald  
**SAMPL** Statistical assessment of the modeling of proteins and ligands  
**SCAR** Suppressor of cyclic AMP repressor  
**Ser** Serine  
**SPR** Surface plasmon resonance  
**TEMOA** Tetra-endo-methyl-octa-acid  
**Thr** Threonine  
**TIP3P** Transferable intermolecular potential with 3 points  
**Trp** Tryptophan  
**Tyr** Tyrosine  
**Val** Valine  
**vdW** Van der Waals  
**VMD** Visual molecular dynamics  
**WASP** Wiskott-Aldrich syndrome protein

# Περιεχόμενα

|   |      |
|---|------|
| Ευχαριστίες .....   | ii   |
| ΠΕΡΙΛΗΨΗ.....   | iv   |
| ABSTRACT.....   | vi   |
| Συντομογραφίες.....   | viii |
| 1. ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ .....  | 1    |
| 1.1 Η σύγχρονη διαδικασία ανάπτυξης φαρμακευτικών ενώσεων .....                               | 1    |
| 1.1.1 Φάσεις ανάπτυξης νέων φαρμάκων .....  | 1    |
| 1.1.2 Σχεδιασμός φαρμάκων μέσω υπολογιστή .....   | 2    |
| 1.1.3 Βελτιστοποίηση ενώσεων-οδηγών μέσω υπολογιστή .....                                     | 3    |
| 1.2 Ελεύθερη ενέργεια πρόσδεσης .....   | 4    |
| 1.2.1 Ελεύθερη ενέργεια Gibbs .....   | 4    |
| 1.2.2 Εκτίμηση της ελεύθερης ενέργειας πρόσδεσης μέσω υπολογιστικών μεθόδων .....             | 7    |
| 1.2.3 Η μέθοδος διατάραξης της ελεύθερης ενέργειας για τη βελτιστοποίηση ενώσεων οδηγών ..... | 11   |
| 1.2.4 Πειραματικές τεχνικές για τον προσδιορισμό ελεύθερης ενέργειας πρόσδεσης .....          | 14   |
| 1.3 Διαμοριακές αλληλεπιδράσεις υποδοχέα-προσδέτη .....                                       | 18   |
| 1.3.1 Ιοντικές αλληλεπιδράσεις.....   | 18   |
| 1.3.2 Δεσμοί υδρογόνου .....  | 19   |
| 1.3.3 Δεσμοί Αλογόνου .....   | 20   |
| 1.3.4 van der Waals αλληλεπιδράσεις.....  | 20   |
| 1.3.5 Υδρόφοβες αλληλεπιδράσεις .....   | 21   |
| 1.3.6 π-αλληλεπιδράσεις .....   | 22   |
| 1.4 Οι προκλήσεις για τη στατιστική αξιολόγηση της ελεύθερης ενέργειας (SAMPL challenges) ..  | 23   |
| 1.4.1 Επισκόπηση των προκλήσεων SAMPL.....  | 23   |
| 1.4.2 Ο διαγωνισμός SAMPL6 SAMPLing.....  | 25   |
| 1.5 Η περίπτωση του πρωτεϊνικού συμπλέγματος Agr2/3 .....                                     | 26   |
| 1.5.1 Δομή.....   | 26   |
| 1.5.2 Βιολογική Λειτουργία .....  | 27   |
| 1.5.3 Συσχέτιση του συμπλέγματος Agr2/3 με τη μετάσταση του καρκίνου .....                    | 28   |
| 1.5.4 Αναστολή του πρωτεϊνικού συμπλέγματος Agr2/3 .....                                      | 30   |
| 1.5.5 Μηχανισμός δράσης υπαρχόντων αναστολέων .....   | 31   |
| 1.6 Σκοπός της εργασίας.....  | 32   |
| 2. ΜΕΘΟΔΟΛΟΓΙΑ .....  | 35   |
| 2.1 Προσομοίωση μοριακού συστήματος.....  | 35   |
| 2.2 Βασικές αρχές προσομοιώσεων Μοριακής Δυναμικής.....                                       | 38   |

|   |     |
|---|-----|
| 2.3 Προετοιμασία του υποδοχέα για προσομοίωση .....   | 42  |
| 2.4 Προετοιμασία του προσδέτη για προσομοίωση .....   | 43  |
| 2.5 Υπολογισμός απόλυτης ελεύθερης ενέργειας πρόσδεσης με τη μέθοδο της Διατάραξης<br>Ελεύθερης Ενέργειας ..... | 43  |
| 2.6 Υπολογισμός σχετικής ελεύθερης ενέργειας πρόσδεσης με τη μέθοδο της Διατάραξης<br>Ελεύθερης Ενέργειας ..... | 45  |
| 2.7 Σύγκλιση υπολογισμών Διατάραξης Ελεύθερης Ενέργειας .....   | 46  |
| 2.8 Υπερυπολογιστικά συστήματα .....  | 48  |
| 3. ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ.....   | 51  |
| 3.1 Υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης για το διαγωνισμό SAMPL6<br>SAMPLing .....               | 51  |
| 3.1.1 Τα συστήματα υποδοχέα-μορίου προσδέτη .....   | 51  |
| 3.1.2 Προετοιμασία των συστημάτων για προσομοίωση .....   | 53  |
| 3.1.3 Προσομοιώσεις διατάραξης ελεύθερης ενέργειας πρόσδεσης με το υπολογιστικό πακέτο<br>NAMD.....             | 54  |
| 3.1.4 Υπολογισμός απόλυτης ελεύθερης ενέργειας πρόσδεσης .....  | 56  |
| 3.2 Υπολογισμοί σχετικής ελεύθερης ενέργειας πρόσδεσης για τον αναστολέα της Arp2/3 .....                       | 57  |
| 3.2.1 Προετοιμασία της δομής της Arp2/3 για προσομοίωση.....  | 57  |
| 3.2.2 Προετοιμασία της δομής των ανάλογων μορίων της ένωσης οδηγού για προσομοίωση .                            | 58  |
| 3.2.3 Προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο NAMD.....                                      | 59  |
| 3.2.4 Προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο GROMACS .....                                  | 65  |
| 3.3 Δοκιμασίες κλιμάκωσης στο υπερυπολογιστικό σύστημα ARIS .....   | 68  |
| 3.3.1 NAMD.....   | 68  |
| 3.3.2 GROMACS.....  | 71  |
| 4. Αποτελέσματα – Συζήτηση Αποτελεσμάτων .....  | 73  |
| 4.1 Αποτελέσματα μοριακών προσομοιώσεων για την πρόκληση SAMPL6 SAMPLing .....                                  | 73  |
| 4.1.1 Πρόβλεψη απόλυτης ελεύθερης ενέργειας πρόσδεσης του συστήματος CB8-G3.....                                | 73  |
| 4.1.2 Μελέτη σύγκλισης απόλυτης ελεύθερης ενέργειας πρόσδεσης των συστημάτων OA-<br>G3/G6.....                  | 78  |
| 4.2 Αποτελέσματα μοριακών προσομοιώσεων για τις ανάλογες ενώσεις του CK-666 .....                               | 85  |
| 4.2.1 Εκτίμηση σχετικής ελεύθερης ενέργειας πρόσδεσης με NAMD και σύγκριση με<br>πειραματικά δεδομένα .....     | 85  |
| 4.2.2 Εκτίμηση σχετικής ελεύθερης ενέργειας πρόσδεσης με GROMACS και σύγκριση με<br>πειραματικά δεδομένα .....  | 99  |
| 5. Συμπεράσματα.....  | 117 |
| 6. Προτάσεις για μελλοντική έρευνα.....   | 119 |
| Ευρετήριο σχημάτων .....  | 120 |

|                         |     |
|-------------------------|-----|
| Ευρετήριο πινάκων ..... | 122 |
| Βιβλιογραφία .....      | 123 |



# 1. ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

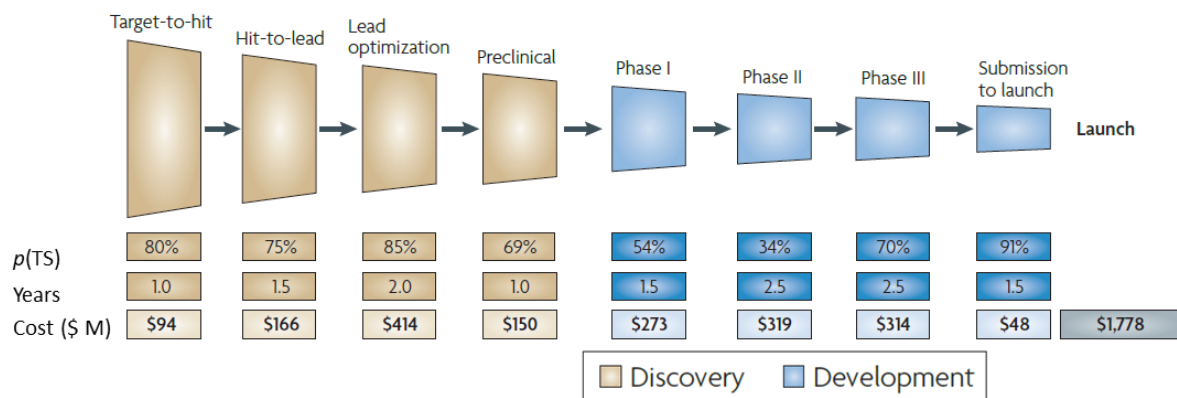
## 1.1 Η σύγχρονη διαδικασία ανάπτυξης φαρμακευτικών ενώσεων

### 1.1.1 Φάσεις ανάπτυξης νέων φαρμάκων

Ο σχεδιασμός φαρμάκων είναι η δημιουργία μικρών συνθετικών μορίων ή αντισωμάτων για την πρόληψη και θεραπεία ασθενειών. Τα φάρμακα δρουν πάνω σε βιομόρια του οργανισμού μας, όπως είναι οι πρωτεΐνες, τα οποία εμπλέκονται σε ασθένειες και ονομάζονται φαρμακευτικοί στόχοι. Η ανάπτυξη ενός νέου φαρμάκου είναι μία απαιτητική διαδικασία η οποία προϋποθέτει μεγάλη επένδυση σε χρόνο και πόρους, ενώ ταυτόχρονα καταγράφονται πολύ υψηλά ποσοστά αποτυχίας [1]. Για τον αποδοτικότερο σχεδιασμό των φαρμάκων, οι φαρμακοβιομηχανίες χρησιμοποιούν την μέθοδο High Throughput Screening (HTS), δηλαδή την πειραματική σάρωση μεγάλων χημειοθηκών ενάντια σε ένα φαρμακευτικό στόχο. Λόγω της ευρείας εφαρμογής αυτής της μεθόδου, η οποία ουσιαστικά βασίζεται στην τυχαιότητα, ο σχεδιασμός φαρμάκων παραμένει χρονοβόρος και δαπανηρός. Προκειμένου να επιταχυνθεί η διαδικασία με οικονομικά αποδοτικότερο τρόπο και να ελαχιστοποιηθεί το ποσοστό αποτυχίας, οι φαρμακοβιομηχανίες σήμερα έχουν στραφεί σε τεχνικές ορθολογικού σχεδιασμού φαρμάκων (rational drug design). Ο ορθολογικός σχεδιασμός φαρμάκων στηρίζεται στη χρήση της τρισδιάστατης δομής του φαρμακολογικού στόχου (structure-based drug design) ή στην χρήση δεδομένων γνωστών προσδετών στο φαρμακολογικό στόχο (ligand-based drug design) αν η δομή δεν είναι γνωστή. Από τη διαδικασία αυτή προκύπτουν ενώσεις-οδηγοί, δηλαδή χημικά μόρια που προσδένονται στο φαρμακολογικό στόχο και αποτελούν πρόδρομα μόρια των υποψηφίων φαρμάκων. Έτσι, ο ορθολογικός σχεδιασμός φαρμάκων οδήγησε σε επιτάχυνση του σχεδιασμού φαρμάκων με οικονομικά και χρονικά αποδοτικότερο τρόπο. Σήμερα η κατασκευή υπολογιστικών μοντέλων που προσομοιώνουν τις αλληλεπιδράσεις μεταξύ των φαρμακευτικών στόχων και των υποψηφίων φαρμάκων επιτρέπει την επιλογή κατάλληλων χημικών ενώσεων μέσα από χιλιάδες ενώσεις, τη βελτιστοποίηση της δομής τους καθώς και την κατασκευή νέων χημειότυπων [2].

Πρώτο βήμα στη διαδικασία σχεδιασμού φαρμάκων αποτελεί η επιλογή και η επικύρωση του φαρμακευτικού στόχου. Στη συνέχεια, πραγματοποιούνται υπολογιστικές προσομοιώσεις με σκοπό την εύρεση πρωτεϊνικών περιοχών στις οποίες μπορούν να προσδεθούν μόρια εάν το ενεργό κέντρο της πρωτεΐνης δεν είναι γνωστό. Ύστερα, σχεδιάζονται κατάλληλα χημικά μόρια τα οποία εκτιμάται πως θα προσδεθούν στις περιοχές που αναγνωρίστηκαν. Οι ενώσεις που παρουσιάζουν βιολογική δραστηριότητα έναντι του φαρμακευτικού στόχου (hit compounds), προχωρούν στην επόμενη φάση του σχεδιασμού φαρμάκων κατά την οποία σχεδιάζονται οι ενώσεις-οδηγοί (lead compounds). Ακολουθεί η βελτιστοποίηση των ενώσεων-οδηγών (lead optimization), που αποσκοπεί στο σχεδιασμό ανάλογων ενώσεων με βελτιωμένες φαρμακοκινητικές και τοξικολογικές, ιδιότητες και πιο ισχυρή πρόσδεση στην πρωτεΐνη. Επόμενο βήμα αποτελεί η οργανική σύνθεση των ανάλογων ενώσεων και η *in vitro* αξιολόγηση των μορίων που παρουσιάζουν τα επιθυμητά χαρακτηριστικά. Η αξιολόγηση αρχικά περιλαμβάνει δοκιμές με την ένωση-οδηγό και την πρωτεΐνη-φαρμακολογικό στόχο δίχως κύτταρα, ενώ στη συνέχεια οι ενώσεις-οδηγοί αξιολογούνται σε κυτταροκαλλιέργειες. Η φάση αυτή στοχεύει στον προσδιορισμό της σχέσης δομής-δράσης, δηλαδή στην αναγνώριση των χημικών ομάδων που ευθύνονται για τη βιολογική δραστηριότητά της ένωσης-οδηγού [3], [4].

Εφόσον οι δοκιμασίες αυτές είναι επιτυχείς, λαμβάνει χώρα το προκλινικό στάδιο (preclinical) κατά το οποίο διεξάγονται μελέτες του υποψήφιου φαρμάκου προκειμένου να καθοριστούν οι μέγιστες ασφαλείς συγκεντρώσεις στα ζώα, αλλά και να προσδιοριστεί η πιθανή τοξικότητα του υποψήφιου φαρμάκου. Συνήθως, οι προκλινικές μελέτες διαρκούν μικρό χρονικό διάστημα, ωστόσο παρέχουν λεπτομερείς πληροφορίες σχετικά με τις ιδιότητες Απορρόφησης, Κατανομής, Μεταβολισμού και Απέκκρισης (Absorption, Distribution, Metabolism, Excretion, ADME) του φαρμάκου, καθορίζοντας έτσι τα επίπεδα δοσολογίας και τοξικότητας [5]. Εάν το υποψήφιο φάρμακο κριθεί κατάλληλο για δοκιμή σε ανθρώπους, ακολουθεί το στάδιο των κλινικών μελετών. Η πρώτη φάση διαρκεί μερικούς μήνες και έχει κύριο στόχο τον καθορισμό της ασφάλειας και της επιτρεπόμενης δοσολογίας του φαρμάκου σε περίπου 20 έως 100 υγιείς εθελοντές. Η δεύτερη φάση αποσκοπεί στην εκτίμηση της αποτελεσματικότητας και των παρενεργειών του φαρμάκου και περιλαμβάνει δοκιμές σε εκατοντάδες ασθενείς (συνήθως 100-300) για χρονικό διάστημα μέχρι δύο χρόνια. Τέλος, η τρίτη φάση των κλινικών δοκιμών διαρκεί από 1 έως 4 έτη και είναι αρκετά πιο ευρεία καθώς περιλαμβάνει 300 έως 3000 ασθενείς. Ο βασικός στόχος της φάσης αυτής αφορά την εκτίμηση της αποτελεσματικότητας του φαρμάκου αλλά και την παρακολούθηση των ανεπιθύμητων παρενεργειών [6]. Όλες οι παραπάνω φάσεις συνοψίζονται στο Σχήμα 1.1 και όπως γίνεται αντιληπτό ο σχεδιασμός ενός καινοτόμου φαρμάκου απαιτεί περίπου δώδεκα χρόνια και κοστίζει κατά μέσο όρο 1.8 δισεκατομμύρια δολάρια ΗΠΑ.



**Σχήμα 1.1** Οι κυριότερες φάσεις για το σχεδιασμό φαρμακευτικών ενώσεων. Κάθε στάδιο καθορίζεται από την πιθανότητα τεχνικής επιτυχίας  $p(TS)$ , τη χρονική διάρκεια που απαιτεί (years) και το αντίστοιχο κόστος (cost). Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [7].

### 1.1.2 Σχεδιασμός φαρμάκων μέσω υπολογιστή

Ο σχεδιασμός ενός καινοτόμου μορίου με φαρμακολογική δράση αποσκοπεί στη ρύθμιση της λειτουργίας ενός φαρμακευτικού στόχου, ο οποίος είναι συνήθως μία πρωτεΐνη [8]. Τις τελευταίες δεκαετίες, έχει σημειωθεί σημαντική αύξηση στον αριθμό επικυρωμένων πρωτεϊνικών στόχων που σχετίζονται με κάποια ασθένεια [9] και επομένως έχουν καταβληθεί συστηματικές προσπάθειες για τη θεραπεία των ασθενειών που οφείλονται σε πρωτεΐνες. Επίσης, η πρόοδος στα πειράματα High Throughput Screening επέτρεψε την ταυτόχρονη αξιολόγηση χιλιάδων μορίων με τη χρήση ρομποτικού αυτοματισμού [10]. Ωστόσο, τα πειράματα αυτά εξακολουθούν να είναι χρονοβόρα και δαπανηρά, καθώς η αξία κτήσης και το λειτουργικό κόστος είναι απαγορευτικά για τα περισσότερα εργαστήρια. Προκειμένου να γίνει πιο αποδοτική η διαδικασία σχεδιασμού φαρμάκων, καθίσταται



αναγκαία η εφαρμογή τεχνικών ορθολογικού σχεδιασμού φαρμάκων με τη βοήθεια υπολογιστή (Computer Aided Drug Design, CADD) [1]. Οι τεχνικές αυτές αξιοποιούν μεθόδους Μοριακής Μοντελοποίησης με τις οποίες προσομοιώνονται οι βιολογικές συνθήκες στο εσωτερικό του οργανισμού.

Τα υποψήφια φάρμακα μπορούν να σχεδιαστούν αρχικά μέσω υπολογιστή και να εκτιμηθεί πόσο ισχυρά προσδένονται στο φαρμακευτικό στόχο καθώς επίσης και να προβλεφθούν οι ιδιότητες ADME. Υπάρχουν δύο ευρείες υπολογιστικές προσεγγίσεις για τον υπολογιστικό σχεδιασμό φαρμάκων, ο σχεδιασμός φαρμάκων με βάση την τρισδιάστατη δομή του φαρμακολογικού στόχου και ο σχεδιασμος με χρήση δεδομένων γνωστών προσδετών στο φαρμακολογικό στόχο. Η πρώτη προσέγγιση ακολουθείται όταν η τρισδιάστατη δομή του φαρμακευτικού στόχου είναι γνωστή από πειραματικές (κρυσταλλογραφία ακτίνων Χ, φασματοσκοπία NMR) ή υπολογιστικές τεχνικές (homology modeling, de novo modeling). Με αφετηρία την τρισδιάστατη δομή χρησιμοποιούνται επιπρόσθετες υπολογιστικές τεχνικές όπως η μοριακή πρόσδεση ή ο εκ νέου (de novo) σχεδιασμός προσδέτη, και σχεδιάζονται τα υποψήφια φάρμακα. Η εναλλακτική προσέγγιση ακολουθείται όταν δεν είναι διαθέσιμη η τρισδιάστατη δομή του φαρμακευτικού στόχου και προκύπτει από την ανάλυση δεδομένων δομής-δράσης μορίων που έχουν προηγουμένως ταυτοποιηθεί πως παρουσιάζουν δραστηριότητα έναντι του φαρμακευτικού στόχου. Στη συνέχεια με το υπολογιστικό φιλτράρισμα των μορίων δίνεται προτεραιότητα στη χημική σύνθεση αυτών που παρουσιάζουν τις επιθυμητές ιδιότητες, ενώ ταυτόχρονα απορρίπτονται τα μόρια που πιθανόν να οδηγήσουν σε αποτυχία στις μετέπειτα φάσεις του σχεδιασμού φαρμάκων [5].

### 1.1.3 Βελτιστοποίηση ενώσεων-οδηγών μέσω υπολογιστή

Η έκβαση του επιθυμητού θεραπευτικού αποτελέσματος εξαρτάται από το πόσο ισχυρά προσδένεται ένα χημικό μόριο στο φαρμακευτικό στόχο. Η πρόσδεση πραγματοποιείται με την εξισορρόπηση ευνοϊκών διαμοριακών αλληλεπιδράσεων και εντροπικών ποινών μεταξύ της ενώσεως-οδηγού και της πρωτεΐνης, οι οποίες συνεισφέρουν στην ελεύθερη ενέργεια πρόσδεσης. Η ελεύθερη ενέργεια πρόσδεσης καθορίζει το πόσο ισχυρά θα δεσμευθεί η ένωση-οδηγός στην συγκεκριμένη πρωτεΐνη. Συνεπώς, ο σχεδιασμός ενώσεων-οδηγών οι οποίες εμφανίζουν ισχυρή ελεύθερη ενέργεια πρόσδεσης στο φαρμακευτικό στόχο αποτελεί έναν από τους πρωταρχικούς στόχους του CADD [1], [11].

Υπάρχει πληθώρα υπολογιστικών μεθόδων για την εκτίμηση της ελεύθερης ενέργειας πρόσδεσης. Ορισμένες από τις προσεγγίσεις αυτές περιλαμβάνουν τη Μοριακή Πρόσδεση (Molecular Docking), τη Γραμμική Ενέργεια Αλληλεπίδρασης (LIE), τη Μοριακή Μηχανική με Poisson-Boltzmann/Generalized Born και Επιδιάλυση Επιφάνειας (MM/PBSA, MM/GBSA), και τη Διατάραξη Ελεύθερης Ενέργειας (FEP) [12]. Κάθε μια από τις μεθοδολογίες αυτές εμφανίζει διαφορές σχετικά με την ακρίβεια και την αποτελεσματικότητα, με τη μέθοδο FEP να αποτελεί μια από τις πιο έγκυρες προσεγγίσεις για τον υπολογισμό της ελεύθερης ενέργειας πρόσδεσης προσδέτη-πρωτεΐνης. Στα πλαίσια της βελτιστοποίησης ενώσεων-οδηγών, το θερμοδυναμικό μέγεθος το οποίο συνήθως εκτιμάται είναι η σχετική διαφορά στην ελεύθερη ενέργεια πρόσδεσης μεταξύ δύο ενώσεων, και όχι η απόλυτη ελεύθερη ενέργεια πρόσδεσης κάθε μορίου. Έτσι, η μέθοδος FEP εφαρμόζεται για να

προβλεφθεί το πόσο ισχυρά προσδένεται ένα μόριο σε σχέση με ένα άλλο σε μία συγκεκριμένη πρωτεΐνη. Ωστόσο, μέχρι και σήμερα δεν έχει αποδειχθεί ότι με τη μέθοδο FEP μπορούν να εξαχθούν αποτελέσματα με μεγάλη ακρίβεια για ένα ευρύ φάσμα προσδετών και πρωτεϊνικών στόχων, ούτως ώστε να ενσωματωθεί συστηματικά σε ερευνητικά προγράμματα της φαρμακοβιομηχανίας. Τις τελευταίες δεκαετίες, η κοινότητα της υπολογιστικής χημείας καταβάλλει πυρετώδεις προσπάθειες για την περαιτέρω βελτίωση της Διατάραξης Ελεύθερης Ενέργειας ώστε να μπορεί να παράγει αξιόπιστα αποτελέσματα σε ένα σύντομο χρονικό διάστημα [11].

## 1.2 Ελεύθερη ενέργεια πρόσδεσης

### 1.2.1 Ελεύθερη ενέργεια Gibbs

Στη θερμοδυναμική, η ελεύθερη ενέργεια είναι το μέγεθος που καθορίζει εάν μία διεργασία συμβαίνει αυθόρμητα ή όχι. Αυτό συμβαίνει επειδή χωρίς εξωτερική πηγή ενέργειας ένα σύστημα μεταβαίνει στην κατάσταση με τη χαμηλότερη ελεύθερη ενέργεια. Ως εκ τούτου, η ελεύθερη ενέργεια καθορίζει την αυθόρμητη (ή όχι) πρόσδεση ενός μορίου σε ένα μακρομόριο, την αναδίπλωση των πρωτεϊνών, καθώς και πολλές άλλες διεργασίες στο εσωτερικό των κυττάρων [13]. Η ελεύθερη ενέργεια μπορεί να περιγραφεί είτε κατά Helmholtz είτε κατά Gibbs. Η ελεύθερη ενέργεια κατά Helmholtz ορίζεται σύμφωνα την Εξίσωση 1.1:

$$A = U - TS \quad (1.1)$$

Όπου,  $U$  η εσωτερική ενέργεια,  $T$  η θερμοκρασία, και  $S$  η εντροπία του συστήματος.

Η ελεύθερη ενέργεια κατά Gibbs ορίζεται ως:

$$G = H - TS \quad (1.2)$$

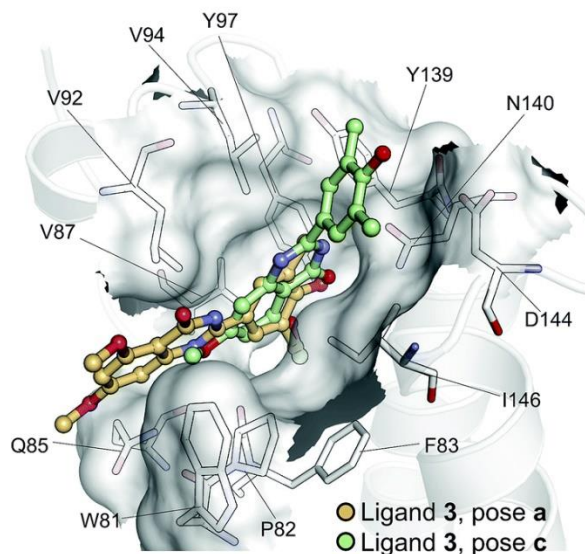
όπου,  $H$ , η ενθαλπία του συστήματος, η οποία ορίζεται ως  $H = U + pV$ ,  $p$ , η πίεση και  $V$ , ο όγκος του συστήματος.

Σημειώνεται πως και οι δύο ορισμοί περιέχουν έναν όρο που εξαρτάται από την εσωτερική ενέργεια του συστήματος και έναν άλλο που εξαρτάται από την εντροπία. Επομένως, η ελεύθερη ενέργεια επηρεάζεται τόσο από τις αλλαγές στις αλληλεπιδράσεις μεταξύ των ατόμων του συστήματος, όσο και από τις εντροπικές αλλαγές. Πιο συγκεκριμένα, αλλαγές στην εντροπία του διαλύτη μπορούν να παίξουν σημαντικό ρόλο σε περιπτώσεις, όπως το φαινόμενο της υδροφοβικότητας (βλ. Σχήμα 1.16) όπου τα μόρια του διαλύτη απελευθερώνονται κατά τον σχηματισμό μιας υδρόφοβης πρωτεΐνης [14]. Η διαφορά μεταξύ της ελεύθερης ενέργειας κατά Helmholtz και κατά Gibbs συνίσταται στον όρο  $pV$ , με την ελεύθερη ενέργεια Helmholtz να εφαρμόζεται υπό σταθερό όγκο. Η ελεύθερη ενέργεια Gibbs προϋποθέτει σταθερή τιμή πίεσης, ενώ ο όρος  $pV$  ποσοτικοποιεί το έργο που σχετίζεται με την αλλαγή του όγκου [13].

Σε ένα τυπικό εργαστηριακό πείραμα συνήθως επικρατούν ισοβαρείς συνθήκες και συνεπώς η βαθμίδα της ελεύθερης ενέργειας Gibbs λειτουργεί ως κινητήρια δύναμη για ένα σύστημα [14]. Επιπλέον, η ελεύθερη ενέργεια Gibbs μετρά την ικανότητα ενός θερμοδυναμικού συστήματος να εκτελέσει μέγιστο έργο σε συνθήκες σταθερής πίεσης και θερμοκρασίας. Αντίστοιχα με οποιαδήποτε αυθόρμητη διεργασία, η πρόσδεση ενός μορίου σε μία πρωτεΐνη πραγματοποιείται μόνο όταν η

αλλαγή στην ελεύθερη ενέργεια Gibbs ( $\Delta G$ ) του συστήματος είναι αρνητική όταν αυτό φτάνει σε ισορροπία υπό σταθερή θερμοκρασία και πίεση. Η ισχύς της πρόσδεσης προσδέτη-πρωτεΐνης καθορίζεται από το εύρος της αρνητικής  $\Delta G$ , δηλαδή όσο πιο αρνητική είναι η  $\Delta G$  τόσο πιο ισχυρή είναι η πρόσδεση. Συνεπώς, μπορεί να θεωρηθεί πως η  $\Delta G$  καθορίζει τη σταθερότητα ενός συμπλόκου πρωτεΐνης-προσδέτη ή εναλλακτικά δείχνει το πόσο ισχυρά δεσμεύεται ένας προσδέτης σε μία συγκεκριμένη πρωτεΐνη. Επιπρόσθετα, η ελεύθερη ενέργεια Gibbs είναι μία σημειακή συνάρτηση και ορίζεται αποκλειστικά από την αρχική και τελική κατάσταση του συστήματος, ανεξαρτήτως της διαδρομής που συνδέει τις δύο αυτές καταστάσεις [15]. Στις περισσότερες περιπτώσεις οι υπολογισμοί ελεύθερης ενέργειας εκτιμούν τη διαφορά ελεύθερης ενέργειας μεταξύ δύο καταστάσεων ενός μοριακού συστήματος μέσω υπολογισμού του αντιστρεπτού έργου που απαιτείται για τη μετάβαση από την αρχική στην τελική κατάσταση.

Η πλήρης κατανόηση των περισσότερων βιοχημικών διεργασιών απαιτεί διερεύνηση της συμπεριφοράς της ελεύθερης ενέργειας του εξεταζόμενου συστήματος. Για παράδειγμα μέσω υπολογισμών ελεύθερης ενέργειας μπορούν να προβλεφθούν οι πιθανές διαμορφώσεις που μπορεί να λάβει ένας προσδέτης ο οποίος δεσμεύεται σε μία πρωτεΐνη (Σχήμα 1.2). Υπάρχουν δύο ευρείες προσεγγίσεις που αφορούν την πρόσδεση φαρμακευτικών μορίων σε μία πρωτεΐνη, ο υπολογισμός της απόλυτης ελεύθερης ενέργειας πρόσδεσης και ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης [16].



**Σχήμα 1.2** Μέσω υπολογισμών ελεύθερης ενέργειας καθίσταται δυνατή η πρόβλεψη των πιθανών διαμορφώσεων (εδώ pose a και pose c) ενός προσδέτη σε μία πρωτεΐνη. Η εικόνα λήφθηκε από τη βιβλιογραφική πηγή [17].

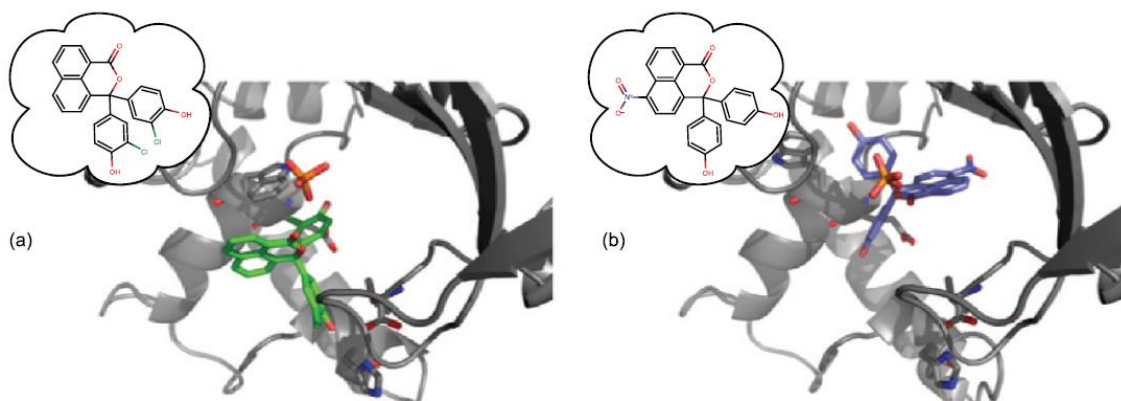
### 1.2.1.1. Απόλυτη ελεύθερη ενέργεια πρόσδεσης

Η πρώτη γενική προσέγγιση υπολογίζει απευθείας την ελεύθερη ενέργεια πρόσδεσης μέσω υπολογισμού του αντιστρεπτού έργου μεταφοράς του μορίου-προσδέτη από τη θέση πρόσδεσης (αρχική κατάσταση) στο διάλυμα (τελική κατάσταση) και ονομάζεται υπολογισμός απόλυτης ελεύθερης ενέργειας πρόσδεσης. Η διαδρομή που συνδέει την αρχική με την τελική κατάσταση μπορεί να είναι φυσικά πραγματοποιήσιμη (physical pathway) ή μπορεί να μελετηθεί μόνο *in silico*, οπότε ονομάζεται «αλχημική» διαδρομή (alchemical pathway). Οι μέθοδοι που ακολουθούν τη

φυσική διαδρομή παρέχουν την απόλυτη ελεύθερη ενέργεια πρόσδεσης με τον υπολογισμό του αναστρέψιμου έργου που απαιτείται για την απομάκρυνση του προσδέτη από τη θέση δέσμευσης. Αντίθετα, οι μέθοδοι που ακολουθούν την αλχημική διαδρομή εξαφανίζουν τεχνητά τον προσδέτη από τη θέση δέσμευσης και στη συνέχεια τον εμφανίζουν στο διάλυμα. Δεδομένου ότι η ελεύθερη ενέργεια είναι σημειακή συνάρτηση, οι προσεγγίσεις της αλχημικής και φυσικής οδού παρέχουν συγκρίσιμα αποτελέσματα όταν εφαρμόζονται για το ίδιο σύστημα [18]. Έχει αποδειχθεί πως οι μέθοδοι που ακολουθούν την αλχημική διαδρομή μπορούν να υπολογίσουν με ακρίβεια την ελεύθερη ενέργεια πρόσδεσης για μικρά μόρια. Παρ' όλα αυτά, σε περιπτώσεις όπου εξαφανίζεται ένα μεγαλύτερο μακρομόριο, προκύπτουν ζητήματα σύγκλισης των αλχημικών υπολογισμών. Στην περίπτωση αυτή, την καλύτερη εναλλακτική προσφέρουν οι μέθοδοι που βασίζονται στη φυσική διαδρομή σε συνδυασμό με προηγμένες τεχνικές δειγματοληψίας. Για τον ακριβή προσδιορισμό της ελεύθερης ενέργειας πρόσδεσης θα πρέπει να εφαρμοστούν κατάλληλα δυναμικά περιορισμού, τα οποία ωστόσο απαιτούν ιδιαίτερη εμπειρία και προσοχή για τη σωστή εφαρμογή τους [19].

### 1.2.1.2. Σχετική ελεύθερη ενέργεια πρόσδεσης

Η δεύτερη γενική προσέγγιση υπολογίζει τη σχετική διαφορά μεταξύ των ελεύθερων ενεργειών πρόσδεσης δύο διαφορετικών προσδετών για τον ίδιο υποδοχέα. Αυτό πραγματοποιείται υπολογίζοντας το έργο που απαιτείται για την τεχνητή μετατροπή ενός προσδέτη στον άλλο, αρχικά στην δεσμευμένη κατάσταση και στη συνέχεια ελεύθερο σε διάλυμα. Εφόσον οι μετατροπές αυτές δεν είναι υλοποιήσιμες από φυσική άποψη, αυτοί οι υπολογισμοί καλούνται και πάλι αλχημικοί [20]. Οι αλχημικοί υπολογισμοί σχετικής ελεύθερης ενέργειας παρέχουν αξιόπιστα αποτελέσματα σε σχέση με αντίστοιχα πειραματικά δεδομένα εάν οι δύο προσδέτες έχουν παρόμοια χημική δομή. Ωστόσο, εμφανίζονται σημαντικές αποκλίσεις από τα πειραματικά αποτελέσματα εάν οι δύο προσδέτες διαφέρουν πολύ από χημική σκοπιά ή εάν υπάρχει ένα υψηλό ενεργειακό φράγμα που διαχωρίζει τις πιο σταθερές διαμορφώσεις που λαμβάνουν τα δύο μόρια όταν είναι προσδεμένα στην ίδια πρωτεΐνη. Για παράδειγμα, ο Stout και οι συνεργάτες του [21] ανέπτυξαν μια σειρά αναστολέων θυμιδυλικής συνθετάσης με βάση τα παράγωγα φαινολοφθαλείνης και φθαλίνης. Διαπίστωσαν ότι η αλλαγή ενός πενταμελούς δακτυλίου σε ένα εξαμελή δακτύλιο μεταβάλλει δραματικά τον τρόπο δέσμευσης. Στη συνέχεια αντιλήφθηκαν πως η προσθήκη μιας νιτρο-ομάδας, με ταυτόχρονη απομάκρυνση δύο χλωρίων, στην προκύπτουσα ένωση εισήγαγε έναν ακόμη μη αναμενόμενο τρόπο δέσμευσης (Σχήμα 1.3) [22].



**Σχήμα 1.3** Οι υπολογισμοί σχετικής ελεύθερης ενέργειας υποδεικνύουν αλλαγές στον τρόπο πρόσδεσης παρόμοιων αναστολέων. **(a)** Αρχική διαμόρφωση του αναστολέα κατά τη δέσμευσή του στην θυμιδυλική

συνθετάση. **(b)** Τροποποιημένος τρόπος πρόσδεσης του παράγωγου αναστολέα στην πρωτεΐνη στόχο με την προσθήκη μιας νιτρο-ομάδας και διαγραφή των δύο χλωρίων. Οι κωδικοί PDB είναι 1TSL και 1TSM91, αντίστοιχα. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [22].

Παρά τα ζητήματα αυτά, οι υπολογισμοί αληθινής σχετικής ελεύθερης ενέργειας αποτελούν σήμερα μια από τις καλύτερα αυτοματοποιημένες και συχνά χρησιμοποιούμενες μεθοδολογίες εκτίμησης ελεύθερης ενέργειας [12]. Ένα ιδιαίτερα ενδιαφέρον και ολοκληρωμένο σύνολο μελετών αποτελεί το έργο από το εργαστήριο του W.L Jorgensen για τη δέσμευση των μη νουκλεοσιδικών αναστολέων της αντίστροφης μεταγραφάσης (NNRTIs) του HIV-1 [23], [24]. Τα αποτελέσματα των αληθινών υπολογισμών σχετικής ελεύθερης ενέργειας πρόσδεσης ήταν σε συμφωνία (σε εύρος 1-2 kcal/mol) με τα πειραματικά αποτελέσματα, ενώ στη συνέχεια τα ευρήματα των μελετών επιβεβαιώθηκαν και από κρυσταλλογραφικά πειράματα [23].

## 1.2.2 Εκτίμηση της ελεύθερης ενέργειας πρόσδεσης μέσω υπολογιστικών μεθόδων

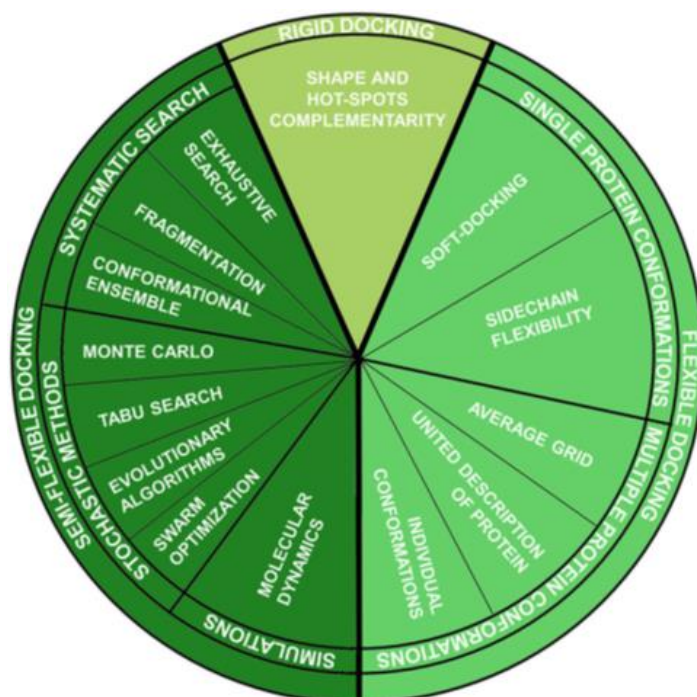
Στη συνέχεια συνοψίζονται μερικές από τις συνηθέστερες υπολογιστικές προσεγγίσεις που χρησιμοποιούνται για την εκτίμηση της ελεύθερης ενέργειας πρόσδεσης προσδέτη-πρωτεΐνης [12].

### **Μοριακή Πρόσδεση**

Οι τεχνικές Μοριακής Πρόσδεσης (Molecular Docking) στοχεύουν στην πρόβλεψη του πιο πιθανού τρόπου δέσμευσης ενός προσδέτη σε μία πρωτεΐνη. Με βάση την τρισδιάστατη δομή του φαρμακευτικού στόχου, η οποία μπορεί να έχει επιλυθεί πειραματικά (π.χ. κρυσταλλογραφία ακτίνων Χ) ή υπολογιστικά (π.χ. μοντελοποίηση ομολογίας) [25], παράγεται ένας αριθμός πιθανών διαμορφώσεων που μπορεί να λάβει ο προσδέτης όταν βρίσκεται στη θέση δέσμευσης της πρωτεΐνης. Η Μοριακή Πρόσδεση αποτελείται κυρίως από δύο μέρη: ένα λογισμικό το οποίο αναζητά τις πιθανές διαμορφώσεις που μπορεί να λάβει ο προσδέτης με βάση το διαθέσιμο χημικό χώρο (δειγματοληψία) και μία συνάρτηση βαθμολόγησης (scoring function), η οποία κατατάσσει τις πιθανές διαμορφώσεις που έχουν βρεθεί [26]. Οι συναρτήσεις βαθμολόγησης ουσιαστικά χρησιμεύουν για τη διάκριση των ευνοϊκών τρόπων σύνδεσης των μορίων στον υποδοχέα. Κατά κύριο λόγο οι συναρτήσεις βαθμολόγησης προκύπτουν από την Κλασική Μηχανική, από διάφορες εμπειρικές συναρτήσεις, καθώς και από συναρτήσεις που βασίζονται στη συχνότητα επαφής προσδέτη-πρωτεΐνης.

Ο πρώτος αλγόριθμος μοριακής πρόσδεσης αναπτύχθηκε στη δεκαετία του 1980 από τον Kuntz και τους συνεργάτες του [27], όπου ο υποδοχέας και ο προσδέτης προσεγγίστηκαν από μία σειρά σφαιρών. Αυτή η μέθοδος ανήκει στις πλήρως άκαμπτες τεχνικές Μοριακής Πρόσδεσης (fully-rigid docking) κατά τις οποίες, τόσο ο προσδέτης όσο και πρωτεΐνη θεωρούνται άκαμπτες οντότητες και κατά τη δειγματοληψία του χημικού χώρου εξετάζονται μόνο οι τρεις μεταφορικοί και τρεις περιστροφικοί βαθμοί ελευθερίας. Εναλλακτικά, υπάρχει και η λεγόμενη μερικώς εύκαμπτη πρόσδεση (semi-flexible docking) όπου μόνο ο προσδέτης είναι εύκαμπτος, ενώ η πρωτεΐνη είναι άκαμπτη. Στην περίπτωση αυτή, εκτός από τους έξι μεταφορικούς και περιστροφικούς βαθμούς ελευθερίας, λαμβάνονται υπόψιν και οι βαθμοί ελευθερίας που χαρακτηρίζουν τις διαμορφώσεις του προσδέτη. Τέλος, η προσέγγιση της εύκαμπτης πρόσδεσης (flexible docking) βασίζεται στην ιδέα ότι μια πρωτεΐνη δεν είναι μια παθητική άκαμπτη οντότητα κατά τη διάρκεια της πρόσδεσης και

θεωρείται πως ο προσδέτης και η πρωτεΐνη είναι ευέλικτα μόρια [28]. Στο Σχήμα 1.4 συνοψίζονται οι μέθοδοι Μοριακής Πρόσδεσης με βάση τις παραπάνω θεωρήσεις και τους αλγορίθμους δειγματοληψίας του χημικού χώρου.



**Σχήμα 1.4** Μοριακές τεχνικές σύνδεσης οργανωμένες σύμφωνα με την ευελιξία των προσδετών-πρωτεϊνών και τον τρόπο αναζήτησης των πιθανών διαμορφώσεων. Η εικόνα λήφθηκε από τη βιβλιογραφική πηγή [28].

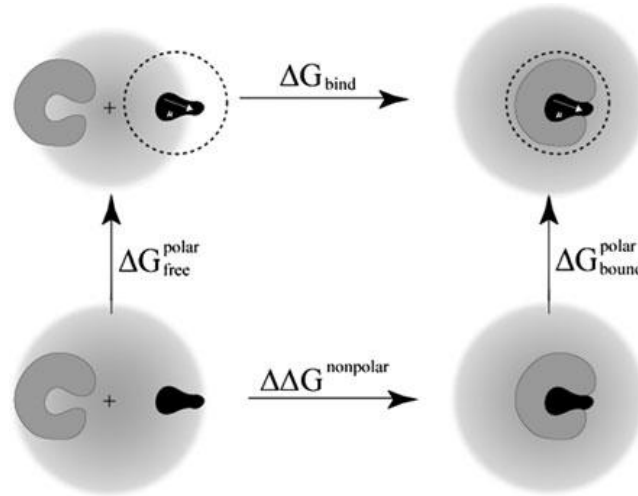
### Γραμμική Ενέργεια Αλληλεπίδρασης

Η μέθοδος της Γραμμικής Ενέργειας Αλληλεπίδρασης (Linear Interaction Energy, LIE) αναπτύχθηκε από τον Åqvist και τους συνεργάτες του [29] και αποτελεί μία πολύ ελκυστική μέθοδο για τη φάση της βελτιστοποίησης των μορίων-οδηγών καθώς προσφέρει έναν ικανοποιητικό συμβιβασμό μεταξύ ακρίβειας και ταχύτητας. Η διαδικασία της πρόσδεσης ενός μορίου σε ένα βιολογικό μακρομόριο μπορεί να θεωρηθεί ως μία διεργασία, κατά την οποία ο προσδέτης (ligand, l) μεταφέρεται από ένα μέσο σε ένα άλλο. Στο αρχικό μέσο (συνήθως νερό) ο προσδέτης βρίσκεται σε ελεύθερη κατάσταση (free, f), ενώ στο τελικό μέσο (συνήθως πρωτεΐνη σε περιβάλλον νερού) ο προσδέτης είναι δεσμευμένος (bound, b). Συνεπώς, για την περιγραφή της ελεύθερης ενέργειας πρόσδεσης θα πρέπει να ληφθούν υπόψη και οι δύο καταστάσεις στις οποίες βρίσκεται ο προσδέτης, τόσο η αρχική ελεύθερη, όσο και η τελική δεσμευμένη κατάσταση [30]. Η μέθοδος LIE εκτιμά την ελεύθερη ενέργεια πρόσδεσης ως την ελεύθερη ενέργεια μεταφοράς του προσδέτη από το περιβάλλον νερού, στο περιβάλλον της επιδιαλυτωμένης πρωτεΐνης (Εξίσωση 1.3).

$$\Delta G_{bind}(l) = \Delta G_{sol}^b(l) - \Delta G_{sol}^f(l) \quad (1.3)$$

Προκειμένου να υπολογιστεί η ελεύθερη ενέργεια της πρόσδεσης ως συνάρτηση των δύο καταστάσεων που βρίσκεται ο προσδέτης, θα πρέπει κανείς να σχεδιάσει το θερμοδυναμικό κύκλο που απεικονίζεται στο Σχήμα 1.5. Στον κύκλο αυτό, οι άνω γωνίες αντιπροσωπεύουν τις φυσικές καταστάσεις (αριστερά: επιδιαλυτωμένος σε νερό, δεξιά: δεσμευμένος στην πρωτεΐνη), ενώ οι δύο

κάτω γωνίες αναπαριστούν δύο μη φυσικές, ενδιάμεσες καταστάσεις όπου ο προσδέτης είναι απαλλαγμένος από οποιαδήποτε ηλεκτροστατική αλληλεπίδραση στην ελεύθερη (αριστερά) και δεσμευμένη (δεξιά) κατάσταση.



**Σχήμα 1.5** Ο θερμοδυναμικός κύκλος που χρησιμοποιείται για την εκτίμηση των ελεύθερων ενεργειών πρόσδεσης με τη μέθοδο γραμμικής ενέργειας αλληλεπίδρασης βάσει της Εξίσωσης 1.4 [30].

Η επίλυση του θερμοδυναμικού κύκλου του Σχήματος 1.5 οδηγεί στην ακόλουθη Εξίσωση 1.4:

$$\Delta G_{bind} = (\Delta G_{bound}^{polar} - \Delta G_{free}^{polar}) + \Delta \Delta G_{bind}^{nonpolar} = \Delta \Delta G_{bind}^{polar} + \Delta \Delta G_{bind}^{nonpolar} \quad (1.4)$$

Στην παραπάνω εξίσωση, η ελεύθερη ενέργεια πρόσδεσης μπορεί να εκφραστεί ως άθροισμα των πολικών και μη πολικών συνιστωσών της ελεύθερης ενέργειας. Η συνεισφορά του πολικού όρου, υπολογίζεται μέσω μιας προσέγγισης που πηγάζει από τη θεωρία γραμμικής απόκρισης για τις ηλεκτροστατικές δυνάμεις [30]. Έτσι, στη γενική του μορφή ο πολικός όρος της Εξίσωσης 1.4 γράφεται ως:

$$\Delta G_{bind}^{polar} = \beta (\langle U_{l-s}^{el} \rangle_b - \langle U_{l-s}^{el} \rangle_f) = \beta \Delta \langle U_{l-s}^{el} \rangle \quad (1.5)$$

Όπου,  $\beta = 1/(k_B T)$ , οι  $\langle \rangle$  συμβολίζουν μέσο όρο των ενεργειακών αλληλεπιδράσεων προσδέτη-περιβάλλοντος (ligand-surrounding, l-s). Ο μη πολικός όρος της Εξίσωσης 1.4 υπολογίζεται με βάση την παρακάτω σχέση:

$$\Delta G_{bind}^{nonpolar} = a (\langle U_{l-s}^{vdW} \rangle_b - \langle U_{l-s}^{vdW} \rangle_f) = a \Delta \langle U_{l-s}^{vdW} \rangle + \gamma \quad (1.6)$$

Όπου, η παράμετρος  $a$  προέρχεται από έναν εμπειρικά προσδιοριζόμενο παράγοντα κλιμάκωσης και  $\gamma$  είναι μία σταθερά. Λαμβάνοντας υπόψιν όλες τις παραπάνω σχέσεις, η πλήρης εξίσωση με βάση την οποία η μέθοδος LIE εκτιμά την ελεύθερη ενέργεια πρόσδεσης είναι:

$$\Delta G_{bind} = a \Delta \langle U_{l-s}^{vdW} \rangle + \beta \Delta \langle U_{l-s}^{el} \rangle + \gamma \quad (1.7)$$

Σημειώνεται ότι με την Εξίσωση 1.7, μπορεί κανείς να υπολογίσει την ελεύθερη ενέργεια πρόσδεσης με βάση τον μέσο όρο των δυναμικών ενεργειών που χαρακτηρίζουν το σύστημα περιβάλλον-

προσδέτης, οι οποίες συλλέγονται μόνο για τις δύο φυσικές καταστάσεις που βρίσκεται ο προσδέτης (άνω γωνίες στο Σχήμα 1.5). Αυτό κάνει μια ουσιαστική διαφορά στην ακρίβεια της μεθόδου σε σύγκριση με άλλες μεθόδους για την εκτίμηση των ελεύθερων ενεργειών πρόσδεσης, όπως η Διατάραξη της Ελεύθερης Ενέργειας, στις οποίες οι ενδιάμεσες μη-φυσικές καταστάσεις προσομοιώνονται ξεχωριστά [30].

### **Η μέθοδος MM/PBSA**

Μία από τις πιο δημοφιλείς μεθόδους για τον υπολογισμό της ελεύθερης ενέργειας πρόσδεσης ενός μορίου σε έναν υποδοχέα, αποτελεί η μέθοδος της Μοριακής Μηχανικής με Poisson-Boltzmann με επιδιαλύτωση της επιφάνειας (Molecular Mechanics with Poisson-Boltzmann and surface area solvation, MM/PBSA) [12]. Η μέθοδος MM/PBSA αναπτύχθηκε από τον Kollman και τους συνεργάτες του [31] στα τέλη της δεκαετίας του 1990 και αποτελεί ένα συμβιβασμό μεταξύ ακρίβειας και ταχύτητας για τον υπολογισμό της ελεύθερης ενέργειας πρόσδεσης. Η MM-PBSA απαιτεί την προσομοίωση μόνο των δεσμευμένων και των μη δεσμευμένων καταστάσεων [12]. Η ελεύθερη ενέργεια πρόσδεσης μπορεί να γραφτεί ως μια διαφορά στις ελεύθερες ενέργειες επιδιαλύτωσης κάθε ουσίας με βάση την Εξίσωση 1.8:

$$\Delta G_{Bind} = \Delta G_{PL-solv} - \Delta G_{L-solv} + \Delta G_{P-solv} \quad (1.8)$$

Κάθε μία από τις ενέργειες επιδιαλύτωσης μπορεί να γραφτεί ως εξής:

$$\Delta G_{solv} = \Delta H_{solv} - T\Delta S_{solv} \quad (1.9)$$

Υπολογίζοντας κατά μέσο όρο τις συντεταγμένες του διαλύτη σε όλες τις διαμορφώσεις, η Εξίσωση 1.9 προσεγγίζεται με βάση την Εξίσωση 1.10:

$$\Delta G_{X-solv} = \langle E_{X-MM} \rangle + \Delta G_{X-solvent} - T\Delta S_{X-MM} \quad (1.10)$$

Με  $\langle E_X \rangle$  συμβολίζεται η μέση δυναμική ενέργεια του μορίου X (δίχως νερό),  $\Delta S_X$  είναι η εσωτερική εντροπία του X (δίχως νερό), και  $\Delta G_X$  είναι η ενέργεια και η εντροπία λόγω της επιδιαλύτωσης του X. Αυτές οι ενέργειες επιδιαλύτωσης για την πρωτεΐνη (protein, P), τον προσδέτη (ligand, L) και το σύμπλοκο πρωτεΐνης-προσδέτη (protein-ligand, PL) μπορούν στη συνέχεια να συνδυαστούν για τον υπολογισμό της συνολικής ελεύθερης ενέργειας πρόσδεσης. Η δυναμική ενέργεια  $E_{X-MM}$  αφορά μόνο την πρωτεΐνη και τον προσδέτη και υπολογίζεται με βάση την Εξίσωση 1.11.

$$E_{X-MM} = E^{elec} + E^{vdW} + E^{int} \quad (1.11)$$

Όπου,  $E^{elec}$  είναι η ηλεκτροστατική ενέργεια,  $E^{vdW}$  η ενέργεια των δυνάμεων van der Waals και  $E^{int}$  αποτελείται από όρους εσωτερικής ενέργειας τόσο για τον προσδέτη όσο και για την πρωτεΐνη. Ο όρος της ενέργειας επιδιαλύτωσης  $\Delta G_{X-solv}$  χωρίζεται σε άθροισμα δύο στοιχείων, ένα λόγω ηλεκτροστατικών και ένα λόγω μη πολικών αλληλεπιδράσεων:

$$\Delta G_{X-solv} = \Delta G_{PBSA} = \Delta G_{PB} + \Delta G_{SA} \quad (1.12)$$



Ο όρος  $\Delta G_{PB}$  αντιπροσωπεύει την πολική συνεισφορά, ενώ ο όρος  $\Delta G_{SA}$  αντιστοιχεί στο μη πολικό μέρος της ελεύθερης ενέργειας επιδιάλυτωσης. Ο πολικός όρος στην Εξίσωση 1.12 αντιπροσωπεύει την ενέργεια που αποθηκεύεται στο διηλεκτρικό συνεχές σε απόκριση της παρουσίας της κατανομής φορτίου του διαλύτη και τυπικά λαμβάνεται από την επίλυση της εξίσωσης Poisson-Boltzmann (PB). Ο τελευταίος όρος της Εξίσωσης 1.12 αφορά το μη πολικό κομμάτι της ελεύθερης ενέργειας επιδιάλυτωσης και συνήθως θεωρείται ανάλογος με την επιφάνεια που είναι εκτεθειμένη στο διαλύτη:

$$G^{SA} = \gamma \Delta SA \quad (1.13)$$

Με  $\Delta SA$  αντιπροσωπεύεται η μεταβολή της μοριακής επιφάνειας κατά τη σύνδεση, ενώ με τον όρο  $\gamma$  είναι η ελεύθερη ενέργεια επιφάνειας για το σχηματισμό της αντίστοιχης κοιλότητας στο νερό [32]. Η Εξίσωση 1.13 προέρχεται από εμπειρικά δεδομένα για γραμμικούς, κυκλικούς και διακλαδωμένους υδρογονάνθρακες και όπως είναι αναμενόμενο περιέχει αρκετές απλοποιήσεις.

Ο τελευταίος όρος στο δεξί μέλος της Εξίσωσης 1.10 αναπαριστά το εντροπικό κόστος περιορισμού του ελεύθερου προσδέτη, το οποίο αντιπροσωπεύει ένα σημαντικό κλάσμα της ολικής μεταβολής της εντροπίας της διαλυμένης ουσίας κατά το σχηματισμό του συμπλόκου προσδέτη-πρωτεΐνης. Επιπρόσθετα, για την πιο ακριβή εκτίμηση της εντροπίας του προσδέτη μπορούν να χρησιμοποιηθούν μέθοδοι οι οποίες βασίζονται στην ανάλυση κανονικών δονήσεων (normal mode analysis), ωστόσο είναι υπολογιστικά δαπανηρές. Εναλλακτικά, μπορεί να γίνει χρήση εμπειρικών εκτιμήσεων που βασίζονται στον υπολογισμό της εντροπίας που απαιτείται για τον περιορισμό της περιστροφής του προσδέτη γύρω από κάθε δεδομένο περιστροφικό βαθμό ελευθερίας. Παρ' όλα αυτά, καμία από αυτές τις προσεγγίσεις δεν παράγει αξιόπιστες εκτιμήσεις για την εντροπία του προσδέτη. Δεδομένου ότι η επίλυση της εξίσωσης Poisson-Boltzmann είναι υπολογιστικά απαιτητική, ορισμένες ερευνητικές ομάδες έχουν καταβάλει σημαντικές προσπάθειες για την ανάπτυξη ταχύτερων προσεγγίσεων, όπως οι γενικευμένες προσεγγίσεις Born (Generalized Born, GB) [12], δίχως ωστόσο κάποια επιτυχία [33].

### 1.2.3 Η μέθοδος διατάραξης της ελεύθερης ενέργειας για τη βελτιστοποίηση ενώσεων οδηγών

Οι μέθοδοι που αναλύθηκαν παραπάνω δεν επαρκούν για μια πλήρη μοριακή περιγραφή των φαινομένων που συμβαίνουν κατά την πρόσδεση ενός προσδέτη σε μία πρωτεΐνη, όπως ο σχηματισμός δικτύων δεσμών υδρογόνου στις θέσεις πρόσδεσης. Στη βιβλιογραφία μάλιστα έχουν αναφερθεί πολλές περιπτώσεις συστημάτων πρωτεΐνης-προσδέτη όπου οι ακριβείς θέσεις των μορίων νερού γύρω από τη θέση πρόσδεσης παίζουν σημαντικό ρόλο κατά την πρόσδεση των μορίων στις πρωτεΐνες [12].

Η πιο έγκυρη προσέγγιση για την περιγραφή της συμπεριφοράς των ατόμων ενός συστήματος προσδέτη-πρωτεΐνης και την ακριβή εκτίμηση της απόλυτης ή σχετικής ελεύθερης ενέργειας πρόσδεσης, είναι οι υπολογιστικά δαπανηρές προσομοιώσεις ελεύθερης ενέργειας. Μία από τις πιο καλά καθιερωμένες μεθόδους για το σκοπό αυτό είναι η Διατάραξη Ελεύθερης Ενέργειας (Free Energy Perturbation, FEP). Ο όρος «Διατάραξη» στη μεθοδολογία FEP αναφέρεται στην αλλαγή της

χημικής ταυτότητας ενός μορίου, αφού οι προσομοιώσεις συχνά περιλαμβάνουν μετατροπή μιας χημικής οντότητας σε μία άλλη. Συνήθως οι ενεργειακοί όροι που διαταράσσονται αφορούν τις ηλεκτροστατικές και τις van der Waals αλληλεπιδράσεις του μορίου-προσδέτη.

Η θεωρία της Διατάραξης της Ελεύθερης Ενέργειας σε συνδυασμό με τεχνικές δειγματοληψίας Μοριακής Δυναμικής και Monte Carlo, παρουσιάστηκε για πρώτη φορά από τον Zwanzig το 1954 [34]. Με βάση την προσέγγιση αυτή, η διαφορά της ελεύθερης ενέργειας κατά Gibbs μεταξύ μιας αρχικής (αναφοράς) και τελικής (στόχου) κατάστασης ενός συστήματος είναι συνάρτηση της ενεργειακής διαφοράς που υπολογίζεται με δειγματοληψία μόνο της αρχικής κατάστασης (Εξίσωση 1.14).

$$\Delta G = G_1 - G_0 = -k_B T \ln \left\langle \exp \left[ \frac{-(E_1(x) - E_0(x))}{k_B T} \right] \right\rangle_0 \quad (1.14)$$

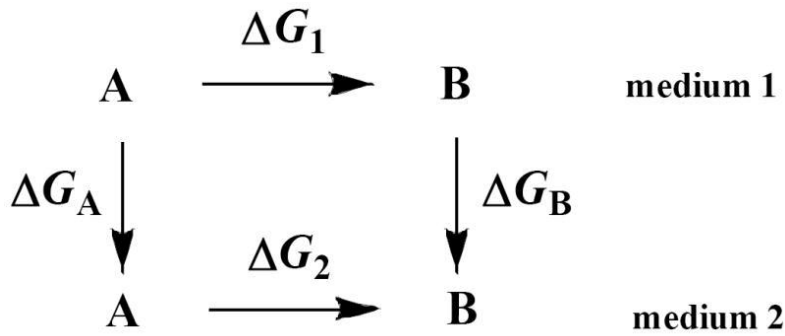
Στην Εξίσωση 1.14 με 0 και 1 συμβολίζονται η αρχική και τελική κατάσταση, αντίστοιχα, ενώ με E συμβολίζεται η ενέργεια του κάθε συστήματος. Ο Zwanzig στη συνέχεια προσπάθησε να παράγει μια έκφραση για τη μεταβολή της ελεύθερης ενέργειας ως δυναμοσειρά. Καθορίζοντας την ενέργεια της κατάστασης στόχου,  $E_1$ , ως άθροισμα της ενέργειας μιας κατάστασης αναφοράς,  $E_0$  και ενός μικρού δυναμικού διαταραχής,  $V$  (Εξίσωση 1.15), προκύπτει η Εξίσωση 1.16 επεκτείνοντας την Εξίσωση 1.14 στη δεύτερη τάξη.

$$E_1 = E_0 + V \quad (1.15)$$

$$\Delta G = G_1 - G_0 = \langle V \rangle - \frac{1}{2k_B T} (\langle V^2 \rangle - \langle V \rangle^2) \quad (1.16)$$

Εδώ, η διαφορά ελεύθερης ενέργειας μεταξύ των καταστάσεων αναφοράς και στόχου εκτιμάται στην πρώτη τάξη από τον μέσο όρο του συνόλου αναφοράς του δυναμικού διαταραχής και στην δεύτερη τάξη συμπεριλαμβάνοντας τη μέση τετραγωνική απόκλιση του δυναμικού διαταραχής από τον μέσο όρο. Αυτή η προσέγγιση για τον υπολογισμό των διαφορών ελεύθερης ενέργειας έχει αναγνωριστεί ως ιδιαίτερα σημαντική για συστήματα με διακυμάνσεις που χαρακτηρίζονται από κατανομή πιθανότητας κατά Gauss, ενώ βρίσκει εφαρμογές σε διάφορες περιπτώσεις κατά τη φόρτιση των ιόντων σε διάλυμα, την εκτίμηση των εγγενών  $\rho_{Ka}$  και ο υπολογισμός της συνεισφοράς του διαλύτη στις διεγερμένες καταστάσεις της τρυπτοφάνης [35].

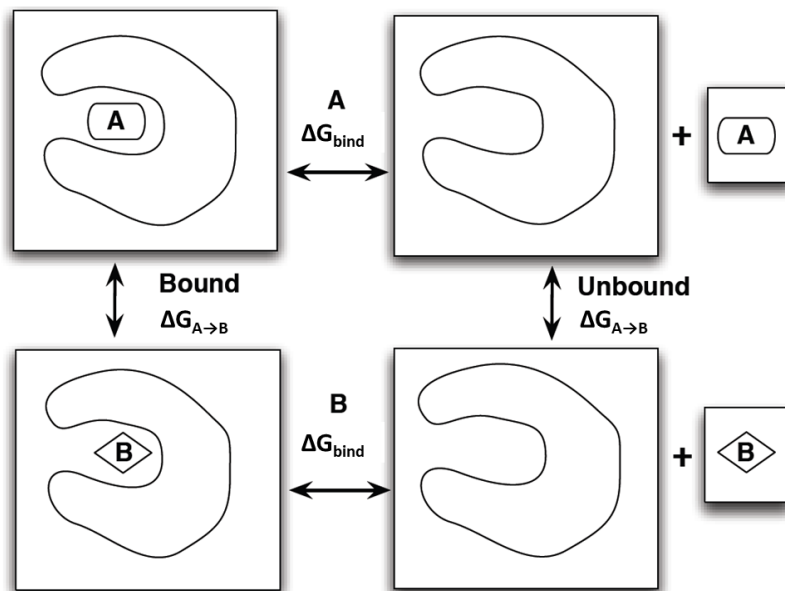
Όταν η Εξίσωση 1.14 εφαρμόζεται κατά τη χημική ισορροπία, συνήθως οι καταστάσεις 0 και 1 αφορούν διαφορετικά μόρια A και B, και εξετάζεται η αλλαγή της ελεύθερης ενέργειας για τη μετατροπή του A σε B σε δύο διαφορετικά μέσα. Για το σκοπό αυτό κατασκευάζεται ένας θερμοδυναμικός κύκλος όπως στο Σχήμα 1.6, όπου  $\Delta G_A$  και  $\Delta G_B$  είναι οι ελεύθερες ενέργειες μεταφοράς A και B από το μέσο 1 στο μέσο 2. Τα  $\Delta G_1$  και  $\Delta G_2$  υπολογίζονται και το αποτέλεσμα της αλλαγής του μέσου δίνεται από τη σχετική ελεύθερη ενέργεια  $\Delta \Delta G = \Delta G_2 - \Delta G_1 = \Delta G_B - \Delta G_A$ . Οι πρώτοι υπολογισμοί FEP όπως παρουσιάζονται στο Σχήμα 1.6 πραγματοποιήθηκαν για πρώτη φορά το 1985 από τον W.L Jorgensen με σκοπό τον υπολογισμό των σχετικών ελεύθερων ενεργειών επιδιалύτωσης αιθανίου (A) και μεθανόλης (B) με το μέσο 1 και 2 να αντιπροσωπεύουν την αέρια και υγρή φάση, αντίστοιχα [36].



**Σχήμα 1.6** Θερμοδυναμικός κύκλος για την εύρεση της αλλαγής της ελεύθερης ενέργειας για τη μετατροπή μιας ουσίας A σε μία ουσία σε δύο διαφορετικά περιβάλλοντα. Η εικόνα λήφθηκε από τη βιβλιογραφική πηγή [35].

Αναλόγως με το τι επιθυμεί να επιτύχει κανείς μέσα από την μοριακή προσομοίωση και τον υπολογισμό της μεταβολής της ελεύθερης ενέργειας μέσω της Διατάραξης Ελεύθερης Ενέργειας, μπορεί να κατασκευάσει έναν αντίστοιχο θερμοδυναμικό κύκλο ο οποίος θα καθορίζει τις τελικές καταστάσεις της προσομοίωσης. Για παράδειγμα, κατά το στάδιο της βελτιστοποίησης της ένωση-οδηγού συνήθως πρέπει κανείς να γνωρίζει εάν μια ένωση A ή μια ένωση B προσδέεται ισχυρότερα στο φαρμακευτικό στόχο. Έστω ότι η A αποτελεί την ένωση-οδηγό και B είναι μία ένωση χημικά παρόμοια με την A. Ο θερμοδυναμικός κύκλος για τη μετατροπή της ένωσης A στην ένωση B απεικονίζεται στο Σχήμα 1.7 και ο τρόπος για να ελέγξει κανείς ποια από τις δύο ενώσεις προσδέεται ισχυρότερα στο στόχο είναι μέσω της Εξίσωσης 1.17. Έτσι, εάν  $\Delta\Delta G_{bind} < 0$ , τότε η πρόσδεση του B είναι πιο ευνοϊκή από την πρόσδεση του A.

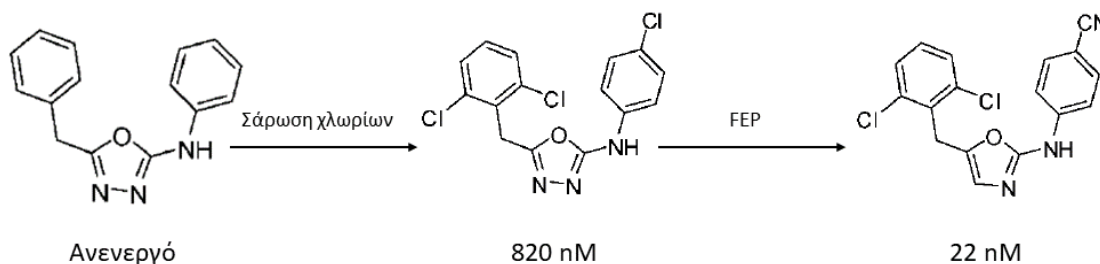
$$\Delta\Delta G_{bind} = \Delta G_{bind}^B - \Delta G_{bind}^A = \Delta G_{A \rightarrow B}^{bound} - \Delta G_{A \rightarrow B}^{unbound} \quad (1.17)$$



**Σχήμα 1.7** Θερμοδυναμικός κύκλος για την εκτίμηση της σχετικής ελεύθερης ενέργειας πρόσδεσης. Για το σκοπό αυτό θα μπορούσε κανείς να αφαιρέσει και τα δύο μόρια από τη θέση πρόσδεσης, κάτι που είναι χρονικά απαγορευτικό ή θα μπορούσε απλά να μετατρέψει το μόριο A στο B εντός και εκτός της θέσης πρόσδεσης και να υπολογίσει ευκολότερα τη διαφορά  $\Delta\Delta G$  [37].

Χαρακτηριστικό παράδειγμα χρήσης της Διατάραξης Ελεύθερης Ενέργειας για τη βελτιστοποίηση μίας ένωσης-οδηγού αποτελεί ο μετασχηματισμός του σκελετού της 2-ανιλινυλο-5-βενζυλοξαδιαζόλης σε δύο ισχυρούς αναστολείς της αντίστροφης μεταγραφάσης του ιού HIV. Πιο συγκεκριμένα, η ομάδα του W.L Jorgensen πραγματοποίησε μια καθοδηγούμενη από την FEP «σάρωση χλωρίων» για να αναγνωρίσει υποσχόμενες περιοχές για την υποκατάσταση των αρυλικών υδρογόνων από μόρια χλωρίου. Τα αποτελέσματα της σάρωσης ανέδειξαν δύο πιθανούς αναστολείς με τιμές  $EC_{50}$ <sup>1</sup> 820 και 310 nM για την προστασία ανθρώπινων T-λεμφοκυττάρων από τον ιό HIV-1. Μετέπειτα υπολογισμοί Διατάραξης Ελεύθερης ενέργειας οδήγησαν σε περαιτέρω τροποποιήσεις οι οποίες ανέδειξαν δύο οξαζόλες που αναγνωρίστηκαν ως πιθανά φάρμακα για την αντιμετώπιση του HIV με τιμές  $EC_{50}$  στην περιοχή 10-20 nM. Στο Σχήμα 1.8 συνοψίζεται η συνολική πορεία για τη βελτιστοποίηση του ανενεργού σκελετού 2-ανιλινυλο-5-βενζυλοξαδιαζόλης σε ένα υποψήφιο φάρμακο.

Ακολούθησε χημική σύνθεση των δύο ενώσεων και ύστερα εξετάστηκε εάν διαθέτουν τις απαραίτητες ADME ιδιότητες χρησιμοποιώντας επιπρόσθετα υπολογιστικά εργαλεία. Συνολικά, οι σχετικές ελεύθερες ενέργειες που εκτίμησε η μέθοδος FEP επιβεβαιώθηκαν από τα αντίστοιχα πειραματικά αποτελέσματα. Επομένως αντιλαμβάνεται κανείς πως η βελτιστοποίηση μορίων οδηγών καθοδηγούμενη από τη Διατάραξη Ελεύθερης Ενέργειας αποτελεί ένα πολύτιμο εργαλείο για τον ορθολογικό σχεδιασμό φαρμάκων [38].



**Σχήμα 1.8** Βελτιστοποίηση του σκελετού της ανενεργής 2-ανιλινυλο-5-βενζυλοξαδιαζόλης με υπολογισμούς Διατάραξης Ελεύθερης Ενέργειας [38].

#### 1.2.4 Πειραματικές τεχνικές για τον προσδιορισμό ελεύθερης ενέργειας πρόσδεσης

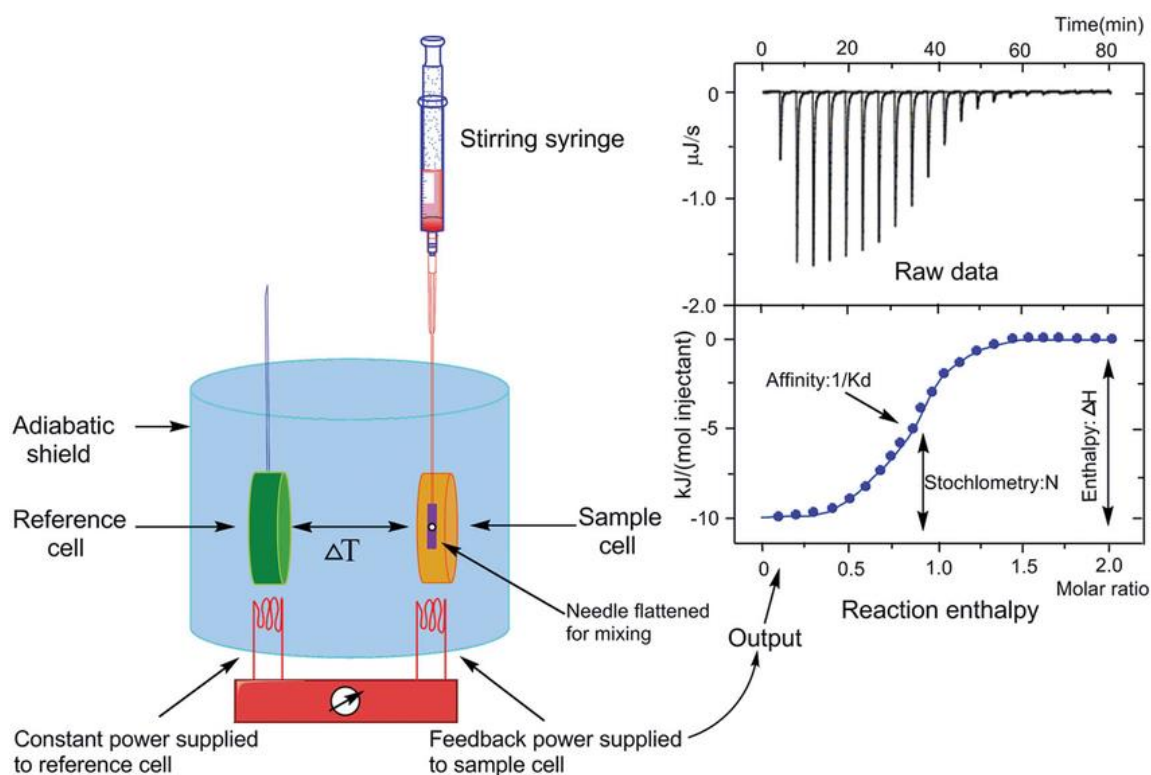
Στην προηγούμενη ενότητα αναλύθηκαν μερικές από τις πιο δημοφιλείς υπολογιστικές μεθόδους για την πρόβλεψη της ελεύθερης ενέργειας πρόσδεσης ενός μορίου σε ένα φαρμακευτικό στόχο. Παρά την τεράστια προσπάθεια που έχει καταβληθεί τις τελευταίες δεκαετίες από τον κλάδο της υπολογιστικής χημείας, οι υπολογιστικές μέθοδοι για τον προσδιορισμό της ελεύθερης ενέργειας πρόσδεσης δεν έχουν κατορθώσει να αναπαράγουν τις αντίστοιχες πειραματικές μετρήσεις με την απαιτούμενη ακρίβεια. Στη βιβλιογραφία υπάρχουν πολλές πειραματικές δοκιμές που μπορούν να εφαρμοστούν για την εκτίμηση της ελεύθερης ενέργειας πρόσδεσης πρωτεΐνης-προσδέτη. Έτσι λοιπόν κρίνεται σκόπιμη η σύντομη παρουσίαση των κυριότερων πειραματικών τεχνικών που χρησιμοποιούνται για την εκτίμηση της ελεύθερης ενέργειας πρόσδεσης.

<sup>1</sup> Ο όρος  $EC_{50}$  ή αλλιώς μέση αποτελεσματική συγκέντρωση ενός φαρμάκου αναφέρεται στη συγκέντρωση του φαρμάκου όπου παρατηρείται το 50% του μέγιστου θεραπευτικού αποτελέσματος.

## Ισοθερμική Θερμιδομετρία Τιτλοδότησης

Ισοθερμική Θερμιδομετρία Τιτλοδότησης (Isothermal Titration Calorimetry, ITC) είναι μία φυσική, ποσοτική τεχνική που μετρά άμεσα την ανταλλαγή θερμότητας κατά τη διάρκεια σχηματισμού συμπλόκου ανάμεσα στην πρωτεΐνη και το μόριο, σε σταθερή θερμοκρασία. Αποτελείται από δύο κελιά (cells) που περικλείονται από ένα αδιαβατικό πλαίσιο. Οι ενώσεις που πρόκειται να μελετηθούν τοποθετούνται στο κελί δείγματος (sample cell), ενώ το άλλο κελί (reference cell) χρησιμοποιείται ως control και περιέχει το ρυθμιστικό διάλυμα στο οποίο διαλύεται το δείγμα. Ένα τυπικό πείραμα ITC περιέχει τρία στάδια. Το πρώτο αφορά την τιτλοδότηση του μορίου-προσδέτη στο διάλυμα που περιέχει την πρωτεΐνη. Ύστερα, η θερμότητα που εκλύεται ή απορροφάται κατά την πρόσδεση του μορίου στην πρωτεΐνη μετράται και τέλος τα δεδομένα της ITC επεξεργάζονται και προσαρμόζονται για να ληφθούν μεγέθη όπως η σταθερά πρόσδεσης ( $K_d$ ), η ελεύθερη ενέργεια πρόσδεσης Gibbs ( $\Delta G$ ), η ενθαλπία πρόσδεσης ( $\Delta H$ ), η εντροπία πρόσδεσης ( $\Delta S$ ) και η στοιχειομετρία ( $N$ ) της αντίδρασης (Σχήμα 1.9). Συγκεκριμένα, οι μεταβολές της ελεύθερης ενέργειας Gibbs και της εντροπίας μπορούν να προσδιοριστούν χρησιμοποιώντας την Εξίσωση 1.18. Οι συνολικές μετρήσεις συνίστανται από την χρονικά εξαρτώμενη ισχύ που απαιτείται για τη διατήρηση ίσων θερμοκρασιών μεταξύ των δύο κελιών σε κάθε τιτλοδότηση [39].

$$\Delta G = -RT \ln(K_d) = \Delta H - T\Delta S \quad (1.18)$$



**Σχήμα 1.9** Βασική αρχή της ισοθερμικής θερμιδομετρίας τιτλοδότησης. Σχηματική αναπαράσταση του ισοθερμικού θερμιδομέτρου τιτλοδότησης (αριστερά) και ένα χαρακτηριστικό πείραμα τιτλοδότησης (άνω δεξιά) με την αξιολόγηση (κάτω δεξιά). Στην (άνω δεξιά) εικόνα, το θερμόγραμμα τιτλοδότησης αντιπροσωπεύεται ως θερμότητα ανά μονάδα χρόνου που απελευθερώνεται μετά από κάθε έγχυση του προσδέτη στην πρωτεΐνη (μαύρο), καθώς επίσης και η αραίωση του προσδέτη σε ρυθμιστικό (μπλε). Στην εικόνα (κάτω δεξιά) αναπαρίσταται η εξάρτηση της απελευθερωμένης θερμότητας σε κάθε έγχυση ως συνάρτηση του λόγου μεταξύ της συνολικής συγκέντρωσης του προσδέτη και της συνολικής συγκέντρωσης πρωτεΐνης. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [40].

## Πόλωση Φθορισμού

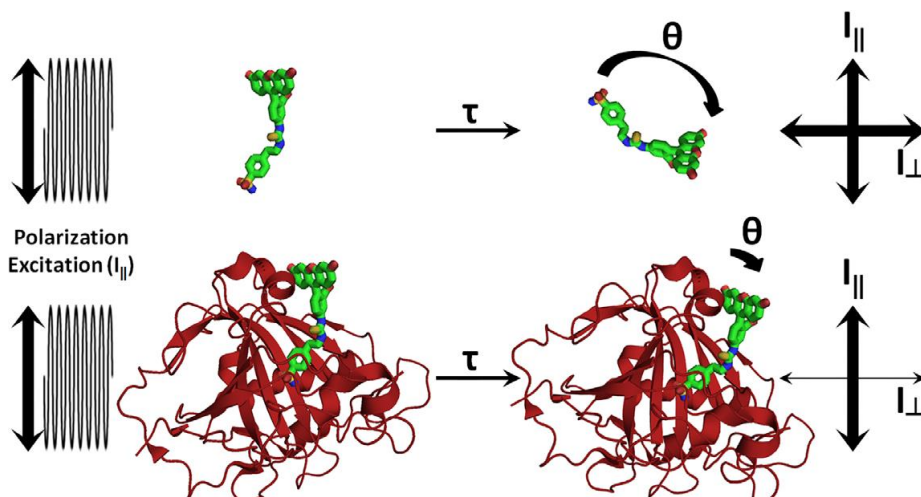
Το 1926, ο Francis Perrin περιέγραψε τη θεωρητική βάση της πόλωσης του φθορισμού παρατηρώντας πως η εκπομπή από ένα μικρό φθορίζον μόριο που διεγείρεται από το πολωμένο φως, είναι σε μεγάλο βαθμό μη πολωμένο λόγω περιστροφικής διάχυσης κατά τη διάρκεια του φθορισμού. Ο Perrin εξήγαγε την Εξίσωση (1.19), όπου  $FP$  είναι η παρατηρούμενη πόλωση,  $R$  είναι η παγκόσμια σταθερά των αερίων,  $T$  είναι η απόλυτη θερμοκρασία,  $\eta$  είναι το ιξώδες του διαλύματος,  $FP_0$  είναι η εγγενής πολικότητα (η τιμή πόλωσης απουσία μοριακής περιστροφής ή  $FP$  όταν  $T/\eta \rightarrow 0$ ),  $V$  είναι ο γραμμομοριακός όγκος και  $\tau$  είναι ο χρόνος της διεγερμένης κατάστασης του φθορισμού. Η Εξίσωση 1.19 δείχνει σαφώς πώς η παρατηρούμενη πόλωση φθορισμού εξαρτάται από τον μοριακό όγκο των φθορίζοντων ουσιών όταν οι συνθήκες θερμοκρασίας και διαλύματος είναι σταθερές. Επομένως, η πόλωση του φθορισμού μπορεί να χρησιμοποιηθεί για να παρατηρηθούν βιολογικές διεργασίες που συνεπάγονται αλλαγές στο μοριακό βάρος, όπως η πρόσδεση πρωτεΐνης-προσδέτη ή η ενζυματική διάσπαση ενός υποστρώματος. Ο χρόνος περιστροφικής συσχέτισης ( $\theta$ ) για ένα μόριο, που φαίνεται στην Εξίσωση 1.20 συμβάλλει στην (1.19) λαμβάνοντας υπόψιν ότι η  $FP$  αυξάνεται με την αύξηση του  $\theta$ . Από την Εξίσωση 1.19 είναι επίσης σαφές ότι η  $FP$  μειώνεται όσο αυξάνεται η  $\tau$ .

$$\frac{1}{FP} - \frac{1}{3} = \left( \frac{1}{FP_0} - \frac{1}{3} \right) \left( 1 + \frac{RT}{\eta V} \tau \right) \quad (1.19)$$

$$\theta = \frac{\eta V}{RT} \quad (1.20)$$

Στη δεκαετία του 1950, ο Weber επέκτεινε το θεωρητικό έργο του Perrin αναπτύσσοντας το πρώτο όργανο μέτρησης της πόλωσης του φθορισμού το οποίο χρησιμοποίησε για τη μελέτη των πρωτεϊνών. Για τη μέτρηση της πόλωσης του φθορισμού στο εργαστήριο, ένα φθορίζον δείγμα διεγείρεται από το πολωμένο φως και οι εντάσεις εκπομπής συλλέγονται από διαύλους παράλληλους ( $I_{||}$ ) και κάθετους ( $I_{\perp}$ ) στον ηλεκτρικό φορέα του φωτός διέγερσης (Σχήμα 1.10). Αυτές οι μετρήσεις μπορούν να χρησιμοποιηθούν για τον υπολογισμό είτε της πόλωσης φθορισμού ( $FP$ ) είτε της ανισοτροπίας φθορισμού. Από την Εξίσωση 1.21 υπολογίζεται η  $FP$ , ως η διαφορά μεταξύ της παράλληλης ( $I_{||}$ ) και της κάθετης έντασης εκπομπών κανονικοποιημένων με τη συνολική ένταση φθορισμού της δέσμης εκπομπών [41].

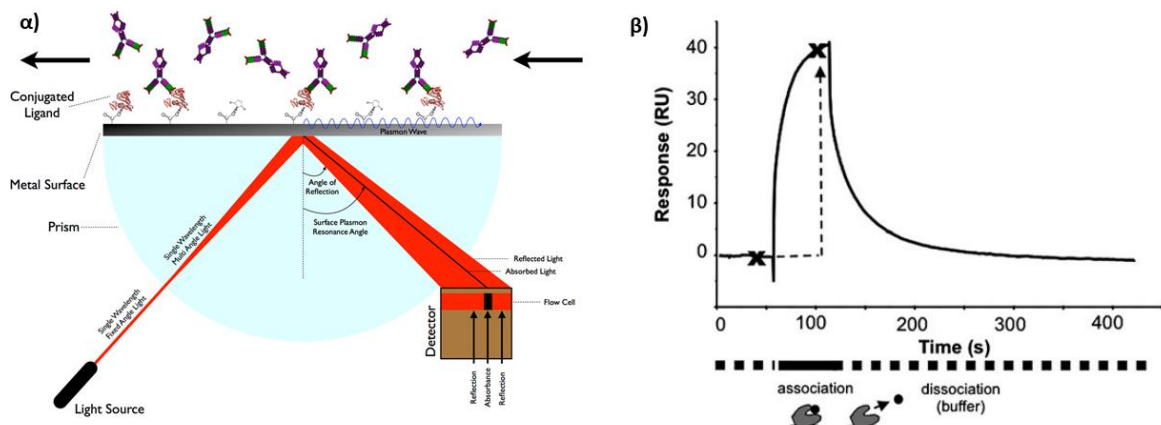
$$FP = \frac{I_{||} - I_{\perp}}{I_{||} + I_{\perp}} \quad (1.21)$$



**Σχήμα 1.10** Η βασική αρχή της πόλωσης του φθορισμού. Ένας φθορίζων ανιχνευτής διεγείρεται με πολωμένο φως ( $I_{||}$ ) και εκπέμπει μετά τη διάρκεια ζωής φθορισμού ( $\tau$ ) του ανιχνευτή. Η ένταση εκπομπής μετράται σε προσανατολισμούς που είναι παράλληλοι ( $I_{||}$ ) και κάθετοι ( $I_{\perp}$ ) στο επίπεδο διέγερσης. Ο βαθμός στον οποίο ο εκπεμπόμενος φθορισμός παραμένει πολωμένος ( $I_{||} \gg I_{\perp}$ ) εξαρτάται από το  $\tau$ , καθώς και από τον μοριακό μέγεθος του ανιχνευτή το οποίο καθορίζει το χρόνο περιστροφικής συσχέτισης ( $\theta$ ). **(Πάνω)** Ένας μη δεσμευμένος αναστολέας φθορισμού (πράσινος) έχει υψηλή περιστροφική διάχυση κατά τη διάρκεια  $\tau$ , με αποτέλεσμα ένα σήμα αποπολωμένης εκπομπής ( $I_{||} \approx I_{\perp}$ ). **(Κάτω)** Ο αναστολέας φθορισμού προσδένεται σε μεγαλύτερη πρωτεΐνη, με αποτέλεσμα μια μειωμένη περιστροφική διάχυση κατά τη διάρκεια του  $\tau$  και ένα σήμα πολωμένης εκπομπής ( $I_{||} \gg I_{\perp}$ ). Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [41].

## Συντονισμός Επιφανειακών Πλασμονίων

Ο Συντονισμός Επιφανειακών Πλασμονίων (Surface Plasmon Resonance, SPR) είναι μία οπτική τεχνική που χρησιμοποιείται για τη μελέτη των αλληλεπιδράσεων ανάμεσα σε βιομόρια. Τα αλληλεπιδρώντα μόρια μπορεί να είναι πρωτεΐνες, πεπτίδια, λιπίδια, ιοί, νουκλεϊκά οξέα ή μικρά οργανικά μόρια όπως διάφορα υποψήφια φάρμακα. Συγκεκριμένα, η μέθοδος SPR μετρά τη μεταβολή του δείκτη διάθλασης κοντά σε μία επιφάνεια ενός αισθητήρα και είναι ικανή να ποσοτικοποιήσει την κινητική πρόσδεσης ενός μορίου σε μία πρωτεΐνη σε πραγματικό χρόνο. Σε ένα όργανο SPR η επιφάνεια του αισθητήρα είναι συνήθως μία λεπτή μεμβράνη χρυσού, τοποθετημένη σε ένα γυάλινο στήριγμα, το οποίο σχηματίζει το επίπεδο ροής μιας κυψελίδας μέσω της οποίας ένα υδατικό διάλυμα ρέει συνεχώς (Σχήμα 1.11α). Τα μόρια της πρωτεΐνης είναι ακινητοποιημένα στην επιφάνεια του αισθητήρα και ο προσδέτης εγχέεται στο υδατικό διάλυμα για να ανιχνεύσει την αντίδραση πρόσδεσης. Σε μία δοκιμή SPR μία ακτίνα φωτός προσπίπτει σε μία μεταλλική μεμβράνη μέσω ενός πρίσματος και η ανακλώμενη δέσμη συλλέγεται και αναλύεται. Καθώς τα μόρια προσδένονται στα ακινητοποιημένα μόρια της πρωτεΐνης, παρατηρείται αύξηση στο δείκτη διάθλασης (εκφραζόμενη σε μονάδες απόκρισης, response units, RU) (Σχήμα 1.11β). Μετά από έναν επιθυμητό χρόνο αντίδρασης (δηλαδή όταν όλες οι θέσεις πρόσδεσης είναι κατειλημμένες), ένα διάλυμα που δεν περιέχει μόριο εγχέεται μέσω της κυψελίδας ώστε να διαχωριστεί το σύμπλεγμα πρωτεΐνης-προσδέτη. Καθώς ο προσδέτης αποσυνδέεται από την ακινητοποιημένη πρωτεΐνη, παρατηρείται μείωση στις RU. Όλα τα βήματα στην προετοιμασία και την ανάλυση της επιφάνειας παρακολουθούνται μέσα από ένα σενσογράφημα (sensogram), το οποίο καταγράφει το σχηματισμό και τη διάσπαση συμπλοκών σε όλη την πορεία της αλληλεπίδρασης, με την κινητική (ταχύτητα σύνδεσης και αποσύνδεσης) να αποκαλύπτεται από το σχήμα της καμπύλης σύνδεσης.



**Σχήμα 1.11 (α)** Τυπική διάταξη ενός οργάνου που χρησιμοποιείται κατά τον Συντονισμό Επιφανειακών Πλασμονίων. **(β)** Σχηματική απεικόνιση ενός σενσογραφήματος. Οι ράβδοι κάτω από την καμπύλη αισθητήρα

δείχνουν τις λύσεις που περνούν πάνω από την επιφάνεια του αισθητήρα. Η διακεκομμένη ράβδος υποδεικνύει τη διέλευση του ρυθμιστικού διαλύματος, ενώ η συμπαγής ράβδος υποδεικνύει έγχυση ενός διαλύματος προσδέτη που έχει σαν αποτέλεσμα την πρόσδεση αυτού του προσδέτη με έναν ακινητοποιημένο στόχο [42].

### 1.3 Διαμοριακές αλληλεπιδράσεις υποδοχέα-προσδέτη

Ένα σύστημα πρωτεΐνης-προσδέτη-διαλύτη είναι ένα θερμοδυναμικό σύστημα που αποτελείται από τη διαλυμένη ουσία (δηλαδή τα μόρια πρωτεΐνης και προσδέτη) και το διαλύτη (δηλαδή το νερό και τα ιόντα). Ως αποτέλεσμα, οι κινητήριες δυνάμεις που καθορίζουν τη συσχέτιση μεταξύ πρωτεΐνης και προσδετών είναι ένα σύνθετο αποτέλεσμα διαφόρων αλληλεπιδράσεων και ενεργειακών ανταλλαγών μεταξύ της πρωτεΐνης, του προσδέτη, του νερού και των ιόντων. Όπως προαναφέρθηκε και στην προηγούμενη ενότητα, η ελεύθερη ενέργεια Gibbs, η οποία είναι ένα δυναμικό που μετρά την ικανότητα ενός συστήματος να εκτελέσει μέγιστο έργο σε σταθερή θερμοκρασία και πίεση, είναι μία από τις σημαντικότερες θερμοδυναμικές ποσότητες για τον χαρακτηρισμό της κινητήριας δύναμης. Όπως και κάθε αυθόρμητη διεργασία, η σύνδεση πρωτεΐνης-προσδέτη προκύπτει μόνο όταν η αλλαγή στην ελεύθερη ενέργεια Gibbs ( $\Delta G$ ) του συστήματος είναι αρνητική όταν το σύστημα φτάσει σε κατάσταση ισορροπίας σε σταθερή πίεση και θερμοκρασία.

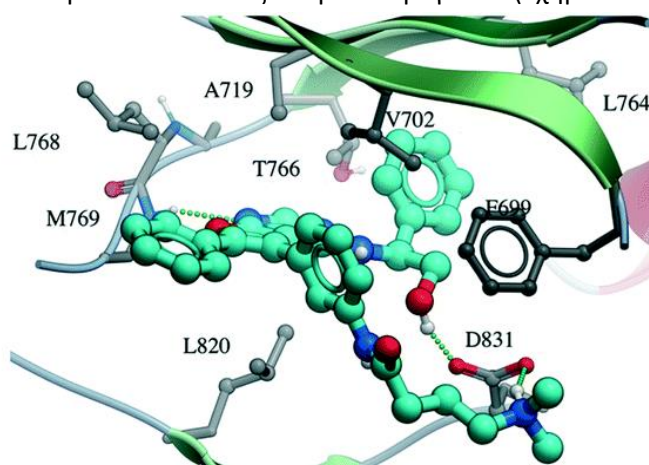
Η ελεύθερη ενέργεια πρόσδεσης αποτελείται από έναν ενθαλπικό και έναν εντροπικό όρο οι οποίοι αναπαριστούν τη μεταβολή στην ενθαλπία ( $\Delta H$ ) και την εντροπία ( $\Delta S$ ) του συστήματος κατά την πρόσδεση ενός προσδέτη σε μία πρωτεΐνη. Η ενθαλπία αποτελεί μία μέτρηση της συνολικής ενέργειας του συστήματος, δηλαδή περιλαμβάνει το άθροισμα της εσωτερικής ενέργειας όλων των μορίων καθώς και το απαιτούμενο έργο για το καθορισμό του συστήματος αυτού. Ο όρος  $\Delta H$  είναι αρνητικός για εξώθερμες, δηλαδή όταν πραγματοποιείται ο σχηματισμός ενεργειακά ευνοϊκών μη ομοιοπολικών δεσμών μεταξύ ατόμων, και θετικός για ενδόθερμες διεργασίες όταν διαταράσσονται οι ενεργειακά ευνοϊκές μη ομοιοπολικές αλληλεπιδράσεις. Έτσι, μπορεί να θεωρηθεί πως ο όρος  $\Delta H$  ή η ενθαλπία πρόσδεσης, αντανακλά στην ενεργειακή μεταβολή του συστήματος όταν ο προσδέτης δεσμεύεται σε μία πρωτεΐνη. Η ενθαλπία πρόσδεσης καθορίζεται από τις αλλαγές στην ενέργεια του συστήματος οι οποίες προκύπτουν από τους σχηματισμούς των μη ομοιοπολικών διαμοριακών αλληλεπιδράσεων πρωτεΐνης-προσδέτη [15]. Οι διαμοριακές αλληλεπιδράσεις περιλαμβάνουν ιοντικές αλληλεπιδράσεις, δεσμούς υδρογόνου, δεσμούς αλογόνου, van der Waals αλληλεπιδράσεις, υδρόφοβες αλληλεπιδράσεις και π αλληλεπιδράσεις, και θα παρουσιαστούν συνοπτικά στη συνέχεια.

#### 1.3.1 Ιοντικές αλληλεπιδράσεις

Οι ιονικές αλληλεπιδράσεις συνιστούν μία κατηγορία ισχυρών δυνάμεων που περιγράφονται από το νόμο του Coulomb και προκύπτουν από την ηλεκτροστατική έλξη μεταξύ δύο ομάδων αντίθετου φορτίου. Τέτοιες αλληλεπιδράσεις παρατηρούνται μεταξύ φορτισμένων ατόμων ενός μορίου-προσδέτη και φορτισμένων ομάδων ενός υποδοχέα [14]. Εάν για παράδειγμα ο υποδοχέας είναι μια πρωτεΐνη, ιοντικές αλληλεπιδράσεις μπορούν να παρατηρηθούν μεταξύ ενός προσδέτη με κάποια από τις φορτισμένες πλευρικές αλυσίδες των αμινοξέων της πρωτεΐνης όπως οι θετικά φορτισμένες λυσίνη, ιστιδίνη και αργινίνη, και τα αρνητικά γλουταμινικό και ασπαρτικό οξύ. Μια ειδική κατηγορία



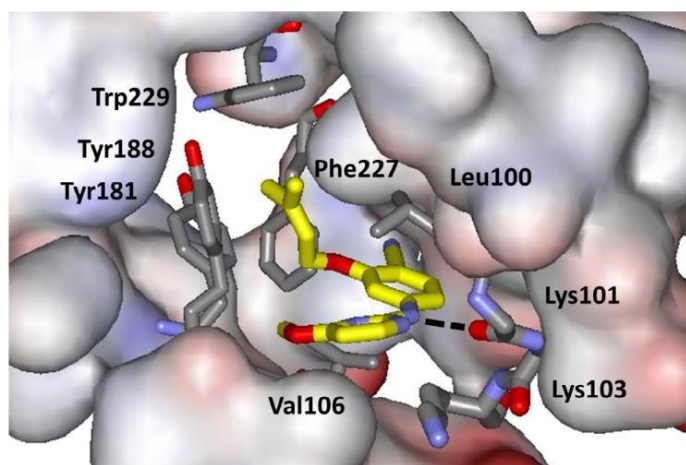
ιοντικών αλληλεπιδράσεων είναι και οι γέφυρες άλατος οι οποίες είναι ουσιαστικά ένας συνδυασμός ελκτικών ιοντικών αλληλεπιδράσεων και ενός δεσμού υδρογόνου (Σχήμα 1.12).



**Σχήμα 1.12** Γέφυρα άλατος (πράσινη διακεκομμένη γραμμή) στην EGFR κινάση μεταξύ του προσδέτη και του ασπαρτικού οξέος στη θέση 831. Ο κωδικός του συμπλόκου στη Protein Data Bank (PDB) είναι 4JRV [43].

### 1.3.2 Δεσμοί υδρογόνου

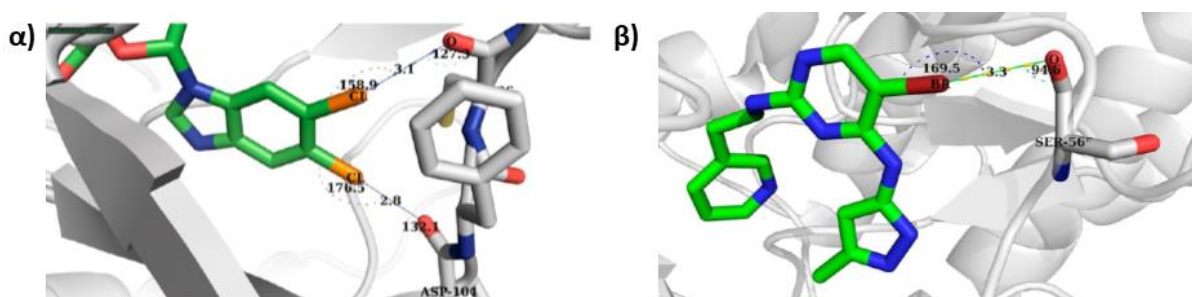
Ως δεσμός υδρογόνου ορίζεται η αλληλεπίδραση ενός ατόμου υδρογόνου (H) με ένα ηλεκτραρνητικό άτομο, όπως είναι το φθόριο (F), το οξυγόνο (O), ή το άζωτο (N). Δημιουργείται λόγω της μετατόπισης του ζεύγους ηλεκτρονίων προς το ηλεκτραρνητικό άτομο, που οδηγεί στη δημιουργία ενός τοπικού θετικού φορτίου στο H. Έπειτα, το θετικό φορτίο του υδρογόνου αλληλεπιδρά με το ηλεκτραρνητικό άτομο ενός γειτονικού μορίου. Ο δεσμός υδρογόνου χαρακτηρίζεται ως ένας ισχυρός δεσμός και απαιτείται αρκετή ενέργεια για τη διάσπασή του. Γενικά, ο δεσμός υδρογόνου αποτελεί ένα από τα σπουδαιότερα είδη αλληλεπιδράσεων καθώς συμμετέχει στη συγκρότηση, τη διαμόρφωση και τη σταθεροποίηση σύνθετων αρχιτεκτονικών δομών, όπως η δομή μιας πρωτεΐνης σε σύμπλοκο με τον αναστολέα (Σχήμα 1.13).



**Σχήμα 1.13** Η απεικόνιση από την κρυσταλλική δομή του ενός πιθανού αναστολέα της αντίστροφης μεταγραφάσης του HIV. Τα άτομα άνθρακα του αναστολέα απεικονίζονται με κίτρινο, ο δεσμός υδρογόνου του αναστολέα με την Lys101 απεικονίζεται με διακεκομμένη γραμμή, ενώ η πλευρική αλυσίδα της Lys101 έχει αφαιρεθεί για λόγους σαφήνειας [44]

### 1.3.3 Δεσμοί Αλογόνου

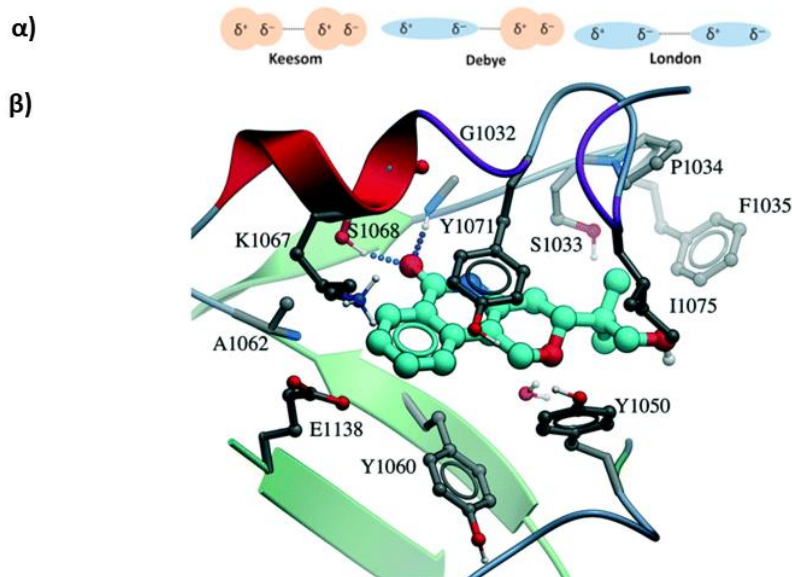
Οι δεσμοί αλογόνου μοιάζουν αρκετά με τους δεσμούς υδρογόνου, μιας και χαρακτηρίζονται από έναν δότη και έναν δέκτη ηλεκτρονίων. Στην προκειμένη περίπτωση το ρόλο του δέκτη έχει το αλογόνο αντί για το υδρογόνο. Σε αυτή την κατηγορία αλληλεπιδράσεων ένα αλογόνο δρα ως ηλεκτρονιόφιλο και σχηματίζει μια ασθενή ηλεκτροστατική αλληλεπίδραση με ένα πυρηνόφιλο άτομο. Το πυρηνόφιλο αυτό άτομο τείνει να είναι έντονα ηλεκτραρνητικό (όπως τα άτομα F, O, N) ή ακόμα και κάποιο ανιόν. Τα αλογόνα που συμμετέχουν στο σχηματισμό δεσμών αλογόνου είναι το ιώδιο (I), το βρώμιο (Br), το χλώριο (Cl) και μερικές φορές το φθόριο (F). Αυτά τα αλογόνα μπορούν να δράσουν ως δότες δεσμού αλογόνου (δέκτες ηλεκτρονίου), ωστόσο ακολουθούν μια σειρά σχετικά με την ισχύ των δεσμών που σχηματίζουν σύμφωνα με την οποία ισχύει ότι  $F < Cl < Br < I$ , με το ιώδιο να σχηματίζει τους ισχυρότερους δεσμούς. Συνήθως, η εισαγωγή ατόμων αλογόνου στις ενώσεις-οδηγούς χρησιμοποιείται για τη διερεύνηση των στερεοχημικών και ηλεκτρονιακών επιδράσεων τους. Στο Σχήμα 1.14 παρουσιάζονται μερικά παραδείγματα δεσμών αλογόνου σε σύμπλοκα προσδέτη-πρωτεΐνης [45].



**Σχήμα 1.14** Παραδείγματα δεσμών αλογόνου σε σύμπλοκα προσδέτη-πρωτεΐνης. **(α)** Ο προσδέτης σχηματίζει αλληλεπιδράσεις  $Cl \cdots O$  με την Cys106 και το Asp104. **(β)** Ο προσδέτης σχηματίζει αλληλεπίδραση  $Br \cdots O$  με το καρβοξυλικό οξύ της Val127. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [45].

### 1.3.4 van der Waals αλληλεπιδράσεις

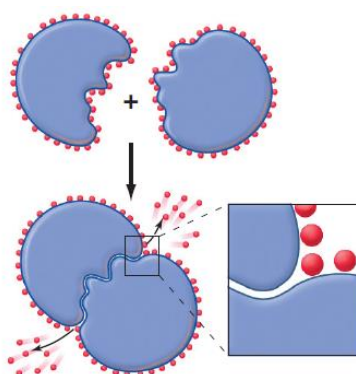
Οι δυνάμεις van der Waals (vdW) μπορούν να κατηγοριοποιηθούν με βάση τα μόρια που εμπλέκονται στην αλληλεπίδραση σε δυνάμεις Keesom, Debye και London (ή διασποράς). Οι αλληλεπιδράσεις αυτές παρουσιάζονται συνοπτικά στο Σχήμα 1.5 α. Αρχικά, οι αλληλεπιδράσεις Keesom, ή αλληλεπιδράσεις διπόλου-διπόλου, είναι ηλεκτροστατικές αλληλεπιδράσεις που εμφανίζονται μεταξύ ατόμων με διπολική ροπή. Τέτοιου είδους αλληλεπιδράσεις σχηματίζονται μεταξύ των αμινοξέων μιας πρωτεΐνης με πολικές πλευρικές αλυσίδες και των πολικών μορίων-αναστολέων. Έπειτα, οι δυνάμεις διπόλου- επαγόμενου διπόλου, ή δυνάμεις Debye, σχηματίζονται όταν ένα μη πολικό μόριο προσεγγίζεται από ένα πολικό με αποτέλεσμα την επαγωγή διπολικής ροπής στο μη πολικό μόριο. Τέλος, οι αλληλεπιδράσεις στιγμιαίου διπόλου-επαγόμενου διπόλου ονομάζονται και δυνάμεις διασποράς London και δημιουργούνται σε μη πολικά μόρια με τη στιγμιαία συσσώρευση ηλεκτρονίων σε ένα σημείο ενός μορίου έτσι ώστε το μόριο να αποκτά στιγμιαία διπολική ροπή. Αξίζει να σημειωθεί ότι παρά τον ασθενή χαρακτήρα των αλληλεπιδράσεων van der Waals, καθώς η έντασή τους μειώνεται με την αύξηση της απόστασης μεταξύ των ατόμων, ο μεγάλος αριθμός των ατόμων που αλληλεπιδρούν οδηγεί σε σημαντική συνεισφορά των δυνάμεων αυτών στη σταθερότητα του συμπλόκου ανάμεσα στην πρωτεΐνη και το μόριο-προσδέτη (Σχήμα 1.15.β) [14].



**Σχήμα 1.15 (α)** Απεικόνιση των τριών συνιστωσών των αλληλεπιδράσεων van der Waals [46]. **(β)** Οι δυνάμεις van der Waals σταθεροποιούν τον προσδέτη στην κοιλότητα πρόσδεσης της πρωτεΐνης TNKS2 (Κωδικός PDB:5C5P) [43].

### 1.3.5 Υδρόφοβες αλληλεπιδράσεις

Οι υδρόφοβες αλληλεπιδράσεις χαρακτηρίζονται ως η «επιθυμία» των μη-πολικών μορίων να συσσωματώνονται σε υδατικά διαλύματα προκειμένου να μην εκτεθούν στο νερό. Αυτό το φαινόμενο οδηγεί στην ύπαρξη μιας ελάχιστα εκτεθειμένης επιφάνειας των μη πολικών μορίων στα πολικά μόρια νερού και χρησιμοποιείται συνήθως στη βιοχημεία για να μελετήσει την πρόσδεση μικρών μορίων σε πρωτεΐνες. Οι πλευρικές αλυσίδες υδρόφοβων αμινοξέων μπορούν να αλληλεπιδρούν μεταξύ τους. Λόγω αλληλεπιδράσεων των πλευρικών αλυσίδων, η αλληλουχία και η θέση των αμινοξέων σε μία συγκεκριμένη πρωτεΐνη καθορίζουν την αναδίπλωση αυτής της πρωτεΐνης [47].



**Σχήμα 1.16** Το φαινόμενο της υδρόφοβικότητας προκύπτει όταν δύο συμπληρωματικές, μη πολικές επιφάνειες έρχονται σε επαφή και διώχνουν τα μόρια νερού που προηγουμένως συνδέονταν με αυτές. Η αυξημένη διαταραχή του νερού αυξάνει την εντροπία και μειώνει τη συνολική ελεύθερη ενέργεια του συστήματος. Επιπλέον, οι δυνάμεις van der Waals μεταξύ των κοντινών ατόμων στις συμπληρωματικές επιφάνειες σταθεροποιούν τις υδρόφοβες αλληλεπιδράσεις [47].

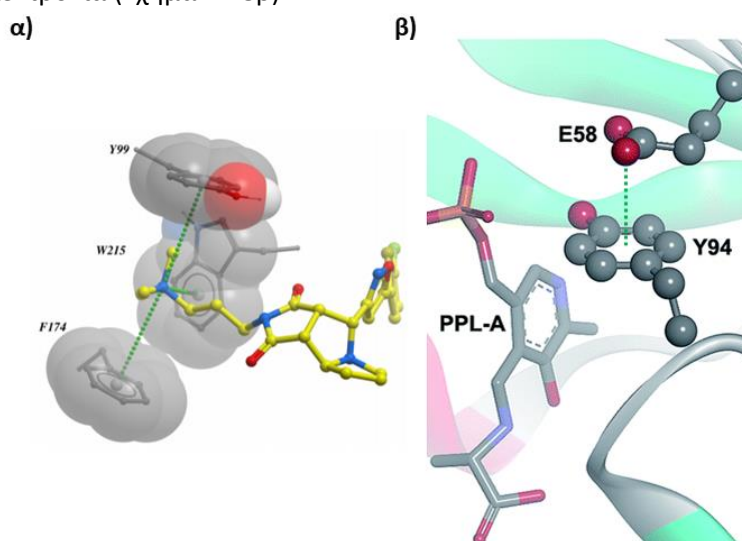
### 1.3.6 π-αλληλεπιδράσεις

Οι π αλληλεπιδράσεις αφορούν τις ελκτικές δυνάμεις που αναπτύσσονται μεταξύ των π-συστημάτων συζευγμένων μορίων και συγκεκριμένα αρωματικών δακτυλίων. Οι αλληλεπιδράσεις αυτές μπορούν να διακριθούν στις εξής κατηγορίες: α) π-π, κατιόν-π, β) ανιόν-π και γ) δίπολο-π. Οι π-π αλληλεπιδράσεις αναπτύσσονται μεταξύ των π-τροχιακών ενός μοριακού συστήματος. Για παράδειγμα, δύο γειτονικά άτομα βενζολίου μπορούν να διαταχθούν στο χώρο σε τρεις πιθανές διαμορφώσεις, σχηματίζοντας π-π αλληλεπιδράσεις. Οι δύο από αυτές είναι ενεργειακά ευνοϊκές και ονομάζονται displaced και edge-to-face, ενώ η τρίτη, διαμόρφωση sandwich, δεν είναι σταθερή εξαιτίας υψηλής ηλεκτροστατικής άπωσης των ηλεκτρονίων στα π τροχιακά των δύο δακτυλίων (Σχήμα 1.17) [48].



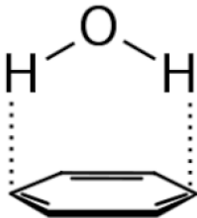
**Σχήμα 1.17** π-π αλληλεπιδράσεις. Από τις τρεις διαμορφώσεις μόνο οι δύο ευνοούνται ενεργειακά, η displaced και η edge-to-face. Η διαμόρφωση sandwich δεν συναντάται γιατί τα ηλεκτρονιακά νέφη στα κέντρα των βενζολικών δακτυλίων απωθούν το ένα το άλλο. Ανάμεσα στις δύο πρώτες διαμορφώσεις η edge-to-face είναι η πιο ευνοϊκή ενώ η displaced συναντάται μόνο όταν η edge-to-face δεν μπορεί να ληφθεί λόγω γεωμετρίας.

Οι αλληλεπιδράσεις κατιόν-π συναντώνται όταν το θετικό φορτίο ενός κατιόντος αλληλοεπιδρά με τα ηλεκτρόνια του π-συστήματος ενός μορίου. Αυτό το είδος αλληλεπίδρασης είναι αρκετά ισχυρό, πιο ισχυρό και από το δεσμό υδρογόνου ανάμεσα σε δύο μόρια (Σχήμα 1.18α) Οι ανιόν-π αλληλεπιδράσεις αφορούν την περίπτωση όπου ένα ανιόν βρίσκεται πάνω από π-σύστημα φτωχό σε ηλεκτρόνια, που συνήθως δημιουργείται με την τοποθέτηση στο συζευγμένο μόριο υποκαταστατών που αποσύρουν ηλεκτρόνια (Σχήμα 1.18β)



**Σχήμα 1.18** Με πράσινη διακεκομμένη γραμμή παρουσιάζεται η αλληλεπίδραση (α) κατιόν-π μεταξύ του τεταρτοταγούς αμμωνίου ενός προσδέτη και των αρωματικών αμινοξέων της πρωτεΐνης (PDB: 2JKH) [43] και (β) η αλληλεπίδραση ανιόν-π μεταξύ του Glu58 και της Trp94 (PDB: 1WC7) [48].

Τέλος, οι αλληλεπιδράσεις διπόλου-π, σχηματίζονται ανάμεσα σε πολικά μόρια όπως το νερό και σε π-συστήματα αρωματικών δακτυλίων, όπως το βενζόλιο (Σχήμα 1.19). Κάθε μόριο που έχει έναν δότη δεσμών υδρογόνου, σχηματίζει ηλεκτροστατικές αλληλεπιδράσεις με το πλούσιο σε ηλεκτρόνια π-σύστημα ενός συζευγμένου μορίου.



**Σχήμα 1.19** Αλληλεπιδράσεις διπόλου -π μεταξύ ενός μορίου νερού και ενός βενζολίου.

## 1.4 Οι προκλήσεις για τη στατιστική αξιολόγηση της ελεύθερης ενέργειας (SAMPL challenges)

### 1.4.1 Επισκόπηση των προκλήσεων SAMPL

Οι προκλήσεις SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) είναι μία σειρά παγκόσμιων προκλήσεων που απευθύνονται στην κοινότητα της υπολογιστικής χημείας και αποσκοπούν στην εξέλιξη των υπολογιστικών τεχνικών που χρησιμοποιούνται στον ορθολογικό σχεδιασμό φαρμάκων. Οι μεθοδολογίες για το σχεδιασμό φαρμάκων με τη βοήθεια υπολογιστή έχουν βελτιωθεί σημαντικά με την πάροδο του χρόνου χάρη στην ταχεία ανάπτυξη των Υπολογιστών Υψηλής Απόδοσης (High Performance Computing, HPC). Ωστόσο, η εφαρμογή τους στη φαρμακοβιομηχανία εξακολουθεί να είναι περιορισμένη, κυρίως λόγω της ανεπαρκούς ακρίβειας και αναπαραγωγιμότητας που παρουσιάζουν [12], [20].

Για να ξεπεραστεί αυτό το πρόβλημα, οι προκλήσεις του SAMPL οργανώθηκαν ως «τυφλές» δοκιμές (blind challenges). Πιο συγκεκριμένα παρέχεται ένα ευρύ φάσμα βιολογικά σχετικών συστημάτων καθένα από τα οποία χαρακτηρίζεται από διαφορετικά μεγέθη και δυσκολίες, συμπεριλαμβανομένων πρωτεϊνών, συμπλόκων υπερμοριακών δεκτών με μικρά μόρια προσδέτες (τα οποία συχνά καλούνται συστήματα host-guest) και μικρών φαρμακευτικών μορίων. Οι συμμετέχοντες καλούνται να εκτιμήσουν υπολογιστικά ορισμένες σχετικές θερμοδυναμικές ιδιότητες του συστήματος όπως την ελεύθερη ενέργεια πρόσδεσης, την ελεύθερη ενέργεια επιδιάλυτωσης και τους συντελεστές κατανομής. Τα πειραματικά δεδομένα για τα μεγέθη αυτά, έχουν συλλεχθεί από ακαδημαϊκά ή βιομηχανικά ερευνητικά εργαστήρια και οι μετρήσεις εκδίδονται λίγο μετά την προθεσμία υποβολής, αποκαλύπτοντας έτσι την πραγματική εκτιμητική ισχύ των μεθοδολογιών που χρησιμοποιήθηκαν. Οι ερευνητές μπορούν στη συνέχεια να συγκρίνουν τα πειραματικά δεδομένα με τις υπολογιστικές εκτιμήσεις. Βασική έμφαση δίδεται στα διδάγματα που αντλήθηκαν, επιτρέποντας έτσι στους συμμετέχοντες των μελλοντικών προκλήσεων να επωφεληθούν από τα συμπεράσματα και τις προτάσεις των προγενέστερων προκλήσεων [20].

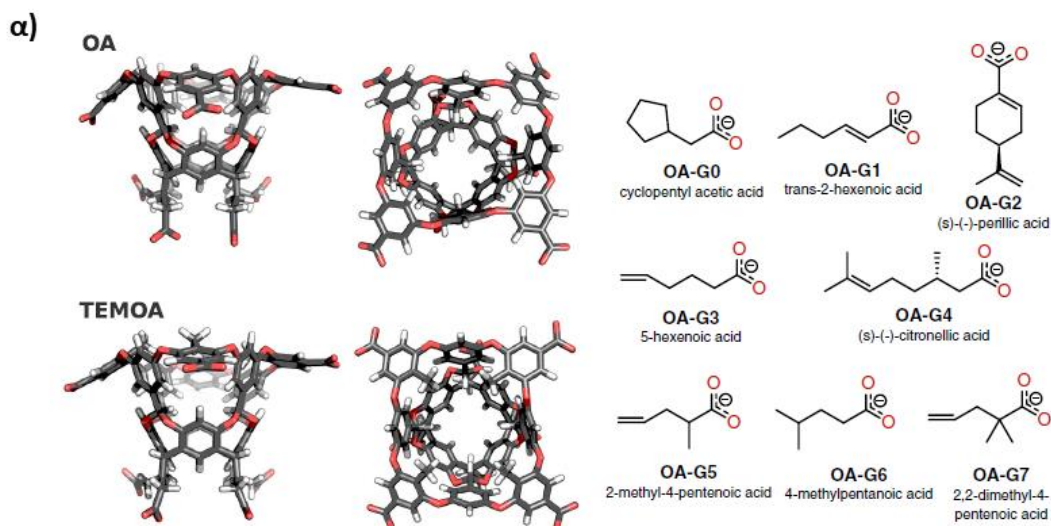
Από το 2008, οι προκλήσεις SAMPL έχουν προσελκύσει το ενδιαφέρον επιστημόνων που ασχολούνται με τον CADD σε όλο τον κόσμο. Οι προκλήσεις SAMPL επικεντρώνονται ιστορικά στις ιδιότητες των συστημάτων προσδετών-δεκτών και στα μικρά μόρια που μοιάζουν με φάρμακα. Αυτά τα απλά μοντέλα απαιτούν σημαντικά λιγότερους υπολογιστικούς πόρους για την προσομοίωση, σε σύγκριση με συστήματα πρωτεϊνών, και έτσι επιτρέπουν πολύ ταχύτερη σύγκλιση. Παλαιότερες προκλήσεις σε τέτοια συστήματα αποκάλυψαν τους περιορισμούς και τα ελαττώματα των σύγχρονων τεχνικών μοριακής μοντελοποίησης, διευκόλυναν την ανάπτυξη μοντέλων διαλυτών και υπογράμμισαν τη σημασία του σωστού χειρισμού των καταστάσεων πρωτονίωσης και της επίδρασης των αλάτων [49], [50].

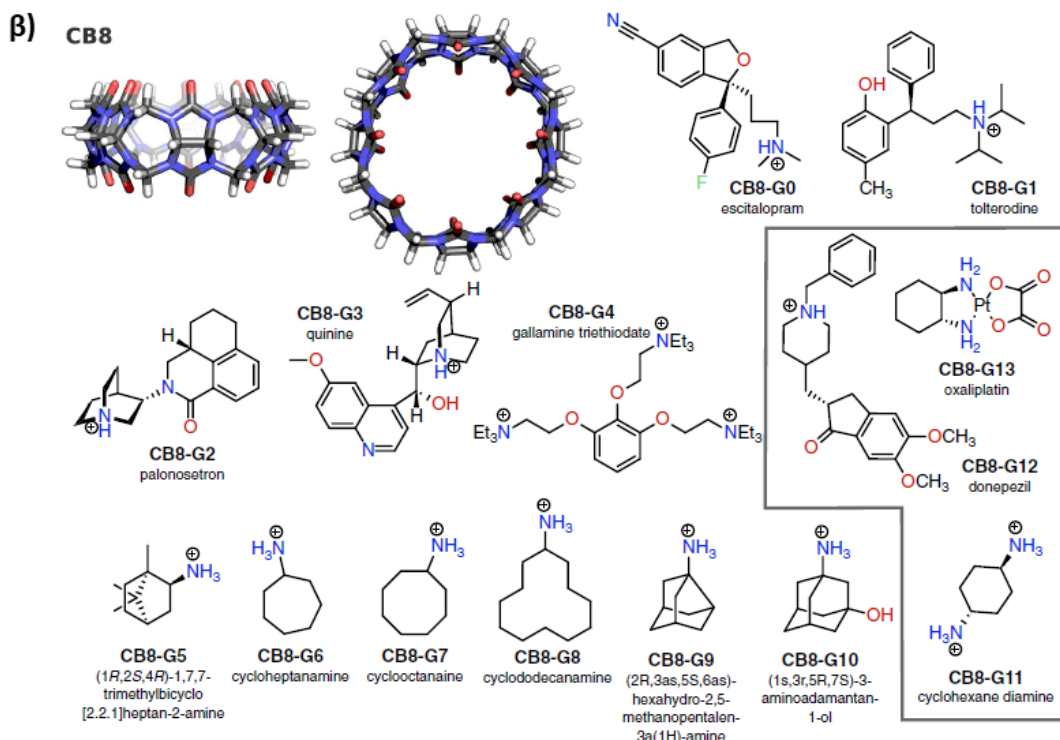
Η πρώτη πρόκληση SAMPL, SAMPL0 (2008) επικεντρώθηκε στις προβλέψεις των ελεύθερων ενεργειών επιδιалύτωσης 17 μικρών μορίων. Συγκεκριμένα, μια ερευνητική ομάδα στο Πανεπιστήμιο του Stanford και επιστήμονες της εταιρείας OpenEye Scientific Software πραγματοποίησαν τους υπολογισμούς. Παρά την ανεπίσημη μορφή, το SAMPL0 έθεσε τις βάσεις για τις ακόλουθες προκλήσεις SAMPL [51]. Οι προκλήσεις SAMPL1 (2009) και SAMPL2 (2010) οργανώθηκαν από την OpenEye και συνέχισαν να επικεντρώνονται στην πρόβλεψη των ελεύθερων ενεργειών επιδιалύτωσης μικρών μορίων που μοιάζουν με φάρμακα. Προσπάθειες έγιναν επίσης για την πρόβλεψη ελεύθερων ενεργειών πρόσδεσης, πιθανών τρόπων δέσμευσης φαρμάκων σε πρωτεΐνες καθώς και αναλογιών ταυτομερών [52], [53]. Τα «τυφλά» σύνολα δεδομένων για τις ελεύθερες ενέργειες πρόσδεσης συστημάτων προσδέτη-δέκτη εισήχθησαν για πρώτη φορά στην πρόκληση SAMPL3 (2011-2012) αποτελώντας μια προσθήκη στον υπολογισμό ελεύθερων ενεργειών επιδιалύτωσης μικρών μορίων καθώς και τον υπολογισμό ελεύθερων ενεργειών πρόσδεσης για 500 αναστολείς που μοιάζουν με την τυροσίνη. Τα τρία μόρια δέκτες (hosts) που χρησιμοποιήθηκαν ήταν από την οικογένεια κυκλικών μακρομορίων cucurbituril. Η πρόκληση SAMPL3 έλαβε 103 υποβολές από 23 ερευνητικές ομάδες παγκοσμίως [54]. Διαφορετικά από τις προηγούμενες προκλήσεις SAMPL, η πρόκληση SAMPL4 (2013-2014) συντονίστηκε από ακαδημαϊκούς ερευνητές, με υλικοτεχνική υποστήριξη από την εταιρεία Open Eye. Τα δεδομένα του διαγωνισμού SAMPL4 αφορούσαν ελεύθερες ενέργειες πρόσδεσης συστημάτων προσδέτη-δέκτη και αναστολείς ιντεγκράσης του HIV, καθώς επίσης και από ελεύθερες ενέργειες επιδιалύτωσης μικρών μορίων. Τα μόρια-δέκτες συμπεριλάμβαναν το cucurbit[7]uril (CB7) και τα οκτα-οξέα (OA). Η πρόκληση επιδιалύτωσης SAMPL4 έλαβε 49 υποβολές από 19 ομάδες [55]. Στην πρόκληση SAMPL5 (2015-2016), διαχωρίστηκαν οι προκλήσεις πρωτεΐνης-προσδέτη, ενώ οι συμμετέχοντες κλήθηκαν να προβλέψουν τις ελεύθερες ενέργειες πρόσδεσης τριών συστημάτων προσδετών-δεκτών, τα οποία περιλάμβαναν ένα ακυκλικό παράγωγο CB7 και δύο δέκτες από την οικογένεια οκτα-οξέων. Επίσης, οι συμμετέχοντες ενθαρρύνθηκαν να υποβάλουν προβλέψεις για τον υπολογισμό της ενθαλπίας πρόσδεσης. Κατά την πρόκληση δοκιμάστηκε ένα ευρύ φάσμα υπολογιστικών προσεγγίσεων, συμπεριλαμβανομένης της θεωρίας Density Function Theory (DFT), της Μοριακής Δυναμικής, της Μοριακής Πρόσδεσης και της μεθόδου metadynamics. Οι μέθοδοι που εμφάνισαν την υψηλότερη απόδοση στην πρόκληση με τα συστήματα προσδέτη-δέκτη, έδωσαν ατελείς συσχετίσεις με πειραματικά δεδομένα, καθώς και συστηματικές αποκλίσεις σε σχέση με τα αντίστοιχα πειράματικά δεδομένα. Τέλος, στο διαγωνισμό SAMPL5 πραγματοποιήθηκαν προβλέψεις του συντελεστή κατανομής οκτανόλης/νερού (logP) για 53 μικρά μόρια και συνολικά λήφθηκαν 76 υποβολές από 18 ομάδες ερευνητών [50].

## 1.4.2 Ο διαγωνισμός SAMPL6 SAMPLing

Ο διαγωνισμός SAMPL6 ξεκίνησε τον Σεπτέμβριο του 2017 και χωρίζεται σε δύο φάσεις. Το πρώτο μέρος του διαγωνισμού αφορά την πρόβλεψη φυσικοχημικών ιδιοτήτων όπως η σταθερά διάστασης οξέων ( $pK_a$ ) και οι συντελεστές κατανομής οκτανόλης/νερού, για μία σειρά συστημάτων προσδέτη-δέκτη. Το δεύτερο μέρος του διαγωνισμού SAMPL6 περιλαμβάνει την πρόκληση SAMPLing. Ο σκοπός της πρόκλησης αυτής είναι η αξιολόγηση και η σύγκριση των επιδόσεων διαφορετικών μεθοδολογιών για τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης τριών συστημάτων προσδετών-δεκτών τα οποία προέρχονται από την πρώτη φάση της πρόκλησης SAMPL6. Για το σκοπό αυτό θα εκτελεστούν μεγάλες προσομοιώσεις, της τάξης των εκατοντάδων ns ανά σύστημα προσδέτη-δέκτη, με διαφορετικές μεθοδολογίες. Στη συνέχεια θα γίνει προσπάθεια αξιολόγησης των αποτελεσμάτων ως προς το πόσο συγκλίνουν οι διαφορετικές μεθοδολογίες των συμμετεχόντων. Τα συστήματα της ευρύτερης πρόκλησης SAMPL6 περιλαμβάνουν τους μακρομοριακούς δέκτες cucurbit[8]uril (CB8), οκτα-οξύ (OA) και τετρα-ενδο-μεθυλο-οκτα-οξύ (TEMOA) καθώς και μια σειρά μικρών φαρμακευτικών μορίων τα οποία συνοψίζονται στο Σχήμα 1.20 [49].

Σε γενικές γραμμές η πρόκληση SAMPLing είναι αρκετά πειραματική, καθώς ενδέχεται οι διαφορετικές μεθοδολογίες να μην συμφωνούν ακόμη και όταν έχει επιτευχθεί η σύγκλιση της ελεύθερης ενέργειας για κάθε μέθοδο. Έτσι, απαιτείται η συνεργασία μεταξύ των διάφορων ομάδων προκειμένου να εντοπισθούν τα αίτια της απόκλισης. Εάν επιτευχθεί συμφωνία, τότε θα συγκριθεί το ποσοστό σύγκλισης μεταξύ των διάφορων μεθόδων, ενώ ταυτόχρονα θα διερευνηθεί ποιες από αυτές ήταν πιο αποτελεσματικές. Η ανάλυση αυτής της πρόκλησης θα επικεντρωθεί περισσότερο από το συνηθισμένο στα διδάγματα που μπορούμε να λάβουμε από αυτή, παρά στο ποιες από τις μεθόδους απέδωσαν καλύτερα [49], [56].





**Σχήμα 1.20** Παρουσίαση του συνόλου δεδομένων για τα συστήματα προσδέτη-δέκτη της πρόκλησης SAMPL6. Οι τρισδιάστατες δομές των τριών δεκτών (OA, TEMOA και CB8) παρουσιάζονται στην άνω αριστερά γωνία των σχημάτων **(α)** και **(β)**. Τα άτομα άνθρακα απεικονίζονται σε γκρι, τα υδρογόνα σε λευκό, τα άτομα αζώτου σε μπλε χρώμα και τα οξυγόνα με κόκκινο χρώμα. Οι προσδέτες (guests) κάθε συστήματος δέκτη απεικονίζονται ως δισδιάστατες χημικές δομές. Οι καταστάσεις πρωτονίωσης των προσδετών αντιστοιχούν στην προβλεπόμενη κυρίαρχη μικροκατάσταση στο πειραματικό pH στο οποίο υπολογίσθηκαν οι ελεύθερες ενέργειες πρόσδεσης [49].

## 1.5 Η περίπτωση του πρωτεϊνικού συμπλέγματος Arp2/3

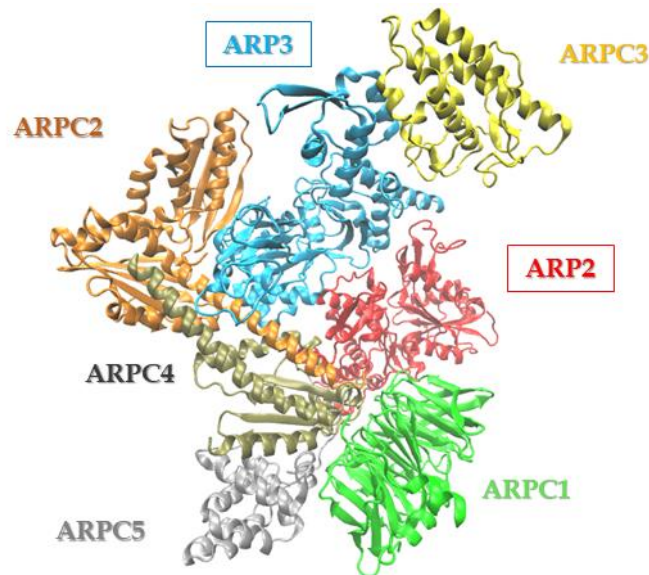
### 1.5.1 Δομή

Οι πρωτεΐνες Arp2 και Arp3 είναι δύο μέλη μιας οικογένειας πρωτεϊνών που σχετίζονται με τον πολυμερισμό των ινιδίων ακτίνης. Τα ινίδια ακτίνης είναι πολυμερικές δομές αποτελούμενες από μονομερή της πρωτεΐνης ακτίνης τα οποία συναντώνται σε όλα τα ευκαρυωτικά κύτταρα και διαδραματίζουν σημαντικό ρόλο σε πολλές κινήσεις τους. Οι Arp2 και Arp3 συνδέονται μαζί με άλλες πέντε πρωτεΐνες για τον σχηματισμό του πρωτεϊνικού συμπλέγματος Arp2/3, το οποίο συναντάται σε μεγάλο εύρος οργανισμών, από τις ζύμες ως τα θηλαστικά κύτταρα. Το σύμπλεγμα Arp2/3 απομονώθηκε αρχικά από τον μικροοργανισμό *Acanthamoeba castellanii* και αποδείχθηκε ότι αποτελείται από μία σταθερή συγκρότηση επτά υπομονάδων: τις Arp2 και Arp3 και πέντε μικρότερες πρωτεΐνες (ARPC1, ARPC2, ARPC3, ARPC4 και ARPC5) [57].

Οι πρώιμες προσπάθειες για την απόκτηση πληροφοριών για τη δομή του συμπλέγματος Arp2/3 βασίστηκαν στη γενετική, τη βιοχημική ανασύσταση και τη χημική διασταύρωση *in vitro*. Οι λεπτομέρειες της δομής του συμπλέγματος Arp2/3 αποκαλύφθηκαν από την ομάδα του Robinson το 2001, όταν επιλύθηκε η κρυσταλλική δομή της πρωτεΐνης σε ανάλυση 2.0 Å [58]. Το σύμπλεγμα



Arp2/3 που απομονώθηκε στον κρύσταλλο θεωρείται πως βρίσκεται σε αδρανή διαμόρφωση για δύο κύριους λόγους. Αρχικά, στο σύμπλεγμα δεν είναι δεσμευμένος κανένας από τους παράγοντες που είναι απαραίτητοι για την ενεργοποίησή του. Επιπλέον, οι Arp2 και Arp3 βρίσκονται πολύ μακριά η μία από την άλλη, το οποίο είναι ασυμβίβαστο με την κύρια υπόθεση που αναφέρει πως οι δύο πρωτεΐνες βρίσκονται σε κοντινή απόσταση, λειτουργώντας ως ένα ετεροδιμερές που δρα ως πυρήνας συσσωμάτωσης μονομερών ακτίνης. Στο Σχήμα 1.20 παρουσιάζεται η δομή του πρωτεϊνικού συμπλέγματος Arp2/3 [57].



**Σχήμα 1.21** Αναπαράσταση της οργάνωσης του Arp2/3 συμπλέγματος σε υπομονάδες. Κρυσταλλική δομή του πρωτεϊνικού συμπλέγματος Arp2/3 όπως απομονώθηκε από τον οργανισμό *bos taurus* (Κωδικός PDB: 1A8K) [58].

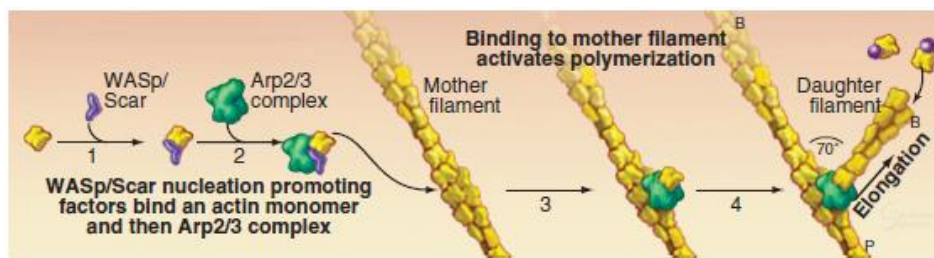
### 1.5.2 Βιολογική Λειτουργία

Ο κυτταροσκελετός της ευκαρυωτικής ακτίνης παίζει σημαντικό ρόλο σε διαφορετικές κυτταρικές διεργασίες που είναι απαραίτητες για την επιβίωση του κυττάρου όπως η κυτταρική μετανάστευση, η ενδοκυττάρωση, η κυκλοφορία των κυστιδίων και η κυτταροκίνηση. Το κύριο συστατικό του κυτταροσκελετού της ακτίνης είναι η μονομερής σφαιρική (globular) G-ακτίνη, μια ATPάση μοριακού βάρους 43-kDa που μπορεί να αυτοσυναρμολογηθεί σε νηματοειδή (filamentous) F-ακτίνη. Επειδή τα μονομερή ακτίνης προσανατολίζονται όλα προς την ίδια κατεύθυνση, τα ινίδια έχουν διαφορετική πολικότητα στα δύο άκρα τους. Γι' αυτό το λόγο διαχωρίζονται μεταξύ τους με τους όρους (+) άκρο (ή αγκαθωτό άκρο, barbed end) και (-) άκρο (ή αιχμηρό άκρο, pointed end) [57].

Η κύρια λειτουργία της Arp2/3 είναι η δημιουργία εναρκτήριων σημείων ανάπτυξης νέων διακλαδώσεων από μόρια ακτίνης στα ήδη υπάρχοντα «μητρικά» νημάτια F- ακτίνης, μία διαδικασία που απαιτεί συγκεντρώσεις ATP της τάξης του  $\mu\text{M}$ . Αξίζει να αναφερθεί ότι στην ανάπτυξη νέων ινιδίων ακτίνης συμμετέχουν και άλλες πρωτεΐνες της οικογένειας ABP (actin-binding proteins), αλλά καμία με τον τρόπο που ενεργεί η Arp2/3. Με άλλα λόγια, καμία άλλη πρωτεΐνη δε δημιουργεί διακλαδισμένα νημάτια ακτίνης, μία ιδιότητα στην οποία, όπως θα εξηγηθεί στη συνέχεια, οφείλει η Arp2/3 το ερευνητικό ενδιαφέρον που έχει συγκεντρωθεί γύρω της [47]. Ερεθίσματα εκτός του

κυττάρου, όπως αυξητικοί παράγοντες, μεταδίδονται μέσω της πλασματικής μεμβράνης με τη βοήθεια διαμεμβρανικών πρωτεϊνών που καλούνται GPCRs (G Protein Coupled Receptors), όπως ο Cdc42. Ακολούθως προκαλείται η ενεργοποίηση πρωτεϊνών όπως οι WASP (Wiskott-Aldrich Syndrome Protein), N-WASP (το ομόλογο της WASP στους νευρώνες), SCAR (suppressor of cyclic AMP repressor) που αποτελούν τους λεγόμενους παράγοντες προώθησης της δημιουργίας νέων ινιδίων ακτίνης (Nucleation Promoting Factors, NPFs). Η ενεργοποιημένη δομή της WASP περιέχει σημεία πρόσδεσης με το μονομερές της ακτίνης αλλά και με την Arp2/3. Όταν στην Arp2/3 προσδένεται η ενεργοποιημένη WASP τότε η πρώτη αλλάζει διαμόρφωση έτσι ώστε να δράσει ως πυρήνας πολυμερισμού «θυγατρικών» ινιδίων ακτίνης σε γωνία  $\sim 70^\circ$  σε σχέση με την ευθεία του «μητρικού» ινιδίου (Σχήμα 1.22) [57].

Κατά συνέπεια, η Arp2/3 συνδέει το (-) άκρο του νέου «θυγατρικού» ινιδίου στο «μητρικό» ινίδιο, αφήνοντας το (+) άκρο του «θυγατρικού» διαθέσιμο για επιμήκυνση. Για την ενεργοποίηση του συμπλέγματος Arp2/3 είναι απαραίτητη η ύπαρξη μιας VCA (ή WCA) περιοχής, η οποία είναι κοινή στην οικογένεια πρωτεϊνών WASP, ενός μονομερούς G-ακτίνης και ενός μητρικού ινιδίου ακτίνης. Το πιο απλό μοντέλο δράσης της VCA περιοχής είναι το εξής: η VCA περιοχή παραδίδει ένα μονομερές ακτίνης στο σύμπλεγμα Arp2/3, σχηματίζοντας ένας τριμερές των Arp2 και Arp3 υπομονάδων και του μονομερούς ακτίνης. Η δημιουργία αυτού του τριμερούς οδηγεί προς την αυθόρμητη ανάπτυξη του ινιδίου.



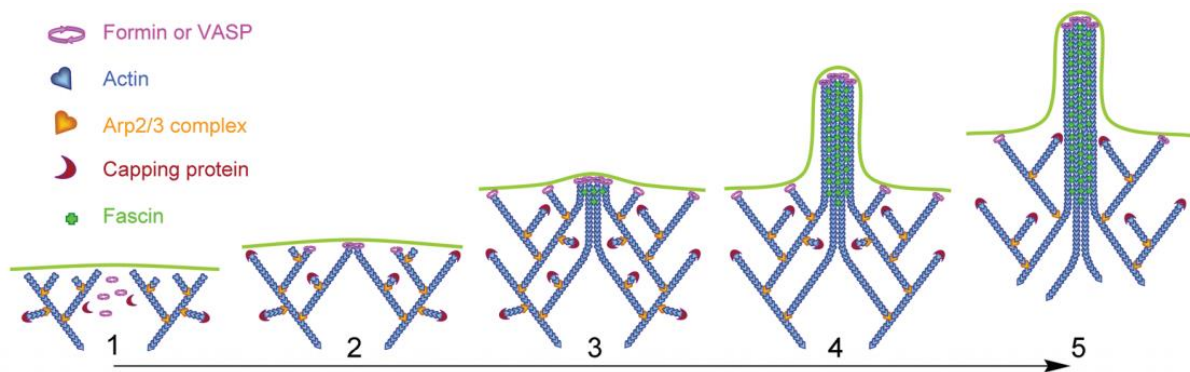
**Σχήμα 1.22** Στάδια σχηματισμού διακλαδώσεων: **(1)** Ένας NPF δεσμεύει ένα μονομερές ακτίνης. **(2)** Αυτό το σύμπλοκο δεσμεύει δύο θέσεις στο σύμπλεγμα Arp2/3, φέρνοντας μαζί δύο υπομονάδες ακτίνης με Arp2 και Arp3. **(3)** Το τριμερές σύμπλεγμα δεσμεύεται στην πλευρά ενός νήματος ακτίνης, ολοκληρώνοντας τη διαδικασία ενεργοποίησης. **(4)** Ένα καινούργιο θυγατρικό νήμα μεγαλώνει στο αγκαθωτό άκρο (+) του από την πλευρά του παλαιότερου μητρικού νήματος. Στο παραπάνω σχήμα με B και P συμβολίζεται το αγκαθωτό και το αιχμηρό άκρο, αντίστοιχα. Η εικόνα λήφθηκε από τη βιβλιογραφική πηγή [47].

### 1.5.3 Συσχέτιση του συμπλέγματος Arp2/3 με τη μετάσταση του καρκίνου

Η μετανάστευση κυττάρων (cell migration) παρατηρείται σε πολλές βιολογικές διαδικασίες, όπως η εμβρυονική μορφογένεση, η ανοσοποιητική απόκριση και η αποκατάσταση και ανάπτυξη ιστών. Μη φυσιολογική ρύθμιση της μετανάστευσης κυττάρων οδηγεί σε επιδείνωση πολλών ασθενειών, όπως σε διήθηση, ενδαγγείωση και μετάσταση του καρκίνου. Η μετανάστευση κυττάρων απαρτίζεται από δύο βασικές διαδικασίες: 1) την προσκόλληση του μεταναστεύοντος κυττάρου σε ένα συγκεκριμένο υπόστρωμα και 2) τη δημιουργία προεκβολών στην κυτταρική μεμβράνη, όπως λαμελλιπόδια (lamellipodia), φιλοπόδια (filopodia), ποδοσωμάτια (podosomes) και ινβαδοπόδια (invadopodia). Σύμφωνα με μελέτες έχει παρατηρηθεί έντονη συσχέτιση της Arp2/3 με τη δημιουργία αυτών των προεκβολών [57]. Μία ένδειξη που συνδέει ισχυρά την Arp2/3 με διάφορα είδη καρκίνου είναι η

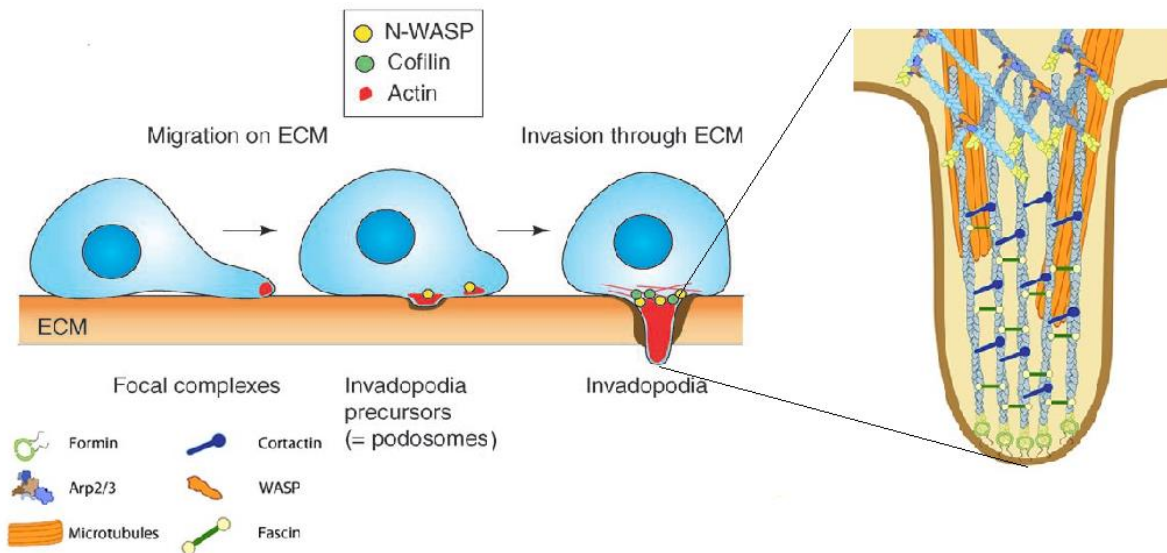
υπερέκφραση του γονιδίου που την κωδικοποιεί στα καρκινικά κύτταρα. Συγκεκριμένα, in vivo πειράματα καθώς και κλινικές μελέτες σε καρκίνους της ουροδόχου κύστης, του μαστού και του παχέος εντέρου υπέδειξαν ότι τα γονίδια του συμπλέγματος Arp2/3 καθώς και των πρωτεϊνών CP (capping proteins) και κοφιλίνη, οι οποίες σχετίζονται με τη δραστηριότητα της Arp2/3, βρίσκονταν σε υπερέκφραση [59], [60], [61].

Τα ινίδια ακτίνης βρίσκονται πολύ κοντά στην κυτταρική μεμβράνη και η προσθήκη νέων μονομερών ακτίνης στις διακλαδώσεις των ινιδίων προκαλεί πίεση στην κυτταρική μεμβράνη με αποτέλεσμα την ώθησή της προς το εξωτερικό του κυττάρου. Αυτό έχει ως αποτέλεσμα τη δημιουργία προεκβολών, όπως τα λαμελλιπόδια και τα φιλοπόδια, στην εξωτερική επιφάνεια του κυττάρου οι οποίες έχουν συσχετιστεί με την ικανότητα του κυττάρου να κινείται (Σχήμα 1.22) [62].



**Σχήμα 1.23** Μοντέλο έναρξης σχηματισμού φιλοποδίων: **(1)** Σχηματισμός διακλαδισμένου δικτύου ακτίνης στο λαμελλοπόδιο από την Arp2/3. **(2)** Άλλες πρωτεΐνες οι οποίες ονομάζονται συντελεστές επιμήκυνσης (VASP ή formin) διατηρούν τη συνεχή επιμήκυνση ορισμένων άκρων. **(3)** Η αλληλεπίδραση μεταξύ των αγκαθωτών (-) άκρων, πιθανώς επίσης με τη μεσολάβηση των VASP ή formin, έχει ως αποτέλεσμα τη συγχρονισμένη παράλληλη επιμήκυνση πολλών συγγενών νηματιών. **(4)** Τα παράλληλα νηματία είναι διασταυρωμένα συνδεδεμένα με μία άλλη πρωτεΐνη την fascin με αποτέλεσμα το σχηματισμό δέσμης ακτίνης. **(5)** Με την πάροδο του χρόνου, το σύμπλεγμα Arp2/3 που πυρηνοποιείται τα νηματοειδή νήματα διασπάται αφήνοντας τα ελεύθερα αιχμηρά άκρα του συσσωματωμένου νήματος ακτίνης [62].

Επιπλέον, τα καρκινικά κύτταρα προκειμένου να κινηθούν μέσω του φυσικού εμποδίου που θέτει η εξωκυττάρια ουσία ή μήτρα (extracellular matrix, ECM), εκβάλλουν ινβαδοπόδια τα οποία έχουν την ικανότητα να αποδομούν και να επαναπροσδιορίζουν τη δομή της ECM. Αυτός ο τύπος προεκβολής είναι πολύ σημαντικός για τη διήθηση των καρκινικών κυττάρων μέσω της μεμβράνης που καλύπτει τα αγγεία του αίματος και την είσοδό τους στην κυκλοφορία. Τα ινβαδοπόδια είναι πλούσια σε ακτίνη και εκτείνονται κάθετα προς την κοιλότητα την κυτταρικής μεμβράνης. Τα ποδοσωμάτια είναι δυναμικές, πλούσιες σε ακτίνη δομές, με δυνατότητα προσκόλλησης στους τριγύρω ιστούς, τα οποία παρουσιάζουν ομοιότητα με τα ινβαδοπόδια. Έχει προταθεί ότι τα ποδοσωμάτια είναι πρόδρομες δομές οι οποίες ωριμάζουν σε πιο σταθερά ινβαδοπόδια κάτω από συγκεκριμένες συνθήκες. Στο Σχήμα 1.24 φαίνεται πως το σύμπλεγμα Arp2/3 συμβάλλει στη δημιουργία των ινβαδοποδίων [59].



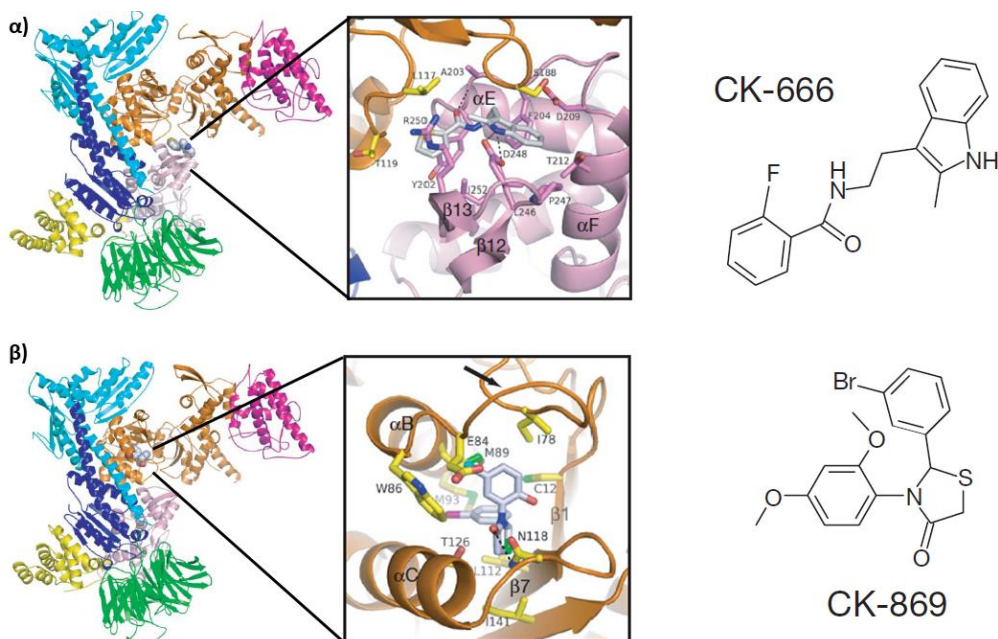
**Σχήμα 1.24** Μοντέλο της μετανάστευσης των καρκινικών κυττάρων πάνω και διαμέσου της ECM. Έντονα κινητικά καρκινικά κύτταρα χρησιμοποιούν δυναμικές δομές προσκόλλησης στην ECM, που καλούνται εστιακά συμπλέγματα (focal complexes), καθώς και ποδοσωμάτια και ινβαδοπόδια για διείσδυση στην ECM. Οι πρωτεΐνες N-WASP και κοφιλίνη πυροδοτούν την ενεργοποίηση της Arp2/3 ώστε να ξεκινήσει τον πολυμερισμό των διακλαδισμένων ινιδίων ακτίνης στη βάση των προεκβολών [59].

#### 1.5.4 Αναστολή του πρωτεϊνικού συμπλέγματος Arp2/3

Λαμβάνοντας υπόψιν το πως η Arp2/3 συμβάλλει στην μετανάστευση των καρκινικών κυττάρων έγινε αντιληπτό πως οι αναστολείς του συμπλέγματος θα εξυπηρετούν ως πιθανοί αντικαρκινικοί παράγοντες. Ο B.J. Nolen και οι συνεργάτες του [63] ανακάλυψαν δύο νέα μόρια αναστολείς του συμπλέγματος Arp2/3, με τα κωδικά ονόματα CK-636 και CK-548 μέσω σάρωσης μιας βιβλιοθήκης 400.000 πιθανών μορίων. Πειράματα στην πρωτεΐνη Arp2/3 του οργανισμού *Bos taurus* έδειξαν ότι τα μόρια ανέστειλαν τη δράση του συμπλόκου με τις τιμές  $IC_{50}^2$  να είναι 32  $\mu$ M και 11  $\mu$ M για το CK-636 και το CK-548 αντίστοιχα [63].

Στη συνέχεια, χρησιμοποίησαν τον μικροοργανισμό *Listeria monocytogenes* για την εύρεση πιο ισχυρών αναστολέων σε σχέση με τους CK-636 και CK-548. Η έρευνα απέδωσε καθώς βρέθηκαν δύο νέοι πιο δραστικοί αναστολείς με τα κωδικά ονόματα CK-666 (2-Fluoro-N-[2-(2-methyl-1H-indol-3-yl)ethyl]-benzamide) και CK-869 (2-(3-Bromophenyl)-3-(2,4-dimethoxyphenyl)-4-thiazolidinone). Ο CK-666 αναστέλλει τη δράση της Arp2/3 του οργανισμού *Bos taurus* με τιμή  $IC_{50}$  12 $\mu$ M [64]. Ο CK-869 παρουσιάζει ίδιας κλίμακας αναστολή της βοοειδούς Arp2/3, ωστόσο αναστέλλει πιο αποτελεσματικά τον πολυμερισμό της ακτίνης στη *Listeria monocytogenes*. Είναι σημαντικό να αναφερθεί πως όλοι οι αναστολείς παρουσιάζουν εκλεκτικότητα προς την Arp2/3 καθώς αλληλεπιδρούν με αυτή και όχι με τη G- ακτίνη, τη WASP ή τη N-WASP. Το σύμπλοκο της Arp2/3 μαζί με τους αναστολείς CK-666 και CK-869, καθώς και η δισδιάστατη δομή αυτών παρουσιάζονται στο Σχήμα 1.25 [63].

<sup>2</sup> Η ποσότητα  $IC_{50}$  ονομάζεται μέση ανασταλτική συγκέντρωση (half maximal inhibitory concentration) και υποδηλώνει την ποσότητα της ουσίας η οποία είναι απαραίτητη για να προκαλέσει 50% αναστολή της δράσης ενός φαρμακευτικού στόχου

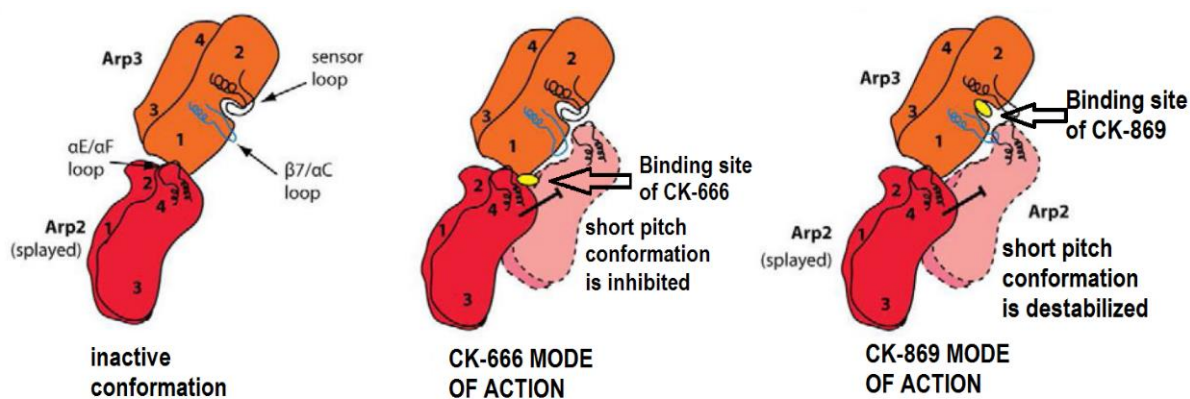


**Σχήμα 1.25** Κρυσταλλικές δομές του συμπλόκου Arp2/3 με τους αναστολείς CK-666 (α) και CK-869 (β). Χρώματα: Arp3 πορτοκαλί, Arp2 ροζ, ARPC1 πράσινο, ARPC2 γαλάζιο, ARPC3 φούξια, ARPC4 μπλε. Στο δεξιό μέλος κάθε σχήματος παρουσιάζεται λεπτομερώς οι λεπτομέρειες της θέσης πρόσδεσης του κάθε αναστολέα. Οι κωδικοί PDB για τις κρυσταλλικές δομές του συμπλόκου της Arp2/3 με τους CK-666 και CK-869 είναι 3UKR και 3UKU, αντίστοιχα [63].

### 1.5.5 Μηχανισμός δράσης υπαρχόντων αναστολέων

Κάθε αναστολέας υιοθετεί διαφορετικό μηχανισμό απενεργοποίησης της Arp2/3. Ο CK-666 προσδένεται στη διεπιφάνεια ανάμεσα στις υπομονάδες Arp2 και Arp3, η οποία υφίσταται μόνο στην ανενεργή μορφή της Arp2/3, αποτρέποντας την κίνηση της Arp2 προς την Arp3 που είναι απαραίτητη για την ενεργοποίηση του συμπλέγματος. Με αυτό τον τρόπο ο CK-666 εμποδίζει τη λήψη της ενεργούς short-pitch διαμόρφωσης από την Arp2/3 και ταυτόχρονα σταθεροποιεί την ανενεργή «ανοιγμένη» διαμόρφωσή της (splayed conformation) [65].

Η πρόσδεση του CK-869 στην υπομονάδα Arp3, αποσταθεροποιεί τη διαδικασία προσέγγισης των δύο υπομονάδων Arp2 και Arp3, μην επιτρέποντας στην Arp2/3 να λάβει την ενεργή διαμόρφωση της. Προσδένεται στην κοιλότητα στην οποία φυσιολογικά στερεώνεται μία θηλιά (loop) της Arp2 και έμμεσα διαταράσσει δύο θηλιές μεγάλης σημασίας για τη διεπιφάνεια των Arp2 και Arp3 που υφίσταται στη ενεργή διαμόρφωση της πρωτεΐνης. Συγκεκριμένα, όταν ο CK-869 είναι προσδεμένος στην κοιλότητα της Arp3, λαμβάνει χώρα στερεοχημική παρεμπόδιση μεταξύ μίας θηλιάς της Arp2 και μίας άλλης της Arp3 [65]. Οι παραπάνω μηχανισμοί δράσης συνοψίζονται στο Σχήμα 1.26.



**Σχήμα 1.26** Μοντέλο μηχανισμού αναστολής της Arp2/3 από τους CK-666 και CK-869 αναστολείς. **Αριστερά:** Ανενεργή απλωμένη (splayed) διαμόρφωση της Arp2/3. **Κέντρο:** Ο CK-666 σταθεροποιεί την ενεργή διαμόρφωση του συμπλέγματος με την πρόσδεσή του στη διεπιφάνεια μεταξύ των Arp2 και Arp3 υπομονάδων, η οποία υπάρχει μόνο στην ανενεργή διαμόρφωση. **Δεξιά:** Ο CK-869 εξαναγκάζει μια πιο ανοιχτή διαμόρφωση της sensor θηλιάς της Arp3 και τη μετακίνηση της β7/αC θηλιάς της Arp3 προς τη υποπεριοχή 2 της Arp3. Αυτό προκαλεί αποσταθεροποίηση της ενεργής διαμόρφωσης καθώς συμβαίνει στερεοχημική παρεμπόδιση μεταξύ της β7/αC θηλιάς της Arp3 και της αE/αF θηλιάς της Arp2. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [65].

## 1.6 Σκοπός της εργασίας

Μία από τις πιο απαιτητικές διαδικασίες κατά τη φάση βελτιστοποίησης των ενώσεων-οδηγών κατά το σχεδιασμό φαρμάκων είναι να προβλεφθεί με ακρίβεια η ελεύθερη ενέργεια πρόσδεσης των ενώσεων στο φαρμακευτικό στόχο. Μια από τις πιο αξιόπιστες προσεγγίσεις για τον υπολογισμό της ελεύθερης ενέργειας πρόσδεσης είναι οι υπολογισμοί Διατάραξης Ελεύθερης Ενέργειας (FEP) με Μοριακές Δυναμικές προσομοιώσεις για τον καθορισμό είτε της απόλυτης ή της σχετικής ελεύθερης ενέργειας πρόσδεσης, καθώς θεμελιώνεται σε βασικές αρχές της στατιστικής μηχανικής. Ωστόσο, το υψηλό υπολογιστικό κόστος, η περιορισμένη ακρίβεια των μεθόδων μοριακής μοντελοποίησης και οι τεχνικές προκλήσεις για τη προετοιμασία, την εκτέλεση και την ανάλυση προσομοιώσεων ελεύθερης ενέργειας, έχουν παρεμποδίσει την επιτυχή εφαρμογή της μεθόδου στη φαρμακοβιομηχανία. Προκειμένου να επιλυθούν αυτά τα ζητήματα και να βελτιωθεί η εκτιμητική ικανότητα των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας, έχουν πραγματοποιηθεί μελέτες μεγάλης κλίμακας και παγκόσμιες προκλήσεις, όπου δοκιμάζεται η εκτιμητική ισχύς της FEP σε σύγκριση με άλλες μεθόδους και με πειραματικά δεδομένα. Ταυτόχρονα, έχει αναπτυχθεί ένας σημαντικός αριθμός εφαρμογών και μεθοδολογιών με σκοπό την αυτοματοποίηση της προετοιμασίας και της εκτέλεσης των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας.

Για το σκοπό αυτό, λάβαμε συμμετοχή στην πρόκληση SAMPL6 SAMPLing, η οποία επικεντρώνεται στην αξιολόγηση της σύγκλισης και της αναπαραγωγιμότητας των προβλέψεων της απόλυτης ελεύθερης ενέργειας πρόσδεσης. Στο πλαίσιο αυτό πραγματοποιήθηκαν υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης για τρία συστήματα προσδετών-δεκτών χρησιμοποιώντας το υπολογιστικό πακέτο NAMD. Επιπλέον, αξιολογήθηκε η σύγκλιση των υπολογισμών ελεύθερης ενέργειας, ενώ τα αποτελέσματα συγκρίθηκαν με άλλα πακέτα προσομοίωσης και πειραματικά δεδομένα. Στη συνέχεια, μελετήθηκε μια σειρά 15 ανάλογων ενώσεων του μορίου CK-666 που αναστέλλει τη δράση του πρωτεϊνικού συμπλέγματος Arp2/3. Τέλος, έγινε σύγκριση των διαφορών

της σχετικής ελεύθερης ενέργειας πρόσδεσης με τη χρήση δύο πακέτων Μοριακής Δυναμικής (NAMD, GROMACS) καθώς και πειραματικών δεδομένων που έχουν ληφθεί από εξωτερικούς συνεργάτες. Η παρούσα διπλωματική παρουσιάζει τα αποτελέσματα αυτά.

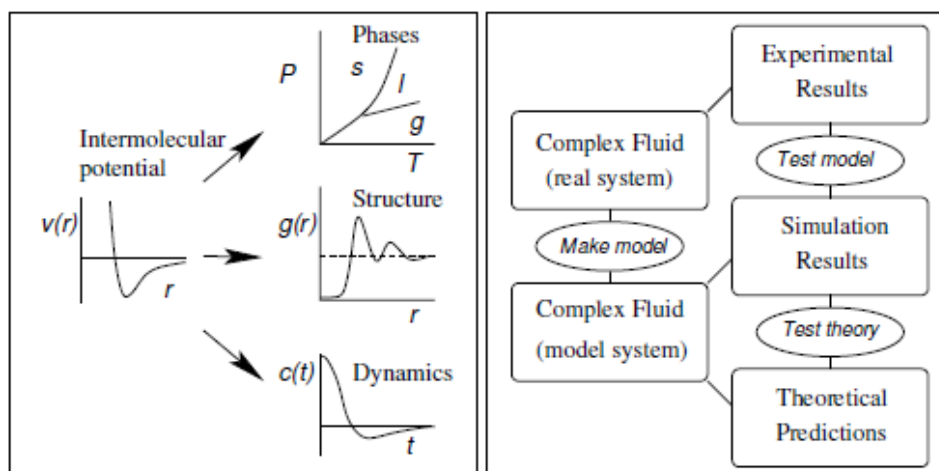




## 2. ΜΕΘΟΔΟΛΟΓΙΑ

### 2.1 Προσομοίωση μοριακού συστήματος

Η Μοριακή Μοντελοποίηση περιλαμβάνει θεωρητικές και υπολογιστικές μεθόδους οι οποίες χρησιμοποιούνται προκειμένου να μοντελοποιήσουν τη συμπεριφορά των μορίων. Η προσέγγιση αυτή έχει ευρεία εφαρμογή σε πολλούς τομείς όπως η υπολογιστική χημεία, ο σχεδιασμός φαρμάκων, η υπολογιστική βιολογία και η επιστήμη των υλικών. Το κοινό χαρακτηριστικό των μεθόδων Μοριακής Μοντελοποίησης είναι η περιγραφή σε ατομικό επίπεδο των μοριακών συστημάτων, τα οποία κυμαίνονται από μικρά χημικά μόρια έως μεγάλα βιολογικά συγκροτήματα [66]. Για την κατανόηση των ιδιοτήτων των μορίων που σχετίζονται με τη δομή και τις μεταξύ τους αλληλεπιδράσεις, πραγματοποιούνται προσομοιώσεις των μοριακών συστημάτων με τη χρήση ηλεκτρονικού υπολογιστή. Οι προσομοιώσεις λειτουργούν ως γέφυρα μεταξύ της μικροσκοπικής κλίμακας μήκους και χρόνου και του μακροσκοπικού κόσμου του εργαστηρίου (Σχήμα 2.1). Μέσω των μοριακών προσομοιώσεων, παρέχεται η δυνατότητα προσομοίωσης συνθηκών που είναι δύσκολες έως αδύνατες να πραγματοποιηθούν στο εργαστήριο, όπως μελέτη ενός συστήματος σε ακραίες θερμοκρασίες ή πιέσεις [67].



**Σχήμα 2.1** Οι προσομοιώσεις λειτουργούν ως γέφυρα μεταξύ: μικροσκοπικών και μακροσκοπικών ιδιοτήτων (αριστερά), θεωρητικών προσεγγίσεων και πειράματος (δεξιά). Το σχήμα λήφθηκε από τη βιβλιογραφική αναφορά [67].

Όπως είναι γνωστό, η ύλη απαρτίζεται από άτομα, τα οποία αποτελούνται από τα ηλεκτρόνια και από έναν πυρήνα πρωτονίων και νετρονίων. Σε μακροσκοπικό επίπεδο η κίνηση των σωματιδίων ενός συστήματος μπορεί να περιγραφεί πλήρως από την Κλασική Μηχανική, επιλύοντας τις εξισώσεις κίνησης του Νεύτωνα. Ωστόσο, οι νόμοι της Κλασικής Μηχανικής αδυνατούν να περιγράψουν την κίνηση τόσο μικρών σωματιδίων όπως τα ηλεκτρόνια. Για το σκοπό αυτό εφαρμόζονται οι εξισώσεις της Κβαντικής Μηχανικής, η οποία αναγνωρίζει το δυϊσμό των σωματιδίων αυτών θεωρώντας ότι ένα σωματίδιο αντί να ταξιδεύει κατά μήκος μιας καθορισμένης τροχιάς, διανέμεται διαμέσου του χώρου σαν ένα κύμα [14]. Τα μοντέλα της Κβαντικής Μηχανικής πηγάζουν από την εξίσωση του Schrödinger που πρωτοεμφανίστηκε στα τέλη της δεκαετίας του

1920, η οποία αντιμετωπίζει τα μόρια ως συλλογές πυρήνων και ηλεκτρονίων, δίχως καμιά αναφορά σε «χημικούς δεσμούς».

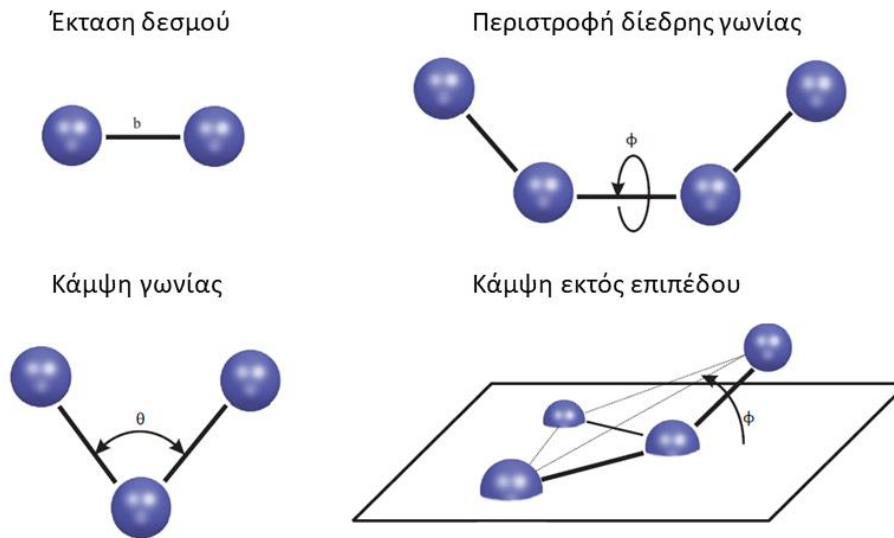
Ωστόσο, στην περίπτωση όπου το μοριακό σύστημα που πρόκειται να μελετηθεί είναι βιολογικό, η Κβαντική Μηχανική αδυνατεί να το περιγράψει, εξαιτίας του μεγάλου μεγέθους που χαρακτηρίζει ένα τέτοιο σύστημα και της τάξης μεγέθους του χρόνου στον οποίο πραγματοποιούνται οι διάφορες βιολογικές δράσεις. Συνεπώς για τον υπολογιστικό χειρισμό τέτοιων πολύπλοκων συστημάτων, έχουν αναπτυχθεί διάφορες προσεγγιστικές μέθοδοι οι οποίες βασίζονται στη Μοριακή Μηχανική [68]. Μία από αυτές τις μεθοδολογίες είναι και η Μοριακή Δυναμική, μία προσέγγιση με βάση την οποία η ενέργεια ενός συστήματος προκύπτει μέσω αριθμητικής ολοκλήρωσης του νόμου κίνησης του Νεύτωνα. Έτσι παράγονται πληροφορίες του συστήματος σε μικροσκοπικό επίπεδο, οι οποίες ύστερα συνδέονται με τις μακροσκοπικές ιδιότητες μέσω της Στατιστικής Μηχανικής [69]. Προκειμένου να εφαρμοστεί η Μοριακή Δυναμική για την ακριβή περιγραφή των βιολογικών μοριακών συστημάτων, είναι απαραίτητη η χρήση απλοποιημένων εξισώσεων, οι οποίες θα παρέχουν αποτελέσματα που συμφωνούν με τα πειραματικά, αλλά και θα επιλύονται υπολογιστικά σε εύλογο χρονικό διάστημα. Μία από τις κυριότερες απλοποιήσεις που περιλαμβάνει η Μοριακή Δυναμική είναι η προσέγγιση Born – Oppenheimer, σύμφωνα με την οποία διαχωρίζονται οι κινήσεις των μορίων στο ηλεκτρονιακό (δονητικό) και στο πυρηνικό (περιστροφικό) μέρος. Επιπλέον, οι θέσεις των πυρήνων θεωρούνται πρακτικά αμετάβλητες σε σχέση με αυτές των ηλεκτρονίων, μιας και οι πυρήνες είναι πολύ βαρύτεροι σε σχέση με τα ηλεκτρόνια και επομένως κινούνται πολύ αργά σε σχέση με αυτά [68].

Επιπρόσθετα, για την προσομοίωση των βιολογικών συστημάτων εφαρμόζεται μία ακόμη απλοποίηση για την περιγραφή των ενδομοριακών και διαμοριακών αλληλεπιδράσεων των ατόμων του συστήματος. Ο συνήθης τρόπος για τη μελέτη τέτοιων συστημάτων είναι η χρήση ενός πεδίου δυνάμεων (force field) το οποίο αποτελεί τη μαθηματική έκφραση του δυναμικού το οποίο ασκείται σε όλα τα άτομα ενός συστήματος [70]. Πιο συγκεκριμένα, η δυναμική ενέργεια του συστήματος περιγράφεται μέσω εμπειρικών συναρτήσεων, οι οποίες εκφράζονται ως άθροισμα ενδομοριακών (Εξίσωση 2.1) και διαμοριακών (Εξίσωση 2.2) ενεργειακών όρων:

$$\begin{aligned}
 U_{bonded} &= U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers} \\
 &= \sum_{bonds} \frac{1}{2} k_b (b - b_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} k_\phi (1 + \cos[n\phi - \gamma]) \\
 &\quad + \sum_{impropers} \frac{1}{2} k_\omega (\omega - \omega_0)^2
 \end{aligned} \tag{2.1}$$

Ο όρος  $U_{bonds}$  περιγράφει τις δυνάμεις που δρουν μεταξύ δύο ομοιοπολικά συνδεδεμένων ατόμων με βάση του νόμου του Hooke, όπου  $k_b$  η σταθερά δύναμης του ελατηρίου και  $b_0$  η τιμή ισορροπίας γύρω από την οποία ταλαντώνεται ο δεσμός. Ο δεύτερος όρος  $U_{angles}$  αφορά την ενέργεια των γωνιών κάμψης που σχηματίζονται μεταξύ τριών διαδοχικά ενωμένων ατόμων με ομοιοπολικό δεσμό, με  $k_\theta$  τη σταθερά κάμψης του δεσμού και  $\theta_0$  την απόκλιση από την ιδανική γωνία κάμψης. Ο όρος  $U_{dihedrals}$  αποτελεί μία συνάρτηση ταλάντωσης για τις διέδρες γωνίες που σχηματίζονται μεταξύ τεσσάρων διαδοχικών ατόμων που ανήκουν σε διαφορετικά επίπεδα, με  $\phi$  τη διέδρη γωνία,  $n$  την περιοδικότητα,  $\gamma$  τη γωνία με τη χαμηλότερη ενέργεια και  $k_\phi$  την σταθερά δύναμης. Τέλος, ο όρος  $U_{impropers}$  περιγράφει τις εκτός επιπέδου κινήσεις των ατόμων με  $\omega$  τη γωνία ανάμεσα στο επίπεδο

που ορίζουν το κεντρικό άτομο και δύο περιφερειακά άτομα και στο επίπεδο που ορίζουν τα περιφερειακά άτομα. Οι παραπάνω ενδομοριακοί όροι συνοψίζονται στο Σχήμα 2.2.



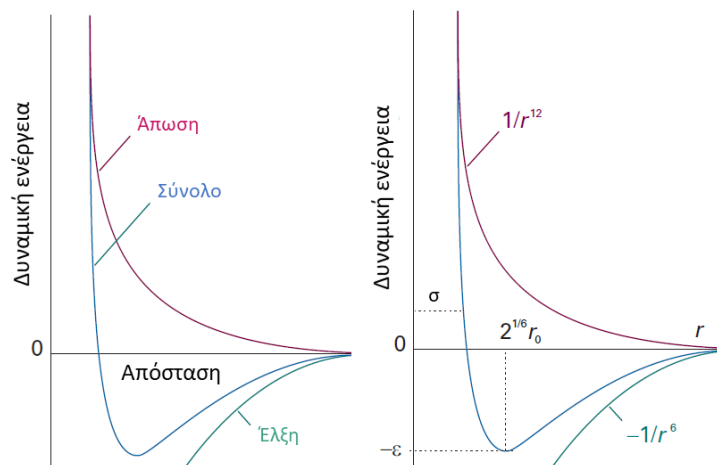
**Σχήμα 2.2** Σχηματική αναπαράσταση των ενδομοριακών αλληλεπιδράσεων μεταξύ των ατόμων ενός συστήματος. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [71] και μεταφράστηκε στα ελληνικά.

Το διαμοριακό (non-bonded) μέρος της δυναμικής ενέργειας ορίζεται με βάση την Εξίσωση 2.2 και αποτελεί το άθροισμα των ηλεκτροστατικών δυνάμεων που περιγράφονται από το Νόμο του Coulomb και των van der Waals αλληλεπιδράσεων που περιγράφονται από το δυναμικό Lennard-Jones.

$$U_{non-bonded} = U_{electrostatic} + U_{vdW}$$

$$= \sum_{i,j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + \sum_{i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.2)$$

Ο πρώτος όρος της Εξίσωσης 2.2 περιλαμβάνει τις μεγάλης εύρους ηλεκτροστατικές αλληλεπιδράσεις (long-range electrostatic interactions) ανάμεσα σε άτομα που φέρουν καθαρό ή μερικό φορτίο. Στο νόμο του Coulomb ως  $r_{ij}$  ορίζεται η απόσταση μεταξύ δύο ατόμων που φέρουν φορτία  $q_i$  και  $q_j$  αντίστοιχα, ενώ με  $\epsilon_0$  συμβολίζεται η διηλεκτρική σταθερά του κενού. Οι δυνάμεις van der Waals περιγράφουν τις αλληλεπιδράσεις μικρού εύρους (short-range interactions) ανάμεσα σε άτομα που βρίσκονται σε κοντινή απόσταση. Περιγράφονται από το δυναμικό Lennard-Jones στο οποίο εμπεριέχεται ένα ελκτικό και ένα απωστικό κομμάτι (Σχήμα 2.3). Συγκεκριμένα, οι ελκτικές δυνάμεις οφείλονται στις δυνάμεις London και το απωστικό κομμάτι απορρέει από την απαγορευτική αρχή του Pauli. Επομένως, στο δεύτερο όρο της Εξίσωσης 2.2 με  $\epsilon_{ij}$  συμβολίζεται το φρέαρ του δυναμικού και  $\sigma_{ij}$  είναι η απόσταση στην οποία το ενδομοριακό δυναμικό μεταξύ δύο ατόμων είναι μηδέν. Με άλλα λόγια, το  $\sigma_{ij}$  είναι μια μέτρηση του πόσο κοντά δύο μη ενωμένα σωματίδια μπορούν να φτάσουν. Γι' αυτό το λόγο συχνά αναφέρεται και ως ακτίνα van der Waals [72].



**Σχήμα 2.3** Σχηματική αναπαράσταση του δυναμικού Lennard – Jones. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [14].

## 2.2 Βασικές αρχές προσομοιώσεων Μοριακής Δυναμικής

Η Μοριακή Δυναμική μπορεί πλέον να εφαρμοστεί συστηματικά στην έρευνα για τη μελέτη ενός μεγάλου φάσματος δυναμικών ιδιοτήτων και διεργασιών από ερευνητές σε τομείς όπως η δομική βιοχημεία, η βιοφυσική, η μοριακή βιολογία, η φαρμακευτική χημεία και η βιοτεχνολογία [73]. Η ιδέα για τη μέθοδο της Μοριακής Δυναμικής γεννήθηκε από την κοινότητα της θεωρητικής φυσικής κατά τη δεκαετία του 1950. Πιο συγκεκριμένα, το 1957 οι Alder και Wainwright [74] πραγματοποίησαν την πρώτη προσομοίωση Μοριακής Δυναμικής χρησιμοποιώντας το λεγόμενο μοντέλο σκληρής σφαίρας, στο οποίο τα άτομα αλληλεπιδρούσαν μόνο μέσω τέλειων συγκρούσεων. Ύστερα, ο Rahman [75] εφάρμοσε ένα ομαλό και συνεχές δυναμικό ώστε να μιμηθεί πραγματικές ατομικές αλληλεπιδράσεις. Κατά τη διάρκεια της δεκαετίας του 1970, καθώς οι υπολογιστές έγιναν περισσότερο διαδεδομένοι, αναπτύχθηκαν προσομοιώσεις Μοριακής Δυναμικής για πιο σύνθετα συστήματα, με αποκορύφωμα το 1976 όπου πραγματοποιήθηκε η πρώτη προσομοίωση πρωτεΐνης [76], [77].

Για να ξεκινήσει μια προσομοίωση Μοριακής Δυναμικής είναι απαραίτητη η ύπαρξη ενός αρχικού μοντέλου του συστήματος το οποίο πρόκειται να μελετηθεί. Συνήθως, η δομή αυτή λαμβάνεται από πειραματικές τεχνικές και περιέχει τις αρχικές συντεταγμένες των ατόμων του συστήματος. Στην περίπτωση των βιολογικών συστημάτων, η αρχική δομή είναι μια τρισδιάστατη δομή που έχει επιλυθεί με τεχνικές όπως η διάθλαση ακτίνων Χ, η φασματοσκοπία NMR και έχει κατατεθεί στην βάση δεδομένων Protein Data Bank (PDB) [78], [79]. Οι προσομοιώσεις Μοριακής Δυναμικής περιλαμβάνουν τον αριθμητικό υπολογισμό των στιγμιαίων δυνάμεων και των επακόλουθων κινήσεων που υπάρχουν σε ένα σύστημα Μοριακής Μηχανικής. Το σύστημα Μοριακής Μηχανικής αποτελείται από ένα σύνολο σωματιδίων που κινούνται ως απόκριση στις αλληλεπιδράσεις τους σύμφωνα με τις εξισώσεις κίνησης που ορίζονται στην Νευτώνεια μηχανική. Έτσι, για ένα άτομο μάζας  $m_i$ , με θέση που ορίζεται από το τρισδιάστατο διάνυσμα  $r_i$ , η σχέση μεταξύ της ταχύτητας και της ορμής του ατόμου, ορίζεται από την Εξίσωση 2.3.

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} \quad (2.3)$$

Η δύναμη  $F_i$  που ασκείται στο άτομο  $i$  από το υπόλοιπο σύστημα δίνεται από την αρνητική βαθμίδα της συνάρτησης δυναμικού σε σχέση με τη θέση του ατόμου (Εξίσωση 2.4).

$$F_i = -\frac{dV}{dr_i} \quad (2.4)$$

Έτσι ο νόμος κίνησης του Νεύτωνα δίνεται από την Εξίσωση 2.5

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad (2.5)$$

Η ολοκλήρωση της παραπάνω εξίσωσης παράγει τη χρονική εξέλιξη της θέσης των ατόμων του συστήματος, που ονομάζεται τροχιά (trajectory). Η τροχιά της μοριακής δυναμικής, περιγράφει την εξέλιξη του συστήματος σε σχέση με τον χρόνο προσομοίωσης και η γνώση αυτής, επιτρέπει τον υπολογισμό των μακροσκοπικών φυσικών ιδιοτήτων του βιολογικού συστήματος. Ωστόσο, η πολυπλοκότητα της Εξίσωσης 2.5 έχει ως αποτέλεσμα την αδυναμία εύρεσης αναλυτικής λύσης. Έτσι, απαιτείται η χρήση αλγορίθμων οι οποίοι ολοκληρώνουν αριθμητικά την παραπάνω εξίσωση [73]. Σήμερα, υπάρχουν πολλοί αλγόριθμοι για την ολοκλήρωση των εξισώσεων κίνησης και οι περισσότεροι από αυτούς, βασίζονται σε μεθόδους πεπερασμένων διαφορών όπου η ολοκλήρωση γίνεται σε μικρά βήματα, κάθε ένα από τα οποία διαχωρίζεται χρονικά από μια συγκεκριμένη περίοδο  $\delta t$  [80], [81], [82]. Ο πιο συχνά χρησιμοποιούμενος αλγόριθμος στις προσομοιώσεις Μοριακής Δυναμικής, είναι ο αλγόριθμος Verlet εξαιτίας της απλότητας και της σταθερότητας που παρουσιάζει. Ο αλγόριθμος αυτός, χρησιμοποιεί τις ατομικές θέσεις και τις επιταχύνσεις στο χρόνο  $t$  και τις θέσεις από το προηγούμενο βήμα,  $r(t - \delta t)$ , για τον προσδιορισμό των νέων θέσεων σε  $t + \delta t$  (Εξίσωση 2.6).

$$r(t + \delta t) \cong 2r(t) - r(t - \delta t) + \frac{d^2 r}{dt^2} \delta t^2 \quad (2.6)$$

Το βήμα ολοκλήρωσης  $\delta t$  καθορίζεται πρακτικά από την πιο γρήγορη δόνηση του βιομοριακού συστήματος, δηλαδή από τη συχνότητα δόνησης του δεσμού C-H ( $\delta t = 1 \text{ fs} = 10^{-15} \text{ s}$ ). Ωστόσο, ένα πολύ μικρό βήμα, αυξάνει τον απαιτούμενο υπολογιστικό χρόνο. Μια συνήθης παραδοχή στη Μοριακή Δυναμική είναι η θεώρηση ότι οι ενδομοριακοί δεσμοί δεν ταλαντώνονται, αλλά έχουν σταθερό μήκος. Αυτό επιτυγχάνεται με την εφαρμογή περιορισμών στο μήκος του δεσμού και στη γωνία των δεσμών. Για το σκοπό αυτό έχουν αναπτυχθεί διάφοροι αλγόριθμοι, όπως ο SHAKE, η εφαρμογή των οποίων επιτρέπει την αύξηση του χρονικού διαστήματος  $\delta t$  και έχει ως αποτέλεσμα τη μείωση του απαιτούμενου υπολογιστικού χρόνου.

Οι Νευτώνειες εξισώσεις είναι οι φυσικά ακριβείς εξισώσεις που περιγράφουν την κίνηση των ατόμων στο σύστημα. Ωστόσο, στις προσομοιώσεις Μοριακής Δυναμικής εμφανίζεται το πρόβλημα ότι η θερμοκρασία του συστήματος πρέπει να επαναρυθμίζεται περιοδικά έτσι ώστε να προσομοιωθούν οι βιολογικές συνθήκες. Για την αποφυγή αυτού του προβλήματος έχουν αναπτυχθεί διάφορα σύνολα εξισώσεων, τα οποία ελέγχουν τη θερμοκρασία και την πίεση του συστήματος. Η μετατροπή των μικροσκοπικών παραμέτρων (ατομική θέση, ταχύτητα) σε μακροσκοπικά φυσικά μεγέθη (θερμοκρασία, πίεση, ενέργεια) ακολουθεί τις αρχές της Στατιστικής Μηχανικής [72]. Η συσχέτιση των μικροσκοπικών με τις μακροσκοπικές ιδιότητες ενός συστήματος

βασίζεται στη χρήση στατιστικών συνόλων. Ένα στατιστικό σύνολο αποτελεί μια συλλογή σημείων στο χώρο φάσεων (phase space) που ικανοποιεί τις συνθήκες μιας συγκεκριμένης θερμοδυναμικής κατάστασης. Τα στατιστικά σύνολα περιγράφουν μια συλλογή όλων των πιθανών συστημάτων που μπορούν να έχουν διαφορετικές μικροσκοπικές καταστάσεις αλλά είναι ταυτόσημες μακροσκοπικά (ή θερμοδυναμικά). Στη συνέχεια παρουσιάζεται μια σύνοψη των διαφορετικών στατιστικών συνόλων με διαφορετικά χαρακτηριστικά [73].

### **Μικρο-κανονική κατανομή (Microcanonical – NVE)**

Στο μικρο-κανονικό στατιστικό σύνολο, οι τρεις σταθερές παράμετροι είναι  $N$ ,  $V$ ,  $E$  δηλαδή, ο αριθμός των σωματιδίων, ο όγκος και η ενέργεια του συστήματος. Αυτό είναι το φυσικό σύνολο για προσομοιώσεις Μοριακής Δυναμικής, όπου οι Νευτώνειες εξισώσεις κίνησης μπορούν να εφαρμοστούν αμετάβλητες στο σύστημα. Η κατανομή αυτή περιγράφει ένα απομονωμένο σύστημα.

### **Κανονική κατανομή (Canonical – NVT)**

Στο κανονικό σύνολο, αριθμός σωματιδίων ( $N$ ), ο όγκος ( $V$ ) και η θερμοκρασία ( $T$ ) είναι οι σταθερές παράμετροι του συστήματος. Δεδομένου ότι η θερμοκρασία του συστήματος πρέπει να διατηρηθεί σταθερή για αυτό το στατιστικό σύνολο, πρέπει να εισαχθεί ένας θερμοστάτης. Έτσι η Χαμιλτονιανή (Hamiltonian,  $H$ ) του συστήματος μετατρέπεται στην Εξίσωση 2.7:

$$H = K + V + K_s + V_s \quad (2.7)$$

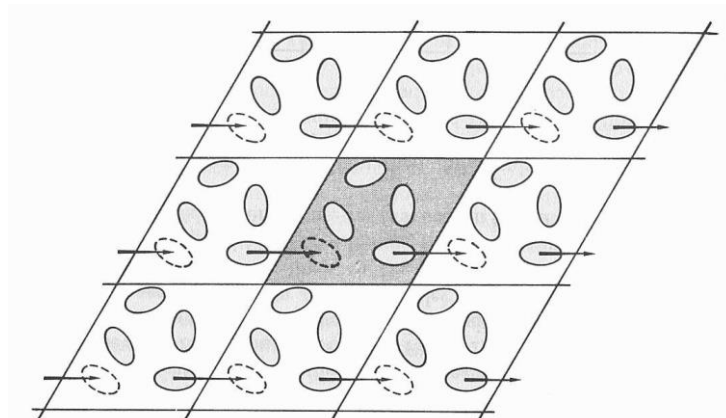
όπου  $K$  είναι η κινητική ενέργεια του συστήματος,  $V$  η δυναμική ενέργεια και  $K_s$ ,  $V_s$  η κινητική και δυναμική ενέργεια που συνδέεται με τον θερμοστάτη.

### **Ισόθερμη – ισοβαρής κατανομή (Isothermal–isobaric – NPT)**

Σε αυτό το στατιστικό σύνολο, η πίεση και η θερμοκρασία διατηρούνται σταθερές, δηλαδή οι παράμετροι  $N$ ,  $P$  και  $T$  παραμένουν αμετάβλητοι. Στην περίπτωση αυτή, ο νόμος κίνησης του Νεύτωνα γίνεται ακόμα πιο πολύπλοκος επειδή εκτός από το θερμοστάτη, το σύστημα συνδέεται και με βαροστάτη ούτως ώστε η κατανομή να προσομοιάζει καλύτερα τις κανονικές πειραματικές συνθήκες.

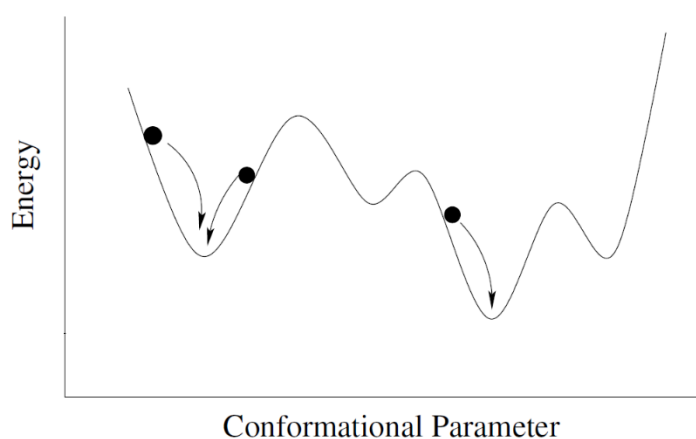
Επιπρόσθετα, για να αποφευχθούν τεχνητά οριακά φαινόμενα στις προσομοιώσεις Μοριακής Δυναμικής τα οποία μπορεί να έχουν ως αποτέλεσμα τη δημιουργία διεπιφανειών στα όρια του συστήματος με το περιβάλλον του, χρησιμοποιούνται περιοδικές οριακές συνθήκες (periodic boundary conditions). Με την προσέγγιση αυτή θεωρείται ουσιαστικά ότι το κουτί της προσομοίωσης είναι περιοδικά συνεχές. Έτσι, η μπροστά του επιφάνεια είναι απολύτως ίδια με την πίσω, η δεξιά του απολύτως ίδια με την αριστερή του και η άνω του επιφάνεια απολύτως ίδια με την κάτω (Σχήμα 2.4). Συνεπώς όταν ένα μόριο εγκαταλείπει το κεντρικό κουτί, τότε μια περιοδική του εικόνα εισέρχεται ακριβώς από την απέναντι πλευρά. Κατά τις μετακινήσεις των μορίων η αριθμητική πυκνότητα στο κεντρικό κουτί παραμένει σταθερή. Όπως φαίνεται και στο Σχήμα 2.4, το κεντρικό κουτί είναι το σκιασμένο με βάση το οποίο κατασκευάζονται οκτώ κλώνοι του προς όλες τις

κατευθύνσεις. Τα άτομα στα περιβάλλοντα εικονικά συστήματα αλληλεπιδρούν με τα άτομα στο πραγματικό σύστημα [70].



**Σχήμα 2.4** Δισδιάστατη απεικόνιση ενός περιοδικού συστήματος το οποίο αντιπροσωπεύει τις περιοδικές οριακές συνθήκες που χρησιμοποιούνται στις προσομοιώσεις Μοριακής Δυναμικής. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [72].

Προτού ξεκινήσει μια προσομοίωση Μοριακής Δυναμικής, συνίσταται η ελαχιστοποίηση της ενέργειας, με σκοπό την απομάκρυνση αλληλεπιδράσεων που τυχόν υπάρχουν στο σύστημα και μπορούν να προκαλέσουν διαστρέβλωση της δομής του. Στο πλαίσιο των βιομορίων, η συνάρτηση που πρέπει να ελαχιστοποιηθεί είναι η δυναμική ενέργεια. Η δυναμική ενέργεια των βιολογικών συστημάτων είναι μια περίπλοκη, πολυδιάστατη συνάρτηση των καρτεσιανών συντεταγμένων του συστήματος. Το ενεργειακό διάγραμμα ενός βιομορίου χαρακτηρίζεται από ένα μεγάλο αριθμό ελαχίστων, που αντιστοιχούν σε ενδιάμεσες διαμορφώσεις του συστήματος. Η ελαχιστοποίηση της ενέργειας περιλαμβάνει την αναζήτηση του ενεργειακού τοπίου (energy landscape) του μορίου με σκοπό την εύρεση ενός τοπικού ελαχίστου (Σχήμα 2.5).



**Σχήμα 2.5** Σχηματική αναπαράσταση μιας μονοδιάστατης επιφάνειας της δυναμικής ενέργειας. Η ελαχιστοποίηση της ενέργειας αποσκοπεί στην εύρεση του πλησιέστερου τοπικού ελαχίστου της ενέργειας. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [72].

Η ενέργεια σε αυτό το τοπικό ελάχιστο μπορεί να είναι πολύ υψηλότερη από την ενέργεια του ολικού ελαχίστου. Από φυσικής σκοπιάς, η ελαχιστοποίηση της ενέργειας αντιστοιχεί σε στιγμιαία «ψύξη» του συστήματος στους 0 K. Οι γεωμετρίες του συστήματος που αντιστοιχούν στα σημεία τοπικού ελαχίστου υπολογίζονται μέσω αλγορίθμων ελαχιστοποίησης της επιφάνειας της ενέργειας. Στα

βιολογικά συστήματα απαιτείται η χρήση αριθμητικών μεθόδων, καθώς η Επιφάνεια της Δυναμικής Ενέργειας (Potential Energy Surface, PES), δεν μπορεί να εκφραστεί από μια αναλυτική συνάρτηση. Οι περισσότεροι αλγόριθμοι ελαχιστοποίησης ενέργειας μπορούν να εντοπίσουν μόνο το ελάχιστο που βρίσκεται πλησιέστερα στο σημείο εκκίνησης. Για να είναι εφικτός ο υπολογισμός περισσότερων ελαχίστων σε μία επιφάνεια της δυναμικής ενέργειας, απαιτείται η επιλογή διαφορετικών σημείων εκκίνησης [70], [72].

## 2.3 Προετοιμασία του υποδοχέα για προσομοίωση

Για τον επιτυχή σχεδιασμό ενός φαρμάκου με γνωστή τη δομή του υποδοχέα, είναι απαραίτητο η τρισδιάστατη δομή του φαρμακευτικού στόχου να ανταποκρίνεται στην πραγματική του δομή. Όπως προαναφέρθηκε, η Protein Data Bank (PDB) είναι μια βάση δεδομένων που περιέχει τρισδιάστατα δομικά δεδομένα για μεγάλα βιομόρια όπως πρωτεΐνες και νουκλεϊκά οξέα [78]. Τα δεδομένα αυτά, τυπικά προέρχονται από κρυσταλλογραφία ακτίνων Χ, φασματοσκοπία NMR, φασματοσκοπία σκέδασης νετρονίων ή κρυογονικής ηλεκτρονικής μικροσκοπίας (Cryogenic electron microscopy, cryo-EM) [83]. Ωστόσο, οι δομές αυτές δεν μπορούν να αξιοποιηθούν απευθείας λόγω αρκετών ελλείψεων και αποκλίσεων που παρουσιάζουν από την πραγματική δομή. Οι κυριότερες από αυτές συνοψίζονται παρακάτω.

Αρχικά, οι περισσότερες δομές προέρχονται συνήθως από πειράματα κρυσταλλογραφίας ακτίνων Χ, στα οποία απουσιάζουν τα υδρογόνα. Αυτό συμβαίνει καθώς οι κρυσταλλικές δομές προσδιορίζονται με περίθλαση των ακτίνων Χ κατά την οποία μετράται η πυκνότητα των ηλεκτρονίων. Όσο βαρύτερο είναι ένα άτομο και όσο περισσότερα ηλεκτρόνια έχει, τόσο ισχυρότερη είναι η συμβολή του στην περίθλαση. Αυτό σημαίνει πως ειδικά με την παρουσία βαρέων ατόμων, τα ελαφρά άτομα είναι πιο δύσκολο να εντοπιστούν. Το ελαφρύτερο άτομο όλων είναι το υδρογόνο μιας και έχει μόνο ένα ηλεκτρόνιο, τοποθετημένο μακριά από τον πυρήνα. Επομένως, τα άτομα υδρογόνου είναι δύσκολα ανιχνεύσιμα με μεθόδους περίθλασης ακτίνων Χ και συνεπώς θα πρέπει να προστεθούν στην δομή του υποδοχέα προκειμένου να ξεκινήσει η προσομοίωση του συστήματος [84].

Κατά την κρυσταλλογραφία των πρωτεϊνών πολλές φορές συν-κρυσταλλώνονται ιόντα και διάφορα μόρια, όπως μόρια νερού, πεπτίδια και μικρά οργανικά μόρια, τα οποία εμφανίζονται στις τελικές δομές. Αυτά τα επιπλέον στοιχεία ορισμένες φορές δίνουν σημαντικές πληροφορίες, όπως στην περίπτωση των μορίων νερού, αλλά στις περισσότερες περιπτώσεις πρέπει να αφαιρεθούν [85]. Επιπρόσθετα, η κρυσταλλογραφία ακτίνων Χ δεν παρέχει πληροφορίες σχετικά με τους δεσμούς μεταξύ ατόμων, τα τοπολογικά χαρακτηριστικά της δομής και τα φορτία των ατόμων. Επίσης, ορισμένα τερματικά αμινικά άκρα, καθώς και αμινοξέα ασπαραγίνης, μπορεί να είναι λάθος καταταξιωμένα, επειδή η κρυσταλλογραφία συχνά αδυνατεί να διακρίνει ξεκάθαρα το O και το NH<sub>2</sub>. Η προσθήκη αυτών των ελλείψεων είναι εύκολη για την περίπτωση των αμινοξέων, ωστόσο για μόρια όπως οι προσδέτες, είναι μια διαδικασία πιο περίπλοκη και απαιτεί την προσθήκη τους από τον χρήστη. Σε πολλές πρωτεϊνικές δομές που λαμβάνονται από την Protein Data Bank ενδέχεται να απουσιάζουν ορισμένες πλευρικές αλυσίδες ή και ολόκληρες θηλιές της δευτεροταγούς δομής τους, συνήθως επειδή είναι πολύ ευκίνητες και δεν είναι εφικτό να φανούν κατά την κρυσταλλογραφία. Η συμπλήρωση των πλευρικών αλυσίδων είναι σημαντική καθώς ένας προσδέτης μπορεί να



αλληλεπιδρά με αυτές. Επίσης, παράλειψη θηλιών μπορεί να οδηγήσει σε μεγάλα σφάλματα κατά τις προσομοιώσεις Μοριακής Δυναμικής [2].

Τέλος, από τις δομές συχνά απουσιάζουν οι σωστές καταστάσεις πρωτονίωσης και ταυτομέρειας των συν-κρυσταλλομένων μορίων. Πιο συγκεκριμένα, είναι απαραίτητο να προβλεφθούν οι καταστάσεις πρωτονίωσης αμινοξέων όπως η ιστιδίνη, το ασπαρτικό οξύ και το γλουταμινικό οξύ καθώς και τα ταυτομερή της ιστιδίνης. Επίσης, συνήθως παρατηρούνται στερεοχημικές παρεμποδίσεις μεταξύ κάποιων ατόμων με αποτέλεσμα να απαιτείται η διόρθωση της κρυσταλλικής δομής [85].

## 2.4 Προετοιμασία του προσδέτη για προσομοίωση

Εξίσου σημαντική με την προετοιμασία του υποδοχέα, είναι και η προετοιμασία των μορίων-προσδετών που πρόκειται να χρησιμοποιηθούν στις προσομοιώσεις Μοριακής Δυναμικής. Αυτό συμβαίνει καθώς ένας προσδέτης είναι πιθανόν να απαντάται με δύο μορφές ή/και να εμφανίζει διαφορετική κατάσταση πρωτονίωσης ανάλογα με το pH στο οποίο βρίσκεται. Επιπλέον, είναι σημαντικό να εντοπίζονται στερεογονικά κέντρα, αλλά και να προβλέπονται τα διάφορα χειρόμορφα/εναντιομερή μόρια, αφού πολλά από τα φάρμακα βρίσκονται σε μορφή ρακεμικών μιγμάτων (μίγματα εναντιομερών). Τέλος, προβλέπονται οι διάφορες διαμορφώσεις (διαμορφομερή) κορεσμένων δακτυλίων (ανάκλιντρα, λουτήρες, ισημερινές-αξονικές θέσεις). Η εκτίμηση αυτών των ιδιοτήτων για έναν προσδέτη είναι ουσιαστική για τις προσομοιώσεις Μοριακής Δυναμικής. Η μη πρόβλεψη αυτών μπορεί να οδηγήσει σε μετέπειτα λανθασμένους υπολογισμούς, καθώς δεν θα έχουν εξερευνηθεί επαρκώς οι πιθανές διαμορφώσεις που μπορεί να λάβει ο προσδέτης [86].

## 2.5 Υπολογισμός απόλυτης ελεύθερης ενέργειας πρόσδεσης με τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας

Η πρόσδεση ενός μορίου σε έναν πρωτεϊνικό στόχο παίζει καταλυτικό ρόλο στη θεραπεία ασθενειών. Στην προκειμένη περίπτωση, εστιάζουμε στο σενάριο όπου ένας προσδέτης σχηματίζει ένα αντιστρεπτό, μη ομοιοπολικό σύμπλοκο σε στοιχειομετρική αναλογία, με ένα βιολογικό μακρομόριο το οποίο είναι σε θερμοδυναμική ισορροπία με τα μη δεσμευμένα μόρια που βρίσκονται στο διάλυμα. Η ελεύθερη ενέργεια πρόσδεσης μπορεί να υπολογιστεί λαμβάνοντας υπόψιν ότι ο προσδέτης μπορεί να βρεθεί σε δύο καταστάσεις. Η δεσμευμένη κατάσταση (complex state) είναι αυτή κατά την οποία ο προσδέτης είναι σε σύμπλοκο με το βιολογικό μακρομόριο, ενώ στην ελεύθερη κατάσταση (solvent state) ο προσδέτης βρίσκεται ελεύθερος στο διάλυμα σε ισορροπία με τα μόρια του διαλύτη [87].

Η διαφορά της ελεύθερης ενέργειας μεταξύ αυτών των δύο καταστάσεων μπορεί να υπολογιστεί μέσω δειγματοληψίας (sampling) πολλών διαφορετικών διαμορφώσεων (μικροκαταστάσεων), υπολογισμού των ενεργειών σε κάθε μικροκατάσταση, και στη συνέχεια υπολογίζοντας τη συνάρτηση κατανομής Z, η οποία μπορεί να χρησιμοποιηθεί για την εξαγωγή πολλών σημαντικών ιδιοτήτων, μεταξύ άλλων και της σταθεράς πρόσδεσης ενός μορίου στον φαρμακευτικό στόχο.

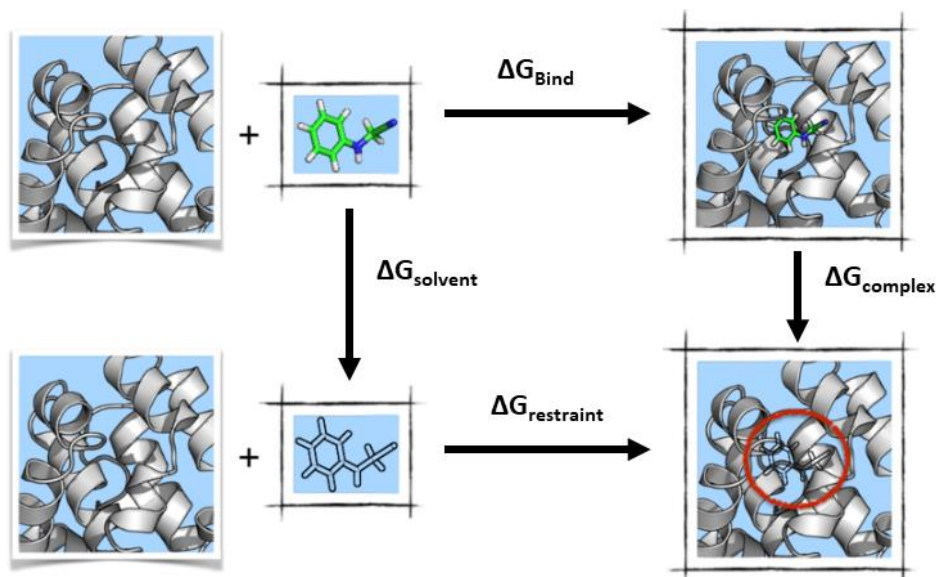
$$RT \ln(K_D) = \Delta G_{Bind}^0 = -RT \ln \left( \frac{Z_{Complex}}{Z_{Solvent}} \right) \quad (2.8)$$

$$Z_{state} = \sum_i^{state} \exp \left( -\frac{E_i}{k_B T} \right) \quad (2.9)$$

Οι Εξισώσεις 2.8 και 2.9 επιτρέπουν τον ακριβή υπολογισμό της απόλυτης ενέργειας πρόσδεσης ενός μορίου-προσδέτη, εάν η ενέργεια για όλες τις διαμορφώσεις των δύο τελικών καταστάσεων είναι γνωστή. Ο ακριβής υπολογισμός όλων των ενεργειακών καταστάσεων μπορεί να γίνει με ακρίβεια μέσω κβαντομηχανικών υπολογισμών. Ωστόσο, το γεγονός ότι τα περισσότερα βιολογικά συστήματα απαρτίζονται από 10 έως 100 χιλιάδες άτομα, καθιστά τους κβαντομηχανικούς υπολογισμούς μια επίπονη υπολογιστικά διαδικασία, μιας και απαιτούνται χρόνοι προσομοίωσης της τάξης των ms έως ms, ανάλογα με το σύστημα. Για την αντιμετώπιση αυτού του ζητήματος, μπορεί κανείς να εισάγει μία κλασική περιγραφή του συστήματος, μέσω πεδίων δυνάμεων Μοριακής Δυναμικής και να εφαρμόσει τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας.

Με τη Διατάραξη Ελεύθερης Ενέργειας (Free Energy Perturbation, FEP) διαταράσσεται το διαμοριακό κομμάτι της συνολικής ενέργειας του συστήματος και πραγματοποιείται μετάβαση από την κατάσταση A σε μία κατάσταση B. Οι ενεργειακοί όροι που συνήθως διαταράσσονται αφορούν τις ενέργειες van der Waals και τις ηλεκτροστατικές αλληλεπιδράσεις. Η σύνδεση της ελεύθερης ενέργειας Gibbs με αυτούς τους ενεργειακούς όρους γίνεται μέσω της εξίσωσης του Zwanzig (βλ. Εξίσωση 1.14). Έτσι, προσδευτικά οι ηλεκτροστατικές και οι van der Waals αλληλεπιδράσεις του συστήματος A μετατρέπονται στις αλληλεπιδράσεις του συστήματος B. Για το σκοπό αυτό, τα δύο συστήματα συνδέονται με τη χρήση μη-φυσικών ενδιάμεσων καταστάσεων, οι οποίες χαρακτηρίζονται από μια μεταβλητή που συνήθως συμβολίζεται με  $\lambda$ . Η μεταβλητή αυτή ορίζεται με τέτοιο τρόπο ώστε μια αλλαγή του  $\lambda$  από το μηδέν στο ένα, να αντιστοιχεί σε μία αλλαγή του συστήματος από το σύστημα A στο σύστημα B [88]. Επειδή οι ενδιάμεσες καταστάσεις που χρησιμοποιούνται δεν έχουν κάποιο φυσικό νόημα, οι υπολογισμοί Διατάραξης Ελεύθερης Ενέργειας συχνά χαρακτηρίζονται ως «αλχημικοί».

Ένα πρόβλημα που ενδέχεται να προκύψει κατά τον υπολογισμό των απόλυτων ελεύθερων ενεργειών πρόσδεσης με τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας είναι το μόριο-προσδέτης να εγκαταλείψει τη θέση πρόσδεσης, γεγονός που οδηγεί σε σημαντικά σφάλματα κατά τον υπολογισμό της ελεύθερης ενέργειας. Το ζήτημα αυτό μπορεί να εμφανιστεί προς το τέλος της αλχημικής προσομοίωσης, καθώς στο σημείο αυτό ο προσδέτης είναι ασθενώς δεσμευμένος από τον υποδοχέα, μιας και έχουν αποσυνδεθεί σχεδόν πλήρως οι ηλεκτροστατικές και οι van der Waals αλληλεπιδράσεις του. Μία προσέγγιση για την αντιμετώπιση του ζητήματος αυτού είναι η επιβολή περιορισμών (restraints), οι οποίοι θα έχουν ως αποτέλεσμα την παραμονή του προσδέτη κοντά στη θέση δέσμευσης του υποδοχέα. Οι περιορισμοί αυτοί επιβάλλονται μεταξύ ορισμένων ατόμων του συστήματος προσδέτη-υποδοχέα δίχως να υπάρχει κάποιος γενικός κανόνας ως προς την επιλογή τόσο των ατόμων, όσο και του είδους των περιορισμών (γωνία, απόσταση, δίεδρη γωνία, κλπ.) [20]. Στο Σχήμα 2.7 φαίνεται πώς διαμορφώνεται ο θερμοδυναμικός κύκλος για τον υπολογισμό των απόλυτων ελεύθερων ενεργειών πρόσδεσης με τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας.



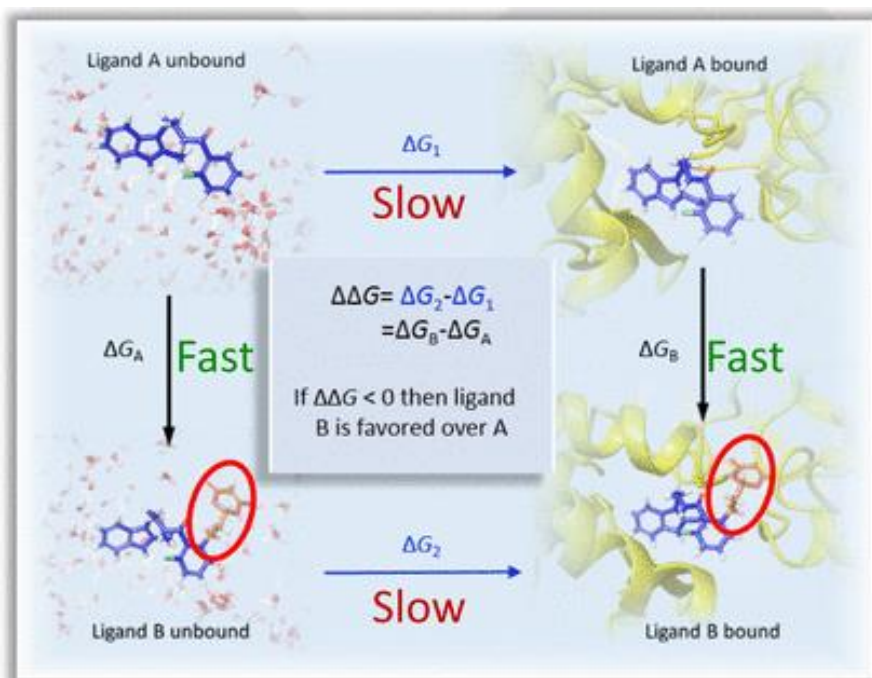
**Σχήμα 2.6** Παράδειγμα θερμοδυναμικού κύκλου που χρησιμοποιείται για τους αλχημικούς υπολογισμούς απόλυτης ελεύθερης ενέργειας πρόσδεσης με τη μέθοδο FEP. Η ελεύθερη ενέργεια πρόσδεσης υπολογίζεται ως  $\Delta G_{\text{bind}} = \Delta G_{\text{solvent}} - \Delta G_{\text{complex}}$ . Για να μην εκτραπεί ο προσδέτης από την κοιλότητα δέσμευσης της πρωτεΐνης επιβάλλονται περιορισμοί θέσης οι οποίοι συνεισφέρουν  $\Delta G_{\text{restraint}}$  στη συνολική ενέργεια πρόσδεσης [20].

## 2.6 Υπολογισμός σχετικής ελεύθερης ενέργειας πρόσδεσης με τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας

Οι αλχημικοί υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης έχουν περιορισμένη εφαρμογή, καθώς απαιτούν μεγάλους χρόνους προσομοίωσης για τη δειγματοληψία των πιθανών διαμορφώσεων. Για να αντιμετωπιστεί το ζήτημα αυτό, λαμβάνεται υπόψιν μόνο η σχετική διαφορά ελεύθερης ενέργειας μεταξύ δύο συγγενικών προσδετών, με την προϋπόθεση ότι η συνολική θέση πρόσδεσης του μορίου διατηρείται. Η προσέγγιση αυτή δεν είναι ιδιαίτερα περιοριστική στο πλαίσιο της βελτιστοποίησης ενώσεων-οδηγών, όπου σύνθητες ζητούμενο αποτελεί ο υπολογισμός της σχετικής ελεύθερης ενέργειας μεταξύ δύο παρόμοιων ενώσεων. Στο πλαίσιο αυτό, πραγματοποιείται μια απλοποίηση που προέρχεται από τις Εξισώσεις 2.8 και 2.9 και μπορεί να μειώσει την ανάγκη εκτενούς δειγματοληψίας, αποφεύγοντας τον υπολογισμό των απόλυτων ελεύθερων ενεργειών (Εξίσωση 2.10).

$$\begin{aligned} \Delta \Delta G_{A,B}^0 &= \Delta G_{B, \text{Bind}}^0 - \Delta G_{A, \text{Bind}}^0 = -RT \ln \left( \frac{Z_{\text{Complex}}^B Z_{\text{Solvent}}^A}{Z_{\text{Complex}}^A Z_{\text{Solvent}}^B} \right) \\ &= \Delta G_{A,B}^0 \text{ Complex} - \Delta G_{A,B}^0 \text{ Solvent} \end{aligned} \quad (2.10)$$

Αυτή η στρατηγική εξαλείφει το σφάλμα που σχετίζεται με τη δειγματοληψία μεγάλων τμημάτων του χημικού χώρου, κάτι που θα απαιτούνταν για τον υπολογισμό των απόλυτων ελεύθερων ενεργειών πρόσδεσης, καθώς εξαρτάται μόνο από την ελεύθερη αλλαγή ενέργειας των μετασχηματισμένων τμημάτων του συστήματος A στο σύστημα B στη δεσμευμένη και μη δεσμευμένη κατάσταση. Στο Σχήμα 2.7 φαίνεται πώς διαμορφώνεται ο θερμοδυναμικός κύκλος για τον υπολογισμό της σχετικής ελεύθερης ενέργειας πρόσδεσης ( $\Delta \Delta G$ ) με τη μέθοδο της Διατάραξης Ελεύθερης Ενέργειας.



**Σχήμα 2.7** Θερμοδυναμικός κύκλος για υπολογισμούς σχετικής ελεύθερης ενέργειας πρόσδεσης. Η σχετική διαφορά της ελεύθερης ενέργειας πρόσδεσης μεταξύ των μορίων A και B ( $\Delta\Delta G$ ) μπορεί να υπολογισθεί μέσω δύο δυνατών διαδρομών. Η προσομοίωση της άμεσης διαδρομής (οριζόντιων γραμμών) του  $\Delta G_2 - \Delta G_1$  λόγω της μεγάλης διαφοράς μεταξύ των τελικών καταστάσεων (νερό έναντι πρωτεΐνης). Η αλχημική διαδρομή (κάθετες γραμμές) απαιτεί πολύ μικρότερες διαταραχές στο σύστημα και επομένως τείνει να συγκλίνει πολύ πιο γρήγορα. Εδώ, ο προσδέτης A διαταράσσεται στον προσδέτη B στην δεσμευμένη κατάσταση (δεξιά κατακόρυφη γραμμή) και στην μη δεσμευμένη κατάσταση (αριστερή κατακόρυφη γραμμή), η διαφορά της οποίας είναι ίδια με την άμεση διαδρομή (λόγω του κλειστού θερμοδυναμικού κύκλου). Η εικόνα λήφθηκε από τη βιβλιογραφική πηγή [89].

## 2.7 Σύγκλιση υπολογισμών Διατάραξης Ελεύθερης Ενέργειας

Στο πλαίσιο των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας η διαφορά ελεύθερης ενέργειας Gibbs μεταξύ δύο καταστάσεων 0 και 1 δίνεται από την εξίσωση του Zwanzig:

$$\Delta G_{0 \rightarrow 1} = -k_B T \ln \left\langle \exp \left[ \frac{-(H_1(x) - H_0(x))}{k_B T} \right] \right\rangle_0 \quad (2.11)$$

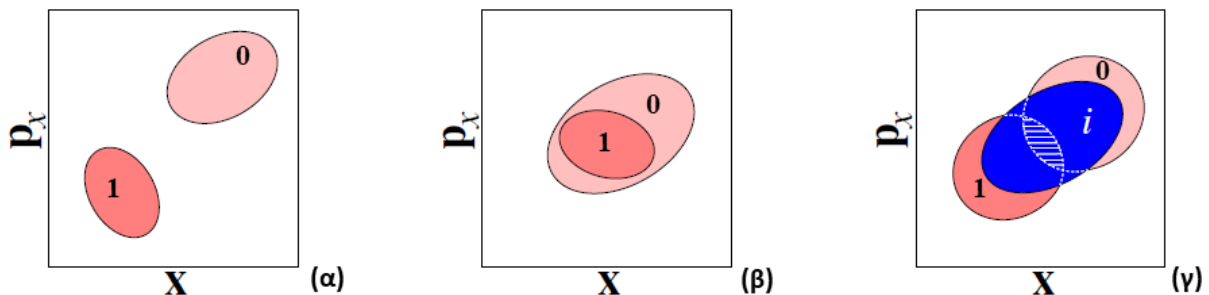
Στην Εξίσωση 2.11, με  $\Delta G_{0 \rightarrow 1}$  συμβολίζεται η ελεύθερη ενέργεια που απαιτείται για τη μετάβαση από την κατάσταση 0 στην κατάσταση 1,  $k_B$  είναι η σταθερά Boltzmann,  $T$  είναι η θερμοκρασία,  $H_0(x)$  και  $H_1(x)$  είναι οι Χαμιλτονιανές που περιγράφουν την ενέργεια των καταστάσεων 0 και 1, ενώ ο όρος  $\langle \dots \rangle_0$  υποδεικνύει ένα σύνολο του μέσου όρου (ensemble average) των διαμορφώσεων που είναι αντιπροσωπευτικές της κατάστασης αναφοράς 0.

Η σύγκλιση της Εξίσωσης 2.11 υποδηλώνει ότι οι διαμορφώσεις χαμηλής ενέργειας (low-energy configurations) της κατάστασης στόχου, δηλαδή της κατάστασης 1, είναι επίσης διαμορφώσεις της αρχικής κατάστασης αναφοράς 0, γεγονός που οδηγεί σε ικανοποιητική επικάλυψη των αντίστοιχων συνόλων (Σχήμα 2.8β). Ο μετασχηματισμός μεταξύ των δύο θερμοδυναμικών καταστάσεων αντικαθίσταται από μια σειρά μετασχηματισμών μεταξύ μη φυσικών, ενδιάμεσων καταστάσεων

κατά μήκος μιας διαδρομής που συνδέει το 0 προς το 1. Αυτή η διαδρομή συχνά καλείται αλχημική και χαρακτηρίζεται από την παράμετρο  $\lambda$ , η οποία λαμβάνει τιμές από 0 έως 1. Με τον τρόπο αυτό η Χαμιλτονιανή του συστήματος και επομένως η ελεύθερη ενέργεια, μετατρέπεται σε μια συνεχή συνάρτηση αυτής της παραμέτρου μεταξύ των καταστάσεων αναφοράς (κατάσταση 0) και στόχου (κατάσταση 1). Η εξάρτηση της ελεύθερης ενέργειας Gibbs του συστήματος από την παράμετρο  $\lambda$  διακρίνεται στην Εξίσωση 2.11.

$$\Delta G_{0 \rightarrow 1} = -\frac{1}{\beta} \sum_{i=1}^N \ln \langle \exp\{-\beta[H(x, \lambda_{i+1}) - H(x, \lambda_i)]\} \rangle_i \quad (2.11)$$

Όπου,  $\beta^{-1} \equiv k_B T$ ,  $N$  ο αριθμός των ενδιάμεσων καταστάσεων μεταξύ της αρχικής και τελικής κατάστασης και  $H$  οι αντίστοιχες Χαμιλτονιανές.



**Σχήμα 2.8** Σύγκλιση υπολογισμών Διατάραξης Ελεύθερης Ενέργειας. **(α)** Αν τα σύνολα που αντιπροσωπεύουν τις καταστάσεις 0 και 1 είναι πολύ διαφορετικά, η Εξίσωση (2.11) δεν θα συγκλίνει. **(β)** Εάν, οι διαμορφώσεις της κατάστασης 1 σχηματίζουν ένα υποσύνολο του συνόλου των διαμορφώσεων που χαρακτηρίζουν την κατάσταση 0, η προσομοίωση αναμένεται να συγκλίνει. **(γ)** Οι δυσκολίες που υπάρχουν στην περίπτωση (α) μπορεί να αντιμετωπιστούν με την εισαγωγή αλληλεπικαλυπτόμενων ενδιάμεσων καταστάσεων που συνδέουν το 0 με το 1. Θα πρέπει να αναφερθεί ότι πρακτικά η κινητική συμβολή  $K(p_x)$  θεωρείται ότι είναι ίδια για την κατάσταση 0 και την κατάσταση 1 [16].

Εάν η επικάλυψη των δύο συνόλων μεταξύ διαδοχικών ενδιάμεσων είναι πολύ υψηλή, τότε η Εξίσωση 2.11 συγκλίνει. Για παράδειγμα, αν ένας αιθέρας αλλάξει σε θειοαιθέρα, υπάρχει σχετικά μικρή αλλαγή στο χώρο φάσης που γίνεται η δειγματοληψία, και η Εξίσωση 2.11 θα συγκλίνει χρησιμοποιώντας ένα μικρό αριθμό ενδιάμεσων. Ωστόσο, εάν εξαφανιστεί ή εμφανιστεί ένα ολόκληρο βαρύ άτομο (δηλ. όχι υδρογόνο) ή εάν η φόρτιση ενός ατόμου αλλάξει σημαντικά, η επικάλυψη δεν θα είναι σημαντική και η (2.11) θα είναι αναποτελεσματική χωρίς μεγάλο αριθμό ενδιάμεσων. Για τα περισσότερα συστήματα που εξετάζονται στο πλαίσιο της απόλυτης ή σχετικής ελεύθερης ενέργειας πρόσδεσης, οι υπολογισμοί με την εξίσωση του Zwanzig είναι μη αποδοτικοί, καθώς απαιτούν την προσομοίωση πολλών καταστάσεων [16], [12].

Υπάρχουν αρκετές προσεγγίσεις για την αντιμετώπιση των ζητημάτων σύγκλισης της Εξίσωσης 2.11. Μεταξύ των προτεινόμενων ορθών πρακτικών για τους υπολογισμούς FEP, συνιστάται ιδιαίτερα να εκτιμηθούν οι διαφορές ελεύθερης ενέργειας χρησιμοποιώντας μια αμφίδρομη προσέγγιση, δηλαδή εκτός από την forward προσομοίωση ( $0 \rightarrow 1$ ) της αλλαγής της ελεύθερης ενέργειας, συνιστάται και η δειγματοληψία της backward προσομοίωσης, δηλαδή ο υπολογισμός της Διατάραξης της Ελεύθερης Ενέργειας από την κατάσταση στόχο προς την κατάσταση αναφοράς ( $1 \rightarrow 0$ ). Η αξιολόγηση της σύγκλισης των υπολογισμών FEP μπορεί να επιτευχθεί με τη μέτρηση του βαθμού αλληλοεπικάλυψης μεταξύ των κατανομών πιθανοτήτων,  $P_0(\Delta U)$  και  $P_1(\Delta U)$  που χαρακτηρίζουν τους forward και backward μετασχηματισμούς, αντίστοιχα [90]. Το 1976 ο Charles Bennett δημιούργησε

μία από τις πιο αποτελεσματικές προσεγγίσεις (Εξίσωση 2.12 και 2.13) για τον υπολογισμό της μεταβολής της ελεύθερης ενέργειας η οποία μάλιστα ελαχιστοποιεί τη διασπορά της ελεύθερης ενέργειας. Η μεθοδολογία αυτή χρησιμοποιείται εκτενώς μέχρι έως σήμερα και ονομάζεται Bennett's Acceptance Ratio (BAR) [91].

$$\exp(\beta \Delta G^{\text{BAR}}) = \frac{\langle f[-\beta(\Delta U - c)] \rangle_1}{\langle f[+\beta(\Delta U - c)] \rangle_0} \exp(+\beta c) \quad (2.12)$$

Όπου,  $f(x) = 1/[1+\exp(c)]$ ,  $c = \Delta G + (1/\beta) \ln(N_1/N_0)$ ,  $N_0$  και  $N_1$  είναι ο αριθμός των διαμορφώσεων που έχει ληφθεί δείγμα από τις καταστάσεις αναφοράς και στόχου, αντίστοιχα.

$$\sigma_{\Delta A}^2 \text{BAR} = \frac{1}{N_0 \beta^2} \left[ \frac{\langle f^2(x) \rangle_0}{\langle f(x) \rangle_0^2} - 1 \right] + \frac{1}{N_1 \beta^2} \left[ \frac{\langle f^2(-x) \rangle_1}{\langle f(-x) \rangle_1^2} - 1 \right] \quad (2.13)$$

Εδώ,  $x = \beta (\Delta U - c)$ .

## 2.8 Υπερυπολογιστικά συστήματα

Οι “Υπερυπολογιστές” είναι υπολογιστικά συστήματα που αξιοποιούνται σε επιστημονικές εφαρμογές οι οποίες απαιτούν την εκτέλεση πολλών εκατομμυρίων πράξεων ή την επεξεργασία μεγάλου όγκου δεδομένων. Λόγω αυτών των απαιτήσεων, τέτοιου είδους προβλήματα είτε θα χρειάζονταν απαγορευτικά μεγάλο χρόνο για να ολοκληρωθούν σε έναν απλό υπολογιστή γραφείου είτε λόγω περιορισμένων πόρων (π.χ. κεντρική μνήμη, αποθηκευτικός χώρος) δεν είναι εφικτό να πραγματοποιηθούν καθόλου. Οι υπερυπολογιστές ξεπερνούν τους περιορισμούς αυτούς χρησιμοποιώντας εξειδικευμένο υλικό τελευταίας τεχνολογίας κάθε χρονική στιγμή, εκμεταλλευόμενοι παράλληλα την υπολογιστική ισχύ από πολλαπλές υπολογιστικές μονάδες. Ένας υπερυπολογιστής σήμερα είναι στην πραγματικότητα ένα σύστημα από εκατοντάδες ή και χιλιάδες υπολογιστές, οι οποίοι αναφέρονται συνήθως ως «κόμβοι», που επικοινωνούν μεταξύ τους χρησιμοποιώντας ένα πολύ γρήγορο δίκτυο και οι οποίοι συνεργατικά μπορούν να επιλύουν προβλήματα με μεγάλη ταχύτητα. Σήμερα οι υπερυπολογιστές χρησιμοποιούνται για να επιλύσουν μερικά από τα σημαντικότερα προβλήματα της ανθρωπότητας, όπως η προέλευση του σύμπαντος, η ανακάλυψη νέων φαρμάκων καθώς και η έρευνα της κλιματικής αλλαγής.

Στην Ελλάδα το Εθνικό Δίκτυο Έρευνας και Τεχνολογίας έχει αναπτύξει και λειτουργεί το υπερυπολογιστικό σύστημα Advanced Research Information System (ARIS) το οποίο αξιοποιείται σε ένα μεγάλο εύρος εφαρμογών, μεταξύ αυτών και προσομοιώσεις μοριακών συστημάτων. Ο ARIS είναι το ισχυρότερο υπολογιστικό σύστημα στην Ελλάδα και τέθηκε σε λειτουργία τον Ιούλιο του 2015 από την ΕΔΕΤ Α.Ε. προσφέροντας ένα ισχυρό εργαλείο έρευνας στην Ελληνική επιστημονική κοινότητα. Το υπολογιστικό σύστημα ARIS σήμερα έχει μέγιστη θεωρητική υπολογιστική ισχύ 444 TFlor/s, μπορεί δηλαδή να εκτελεί 444 τρισεκατομμύρια μαθηματικές πράξεις το δευτερόλεπτο, και προσφέρει πολλαπλές δυνατότητες επεξεργασίας δεδομένων. Ο ARIS συνδυάζει 4 διαφορετικές αρχιτεκτονικές οι οποίες είναι διαμοιρασμένες σε αντίστοιχες “νησίδες κόμβων” [92]. Αναλυτικά, η υποδομή αποτελείται από:

A) Μία νησίδα η οποία διαθέτει 426 υπολογιστικούς κόμβους (thin nodes). Κάθε κόμβος διαθέτει δύο επεξεργαστές (Ivy Bridge - Intel Xeon E5-2680v2) και κάθε επεξεργαστής περιέχει 10

επεξεργαστικούς πυρήνες προσφέροντας έτσι συνολικά 8,520 πυρήνες (CPU cores). Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές υψηλής παραλληλίας που μπορούν να σπάσουν τα δεδομένα τους σε πολλά μικρά κομμάτια πριν τα επεξεργαστούν.

Β) Μια νησίδα κόμβων μεγάλης μνήμης (fat nodes) που αποτελείται από 44 κόμβους. Κάθε κόμβος προσφέρει 4 επεξεργαστές, 40 πυρήνες και 512 GB κεντρικής μνήμης ανά κόμβο. Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές που χρειάζονται πολύ μεγάλη κεντρική μνήμη και όχι τόσο για υψηλή κλιμάκωση.

Γ) Μια νησίδα κόμβων επιταχυντών GPU (gpu nodes) που αποτελείται από 44 κόμβους. Κάθε κόμβος περιέχει 2 επεξεργαστές με 10 πυρήνες ανά επεξεργαστή, 64 GB μνήμης και 2 κάρτες γραφικών GPU NVidia K40. Οι κόμβοι αυτοί είναι κατάλληλοι για εφαρμογές που υλοποιούν υπολογιστικές πράξεις που μπορούν να αξιοποιήσουν τις κάρτες γραφικών ως συνεπεξεργαστές για επιτάχυνση των υπολογισμών.

Δ) Μια νησίδα κόμβων επιταχυντών Xeon Phi (phi nodes) που αποτελείται από 18 κόμβους, καθένας εκ των οποίων περιέχει 2 επεξεργαστές με 10 πυρήνες, 64 GB μνήμης και 2 συνεπεξεργαστές Intel Xeon Phi 7120P. Είναι κατάλληλη για παράλληλες εφαρμογές που αξιοποιούν την τεχνολογία συνεπεξεργαστών της Intel Xeon Phi [92].





### 3. ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

#### 3.1 Υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης για το διαγωνισμό SAMPL6 SAMPLing

##### 3.1.1 Τα συστήματα υποδοχέα-μορίου προσδέτη

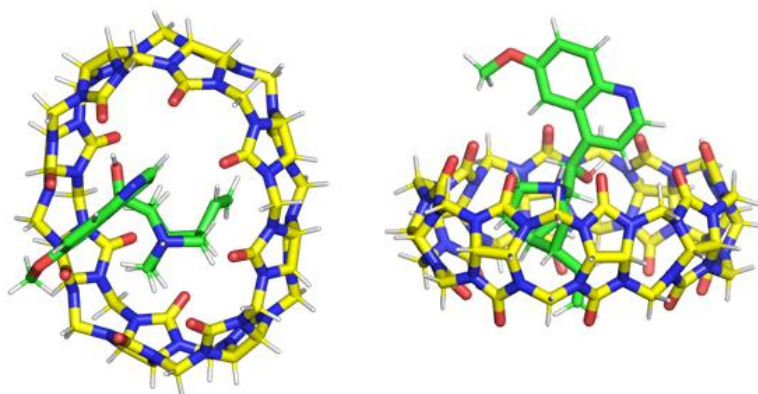
Οι μέθοδοι Μοριακής Μοντελοποίησης παίζουν ολοένα και σπουδαιότερο ρόλο στην εύρεση και το σχεδιασμό μικρών φαρμακευτικών μορίων που δεσμεύονται ισχυρά σε έναν πρωτεϊνικό υποδοχέα. Η δυνατότητα αξιόπιστης πρόβλεψης της ελεύθερης ενέργειας πρόσδεσης ενός μορίου-προσδέτη σε ένα φαρμακευτικό στόχο μπορεί να επιταχύνει ουσιαστικά το σχεδιασμό νέων φαρμάκων, επικεντρώνοντας τις προσπάθειες χημικής σύνθεσης σε μόρια τα οποία δεσμεύονται ισχυρά από την πρωτεΐνη του ενδιαφέροντος. Ωστόσο, πρόσφατες μελέτες υποδεικνύουν πως οι υπάρχουσες μέθοδοι είναι ικανές προβλέψουν την ελεύθερη ενέργεια πρόσδεσης μικρών μορίων με σφάλμα της τάξης του 1-2 kcal/mol από τα αντίστοιχα πειραματικά αποτελέσματα [17], [11]. Αυτή η απόκλιση μπορεί να αποδοθεί στους περιορισμούς που πηγάζουν από τη χρήση των πεδίων δυνάμεων της Μοριακής Μηχανικής, ενώ τα τελευταία έτη έχει επισημανθεί ο ρόλος επιπρόσθετων παραγόντων, όπως η πρόβλεψη των καταστάσεων πρωτονίωσης [49].

Τα υπερμοριακά σύμπλοκα δεκτών-προσδετών έχουν αναδειχθεί ως ένα πρακτικό μοντέλο για την ποσοτική εκτίμηση των σφαλμάτων μοντελοποίησης που χαρακτηρίζουν την αλληλεπίδραση μικρών μορίων-προσδετών με διάφορους υποδοχείς. Οι υπερμοριακοί δέκτες, όπως οι οικογένειες των cucurbiturils, cavitands, και των κυκλοδεξτρινών, είναι συστήματα που μπορούν να δεσμεύσουν μικρά φάρμακα παρουσιάζοντας παρόμοιες ελεύθερες ενέργειες πρόσδεσης με αυτές των συμπλόκων πρωτεΐνης-προσδέτη.

Οι συνήθεις μακρομοριακοί υποδοχείς, όπως οι πρωτεΐνες, χαρακτηρίζονται από σχετικά περίπλοκη αρχιτεκτονική, με αποτέλεσμα να υπάρχει η ανάγκη για προσομοιώσεις μεγάλης κλίμακας (μs έως ms) ώστε να επιτευχθεί η απαραίτητη εξισορρόπηση της δομής τους. Αντίθετα, οι υπερμοριακοί δέκτες αποτελούνται από λίγα άτομα, διακρίνονται από απλή δομή, και επομένως με μικρής κλίμακας προσομοιώσεις (ns) μπορεί να επιτευχθεί η απαραίτητη εξισορρόπηση. Έτσι, μπορεί να θεωρηθεί πως πιθανά ζητήματα σύγκλισης των υπολογισμών ελεύθερης ενέργειας, δε θα αποδίδονται στην μικρή κλίμακα των προσομοιώσεων [18]. Επιπρόσθετα, το μικρό μέγεθος αυτών των συστημάτων επιτρέπει σε πολλές μεθοδολογίες να επωφεληθούν από τους γρήγορους χρόνους προσομοίωσης και έτσι αξιολογούνται ταχύτερα τα σφάλματα που σχετίζονται με το πεδίο δυνάμεων. Ακόμα, οι υψηλές διαλυτότητες αυτών των συστημάτων επιτρέπουν την ακριβή εκτίμηση της ελεύθερης ενέργειας πρόσδεσης μέσω πειραματικών μεθόδων όπως η θερμοδομετρία ισοθερμικής τιτλοδότησης (ITC) [49], [93]. Επιπλέον, η σταθερότητα των υπερμοριακών δεκτών σε ακραίες τιμές pH επιτρέπει τον έλεγχο των καταστάσεων πρωτονίωσης, κάτι που δεν είναι δυνατό με τα συστήματα πρωτεϊνών-προσδέτη, επιτρέποντας ακόμα και το να μη ληφθούν υπόψιν οι καταστάσεις πρωτονίωσης, εάν είναι επιθυμητό. Συνολικά, αυτές οι ιδιότητες έχουν αναδείξει τα συστήματα δεκτών-προσδετών ως αξιόπιστα εργαλεία για την ανάδειξη των προβλημάτων σύγχρονων μεθόδων Μοριακής Μοντελοποίησης μέσω παγκόσμιων διαγωνισμών, όπως η πρόκληση SAMPL6 [49], [50], [55].

### 3.1.1.1 Το σύστημα CB8-G3

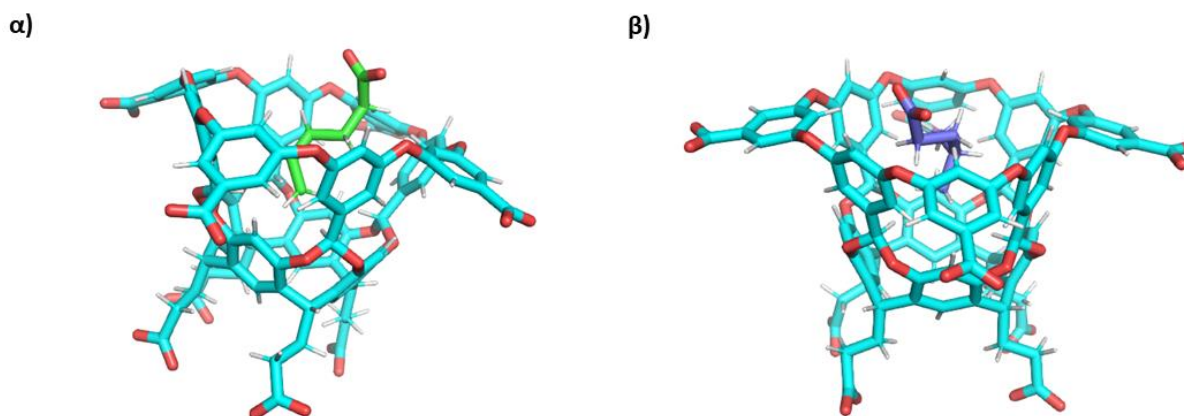
Το cucurbit[8]uril (CB8), είναι ένας συμμετρικός, δακτυλιοειδής δέκτης ο οποίος περιλαμβάνει οκτώ όμοια μονομερή γλυκολουριδίου συνδεδεμένα με ζεύγη γεφυρών μεθυλενίου. Γενικότερα, τα κυκλικά μακρομόρια cucurbiturils λειτουργούν ως μοριακοί υποδοχείς οι οποίοι δημιουργούν σύμπλοκα με ουδέτερα ή θετικά φορτισμένα οργανικά μόρια, μέσω αλληλεπιδράσεων κατιόντος-διπόλου και υδρόφοβων δυνάμεων. Επιπλέον, τα cucurbiturils είναι εμπορικά διαθέσιμα, μη τοξικά, μη αντιδραστικά και επαρκώς διαλυτά σε υδατικό περιβάλλον. Τα χαρακτηριστικά αυτά τα καθιστούν ιδανικούς υποψήφιους για ποικίλλες εφαρμογές στην κατάλυση και τη μεταφορά φαρμάκων [94]. Η κινίνη (quinine), η οποία στο συγκεκριμένο σύστημα δέκτη-προσδέτη συμβολίζεται με G3, είναι ένα φάρμακο που χρησιμοποιείται για τη θεραπεία της ελονοσίας και της μεμπεσίωσης [95].



**Σχήμα 3.1** Το σύστημα προσδέτη-δέκτη CB8-G3. Η πρόκληση SAMPL6 SAMPLing αποσκοπεί στον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης της κινίνης (G3) στον μακρομοριακό υποδοχέα cucurbit[8]uril (CB8).

### 3.1.1.2. Τα συστήματα OA-G3/G6

Τα οκτα-οξέα (octa-acids, OA) είναι συνθετικοί δέκτες με βαθιές, σε σχήμα καλαθιού, υδρόφοβες θέσεις πρόσδεσης. Οι οκτώ καρβοξυλικές όξινες ομάδες καθιστούν αυτούς τους δέκτες υδατοδιαλυτούς, δίχως να αλληλεπιδρούν άμεσα με δεσμευμένους προσδέτες μιας και τους προβάλλουν προς το διαλύτη. Τα οκτα-οξέα τείνουν να δεσμεύουν προσδέτες που διαθέτουν ένα υδρόφοβο τμήμα, το οποίο εισέρχεται εντός της κοιλότητας του δέκτη και ένα υδρόφιλο τμήμα που προεξέχει και αλληλεπιδρά άμεσα με τα μόρια του διαλύτη. Συνεπώς, δεσμεύουν θετικά και αρνητικά φορτισμένα οργανικά μόρια, καθώς και ουδέτερες ενώσεις με ποικίλους βαθμούς πολικότητας [18]. Στο πλαίσιο της πρόκλησης SAMPL6 SAMPLing, χρησιμοποιήθηκαν δύο συστήματα που είχαν ως δέκτη το OA. Το πρώτο σύστημα συμβολίζεται με OA-G3 με το μόριο προσδέτη να είναι το 5-εξενοϊκό οξύ (G3). Το δεύτερο σύστημα στο οποίο ως μόριο δέκτης χρησιμοποιήθηκε το σύστημα OA, συμβολίζεται με OA-G6 και το μόριο προσδέτης είναι το 4-μεθυλοπεντανοϊκό οξύ [49]. Τα δύο οξέα αυτά μοιάζουν με μόρια-θραύσματα (fragment-like molecules) και μπορούν να χρησιμοποιηθούν ως βάσεις για την εύρεση και το σχεδιασμό νέων ενώσεων-οδηγών [96].



**Σχήμα 3.2** Τα σύστημα προσδέτη-δέκτη OA-G3 (**α**) και OA-G6 (**β**). Η πρόκληση SAMPL6 SAMPLing αποσκοπεί στον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης του 5-εξενοϊκού (G3) και του 4-μεθυλοπεντανοϊκού οξέος (G6) στον μακρομοριακό δέκτη οκτα-οξύ (OA).

### 3.1.2 Προετοιμασία των συστημάτων για προσομοίωση

Τα απαραίτητα αρχεία για τους υπολογισμούς απόλυτης ελεύθερης ενέργειας πρόσδεσης προετοιμάστηκαν από τους διοργανωτές του διαγωνισμού SAMPL6. Συγκεκριμένα, προετοιμάστηκαν τα αρχεία συντεταγμένων PDB για τα άτομα του complex και solvent σκέλους της προσομοίωσης, καθώς και αρχεία που περιέχουν επιπρόσθετες πληροφορίες για τη φύση των μορίων, όπως τα αρχεία παραμέτρων (Parameter, PRM), τοπολογίας (Residue Topology File, RTF) και συνδεσιμότητας των ατόμων (Protein Structure File, PSF). Τα αρχεία παραμέτρων και τοπολογίας των ατόμων περιέχουν τις κατάλληλες παραμέτρους για τον υπολογισμό της δυναμικής ενέργειας των μορίων. Το αρχείο RTF είναι συνδεδεμένο με το αρχείο PRM και περιέχει πληροφορίες όπως τύπος ατόμου, μάζα ατόμου, δότες/δέκτες δεσμών υδρογόνου αλλά και τα μερικά ατομικά φορτία για το κάθε άτομο του συστήματος. Παρακάτω περιγράφονται τα βήματα που ακολουθήθηκαν από τους διοργανωτές για την παραγωγή των αρχείων εισόδου της προσομοίωσης.

Αρχικά, οι καταστάσεις πρωτονίωσης των συστημάτων δέκτη-προσδέτη καθορίστηκαν από το εργαλείο Erik 4.0013 [97] της σουίτας Schrödinger2017-2, σε pH=7.4 για το CB8-G3 και σε pH=11.7 για τα συστήματα OA-G3 και OA-G6. Αυτές οι τιμές αντιπροσωπεύουν το pH του ρυθμιστικού διαλύματος που χρησιμοποιήθηκε κατά τις πειραματικές μετρήσεις του διαγωνισμού. Για κάθε σύστημα δέκτη-προσδέτη παράχθηκαν 5 αντίγραφα (replicas) με τη βοήθεια του εργαλείου FRED [98], κάθε ένα από τα οποία διέφερε ως προς την αρχική διαμόρφωση που λάμβανε ο προσδέτης όταν βρισκόταν σε σύμπλοκο με τον δέκτη.

Επόμενο βήμα για την προετοιμασία των συστημάτων για προσομοίωση αποτέλεσε η παραμετροποίηση, δηλαδή η ανάθεση των αριθμητικών σταθερών που απαιτούνται για τον υπολογισμό δυνάμεων και ενεργειών. Οι σταθερές αυτές περιλαμβάνουν μήκη δεσμών, γωνίες έκτασης δεσμών, γωνίες κάμψης, όρους των διεδρων γωνιών, σταθερές δύναμης και παραμέτρους Lennard-Jones. Οι παράμετροι που απαιτούνται για τον υπολογισμό της δυναμικής ενέργειας των ατόμων του συστήματος μπορεί να προέρχονται από πειράματα, από κβαντικομηχανικούς

υπολογισμούς ή και τα δύο. Στην παρούσα περίπτωση, η παραμετροποίηση όλων των συστημάτων του διαγωνισμού SAMPL6 SAMPLing έγινε με το πεδίο δυνάμεων GAFF v1.8 [99] με χρήση του εργαλείου antechamber [100]. Επιπλέον, τα φορτία των συστημάτων υπολογίστηκαν με τη βοήθεια του εργαλείου QUACPAC του λογισμικού OpenEye[101], μέσω της μεθόδου AM1-BCC [102] η οποία βασίζεται σε ημι-εμπειρικά μοντέλα κβαντικής μηχανικής [102]. Στα συστήματα προστέθηκαν TIP3P (Transferable Intermolecular Potential with 3 Points) [103] μόρια νερού σε κυβικό κουτί ακμής 12 Å, με χρήση του πακέτου tleap που είναι ενσωματωμένο στο λογισμικό AmberTools16 [100]. Για να διατηρηθεί ουδέτερο το φορτίο των συστημάτων χρησιμοποιήθηκαν ιόντα χλωρίου και νατρίου τα οποία παραμετροποιήθηκαν με το μοντέλο Joung-Cheatham [104]. Στη συνέχεια, προστέθηκαν περισσότερα ιόντα Na<sup>+</sup> και Cl<sup>-</sup> με αποτέλεσμα να αυξηθεί η ιοντική ισχύς των συστημάτων CB8-G3 και OA-G3/G6 σε 150mM και 60mM, αντίστοιχα, σε μία προσπάθεια να προσομοιωθούν πιστά οι πειραματικές συνθήκες [49].

Στη συνέχεια, η ενέργεια των συστημάτων ελαχιστοποιήθηκε με χρήση του αλγορίθμου L-BFGS [105], [106] και η εξισορρόπηση αυτών πραγματοποιήθηκε για 1 ns με χρήση δυναμικής Langevin[107]. Η προσομοίωση έγινε στους 298.15 K, ενώ η πίεση διατηρήθηκε σταθερή στα 1 atm με χρήση βαροστάτη Monte Carlo με τη βοήθεια του πακέτου OpenMM 7.1.1 [108] και του λογισμικού OpenMMTools [108]. Για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλου εύρους χρησιμοποιήθηκε η μέθοδος PME, με τις van der Waals και τις ηλεκτροστατικές αλληλεπιδράσεις να μη λαμβάνονται υπόψιν ύστερα από 10 Å. Τέλος, με χρήση των εργαλείων InterMol και ParmEd [49], δημιουργήθηκαν τα αρχεία εισόδου PDB, RTF, PRM και PSF τα οποία είναι απαραίτητα για την προσομοίωση των συστημάτων με το υπολογιστικό πακέτο NAMD.

### 3.1.3 Προσομοιώσεις διατάραξης ελεύθερης ενέργειας πρόσδεσης με το υπολογιστικό πακέτο NAMD

Συνολικά πραγματοποιήθηκαν 15 προσομοιώσεις (3 συστήματα δέκτη-προσδέτη, 5 αντίγραφα ανά σύστημα) με το υπολογιστικό πακέτο NAMD (έκδοση 2.12) [109], με σκοπό τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης των συστημάτων προσδέτη-δέκτη. Πριν την έναρξη των προσομοιώσεων, πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας κάθε αντίγραφου για κάθε σύστημα, προκειμένου να βρεθεί σε ενεργειακά ευνοϊκή κατάσταση. Η ελαχιστοποίηση έγινε για 100,000 βήματα προσομοίωσης με τη χρήση ενός αλγορίθμου συζευγμένης κλίσης και αναζήτησης γραμμής. Η ελαχιστοποίηση της ενέργειας εξασφάλισε ότι έχουμε μια λογική αρχική δομή, από άποψη γεωμετρίας και προσανατολισμού του διαλύτη και της διαλυμένης ουσίας.

Για να ξεκινήσει η προσομοίωση Μοριακής Δυναμικής, θα πρέπει να εξισορροπηθεί ο διαλύτης και τα ιόντα γύρω από την εκάστοτε διαλυμένη ουσία. Για το σκοπό αυτό, κάθε σύστημα εξισορροπήθηκε για 1 ns υπό σύνολο NPT (298.15 K και 1 atm) χρησιμοποιώντας βήμα 1 fs. Το υπολογιστικό πακέτο NAMD χρησιμοποιεί τον αλγόριθμο Verlet για την ολοκλήρωση του νόμου κίνησης του Νεύτωνα. Το πεδίο δυνάμεων που χρησιμοποιήθηκε για τις προσομοιώσεις ήταν το GAFF v.1.8, ενώ τα μόρια νερού που περιβάλλουν τα συστήματα προσομοιώθηκαν με το μοντέλο TIP3P. Η συμπεριφορά των ατόμων κοντά στα όρια του κουτιού της προσομοίωσης αναπαραστάθηκε με την εφαρμογή περιοδικών οριακών συνθηκών (periodic boundary conditions), ενώ χρησιμοποιήθηκε η μέθοδος Particle Mesh Ewald (PME) για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων

μεγάλου εύρους, με τις εξής παραμέτρους: PME tolerance =  $10^{-6}$ , PME spline order = 4 and PME grid =  $48 \times 48 \times 48$  [110]. Ο αλγόριθμος SHAKE εφαρμόστηκε για τον περιορισμό όλων των δεσμών που περιλάμβαναν άτομα υδρογόνου [111].

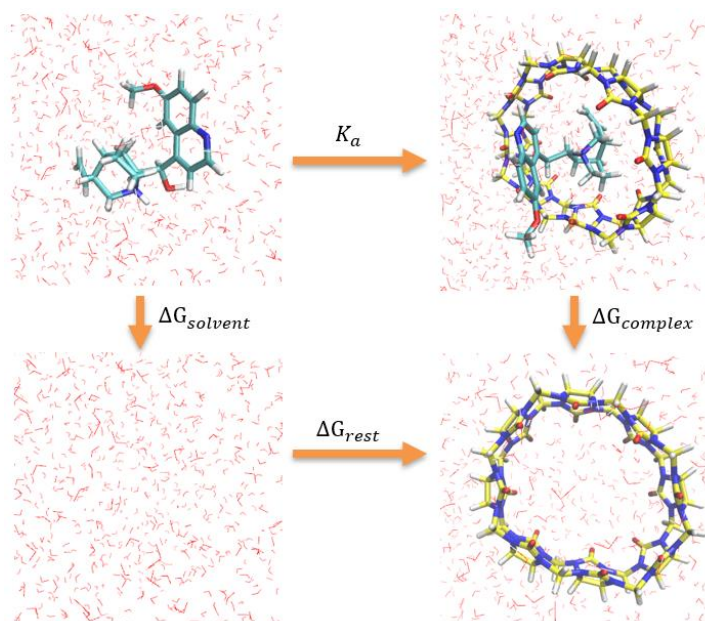
Οι παραγωγικές φάσεις των προσομοιώσεων πραγματοποιήθηκαν υπό σύνολο NPT με πίεση 1 atm και θερμοκρασία 298.15 K. Η πίεση ρυθμίζεται με δυναμική προσαρμογή του μεγέθους της στοιχειώδους κυψελίδας (unit cell) αλλάζοντας την κλίμακα (rescaling) όλων σχεδόν των ατομικών συντεταγμένων κατά τη διάρκεια της προσομοίωσης. Στο πακέτο NAMD, αυτό επιτυγχάνεται με τη χρήση της τροποποιημένης μεθόδου Nosé-Hoover, με την οποία χρησιμοποιείται δυναμική Langevin για τον έλεγχο των διακυμάνσεων του βαρομέτρου. Οι παράμετροι του πιστονιού Nosé-Hoover πίεσης ήταν περίοδος ταλάντωσης 100 fs και κλίμακα χρόνου απόσβεσης 50 fs [112]. Ο έλεγχος θερμοκρασίας επιτεύχθηκε με τη χρήση δυναμικής Langevin, με τον συντελεστή σύζευξης Langevin να είναι  $1.0 \text{ ps}^{-1}$ . Επιπλέον, οι van der Waals και οι ηλεκτροστατικές αλληλεπιδράσεις δε λήφθηκαν υπόψιν μετά τα 10 Å.

Για την αποφυγή των καταστάσεων που συχνά αποκαλούνται «καταστροφές τελικού σημείου», θα πρέπει τα αναπτυσσόμενα σωματίδια να μην επικαλύπτονται με ήδη υπάρχοντα σωματίδια, καθώς στην περίπτωση αυτή το δυναμικό van der Waals θα προσεγγίσει το άπειρο, μιας και η απόσταση αλληλεπίδρασης των ατόμων τείνει στο μηδέν. Για την αποφυγή αυτού του ζητήματος, χρησιμοποιείται ένα τροποποιημένο δυναμικό van der Waals (soft-core potential), το οποίο απωθεί τα επικαλυπτόμενα σωματίδια στις ακραίες τιμές της παραμέτρου  $\lambda$ , δηλαδή όταν  $\lambda \rightarrow 0$  ή  $\lambda \rightarrow 1$ , καθώς τότε είναι πιθανότερη η αλληλοεπικάλυψη των σωματιδίων. Στην προκειμένη περίπτωση επιλέχθηκαν τα 5 Å ως συντελεστής μετατόπισης ακτίνας. Το μέγεθος αυτό είναι υπεύθυνο για την κατασκευή των τροποποιημένων vdW αλληλεπιδράσεων κατά τη διάρκεια των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας με το πακέτο NAMD. Έτσι, διασφαλίζεται ότι το δυναμικό vdW είναι πεπερασμένο παντού ακόμα και για μικρές τιμές  $\lambda$ , γεγονός που βελτιώνει σημαντικά την ακρίβεια και τη σύγκλιση των υπολογισμών FEP, ενώ ταυτόχρονα εμποδίζει τα επικαλυπτόμενα σωματίδια να καταστήσουν ασταθή την προσομοίωση [90], [109].

Στο υπολογιστικό πακέτο NAMD η αρχική κατάσταση, δηλαδή η κατάσταση όπου ο προσδέτης είναι συνδεδεμένος με τον δέκτη, ορίζεται ως  $\lambda=0$ , ενώ η τελική κατάσταση, δηλαδή όταν ο προσδέτης έχει εξαφανιστεί από το σύστημα, ορίζεται για  $\lambda=1$ . Καθώς εξελίσσεται η παραγωγική φάση της προσομοίωσης Μοριακής Δυναμικής, η συνάρτηση της δυναμικής ενέργειας που χαρακτηρίζει την κατάσταση για  $\lambda=0$ , μετατρέπεται σταδιακά σε εκείνη που αντιπροσωπεύει την κατάσταση για  $\lambda=1$ . Για τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης χρησιμοποιούνται συνολικά 32 ενδιάμεσες  $\lambda$  καταστάσεις, με τη μεταβλητή  $\lambda$  να αλλάζει από το 0 προς το 1 με σταθερό βήμα 0.03125. Κάθε ενδιάμεση κατάσταση προσομοιώθηκε ανεξάρτητα για 20 ns και χρονικό βήμα 2 fs, ενώ σε κάθε μία πραγματοποιήθηκαν επιπλέον 2 ns εξισορρόπησης προτού αρχίσει ο υπολογισμός της ελεύθερης ενέργειας.

Οι ηλεκτροστατικές αλληλεπιδράσεις του συστήματος CB8-G3 αποσυνδέονται γραμμικά από την προσομοίωση για το εύρος τιμών  $\lambda$  0 – 0.9. Για  $\lambda=0$ , οι ηλεκτροστατικές αλληλεπιδράσεις είναι πλήρως συζευγμένες στην προσομοίωση και στη συνέχεια μειώνονται γραμμικά με την αύξηση του  $\lambda$ , με τέτοιο τρόπο ώστε σε τιμές  $\lambda$  μεγαλύτερες ή ίσες από 0.9, οι ηλεκτροστατικές αλληλεπιδράσεις να είναι πλήρως αποσυνδεδεμένες από την προσομοίωση. Όσον αφορά τα συστήματα δεκτών OA, οι ηλεκτροστατικές τους αλληλεπιδράσεις αποσυνδέθηκαν στην περιοχή τιμών  $\lambda$  0 – 0.5. Για όλα τα συστήματα δέκτη-προσδέτη, οι αλληλεπιδράσεις van der Waals αποσυνδέονται γραμμικά από

την προσομοίωση για τιμές  $\lambda$  μεταξύ 0 – 1 και σε κάθε περίπτωση, οι αλληλεπιδράσεις vdW είναι πλήρως αποσυνδεδεμένες από την προσομοίωση μόνο όταν  $\lambda=1$ . Κατά τη διάρκεια του complex σκέλους της προσομοίωσης, εφαρμόστηκε αρμονικός περιορισμός με σταθερά δύναμης 100 kcal/mol/Å<sup>2</sup> προκειμένου να αποτραπεί η απομάκρυνση του προσδέτη από τον δέκτη, όταν οι αλληλεπιδράσεις του προσδέτη είναι ασθενώς συζευγμένες στην προσομοίωση. Στο Σχήμα 3.3 παρουσιάζεται ως παράδειγμα ο θερμοδυναμικός κύκλος που χρησιμοποιείται για τους αλχημικούς υπολογισμούς Διατάραξης Ελεύθερης Ενέργειας του συστήματος CB8-G3.



**Σχήμα 3.3** Θερμοδυναμικός κύκλος που χρησιμοποιείται για τον αλχημικό μετασχηματισμό. Η μεθοδολογία FEP μπορεί να χρησιμοποιηθεί για την εξαφάνιση ή τη δημιουργία του προσδέτη (εδώ, κίνηνη) τόσο στην solvent (πάνω αριστερά) όσο και στην complex (πάνω δεξιά) κατάσταση. Από αυτό προκύπτει ότι  $\Delta G_{binding} = -1/\beta \ln K_a = \Delta G_{solvent} - \Delta G_{complex}$ . Για να αποφευχθεί η μετακίνηση του προσδέτη όταν είναι μόνο ασθενώς συνδεδεμένος με τον δέκτη (cucurbit[8]uril), δηλαδή στο τέλος της εξαφάνισης ή στην αρχή της δημιουργίας, θα πρέπει να επιβληθούν περιορισμοί θέσης, που συνεισφέρουν  $\Delta G_{rest}$  στην ολική ελεύθερη ενέργεια πρόσδεσης.

### 3.1.4 Υπολογισμός απόλυτης ελεύθερης ενέργειας πρόσδεσης

Μετά την παραγωγική φάση των προσομοιώσεων Μοριακής Δυναμικής για κάθε φάση και σύστημα, ακολουθεί ο υπολογισμός της απόλυτης ελεύθερης ενέργειας πρόσδεσης με βάση την Εξίσωση 3.1:

$$\Delta G_{bind} = \Delta G_{solvent} - \Delta G_{complex} + \Delta G_{rest} \quad (3.1)$$

Η εξίσωση αυτή απορρέει από το θερμοδυναμικό κύκλο του Σχήματος 3.3, ενώ θα πρέπει να σημειωθεί πως ο όρος  $\Delta G_{rest}$  έχει αμελητέα επίδραση στη συνολική ελεύθερη ενέργεια πρόσδεσης, επομένως δε θα ληφθεί υπόψιν για τον υπολογισμό της ελεύθερης ενέργειας.

Για την ανάλυση των τροχιών που λήφθηκαν από τις προσομοιώσεις της Μοριακής Δυναμικής και την εξαγωγή της διαφοράς ελεύθερης ενέργειας σε κάθε φάση, χρησιμοποιείται το εργαλείο ParseFEP [113] το οποίο είναι ενσωματωμένο στο λογισμικό VMD (έκδοση 1.9.4a29). Το ParseFEP υπολογίζει την ελεύθερη ενέργεια κάθε φάσης με τη μέθοδο BAR [91] για την οποία έγινε λόγος στο

Κεφάλαιο 2.7. Επομένως, η κάθε φάση του κάθε συστήματος προσδέτη-δέκτη (complex και solvent) θα πρέπει να προσομοιωθεί δύο φορές: μία για  $\lambda 0 \rightarrow 1$  (forward προσομοίωση) και μία για  $\lambda 1 \rightarrow 0$  (backward προσομοίωση). Παρ' όλο που είναι υπολογιστικά δαπανηρή μέθοδος, η μέθοδος BAR μπορεί να εκτιμήσει με ακρίβεια την αλλαγή της ελεύθερης ενέργειας, προϋποθέτοντας επαρκή δειγματοληψία των πιθανών διαμορφώσεων του εκάστοτε συστήματος. Το εργαλείο ParseFEP συνδυάζει τα αρχεία εξόδου των forward και backward αλχημικών προσομοιώσεων (FEPOUT files), υπολογίζει αριθμητικά τις ελεύθερες ενέργειες ( $\Delta G$ ) κάθε φάσης και τις συνοδεύει με την αντίστοιχη στατιστική αβεβαιότητα  $\sigma_{\text{BAR}}$  με βάση τις Εξισώσεις 2.12 και 2.13, αντίστοιχα. Ως αρχική εκτίμηση του μεγέθους  $\Delta G$  που περιέχεται στον όρο  $c$  της Εξίσωσης 2.12 χρησιμοποιείται ελεύθερη ενέργεια που έχει υπολογισθεί από την εξίσωση του Zwanzig, ενώ στη συνέχεια μέσω επαναληπτικής διαδικασίας υπολογίζεται η  $\Delta G^{\text{BAR}}$ . Εκτός της ελεύθερης ενέργειας, το ParseFEP παρέχει τις γραφικές παραστάσεις του ιστογράμματος κατανομής πιθανότητας της δυναμικής ενέργειας και της εξέλιξης της ελεύθερης ενέργειας κατά τη διάρκεια της προσομοίωσης, για κάθε ενδιάμεση κατάσταση  $\lambda$  τόσο για την forward, όσο και για την backward προσομοίωση. Η κατανομή πιθανότητας κάθε ενδιάμεσης κατάστασης προσεγγίζεται μέσω σειρών Gram-Charlier. Με τον τρόπο αυτό καθίσταται δυνατή η παρακολούθηση της σύγκλισης της ελεύθερης ενέργειας, καθώς και η ανίχνευση πιθανών προβληματικών καταστάσεων, οι οποίες θα αντικατοπτρίζονται ως μη ικανοποιητική αλληλοεπικάλυψη των ιστογραμμάτων της κατανομής πιθανοτήτων της δυναμικής ενέργειας των αμφίδρομων (forward και backward) υπολογισμών Διατάραξης Ελεύθερης Ενέργειας [113]. Τέλος, για την αξιολόγηση της σύγκλισης της απόλυτης ελεύθερης ενέργειας πρόσδεσης του πακέτου NAMD σε σχέση με τις υπόλοιπες μεθοδολογίες που έλαβαν μέρος στο διαγωνισμό SAMPL6 SAMPLing, υπολογίστηκε η ελεύθερη ενέργεια και η τυπική απόκλιση για κάθε 1%, 2%, 3%, ..., 100% της προσομοίωσης.

## 3.2 Υπολογισμοί σχετικής ελεύθερης ενέργειας πρόσδεσης για τον αναστολέα της Agr2/3

### 3.2.1 Προετοιμασία της δομής της Agr2/3 για προσομοίωση

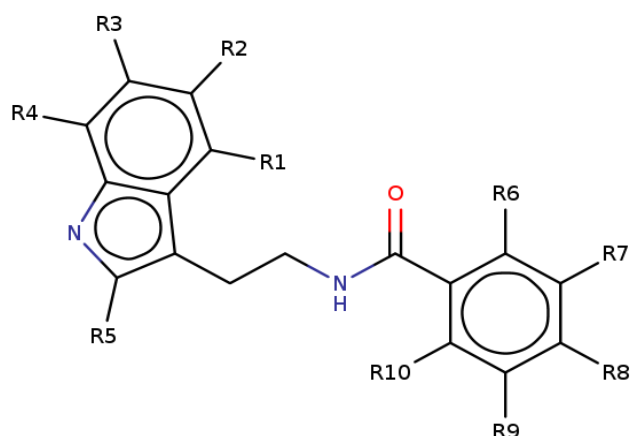
Αρχικό σημείο για την παρούσα μελέτη αποτελεί η δομή του πρωτεϊνικού συμπλέγματος Agr2/3 σε σύμπλοκο με τον αναστολέα CK-666 που είναι διαθέσιμη στην Protein Data Bank με το κωδικό όνομα 3UKR. Η πρωτεΐνη Agr2/3 προέρχεται από τον οργανισμό *Bos taurus* και η δομή της έχει επιλυθεί πειραματικά μέσω Κρυσταλλογραφίας Ακτίνων Χ (X-Ray Diffraction) σε ανάλυση 2.48 Å. Το σύμπλοκο απομονώθηκε σε pH = 7.5 και θερμοκρασία 277.0 K, ενώ οι διαστάσεις της κρυσταλλογραφικής κυψελίδας ήταν 111 Å x 130 Å x 205 Å [64]. Η δομή αυτή απαρτίζεται από 7 αλυσίδες, κάθε μία από τις οποίες αντιστοιχεί σε μία υπομονάδα του πρωτεϊνικού συμπλέγματος Agr2/3: α) chain A, Agr3, β) chain B, Agr2, γ) chain C, ARPC1, δ) chain D, ARPC2, ε) chain E, ARPC3, στ) chain F, ARPC4 και η) chain G, ARPC5. Από το σύμπλοκο αυτό απομονώθηκαν οι δύο πρώτες αλυσίδες που αντιστοιχούν στις πρωτεΐνες Agr3 και Agr2, και αποτελούνται από 418 και 394 αμινοξέα, αντίστοιχα. Η ένωση οδηγός CK-666 είναι προσδεσμένη στη διεπιφάνεια των δύο αυτών πρωτεϊνών, επομένως η μελέτη των μεταξύ τους αλληλεπιδράσεων μπορεί να πραγματοποιηθεί απουσία των υπόλοιπων υπομονάδων του πρωτεϊνικού συμπλέγματος.

Η δομή των Arp2 και Arp3 σε σύμπλοκο με τον αναστολέα CK-666 προετοιμάστηκαν τη χρήση του υπολογιστικού εργαλείου Protein Preparation Wizard της σουίτας Schrödinger [85]. Συνοπτικά: 1) οριοθετήθηκαν οι δεσμοί όλων των ατόμων μέσω ανάθεσης των καταστάσεων πρωτονίωσης, 2) προστέθηκαν τα υδρογόνα που απουσίαζαν, 3) δημιουργήθηκαν οι δισουλφιδικοί δεσμοί μεταξύ των ατόμων θείου της κυστεΐνης που βρέθηκαν σε απόσταση 3.2 Å, 4) ανατέθηκαν οι καταστάσεις ιονισμού των αμινοξέων σε τιμή pH=7.0 με τη χρήση του λογισμικού PROPKA και 5) ελαχιστοποιήθηκε η συνολική ενέργεια της συνολικής δομής. Επιπλέον, από την τριτοταγή δομή της Arp3 απουσίαζαν 18 αμινοξέα (1-2, 41-51, 355-358 και 418), ενώ από τη δομή της Arp2 απουσίαζαν 201 αμινοξέα (1-153 και 347-393) λόγω υψηλής κινητικότητας που δεν επέτρεψε μεγάλη ευκρίνεια στο ηλεκτρονικό τους νέφος. Τα αμινοξέα αυτά προστέθηκαν με τη χρήση του λογισμικού Prime [114]. Στη συνέχεια, ο αναστολέας CK-666 διαχωρίστηκε από την πρωτεΐνη προκειμένου να σχηματιστούν και να προετοιμαστούν οι ανάλογες ενώσεις, ενώ στην πρωτεϊνική δομή αφαιρέθηκαν τα υδρογόνα με το λογισμικό Chimera [115], κάτι το οποίο είναι προαπαιτούμενο για τα μετέπειτα στάδια.

### 3.2.2 Προετοιμασία της δομής των ανάλογων μορίων της ένωσης οδηγού για προσομοίωση

Επόμενο βήμα αποτελεί ο σχεδιασμός και η προετοιμασία της δομής των ανάλογων μορίων της ένωσης οδηγού CK-666. Αρχικά, η δομή των ανάλογων ενώσεων δημιουργήθηκε με το εργαλείο 3D-Builder της σουίτας Schrödinger, χρησιμοποιώντας ως πρότυπο τη δομή του CK-666 που εξήχθη από την προηγούμενη φάση της προετοιμασίας. Με τον τρόπο αυτό, το κοινό μέρος της ανάλογης ένωσης και του CK-666 θα έχει τις ίδιες συντεταγμένες και επομένως τα δύο μόρια θα καταλαμβάνουν την ίδια περιοχή στη θέση πρόσδεσης. Στη συνέχεια προετοιμάστηκαν μόνο οι ανάλογες ενώσεις με χρήση του εργαλείου LigPrep της σουίτας Schrödinger [86]. Πιο συγκεκριμένα, διατηρήθηκε η ίδια χειρομορφία και το ίδιο φορτίο με την ένωση οδηγό, ενώ στη συνέχεια η δομή του κάθε ανάλογου μορίου ευθυγραμμίστηκε με το CK-666. Η ευθυγράμμιση των μορίων έγινε μέσω του εργαλείου Flexible Ligand Alignment και με χρήση των επιλογών Common scaffold alignment και Maximum common substructure. Με τον τρόπο αυτό, και σε συνδυασμό με τη διατήρηση της χειρομορφίας και των συντεταγμένων, εξασφαλίζεται πως η ανάλογη ένωση θα διατηρεί τον ίδιο τρόπο σύνδεσης με το CK-666. Έτσι, τα προφίλ της δυναμικής ενέργειας των δύο μορίων θα αλληλεπικαλύπτονται και επομένως ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης αναμένεται να συγκλίνει. Όσον αφορά τα φορτία, θα πρέπει να διατηρούνται συνολικά ίδια μεταξύ της ανάλογης ένωσης και του μορίου οδηγού, καθώς η εισαγωγή φορτίου στη μέθοδο FEP ενδεχομένως να έχει ως αποτέλεσμα την ασθενή σύγκλιση των υπολογισμών. Στο Σχήμα 3.4 παρατίθεται η γενική χημική δομή των ανάλογων ενώσεων, ενώ στον Πίνακα 3.1 συνοψίζονται οι αλλαγές στην κάθε δομή συγκριτικά με την ένωση οδηγό CK-666.





**Σχήμα 3.4** Γενική δομή των ανάλογων μορίων της ένωσης οδηγού. Με R1, R2, ..., R10 σημειώνονται οι θέσεις υποκατάστασης.

**Πίνακας 3.1** Σύνοψη των 15 ανάλογων ενώσεων του μορίου CK-666. Οι διαφορετικοί υποκαταστάτες R1, R2, ..., R10 σημειώνονται με κόκκινο.

| Χημική ένωση | R1 | R2               | R3 | R4 | R5              | R6 | R7               | R8                              | R9 | R10 |
|--------------|----|------------------|----|----|-----------------|----|------------------|---------------------------------|----|-----|
| CK-666       | H  | H                | H  | H  | CH <sub>3</sub> | F  | H                | H                               | H  | H   |
| AI-003       | H  | H                | H  | OH | CH <sub>3</sub> | F  | H                | H                               | H  | H   |
| AI-007       | H  | H                | H  | Cl | CH <sub>3</sub> | F  | H                | H                               | H  | H   |
| AI-015       | H  | H                | H  | H  | CH <sub>3</sub> | F  | H                | OH                              | H  | H   |
| AI-062       | H  | H                | H  | H  | CH <sub>3</sub> | H  | H                | OCH <sub>3</sub>                | H  | H   |
| AI-064       | H  | H                | H  | H  | CH <sub>3</sub> | H  | H                | CH <sub>3</sub>                 | H  | H   |
| AI-065       | H  | F                | H  | H  | CH <sub>3</sub> | Cl | H                | Cl                              | H  | H   |
| AI-066       | H  | OCF <sub>3</sub> | H  | H  | CH <sub>3</sub> | H  | H                | CO <sub>2</sub> CH <sub>3</sub> | H  | H   |
| AI-067       | H  | H                | H  | H  | CH <sub>3</sub> | H  | H                | F                               | H  | H   |
| AI-068       | H  | H                | H  | H  | CH <sub>3</sub> | H  | Cl               | Cl                              | H  | H   |
| AI-071       | H  | H                | H  | H  | CH <sub>3</sub> | F  | H                | H                               | H  | F   |
| AI-078       | H  | H                | H  | H  | CH <sub>3</sub> | H  | F                | F                               | H  | H   |
| AI-079       | H  | H                | H  | H  | CH <sub>3</sub> | H  | H                | CF <sub>3</sub>                 | H  | H   |
| AI-086       | H  | H                | H  | H  | CH <sub>3</sub> | H  | OCH <sub>3</sub> | H                               | H  | H   |
| AI-093       | H  | H                | H  | Cl | CH <sub>3</sub> | F  | H                | OH                              | H  | H   |
| AI-094       | H  | H                | H  | OH | CH <sub>3</sub> | F  | H                | OH                              | H  | H   |

### 3.2.3 Προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο NAMD

#### 3.2.3.1 Προετοιμασία του συστήματος πρωτεΐνης-προσδέτη για υπολογισμούς Διατάραξης Ελεύθερης Ενέργειας

Έχοντας ακολουθήσει τα παραπάνω βήματα για την προετοιμασία των κύριων δομών της πρωτεΐνης και των ανάλογων ενώσεων, καταλήγει κανείς με δύο αρχεία PDB (ένα για το μόριο, ένα για την δομή της πρωτεΐνης) τα οποία περιέχουν τις καρτεσιανές συντεταγμένες όλων των ατόμων του συστήματος. Ωστόσο για να προχωρήσει κανείς στην προσομοίωση της Μοριακής Δυναμικής θα πρέπει να δημιουργήσει αρχεία τα οποία περιέχουν επιπρόσθετες πληροφορίες για τη φύση των μορίων, όπως για παράδειγμα πληροφορίες σχετικά με τις παραμέτρους που είναι απαραίτητες για

τον υπολογισμό της ενέργειας των μορίων, τα φορτία των ατόμων, αλλά και αρχεία σχετικά με τη σύνθεση και τη συνδεσιμότητα των ατόμων.

Η παραμετροποίηση των πρωτεϊνικών μορίων, είναι μία σχετικά απλή υπόθεση καθώς όλες οι πρωτεΐνες περιλαμβάνουν 20 γνωστά αμινοξέα και έτσι παράμετροι όπως μερικά φορτία και πληροφορίες σχετικά με δεσμούς, γωνίες και τις δίεδρες γωνίες μπορούν να εξαχθούν εύκολα σε κάθε περίπτωση. Ωστόσο, μία πιο περίπλοκη υπόθεση είναι η παραμετροποίηση μικρών μορίων μιας και αυτά δεν έχουν συγκεκριμένες παραμέτρους και επομένως θα πρέπει κάθε φορά να παράγονται εκ νέου τα κατάλληλα αρχεία. [116], [117]. Στο πλαίσιο αυτό, η ομάδα του W.L. Jorgensen ανέπτυξε μία διαδικτυακή εφαρμογή (Σχήμα 3.5) που αναλαμβάνει την παραμετροποίηση των μικρών οργανικών μορίων και προσδετών (έως 200 άτομα). Ο ιστότοπος LigParGen αναλαμβάνει την παραμετροποίηση των μικρών μορίων, με βάση το πεδίο δυνάμεων OPLS-AA (All Atoms Optimized Potentials for Liquid Simulations), καθώς και την εκτίμηση των μερικών φορτίων χρησιμοποιώντας ένα εκ των μοντέλων 1.14\*CM1A ή 1.14\*CM1A-LBCC. Στη συνέχεια, δημιουργεί τα αντίστοιχα αρχεία σε διάφορες μορφές (formats) που αναγνωρίζουν τα πιο δημοφιλή πακέτα προσομοιώσεων Μοριακής Δυναμικής και Monte Carlo. Έτσι, με βάση τα αρχεία PDB για τα ανάλογα μόρια της ένωσης οδηγού και με τη χρήση του LigParGen, επιλέγοντας το μοντέλο φορτίων 1.14\*CM1A, παράγονται τα αρχεία παραμέτρων (PRM) και τοπολογίας ατόμων (RTF), τα οποία περιέχουν όλες τις απαραίτητες παραμέτρους για τον υπολογισμό της δυναμικής ενέργειας των μορίων [118].

LigParGen is a web-based service that provides force field (FF) parameters for organic molecules or ligands, offered by the Jorgensen group. LigParGen provides bond, angle, dihedral, and Lennard-Jones OPLS-AA parameters with 1.14\*CM1A or 1.14\*CM1A-LBCC partial atomic charges. Server provides parameter and topology files for commonly used molecular dynamics and Monte Carlo packages OpenMM, Gromacs, NAMD, CHARMM, LAMMPS, TINKER, CNS/X-PLOR, Q, DESMOND, BOSS and MCPRO. Also, the PQR file is generated. Supported input formats: SMILES, MOL and PDB. Maximum ligand size allowed is 200 atoms. Check this link to use LigParGen software from command-line in your local computer. Please, report any issue clicking in the following link: LigParGen issues

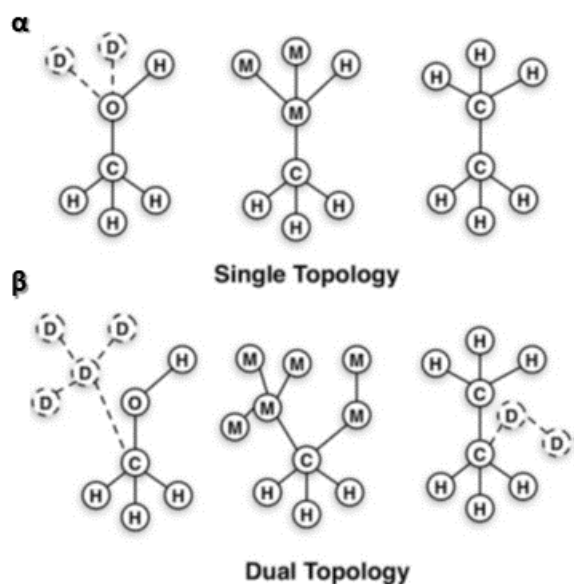
**Step 1: Input structure**  
SMILES  
Enter SMILES Code  
OR upload MOL / PDB file (Structures MUST include all hydrogens)  
Choose File No file chosen

**Step 2: Options**  
Molecule Optimization Iterations 0  
Select charge model:  
 1.14\*CM1A-LBCC (Neutral molecules)  
 1.14\*CM1A (Neutral or Charged molecules)  
Molecule charge 0  
Submit Molecule Sample Benzene

<sup>1</sup> For charged molecules, CM1A charges are NOT scaled by a factor 1.14

**Σχήμα 3.5** Ο ιστότοπος LigParGen αναλαμβάνει την παραμετροποίηση μικρών οργανικών μορίων [118].

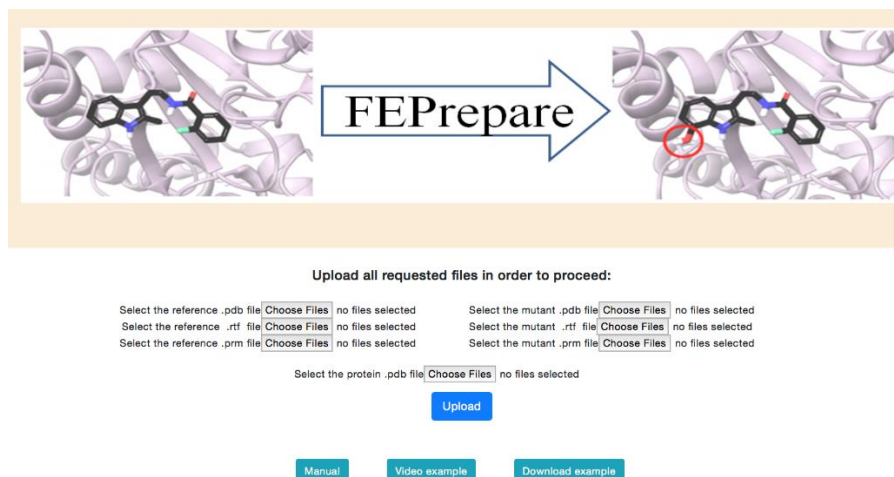
Επόμενο βήμα προετοιμασίας του συστήματος για τις προσομοιώσεις Μοριακής Δυναμικής και τον υπολογισμό της σχετικής ελεύθερης ενέργειας πρόσδεσης, αποτελεί η δημιουργία των καταστάσεων που αναπαρίστανται στο θερμοδυναμικό κύκλο του Σχήματος 2.6. Με άλλα λόγια, θα πρέπει να δημιουργηθούν δύο συστήματα: ένα στο οποίο το μόριο προσδέτης θα είναι σε σύμπλοκο με τον υποδοχέα (complex), στην προκειμένη περίπτωση την πρωτεΐνη Agr2/3, σε ένα περιβάλλον με μόρια νερού και ιόντα, και ένα σύστημα στο οποίο ο προσδέτης θα περιβάλλεται μόνο από τον διαλύτη (solvent). Εκτός αυτού, θα πρέπει να κατασκευαστεί η αληθινή διαδρομή με βάση την οποία ο αναστολέας CK-666 θα μετατρέπεται σε μία από τις ανάλογες ενώσεις στις δύο αυτές καταστάσεις (complex και solvent). Η δημιουργία της αληθινής διαδρομής μπορεί να γίνει με δύο τρόπους: την προσέγγιση μονής τοπολογίας (single topology) και την προσέγγιση διπλής τοπολογίας (dual topology).



**Σχήμα 3.6** Προσέγγιση μονής (α) και διπλής τοπολογίας (β) για την κατασκευή μιας αληθινής διαδρομής μεταξύ αιθανίου και αιθανόλης. Το D αντιπροσωπεύει άτομα τα οποία δεν έχουν αλληλεπιδράσεις (dummy atoms), ενώ το M αναπαριστά μη φυσικά ενδιάμεσα άτομα. Η εικόνα λήφθηκε από τη βιβλιογραφική αναφορά [37]

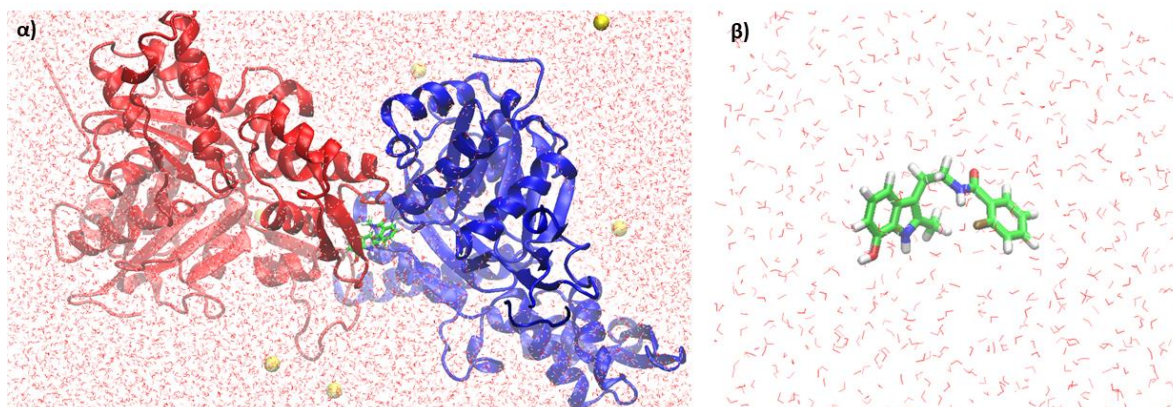
Για την κατασκευή της αληθινής διαδρομής και οι δύο προσεγγίσεις χρησιμοποιούν ορισμένα σωματίδια τα οποία κατά κανόνα δεν αλληλεπιδρούν με το υπόλοιπο σύστημα και ονομάζονται dummy atoms. Στην προσέγγιση μονής τοπολογίας τόσο τα μη-αλληλεπιδρώντα σωματίδια, όσο και τα υπόλοιπα άτομα του συστήματος είναι πάντοτε παρόντα και επιτρέπεται μόνο η άμεση μετατροπή του ενός ατόμου σε ένα άλλο. Στο παράδειγμα του Σχήματος 3.6α, στη μία τελική κατάσταση δύο άτομα υδρογόνου μετατρέπονται σε dummy atoms τα οποία δεν αλληλεπιδρούν με το υπόλοιπο σύστημα, ενώ το άνω βαρύ άτομο είναι το οξυγόνο. Στην άλλη τελική κατάσταση, όλα τα άτομα αλληλεπιδρούν μεταξύ τους, ενώ το άνω βαρύ άτομο έχει μετατραπεί σε άνθρακα. Η εναλλακτική προσέγγιση διπλής τοπολογίας διαφέρει στο ότι κανένα άτομο δεν αλλάζει τύπο και απλώς μετατρέπονται από dummy atoms σε άτομα που αλληλεπιδρούν πλήρως με το περιβάλλον τους. Στην περίπτωση του Σχήματος 3.6β, στην τελική κατάσταση όπου είναι σχηματισμένη η μεθανόλη, η μεθυλομάδα δεν αλληλεπιδρά με το υπόλοιπο σύστημα. Αντίθετα στην τελική κατάσταση όπου σχηματίζεται το αιθάνιο, η υδροξυλομάδα δεν αλληλεπιδρά με το υπόλοιπο σύστημα [37].

Το υπολογιστικό πακέτο NAMD [109] επιτρέπει τη χρήση της προσέγγισης της διπλής τοπολογίας, ενώ το πακέτο GROMACS [119], για το οποίο θα γίνει λόγος παρακάτω, χρησιμοποιεί ένα συνδυασμό αυτών των δύο προσεγγίσεων. Η δημιουργία των κατάλληλων αρχείων που περιέχουν τις παραμέτρους, την τοπολογία και τις συντεταγμένες των τελικών καταστάσεων που περιγράφονται στο Σχήμα 2.6 γίνεται μέσω της διαδικτυακής εφαρμογής FEPprepare [120]. Ο ισότοπος FEPprepare απαιτεί ως είσοδο τα αρχεία PDB, PRM, και RTF για την ένωση-οδηγό και το ανάλογο μόριο (τα οποία έχουν ήδη παραχθεί από τον LigParGen), το αρχείο PDB με τη δομή της πρωτεΐνης απαλλαγμένο από υδρογόνα, και στη συνέχεια παράγει τα απαραίτητα αρχεία για την προσομοίωση Διατάραξης Ελεύθερης Ενέργειας με το πακέτο NAMD.



**Σχήμα 3.7** Ο ιστότοπος FEPPrepare αναλαμβάνει την προετοιμασία των αρχείων για προσομοιώσεις με το υπολογιστικό πακέτο NAMD [120].

Ο FEPPrepare χρησιμοποιεί μια σειρά κώδικα Python για να συνδυάσει κατάλληλα τα αρχεία ούτως ώστε να εφαρμοστεί η διπλή τοπολογία, και στη συνέχεια επικαλείται τα πακέτα psfgen, solvate και autoionize του λογισμικού VMD (Visual Molecular Dynamics) [121]. Το πακέτο psfgen χρησιμοποιεί τα αρχεία τοπολογίας που έχουν παραχθεί για τα δύο συστήματα (complex και solvent) και αναλαμβάνει τη δημιουργία των αρχείων PSF (Protein Structure File) τα οποία περιέχουν πληροφορίες σχετικά με τη συνδεσιμότητα των ατόμων. Στη συνέχεια τα πακέτα solvate και autoionize αναλαμβάνουν την τοποθέτηση του κάθε συστήματος σε ένα πλαίσιο μορίων νερού και ιόντων, προκειμένου να προσομοιωθούν οι βιολογικές συνθήκες (Σχήμα 3.5). Το πακέτο solvate αναλαμβάνει την επιδιαιλύτωση των συστημάτων με μόρια νερού, τα οποία τοποθετούνται σε απόσταση 10 Å από το άτομο με την μεγαλύτερη συντεταγμένη προς κάθε κατεύθυνση, ενώ το πακέτο autoionize τοποθετεί τα απαραίτητα ιόντα ώστε να διατηρηθεί ηλεκτρικά ουδέτερο το περιβάλλον των συστημάτων [121]. Αξίζει να αναφερθεί πώς το μέγεθος του πλαισίου νερού επιλέχθηκε ώστε να μην επιτρέπει στην πρωτεΐνη να αλληλεπιδρά με την «εικόνα» της στο δίπλα κουτί, καθώς χρησιμοποιούνται περιοδικές οριακές συνθήκες. Στην προκειμένη περίπτωση τοποθετήθηκαν ιόντα  $\text{Na}^+$ , ενώ η προσομοίωση των ιδιοτήτων των μορίων του νερού έγινε με το μοντέλο TIP3P [103].



**Σχήμα 3.8** Απεικόνιση των προετοιμασμένων δομών. **(α)** Σύμπλοκο πρωτεΐνης-προσδέτη σε περιβάλλον νερού και ιόντων νατρίου (complex). **(β)** Η ένωση οδηγός ελεύθερη στο αντίστοιχο περιβάλλον (solvent). Τα μόρια νερού παριστάνονται με κόκκινο χρώμα, ενώ τα ιόντα νατρίου απεικονίζονται με κίτρινο.

### 3.2.3.2 Προσομοιώσεις Μοριακής Δυναμικής στις ανάλογες ενώσεις

Μετά την προετοιμασία των απαραίτητων αρχείων ακολούθησαν οι προσομοιώσεις Μοριακής Δυναμικής, με σκοπό να εξαχθούν οι τροχιές κίνησης των ατόμων συναρτήσει του χρόνου και στη συνέχεια να υπολογισθεί η σχετική ελεύθερη ενέργεια πρόσδεσης των ανάλογων ενώσεων. Οι προσομοιώσεις πραγματοποιήθηκαν με τη χρήση της έκδοσης 2.12 του NAMD [109]. Συνολικά έλαβαν χώρα 30 προσομοιώσεις, μιας οι ανάλογες ενώσεις είναι 15 και κάθε μόριο πρέπει να προσομοιωθεί σε δύο φάσεις (complex και solvent).

Πριν την έναρξη των προσομοιώσεων, πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας κάθε συστήματος, προκειμένου να βρεθεί σε ενεργειακά ευνοϊκή κατάσταση. Η ελαχιστοποίηση έγινε για 50,000 βήματα προσομοίωσης με τη χρήση ενός αλγορίθμου συζευγμένης κλίσης και αναζήτησης γραμμής. Στη συνέχεια, ακολούθησαν δύο φάσεις εξισορρόπησης. Η πρώτη φάση διεξάγεται υπό σύνολο NVT για χρονική διάρκεια 0.5 ns με σκοπό την εξισορρόπηση της θερμοκρασίας. Η δεύτερη φάση περιλαμβάνει εξισορρόπηση NPT για επίσης 0.5 ns, η οποία αποσκοπεί στην εξισορρόπηση της πίεσης του συστήματος. Το χρονικό βήμα και στις δυο περιπτώσεις ήταν 1 fs.

Το υπολογιστικό πακέτο NAMD χρησιμοποιεί τον αλγόριθμο Verlet για την ολοκλήρωση του νόμου κίνησης του Νεύτωνα. Το πεδίο δυνάμεων που χρησιμοποιήθηκε για την πρωτεΐνη και τις ανάλογες ενώσεις είναι το OPLS [122] και το OPLS-AA [118] (από τον LigParGen), αντίστοιχα. Τα μόρια νερού που περιβάλλουν τις πρωτεϊνικές δομές προσομοιώθηκαν με το μοντέλο TIP3P [103]. Επιπλέον, εφαρμόστηκαν περιοδικές οριακές συνθήκες, ενώ χρησιμοποιήθηκε η μέθοδος PME για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλου εύρους, με τις εξής παραμέτρους: PME tolerance =  $10^{-6}$ , PME spline order = 4 and PME grid spacing = 1.0 Å [110]. Επίσης, ο αλγόριθμος SHAKE εφαρμόστηκε για τον περιορισμό όλων των δεσμών που περιλάμβαναν άτομα υδρογόνου [111].

Οι παραγωγικές φάσεις των προσομοιώσεων πραγματοποιήθηκαν υπό σύνολο NPT με πίεση 1 atm και θερμοκρασία 310 K. Η πίεση ρυθμίζεται με χρήση τροποποιημένης μεθόδου Nosé-Hoover, με την οποία χρησιμοποιείται δυναμική Langevin για τον έλεγχο των διακυμάνσεων του βαρομέτρου. Οι παράμετροι του πιστονιού Nosé-Hoover πίεσης ήταν περίοδος ταλάντωσης 100 fs και κλίμακα χρόνου απόσβεσης 50 fs [112]. Ο έλεγχος θερμοκρασίας επιτεύχθηκε με χρήση δυναμικής Langevin, με τον συντελεστή σύζευξης Langevin να είναι  $1.0 \text{ ps}^{-1}$ . Επιπρόσθετα, οι ηλεκτροστατικές και οι van der Waals αλληλεπιδράσεις δε λαμβάνονται υπόψιν ύστερα από 12 Å.

Όπως προαναφέρθηκε, το NAMD χρησιμοποιεί την προσέγγιση διπλής τοπολογίας, στην οποία καθορίζονται ταυτόχρονα τόσο η αρχική κατάσταση ( $\lambda=0$ ) όσο και η τελική κατάσταση ( $\lambda=1$ ). Καθώς εξελίσσεται η προσομοίωση Μοριακής Δυναμικής, η συνάρτηση της δυναμικής ενέργειας χαρακτηριστική του  $\lambda=0$  είναι μετατρέπεται σταδιακά σε αυτή που αντιπροσωπεύει την κατάσταση για  $\lambda=1$ . Για την αλχημική μετατροπή της ένωσης οδηγού CK-666 σε κάθε ένα από τα ανάλογα μόρια, χρησιμοποιήθηκαν συνολικά 16 ενδιάμεσες καταστάσεις, δηλαδή η μεταβλητή  $\lambda$  αλλάζει από το 0 προς το 1 με σταθερό βήμα 0.0625. Κάθε ενδιάμεση κατάσταση προσομοιώθηκε για 5 ns με χρονικό βήμα 1 fs. Σε κάθε ενδιάμεση  $\lambda$  κατάσταση, πραγματοποιούνται 500,000 βήματα εξισορρόπησης πριν αρχίσει η παραγωγική φάση της προσομοίωσης κατά την οποία υπολογίζεται η ελεύθερη ενέργεια του κάθε συστήματος.

Οι ηλεκτροστατικές αλληλεπιδράσεις ενός σωματιδίου που εμφανίζεται, «εισέρχονται» στην προσομοίωση γραμμικά για το εύρος τιμών  $\lambda$  0.5–1.0. Σε τιμές  $\lambda$  μικρότερες ή ίσες από το εύρος αυτό,

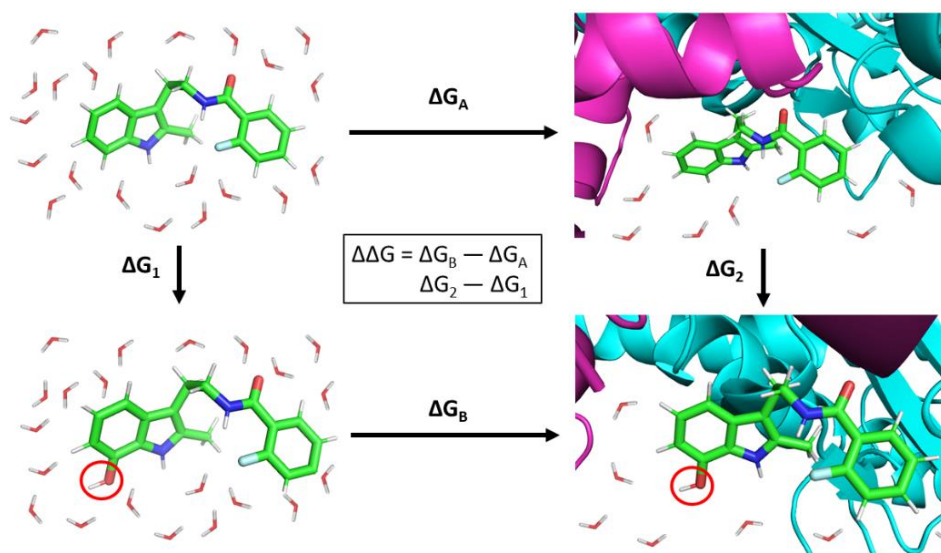
οι ηλεκτροστατικές αλληλεπιδράσεις του εμφανιζόμενου σωματιδίου είναι πλήρως αποσυνδεδεμένες από την προσομοίωση. Στα σωματίδια που εξαφανίζονται, οι ηλεκτροστατικές τους αλληλεπιδράσεις αποσυνδέονται γραμμικά από την προσομοίωση για εύρος τιμών  $\lambda$  0–0.5. Έτσι στην προκειμένη περίπτωση για  $\lambda=0$  οι ηλεκτροστατικές αλληλεπιδράσεις είναι πλήρως συζευγμένες στην προσομοίωση, ενώ στη συνέχεια μειώνονται γραμμικά μέχρι την τιμή  $\lambda=0.5$ .

Οι van der Waals αλληλεπιδράσεις των σωματιδίων που εμφανίζονται είναι πλήρως αποσυνδεδεμένες στην αρχή της προσομοίωσης, δηλαδή όταν για τιμή  $\lambda=0$ . Η σύζευξη των van der Waals αλληλεπιδράσεων στην προσομοίωση γίνεται με αύξηση της τιμής του  $\lambda$ , με τέτοιο τρόπο ούτως ώστε οι vdW αλληλεπιδράσεις των εμφανιζόμενων σωματιδίων να είναι πλήρως συζευγμένες στην προσομοίωση όταν  $\lambda=1$ . Το ακριβώς αντίστροφο ισχύει για τα σωματίδια που εξαφανίζονται κατά τη διάρκεια της προσομοίωσης. Τέλος, για να εξασφαλιστεί ότι το δυναμικό θα είναι πεπερασμένο καθ' όλη τη διάρκεια της προσομοίωσης εφαρμόστηκε το τροποποιημένο van der Waals δυναμικό, επιλέγοντας ως συντελεστή μετατόπισης ακτίνας  $\lambda$  τα 5 Å. [109], [90].

### 3.2.3.3. Υπολογισμός σχετικής ελεύθερης ενέργειας πρόσδεσης μεταξύ ανάλογων ενώσεων

Μετά την παραγωγική φάση των προσομοιώσεων Μοριακής Δυναμικής, ακολουθεί ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης. Στο Σχήμα 3.9 φαίνεται ο θερμοδυναμικός κύκλος που χρησιμοποιείται για την περίπτωση των ανάλογων ενώσεων του αναστολέα CK-666, ενώ η σχετική ελεύθερη ενέργεια πρόσδεσης προκύπτει από την Εξίσωση 3.2. Έτσι, εάν με βάση την Εξίσωση 3.2 η σχετική ελεύθερη ενέργεια  $\Delta\Delta G$  μεταξύ του CK-666 και μίας από τις ανάλογες ενώσεις βρεθεί αρνητική, τότε σημαίνει πως ευνοείται η πρόσδεση της ανάλογης ένωσης έναντι του CK-666.

$$\Delta\Delta G = \Delta G_{\text{complex}} - \Delta G_{\text{solvent}} \quad (3.2)$$



**Σχήμα 3.9** Θερμοδυναμικός κύκλος για υπολογισμούς σχετικής ελεύθερης ενέργειας πρόσδεσης για τις ανάλογες ενώσεις του CK-666. Η σχετική διαφορά της ελεύθερης ενέργειας πρόσδεσης μεταξύ του CK-666 (επάνω) και του κάθε ανάλογου μορίου, στην προκειμένη περίπτωση του ai003 (κάτω) υπολογίζεται με βάση την αλχημική διαδρομή που αναπαρίσταται από τις κάθετες γραμμές.

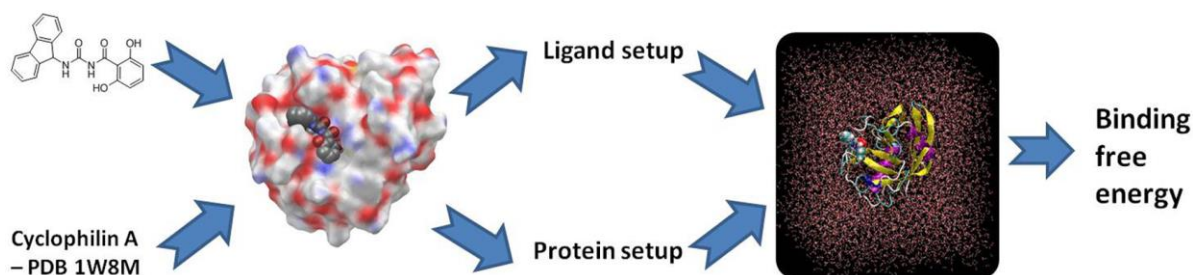
Για την ανάλυση των τροχιών που λήφθηκαν από τις προσομοιώσεις της Μοριακής Δυναμικής με το πακέτο NAMD και την εξαγωγή της διαφοράς ελεύθερης ενέργειας για κάθε φάση θα χρησιμοποιηθεί το εργαλείο ParseFEP, το οποίο είναι ενσωματωμένο στο λογισμικό VMD (έκδοση 1.9.4a29) και υπολογίζει την ελεύθερη ενέργεια κάθε φάσης με τη μέθοδο BAR [91]. Συνεπώς, η κάθε φάση του κάθε συστήματος προσδέτη-δέκτη θα πρέπει να προσομοιωθεί δύο φορές: μία για  $\lambda$  0→1 (forward προσομοίωση) και μία  $\lambda$  1→0 (backward προσομοίωση). Στη συνέχεια, συνδυάζονται τα αρχεία εξόδου των forward και backward αλχημικών προσομοιώσεων και υπολογίζονται οι ελεύθερες ενέργειες  $\Delta G$  κάθε φάσης, καθώς και η αντίστοιχη στατιστική αβεβαιότητα  $\sigma_{\text{BAR}}$ . Επιπλέον, για την κάθε ενδιάμεση κατάσταση  $\lambda$  που προσομοιώνεται (forward και backward), παρέχονται οι γραφικές παραστάσεις κατανομής πιθανότητας της δυναμικής ενέργειας και της εξέλιξης της ελεύθερης ενέργειας κατά τη διάρκεια της προσομοίωσης.

### 3.2.4 Προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο GROMACS

#### 3.2.4.1 Προετοιμασία του συστήματος πρωτεΐνης-προσδέτη για υπολογισμούς Διατάραξης Ελεύθερης Ενέργειας

Έχοντας ακολουθήσει τα βήματα που περιγράφονται στα Κεφάλαια 3.2.1 και 3.2.2 για την προετοιμασία των κύριων δομών της πρωτεΐνης και των ανάλογων ενώσεων, καταλήγει κανείς με δύο αρχεία συντεταγμένων PDB, όπου το ένα να αφορά την ανάλογη ένωση και το άλλο σχετίζεται με την δομή της πρωτεΐνης. Όπως προαναφέρθηκε, για να προχωρήσει κανείς στην προσομοίωση της Μοριακής Δυναμικής θα πρέπει να παραχθούν τα απαραίτητα αρχεία παραμέτρων, καθώς και να προετοιμαστούν τα δύο αρχεία PDB κατάλληλα ούτως ώστε να έχει δημιουργήσει τα συστήματα που απεικονίζονται στο Σχήμα 3.9. Στην περίπτωση των προσομοιώσεων Μοριακής Δυναμικής με το υπολογιστικό πακέτο GROMACS, την προετοιμασία αυτή την αναλαμβάνει το εργαλείο FESetup το οποίο δημιουργήθηκε το 2015 από τον Loeffler και τους συνεργάτες του [123]. Το FESetup (Σχήμα 3.10) είναι ένα εργαλείο γραμμένο σε κώδικα Python που στοχεύει στην επιτάχυνση και διευκόλυνση της προετοιμασίας των προσομοιώσεων Διατάραξης Ελεύθερης Ενέργειας για διάφορα υπολογιστικά πακέτα Μοριακής Δυναμικής όπως GROMACS, AMBER και SOMD.

Το FESetup χρησιμοποιεί τα εξωτερικά προγράμματα του υπολογιστικού πακέτου AmberTools [100], antechamber και leap, για τη δημιουργία των αρχείων τοπολογίας και την παραμετροποίηση των προσδετών. Στην προκειμένη περίπτωση χρησιμοποιήθηκε το πεδίο δυνάμεων GAFF (General Amber Force Field) [99] για την παραμετροποίηση των προσδετών, ωστόσο υποστηρίζονται όλα τα πεδία δυνάμεων AMBER (όχι όμως το OPLS). Τα φορτία των μορίων υπολογίστηκαν μέσω της ημι-εμπειρικής μεθόδου AM1-BCC [102]. Μετά την παραμετροποίηση των μορίων και την ανάθεση των μερικών φορτίων, χρησιμοποιείται το εργαλείο leap για τη δημιουργία του κουτιού προσομοίωσης, καθώς και για τη δημιουργία των απαραίτητων αρχείων τοπολογίας και συντεταγμένων.



**Σχήμα 3.10** Το υπολογιστικό εργαλείο FESetup αυτοματοποιεί την προετοιμασία των αλχημικών προσομοιώσεων ελεύθερης ενέργειας [123].

Για τη δημιουργία της αλχημικής διαδρομής με βάση την οποία ο αναστολέας θα μετατρέπεται σε μία από τις ανάλογες ενώσεις, το υπολογιστικό πακέτο GROMACS χρησιμοποιεί μια υβριδική προσέγγιση η οποία επιτρέπει την ταυτόχρονη αντιστοίχιση μονής και διπλής τοπολογίας στην ίδια περιοχή. Το εργαλείο FESetup χαρτογραφεί την περιοχή αυτή και δημιουργεί το υβριδικό μόριο (morph) με χρήση ενός αλγορίθμου Maximum Common Substructure Search (MCSS) και συγκεκριμένα του fmcs που είναι ουσιαστικά ένας συνδεδεμένος MCSS. Όσον αφορά την πρωτεΐνη, το εργαλείο FESetup αναλαμβάνει την παραμετροποίηση της με το πεδίο δυνάμεων AMBER ff14 SB [124]. Στη συνέχεια το εργαλείο FESetup συνδυάζει την δομή της Apr2/3 με το υβριδικό μόριο και προσθέτει τα απαραίτητα μόρια νερού και ιόντα, ούτως ώστε να προσομοιωθούν οι κυτταρικές συνθήκες. Στην περίπτωση αυτή, αλλά και στην περίπτωση του επιδιαλυτωμένου προσδέτη απουσία της πρωτεΐνης, η προσθήκη διαλύτη γίνεται σε απόσταση 12 Å από το άτομο με την μεγαλύτερη συντεταγμένη προς κάθε κατεύθυνση [123]. Όπως και στις προσομοιώσεις με το NAMD, τοποθετήθηκαν ιόντα Na<sup>+</sup>, ενώ η προσομοίωση των ιδιοτήτων των μορίων του νερού έγινε με το μοντέλο TIP3P[103].

### 3.2.4.2 Προσομοιώσεις Μοριακής Δυναμικής στις ανάλογες ενώσεις

Μετά την προετοιμασία των απαραίτητων αρχείων ακολούθησαν οι προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο GROMACS (έκδοση 5.1.4) [119], ώστε να υπολογισθεί η σχετική ελεύθερη ενέργεια πρόσδεσης των ανάλογων ενώσεων. Υπενθυμίζεται πως συνολικά έλαβαν χώρα 30 προσομοιώσεις (15 ανάλογες ενώσεις, 2 φάσεις ανά ένωση). Για την αλχημική μετατροπή της ένωσης οδηγού CK-666 σε κάθε ένα από τα ανάλογα μόρια, χρησιμοποιήθηκαν συνολικά 21 ενδιάμεσες καταστάσεις. Κάθε ενδιάμεση λ-κατάσταση προσομοιώθηκε ξεχωριστά και περιλάμβανε 4 διακριτά στάδια: 1) την ελαχιστοποίηση της ενέργειας, 2) την NVT εξισορρόπηση, 3) την NPT εξισορρόπηση και 4) την παραγωγική φάση, κατά την οποία πραγματοποιείται ο υπολογισμός της Διατάραξης Ελεύθερης Ενέργειας.

Η ελαχιστοποίηση της ενέργειας πραγματοποιήθηκε για κάθε ένα από τα συστήματα, προκειμένου να βρεθεί σε ενεργειακά ευνοϊκή κατάσταση. Η ελαχιστοποίηση της ενέργειας του κάθε συστήματος έγινε με τη χρήση ενός αλγορίθμου steepest descent, ο οποίος έχει ως αρχικό βήμα 0.01 nm και θεωρεί πως έχει επιτευχθεί σύγκλιση όταν η μέγιστη δύναμη που ασκείται σε κάποιο άτομο είναι μικρότερη από 1000 kJ/(mol·nm). Στη συνέχεια ακολουθούν δύο φάσεις εξισορρόπησης με την πρώτη φάση να διεξάγεται υπό σύνολο NVT με σκοπό την εξισορρόπηση της θερμοκρασίας, ενώ η δεύτερη φάση περιλαμβάνει εξισορρόπηση NPT και στοχεύει στην εξισορρόπηση της πίεσης του συστήματος. Κάθε μία από τις προσομοιώσεις πραγματοποιήθηκε για 0.5 ns και το χρονικό βήμα ήταν 1 fs.

Για την ολοκλήρωση του νόμου κίνησης του Νεύτωνα με το υπολογιστικό πακέτο GROMACS, χρησιμοποιήθηκε ο στοχαστικός αλγόριθμος leap-frog. Το πεδίο δυνάμεων που χρησιμοποιήθηκε για την παραμετροποίηση της πρωτεΐνης και των ανάλογων ενώσεων είναι το AMBER ff14 SB [124] και το GAFF [99], αντίστοιχα. Η συμπεριφορά των ατόμων κοντά στα όρια του κουτιού αναπαραστάθηκε με την εφαρμογή περιοδικών οριακών συνθηκών, ενώ χρησιμοποιήθηκε η μέθοδος PME για τον υπολογισμό των ηλεκτροστατικών αλληλεπιδράσεων μεγάλου εύρους, με τις παρακάτω παραμέτρους: PME tolerance = 10<sup>-6</sup>, PME spline order = 6 and fourierspacing 0.12 nm [110]. Επίσης,



εφαρμόστηκε ο αλγόριθμος LINCS για τον περιορισμό όλων των δεσμών που περιλάμβαναν άτομα υδρογόνου [125].

Οι παραγωγικές φάσεις των προσομοιώσεων πραγματοποιήθηκαν υπό σύνολο NPT με πίεση 1 atm και θερμοκρασία 310 K. Η πίεση του συστήματος διατηρήθηκε σταθερή με χρήση του βαροστάτη Parinello-Rahman, με σταθερά χρόνου 1 ps και συμπιεστότητα  $4.5 \cdot 10^{-5} \text{ bar}^{-1}$  [126], ενώ ο έλεγχος της θερμοκρασίας επιτεύχθηκε με τη χρήση δυναμικής Langevin, με συντελεστή σύζευξης Langevin 1.0 ps<sup>-1</sup>. Όπως και στην περίπτωση του NAMD, οι ηλεκτροστατικές και οι van der Waals αλληλεπιδράσεις δε λαμβάνονται υπόψιν ύστερα από 12 Å.

Το υπολογιστικό πακέτο GROMACS εφαρμόζει μια υβριδική προσέγγιση για την αλχημική μετατροπή του αναστολέα CK-666 σε μία από τις ανάλογες ενώσεις. Έτσι, στην περίπτωση αυτή τόσο στα complex, όσο και στα solvent συστήματα, οι van der Waals και οι ηλεκτροστατικές αλληλεπιδράσεις διαταράχθηκαν ξεχωριστά σε ένα εύρος 21 λ-καταστάσεων. Αρχικά, έγινε μετατροπή των van der Waals αλληλεπιδράσεων (λ: 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00) και ύστερα ακολούθησε η μετατροπή των ηλεκτροστατικών αλληλεπιδράσεων (λ: 0.00, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00) της κατάστασης αναφοράς σε αυτές που αντιστοιχούν στην κατάσταση στόχου. Επιπρόσθετα, στο υπολογιστικό πακέτο GROMACS απαιτείται ο καθορισμός των λ καταστάσεων στις οποίες θα μετασχηματιστούν οι ενδομοριακές αλληλεπιδράσεις και οι μάζες των ατόμων, από αυτές που αντιστοιχούν στην αρχική κατάσταση A προς αυτές που αφορούν την κατάσταση B. Η αλχημική μετατροπή των παραμέτρων αυτών επιλέχθηκε να πραγματοποιηθεί ταυτόχρονα με την διαταραχή των παραμέτρων van der Waals. Τέλος, χρησιμοποιήθηκε ένα τροποποιημένο van der Waals δυναμικό, για την αποφυγή των ανεπιθύμητων καταστάσεων που περιεγράφηκαν στο Κεφάλαιο 3.1.3, με τις εξής παραμέτρους: soft-core alpha = 0.3, soft-core power = 1, soft-core sigma = 0.3. [90], [119].

### 3.2.4.3 Υπολογισμός σχετικής ελεύθερης ενέργειας πρόσδεσης μεταξύ ανάλογων ενώσεων

Μετά την παραγωγική φάση των προσομοιώσεων Μοριακής Δυναμικής για τις complex και solvent φάσεις κάθε συστήματος, ακολουθεί ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης (Εξίσωση 3.2). Για την ανάλυση των τροχιών που λήφθηκαν από τις προσομοιώσεις της Μοριακής Δυναμικής και την εξαγωγή της διαφοράς ελεύθερης ενέργειας για κάθε φάση, θα χρησιμοποιηθεί η μέθοδος BAR [91] η οποία είναι ενσωματωμένη στο υπολογιστικό πακέτο GROMACS μέσω του εργαλείου gmx bar.

Ο τρόπος με τον οποίο είναι εφαρμοσμένη η μεθοδολογία BAR στο πακέτο GROMACS, δεν απαιτεί την ξεχωριστή δειγματοληψία των forward και backward προσομοιώσεων κάθε φάσης της κάθε ανάλογης ένωσης. Στην περίπτωση αυτή, κάθε ενδιάμεση λ κατάσταση προσομοιώνεται διακριτά και σε κάθε μία υπολογίζεται η ενεργειακή διαφορά της σε σχέση με τις «γειτονικές» ενδιάμεσες καταστάσεις. Με άλλα λόγια εάν λαμβάνει χώρα η προσομοίωση για την 7<sup>η</sup> ενδιάμεση κατάσταση, τότε θα υπολογισθούν οι ενεργειακές διαφορές αυτής, τόσο σε σχέση με την 6<sup>η</sup> όσο και σε σχέση με την 8<sup>η</sup> ενδιάμεση λ κατάσταση. Έτσι, η μέθοδος BAR υπολογίζει ένα λόγο σταθμισμένου μέσου όρου της Χαμιλτονιανής διαφοράς της κατάστασης B δεδομένης της κατάστασης A και αντίστροφα. Επίσης, όπως και στην περίπτωση του εργαλείου ParseFEP, συνδυάζονται κατάλληλα τα αρχεία εξόδου των αλχημικών προσομοιώσεων (εδώ XVG files), υπολογίζονται οι ελεύθερες ενέργειες ΔG κάθε φάσης και εκτιμάται η αντίστοιχη στατιστική αβεβαιότητα σ<sub>BAR</sub> [119].

Τέλος, για τον έλεγχο και την παρακολούθηση της σύγκλισης της ελεύθερης ενέργειας, παρέχονται δύο ειδών αρχεία εξόδου από το πακέτο gmx bar. Το πρώτο κομμάτι παρέχει λεπτομερή εκτίμηση των διαφορών ελεύθερης ενέργειας για κάθε ενδιάμεση κατάσταση, μετρήσεις της σχετικής εντροπίας, ενώ και τα δύο μεγέθη που υπολογίζονται συνοδεύονται από μία εκτίμηση σφάλματος. Η σχετική εντροπία μίας κατάστασης A σε σχέση με μία κατάσταση B, ερμηνεύεται ως το μέτρο της επικάλυψης του χώρου φάσης: η σχετική εντροπία  $s_A$  των δειγμάτων της κατάστασης  $\lambda_B$  στο στατιστικό σύνολο της κατάστασης  $\lambda_A$  (και αντίστροφα για το  $s_B$ ) είναι ένα μέτρο της διαφοράς μεταξύ των κατανομών Boltzmann των δύο καταστάσεων, και θα πρέπει να τείνει στο μηδέν για ταυτόσημες κατανομές. Επιπλέον, για να αποκτήσει κανείς μια οπτική εκτίμηση της επικάλυψης του χώρου φάσεων, το πακέτο gmx bar παρέχει τις γραφικές παραστάσεις των κατανομών της βαθμίδας  $\Delta H/\Delta L$  και των Χαμιλτονιανών διαφορών ( $\Delta H$ ) μεταξύ των «γειτονικών»  $\lambda$  ενδιάμεσων καταστάσεων, σε συνάρτηση με τον αριθμό δειγμάτων που λαμβάνεται από κάθε κατάσταση [127]. Όσον αφορά το δεύτερο κομμάτι των αρχείων εξόδου του πακέτου gmx bar, παρέχονται οι τελικές τιμές της εκτιμώμενης διαφοράς ελεύθερης ενέργειας ( $\Delta G$ ) μεταξύ της κατάστασης αναφοράς και στόχου συνοδευόμενες με το αντίστοιχο σφάλμα.

Στην περίπτωση του υπολογιστικού πακέτου GROMACS, για να θεωρηθεί αξιόπιστη η εκτίμηση  $\Delta G$ , απαιτείται επαρκής δειγματοληψία σε κάθε ενδιάμεση κατάσταση (με βάση τις κατανομές  $\Delta H/\Delta L$ ) και μεταξύ γειτονικών ενδιάμεσων καταστάσεων (κατανομή  $\Delta H$ ). Εάν οι υπολογισμοί οδηγούν σε μη αλληλεπικαλυπτόμενα ιστογράμματα, τιμές σχετικής εντροπίας μακριά από το μηδέν και μεγάλες εκτιμήσεις σφάλματος, είναι πιθανό να απαιτείται μεγαλύτερος χρόνος προσομοίωσης, περισσότερες  $\lambda$  καταστάσεις ή διαφορετικές παράμετροι για την αληθινή μετατροπή του εξεταζόμενου μορίου.

### 3.3 Δοκιμασίες κλιμάκωσης στο υπερυπολογιστικό σύστημα ARIS

Οι προσομοιώσεις Μοριακής Δυναμικής με σκοπό τον υπολογισμό της σχετικής και απόλυτης ελεύθερης ενέργειας πρόσδεσης τόσο για το πακέτο NAMD, όσο και για το πακέτο GROMACS, πραγματοποιήθηκαν με τη βοήθεια του ελληνικού υπερυπολογιστικού συστήματος ARIS. Προτού ξεκινήσουν οι τελικές προσομοιώσεις κάθε συστήματος, είναι απαραίτητο να πραγματοποιηθούν οι κατάλληλες δοκιμασίες κλιμάκωσης (scaling tests). Μέσω αυτών, μπορεί κανείς να υπολογίσει τον συνδυασμό κόμβων (nodes) και επεξεργαστών ανά διεργασία (CPUs per task) που είναι πιο αποδοτικός για κάθε σύστημα που πρόκειται να προσομοιωθεί.

#### 3.3.1 NAMD

##### 3.3.1.1 Δοκιμασία κλιμάκωσης του συστήματος CB8-G3-0

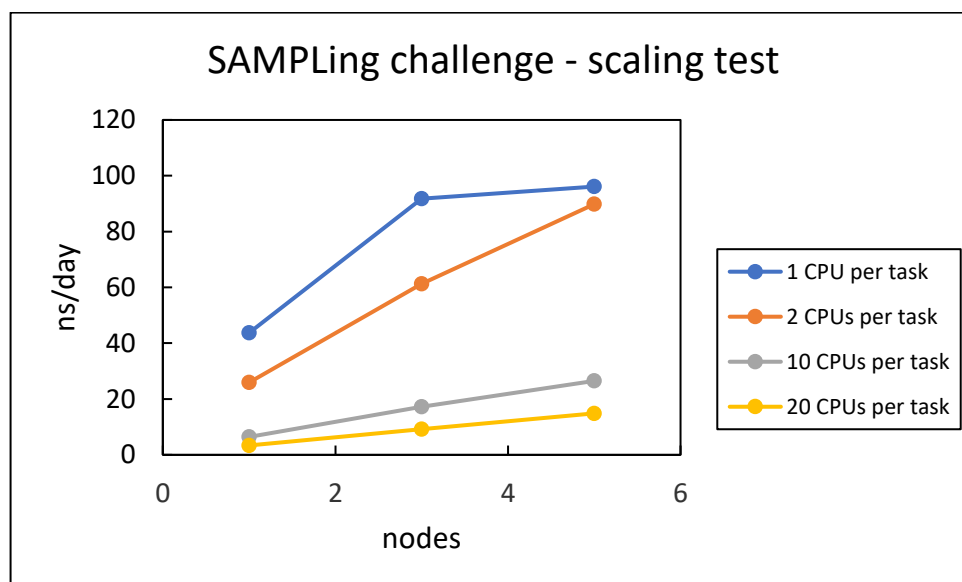
Η πρώτη δοκιμασία κλιμάκωσης έλαβε χώρα για τους υπολογισμούς απόλυτης ελεύθερης ενέργειας με το υπολογιστικό πακέτο NAMD για το σύστημα προσδέτη-δέκτη CB8-G3-0, δηλαδή για το πρώτο αντίγραφο του συστήματος CB8-G3. Όλα τα συστήματα προσδέτη-δέκτη κρίνεται πως ακολουθούν την ίδια συμπεριφορά ως προς την αποδοτικότητα, καθώς αποτελούνται από ίδιας τάξης αριθμό

ατόμων και για το λόγο αυτό δεν πραγματοποιήθηκαν επιπλέον δοκιμασίες κλιμάκωσης για τα συστήματα OA-G3/G6.

Η δοκιμασία κλιμάκωσης για το σύστημα CB8-G3-0, πραγματοποιήθηκε για 0.2 ns για την complex φάση του συστήματος αυτού (~7,000 άτομα) για διαφορετικούς συνδυασμούς nodes και CPUs ανά διεργασία. Η δοκιμασία κλιμάκωσης για τη solvent φάση δεν κρίνεται απαραίτητη, καθώς αποτελείται από ίδιας τάξης αριθμό ατόμων με την complex φάση (~6,000 άτομα) και επομένως εκτιμάται πως τα δύο συστήματα θα παρουσιάζουν αντίστοιχη συμπεριφορά ως προς την αποδοτικότητα. Από τα δεδομένα του Πίνακα 3.2 και από το Σχήμα 3.11, γίνεται αντιληπτό πως είναι πιο αποδοτικό τα συστήματα της πρόκλησης SAMPL6 SAMPLing να προσομοιωθούν σε ένα κόμβο με χρήση ενός CPU ανά διεργασία, καθώς η αύξηση του αριθμού των επεξεργαστών οδηγεί σε μείωση της αποδοτικότητας, ενώ αυξάνοντας τους κόμβους δεν παρατηρείται η αντίστοιχη αύξηση σε αποδοτικότητα. Ένας συνδυασμός κρίνεται αποδοτικός όταν η χρήση πολλαπλάσιων κόμβων, έχει την αντίστοιχη επίδραση στο χρόνο που απαιτείται για την προσομοίωση. Για παράδειγμα για να χρησιμοποιηθούν 3 κόμβοι αντί για 1, θα έπρεπε η αποδοτικότητα να ήταν της τάξης των 120 ns/day.

**Πίνακας 3.2** Οι συνδυασμοί κόμβων και CPUs ανά διεργασία καθώς και η αντίστοιχη αποδοτικότητα σε ns/day για την complex φάση του συστήματος CB8-G3-0.

| Κόμβοι | CPUs ανά διεργασία |      |      |      |
|--------|--------------------|------|------|------|
|        | 1                  | 2    | 10   | 20   |
| 1      | 43.7               | 25.9 | 6.4  | 3.4  |
| 3      | 91.7               | 61.2 | 17.2 | 9.2  |
| 5      | 96.1               | 89.8 | 26.5 | 14.9 |



**Σχήμα 3.11** Δοκιμασία κλιμάκωσης με το πακέτο NAMD για την για την complex φάση του συστήματος CB8-G3-0.

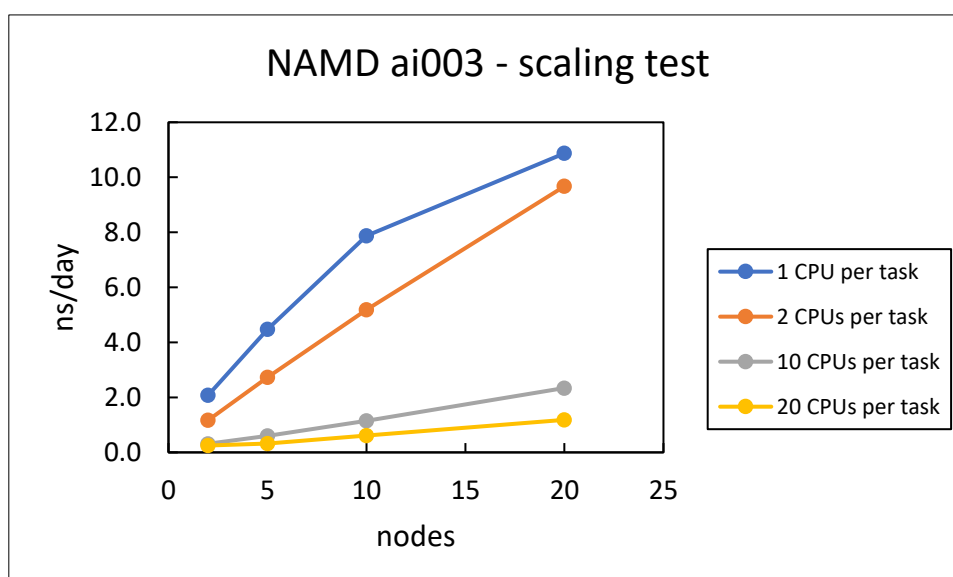
### 3.3.1.2 Δοκιμασία κλιμάκωσης του συστήματος Arg2/3 – ανάλογο μόριο ai003

Η δοκιμασία κλιμάκωσης για τους υπολογισμούς σχετικής ελεύθερης ενέργειας με το υπολογιστικό πακέτο NAMD πραγματοποιήθηκε για την ανάλογη ένωση ai003. Για το σκοπό αυτό, προσομοιώθηκε για 0.1 ns η complex φάση του συστήματος αυτού, η οποία αποτελείται από ~115,000 άτομα, για διαφορετικούς συνδυασμούς κόμβων και CPUs ανά διεργασία (Πίνακας 3.3).

Στον Πίνακα 3.3 συνοψίζονται οι συνδυασμοί κόμβων και CPUs ανά διεργασία καθώς και η αντίστοιχη αποδοτικότητα σε ns/day, για την complex φάση της ανάλογης ένωσης ai003. Από το Σχήμα 3.12, γίνεται αντιληπτό πως οι complex φάσεις των ανάλογων ενώσεων του CK-666 θα πρέπει να προσομοιωθούν σε 5 nodes και 1 CPU ανά διεργασία, αφού η χρήση περισσότερων επεξεργαστών οδηγεί σε μείωση της αποδοτικότητας, ενώ μετά τη χρήση 5 κόμβων δεν παρατηρείται αντίστοιχη αύξηση στην αποδοτικότητα του συστήματος. Όσον αφορά την solvent φάση της προσομοίωσης, δεν κρίθηκε απαραίτητη η πραγματοποίηση επιπλέον δοκιμασίας κλιμάκωσης επειδή το σύστημα αυτό αποτελείται από < 2,000 άτομα. Έτσι, τα solvent συστήματα κάθε ανάλογης ένωσης προσομοιωθήκαν σε 1 κόμβο με 1 CPU ανά διεργασία.

**Πίνακας 3.3** Οι συνδυασμοί κόμβων και CPUs ανά διεργασία καθώς και η αντίστοιχη αποδοτικότητα σε ns/day για την complex φάση της ανάλογης ένωσης ai003.

| CPUs ανά διεργασία |      |     |     |     |
|--------------------|------|-----|-----|-----|
| Κόμβοι             | 1    | 2   | 10  | 20  |
| 2                  | 2.1  | 1.2 | 0.3 | 0.2 |
| 5                  | 4.5  | 2.7 | 0.6 | 0.3 |
| 10                 | 7.9  | 5.2 | 1.1 | 0.6 |
| 20                 | 10.9 | 9.7 | 2.3 | 1.2 |



**Σχήμα 3.12** Δοκιμασία κλιμάκωσης με το πακέτο NAMD για την για την complex φάση της ανάλογης ένωσης ai003.

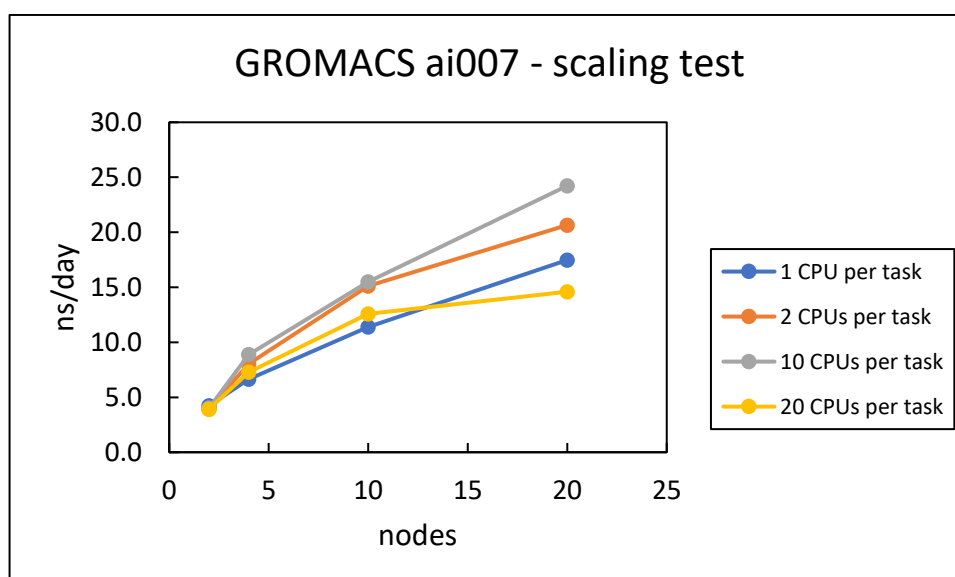
### 3.3.2 GROMACS

#### 3.3.2.1. Δοκιμασία κλιμάκωσης του συστήματος Arg2/3 – ανάλογο μόριο ai007

Τέλος, πραγματοποιήθηκε η δοκιμασία κλιμάκωσης για τους υπολογισμούς σχετικής ελεύθερης ενέργειας πρόσδεσης για το υπολογιστικό πακέτο GROMACS για την complex φάση της ανάλογης ένωσης ai007. Η προσομοίωση κλιμάκωσης έγινε για 0.1 ns στην complex φάση του συστήματος αυτού (~115,000 άτομα) για διαφορετικούς συνδυασμούς nodes και CPUs ανά διεργασία (Πίνακας 3.4). Όπως και στην περίπτωση του NAMD, για τη solvent φάση της προσομοίωσης δεν κρίθηκε απαραίτητη η πραγματοποίηση επιπλέον δοκιμασίας κλιμάκωσης επειδή το σύστημα αυτό αποτελείται από <2,000 άτομα. Έτσι και για το υπολογιστικό πακέτο GROMACS, τα solvent συστήματα κάθε ανάλογης ένωσης προσομοιωθήκαν σε 1 κόμβο με 1 CPU ανά διεργασία. Από το Σχήμα 3.13, γίνεται αντιληπτό πως οι complex φάσεις των ανάλογων ενώσεων του CK-666 θα πρέπει να προσομοιωθούν σε 4 nodes και 10 CPU ανά διεργασία.

**Πίνακας 3.4** Οι συνδυασμοί κόμβων και CPUs ανά διεργασία καθώς και η αντίστοιχη αποδοτικότητα σε ns/day για την complex φάση της ανάλογης ένωσης ai007.

| Κόμβοι | CPUs ανά διεργασία |      |      |      |
|--------|--------------------|------|------|------|
|        | 1                  | 2    | 10   | 20   |
| 2      | 4.2                | 3.9  | 4.0  | 3.9  |
| 4      | 6.7                | 8.0  | 8.9  | 7.3  |
| 10     | 11.4               | 15.1 | 15.5 | 12.6 |
| 20     | 17.5               | 20.7 | 24.2 | 14.6 |



**Σχήμα 3.13** Δοκιμασία κλιμάκωσης με το πακέτο GROMACS για την για την complex φάση της ανάλογης ένωσης ai007.



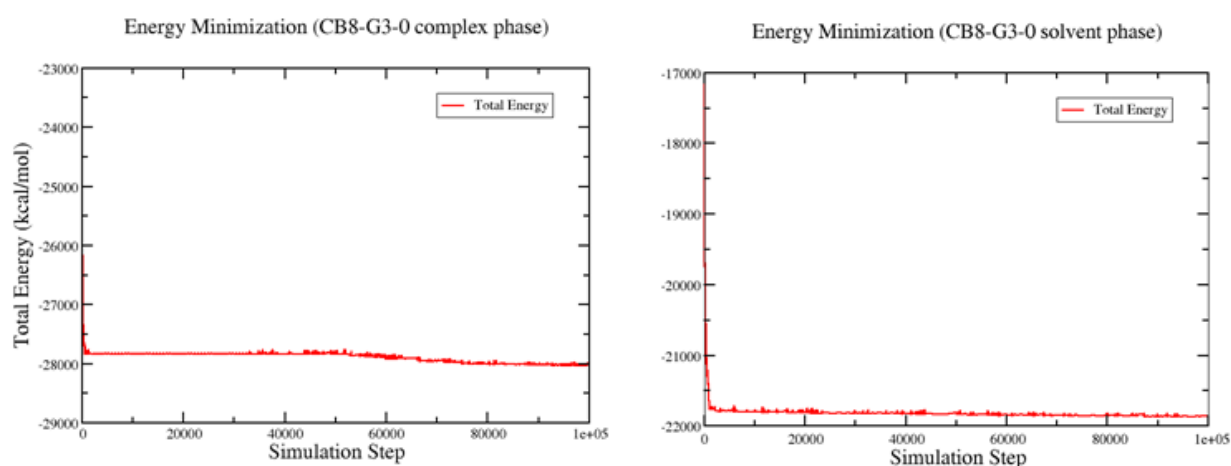
## 4. Αποτελέσματα – Συζήτηση Αποτελεσμάτων

### 4.1 Αποτελέσματα μοριακών προσομοιώσεων για την πρόκληση SAMPL6 SAMPLing

#### 4.1.1 Πρόβλεψη απόλυτης ελεύθερης ενέργειας πρόσδεσης του συστήματος CB8-G3

Οι διοργανωτές της πρόκλησης SAMPL6 SAMPLing παρέιχαν τα απαραίτητα αρχεία εισόδου για την έναρξη των προσομοιώσεων για πέντε αντίγραφα κάθε συστήματος προσδέτη-δέκτη για τις complex και solvent φάσεις. Τα αντίγραφα του ίδιου συστήματος, στην προκειμένη περίπτωση του συστήματος CB8-G3, διαφέρουν μεταξύ τους ως προς την αρχική διαμόρφωση που λαμβάνει ο προσδέτης (εδώ η κινίνη, G3) στο μακρομοριακό δέκτη (εδώ το cucurbit[8]uril, CB8). Επομένως, για επαρκώς μεγάλη προσομοίωση, θα πρέπει η εκτιμώμενη απόλυτη ελεύθερη ενέργεια πρόσδεσης του G3 στο CB8 να είναι ίδια και για τα πέντε αντίγραφα του συστήματος.

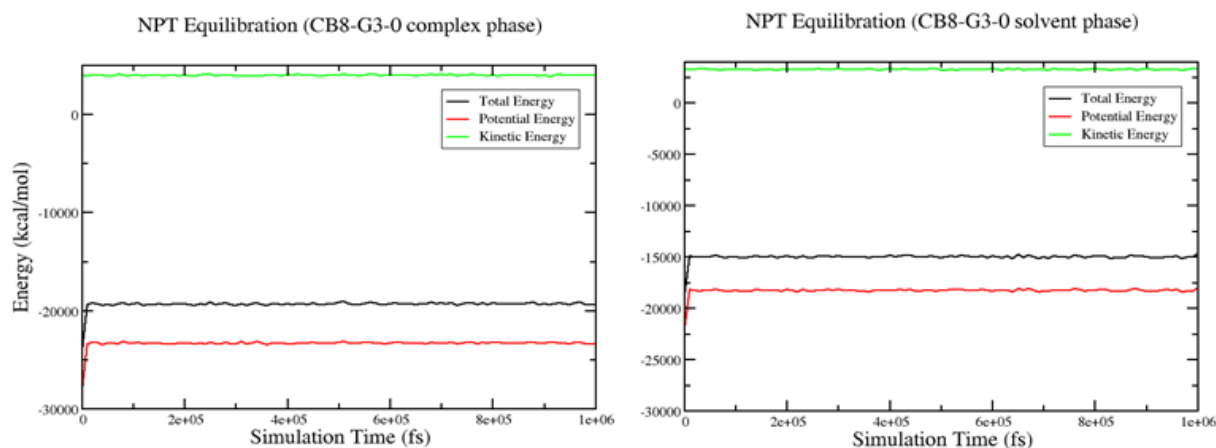
Αρχικά, σε κάθε δομή πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας για 100,000 βήματα προσομοίωσης. Το γεγονός ότι η ελαχιστοποίηση της ενέργειας του συστήματος επιτεύχθηκε επιβεβαιώνεται από το Σχήμα 4.1, όπου απεικονίζεται η συνολική ενέργεια του συστήματος ως συνάρτηση των βημάτων προσομοίωσης. Το Σχήμα 4.1 κατασκευάστηκε από τα αποτελέσματα της ελαχιστοποίησης ενέργειας για την complex και solvent φάση του πρώτου αντίγραφου του συστήματος CB8-G3, δηλ. αφορά το σύστημα CB8-G3-0. Το διάγραμμα αυτό αποτελεί ένα ενδεικτικό παράδειγμα καθώς όλα τα υπόλοιπα συστήματα CB8-G3 και OA-G3/G6, ακολουθούν το ίδιο μοτίβο σύγκλισης της ενέργειας κατά την ελαχιστοποίηση, και για τις δύο φάσεις.



**Σχήμα 4.1** Διάγραμμα της συνολικής ενέργειας του συστήματος συναρτήσει των βημάτων προσομοίωσης κατά την ελαχιστοποίηση της ενέργειας για την complex (επάνω) και solvent (κάτω) φάση της προσομοίωσης του συστήματος CB8-G3-0.

Στη συνέχεια ακολούθησε εξισορρόπηση υπό σύνολο NPT ( $T=298.15$  K,  $p=1$  atm) για χρονική διάρκεια 1 ns και βήμα 1 fs. Αφού μελετήθηκαν οι ιδιότητες των συστημάτων, όπως η θερμοκρασία και η ενέργεια ως προς το χρόνο και παρατηρήθηκε ότι δεν υπάρχουν μεγάλες διακυμάνσεις στις τιμές

τους, ακολούθησαν οι παραγωγικές προσομοιώσεις για όλα τα συστήματα. Στο Σχήμα 4.2 παρουσιάζεται ενδεικτικά η γραφική παράσταση της συνολικής, κινητικής και δυναμικής ενέργειας για την NPT εξισορρόπηση, συναρτήσει του χρόνου προσομοίωσης για τις δύο φάσεις του συστήματος CB8-G3-0. Το διάγραμμα αυτό αποτελεί ένα ενδεικτικό παράδειγμα καθώς τα υπόλοιπα συστήματα CB8-G3 και OA-G3/G6, ακολουθούν το ίδιο μοτίβο σύγκλισης της ενέργειας κατά την ελαχιστοποίηση, και για τα δύο σκέλη προσομοίωσης.

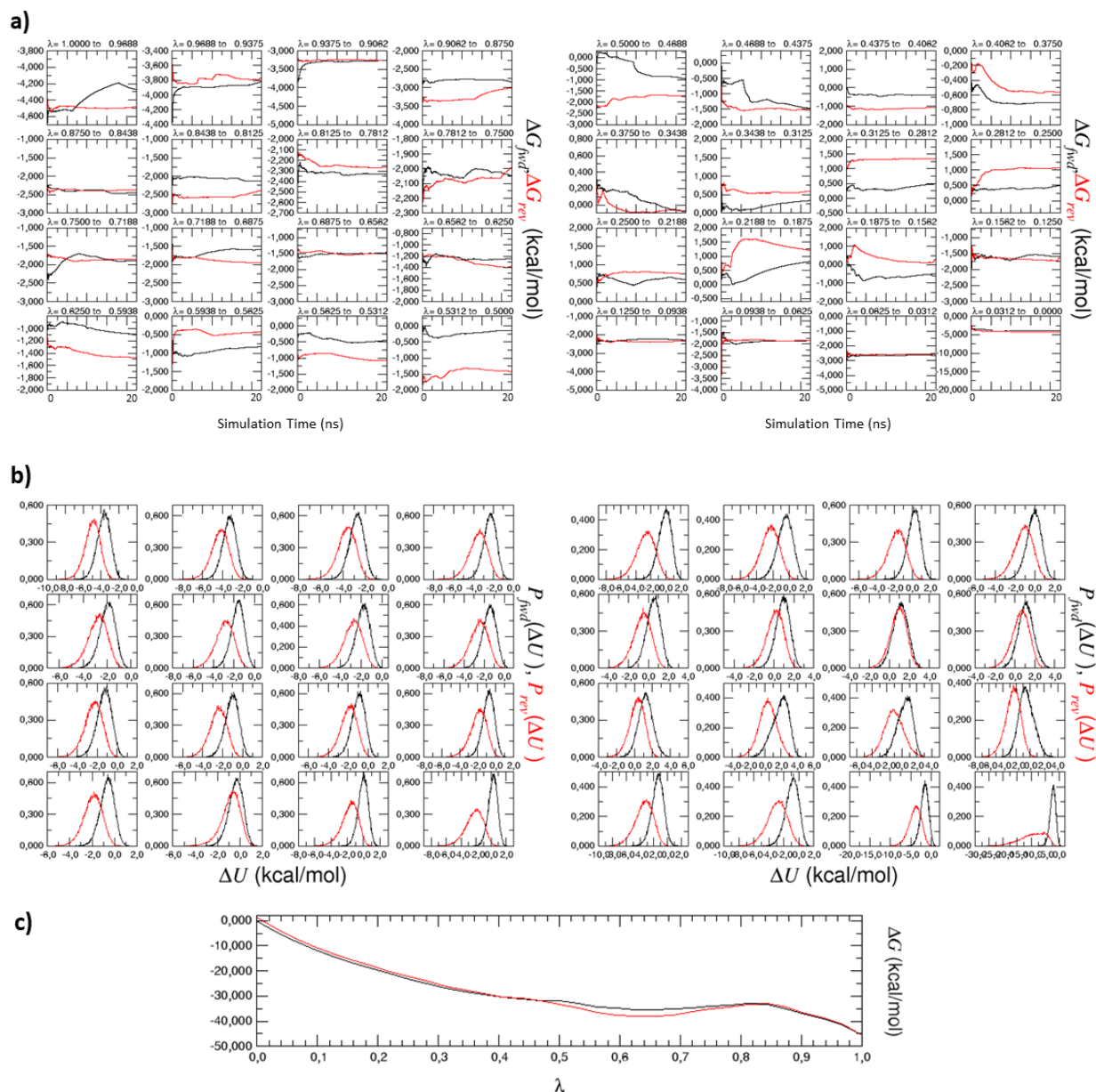


**Σχήμα 4.2** Διαγράμματα της κινητικής, δυναμικής και συνολικής ενέργειας του συστήματος συναρτήσει του χρόνου προσομοίωσης για την complex (επάνω) και solvent (κάτω) φάση της προσομοίωσης του συστήματος CB8-G3-0.

Στη συνέχεια, ακολούθησαν οι παραγωγικές φάσεις των προσομοιώσεων με το υπολογιστικό πακέτο NAMD για τα πέντε αντίγραφα του συστήματος CB8-G3. Ο υπολογισμός της απόλυτης ελεύθερης ενέργειας πρόσδεσης γίνεται με βάση την Εξίσωση 3.1, αγνοώντας την επίδραση του όρου  $\Delta G_{rest}$ . Για την αληθινή «εξαφάνιση» του προσδέτη από το μακρομοριακό δέκτη, χρησιμοποιήθηκαν 32 ενδιάμεσες καταστάσεις, με τη μεταβλητή  $\lambda$  να αλλάζει από το 0 προς το 1 με σταθερό βήμα 0.03125. Η προσομοίωση των καταστάσεων αυτών έγινε για 20 ns/κατάσταση με χρονικό βήμα 2fs, ενώ για την κάθε  $\lambda$  κατάσταση πραγματοποιήθηκε forward και backward προσομοίωση.

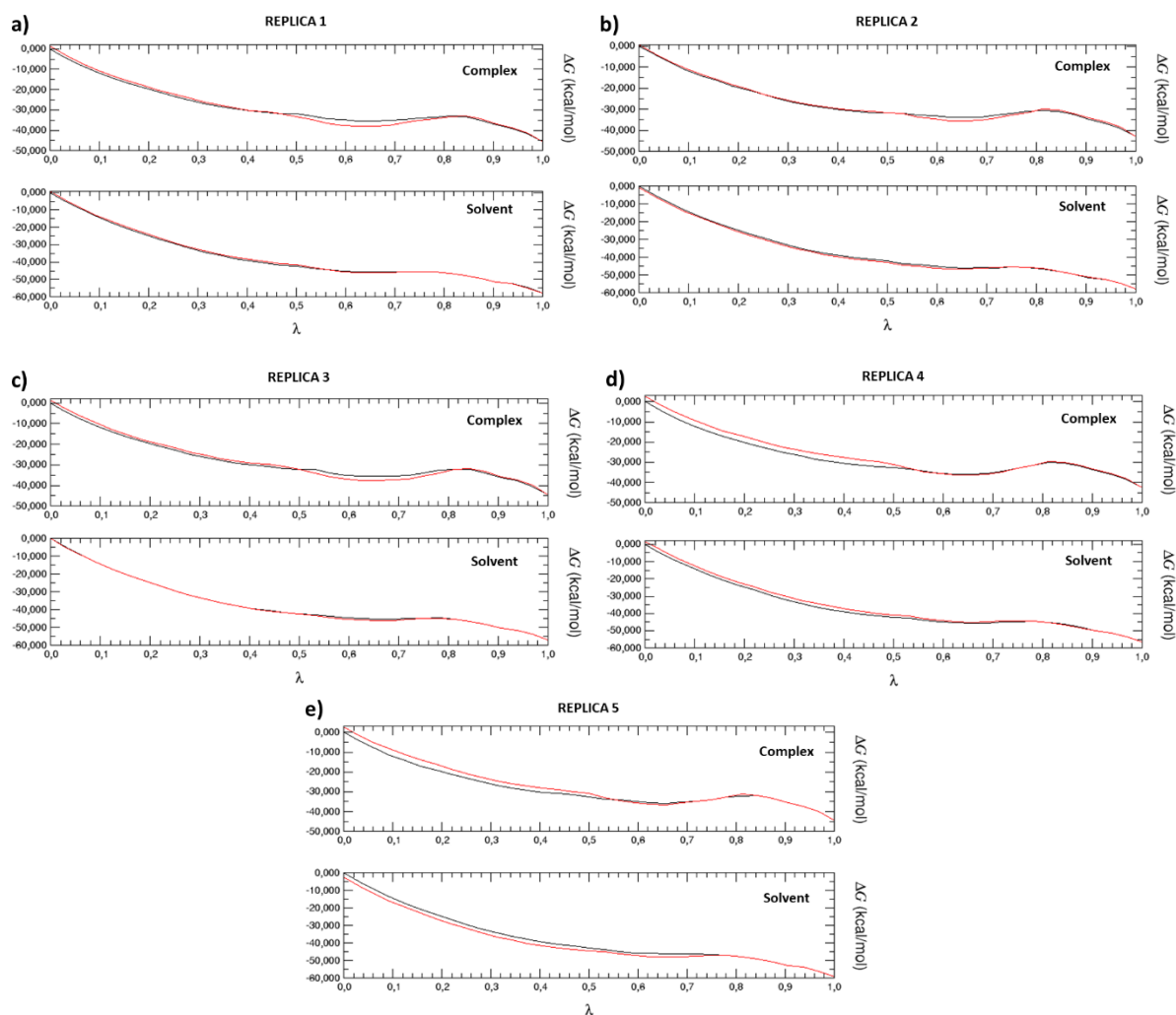
Για την ανάλυση των τροχιών που λήφθηκαν από τις προσομοιώσεις Μοριακής Δυναμικής και την εξαγωγή της διαφοράς ελεύθερης ενέργειας σε κάθε φάση, χρησιμοποιείται το εργαλείο ParseFEP του λογισμικού VMD. Τα αρχεία εξόδου των forward και backward προσομοιώσεων συνδυάζονται και με τη βοήθεια του εργαλείου ParseFEP, υπολογίζεται η ελεύθερη ενέργεια κάθε φάσης με τη μέθοδο BAR. Επιπρόσθετα, παρέχονται οι γραφικές παραστάσεις της κατανομής πιθανότητας της δυναμικής ενέργειας και της εξέλιξης της ελεύθερης ενέργειας κατά τη διάρκεια της προσομοίωσης που χαρακτηρίζει τόσο τον forward, όσο και τον backward μετασχηματισμό, κάθε ενδιάμεσης κατάστασης  $\lambda$ . Εκτός αυτών, παρέχεται ένα επιπλέον συγκεντρωτικό γράφημα στο οποίο αναπαρίσταται η συνολική εξέλιξη της ελεύθερης ενέργειας Gibbs συναρτήσει των  $\lambda$  καταστάσεων. Στο Σχήμα 4.3 απεικονίζονται όλα τα γραφήματα εξόδου που παρέχονται από το εργαλείο ParseFEP, για την complex φάση του συστήματος CB8-G3-0.





**Σχήμα 4.3** Τυπικά διαγράμματα παρακολούθησης της εξέλιξης της ελεύθερης ενέργειας από το εργαλείο ParseFEP για την complex φάση του συστήματος CB8-G3-0. **(a)** Χρονική εξέλιξη της ελεύθερης ενέργειας, **(b)** ιστογράμματα των πιθανοτήτων κατανομής της δυναμικής ενέργειας, και **(c)** συνολική μεταβολή της ελεύθερης ενέργειας ως συνάρτηση των ενδιάμεσων  $\lambda$  καταστάσεων. Οι μαύρες και κόκκινες συμπαγείς γραμμές υποδηλώνουν τη συνολική αλλαγή ελεύθερης ενέργειας για τον προς τα εμπρός (forward, fwd) και τον αντίστροφο (reverse/backward, rev) μετασχηματισμό, αντίστοιχα.

Στη συνέχεια, κρίνεται σκόπιμη μόνο η παρουσίαση των γραφημάτων που απεικονίζουν τη συνολική μεταβολή της ελεύθερης ενέργειας των ενδιάμεσων  $\lambda$  καταστάσεων, καθώς ουσιαστικά ενσωματώνουν τις πληροφορίες που περιέχονται στα παραπάνω διαγράμματα. Το Σχήμα 4.44 απεικονίζει την εξέλιξη της συνολικής μεταβολής της ελεύθερης ενέργειας για κάθε φάση και αντίγραφο του συστήματος CB8-G3 που προσομοιώθηκε.



**Σχήμα 4.4** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για όλα τα αντίγραφα του συστήματος CB8-G3 με χρήση 32 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 20 ns/λ κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. **(a)** Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας για το σύστημα CB8-G3-0 βρέθηκε  $-58.13 \pm 0.18$  kcal/mol και  $-46.40 \pm 0.25$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(b)** Η προβλεπόμενη μεταβολή ελεύθερης ενέργειας του συστήματος CB8-G3-1 υπολογίζεται ίση με  $-57.49 \pm 0.18$  kcal/mol και  $-43.60 \pm 0.25$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(c)** Στο σύστημα CB8-G3-2 η μεταβολή της ελεύθερης ενέργειας εκτιμάται στα  $-56.86 \pm 0.24$  kcal/mol και  $-45.42 \pm 0.12$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(d)** Η προβλεπόμενη μεταβολή ελεύθερης ενέργειας του συστήματος CB8-G3-3 υπολογίζεται ίση με  $-57.54 \pm 0.14$  kcal/mol και  $-43.84 \pm 0.16$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(e)** Η μεταβολή της ελεύθερης ενέργειας του συστήματος CB8-G3-4 εκτιμάται στα  $-57.98 \pm 0.11$  kcal/mol και  $-45.54 \pm 0.19$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα.

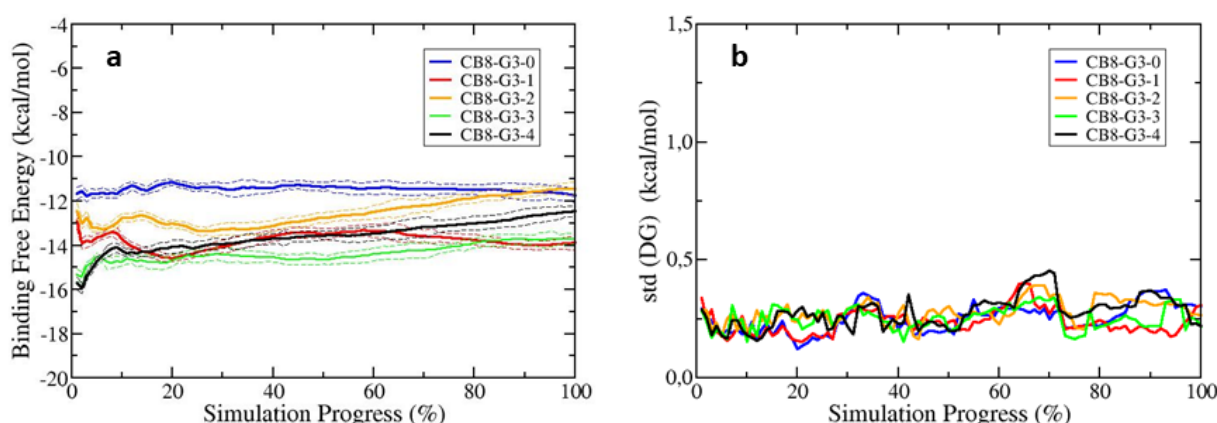
Στον Πίνακα 4.1 συνοψίζονται οι υπολογισμένες ελεύθερες ενέργειες για την προσομοίωση με το υπολογιστικό πακέτο NAMD για την κάθε φάση και αντίγραφο του συστήματος CB8-G3. Επιπλέον παρατίθενται τα αποτελέσματα για την απόλυτη ελεύθερη ενέργεια πρόσδεσης  $\Delta G_{\text{Bind}}$ . Το σφάλμα για την απόλυτη ελεύθερη ενέργεια πρόσδεσης εκτιμήθηκε από την Εξίσωση 4.1, με βάση το σφάλμα της μεθόδου BAR για το κάθε σκέλος της προσομοίωσης.

$$error = \sqrt{(\sigma_{BAR}^{complex})^2 + (\sigma_{BAR}^{solvent})^2} \quad (4.1)$$

Επιπρόσθετα, μετρήθηκε το υπολογιστικό κόστος των προσομοιώσεων, ως ο συνολικός απαιτούμενος χρόνος για την προσομοίωση κάθε φάσης (solvent και complex) και ενδιάμεσης λ κατάστασης. Τέλος, για την παρακολούθηση της συμπεριφοράς των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας κατά τη διάρκεια της προσομοίωσης, εκτιμήθηκε η απόλυτη ελεύθερη ενέργεια πρόσδεσης καθώς και το αντίστοιχο σφάλμα για 1%, 2%, ..., 100% της προσομοίωσης κάθε αντιγράφου του συστήματος CB8-G3 (Σχήμα 4.5).

**Πίνακας 4.1** Σύνοψη των υπολογισμών απόλυτης ελεύθερης ενέργειας για την κάθε αρχική διαμόρφωση του συστήματος CB8-G3.

| Αρχική διαμόρφωση | $\Delta G_{\text{solvent}}$<br>(kcal/mol) | $\Delta G_{\text{complex}}$<br>(kcal/mol) | $\Delta G_{\text{Bind}}$<br>(kcal/mol) | Συνολικός χρόνος<br>προσομοίωσης (h) |
|-------------------|---|---|--|--------------------------------------|
| <b>CB8-G3-0</b>   | $-46.40 \pm 0.25$                         | $-58.13 \pm 0.18$                         | $-11.74 \pm 0.31$                      | 695                                  |
| <b>CB8-G3-1</b>   | $-43.60 \pm 0.25$                         | $-57.49 \pm 0.18$                         | $-13.88 \pm 0.31$                      | 691                                  |
| <b>CB8-G3-2</b>   | $-45.42 \pm 0.12$                         | $-56.86 \pm 0.24$                         | $-11.44 \pm 0.27$                      | 681                                  |
| <b>CB8-G3-3</b>   | $-43.84 \pm 0.16$                         | $-57.54 \pm 0.14$                         | $-13.71 \pm 0.22$                      | 685                                  |
| <b>CB8-G3-4</b>   | $-45.54 \pm 0.19$                         | $-57.98 \pm 0.11$                         | $-12.45 \pm 0.22$                      | 707                                  |



**Σχήμα 4.5** Εξέλιξη της ελεύθερης ενέργειας πρόσδεσης και τυπικής απόκλισης κάθε αρχικής διαμόρφωσης CB8-G3 καθ' όλη τη διάρκεια της προσομοίωσης. **(a)** Αντιπαράθεση των απόλυτων ελεύθερων ενεργειών πρόσδεσης για τα 5 αντίγραφα του συστήματος CB8-G3. Οι έντονες συνεχείς γραμμές υποδηλώνουν την ελεύθερη ενέργεια πρόσδεσης, ενώ οι διακεκομμένες υποδεικνύουν την αβεβαιότητα της εκτίμησης. **(b)** Τυπική απόκλιση των εκτιμώμενων ελεύθερων ενεργειών για όλες τις αρχικές διαμορφώσεις του συστήματος CB8-G3.

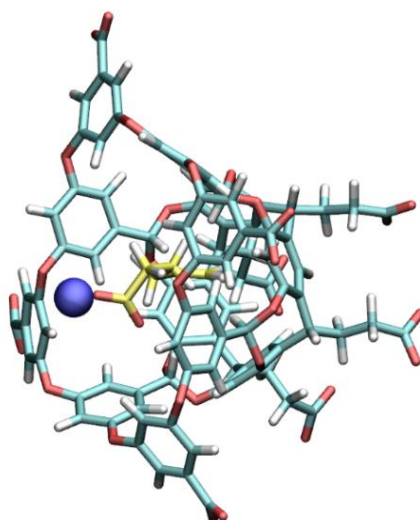
Από το Σχήμα 4.5 μπορεί κανείς να αντιληφθεί πως οι προσομοιώσεις με το NAMD για τα αντίγραφα του CB8-G3 συστήματος παρείχαν αξιόπιστες εκτιμήσεις για τη μεταβολή της ελεύθερης ενέργειας πρόσδεσης. Παρ' όλα αυτά, από τα δεδομένα του Πίνακα 4.1 και το Σχήμα 4.5 παρατηρεί κανείς πως η απόλυτη ελεύθερη ενέργεια πρόσδεσης δεν τείνει σε μία σταθερή τιμή και για τα πέντε αντίγραφα του συστήματος CB8-G3. Η μέση τιμή των προβλέψεων του NAMD για το σύστημα CB8-G3 ανέρχεται στα  $-12.6 \pm 0.6$  kcal/mol και είναι συνεπής σε εύρος, 1-2 kcal/mol με τους υπολογισμούς αναφοράς που πραγματοποιήθηκαν από τους διοργανωτές. Πιο συγκεκριμένα, οι διοργανωτές χρησιμοποίησαν το υπολογιστικό πακέτο YANK και εκτίμησαν ότι η μέση τιμή της απόλυτης ελεύθερης πρόσδεσης του συστήματος CB8-G3 ισούται με  $-10.9 \pm 0.1$  kcal/mol. Ωστόσο και τα δύο υπολογιστικά πακέτα υπερεκτίμησαν την απόλυτη ελεύθερη ενέργεια πρόσδεσης της κινίνης στον δέκτη cucurbit[8]uril

καθώς η πειραματικά υπολογισμένη ελεύθερη ενέργεια πρόσδεσης ανέρχεται στα  $-6.5 \pm 0.1$  kcal/mol. Αξίζει να αναφερθεί πως καμία από τις υπόλοιπες ομάδες που έλαβαν μέρος στο διαγωνισμό, δεν κατάφεραν να πλησιάσουν σε εύρος 1-2 kcal/mol την πειραματική τιμή, γεγονός που υποδεικνύει την ύπαρξη συστηματικού σφάλματος το οποίο πιθανόν οφείλεται στην παραμετροποίηση του μορίου προσδέτη. Ενδεικτικά αναφέρεται πως οι υπολογισμοί με το πακέτο AMBER οδήγησαν σε πιο κοντινή εκτίμηση των πειραματικών δεδομένων, με την ελεύθερη ενέργεια πρόσδεσης να ανέρχεται περίπου στα -10.5 kcal/mol.

#### 4.1.2 Μελέτη σύγκλισης απόλυτης ελεύθερης ενέργειας πρόσδεσης των συστημάτων OA-G3/G6

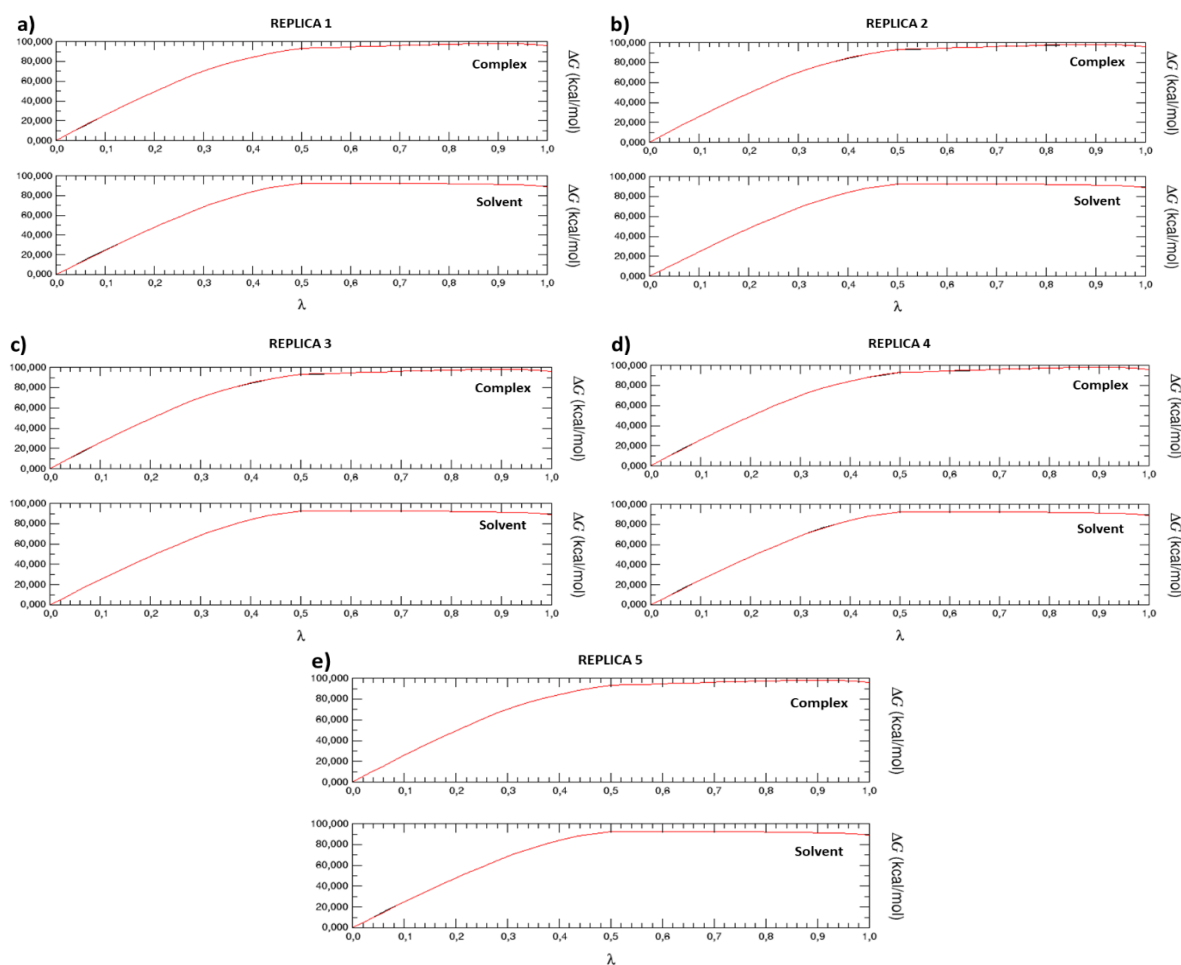
Στη συνέχεια έλαβαν χώρα οι προσομοιώσεις Μοριακής Δυναμικής με σκοπό τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης για τα συστήματα OA-G3 και OA-G6. Όπως και στην περίπτωση του συστήματος CB8-G3, οι διοργανωτές της πρόκλησης SAMPL6 SAMPLing παρέιχαν τα απαραίτητα αρχεία για την έναρξη των προσομοιώσεων για πέντε αντίγραφα του κάθε συστήματος προσδέτη-δέκτη τόσο για την complex, όσο και για τη solvent φάση.

Αρχικά, στα συστήματα OA-G3/G6 πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας και ύστερα ακολούθησε φάση εξισορρόπησης NPT ( $T=298.15$  K,  $p=1$  atm) για χρονική διάρκεια 1 ns και βήμα 1 fs. Τα ενεργειακά διαγράμματα τόσο για την ελαχιστοποίηση και την εξισορρόπηση για όλα τα αντίγραφα των συστημάτων OA-G3/G6, ακολουθούν το ίδιο μοτίβο με αυτά που παρουσιάστηκαν στα Σχήματα 4.41 και 4.42 και έτσι δεν κρίνεται σκόπιμη η παρουσίασή τους. Όσον αφορά τις παραγωγικές φάσεις των προσομοιώσεων των συστημάτων OA-G3 και OA-G6, εφαρμόστηκε το ίδιο πρωτόκολλο που χρησιμοποιήθηκε για τον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης του CB8-G3, δηλαδή χρησιμοποιήθηκαν 32 ενδιάμεσες λ-καταστάσεις με χρόνο προσομοίωσης 20 ns / λ-κατάσταση. Αξίζει να σημειωθεί πως στην περίπτωση των OA-G3/G6 οι ηλεκτροστατικές αλληλεπιδράσεις αποσυνδέθηκαν εύρος λ από 0 – 0.5, αντίθετα με την περίπτωση του CB8-G3 όπου η αποσύνδεση πραγματοποιήθηκε σε εύρος λ 0 – 0.9. Αυτό πραγματοποιήθηκε με σκοπό την επίλυση μιας προβληματικής κατάστασης που παρατηρήθηκε στα συστήματα των όκτα-οξέων. Συγκεκριμένα, παρατηρήθηκε ότι ιόντα νατρίου προσεγγίζουν σε πολύ μικρή απόσταση από την καρβοξυλομάδα των G3 και G6, εισάγοντας μια σημαντική ασυμφωνία μεταξύ των forward και backward μετασχηματισμών στο solvent και το complex σκέλος της προσομοίωσης (Σχήμα 4.46).

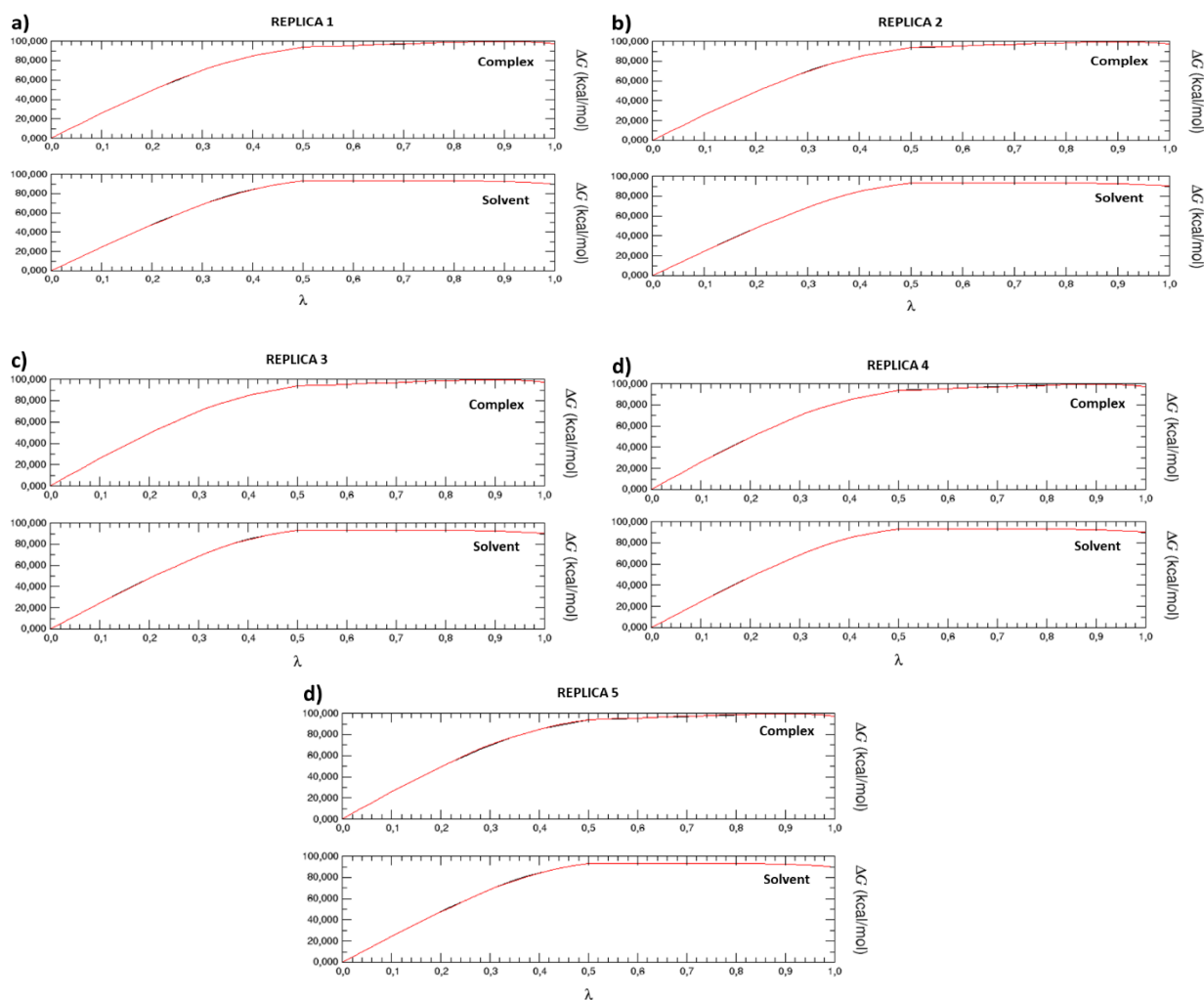


**Σχήμα 4.6** Η συνεχής πρόσδεση ενός ιόντος νατρίου (μπλε σφαίρα) στην ομάδα καρβοξυλίου του 5-εξεοϊκού οξέος, εισάγει σημαντικό σφάλμα στον υπολογισμό της απόλυτης ελεύθερης ενέργειας πρόσδεσης των OA-G3/G6 συστημάτων.

Για την ανάλυση των τροχιών που λήφθηκαν και την εξαγωγή της διαφοράς ελεύθερης ενέργειας σε κάθε φάση χρησιμοποιείται το εργαλείο ParseFEP που είναι ενσωματωμένο στο λογισμικό VMD και επικαλείται τη μέθοδο BAR. Ο υπολογισμός της απόλυτης ελεύθερης ενέργειας πρόσδεσης γίνεται με βάση την Εξίσωση 3.1, αγνοώντας την επίδραση του όρου  $\Delta G_{rest}$ . Στα Σχήματα 4.7 και 4.8 παρουσιάζεται η συνολική μεταβολή της ελεύθερης ενέργειας των ενδιάμεσων  $\lambda$  καταστάσεων για κάθε ένα από τα αντίγραφα των OA-G3/G6. Γίνεται αντιληπτό πως το τροποποιημένο πρωτόκολλο ήταν ακριβές για τα συστήματα OA-G3/G6, καθώς τόσο οι forward όσο και οι backward μετασχηματισμοί συγκλίνουν στην ίδια τιμή  $\Delta G$ .



**Σχήμα 4.7** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για όλα τα αντίγραφα του συστήματος OA-G3 με χρήση 32 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 20 ns/λ κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. **(a)** Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας για το σύστημα OA-G3-0 βρέθηκε  $89.14 \pm 0.01$  kcal/mol και  $96.00 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(b)** Στο σύστημα OA-G3-1 η μεταβολή της ελεύθερης ενέργειας εκτιμάται στα  $89.18 \pm 0.02$  kcal/mol και  $96.00 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(c)** Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας του συστήματος OA-G3-2 υπολογίζεται ίση με  $89.17 \pm 0.01$  kcal/mol και  $95.95 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(d)** Η προβλεπόμενη μεταβολή ελεύθερης ενέργειας του συστήματος OA-G3-3 υπολογίζεται ίση με  $89.19 \pm 0.01$  kcal/mol και  $95.89 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(e)** Η μεταβολή της ελεύθερης ενέργειας του συστήματος OA-G3-4 εκτιμάται στα  $89.13 \pm 0.01$  kcal/mol και  $-95.98 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα.

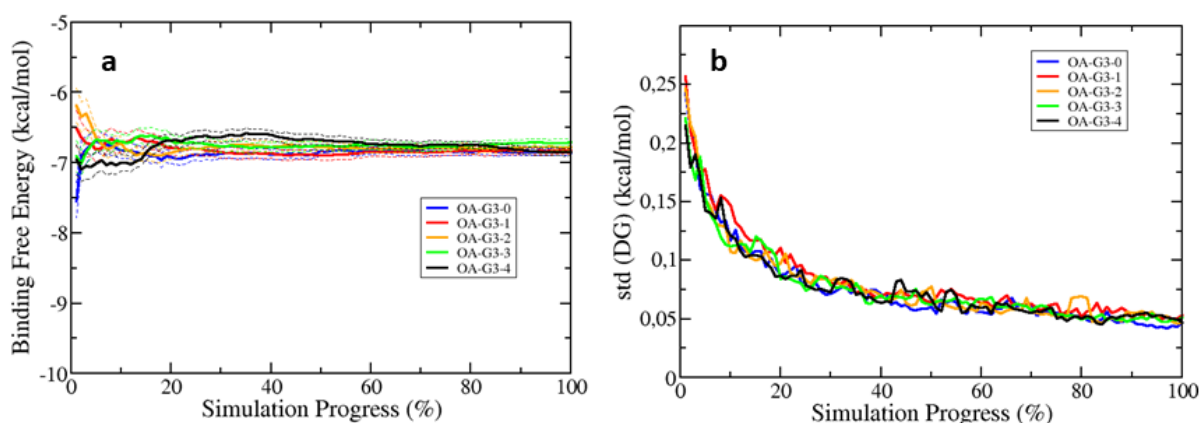


**Σχήμα 4.8** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για όλα τα αντίγραφα του συστήματος OA-G6 με χρήση 32 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 20 ns/λ κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. **(a)** Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας του συστήματος OA-G6-0 υπολογίζεται ίση με  $90.12 \pm 0.02$  kcal/mol και  $97.29 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(b)** Στο σύστημα OA-G6-1 η μεταβολή της ελεύθερης ενέργειας εκτιμάται ίση με  $90.13 \pm 0.02$  kcal/mol και  $97.44 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(c)** Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας του συστήματος OA-G6-2 υπολογίζεται ίση με  $90.14 \pm 0.01$  kcal/mol και  $97.47 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(d)** Η μεταβολή ελεύθερης ενέργειας του συστήματος OA-G6-3 βρέθηκε ίση με  $90.16 \pm 0.01$  kcal/mol και  $97.44 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα. **(e)** Η μεταβολή της ελεύθερης ενέργειας του συστήματος OA-G6-4 εκτιμάται στα  $90.11 \pm 0.01$  kcal/mol και  $-95.98 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις της προσομοίωσης, αντίστοιχα.

Στον Πίνακα 4.2 και 4.3 συνοψίζονται οι υπολογισμένες ελεύθερες ενέργειες για την προσομοίωση με το υπολογιστικό πακέτο NAMD για την κάθε φάση και αντίγραφο των συστημάτων OA-G3 και OA-G6, αντίστοιχα. Επιπλέον παρατίθενται τα αποτελέσματα για την απόλυτη ελεύθερη ενέργεια πρόσδεσης  $\Delta G_{\text{Bind}}$  και το υπολογιστικό κόστος των προσομοιώσεων. Τέλος, για την παρακολούθηση της συμπεριφοράς της μεταβολής της ελεύθερης ενέργειας κατά τη διάρκεια της προσομοίωσης, εκτιμήθηκε η απόλυτη ελεύθερη ενέργεια πρόσδεσης καθώς και το αντίστοιχο σφάλμα για 1%, 2%, ..., 100% της προσομοίωσης κάθε αντιγράφου των συστημάτων OA-G3/G6 (Σχήματα 4.9 και 4.10).

**Πίνακας 4.2** Σύνοψη των υπολογισμών απόλυτης ελεύθερης ενέργειας για την κάθε αρχική διαμόρφωση του συστήματος OA-G3.

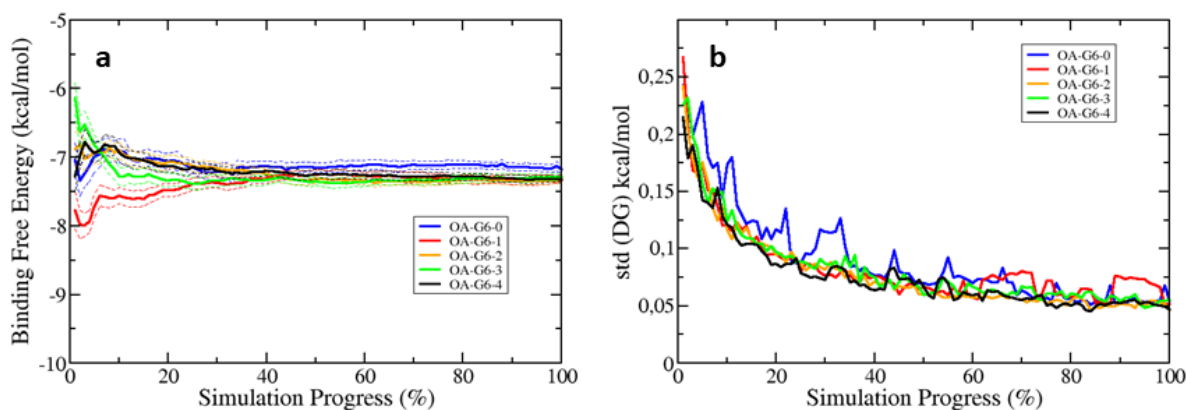
| Αρχική διαμόρφωση | $\Delta G_{\text{solvent}}$<br>(kcal/mol) | $\Delta G_{\text{complex}}$<br>(kcal/mol) | $\Delta G_{\text{Bind}}$<br>(kcal/mol) | Συνολικός χρόνος<br>προσομοίωσης (h) |
|-------------------|---|---|--|--------------------------------------|
| OA-G3-0           | 96.00 ± 0.04                              | 89.14 ± 0.01                              | -6.85 ± 0.05                           | 632                                  |
| OA -G3-1          | 96.00 ± 0.05                              | 89.18 ± 0.02                              | -6.82 ± 0.05                           | 645                                  |
| OA -G3-2          | 95.95 ± 0.05                              | 89.17 ± 0.01                              | -6.78 ± 0.05                           | 645                                  |
| OA -G3-3          | 95.89 ± 0.05                              | 89.19 ± 0.01                              | -6.71 ± 0.05                           | 652                                  |
| OA -G3-4          | 95.98 ± 0.05                              | 89.13 ± 0.01                              | -6.85 ± 0.05                           | 637                                  |



**Σχήμα 4.9** Εξέλιξη της ελεύθερης ενέργειας πρόσδεσης και τυπικής απόκλισης κάθε αρχικής διαμόρφωσης OA-G3 καθ' όλη τη διάρκεια της προσομοίωσης. **(α)** Αντιπαράθεση των απόλυτων ελεύθερων ενεργειών πρόσδεσης για τα 5 αντίγραφα του συστήματος OA-G3. Οι έντονες συνεχείς γραμμές υποδηλώνουν την ελεύθερη ενέργεια πρόσδεσης, ενώ οι διακεκομμένες υποδεικνύουν την αβεβαιότητα της εκτίμησης. **(β)** Τυπική απόκλιση των εκτιμώμενων ελεύθερων ενεργειών για όλες τις αρχικές διαμορφώσεις του συστήματος OA-G3.

**Πίνακας 4.3** Σύνοψη των υπολογισμών απόλυτης ελεύθερης ενέργειας για την κάθε αρχική διαμόρφωση του συστήματος OA-G6.

| Αρχική διαμόρφωση | $\Delta G_{\text{solvent}}$<br>(kcal/mol) | $\Delta G_{\text{complex}}$<br>(kcal/mol) | $\Delta G_{\text{Bind}}$<br>(kcal/mol) | Συνολικός χρόνος<br>προσομοίωσης (h) |
|-------------------|---|---|--|--------------------------------------|
| OA-G6-0           | 97.29 ± 0.05                              | 90.12 ± 0.02                              | -7.17 ± 0.06                           | 639                                  |
| OA -G6-1          | 97.44 ± 0.05                              | 90.13 ± 0.02                              | -7.31 ± 0.05                           | 642                                  |
| OA -G6-2          | 97.47 ± 0.04                              | 90.14 ± 0.01                              | -7.33 ± 0.05                           | 637                                  |
| OA -G6-3          | 97.44 ± 0.05                              | 90.16 ± 0.01                              | -7.28 ± 0.06                           | 639                                  |
| OA -G6-4          | 95.98 ± 0.05                              | 90.11 ± 0.01                              | -7.32 ± 0.05                           | 636                                  |



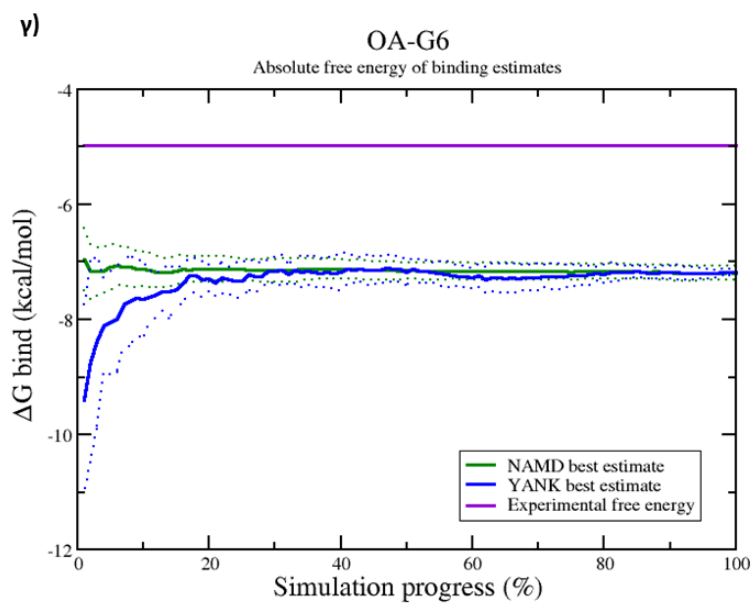
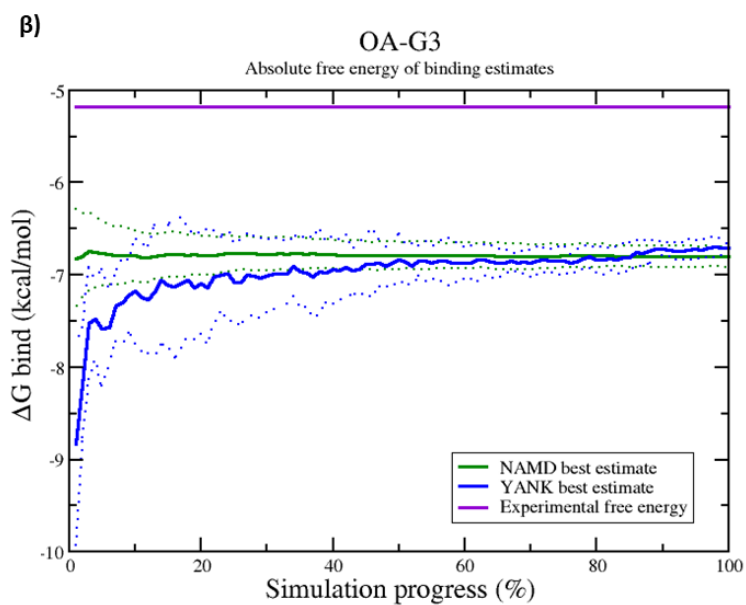
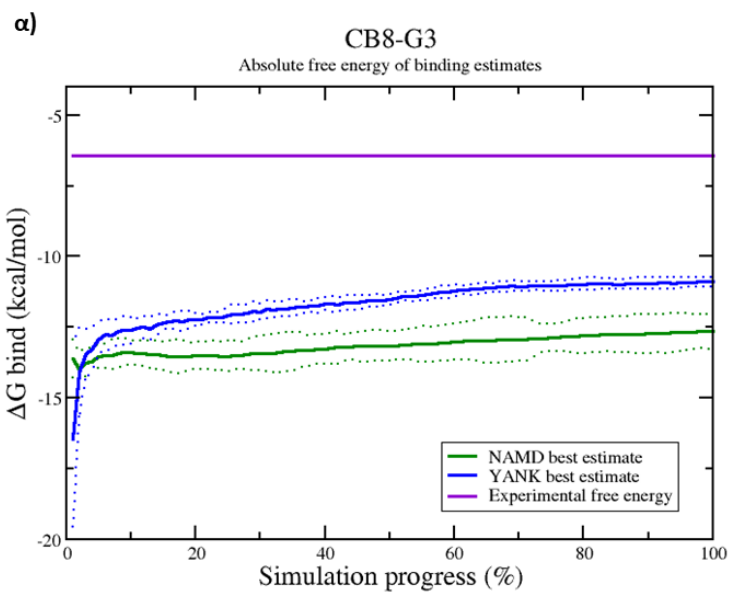
**Σχήμα 4.10** Εξέλιξη της ελεύθερης ενέργειας πρόσδεσης και τυπικής απόκλισης κάθε αρχικής διαμόρφωσης OA-G6 καθ' όλη τη διάρκεια της προσομοίωσης. **(a)** Αντιπαράθεση των απόλυτων ελεύθερων ενεργειών πρόσδεσης για τα 5 αντίγραφα του συστήματος OA-G6. Οι έντονες συνεχείς γραμμές υποδηλώνουν την ελεύθερη ενέργεια πρόσδεσης, ενώ οι διακεκομμένες υποδεικνύουν την αβεβαιότητα της εκτίμησης. **(b)** Τυπική απόκλιση των εκτιμώμενων ελεύθερων ενεργειών για όλες τις αρχικές διαμορφώσεις του συστήματος OA-G6.

Από τα Σχήματα 4.7 και 4.8 γίνεται αντιληπτό πως το υπολογιστικό πακέτο NAMD εκτίμησε με μεγάλη ακρίβεια την πρόσδεση του 5-εξενοϊκού (G3) και του 4-μεθυλοπεντανοϊκού οξέος (G6) στον μακρομοριακό δέκτη OA, καθώς οι forward και backward μετασχηματισμοί πρακτικά ταυτίζονται. Από τα δεδομένα των Πινάκων 4.9 και 4.10, αντιλαμβάνεται κανείς πως οι υπολογισμοί NAMD/FEP συνέκλιναν σε μία πρακτικά σταθερή τιμή για τα αντίγραφα και των δύο συστημάτων. Συγκεκριμένα, η απόλυτη ελεύθερη ενέργεια πρόσδεσης εκτιμήθηκε στα  $-6.8 \pm 0.1$  kcal/mol και  $-7.3 \pm 0.1$  kcal/mol για τα συστήματα OA-G3 και OA-G6, αντίστοιχα. Τα αποτελέσματα του NAMD και για τα δύο συστήματα, είναι σε συμφωνία με τους υπολογισμούς αναφοράς που πραγματοποιήθηκαν από τους διοργανωτές της πρόκλησης SAMPL6 SAMPLING με το υπολογιστικό πακέτο YANK. Συγκεκριμένα οι προσομοιώσεις με το πακέτο YANK εκτίμησαν την ελεύθερη ενέργεια πρόσδεσης στα  $-6.7 \pm 0.1$  kcal/mol και  $-7.2 \pm 0.1$  kcal/mol για τα συστήματα OA-G3 και OA-G6, αντίστοιχα.

Αξίζει να αναφερθεί πως και τα δύο υπολογιστικά πακέτα κατέληξαν σε βελτιωμένη συσχέτιση με τα πειραματικά αποτελέσματα, σε σχέση με το σύστημα CB8-G3, καθώς οι μετρήσεις με τη μέθοδο ITC εκτίμησαν την μεταβολή της ελεύθερης ενέργειας στα  $-5.2$  kcal/mol και  $-4.97$  kcal/mol για τα συστήματα OA-G3 και OA-G6, αντίστοιχα. Σε γενικές γραμμές, οι υπόλοιπες μεθοδολογίες ελεύθερης ενέργειας που εφαρμόστηκαν από τις υπόλοιπες ομάδες που συμμετείχαν στην πρόκληση SAMPL6 SAMPLING, κατέληξαν σε παρόμοια αποτελέσματα με αυτά των προβλέψεων από τα υπολογιστικά πακέτα NAMD και YANK. Η απόκλιση μεταξύ εκτιμήσεων υπολογιστικής και πειραματικής ελεύθερης ενέργειας μπορεί να προκύψει εξαιτίας διάφορων λόγων, όπως οι παράμετροι πεδίου δυνάμεων, το χρησιμοποιούμενο μοντέλο νερού και η ανεπάρκεια των αλγορίθμων δειγματοληψίας.

Στο Σχήμα 4.11 παρατίθεται η εξέλιξη της καλύτερης εκτίμησης της απόλυτης ελεύθερης ενέργειας πρόσδεσης, η οποία έχει προκύψει από τη μέση τιμή των υπολογισμών για όλες τις αρχικές διαμορφώσεις κάθε συστήματος, για τα πακέτα NAMD και YANK. Επιπλέον, για κάθε σύστημα παρατίθενται και τα αντίστοιχα πειραματικά αποτελέσματα.





**Σχήμα 4.11** Σύγκριση των καλύτερων εκτιμήσεων της απόλυτης ελεύθερης ενέργειας πρόσδεσης μεταξύ των πακέτων NAMD και YANK, για την πρόσδεση (**α**) της κινίνης στον δέκτη cucurbit[8]uril, (**β**) του 5-εξενοϊκού και (**γ**) του 4-μεθυλοπεντανοϊκού οξέος στον δέκτη OA. Οι πειραματικές μετρήσεις ελεύθερης ενέργειας πρόσδεσης κάθε συστήματος απεικονίζονται με μωβ, ενώ με πράσινο και μπλε αναπαρίσταται η μεταβολή της ελεύθερης ενέργειας που υπολογίσθηκε με τα πακέτα NAMD και YANK, αντίστοιχα. Οι διακεκομμένες γραμμές αντιπροσωπεύουν την τυπική απόκλιση κάθε υπολογιστικής εκτίμησης της απόλυτης ελεύθερης ενέργειας πρόσδεσης.

Τέλος, η ασυμφωνία μεταξύ των αμφίδρομων μετασχηματισμών που παρατηρήθηκε για το σύστημα CB8-G3 (βλ. Σχήμα 4.4) είναι ένα φαινόμενο το οποίο έχει αναφερθεί πρόσφατα από τον de Vaca Cabeza και τους συνεργάτες του [128]. Στο συγκεκριμένο άρθρο, οι συγγραφείς προτείνουν πως στους υπολογισμούς απόλυτης ελεύθερης ενέργειας πρόσδεσης, το στάδιο κατά το οποίο «δημιουργείται» ο προσδέτης, οδηγεί σε αποτελέσματα ελεύθερης ενέργειας πιο συνεπή με τα πειραματικά σε σχέση με το στάδιο της «εξαφάνισης» του μορίου, δίχως να υπάρχει η ανάγκη για επιπρόσθετη επεξεργασία με μεθόδους όπως η BAR. Στην προκειμένη περίπτωση η δημιουργία (creation) του προσδέτη λαμβάνει χώρα κατά τον backward μετασχηματισμό, ενώ η εξαφάνισή (annihilation) του κατά την forward προσομοίωση. Για τον έλεγχο της υπόθεσης αυτής με βάση τα δεδομένα που λήφθηκαν κατά τις προσομοιώσεις απόλυτης ενέργειας, υπολογίσθηκε ξεχωριστά η αλλαγή της ελεύθερης ενέργειας για κάθε ένα στάδιο. Πράγματι, η χρήση της εξίσωσης του Zwanzig για το στάδιο της δημιουργίας του μορίου για τα περισσότερα αντίγραφα των συστημάτων προσδέτη-δέκτη (Πίνακας 4.4 έως 4.6) κατέληξε σε προβλέψεις πιο κοντινές με την μέθοδο BAR, σε σχέση με το στάδιο της εξαφάνισης, επιβεβαιώνοντας έτσι τα συμπεράσματα της αναφοράς [128].

**Πίνακας 4.4** Εκτιμήσεις ελεύθερης ενέργειας πρόσδεσης για το σύστημα CB8-G3 με χρήση της μεθόδου BAR, και μέσω των μετασχηματισμών δημιουργίας (creation) και εξαφάνισης (annihilation).

| Αντίγραφο<br>συστήματος CB8-G3 | BAR<br>(kcal/mol) | Creation<br>(kcal/mol) | Annihilation<br>(kcal/mol) |
|--------------------------------|-------------------|------------------------|----------------------------|
| 1                              | -11.74 ± 0.31     | -11.55                 | -12.32                     |
| 2                              | -13.89 ± 0.31     | -13.72                 | -14.88                     |
| 3                              | -11.44 ± 0.27     | -11.30                 | -12.54                     |
| 4                              | -13.71 ± 0.22     | -13.30                 | -14.14                     |
| 5                              | -12.45 ± 0.22     | -14.78                 | -9.73                      |

**Πίνακας 4.5** Εκτιμήσεις ελεύθερης ενέργειας πρόσδεσης για το σύστημα OA-G3 με χρήση της μεθόδου BAR, και μέσω των μετασχηματισμών δημιουργίας (creation) και εξαφάνισης (annihilation).

| Αντίγραφο<br>συστήματος OA-G3 | BAR<br>(kcal/mol) | Creation<br>(kcal/mol) | Annihilation<br>(kcal/mol) |
|-------------------------------|-------------------|------------------------|----------------------------|
| 1                             | -6.85 ± 0.05      | -6.80                  | -6.94                      |
| 2                             | -6.82 ± 0.05      | -6.82                  | -6.88                      |
| 3                             | -6.78 ± 0.05      | -6.74                  | -6.87                      |
| 4                             | -6.71 ± 0.05      | -6.63                  | -6.80                      |
| 5                             | -6.85 ± 0.05      | -6.82                  | -6.93                      |

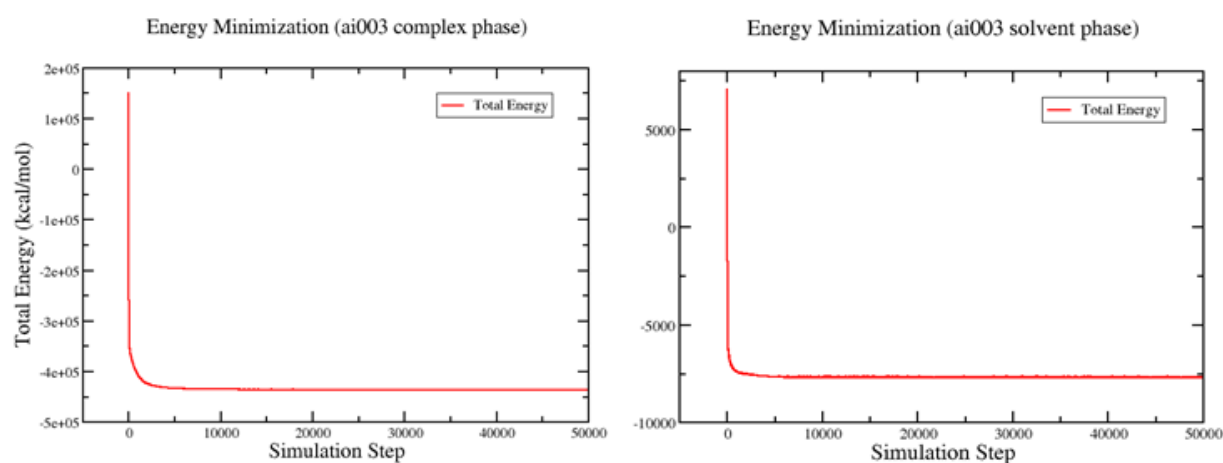
**Πίνακας 4.6** Εκτιμήσεις ελεύθερης ενέργειας πρόσδεσης για το σύστημα OA-G6 με χρήση της μεθόδου BAR, και μέσω των μετασχηματισμών δημιουργίας (creation) και εξαφάνισης (annihilation).

| Αντίγραφο συστήματος OA-G6 | BAR (kcal/mol) | Creation (kcal/mol) | Annihilation (kcal/mol) |
|----------------------------|----------------|---------------------|-------------------------|
| 1                          | -7.17 ± 0.06   | -7.10               | -7.30                   |
| 2                          | -7.31 ± 0.05   | -7.33               | -7.31                   |
| 3                          | -7.33 ± 0.05   | -7.25               | -7.48                   |
| 4                          | -7.28 ± 0.06   | -7.26               | -7.24                   |
| 5                          | -7.32 ± 0.05   | -7.31               | -7.46                   |

## 4.2 Αποτελέσματα μοριακών προσομοιώσεων για τις ανάλογες ενώσεις του CK-666

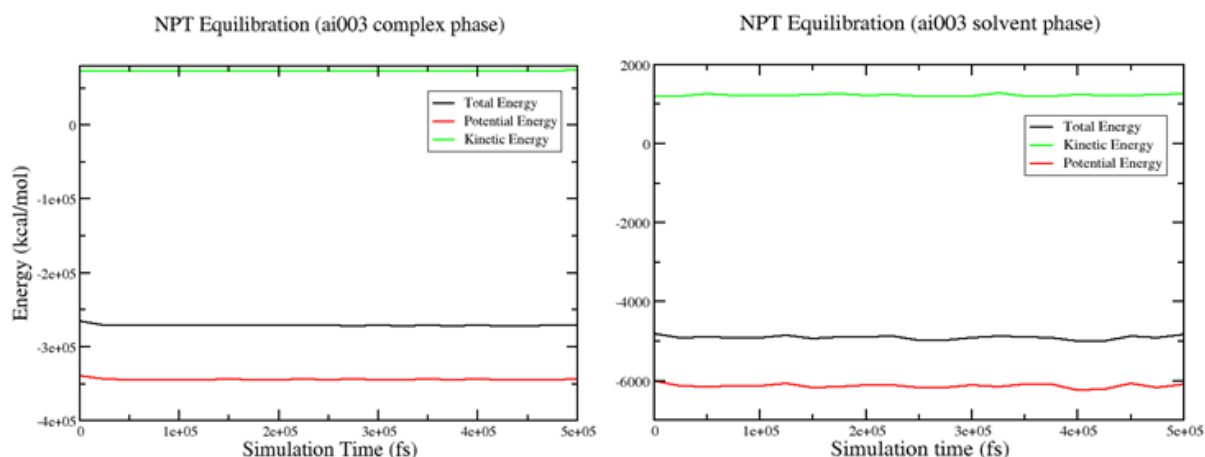
### 4.2.1 Εκτίμηση σχετικής ελεύθερης ενέργειας πρόσδεσης με NAMD και σύγκριση με πειραματικά δεδομένα

Τα δεκαπέντε ανάλογα μόρια της ένωσης οδηγού CK-666 τα οποία προετοιμάστηκαν με τη διαδικασία που περιγράφεται στο Κεφάλαιο 3.2.3.1, υποβλήθηκαν σε προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο NAMD. Αρχικά, σε κάθε δομή πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας για 50,000 βήματα προσομοίωσης. Ο αριθμός των βημάτων ήταν αρκετός ώστε να συγκλίνει η ενέργεια του κάθε συστήματος σε μία σταθερή τιμή. Η σύγκλιση επιβεβαιώνεται στο Σχήμα 4.12, όπου απεικονίζονται η συνολική ενέργεια του συστήματος ως συνάρτηση των βημάτων προσομοίωσης. Σημειώνεται ότι το Σχήμα 4.12 κατασκευάστηκε από τα αποτελέσματα της ελαχιστοποίησης ενέργειας για την complex και solvent φάση της ανάλογης ένωσης ai003. Το διάγραμμα αυτό αποτελεί ένα ενδεικτικό παράδειγμα, καθώς τα υπόλοιπα συστήματα ανάλογων ενώσεων ακολουθούν το ίδιο μοτίβο σύγκλισης της ενέργειας κατά την ελαχιστοποίηση, και για τις δύο φάσεις.



**Σχήμα 4.12** Διάγραμμα της συνολικής ενέργειας του συστήματος συναρτήσει των βημάτων προσομοίωσης κατά την ελαχιστοποίηση της ενέργειας με το υπολογιστικό πακέτο NAMD για την complex (αριστερά) και solvent (δεξιά) φάση της προσομοίωσης της ανάλογης ένωσης ai003.

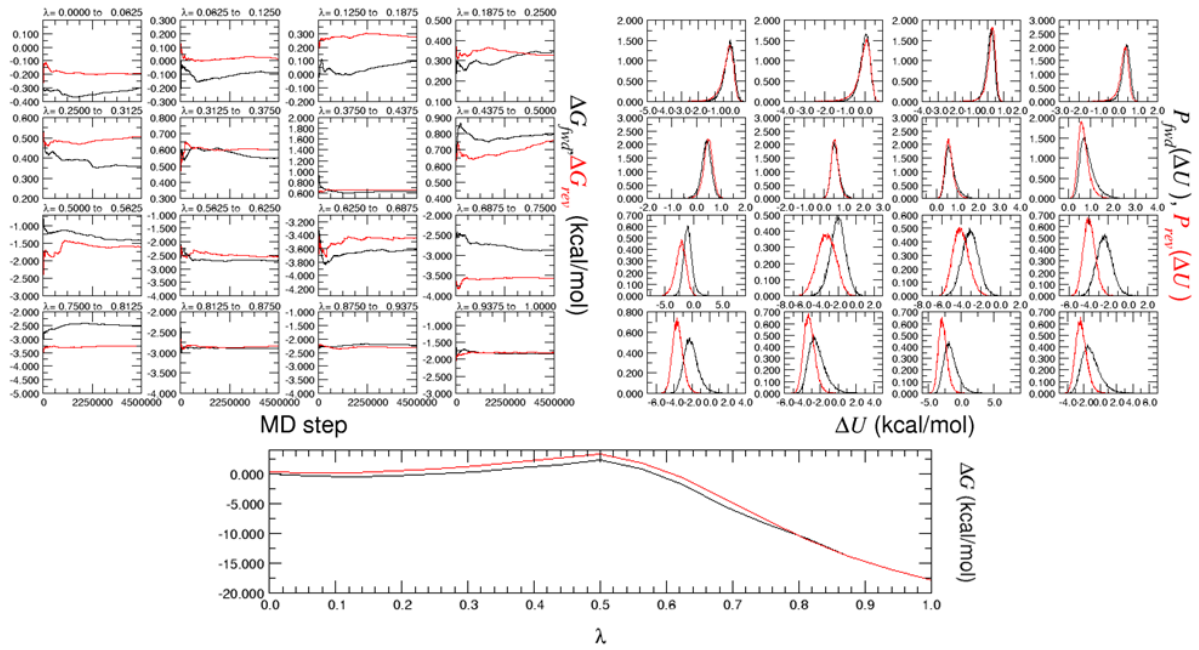
Στη συνέχεια ακολούθησαν δύο φάσεις εξισορρόπησης, αρχικά NVT και ύστερα NPT, κάθε μία από τις οποίες είχε διάρκεια προσομοίωσης 0.5 ns, ενώ και στις δύο περιπτώσεις το βήμα ήταν 1 fs. Έπειτα από την ολοκλήρωση των σταδίων αυτών, μελετήθηκαν οι ιδιότητες των συστημάτων όπως η θερμοκρασία και η ενέργεια ως προς το χρόνο και παρατηρήθηκε ότι δεν υπάρχουν μεγάλες διακυμάνσεις στις τιμές τους. Επομένως, η εξισορρόπηση πραγματοποιήθηκε με επιτυχία και συνεπώς μπορούν να ακολουθήσουν οι παραγωγικές προσομοιώσεις για όλα τα συστήματα. Στο Σχήμα 4.13 παρουσιάζεται ενδεικτικά η γραφική παράσταση της συνολικής, κινητικής και δυναμικής ενέργειας για την NPT εξισορρόπηση, συναρτήσει του χρόνου προσομοίωσης για τις δύο φάσεις της ανάλογης ένωσης ai003.



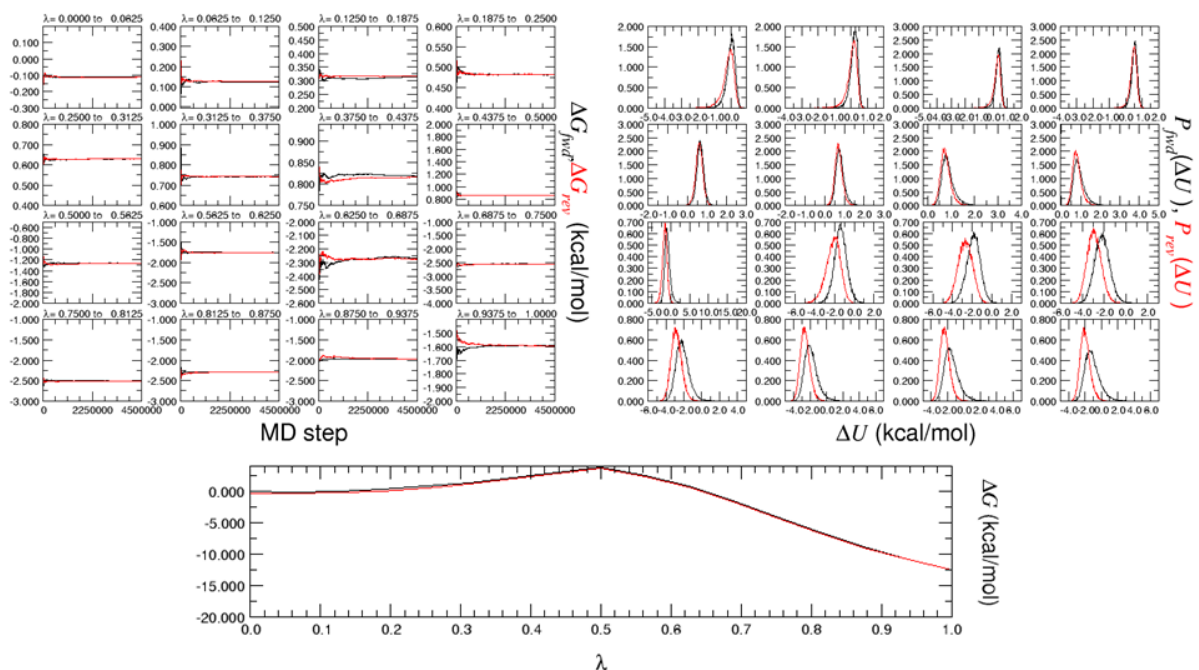
**Σχήμα 4.13** Διαγράμματα της κινητικής, δυναμικής και συνολικής ενέργειας του συστήματος συναρτήσει του χρόνου προσομοίωσης για την complex (αριστερά) και solvent (δεξιά) φάση της προσομοίωσης της ανάλογης ένωσης ai003.

Στη συνέχεια, ακολούθησαν οι παραγωγικές φάσεις των προσομοιώσεων για τις 15 ανάλογες ενώσεις του CK-666. Ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης γίνεται με βάση την Εξίσωση 3.2. Για την αλχημική μετατροπή της ένωσης οδηγού CK-666 σε κάθε ένα από τα ανάλογα μόρια, χρησιμοποιήθηκαν 16 συνολικά ενδιάμεσες καταστάσεις, με τη μεταβλητή  $\lambda$  να αλλάζει από το 0 προς το 1 με σταθερό βήμα 0.0625. Η προσομοίωση των καταστάσεων αυτών έγινε για 5 ns/κατάσταση με χρονικό βήμα 1fs, ενώ κάθε  $\lambda$  κατάσταση προσομοιώθηκε forward και backward.

Για την ανάλυση των  $\tau$  προσομοιώσεων Μοριακής Δυναμικής και την εξαγωγή της διαφοράς ελεύθερης ενέργειας σε κάθε φάση, χρησιμοποιείται το εργαλείο ParseFEP το οποίο είναι ενσωματωμένο στο λογισμικό VMD και επικαλείται τη μέθοδο BAR. Στο Σχήμα 4.14 και Σχήμα 4.15 απεικονίζονται όλα τα γραφήματα εξόδου που παρέχονται από το ParseFEP για την περίπτωση της ανάλογης ένωσης ai094 για το complex και για το solvent σκέλος της προσομοίωσης, αντίστοιχα.



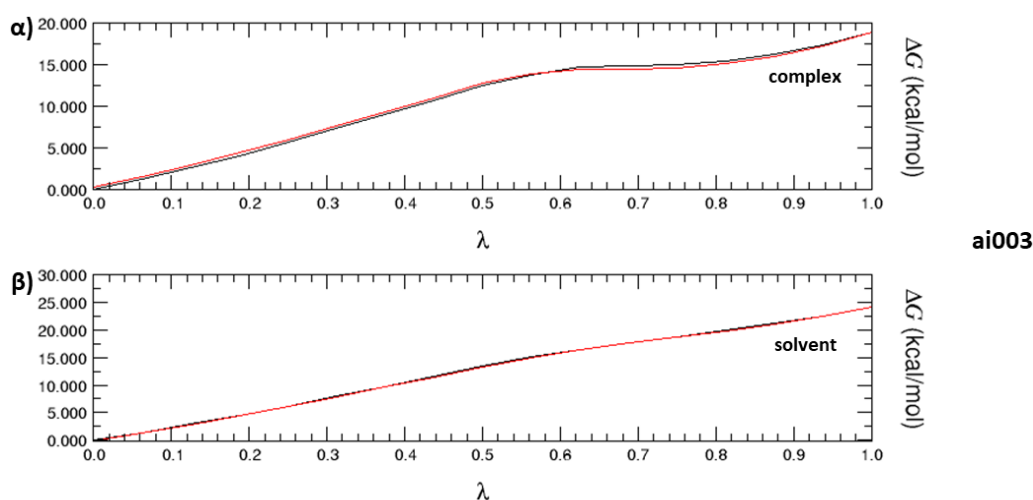
**Σχήμα 4.14** Τυπικά διαγράμματα παρακολούθησης της εξέλιξης της ελεύθερης ενέργειας από το εργαλείο ParseFEP. Οι γραφικές παραστάσεις περιλαμβάνουν τη χρονική εξέλιξη της ελεύθερης ενέργειας (άνω αριστερά), ιστογράμματα των πιθανοτήτων κατανομής της δυναμικής ενέργειας (άνω δεξιά) και τη συνολική μεταβολή της ελεύθερης ενέργειας ως συνάρτηση των ενδιάμεσων  $\lambda$  καταστάσεων (κάτω). Οι μαύρες και κόκκινες συμπαγείς γραμμές υποδηλώνουν τη συνολική αλλαγή ελεύθερης ενέργειας για τον προς τα εμπρός (forward, fwd) και τον αντίστροφο (reverse/backward, rev) μετασχηματισμό, αντίστοιχα. Τα γραφήματα αυτά λήφθηκαν από τις αληθινές προσομοιώσεις του complex σκέλους της ανάλογης ένωσης αι094.



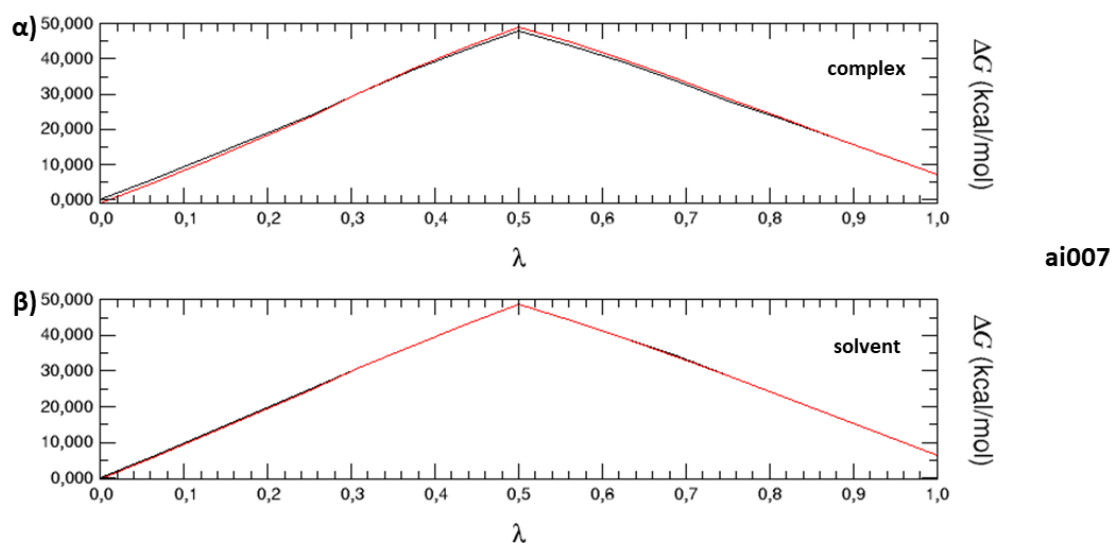
**Σχήμα 4.15** Τυπικά διαγράμματα παρακολούθησης της εξέλιξης της ελεύθερης ενέργειας από το εργαλείο ParseFEP. Οι γραφικές παραστάσεις περιλαμβάνουν τη χρονική εξέλιξη της ελεύθερης ενέργειας (άνω αριστερά), ιστογράμματα των πιθανοτήτων κατανομής της δυναμικής ενέργειας (άνω δεξιά) και τη συνολική μεταβολή της ελεύθερης ενέργειας ως συνάρτηση των ενδιάμεσων  $\lambda$  καταστάσεων (κάτω). Οι μαύρες και κόκκινες συμπαγείς γραμμές υποδηλώνουν τη συνολική αλλαγή ελεύθερης ενέργειας για τον προς τα εμπρός

(forward, fwd) και τον αντίστροφο (reverse/backward, rev) μετασχηματισμό, αντίστοιχα. Τα γραφήματα αυτά λήφθηκαν από τις αλχημικές προσομοιώσεις του solvent σκέλους της ανάλογης ένωσης ai094.

Στη συνέχεια κρίνεται σκόπιμη μόνο η παρουσίαση των γραφημάτων που απεικονίζουν τη συνολική μεταβολή της ελεύθερης ενέργειας συναρτήσει των ενδιάμεσων λ καταστάσεων, καθώς ουσιαστικά ενσωματώνουν τις πληροφορίες που περιέχονται στα διαγράμματα της χρονικής εξέλιξης της ελεύθερης ενέργειας και των ιστογραμμάτων των πιθανοτήτων κατανομής της δυναμικής ενέργειας για κάθε ενδιάμεση λ κατάσταση. Τα Σχήματα 4.16 έως 4.30 απεικονίζουν την εξέλιξη της συνολικής μεταβολής της ελεύθερης ενέργειας για κάθε φάση και ανάλογη ένωση που προσομοιώθηκε.

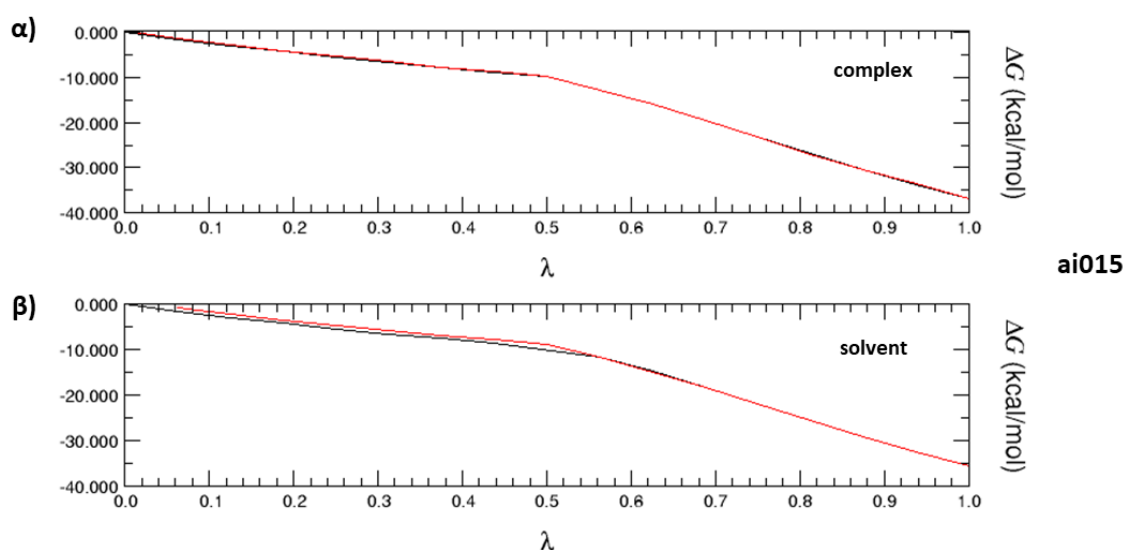


**Σχήμα 4.16** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai003 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $18.64 \pm 0.05$  kcal/mol και  $24.18 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

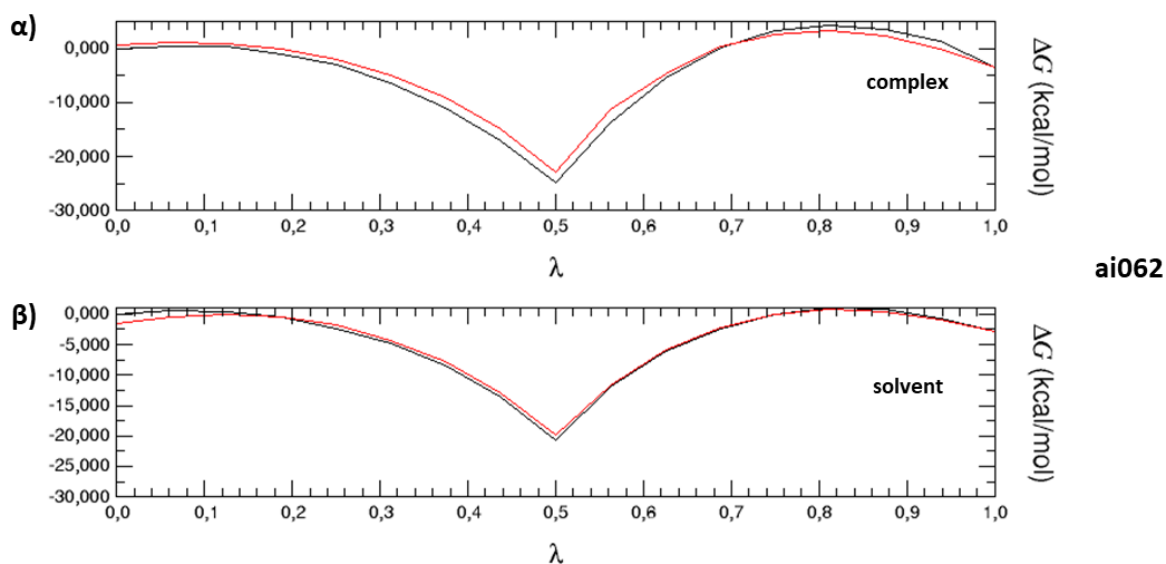


**Σχήμα 4.17** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai007 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα,

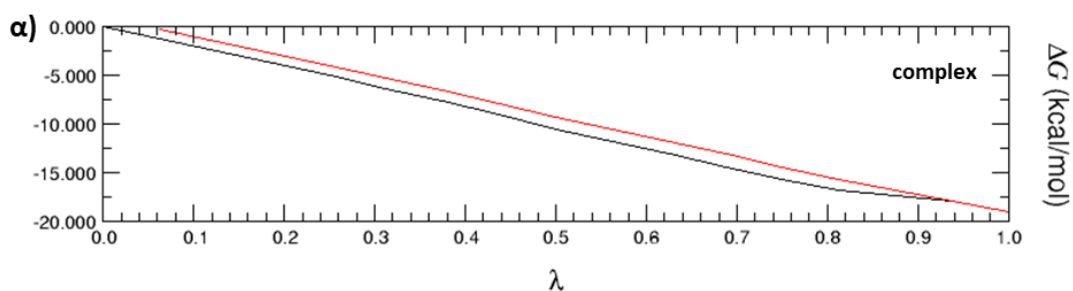
αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $7.69 \pm 0.10$  kcal/mol και  $6.67 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



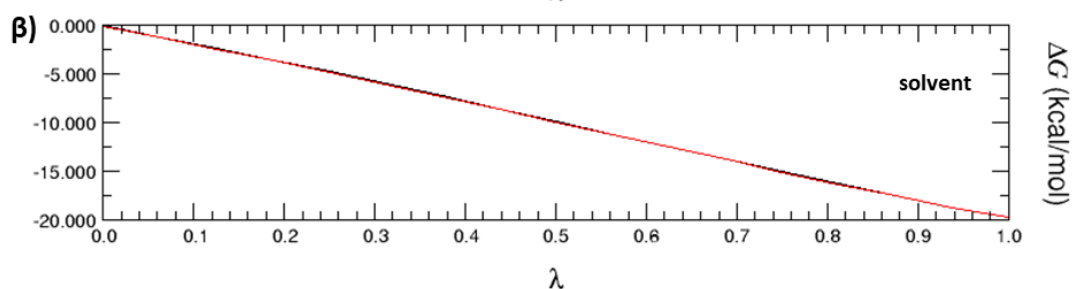
**Σχήμα 4.18** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai015 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-36.84 \pm 0.04$  kcal/mol και  $-35.53 \pm 0.02$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



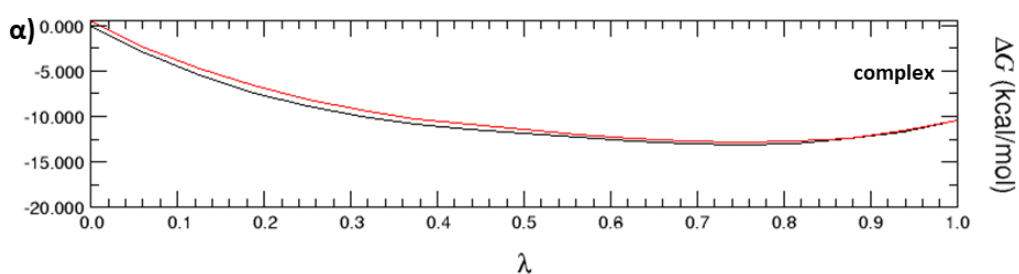
**Σχήμα 4.19** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai062 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-6.27 \pm 0.16$  kcal/mol και  $-3.94 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



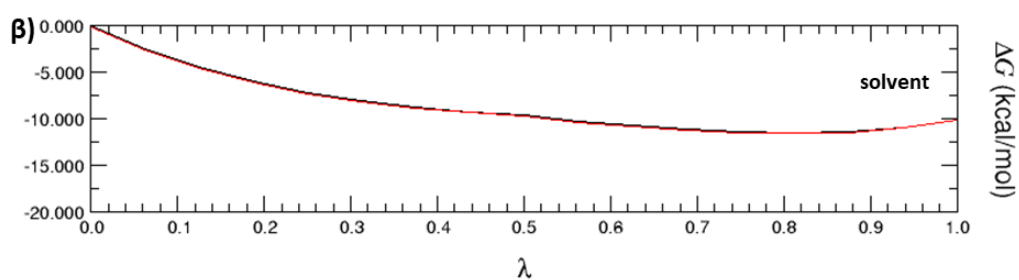
ai064



**Σχήμα 4.20** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai064 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-18.94 \pm 0.02$  kcal/mol και  $-19.70 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

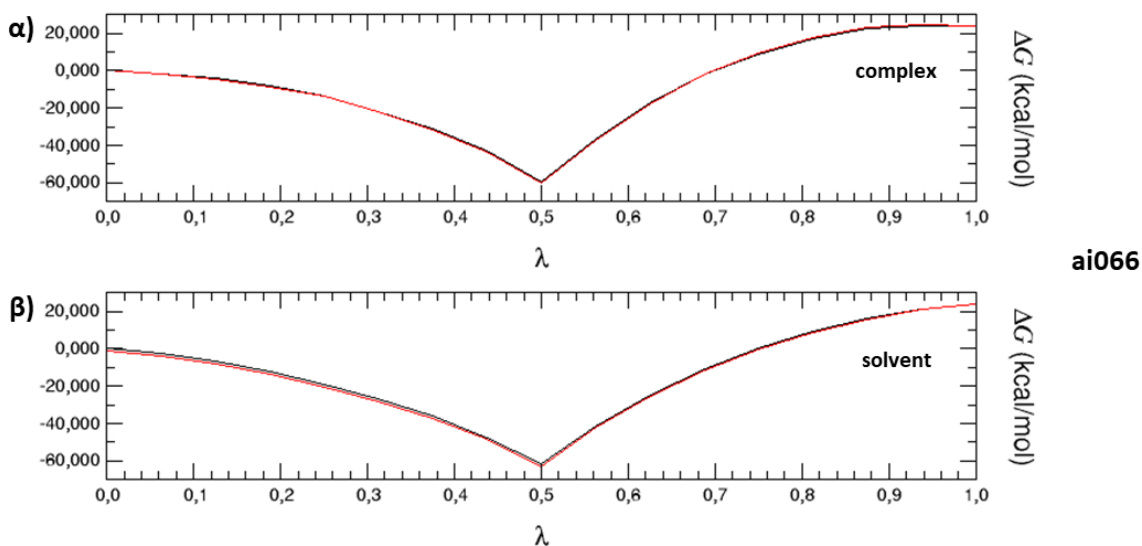


ai065

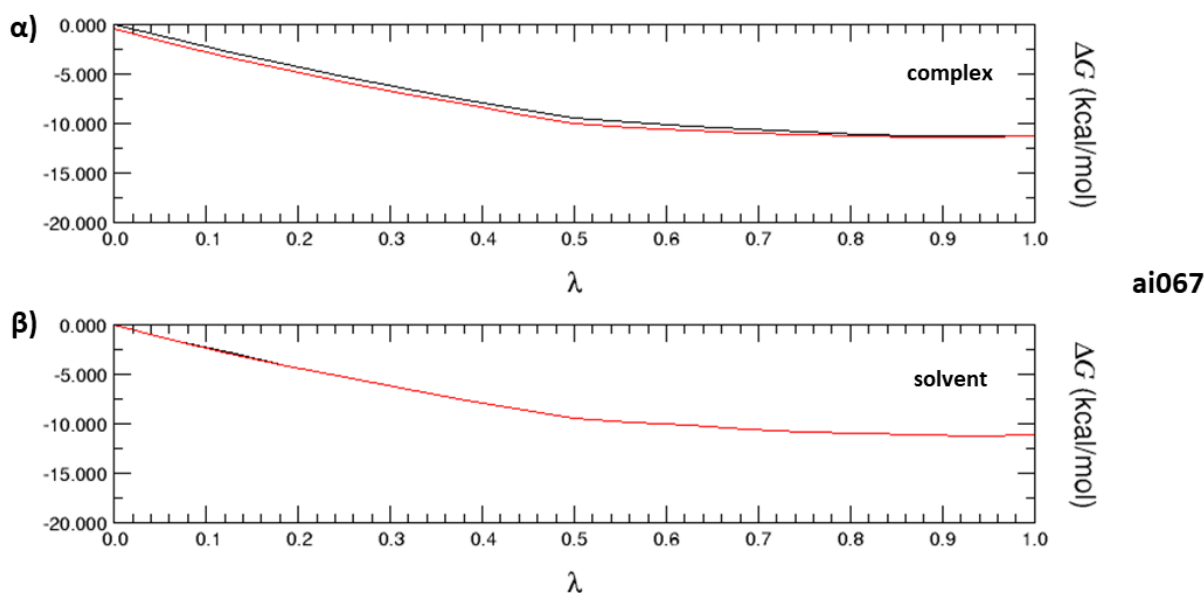


**Σχήμα 4.21** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai065 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-10.48 \pm 0.06$  kcal/mol και  $-10.05 \pm 0.02$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

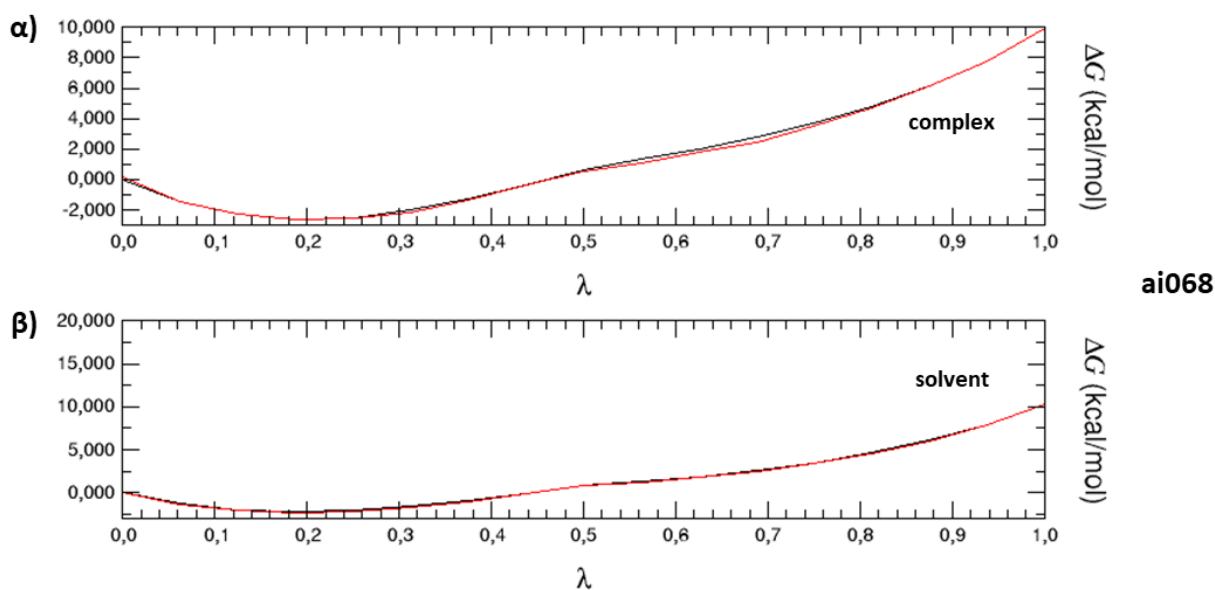




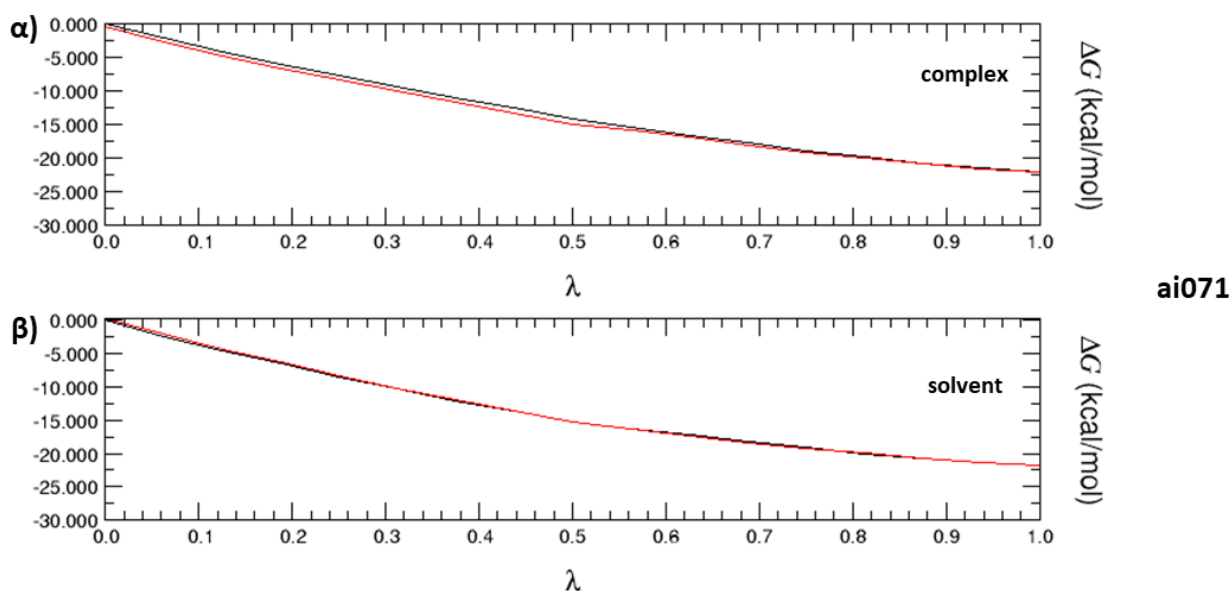
**Σχήμα 4.22** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai066 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $20.88 \pm 0.09$  kcal/mol και  $21.08 \pm 0.08$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



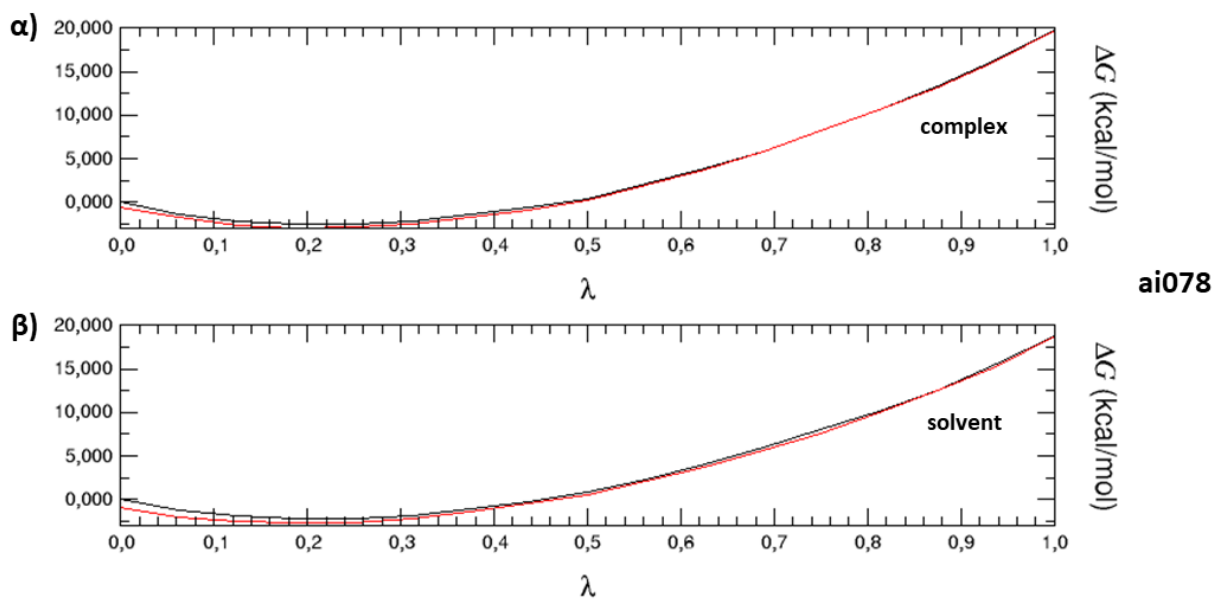
**Σχήμα 4.23** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai067 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-11.05 \pm 0.01$  kcal/mol και  $-11.14 \pm 0.00$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



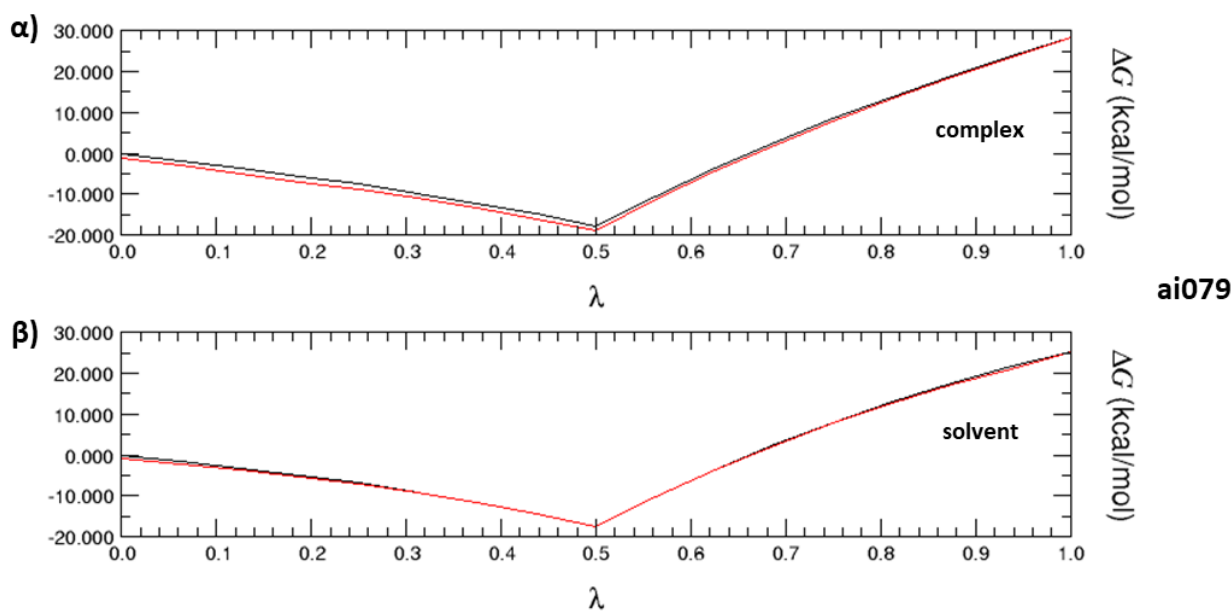
**Σχήμα 4.24** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai068 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $10.57 \pm 0.04$  kcal/mol και  $10.77 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



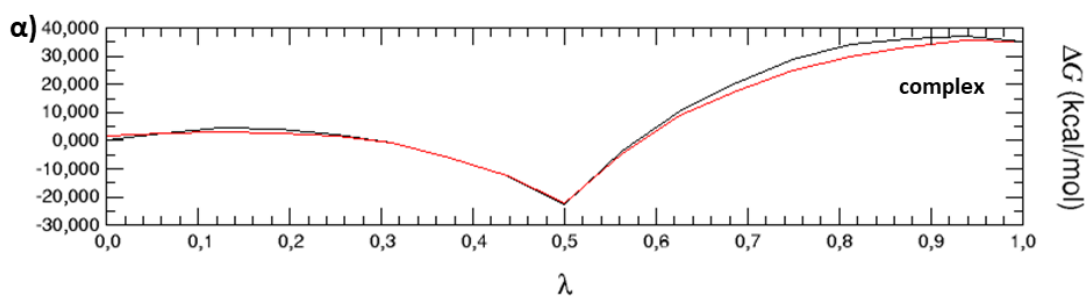
**Σχήμα 4.25** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai071 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-21.78 \pm 0.04$  kcal/mol και  $-21.81 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



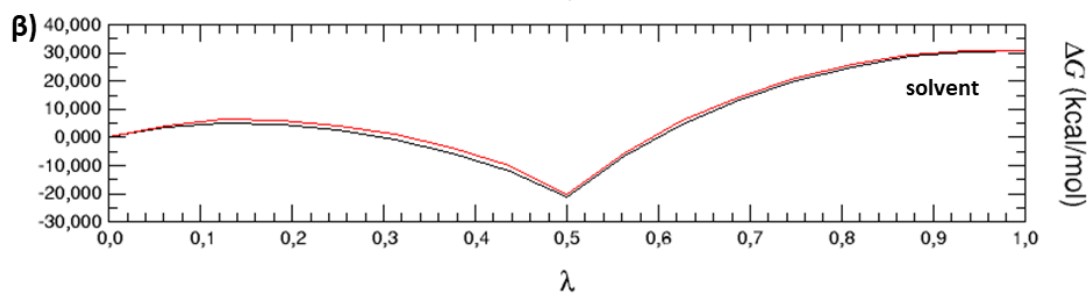
**Σχήμα 4.26** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai078 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $20.49 \pm 0.06$  kcal/mol και  $19.84 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



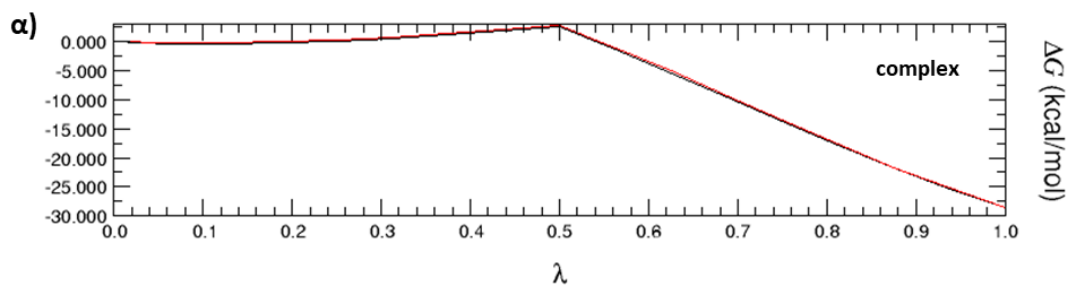
**Σχήμα 4.27** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai079 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $29.04 \pm 0.09$  kcal/mol και  $25.72 \pm 0.03$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



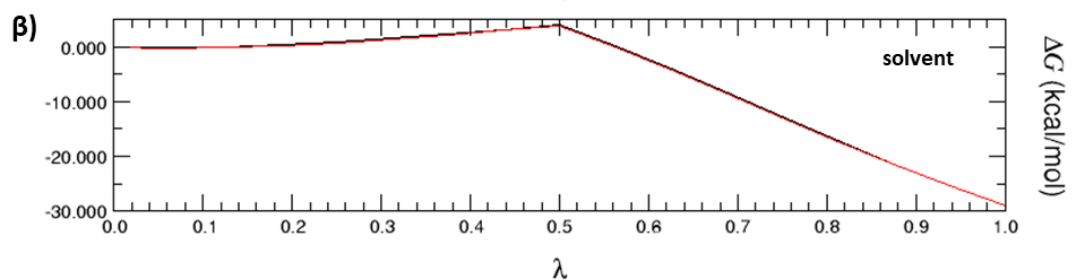
ai086



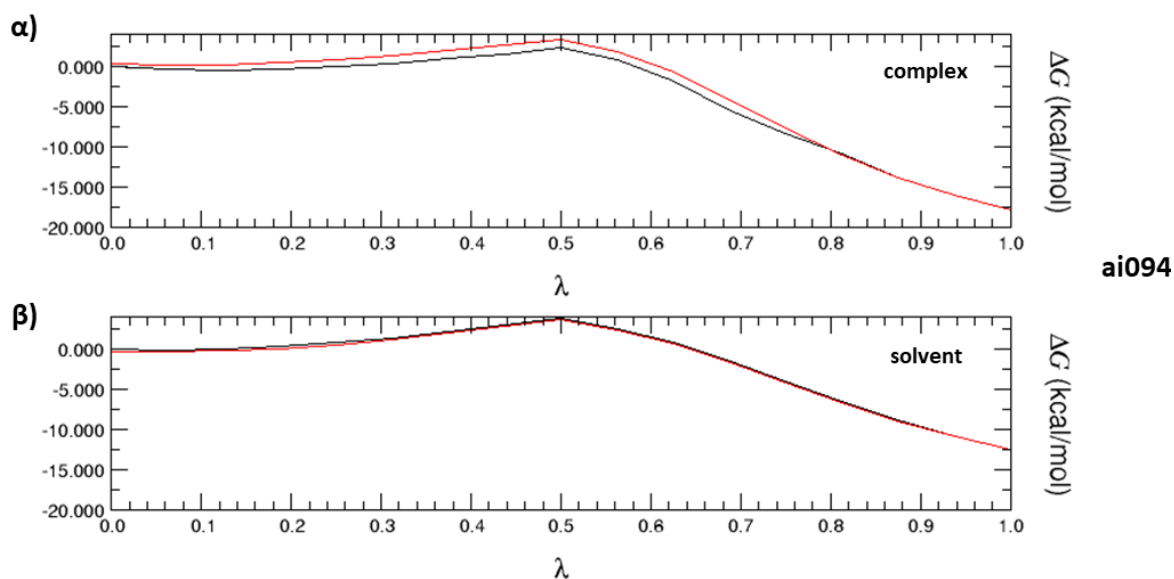
**Σχήμα 4.28** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai086 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $33.26 \pm 0.35$  kcal/mol και  $27.20 \pm 0.10$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



ai093



**Σχήμα 4.29** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai093 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-28.65 \pm 0.04$  kcal/mol και  $-28.87 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



**Σχήμα 4.30** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai094 με χρήση 16 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις complex και solvent προσομοιώσεις αναπαρίσταται στα επάνω και κάτω γραφήματα, αντίστοιχα. Οι μαύρες και κόκκινες συμπαγείς γραμμές αντιστοιχούν στην forward και backward προσομοίωση, αντίστοιχα. Η εκτιμώμενη μεταβολή της ελεύθερης ενέργειας προσομοιώσεις υπολογίζεται ίση με  $-18.12 \pm 0.05$  kcal/mol και  $-12.35 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

Στον Πίνακα 4.7 συνοψίζονται οι υπολογισμένες ελεύθερες ενέργειες για την προσομοίωση με το υπολογιστικό πακέτο NAMD για την κάθε φάση και ανάλογη ένωση. Επιπλέον παρατίθενται τα αποτελέσματα για την σχετική ελεύθερη ενέργεια πρόσδεσης  $\Delta\Delta G$ , η οποία έχει υπολογισθεί με βάση την Εξίσωση 3.2, ενώ το αντίστοιχο σφάλμα υπολογίστηκε με βάση την Εξίσωση 4.1.

**Πίνακας 4.7** Αποτελέσματα σχετικής ελεύθερης ενέργειας για την κάθε ανάλογη ένωση με το υπολογιστικό πακέτο NAMD.

| Ανάλογη Ένωση | $\Delta G_{\text{complex}}$ [kcal/mol] | $\Delta G_{\text{solvent}}$ [kcal/mol] | $\Delta\Delta G_{\text{NAMD}}$ [kcal/mol] |
|---------------|--|--|---|
| CK-666        | $0.00 \pm 0.00$                        | $0.00 \pm 0.00$                        | $0.00 \pm 0.00$                           |
| ai003         | $18.64 \pm 0.05$                       | $24.18 \pm 0.01$                       | $-5.54 \pm 0.05$                          |
| ai007         | $7.69 \pm 0.1$                         | $6.67 \pm 0.01$                        | $1.02 \pm 0.1$                            |
| ai015         | $-36.84 \pm 0.04$                      | $-35.53 \pm 0.02$                      | $-1.31 \pm 0.04$                          |
| ai062         | $-6.27 \pm 0.16$                       | $-3.94 \pm 0.04$                       | $-2.33 \pm 0.17$                          |
| ai064         | $-18.94 \pm 0.02$                      | $-19.7 \pm 0.01$                       | $0.76 \pm 0.02$                           |
| ai065         | $-10.48 \pm 0.06$                      | $-10.05 \pm 0.02$                      | $-0.43 \pm 0.06$                          |
| ai066         | $20.88 \pm 0.09$                       | $21.08 \pm 0.08$                       | $-0.2 \pm 0.12$                           |
| ai067         | $-11.05 \pm 0.01$                      | $-11.14 \pm 0$                         | $0.09 \pm 0.01$                           |
| ai068         | $10.57 \pm 0.04$                       | $10.77 \pm 0.04$                       | $-0.2 \pm 0.05$                           |

|              |               |               |              |
|--------------|---------------|---------------|--------------|
| <b>ai071</b> | -21.78 ± 0.04 | -21.81 ± 0.05 | 0.02 ± 0.06  |
| <b>ai078</b> | 20.49 ± 0.06  | 19.84 ± 0.04  | 0.65 ± 0.07  |
| <b>ai079</b> | 29.04 ± 0.09  | 25.72 ± 0.03  | 3.32 ± 0.09  |
| <b>ai086</b> | 33.26 ± 0.35  | 27.2 ± 0.1    | 6.07 ± 0.37  |
| <b>ai093</b> | -28.65 ± 0.04 | -28.87 ± 0.01 | 0.21 ± 0.04  |
| <b>ai094</b> | -18.12 ± 0.05 | -12.35 ± 0.01 | -5.78 ± 0.05 |

Μετά τον υπολογισμό της σχετικής ελεύθερης ενέργειας με το υπολογιστικό πακέτο NAMD, επόμενο βήμα αποτελεί η σύγκριση των εκτιμήσεων  $\Delta\Delta G$  με πειραματικά δεδομένα. Τα πειραματικά δεδομένα αυτά βασίζονται σε *in vitro* μετρήσεις του ρυθμού πολυμερισμού της ακτίνης. Αρχικά, απομονώθηκε ακτίνη από σκελετικό μυ κουνελιού, η οποία στη συνέχεια σημαδεύτηκε με φθορίζουσα ουσία (rylene-labeled actin). Ύστερα, απομονώθηκε το πρωτεϊνικό σύμπλεγμα Agr2/3 από τον οργανισμό *Bos taurus*, καθώς και η περιοχή N-WASP-VCA (VCA) η οποία είναι απαραίτητη για την ενεργοποίηση του πρωτεϊνικού συμπλέγματος. Για τη μέτρηση της ανασταλτικής ισχύος των ανάλογων ενώσεων πραγματοποιήθηκαν *in vitro* δοκιμές φθορισμού κατά τις οποίες παρακολούθηθηκε η εξέλιξη του ρυθμού πολυμερισμού της ακτίνης. Στη συνέχεια τα ληφθέντα δεδομένα ρυθμού προσαρμόστηκαν σε κατάλληλη καμπύλη και υπολογίστηκαν οι τιμές  $IC_{50}$  με βάση την Εξίσωση 4.2 [64]:

$$rate = max - \frac{max - min}{1 + \frac{IC_{50}}{[inhibitor]}} \quad (4.2)$$

Όπου, *rate* ο ρυθμός πολυμερισμού [nM/s], *max* ο μέγιστος ρυθμός πολυμερισμού [nM/s], *min* ο ελάχιστος ρυθμός πολυμερισμού [nM/s],  $IC_{50}$  η μέση ανασταλτική συγκέντρωση [μM], και [inhibitor] η συγκέντρωση του αναστολέα [μM]. Από τις συγκεντρώσεις  $IC_{50}$  τόσο για την ένωση οδηγό CK-666, όσο και για τις ανάλογες ενώσεις, καθίσταται δυνατός ο υπολογισμός της σχετικής ελεύθερης ενέργειας πρόσδεσης ( $\Delta\Delta G$ ) μεταξύ των δύο ενώσεων με βάση την Εξίσωση 4.3.

$$\Delta\Delta G_{experimental} = -RT \ln \frac{IC_{50}^{CK-666}}{IC_{50}^{analog}} \quad (4.3)$$

Όπου, *R* η παγκόσμια σταθερά των αερίων (kcal/mol) και *T*= 310K η θερμοκρασία στην οποία πραγματοποιήθηκαν οι πειραματικές μετρήσεις και οι υπολογιστικές προσομοιώσεις. Τα πειραματικά αποτελέσματα  $IC_{50}$  και  $\Delta\Delta G$  παρατίθενται στον Πίνακα 4.8, ενώ στον Πίνακα 4.9 συνοψίζονται οι πειραματικές και υπολογιστικά εκτιμώμενες τιμές  $\Delta\Delta G$  για το πακέτο NAMD. Σημειώνεται πως για τις ανάλογες ενώσεις που βρέθηκε πως  $IC_{50} > 200 \mu M$  θεωρήθηκε πως  $IC_{50} = 250 \mu M$ , ώστε να επιτραπεί η σύγκριση της πειραματικής σχετικής ελεύθερης ενέργειας με τα υπολογιστικά δεδομένα. Αντίστοιχα, στα μόρια που χαρακτηρίστηκαν ως ανενεργά (inactive) η τιμή  $IC_{50}$  θεωρήθηκε ίση με 3000 μM. Η γραφική σύγκριση μεταξύ των πειραματικών δεδομένων και υπολογιστικά εκτιμώμενων σχετικών ελεύθερων ενεργειών πρόσδεσης απεικονίζεται στο Σχήμα 4.31.

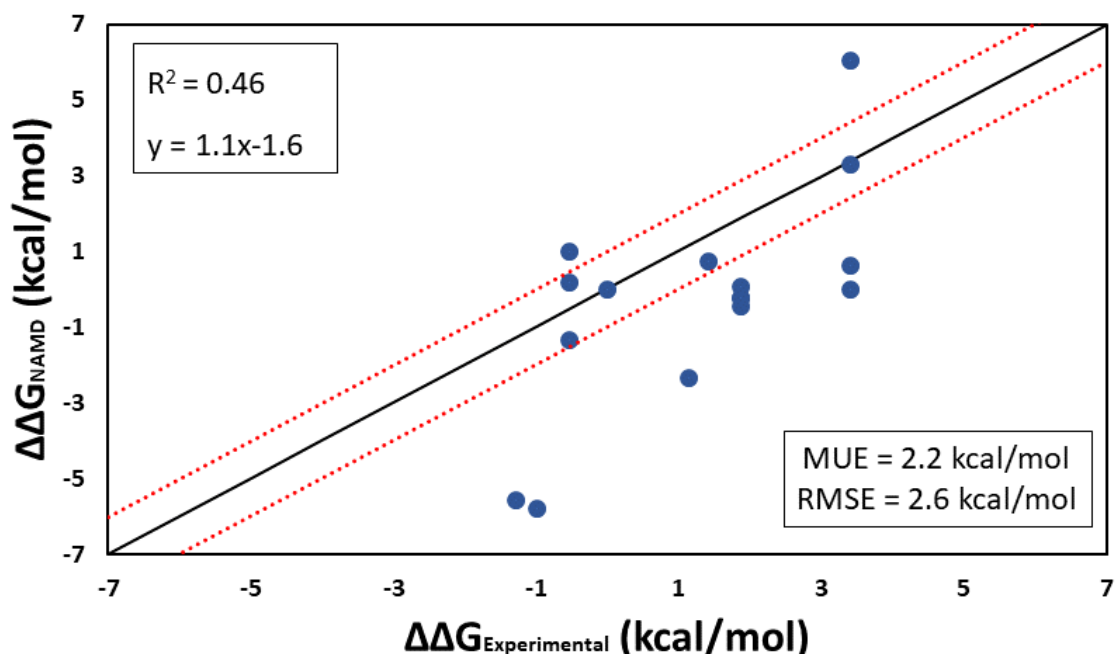
**Πίνακας 4.8** Αποτελέσματα πειραματικής αξιολόγησης των ανάλογων μορίων της ένωσης οδηγού CK-666.

| Ανάλογη Ένωση | IC <sub>50</sub> [μM] | ΔΔG <sub>exp</sub> [kcal/mol] |
|---------------|-----------------------|-------------------------------|
| CK-666        | 12                    | 0.00                          |
| ai003         | 1.5                   | -1.28                         |
| ai007         | 5                     | -0.54                         |
| ai015         | 4.92                  | -0.55                         |
| ai062         | 75                    | 1.13                          |
| ai064         | 120                   | 1.42                          |
| ai065         | >200                  | 1.87                          |
| ai066         | >200                  | 1.87                          |
| ai067         | >200                  | 1.87                          |
| ai068         | >200                  | 1.87                          |
| ai071         | inactive              | 3.40                          |
| ai078         | inactive              | 3.40                          |
| ai079         | inactive              | 3.40                          |
| ai086         | inactive              | 3.40                          |
| ai093         | 5                     | -0.54                         |
| ai094         | 2.4                   | -0.99                         |

**Πίνακας 4.9** Σύνοψη των πειραματικά και υπολογιστικά εκτιμώμενων τιμών σχετικής ελεύθερης ενέργειας πρόσδεσης για τις ανάλογες ενώσεις του CK-666. Οι πειραματικές τιμές λήφθηκαν με in vitro δοκιμές πολυμερισμού ακτίνης, ενώ οι υπολογιστικές τιμές εκτιμήθηκαν με βάση προσομοιώσεις Μοριακής Δυναμικής και τη μεθοδολογία της Διατάραξης Ελεύθερης Ενέργειας (FEP) με το υπολογιστικό πακέτο NAMD.

| Ανάλογη Ένωση | ΔΔG <sub>exp</sub> [kcal/mol] | ΔΔG <sub>NAMD</sub> [kcal/mol] |
|---------------|-------------------------------|--------------------------------|
| CK-666        | 0.00                          | 0.00 ± 0.00                    |
| ai003         | -1.28                         | -5.54 ± 0.05                   |
| ai007         | -0.54                         | 1.02 ± 0.1                     |
| ai015         | -0.55                         | -1.31 ± 0.04                   |
| ai062         | 1.13                          | -2.33 ± 0.17                   |
| ai064         | 1.42                          | 0.76 ± 0.02                    |
| ai065         | 1.87                          | -0.43 ± 0.06                   |
| ai066         | 1.87                          | -0.2 ± 0.12                    |
| ai067         | 1.87                          | 0.09 ± 0.01                    |
| ai068         | 1.87                          | -0.2 ± 0.05                    |
| ai071         | 3.40                          | 0.02 ± 0.06                    |
| ai078         | 3.40                          | 0.65 ± 0.07                    |
| ai079         | 3.40                          | 3.32 ± 0.09                    |
| ai086         | 3.40                          | 6.07 ± 0.37                    |
| ai093         | -0.54                         | 0.21 ± 0.04                    |
| ai094         | -0.99                         | -5.78 ± 0.05                   |

## NAMD vs experimental



**Σχήμα 4.31** Σύγκριση αποτελεσμάτων σχετικής ελεύθερης ενέργειας πρόσδεσης  $\Delta\Delta G$  μεταξύ πειραματικών αποτελεσμάτων και εκτιμήσεων που λήφθηκαν με το υπολογιστικό πακέτο NAMD. Με κόκκινη διακεκομμένη γραμμή συμβολίζεται το τυπικό σφάλμα της μεθόδου FEP το οποίο ανέρχεται στα  $\pm 1$  kcal/mol, ενώ με τη μαύρη έντονη γραμμή απεικονίζεται η ευθεία  $x=y$ .

Με βάση τα Σχήματα 4.16 έως 4.30 αντιλαμβάνεται κανείς πως οι εκτιμήσεις που παρέχει το υπολογιστικό πακέτο NAMD μπορούν να θεωρηθούν αξιόπιστες, καθώς υπάρχει ικανοποιητικός βαθμός επικάλυψης των forward και backward προσομοιώσεων για τις περισσότερες περιπτώσεις. Ενδεχομένως, θα μπορούσε κανείς να πει πως σε δύο μόλις περιπτώσεις (ai086 complex και ai064 complex) υπάρχει ασυμφωνία της τάξης των  $\sim 2$  kcal/mol μεταξύ των αμφίδρομων προσομοιώσεων. Παρ' όλα αυτά, στις περισσότερες περιπτώσεις υπάρχει σύγκλιση των δύο προσομοιώσεων, γεγονός που φανερώνει πως το πρωτόκολλο που ακολουθήθηκε για τις ανάλογες ενώσεις ήταν αρκετό για την εξαγωγή αξιόπιστων εκτιμήσεων σχετικής ελεύθερης ενέργειας.

Από το Σχήμα 4.31 γίνεται αντιληπτό πως μόλις 5/15 ανάλογες ενώσεις βρίσκονται εντός του τυπικού σφάλματος της μεθόδου FEP το οποίο ανέρχεται στα  $\pm 1$  kcal/mol (κόκκινες διακεκομμένες γραμμές), ενώ μόλις σε μία ένωση υπάρχει σχεδόν ταύτιση υπολογιστικών και πειραματικών αποτελεσμάτων, μιας και βρίσκεται επάνω στη διαγώνιο  $x=y$ . Επιπλέον, θα πρέπει να ληφθεί υπόψιν πως οι υπολογισμοί FEP δεν απαιτείται να παραδίδουν τέλεια αποτελέσματα. Σε πρώτο στάδιο θα πρέπει να υπάρχει η δυνατότητα πρόβλεψης του σωστού πρόσημου της σχετικής ελεύθερης ενέργειας πρόσδεσης, με κάποια αξιοπιστία, ούτως ώστε να μπορεί να πραγματοποιηθεί η σύνθεση και δοκιμή των ενώσεων που αναμένεται να προσδένονται ισχυρότερα στο φαρμακολογικό στόχο. Στο πλαίσιο αυτό, οι προσομοιώσεις με το πακέτο NAMD κατάφεραν να προβλέψουν σωστά το 60% των πρόσημων της σχετικής ελεύθερης ενέργειας πρόσδεσης διάφορων μορίων στον πρωτεϊνικό στόχο. Ωστόσο, η τιμή της ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error, RMSE), του μέσου απόλυτου σφάλματος (Mean Unsigned Error, MUE), η μέτρια γραμμική συσχέτιση των πειραματικών με τα υπολογιστικά αποτελέσματα ( $R^2 = 0.46$ ) και το γεγονός ότι σε γενικές γραμμές το NAMD έχει παραδώσει αξιόπιστα αποτελέσματα με βάση τα Σχήματα 4.16 έως 4.30, φανερώνει πως υπάρχουν συστηματικά σφάλματα στις προσομοιώσεις Μοριακής Δυναμικής. Τα σφάλματα αυτά είναι συνδυασμός τριών κύριων συνιστωσών οι οποίες αφορούν την παραμετροποίηση των μορίων,

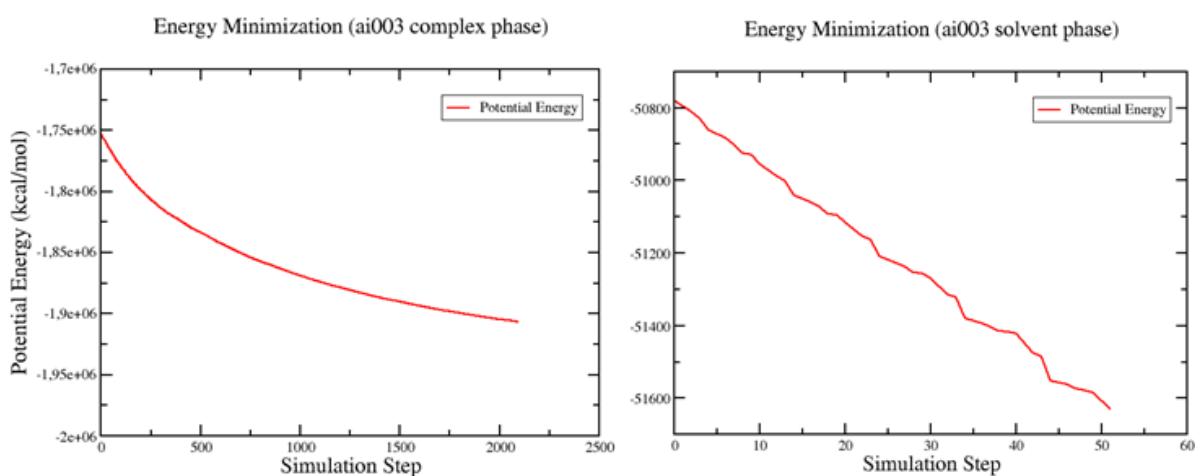


τους αλγορίθμους με βάση τους οποίους γίνεται η δειγματοληψία κατά τις προσομοιώσεις Μοριακής Δυναμικής, καθώς και στον τρόπο με τον οποίο μοντελοποιούνται οι ιδιότητες του διαλύτη.

#### 4.2.2 Εκτίμηση σχετικής ελεύθερης ενέργειας πρόσδεσης με GROMACS και σύγκριση με πειραματικά δεδομένα

Τέλος, πραγματοποιήθηκαν προσομοιώσεις Μοριακής Δυναμικής με το υπολογιστικό πακέτο GROMACS για τις ίδιες δεκαπέντε ανάλογες ενώσεις της ένωσης οδηγού CK-666. Η διαδικασία προετοιμασίας των μορίων για τις προσομοιώσεις με το υπολογιστικό πακέτο GROMACS περιγράφεται εκτενώς στο Κεφάλαιο 3.2.4.1. Για την αλχημική μετατροπή της ένωσης οδηγού CK-666 σε κάθε ένα από τα ανάλογα μόρια, χρησιμοποιήθηκαν συνολικά 21 ενδιάμεσες καταστάσεις. Κάθε ενδιάμεση λ-κατάσταση προσομοιώθηκε ξεχωριστά και περιλάμβανε 4 διακριτά στάδια: 1) την ελαχιστοποίηση της ενέργειας, 2) την NVT εξισορρόπηση, 3) την NPT εξισορρόπηση και 4) την παραγωγική φάση, κατά την οποία πραγματοποιείται ο υπολογισμός της ελεύθερης ενέργειας.

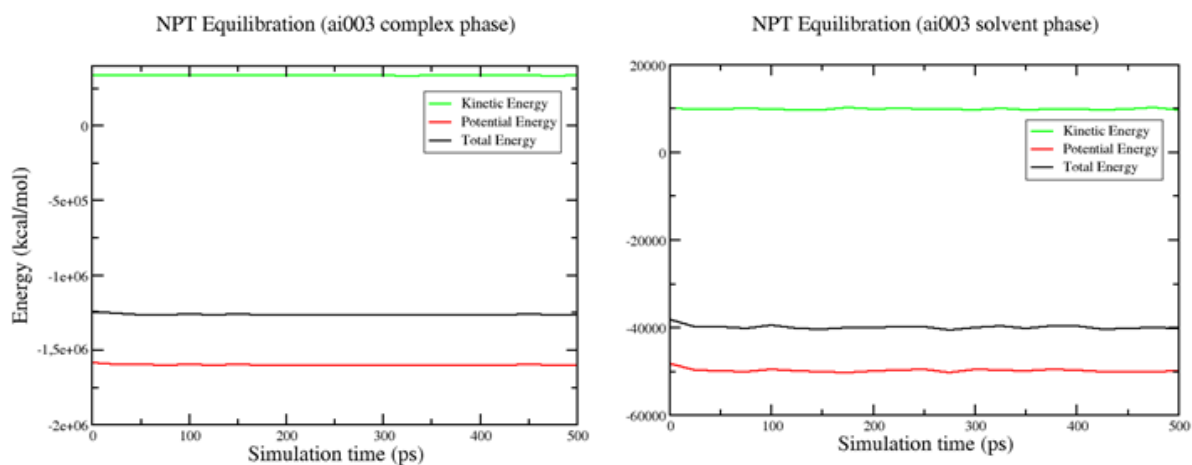
Όπως και στην περίπτωση του NAMD, σε κάθε δομή πραγματοποιήθηκε ελαχιστοποίηση της ενέργειας προκειμένου να εξαλειφθούν πιθανές στερεοχημικές παρεμποδίσεις. Στο Σχήμα 4.32 απεικονίζεται ένα ενδεικτικό γράφημα της δυναμικής ενέργειας συναρτήσει του βήματος προσομοίωσης για τα δύο σκέλη της προσομοίωσης της ανάλογης ένωσης ai003.



**Σχήμα 4.32** Διάγραμμα της δυναμικής ενέργειας του συστήματος συναρτήσει των βημάτων προσομοίωσης κατά την ελαχιστοποίηση της ενέργειας με το υπολογιστικό πακέτο GROMACS για την complex (αριστερά) και solvent (δεξιά) φάση της προσομοίωσης της ανάλογης ένωσης ai003.

Υπενθυμίζεται πως στο υπολογιστικό πακέτο GROMACS, χρησιμοποιείται ένας αλγόριθμος steepest descent ο οποίος συγκλίνει όταν η δύναμη σε κάποιο άτομο δεν ξεπερνά τα 1000 kJ/(mol·nm), και για το λόγο αυτό τα βήματα προσομοίωσης δεν είναι σταθερά όπως συμβαίνει στην περίπτωση της ελαχιστοποίησης ενέργειας με το NAMD. Στη συνέχεια ακολούθησαν δύο φάσεις εξισορρόπησης, αρχικά NVT και ύστερα NPT, κάθε μία από τις οποίες είχε διάρκεια προσομοίωσης 0.5 ns. Έπειτα από την ολοκλήρωση των σταδίων αυτών, μελετήθηκαν ιδιότητες των συστημάτων και παρατηρήθηκε ότι δεν υπάρχουν μεγάλες διακυμάνσεις στις τιμές τους. Συνεπώς θεωρείται η εξισορρόπηση ήταν επιτυχής και έτσι μπορούν να πραγματοποιηθούν οι παραγωγικές προσομοιώσεις για όλα τα συστήματα. Στο Σχήμα 4.33 παρουσιάζεται ενδεικτικά η γραφική παράσταση της συνολικής,

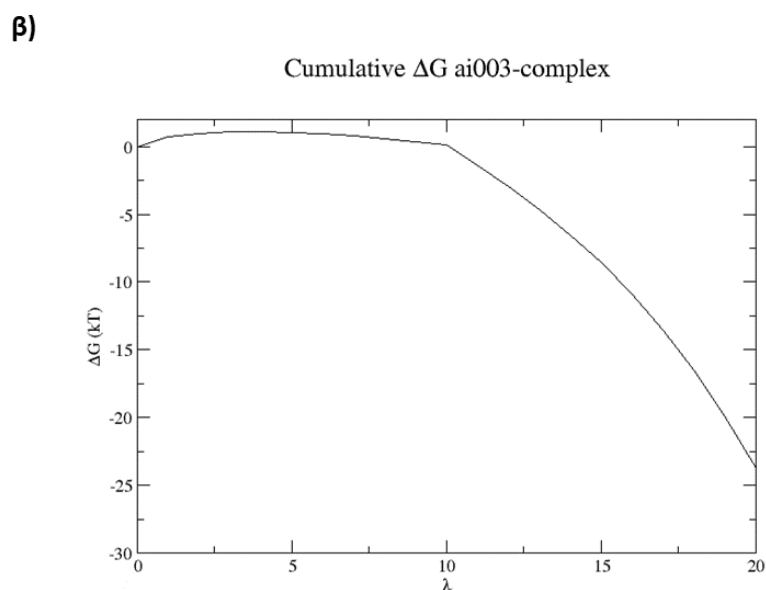
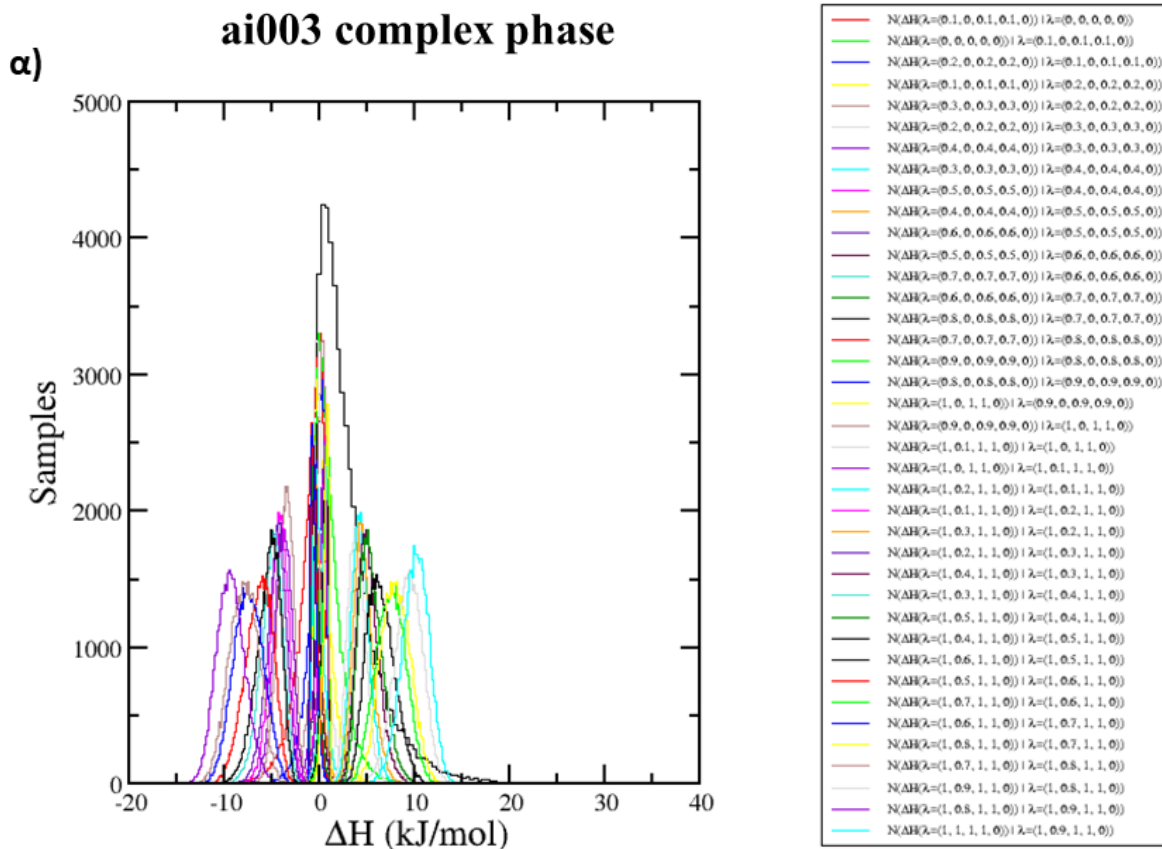
κινητικής και δυναμικής ενέργειας για την NPT εξισορρόπηση, συναρτήσει του χρόνου προσομοίωσης για τις δύο φάσεις της ανάλογης ένωσης ai003.



**Σχήμα 4.33** Διαγράμματα της κινητικής, δυναμικής και συνολικής ενέργειας του συστήματος συναρτήσει του χρόνου προσομοίωσης για την complex (επάνω) και solvent (κάτω) φάση της προσομοίωσης της ανάλογης ένωσης ai003.

Στη συνέχεια, ακολούθησαν οι παραγωγικές φάσεις των προσομοιώσεων οι οποίες αποσκοπούν στον υπολογισμό της σχετικής ελεύθερης ενέργειας πρόσδεσης μεταξύ της ένωσης οδηγού CK-666 και δεκαπέντε ανάλογων μορίων. Για την αλχημική μετατροπή της ένωσης οδηγού CK-666 σε κάθε ένα από τα ανάλογα μόρια, τόσο όταν ο πιθανός αναστολέας βρίσκεται σε σύμπλοκο με την πρωτεΐνη σε περιβάλλον διαλύτη (complex), όσο και όταν υπάρχει ελεύθερος στο περιβάλλον του διαλύτη (solvent), χρησιμοποιήθηκαν 21 διακριτές λ ενδιάμεσες καταστάσεις. Κάθε μία από τις παραγωγικές φάσεις προσομοιώνεται για 5 ns με χρονικό βήμα 1 fs. Για την ανάλυση των τροχιών που λήφθηκαν από τις προσομοιώσεις της Μοριακής Δυναμικής και την εξαγωγή της διαφοράς ελεύθερης ενέργειας για κάθε φάση, θα χρησιμοποιηθεί η μέθοδος BAR η οποία είναι ενσωματωμένη στο εργαλείο gmx.

Για τον έλεγχο και την παρακολούθηση της σύγκλισης της ελεύθερης ενέργειας, παρέχονται αποτελέσματα της σχετικής εντροπίας μεταξύ των δύο καταστάσεων A και B, η οποία ιδανικά θα πρέπει να τείνει στο μηδέν, οι γραφικές παραστάσεις των Χαμιλτονιανών διαφορών μεταξύ των «γειτονικών» λ ενδιάμεσων καταστάσεων, σε συνάρτηση με τον αριθμό δειγμάτων που λαμβάνεται από κάθε κατάσταση. Επιπλέον, παρέχεται και το γράφημα της εξέλιξης της ελεύθερης ενέργειας συναρτήσει του αριθμού των λ καταστάσεων. Για να θεωρηθεί αξιόπιστη η εκτίμηση  $\Delta G$ , θα πρέπει να έχει γίνει επαρκής δειγματοληψία μεταξύ γειτονικών ενδιάμεσων καταστάσεων. Αυτό το αντιλαμβάνεται κανείς όταν οι προσομοιώσεις οδηγούν σε αλληλεπικαλυπτόμενα ιστογράμματα Χαμιλτονιανών διαφορών, χαμηλές εκτιμήσεις σφάλματος από τη μέθοδο BAR, καθώς και τιμές σχετικής εντροπίας κοντά στο μηδέν. Στο Σχήμα 4.34 και Σχήμα 4.35 απεικονίζονται τα γραφήματα των κατανομών των τιμών  $\Delta H$  μεταξύ των γειτονικών λ καταστάσεων, ενώ στους Πίνακες 4.10 και 4.11 παρατίθενται τα αποτελέσματα σχετικής εντροπίας για την περίπτωση της ανάλογης ένωσης ai003 και για τις δύο φάσεις της προσομοίωσης.

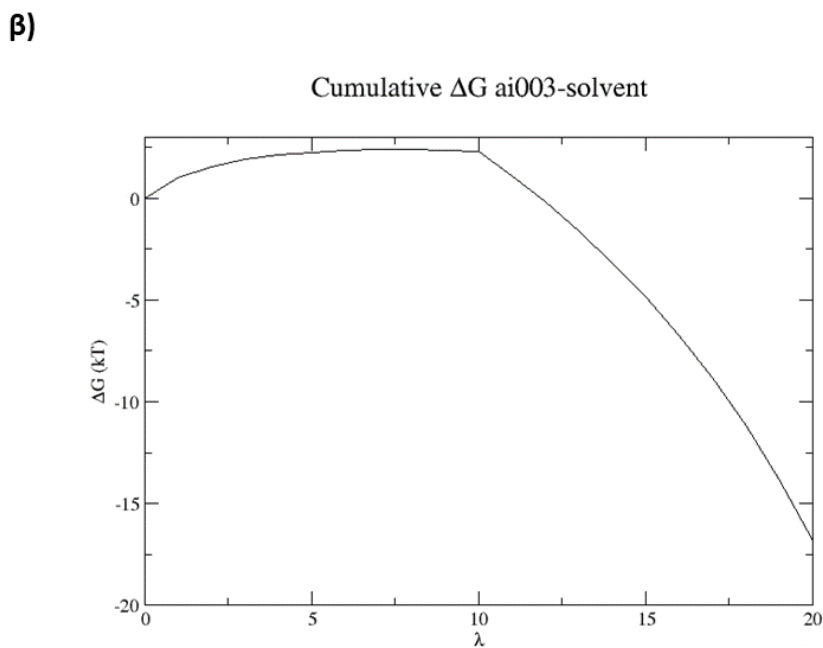
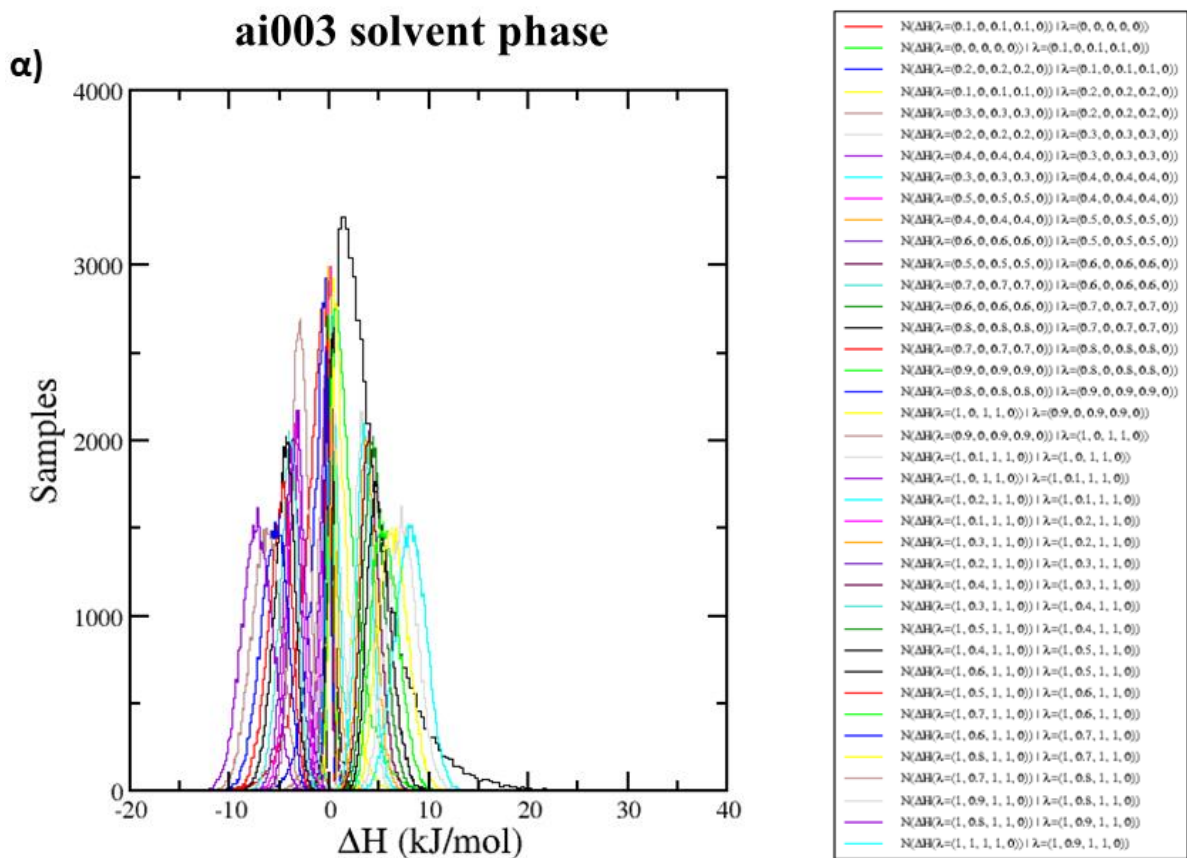


**Σχήμα 4.34** Τυπικά διαγράμματα παρακολούθησης της εξέλιξης της ελεύθερης ενέργειας από το υπολογιστικό πακέτο GROMACS. **(α)** Κατανομές των τιμών  $\Delta H$  (όπου  $H$ , η Χαμιλτονιανή του συστήματος) μεταξύ των γειτονικών  $\lambda$  καταστάσεων. Ο όρος “ $N(\Delta H(\lambda=(0.1, 0, 0.1, 0.1, 0)) | \lambda=(0, 0, 0, 0, 0))$ ” υποδεικνύει την κατανομή των τιμών  $\Delta H$  που έχουν υπολογισθεί στην τιμή του  $\lambda = 0.1$  από τις διαμορφώσεις στην τροχιά για  $\lambda = 0$ . Υπενθυμίζεται πως στην αρχή της προσομοίωσης διαταράσσονται οι μάζες, οι όροι van der Waals και οι διαμοριακές αλληλεπιδράσεις, για το λόγο αυτό υπάρχουν 3 όροι για  $\lambda = 0.1$  εντός της κατανομής. Με την πάροδο της προσομοίωσης οι παράμετροι αυτοί έχουν μετατραπεί πλήρως στις αντιπροσωπευτικές της κατάστασης στόχου και ακολουθεί η διαταραχή των ηλεκτροστατικών αλληλεπιδράσεων. Αξίζει να σημειωθεί πως υπάρχει και ένας όρος που δεν διαταράσσεται ( $\lambda = 0$ , σε κάθε περίπτωση) και αφορά αλληλεπιδράσεις περιορισμού (restraint interactions) που επιβάλλονται στον προσδέτη για να μην εκτραπεί από τη θέση

πρόσδεσης. **(β)** Παρακολούθηση της μεταβολής της ελεύθερης ενέργειας συναρτήσει του αριθμού των ενδιάμεσων λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Ενδεικτικά αναφέρεται πως για T = 310 K  $1kT = 0.615$  kcal/mol. Τα αποτελέσματα λήφθηκαν για την παραγωγική φάση του complex σκέλους της ανάλογης ένωσης αι003.

**Πίνακας 4.10** Αποτελέσματα σχετικής εντροπίας μεταξύ των δύο καταστάσεων A και B για την complex φάση της ανάλογης ένωσης αι003. Με  $S_A$  συμβολίζεται η σχετική εντροπία των δειγμάτων του  $\lambda_B$  στο σύνολο του  $\lambda_A$ , ενώ με  $S_B$  το αντίστροφο.

| $\lambda_A$ | $\lambda_B$ | $S_A$ (kcal/mol) | $S_B$ (kcal/mol) |
|-------------|-------------|------------------|------------------|
| 0           | 1           | $0.28 \pm 0.01$  | $0.17 \pm 0.00$  |
| 1           | 2           | $0.09 \pm 0.01$  | $0.07 \pm 0.01$  |
| 2           | 3           | $0.04 \pm 0.00$  | $0.03 \pm 0.00$  |
| 3           | 4           | $0.03 \pm 0.00$  | $0.02 \pm 0.00$  |
| 4           | 5           | $0.02 \pm 0.00$  | $0.02 \pm 0.00$  |
| 5           | 6           | $0.01 \pm 0.00$  | $0.01 \pm 0.00$  |
| 6           | 7           | $0.01 \pm 0.00$  | $0.01 \pm 0.00$  |
| 7           | 8           | $0.01 \pm 0.00$  | $0.01 \pm 0.00$  |
| 8           | 9           | $0.01 \pm 0.00$  | $0.01 \pm 0.00$  |
| 9           | 10          | $0.01 \pm 0.00$  | $0.01 \pm 0.00$  |
| 10          | 11          | $0.03 \pm 0.01$  | $0.04 \pm 0.01$  |
| 11          | 12          | $0.04 \pm 0.01$  | $0.04 \pm 0.01$  |
| 12          | 13          | $0.03 \pm 0.00$  | $0.04 \pm 0.00$  |
| 13          | 14          | $0.06 \pm 0.01$  | $0.06 \pm 0.01$  |
| 14          | 15          | $0.04 \pm 0.02$  | $0.04 \pm 0.02$  |
| 15          | 16          | $0.12 \pm 0.02$  | $0.12 \pm 0.02$  |
| 16          | 17          | $0.13 \pm 0.00$  | $0.13 \pm 0.00$  |
| 17          | 18          | $0.06 \pm 0.02$  | $0.06 \pm 0.02$  |
| 18          | 19          | $0.16 \pm 0.02$  | $0.16 \pm 0.02$  |
| 19          | 20          | $0.09 \pm 0.01$  | $0.09 \pm 0.00$  |

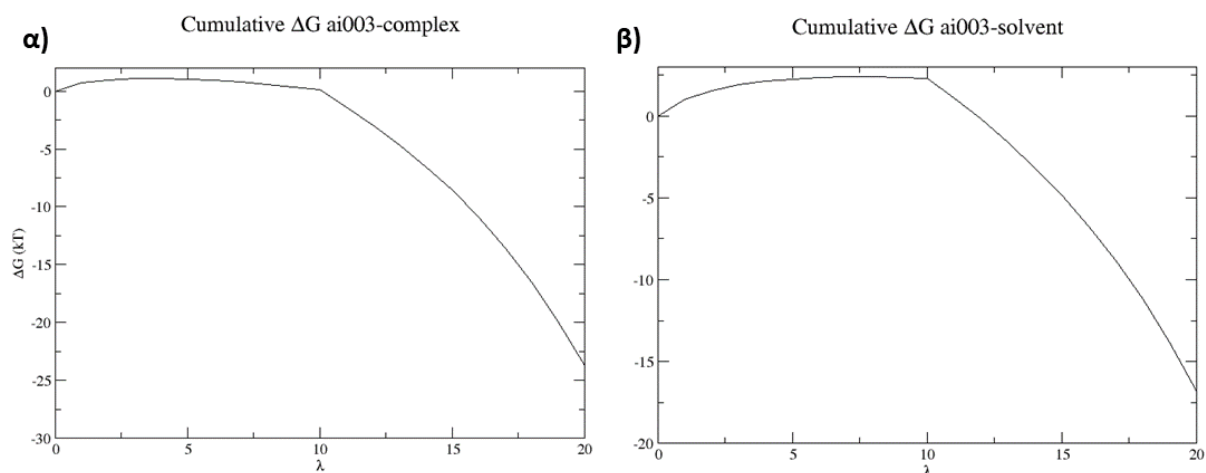


**Σχήμα 4.35** Τυπικά διαγράμματα παρακολούθησης της εξέλιξης της ελεύθερης ενέργειας από το υπολογιστικό πακέτο GROMACS. **(α)** Κατανομές των τιμών  $\Delta H$  (όπου  $H$ , η Χαμιλτονιανή του συστήματος) μεταξύ των γειτονικών  $\lambda$  καταστάσεων. Ο όρος " $N(\Delta H(\lambda=(0.1, 0, 0.1, 0.1, 0)) | \lambda=(0, 0, 0, 0))$ " υποδεικνύει την κατανομή των τιμών  $\Delta H$  που έχουν υπολογισθεί στην τιμή του  $\lambda = 0.1$  από τις διαμορφώσεις στην τροχιά για  $\lambda = 0$ . **(β)** Παρακολούθηση της μεταβολής της ελεύθερης ενέργειας συναρτήσει του αριθμού των ενδιάμεσων  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Για  $T = 310$  K η τιμή kT αντιστοιχεί σε 0.615 kcal/mol. Τα αποτελέσματα λήφθηκαν για την παραγωγική φάση του solvent σέκλους της ανάλογης ένωσης ai003.

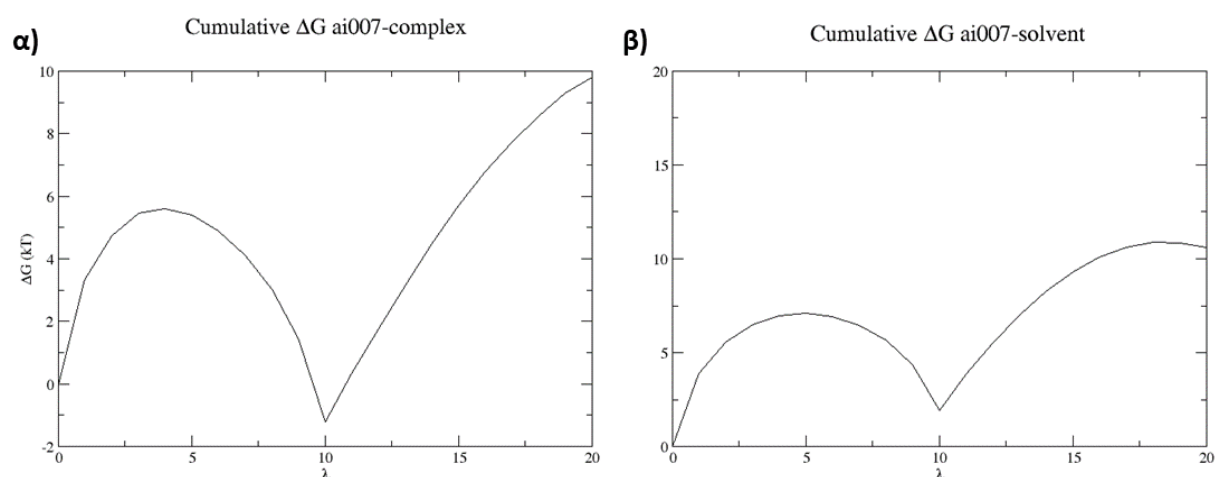
**Πίνακας 4.11** Αποτελέσματα σχετικής εντροπίας μεταξύ των δύο καταστάσεων A και B για την solvent φάση της ανάλογης ένωσης αι003. Με  $S_A$  συμβολίζεται η σχετική εντροπία των δειγμάτων του  $\lambda_B$  στο σύνολο του  $\lambda_A$ , ενώ με  $S_B$  το αντίστροφο.

| $\lambda_A$ | $\lambda_B$ | $S_A$ (kcal/mol) | $S_B$ (kcal/mol) |
|-------------|-------------|------------------|------------------|
| 0           | 1           | 0.52 ± 0.01      | 0.32 ± 0.00      |
| 1           | 2           | 0.16 ± 0.00      | 0.12 ± 0.00      |
| 2           | 3           | 0.08 ± 0.00      | 0.07 ± 0.00      |
| 3           | 4           | 0.05 ± 0.00      | 0.04 ± 0.00      |
| 4           | 5           | 0.04 ± 0.00      | 0.03 ± 0.00      |
| 5           | 6           | 0.03 ± 0.00      | 0.02 ± 0.00      |
| 6           | 7           | 0.02 ± 0.00      | 0.02 ± 0.00      |
| 7           | 8           | 0.02 ± 0.00      | 0.02 ± 0.00      |
| 8           | 9           | 0.01 ± 0.00      | 0.01 ± 0.00      |
| 9           | 10          | 0.01 ± 0.00      | 0.01 ± 0.00      |
| 10          | 11          | 0.06 ± 0.01      | 0.06 ± 0.01      |
| 11          | 12          | 0.06 ± 0.01      | 0.06 ± 0.01      |
| 12          | 13          | 0.06 ± 0.01      | 0.06 ± 0.01      |
| 13          | 14          | 0.06 ± 0.00      | 0.06 ± 0.00      |
| 14          | 15          | 0.08 ± 0.00      | 0.09 ± 0.00      |
| 15          | 16          | 0.09 ± 0.00      | 0.10 ± 0.00      |
| 16          | 17          | 0.11 ± 0.00      | 0.12 ± 0.00      |
| 17          | 18          | 0.14 ± 0.00      | 0.15 ± 0.00      |
| 18          | 19          | 0.17 ± 0.01      | 0.17 ± 0.01      |
| 19          | 20          | 0.16 ± 0.01      | 0.17 ± 0.01      |

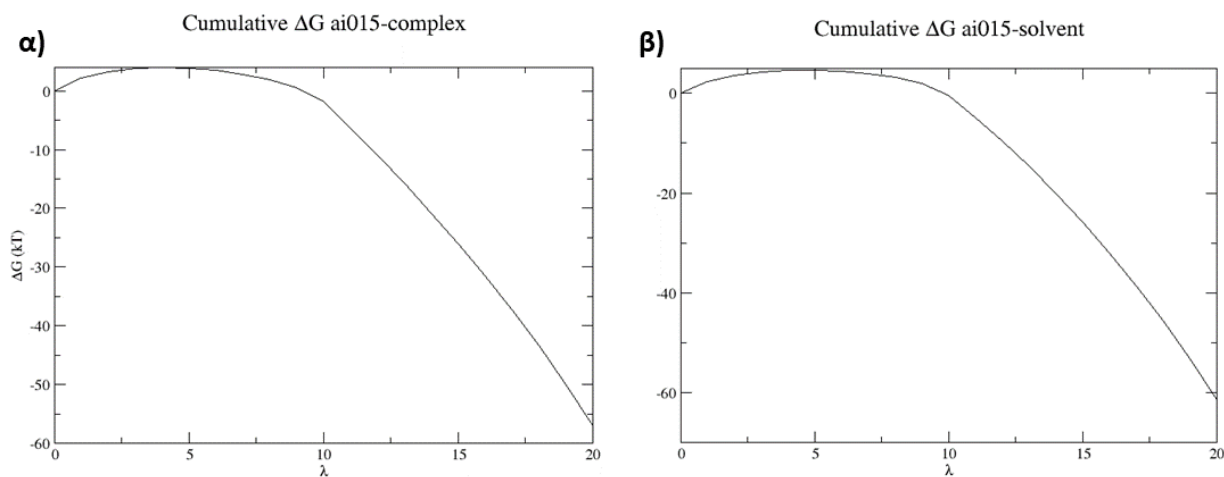
Αξίζει να αναφερθεί πως για όλα τα συστήματα ανάλογων ενώσεων του CK-666, ο βαθμός επικάλυψης των κατανομών  $\Delta H$  και της σχετικής εντροπίας παρουσιάζουν παρόμοια συμπεριφορά με αυτή της ανάλογης ένωσης αι003, με το εύρος της  $\Delta H$  να κυμαίνεται από -10 έως 20 kJ/mol. Έτσι, κρίνεται σκόπιμη μόνο η παρουσίαση των γραφημάτων που απεικονίζουν τη συνολική μεταβολή της ελεύθερης ενέργειας συναρτήσει του αριθμού των ενδιάμεσων  $\lambda$  καταστάσεων. Στα Σχήματα 4.36 έως 4.50 απεικονίζεται η εξέλιξη της συνολικής μεταβολής της ελεύθερης ενέργειας για κάθε φάση και ανάλογη ένωση που προσομοιώθηκε.



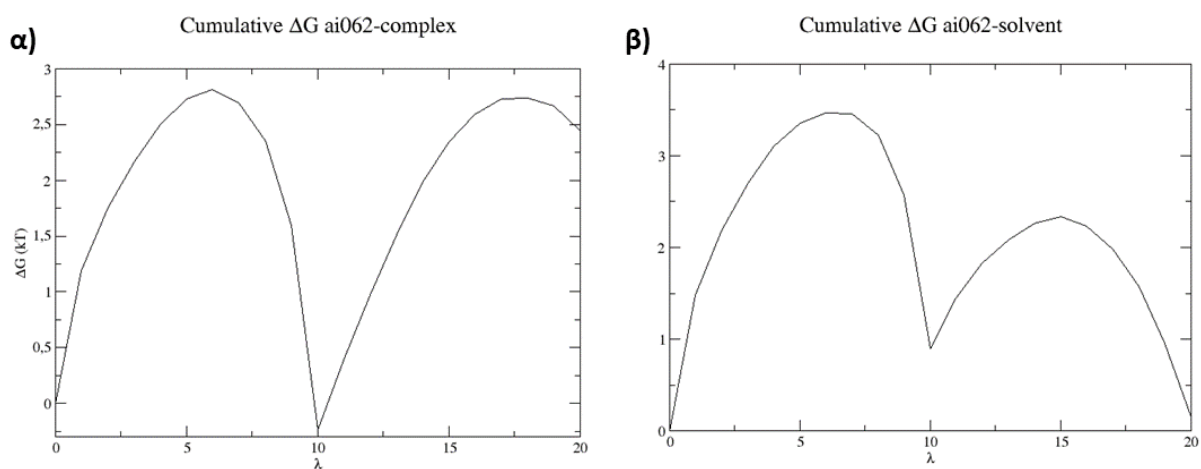
**Σχήμα 4.36** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai003 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $-14.60 \pm 0.06$  kcal/mol και  $-10.37 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



**Σχήμα 4.37** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai007 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $6.04 \pm 0.09$  kcal/mol και  $6.52 \pm 0.08$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

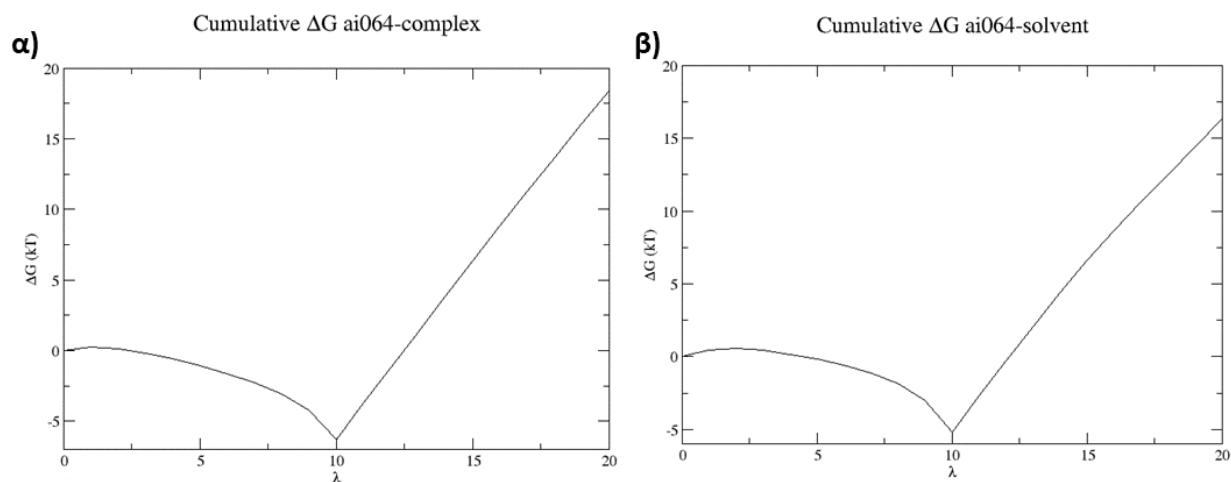


**Σχήμα 4.38** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai015 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες  $kT$  αντιπροσωπεύουν ενέργεια ανά mol, με το  $k$  να είναι η παγκόσμια σταθερά των αερίων και  $T$  η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $-35.07 \pm 0.05$  kcal/mol και  $-37.77 \pm 0.10$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

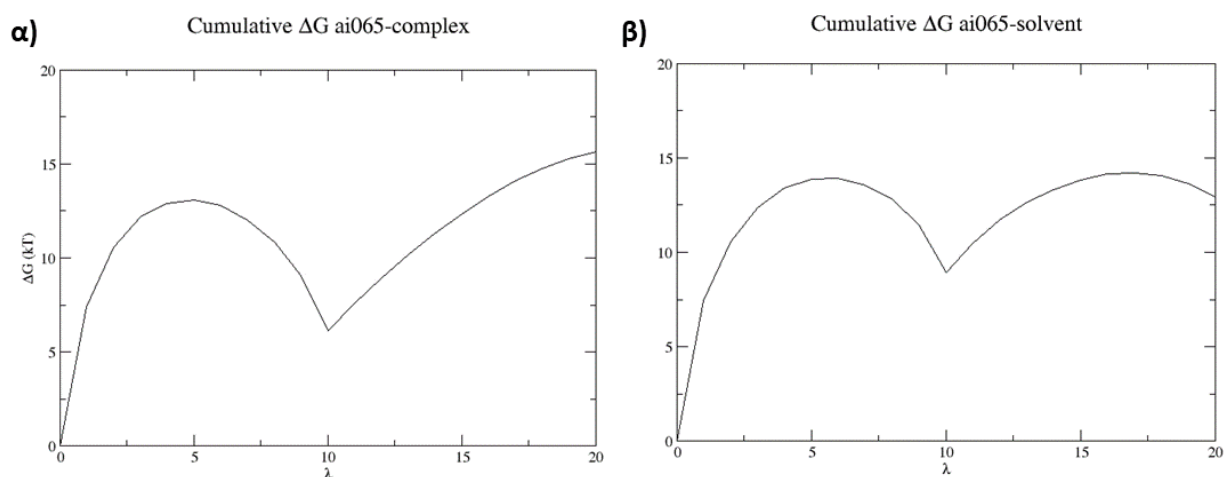


**Σχήμα 4.39** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai062 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες  $kT$  αντιπροσωπεύουν ενέργεια ανά mol, με το  $k$  να είναι η παγκόσμια σταθερά των αερίων και  $T$  η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $1.50 \pm 0.03$  kcal/mol και  $0.09 \pm 0.08$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

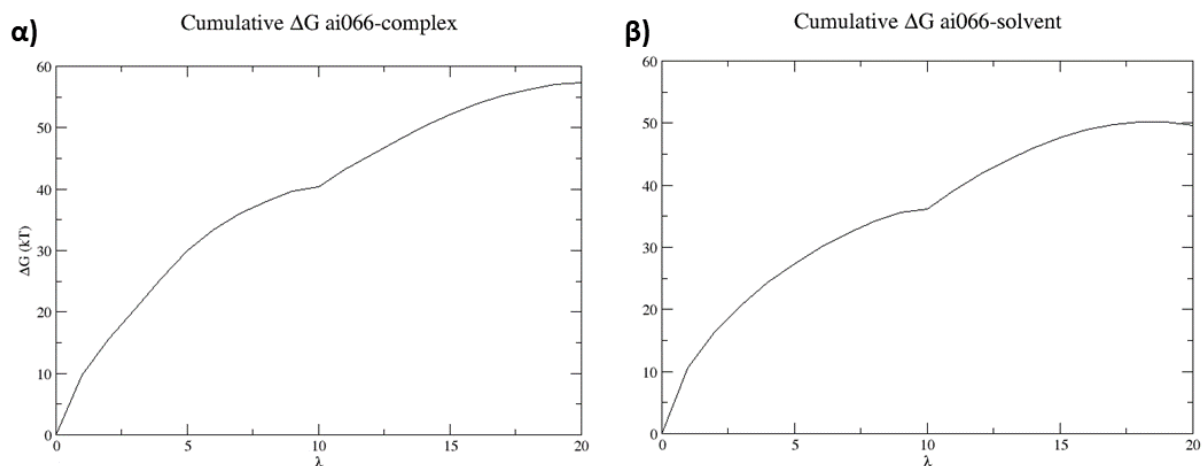




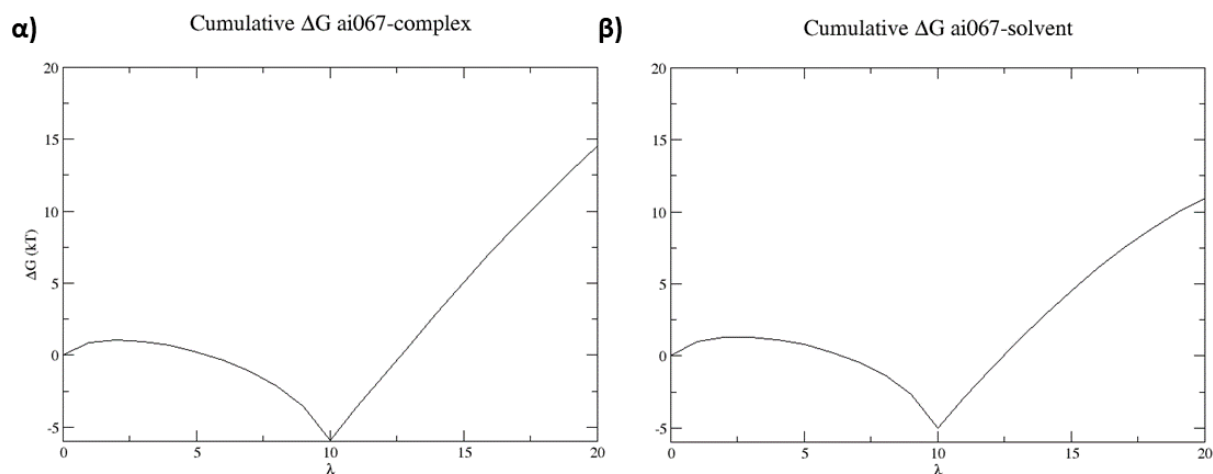
**Σχήμα 4.40** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai064 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις **(α)** complex και **(β)** solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $11.34 \pm 0.07$  kcal/mol και  $10.06 \pm 0.01$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



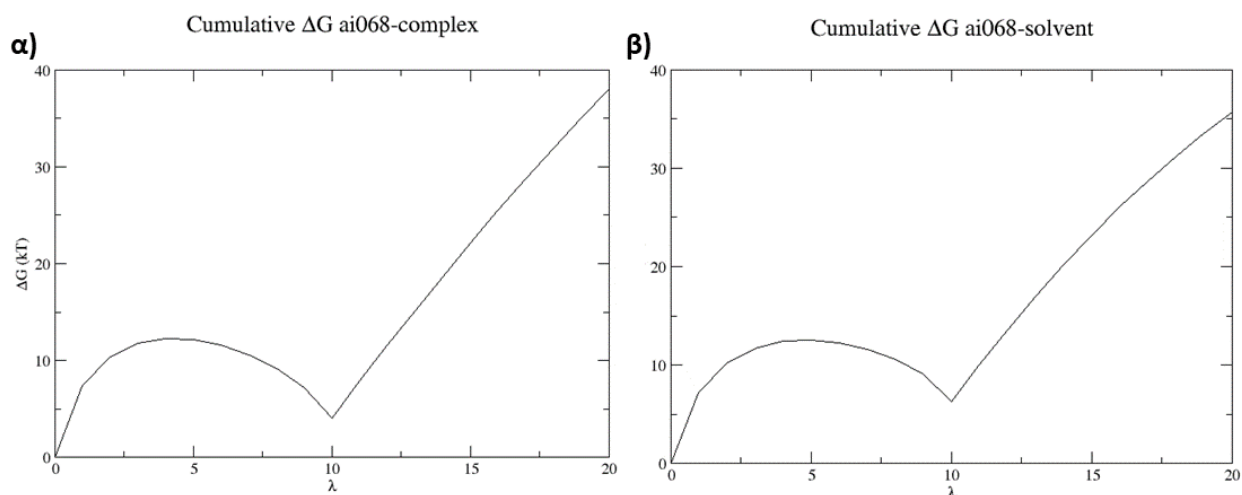
**Σχήμα 4.41** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai065 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις **(α)** complex και **(β)** solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $9.65 \pm 0.09$  kcal/mol και  $7.98 \pm 0.04$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



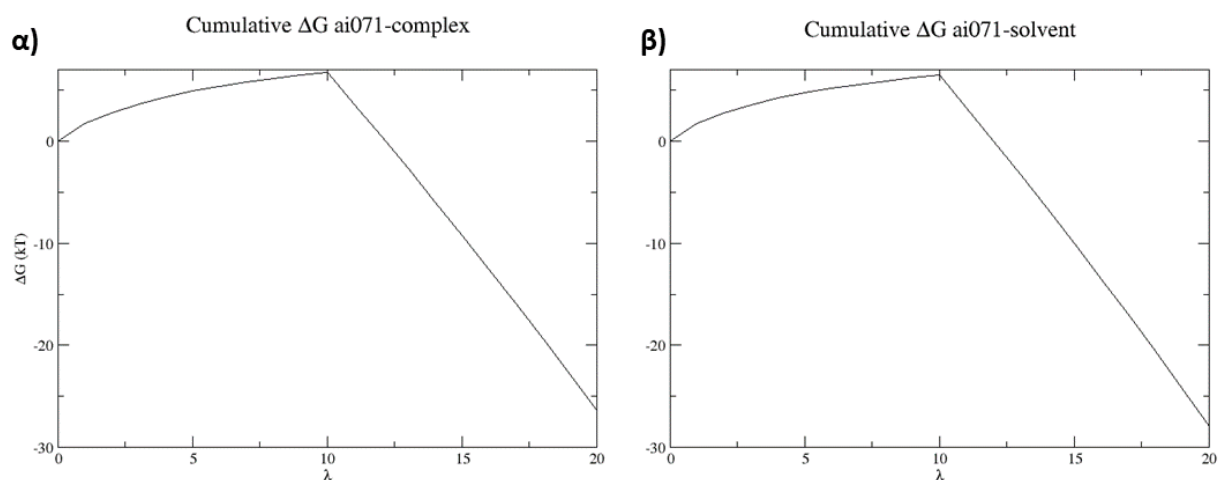
**Σχήμα 4.42** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai066 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις **(α)** complex και **(β)** solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $35.32 \pm 0.25$  kcal/mol και  $30.54 \pm 0.1$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



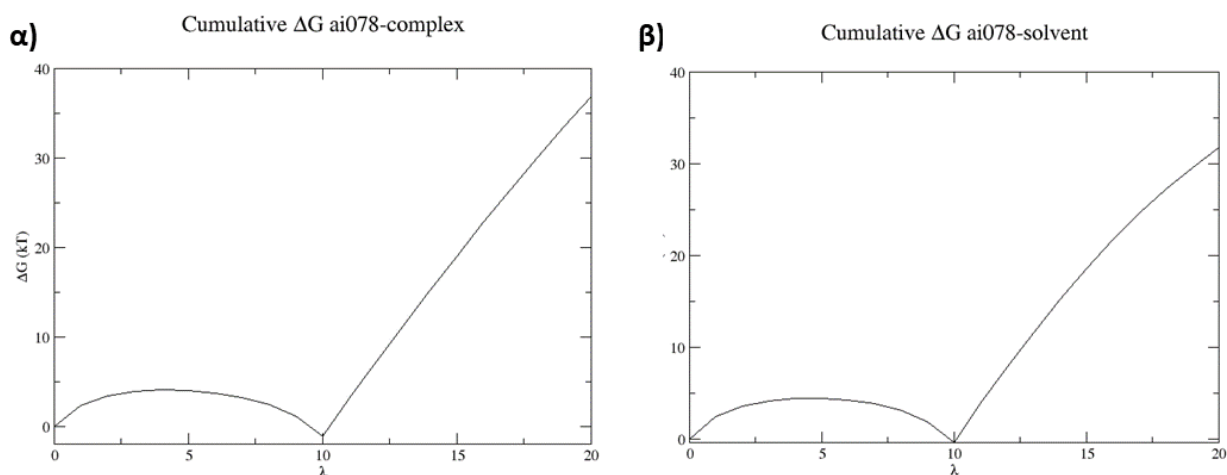
**Σχήμα 4.43** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai067 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις **(α)** complex και **(β)** solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $8.94 \pm 0.04$  kcal/mol και  $6.76 \pm 0.08$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



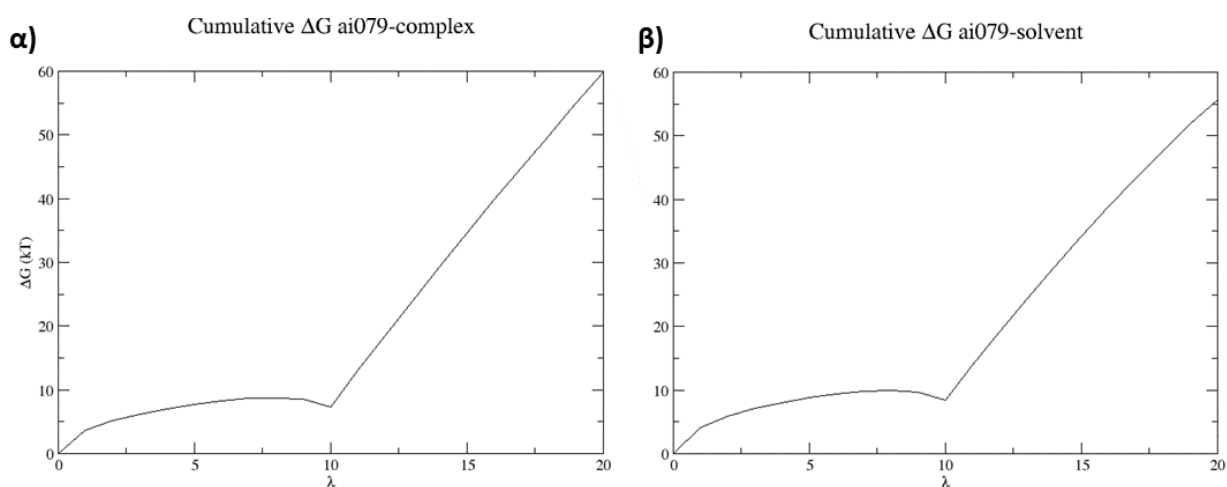
**Σχήμα 4.44** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai068 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $23.42 \pm 0.04$  kcal/mol και  $21.99 \pm 0.15$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



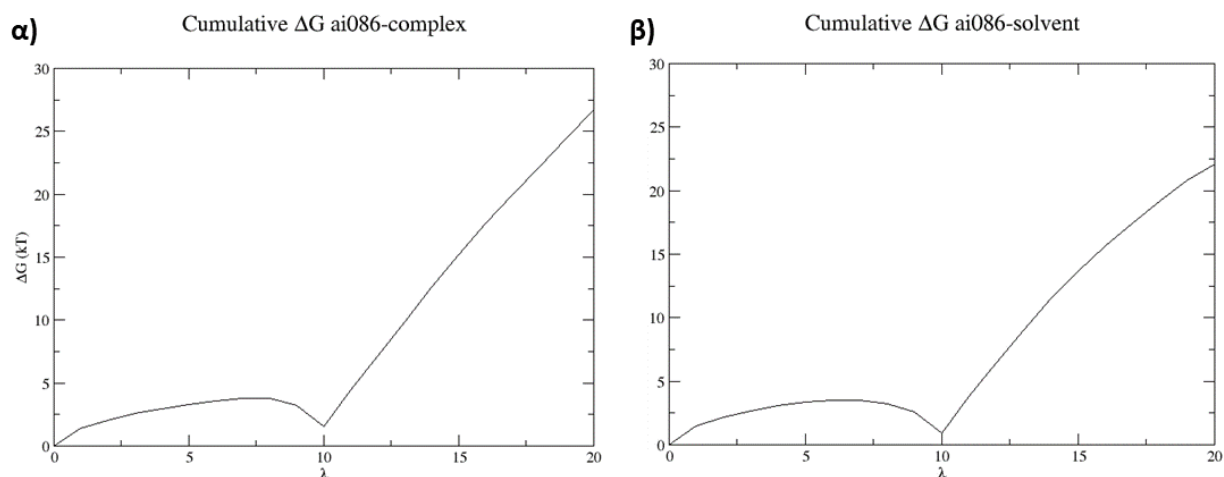
**Σχήμα 4.45** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai071 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των  $\lambda$  καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $-16.23 \pm 0.04$  kcal/mol και  $-17.19 \pm 0.02$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



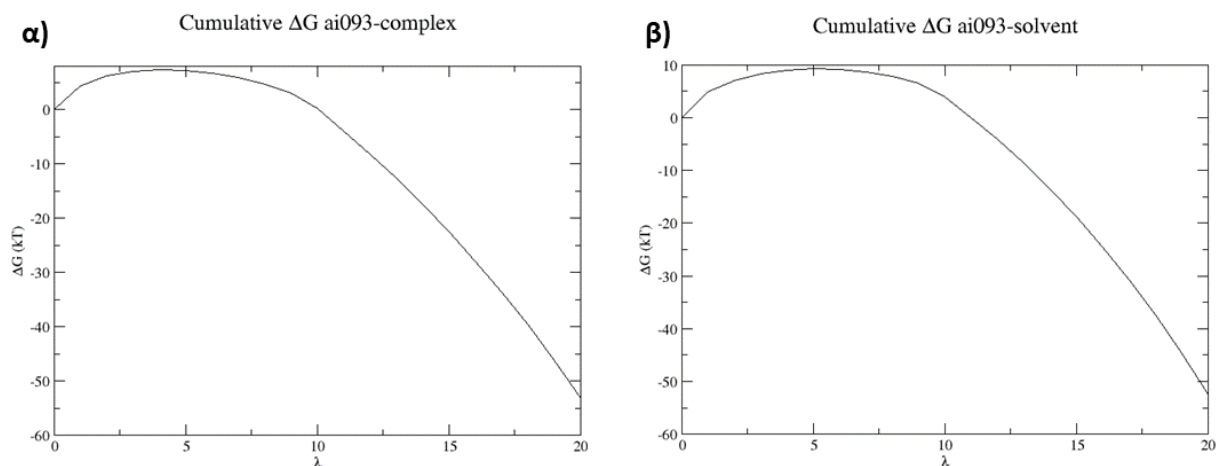
**Σχήμα 4.46** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai078 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $22.69 \pm 0.08$  kcal/mol και  $19.56 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



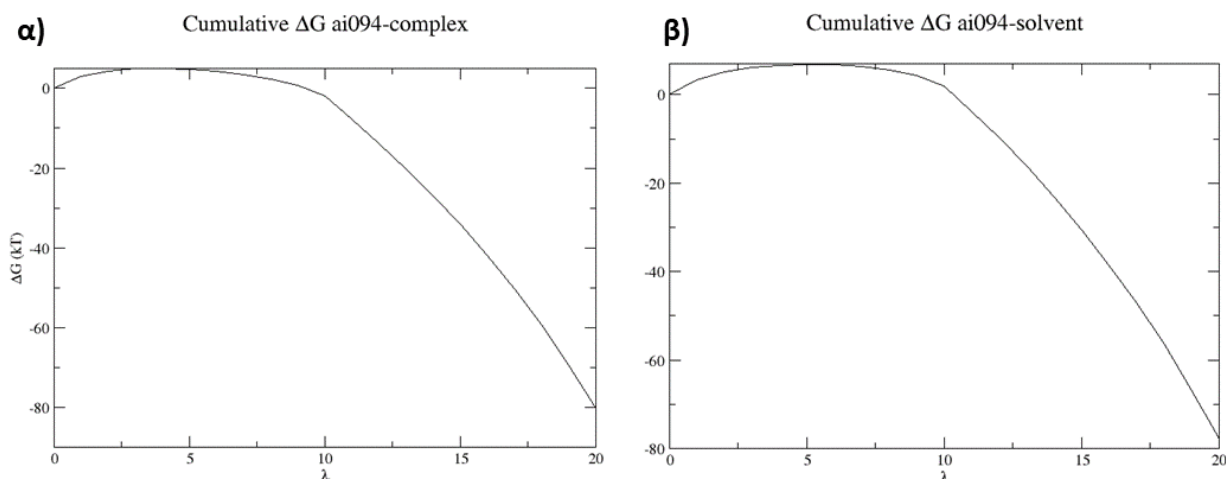
**Σχήμα 4.47** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai079 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $36.83 \pm 0.05$  kcal/mol και  $34.31 \pm 0.05$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



**Σχήμα 4.48** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai086 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $16.44 \pm 0.21$  kcal/mol και  $13.60 \pm 0.12$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



**Σχήμα 4.49** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai093 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $-32.71 \pm 0.15$  kcal/mol και  $-32.31 \pm 0.07$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.



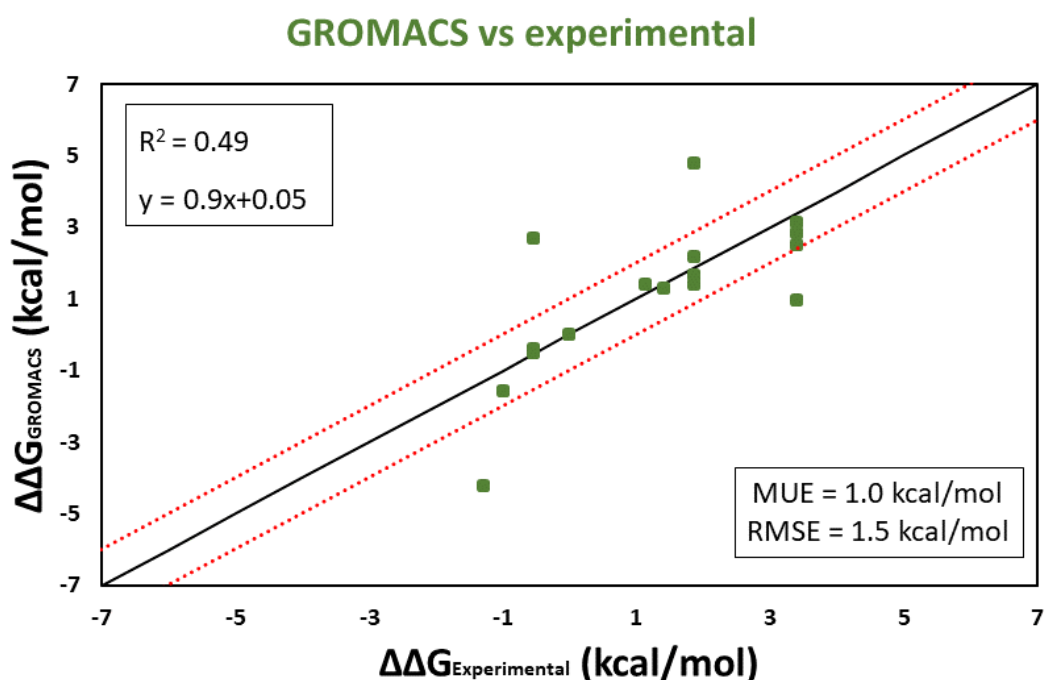
**Σχήμα 4.50** Γραφική αναπαράσταση της εξέλιξης της ελεύθερης ενέργειας για την ανάλογη ένωση ai094 με χρήση 21 ενδιάμεσων καταστάσεων και χρόνο προσομοίωσης 5 ns/κατάσταση. Η μεταβολή της ελεύθερης ενέργειας για τις (α) complex και (β) solvent προσομοιώσεις ως συνάρτηση του αριθμού των λ καταστάσεων. Οι μονάδες kT αντιπροσωπεύουν ενέργεια ανά mol, με το k να είναι η παγκόσμια σταθερά των αερίων και T η απόλυτη θερμοκρασία. Έτσι για  $T = 310$  K, προκύπτει πως η μεταβολή της ελεύθερης ενέργειας υπολογίζεται ίση με  $-49.46 \pm 0.07$  kcal/mol και  $-47.89 \pm 0.07$  kcal/mol για τις complex και solvent φάσεις, αντίστοιχα.

Στον Πίνακα 4.12 συνοψίζονται οι υπολογισμένες ελεύθερες ενέργειες για την προσομοίωση με το υπολογιστικό πακέτο GROMACS για την κάθε φάση και ανάλογη ένωση. Επιπλέον παρατίθενται τα αποτελέσματα για την σχετική ελεύθερη ενέργεια πρόσδεσης  $\Delta\Delta G$ , η οποία έχει υπολογισθεί με βάση την Εξίσωση 3.2, ενώ το αντίστοιχο σφάλμα εκτιμήθηκε με βάση την Εξίσωση 4.1. Τα πειραματικά δεδομένα για τα ανάλογα μόρια της ένωσης οδηγού συνοψίζονται στον Πίνακα 4.8. ενώ στον Πίνακα 4.13 συνοψίζονται οι πειραματικές και υπολογιστικά εκτιμώμενες τιμές  $\Delta\Delta G$  για το πακέτο GROMACS. Τέλος, γίνεται γραφική σύγκριση μεταξύ των πειραματικών αποτελεσμάτων και υπολογιστικά εκτιμώμενων σχετικών ελεύθερων ενεργειών πρόσδεσης στο Σχήμα 4.51.

**Πίνακας 4.12** Αποτελέσματα σχετικής ελεύθερης ενέργειας για την κάθε ανάλογη ένωση με το υπολογιστικό πακέτο GROMACS.

| Ανάλογη Ένωση | $\Delta G_{\text{complex}}$ [kcal/mol] | $\Delta G_{\text{solvent}}$ [kcal/mol] | $\Delta\Delta G_{\text{GROMACS}}$ [kcal/mol] |
|---------------|--|--|--|
| CK-666        | $0.00 \pm 0.00$                        | $0.00 \pm 0.00$                        | $0.00 \pm 0.00$                              |
| ai003         | $-14.6 \pm 0.06$                       | $-10.37 \pm 0.01$                      | $-4.22 \pm 0.06$                             |
| ai007         | $6.04 \pm 0.09$                        | $6.52 \pm 0.08$                        | $-0.48 \pm 0.12$                             |
| ai015         | $-35.07 \pm 0.05$                      | $-37.77 \pm 0.1$                       | $2.7 \pm 0.11$                               |
| ai062         | $1.5 \pm 0.03$                         | $0.09 \pm 0.08$                        | $1.41 \pm 0.09$                              |
| ai064         | $11.34 \pm 0.07$                       | $10.06 \pm 0.01$                       | $1.29 \pm 0.07$                              |
| ai065         | $9.65 \pm 0.09$                        | $7.98 \pm 0.04$                        | $1.67 \pm 0.1$                               |
| ai066         | $35.32 \pm 0.25$                       | $30.54 \pm 0.11$                       | $4.79 \pm 0.28$                              |
| ai067         | $8.94 \pm 0.04$                        | $6.76 \pm 0.08$                        | $2.18 \pm 0.09$                              |

|       |               |               |             |
|-------|---------------|---------------|-------------|
| ai068 | 23.42 ± 0.04  | 21.99 ± 0.15  | 1.43 ± 0.15 |
| ai071 | -16.23 ± 0.04 | -17.19 ± 0.02 | 0.96 ± 0.04 |
| ai078 | 22.69 ± 0.08  | 19.56 ± 0.05  | 3.13 ± 0.09 |
| ai079 | 36.83 ± 0.05  | 34.31 ± 0.05  | 2.52 ± 0.07 |
| ai086 | 16.44 ± 0.21  | 13.6 ± 0.12   | 2.84 ± 0.24 |
| ai093 | -32.71 ± 0.15 | -32.31 ± 0.07 | -0.4 ± 0.16 |
| ai094 | -49.46 ± 0.07 | -47.89 ± 0.07 | -1.57 ± 0.1 |



**Σχήμα 4.51** Σύγκριση αποτελεσμάτων σχετικής ελεύθερης ενέργειας πρόσδεσης  $\Delta\Delta G$  μεταξύ πειραματικών αποτελεσμάτων και εκτιμήσεων που λήφθηκαν με το υπολογιστικό πακέτο GROMACS. Με κόκκινη διακεκομμένη γραμμή συμβολίζεται το τυπικό σφάλμα της μεθόδου FEP το οποίο ανέρχεται στα  $\pm 1$  kcal/mol, ενώ με τη μαύρη έντονη γραμμή απεικονίζεται η ευθεία  $x=y$ .

Από το παραπάνω διάγραμμα γίνεται αντιληπτό πως το υπολογιστικό πακέτο GROMACS σε συνδυασμό με το πεδίο δυνάμεων GAFF και το μοντέλο AM1-BCC για τα μερικά φορτία των ατόμων, παρείχαν ακριβή αποτελέσματα σχετικής ελεύθερης ενέργειας πρόσδεσης για τις ανάλογες ενώσεις του μορίου CK-666. Οι υπολογισμοί FEP με το πακέτο GROMACS συμφωνούν με τα πειραματικά αποτελέσματα, μιας και 11/15 ενώσεις βρίσκονται εντός της περιοχής  $\pm 1$  kcal/mol (κόκκινες διακεκομμένες γραμμές), η οποία αποτελεί το τυπικό σφάλμα των υπολογισμών ελεύθερης ενέργειας. Οι προσομοιώσεις με το πακέτο GROMACS οδήγησαν στη σωστή πρόβλεψη του πρόσημου της  $\Delta\Delta G$  για 14/15 ανάλογες ενώσεις, ενώ οι τιμές MUE = 1 kcal/mol και RMSE = 1.5 kcal/mol (Σχήμα 4.51) υποδεικνύουν πως οι προσομοιώσεις που πραγματοποιήθηκαν δίνουν αποτελέσματα εντός του τυπικού σφάλματος της μεθόδου. Στον Πίνακα 4.13 συνοψίζονται οι τιμές σχετικής ελεύθερης ενέργειας πρόσδεσης, που εκτιμήθηκαν πειραματικά και υπολογιστικά και για τα δύο πακέτα Μοριακής Δυναμική, για όλα τα ανάλογα μόρια της ένωσης οδηγού.

**Πίνακας 4.13** Σύγκριση μεταξύ πειραματικών δεδομένων σχετικής ελεύθερης ενέργειας πρόσδεσης και αποτελεσμάτων των προσομοιώσεων που πραγματοποιήθηκαν με τα υπολογιστικά πακέτα NAMD και GROMACS.

| Ανάλογη Ένωση | $\Delta\Delta G_{\text{GROMACS}}$ [kcal/mol] | $\Delta\Delta G_{\text{NAMD}}$ [kcal/mol] | $\Delta\Delta G_{\text{experimental}}$ [kcal/mol] |
|---------------|--|---|---|
| CK-666        | 0.00 ± 0.00                                  | 0.00 ± 0.00                               | 0.00  |
| ai003         | -4.22 ± 0.06                                 | -5.54 ± 0.05                              | -1.28   |
| ai007         | -0.48 ± 0.12                                 | 1.02 ± 0.1                                | -0.54   |
| ai015         | 2.7 ± 0.11                                   | -1.31 ± 0.04                              | -0.55   |
| ai062         | 1.41 ± 0.09                                  | -2.33 ± 0.17                              | 1.13  |
| ai064         | 1.29 ± 0.07                                  | 0.76 ± 0.02                               | 1.42  |
| ai065         | 1.67 ± 0.1                                   | -0.43 ± 0.06                              | 1.87  |
| ai066         | 4.79 ± 0.28                                  | -0.2 ± 0.12                               | 1.87  |
| ai067         | 2.18 ± 0.09                                  | 0.09 ± 0.01                               | 1.87  |
| ai068         | 1.43 ± 0.15                                  | -0.2 ± 0.05                               | 1.87  |
| ai071         | 0.96 ± 0.04                                  | 0.02 ± 0.06                               | 3.40  |
| ai078         | 3.13 ± 0.09                                  | 0.65 ± 0.07                               | 3.40  |
| ai079         | 2.52 ± 0.07                                  | 3.32 ± 0.09                               | 3.40  |
| ai086         | 2.84 ± 0.24                                  | 6.07 ± 0.37                               | 3.40  |
| ai093         | -0.4 ± 0.16                                  | 0.21 ± 0.04                               | -0.54   |
| ai094         | -1.57 ± 0.1                                  | -5.78 ± 0.05                              | -0.99   |

Με βάση τα δεδομένα του Πίνακα 4.13 και σε συνδυασμό με τα διαγράμματα που απεικονίζονται στα Σχήματα 4.31 και 4.51 γίνεται αντιληπτό πως οι προσομοιώσεις με το υπολογιστικό πακέτο GROMACS οδήγησαν σε πιο ακριβείς εκτιμήσεις  $\Delta\Delta G$  σε σύγκριση με το NAMD. Για την κατανόηση των αιτιών στις οποίες οφείλονται οι αρκετά διαφορετικές εκτιμήσεις ανάμεσα στα δύο πακέτα, θα πρέπει να κανείς να εστιάσει την προσοχή του στον τρόπο με τον οποίο προετοιμάστηκαν οι ανάλογες ενώσεις για τις αντίστοιχες προσομοιώσεις.

Στην περίπτωση του NAMD, η παραμετροποίηση των μορίων έγινε με το πεδίο δυνάμεων OPLS-AA, ενώ για το πακέτο GROMACS η παραμετροποίηση πραγματοποιήθηκε με το GAFF. Έτσι, ένας από τους λόγους ασυμφωνίας των προσομοιώσεων με τα δύο πακέτα εντοπίζεται στη χρήση διαφορετικού πεδίου δυνάμεων. Επιπλέον, τα μοντέλα με βάση τα οποία υπολογίζονται τα μερικά φορτία των ατόμων αποτελούν μια εξίσου σημαντική πηγή σφάλματος, καθώς τα περισσότερα από αυτά προέρχονται από ημι-εμπειρικές προσεγγίσεις οι οποίες θεωρούν σταθερό σημειακό φορτίο



(point charge) και αγνοούν σε μεγάλο βαθμό τις αλληλεπιδράσεις των ατόμων με το περιβάλλον [117].

Στην προκειμένη περίπτωση, παρατηρήθηκε πως το μοντέλο 1.14\*CM1A το οποίο χρησιμοποιήθηκε στις προσομοιώσεις με το πακέτο NAMM, εκτίμησε το φορτίο του αμιδικού αζώτου των ανάλογων ενώσεων περίπου -1, ενώ στην περίπτωση του μοντέλου AM1-BCC το οποίο χρησιμοποιήθηκε για τις προσομοιώσεις με το πακέτο GROMACS, η τιμή του συγκεκριμένου φορτίου βρέθηκε ίση με -0.5. Επομένως, η υπερεκτίμηση του αρνητικού φορτίου από το μοντέλο 1.14\*CM1A, οδηγεί στο συμπέρασμα πως η χρήση διαφορετικών μοντέλων για την εκτίμηση των φορτίων αποτελεί την κυριότερη αιτία ασυμφωνίας μεταξύ των δύο υπολογιστικών πακέτων. Λαμβάνοντας υπόψιν όλα τα παραπάνω, γίνεται σαφές πως για την περίπτωση της ένωσης-οδηγού CK-666 απαιτείται ιδιαίτερη μεταχείριση κατά την παραμετροποίηση και την ανάθεση των φορτίων.

Επιπρόσθετα, μετά το πέρας των υπολογισμών σχετικής ελεύθερης ενέργειας είναι εύλογο να αναρωτηθεί κανείς εάν η εκτιμώμενη αλλαγή συμβαδίζει με τις νέες αλληλεπιδράσεις που σχηματίζονται μεταξύ της ανάλογης ένωσης και της Arg2/3. Ο αναστολέας CK-666 αλληλεπιδρά με την πρωτεΐνη μέσω δεσμών υδρογόνου και van der Waals αλληλεπιδράσεων. Συγκεκριμένα, σχηματίζει δύο δεσμούς υδρογόνου, το αμιδικό οξυγόνο του CK-666 με το σκελετό της Ala203 και το άζωτο του ινδολικού δακτυλίου με την πλευρική αλυσίδα της Asp248. Οι van der Waals αλληλεπιδράσεις είναι πολυάριθμες και περιλαμβάνουν τα εξής αμινοξέα της πρωτεΐνης: Ile252, Tyr202, Thr119, Leu246, Arg250, Asp248, Leu117, Ser188, Gly187, Asp209, Phe204, Ala203. Εξαιτίας του μεγάλου αριθμού των ανάλογων ενώσεων, θα παρουσιαστούν ενδεικτικά οι περιπτώσεις των ai003 και ai094.

Κατά τις προσομοιώσεις του ai003 και με τα δύο πακέτα παρατηρήθηκε συχνός σχηματισμός δεσμών υδρογόνου καθ' όλη τη διάρκεια της προσομοίωσης τόσο με την ίδια την πρωτεΐνη, όσο και με μόρια νερού. Συγκεκριμένα, το υδρογόνο της -OH ομάδας που προστέθηκε στη θέση R4 βρέθηκε να συμμετέχει σε δεσμούς υδρογόνου με το ένα εκ των δύο οξυγόνων της καρβοξυλομάδας της πλευρικής αλυσίδας της Asp248. Συχνό ήταν το φαινόμενο κατά το οποίο και τα δύο οξυγόνα της καρβοξυλομάδας της Asp248 συμμετείχαν ταυτόχρονα σε δύο δεσμούς υδρογόνου τόσο με το προαναφερθέν υδρογόνο, όσο και με το υδρογόνο της -NH ομάδας του ινδολικού δακτυλίου. Επιπλέον, στους δεσμούς υδρογόνου που παρατηρήθηκαν σε όλο το εύρος του ανάλογου μορίου ανήκει ο δεσμός που βρέθηκε μεταξύ του φθορίου της θέσης R6 και του υδρογόνου της πλευρικής ομάδας της Arg250, ενώ το υδρογόνο της θέσης R7 συμμετείχε σε δεσμούς υδρογόνου με το άζωτο και το οξυγόνο που λάμβαναν μέρος στο σχηματισμό του πεπτιδικού δεσμού μεταξύ της Ile251 και Ile252.

Τέλος, στην ανάλογη ένωση ai094 τα υδρογόνα στις θέσεις R4 και R8 έχουν αντικατασταθεί από -OH. Έτσι, το μόριο αυτό περιλαμβάνει την τροποποίηση που πραγματοποιήθηκε στην ένωση ai003 και επιπλέον διαθέτει μια υδροξυλομάδα στη θέση R8. Κατά τις προσομοιώσεις παρατηρήθηκαν παρόμοιες αλληλεπιδράσεις με αυτές που περιγράφηκαν παραπάνω στην περιοχή R4 και επιπρόσθετα διαπιστώθηκε ο συχνός σχηματισμός δεσμού υδρογόνου μεταξύ του οξυγόνου της -OH ομάδας στη θέση R8 και ενός υδρογόνου του σκελετού της Thr119. Συνεπώς γίνεται αντιληπτό πως και για τα δύο μόρια σχηματίζονται πιο ισχυρές αλληλεπιδράσεις σε σχέση με την ένωση οδηγό, επομένως μπορεί να δικαιολογηθεί το αρνητικό πρόσημο της ΔΔG. Στις υπόλοιπες ενώσεις δεν παρατηρήθηκε ο εκτενής σχηματισμών δεσμών υδρογόνου όπως στις παραπάνω περιπτώσεις, με αποτέλεσμα κύρια συνεισφορά στη σχετική ελεύθερη ενέργεια πρόσδεσης να έχουν οι van der Waals αλληλεπιδράσεις.



## 5. Συμπεράσματα

Οι υπολογισμοί Διατάραξης Ελεύθερης Ενέργειας (Free Energy Perturbation, FEP) στο πλαίσιο των προσομοιώσεων Μοριακής Δυναμικής, μπορούν να εκτιμήσουν τις πειραματικές τιμές απόλυτης ή σχετικής ελεύθερης ενέργειας πρόσδεσης ενός φαρμάκου σε έναν στόχο με ανακρίβεια της τάξης των 1 – 2 kcal/mol. Με τον τρόπο αυτό, μπορούν να κατευθυνθούν οι προσπάθειες οργανικής σύνθεσης μόνο προς υποψήφια φάρμακα τα οποία αναμένεται να δεσμεύονται ισχυρά στο φαρμακευτικό στόχο, καθιστώντας έτσι το στάδιο της βελτιστοποίησης των ενώσεων-οδηγών πιο αποδοτικό από χρονική και οικονομική σκοπιά. Ωστόσο, το υψηλό υπολογιστικό κόστος, οι παραδοχές που χαρακτηρίζουν τα σύγχρονα πεδία δυνάμεων, και οι τεχνικές δυσκολίες που διέπουν την προετοιμασία των υπολογισμών λειτουργούν ως φραγμός για την επιτυχή εφαρμογή της μεθοδολογίας FEP στη βιομηχανική ή την ακαδημαϊκή φαρμακευτική έρευνα. Στο πλαίσιο αυτό, στην παρούσα διπλωματική εργασία επιχειρήθηκε η διερεύνηση της εκτιμητικής ισχύος της μεθοδολογίας της Διατάραξης Ελεύθερης Ενέργειας, στο πλαίσιο των υπολογισμών απόλυτης και σχετικής ελεύθερης ενέργειας πρόσδεσης.

Αρχικά, λάβαμε μέρος στην πρόκληση SAMPL6 SAMPLing, η οποία επικεντρώνεται στην αξιολόγηση της σύγκλισης και της αναπαραγωγιμότητας των προβλέψεων της απόλυτης ελεύθερης ενέργειας πρόσδεσης. Για το σκοπό αυτό, πραγματοποιήθηκαν υπολογισμοί απόλυτης ελεύθερης ενέργειας πρόσδεσης τριών συστημάτων προσδετών-δεκτών CB8-G3, OA-G3 και OA-G6, χρησιμοποιώντας το πακέτο NAMD. Η μέση τιμή των προβλέψεων του NAMD για το σύστημα CB8-G3 ανέρχεται στα  $-12.6 \pm 0.6$  kcal/mol και είναι συνεπής σε εύρος, 1-2 kcal/mol, με τους υπολογισμούς αναφοράς ( $-10.9 \pm 0.1$  kcal/mol) που έγιναν από τους διοργανωτές με το πακέτο YANK. Ωστόσο, και τα δύο πακέτα υπερεκτίμησαν την ελεύθερη ενέργεια πρόσδεσης της κινίνης στον δέκτη CB8, καθώς η πειραματικά υπολογισμένη ελεύθερη ενέργεια πρόσδεσης ανέρχεται στα  $-6.5 \pm 0.1$  kcal/mol. Όσον αφορά τα συστήματα του μακρομοριακού δέκτη OA, η απόλυτη ελεύθερη ενέργεια πρόσδεσης εκτιμήθηκε  $-6.8 \pm 0.1$  kcal/mol και  $-7.3 \pm 0.1$  kcal/mol για την περίπτωση του 5-εξενοϊκού (OA-G3) και 4-μεθυλοπεντανοϊκού οξέος (OA-G6), αντίστοιχα. Τα αποτελέσματα του NAMD και για τα δύο συστήματα, είναι σε συμφωνία με τους υπολογισμούς αναφοράς που πραγματοποιήθηκαν από τους διοργανωτές της πρόκλησης, μιας και οι προσομοιώσεις με το υπολογιστικό πακέτο YANK εκτίμησαν την ελεύθερη ενέργεια πρόσδεσης  $-6.7 \pm 0.1$  kcal/mol και  $-7.2 \pm 0.1$  kcal/mol για τα συστήματα OA-G3 και OA-G6, αντίστοιχα. Επιπρόσθετα, υπήρξε βελτιωμένη συσχέτιση με τα πειραματικά αποτελέσματα, σε σχέση με το σύστημα CB8-G3, καθώς οι αντίστοιχες πειραματικές μετρήσεις εκτίμησαν την μεταβολή της ελεύθερης ενέργειας στα  $-5.2$  kcal/mol και  $-4.97$  kcal/mol για τα συστήματα OA-G3 και OA-G6, αντίστοιχα. Αξίζει να αναφερθεί πως οι εκτιμήσεις των άλλων υπολογιστικών πακέτων όπως το GROMACS, AMBER και SOMD από τις υπόλοιπες ομάδες που έλαβαν μέρος στο διαγωνισμό, βρίσκονταν στο ίδιο εύρος με τις προβλέψεις των NAMD και YANK.

Στο πλαίσιο των υπολογισμών σχετικής ελεύθερης ενέργειας πρόσδεσης, μελετήθηκε μία σειρά 15 ανάλογων μορίων της ένωσης-οδηγού CK-666, η οποία αποτελεί έναν υποψήφιο αναστολέα της πρωτεΐνης Apr2/3, μέσω προσομοιώσεων Μοριακής Δυναμικής με τη χρήση των υπολογιστικών πακέτων NAMD και GROMACS. Οι υπολογισμοί σχετικής ελεύθερης ενέργειας πρόσδεσης φανέρωσαν, πως το υπολογιστικό πακέτο GROMACS σε συνδυασμό με το πεδίο δυνάμεων GAFF και το μοντέλο φορτίων AM1-BCC, οδήγησε σε καλύτερα συσχετιζόμενα αποτελέσματα με τις αντίστοιχες πειραματικές μετρήσεις, σε σύγκριση με το υπολογιστικό πακέτο NAMD, στο οποίο ως πεδίο δυνάμεων και μοντέλο φορτίων χρησιμοποιήθηκαν τα OPLS-AA και 1.14\*CM1A. Πιο

συγκεκριμένα, οι υπολογισμοί FEP με το πακέτο GROMACS ανέδειξαν 11 ενώσεις εντός της περιοχής  $\pm 1$  kcal/mol, η οποία αποτελεί το τυπικό σφάλμα των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας, ενώ ταυτόχρονα προβλέφθηκε σωστά το πρόσημο της σχετικής ελεύθερης ενέργειας για 14/15 ανάλογες ενώσεις. Από την άλλη πλευρά, οι υπολογισμοί με το πακέτο NAMD ανέδειξαν μόλις 5 ενώσεις εντός της περιοχής  $\pm 1$  kcal/mol, ενώ το πρόσημο της σχετικής ελεύθερης ενέργειας προβλέφθηκε σωστά για 9/15 ανάλογες ενώσεις. Επιπλέον, οι τιμές του μέσου απόλυτου σφάλματος (MUE) και της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE) υπολογίστηκαν 1.0 και 1.5 kcal/mol για το πακέτο GROMACS, ενώ για το υπολογιστικό πακέτο NAMD οι αντίστοιχες τιμές βρέθηκαν 2.2 και 2.6 kcal/mol. Εκτιμάται πως η ασυμφωνία των προσομοιώσεων που πραγματοποιήθηκαν με τα δύο πακέτα οφείλονται σε δύο κύριες αιτίες. Ο πρώτος λόγος αφορά τη χρήση διαφορετικού πεδίου δυνάμεων, μιας και κάθε ένα από αυτά εκτιμά διαφορετικές παραμέτρους για τον υπολογισμό της δυναμικής ενέργειας των προσδετών. Ο δεύτερος λόγος σχετίζεται με τη χρήση του μοντέλου 1.14\*CM1A για τον υπολογισμό των μερικών ατομικών φορτίων κατά τις προσομοιώσεις με το πακέτο NAMD. Συγκεκριμένα, το μοντέλο 1.14\*CM1A υπολόγισε το αρνητικό φορτίο του αμιδικού αζώτου περίπου -1, ενώ στην περίπτωση του μοντέλου AM1-BCC το συγκεκριμένο φορτίο βρέθηκε περίπου -0.5. Τόσο στο πλαίσιο του διαγωνισμού SAMPL6 SAMPLing, όσο και στην περίπτωση των ανάλογων μορίων της ένωσης-οδηγού, έγινε φανερό πως η απόκλιση από τα αντίστοιχα πειραματικά δεδομένα κυμαίνεται στις πιο αισιόδοξες περιπτώσεις από 1 έως 2 kcal/mol. Επομένως, σε κάθε περίπτωση αντιλαμβάνεται κανείς πως στις προσομοιώσεις εισέρχονται αναπόφευκτα ευρύτερα συστηματικά σφάλματα, τα οποία σχετίζονται με τη μοντελοποίηση των βιολογικών συστημάτων (πεδίο δυνάμεων, μοντέλο νερού) και τους αλγορίθμους δειγματοληψίας.

## 6. Προτάσεις για μελλοντική έρευνα

Αρχικά, στο πλαίσιο των υπολογισμών απόλυτης ελεύθερης ενέργειας πρόσδεσης που πραγματοποιήθηκαν για την πρόκληση SAMPL6 SAMPLing, θα μπορούσε κανείς να τροποποιήσει το πρωτόκολλο που ακολουθήθηκε για το σύστημα CB8-G3 ούτως ώστε οι ηλεκτροστατικές και οι van der Waals αλληλεπιδράσεις να διαταράσσονται στο ίδιο εύρος ενδιάμεσων καταστάσεων με την περίπτωση των συστημάτων OA-G3/G6, και να ελέγξει εάν οδηγείται σε παρόμοια αποτελέσματα. Γενικότερα, οι «τυφλές» προκλήσεις που απευθύνονται στην κοινότητα της υπολογιστικής χημείας, όπως ο διαγωνισμός SAMPL6 SAMPLing, αναμένεται να διαδραματίσουν καταλυτικό ρόλο για την πρόοδο των υπολογιστικών τεχνικών που χρησιμοποιούνται στον ορθολογικό σχεδιασμό φαρμάκων. Στο πλαίσιο των υπολογισμών σχετικής ελεύθερης ενέργειας πρόσδεσης θα πρέπει να χρησιμοποιηθεί το ίδιο πεδίο δυνάμεων και μοντέλο περιγραφής των ατομικών φορτίων ούτως ώστε να πραγματοποιηθεί μία πιο άμεση σύγκριση μεταξύ της εκτιμητικής ισχύος των δύο δημοφιλών υπολογιστικών πακέτων Μοριακής Δυναμικής NAMD και GROMACS. Επιπλέον, θα πρέπει να διερευνηθεί η προβλεπτική ικανότητα προσομοιώσεων Διατάραξης Ελεύθερης Ενέργειας των ανάλογων ενώσεων του CK-666 με επιπλέον υπολογιστικά πακέτα όπως τα MCPRO, FEP+ και AMBER, αλλά και διαφορετικά πεδία δυνάμεων όπως το GAFF2. Επιπρόσθετα, θα πρέπει να πραγματοποιηθούν οι αντίστοιχοι υπολογισμοί για ένα ευρύτερο σύνολο ανάλογων μορίων της ένωσης οδηγού, καθώς μία τέτοια μελέτη μεγάλης κλίμακας θα αποτελούσε ένα χρήσιμο εργαλείο για την ακαδημαϊκή και για τη βιομηχανική κοινότητα για την περεταίρω διερεύνηση της ακρίβειας των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας.

Όπως έγινε ξεκάθαρο και στις δύο περιπτώσεις που μελετήθηκαν, η εξέλιξη των σύγχρονων πεδίων δυνάμεων αναμένεται να ωφελήσει την ακρίβεια και την αξιοπιστία των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας. Καινοτόμες προσεγγίσεις που στηρίζονται τόσο στη διαθεσιμότητα μεγάλου όγκου πειραματικών και κβαντομηχανικών δεδομένων όσο και στις προσεγγίσεις της μηχανικής μάθησης, θα δίνουν τη δυνατότητα δημιουργίας πεδίων δυνάμεων τα οποία θα είναι εξειδικευμένα για τις διάφορες οικογένειες προσδετών που μελετώνται, εξαλείφοντας έτσι το μεγαλύτερο μέρος των συστηματικών σφαλμάτων που εισέρχονται στις μοριακές προσομοιώσεις. Παρόμοιο αντίκτυπο αναμένεται να έχει και η καθιέρωση προηγμένων τεχνικών δειγματοληψίας, όπως η τεχνική Replica Exchange with Solute Tempering, καθώς θα δίνουν της δυνατότητα εξερεύνησης μεγαλύτερο μέρος του χημικού χώρου, βελτιώνοντας έτσι τη σύγκλιση των υπολογισμών Διατάραξης Ελεύθερης Ενέργειας. Τέλος, εξίσου σημαντική είναι η αυτοματοποίηση των υπολογισμών αυτών τόσο ως προς την προετοιμασία όσο και προς την ανάλυση, ούτως ώστε να γίνουν διαθέσιμο εργαλείο στο ευρύτερο κοινό των ερευνητών. Ωστόσο, θα πρέπει να αναφερθεί πως μία εξαιρετικά ακριβής και αξιόπιστη μεθοδολογία FEP δεν είναι, σε καμία περίπτωση, αντικαταστάτης μιας δημιουργικής και τεχνικά ισχυρής ομάδας φαρμακευτικής χημείας. Αντίθετα, η υπολογιστική προσέγγιση που περιγράφεται εδώ μπορεί να θεωρηθεί ως ένα εργαλείο που επιτρέπει στους συνθετικούς χημικούς να ακολουθήσουν τροποποιήσεις και νέες συνθετικές κατευθύνσεις που θα θεωρούνταν επίφοβες χωρίς υπολογιστική επικύρωση ή να απορρίπτονται ενώσεις που είναι απίθανο να ικανοποιήσουν την επιθυμητή ελεύθερη ενέργεια πρόσδεσης. Οι συνεχείς βελτιώσεις στις τεχνικές που περιγράφηκαν παραπάνω θα καταστήσουν τις υπολογιστικές προσεγγίσεις ολοένα και πιο ελκυστικές, οδηγώντας σε βελτιωμένα εργαλεία προσομοίωσης και τελικά σε ένα αναπόσπαστο μέρος της φαρμακοβιομηχανίας.

## Ευρετήριο σχημάτων

|                  |    |
|------------------|----|
| Σχήμα 1.1 .....  | 2  |
| Σχήμα 1.2 .....  | 5  |
| Σχήμα 1.3 .....  | 6  |
| Σχήμα 1.4 .....  | 8  |
| Σχήμα 1.5 .....  | 9  |
| Σχήμα 1.6 .....  | 13 |
| Σχήμα 1.7 .....  | 13 |
| Σχήμα 1.8 .....  | 14 |
| Σχήμα 1.9 .....  | 15 |
| Σχήμα 1.10 ..... | 17 |
| Σχήμα 1.11 ..... | 17 |
| Σχήμα 1.12 ..... | 19 |
| Σχήμα 1.13 ..... | 19 |
| Σχήμα 1.14 ..... | 20 |
| Σχήμα 1.15 ..... | 21 |
| Σχήμα 1.16 ..... | 21 |
| Σχήμα 1.17 ..... | 22 |
| Σχήμα 1.18 ..... | 22 |
| Σχήμα 1.19 ..... | 23 |
| Σχήμα 1.20 ..... | 26 |
| Σχήμα 1.21 ..... | 27 |
| Σχήμα 1.22 ..... | 28 |
| Σχήμα 1.23 ..... | 29 |
| Σχήμα 1.24 ..... | 30 |
| Σχήμα 1.25 ..... | 31 |
| Σχήμα 1.26 ..... | 32 |
| Σχήμα 2.1 .....  | 35 |
| Σχήμα 2.2 .....  | 37 |
| Σχήμα 2.3 .....  | 38 |
| Σχήμα 2.4 .....  | 41 |
| Σχήμα 2.5 .....  | 41 |
| Σχήμα 2.6 .....  | 45 |
| Σχήμα 2.7 .....  | 46 |
| Σχήμα 2.8 .....  | 47 |
| Σχήμα 3.1 .....  | 52 |
| Σχήμα 3.2 .....  | 53 |
| Σχήμα 3.3 .....  | 56 |
| Σχήμα 3.4 .....  | 59 |
| Σχήμα 3.5 .....  | 60 |
| Σχήμα 3.6 .....  | 61 |
| Σχήμα 3.7 .....  | 62 |
| Σχήμα 3.8 .....  | 62 |
| Σχήμα 3.9 .....  | 64 |

|                  |     |
|------------------|-----|
| Σχήμα 3.10 ..... | 66  |
| Σχήμα 3.11 ..... | 69  |
| Σχήμα 3.12 ..... | 70  |
| Σχήμα 3.13 ..... | 71  |
| Σχήμα 4.1 .....  | 73  |
| Σχήμα 4.2 .....  | 74  |
| Σχήμα 4.3 .....  | 75  |
| Σχήμα 4.4 .....  | 76  |
| Σχήμα 4.5 .....  | 77  |
| Σχήμα 4.6 .....  | 79  |
| Σχήμα 4.7 .....  | 79  |
| Σχήμα 4.8 .....  | 80  |
| Σχήμα 4.9 .....  | 81  |
| Σχήμα 4.10 ..... | 82  |
| Σχήμα 4.11 ..... | 84  |
| Σχήμα 4.12 ..... | 85  |
| Σχήμα 4.13 ..... | 86  |
| Σχήμα 4.14 ..... | 87  |
| Σχήμα 4.15 ..... | 87  |
| Σχήμα 4.16 ..... | 88  |
| Σχήμα 4.17 ..... | 88  |
| Σχήμα 4.18 ..... | 89  |
| Σχήμα 4.19 ..... | 89  |
| Σχήμα 4.20 ..... | 90  |
| Σχήμα 4.21 ..... | 90  |
| Σχήμα 4.22 ..... | 91  |
| Σχήμα 4.23 ..... | 91  |
| Σχήμα 4.24 ..... | 92  |
| Σχήμα 4.25 ..... | 92  |
| Σχήμα 4.26 ..... | 93  |
| Σχήμα 4.27 ..... | 93  |
| Σχήμα 4.28 ..... | 94  |
| Σχήμα 4.29 ..... | 94  |
| Σχήμα 4.30 ..... | 95  |
| Σχήμα 4.31 ..... | 98  |
| Σχήμα 4.32 ..... | 99  |
| Σχήμα 4.33 ..... | 100 |
| Σχήμα 4.34 ..... | 101 |
| Σχήμα 4.35 ..... | 103 |
| Σχήμα 4.36 ..... | 105 |
| Σχήμα 4.37 ..... | 105 |
| Σχήμα 4.38 ..... | 106 |
| Σχήμα 4.39 ..... | 106 |
| Σχήμα 4.40 ..... | 107 |
| Σχήμα 4.41 ..... | 107 |
| Σχήμα 4.42 ..... | 108 |
| Σχήμα 4.43 ..... | 108 |
| Σχήμα 4.44 ..... | 109 |

|                  |     |
|------------------|-----|
| Σχήμα 4.45 ..... | 109 |
| Σχήμα 4.46 ..... | 110 |
| Σχήμα 4.47 ..... | 110 |
| Σχήμα 4.48 ..... | 111 |
| Σχήμα 4.49 ..... | 111 |
| Σχήμα 4.50 ..... | 112 |
| Σχήμα 4.51 ..... | 113 |

## Ευρετήριο πινάκων

|                    |     |
|--------------------|-----|
| Πίνακας 3.1 .....  | 59  |
| Πίνακας 3.2 .....  | 69  |
| Πίνακας 3.3 .....  | 70  |
| Πίνακας 3.4 .....  | 71  |
| Πίνακας 4.1 .....  | 77  |
| Πίνακας 4.2 .....  | 80  |
| Πίνακας 4.3 .....  | 81  |
| Πίνακας 4.4 .....  | 84  |
| Πίνακας 4.5 .....  | 84  |
| Πίνακας 4.6 .....  | 85  |
| Πίνακας 4.7 .....  | 95  |
| Πίνακας 4.8 .....  | 97  |
| Πίνακας 4.9 .....  | 97  |
| Πίνακας 4.10 ..... | 102 |
| Πίνακας 4.11 ..... | 104 |
| Πίνακας 4.12 ..... | 112 |
| Πίνακας 4.13 ..... | 114 |



## Βιβλιογραφία

1. Merz Jr, K.M., D. Ringe, and C.H. Reynolds, *Drug design: structure-and ligand-based approaches*. 2010: Cambridge University Press.
2. Lionta, E., et al., *Structure-based virtual screening for drug discovery: principles, applications and recent advances*. Current topics in medicinal chemistry, 2014. **14**(16): p. 1923-1938.
3. Vohora, D. and G. Singh, *Pharmaceutical Medicine and Translational Clinical Research*. 2018: Academic Press.
4. Bleicher, K.H., et al., *A guide to drug discovery: hit and lead generation: beyond high-throughput screening*. Nature reviews Drug discovery, 2003. **2**(5): p. 369.
5. Jorgensen, W.L. and E.M. Duffy, *Prediction of drug solubility from structure*. Advanced drug delivery reviews, 2002. **54**(3): p. 355-366.
6. *Step 3: Clinical Research*. 2018.
7. Paul, S.M., et al., *How to improve R&D productivity: the pharmaceutical industry's grand challenge*. Nature reviews Drug discovery, 2010. **9**(3): p. 203.
8. Hopkins, A.L. and C.R. Groom, *The druggable genome*. Nature reviews Drug discovery, 2002. **1**(9): p. 727.
9. Brown, D. and G. Superti-Furga, *Rediscovering the sweet spot in drug discovery*. Drug discovery today, 2003. **8**(23): p. 1067-1077.
10. Macarron, R., et al., *Impact of high-throughput screening in biomedical research*. Nature reviews Drug discovery, 2011. **10**(3): p. 188.
11. Wang, L., et al., *Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field*. Journal of the American Chemical Society, 2015. **137**(7): p. 2695-2703.
12. Shirts, M.R., D.L. Mobley, and S.P. Brown, *Free-energy calculations in structure-based drug design*. Drug design: structure-and ligand-based approaches, 2010: p. 61-86.
13. Gapsys, V., et al., *Calculation of binding free energies*, in *Molecular Modeling of Proteins*. 2015, Springer. p. 173-209.
14. Atkins, P., J. De Paula, and J. Keeler, *Atkins' physical chemistry*. 2018: Oxford university press.
15. Du, X., et al., *Insights into protein–ligand interactions: mechanisms, models, and methods*. International journal of molecular sciences, 2016. **17**(2): p. 144.
16. Chipot, C. and A. Pohorille, *Free energy calculations*. 2007: Springer.
17. Aldeghi, M., et al., *Accurate calculation of the absolute free energy of binding for drug molecules*. Chemical science, 2016. **7**(1): p. 207-218.
18. Mobley, D.L. and M.K. Gilson, *Predicting binding free energies: frontiers and benchmarks*. Annual review of biophysics, 2017. **46**: p. 531-558.
19. Gumbart, J.C., B. Roux, and C. Chipot, *Efficient determination of protein–protein standard binding free energies from first principles*. Journal of chemical theory and computation, 2013. **9**(8): p. 3789-3798.
20. Chodera, J.D., et al., *Alchemical free energy methods for drug discovery: progress and challenges*. Current opinion in structural biology, 2011. **21**(2): p. 150-160.
21. Stout, T.J., et al., *Structure-based design of inhibitors specific for bacterial thymidylate synthase*. Biochemistry, 1999. **38**(5): p. 1607-1617.
22. Mobley, D.L. and P.V. Klimovich, *Perspective: Alchemical free energy calculations for drug discovery*. The Journal of chemical physics, 2012. **137**(23): p. 230901.
23. Rizzo, R.C., et al., *Validation of a model for the complex of HIV-1 reverse transcriptase with sustiva through computation of resistance profiles*. Journal of the American Chemical Society, 2000. **122**(51): p. 12898-12900.

24. Kim, J.T., et al., *FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase*. Journal of the American Chemical Society, 2006. **128**(48): p. 15372-15373.
25. Salmaso, V., et al., *Combining self-and cross-docking as benchmark tools: the performance of DockBench in the D3R Grand Challenge 2*. Journal of computer-aided molecular design, 2018. **32**(1): p. 251-264.
26. Abagyan, R. and M. Totrov, *High-throughput docking for lead generation*. Current opinion in chemical biology, 2001. **5**(4): p. 375-382.
27. Kuntz, I.D., et al., *A geometric approach to macromolecule-ligand interactions*. Journal of molecular biology, 1982. **161**(2): p. 269-288.
28. Salmaso, V. and S. Moro, *Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: an overview*. Frontiers in pharmacology, 2018. **9**.
29. Åqvist, J., C. Medina, and J.-E. Samuelsson, *A new method for predicting binding affinity in computer-aided drug design*. Protein Engineering, Design and Selection, 1994. **7**(3): p. 385-391.
30. Gutiérrez-de-Terán, H. and J. Åqvist, *Linear interaction energy: method and applications in drug design*, in *Computational drug discovery and design*. 2012, Springer. p. 305-323.
31. Srinivasan, J., et al., *Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices*. Journal of the American Chemical Society, 1998. **120**(37): p. 9401-9409.
32. Sharp, K.A., et al., *Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects*. Science, 1991. **252**(5002): p. 106-109.
33. Geney, R., et al., *Investigation of salt bridge stability in a generalized born solvent model*. Journal of Chemical Theory and Computation, 2006. **2**(1): p. 115-127.
34. Zwanzig, R.W., *High-temperature equation of state by a perturbation method. I. Nonpolar gases*. The Journal of Chemical Physics, 1954. **22**(8): p. 1420-1426.
35. Jorgensen, W.L. and L.L. Thomas, *Perspective on free-energy perturbation calculations for chemical equilibria*. Journal of chemical theory and computation, 2008. **4**(6): p. 869-876.
36. Jorgensen, W.L. and C. Ravimohan, *Monte Carlo simulation of differences in free energies of hydration*. The Journal of chemical physics, 1985. **83**(6): p. 3050-3054.
37. Shirts, M.R. and D.L. Mobley, *An introduction to best practices in free energy calculations*, in *Biomolecular Simulations*. 2013, Springer. p. 271-311.
38. Zeevaart, J.G., et al., *Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations*. Journal of the American Chemical Society, 2008. **130**(29): p. 9492-9499.
39. Freyer, M.W. and E.A. Lewis, *Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions*. Methods in cell biology, 2008. **84**: p. 79-113.
40. Song, C., S. Zhang, and H. Huang, *Choosing a suitable method for the identification of replication origins in microbial genomes*. Frontiers in microbiology, 2015. **6**: p. 1049.
41. Hall, M.D., et al., *Fluorescence polarization assays in high-throughput screening and drug discovery: a review*. Methods and applications in fluorescence, 2016. **4**(2): p. 022001.
42. Frostell, Å., L. Vinterbäck, and H. Sjöbom, *Protein–Ligand Interactions Using SPR Systems*, in *Protein-Ligand Interactions*. 2013, Springer. p. 139-165.
43. de Freitas, R.F. and M. Schapira, *A systematic analysis of atomic protein–ligand interactions in the PDB*. Medchemcomm, 2017. **8**(10): p. 1970-1981.
44. Jorgensen, W.L., *Efficient drug lead discovery and optimization*. Accounts of chemical research, 2009. **42**(6): p. 724-733.
45. Sirimulla, S., et al., *Halogen interactions in protein–ligand complexes: implications of halogen bonding for rational drug design*. Journal of chemical information and modeling, 2013. **53**(11): p. 2781-2791.

46. Qiu, Y., et al., *Developing solid oral dosage forms: pharmaceutical theory and practice*. 2016: Academic press.
47. Pollard, T.D., et al., *Cell Biology E-Book*. 2016: Elsevier Health Sciences.
48. Lucas, X., et al., *A thorough anion- $\pi$  interaction study in biomolecules: on the importance of cooperativity effects*. *Chemical science*, 2016. **7**(2): p. 1038-1050.
49. Rizzi, A., et al., *Overview of the SAMPL6 host-guest binding affinity prediction challenge*. *Journal of computer-aided molecular design*, 2018. **32**(10): p. 937-963.
50. Yin, J., et al., *Overview of the SAMPL5 host-guest challenge: Are we doing better?* *Journal of computer-aided molecular design*, 2017. **31**(1): p. 1-19.
51. Nicholls, A., et al., *Predicting small-molecule solvation free energies: an informal blind test for computational chemistry*. *Journal of medicinal chemistry*, 2008. **51**(4): p. 769-779.
52. Guthrie, J.P., *A blind challenge for computational solvation free energies: introduction and overview*. 2009, ACS Publications.
53. Geballe, M.T., et al., *The SAMPL2 blind prediction challenge: introduction and overview*. *Journal of computer-aided molecular design*, 2010. **24**(4): p. 259-279.
54. Muddana, H.S., et al., *Blind prediction of host-guest binding affinities: a new SAMPL3 challenge*. *Journal of computer-aided molecular design*, 2012. **26**(5): p. 475-487.
55. Muddana, H.S., et al., *The SAMPL4 host-guest blind prediction challenge: an overview*. *Journal of computer-aided molecular design*, 2014. **28**(4): p. 305-317.
56. Papadourakis, M., S. Bosisio, and J. Michel, *Blinded predictions of standard binding free energies: lessons learned from the SAMPL6 challenge*. *Journal of computer-aided molecular design*, 2018. **32**(10): p. 1047-1058.
57. Goley, E.D. and M.D. Welch, *The ARP2/3 complex: an actin nucleator comes of age*. *Nature reviews Molecular cell biology*, 2006. **7**(10): p. 713.
58. Robinson, R.C., et al., *Crystal structure of Arp2/3 complex*. *science*, 2001. **294**(5547): p. 1679-1684.
59. Wang, W., et al., *Tumor cells caught in the act of invading: their strategy for enhanced cell motility*. *Trends in cell biology*, 2005. **15**(3): p. 138-145.
60. Wyckoff, J.B., J.E. Segall, and J.S. Condeelis, *The collection of the motile population of cells from a living tumor*. *Cancer research*, 2000. **60**(19): p. 5401-5404.
61. Otsubo, T., et al., *Involvement of Arp2/3 complex in the process of colorectal carcinogenesis*. *Modern pathology*, 2004. **17**(4): p. 461.
62. Yang, C. and T. Svitkina, *Filopodia initiation: focus on the Arp2/3 complex and formins*. *Cell adhesion & migration*, 2011. **5**(5): p. 402-408.
63. Nolen, B., et al., *Characterization of two classes of small molecule inhibitors of Arp2/3 complex*. *Nature*, 2009. **460**(7258): p. 1031.
64. Baggett, A.W., et al., *Structural characterization and computer-aided optimization of a small-molecule inhibitor of the Arp2/3 complex, a key regulator of the actin cytoskeleton*. *ChemMedChem*, 2012. **7**(7): p. 1286-94.
65. Hetrick, B., et al., *Small molecules CK-666 and CK-869 inhibit actin-related protein 2/3 complex by blocking an activating conformational change*. *Chemistry & biology*, 2013. **20**(5): p. 701-712.
66. Somer Jr, F.L., *Molecular Modelling for Beginners (Alan Hinchliffe)*. 2004, ACS Publications.
67. Allen, M.P., *Introduction to molecular dynamics simulation*. *Computational soft matter: from synthetic polymers to proteins*, 2004. **23**: p. 1-28.
68. Hehre, W.J., *A guide to molecular mechanics and quantum chemical calculations*. Vol. 2. 2003: Wavefunction Irvine, CA.
69. Frenkel, D. and B. Smit, *Understanding molecular simulation: from algorithms to applications*. Vol. 1. 2001: Elsevier.
70. Field, M.J., *A practical introduction to the simulation of molecular systems*. 1999: Cambridge University Press.

71. Alekozai, E.M., *Enhanced Multiscale Sampling of the Cel7A-Cellulose Interaction*. 2013.
72. Cournia, Z., *Structure and Dynamics of Biomembranes containing Cholesterol and other Biologically-Important Sterols: a computational perspective*. 2006.
73. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins*. Chemical reviews, 2006. **106**(5): p. 1589-1615.
74. Alder, B.J. and T.E. Wainwright, *Phase transition for a hard sphere system*. The Journal of chemical physics, 1957. **27**(5): p. 1208-1209.
75. Rahman, A., *Correlations in the motion of atoms in liquid argon*. Physical review, 1964. **136**(2A): p. A405.
76. McCammon, J., *Molecular dynamics study of the bovine pancreatic trypsin inhibitor*. Models for Protein Dynamics, 1976: p. 137.
77. McCammon, J.A., B.R. Gelin, and M. Karplus, *Dynamics of folded proteins*. Nature, 1977. **267**(5612): p. 585.
78. Berman, H.M., et al., *The protein data bank*, in *Protein Structure*. 2003, CRC Press. p. 394-410.
79. Berman, H., K. Henrick, and H. Nakamura, *Announcing the worldwide protein data bank*. Nature Structural & Molecular Biology, 2003. **10**(12): p. 980.
80. Beeman, D., *Some multistep methods for use in molecular dynamics calculations*. Journal of computational physics, 1976. **20**(2): p. 130-139.
81. Swope, W.C., et al., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*. The Journal of Chemical Physics, 1982. **76**(1): p. 637-649.
82. Verlet, L., *Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*. Physical review, 1967. **159**(1): p. 98.
83. Berman, H., et al., *The protein data Bank nucleic acids research*, 28: 235-242. URL: www.rcsb.org Citation, 2000.
84. Parker, M., *Protein structure from X-ray diffraction*. Journal of biological physics, 2003. **29**(4): p. 341-362.
85. Sastry, G.M., et al., *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments*. Journal of computer-aided molecular design, 2013. **27**(3): p. 221-234.
86. Schrödinger, *Schrödinger Release 2019-2: LigPrep*. 2019, New York, NY: Schrödinger LLC.
87. Gilson, M.K., et al., *The statistical-thermodynamic basis for computation of binding affinities: a critical review*. Biophysical journal, 1997. **72**(3): p. 1047-1069.
88. Cournia, Z., B. Allen, and W. Sherman, *Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations*. J Chem Inf Model, 2017. **57**(12): p. 2911-2937.
89. Athanasiou, C., et al., *Using physics-based pose predictions and free energy perturbation calculations to predict binding poses and relative binding affinities for FXR ligands in the D3R Grand Challenge 2*. J Comput Aided Mol Des, 2018. **32**(1): p. 21-44.
90. Pohorille, A., C. Jarzynski, and C. Chipot, *Good practices in free-energy calculations*. The Journal of Physical Chemistry B, 2010. **114**(32): p. 10235-10253.
91. Bennett, C.H., *Efficient Estimation of Free-Energy Differences from Monte-Carlo Data*. J Comput Phys, 1976. **22**(2): p. 245-268.
92. *Εισαγωγή στους υπερυπολογιστές και το σύστημα ARIS*. 2015 [cited 2019 20/08/2019]; Available from: <https://hpc.grnet.gr/supercomputer/>.
93. Sullivan, M.R., et al., *Binding of carboxylate and trimethylammonium salts to octa-acid and TEMOA deep-cavity cavitands*. Journal of computer-aided molecular design, 2017. **31**(1): p. 21-28.
94. Pazos, E., et al., *Cucurbit [8] uril (CB [8])-Based Supramolecular Switches*. Angewandte Chemie International Edition, 2019. **58**(2): p. 403-416.

95. Organization, W.H., *Guidelines for the treatment of malaria*. 2015: World Health Organization.
96. Price, A.J., S. Howard, and B.D. Cons, *Fragment-based drug discovery and its application to challenging drug targets*. *Essays in biochemistry*, 2017. **61**(5): p. 475-484.
97. Greenwood, J.R., et al., *Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution*. *Journal of computer-aided molecular design*, 2010. **24**(6-7): p. 591-604.
98. McGann, M., *FRED pose prediction and virtual screening accuracy*. *Journal of chemical information and modeling*, 2011. **51**(3): p. 578-596.
99. Wang, J., et al., *Development and testing of a general amber force field*. *Journal of computational chemistry*, 2004. **25**(9): p. 1157-1174.
100. Case, D., et al., *AMBER16, 2016*. San Francisco.
101. QUACPAC 2.0.1.2: OpenEye Scientific Software, S.F., NM.
102. Jakalian, A., et al., *Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method*. *Journal of computational chemistry*, 2000. **21**(2): p. 132-146.
103. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. *The Journal of chemical physics*, 1983. **79**(2): p. 926-935.
104. Joung, I.S. and T.E. Cheatham III, *Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations*. *The journal of physical chemistry B*, 2008. **112**(30): p. 9020-9041.
105. Byrd, R.H., et al., *A limited memory algorithm for bound constrained optimization*. *SIAM Journal on Scientific Computing*, 1995. **16**(5): p. 1190-1208.
106. Zhu, C., et al., *Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization*. *ACM Transactions on Mathematical Software (TOMS)*, 1997. **23**(4): p. 550-560.
107. Loncharich, R.J., B.R. Brooks, and R.W. Pastor, *Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanine-N'-methylamide*. *Biopolymers: Original Research on Biomolecules*, 1992. **32**(5): p. 523-535.
108. Eastman, P., et al., *OpenMM 7: Rapid development of high performance algorithms for molecular dynamics*. *PLoS computational biology*, 2017. **13**(7): p. e1005659.
109. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. *J Comput Chem*, 2005. **26**(16): p. 1781-802.
110. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N · log (N) method for Ewald sums in large systems*. *The Journal of chemical physics*, 1993. **98**(12): p. 10089-10092.
111. Ryckaert, J.-P., G. Ciccotti, and H.J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. *Journal of computational physics*, 1977. **23**(3): p. 327-341.
112. Feller, S.E., et al., *Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method*. *Journal of Chemical Physics*, 1995. **103**(11): p. 4613-4621.
113. Liu, P., et al., *A Toolkit for the Analysis of Free-Energy Perturbation Calculations*. *J Chem Theory Comput*, 2012. **8**(8): p. 2606-16.
114. Jacobson, M.P., et al., *A hierarchical approach to all-atom protein loop prediction*. *Proteins: Structure, Function, and Bioinformatics*, 2004. **55**(2): p. 351-367.
115. Pettersen, E.F., et al., *UCSF Chimera—a visualization system for exploratory research and analysis*. *Journal of computational chemistry*, 2004. **25**(13): p. 1605-1612.
116. Jorgensen, W.L. and J. Tirado-Rives, *Potential energy functions for atomic-level simulations of water and organic and biomolecular systems*. *Proceedings of the National Academy of Sciences*, 2005. **102**(19): p. 6665-6670.
117. Dodda, L.S., et al., *1.14\* CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations*. *The Journal of Physical Chemistry B*, 2017. **121**(15): p. 3864-3870.

118. Dodda, L.S., et al., *LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands*. *Nucleic acids research*, 2017. **45**(W1): p. W331-W336.
119. Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation*. *Computer physics communications*, 1995. **91**(1-3): p. 43-56.
120. Zavitsanou, S.N., *Automating Free Energy Perturbation Calculations for Drug Design*, in *Department of Informatics and Telecommunication*. 2018, National Kapodistrian University of Athens: Athens.
121. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. *Journal of molecular graphics*, 1996. **14**(1): p. 33-38.
122. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin*. *Journal of the American Chemical Society*, 1988. **110**(6): p. 1657-1666.
123. Loeffler, H.H., J. Michel, and C. Woods, *FESetup: automating setup for alchemical free energy simulations*. 2015, ACS Publications.
124. Maier, J.A., et al., *ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB*. *Journal of chemical theory and computation*, 2015. **11**(8): p. 3696-3713.
125. Hess, B., et al., *LINCS: a linear constraint solver for molecular simulations*. *Journal of computational chemistry*, 1997. **18**(12): p. 1463-1472.
126. Parrinello, M. and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*. *Journal of Applied physics*, 1981. **52**(12): p. 7182-7190.
127. Wu, D. and D.A. Kofke, *Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations*. *The Journal of chemical physics*, 2005. **123**(8): p. 084109.
128. de Vaca Cabeza, I., et al., *Robust FEP Protocols for Creating Molecules in Solution*. *Journal of chemical theory and computation*, 2019.