NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCE

DEPARTMENT OF CHEMISTRY

INTERDISCIPLINARY POSTGRADUATE PROGRAM "CHEMISTRY"

SPECIALIZATION "PHYSICAL CHEMISTRY"

MASTER THESIS

Statistical Mechanical Study of β-N-methylamino-L-alanine and its carbamate adducts as potential inhibitors of the AMPA glutamate receptor using Molecular Dynamics Simulations

DIAKOGIANNAKI ISIDORA

CHEMIST

ATHENS

JUNE 2021

MASTER THESIS

Statistical Mechanical Study of β-N-methylamino-L-alanine and its carbamate adducts as potential inhibitors of the AMPA glutamate receptor using Molecular Dynamics Simulations

DIAKOGIANNAKI ISIDORA

S.N.: 191901

SUPERVISOR: Dr. Andreas Koutselos, Professor
National and Kapodistrian University of Athens

EXAMINATION COMITTEE

Dr. Andreas Koutselos, Professor
National and Kapodistrian University of Athens
(NKUA)
Dr. Zoe Cournia, Researcher - Associate Professor Level,
Biomedical Research Foundation of the Academy of Athens
(BRFAA)
Dr. Dimitra Tzeli, Associate Professor
National and Kapodistrian University of Athens
(NKUA)

25/06/2021

ΕΡΕΥΝΗΤΙΚΗ ΕΡΓΑΣΙΑ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Στατιστική Μηχανική Μελέτη μέσω Μοριακής Δυναμικής Προσομοίωσης της δράσης της β-N-μεθυλάμινο-L-αλανίνης και των καρβαμιδικών παραγώγων της ως αναστολείς του γλουταμινικού υποδοχέα AMPA

ΔΙΑΚΟΓΙΑΝΝΑΚΗ ΙΣΙΔΩΡΑ

Α.Μ.: 191901

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Δρ. Ανδρέας Κούτσελος, Καθηγητής
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Δρ. Ανδρέας Κούτσελος, Καθηγητής,
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
(ΕΚΠΑ)
Δρ. Ζωή Κούρνια, Κύρια Ερευνήτρια – Αναπληρώτρια Καθηγήτρια,
Ίδρυμα Ιατροβιολογικών Ερευνών, Ακαδημίας Αθηνών
(ΙΙΒΕΑΑ)
Δρ. Δήμητρα Τζέλη, Αναπληρώτρια Καθηγήτρια,
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών
(ΕΚΠΑ)

ΗΜΕΡΟΜΗΝΙΑ ΕΞΕΤΑΣΗΣ 25/06/2021

ABSTRACT

BMAA is a neurotoxic non-protein amino acid, which may reach the human body through the food chain. When BMAA interacts with bicarbonate in the human body, carbamate adducts are produced, which share high structural similarity with the neurotransmitter glutamate. It is believed that BMAA and its carbamate adducts bind in the glutamate binding site of glutamate receptors (GluR) causing the neurodegenerative effects of diseases such as Amyotrophic Lateral Sclerosis and Parkinson's Disease. The mechanism of BMAA action and its carbamate adducts bound to GluR has not been yet elucidated. In this work, we investigate the binding affinity of BMAA and its carbamate adducts to glutamate receptors in comparison to the natural agonist, glutamate, in order to understand whether these can act as glutamate inhibitors. Initially, we perform Molecular Dynamics (MD) simulations of BMAA and its carbamate adducts bound to the AMPA glutamate receptor in order to examine the stability of the ligands in the S1/S2 ligand-binding domain of the AMPA receptor. In addition, we perform MD simulations coupled with Free Energy Perturbation calculations to calculate the difference in the free energy of binding of BMAA to glutamate receptors compared to glutamate. Our findings indicate that one of the adducts of BMAA, β-carbamate adduct, has enhanced stability in the binding site of the AMPA receptor compared to the natural agonist, glutamate. In addition, the results from MD/FEP calculations reveal that glutamate and β-carbamate adduct of BMAA have comparable binding affinity in the AMPA glutamate receptor.

SUBJECT AREA: Computational Chemistry

KEYWORDS: BMAA, AMPA glutamate receptor, neurodegenerative diseases, MD simulations, FEP

ΠΕΡΙΛΗΨΗ

Το BMAA είναι ένα μη πρωτεϊνικό αμινοξύ που μπορεί να καταλήξει στον ανθρώπινο οργανισμό μέσω της τροφικής αλυσίδας. Όταν το BMAA αλληλεπιδρά με το διττανθρακικό ιόν στον άνθρωπο παράγονται καρβαμιδικά παράγωγα που έχουν παρόμοια δομή με το νευροδιαβιβαστή γλουταμινικό ιόν. Πιστεύεται ότι το BMAA και τα καρβαμιδικά παράγωγά του προσδένονται στο σημείο πρόσδεσης γλουταμινικών υποδοχέων προκαλώντας νευρικά συμπτώματα ασθενειών όπως το Πάρκινσον και η Αμυοτροφική Πλευρική Σκλήρυνση. Ο μηχανισμός δράσης του BMAA και των καρβαμιδικών παραγώγων του προσδεμένων σε γλουταμινικούς υποδοχείς δεν έχει μελετηθεί έως τώρα. Σε αυτήν την εργασία, ερευνούμε τη δεσμευτική συγγένεια του BMAA και των καρβαμιδικών παραγώγων του σε γλουταμινικούς υποδοχείς σε σύγκριση με τον φυσιολογικό αγωνιστή, το γλουταμινικό ιόν, προκειμένου να γίνει κατανοητό εάν αυτά μπορούν να δράσουν ως αναστολείς γλουταμινικών υποδοχέων. Αρχικά, εφαρμόζουμε προσομοιώσεις Μοριακής Δυναμικής του BMAA και των καρβαμιδικών παραγώγων του στον γλουταμινικό υποδοχέα AMPA, έτσι ώστε να μελετηθεί η σταθερότητα των μορίων στην S1/S2 περιοχή πρόσδεσης του υποδοχέα AMPA. Επιπλέον, χρησιμοποιούμε προσομοιώσεις Μοριακής Δυναμικής σε συνδυασμό με υπολογισμούς Διατάραξης Ελεύθερης Ενέργειας προκειμένου να υπολογιστεί η διαφορά της ελεύθερης ενέργειας πρόσδεσης του BMAA σε σχέση με το γλουταμινικό ιόν στον AMPA υποδοχέα. Τα αποτελέσματα δείχνουν ότι ένα από τα παράγωγα του BMAA, το β-καρβαμιδικό ιόν, παρουσιάζει ιδιαίτερη σταθερότητα στον υποδοχέα AMPA σε σύγκριση με τον φυσιολογικό αγωνιστή, το γλουταμινικό ιόν. Επιπλέον, τα αποτελέσματα από τους υπολογισμούς FEP υποδεικνύουν ότι το γλουταμινικό ιόν και το β-καρβαμιδικό παράγωγο του BMAA έχουν συγκρίσιμη συγγένεια πρόσδεσης στον υποδοχέα AMPA.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Υπολογιστική Χημεία

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: BMAA, γλουταμινικός υποδοχέας AMPA, νευροεκφυλιστικές ασθένειες, Μοριακή Δυναμική, Διατάραξη Ελεύθερης Ενέργειας

To my family

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude and appreciation to those who have helped me complete this thesis. First and foremost, I would like to thank my supervisors, Dr. Zoe Cournia, Researcher - Associate Professor Level, at the Biomedical Research Foundation of the Academy of Athens (BRFAA) and Dr. Andreas Koutselos, Professor at the National and Kapodistrian University of Athens for their leadership, encouragement, and thoughtful advice throughout all these months. All the experiences have been immensely valuable for me and my academic steps.

I would like to take this opportunity and express the deepest appreciation to Dr. Michail Papadourakis, member of the Cournia lab, for his useful advices and guidance in setting up and running Molecular Dynamics simulations as well as Free Energy Perturbation calculations.

Moreover, I would like to acknowledge my fellow lab mates Dr. Ioannis Andreadelis, Alexios Chatzigoulas, Dr. Sofia Kiriakidi and Zoe Diamanti for the pleasant working environment as well as all the constructive conversations and suggestions throughout my thesis.

In addition, I would like to thank my family and all my friends for supporting me spiritually throughout this thesis.

Last, I would like to acknowledge the Greek Research & Technology Network (GRNET) in the National HPC facility ARIS, for awarding the project with computational time, as well as the Biomedical Research Foundation Academy of Athens for the equipment that provided to complete my thesis.

Contents

LIST OF FIGURES

LIST OF TABLES

PREFACE

The master thesis “Statistical Mechanical Study of β-N-methylamino-L-alanine and its carbamate adducts as potential inhibitors of the AMPA glutamate receptor using Molecular Dynamics Simulations” has been conducted in collaboration with the Biomedical Research Foundation Academy of Athens for the completion of the Postgraduate Program “Physical Chemistry”, Department of Chemistry at the National and Kapodistrian University of Athens.

In the first chapter β-N-methylamino-L-alanine (BMAA) is introduced, as well as the mechanism of BMAA neurotoxic activity, which it is believed that provokes the neurotoxic effects of diseases such as Amyotrophic Lateral Sclerosis and Parkinson.

Subsequently, in the second chapter, the theoretical background of the methods used in the present thesis are described. At the beginning, Molecular Dynamics (MD) simulations theory is introduced, followed by the Free Energy Perturbation (FEP) approach coupled with MD simulations.

In the third chapter, are presented the results that were obtained. We start by the description of the procedure of setting up the systems, the parameters used, along with the results from both MD simulations and FEP calculations coupled with MD simulations.

Finally, the conclusions arising from this work are presented followed by future perspectives of this project.

1 INTRODUCTION

1.1 BMAA: a neurotoxic non-protein amino acid linked to neurodegenerative diseases

β-N-methylamino-L-alanine or (S)-2-amino-3-(methylamino)-propanoic acid (IUPAC name), abbreviated as BMAA, is a non-proteinogenic amino acid (Figure 1b) [1], which indicates that it is not involved in the translational process that assemble proteins. There are about 140 known non-proteinogenic amino acids with multiple analogues or variants and BMAA is an alanine variant non-proteinogenic amino acid (Figure 1c). Many of these compounds serve important functions in organisms, e.g., the D-amino acids, which are found on bacterial cell walls, they can be incorporated into proteins. However, some non-proteinogenic amino acids like BMAA have been associated with toxicity [3,4].

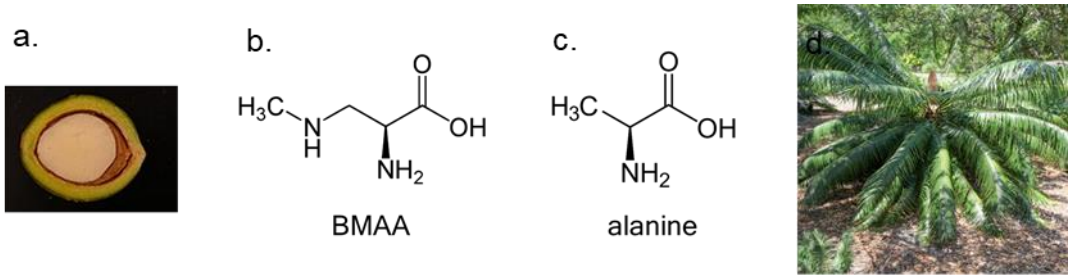

Figure 1. a) cycad seed, b) structure of β-N-methylamino-L-alanine, c) structure of alanine amino acid, d) cycad tree

It was first in the Chamorro population of the western Pacific islands of Guam and Rota, that an unusually high incidence of neurological effects was observed in the population, about 100 times higher than anywhere else in the world [5]. Chamorro patient symptoms were characterized by slowed movements, tremor, and rigidity, which are signs of parkinsonism as well as progressive limb weakness, which common with amyotrophic lateral sclerosis (ALS). Other patients developed cognitive dysfunction typical of dementia found in patients with Alzheimer's. In addition, this incident was observed among the natives in Irian Jaya region and also

amongst the natives of the Kii Peninsula of Japan [6,7]. The common factor in these populations was the regular consumption of cycad seeds (Figure 1a, 1d) containing BMAA either for food or for medicinal purposes and it was therefore proposed by epidemiological studies that the diseases represent neurological damage caused by the toxin in cycad seeds.

1.2 Pathways of BMAA for human exposure

BMAA is produced by a range of ecologically diverse phytoplankton groups such as cyanobacteria (including the common genera Anabaena and Nostoc), diatoms, and dinoflagellates. It has been demonstrated to bioaccumulate in various aquatic ecosystems, ranging from the subtropical marine aquatic ecosystems of Florida, USA, to the temperate brackish ecosystem of the Baltic Sea, and from the limnic aquatic ecosystems of the subtropical Lake Taihu, China, to the temperate Lake Finjasjön in Sweden. It has been demonstrated not only in aquatic ecosystems, but also in terrestrial ecosystems and specifically in cycad trees, on which roots live cyanobacteria of the genus Nostoc [2].

Cycad seeds were a common food source included in flour or eaten by animals consumed by humans in Guam. Thus, BMAA may end up in humans through the food chain not only in terrestrial ecosystems, but also through food chain in any aquatic system. Specifically, the Chamorros were using cycad seeds as food source, but they were aware that the cycad seeds were toxic and generally washed the dried seed ‘chips’ with unheated water multiple times over several days, a procedure that should remove most of the toxic BMAA. It is true that this amount of BMAA was negligible to provoke such neurological effects, however it was also the flying foxes and bats, which feed of cycad seeds and are consumed by humans. Bats can triple their weight in just one evening of feeding on cycad seeds and the natives would regularly feast on when cooked in coconut cream and fur, while in 1970’s there was near disappearance of the Guam flying fox population due to over-hunting (Figure 2). In two other locations, Irian Jaya and the Kii Peninsula of Japan, raw seeds of the Cycas revolute Thunb are used by natives as topical medicines for cuts, hemorrhoids, and

open sores [7]. Furthermore, BMAA is transferred from cyanobacteria via zooplankton to organisms at higher trophic levels (fishes and humans, Figure 2) inhabiting both pelagic and benthic water masses, thereby representing different niches in a cyanobacterial based food web [8].

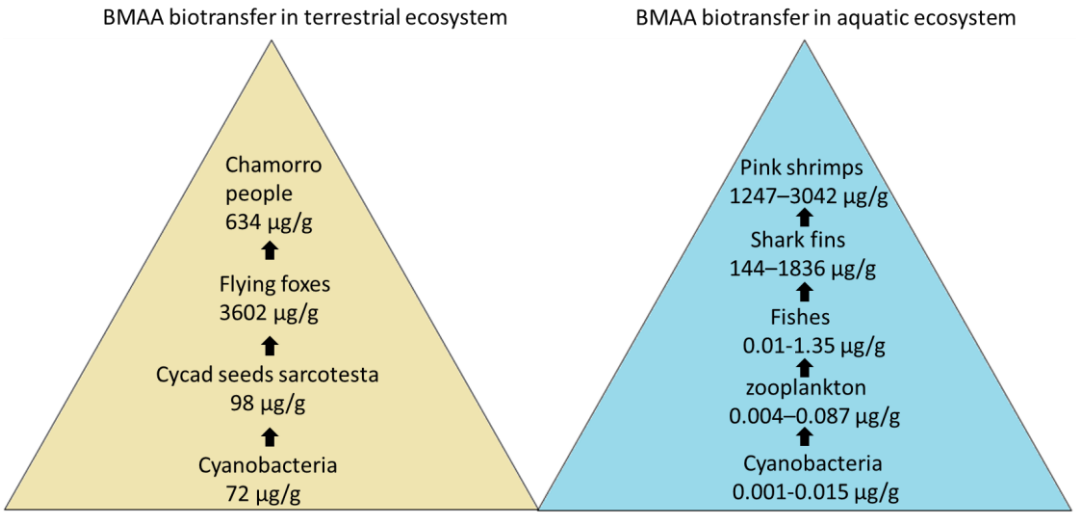

Figure 2. BMAA biotransfer in terrestrial and aquatic ecosystem. BMAA levels are expressed as µg BMAA/g dry weight [8].

1.3 The mechanism of BMAA neurotoxic activity

Investigations concerning the dependence of BMAA in presence of bicarbonate (HCO_3^-), which is produced from the interaction of carbon diocide with water, revealed that neurotoxicity of BMAA depended on the presence of bicarbonate, in murine cortical cells. Specifically, BMAA was nontoxic in physiological salt solution, but became toxic in the presence of bicarbonate. Bicarbonate unmasked BMAA toxicity by presenting morphological changes in neurons. The explanation is that bicarbonate interacts directly with BMAA to produce a structure suitable for glutamate receptor activation [9,10] In addition, BMAA has been detected in post-mortem brain and spinal cord tissues of Amyotrophic Lateral Sclerosis, Parkinson's and Alzheimer's patients [1].

Carbamylation reaction

The presence of bicarbonate ions at close to physiological pH, created ideal conditions for the formation of carbamate adducts, α-carbamate adduct and β-carbamate adduct of BMAA in a 86:14 ratio, respectively, through a reaction called carbamylation (Figure 3).

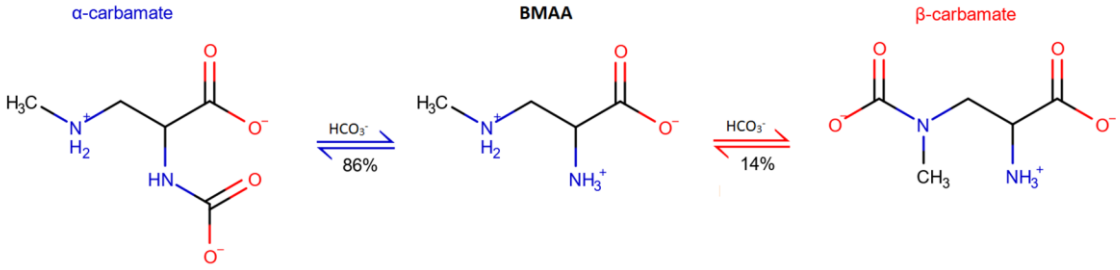

Figure 3. BMAA and its carbamate adducts.

Under physiological conditions (pH ca. 7), the amines are mostly protonated and the CO_2 exists in its hydrated form, HCO_3^-. Though these conditions are unfavorable for carbamylation, the favorable ΔG can be attributed to the pKa values of the protonated amines increasing the nucleophilicity. In this way, amines get deprotonated, and these conditions are suitable for carbamate formation. The presence of an electrondeficient HCO_3^- allows the nucleophilic amine to attack the electrophilic carbon leading to the carbamate formation, followed by stabilization of the carbamate products by noncovalent intermolecular interactions [11,12]

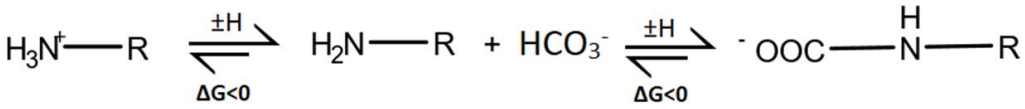

Figure 4. The formation of carbamylation reaction.

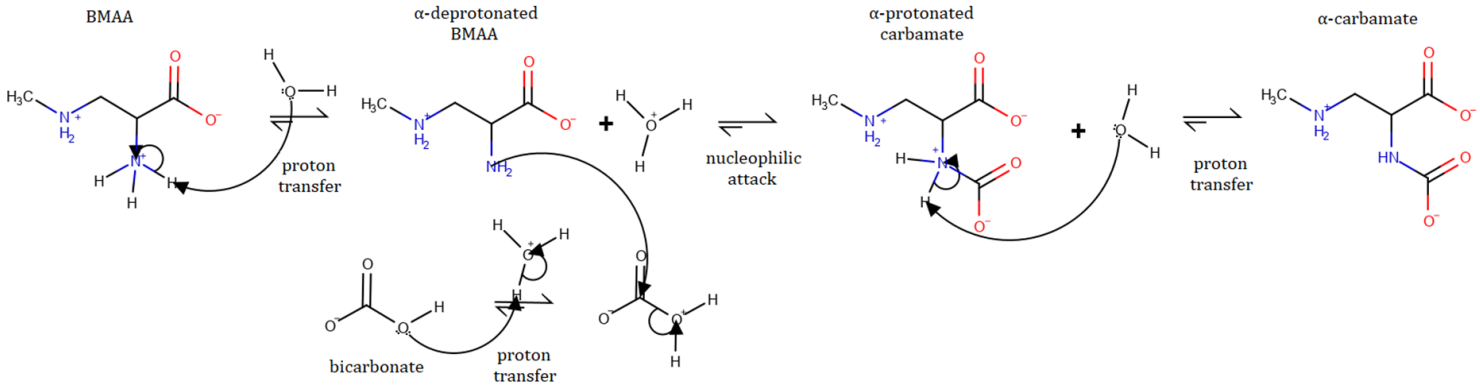

Figure 5. Example of the mechanism of carbamylation reaction to produce α-carbamate adduct of BMAA.

1.4 BMAA adducts with toxic activity and structural similarity with glutamate

The formation of the carbamate adducts of BMAA in the presence of HCO_3^- share a high structural similarity with glutamate. Glutamate is an important neurotransmitter, of which action on glutamate receptors is necessary for the transduction of signals in the nervous system (Figure 6). This, leads to a possible explanation of the mechanism by which BMAA may bind to glutamate receptors. [12].

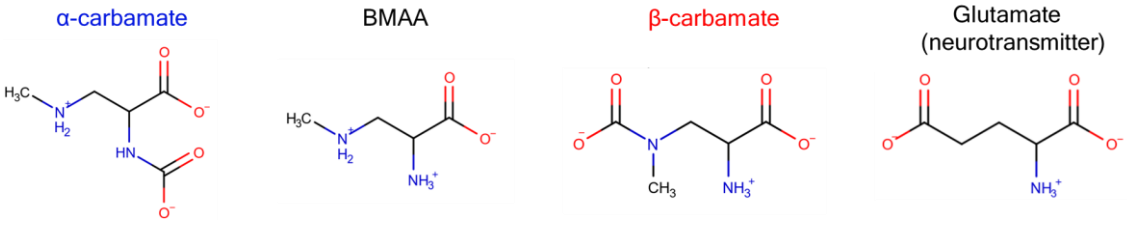

Figure 6. Structures of α-carbamate, BMAA, β-carbamate and glutamate.

Isomers of BMAA

The team of Jiang by performing a database search (Scifinder, Pubmend) were able to find isomers of BMAA. From the 260 theoretical compounds they ended up to 7 molecules that satisfy the criteria for chemical stability. Here are presented three of them, which have been further studied (Figure 7) [13]. DAB, is believed to be a hepatotoxic and neurotoxic non-protein amino acid that is found in many prokaryotic and eukaryotic organisms and it has caused neurological effects in animals [12,

13]. AEG induces oxidate stress in metabotropic glutamate receptors in cells [14], and BAMA has been found in mussels and oysters, however its toxicity hasn't been further studied [13].

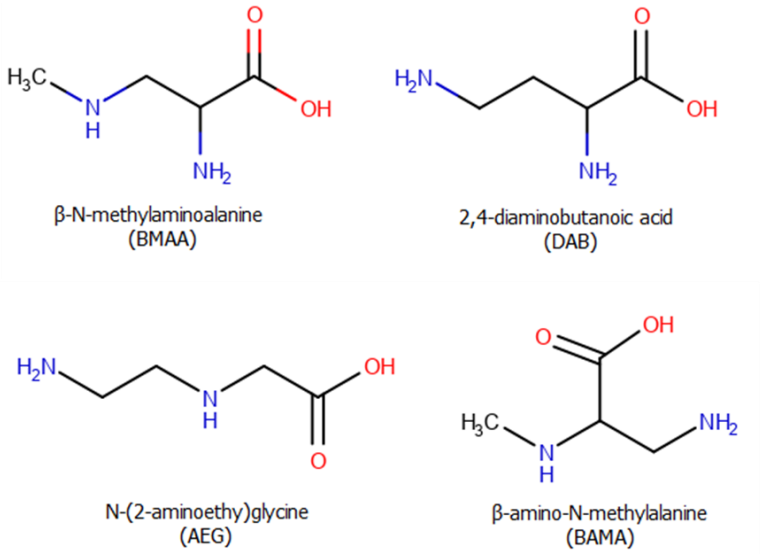

Figure 7. Isomers of BMAA.

1.5 Glutamate receptors and potential inhibition by BMAA and its adducts

The activation of glutamate receptors is important in the development and function of the nervous system, while they are also essential in memory and learning. Dysfunction of glutamate receptors leads to exitotoxic cell death [15]. Their natural agonist is glutamate but they can be activated also by structurally similar substances, such as quinoxalinedione derivatives [16]. Glutamate receptors are subdivided into two categories, the ionotropic receptors and metabotropic receptors (Figure 8).

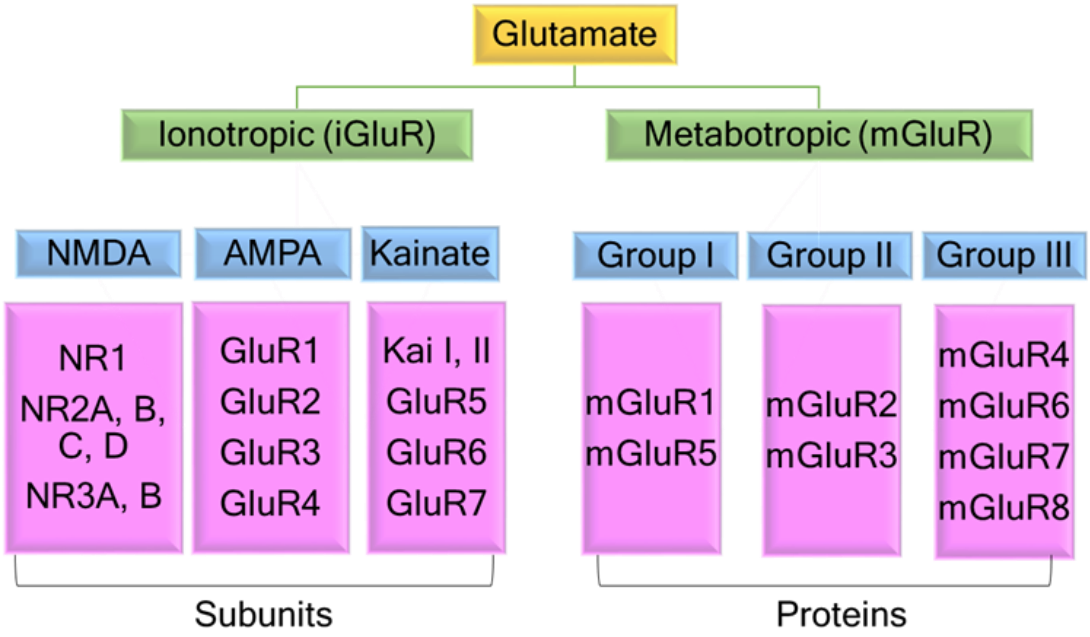

Figure 8. Subunits and proteins of glutamate receptors.

1.5.1 Ionotropic receptors

Ionotropic receptors (iGluR) have multiple subunits, each of them having an amino terminal domain (ATD), a ligand binding domain (LBD) and a transmembrane domain (TMD) (Figure 9). Examples of iGluRs include NMDA, AMPA and Kainate receptors. In addition, iGluRs have a membrane-spanning domain that forms ion channel. Particularly, NMDA receptors have calcium conductivity, while AMPA and Kainate receptors primarily mediate sodium influx [17].

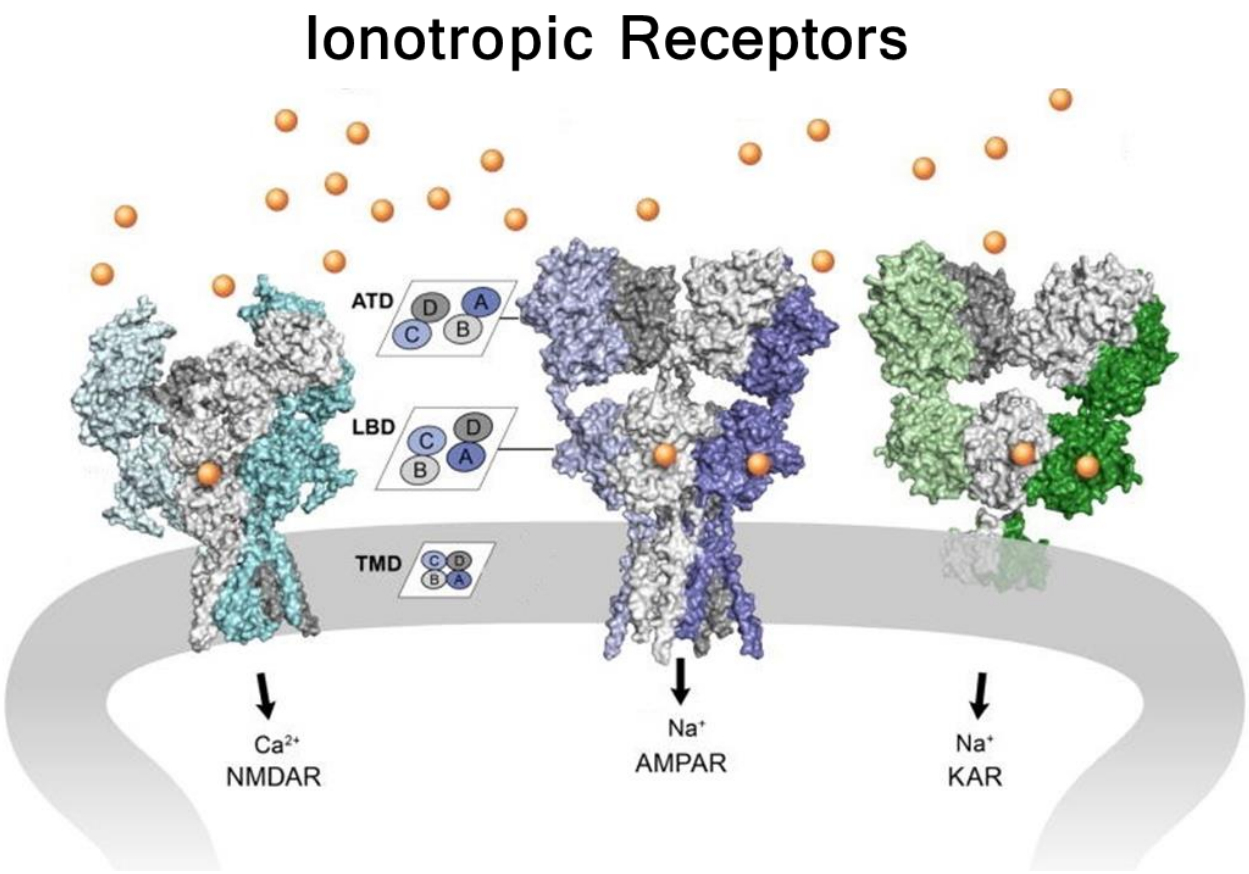

Figure 9. Ionotropic glutamate receptors, NMDA, AMPA and Kainate.

1.5.2 Metabotropic receptors

Metabotropic receptors (mGluR) have eight proteins and in contrast with iGluR, they have monomers instead of multiple subunits. They have also an ATD, a LBD and a TMD (Figure 10), however, mGluRs do not form ion channels, but they bind with g-protein and they form signals which are various second-messenger systems [18].

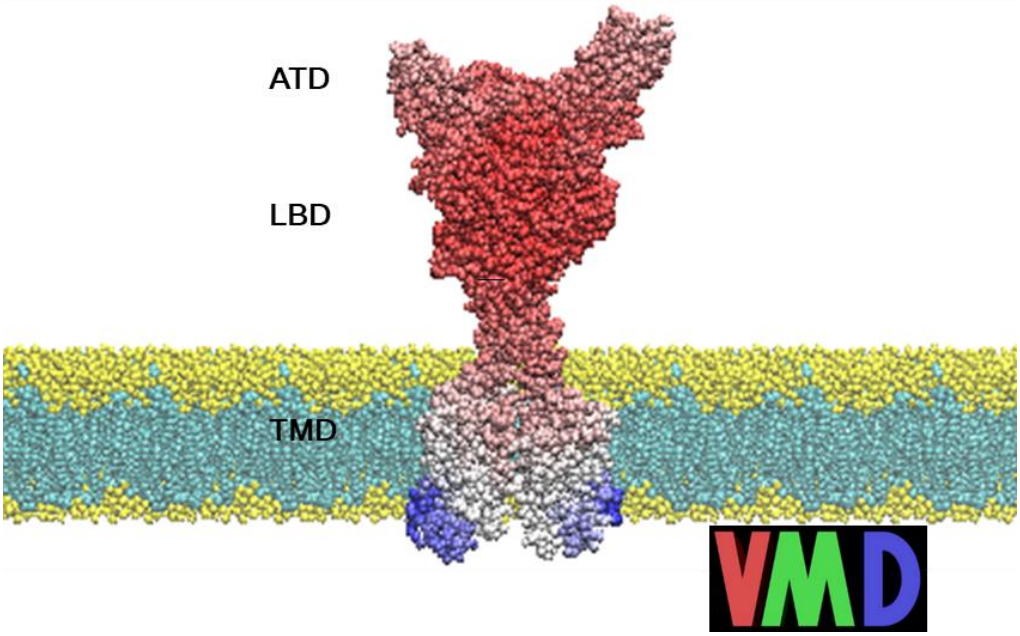

Figure 10. Domains of metabotropic receptors.

1.6 Study objectives

As previously mentioned, there have been investigations reporting the structural similarity of BMAA and its adducts with the neurotransmitter glutamate, the toxic activity of BMAA, suggesting that BMAA and its carbamate adducts bind in glutamate's binding site [3,4,10,12]. However, the role of BMAA and its carbamates adducts bound to GluR has not been reported. The aim of this thesis is to investigate the binding affinity of BMAA and its carbamates to GluR in comparison with the natural agonist, glutamate, in order to understand whether these can act as glutamate inhibitors. In the eventuality that they have better affinity than glutamate, it would give an explanation of whether BMAA provoke dysfunction on neurons, leading to neurodegenerative diseases. Accordingly, our research will focus on the following objectives in increasing order of complexity. First, we will study the stability of BMAA and its adducts in the AMPA receptor using long atomistic Molecular Dynamics simulations. Subsequently, we

will calculate the difference of the free energy of binding between glutamate and the molecule that will present the highest stability in the receptor of interest. For this purpose, we will use free energy perturbation (FEP) calculations coupled with MD simulations.

2 METHODOLOGY

2.1 Computer-aided drug design

Computer aided drug design is a powerful technique that uses approximations and restrictions in order to find promising drug candidates. This approach has become an essential part in the drug design process due to its ability to accelerate drug discovery by utilizing existing knowledge on receptor-ligand interactions, energy and structural optimization. With this technique, bioactive molecules with desired properties can be discovered [19].

In computer-aided drug design, the most commonly used methods is the structure-based drug design and the ligand-based drug design. Structure-based drug design methods analyze the conformation of the macromolecular target and finds significant binding sites and interactions for this target [20]. Ligand-based drug design methods focus on existing ligands for the target, which are used as guides in discovering new drugs with improved properties.

Nowadays, there are multiple computational instruments provided to assist drug design. The main challenge of is to use them wisely in order to provide a useful output corresponding to the current needs in a reasonable time [19].

2.1.1 Molecular Docking

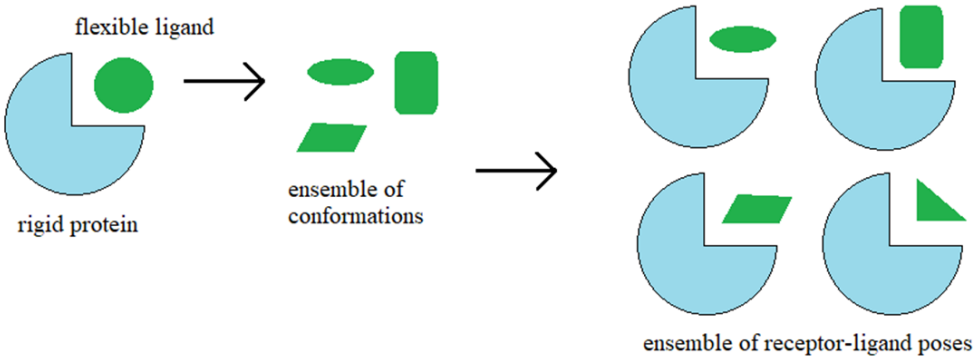

Figure 11. Illustration of Molecular Docking procedure.

Molecular docking is a tool to find the most favorable position of one molecule in a receptor when they are brought together to form a complex. It plays an important role in structure-based drug design and in the research of protein structures and functions. Docking is achieved through two interrelated steps. First, by sampling conformations of the ligand in the active site of the protein (most of the times, it is assumed that the receptor is rigid, while the ligand if flexible). As a result, an ensemble of conformations for the ligand is generated. In the following step, it is evaluated, which of this ensemble of conformations fits better energetically and geometrically to the receptor of interest by giving a score using scoring functions. This score is an approximation of ligand binding free energy to receptors.

$$\Delta G_{bind} = \Delta G_{solv.} + \Delta G_{conf.} + \Delta G_{int.} + \Delta G_{rot.} + \Delta G_{t/r} + \Delta G_{vib.} \quad (2.1.)$$

where, ΔG_{solv} is the contribution of the solvation free energy. ΔG_{conf} is the contribution of conformational changes in the ligand and in the protein. ΔG_{int} is the free energy, due to ligand-protein interactions. ΔG_{rot} (entropic contribution) is the free energy loss arising from freezing internal rotations of protein and the ligand. $\Delta G_{t/r}$ is the penalty for the loss of translational and rotational free energy caused by the association of two-bodies (ligand and receptor) to create a single body (complex). ΔG_{vib} is the free energy of the system due to changes in vibrational modes. Some of the ligand conformations are rejected because of high-energy clashes with the protein [21,22]

2.2 Molecular Dynamics (MD) simulations

In this section, the scope of MD is highlighted, as well as how simulations are performed and how macroscopic properties for the systems can be calculated through statistical mechanics. The study of molecular systems' properties is achieved with the aid of computer models. The size of systems, as well as the time scales in which interesting phenomena occur differs and therefore different techniques are used (Figure 12). For instance, Quantum Mechanics (QM) describe the electrons of a system [23], Brownian Dynamics (BD) have been widely used for microsized particle simulation in

the Cartesian space [24] and MD uses an empirical potential to calculate the energy of a system based on the nuclear positions only. MD simulations numerically integrate Newton's equations of motion in order to describe the dynamics of the system and to generate information about the system on the microscopic level. Subsequently, the microscopic information is linked to the macroscopic observables through statistical mechanics.

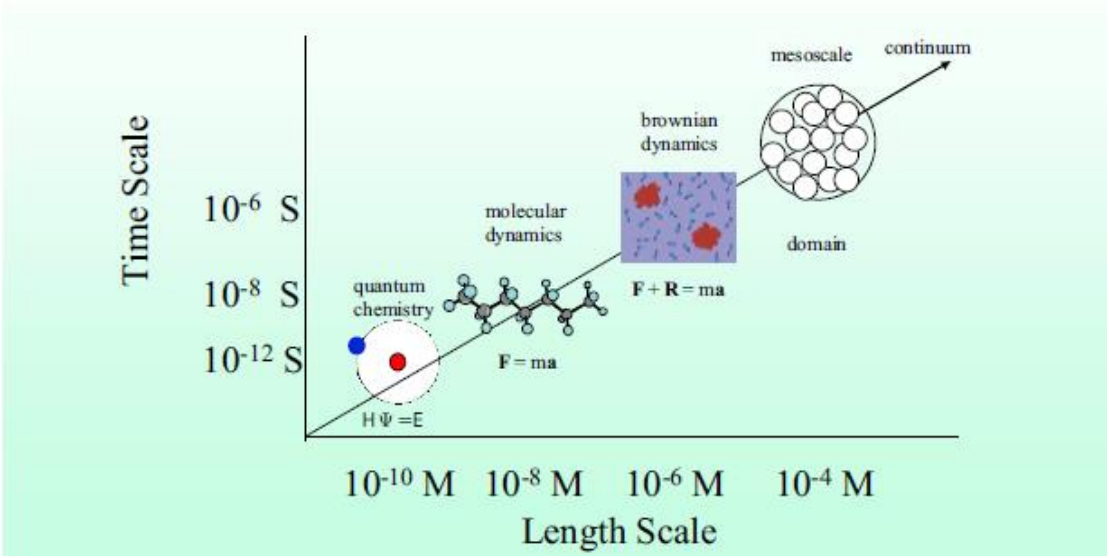

Figure 12. Graphical representation of the different time and length scales achieved by different methods in molecular modeling. Adopted from Cournia [27]

MD is based on different approximations, such as the Born-Oppenheimer approximation and the potential energy function. According to Born-Oppenheimer approximation, electron motion in molecules is so much faster that nuclear motion and therefore can be considered to be fixed. In addition, the empirical potential energy function describes the intra- and inter-molecular interactions that occur between the nuclei and it is further discussed in the next section [25, 26].

2.2.1 The Empirical Potential Energy Function (force field)

The energy of a system is represented by the Hamiltonian:

$$H = T + V \quad (2.2.)$$

where T is the Kinetic Energy and V is the Dynamical (potential) Energy. In classical mechanics, the kinetic energy of a system can be described by $\frac{p^2}{2m}$. However, the description of the potential energy term, V, is more complicated, due to the fact that the intermolecular interactions need to be taken into consideration. Therefore, the potential energy, $V(r)$, can be defined as the non-bonded and bonded energy terms of the system [28].

$$V(r) = V_{bonded} + V_{non-bonded} \quad (2.3.)$$

the bonded terms involve the simple covalent binding in addition to the complex hybridization and π-orbital effects, while the non-bonded interactions are described by the intermolecular van der Waals and electrostatic interactions. In order to simulate the bonded and non-bonded energy terms of a molecular system in a classical mechanic's framework, an empirical potential energy function that describes these interactions is introduced with a set of equations, also known as force field [29, 30]. The different equations for bonded and non-bonded interactions that comprise the potential energy function are presented below.

the V_{bonded} term is the summation of the terms that correspond to the types of the atom movement. This includes terms for bonds, angles, proper dihedrals and improper dihedrals (Figure 13):

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals} + V_{impropers} \quad (2.4.)$$

Non-bonded interactions involve two types of interactions: the van der Waals interaction energy and the electrostatic interaction energy:

$$V_{non-bonded} = V_{vdw} + V_{electrostatic} \text{ (2.5.)}$$

These interactions are the most computationally demanding due to the fact that they include long-range interactions between the atoms in the system.

BONDED INTERACTIONS

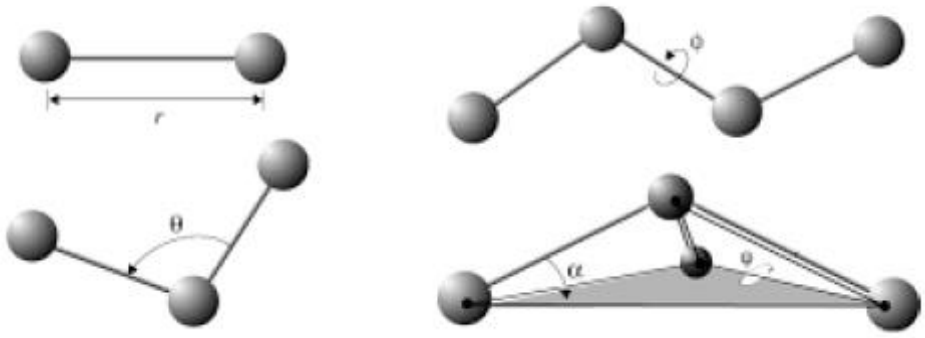

Figure 13. Symbolic representation of the bonded interactions: bond stretching r (upper left), bond angle bending θ (bottom left), proper dihedral φ (upper right) and improper dihedral ψ (bottom right) and the small out-of-plane angle α. Adopted from Adcock and McCammon [26].

Bond Stretching

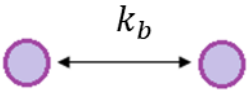

Figure 14. Definition of bond stretching between two atoms.

In molecules, atoms are connected with covalent bonds which vibrate (Figure 14). The bond vibrations are described by a harmonic potential according to Hooke's law:

$$F = -kx \text{ (2.6.)}$$

Taking into consideration that the force can be written as the gradient of the potential energy of the system, we conclude to this formula:

$$F = -kx = -\nabla V \text{ (2.7.)},$$

And hence the potential of bond stretching becomes:

$$V = -kx^2, \text{ (2.8.)}$$

which is expressed to describe the force constant k_b of the vibration of the spring and the distance between the two atoms, b, from the average distance, b_0, of the two bonded atoms:

$$V_{bonds} = k_b(b - b_0)^2 \text{ (2.9.)}$$

Bond Angle Bending

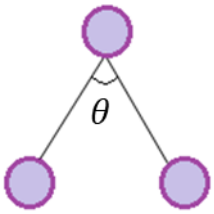

Figure 15. Definition of bond bending between three atoms.

The angle bending terms describe the force originating from the deformation of the valence angles between three covalently bonded atoms (Figure 15). This conformational change can be described again using Hooke's law (eq. 2.8) ensuing into an angle bending term, which is expressed as:

$$V_{angles} = k_\theta(\theta - \theta_0)^2 \text{ (2.10.)}$$

where k_θ is a force constant, θ is the angle between three atoms and θ_0 is the reference angle.

Dihedral Torsion Term

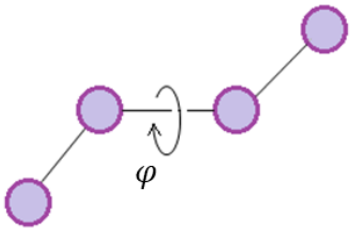

Figure 16. Definition of dihedral torsion between four atoms.

Torsional terms describe the barriers to rotations existing between four covalently bonded atoms. They are weaker than the bond stretching and angle bending terms, however, the need for the proper torsional potential arises, as bond-stretching or angle-bending potentials cannot describe the energetics of rotatable bonds, since they appear in four covalently bonded atoms (Figure 16).

The rotation along a covalent bond is characterized by periodicity, that's why the torsional potentials can be outlined by a cosine function as presented below:

$$V_{dihedrals} = k_{\varphi}(1 - \cos(n\varphi - \delta) \quad (2.11.)$$

where $k\varphi$ is a force constant, n is the periodicity, φ is the angle between the planes formed by the first and the last three of the four atoms, and δ sets the minimum energy angle.

Improper Dihedral Term

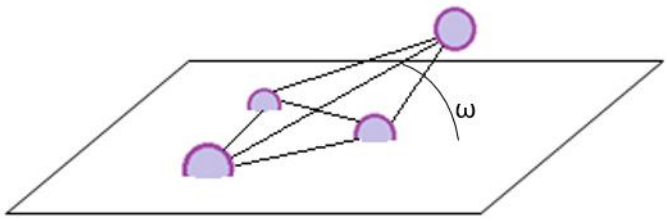

Figure 17. The improper dihedral.

The improper dihedral term is designed both to maintain chirality about a tetrahedral heavy atom and to maintain planarity about certain atoms. The improper angle is defined as the angle between the plane formed by the central atom and two peripheral atoms and the plane formed by the peripheral atoms (Figure 17). The improper potential is expressed by a harmonic function:

$$V_{impropers} = k_{\omega}(\omega - \omega_0)^2 \quad (2.12.)$$

where ω is the angle between the planes as described above.

NON-BONDED INTERACTIONS

Van Der Waals Interactions

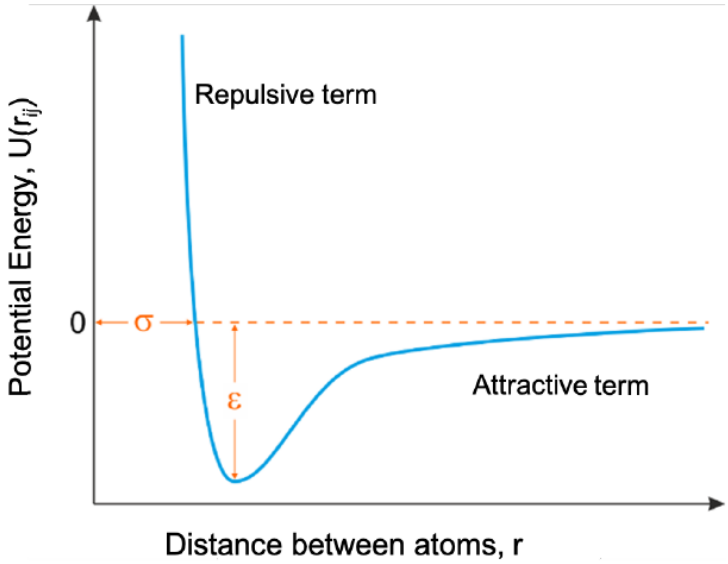

Figure 18. The Lennard-Jones potential. σ is the collision parameter and ε the well depth. Adopted from Chemistry Dictionary & Glossary [31].

The van der Waals interaction occurs on atoms in close proximity. It is strongly repulsive at short range and attractive when there are vacillations in the charge dispersion in the electron clouds [32]. It is described by a Lennard-Jones potential:

$$4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.13.)$$

where $r_{i,j}$ is the distance between two atoms. It is parametrized by σ, which is the collision parameter (the separation for which the energy is zero) and the depth of the potential well, ε (Figure 18).

Electrostatic Interactions

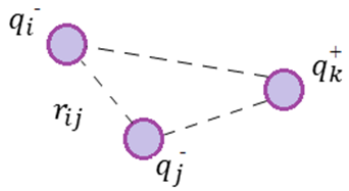

Figure 19. Schematic representation of electrostatic interactions between three charged atoms q_i, q_k, q_j.

Two non-bonded atoms having a charge or partial charge may interact electrostatically with each other. (Figure 19) [33]. This long-distance electrostatic interaction between two atoms is described by Coulomb's law:

$$V_{electrostatic}=\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.14.)$$

Equation of the potential energy function

Finally, the equation of the potential energy function that represents the force field is:

$$V(r)=\sum_{bonds}k_b(b-b_0)^2+\sum_{angles}k_\theta(\theta-\theta_0)^2+\sum_{dihedrals}k_\varphi(1+\cos[n\varphi-\delta])+\sum_{impropers}k_\omega(\omega-\omega_0)^2+\sum_{electrostatic}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}+\sum_{vdw}4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (2.15.)$$

In MD simulations the long-range interactions are the most time-consuming ones. Assuming having a box consisting of 500 molecules, the range of these forces is greater than half of the box length. The most commonly used method in MD simulations, in order to correctly account of the electrostatic interactions, is the Ewald sum, which includes the interaction of an ion or molecule with all its periodic images [34,35].

More precisely, simulations systems with periodic boundary conditions can be a good approximation to the behavior of a small subsystem in a larger bulk phase. During the simulation it is assumed that as a molecule leaves the central box, its periodic image will enter the central box from the other side in exactly the same way (Figure 20). At the boundary of the central box there are no walls and no surface molecules. In addition, it is undesirable for a single particle to interact with the same particle multiple times. In order to prevent this, a cut-off of many non-bonded interactions should be chosen that is less than half the length of the simulation box in any dimension. In this way, a natural lower limit to the size of a periodic simulation box is applied, as the box must be large enough to capture all of the most significant non-bonded interactions [36].

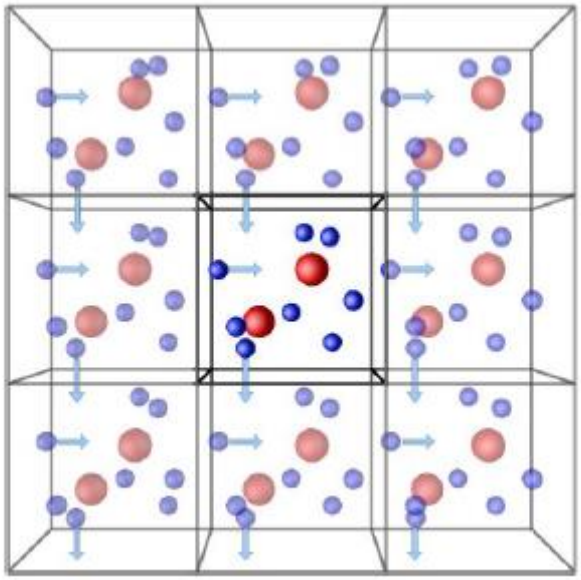

Figure 20. A periodic system illustrating the periodic boundary conditions used in MD simulations [52].

2.2.2 Energy Minimization

The energy landscape of a biomolecule includes a large number of minima. Prior to beginning a MD simulation, the system should undergo an energy minimization and reach a local energy minimum (Figure 21). Through this process, steric clashes between atoms, distorted bond angles or lengths, as well as any non-physical van der Waals contacts, will be relieved and it will be ensured that there is a realistic starting structure. Otherwise, these interactions may lead to an unbalanced simulation [37].

Given a function *f*, which depends on the variables x_1, x_2, …, x_n, a minimum or maximum of *f* is defined as a point where the first derivative of the function with respect to each of the variables is zero and the second derivative is positive:

$$\frac{\theta f}{\theta x_i} = 0, \ \frac{\theta^2 f}{\theta x_i^2} > 0 \quad \text{(2.16.)}$$

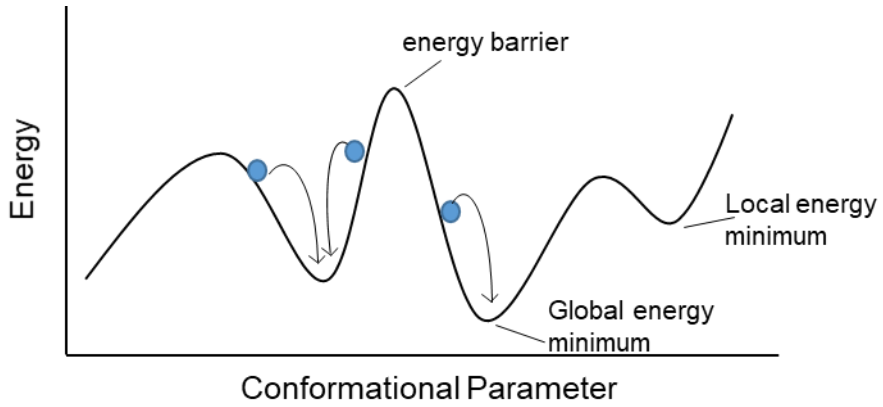

Figure 21. Representation of one-dimensional potential energy surface. Minimization methods move downhill to the nearest minimum.

Most common minimization algorithms use derivatives of the energy with respect to the coordinates in order to predict the location of the closest minimum [38]. The three main algorithms that are used are: Steepest Descent, Conjugate Gradient and Newton-Raphson algorithms. The appropriate selection of the method is based on the quantity of iterations essential to converge, as well as on the quantity of function assessments necessary per iteration [37].

Steepest Descent

The basis for the Steepest Descent method, also known as the gradient descent method, is the simple observation that a continuous function should decrease, at least initially, if one takes a step along the direction of the negative gradient. The search direction is given by:

$$F(r) = -\frac{dV(r)}{dr} \quad (2.17.)$$

The step length is predicted with a line search, which utilizes the direction vector obtained from the first derivative of the potential function, and in this way, locates the optimal step size close to this direction vector. The moment this restricted minimum close to the direction of the derivative is settled, the step can be acquired. In order to find the best possible step size, the line search needs an enormous number of function assessments. The two main computational advantages of the steepest descent algorithm are the simplicity with which a computer algorithm can be implemented and the low storage requirements necessary [38,39].

Conjugate gradient and Newton-Raphson

Conjugate Gradient and Newton-Raphson algorithms are more efficient methods for energy minimization. The first, utilizes information from first derivatives to find the most optimal direction for a line search, while the Newton-Raphson method makes use of the second derivatives in combination with the first ones in order to accomplish the same result with higher accuracy. In addition, the Newton-Raphson takes advantage of the curvature to forecast the point where the gradient of the function will be changed into another direction [38].

$$V(r) = V(r_0) + V'^{(r_0)}(r - r_0) + \frac{V"(r_0)(r - r_0)^2}{2} + \cdots$$

$$\frac{dV}{dr} = 0;\ \frac{d^2V}{dr^2} > 0 \quad (2.18.)$$

2.2.3 MD Formalism

As mentioned previously, MD simulations are based on Newton's second law of motion:

$$F_i = m_i a_i = m_i \frac{\vartheta^2 r_i}{\vartheta t^2}, i = 1, \dots, N \text{ (2.19.)}$$

where m_i is the mass of each atom i and a_i is the acceleration on atom i. Forces can be expressed as the negative derivatives of the potential energy function $V(r_1, r_2, r_3, \dots, r_N)$:

$$F_i = -\frac{\vartheta V}{\vartheta r_i} \text{ (2.20.)}$$

By combining the two equations:

$$-\frac{\vartheta V}{\vartheta r_i} = m_i \frac{\vartheta^2 r_i}{\vartheta t^2}, \text{ (2.21.)}$$

and by solving this differential equation a trajectory of the system is obtained.

To begin a MD simulation, an initial structure of the system is required. Most often, the initial configuration is an X-ray crystallography or NMR structure that can be easily obtained by the Protein Data Bank (PDB). The initial coordinates are defined from these experimental structures, while velocities can be obtained from a distribution, e.g. Maxwell – Boltzmann distribution, which is expressed as:

$$\frac{dN}{N} = \sqrt{\frac{m}{2\pi k_b T}} e^{-mu^2/2k_b T} du \text{ (2.22.)}$$

where dN/N is the fraction of molecules moving at velocity u to $u+du$, m is the mass of the molecule, k_b is the Boltzmann constant and T is the temperature [40].

The equations are solved synchronously in short time steps, while the parameters such as the temperature and pressure maintain constant at the

desired values throughout the simulation. The different coordinates, related to time, represent the trajectories of the system.

MD simulations evolved into a mature technique that can be used effectively to understand macromolecular structure-to-function relationships and along with the continuous evolution of technology that provides faster and cheaper computers the time and the cost for performing a MD simulation has been significantly decreased. The time scale of solvated protein simulations is up to nanoseconds, however, simulations up to milliseconds have been also recorded [41].

Numerical methods for integrating Newton's equations of motion

The potential energy function, which was analyzed in 2.2.1, is not appropriate to provide an analytical solution in order to solve the equations of motion. Therefore, we conclude to treat the classical equations numerically. Many algorithms have been developed to integrate the equations of motion and some of those algorithms are presented below [42].

Generally, all the integration algorithms predict the positions, velocities and accelerations of the atoms using the Taylor series expansion:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}\alpha(t)\delta t^2 + \cdots$$
$$v(t+\delta t) = r(t) + \alpha(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \cdots$$
$$\alpha(t+\delta t) = \alpha(t) + b(t)\delta t + \cdots \qquad (2.23.)$$

where r is the position, v is the velocity (also can be expressed as the first derivative with respect to time) and α is the acceleration (second derivative with respect to time) [42]. δt is a short time interval and is defined by the fastest vibration of the system, which is the C-H bond (δt=1 fs= 10^{-15} s).

The total force remains constant during each time step and is equal to the vector sum of its interactions with other particles at time *t*. From the force, the acceleration, positions and velocities of the particles are calculated, and the available information at time t are used to calculate the system's positions and velocities at time t+δt.

Verlet Algorithm

The most common integration algorithm is the Verlet algorithm, which doesn't requisite the velocities, but requires a single force calculation per cycle and uses the following two equations:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}\alpha(t)\delta t^2 + \cdots$$
$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}\alpha(t)\delta t^2 + \cdots \quad (2.24.)$$

The sum of these two is the basic formalism of this algorithm:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \alpha(t)\delta t^2 \quad (2.25.)$$

The Verlet algorithm utilizes the positions (r) and accelerations (a) at time t, as the updated positions at time t + δt are calculated from the positions at time t – δt. The velocities can be calculated from the equation below:

$$V(t) = \frac{r(t+\delta t) - r(t-\delta t)}{2\delta t} \quad (2.26.)$$

The Verlet algorithm's advantage is that it requires small amount of storage memory; however, it may generate a lot of errors [42].

Leap-Frog Algorithm

The Leap–Frog algorithm, which is a variation of the Verlet algorithm, uses these two equations:

$$r(t+\delta t) = r(t) + v\left(t + \frac{1}{2}\delta t\right)$$
$$v\left(t + \tfrac{1}{2}\delta t\right) = v\left(t - \tfrac{1}{2}\delta t\right) + \alpha(t)\delta t \quad (2.27.)$$

In particular, it calculates the velocities at time t + 1/2δt, and these velocities are used to calculate the positions at time t + δt. This algorithm calculates the velocities directly, but not at the same time with the positions.

The advantage of this algorithm lies on the fact that it generates less errors than the Verlet algorithm. [42]

2.2.4 Introduction to Statistical Mechanics

MD simulations can be used to connect the microscopic properties to macroscopic properties using statistical ensembles *via* statistical mechanics. A statistical ensemble is a set of representative points in the 6N-dimensional phase space (3N for the spatial coordinates, *r*, and 3N for the momentum, *p*) [44]. The "ergodic hypothesis", which derives from the greek word έργον (energy) and οδός (path), states that the ensemble average and the time average is the same over an infinite period of time. Using means of a distribution function that describes the ensemble behavior, an average ensemble property (e.g. energy) can be calculated. The distribution function is the probability that at some time, under specific thermodynamic conditions, the system has a particular energy. As a result, by calculating the values of the property at different times t, the property average will be computed [43].

The partition function

To calculate the distribution function of an ensemble, the type of the ensemble should be specified. The canonical ensemble (NVT) is the set of the possible positions and momenta for all particles such that the number of particles, N, the volume, V, and the temperature, T, are fixed and it specifies variation of energy.

The probability function, P(Γ) in the NVT ensemble is described by the Gibbs's distribution:

$$P_i(\Gamma) = \frac{e^{-\beta H_i(r^N, p^N)}}{\sum_i e^{-\beta H_i(r^N, p^N)}} \quad (2.28.)$$

where r^N, p^N are a specific set of positions and momentum respectively, H_i is the Hamiltonian function that describes the total energy in state *i*, $\beta = {}^{1}/_{k_B T}$, T is the temperature and k_B is the Boltzmann constant (k_B=1.38 x 10^{-23} JK^{-1}) [45].

The denominator of Equation 2.28. is known as the partition function and usually is symbolized as Q_{NVT}.

$$Q_{NVT} = \sum_i e^{-\beta H_i(r^N, p^N)} \quad (2.29)$$

The partition function computes the number of the microstates accessible to the system at a specific temperature [46].

Due to the fact that the set of microstates is uncountable, the partition function is expressed as an integral:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int e^{-\beta H(r^N, p^N)} dr^N dp^N \quad (2.30.)$$

Where *h* is Plank's constant (h=6.6210^{-34}), $^1/_{h^{3N}}$ and is used in order to make the quantity dimensionless. $^1/_{N!}$ takes into consideration that the N particles are indistinguishable. Assuming that the kinetic (K) and potential energy (U) terms of the Hamiltonian are separable, Equation 2.30. will become:

$$H(r, p) = K + U$$

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int e^{-\beta U(r^N)} dr^N e^{-\beta K(p^N)} dp^N = Z_{NVT} Q_p \quad (2.31.)$$

Where Z_{NVT} is the configuration integral:

$$Z_{NVT} = \int e^{-\beta U(r^N)} dr^N \quad (2.32.)$$

And Q_p is the momentum integral, which can be analytically obtained using the following formula:

$$Q_p = \frac{V^N}{N!} \left(\frac{\sqrt{2\pi m k_B T}}{h^2} \right)^{3N} \quad (2.33.)$$

Where *m* is the molecular mass.

The partition function Q_{NVT} plays significant role in statistical mechanics, due to the fact that it is associated with the Helmholtz free energy:

$$A_{NVT} = -k_B T ln(Q_{NVT}) \quad (2.34.)$$

The Gibbs free energy can be computed using another ensemble, the isothermal-isobaric ensemble (NPT). In this ensemble, the number of particles, N, the pressure, P and the temperature, T, are constant. The partition function can be written as:

$$\Xi_{NPT} = \iiint e^{-\beta H(r^N, p^N)} e^{\beta pV} V^N dr^N dp^N dV \quad (2.35.)$$

Then, the Gibbs free energy is:

$$G = -k_B T ln(\Xi_{NPT}) \quad (2.36.)$$

Finally, in the same way as in the NVT ensemble, Z_{NPT} will become:

$$Z_{NPT} = \iint e^{-\beta U(r^N) + pV} dr^N dV \quad (2.37.)$$

Consequently, the Gibbs free energy can be computed from Z_{NPT}, which can be calculated from the potential energy of the system. The MD simulations similarly sample the system's potential energy completes the sampling of the system's potential energy [43,45,46] when they are conducted under constant temperature and pressure.

2.2.5 Thermostats and barostats

In this section, we discuss why thermostats, which seek to manage the temperature of a simulation, and barostats, which are used to control the pressure of a simulation are often needed for molecular simulations. We analyze the theoretical background of thermostats and barostats to understand their function.

Thermostats

When performing MD simulations, it is desirable to emulate the laboratory experiments conditions. In order to achieve this, sampling from the canonical ensemble is demanded, where the temperature is constant [69]. Normally, if the temperature of the system must be maintained during the simulation, some thermostat algorithm will be employed.

During a MD simulation the temperature is measured from the kinetic energy using the equation:

$$\frac{3}{2}Nk_bT = \langle \sum_{i=1}^{N} \frac{1}{2} m_i \boldsymbol{v}_i^2 \rangle \qquad (2.40.)$$

Where the angled brackets indicate that the temperature is defined as a time-averaged quantity. However, if we use the equation 2.40. to calculate the temperature for a single snapshot in time of an MD simulation instead of time-averaging, this quantity is referred to as the instantaneous temperature. The instantaneous temperature will undergo fluctuations around the target temperature.

There are several thermostat algorithms in order to control temperature. Some thermostats rescale velocities outside of the molecular dynamics' equations of motion, while others include stochastic collisions between the system and an implicit bath of particles, or they explicitly include additional degrees of freedom in the equations of motion giving the effect of an external heat bath [47]. Some popular thermostats are the Langevin [63], Andersen [71] and Berendsen [72].

Barostats

The thermodynamic properties, which are studied in a laboratory, are measured under open-air conditions, meaning that they are measured at essentially constant pressure and temperature (isothermal - isobaric ensemble). In order to obtain a non-atmospheric pressure, some device, would be needed to control the pressure and volume of the system. As is the case with thermostats, if the pressure must be maintained in a simulation, a barostat algorithm will be employed to sample this ensemble.

Barostat algorithms control pressure alone, so if the target ensemble is isothermal-isobaric, they must be applied along with a thermostat. The pressure of an MD simulation is commonly measured using the Clausius virial theorem (an expectation value relating to positions and forces) and when pairwise interactions and periodic boundary conditions are considered, different approaches are utilized. These formulas give pressure as a time-averaged quantity, so if we use these formulas to compute the pressure for a single snapshot, this quantity, as in the case with temperature, is referred to as instantaneous pressure, which will undergo under fluctuations around the target pressure.

Considering a hypothetical system being compressed by a fictitious piston, which has a mass acting from all directions, it can be considered as applying a uniform compression. The mass of the piston can be adapted such that it will change the frequency that the particles in the system will interact with the system enclosure. These impacts from the particles on the "enclosure" will transfer a stress on the system box from the surroundings and serve as a type of barostat [47]. Some barostats frequently used are the Berendsen [72], Andersen [71] and Parrinello-Rahman [73].

2.2.6 Main steps of Molecular Dynamics simulations

The process of performing an MD simulation generally includes these steps:

1. System preparation
2. Minimization/Relaxation
3. Equilibration
4. Production

System Preparation

This step focuses on preparing the starting state of the desired system. This includes building a starting structure, solvating (if necessary), neutralizing, applying a force field, etc. As this step differs so much depending on the system and what data are available about the starting structure, it is a step that varies a great deal depending on the nature of the system at hand and as a result may require unique tools.

Minimization/Relaxation

The goal of minimization, or relaxation, is to find a local energy minimum to start MD simulation from a realistic structure. This involves standard minimization algorithms such as steepest descent. In this way, we guarantee the removal of any unfavorable van der Waals interactions, which may exist (See also section 2.2.2.).

Equilibration

When running a MD simulation, it is desired some properties to be fixed (i.e. temperature or pressure), as we run under constant NVE or NVT. Therefore, it is required to invest simulation time in order to bring the system to the appropriate state point. For this reason, it is advised that a simulation using thermostat is performed prior to a desired production simulation. The thermostat will maintain the temperature by adding or removing heat.

Production

After bringing the system to equilibration, we may proceed with the collection of data for analysis. Typically, this phase is called “production”. Production simulation run more time than equilibration and production data should never be collected immediately after a change in conditions [47].

2.3 Free Energy Perturbation (FEP) calculations

Suppose you want to compute the binding affinity, or free energy of binding, of a ligand *L* to a receptor *R*, given by:

$$R + L \leftrightharpoons RL. \quad (2.41.)$$

The binding constant (K_b) is given by the law of mass action as the ratio of concentrations of product [*RL*] and reactants [*R*], [*L*]:

$$K_b^o = c^o \frac{[RL]}{[L][R]} \quad (2.42.)$$

The standard state concentration c^o depends on the reference state, but it is usually set to 1 mol/L assuming a constant pressure of 1 atm. Thus, the Gibbs free energy of binding ΔG_{bind} is given by:

$$\Delta G_{bind,L}^o = -k_B T ln K_b^o, \quad (2.43.)$$

where k_B is the Boltzmann constant and T the temperature of the system.

A natural, though computationally expensive, way to estimate the equilibrium constant is by directly simulating several binding and unbinding events and using them to compute the probability of binding the receptor-ligand system in the bound state, *P*(*RL*), or the unbound state, *P*(*R*+*L*). Assuming the volume change upon binding to be negligible, then the Gibbs free energy $\Delta G_{bind,L}$ is approximately equal to the Helmholtz free energy $\Delta A_{bind,L}$, and we are able to simulate the system in a box of volume V to obtain:

$$\Delta G_{bind,L}^o \approx \Delta A_{bind,L}^o = -k_B T ln K_b^o \quad (2.44)$$

The free energy difference between the bound and unbound states can be predicted by sampling many configurations, computing the energies of each microstate, and evaluating the partition function Z (deriving from the German “Zustandssumme”):

$$\Delta\Delta G_{bind,AB} = \Delta G_{bind,B} - \Delta G_{bind,A}$$

$$\approx -k_B T \left(ln \frac{Z(RB)}{Z(R+B)} - ln \frac{Z(RA)}{Z(R+A)} \right)$$

$$Z(state) = \int e^{-\beta U(\vec{q})} d\vec{q}, \quad (2.45.)$$

where Z(state) is the configuration integral of the bound or unbound state, β=1/k_BT and $U(\vec{q})$ is the potential energy of conformation $\vec{q}$ [50,52]

2.3.1 Relative Free Energy Calculations

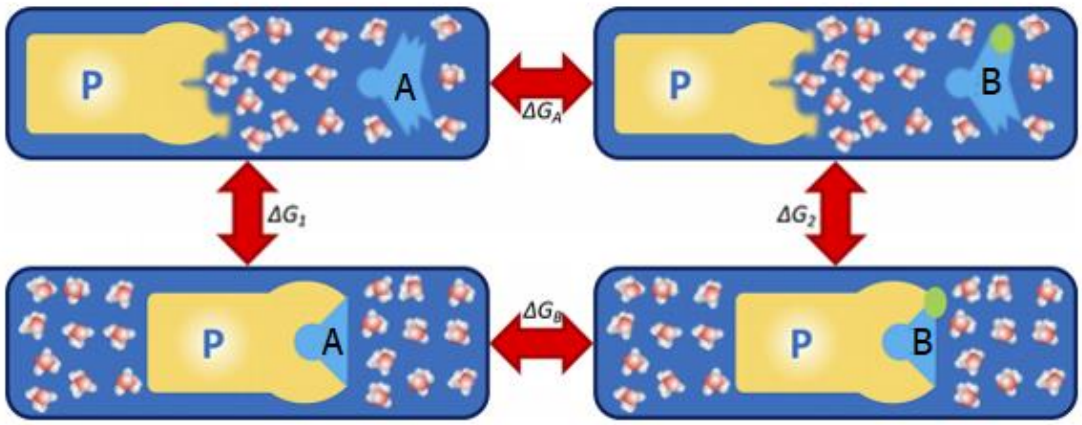

Figure 22. Thermodynamic cycle for relative binding free energy (RBFE) calculations. The binding of ligand A to the receptor is described by the left vertical line and the binding of ligand B to the receptor is described by the right vertical line. Ligand A is perturbed to ligand B in the unbound state (top horizontal line) and the bound state (bottom horizontal line) Adapted from Cournia et al. [49].

Alchemical relative free energy calculations avoid the need to simulate binding and unbinding events by making use of the fact that the free energy is a state function and exploiting the thermodynamic cycle as illustrated in Figure 22. This is apparent after rewriting the Eq. 2.45. as

$$\Delta\Delta G_{bind,AB} \approx -k_B T \left(ln \frac{Z(RB)}{Z(RA)} - ln \frac{Z(R+B)}{Z(R+A)} \right) = -k_B T \left(ln \frac{Z(RB)}{Z(RA)} - ln \frac{Z(B)}{Z(A)} \right)$$

$$= \Delta G_{bound} - \Delta G_{unbound} \quad (2.46.)$$

where $\Delta G_{bound/unbound}$ is the free energy of changing *A* to *B* in the bound/unbound state. Eq. 2.46. and Figure 22 present us that the difference in free energy of binding between (*A*) and (*B*) can be computed by running two independent calculations estimating the free energy cost of perturbating *A* into *B* in the binding pocket (ΔG_{bound} or $\Delta G_{complex}$) and in solvent ($\Delta G_{unbound}$ or $\Delta G_{solvent}$) [50].

2.3.2 Multistate transformation of Free Energy process (parameter λ)

In practice, the mutation of *A* to *B* is defined by a coupling parameter λ controlling the potential energy function $U(q;\lambda)$. Thus the direct transformation of A to B is separated in k intermediate steps such that λ_0 = A = 0.0 and λ_1 = B = 1.0. The first approach to calculate the free energy differences between thermodynamic states, was introduced by Zwanzig in 1955 and it's called Free Energy Perturbation (FEP). If one implements MD simulations using the potential energy function of thermodynamic state A to compute the free energy change of changing A to B, the following equation can be used:

$$\Delta G_{A\to B} = -RTln < e^{\frac{-(U_B(q)-U_A(q))}{RT}} >_A \quad (2.47.)$$

where the angular brackets <> indicate that the quantity inside is averaged over all the configurations of A and weighted by their Boltzmann probabilities. The procedure involves periodically computing the potential energy that B will have for a given q_i value and subtracting this from the potential energy of A at the same q_i. Ultimately, the free energy difference is calculated by using the evaluated potential energies. Hence, the free energy between two thermodynamic states A and B is the Boltzmann weighted probability of the difference of the potential energies between A and B.

An effective way to validate the result of the computation is by performing the reverse process using the following equation, since, as we mentioned before, energy is a state function and $\Delta G_{A\to B} = -\Delta G_{B\to A}$:

$$\Delta G_{B\to A} = -RTln < e^{\frac{-(U_A(q)-U_B(q))}{RT}} >_B \quad (2.48.)$$

Though the free energy changes between the two processes must be equal (if we assume that the number of samples is infinite), this is not the case in practice, due to the fact that datasets are necessarily finite. We are able to calculate the deviations from the expected results using the quantity *h*, which is called hysteresis and it is defined as the absolute value of the sum of the two free energy changes as described below:

$$h = |\Delta G_{B\to A} + \Delta G_{A\to B}| \quad (2.49.)$$

where *h* value should be as low as possible.

The problem with this strategy is the asymmetry in the rate of convergence of the free energy estimate to the true free energy change. This asymmetry can be understood in terms of state space overlap between the low energy configurations of A and B. In other words, the phase-space overlap measures the degree to which the high-probability conformations (or low energy conformations) in one state are also high-probability conformations in the other state. Ideally, the reference state should be the state of higher entropy, as the low energy configurations of the perturbed state is more likely to be a subset of the low energy configurations of this state (Figure 23a). Unfortunately, it is difficult to know a priori, which protein-ligand complex has the higher entropy, therefore it is not easy to determine beforehand in which direction the FEP equation converge more rapidly. The important region of the reference system might overlap with only a part of the important region of the target state (Figure 23b) or do not overlap at all (Figure 23c).

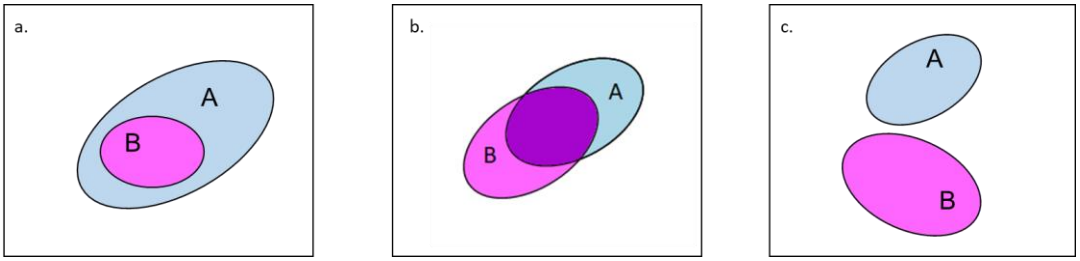

Figure 23. a) The important region of the target system is a subset of the important region of the reference system. b) The important region of the reference system overlaps with only a part of the important region of the target state. c) The important regions do not overlap.

An effective way to deal with this problem is by multi-staging the transformation of A to B. A coupling parameter λ is defined to control this conversion, thus the direct transformation of A to B is separated in k intermediate steps such that λ_0 = A = 0.0 and λ_1 = B = 1.0 as illustrated in Equation:

$$\Delta G_{A\to B} = \sum_{k=0}^{k=n=1} \Delta G(\lambda_k \to \lambda_{k+1}). \tag{2.50.}$$

The result is that the exponential averaging is only implemented between states that have high degree of phase space overlap. Although the number of simulations is increased by a factor k, each of these simulations converge faster and the overall process is more accurate. These intermediate states are unphysical and for this reason the method is called alchemical free energy calculations [50].

2.3.3 Topology approach to relative calculations

There are three approaches to relative calculations, single topology, dual topology, or hybrid topology. The distinction between these can be illustrated by considering a hypothetical transformation from molecule A to molecule B, where both atoms share a common substructure but differ in their substituents [50].

Single

In single topology calculations, the overall transformation is set up to involve as few additional atoms as possible. For instance, considering a transformation of benzene to benzyl alcohol, the benzene would be typically changed into benzyl alcohol by changing one of the hydrogens into a carbon. This site will also be the future home of two additional hydrogen atoms bound to the new carbon, so these must initially be present as non-interacting atoms called “dummy atoms”. Dummy atoms retain their bonded interactions but do not interact with the rest of the system, while bond parameters as well as partial charges between the changing atoms are adjusted accordingly between the initial and final state. Thus, in a single topology calculation, atoms may change their type, ensuring minimal dummy atoms are created [50].

Dual

In contrast, in a dual topology alchemical free energy calculation, no atoms are allowed to change type. This means that the benzene to benzyl alcohol transformation involves starting with benzene plus the non-interacting dummy atoms making up the hydroxyl methyl group, then passing through an intermediate state where some atoms are partially interacting— particularly, those atoms which are becoming dummy atoms or ceasing to be dummy atoms. The transformation finally culminates in a state where benzyl alcohol is present along with the additional dummy atom which was previously a corresponding hydrogen of the benzene [50].

Hybrid

Hybrid topology calculations essentially works by turning one molecule’s interactions with the environment off, while turning the other molecule’s interaction on at the same time. These calculations are in fact two absolute free energy calculations in opposite directions at the same time [50].

2.3.4 Evaluation of Free Energy with MBAR

Free energy calculations can be performed using a variety of schemes such as the thermodynamic integration approach or free energy perturbation using the Zwanzig formula (Equation 2.48.). Another scheme is the Bennett acceptance ratio (BAR), as well as the multistate BAR (MBAR) variation. The latter has been suggested to be the most efficient way to conduct free energy calculations and appears to be the most widely used in practice and has the following equation:

$$\Delta G^{BAR}(A \to B) = -\beta^{-1} \ln \frac{\langle f(\beta[U_{A(q)} - U_{B(q)} - C]\rangle_B}{\langle f(\beta[U_{B(q)} - U_{A(q)} - C]\rangle_A} + C \quad (2.51.)$$

where the numerator of the ratios is the ensemble average of the function *f*. It takes as input β times the difference between U_A and U_B for a given microstate q minus a constant C. This constant is obtained from data sampled from equilibrium distribution *f(B)*. The denominator is the ensemble average of the same function but with opposite sign. For this equation is necessary to assume that the same number of samples is used for both datasets. For a finite number of samples, the statistical optimal choice that minimizes the standard error is provided by the Fermi function as shown in equation:

$$f(x) = \frac{1}{1+e^x} \; and \; C = \Delta G \quad (2.52.)$$

Since ΔG is usually not known in advance, the equation must be solved self-consistently. Firstly, a guess is made for C and then the ratios from Equation 2.48. are solved to obtain ΔG and then one sets C = ΔG. The process is iterated until ΔG is not changing anymore. It can be proved that this procedure always converges to the most accurate ΔG given available data [51].

2.3.5 Convergence of Free Energy calculations

There are several ways to improve the convergence of FEP calculations when it is insufficient. First of all, it is advisable all ligands to be connected in a single graph in a way that each ligand is connected to at least two other ligands with closed cycles. The closure of the cycle contributes to compute sampling error estimates, since energy is a state function and the total free energy change should sum to zero.

In addition, extension of simulation runtimes is a way to improve the simulation convergence, due to the fact that insufficient sampling may lead to error in the calculations. Another method that can be applied in order to increase the convergence of the calculation is the addition of supplemental intermediate molecules. The addition of these compounds implies to a mutation, in which the target state is structurally more similar to the reference state, and as a result the phase space overlap will be increased.

Instead of including extra intermediate molecules, a man could increase the number of λ windows. Poor overlap between of some λ windows is a potential source of error in FEP calculations. This problem can be solved by attaching extra windows between those that present low convergence. Last, the calculations can be run in triplicates. It is possible that one calculation alone won't be reliable. It is suggested to run triplicates so that the result is statistically significant [52].

3 RESULTS

3.1 Selection of target receptor

Criteria for selection of the receptor

For the selection of the appropriate receptor we need some criteria. We desire the receptor to have only glutamate as ligand, good resolution, crystallization conditions close to natural and the receptor to be isolated from human protein. For these simulations, it is suggested that we use the receptor with PDB ID: 1FTJ [53] (Figure 24) for the following reasons. The receptor is the crystal structure of S1/S2 binding core of GluR2 in complex with glutamate. It has been crystallized from rattus norvegicus with T = 277 K, pH = 6.5 and with resolution 1.90 Å. The fact that the receptor hasn't been crystalized from human protein is something that we need to take into consideration, due to the fact that if rat protein and human protein differ to a big extend, the simulation won't reflect the reality.

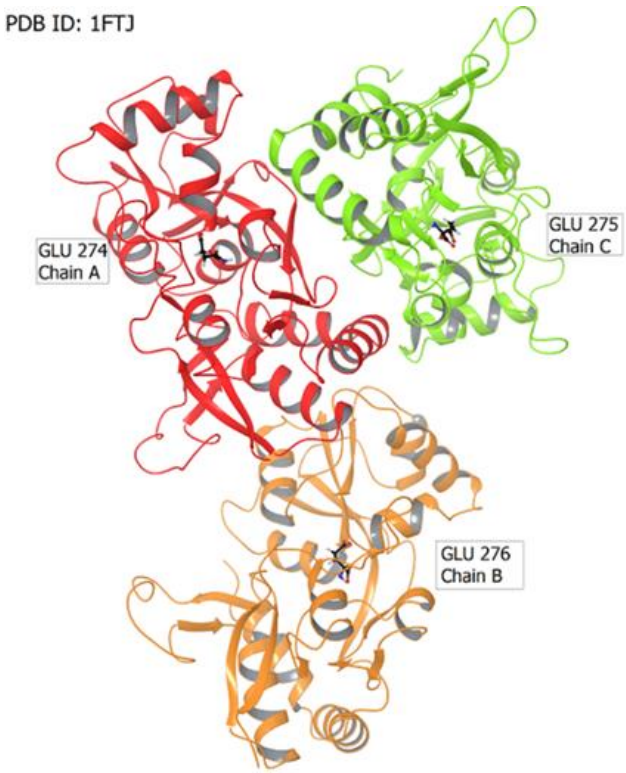

Figure 24. Crystal structure of S1/S2 binding domain of GluR2 with PDB ID: 1FTJ. The AMPA receptor consists of 4 subunits, each one including a molecule of glutamate as substrate in the LBD. This particular structure consists of only three chains. The last chain couldn't be crystalized due to low electron density [53].

Comparison of rat and human AMPA protein using BLAST

The comparison of the sequences of rat protein and human protein using BLAST [54] revealed a 99.5% sequence similarity between the two proteins. In particular, we found out that there are four differences in residues between the two sequences. After visualization in Maestro [55], we located the differences. Only one of them is in the ligand binding domain S1/S2 (G at position 231 of rat is replaced by R at the same position in human) and it is in a distance of ca. 17 Å from the center of mass of glutamate. As a result, crystal structure 1FTJ is the best choice for the calculations.

Domain of AMPA glutamate receptor with PDB ID:1FTJ

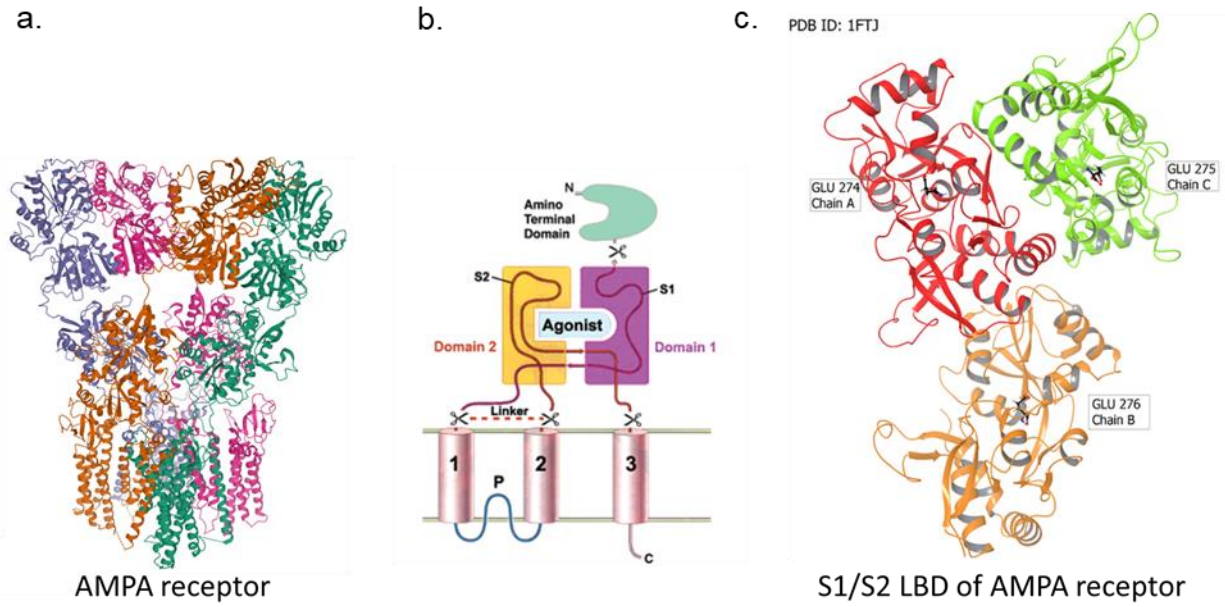

Figure 25. a) Structure of an AMPA receptor (PDB ID:5WEO). b) Simpler illustration of the AMPA gluR. c) S1/S2 LBD of AMPA gluR (PDB ID: 1FTJ).

Our receptor is an AMPA receptor, consisting of 4 subunits (Figure 25a). Figure 25b illustrates a simpler illustration of the receptor's subunits. There is the amino terminal domain, the ligand binding core, which includes 2 domains, the S1 and S2, with the ligand-binding pocket situated between the two domains. 1, 2 and 3 represent the three transmembrane segments to which ligand binding-core is connected.

However, we are interested only in the ligand binding core. Fortunately, this PDB ID includes only the S1/S2 LBD of the AMPA Receptor (Figure 25c) and as a result, this will speed up the calculations. So, scissors in Figure 25b indicate the locations that are modified to generate the ligand-binding core. The isolation was achieved by using a five residue linker with the dipeptide Gly–Thr (GT) dipeptide linker [53]. In particular, if we examine the S1/S2 LBD of our receptor we will notice that the receptor consists of 3 chains, chain A, B and C, each one having a molecule of glutamate as a substrate as shown in figure 25c. The last chain couldn't be crystalized due to the fact that the electron density for this region was not well defined, probably because multiple conformations were occupied. After alignment in Maestro all chains A, B and C are identical, apart from only one residue, Asn 3, that exist in the terminal of chain B (Figure 26). Furthermore, chain A has one Zn^{2+} metal, while chain B and chain C include two Zn^{2+} metals. Zn^{2+} metals are not located inside the binding pocket or close to it. For these simulations, chain A was selected as Heinzelmann *et al.* chose for the calculation of the standard binding free energy of glutamate in the AMPA receptor [56].

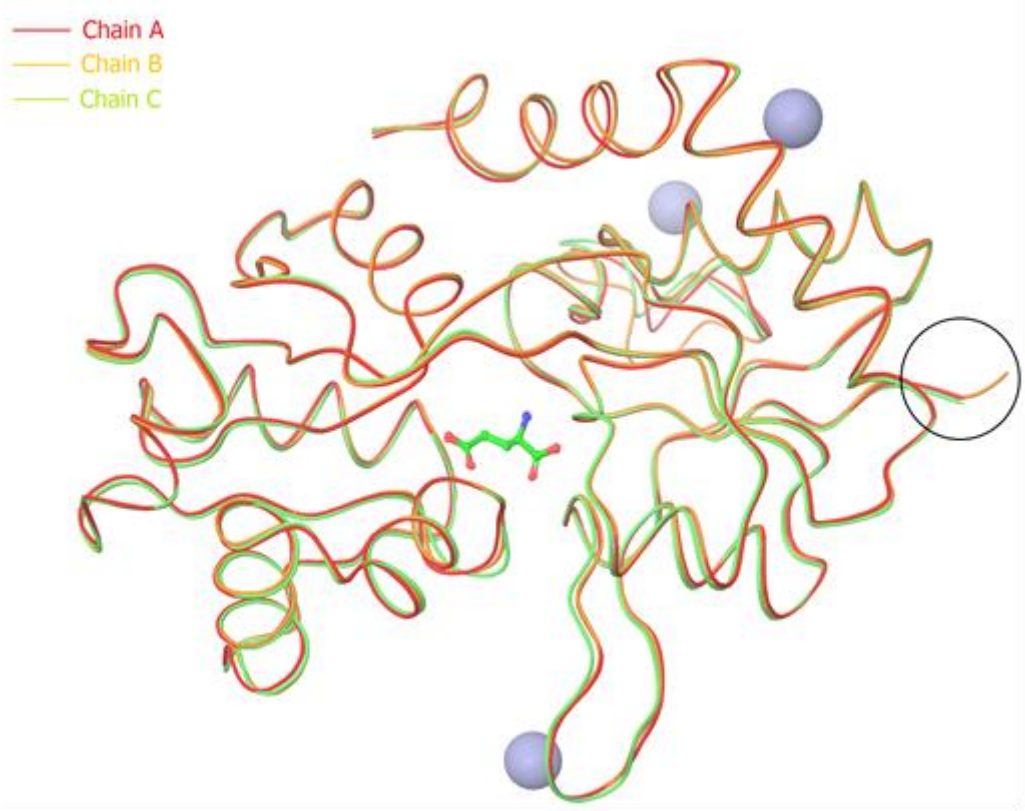

Figure 26. Alignment of Chains A, B and C of 1FTJ in Maestro.

3.2 Simulation of target receptor

3.2.1 Protein Preparation

Chain A in complex with glutamate from crystal structure PDB:1FTJ was loaded in Maestro. The selected structure was prepared with Protein Preparation Wizard tool provided by Schrödinger [57], where the missing side chains were filled in using Prime, water molecules were deleted, and protein was optimized and minimized (Figure 27). Chain A in complex with glutamate and glutamate alone were exported as a .pdb files. Glutamate was uploaded on the CGenFF [58] platform and the .str file was created, which contains bonded information and parameters for glutamate. The force field that was used for the simulations is CHARMM36 [59] both for the protein and the ligand.

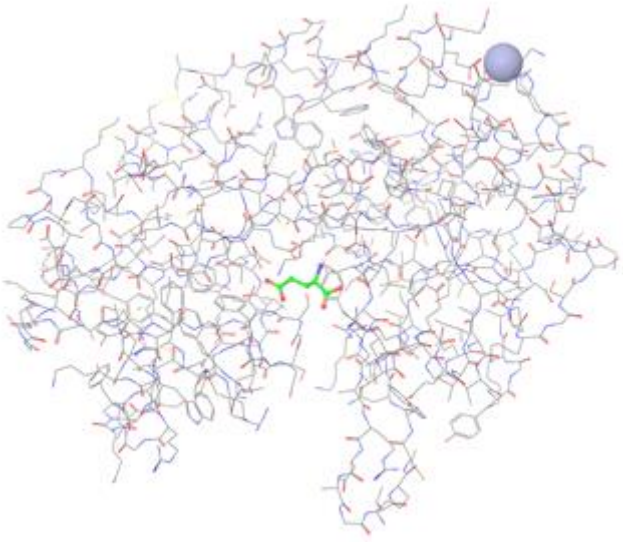

Figure 27. AMPA receptor with PDB ID: 1FTJ after being prepared with Protein Preparation Wizard tool in Maestro.

Subsequently, a tcl script was used in VMD [60], in order to generate the .psf file, to solvate and neutralize the system. The topology files used are:

- top_all36_prot.rtf
- top_all22_prot_metals.inp for zinc
- top_all36_cgenff.rtf
- top_all_lipid.rtf
- top_all36_na.rtf and

- .str file for glutamate

Chain A in complex with glutamate was placed in a water box with dimensions such that there is a layer of water 15 Å in each direction from the atom with the largest coordinate in that direction. The system was neutralized by adding six chlorine ions. We prepared a system consisting of ca. 50,000 atoms, including ca. 15,300 water molecules.

3.2.2 Simulation Parameters

Minimization

Performing an energy minimization will guarantee the removal of any unfavorable van der Waals interactions that may exist, which might otherwise lead to local structural distortion and result in an unstable simulation. The software that was used is NAMD [61]. NAMD incorporates the Particle Mesh Ewald (PME) algorithm, which takes the full electrostatic interactions into account and reduces the computational complexity of electrostatic force evaluation.

The parameters and topology files used are:
par_all36_lipid.prm
par_all36m_prot.prm
par_all36_na.prm
par_all22_prot_metals.inp
toppar_water_ions_namd.str
par_all36_carb.prm
par_all36_cgenff.prm
toppar_all36_na_nad_ppi.str
.str file for ligand

The system was minimized for 50,000 steps.

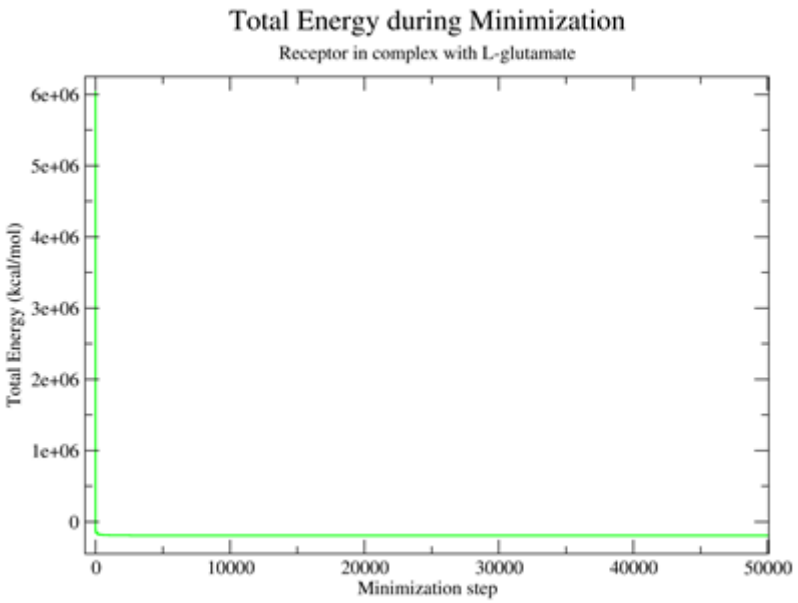

Figure 28. Total Energy of L-glutamate/AMPA in kcal/mol vs minimization step, where we confirm that energy has converged

In figure 28 it is observed that after some steps the energy has reached plateau and so, minimization has converged.

Heating

During the heating phase, initial velocities are assigned at a low temperature, 5 K, and the simulation is started with periodically assigning new velocities at a slightly higher temperature of 10 K and letting the simulation continue. This step is repeated until the desired temperature 310 K is reached.

Heating ran for 10,000 steps for each 10 K of temperature elevation and for a total 0.6 ns.

The parameters that were used are the same with those at the minimization step.

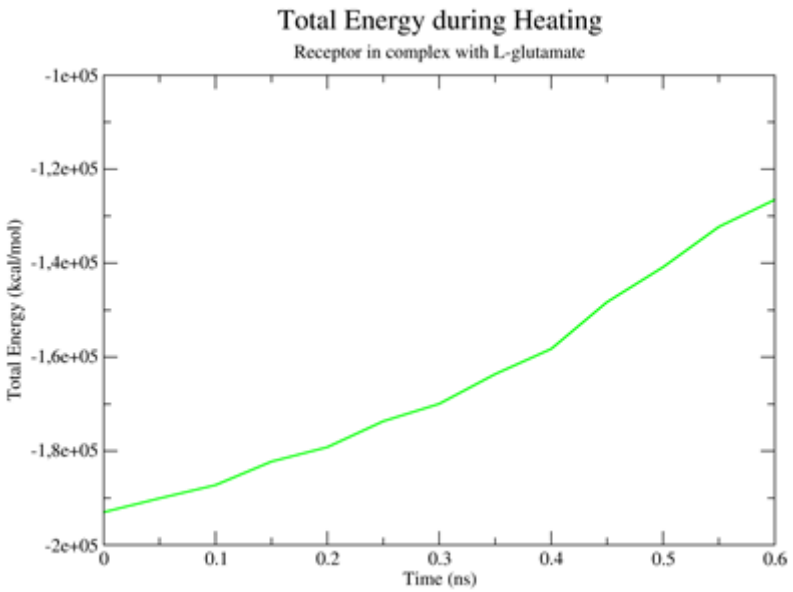

Figure 29. Total energy of L-glutamate/AMPA during the heating phase.

Equilibration

Once the heating process is over and the desired temperature is reached, the simulation is continued and during this phase, properties such as structure, pressure, temperature and the energy are monitored (simulation is performed in the NVT ensemble). As mentioned before, the point of the equilibration phase is to run the simulation until these properties become stable with respect to time.

The temperature was controlled with a langevin thermostat, which was set at 310 K and langevinDamping option, which is about damping coefficient, was set at 1 (damping coefficient of 1/ps) [62].

Equilibration ran for 500,000 steps with a time step of 2 fs (1 ns).
The parameters that were used are the same with those at the minimization step.

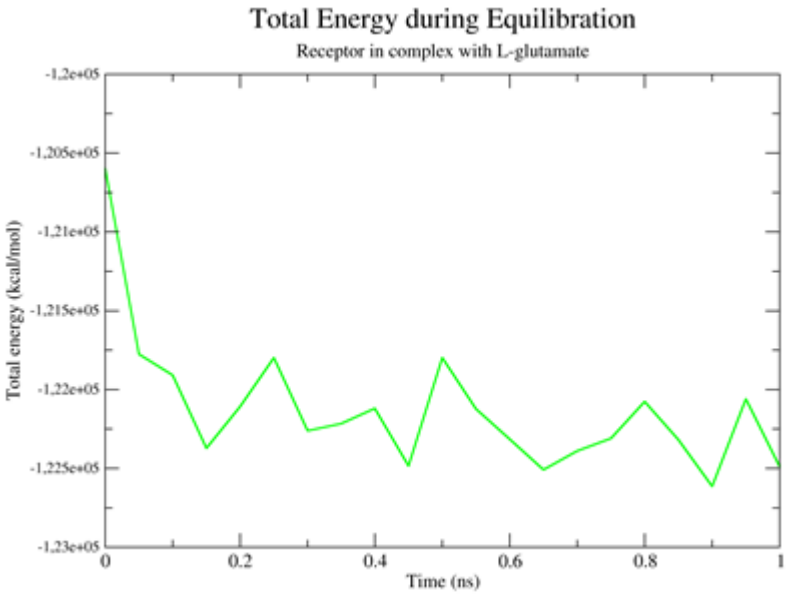

Figure 30. Total energy of L-glutamate/AMPA during equilibration. The average of the total energy is -122126 ± 426 kcal/mol.

Production Run

During this process coordinates of the system at different times are stored in the form of trajectories. These are then used for calculations of mean energy, root mean square (RMS) fluctuations between structures etc. In addition, a Nosé-Hoover barostat was used to control pressure. Option “useGroupPressure” was activated, which is needed for rigid bonds, as well as langevinPiston was. LangevinPistonTarget was used to specify target pressure for Langevin piston method at 1 atm (1atm = 1.01325 bar), langevinPistonPeriod specified barostat oscillation time scale for the Langevin piston method at 100.0 fs, and langevinPistonDecay was used, in order to specify barostat damping time scale for the method at 50.0 fs [63,64].

The simulation was run for 500,000,000 steps with a time step of 2 fs (100 ns).

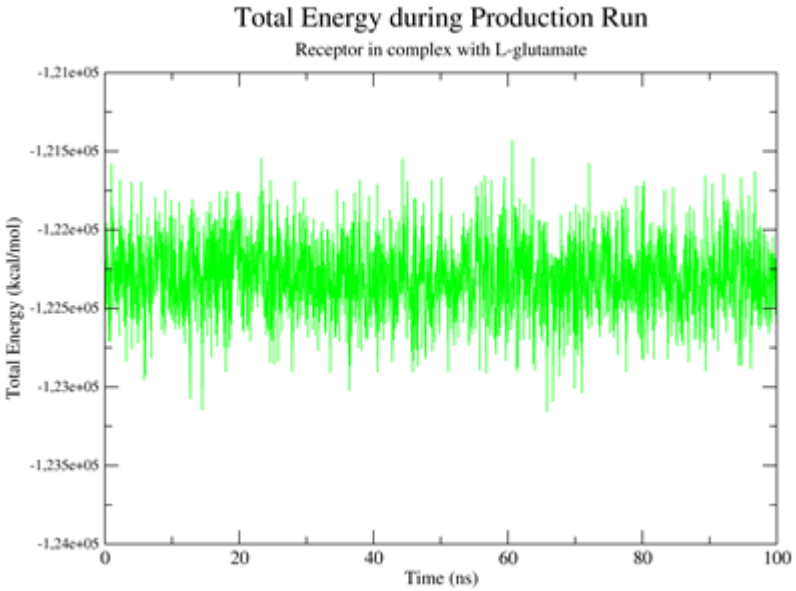

Figure 31. Total energy of L-glutamate/AMPA during the production run.

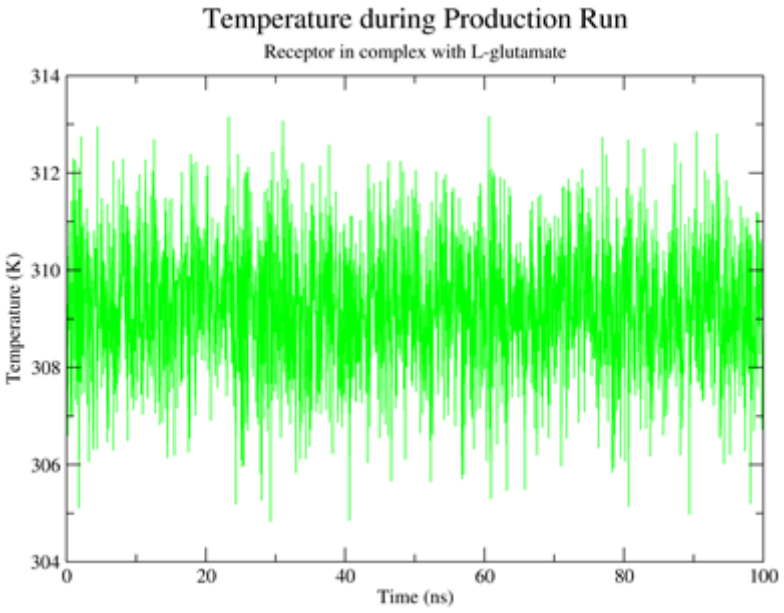

Figure 32. Temperature of the system L-glutamate/AMPA during the production run.

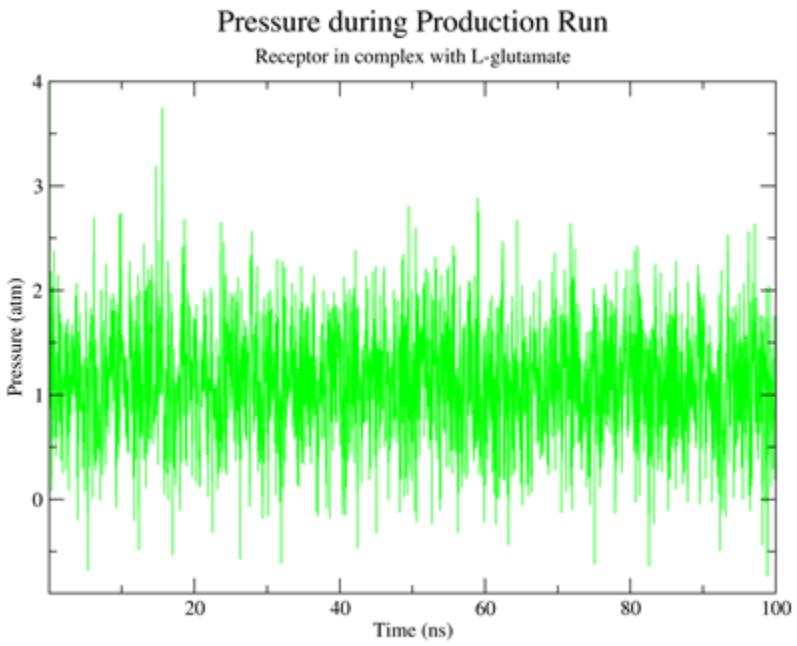

Figure 33. Pressure of the system L-glutamate/AMPA during the production run.

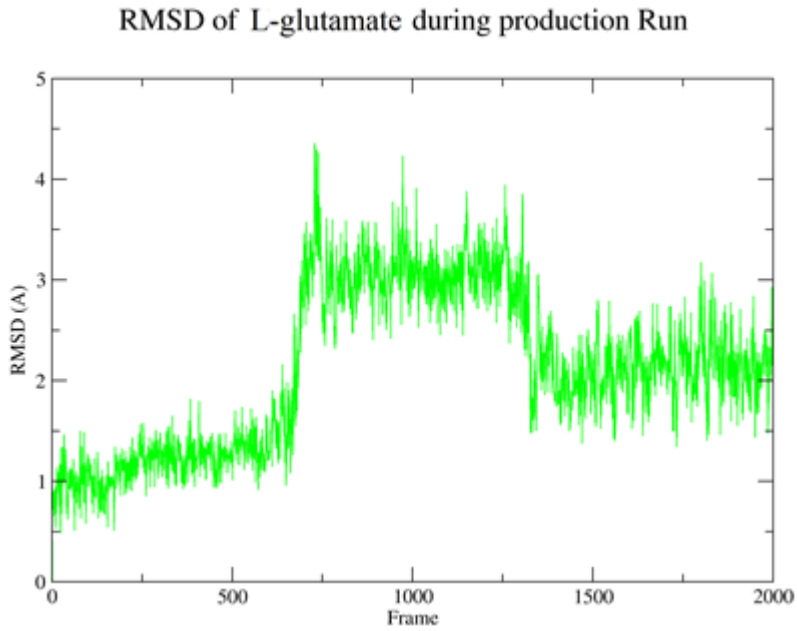

Figure 34. RMSD of L-glutamate in the binding pocket of AMPA receptor during the production run.

Figure 31 shows the total energy of the system during the production run with average - 122299 ± 249 kcal/mol. The average proves that the system is at equilibration. Figure 32 show the fluctuations of the temperature during the production run. The average temperature is 309.2 ± 1.4 K, which reveal that the thermostat worked successfully. Figure 33 show the change of the pressure during the production run, where the average pressure is 1.1 ± 0.6 atm. As a result, the barostat also worked successfully.

Figure 34 illustrates the RMSD (Root Mean Square Deviation) time series of L-glutamate in the binding pocket of the receptor during the production run. RMSD uses the following equation

$$\mathrm{RMSD}(v,w) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\left(v_{i_x} - w_{i_x}\right)^2 + \left(v_{i_y} - w_{i_y}\right)^2 + \left(v_{i_z} - w_{i_z}\right)^2\right]}, \quad (3.3.)$$

where $v_{ix,\,y,\,z}$ are the coordinates of the atoms of one frame and $w_{ix,\,y,\,z}$ are the coordinates of the atoms of another frame, and calculates the difference of x, y and z coordinates. RMSD average for L-glutamate is 2.1 ± 0.8 Å, indicating that the ligand remains in the binding pocket of the receptor after 100 ns.

Note: Part of the proteins left the water box during the production run. In order to avoid such visualization problems, the trajectories were centered in VMD using the following command: pbc wrap -centersel "protein" -center com -compound residue -all.

Analysis of trajectories – Analysis of interactions of last frames

The first cluster representative of the production run was visualized in Maestro in order to inspect the interactions between the receptor and L-glutamate. In addition, in Figure 35 the structure of the receptor in complex with L-glutamate is presented, in order to inspect where the ligand is located most time of the production run and after the production run. In both cases the molecule is located in the binding pocket of the receptor (Figure 35).

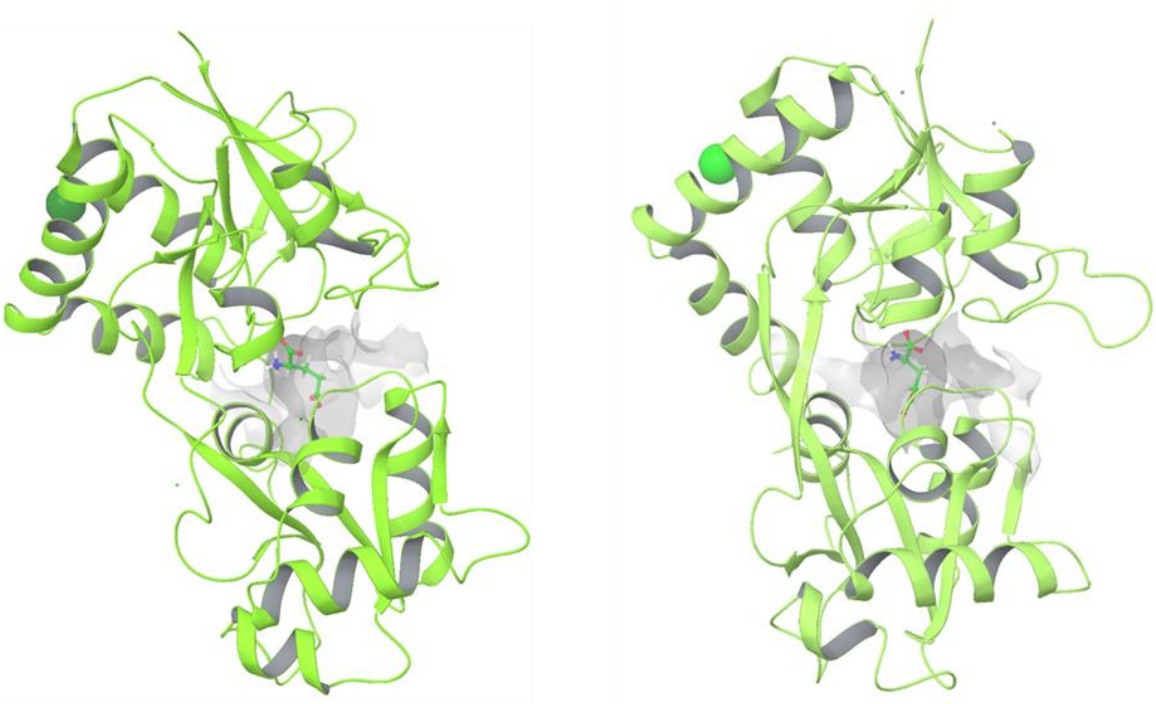

Figure 35. First cluster representative (left) and last frame of the production run (right).

Interactions between receptor and L-glutamate

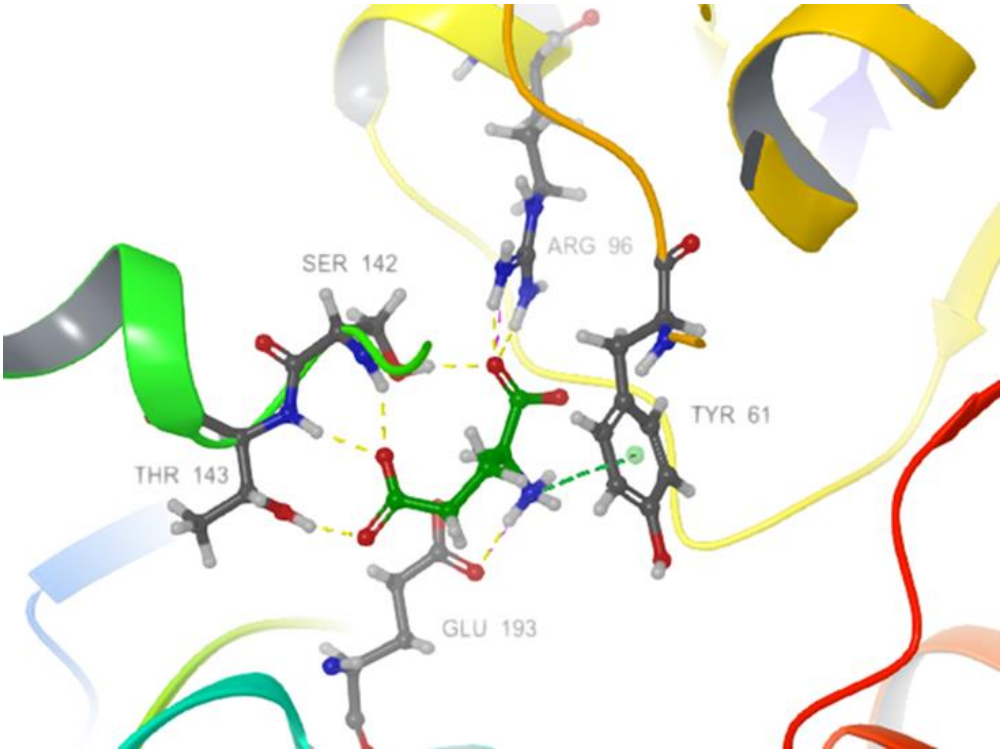

Figure 36. Interactions between L-glutamate and AMPA receptor.

The residues participating in electrostatic interactions are Tyr61 Arg96, Ser142, Thr143 and Glu193.

In figure 36 ion-dipole interactions are observed between the hydrogen of amide group of Ser142 and the carboxyl oxygen of L-glutamate, the hydrogen of the hydroxyl group of Ser142 with the oxygen of the second carboxyl group of the substrate, the hydrogen of the hydroxyl group of Thr143 and the carboxyl oxygen of L-glutamate, and the hydrogen of the amide group of Thr143 and the carboxyl oxygen of the substrate. In addition, there is cation-π interaction between the aromatic benzene ring of Tyr61 and the positive charged nitrogen of L-glutamate as well as salt bridge between the positive charged nitrogen of guanidinium group of Arg96 and the negative charged carboxyl oxygen of the substrate.

Scaling of the Molecular Dynamics simulations code

All the simulations were completed at the Greek Research & Technology Network (GRNET) in the National HPC facility ARIS.

For the plain MD simulations, we present below scalability results for NAMD 2.14 in ARIS (GRNET) for the AMPA glutamate receptor. The multicore NAMD on a single node for 100,000 steps yielded 0.0026 s/step for ca. 50,000 atoms, 0.0014 s/step on two nodes, 0.0008 s/step on four nodes and 0.0007 s/step on 8 nodes (Figure 37). The multimode NAMD scaled almost linearly up to 4 nodes.

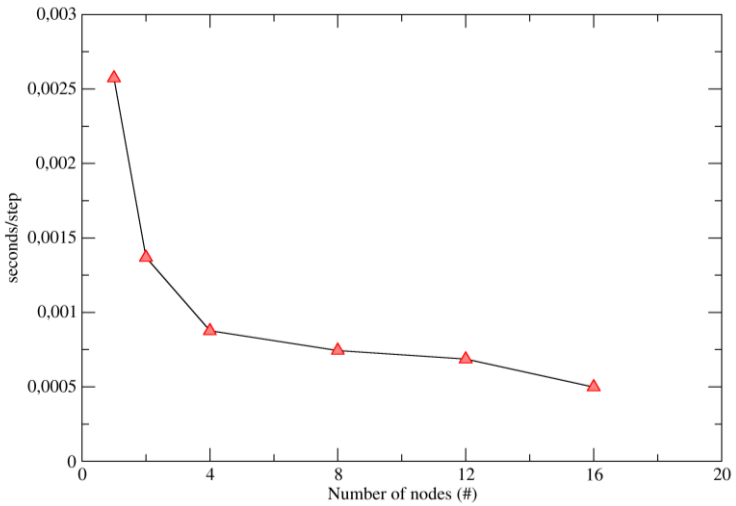

Figure 37. Time scaling

3.2.3 Rotational Energy barrier for the L-glutamate N-C-C-C dihedral

The aim of this step is to examine the stability and the interactions of a different isomer conformation of L-glutamate, by rotation of the dihedral angle N-C-C-C, in the binding pocket of the same receptor. An MD simulation will be performed in order to inspect whether this new conformation leads to a possible better binding on the binding pocket of AMPA glutamate receptor.

Selection of the conformation of L-glutamate by rotation of the dihedral angle N-C-C-C

The trajectory of L-glutamate on the AMPA glutamate receptor, which was obtained from previous step, was visualized in VMD and the desirable dihedral N-C-C-C was selected as shown in Figure 38. Subsequently, the time series of this particular dihedral were plotted (Figure 39) and the probability of each dihedral was found and plotted (Figure 40) as well as the corresponding energies in kcal/mol (Figure 41), which were calculated using the following equation:

$$\Delta G = -k_B T ln P, \quad (3.1)$$

where k_B is the Boltzmann constant, T is temperature and P is the probability.

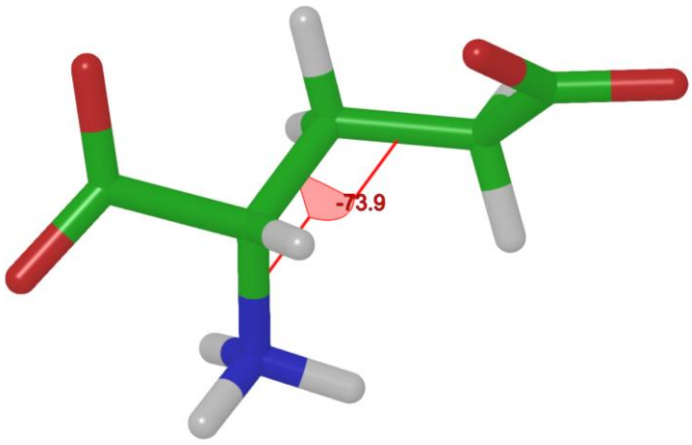

Figure 38. N-C-C-C dihedral of crystal L-glutamate.

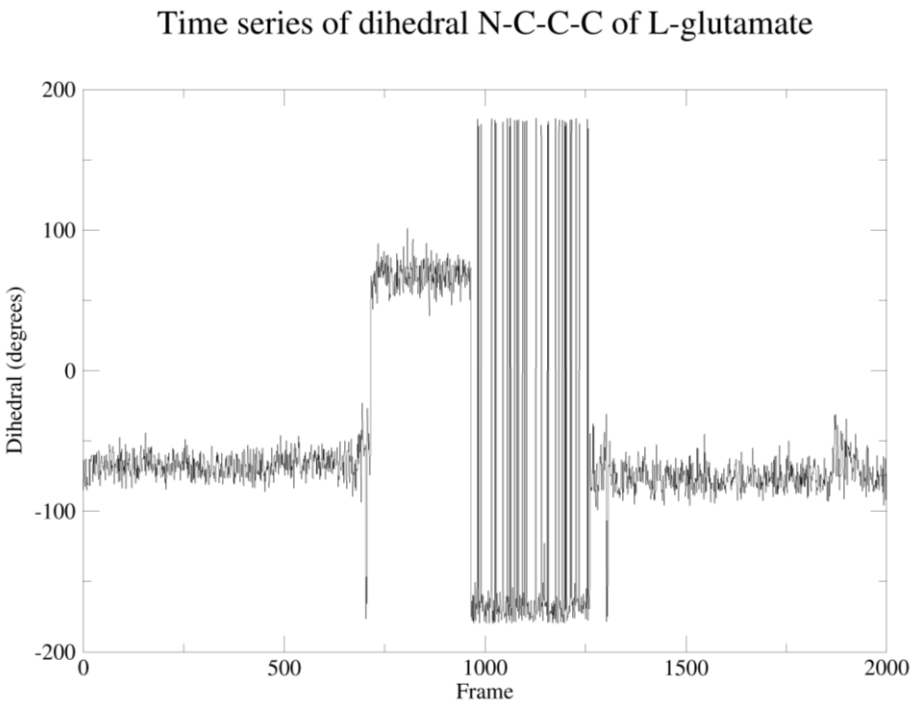

Figure 39. Time series of N-C-C-C dihedral of L-glutamate during the production run.

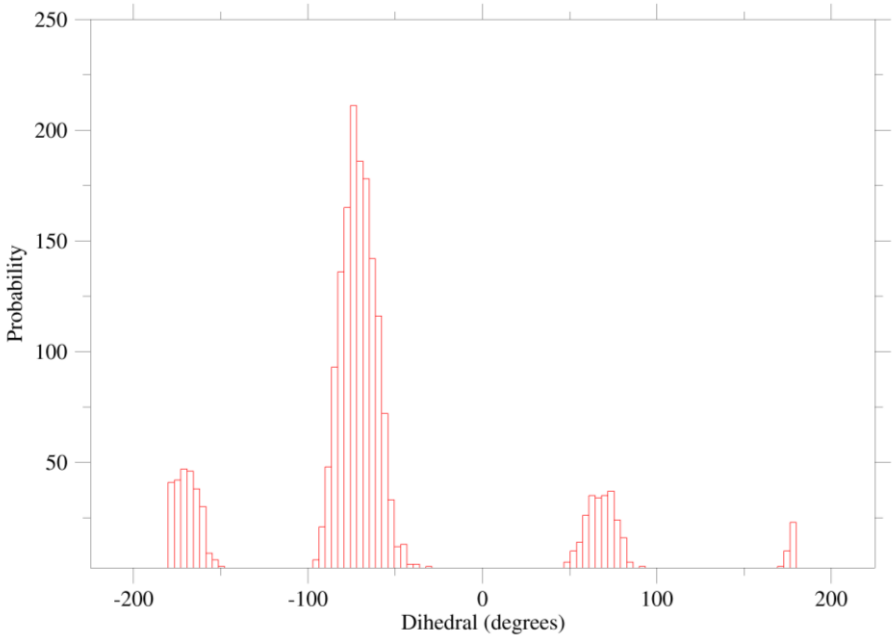

Figure 40. Probability distribution on N-C-C-C dihedral.

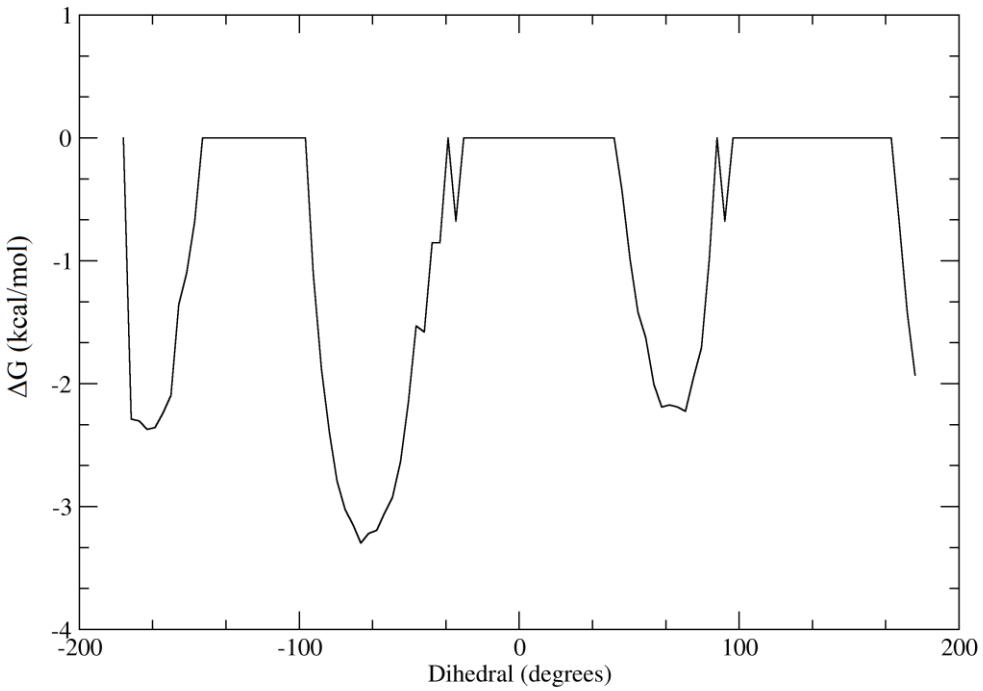

Figure 41. ΔG vs dihedral.

The previous simulation started with L-glutamate's dihedral N-C-C-C = -73.91º (Figure 38), which corresponds to a conformation from the global minimum. We would like the new simulation to start from a conformation with different dihedral, obtained from a local minimum. Frame 842, where L-glutamate's dihedral N-C-C-C = 55.46º is a very good option to start the simulation (Figures 42, 43).

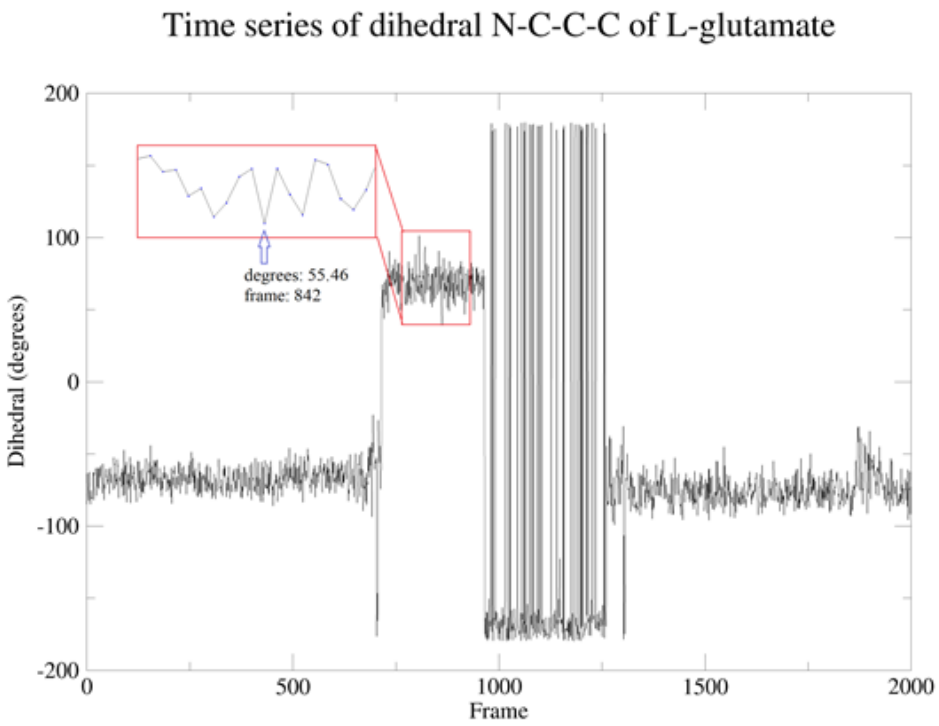

Figure 42. Selection of the frame according to the N-C-C-C dihedral of L-glutamate.

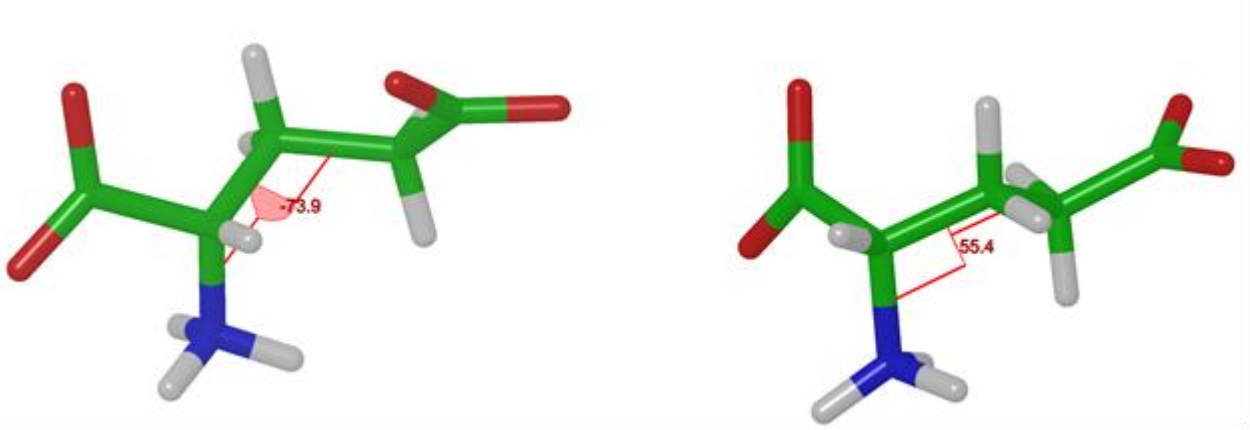

Figure 43. N-C-C-C dihedral of crystal L-glutamate (left) and of frame 842 (right).

The preparation of the system, the simulation as well as the conditions were completed as described previously. The system consists of ca. 55,000 atoms including ca. 17,000 water molecules.

Results

The system was minimized for 50,000 steps.

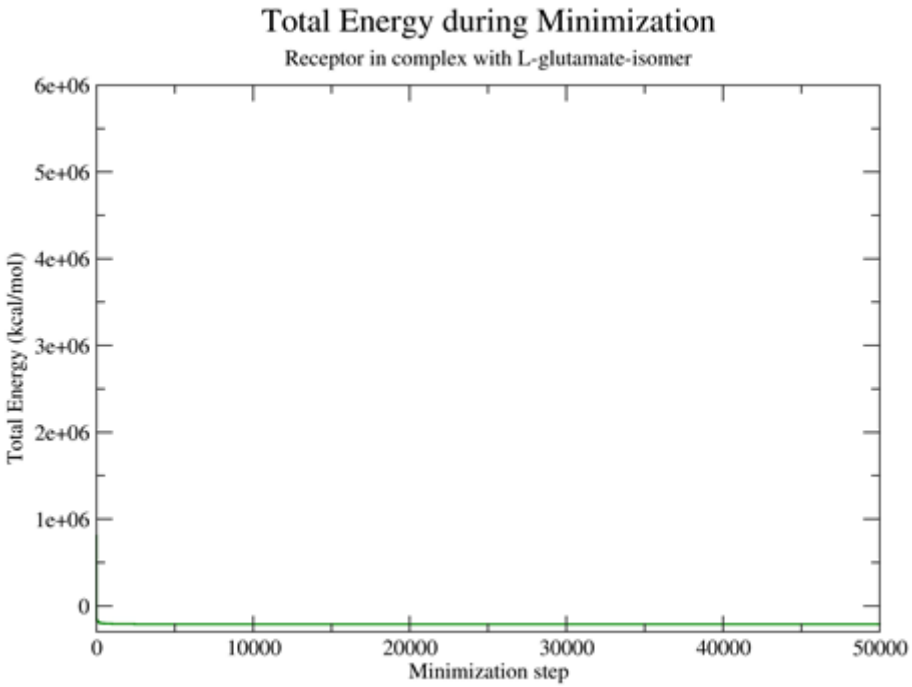

Figure 44. Total energy of L-glutamate-isomer/AMPA during minimization.

In figure 44, it is observed that after some steps the energy has reached plateau and so, minimization has converged.

Heating ran for 10,000 steps for each 10 K of temperature elevation and for total 0.6 ns.

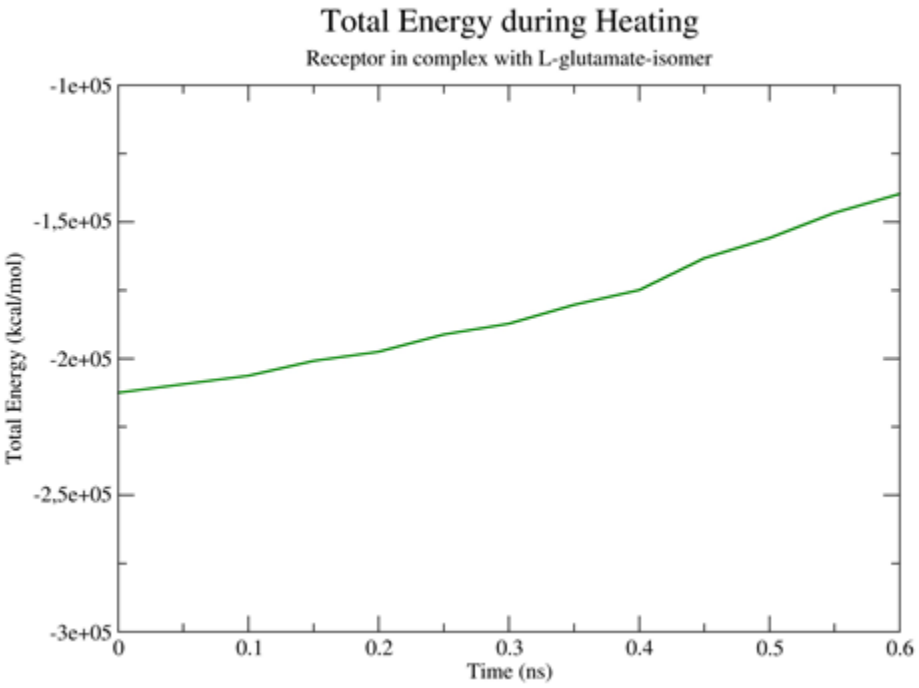

Figure 45. Total energy of L-glutamate-isomer/AMPA during the heating phase.

Figure 45 illustrates the total energy of the system during the simulation.

Equilibration ran for 500,000 steps with a time step of 2 fs (total 1 ns).

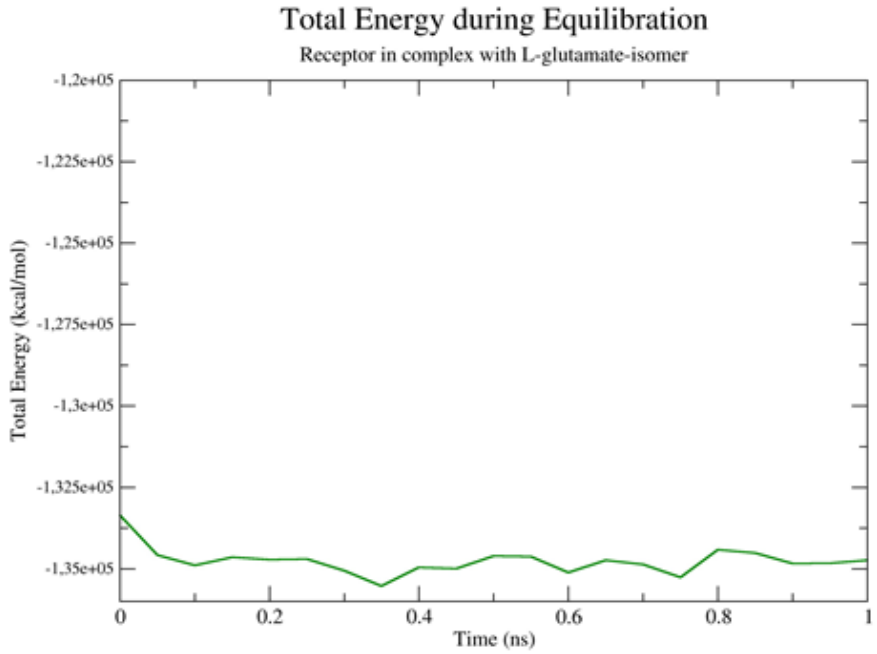

Figure 46. Total energy of L-glutamate-isomer/AMPA during equilibration.

Figure 46 illustrates the total energy vs time during the equilibration phase of the system. Particularly, the average of the total energy is - 134,759 ± 413 kcal/mol. The total energy during equilibration for the L-glutamate was -122,126 ± 426 kcal/mol. The energy difference is due to the fact that the system with the L-glutamate-isomer is consisted of more atoms (ca. 5,000 more atoms).

The simulation was run for 500,000,000 steps with a time step of 2 fs (100 ns).

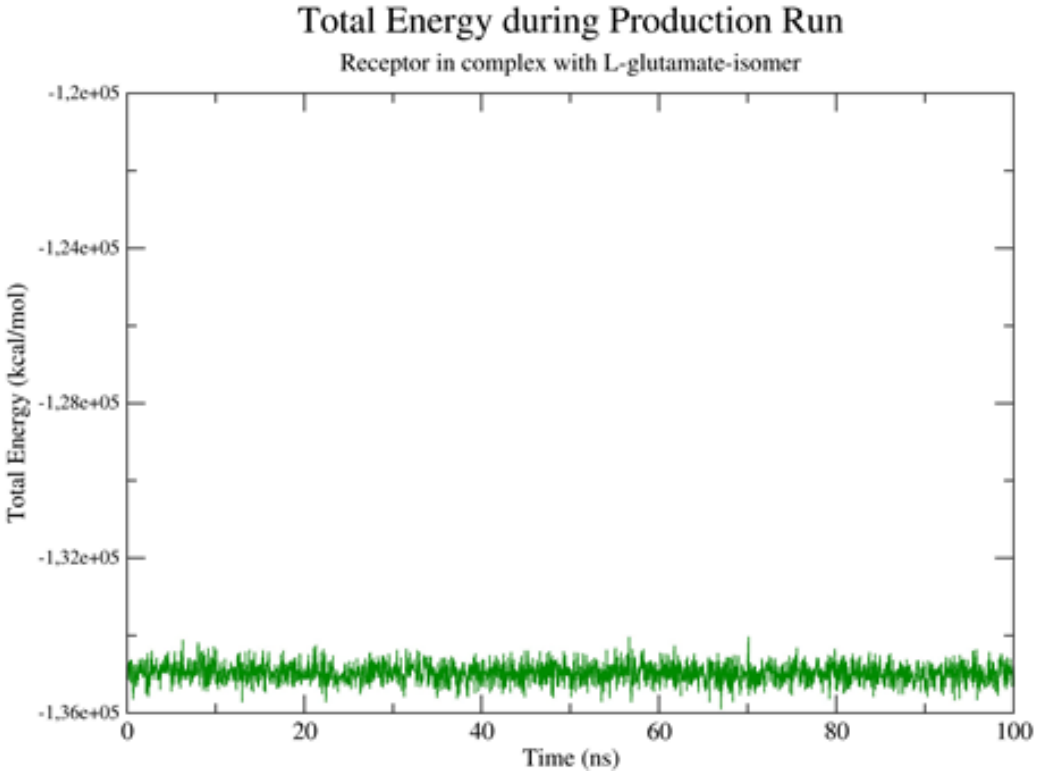

Figure 47. Total energy of L-glutamate-isomer/AMPA during the production run.

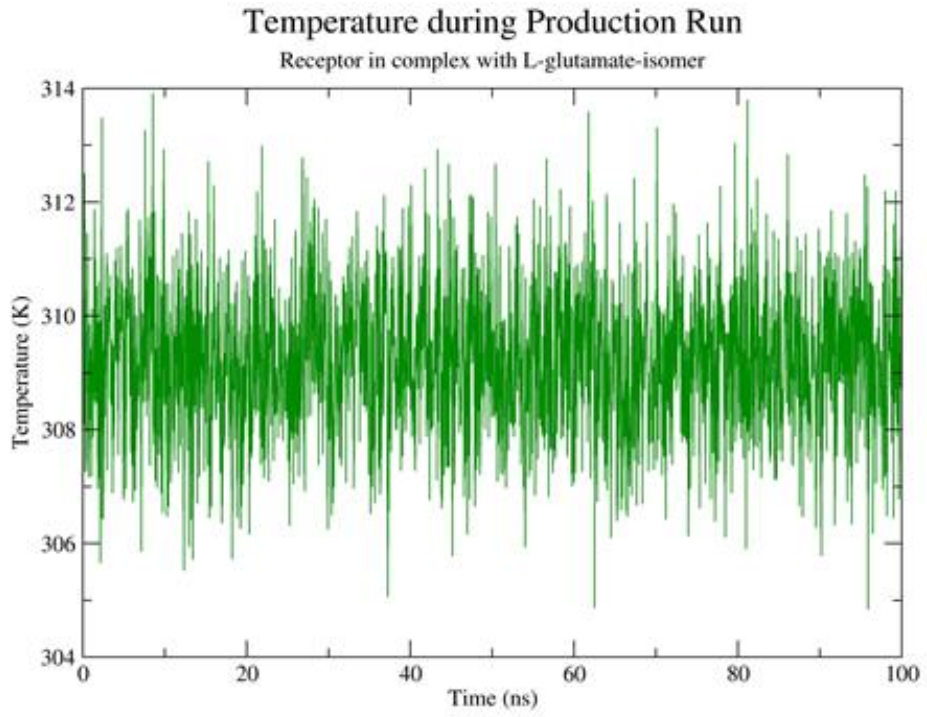

Figure 48. Temperature of system L-glutamate-isomer/AMPA during the production run.

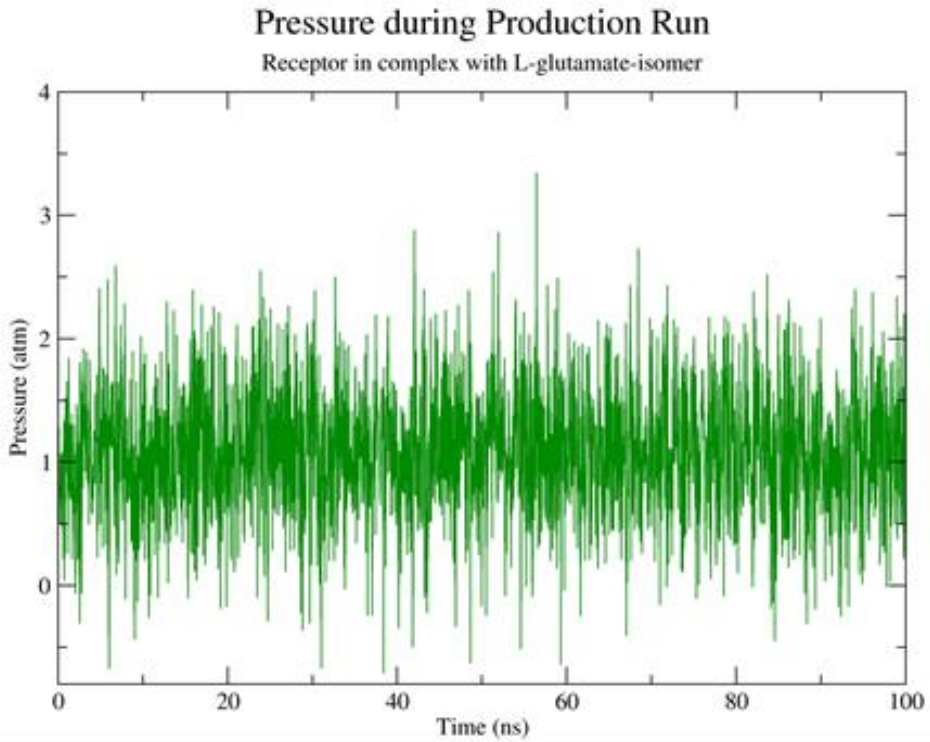

Figure 49. Pressure of system L-glutamate-isomer/AMPA during the production run.

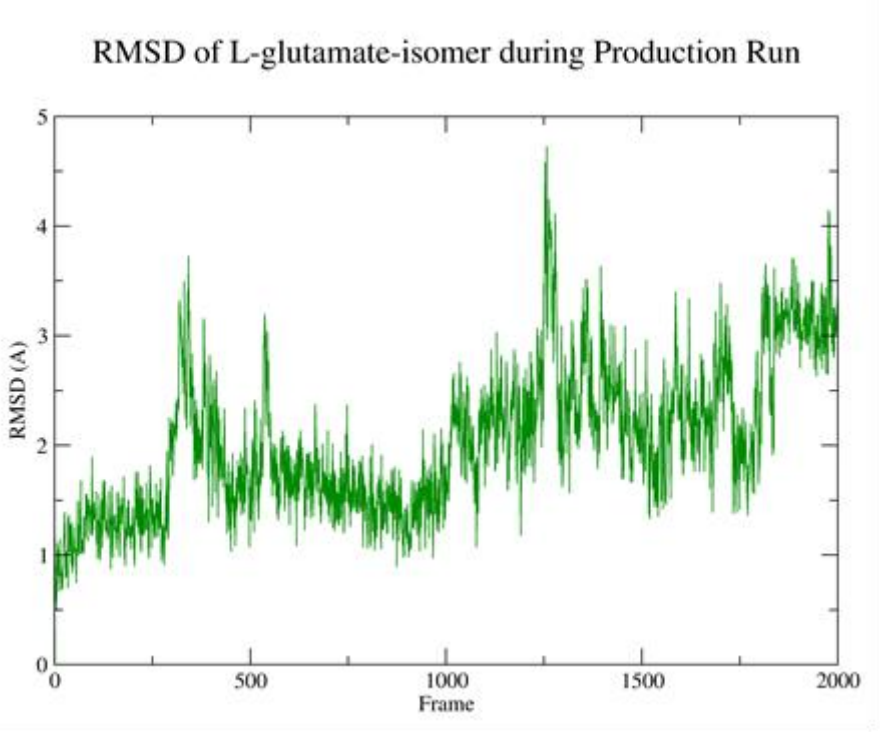

Figure 50. RMSD of L-glutamate-isomer in the binding pocket of the AMPA receptor during the production run.

Figure 47 shows the total energy of the system during the production run with average – 134,967 ± 257 kcal/mol. The total energy of the system with L-glutamate was – 122,299 ± 249 kcal/mol. This difference is due to the different number of atoms of the systems.

Figure 48 shows the change of the temperature during the production run. The average temperature is 309.2 ± 1.3 K, which reveal that the thermostat worked successfully. Figure 49 show the change of the pressure during the production run. The average pressure is 1.0 ± 0.5 atm.

Figure 50 illustrates the RMSD time series of L-glutamate-isomer in the binding pocket of the receptor during the production run. RMSD average for L-glutamate is 2.0 ± 0.7 Å. The RMSD of L-glutamate was 2.1 ± 0.8 Å. As a result, L-glutamate-isomer remains in the binding pocket of the receptor after 100 ns and the RMSD indicates that the rotation of the dihedral angle N-C-C-C, didn't affect the results.

It seems that despite the fact that the initial structure had dihedral N-C-C-C = 55.46º, it changed after the simulations. Specifically, after minimization it changed to N-C-C-C= -76.84º (Figure 51), which is close to the angle of

crystal's structure (-73.9º). During the production run, the dihedral is mostly close to -77º and at the last frames of the simulation -175º (Figure 52). However, this result arises due to the minimization step. Minimization should have been avoided in this step, as it changes the dihedral into the initial angle.

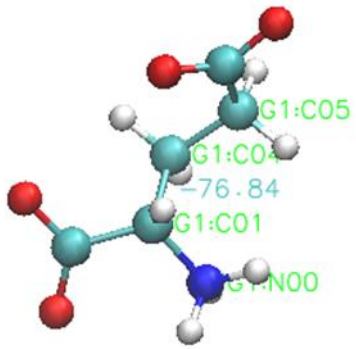

Figure 51. N-C-C-C dihedral of L-glutamate-isomer after minimization.

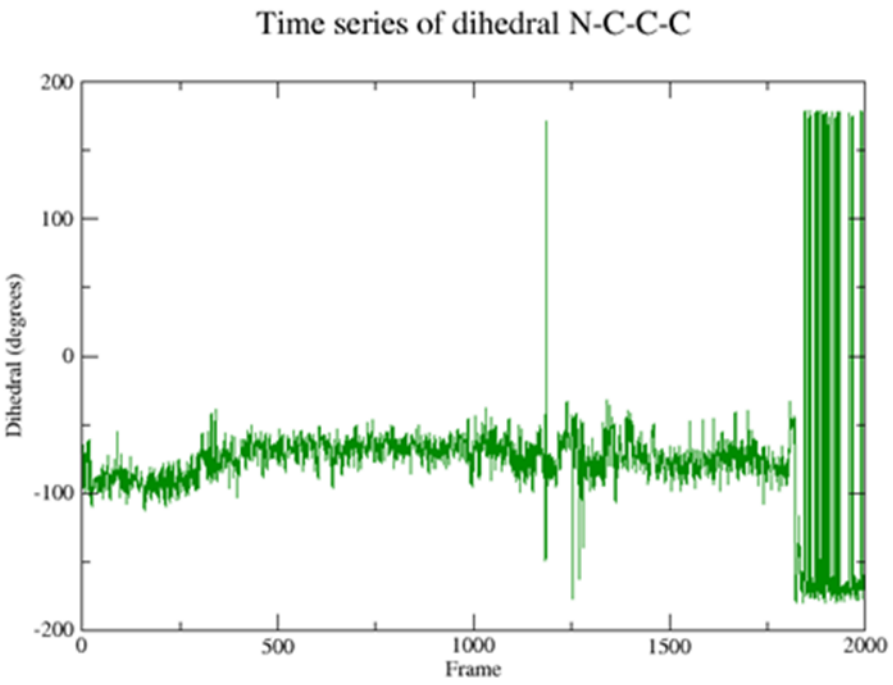

Figure 52. Time series of N-C-C-C dihedral of L-glutamate-isomer during the production run.

3.3 MD Simulations of L-ligands/AMPA receptor

3.3.1 Glutamate and BMAA adduct similarity

In order to prove that our molecules L-α-carbamate, L-β-carbamate, and L-BMAA are similar with respect to glutamate, we used tanimoto coefficient. Tanimoto is a similarity measure for comparing chemical structures represented by means of fingerprints. A fingerprint is a series of bits that represent the presence (1) or absence (0) of chemical substructures in a molecule (i.e. of a fingerprint: 10100010100010000000) [65]. By using the equation

$$T = \frac{N_C}{N_A + N_B - N_C} \quad (3.2.)$$

where N_C is the number of bits ON (1) at Molecule 1 and 2, N_A is the number of bits ON at molecule 1 and N_B is the number of bits at molecule 2, the similarity measure for L-ligands was calculated using ChemBioServer 2.0 platform [66] and the results are presented in the table below.

Table 1. Cross Tanimoto between glutamate, α-carbamate, β-carbamate and BMAA.

Cross Tanimoto				
Molecule	Glutamate	α-carbamate	β-carbamate	BMAA
Glutamate	1	**0.50**	**0.50**	**0.47**
α-carbamate	0.50	1	0.87	0.82
β-carbamate	0.50	0.87	1	0.70
BMAA	0.47	0.82	0.70	1

All ligands have T ≈ 0.50, and as a result L-α-carbamate, L-β-carbamate, and L-BMAA are similar molecules with respect to L-glutamate.

3.3.2 Validation of Glide Docking

For the preparation of the systems, it is necessary to select a method in order to place properly the ligands in the binding pocket of the receptor. For this purpose, we decided to use Glide Docking [67] and we validated the method by self-docking.

Self-docking is a procedure, where two poses, one of a docked molecule and one of a crystal molecule, are compared in order to examine how different they are. We used the receptor of our interest, AMPA, and with Glide we placed the same ligand, glutamate, in the binding pocket. Then, we compared how different were the poses between the docked glutamate and the crystal glutamate with RMSD. For reliable predictions RMSD mustn't exceed 2 Å. The RMSD that we calculated was 1.0 Å (Figure 53), so we can say that glide is a reliable method for docking L-glutamate, and the similar molecules L-α-carbamate, L-β-carbamate, and L-BMAA.

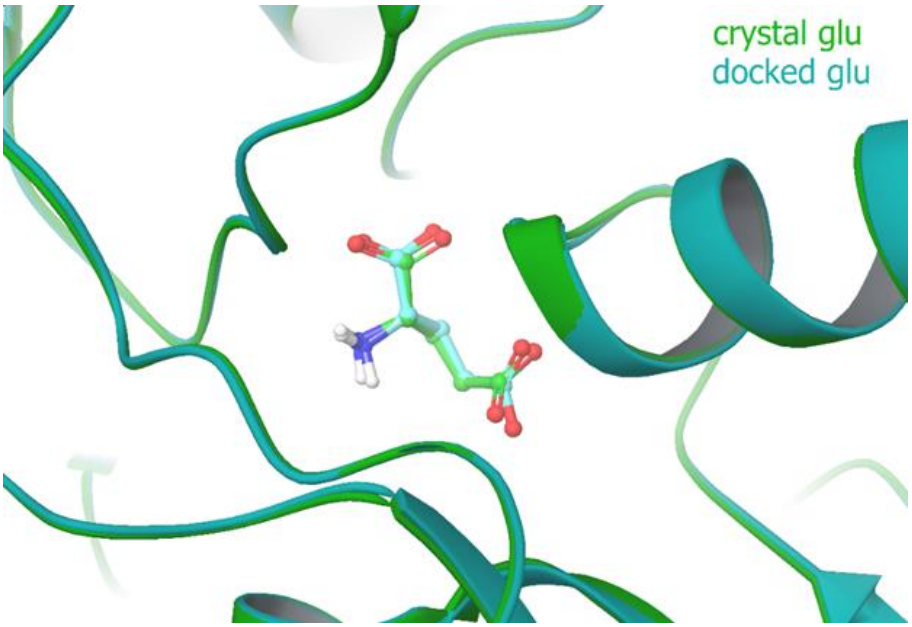

Figure 53. Alignment between crystal glutamate and docked glutamate.

3.3.3 Analysis of L-ligands-AMPA receptor MD simulations

Next, we proceeded with MD simulations of L-α-carbamate, L-β-carbamate, and L-BMAA in the binding pocket of the AMPA receptor. The preparation of the system, the simulation as well as the conditions were completed as described at first place, but before that, L-α-carbamate, L-β-carbamate, and L-BMAA were placed in the binding pocket of the receptor by docking using Glide (Figure 54). The systems are consisted of ca. 50,000 atoms including ca. 17,000 water molecules.

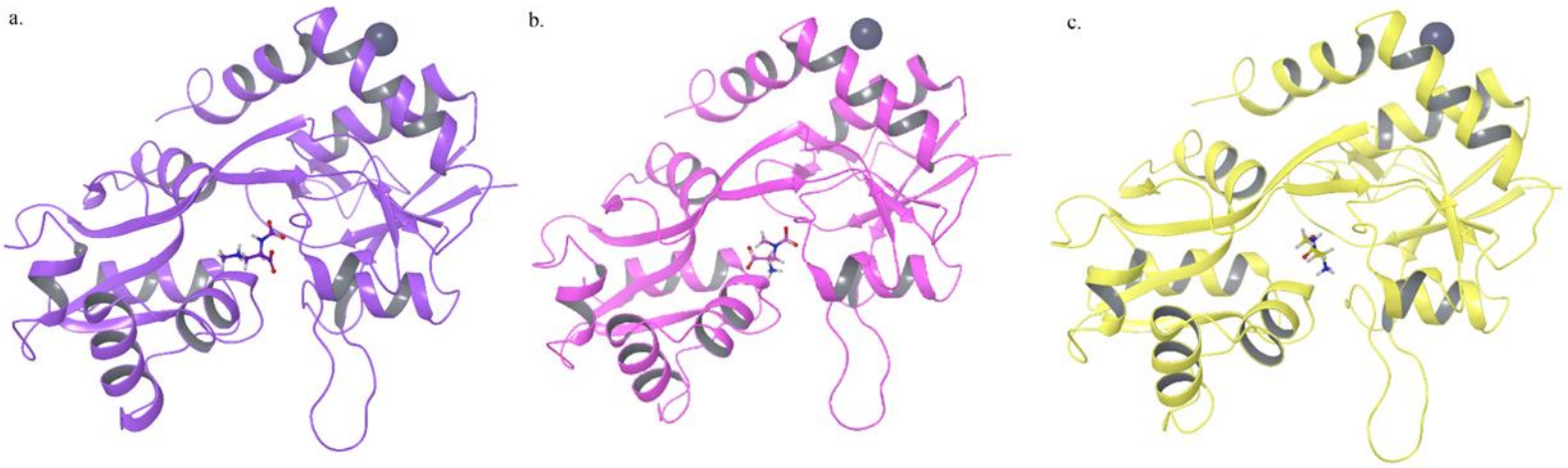

Figure 54. a) L-α-carbamate/AMPA after docking, b) L-β-carbamate/AMPA after docking, and c) L-BMAA/AMPA after docking.

Results

The systems were minimized for 50,000 steps.

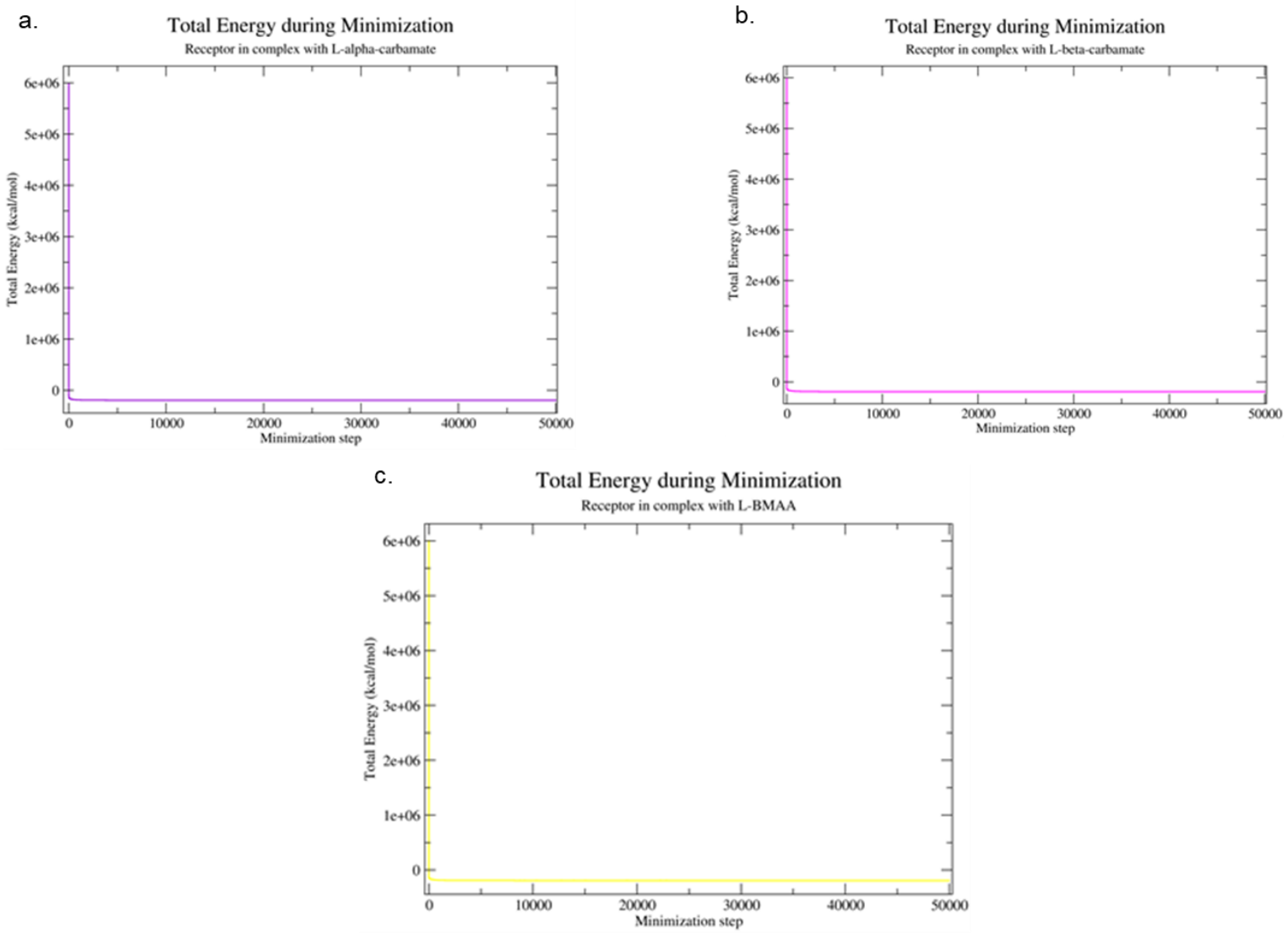

Figure 55. a) Total energy of L-α-carbamate/AMPA during minimization. b) Total energy of L-β-carbamate/AMPA during minimization. c) Total energy of L-BMAA/AMPA during minimization.

In figure 55 it is observed that after some steps the energy has reached plateau and so, minimization has converged in all cases.

Heating ran for 10,000 steps for each 10 K of temperature elevation and for a total 0.6 ns.

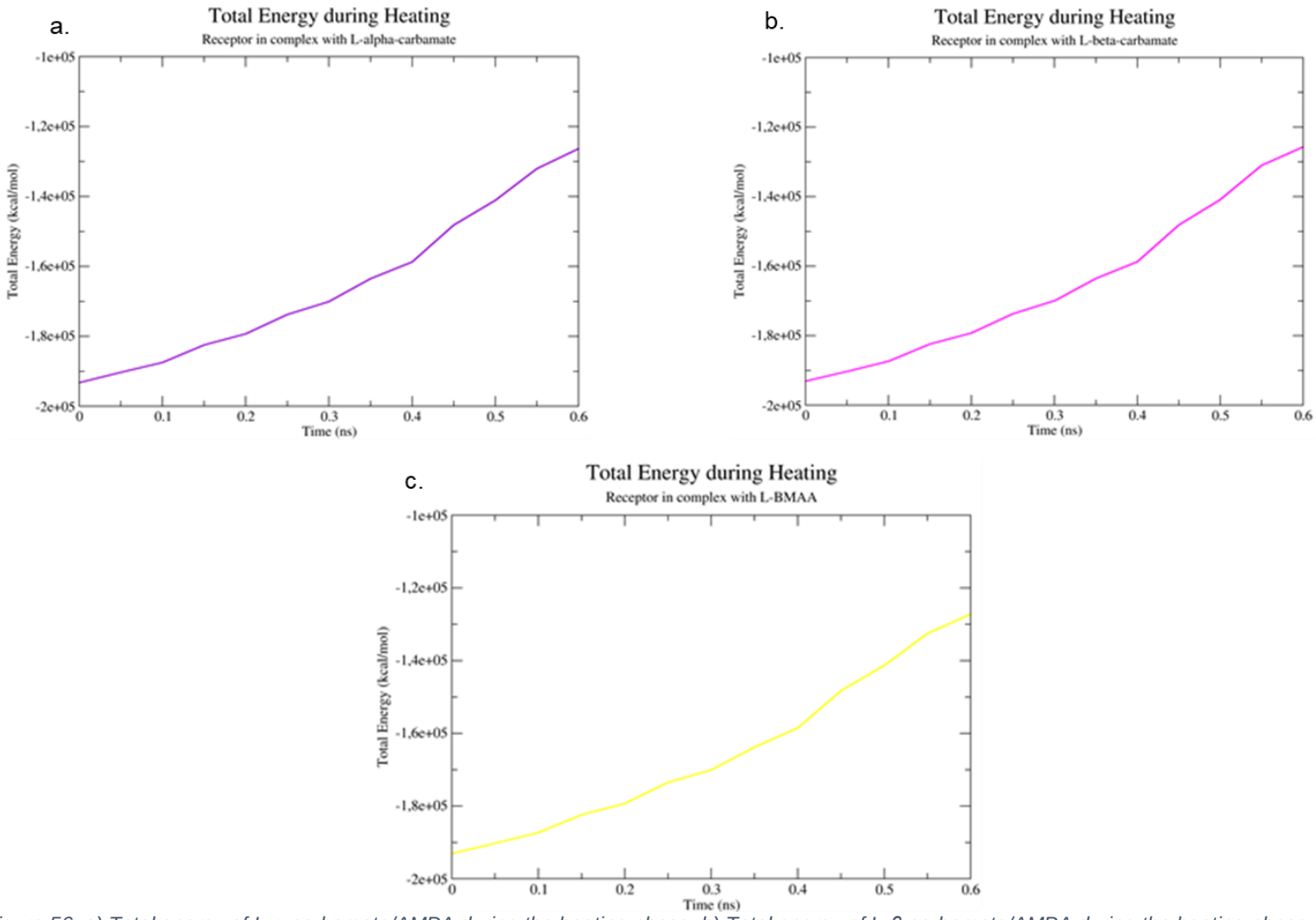

Figure 56. a) Total energy of L-α-carbamate/AMPA during the heating phase. b) Total energy of L-β-carbamate/AMPA during the heating phase. c) Total energy of L-BMAA/AMPA during the heating phase.

Figure 56 illustrates the total energy of the systems during the simulations.

Equilibration ran for 500,000 steps with a time step of 2 fs (1 ns).

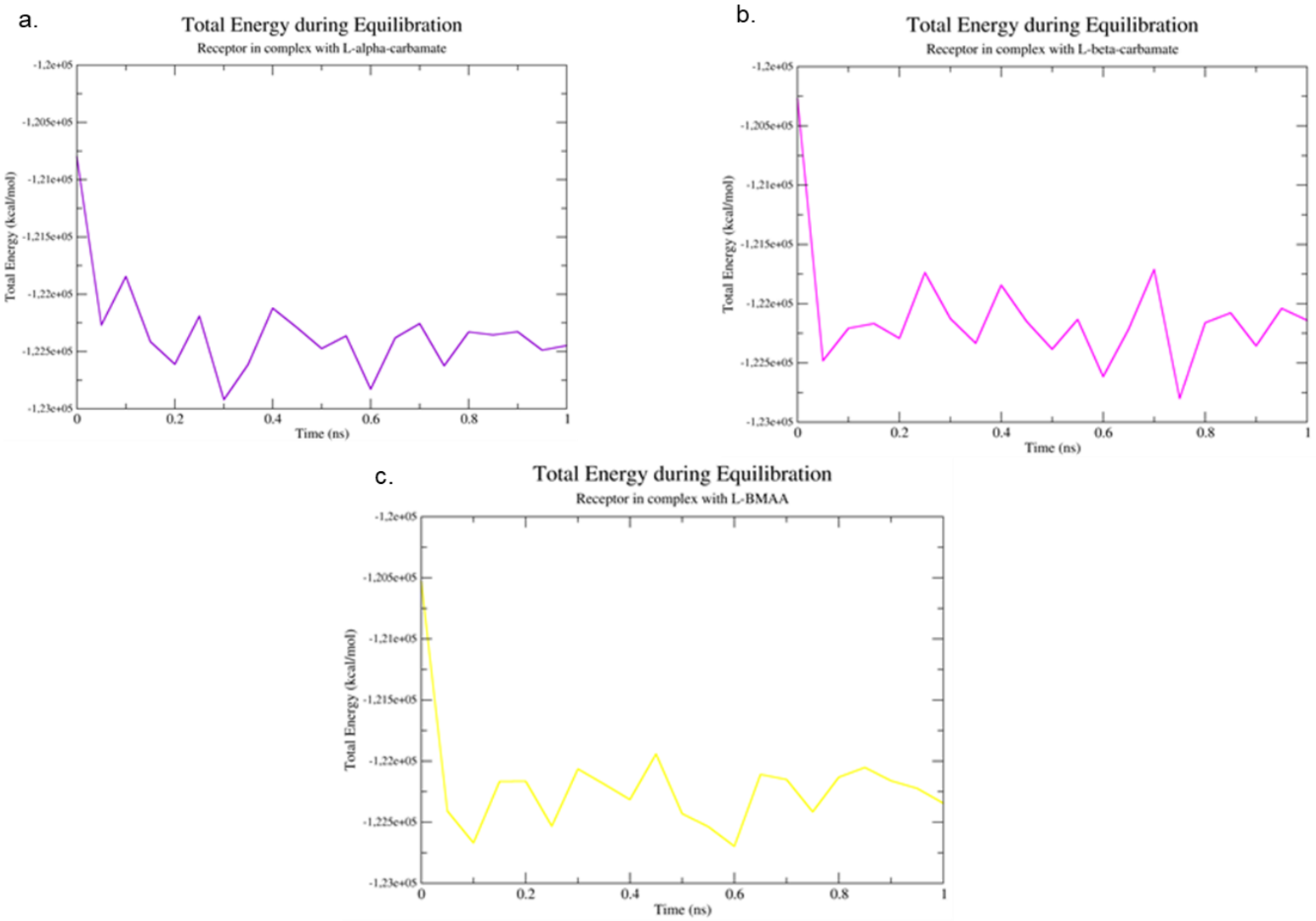

Figure 57. a) Total energy of L-α-carbamate/AMPA during equilibration. b) Total energy of L-β-carbamate/AMPA during equilibration. c) Total energy of L-BMAA /AMPA during equilibration.

Figure 57 illustrates the total energy vs time during the equilibration phase for all three systems. Particularly, the average of the total energy is -122,293 ± 497 kcal/mol for the structure with L-α-carbamate as ligand (Figure 57a), -122,132 ± 376 kcal/mol for the structure with L-β-carbamate as ligand (Figure 57b) and -122,283 ± 509 kcal/mol for the structure with L-BMAA as ligand (Figure 57c). The average of the total energy for L-glutamate was -122126 ± 426 kcal/mol.

The production simulation was run for 500,000,000 steps with a time step of 2 fs (100 ns).

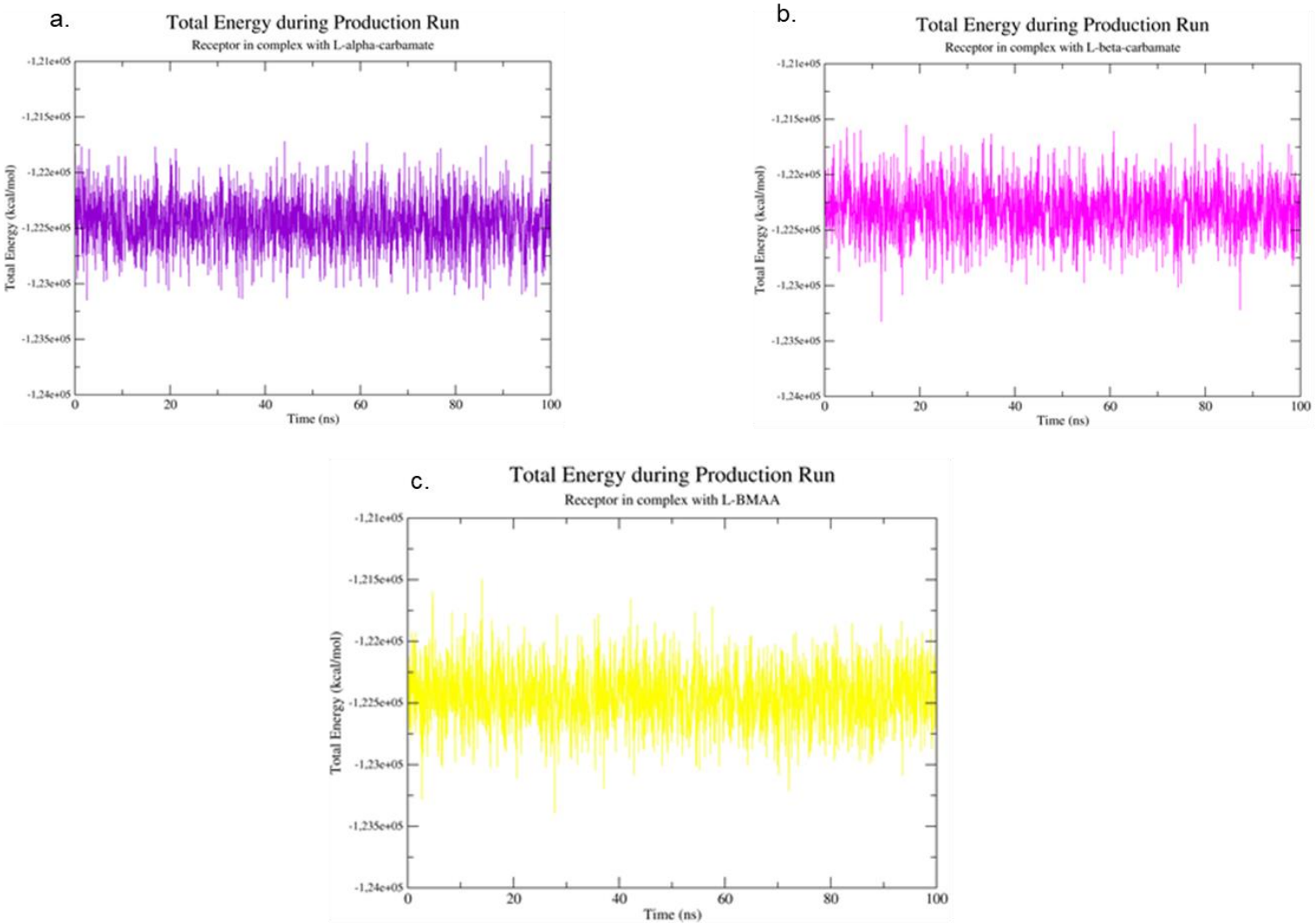

Figure 58. a) Total energy of L-a-carbamate/AMPA during the production run. b) Total energy of L-β-carbamate/AMPA during the production run. c) Total energy of L-BMAA/AMPA during the production run.

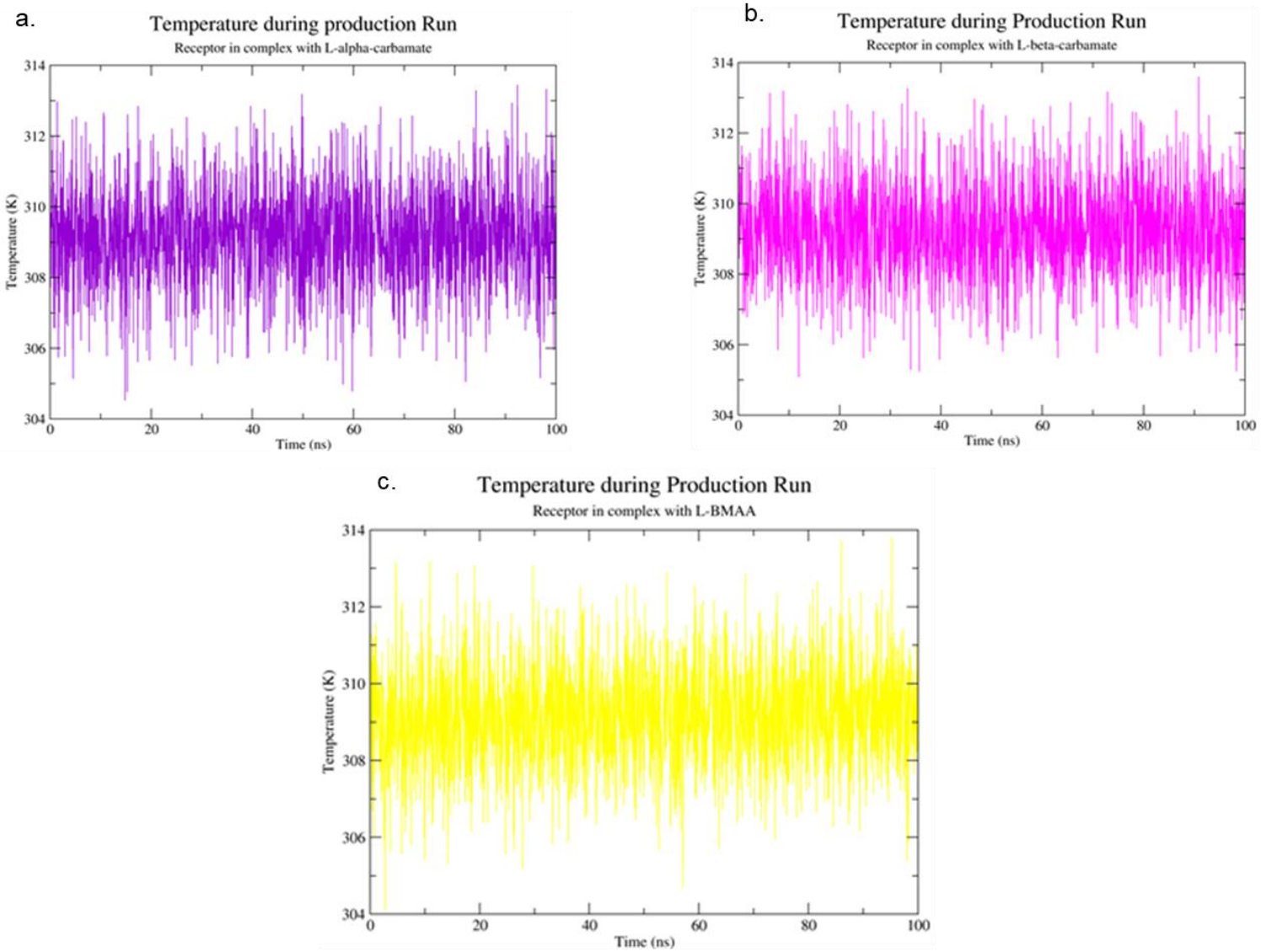

Figure 59. a) Temperature of L-a-carbamate/AMPA during the production run. b) Temperature of L-β-carbamate/AMPA during the production run. c) Temperature of L-BMAA/AMPA during the production run

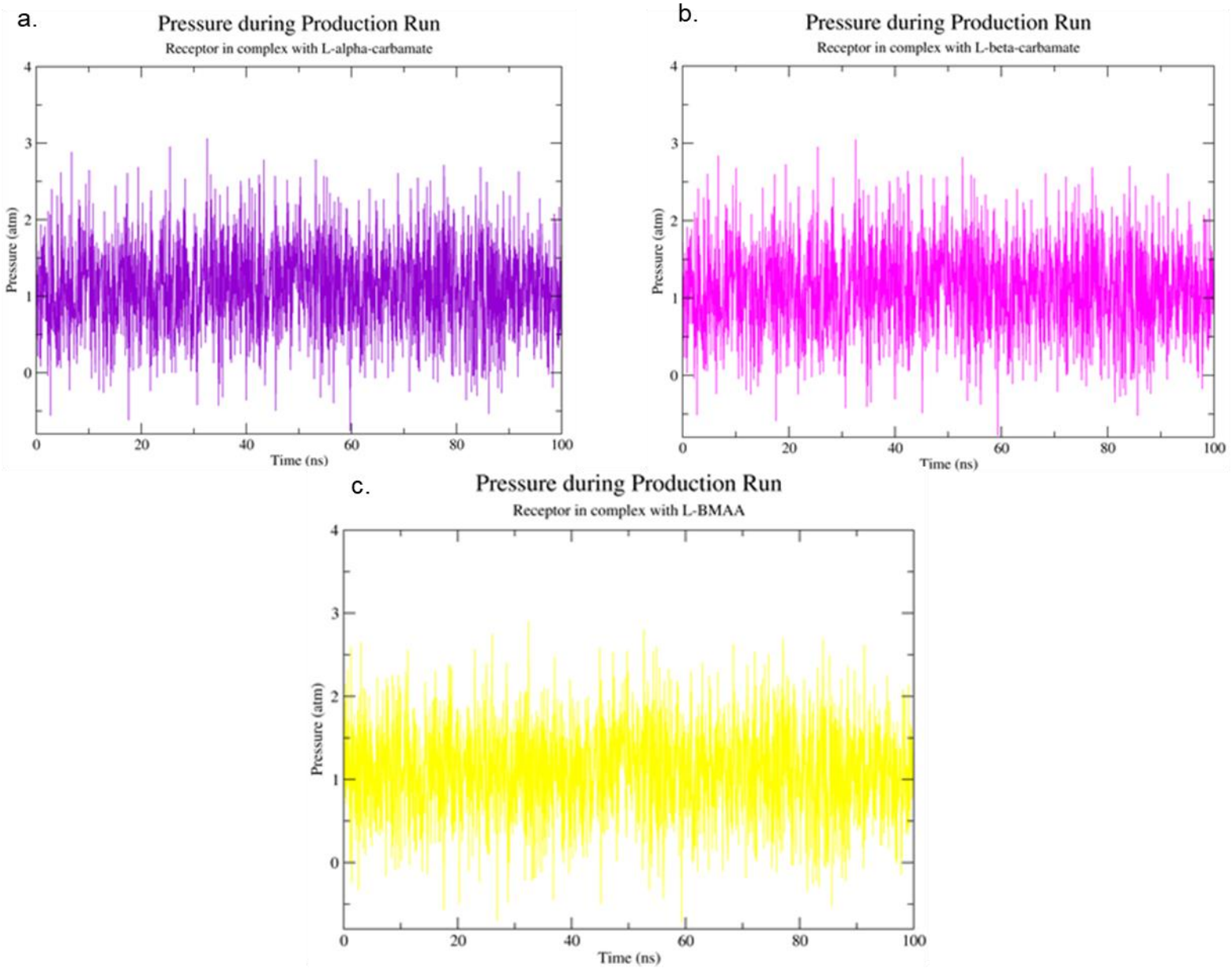

Figure 60. Pressure of L-a-carbamate/AMPA during the production run. b) Pressure of L-β-carbamate/AMPA during the production run. c) Pressure of L-BMAA/AMPA during the production run

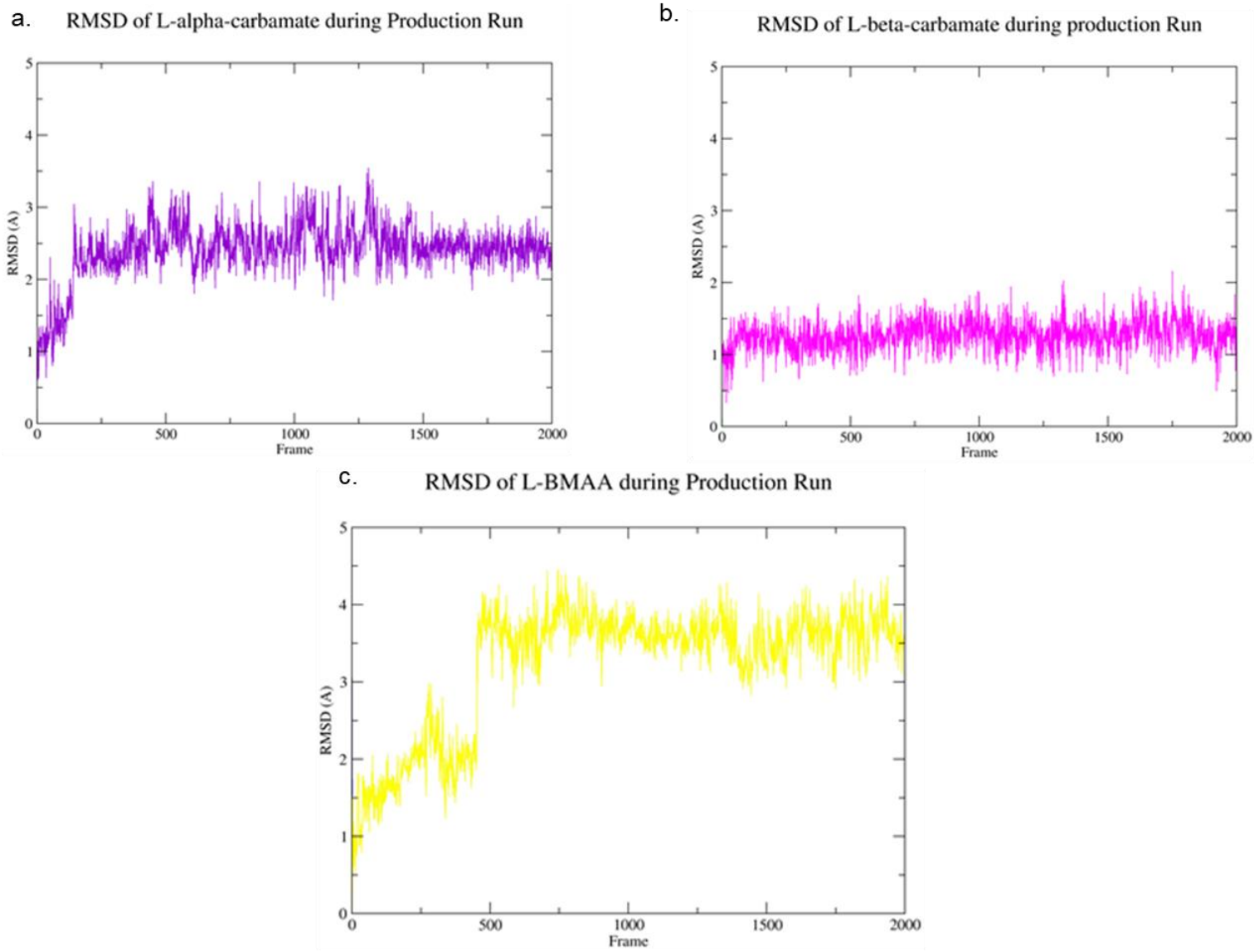

Figure 61. a) RMSD of L-α-carbamate in the binding pocket of AMPA during the production run. b) RMSD of L-β-carbamate in the binding pocket of AMPA during the production run. c) RMSD of L-BMAA in the binding pocket of AMPA during the production run.

Figures 58a, 58b and 58c show the total energies of the three systems (receptor in complex with L-α-carbamate, receptor in complex with L-β-carbamate, and receptor in complex with L-BMAA) during the production run with average -122,442 ± 240 kcal/mol for the structure with L-α-carbamate, -122,312 ± 241 kcal/mol for the structure with L-β-carbamate, and -122,429 ± 236 kcal/mol for the structure with L-BMAA. The averages prove that all systems are at equilibration. The average for L-glutamate was - 122299 ± 249 kcal/mol.

Figures 59a, 59b and 59c show the change of the temperature during the production run. The average temperature is 309.2 ± 1.4 K for all systems,

which reveal that the thermostat worked successfully. Figures 60a, 60b and 60c show the change of the pressure during the production run. The average pressure is 1.1 ± 0.6 atm for all systems and as a result, the barostat also worked successfully.

Figures 61a, 61b, and 61c illustrate the RMSD time series of L-α-carbamate, L-β-carbamate, and L-BMAA respectively in the binding pocket of the receptor during the production run. RMSD average for L-α-carbamate is 2.4 ± 0.4 Å (Figure 61a), RMSD average for L-β-carbamate is 1.2 ± 0.2 Å (Figure 61b), and RMSD average for L-BMAA is 3.0 ± 0.8 Å (Figure 61c). The RMSD for L-glutamate was 2.1 ± 0.8 Å. As a result, all three molecules remain in the binding pocket of the receptor for 100 ns with the best RMSD being for L-β-carbamate, then for L-glutamate, L-α-carbamate, and last for L-BMAA. This indicates that L-β-carbamate, which shares the highest structural similarity among the three ligands (Figure 62), is more stable than the natural substrate, glutamate. In addition, L-BMAA for which it is reported that is not the one that binds in glutamate's binding site [9,12], is the least stable molecule.

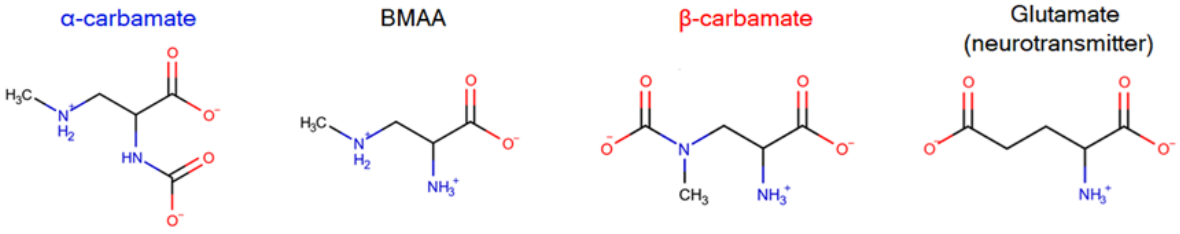

Figure 62. Structures of a-carbamate, BMAA, β-carbamate and glutamate.

Analysis of trajectories – Analysis of interactions of last frames

The first cluster representatives of the production runs were visualized in Maestro in order to inspect the interactions between the receptor and L-α-carbamate, L-β-carbamate and L-BMAA. In addition, in figures 64, 65, and 66 the structures of the receptor in complex with L-α-carbamate, L-β-carbamate and L-BMAA are presented respectively, in order to inspect where the ligands are located most time of the production run (left) and after the production run (right). In all cases the molecules are located in the binding pocket of the receptor.

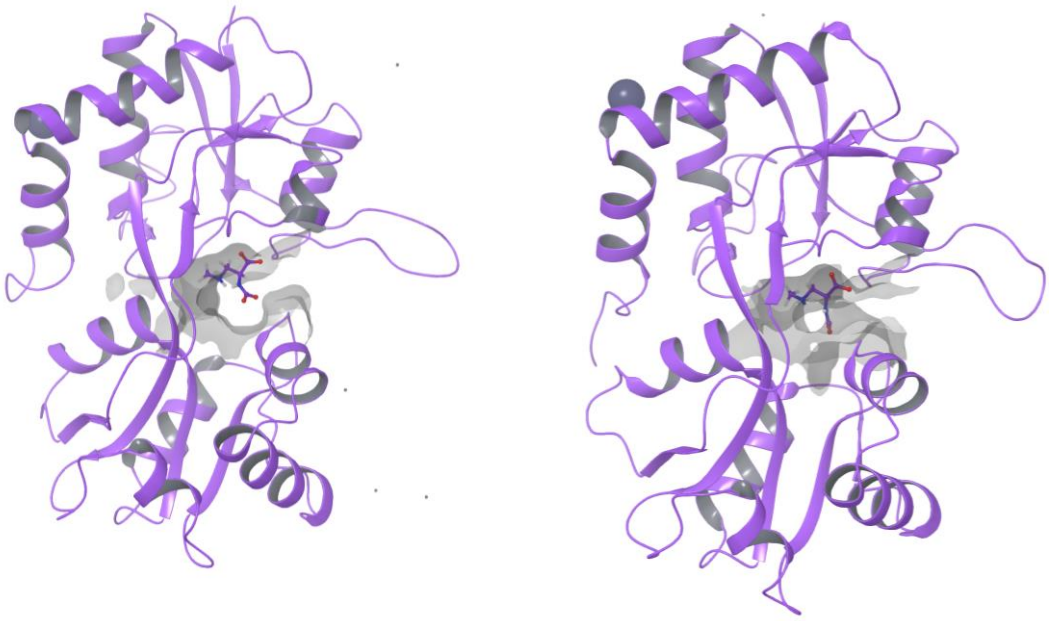

Figure 64. First cluster representative of L-α-carbamate/AMPA (left) and last frame of L-α-carbamate/AMPA of the production run.

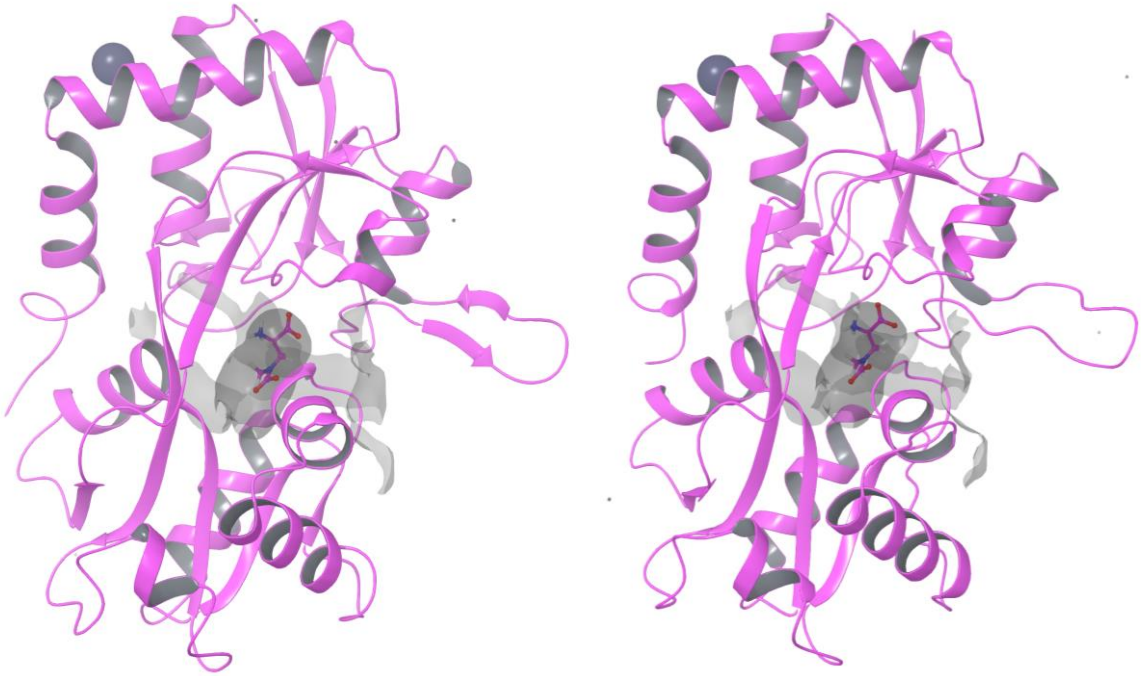

Figure 65. First cluster representative of L-β-carbamate/AMPA (left) and last frame of L-β-carbamate/AMPA of the production run.

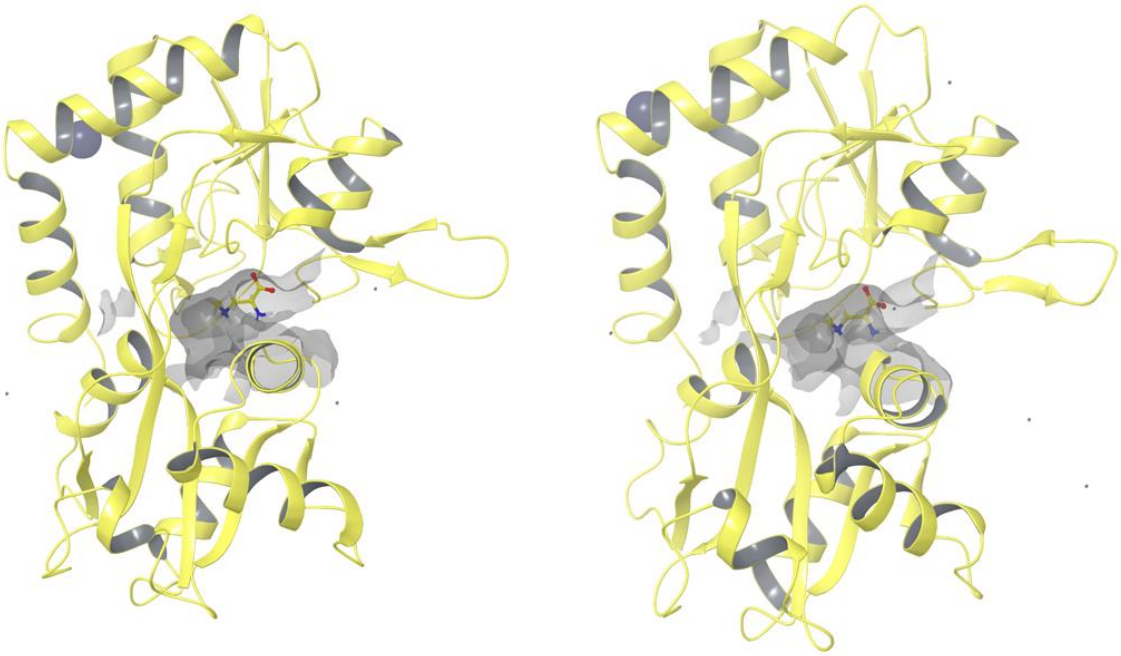

Figure 66. First cluster representative of L-BMAA/AMPA (left) and last frame of L-BMAA /AMPA of the production run.

Interactions between receptor and L-α-carbamate

The residues participating in electrostatic interactions are Thr91, Arg96, Ser142 and Glu193.

In Figure 67 there are two ion-dipole interactions and two salt bridges being observed. One ion-dipole interaction between the hydrogen of the hydroxyl group Thr91 and the carboxyl oxygen of L-α-carbamate, and another one between the hydrogen of the amide group of Ser142 and the oxygen of the second carboxyl group of the ligand. The salt bridges are being observed between the positive charged nitrogen of guanidinium group of Arg96 and the negative charged carboxyl oxygen of the ligand, and between the negative charged carboxyl oxygen of Glu193 and the positive charged nitrogen of L-α-carbamate.

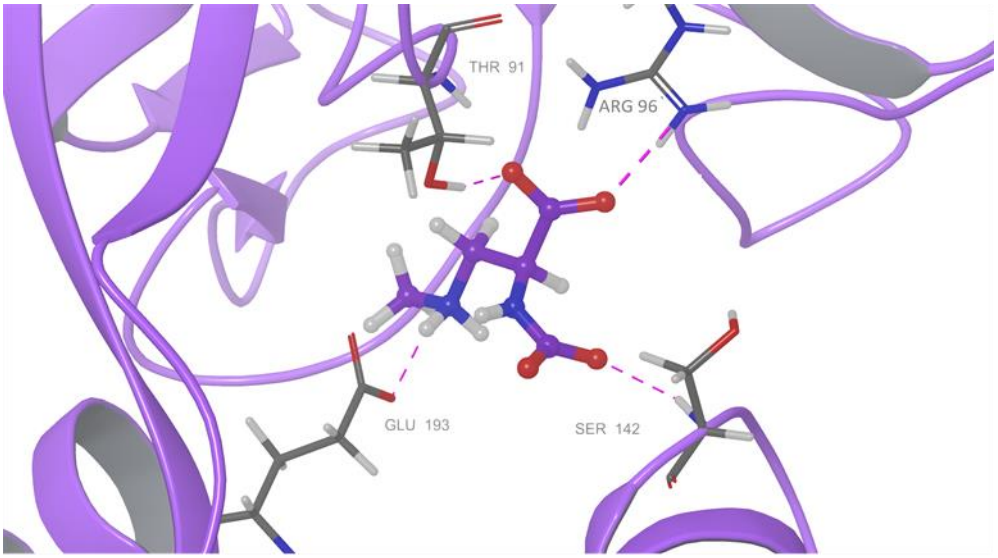

Figure 67. Interactions between L-α-carbamate and AMPA.

Interactions between receptor and L-β-carbamate

The residues participating in electrostatic interactions are Arg96, Thr91, Pro89, Ser142, Thr143 and Glu193.

In Figure 68 ion dipole interactions are depicted between the hydrogen of the amide group of Thr91 and the carboxyl oxygen of L-β-carbamate, hydroxyl hydrogen of Ser142 and the second carboxyl oxygen of the ligand, the hydrogen of the amide group of Ser142 and the oxygen of the second carboxyl group of the ligand, the hydrogen of the amide group of Thr142 and the carboxyl oxygen of the ligand, the hydrogen of the hydroxyl group of Thr142 and the second carboxyl oxygen of the ligand, between the carboxyl oxygen Glu193 and the hydrogen of the amide group L-β-carbamate, and between the carbonyl oxygen of Pro89 and the hydrogen of the amide group of the ligand. In addition, a salt bridge is forming between the positively charged nitrogen of guanidinium group of Arg96 and the negatively charged carboxyl oxygen of L-β-carbamate.

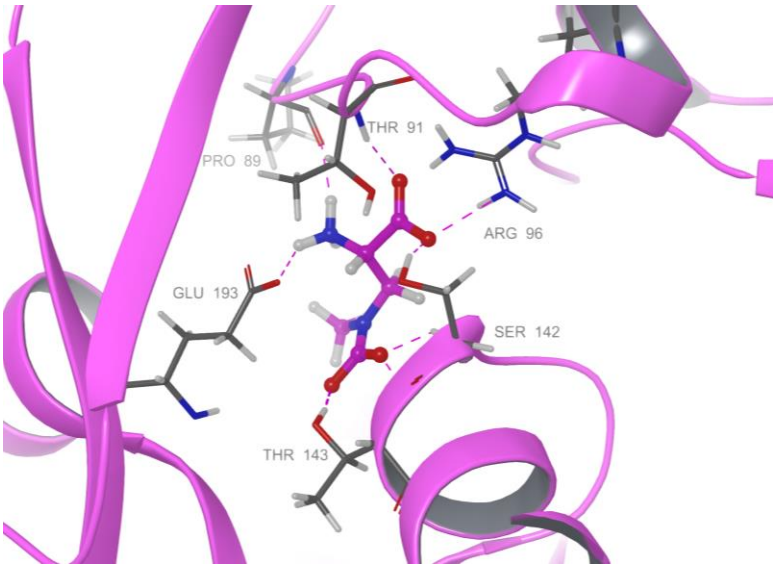

Figure 68. Interactions between L-β-carbamate and AMPA.

Interactions between receptor and L-BMAA

The residues participating in electrostatic interactions are Thr91 and Arg96.

There are only two electrostatic interactions between L-BMAA and the receptor depicted (Figure 69). An ion-dipole interaction is forming between the hydrogen of the amide group of Thr91 and the carboxyl oxygen of the ligand, and a salt bridge between the positive charged nitrogen of guanidinium group of Arg96 and the negative charged oxygen of the ligand.

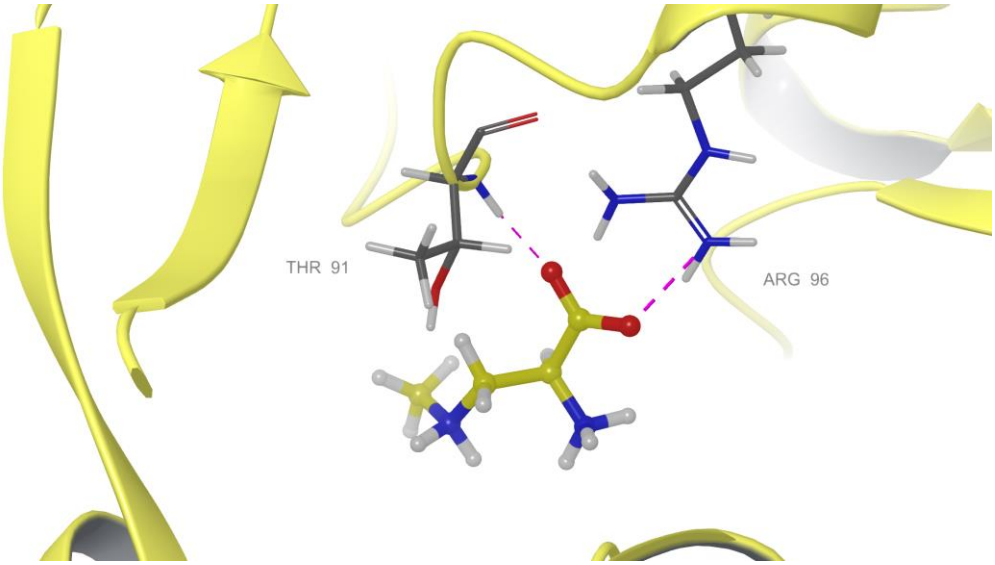

Figure 69. Interactions between L-BMAA and AMPA.

After inspecting the electrostatic interactions, it is known that L-β-carbamate has the greatest number of electrostatic interactions, then L-glutamate, then L-α-carbamate, and last L-BMAA with only two electrostatic interactions being observed. It is now understandable why L-β-carbamate is the molecule that remains most stable in the binding pocket of the receptor (1.2 ± 0.2 Å) and L-BMAA the one that remains the least (3.0 ± 0.8 Å).

3.4 MD simulations of D-ligands/AMPA receptor

Even though the L-ligands are the molecules that are believed to bind into glutamate's binding pocket, it is biologically interesting to study also the D-ligands in the AMPA receptor, since there could be racemic mixture of L-BMAA and D-BMAA. We will examine their stability in the binding pocket and inspect whether they have better affinity with respect to the L-ligands. In this step, we decided to use during the preparation of the systems, an extra method for placing the molecules in the binding pocket of the AMPA receptor. So, we have run MD simulations preparing the systems using superimposition/minimization with respect to L-glutamate and Glide Docking.

3.4.1 Analysis of D-ligands-AMPA receptor MD simulations from preparation with superimposition and minimization

Preparation of the system with superimposition and minimization

Chain A from crystal structure PDB:1FTJ and L-α-carbamate of BMAA were loaded in Maestro. Glutamate from the crystal structure PDB:1FTJ and D-α-carbamate were superimposed in Maestro using as reference structure glutamate, in order to place D-α-carbamate in the binding site of the receptor. Subsequently, D-α-carbamate and chain A without glutamate were merged, and the complex (chain A from PDB ID: 1FTJ and D-α-carbamate) was minimized using the Protein Preparation Wizard tool, provided by Maestro, so that the ligand could perform the appropriate interactions in the binding pocket of the receptor. The rest of the steps concerning the preparation of the systems were completed as described for the previous simulations. The same

procedure was followed also for D-β-carbamate, and D-BMAA. Three systems of ca. 50,000 were prepared, including ca. 15,300 water molecules each.

Results

The systems were minimized for 50,000 steps.

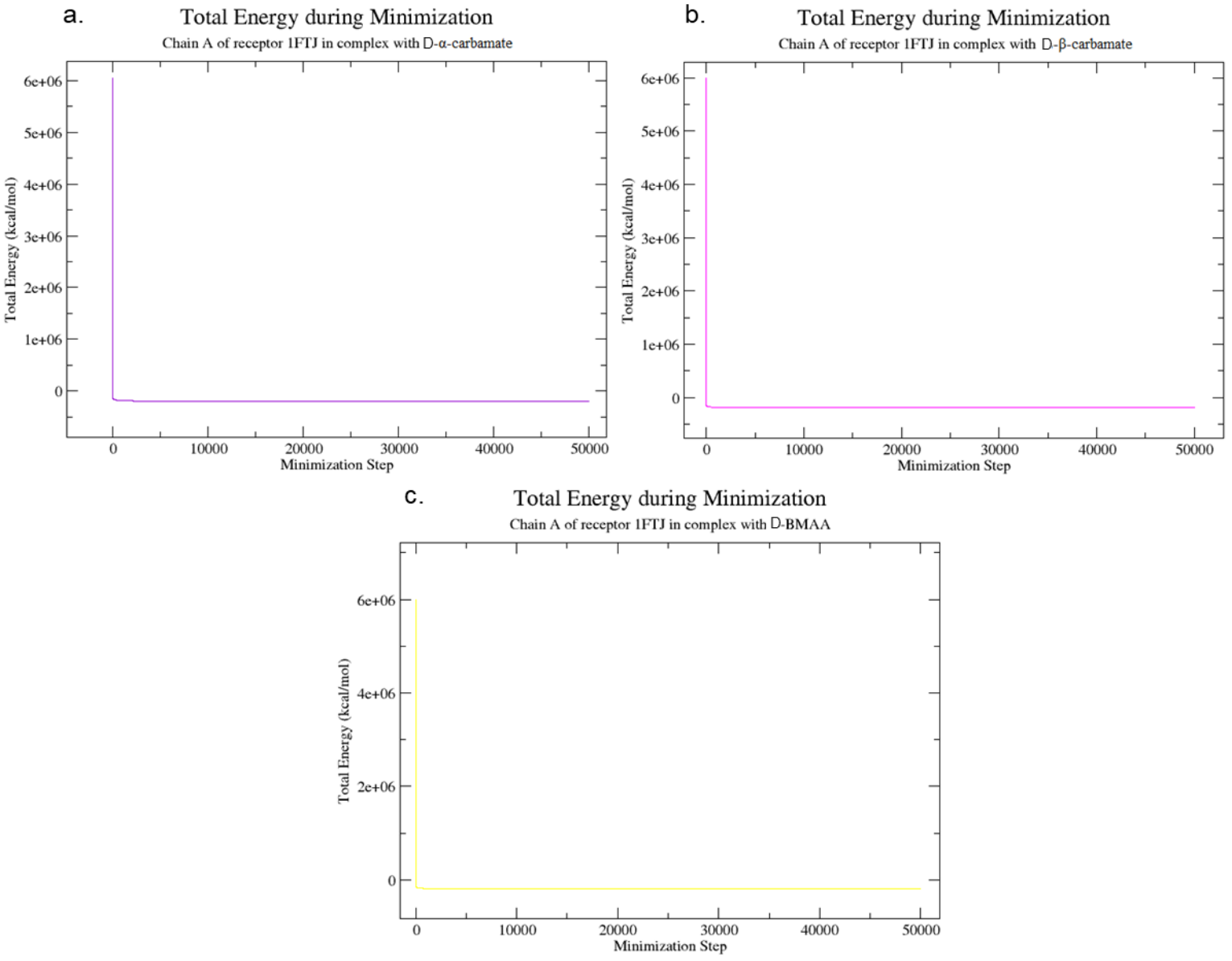

Figure 70. a) Total energy of D-α-carbamate/AMPA during minimization. b) Total energy of D-β-carbamate/AMPA during minimization. c) Total energy of D-BMAA/AMPA during minimization.

In Figures 70a, 70b, and 70c it is observed that after some steps the energy has reached plateau and so, minimization has converged in all cases.

Heating ran for 10,000 steps for each 10 K of temperature elevation and for a total 0.6 ns.

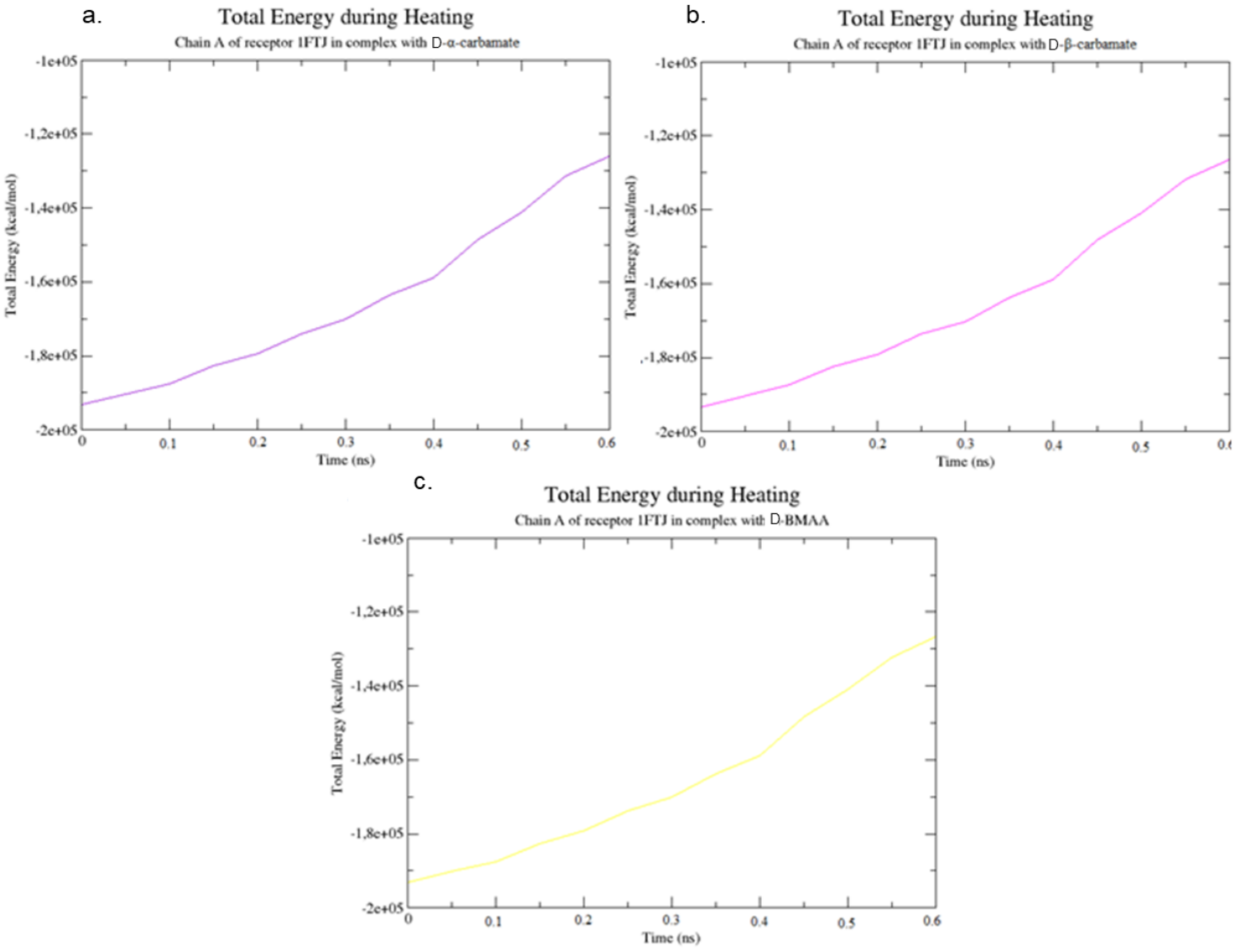

Figure 71. Total energy of D-α-carbamate/AMPA during the heating phase. b) Total energy of D-β-carbamate/AMPA during the heating phase. c) Total energy of D-BMAA/AMPA during the heating phase.

Figure 71 illustrates the total energy of the systems during the simulations.

Equilibration ran for 500,000 steps with a time step of 2 fs (1 ns).

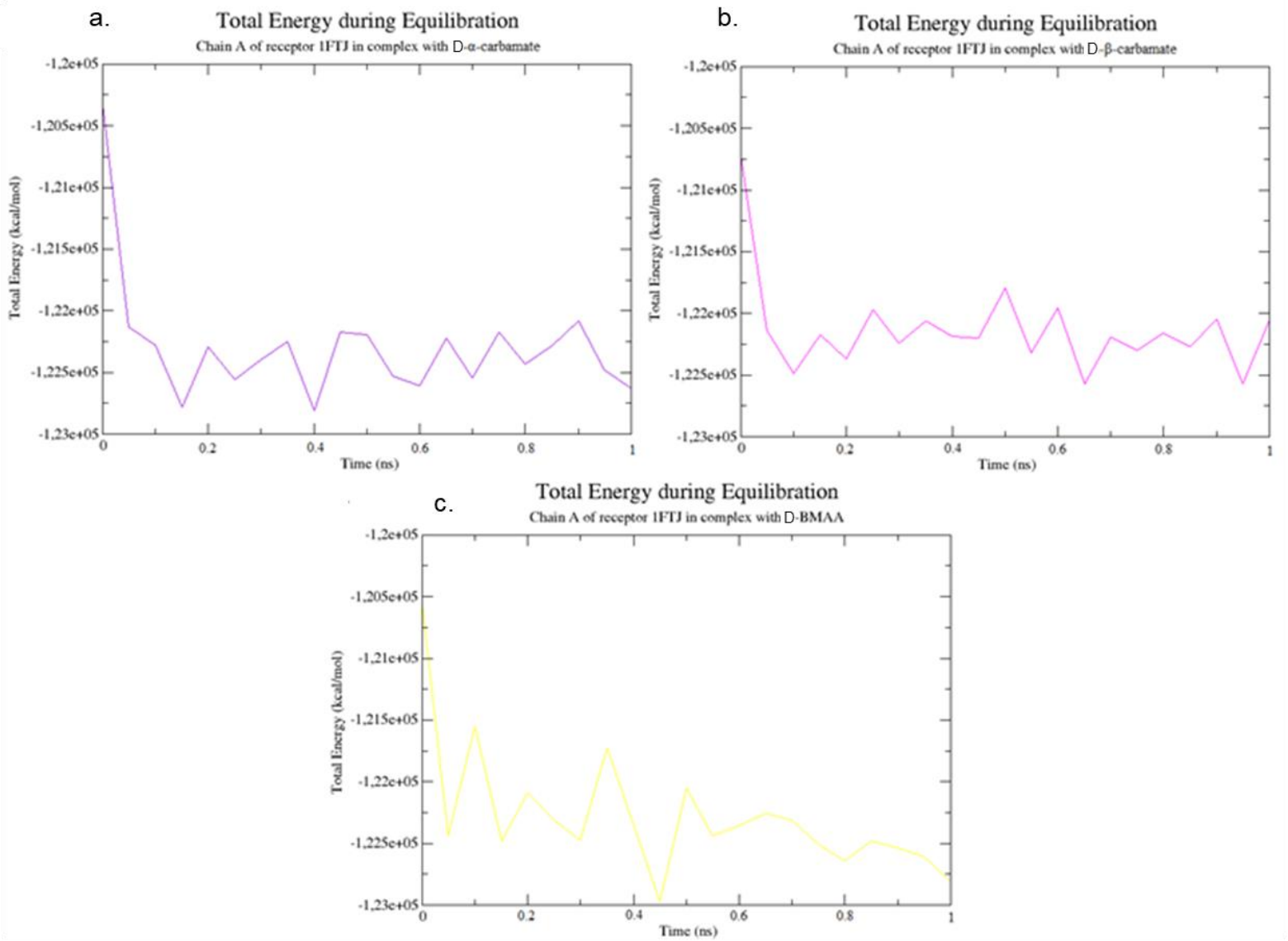

Figure 72. a) Total energy of D-α-carbamate/AMPA during equilibration. b) Total energy of D-β-carbamate/AMPA during equilibration. c) Total energy of D-BMAA/AMPA during equilibration.

Figure 72 illustrates the total energy vs time during the equilibration phase for all three systems. Particularly, the average of the total energy is -122,293 ± 497 kcal/mol for the structure with D-α-carbamate as ligand (Figure 72a), -122,132 ± 376 kcal/mol for the structure with D-β-carbamate as ligand (Figure 72b), and -122,283 ± 509 kcal/mol for the structure with D-BMAA as ligand (Figure 72c).

The simulation was run for 500,000,000 steps with a time step of 2 fs (100 ns).

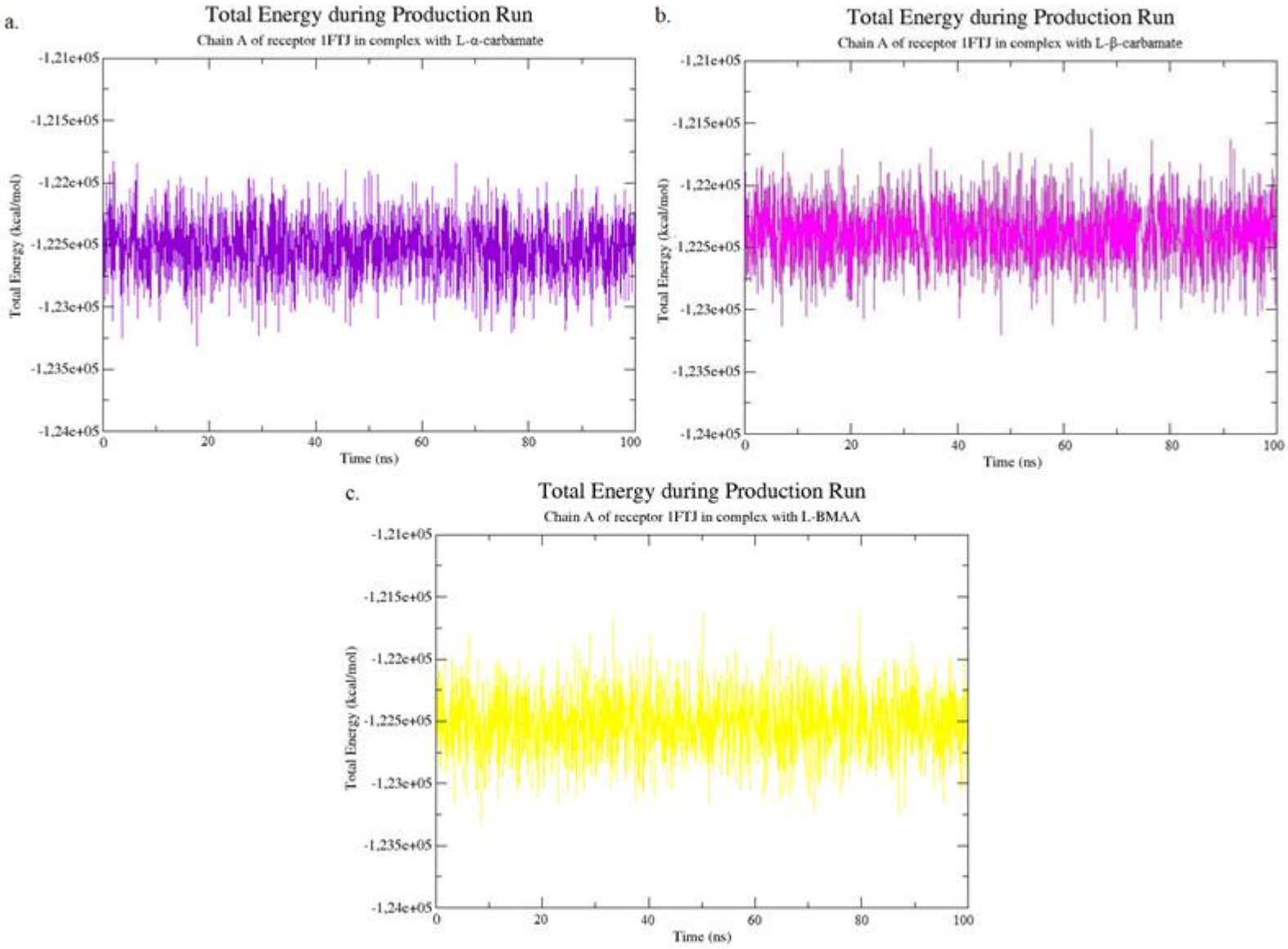

Figure 73. a) Total energy of D-α-carbamate/AMPA during production run. b) Total energy of D-β-carbamate/AMPA during production run. c) Total energy of D-BMAA/AMPA during production run.

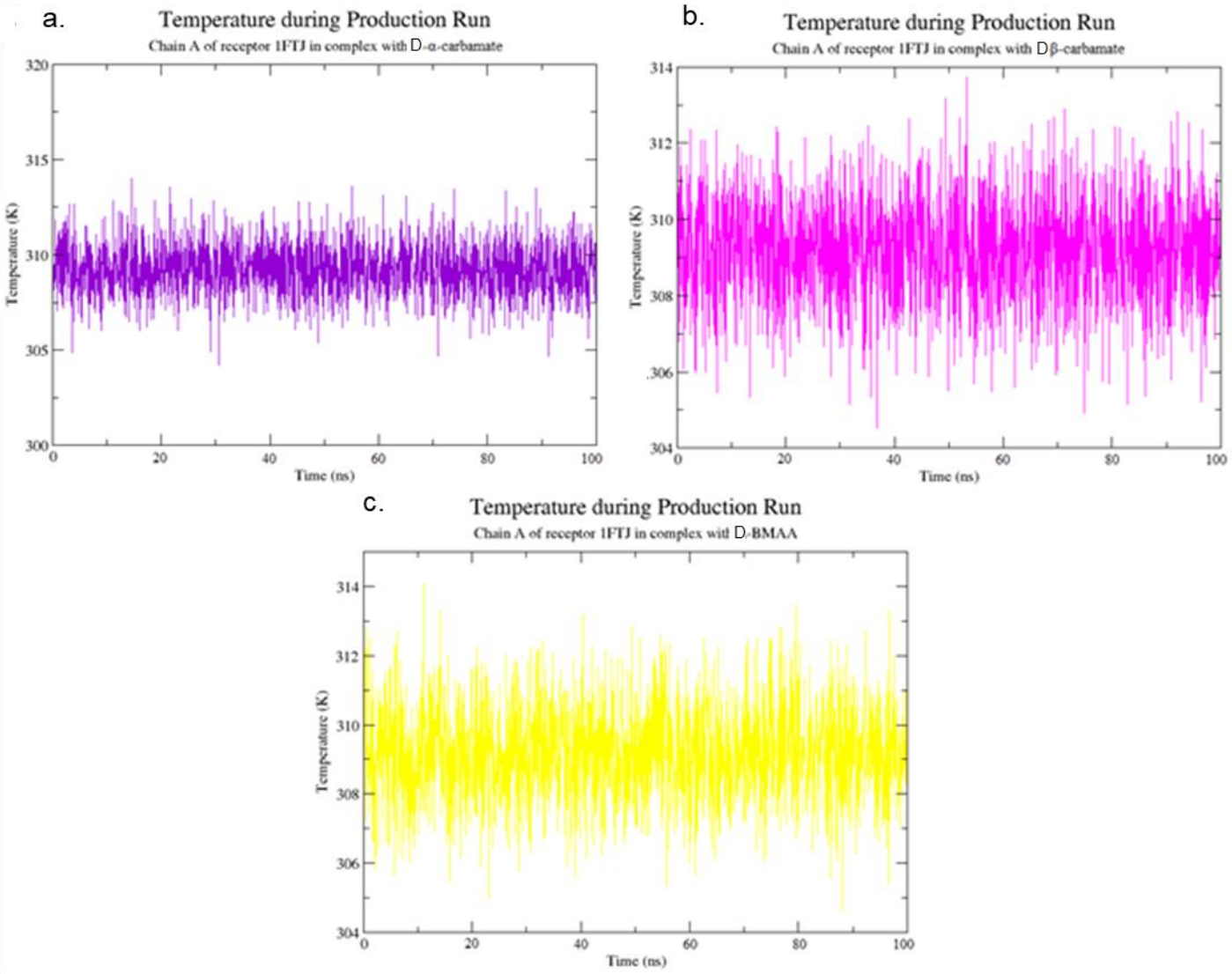

Figure 74. a) Temperature of D-α-carbamate/AMPA during production run. b) Temperature of D-β-carbamate/AMPA during production run. c) Temperature of D-BMAA/AMPA during production run.

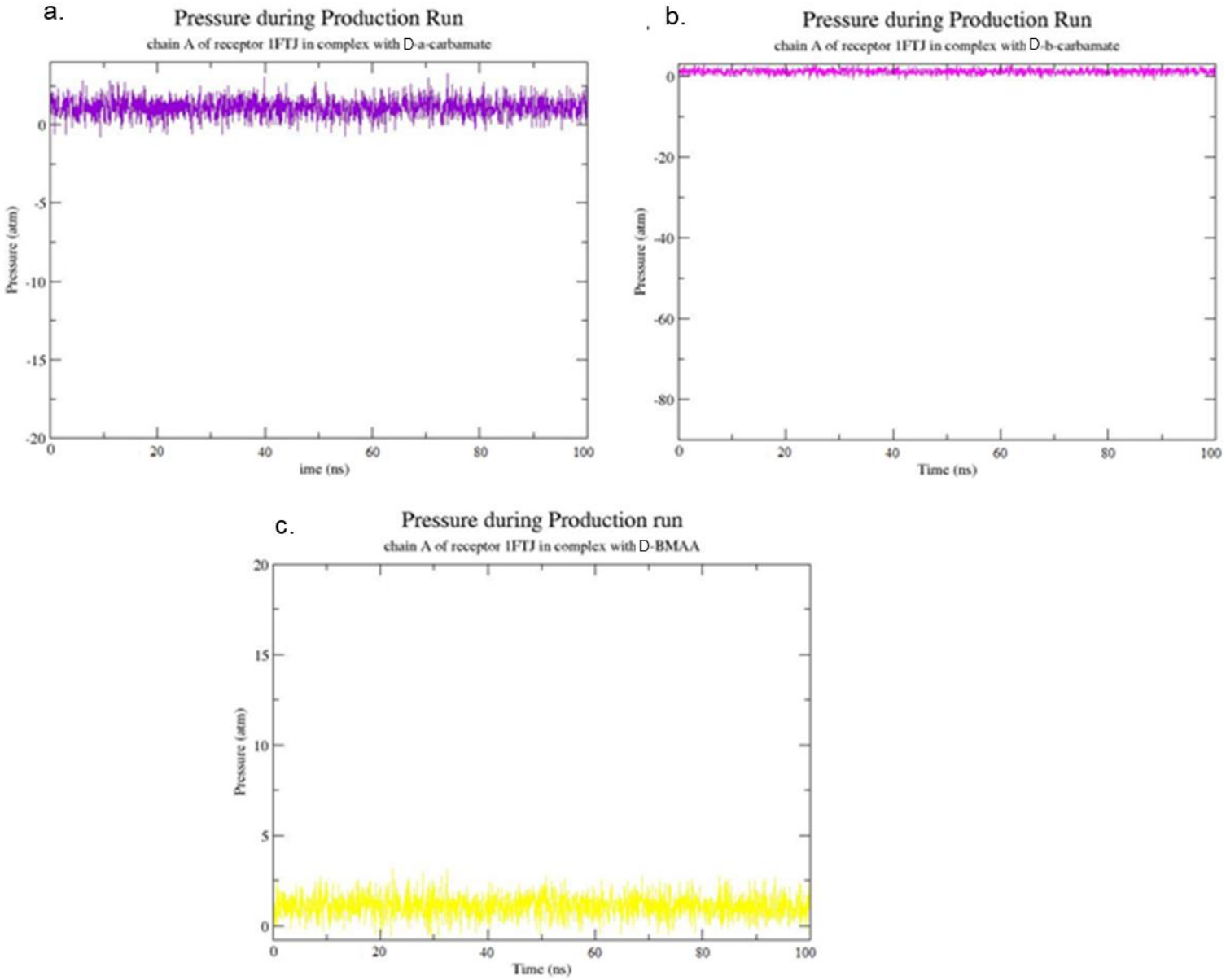

Figure 75. a) Pressure of D-α-carbamate/AMPA during production run. b) Pressure of D-β-carbamate/AMPA during production run. c) Pressure of D-BMAA/AMPA during production run.

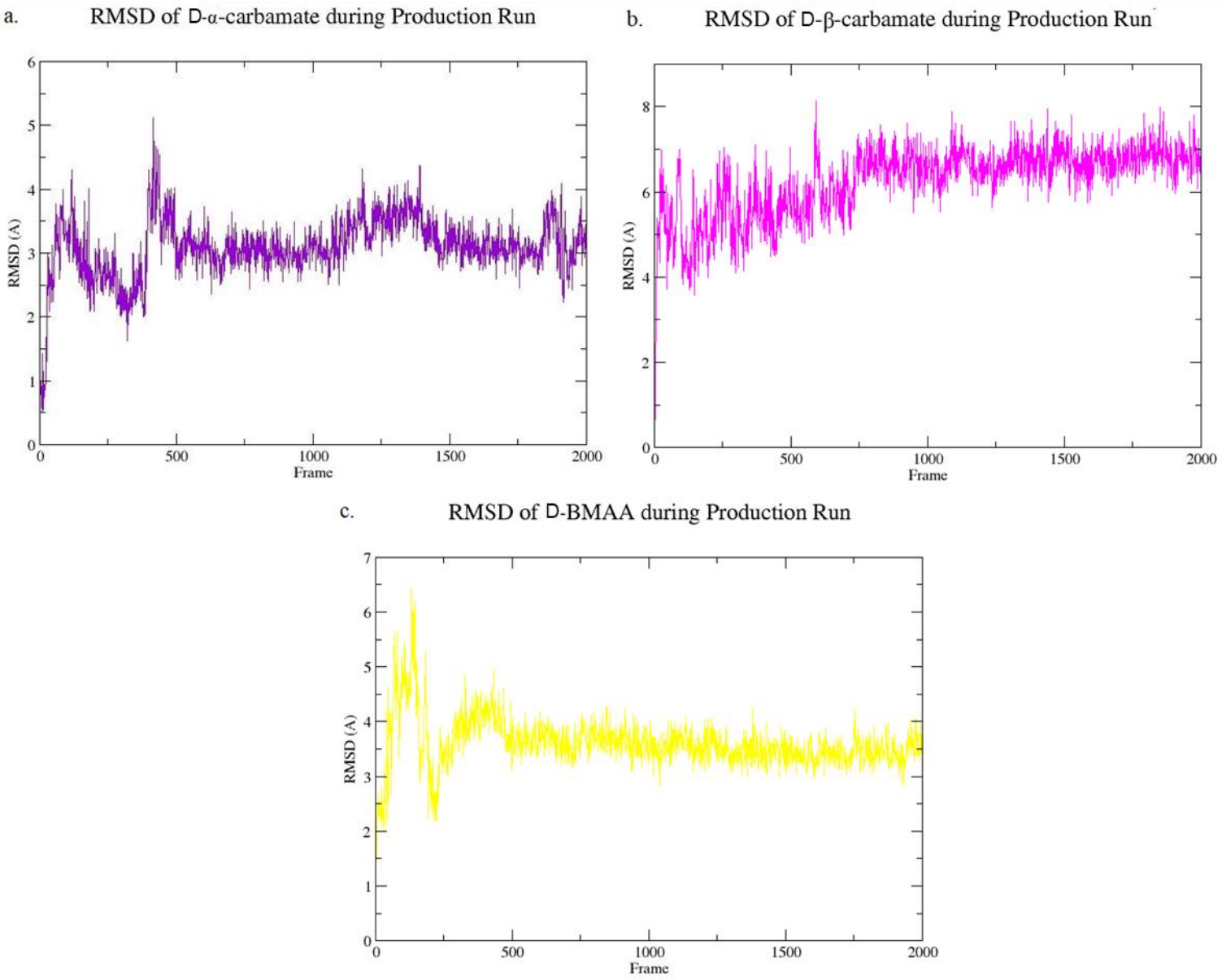

Figure 76. a) RMSD of D-α-carbamate in the binding site of AMPA during production run. b) RMSD of D-β-carbamate in the binding site of AMPA during production run. c) RMSD of D-BMAA in the binding site of AMPA during production run.

Figures 73a, 73b and 73c show the total energies of the three systems (receptor in complex with D-α-carbamate, receptor in complex with D-β-carbamate, and receptor in complex with D-BMAA) during the production run with average -122,530 ± 237 kcal/mol for the structure with D-α-carbamate, -122,382 ± 244 kcal/mol for the structure with D-β-carbamate, and -122,513 ± 243 kcal/mol for the structure with D-BMAA. The averages prove that all systems are at equilibration.

Figures 74a, 74b, and 74c show the change of the temperature during the production run. The average temperature is 309.2 ± 1.4 K for all systems, which reveal that the thermostat worked successfully.

Figures 75a, 75b, and 75c show the change of the pressure during the production run. The average pressure is 1.0 ± 0.4 atm for the system with D-α-carbamate as a ligand, 1.0 ± 0.0 atm for the system with D-β-carbamate as a ligand, and 1.0 ± 0.4 atm for the system with D-BMAA as a ligand. As a result, the barostat also worked successfully.

Figures 76a, 76b and 76c illustrate the RMSD time series of D-α-carbamate, D-β-carbamate, and D-BMAA respectively in the binding pocket of the receptor during the production run. RMSD average for D-α-carbamate is 3.1 ± 0.5 Å (Figure 76a), indicating that the ligand is mobile, however, it did not leave from the binding pocket of the receptor during the simulation. RMSD average for L-β-carbamate is 6.3 ± 0.8 Å (Figure 76b), which means that the ligand was more mobile in the binding site of the receptor than D-α-carbamate (RMSD ca. 3 Å) and glutamate (RMSD ca. 2 Å), and it left from the binding site. RMSD average for D-BMAA is 3.6 ± 0.5 Å (Figure 76c), which indicates that the ligand was also mobile, but it did not leave from the binding pocket of the receptor during the production run.

Analysis of trajectories – Analysis of interactions of last frames

The first cluster representative frames of the production runs were visualized in Maestro in order to inspect the interactions between the receptor and D-α-carbamate, D-β-carbamate, and D-BMAA. In addition, in Figure 77 the structures of the receptor AMPA in complex with D-α-carbamate, D-β-carbamate, and D-BMAA are presented, in order to inspect where the ligands are located after the production run. Particularly, Figure 77a illustrates the alignment of the last frame of production run of the receptor in complex with glutamate (green) and of the receptor in complex with D-α-carbamate (violet). It is observed that at the end of the production run, D-α-carbamate is located in the binding site as RMSD proved (RMSD = 3.3 Å). Figure 77b illustrates the alignment of the last frame of production run of the receptor in complex with glutamate (green) and of the receptor in complex with D-β-carbamate

(magenta). It is noticed that D-β-carbamate is not in the binding site of the receptor, which is confirmed from the RMSD that was calculated, 6.3 Å. In Figure 77c, the alignment of the receptor in complex with glutamate (green) and in complex with D-BMAA (yellow) is presented. D-BMAA is still located in the binding site of the receptor at the end of the production run as was mentioned before (RMSD = 3.6 Å).

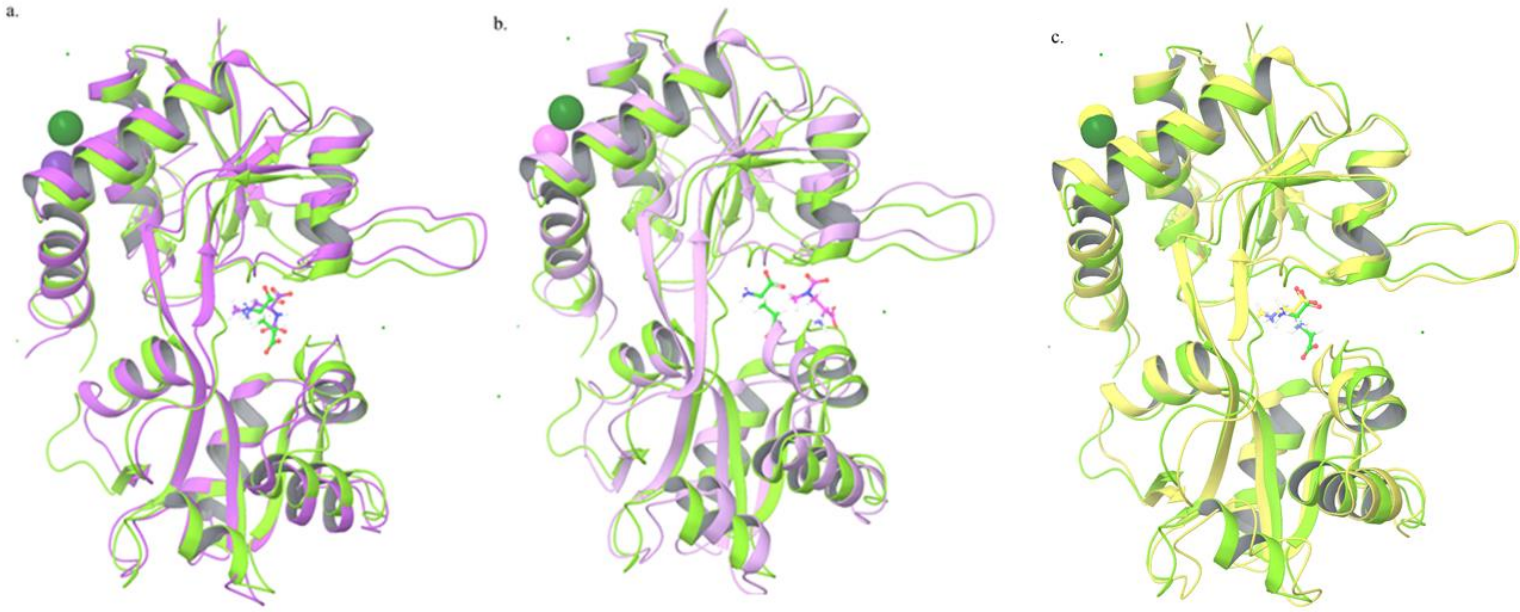

Figure 77. a) The alignment of the last frame of production run of the receptor in complex with glutamate (green) and of the receptor in complex with D-α-carbamate (violet). b) The alignment of the last frame of production run of the receptor in complex with glutamate (green) and of the receptor in complex with D-β-carbamate (pink). c) The alignment of the last frame of production run of the receptor in complex with glutamate (green) and of the receptor in complex with D-BMAA (yellow).

Interactions between receptor and D-α-carbamate

The residues participating in electrostatic interactions are Arg96, Thr91, Tyr61 and Glu193.

Hydrogen bonds are observed, depicted in blue dashed lines, between the two carboxyl oxygen atoms of the ligand and the two hydrogen atoms of the guanidinium group of Arg96, the carboxyl oxygen of the ligand and the hydrogen of the hydroxyl group of Thr91, the carboxyl oxygen of the ligand and the hydrogen of the amino group of Thr91, and the hydrogen of the positive charged amino group of the ligand and the carboxyl oxygen of Glu193. Moreover, salt-bridge is forming, depicted in pink dashed lines,

between the negative charged carboxyl oxygen the ligand and the positive charged nitrogen of guanidinium group of Arg96, and the positive charged nitrogen of the amino group of the ligand and the negative charged oxygen of carboxyl group of Glu193, as well as cation-π interactions are presented, colored in green dashed line, between the positive charged nitrogen of the ligand and the benzene ring of Tyr61 (Figure 78).

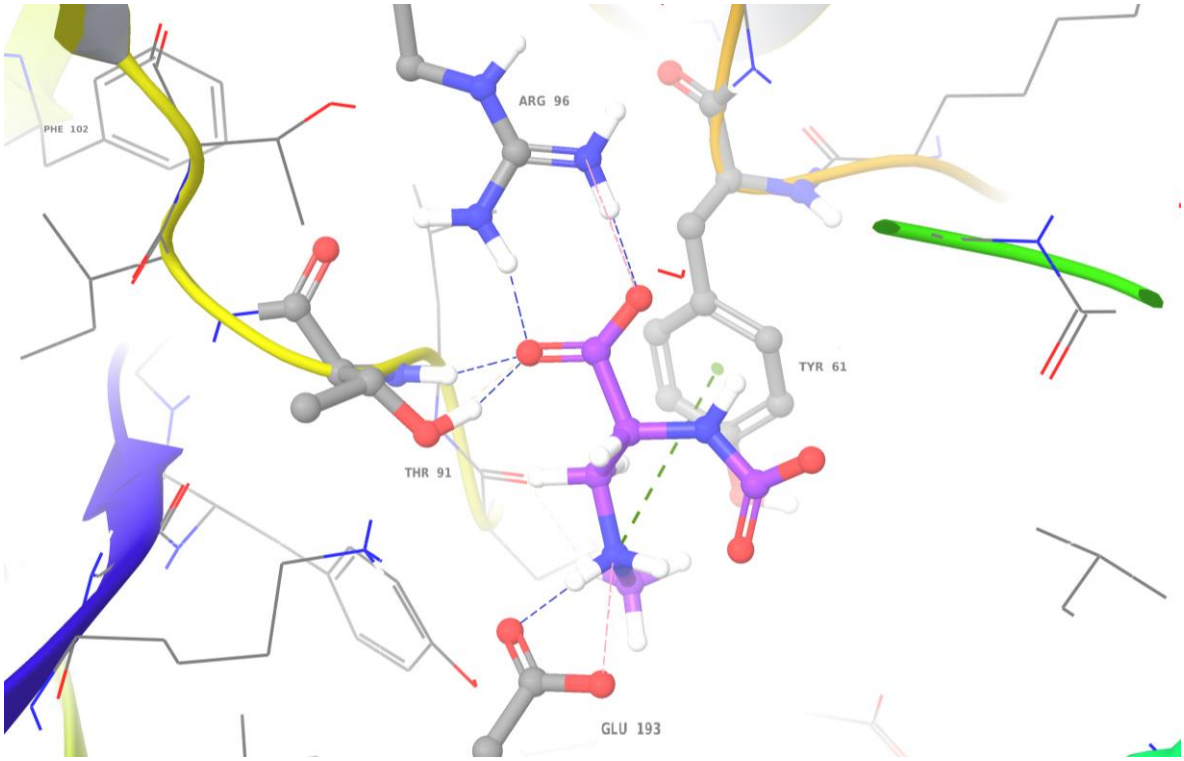

Figure 78. Interactions between the receptor and D-α-carbamate at the last frame of the production run.

Interactions between receptor and D-β-carbamate

The residues participating in electrostatic interactions are Arg96, Ser140, Gly141 and Glu145.

Hydrogen bonds, colored in blue dashed lines, are observed between the carboxyl oxygen of the ligand and the hydrogen of guanidinium group of Arg96, the hydrogen of the positive charged nitrogen of the ligand and the carbonyl oxygen of Gly141, and the hydrogen of the positive charged nitrogen of the ligand and the carbonyl oxygen of Ser140. In addition, salt-bridge are forming, colored in pink dashed lines, between the negative charged oxygen of the ligand and the positive charged nitrogen of guanidinium group of Arg96,

and the positive charged nitrogen of the ligand and the negative charged carboxyl oxygen of Glu145 (Figure 79).

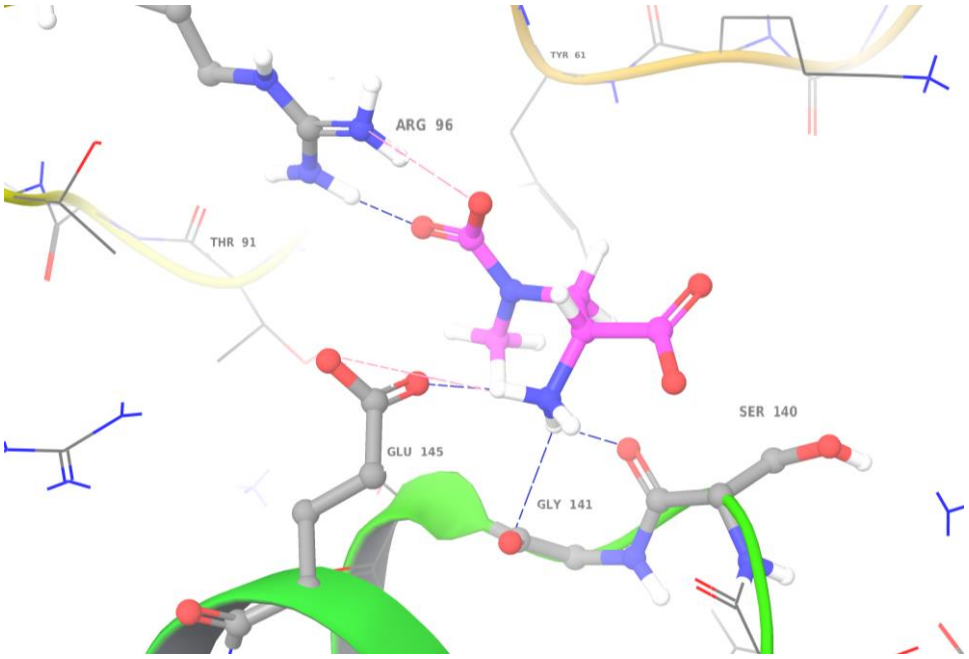

Figure 79. Interactions between the receptor and D-β-carbamate at the last frame of the production run.

Interactions between receptor and D-BMAA

The residues participating in electrostatic interactions are Arg96, Thr91, Tyr61, Glu193 and Ser142.

Hydrogen bonds are being observed, colored in blue dashed lines, between the two carboxyl oxygen atoms of the ligand and two hydrogens of guanidinium group of Arg96, the carboxyl oxygen of the ligand and the hydrogen of the amino group of Thr91, the hydrogen of the terminal amino group of the ligand and the hydroxyl oxygen of Ser142, the hydrogen of the amino group of the ligand and the oxygen of Glu193, and the hydrogen of the positive charged amino group of the ligand and the oxygen of Glu193. In addition, salt-bridge interactions are being observed, depicted in pink dashed lines, between the negative charged carboxyl oxygen of the ligand and the positive charged nitrogen of guanidinium group of Arg96, the two positive charged nitrogen atoms of the ligand and negative charged oxygen of Glu193, as well as cation-π interactions are being presented, in green dashed lines, between the two positive charged nitrogen atoms of D-BMAA and the benzene ring of Tyr61 (Figure 80).

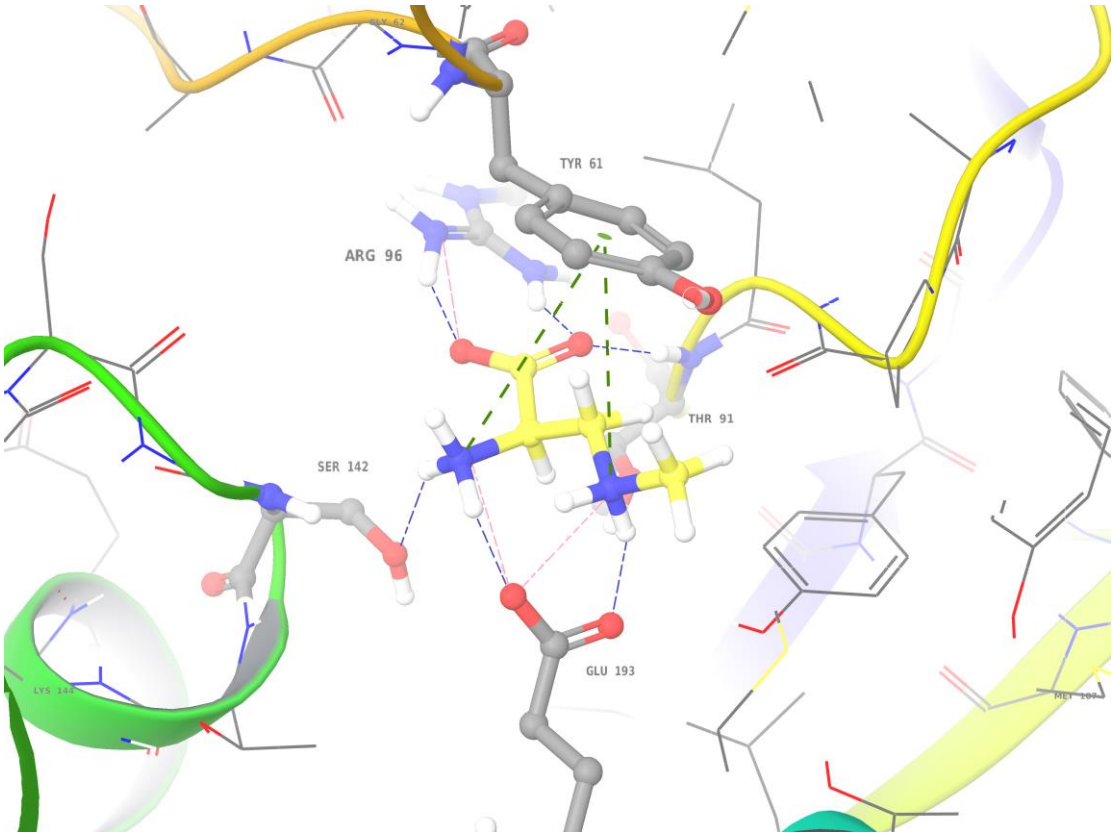

Figure 80. Interactions between the receptor and D-BMAA of the last frame of the production run.

Similarities of interactions of D-α-carbamate, D-β-carbamate and D-BMAA with the receptor with respect to glutamate's

D-α-carbamate and glutamate have common residues, Arg96 and Tyr61, which participate in the same interactions. D-β-carbamate and glutamate have only Arg96 as common residues that participate in the same interactions, and D-BMAA and glutamate have common residues Arg96, Tyr61 and Ser142, which participate in the same interactions. As a result, D-β-carbamate has the least common interactions with glutamate. In particular, for the interaction with residue Tyr61, which exist in all rest of the cases, the benzene ring is in a distance of 7.28 Å (Figure 81), when in all other cases is less than 5 Å (Figure 82).

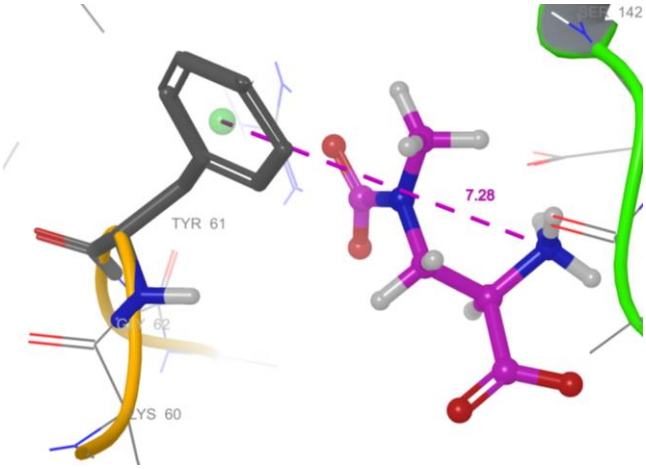

Figure 81. Distance for cation-π interaction between Tyr61 and positive charged nitrogen of L-β-carbamate.

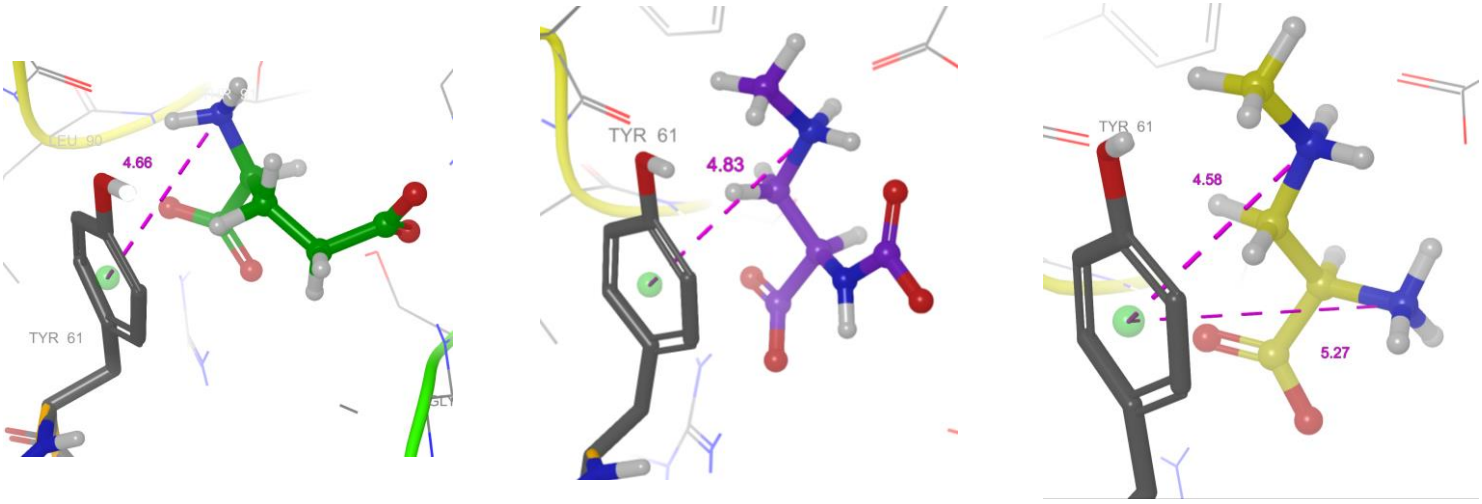

Figure 82. Distance for cation-π interaction between Tyr61 and positive charged nitrogen of glutamate (left), D-β-carbamate (middle) and D-BMAA respectively (right).

Trajectory Snapshots at the end of the simulations

By inspecting the poses of the last frames, it is observed that at the end of the simulation glutamate, D-α-carbamate, and D-BMAA are located in the binding site (Figures 83a, 83b, 83d), while D-β-carbamate seems to leave the receptor, since the protein has opened and a big part of D-β-carbamate interacts with the solution (Figure 83c).

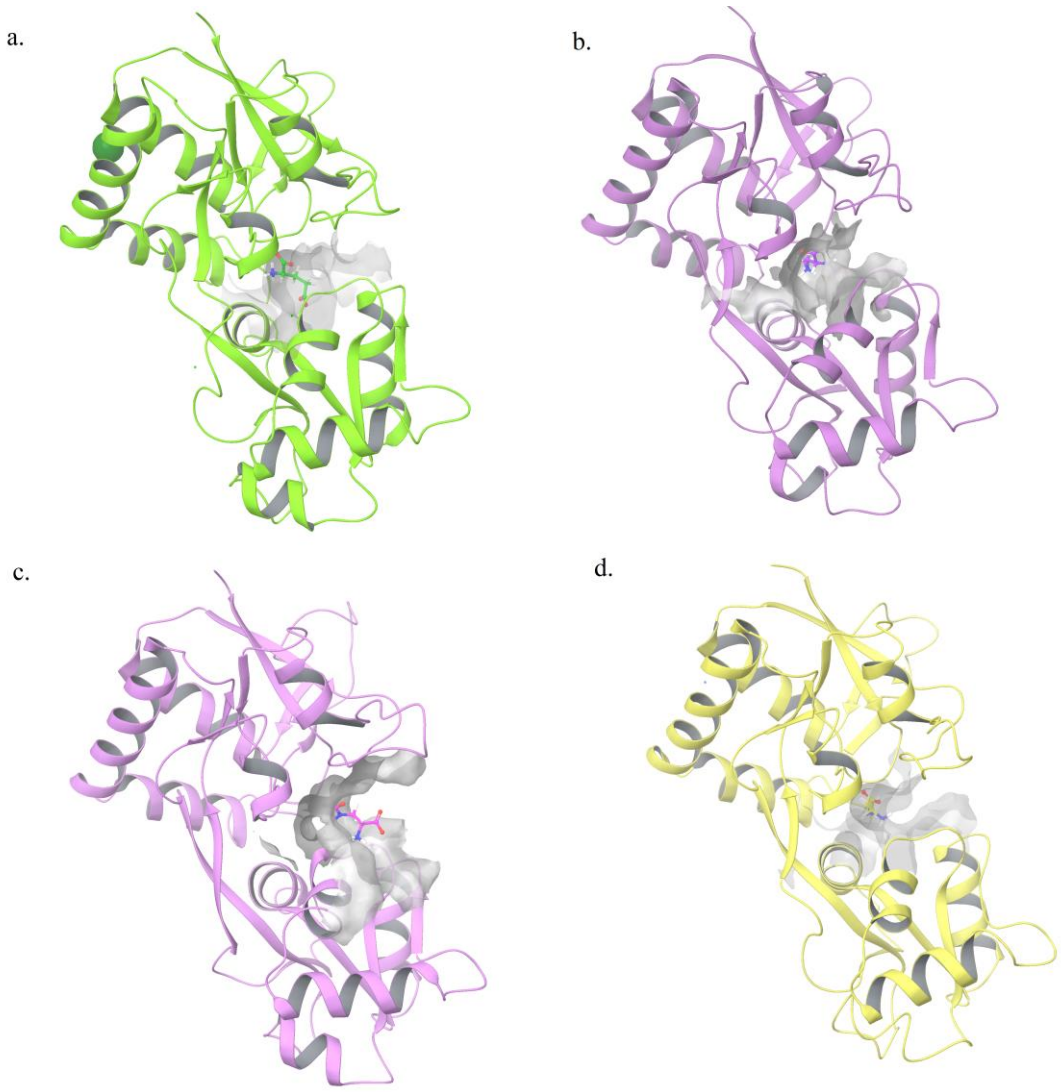

Figure 83. Proteins after production run with binding site surfaces. a) receptor in complex with L-glutamate (green), b) receptor in complex with D-a-carbamate, c) receptor in complex with D-β-carbamate and d) receptor in complex with D-BMAA.

It is remarkable that D-β-carbamate, which shares the highest structural similarities with glutamate between D-α-carbamate and D-BMAA (Figure 62), is the one that leaves the binding site of the receptor. This may be due to the method that was used to place the ligand in the receptor. By superimposing glutamate and D-β-carbamate, the latter might was forced to be in a pose that was less desirable even though the structure was minimized after the placement of D-β-carbamate in the receptor.

3.4.2 Analysis of D-ligands-AMPA receptor MD simulations from preparation with docking

Similarity between glutamate, D-α-carbamate, D-β-carbamate and D-BMAA

Cross tanimoto coefficient was calculated for molecules glutamate, D-α-carbamate, D-β-carbamate and D-BMAA. The results are presented in Table 2 below.

Table 2. Cross tanimoto of glutamate, D-α-carbamate, D-β-carbamate and D-BMAA.

Cross Tanimoto				
Molecule	Glutamate	α-carbamate	β-carbamate	BMAA
Glutamate	1	**0.48**	**0.51**	**0.45**
α-carbamate	0.48	1	0.87	0.82
β-carbamate	0.51	0.87	1	0.70
BMAA	0.45	0.82	0.70	1

The tanimoto that was measured is ca. 0.50 in all cases, so Glide Docking method can be used in order to place the compounds D-α-carbamate, D-β-carbamate, and D-BMAA in the binding pocket of the AMPA receptor.

Preparation of the systems

Chain A from crystal structure PDB:1FTJ and D-α-carbamate of BMAA were loaded in Maestro. D-α-carbamate, D-β-carbamate, and D-BMAA were placed in the binding pocket of the receptor by docking using Glide. The complexes were prepared with Protein Preparation Wizard tool provided by Schrödinger, where the missing side chains were filled in using Prime, water molecules were deleted, and protein was optimized and minimized. The rest of the procedure followed as described previously. The force field that was

used for the simulations is CHARMM36 both for the protein and the ligand. Three systems were prepared each one consisting of ca. 50,000 atoms, including ca. 15,300 water molecules

Results

Minimization run for 50,000 steps.

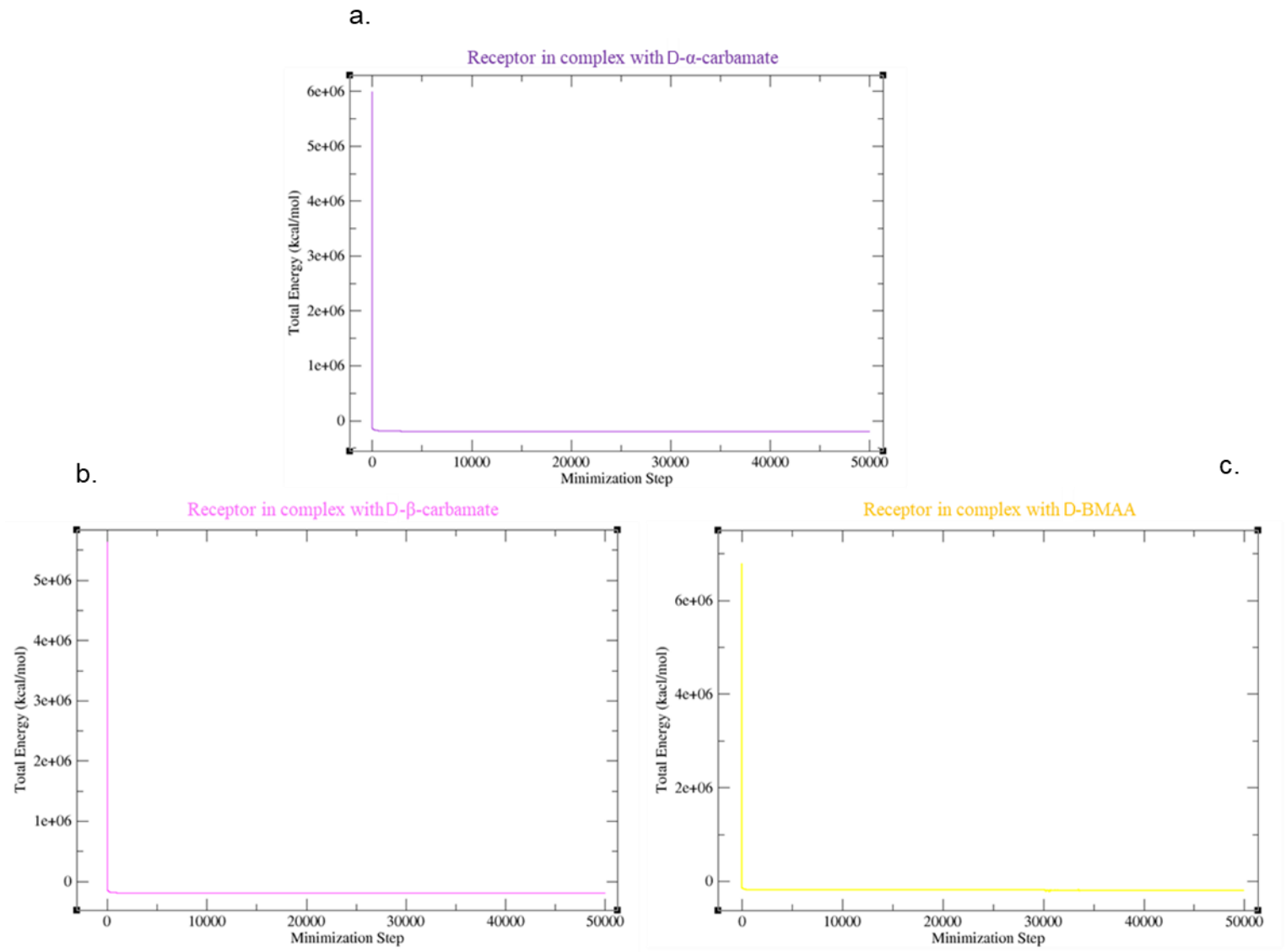

Figure 84. a) Total energy of D-α-carbamate/AMPA during minimization. b) Total energy of D-β-carbamate/AMPA during minimization. c) Total energy of D-BMAA/AMPA during minimization.

In Figure 84 it is observed that after some steps the energy has reached plateau and so, minimization has converged in all cases.

Heating ran for 10,000 steps for each 10 K of temperature elevation and for a total 0.6 ns.

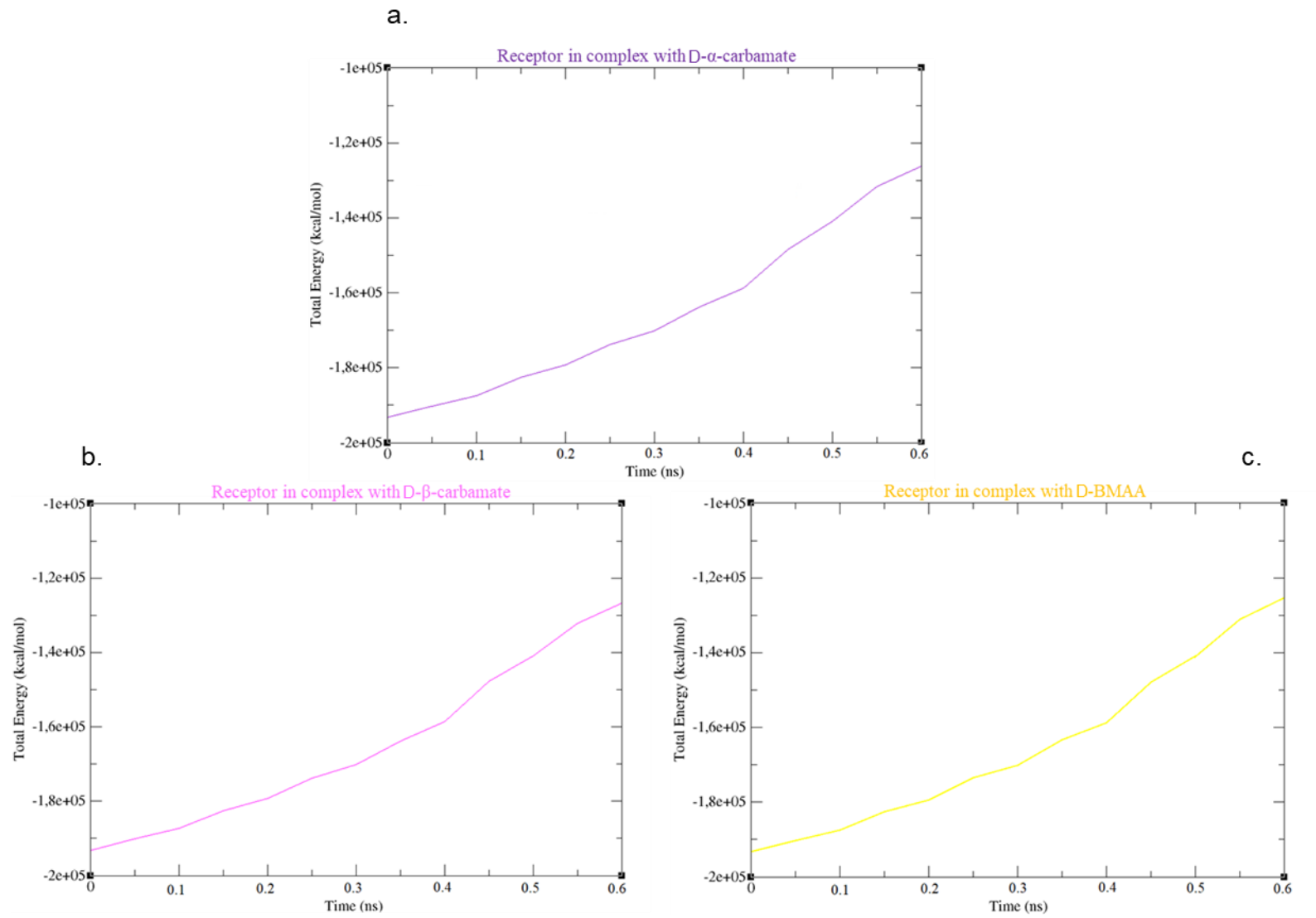

Figure 85. a) Total energy of D-α-carbamate/AMPA during the heating phase. b) Total energy of D-β-carbamate/AMPA during the heating phase. c) Total energy of D-BMAA/AMPA during the heating phase.

Figure 85 illustrates the total energy of the systems during the simulations.

Equilibration ran for 500,000 steps with a time step of 2 fs (1 ns).

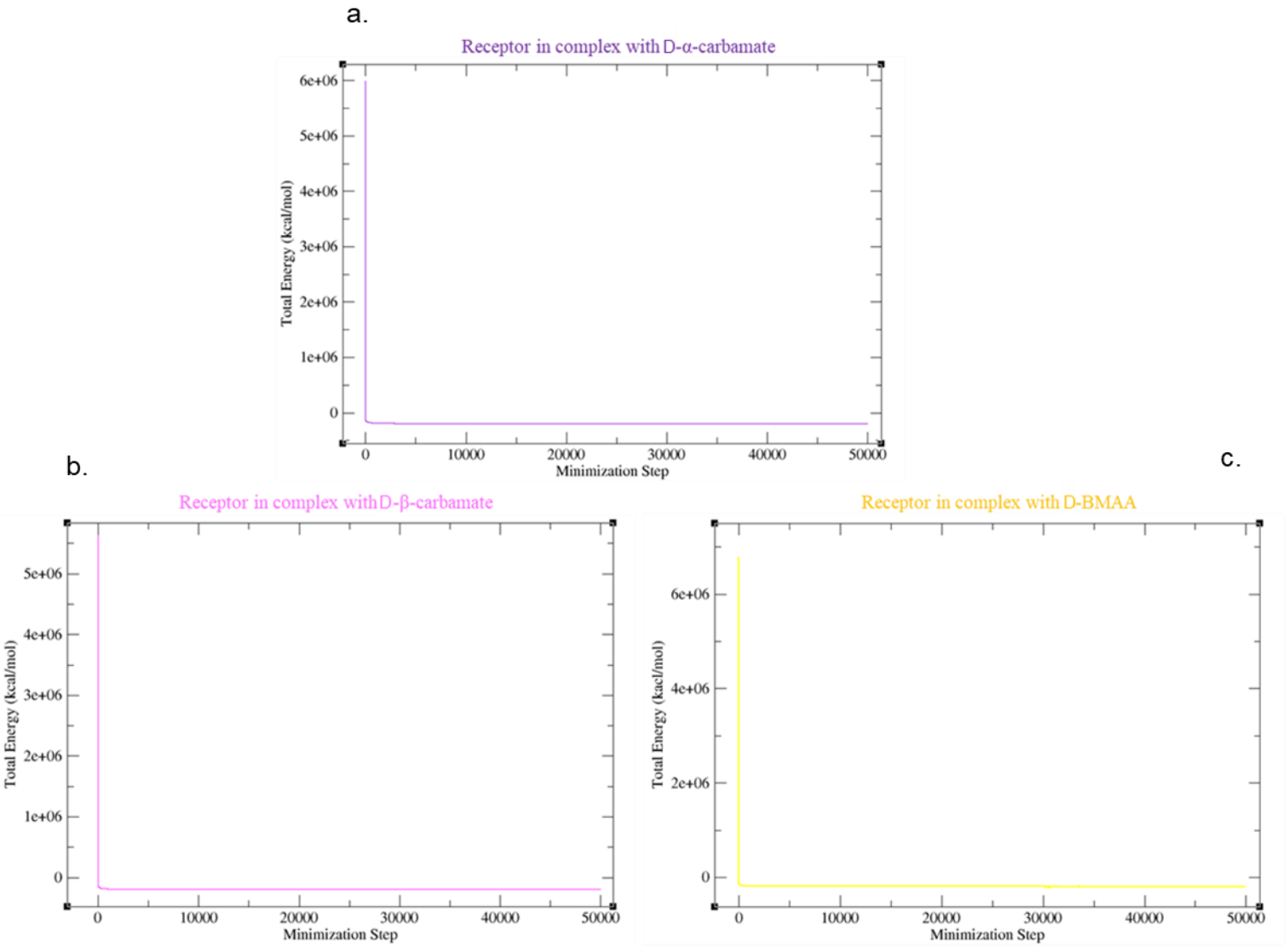

Figure 86. a) Total energy of D-α-carbamate/AMPA during equilibration. b) Total energy of D-β-carbamate/AMPA during equilibration. c) Total energy of D-BMAA/AMPA during equilibration.

Figure 86 illustrates the total energy vs time during the equilibration phase for all three systems. Particularly, the average of the total energy is -122250 ± 461 kcal/mol for the structure with D-α-carbamate as ligand (Figure 86a), -122115 ± 463 kcal/mol for the structure with D-β-carbamate as ligand (Figure 86b), and -122208 ± 474 kcal/mol for the structure with D-BMAA as ligand (Figure 86c).

The simulation was run for 500,000,000 steps with a time step of 2 fs (100 ns).

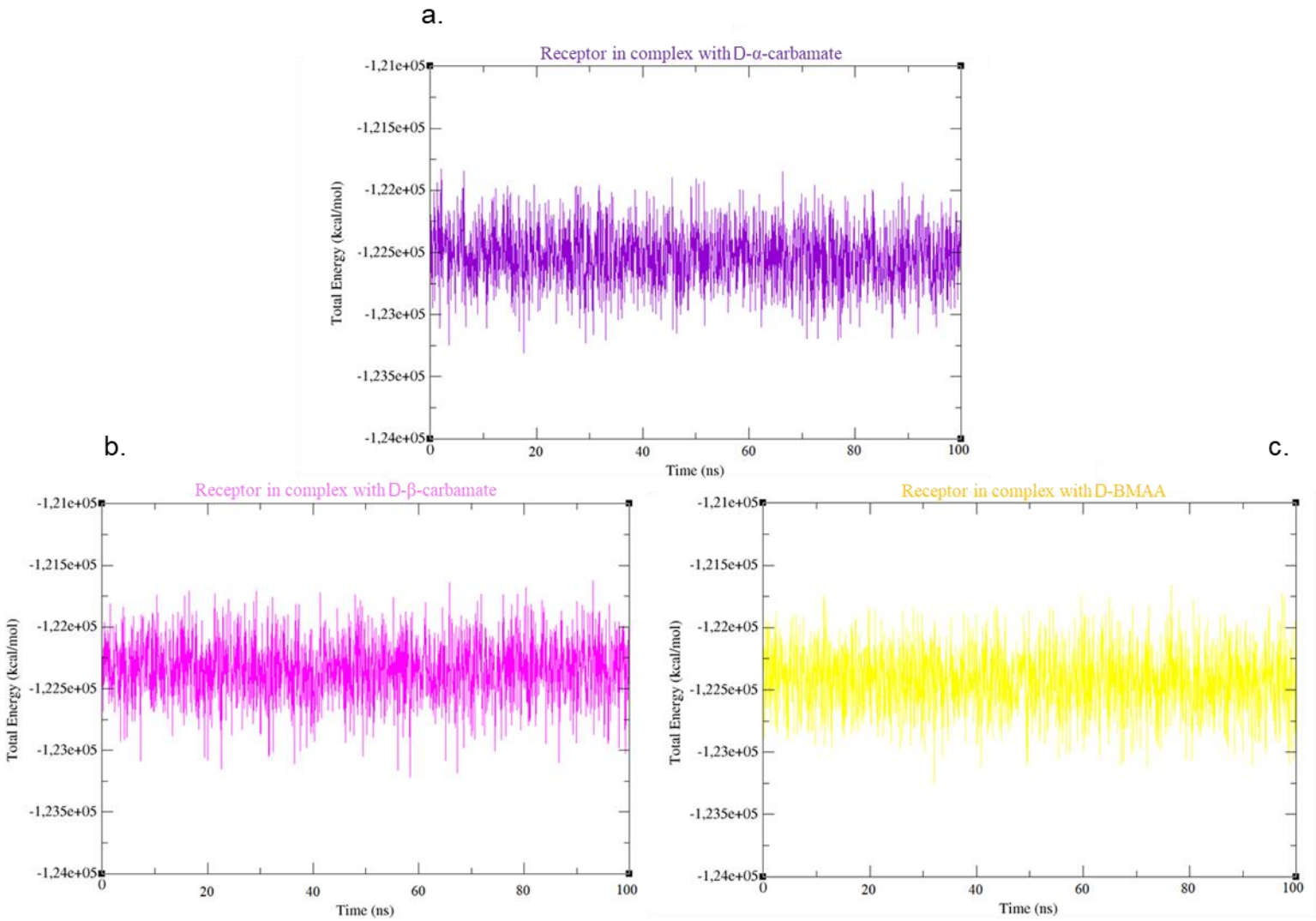

Figure 87. Total energy of D-α-carbamate/AMPA during production run. b) Total energy of D-β-carbamate/AMPA during production run. c) Total energy of D-BMAA/AMPA during production run.

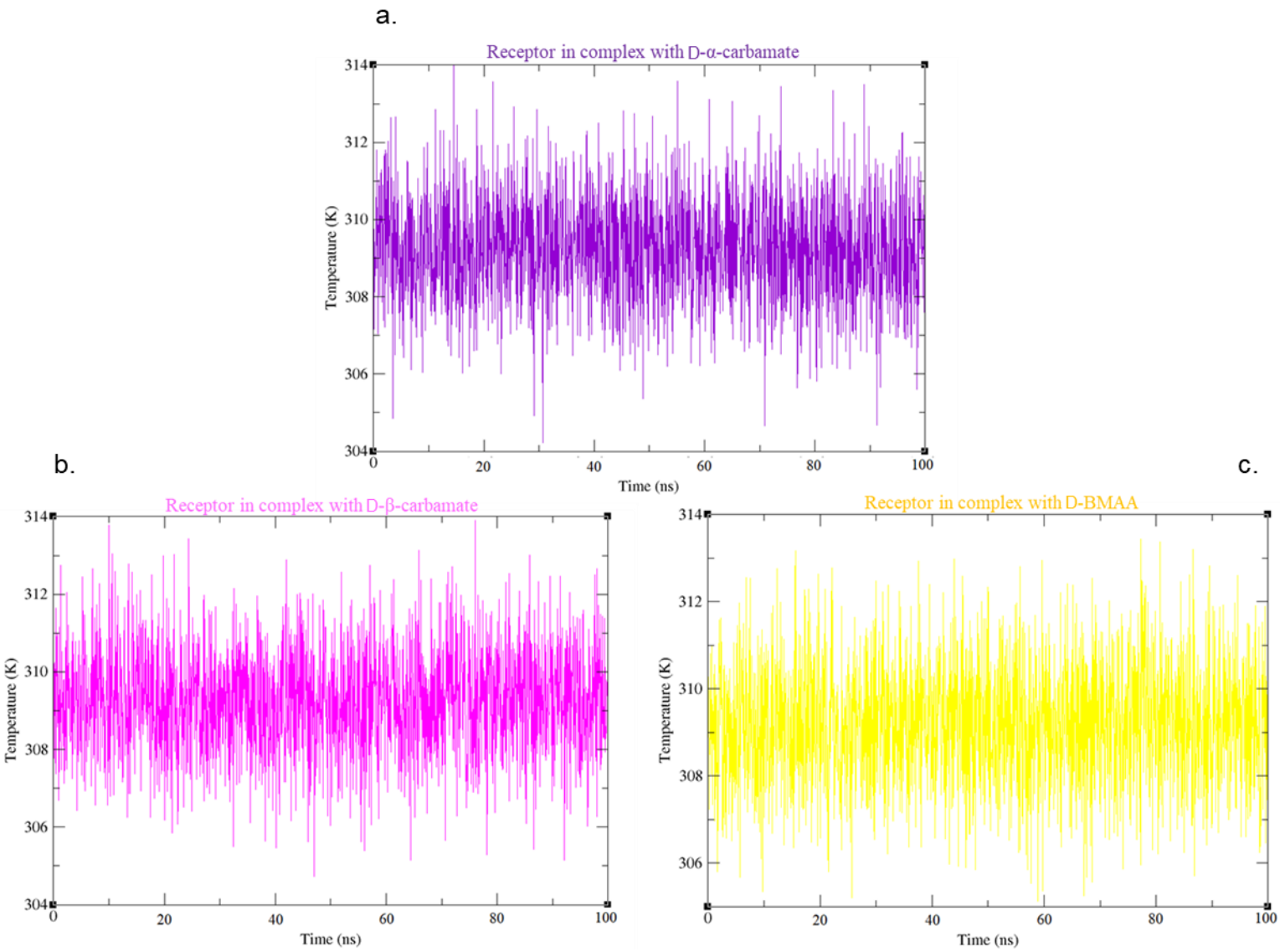

Figure 88. Temperature of D-α-carbamate/AMPA during production run. b) Temperature of D-β-carbamate/AMPA during production run. c) Temperature of D-BMAA/AMPA during production run.

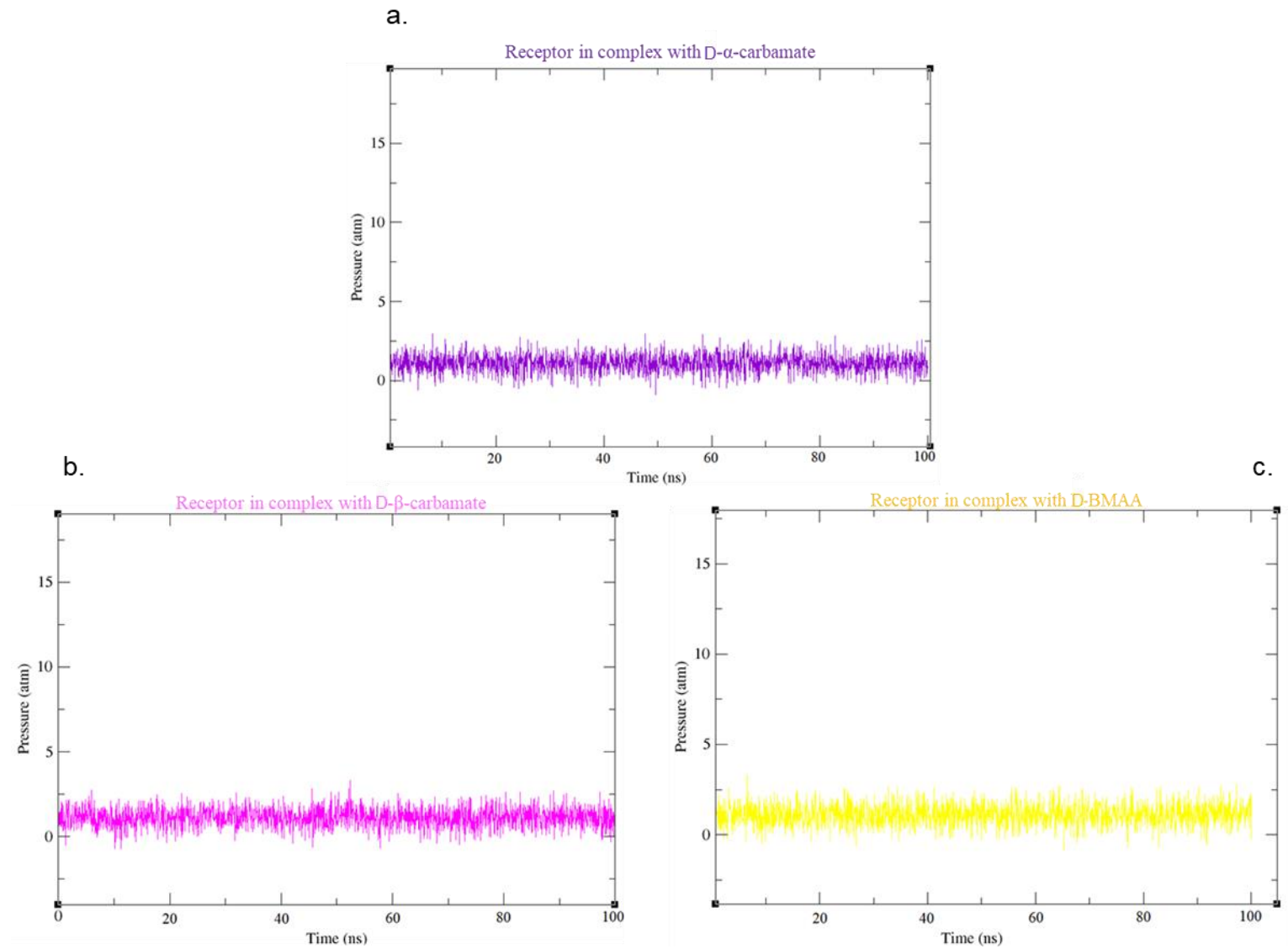

Figure 89. Pressure of D-α-carbamate/AMPA during production run. b) Pressure of D-β-carbamate/AMPA during production run. c) Pressure of D-BMAA/AMPA during production run.

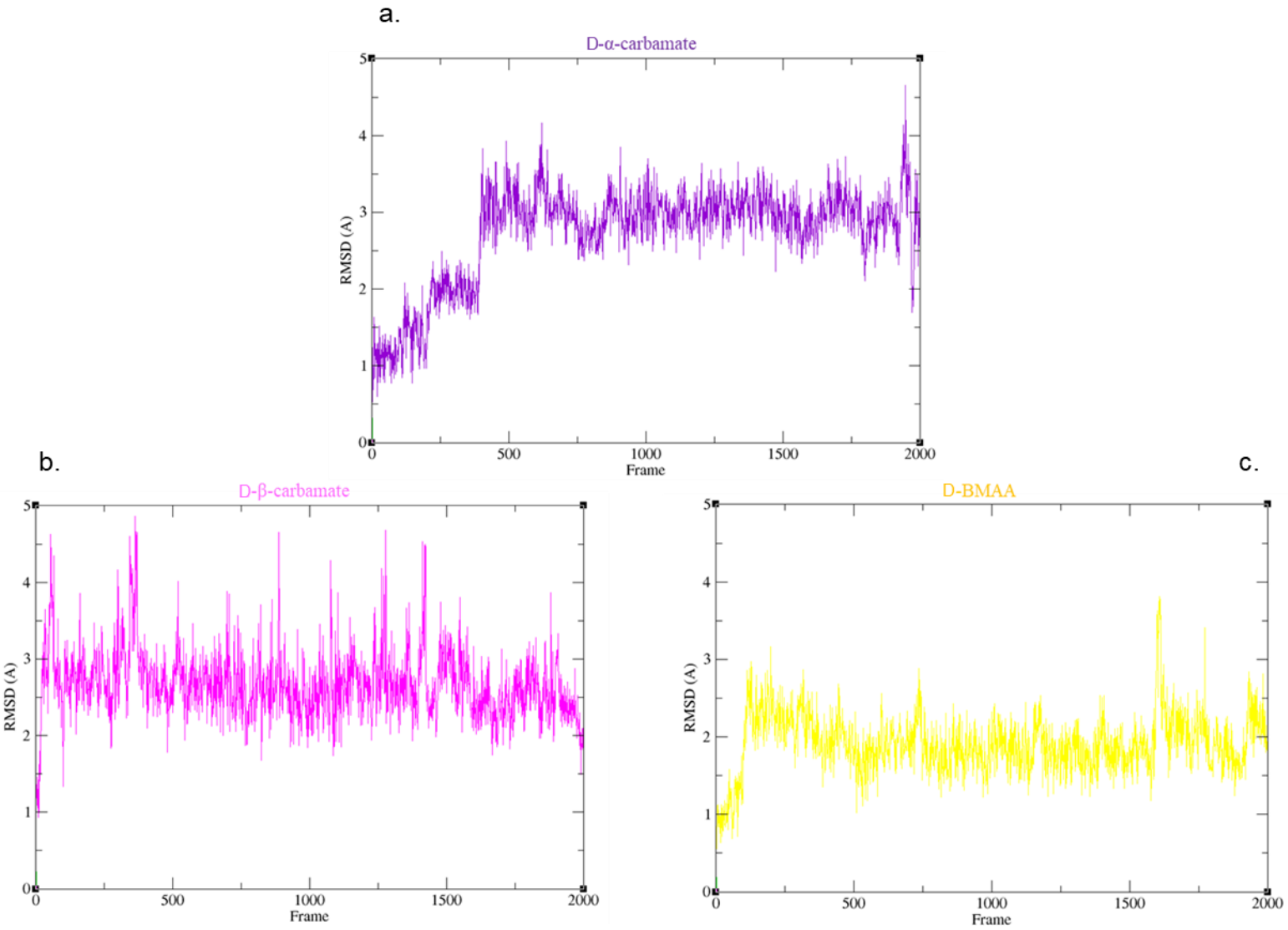

Figure 90. RMSD of D-α-carbamate in the binding pocket of AMPA during production run. b) RMSD of D-β-carbamate in the binding pocket of AMPA during production run. c) RMSD of D-BMAA in the binding pocket of AMPA during production run.

Figures 87a, 87b and 87c show the total energies of the three systems (receptor in complex with D-α-carbamate, receptor in complex with D-β-carbamate, and receptor in complex with D-BMAA) during the production run with average -122250 ± 461 kcal/mol for the structure with D-α-carbamate, -122115 ± 463 kcal/mol for the structure with D-β-carbamate, and -122417 ± 249 kcal/mol for the structure with D-BMAA. The averages prove that all systems are at equilibration.

Figures 88a, 88b and 88c show the change of the temperature during the production run. The average temperature is 309.2 ± 1.4 K for all systems, which reveal that the thermostat worked successfully.

Figures 89a, 89b, and 89c show the change of the pressure during the production run. The average pressure is 1.0 ± 0.7 atm for the system with D-α-carbamate as a ligand, 1.1 ± 0.5 atm for the system with D-β-carbamate as a ligand, and 1.1 ± 0.6 atm for the system with D-BMAA as a ligand. As a result, the barostat also worked successfully.

Figures 90a, 90b and 90c illustrate the RMSD time series of D-α-carbamate, D-β-carbamate and D-BMAA respectively in the binding pocket of the receptor during the production run. RMSD average for D-α-carbamate is 2.7 ± 0.6 Å (Figure 90a), RMSD average for L-β-carbamate is 2.6 ± 0.5 Å (Figure 90b), RMSD average for D-BMAA is 1.9 ± 0.4 Å (Figure 90c), which indicates that the all three ligands are located in the binding pocket of the receptor during the production run (Figure 91).

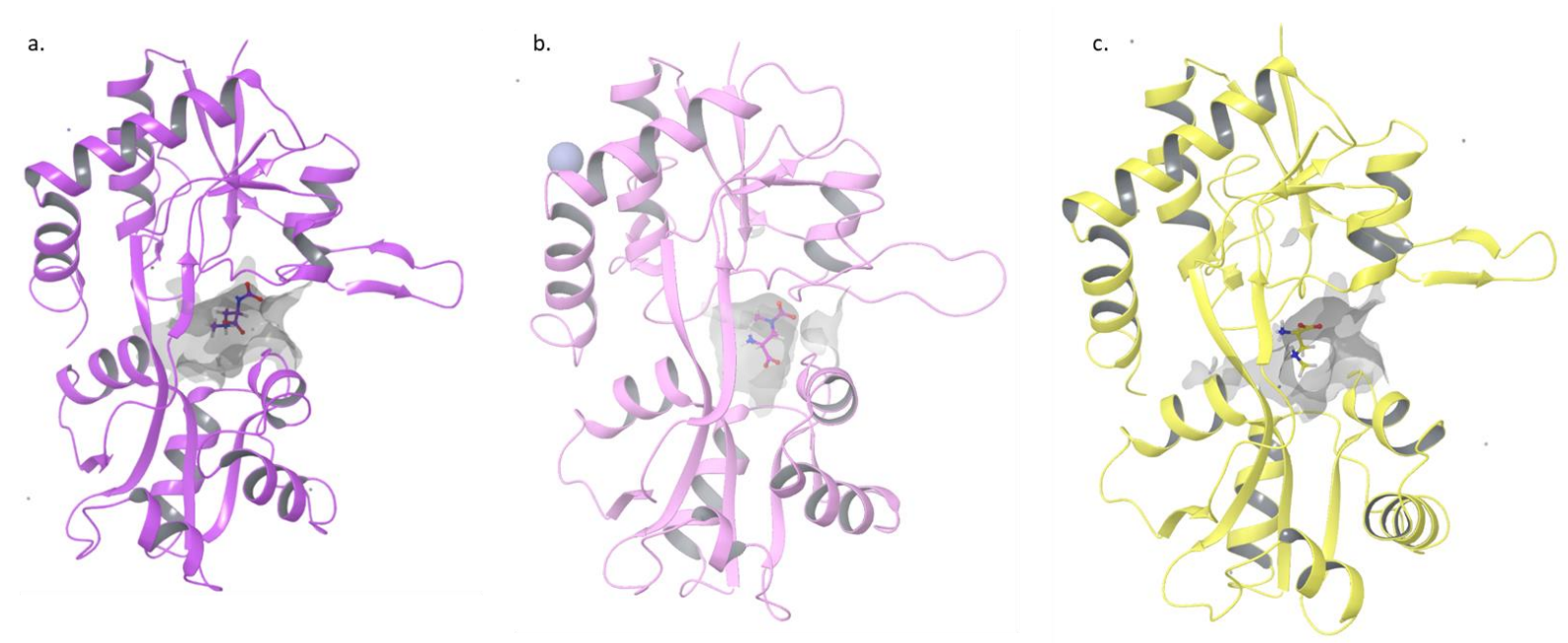

Figure 91. a) Last frame of the production run of D-α-carbamate in AMPA receptor. b) Last frame of the production run of D-β-carbamate in AMPA receptor. c) Last frame of the production run of D-BMAA in AMPA receptor. In all cases, ligands are located in the binding pocket of the receptor.

Using docking as method for the preparation of the systems, all molecules were able to find some interactions and stay located in the binding pocket of AMPA receptor for 100 ns. However, with the superimposition/minimization method D-β-carbamate wasn't able to find appropriate interactions before or during the simulation possibly due to the

fact that it adopted different conformation (rotation of 180º of dihedral N-C-C-C). This is also confirmed from the illustration of the interactions in 2D right after protein preparation with docking (Figure 92 a) and with superimposition/minimization (Figure 92 b). D-β-carbamate has three more electrostatic interactions when system was prepared with docking.

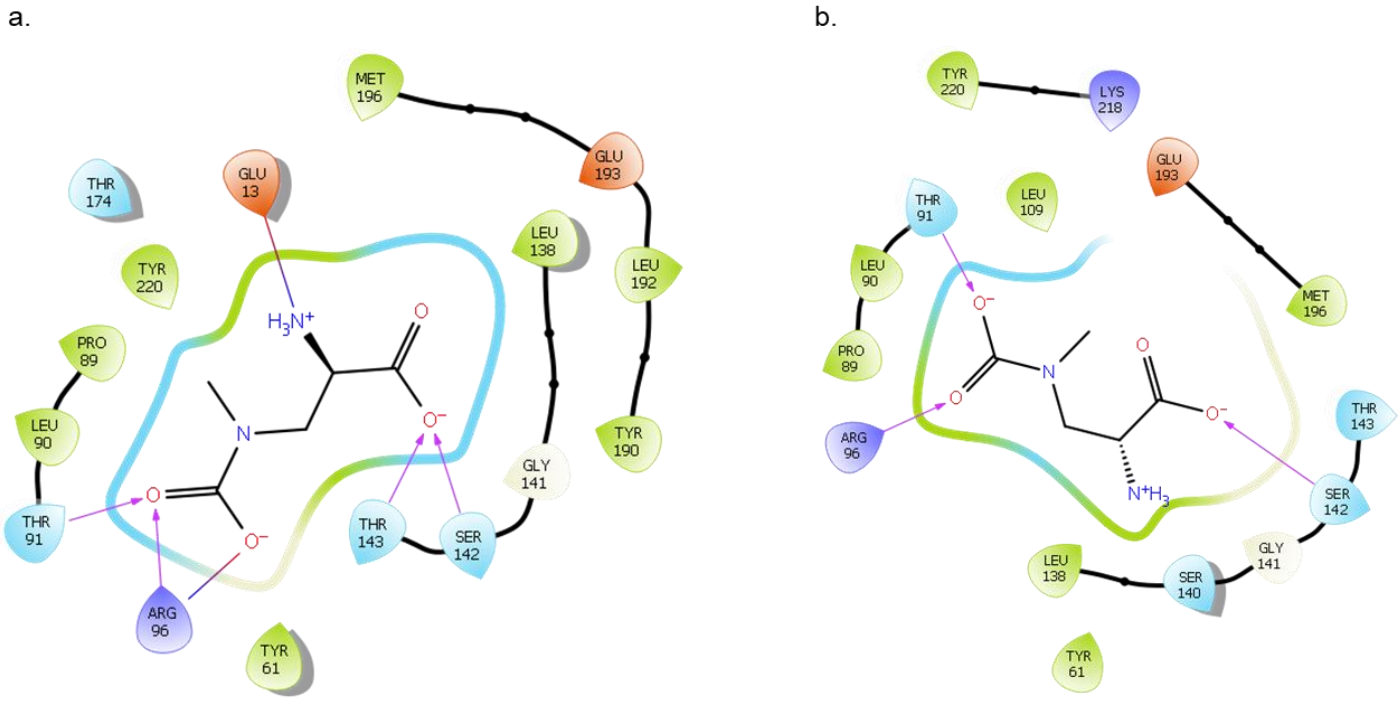

Figure 92. Interactions in 2D of receptor with D-β-carbamate after docking and b) interactions in 2D between the receptor and D-β-carbamate after superimposition and minimization.

3.5 ΔΔG calculation of glutamate and BMAA adducts complexed with the AMPA receptor using FEP/MD simulations

In this section, the difference of the free energy of binding between glutamate and BMAA adducts on the AMPA receptor was studied using FEP/MD simulations. The results concerning the study of the stability of L-glutamate, L-BMAA, L-α-carbamate, and L-β-carbamate, in the binding site of AMPA glutamate receptor using MD simulations, revealed that L-β-carbamate was the most stable molecule in the binding pocket of the receptor (RMSD = 1.2 ± 0.2 Å). The rest of the molecules presented higher mobility (RMSD = 2.1 ± 0.6 Å for L-glutamate, RMSD = 2.4 ± 0.4 Å for L-α-carbamate, and RMSD = 3.0 ± 0.8 Å for L-BMAA). Regarding the free energy perturbation method, it

was decided to study the molecule, which indicated the greater stability in the binding pocket of the AMPA receptor, L-β-carbamate, and investigate the difference in the free energy of binding with respect to glutamate.

3.5.1 Introduction of intermediate molecules and cycle closure between L-glutamate, L-β-carbamate and an intermediate molecule

For the calculation of the difference of the binding free energy between glutamate and β-carbamate of BMAA, an intermediate molecule was used, to which we will refer as intermediate-1. In this way, a cycle closure was designed between glutamate, intermediate-1 and β-carbamate (Figure 93), which will validate the calculations, since the total energy should be equal to zero. To achieve this an intermediate molecule was constructed that will achieve a smooth change between the two states improving the accuracy of the calculation. Three perturbations were completed both in solvent and in complex, which are: glutamate to intermediate-1, intermediate-1 to β-carbamate and glutamate to β-carbamate.

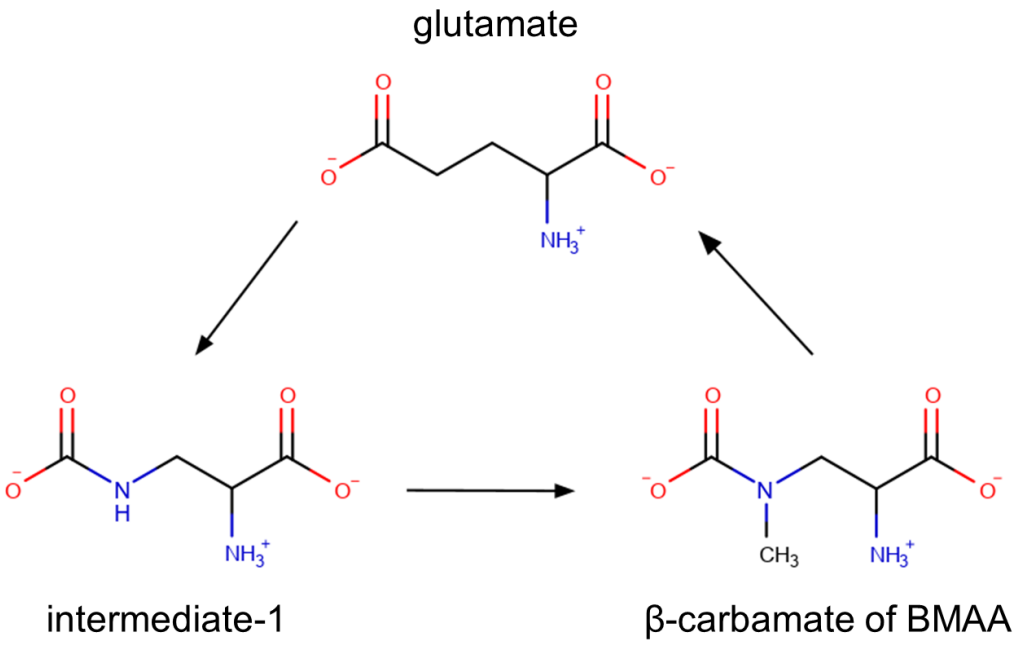

Figure 93. The cycle of glutamate, intermediate-1 and β-carbamate of BMAA.

3.5.2 Preparation of the systems using FEPrepare platform

L-glutamate from the crystal structure with PDB ID:1FTJ and L-β-carbamate of BMAA were loaded in Maestro, the intermediate-1 was created, using L-β-carbamate as reference molecule, and all three were superimposed with respect to glutamate. The molecules were exported as .pdb files, they were uploaded on the LigParGen [67] platform and the .pdb, .prm and .rtf files were created, which contain bonded information and parameters for each molecule. Subsequently, the chosen AMPA receptor with PDB ID:1FTJ was loaded in UCSF Chimera [68], where the hydrogens were excluded and the receptor with no hydrogens or ligand was exported as .pdb file. In order to prepare the systems both in solvent and in complex (Figure 94), the FEPrepare platform [69] was used, where the .pdb, .rtf, .prm files for each molecule and the receptor with no hydrogens were uploaded and the hybrid molecule was created for each system as well as all the appropriate files. The three solvent systems were consisting of ca. 4,000 atoms each, including ca. 1,300 water molecules, and the three complex systems were consisting of ca. 50,000 atoms each, including ca. 15,300 water molecules. The force field used was OPLS-AA [70].

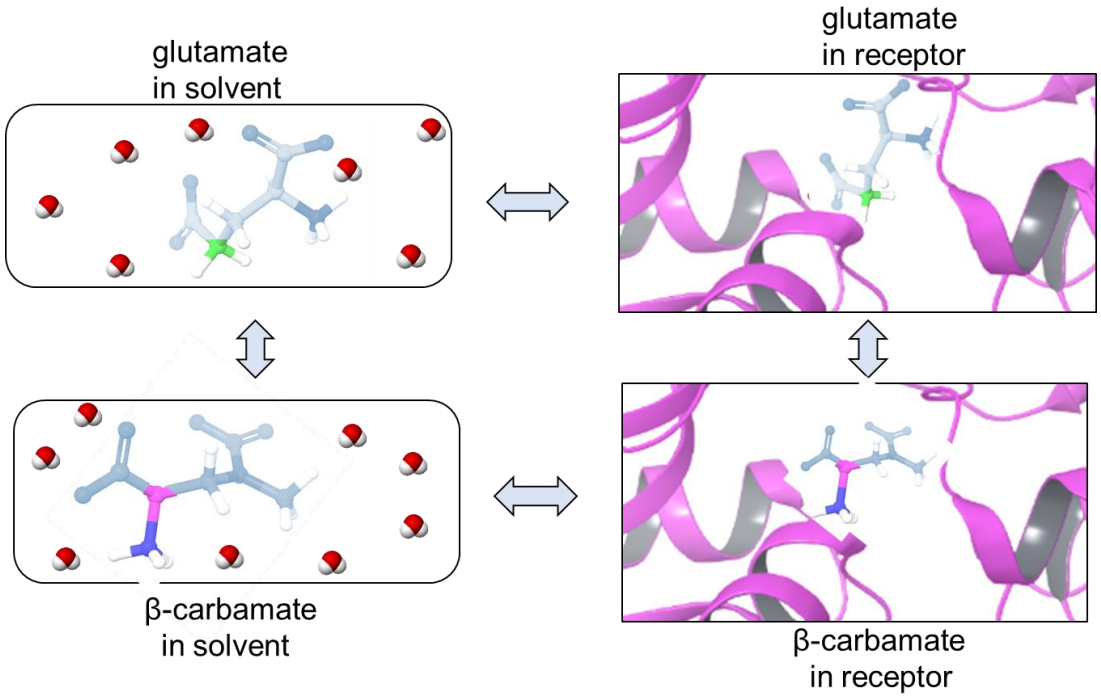

Figure 94. a) Thermodynamic cycle for the perturbation of glutamate to β-carbamate.

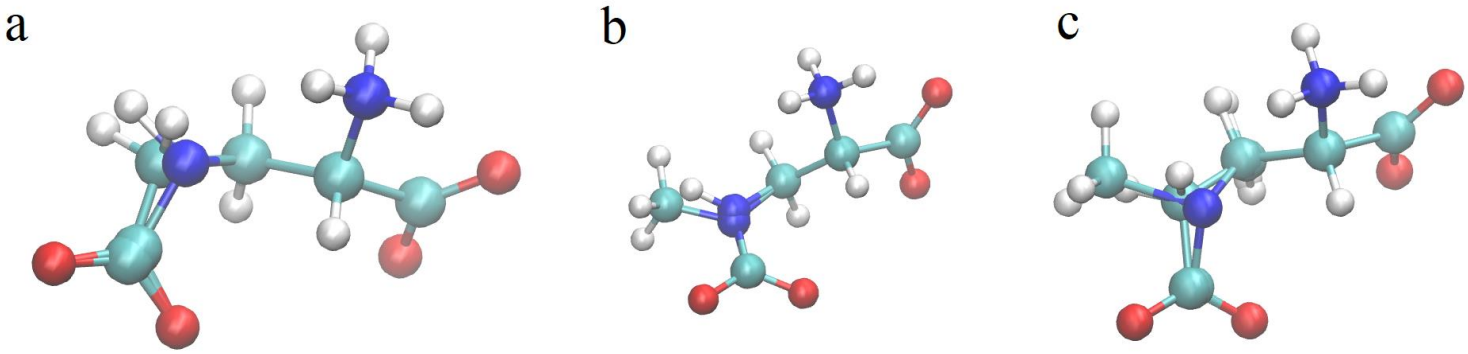

Figure 95. a) Hybrid molecule for the perturbation of glutamate to intermediate-β-carbamate, b) hybrid molecule for the perturbation of intermediate-β-carbamate to β-carbamate and c) hybrid molecule for the perturbation of glutamate to β-carbamate.

3.5.3 Equilibration of the systems

Each system was equilibrated in the NVT ensemble, where the temperature was set at 310 K, which was controlled with a langevin thermostat, and the langevinDamping option, which is about damping coefficient, was set at 1 (damping coefficient of 1/ps). The simulation ran for 500,000 steps with a time step of 2 fs (1 ns). Then, the systems were equilibrated in the NPT ensemble, were the temperature was set at 310 K and was controlled with a langevin thermostat, while a Nosé-Hoover barostat was used to control pressure. Option “useGroupPressure” was activated, which is needed for rigid bonds, as well as langevinPiston was. LangevinPistonTarget was used to specify target pressure for Langevin piston method at 1 atm (1atm = 1.01325 bar), langevinPistonPeriod specified barostat oscillation time scale for the Langevin piston method at 100.0 fs, and langevinPistonDecay was used, in order to specify barostat damping time scale for the method at 50.0 fs.

3.5.4 FEP/MD Simulations for perturbations glutamate to intermediate-1, intermediate-1 to β-carbamate and β-carbamate to glutamate

After the equilibration of the systems in the NPT ensemble, the simulations for the calculation of the binding energies were started. The λ

windows that were used were 16. In particular, λ windows are: λ=0.0000 to λ=0,06250, λ=0,6250 to λ=0.12500, λ=0.12500 to λ=0.18750, λ=0.18750 to λ=0.25000, λ=0.25000 to λ=0.31250, λ=0.31250 to λ=0.37500, λ=0.37500 to λ=0.43750, λ=0.43750 to λ=0.50000, λ=0.50000 to λ=0.56250, λ=0.56250 to λ=0.62500, λ=0.62500 to λ=0.68750, λ=0.68750 to λ=0.75000, λ=0.75000 to λ=81250, λ=0.81250 to λ=0.84500, λ=0.84500 to λ=0.93750, λ=0.93750 to λ=1.00000. Each simulation ran for 5,000,000 steps with a time step of 2 fs (total 10 ns). The energy intervals were 5,000 steps. After the simulations were completed, the .fepout files for the forward process and the backward process both for solvent-system and complex-system and for all three perturbations were obtained.

3.5.5 ΔΔG calculation

For the analysis of the calculations, the “ParseFEP” tool in VMD was used. The temperature was set at 310 K and the Gram-Charlier order was set at 0. Then, the .fepout file for the forward process of the solvent-system of perturbation glutamate → intermediate-1 was uploaded as well as the .fepout file for the backward process. Subsequently, the BAR-estimator option was marked, and the plugin was run. The same procedure was followed for the complex-system and for the rest two perturbations, intermediate-1 → β-carbamate and β-carbamate → glutamate.

- Perturbation glutamate → intermediate-1

Table 3. Binding free energy of perturbation glutamate to intermediate-1

System	*Solvent*	*Complex*
Energy (kcal/mol)	-115.83 ± 0.02	-114.83 ± 0.16
Binding Free Energy = complex – solvent = **1.00 ± 0.16** kcal/mol		

The degree of overlap between the reference (glutamate) and the target (intermediate-1) states of the transformation can be assessed from the diagrams being presented bellow (Figure 96 for the solvent system and Figure 97 for the complex system). Figures 96a and 97a show the time-evolution of

the free energy between two λ windows, figures 96b and 97b are histograms of the probability distributions between two λ windows and figures 96c and 97c illustrate the overall free energy change for the forward (black solid line) and backward (red solid line) transformations. Both solvent and complex systems indicate good overlap, however, the overlap in the solvent-system is better than that in the complex-system (Table 5), as the complex-system is consisted of ca. 46,000 more atoms and has much more degrees of freedom.

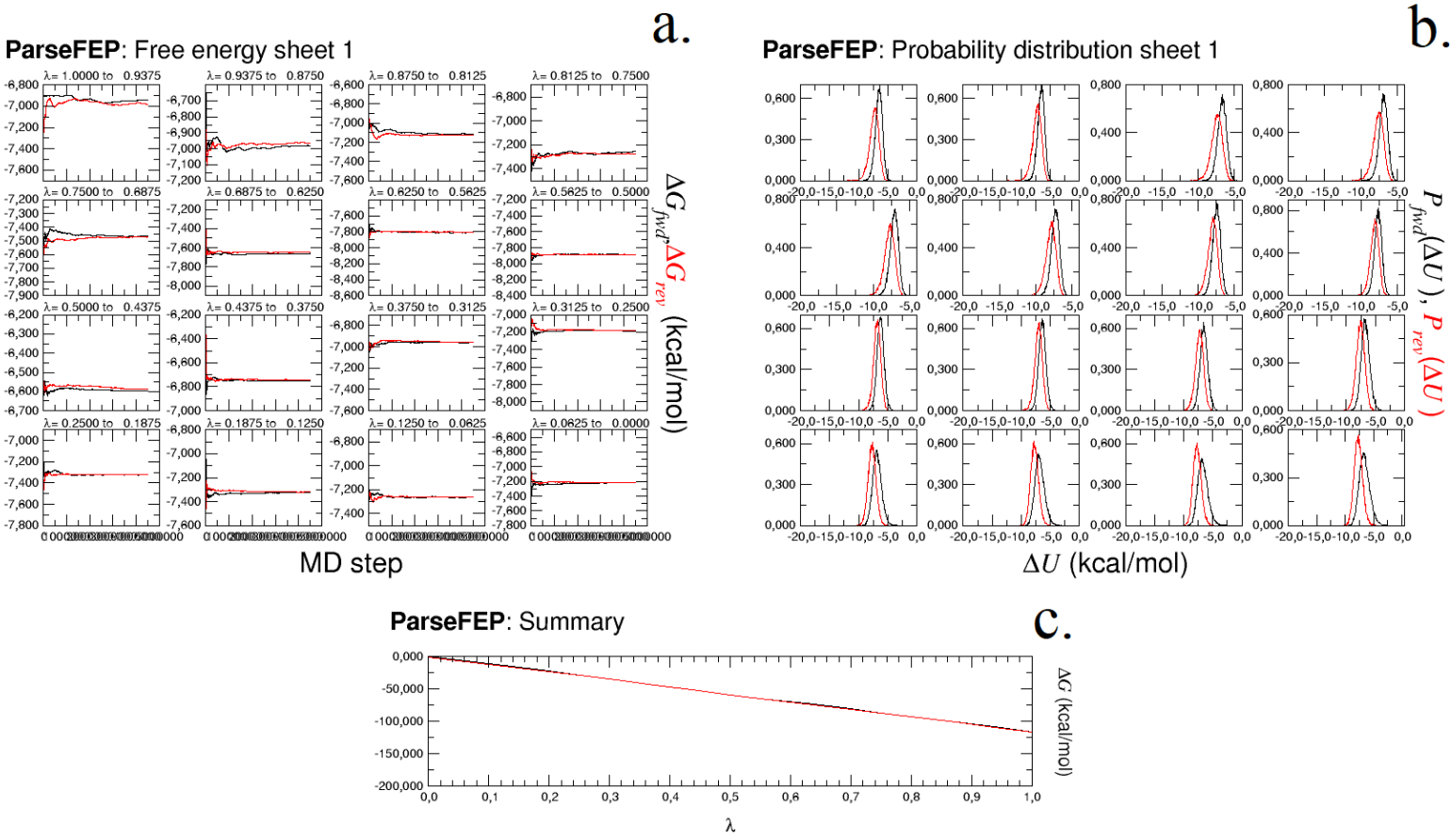

Figure 96. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the solvent-system of perturbation glutamate to intermediate-1.

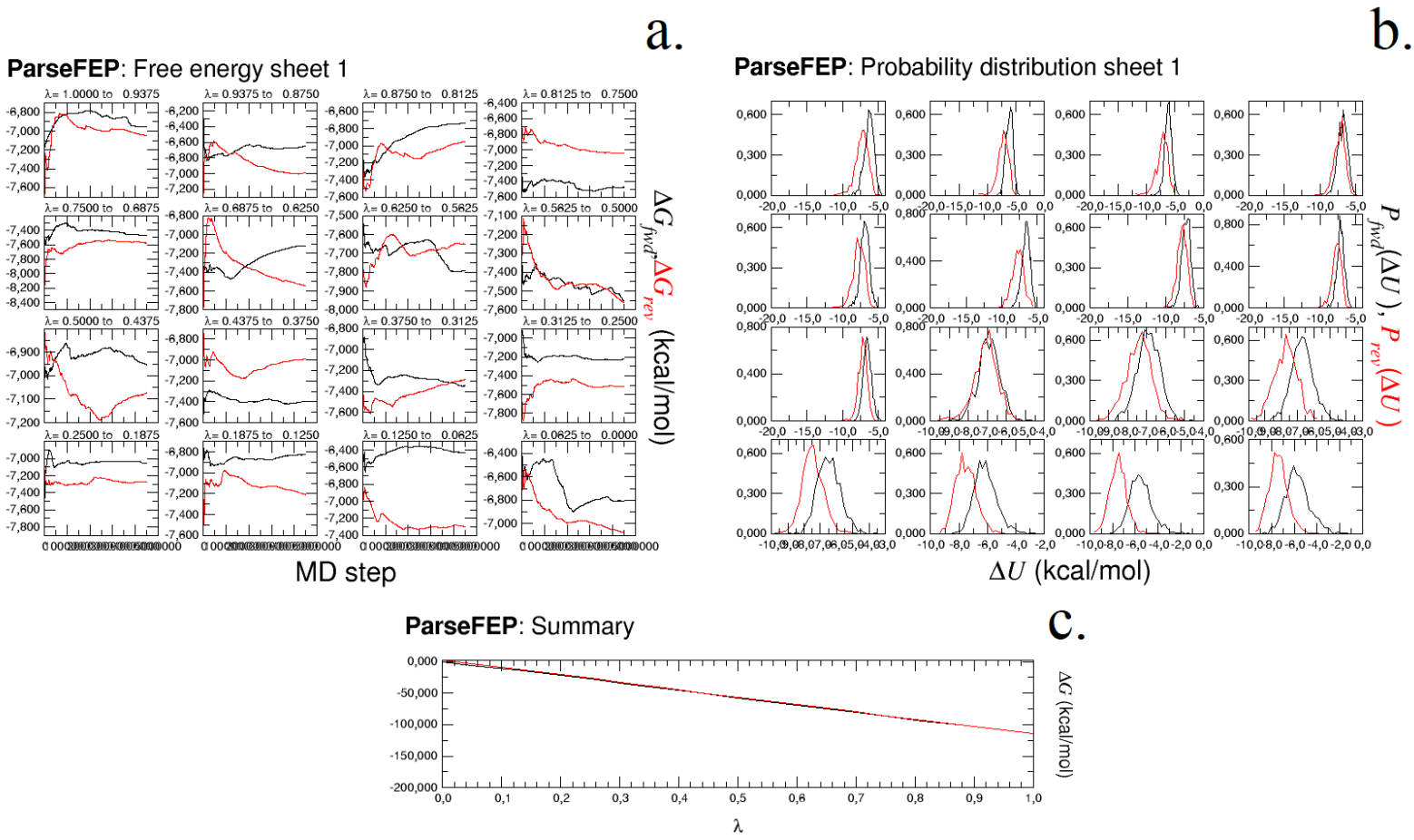

Figure 97. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the complex-system of perturbation glutamate to intermediate-1.

Table 4. Overlap between λ windows of perturbation glutamate to intermediate-1.

		Solvent	complex
λ_K	$\lambda_{K+\delta i}$	P0_overlap_P1 (%)	P0_overlap_P1 (%)
0.00000	0.06250	58.49	45.17
0.06250	0.12500	58.49	45.17
0.12500	0.18750	54.72	46.17
0.18750	0.25000	55.94	77.25
0.25000	0.31250	58.16	50.06
0.31250	0.37500	60.49	36.51
0.37500	0.43750	63.60	66.26
0.43750	0.50000	66.93	57.82
0.50000	0.56250	51.72	62.71

0.56250	0.62500	59.16	87.24
0.62500	0.68750	57.05	66.15
0.68750	0.75000	57.94	45.73
0.75000	0.81250	54.61	47.72
0.81250	0.87500	55.72	36.74
0.87500	0.93750	53.61	23.64
0.93750	1.00000	50.61	34.96

- Perturbation intermediate-1 → β-carbamate of BMAA

Table 5. Binding free energy of perturbation intermediate-1 to β-carbamate of BMAA

System	*Solvent*	*Complex*
Energy (kcal/mol)	67.42 ± 0.08	65.88 ± 0.11
Binding Free Energy = complex – solvent = ***-1.54 ± 0.13*** *kcal/mol*		

The degree of overlap between the reference (intermediate-1) and the target (β-carbamate) states of the transformation can be assessed from the diagrams being presented bellow (Figure 98 for the solvent system and Figure 99 for the complex system). Figures 98a and 99a show the time-evolution of the free energy, figures 98b and 99b are histograms of the probability distributions and figures 98c and 99c illustrate the overall free energy change for the forward (black solid line) and backward (red solid line) transformations. Both solvent and complex systems indicate good overlap, however, the overlap in the solvent-system is better than that in the complex-system (Table 7), as the complex-system is consisted of ca. 46,000 more atoms and has much more degrees of freedom.

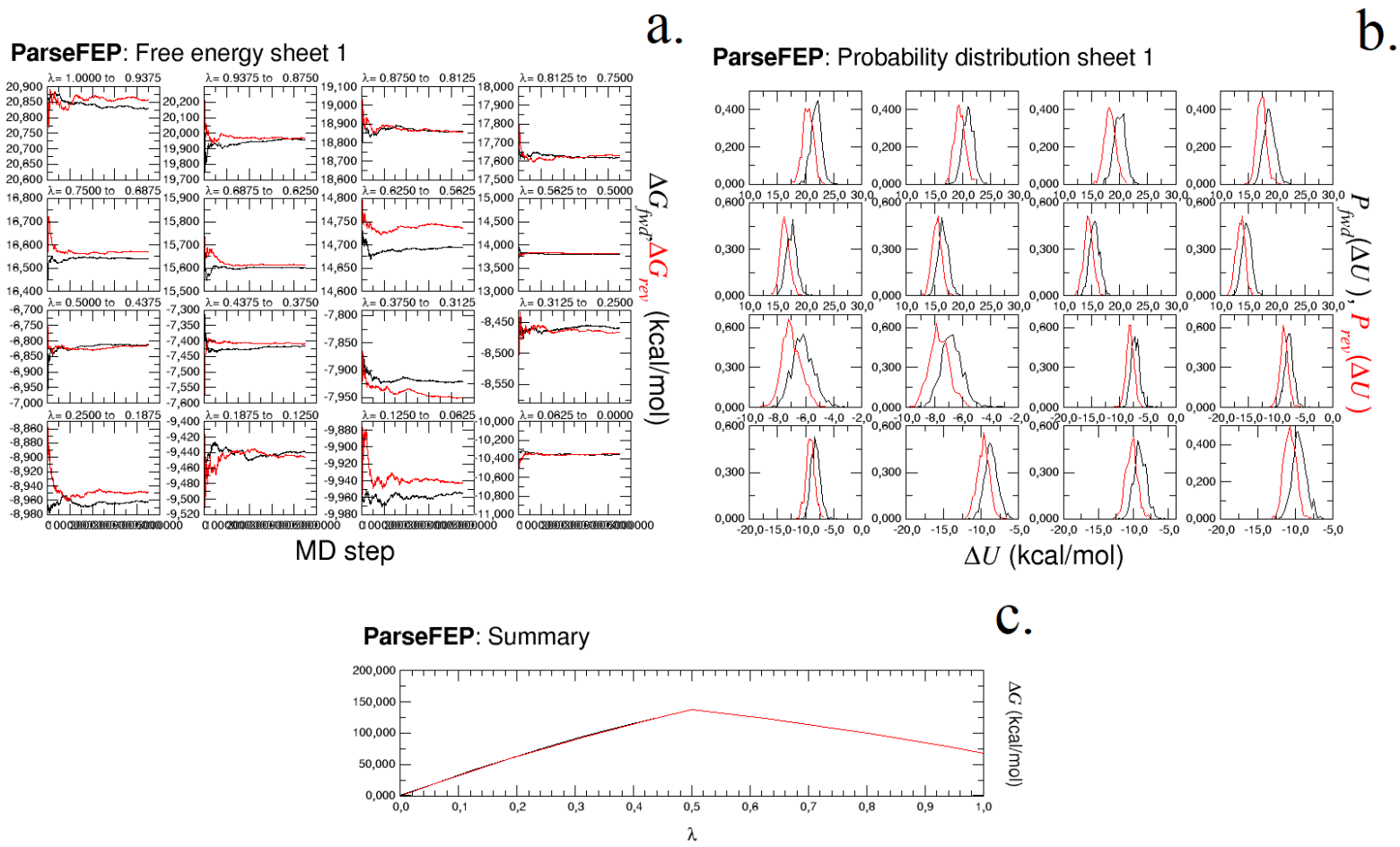

Figure 98. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the solvent-system of perturbation intermediate-1 to β-carbamate of BMAA.

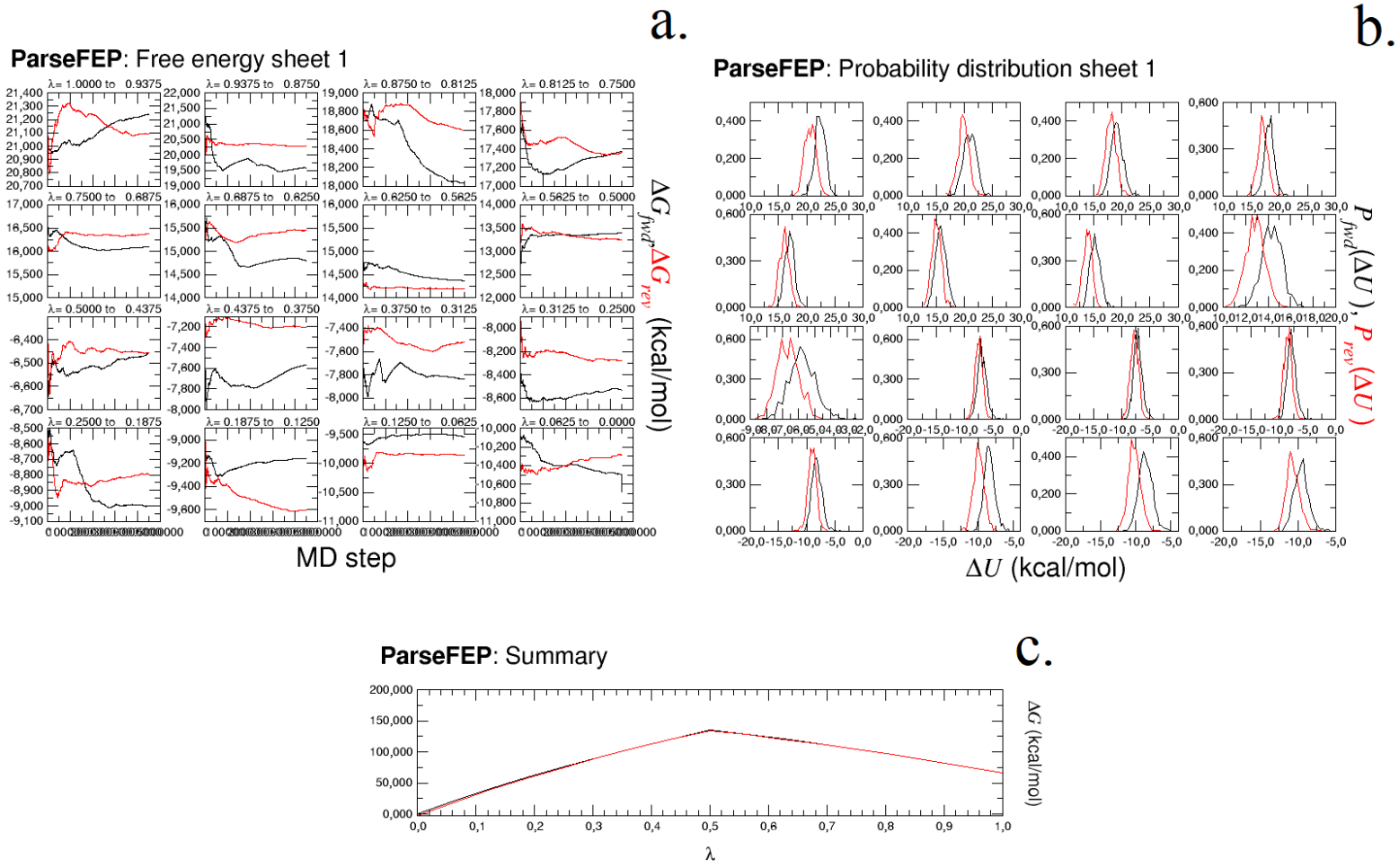

Figure 99. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the complex-system of perturbation intermediate-1 to β-carbamate of BMAA.

Table 6. Overlap between λ windows of perturbation intermediate-1 to β-carbamate of BMAA

		Solvent	complex
λ_K	$\lambda_{K+\delta i}$	P0_overlap_P1 (%)	P0_overlap_P1 (%)
0.00000	0.06250	48.06	58.05
0.06250	0.12500	48.06	58.05
0.12500	0.18750	40.40	59.71
0.18750	0.25000	42.73	47.39
0.25000	0.31250	50.61	56.16
0.31250	0.37500	46.61	66.93
0.37500	0.43750	51.61	44.51
0.43750	0.50000	51.50	39.84

0.50000	0.56250	31.52	28.86
0.56250	0.62500	57.05	73.47
0.62500	0.68750	55.27	69.92
0.68750	0.75000	53.83	67.70
0.75000	0.81250	56.05	62.71
0.81250	0.87500	52.50	31.74
0.87500	0.93750	47.06	40.40
0.93750	1.00000	46.50	53.05

- Perturbation β-carbamate of BMAA → glutamate

Table 7. Binding free energy of perturbation β-carbamate of BMAA to glutamate

System	*Solvent*	*Complex*
Energy (kcal/mol)	55.00 ± 0.06	52.78 ± 0.16
Binding Free Energy = complex – solvent = ***-2.22 ± 0.17*** *kcal/mol*		

The degree of overlap between the reference (glutamate) and the target (β-carbamate) states of the transformation can be assessed from the diagrams being presented bellow (Figure 100 for the solvent system and Figure 101 for the complex system). Figures 100a and 101a show the time-evolution of the free energy, figures 100b and 101b are histograms of the probability distributions and figures 100c and 101c illustrate the overall free energy change for the forward (black solid line) and backward (red solid line) transformations. Both solvent and complex systems indicate good overlap, however, the overlap in the solvent-system is better than that in the complex-system (Table 9), as the complex-system is consisted of ca. 46,000 more atoms and has much more degrees of freedom.

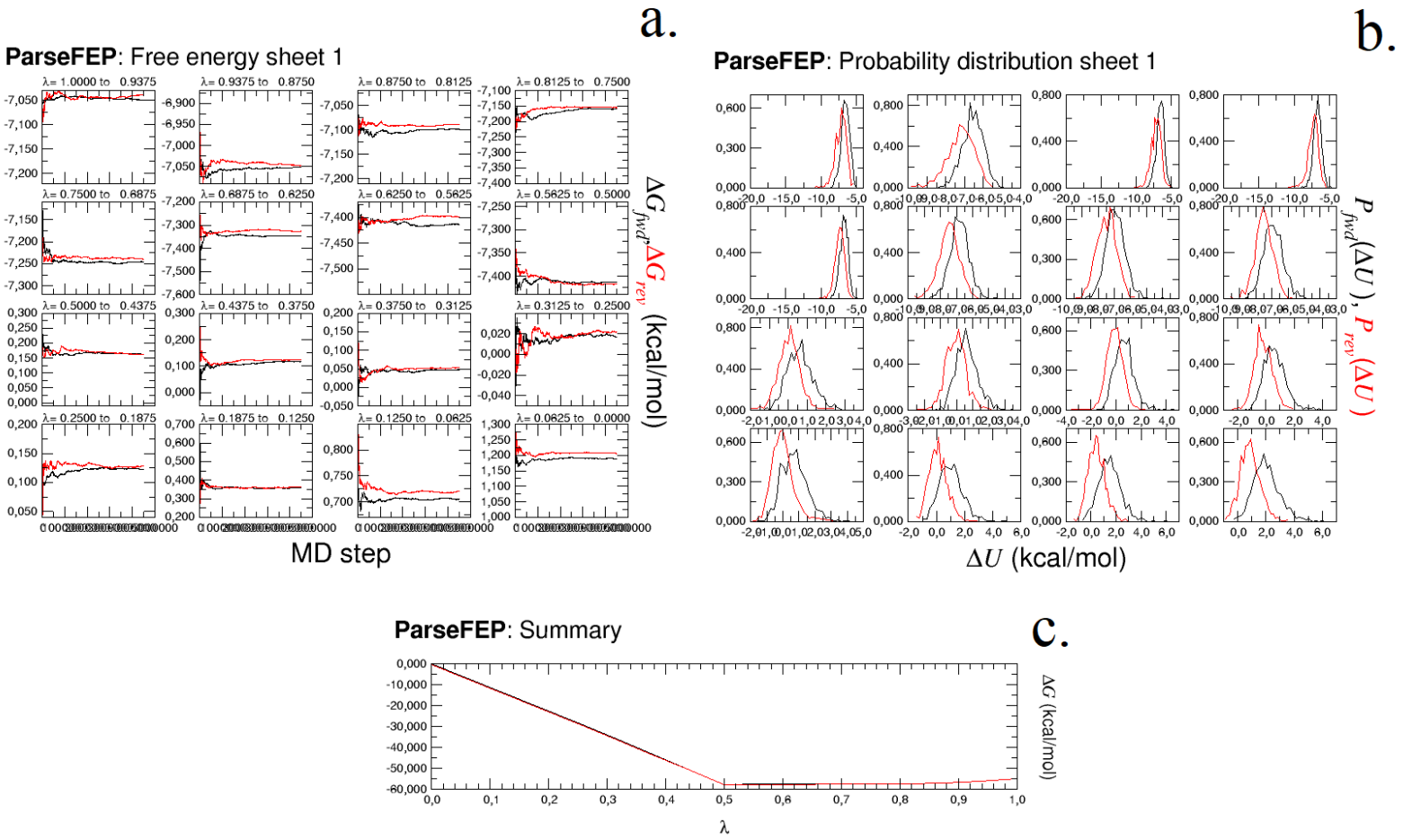

Figure 100. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the solvent-system of perturbation β-carbamate of BMAA to glutamate.

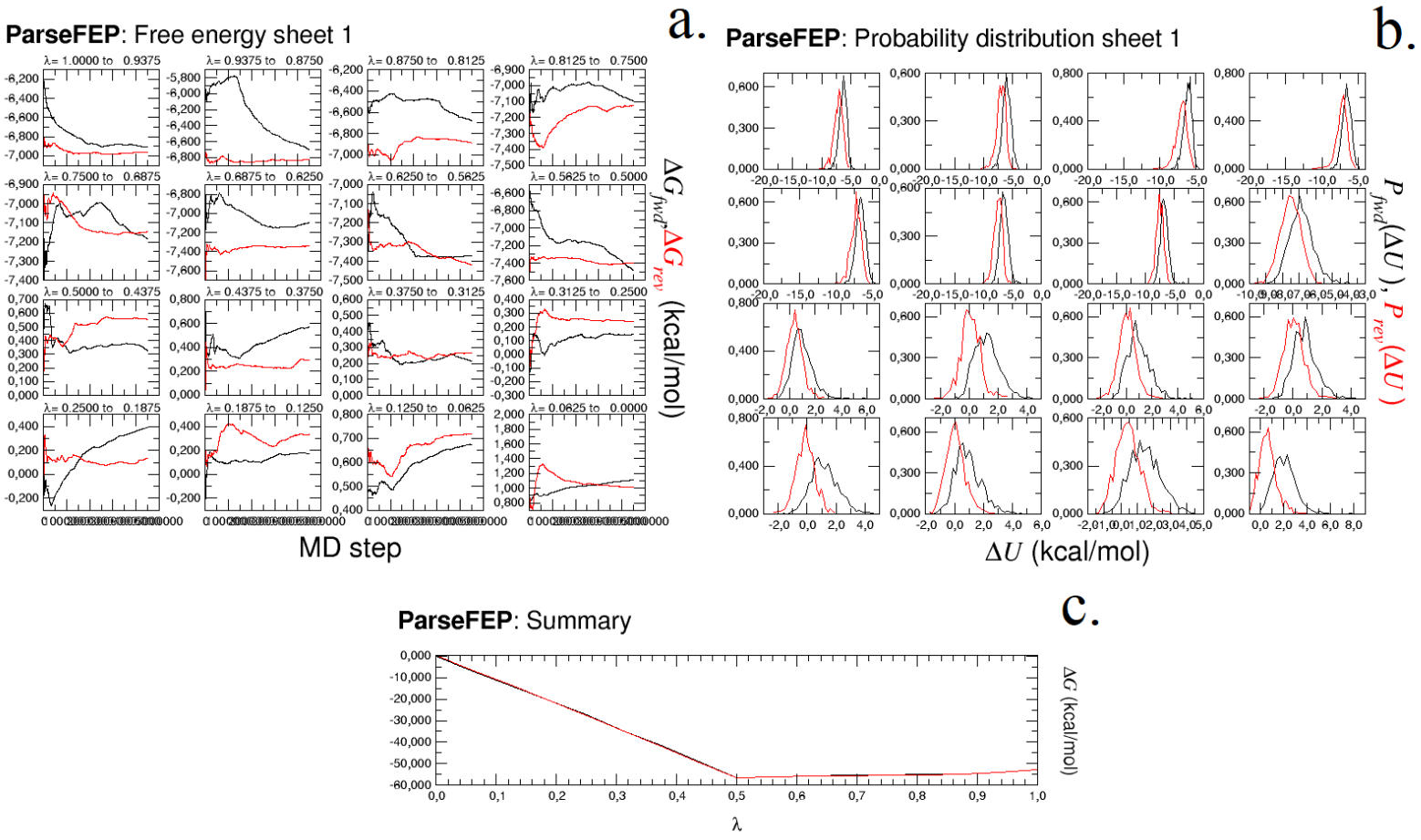

Figure 101. a) Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the complex-system of perturbation β-carbamate of BMAA to glutamate.

Table 8. Overlap between λ windows of perturbation β-carbamate of BMAA to glutamate.

		Solvent	complex
λ_K	$\lambda_{K+\delta i}$	P0_overlap_P1 (%)	P0_overlap_P1 (%)
0.00000	0.06250	58.27	48.39
0.06250	0.12500	58.28	48.39
0.12500	0.18750	58.38	48.61
0.18750	0.25000	63.60	58.16
0.25000	0.31250	58.38	61.60
0.31250	0.37500	25.60	49.83
0.37500	0.43750	67.48	58.49
0.43750	0.50000	63.60	63.37

0.50000	0.56250	60.49	56.16
0.56250	0.62500	60.93	46.28
0.62500	0.68750	58.93	58.05
0.68750	0.75000	59.49	61.04
0.75000	0.81250	58.05	41.51
0.81250	0.87500	56.83	61.04
0.87500	0.93750	52.28	53.61
0.93750	1.00000	49.83	41.62

3.5.6 Error Calculation

Cycle closure error

The total energy of the cycle was calculated by adding all the differences of the binding free energies of the three systems (Figure 102).

$\Delta G = (1.04 \pm 0.17) + (-1.54 \pm 0.13) + (-2.22 \pm 0.17) = -2.72 \pm 0.27$ kcal/mol

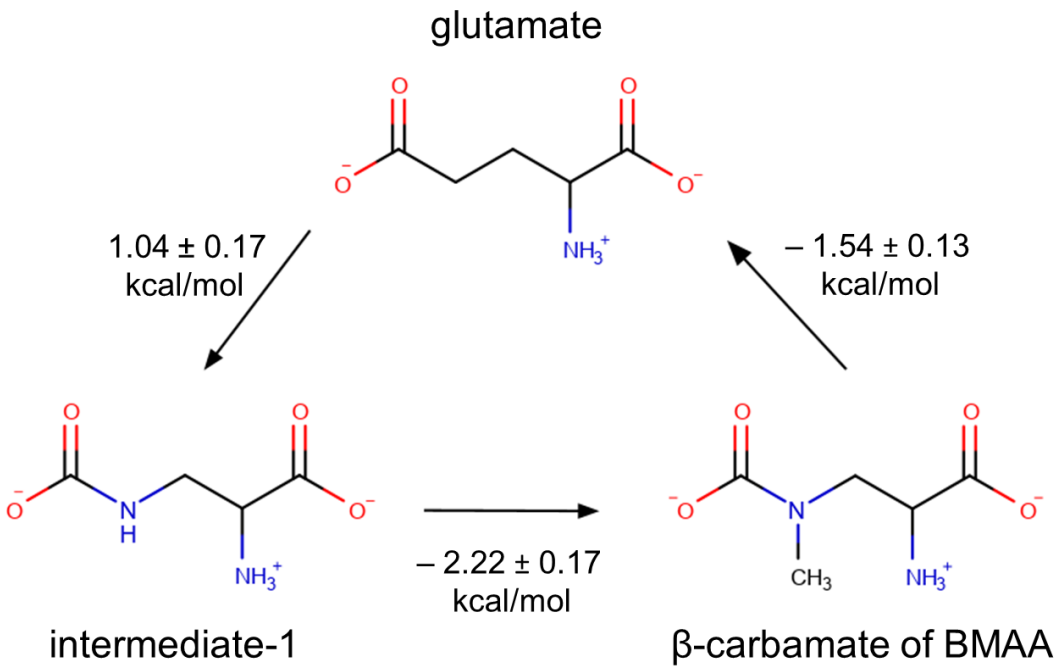

Figure 102. Closed cycle of glutamate, intermediate-1 and β-carbamate with the difference of the binding free energy of each perturbation.

The difference in the binding free energy is negative, indicating that the target molecule, glutamate, has greater affinity than the reference molecule, β-carbamate of BMAA.

The error is ca. -2.70 kcal/mol, while it was expected not to exceed ± 1.50 kcal/mol. There are many ways that can be applied in order to improve the error as described in section 2.3.5. First, we decided to introduce another intermediate molecule for the perturbation β-carbamate to glutamate in the closed cycle, as the convergence of the complex-system of this specific perturbation was medium (ca. 40-60%).

3.5.7 Convergence of the calculations by adding intermediate molecule

We will refer to the new intermediate molecule as intermediate-2. A new closed cycle was designed (Figure 103), where two new perturbations were completed. The first was β-carbamate of BMAA to intermediate-2 and intermediate-2 to glutamate. Intermediate-2 was designed in Maestro, and superimposed with respect to glutamate. The molecule was exported as .pdb file, it was uploaded on the LigParGen platform and the .pdb, .prm and .rtf files were created, which contain bonded information and parameters for each molecule. Subsequently, the systems were prepared both in solvent and in complex using the FEPrepare platform. The two solvent systems were consisting of ca. 4,000 atoms each, including ca. 1,300 water molecules, and the two complex systems were consisting of ca. 50,000 atoms each, including ca. 15,300 water molecules. The force field used was OPLS-AA. Each system was equilibrated and simulated as described previously using also the same λ windows.

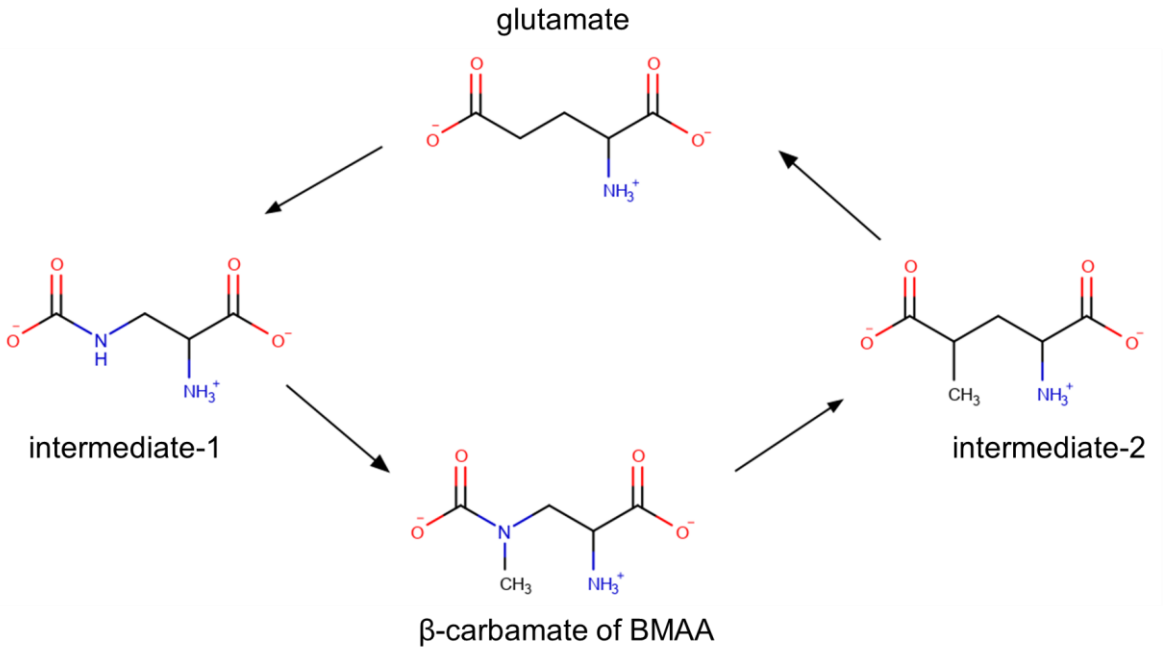

Figure 103. Closed cycle of glutamate, intermediate-1, intermediate-2 and β-carbamate of BMAA.

- Perturbation β-carbamate of BMAA → intermediate-2

Table 9. Binding free energy of perturbation β-carbamate of BMAA to intermediate-2

System	*Solvent*	*Complex*
Energy (kcal/mol)	71.68 ± 0.07	74.04 ± 0.28
Binding Free Energy = complex – solvent = ***2.36 ± 0.29*** *kcal/mol*		

The degree of overlap between the reference (β-carbamate) and the target (intermediate-2) states of the transformation can be assessed from the diagrams being presented bellow (Figure 104 for the solvent system and Figure 105 for the complex system). Figures 104a and 105a show the time-evolution of the free energy, figures 104b and 105b are histograms of the probability distributions and figures 104c and 105c illustrate the overall free energy change for the forward (black solid line) and backward (red solid line) transformations. Both solvent and complex systems indicate good overlap, with solvent system having better overlap than the complex system. The overlap in the complex-system compared to the one of the perturbation β-carbamate to glutamate, generally is improved as shown in Table 11.

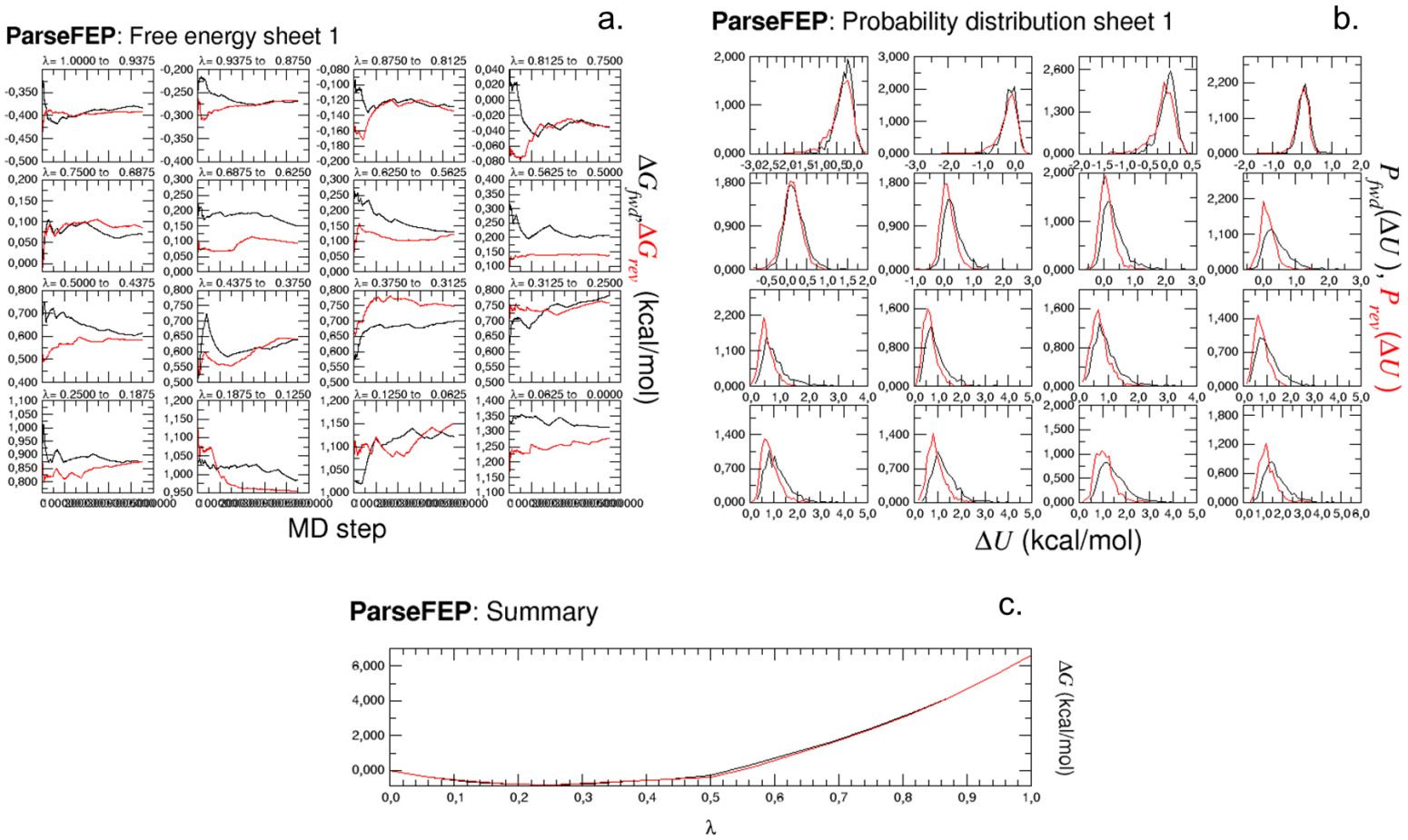

Figure 104. Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the solvent-system of perturbation 6-carbamate to intermediate-2.

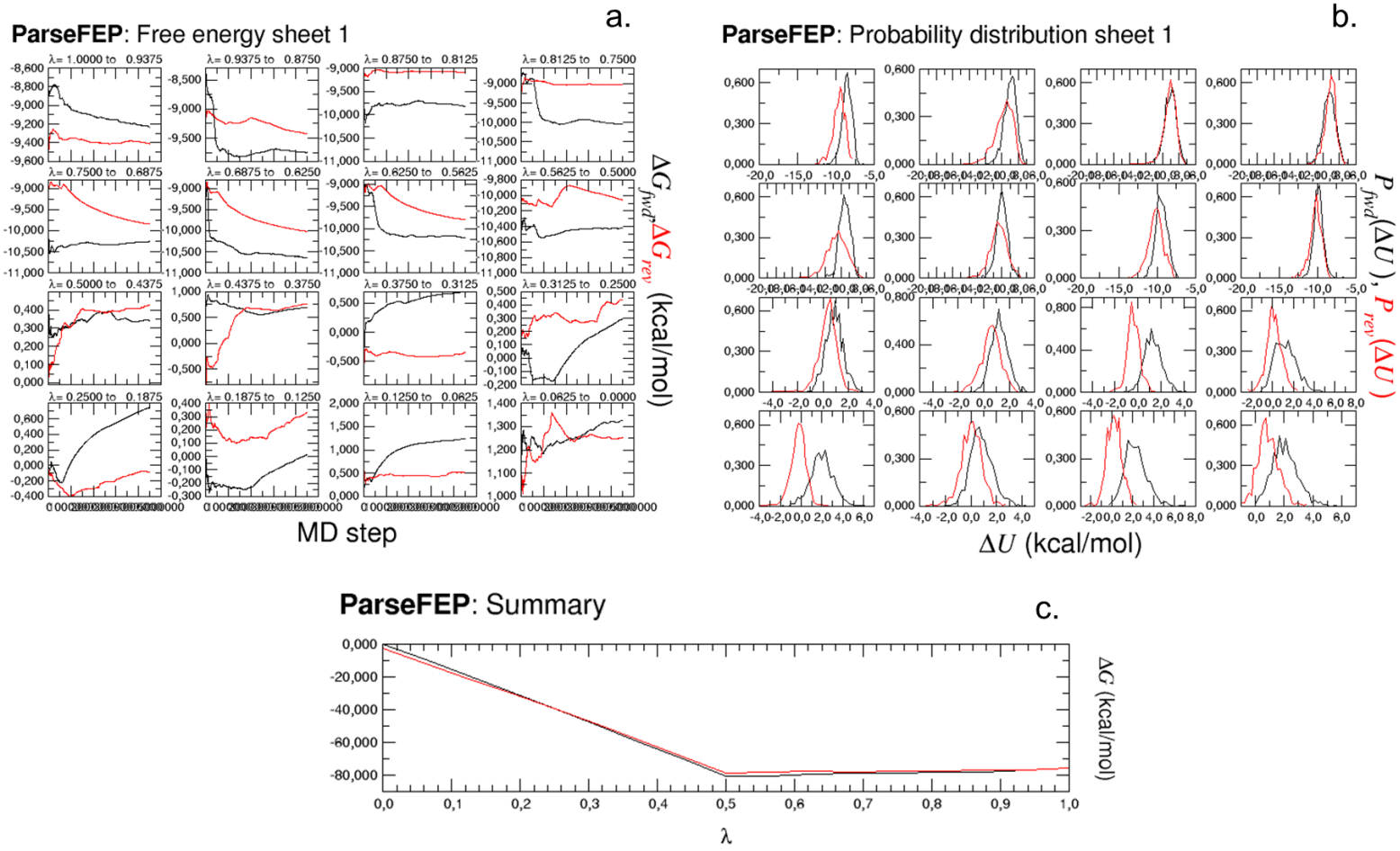

Figure 105. Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the complex-system. of perturbation ß-carbamate to intermediate-2.

Table 10. Overlap between λ windows of perturbations ß-carbamate to glutamate and ß-carbamate to intermediate-2.

		β-carbamate → glutamate		**β-carbamate → intermediate-2**	
		Solvent	Complex	Solvent	complex
λ_K	**$\lambda_{K+\delta i}$**	**P0_overlap_P1 (%)**	**P0_overlap_P1 (%)**	**P0_overlap_P1 (%)**	**P0_overlap_P1 (%)**
0.00000	0.06250	58.27	48.39	63.26	45.06
0.06250	0.12500	58.28	48.39	56.71	66.15
0.12500	0.18750	58.38	48.61	58.60	90.90
0.18750	0.25000	63.60	58.16	57.82	86.57
0.25000	0.31250	58.38	61.60	55.38	51.83
0.31250	0.37500	25.60	49.83	50.39	69.70

0.37500	0.43750	67.48	58.49	57.38	65.04
0.43750	0.50000	63.60	63.37	59.71	77.36
0.50000	0.56250	60.49	56.16	71.59	78.14
0.56250	0.62500	60.93	46.28	58.16	60.49
0.62500	0.68750	58.93	58.05	69.26	17.65
0.68750	0.75000	59.49	61.04	56.49	59.38
0.75000	0.81250	58.05	41.51	60.71	20.09
0.81250	0.87500	56.83	61.04	54.94	64.37
0.87500	0.93750	52.28	53.61	49.61	26.08
0.93750	1.00000	49.83	41.62	48.50	47.39

- Perturbation β-carbamate of intermediate-2 → glutamate

Table 11. Binding free energy of perturbation intermediate-2 to glutamate

System	*Solvent*	*Complex*
Energy (kcal/mol)	-6.72 ± 0.04	-7.47± 0.14
Binding Free Energy = complex – solvent = ***-0.75 ± 0.15*** *kcal/mol*		

The degree of overlap between the reference (intermediate-2) and the target (glutamate) states of the transformation can be assessed from the diagrams being presented bellow (Figure 106 for the solvent system and Figure 107 for the complex system). Figures 106a and 107a show the time-evolution of the free energy, figures 106b and 107b are histograms of the probability distributions and figures 106c and 107c illustrate the overall free energy change for the forward (black solid line) and backward (red solid line) transformations. Both solvent and complex systems indicate good overlap, with solvent system having better overlap than the complex system. The overlap in the complex-system compared to the one of the perturbation β-carbamate to glutamate, was improved in some cases, but deteriorated in other cases as shown in Table 13.

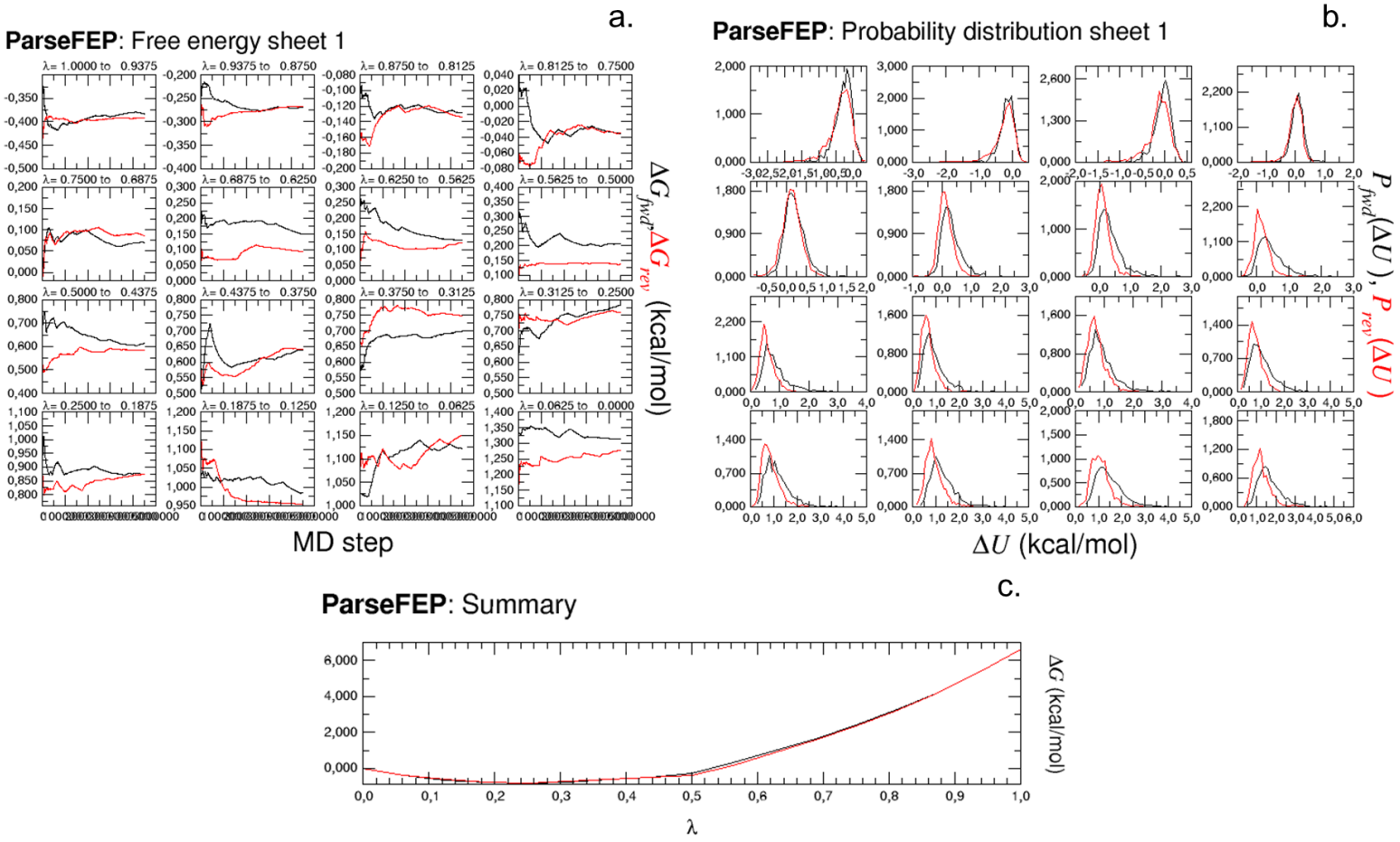

Figure 106. Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the solvent-system of perturbation intermediate-2 to glutamate.

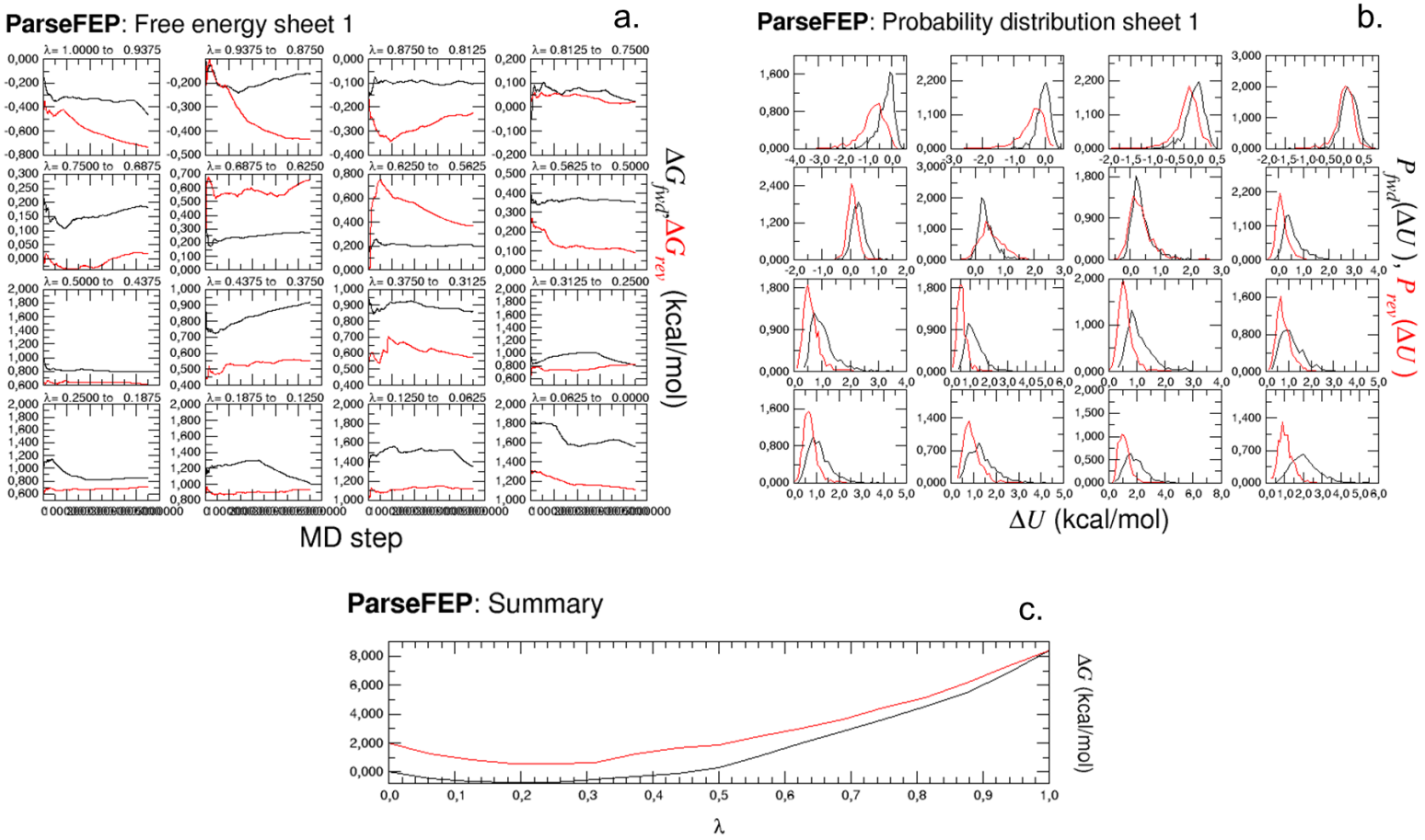

Figure 107. Time-evolution of the free energy, b) histograms of the probability distributions, and c) overall free energy change for the forward (black solid line) and backward (light solid line) transformations for the complex-system of perturbation intermediate-2 to glutamate.

Table 12. Overlap between λ windows of perturbations β-carbamate to glutamate and intermediate-2 to glutamate.

		β-carbamate → glutamate		Intermediate-2 → glutamate	
		Solvent	Complex	Solvent	complex
λ_K	$\lambda_{K+\delta i}$	P0_overlap_P1 (%)	P0_overlap_P1 (%)	P0_overlap_P1 (%)	P0_overlap_P1 (%)
0.00000	0.06250	58.27	48.39	90.57	76.91
0.06250	0.12500	58.28	48.39	87.13	46.73
0.12500	0.18750	58.38	48.61	80.69	63.26
0.18750	0.25000	63.60	58.16	87.13	82.69
0.25000	0.31250	58.38	61.60	86.79	56.16
0.31250	0.37500	25.60	49.83	76.69	76.14
0.37500	0.43750	67.48	58.49	77.80	81.69

0.43750	0.50000	63.60	63.37	69.26	44.06
0.50000	0.56250	60.49	56.16	83.13	49.28
0.56250	0.62500	60.93	46.28	78.47	38.96
0.62500	0.68750	58.93	58.05	83.57	44.17
0.68750	0.75000	59.49	61.04	76.03	72.59
0.75000	0.81250	58.05	41.51	76.58	62.60
0.81250	0.87500	56.83	61.04	71.03	64.59
0.87500	0.93750	52.28	53.61	72.59	50.83
0.93750	1.00000	49.83	41.62	66.93	39.07

Cycle closure error

The total energy of the cycle was calculated by adding all the differences of the binding free energies of the four systems (Figure 108).

$$\Delta G = (1.03 \pm 0.17) + (-1.54 \pm 0.13) + (2.36 \pm 0.29) + (-0.75 \pm 0.15)$$
$$= 1.10 \pm 0.41 \text{ kcal/mol}$$

It is observed that the cycle closure error improved crucially with the addition of the extra intermediate molecule, as the energy was calculated between two states with more similar configurations.

In addition, for the perturbation of β-carbamate to glutamate, which we are interested in, using the intermediate-2 (Figure 108), the difference of the binding free energy is:

$$\Delta\Delta G = (2.36 \pm 0.29) + (-0.75 \pm 0.15) = +1.61 \pm 0.32 \text{ kcal/mol}$$

The difference in the binding free energy is positive, indicating that the reference molecule, β-carbamate of BMAA has greater affinity than the natural agonist, glutamate. However, taking into consideration the error of the method and the standard deviation we can't state that β-carbamate has greater affinity, but it has certainly comparable affinity with glutamate.

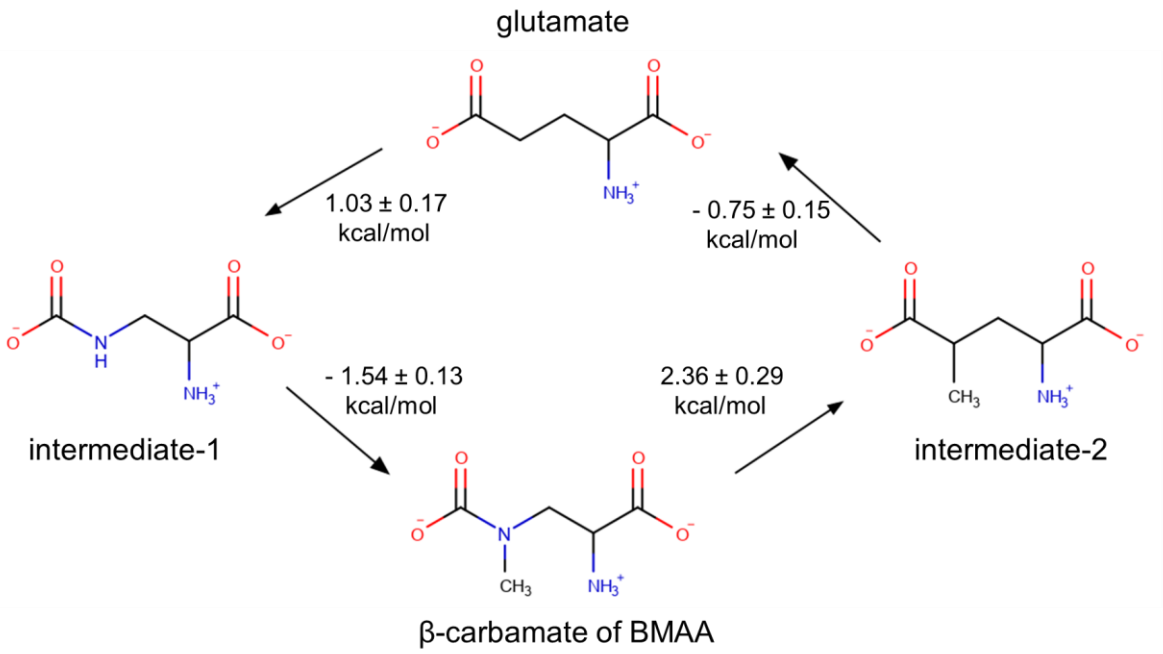

Figure 108. Cycle closure of glutamate, intermediate-1, β-carbamate and intermediate-2 with the difference of the binding free energy for each perturbation.

3.5.8 Convergence of the calculations by increase of the number of λ windows

The addition of an extra intermediate molecule improved remarkably the cycle closure error. Now, we will increase the number of λ windows and we will examine the results in the convergence and the cycle closure error. The addition of the λ windows will take place at the points, where the convergence is clearly low (overlap < 40%). In the first perturbation, glutamate to intermediate-1, in the complex-system, λ windows: λ=0,31250 to λ=0,37500, λ=0,81250 to λ=0,87500, λ=0.87500 to λ=0.93750 and λ=0.93750 to λ=1.0000 have convergence 36.5%, 36.7%, 23.6% and 35.0% respectively according to the ParseFEP.log file (also confirmed from Table 5 and Figure 97b). In addition, in the second perturbation, intermediate-1 to β-carbamate, in complex-system, λ windows: λ=0.5000 to λ=0,56250 and λ=0,81250 to λ=0,87500 have convergence 29.0% and 32.0% respectively (Table 7 and Figure 99b). Last, in perturbation of β-carbamate to intermediate-2, in the complex system, λ windows: λ=0.62500 to λ=0.68750, λ=0.75000 to λ=0.81250 and λ=0.87500 to λ=0.93750 have convergence 17.65%, 20.09%

and 26.08% respectively (Table 11 and Figure 105b). An extra λ window can be added in these cases in order to increase the convergence.

- *Extra λ windows in perturbation glutamate → intermediate-1*

  λ=0.31250 to λ=0.37500 was split: λ=0.31250 to λ=0.34000 and λ=0.34000 to λ=0.37500. The convergence increased from 36.51% to 70.70% and 70.26%.
  λ=0.81250 to λ=0.87500 was split: λ=0.81250 to λ=0.84000 and λ=0.84000 to λ=0.87500. The convergence increased from 36.74% to 46.17% and 51.50%.
  λ=0.87500 to λ=0.93750 was split: λ=0.87500 to λ=0.90000 and λ=0.90000 to λ=0.93750. The convergence increased from 23.64% to 69.92% and 83.80%.
  λ=0.93750 to λ=1.00000 was split: λ=0.93750 to λ=0.97000 and λ=0.97000 to λ=1.00000. The convergence increased from 34.96% to 49.50% and 35.85%.
  Also, the λ of the backward process changed accordingly.

Table 13. Binding free energy of perturbation glutamate to intermediate-1 using extra λ windows.

System	*Solvent*	*Complex*
Energy (kcal/mol)	-115.87 ± 0.07	-114.83 ± 0.13
Binding Free Energy = complex – solvent = ***-1.04 ± 0.17*** *kcal/mol*		

The free energy for perturbating glutamate to intermediate-1 in the complex system is -1.04 ± 0.13 kcal/mol and before insert λ windows it was -1.03 ± 0.17 kcal/mol. The convergence has improved significantly as shown in Table 15, however, the energy didn't change at all.

Table 14. Overlap between λ for perturbation of glutamate to intermediate-1 without extra λ and with the addition of extra λ.

glutamate → intermediate-1					
Without extra λ		**complex**	**With extra λ**		**complex**
λ_K	**$\lambda_{K+\delta i}$**	**P0_overlap_P1 (%)**	λ_K	**$\lambda_{K+\delta i}$**	**P0_overlap_P1 (%)**
0.00000	0.06250	45.17	0.00000	0.06250	45.17
0.06250	0.12500	45.17	0.06250	0.12500	45.17
0.12500	0.18750	46.17	0.12500	0.18750	46.17
0.18750	0.25000	77.25	0.18750	0.25000	77.25
0.25000	0.31250	50.06	0.25000	0.31250	50.06
0.31250	0.37500	**36.51**	0.31250	0.34000	**70.70**
			0.34000	0.37500	**70.26**
0.37500	0.43750	66.26	0.37500	0.43750	66.26
0.43750	0.50000	57.82	0.43750	0.50000	57.82
0.50000	0.56250	62.71	0.50000	0.56250	62.71
0.56250	0.62500	87.24	0.56250	0.62500	87.24
0.62500	0.68750	66.15	0.62500	0.68750	66.15
0.68750	0.75000	45.73	0.68750	0.75000	45.73
0.75000	0.81250	47.72	0.75000	0.81250	47.72
0.81250	0.87500	**36.74**	0.81250	0.84000	**46.17**
			0.84000	0.87500	**51.50**
0.87500	0.93750	**23.64**	0.87500	0.90000	**69.92**
			0.90000	0.93750	**83.80**
0.93750	1.00000	**34.96**	0.93750	0.97000	**49.50**
			0.97000	1.00000	**35.85**

- *Extra λ windows in perturbation intermediate-1 → β-carbamate of BMAA*

  λ=0.50000 to λ=0.56250 was split: λ=0.50000 to λ=0.53000 and λ=0.53000 to λ=0.56250. The convergence changed from 28.86% to 27.75% and 72.25%.
  λ=0.81250 to λ=0.87500 was split: λ=0.81250 to λ=0.84000 and λ=0.84000 to λ=0.87500. The convergence increased from 31.74% to 64.37% and 75.03% (Table 17).
  Also, the λ of the backward process changed accordingly.

Table 15. Binding free energy of perturbation intermediate-1 to β-carbamate of BMAA using extra λ windows.

System	*Solvent*	*Complex*
Energy (kcal/mol)	67.42 ± 0.08	65.99 ± 0.16
Binding Free Energy = complex – solvent = ***-1.43 ± 0.18*** *kcal/mol*		

The free energy for perturbating intermediate-1 to β-carbamate in the complex system is -1.43 ± 0.18 kcal/mol and before insert λ windows it was -1.54 ± 0.13 kcal/mol. The convergence has improved significantly as shown in Table 15, however, the energy changed slightly 0.1 kcal/mol.

Table 16. Overlap between λ for perturbation of intermediate-1 to β-carbamate of BMAA without extra λ and with the addition of λ.

intermediate-1→ β-carbamate of BMAA					
Without extra λ		**Complex**	**With extra λ**		**complex**
λ_K	**$\lambda_{K+\delta i}$**	**P0_overlap_P1 (%)**	**λ_K**	**$\lambda_{K+\delta i}$**	**P0_overlap_P1 (%)**
0.00000	0.06250	58.05	0.00000	0.06250	58.05
0.06250	0.12500	58.05	0.06250	0.12500	58.05
0.12500	0.18750	59.71	0.12500	0.18750	59.71
0.18750	0.25000	47.39	0.18750	0.25000	47.39
0.25000	0.31250	56.16	0.25000	0.31250	56.16
0.31250	0.37500	66.93	0.31250	0.34000	66.93

0.37500	0.43750	44.51	0.37500	0.43750	44.51
0.43750	0.50000	39.84	0.43750	0.50000	39.84
0.50000	0.56250	**28.86**	0.50000	0.53000	**27.75**
			0.53000	0.56250	**72.25**
0.56250	0.62500	73.47	0.56250	0.62500	73.47
0.62500	0.68750	69.92	0.62500	0.68750	69.92
0.68750	0.75000	67.70	0.68750	0.75000	67.70
0.75000	0.81250	62.71	0.75000	0.81250	62.71
0.81250	0.87500	**31.74**	0.81250	0.84000	**64.37**
			0.84000	0.87500	**75.03**
0.87500	0.93750	40.40	0.87500	0.90000	40.40
0.93750	1.00000	53.05	0.93750	0.97000	53.05

- *Extra λ windows in perturbation β-carbamate of BMAA → intermediate-2*

λ=0.62500 to λ=0.68750 was split: λ=0.62500 to λ=0.65000 and λ=0.65000 to λ=0.68750. The convergence increased from 17.65% to 63.15% and 48.95%. λ=0.75000 to λ=0.81250 was split: λ=0.75000 to λ=0.78000 and λ=0.78000 to λ=0.81250. The convergence improved from 31.74% to 64.37% and 75.03% λ=0.87500 to λ=0.93750 was split: λ=0.87500 to λ=0.90000 and λ=0.90000 to λ=0.93750. The convergence increased from 26.08% to 43.62% and 75.80% (Table 19).
In addition, the λ of the backward process changed accordingly.

Table 17. Binding free energy of perturbation β-carbamate of BMAA to intermediate-2 using extra λ windows.

System	*Solvent*	*Complex*
Energy (kcal/mol)	71.68 ± 0.07	74.14 ± 0.27
Binding Free Energy = complex – solvent = ***2.46 ± 0.29*** *kcal/mol*		

The free energy for transforming β-carbamate to intermediate-2 in the complex system is 2.46 ± 0.29 kcal/mol and before insert λ windows it was

2.36 ± 0.29 kcal/mol. The convergence has improved significantly as shown in Table 19, however, the energy changed slightly 0.1 kcal/mol.

Table 18. Overlap between λ for perturbation of intermediate-1 to β-carbamate of BMAA without extra λ and with the addition of λ.

intermediate-1→ β-carbamate of BMAA					
Without extra λ		**Complex**	**With extra λ**		**complex**
λ_{κ}	**$\lambda_{\kappa+\delta i}$**	**P0_overlap_P1 (%)**	**λ_{κ}**	**$\lambda_{\kappa+\delta i}$**	**P0_overlap_P1 (%)**
0.00000	0.06250	45.06	0.00000	0.06250	45.06
0.06250	0.12500	66.15	0.06250	0.12500	66.15
0.12500	0.18750	90.90	0.12500	0.18750	90.90
0.18750	0.25000	86.57	0.18750	0.25000	86.57
0.25000	0.31250	51.83	0.25000	0.31250	51.83
0.31250	0.37500	69.70	0.31250	0.34000	69.70
0.37500	0.43750	65.04	0.37500	0.43750	65.04
0.43750	0.50000	77.36	0.43750	0.50000	77.36
0.50000	0.56250	78.14	0.50000	0.53000	78.14
0.56250	0.62500	60.49	0.56250	0.62500	60.49
0.62500	0.68750	**17.65**	0.62500	0.65000	**63.15**
			0.65000	0.65750	**48.95**
0.68750	0.75000	59.38	0.68750	0.75000	59.38
0.75000	0.81250	**20.09**	0.75000	0.78000	**71.37**
			0.78000	0.81250	**83.57**
0.81250	0.87500	64.37	0.81250	0.84000	64.37
0.87500	0.93750	**26.08**	0.87500	0.90000	**43.62**
			0.90000	0.93750	**75.80**
0.93750	1.00000	47.39	0.93750	1.00000	47.39

Cycle closure error

The total energy of the cycle was calculated by adding all the differences of the binding free energies of the four systems.

$\Delta G = (1.04 \pm 0.17) + (-1.43 \pm 0.18) + (2.46 \pm 0.29) + (-0.75 \pm 0.15)$

$= 1.30 \pm 0.41$ kcal/mol

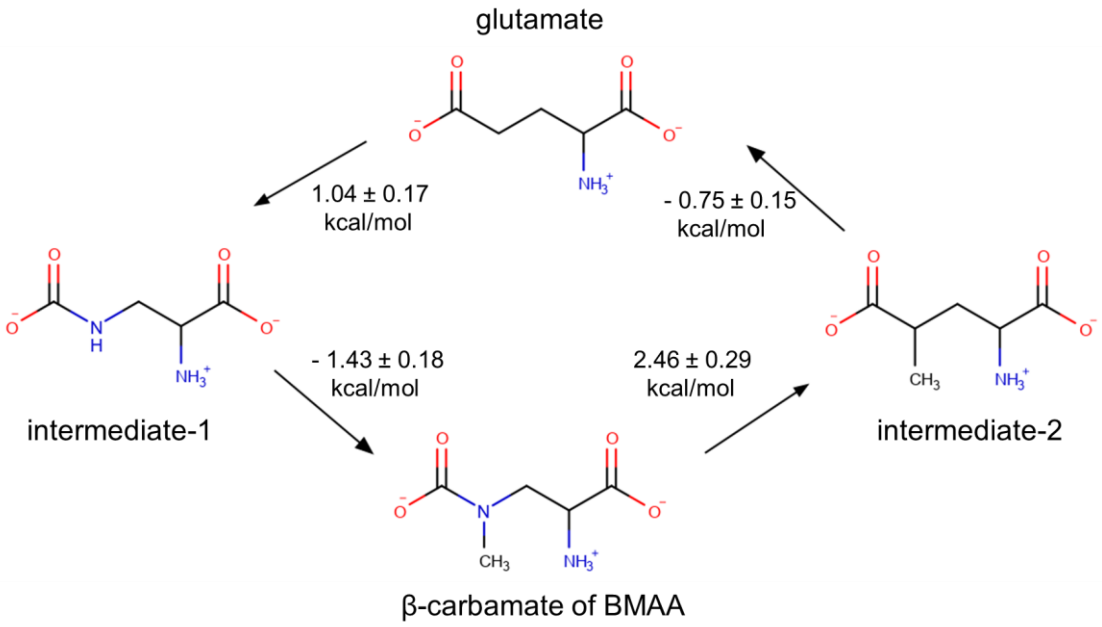

Figure 109. Cycle closure of glutamate, intermediate-1, β-carbamate and intermediate-2 with the difference of the binding free energy for each perturbation after the addition of λ windows.

The cycle closure error is 1.30 ± 0.41 kcal/mol, while before adding the λ windows it was 1.10 ± 0.41 kcal/mol. The change is negligible as it is covered from the standard deviation error.

4 CONCLUSIONS

MD simulations has evolved into a mature theoretical technique that can be used effectively to understand macromolecular structure-to-function relationships, which cannot be achieved by smaller-scale simulations such as quantum mechanical calculations due to restrictions in computational power or existing experimental approaches alone. In this context, with MD simulations we were able to study the stability of glutamate, BMAA and its carbamates, both in their L- and D- form, inside the AMPA glutamate receptor in atomic-level detail. In addition, with MD simulations coupled with FEP calculations, we managed to investigate the binding affinity of BMAA and its carbamate adducts with respect to glutamate in the binding site of the AMPA receptor using a robust method for predicting ligand-protein binding affinities via molecular simulations. Using this method, we inspected whether the L-ligands have better affinity to the receptor with respect to the natural agonist, glutamate, which would might be causing the effects of neurodegenerative diseases.

First, we studied the stability of L-glutamate in the binding pocket using MD simulations. It has been found that L-glutamate remains in the binding pocket of the receptor after 100 ns. Next, we studied the stability of L-α-carbamate, L-β-carbamate and L-BMAA in the AMPA receptor using the docking procedure as preparation method for the systems. The results revealed that all of the molecules, also remain in the binding pocket of the receptor after 100 ns with more stable being L-β-carbamate, then L-α-carbamate, and L-BMAA. Remarkably, L-β-carbamate, also the molecule with the highest structural similarity to glutamate, presented enhanced stability even from the natural L-glutamate, indicating that is the most promising molecule for binding with better affinity to the AMPA receptor.

Subsequently, the stability of D-BMAA and its carbamate adducts was studied in the AMPA receptor. The docking procedure was also used during the preparation of the systems. It was revealed that D-BMAA was the most stable molecule, then D-β-carbamate, and last D-α-carbamate. Afterwards, the superimposition/minimization method was used for the preparation of the

systems. Interestingly, the results were completely different using the two methods. With superimposition/minimization, the most stable molecule was D-α-carbamate, then D-BMAA and last D-β-carbamate. The latter, not only was the less stable molecule between the three, but it left the binding pocket of the receptor. By visualizing D-β-carbamate after the sumperimposition /minimization it was observed that the molecule adopted a different conformation (rotation of 180^{o} of dihedral N-C-C-C) and possibly this is the reason why it could not be able to find the appropriate interactions and remain located in the binding pocket of the receptor during the simulation. However, we will not pursue with this method further.

Finally, in this study, the difference in the free energy of binding between L-β-carbamate and glutamate was calculated using FEP/MD calculations. For the calculation, initially one intermediate molecule was used. The closed cycle error was ca. -2.70 kcal/mol instead of ± 1.50 kcal/mol. In addition, the difference of the free energy of binding of β-carbamate to glutamate was -2.22 ± 0.17 kcal/mol, which indicates that the target molecule, glutamate, has greater affinity. However, the error of the closed cycle is higher than the acceptable value of 1.50 kcal/mol, which is the error of the FEP/MD method, and as a result we investigated another approach. With the addition of a second intermediate molecule in the closed cycle, the result improved crucially, as the cycle closure error was 1.10 ± 0.41 kcal/mol. The difference in the free energy of binding between β-carbamate and glutamate is +1.61 ± 0.32 kcal/mol, indicating that the reference molecule, β-carbamate, has comparable affinity with the natural glutamate in the AMPA receptor. Last, we investigated a potential increase of the simulation convergence using extra λ windows, which led to negligible cycle closed error change.

We conclude that the results from this work provided significant insight of the stability of BMAA and its carbamate adducts in the AMPA gluR. Especially, L-β-carbamate of BMAA, revealed enhanced stability in the binding pocket of the AMPA receptor, comparable affinity with respect to the natural agonist, glutamate in the AMPA gluR, and the inference that it may act as potential inhibitor of the AMPA gluR.

FUTURE PERSPECTIVES

Although the results that we obtained are significant, further investigation is needed to cover all of the aspects of the binding affinity of the other molecules. More calculations need to be completed concerning the calculation of the difference in the free energy of binding between glutamate and L-α-carbamate as well as between glutamate and L-BMAA. This further study would give us a better aspect of the binding affinity of the rest of the molecules.

In addition, QM simulations could be performed to inspect the chemical reaction of BMAA with bicarbonate in human. So far, it is believed that BMAA interacts with bicarbonate to produce carbamate adducts, however, this specific interaction hasn't been studied through MD simulations so far.

Furthermore, as mentioned at the introduction, BMAA has quite a large number of isomers, included molecules that have presented also toxic activity. The study of the affinity of isomers of BMAA, which present toxic activity, on glutamate receptors opens space to explorations of great biological interest.

Finally, absolute free energy calculations can be performed in order to compute the absolute binding free energy of glutamate, and β-carbamate adduct of BMAA on the AMPA gluR. The result regarding the computation of the absolute binding free energy of glutamate can be compared to existed experimental data [74] as well as computational data [56]. This calculation will validate the method and will probably help us to understand which ligand has better affinity to the AMPA receptor.

ABBREVIATIONS - ACRONYMS

ALS	Amyotrophic Lateral Sclerosis
ATD	Amino Terminal Domain
LBD	Ligand Binding Domain
TMD	Transmembrane Domain
iGluR	Ionotropic Glutamate Receptor
mGluR	Metabotropic Glutamate Receptor
MD	Molecular Dynamics
PDB	Protein Data Bank
FEP	Free Energy Perturbation
PME	Particle Mesh Ewald
RMSD	Root Mean Square Deviation

REFERENCES

1. *Porojan, C., et al. (2016). "Overview of the potent cyanobacterial neurotoxin beta-methylamino-L-alanine (BMAA) and its analytical determination." Food Addit Contam Part A Chem Anal Control Expo Risk Assess 33(10): 1570-1586.*
2. *Kim, S. Y. and S. Rydberg (2020). "Transfer of the Neurotoxin Beta-N-methylamino-l-alanine (BMAA) in the Agro-Aqua Cycle." Mar Drugs 18(5).*
3. *Goto, J. J., et al. (2012). "The physiological effect of ingested Beta-N-methylamino-L-alanine on a glutamatergic synapse in an in vivo preparation." Comp Biochem Physiol C Toxicol Pharmacol 156(3-4): 171-177.*
4. *Chiu, A. S., et al. (2011). "Does alpha-amino-beta-methylaminopropionic acid (BMAA) play a role in neurodegeneration?" Int J Environ Res Public Health 8(9): 3728-3746.*
5. *Spencer, P.S., Nunn, P.B., Hugon, J., Ludolph, A.C., Ross, S.M., Roy, D.N., Robertson, R.C., 1987a. Guam amyotrophic lateral sclerosis- parkinsonism-dementia linked to a plant excitant neurotoxin. Science 237, 517-522.*
6. *Spencer, P.S., Ohta, M., Palmer, V.S., 1987b. Cycad use and motor neuron disease in Kii Peninsula of Japan [letter]. Lancet ii, 1462-1463.*
7. *Brownson, D. M., et al. (2002). "The cycad neurotoxic amino acid, β-N-methylamino-L-alanine (BMAA), elevates intracellular calcium levels in dissociated rat brain." J. Ethnopharmacol. 82: 159-167.*
8. *Jonasson, S., et al. (2010). "Transfer of cyanobacterial neurotoxin with a temperate aquatic ecosystem suggest pathwas for human exposure." Proceedings of the National Academy of Sciences of the United States of America 107(20): 9252-9257.*
9. *Weiss, J. H. and D. W. Choi (1988). "Beta-N-methylamino-L-alanine neurotoxicity: requirement for bicarbonate as a cofactor." Science 241(4868): 973-975.*
10. *Allen, C. N., et al. (1993). "β-N-Methylamino-L-Alanine in the presence of bicarbonate is an agonist at non-N-Methyl-D-Aspartate-type receptors." Neuroscience 54(3): 567-574.*
11. *Zimmerman, D., et al. (2016). "Equilibrium Dynamics of beta-N-Methylamino-L-Alanine (BMAA) and Its Carbamate Adducts at Physiological Conditions." PLoS One 11(8): e0160491.*

12. *Diaz-Parga, P., et al. (2018). "Chemistry and Chemical Equilibrium Dynamics of BMAA and Its Carbamate Adducts." Neurotox Res 33(1): 76-86.*
13. *Jiang, L., et al. (2012). "Selective LC-MS/MS method for the identification of BMAA from its isomers in biological samples." Anal Bioanal Chem 403(6): 1719-1730.*
14. *Schneider, T., et al. (2020). "Neurotoxicity of isomers of the environmental toxin L-BMAA." Toxicon 184: 175-179.*
15. *Mayer, M. L. (2005). "Crystal structures of the GluR5 and GluR6 ligand binding cores: molecular mechanisms underlying kainate receptor selectivity." Neuron 45(4): 539-552.*
16. *Catarzi, D., et al. (2006). "Competitive Gly/NMDA Receptor Antagonists." Curr Top Med Chem 6(8): 809-821.*
17. *Zhu, S. and E. Gouaux (2017). "Structure and symmetry inform gating principles of ionotropic glutamate receptors." Neuropharmacology 112(Pt A): 11-15.*
18. *Tsuchiya, D., et al. (2002). "Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd3+." Proceedings of the National Academy of Sciences of the United States of America 99(5): 2660-2665.*
19. *Macalino, S. J., et al. (2015). "Role of computer-aided drug design in modern drug discovery." Arch Pharm Res 38(9): 1686-1701.*
20. *Yu, W. and A. D. MacKerell, Jr. (2017). "Computer-Aided Drug Design Methods." Methods Mol Biol 1520: 85-106.*
21. *Friesner, R. A., et al. (2006). "Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes." J. Med. Chem. 49: 6177-6196.ALLEN, M. P., AND TILDESLEY, D. J. Computer Simulation of Liquids. Oxford University Press Inc., New York, USA, 2001.*
22. *Yoshida, N. (2017). "Role of Solvation in Drug Design as Revealed by the Statistical Mechanics Integral Equation Theory of Liquids." J Chem Inf Model 57(11): 2646-2656.*
23. *Gupta V.P., Principles and Applications of Quantum Chemistry, 1st edition, 2015*
24. *Yang, Y. and B. Li (2019). "A simulation algorithm for Brownian dynamics on complex curved surfaces." J. Chem. Phys. 151.*
25. *LEACH, A. R. Molecular Modelling. Principles and Applications. Longman, Essex, England, 1999.*

26. *S. A. Adcock and J. A. McCammon, "Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins," Chem. Rev., 2006.*
27. *Cournia Z., Structure and Dynamics of Biomembranes containing Cholesterol and other Biologically-Important Sterols. A computational perspective., 2006*
28. *A. D. MacKerell, J. Wiorkiewicz-Kuczera, and M. Karplus, "An all-atom empirical energy function for the simulation of nucleic acids," J. Am. Chem. Soc., vol. 117, no. 48, pp. 11946–11975, Dec. 1995.*
29. *A. D. MacKerell et al., "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †," J. Phys. Chem. B, vol. 102, no. 18, pp. 3586–3616, Apr. 1998.*
30. *"Parameter Files." [Online]. Available: https://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node25.html. [Accessed: 18-Jun-2019].*
31. *Lennard-Jones potential @ Chemistry Dictionary & Glossary.*
32. *M. P. Allen and D. J. Tildesley, Computer simulation of liquids: Second edition. 2017.*
33. *P. G. Huray, Maxwell's equations. Wiley, 2010.*
34. *EWALD, P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Ann. Phys., 1921, 64, 253-287.*
35. *Cournia, Z. The ewald summation method: Calculating long-range interactions. In: Beitr¨age zumWissenschaftlichen Rechnen Ergebnisse des Gaststudentenprogramms 2000 des John von Neumann-Instituts f¨ur Computing, Zentralinstitut f¨ur Angewandte Mathematik, 2000, pp. 27-38.*
36. *D. C. Rapaport and D. C., The art of molecular dynamics simulation. Cambridge University Press, 1995.*
37. *D. Frenkel and B. Smit, Understanding molecular simulation: from algorithms to applications.*
38. *J. C. Smith Michael Grunze, "Structure and Dynamics of Biomembranes containing Cholesterol and other Biologically-Important Sterols. A computational perspective. Referees."*
39. *A. Cauchy. (1847) Methodes generales pour la resolution des syst'emes dequations simultanees,. C.R. Acad. Sci. Par., 25:536–538.*
40. *"1.7: The Maxwell Distribution Laws - Chemistry LibreTexts." [Online]. Available: https://chem.libretexts.org/Courses/University_of_California_Davis/UCD_Chem_107B%3A_Physical_Chemistry_for_Life_Scientists/Chapters/1%3A_Prope*

rties_of_Gases/1.7%3A_The_Maxwell_Distribution_Laws. [Accessed: 11-Jun-2019].

41. *W. Phadungsukanan, M. Kraft, J. A. Townsend, and P. Murray-Rust, "The semantics of Chemical Markup Language (CML) for computational chemistry: CompChem," J. Cheminform., 2012.*
42. *"Theory of Molecular Dynamics Simulations." [Online]. Available: https://embnet.vital-it.ch/MD_tutorial/pages/MD.Part1.html. [Accessed: 07-Jun-2019].*
43. *Moore, C. C. (2015). "Ergodic theorem, ergodic theory, and statistical mechanics." Proceedings of the National Academy of Sciences 112(7): 1907-1911.*
44. *J. Grotendorst and Winter school (2002.02.25-03.01: Kerkrade), Quantum simulations of complex many-body systems: from theory to algorithms : winter school, 25 February - 1 March 2002, Rolduc Conference Centre, Kerkrade, the Netherlands ; lecture notes. NIC-Secretariat, 2002.*
45. *R. W. Hakala, (1961), "A new derivation of the Boltzmann distribution law," J. Chem. Educ 38(1):33*
46. *D. A. McQuarrie (2000), Statistical mechanics. University Science Books.*
47. *Braun, E., et al. (2019). "Best Practices for Foundations in Molecular Simulations [Article v1.0]." Living J Comput Mol Sci 1(1).*
48. *"Periodic Boundary Conditions." [Online]. Available: http://isaacs.sourceforge.net/phys/pbc.html. [Accessed: 18-Jun-2019].*
49. *Cournia, Z., et al. (2020). "Rigorous Free Energy Simulations in Virtual Screening." J Chem Inf Model 60(9): 4153-4169.*
50. *Mey, A. S. J. S., et al. (2020). "Best Practices for Alchemical Free Energy Calculations." LiveCoMS.*
51. *M. R. Shirts and J. D. Chodera. The Journal of Chemical Physics 129 (2008).*
52. *Cournia, Z., et al. (2017). "Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations." J Chem Inf Model 57(12): 2911-2937.*
53. *Armstrong, N. and E. Gouaux (2000). "Mechanisms for Activation and Antagonism of an AMPA-Sensitive Glutamate Receptor: Cry." Neuron 28: 165-181.*
54. *McGinnis, S. and T. L. Madden (2004). "BLAST: at the core of a powerful and diverse set of sequence analysis tools." Nucleic Acids Res 32(Web Server issue): W20-25.*

55. *L. Schrödinger, Maestro Schrödinger Release 2017-1, New York, NY, 2017*
56. *Heinzelmann, G., et al. (2014). "Computation of standard binding free energies of polar and charged ligands to the glutamate receptor GluA2." J Phys Chem B 118(7): 1813-1824.*
57. *G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, J. Comput.-Aided Mol. Des., 2013, 27, 221–234*
58. *Vanommeslaeghe, K., et al. (2010). "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields." J Comput Chem 31(4): 671-690.*
59. *Huang, J., et al. (2017). "CHARMM36m: an improved force field for folded and intrinsically disordered proteins." Nat Methods 14(1): 71-73.*
60. *Humphrey, W., et al. (1996). "VMD: Visual Molecular Dynamics " J. Mol. Graph. 14: 33-38.*
61. *Phillips, J. C., et al. (2005). "Scalable Molecular Dynamics with NAMD." J Comput Chem. 26(16): 1781-1802.*
62. *Davidchack, R. L., et al. (2009). "Langevin thermostat for rigid body dynamics." J Chem Phys 130(23): 234101.*
63. *Feller, S. E., et al. (1995). "Constant pressure molecular dynamics simulation: The Langevin piston method." The Journal of Chemical Physics 103(11): 4613-4621.*
64. *Jakobsen, A. F. (2005). "Constant-pressure and constant-surface tension simulations in dissipative particle dynamics." J Chem Phys 122(12): 124901.*
65. *Racz, A., et al. (2018). "Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints." J Cheminform 10(1): 48.*
66. *Karatzas, E., et al. (2020). "ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing." Bioinformatics 36(8): 2602-2604.*
67. *Dodda, L. S., et al. (2017). "LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands." Nucleic Acids Res 45(W1): W331-W336.*
68. *Pettersen, E. F., et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem 25(13): 1605-1612.*
69. *Zavitsanou V. S. (2018). "Automating Free Energy Perturbation Calculations*
70. *Robertson, M. J., et al. (2015). "Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field." J Chem Theory Comput 11(7): 3499-3509.*

71. *Andersen, H. C. (1980). "Molecular dynamics simulations at constant pressure and/or temperature." The Journal of Chemical Physics 72(4): 2384-2393.*
72. *Berendsen, H. J. C., et al. (1984). "Molecular dynamics with coupling to an external bath." The Journal of Chemical Physics 81(8): 3684-3690.*
73. *Parrinello M, Rahman A. (1981). "Polymorphic transitions in single crystals: A new molecular dynamics method." Journal of Applied Physics 52(12): 7182-7190*
74. *Guo-Qiang, C. and G. Eric (1997). "Overexpression of a glutamate receptor (GluR2) ligand binding domain in Escherichia coli: Application of a novel protein folding screen." Proceedings of the National Academy of Sciences of the United States of America 94: 13431-13436.*