Light-induced conformational changes in proteins

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften
Karlsruher Institut für Technologie (KIT) - Universitätsbereich
genehmigte

DISSERTATION

von
Tino Wolter
aus
Celle

Dekan: Prof. Dr. Peter Roesky
Referent: Prof. Dr. Marcus Elstner
Korreferent: Prof. Dr. Wolfgang Wenzel
Tag der mündlichen Prüfung: 20.12.2013

Druckjahr: 2013

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet, sowie die Satzung der Universität Karlsruhe (TH) zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Tino Wolter
06.11.2013

Contents

1 Introduction 11

1.1 Bacteriorhodopsin 12

1.2 Light-switchable ionotropic glutamate receptors 17

2 Theory 23

2.1 Introduction 23

2.2 Molecular Mechanics/Molecular Dynamics 23

2.3 Quantum Chemistry Methods 25

2.4 QM/MM 33

2.5 Free Energy Calculations 36

2.6 Docking 39

3 Early states in the bR photocycle 43

3.1 K state 43

3.2 L state 53

3.3 Conclusion 61

4 Dynamical properties of the ligand binding domain of iGluR2 63

4.1 Evaluation of known ligands 63

4.2 Energetics of the LBD 77

5 Development of photoswitchable ligands targeting iGluR2 85

5.1 Introduction 85

5.2 Docking 86

5.3 MD simulations 90

5.4 ATA-2 91

5.5 ATA-3 95

5.6 Discussion 101

5.7 Conclusion 104

6 Summary 107

A bR 121

A.1 L state 121

B Evaluation of known ligands 123

B.1 MD simulation of known ligands 123

C Development of photo-switchable ligands targeting iGluR2 125

C.1 Photo-switch mechanism of ATA-2 125

C.2 Photo-switch mechanism of ATA-3 128

Zusammenfassung

In dieser Arbeit werden durch Licht induzierte Konformationsänderungen in zwei verschiedenen Proteinen untersucht. Bacteriorhodopsin (bR), welches in seiner nativen Form lichtsensitiv ist und zum anderen ein ionotroper Glutamatrezeptor (iGluR2), der erst durch das Binden von photoschaltbaren Liganden durch Licht aktivierbar wird.

Bacteriorhodopsin ist durch eine Vielfalt von Experimenten, einschließlich struktureller Aufklärung, optischen und Schwingungspektren, sehr genau untersucht. Mit der Hilfe dieser Daten, sind die großen Konformationänderungen im Protein hinreichend verstanden, während die strukturellen Details noch strittig sind. In diesem wohl definierten Szenario können computergestützte Berechnungen dazu genutzt werden, um Strukturen anhand ihrer Anregungsenergie zu überprüfen. Im Gegensatz dazu sind nur wenige experimentelle Daten für iGluR2 erhältlich. Diese beschreiben eine sehr langsame Domänenbewegung, die durch das Binden eines Ligandens hevorgerufen wird.
Bedingt durch diese unterschiedlichen Phänomene, wurden in dieser Arbeit verschiedenste Methoden verwendet, die von sehr genauen *ab initio* Methoden bis hinzu klassischen Kraftfeldrechnungen reichen.
Mithilfe dieser Methoden wurden verfeinerte Strukturmodelle der Bindungtasche für die frühen Zwischenzustände von bR entwickelt, um einen detailierteren Einblick in die Intermediate zu geben.
Für die Ligandenbindungsdomäne des Glutamatrezeptors wurden innerhalb dieser Arbeit die Dynamiken und die Energetik untersucht. Die daraus resultierenden Erkenntnisse wurden zur Entwicklung eines Mechanismus genutzt, der die Aktivierung des Rezeptors durch Licht mithilfe von photoschaltbaren Liganden erklärt.

Abstract

In this work, light-induced conformational changes in two different proteins are investigated. Bacteriorhodopsin (bR), which is already light-sensitive in its native form and an ionotropic glutamate receptor (iGluR2), that can be light-activated by binding a photo-switchable ligand.

Bacteriorhodopsin is very well studied by a plurality of experiments regarding structural, vibrational and optical data. Based on these data, the coarse motions of the protein are properly understood, while the diminutive structural changes are still under debate. In this well characterized case, computations can be used to evaluate structures based on spectroscopic data. On the other hand, only few experimental data are available for iGluR2, which depict very slow domain motions within the ligand binding domain due to the binding of a ligand.
Due to the varying phenomena, several methods were used in this work, ranging from very accurate *ab initio* methods to classical force field calculations.

With these tools, refined structural models of the bR binding pocket of the early intermediates could be developed.
Regarding the ligand binding domain of the glutamate receptor, the dynamics and energetics were investigated. Based on the resulting insights, a mechanism was developed that explains the light-activation of the receptor caused by a photo-switchable ligand.

Chapter 1

Introduction

Sunlight is the fundamental energy source for life on earth. Beside the important heating of our planet, it is used for energy extraction of phototrophs, mainly by the process of photosynthesis. Furthermore, light is used as a source of information for orientation and adapting to the changing environment. For this purpose bacteria, algae and higher plants have developed sensory pathways with photoreceptors in the first place. A well known and investigated family are the rhodopsins, which not only have a wide distribution in biology from algae and bacteria to higher mammals including humans, but also serve both fields of using light: energy extraction and sensory. All rhodopsins have a retinal as a cofactor which is covalently bound to a lysine via a protonated Schiff base. The retinal is located within a binding pocket formed by the seven transmembrane helices. Due to light excitation the retinal isomerizes and a sequence of structural changes follows. The family of rhodopsins can be divided into two subgroups: (i) microbial opsins (type I) and (ii) animal opsins (type II).[1] Beside the just mentioned similarities, there are several distinct differences. The tertiary structure, which shows a high similarity within the groups, depicts a completely different alignment of the seven transmembrane helices between the groups.[1] This might be explained by high sequence homology within the families,[2] which is nonexistent between the two subgroups. The second main difference is the retinal configuration. While the type I rhodopsins have an all-*trans* retinal, which isomerizes around the C13=C14 double bond to a 13-*cis* configuration (Fig. 1.1), the type II family has an 11-*cis* retinal which changes into an all-*trans* configuration upon photoabsorption. After the photoisomerization the retinal of type I rhodopsins stays attached to the opsin and thermally turns back to the starting configuration.[3] In case of type II rhodopsins the retinal dissociates from the opsin after isomerization and a new 11-*cis* retinal must be attached to the protein. Furthermore, the retinal configuration differs for archaeal and visual rhodopsins for the C6-C7 bond, linking the polyene chain to the β-ionone ring, showing a 6-*s*-*trans* versus a 6-*s*-*cis*, respectively.

Due to the focus on bacteriorhodpsin, a protein of the type I family, in the first part of this thesis, the type II rhodopsins will not be discussed further. Type I rhodopsins are found in

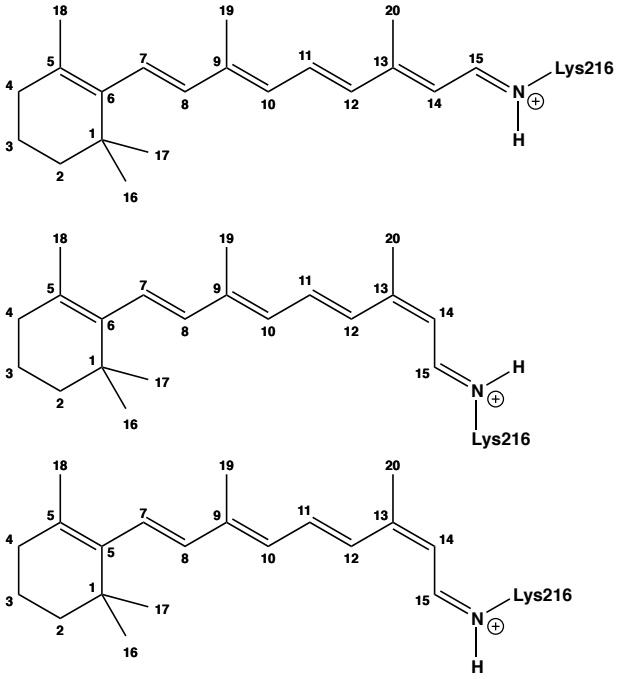


Figure 1.1 – Different configurations of the retinal in bR. Top: all-*trans*; mid: 13-*cis*, 15-*anti*; bottom: 13-*cis*, 15-*syn*

algae, prokaryotes and fungi and are responsible for wide range of functions, like energy extraction, phototaxis and retinal biosynthesis.[4]

1.1 Bacteriorhodopsin

The light-activated ion pump bacteriorhodopsin (bR), found in the early 1970s,[5] participates in the feature of energy extraction in the purple membrane of *Halobacterium salinarum*. Induced by the photoabsorption bacteriorhodopsin translocates a proton across the cell membrane[6] from the cytoplasmic to the extracellular side. The resulting membrane potential is used by the cell for ATP synthesis.[7] The seven transmembrane helices are organized in trimer units forming a two-dimensional hexagonal lattice.[8,9] In the absence of light, a mixture of all-*trans* and 13-*cis*, 15-*syn* retinals with a ratio of 2:1 is present[10,11], while only the proteins with an all-*trans* retinal show proton pump activity under physiological conditions.[12] In the ground state the positive charge of the Schiff base is stabilized by negatively charged counterion residues Asp85 and Asp212. Together with three water molecules a pentagonal water cluster is formed via a hydrogen bonded network (HBN) (see Fig. 1.2).[13,14] The HBN of the binding pocket is connected to the proton release group (PRG) at the extracellular side by Tyr57, Arg82 and additional water molecules.

1.1.1 Bacteriorhodopsin Photocycle

The photocycle is initiated by light absorption of the all-*trans* retinal which is followed by several spectroscopically discriminable intermediates and ends up in the ground state

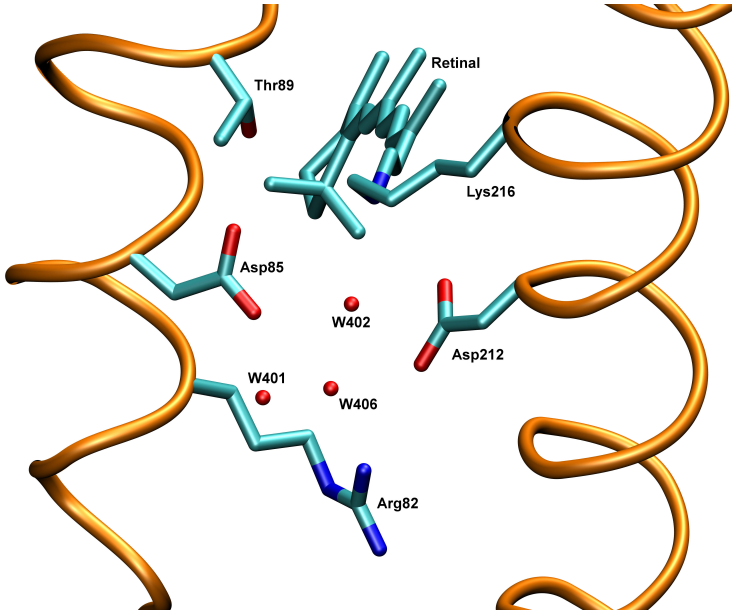


Figure 1.2 – A snapshot of the binding pocket of the bR ground state indicating the hydrogen bonding network at the extracellular side. The picture is based on the crystal structure of ref. 13

again. During the photocycle exactly one proton is pumped against the membrane potential. Since there is no universally accepted photocycle as spectral features depend on factors like pH,[15,16] temperature, excitation conditions[17,18] and existence of spectrally silent transitions[16], only a widely accepted photocycle is shown in Fig. 1.3(a) (for reviews, see refs.[19,20])
Especially for the two early intermediates K and L plenty of experimental data is available. Though, there is no generally accepted description for both states. Since this thesis focuses on these two states, a detailed introduction for the two intermediates is given in Sec. 1.1.2 and 1.1.3

In the following there is a short summary of the main events during the complete photocycle. The photoexcitation leads to the formation of the excited state I_{460} which decays within approximately 500 fs through a conical intersection into a thermal intermediate J_{625} in the ground state,[21,22] which relaxes into the first stable intermediate K_{590} (Fig.1.3; step 1). The K → L transition occurs within approximately 1 μs. During this transition, the possibility of a KL intermediate and its spectral features is discussed in literature.[25,26]

During the next transition L → M (10 μs[16]) the first proton transfer from the Schiff base towards the counterion Asp85 takes place (step 2).[27--30] Additionally, a proton is released on the extracellular side of the protein to the bulk water by a group of amino acids known as the proton release group (PRG) (step 3). In the following the Schiff base is reprotonated from the cytoplasmic side by Asp96 during the M → N transition (step 4). Asp96 itself is subsequently (N → O) reprotonated from the bulk water and the retinal reisomerizes from 13-*cis* to all-*trans* (step 5). The photocycle is completed during the O → bR transition, as

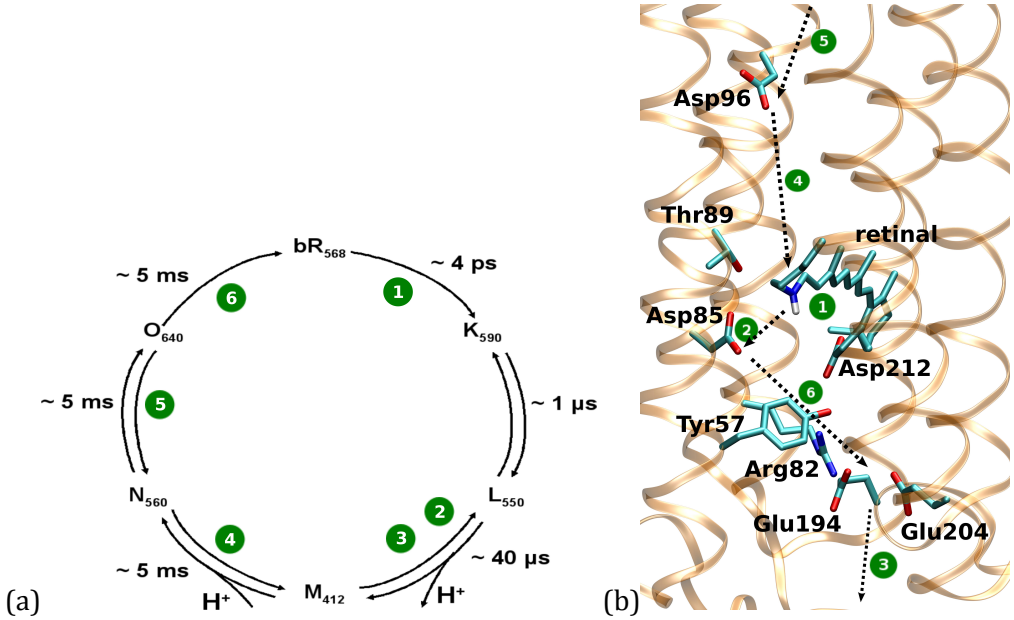


Figure 1.3 – (a) a schematic representation of the bR photocycle. (b) A snaphot of bR showing the important residues for the proton transfer. The arrows indicate the single proton transfer steps. The numbers indicate the sequence.

the proton is transferred from Asp85 to the PRG (step 6).

1.1.2 K state

After the photoinduced isomerization of the retinal the K state rises within a few picoseconds with a 13-*cis*, 15-*anti* retinal configuration.[17] Several X-ray structures (Fig. 1.4)[31--33] and cryo-electron microscopy experiments,[34] suggest that no significant structural changes within the protein occur, and that the small rearrangements are localized in the retinal and its direct surroundings. Unfortunately, the derived models for the K state at low temperatures differ in the details of the binding pocket. While the crystal structure of Edman *et al.*[32](pdb code: 1qko), cryo-trapped at 110 K, proposes a nearly planar retinal with the Schiff base oriented towards the cytoplasmic side, both structures which were trapped at 100 K[31,33] (pdb code: 1ixf and 1m0k, respectively) are indicating a highly twisted retinal with a Schiff base pointing to the carboxylic group of Asp212. A second controversy between the derived models is the location of w402. The structures observed at 100 K show w402 in a position similar to the ground state[13] (pdb code: 1c3w) which is supported by FTIR studies.[35] On the other hand, the structure of ref. 32 proposes the absence of this water molecule which is consistent with the observation from low-temperature FTIR indicating the loss of Schiff base hydrogen bond in K.[36] Another open issue is the possible existence of several K substates. It has been proposed that at 90 K there are at least three K photoproducts which can be distinguished by their UV/Vis absorption maximum.[37] Additionally, diverse spectroscopic experiments indicate that the K state passes trough several substates from the twisted to the relaxed and planar chromophore.[17,38--44]

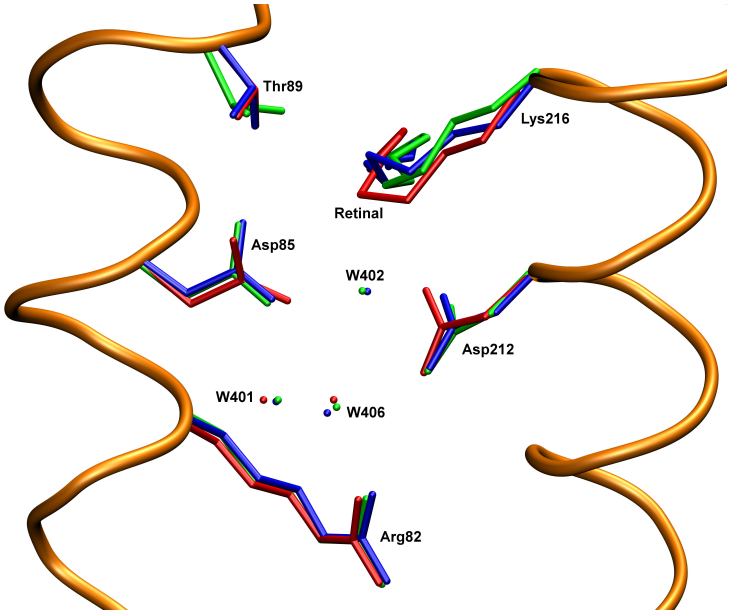


Figure 1.4 – Comparison of crystal structures proposed for the K state. The crystal structures from refs. 33 (1m0k), 32 and 31 depicted in blue, red and green, respectively

While at low temperatures the retinal is highly twisted, time-resolved Raman spectroscopy at room temperature proposed a fast relaxation of the retinal within picoseconds.[39] Furthermore, Shichida *et al.* proposed a new photointermediate (λ_{max}=596 nm) which was called KL.[25] A later study indicated, that the K→KL transition is spectrally silent.[45] Recent time-resolved UV/Vis experiments observed also two spectroscopically distinguishable K substates (K_E and K_L).[46] However, the atomistic details, that produce these different vibrational and visible spectra are not well understood.

1.1.3 L state

To understand the first proton transfer step from the Schiff base to the carboxylic group of Asp85, it is essential to understand the structural features of the L intermediate. Despite the many experimental and theoretical studies, which were performed to reach that goal, it is still under debate what the L state looks like at the atomistic scale. The controversy concerning the structure of the binding pocket is due to the five X-ray structures which propose conflicting details (Fig. 1.5), especially concerning the retinal configuration and neighboring water molecules. In 2000 the first crystal structure, cryotrapped at 170 K, was published (pdb code: 1e0p).[47] It was criticized for possible contamination with K and M states.[48--50] Thus, a new model for L was proposed based on a crystal trapped at 150 K (pdb code: 1vjm).[51] In these two structures the planar retinal has a Schiff base with a cytoplasmic orientation and w402 is absent. The vacancy was interpreted as a coupled movement of w402 and Arg82 towards the extracellular side. Kouyama *et al.* proposed a further structure for the L state (pdb code: 1ucq)[52] which indicated a cytoplasmic oriented Schiff base with a water molecule as a hydrogen bonding partner (termed water B

in this thesis). The idea was raised, that w402 is translocated from the extracellular to the cytoplasmic side during the K→L transition and is coupled to the change of the Schiff base orientation.[52] However this water movement was challenged by theoretical work pointing out a rate-limiting barrier of 13-15 kcal/mol which is not consistent with the known lifetimes. In contrast to the three models proposing a cytoplasmic oriented Schiff base, two X-

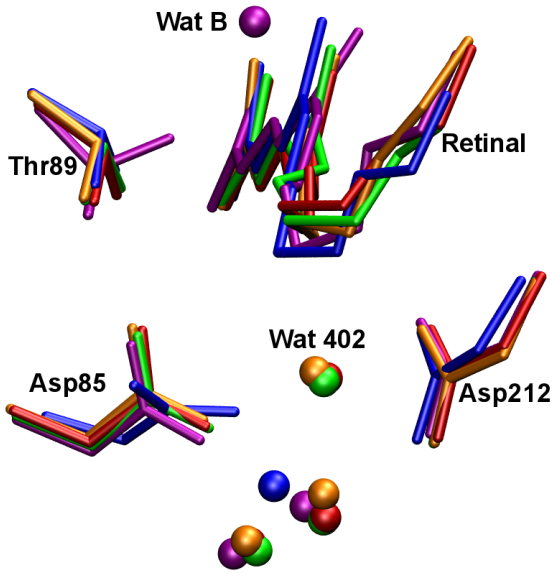


Figure 1.5 – Comparison of crystal structures proposed for the L state. The crystal structures from refs. 51 (1vjm), 52 (1ucq), 53 (1o0a), 54 (2ntw) and 13 (1c3w) are shown in blue, purple, orange, green and red, respectively. For simplicity, the retinal chain is shown without its β-ionone ring. Hydrogen atoms are not depicted.

ray structures indicting an extracellular orientation of the Schiff base were published (pdb code:1o0a[53] and 2ntw[54]). Moreover, w402 is similarly located like in the bR resting state (Fig. 1.5), which led to the idea that w402 could act as a proton mediator during the L→M transition. Nevertheless, the proposed retinal geometries are difficult to correlate with other experiments[55] and computational observations.[56] The model of ref. 53 shows a C14-C15=N-CE dihedral angle of 43.7°, i.e the retinal is 15-*syn*, which is not consistent with several spectroscopic experiments deducing a 15-*anti* geometry.[11,28,57] The chromophore of ref. 54 has a C12-C13=C14-C15 dihedral angle of -107.9°, which means that the retinal has rather an all-*trans* than a 13-*cis* geometry. Beside the problems of different crystal structures which are cryotrapped at low temperatures, the comprehension of the pre-proton transfer structure is further complicated by time-resolved FTIR studies[58,59] indicating different vibrational bands at low temperature compared to room temperature. These differences were assigned to differences in the chromophore geometry and position of water molecules in the vicinity of the Schiff base. The 15-H HOOP (hydrogen-out-of-plane) mode suggests that the retinal is twisted around the C14-C15 bond at room temperature, while its absence at low temperatures indicates a planar retinal.[58,59]

1.2 Light-switchable ionotropic glutamate receptors

In contrast to the rhodopsin family, glutamate receptors are not sensitive to light, rather they respond to binding of glutamate into their binding site. Glutamate is the most important excitatory neurotransmitter in mammalian central nervous system and about half of all synapses release it. The corresponding receptors are very important for clinical neurology, because they are connected to neuronal activity like learning and memory.[60,61] Moreover, glutamate receptors are connected to several neurological disorders, e.g. epilepsy, schizophrenia, Alzheimer's and Parkinson's disease.[62,63] The family of glutamate receptors can be divided into two groups: (i) the ionotropic receptors, which are also termed ligand-gated ion channels, i.e. receptor and ion channel are combined in one protein and (ii) metabotropic or G-coupled receptors, which modulates ion channels by enzymes and second messengers. The former group can be subdivided into three groups which were named after their selective agonists (Fig. 1.6):[64] (i) AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid); (ii) kainate ((2S-3S,4S)-3-(carboxymethyl)-4-prop-1-en-2-yl-pyrrolidine-2-carboxylic acid) and (iii) NMDA (N-methyl-D-aspartate). Each subgroup con-

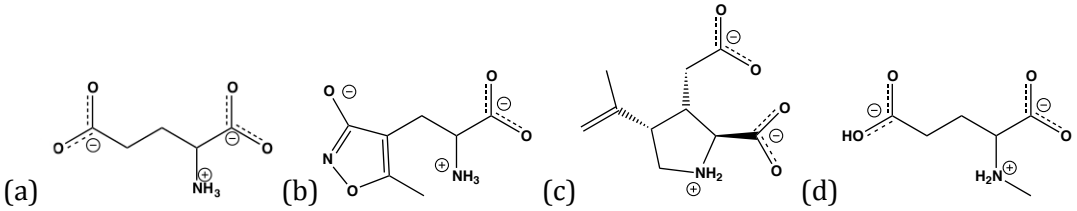


Figure 1.6 – Chemical structures of glutamate (a) and selective agonists:(b) AMPA; (c) kainate and (d) NMDA

sists of nonselective cation channels, permeable to Na^+ and K^+ and with different degrees to Ca^{2+}.[65] Furthermore, the different subgroups show differing action potentials. While AMPA receptors produce a fast increase and decrease of the electrical signal, NMDA receptors show a broad electrical signal with moderate gradients. In general, glutamate receptors are built out of four units which are thought to be dimer of dimers.[66--68] In Table 1.1 the different subunits of the families are shown. The assembled receptors can be of heteromeric or homomeric nature. Each monomer is composed of a ligand binding domain

Table 1.1 – The diversity of subunits to form functional ionotropic glutamate receptors

Receptor	AMPA	Kainate	NMDA
	GluR1	GluR5	NR1
	GluR2	GluR6	NR2a
	GluR3	GluR7	NR2b
	GluR4	KA1	NR2c
		KA2	NR2d

(LBD), that is interrupted by insertion of two transmembrane helices building the ion channel in the full receptor. Additionally, the monomer consists of an extracellular N-terminal domain and a third transmembrane segment followed by a cytoplasmic C-terminal domain (a schematic picture of a monomer is shown in Fig. 1.7). The N-terminus is often termed

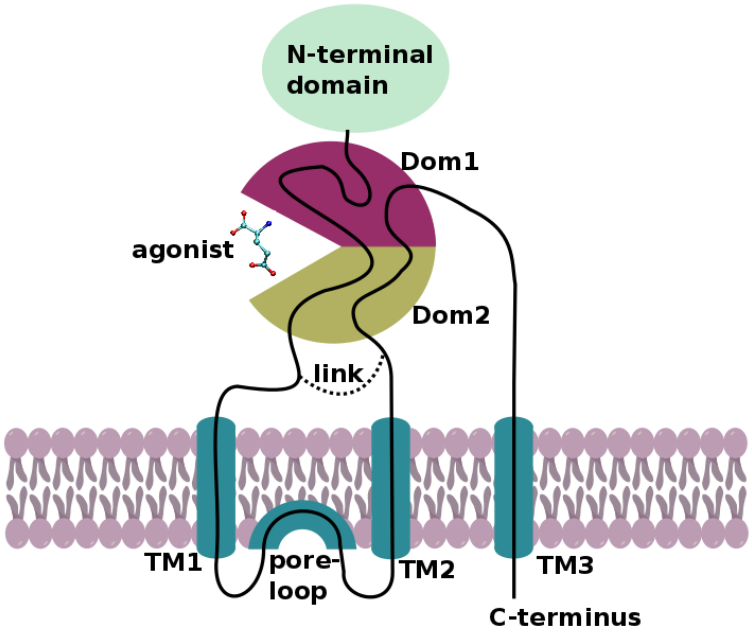


Figure 1.7 – Schematic representation of an iGluR monomer

the X-terminus, because of its unclear function.[69] It is suspected to play a role in receptor assembly, allosteric modulation of the channel and binding of an additional ligand. Nevertheless, experiments with GluR4 with a deleted N-terminal domain, showed no changes in function and assembly.[70] The LBD is the best investigated domain of the receptor, since it could be expressed in *Escherichia coli* as a soluble protein[71] and several X-ray structures of the LBD with different ligands were published in the following.[72--74] Based on these structures the first ideas about the binding mechanism and the connected ion channel opening were proposed. The C-terminus at the intracellular side is an interaction site for different types of proteins and is involved in synaptic plasticity[75] and receptor trafficking.[76]

1.2.1 The ligand binding domain of iGluR2

Due to the available crystal structures with different ligands ranging from agonists over partial agonists to antagonists and the APO state, a mechanism was proposed.[77] This was termed the venus flytrap mechanism. The models derived from the crystals showed a closing of the LBD upon ligand binding, that is supposed to lead to an activation of the channel. The closing seems to depend on the efficacy of the ligand,[72] e.g. the binding of a partial agonist leads to an imperfect closure, i.e. a structure that is half-way between the APO state and structures with an agonist (Fig. 1.8a).[73,78] The LBDs dimerize back to back. Thus, the

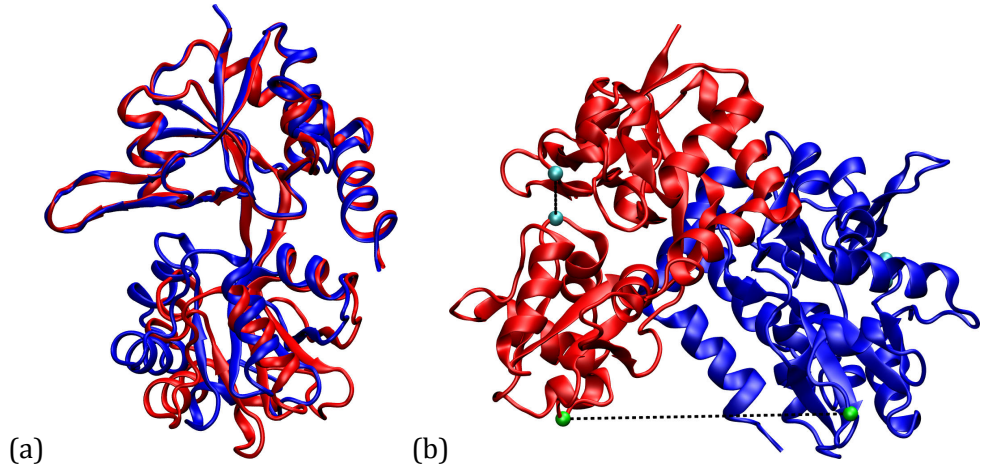


Figure 1.8 – (a) Comparison of the closure between AMPA and APO depicting the flytrap mechanism; (b) Dimer of the LBD. The introduced linker peptides replacing the membrane helices are depicted in green

distance of the linker peptides, which substitute the first two trans membrane helices of the full receptor unit, is increased by the closing of the LBD monomers (Fig. 1.8b). The increase is thought to be coupled with the channel activation by moving the transmembrane helices apart. During the last years, several experiments showed that this model is oversimplified[79--82] and that fully closed LDB can be compatible with partial agonism. Moreover, beside the clamshell motion an orthogonal twisting motion of domain 2 can be involved in the closure of the LBD,[82] which was supported by a principal component analysis of published crystal structures where the twisting motion was identified to be the second eigenvector.[83] Additional studies showed that due to the L650T mutation of iGluR2 AMPA turns from an agonist to a partial agonist, although crystal structure[84] and fluorescence resonance energy transfer[85] studies predict a full closure of the LBD. These contradicting results show that the mechanism of this ligand-gated ion channel opening is not understood to full extent.

1.2.2 Optogenetics meets Glutamate receptors

Beside the intriguing goal to understand the mechanism of glutamate gated ion channels, iGluRs have come into the focus of optogenetics. This relatively new research field focuses on the possibility to use light-gated ion channels, like bR or channelrhodopsins, for controlling defined events at a very specific place in an intact system with a very high time resolution. This means that one can control one specific cell type leaving the rest of the system unaffected. This goal was achieved first time in 2005 by Boyden *et al.*,[86] where channelrhodopsin-2 was introduced into hippocampal neurons, which could be switched with a millisecond precision. Encouraged by this result several successful experiments followed.[87--89] Up to now, optogenetics have been successfully introduced into several living organisms and animals, like *Caenorhabditis elegans*,[87,90] flies,[91,92] zebrafish[93,94] and

rats.[95--97] For reviews, see refs. 98--100.

A different approach to control function by light is to use light-sensitive chemical compounds which interact with the proteins. With this achievement one can bias receptors and channel, that are not light-sensitive in their native form. The oldest variant are caged ligands (Fig. 1.9 top)[101] where a ligand is attached to a protecting group, which hinders the ideal binding of the ligand. This "cage" can be split by light and the ligand receives its original functionality. Though, this process has the disadvantage of irreversibility. To overcome this drawback, one can use photoswitchable attachments which change their configuration reversibly due to photoexcitation. Depending on the configuration, the efficiency of the ligand is altered, e.g. from agonist to an antagonist (Fig. 1.9 mid). These kind of compounds are called photochromic ligands (PCL). Beside the advantages, like ease of application and fast distribution in tissues, their application can be hindered by the low selectivity between receptor subgroups, e.g. AMPA and NMDA receptors. In this case, a

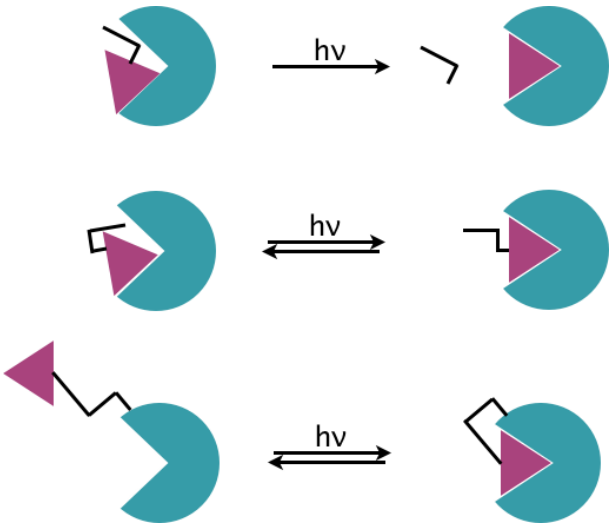


Figure 1.9 – Schematic representation of the different optogenetical approaches. Top: caged ligand, which splits off the blocker by light; mid: a PCL with a photoswitchable attachment; bottom: a PTL, which is covalently attached to the protein surface

further possibility are the photoswitched tethered ligands (PTL) which are covalently attached on the protein surface (Fig. 1.9 bottom). The big advantage of this approach is that the PTLs can be genetically encoded, because they are attached to an engineered cysteine residue by bioconjungation. Therefore, one can actuate one specific receptor within a system and leave the wildtype forms of the receptor unaffected.

The most commonly used PCLs and PTLs are based on azobenzene, because of the well defined geometries of both configurations. It changes from a long and "flat" geometry in *trans* to a bulky geometry in *cis* (Fig. 1.10). The isomerization is very fast and occurs on the femtosecond time scale. Moreover, azobenzene has a high quantum yield. Hence, only low light intensity is needed and the absorption maximum can be tuned by substituents at the benzene rings. These azobenzene-based ligands have been successfully employed

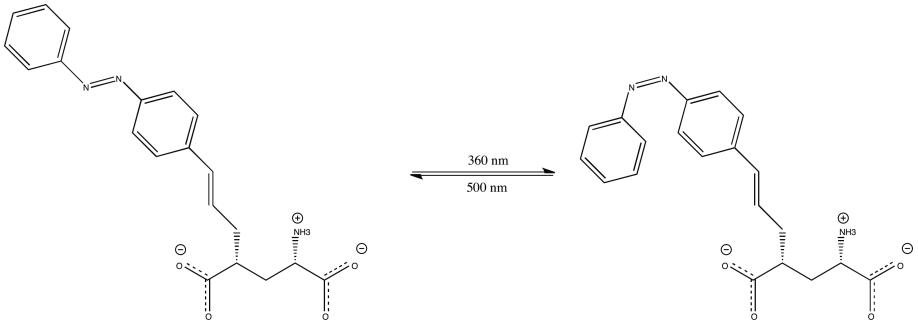


Figure 1.10 – An example for a PCL (4-GluAzo),[102] which shows activity for kainate receptors

in various systems: Voltage-gated potassium channels, that are blocked and unblocked by a PTL[103] or PCL[104]; kainate receptors, which can be switched by a PTL[105] or PCL.[102,106] Despite the huge success of these ligands, it is not known how these photoswitches work in detail: What is the response of the system, i.e. the target protein, after the photoisomerization? What are the energetics? By answering these questions, it would be possible to enhance the properties of these ligands in a more systematic way.

Chapter 2

Theory

2.1 Introduction

Peptides and proteins play a crucial role in most biological processes. The understanding of structural and thermodynamic features of these molecules is the key to their function and role in these processes. Despite the enhancements in experimental techniques like X-ray diffraction experiments and several spectroscopic techniques over the last decades, it is still not possible to monitor the motion of molecules at the atomistic scale. Computer simulations have become a powerful tool to close the gap between experiment and the atomistic resolution.[107--110] One of these simulation tools is the molecular dynamics (MD), which is a numerical protocol to calculate trajectories of biomolecules. From those simulations one can derive properties of the system that are beyond the feasibilities of the experiments. Due to the size and complexity of biomolecules and the time scale that has to be reached, the motion has to be derived from a classical potential, which is described by a classical force field. By solving iteratively Newton's equations of motion trajectories can be generated. Unfortunately, the classical description of the electronic potential has some drawbacks which can arise the need of quantum mechanics, e.g. chemical reactions, charge transfer, electronic excitation.

2.2 Molecular Mechanics/Molecular Dynamics

The basic assumption doing molecular dynamics is that the nuclei behave like classical particles. Therefore, their motion can be describe by Newton's equations of motion:

$$F_n = m_n \cdot a_n = -\nabla_{x_n} V(x_1, x_2, x_3, ..., x_N) \tag{2.1}$$

where F_n is the force acting on nucleus n, m_n and a_n are its mass and acceleration. V is the total potential energy, which is a function of all nuclei coordinates.

Provided the forces can be calculated, one can numerical integrate Newton's equations of motion for the whole system with a small integration time step Δt. The most common

algorithm for this integration is the Verlet leapfrog algorithm:[111]

$$x_n(t_{i+1}) = x_n(t_i) + v_n(t_{i+\frac{1}{2}}) \tag{2.2}$$

Therefor one needs the starting positions $x(t_0)$, which can be obtained by X-ray or NMR structures of the system, and the velocities v_n of every nucleus at half-integer time step.

$$v_n(t_{i+\frac{1}{2}}) = v_n(t_{i-\frac{1}{2}}) + a_n(t_i)\Delta t \tag{2.3}$$

The main task for calculating the trajectory is to derive the potential energy function V_n. Because of the large number of atoms in biomolecular simulations it is quite demanding to computed this energy function. Thus, one has to use approximations for the potential energy of the system. There are two main approaches.

One approach is to describe the potential energy by an interaction-based function. Within the so called molecular mechanics framework the electronic structure of the system is neglected. The total potential energy is formulated as a sum of analytic functions, which describe the different interactions of the atoms. These analytical functions contain empirical parameters that are derived in a way, that the total potential energy is valid.
In biomolecular force fields the interactions are divided into bonded and non-bonded interactions. The bonded interactions are subdivided into interactions between up to four particles, that are connected by bonds, angles or torsions. In most force fields bond and angles are described by a harmonic potential, while a periodic potential is used for the torsions:

$$\begin{aligned} V^b = & \sum_{i,j}^{bonds} \frac{1}{2} k_b^{ij} \left(r_{ij} - r_{ij}^0\right)^2 + \sum_{i,j,k}^{angles} \frac{1}{2} k_a^{ijk} \left(\theta_{ijk} - \theta_{ijk}^0\right)^2 \\ & + \sum_{i,j,k,l}^{torsion} \frac{1}{2} k_{tor}^{ijkl} \left(1 + cos[n_{ijkl}\phi_{ijkl} - \phi_{ijkl}^0]\right) \end{aligned} \tag{2.4}$$

The k's are the empirical derived force constants, r_{ij} is the bond length between two covalently bound atoms, θ_{ijk} is the angle defined by atom i, j and k, and ϕ_{ijkl} is the torsion angle descriminated by the atoms i, j, k and l. The equilibrium values are denoted with a superscripted 0.

The non-bonded potential describes the interaction between two atoms that are more than 3 bonds apart or that are not bonded at all. This interaction energy can be divided into two groups. The Van der Waals interactions, which reflects that atoms and molecules show an attraction at long range, but a repulsive behavior at short range. The attraction, also called dispersion, is generated by the thermally induced electron density fluctuation within an atom (or molecule). If another atom (or molecule) is close-by, the fluctuating

dipoles align and this leads to a lowering in energy. The short range repulsion is generated by the overlap of electron densities, which is energetically unfavorable.

In force fields this many-body problem is describes by a pair potential. In most cases the Lennard-Jones potential is used to mimic dispersion and repulsion:

$$V^{L-J}(x_i, x_j) = \left(\frac{C_{ij}^A}{r_{ij}}\right)^{12} - \left(\frac{C_{ij}^B}{r_{ij}}\right)^{6} \tag{2.5}$$

where r_{ij} is the distance between atom i and j and C_{ij}^A and C_{ij}^B are the repulsion and attraction coefficients, respectively.

The second non-bonded contribution considers the interaction between charged atoms. This has a electrostatic character and is computed by Coulomb's law:

$$V^{Coul}(x_i, x_j) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.6}$$

where the q is a point charge, ϵ_0 is the dielectric constant and r_{ij} is the distance between the charges.

A different way to obtain the potential energy acting on the nuclei is based on quantum mechanics (QM), which can describe the electronic structure of the system explicitly. Therefore, one has to construct the electronic Hamiltonian based on the nuclear configuration and to calculate the electronic wave function. By calculating the expectation value of the Hamiltonian with respect to the electronic wave function one can obtain the total potential energy:

$$V(x_1, x_2, x_3, ..., x_N) = \int \Psi_e^* \hat{H}(x_1, x_2, x_3, ..., x_N)\Psi_e d\tau \tag{2.7}$$

In the next section the evaluation of the electronic wave function is discussed in more detail.

2.3 Quantum Chemistry Methods

2.3.1 Hartree-Fock

Hartree-Fock (HF) is the enhancement of the Hartree approach, which describes the electronic wave function Ψ as a product of one-particle wave functions:

$$\Psi(r_1, ...r_N) = \phi_1(r_1) \cdot \phi_2(r_2)...\phi_N(r_N) \tag{2.8}$$

which describes the electron as non-interacting particles. In HF the Pauli exclusion principle is included by using antisymmetric wave functions. Thus, the N-electron wave function can be written with the help of the Slater-determinant:

$$\Psi(r_1, ...r_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & ... & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & ... & \chi_N(2) \\ ... & ... & ... & ... \\ \chi_1(N) & \chi_2(N) & ... & \chi_N(N) \end{vmatrix} \tag{2.9}$$

where χ is a spin orbital, which is a spatial orbital ϕ multiplies with a spinfuction α or β

$$\chi(1) = \phi(1)\alpha(1) \tag{2.10}$$

Spatial orbital, which are also termed molecular orbitals (MO), are built by a linear combination of atomic orbitals η (LCAO).

$$\phi_i = \sum_j = c_j^i \eta_j \tag{2.11}$$

The total Hartree-Fock energy is given by:

$$E = \sum_i h_i + \frac{1}{2}\sum_{i,j}(J_{ij} - K_{ij}) + V_{nn} \tag{2.12}$$

where h_i is the one-particle energy; V_{nn} is the nucleus interaction; J_{ij} is the classical coulomb repulsion between two electron densities:

$$J_{ij} = \int |\chi_i(i)|^2 r_{ij}^{-1} |\chi_j(j)|^2 dr_i dr_j d\omega_i d\omega_j \tag{2.13}$$

J_{ij} is reduced by K_{ij} representing the fermi hole, i.e. the effect that two electrons with the same spin cannot be at the same spot.

$$K_{ij} = \int \chi_i(i)\chi_j(i) r_{ij}^{-1} \chi_i(j)\chi_j(j) dr_i dr_j d\omega_i d\omega_j. \tag{2.14}$$

Therefore J_{ij} will only be reduced, if the electrons have the same spin. The electron self-interaction J_{ii} is neglected in HF, because of $J_{ii} = K_{ii}$. For a more detailed view on the Hartree-Fock approach, see textbooks on quantum chemistry, e.g. Szabo and Ostlund[112] or Jensen.[113]

DFT

The other main approach in quantum chemistry to describe the electronic structure is the density function theory (DFT). DFT is based on two theorems by Hohenberg and Kohn.[114] The first theorem states that the ground state properties of a many-electron system are determined by an electronic density $\rho(r)$. The second theorem verifies that one gets the lowest energy $E[\rho(r)]$ of the system when the density is the ground state density. Due to the Born-Oppenheimer approximation[115] one describes the nuclei as classical particles and the electrons moving within their potential. Thus, the DFT total energy can be written as:

$$E^{DFT}[\rho(r)] = T[\rho(r)] + E^{ne}[\rho(r)] + E^J[\rho(r)] + \tilde{E}^{xc}[\rho(r)] + E^{nn} \tag{2.15}$$

where T is the kinetic energy of the electrons, E^{ne} the nucleus-electron interaction energy, E^J is the classical part of the electron-electron interaction, $\tilde{E}^{xc}$ is the exchange correlation energy. The exact functionals for the kinetic energy and the exchange correlation is not

known. Kohn and Sham[116] introduced the idea to approximate the kinetic energy by Kohn-Sham orbitals $\Psi_i(r)$ and the separation of the Schrödinger equation into a set of equations describing an imaginary system of non-interacting electrons.

$$\left[-\frac{\nabla^2}{2} + V(r)\right]\Psi_i(r) = \epsilon_i \Psi_i(r) \tag{2.16}$$

Thus, a kinetic energy for a system with non-interacting electrons T^s can be written as:

$$T^s[\rho(r)] = \sum_i n_i \epsilon_i - \int V(r)\rho(r)dr \tag{2.17}$$

Since this leads to a formally not exact theory, the missing energy $(T - T^s)$ is added to the exchange correlation energy.

$$E^{xc} = E^{xc} + (T - T^s) \tag{2.18}$$

The problematic search for an exchange correlation functional remains and several approximative approaches have been developed. Widely used are the following approximations:

- local density approximation (LDA), where the exchange correlation is only depended on the density
- generalized gradient approximation (GGA), where the exchange correlation is additionally depended on the gradient of the density. A widely used GGA functional is PBE.[117]
- hybrid functional, where the exchange part is calculated with the Hartree-Fock formalism, e.g. PBE0[118] or B3LYP.[119]

2.3.2 DFTB

Density functional tight binding (DFTB) is an approximative method that is derived from DFT. DFTB is based on a minimal basis set, i.e. only the valence electrons are described explicitly, while the core electrons are treated by two-center potentials. Crystal field and three-center integrals are neglected to further reduce the computational cost. The two-center Hamiltonian and the overlap matrix is pre-calculated and tabulated for a grid of distances. The remaining energy is approximated and therefore the time consuming step is the diagonalization of the Hamilton matrix.

The starting point of DFTB is to calculate the reference density ρ^0 from a superposition of pre-calculated neutral atomic densities.

$$\rho^0 = \sum_a \rho_a \tag{2.19}$$

The reference density deviates from the DFT density by $\Delta\rho(r)$.

$$\rho(r) = \sum_i n_i \Psi_i^*(r)\Psi_i(r) = \rho^0(r) + \Delta\rho(r) \tag{2.20}$$

The DFT total energy 2.15 can the be rewritten:

$$E[\rho^0(r)+\Delta\rho(r)] = \sum_i n_i \int \Psi_i^*(r)\left(-\frac{\nabla}{2} + V^{ne} + \int \frac{\rho^{0\prime}+\Delta\rho'}{|r-r'|}dr' + V^{xc}[\rho^0+\Delta\rho]\right)\Psi_i(r)dr$$
$$- \frac{1}{2}\int\int \frac{(\rho^{0\prime}+\Delta\rho')(\rho^0+\Delta p)}{|r-r'|}drdr' - \int V^{xc}[\rho^0+\Delta\rho](\rho^0+\Delta\rho)dr$$
$$+ E^{xc}[\rho^0+\Delta\rho] + E^{nn} \tag{2.21}$$

The density fluctuation $\Delta\rho$ of the classical electron-electron interaction and the exchange correlation is neglected from the integral in eq. 2.21 and added again in an additional term using eq. 2.20. In addition, the exchange correlation energy is expanded in a Taylor series. This leads to a new formulation of the DFT energy:

$$E[\rho^0(r)+\Delta\rho(r)] = \sum_i n_i \int \Psi_i^*(r)\left(-\frac{\nabla}{2} + V^{ne} + \int \frac{\rho^{0\prime}}{|r-r'|}dr' + V^{xc}[\rho^0]\right)\Psi_i(r)dr$$
$$- \frac{1}{2}\int\int \frac{\rho^0\rho^{0\prime}}{|r-r'|}dr'dr - \int V^{xc}[\rho^0]\rho^0 + E^{xc}[\rho^0] + E^{nn}$$
$$+ \frac{1}{2}\int\int\left(\frac{1}{|r-r'|} + \left.\frac{\partial^2 E^{xc}[\rho]}{\partial\rho\partial\rho'}\right|_{\rho^0,\rho^{0\prime}}\right)\Delta\rho\Delta\rho' dr'dr$$
$$+ \frac{1}{6}\int\int\int \left.\frac{\partial^3 E^{xc}[\rho]}{\partial\rho\partial\rho'\partial\rho''}\right|_{\rho^0,\rho^{0\prime},\rho^{0\prime\prime}} \Delta\rho\Delta\rho'\Delta\rho'' dr''dr'dr + \ldots \tag{2.22}$$

By truncation of the Taylor series different levels of sophistication can be achieved. Standard DFTB neglects density fluctuation and only the first two lines of eq. 2.22 are considered. DFTB2 approximates the second order terms. Consequently, the newly developed DFTB3 considers also the third order terms for the density fluctuation.

The term in the first line of eq. 2.22 correspondS to the energy of the reference density. In the second line DFT double counting contributions, the interaction between nuclei and the exchange correlation are included. These three terms are summarized approximatively by a repulsive energy E^{rep}:

$$E^{rep} = \frac{1}{2}\sum_{ab} V_{ab}^{rep}[\rho_a^0, \rho_b^0, r_{ab}] \tag{2.23}$$

where V_{ab}^{rep}s are fitted to atom types and distances using several reference systems. Thus, the total energy of DFTB can be written as:

$$E[\rho(r)] = \sum_i \epsilon_i + E^{rep} \tag{2.24}$$

This early truncation of the taylor series is sufficient for describing unpolar systems with no charge transfer between the atoms, e.g. calculation of solids. For systems with heterogeneous electron density distribution one has to include the density fluctuations $\Delta\rho$. This is done in DFTB2 [120] including the terms of the third line of eq 2.22.

$$E^{2nd} = \frac{1}{2}\int\int\left(\frac{1}{|r-r'|} + \left.\frac{\partial^2 E^{xc}[\rho]}{\partial\rho\partial\rho'}\right|_{\rho^0,\rho^{0\prime}}\right)\Delta\rho\Delta\rho' dr' dr \tag{2.25}$$

The density fluctuation is also approximated as atomic contributions:

$$\Delta\rho = \sum_a \Delta\rho_a, \tag{2.26}$$

which neglects multipole interactions. Introducing an analytical function γ, [120] which corresponds to the interaction between the charge fluctuations and further approximations within the second order terms, one can rewrite the second order energy as:

$$E^{2nd} = \frac{1}{2}\sum_{ab}\Delta q_a \Delta q_b \gamma_{ab} \tag{2.27}$$

The γ-function has two limiting cases: (i) For large distances r_{ab} the function is reduced to a coulomb interaction between the two partial charges. (ii) If $a = b$, γ_{ab} describes the on-site self-repulsion, i.e. the electron-electron interactions on atom a

$$\gamma_{aa} = U_a, \tag{2.28}$$

where U_a is the Hubbard parameter, which describes the chemical hardness. For a detail view on γ see ref. 120.

The third order terms (fourth line of 2.22) are included in DFTB3. [121] These are getting important for system with highly localized charges, because the Hubbard in DFTB2 is not dependent on the charge fluctuation of the atom, which is not correct, i.e. the Hubbard differs for a neutral and a charged atom. This is reflected in the Hubbard derivative.

$$U_a^d = \frac{\partial^3 E^{atom}}{\partial q_a^3} \tag{2.29}$$

The DFTB3 total energy is given as:

$$E[\rho(r)] = \sum_i \epsilon_i + E^{rep} + \frac{1}{2}\sum_{ab}\Delta q_a \Delta q_b \gamma_{ab} + \frac{1}{6}U_a^d \Delta q_a^3 \tag{2.30}$$

The Hubbard derivatives are fitted to enhance the description of proton affinities and hydrogen-bonding energies. [122] For a further insight into DFTB3, see review in ref. 123.

DFTB performance

DFTB is 2-3 magnitudes faster than DFT with a medium-sized basis set. Tests showed, that heats of formation, molecular geometries, and other properties are as exact[124,125] as other semi-empirical methods such as PM6[126] or OM2.[127] Especially, the structural ground state properties of protonated Schiff base models are comparable to full DFT including structure, bond length alternation and torsional barriers[128--130] and also the geometry of the binding pocket in bR.[130]

2.3.3 Post-HF Methods

The Hartree Fock approach weakens, if electron correlation plays a role which is only described partly in HF. The correlation between electrons with different spins is missing. To include this effect one can use more advanced methods, which are based on Hartree-Fock, to get the exact energy. The difference between the exact energy and the Hartree-Fock energy is often called correlation energy:

$$E^{corr} = E^{exact} - E^{HF} \tag{2.31}$$

Perturbation Theory

One way to achieve this is to use the perturbation theory, where the Hamilton operator $\hat{H}$ is rewritten by a Hamilton operator describing the unperturbed system $\hat{H}_0$, and one describing the perturbation $\hat{V}$.

$$\hat{H}(\lambda) = \hat{H}_0 + \lambda\hat{V} \tag{2.32}$$

By introducing λ, which can be thought of as the strength of the coupling, the Schrödinger equation can be rewritten

$$\hat{H}(\lambda)\Psi(\lambda) = E(\lambda)\Psi(\lambda) \tag{2.33}$$

and one can expand the energy and according wave function into a series.

$$E(\lambda) = E^{(0)} + \sum_k \lambda^k E^{(K)} \tag{2.34}$$

$$\Psi(\lambda) = \Psi^{(0)} + \sum_k \lambda^k \Psi^{(K)} \tag{2.35}$$

Inserting equations 2.34 and 2.35 into equation 2.33, and sorting terms with the same exponent, leads to a set of equations that can be solved up to the wished magnitude of precision. In the Møller-Plesset perturbation theory the unperturbed hamiltonian is given by the one-particle hamiltonian of HF, while the perturbation is the missing electron correlation. The HF ground state energy is achieved by considering the first order energy term. The energy in second order is given by:

$$E_0^{(2)} = \sum_{n\neq 0} \frac{|\langle\Psi_0^0|\hat{V}|\Psi_n^0\rangle|^2}{E_0^0 - E_n^0}. \tag{2.36}$$

The second order includes the interaction from the groundstate with all excited determinants Ψ_n^0.

Configuration Interaction

The general idea of Configurational Interaction (CI) is that the multi-electron wave function can be built up with the superposition of ground state and excited state determinants. Using a complete orthonormal one-electron basis, like the Hartree-Fock orbitals ϕ_i, one can built the exact N-electron wave function by the sum of all determinants.

$$\Psi(1...N) = a_0\Phi_{HF} + \sum_{r,a} a_a^r \left|\Phi_a^r\right\rangle + \sum_{a<b,r<s} a_{ab}^{rs} \left|\Psi_{ab}^{rs}\right\rangle + \sum_{a<b<c,r<s<t} a_{abc}^{rst} \left|\Psi_{abc}^{rst}\right\rangle + ... \quad (2.37)$$

which includes all electronic excitations from the occupied (a,b,c) into the unoccupied (r,s,t) orbitals. Taking all determinants into account leads to Full-CI and gives the exact energy and equation 2.31 can be rewritten:

$$E^{corr} = E[\Psi(1...N)] - E^{HF} \quad (2.38)$$

Each of these determinants describes exact one electron configuration, e.g. ground state, single or double excitation. Since Full CI is not applicable for most of the interesting systems, one has to choose how many excitation to consider. Taking only the single excitations into account (CI singles) does not improve the ground state, because of Brillouin's theorem. But they interact with the double excitations, and improve their description. The double excitations (CI doubles) improve the description of the ground state. The combination of CI singles and doubles (CISD) describes most (80-90%) of the missing correlation energy for small molecules in the ground state. This decreases for bigger systems, because of the size inconsistency of CI.[112] The main task for the calculation of excitation energies is a balanced description of the correlation energy for both considered states. For example, CISD would calculate a sufficient correlation energy for the ground state but not for the first excited state. Therefore, one has to consider higher terms, which are getting computationally expensive for bigger systems, or use the Multi-reference CI (MRCI) approach. To get a balanced description of the correlation energy for the ground state and the excited state more than one reference configuration is needed. For all additionally considered reference configurations single and double excitation are taken into account, which leads to a balanced description of the states. The choice of the considered references is done by hand or automatically.

2.3.4 SORCI

SORCI[131] (Spectroscopy Oriented Configuration Interaction) is an *ab initio* method to calculate excitation energies. It is a combination of MRCI and perturbation theory. The computational procedure and the used parameters are describe briefly in the following. In the first step Hartree-Fock is used to generate improved virtual orbitals.[132] Subsequent, a reference space is defined by a restricted active space (RAS) or complete active space (CAS). Based on this reference space a CI matrix is generated and diagonalized. Configurations with a weight below $T_{prediag}$ are neglected and the remaining configuration state functions

(CSFs) $\Phi_\mu$built the reference space S, which is re-diagonalized and one gets the energy eigenvalues and the multi-configuration wave function of zeroth order.

$$\Psi_{(I)}^{(0)} = \sum_{\mu \in S} c_{\mu I}^{(0)} \Phi_\mu \tag{2.39}$$

On the basis of the reference space S, all single and double excitation, that are in the DDCI3 space,[133] are included into the CI space R. To reduce computational cost of this too large space, the number of configurations are reduced by a MP2 calculation with respect to $\Psi_{(I)}^{(0)}$ and threshold T_{sel}.

$$\frac{|\langle \Psi_I^{(0)} | \hat{H} | \Phi_\mu \rangle|^2}{\langle \Psi_I^{(0)} | \hat{H}_0 | \Psi_I^{(0)} \rangle - \langle \Phi_\mu | \hat{H}_0 | \Phi_\mu \rangle} \geq T_{Sel} \tag{2.40}$$

Thus, the CI space is divided into a strongly perturbing subspace R' and a weakly perturbing subspace R''. The CI space which consists out of the reference configurations S and R' is diagonalized and give an improved wave function $\Psi_I^{(a)}$ and energy $E_I^{(a)}$. The energy contribution of the unselected space R'' is described on the MP2 level. The resulting energy of the states I,II,... are given as

$$E_{(I)} = E_{(I)}^{(a)} + \sum_{\mu \in R''} E_{(I,\mu)}^{unsel)} + E_{(I)}^{(corr)}, \tag{2.41}$$

where $E_{(I)}^{(\mathrm{corr})}$ is a correction of the Davidson type,[134,135] which reduces the size inconsistency error of truncated CI methods. To reduce the impact of quality and choice of the orbitals in the first step, SORCI works iteratively and built approximate average natural orbitals, that are used in the next iteration step.

Paramter

SORCI was already used for calculating excitation energies of the retinal in rhodopsins. Therefor, the following parameters and thresholds were used. The SV(P)[136] basis set was used with diffuse s- and p-functions on the carboxylic oxygen atoms of the anionic residues. For generation of starting reference space a CAS(4,4) was used after an initial RHF calculation generated improved virtual orbitals. $T_{prediag}$ is set to 10^{-3}. The selection threshold T_{sel} is set to a value of $10^{-6}(E_h)$. This parameters were extensively benchmarked for retinal models by Wanko *et al.*.[129]

Polarization model

The excitation energy of the retinal is strongly influenced by the surrounding binding pocket. Since, it is not possible to include the whole protein into the expensive SORCI calculation, the QM/MM approach is used. In the standard QM/MM approach (see section 2.4) the protein is described as point charges, which polarize the electron density of the QM region, while the point charges are fixed. This leads to a blue shift in excitation energy, due to missing polarization of the protein produced by electron density fluctuations in the QM

region. To take this polarization into account, a polarization model (polar.h) was developed by Wanko *et al.*.[137,138] The classical point charges are replaced by ESP-fitted charges, since the standard CHARMM point charges are pre-polarized to mimic the polarization of other amino acids. In a second step, the newly derived point charges are split into a dipole, which can freely orient and change its moment according to the field. Thus, the dipoles are optimized according to the charge distribution within the QM region and the other dipoles. The ground state S_0 and the first excited state S_1 are calculated in two calculations within the field of the polarized external dipoles. Afterwards the dipoles are polarized by the newly derived charge distribution in the S_0 and S_1. This procedure is iterated until the excitation energy is converged.

Validation for retinal proteins

In combination with SCC-DFTB/MM optimized geometries, it could be demonstrated that SORCI and polar.h are capable to calculate vertical excitation energies comparable to the experiment. In case of the bR ground state this leads to a perfect agreement with the experimental value (2.18 eV). This perfect match is definitely by accident and relies on an error cancellation, since no existing excited state method which is capable of describing such large molecules like the retinal shows an intrinsic accuracy better than 0.1 eV. SCC-DFTB underestimates the bond length alternation (BLA) leading to excitation energies in the gas phase that are about 0.1 eV too low compared to experiment. The neglect of dispersion, on the other hand, leads to excitation energies which are too high. Thus, both effect, the neglect of dispersion and the BLA being underestimated by SCC-DFTB compensate each other. Nevertheless, error cancellation is not unusual in computational chemistry and is very often exploited in applications. This is justifiable as long as the error cancellation is systematic. For application on retinal proteins it means that the changes in retinal geometry and protein environment lead to the correct shifts in excitation energy. In a series of publications it was shown that our setup produce reliable shifts,[129,137--141] e.g. shifts due to mutations in Rhodopsin (Rh) could be identified correctly.[139]

2.4 QM/MM

In most cases biological system were described by classical force fields, as described in section 2.2. This approach reaches its limit when the consideration of the electronic structure and their changes is important, e.g. bond breaking and formation, charge transfer/transport, electronic excitation and polarization. To avoid the computational demanding calculation of the whole system at the quantum mechanical level, Warshel and Levitt developed the mixed quantum mechanic/ molecular mechanic (QM/MM) approach.[142] The system is divided into two subsystems, which are described at different levels of accuracy.

The interaction between these two subsystem is the major issue of this approach and has been solved in different ways.

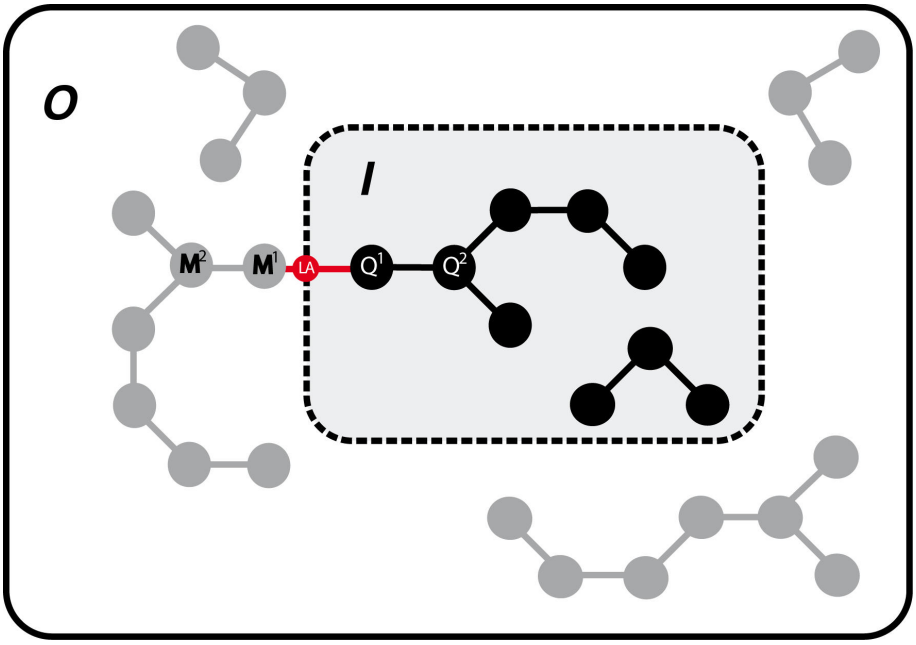


Figure 2.1 – A schematic representation of a system described by the QM/MM approach. **O** describes the subsystem, which is described on the classical level. **I** includes the region, that is described by quantum mechanics

2.4.1 QM/MM energy

Subtractive scheme

A relatively simple approach is the subtractive method, which includes three energy calculation: (i) a MM calculation of the whole system **S**, (ii) a QM calculation of the inner subsystem **I** and (iii) a MM calculation of the same subsystem **I**. The resulting total energy is build by the sum of the first two terms. The double counting of the inner subsystem is neglected by subtracting the third term.

$$E^{\mathbf{S}}_{QM/MM} = E^{\mathbf{S}}_{MM} + E^{\mathbf{I}}_{QM} - E^{\mathbf{I}}_{MM} \tag{2.42}$$

The advantage of this approach is the relatively easy implementation, since no special coupling between the subsystems is needed. But this is also the biggest disadvantage, because the interaction between the subsystems is described at the classical level. In case of the electrostatic interaction this leads to wrong results, if the charges change during a reaction, because the electrostatic interactions are still distinguished based on the not-updated point charges. A second major drawback is, that the quantum mechanical calculation is performed without any interaction with the outer subsystem, i.e. the electron density is not polarized in any way.

Additive scheme

A more sophisticated version is the additive scheme. The total energy of the system is determined by the sum of the MM energy of the outer subsystem **O**, the QM energy of the

inner subsystem **I**, and a coupling term which includes the interaction between **I** and **O**.

$$E^{\mathbf{S}}_{QM/MM} = E^{\mathbf{O}}_{MM} + E^{\mathbf{I}}_{QM} + E^{\mathbf{I,O}}_{QM-MM} \tag{2.43}$$

The coupling term includes electrostatic, van der Waals interactions and contributions from bonds which cross the border between the subsystems.

$$E^{\mathbf{I,O}}_{QM-MM} = E^{el}_{QM-MM} + E^{vdW}_{QM-MM} + E^{b}_{QM-MM} \tag{2.44}$$

Electrostatic interactions

The electrostatic interactions can be described in two fashions. The first is the mechanical embedding, which describes the interactions at the classical level. This leads to the same disadvantages occurring during the subtractive QM/MM scheme. The second method is the electrostatic embedding. Within this approach the QM calculation is performed in the presence of the point charges of the outer subsystem polarizing the electron density. This is achieved by an additional term in the QM-Hamiltonian

$$\hat{H}^{el}_{QM-MM} = -\sum_{i}^{electrons}\sum_{M}\frac{q_M}{|r_i - R_M|} + \sum_{\alpha}\sum_{M}\frac{q_M Z_\alpha}{R_\alpha - R_M|} \tag{2.45}$$

with q_M as classical point charges, Z_α are the nuclear charges of the QM atoms. The plus in accuracy has to be paid with computer time. Furthermore, the MM charges are not effected by the fluctuations of the electron density. To include this one has to use a polarizable embedding which uses a polarizable force field that can react on the changes in electron density.

Van der Waals interactions

The van der Waals interactions between the subsystems are described at the force field level. Thus the additive and the subtractive scheme have the same description with respect to the QM/MM van der Waals coupling. The introduced errors are quite small and can be neglected due to the short range nature of van der Waals interactions.[143].

Bonded Interactions

It is not always possible to divide the subsystem and avoid trans-border bonds. The corresponding bonds have to be cut in some way, which leads to three major problems: (i) The corresponding bond cannot be cut homolytically, because this would lead to unrealistic description of the QM region. The most popular approach is the introduction of a link atom (LA), which is positioned between atom Q^1 and atom M^1 (Fig. 2.1) and is covalently bound to the Q^1. Thus the QM region is saturated. To avoid the introduction of three additional degrees of freedom, the link atom is positioned on the QM-MM bond and fixed at this position.[144]

$$R_{LA}(R_{Q^1}, R_{M^1}) = R_{Q^1} + g(R_{M^1} - R_{Q^1}) \tag{2.46}$$

(ii) In the case of electrostatic embedding, the problem of overpolarization of the electron density at the QM-MM border exists. The link atom and thus the electron density is very

close to M^1. If M^1 is partially charged, the point charge can polarize the electron density in an artificial way. The simplest solution is to delete the point charges of nearby MM atoms, which can lead to severe artifacts.[145,146] This problem can be avoided by deleting the charges of the whole charge group as long as it is neutral. A more sophisticated way is to smear out the charge from M^1 to M^2. To delete the introduced dipole moment a charge pair within the vicinity of M^2 with an opposite dipole moment is placed.[147] A similar approach is the distribution of the M^1 charge on the rest the charge group (*divided frontier charge* DIV).[146] (iii) Where to cut, i.e. where leads a capping to the smallest error? First of all, no QM atoms involved in bond breaking or formation should be included in the coupling term. Due to the dihedrals, they should be at least three bonds away from the border. Furthermore, no conjugated or polarized bonds should be cut. Thus, the best bond to cut is a normal C-C single bond.

2.5 Free Energy Calculations

In an earlier section (see section 2.2) the general ideas of molecular dynamics has been explained. With this tool it is possible to investigate the time propagation of the system up to the microsecond range. While this is already suitable for many topics, one gets into trouble if free energy differences are in the focus of the investigation. In biological process it is quite often that the enthalpic effects are not crucial but entropy. For example, the prediction of free energy of ligands binding to a receptor or an enzyme is of much interest, e.g. rational drug design. In general, one can obtain the free energy difference between two states by normal molecular dynamics. If a constant pressure is provided, the Gibbs-energy is given by:

$$\Delta G = G(\zeta_1) - G(\zeta_0) = -\beta^{-1} \ln \frac{\rho(\zeta_1)}{\rho(\zeta_0)} \tag{2.47}$$

where ζ is the reaction coordinate, ζ_1 and ζ_0 are the relevant minima along the coordinate, $\beta = 1/k_b T$ and $\rho(\zeta)$ is the probability density of the conformational coordinate:

$$\rho(\zeta) = \frac{\int \delta(\zeta' - \zeta)\, \mathrm{e}^{-\beta V(x')}\, dx'}{\int \mathrm{e}^{-\beta V(x')}\, dx'} \tag{2.48}$$

The probabilty distribution $\rho(\zeta)$ represents the frequency of occurrence of the conformations corresponding to ζ in a conical ensemble relative to all other conformations. If one assumes that $\rho(\zeta)$ is normalized in such a way that $\rho(\zeta)_0 = 1$, equation 2.47 can be rewritten

$$\Delta G = -\beta^{-1} \ln \rho\zeta = W(\zeta) \tag{2.49}$$

$W(\zeta)$ is the reversible work (the change in free enthalpy) required when changing the system along reaction coordinate ζ. In general, one can obtain the probability distribution by simply counting the conformations along a standard MD. Unfortunately, the scale of many interesting processes are determined by free energy barriers and the rates to overcome these. This barriers can be up to tens of kilocalories which can lead to the microsecond

time scale. Due to the great leap in computational power these time scales are reachable on standard computer clusters. However, to obtain the correct free energy surface, the system has to reach every spot within the conformational space (at least along the reaction coordinate) several times, e.g overcome the barrier several times in both direction, to conform to the ergodicity theorem (Fig. 2.2a). Additionally, one can be trapped in a local minimum and miss the global minimum along the reaction coordinate, i.e. one gets the wrong idea of the investigated phenomena (Fig. 2.2b). To overcome this problems several

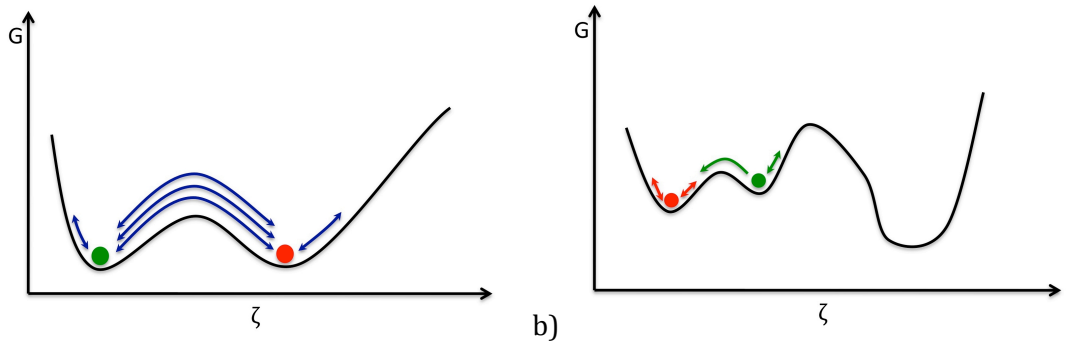


Figure 2.2 – Schematic representation of the problems of standard MD and the estimate of free energy change

approaches has been developed in the past, e.g. Umbrella Sampling,[148] thermodynamic integration,[149] metadynamics,[150] etc.

2.5.1 Umbrella Sampling

The idea of Umbrella Sampling, to overcome the sampling problem at high energy regions, is to add an extra potential to increase the probability to reach these unlikely regions of the conformational space, i.e. one changes the energy surface in such a way that the sampling problem is neglected. By adding an extra potential $U(\zeta)$ one gets a biased probability distribution

$$\rho^*(\zeta) = \frac{\int \delta(\zeta' - \zeta)\, \mathrm{e}^{-\beta[V(x')+U(\zeta')]}\, dx'}{\int \mathrm{e}^{-\beta[V(x')+U(\zeta')]}\, dx'} \tag{2.50}$$

$$= \mathrm{e}^{-\beta U(\zeta)} \cdot \frac{\int \mathrm{e}^{-\beta V(x')}\, dx'}{\int \mathrm{e}^{-\beta[V(x')+U(\zeta')]}\, dx'} \cdot \frac{\int \delta(\zeta' - \zeta)\, \mathrm{e}^{-\beta V(x')}\, dx'}{\int \mathrm{e}^{-\beta V(x')}\, dx'} \tag{2.51}$$

$$= \mathrm{e}^{-\beta U(\zeta)} \cdot \langle \mathrm{e}^{-\beta U} \rangle^{-1} \cdot \rho(\zeta) \tag{2.52}$$

It is quite unlikely to find a single biasing potential for the complete range of the reaction coordinate. The idea of Umbrella Sampling is to create windows along the reaction path. In each window a simulation is performed with an auxiliary, in most cases harmonic, potential $U(\zeta)$ with a minimum in the desired range. The minima are shifted along the reaction coordinate from window to window in such a way, that the complete range is sampled (Fig. 2.3).

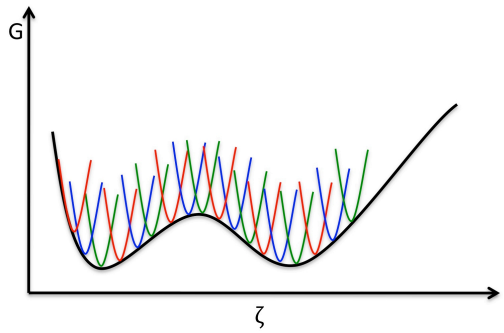


Figure 2.3 – Example for the positioning of the additional potential during Umbrella Sampling

Connecting equation 2.47 and 2.50 one can write

$$\Delta G = -\beta^{-1} \ln \rho^*(\zeta) - U(\zeta) - \underbrace{\beta^{-1} \ln\langle e^{-\beta U}\rangle}_{C_i}, \tag{2.53}$$

where C_i is a different constant for every window. Since one obtains $\rho^*(\zeta)$ during the simulation of every window and $U\zeta$ is known, $\Delta G + C_i$ can be calculated for every window. To connect the resulting potentials of mean force (PMFs), one has to determine the differences between the C_i. Because of the overlapping windows one can merge the PMFs by a method called weighted histogram analysis method (WHAM),[151] which calculates iteratively the C_is and the PMF along the whole reaction coordinate.

2.5.2 Metadynamics

A relatively new approach to distinguish the free energy landscape and also accelerate the corresponding process is Metadynamics.[150] The general idea can be symbolically explained by a walker which is trapped in a valley completely surrounded by high mountains, which stands for the system trapped in a deep free energy minima. The solution to get out is to place a pile sand (Gaussians) where he stands at time t and note position and time. At the beginning, the walker will visit the lower spots more often and will deposit more sand than in higher regions. This leads to a more shallow valley. At some time step t the walker is able to come out of the valley, i.e. the barrier has overcome. Since the position and time of the depositions were noted, one can reconstruct a negative image of the valley profile. Transferring to the molecular system of interest, one assumes that the time-dependent potential build by the sum of Gaussians provides a estimate of the free energy surface of that area.

$$\lim_{t\to\infty} V_G(x,t) \sim -F(x) \tag{2.54}$$

This assumption cannot be derived directly out of any thermodynamic identity, but was postulated on a heuristic basis.[150] The validation was done afterwards[152] and showed that this assumption is correct for systems under the action of a Langevin dynamics.

Algorithm

Like umbrella sampling, thermodynamic integration [153,154] or free energy perturbation, [155] one needs the preliminary designation of one or a set of reaction coordinates, also referred to as collective variables (CV), which are able to characterize the process of interest. The CVs $s_\alpha(x)$ can be any explicit form of the coordinates x. The probability distribution of a CV determines the corresponding equilibrium behavior.

$$P(s) = \frac{\mathrm{e}^{(-\beta F(s))}}{\int \mathrm{e}^{(-\beta F(s))}\, ds} \tag{2.55}$$

where s is the value of $s(x)$. The free energy is given by

$$F(s) = -\beta^{-1} \ln\left(\int \delta(s - s(x))\, \mathrm{e}^{-\beta V(x)}\, dx\right) \tag{2.56}$$

with the potential $V(x)$ describing the system. Like described earlier in this section, one can obtain the free energy surface of the given process by counting the occurrence along the reaction coordinate and the given sampling problems. The way metadynamics antagonises the metastability of local minima and simultaneously reconstructing ($F(s)$) is a history-dependent potential V_G, which modifies the original potential $V(x)$ by adding Gaussians at every time step τ_G.

$$V_G(s(x), t) = \omega \sum_{\substack{t'=\tau_g, 2\tau_G, 3\tau_G \ldots \\ t'<t}} \exp\left(\frac{(-s(x) - s(x_G(t'))^2}{2\delta s^2}\right) \tag{2.57}$$

where ω and δs are the height and the width of the Gaussians and $x_G(t')$ distinguishes the biased trajectory. If more than one CV is used V_G changes accordingly

$$V_G(s(x), t) = \omega \sum_{\substack{t'=\tau_g, 2\tau_G, 3\tau_G \ldots \\ t'<t}} \exp\left(\sum_{\alpha=1}^{d} \frac{(s_\alpha(x) - s_\alpha(x_G(t'))^2}{2\delta s_\alpha^2}\right) \tag{2.58}$$

and one needs to set a width δs_α for every CV α.

The ability of the Metadynamics approach to reconstruct the free energy is given by the approximation that $F(s) = -V_G(s, t)$ in the investigated region. The introduced error depends on the height of the Gaussians and the frequency of deposition and not on the free energy. [156]

$$\epsilon \sim \frac{\omega}{\tau_G} \tag{2.59}$$

2.6 Docking

For many biological processes molecular recognition is a central aspect, e.g. specificity of enzymatic reactions, neurotransmission, immune targeting. With the rapidly increasing number of three dimensional structures of target proteins, determined by X-ray crystallography, NMR experiments or homology modeling, it is possible to investigate the underlying

mechanism. Due to the increased interest in the therapeutic potential of these proteins the necessity of fast evaluation tools is given. Virtual screening has become a powerful tool for searching through large ligand databases too diminish the list of possible candidates for a given target protein and reduce the costs for compound synthesis. Over last two decades several docking approaches has been developed with the goal to be very fast and give reliable evaluation for preferably all ligand types in terms of structural modeling and correct prediction of activity. To tackle this goal, docking is generally divided into two main steps: posing, i.e. determine if a orientation and configuration fit into the binding site and scoring, i.e. evaluation of the ligand protein interactions.

Search Algorithms

The first step within a docking procedure is to generate possible binding position for the ligand. Therefor two aspects have to be taken into account: flexibility of the ligand and, to some extent, the flexibility within the binding pocket. The provision for the ligand flexibility can be divided into three main approaches: (i) stochastic methods: These can be divided into Monte Carlo algorithms, which change randomly the configuration and orientation of the ligand. If the score is better than the the previous one, the position is kept. If it is not a new minimum, the Metropolis criterium is used to decide to reject or keep the position. The other main stochastic methods are based on genetic algorithms. (ii) systematic methods: The most used variation is the incremental search where ligands are grown incrementally into the binding pocket. Therefor the ligand is divided into a rigid core and the flexible sidechains. Once the rigid body is positioned, the flexible side chains are added incrementally. The way how the core is set and the sidechains are added differs between the between the different docking programs. (iii) simulation methods: The most popular simulation approach is the molecular dynamics. The main features of this approach were described earlier in this work. In case of docking, the simulation method is limited, because of the problem too overcome large energy barriers, which leads to a docking results that are close to the starting position. Moreover, molecular dynamics are not suitable for analyzing big ligand data bases.

An additional challenge during the search for binding poses is the protein flexibility, i.e. the change of the binding site due to the ligand binding. This flexibility is included less advanced in todays docking programs. But there are some approaches to include the flexibility to at least part of the protein,[157,158] which includes molecular dynamics, Monte Carlo calculations, rotamer libraries[159] and protein ensembles.[160]

Scoring

Even if the search algorithm predicts the binding pose for a ligand correctly, the overall docking is misleading if the evaluation of the different binding positions is wrong. Therefor a reliable scoring function is very important. Since quantitative free energy calculations are computational demanding and are not easy to use, docking programs use various as-

sumptions and reduce the complexity to evaluate complexes, that are suggested during the searching step. This is mainly the simplification or the neglect of entropic effects. Also this step of docking can be divided into three main approaches: (i) force field-based scoring functions, which quantifies the binding energy by the sum of the ligand-receptor interaction energy and the internal ligand energy, i.e. the energy induced by the binding conformation. In case of a single rigid protein, the internal protein energy can be neglected. Introducing flexible parts within the protein leads to additional energy terms, which are quite similar to the first two. The non-bonded interactions between the ligand and the receptor are most often described like in normal force fields. The van der Waals interactions are calculated by a Lennard-Jones Potential, while the electrostatics are described by Coulombs law. Standard force fields suffer from their parametrization to enthalpic contributions to structure and energetics. They do not include solvation and entropic effects. To diminish these effects one has extended the scoring functions with torsional entropy and hydrogen-bonding terms,[161,162] (ii) empirical scoring functions, which are fitted to reproduce experimental data, e.g. binding energies, with a sum of parameterized functions.[163] Most empirical scoring functions have included approximated terms for entropic effects like entropy penalties on binding based on a weighted sum of the number of rotatable bonds. Other functions includes hydrophobic contributions on the basis of the molecular surface[163] or desolvation effects using continuum electrostatic model.[164] However these corrections are incomplete. (iii) knowledge-based scoring functions, which are also designed to reproduce experimental data, but in comparison to the empirical function they are designed to reproduce experimental structures. This is achieved by simple atomic interaction pair potentials which make the calculation very fast.

Chapter 3

Early states in the bR photocycle

3.1 K state

Despite the large amount of experimental data concerning the K state of the bR photocycle, including X-ray crystallography, FTIR and Resonance Raman spectroscopy there are still open questions concerning the structural features at low temperatures and at physiological conditions. Using QM/MM molecular dynamics at different temperatures and high-*ab initio* methods for calculating excitation energies, we try to give new insights at the atomistic level for low and room temperature K states.

3.1.1 Methods and parameters

Protein model

The starting point for the MD simulations is the crystal structure of ref. 33. Missing coordinates within the structure were modeled using the CHARMM package.[165] The amino acid residues Asp96, Asp115[166,167] and Glu204[168] were considered protonated, all other titrable residues were modeled in their standard protonation state.

Force field and QM/MM parameters

MD simulations were performed using a QM/MM approach. The QM region is described by DFTB3,[120,121] while the remainder of the protein is described by the CHARMM22[169] force field. The QM region includes the sidechains of Asp85, Asp212, Thr89 and Lys216 with the retinal. Additionally, the three water molecules that build the pentagonal water cluster at the extracellular side of the retinal (w401, w402, w406 corresponding to ref. 13) are included. This QM region is denoted here as QM6. The QM/MM border is located between the C_α-C_β bond and the capped QM region is saturated by the link atom scheme[170] and the overpolarization of the electron density is avoided with the "divided frontier charge" scheme.[146]

To obtain the overall shape of the protein, all atoms that are further away than 14 Å from the Schiff base are restrained harmonically to their initial coordinates based on their

B-factors of the crystal structure.

$$f = 0.5 \cdot \frac{3}{2} RT \cdot \frac{8\pi^2}{3B} \tag{3.1}$$

Atoms within 6 Å of the Schiff base are not effected by any restraints. In the region between these two extrema the atoms were restrained harmonically with force constants, which are scaled by a sigmoid function $s(r)$

$$s(r) = \frac{(r - r_1)^2 \cdot (3r_2 - r_1 - 2r)}{(r_2 - r_1)} \tag{3.2}$$

to smooth the transition, with $r_1 = 6$ Å and $r_1 = 14$ Å. The electrostatic shielding effect of the bulk water is modeled by a charge scaling scheme as proposed by Dinner *et al.*.[171]

Molecular dynamics simulations

Based on the QM/MM optimization, one MD simulation at each temperature (50 K, 70 K, 75 K, 80 K, 100 K, 120 K, 150K, 170 K and 300K) was performed including heating, equilibration and production run. The heating of the system to the target temperature was done with steps of 5 K, allowing for 1 ps of equilibration at each step. Equilibration (300 ps) and the production run (5 ns) were performed using the Nose-Hoover[172,173] thermostat with an integration time step of 1 fs.

Computation of excitation energies

The excitation energies are calculated based on optimized structures, which were derived during MD simulations and are calculated with SORCI. The QM region is reduced compared to the MD simulations by neglecting the Thr89 sidechain. Moreover, the protein is not described by the fixed point charges of the CHARMM force field, but with the polarizable model polar.h (see sec. 2.3.4) to consider protein polarization.

Computation of vibrational spectra

To compare the different intermediates we obtain by our QM/MM MD simulations, we computed the IR spectrum of the QM region by the Fourier transform of the dipole autocorrelation function[174] collected by QM/MM MDs. For statistical significance we perform 100 simulations a 40 ps. The nuclear quantum effect on the computed IR spectra is approximated with a harmonic quantum correction factor as done before by Cui and coworkers.[175] This methodology was already used successfully in a previous bacteriorhodopsin study.[168] The band assignment was done by an additionally performed normal mode analysis of the QM region.

3.1.2 Results

Temperature-dependence of the K state retinal dynamics

The MD simulations show that the retinal geometry of the K state is dependent on the temperature. In the low temperature regime from 50 to 120 K, the retinal stays in the

twisted geometry. The probability distributions of the dihedrals along the polyene chain (Fig. 3.1) illustrate that and also shows the expected temperature dependent broadening of the Gaussian shapes. Above 120 K, structural changes within 5 ns could be observed.

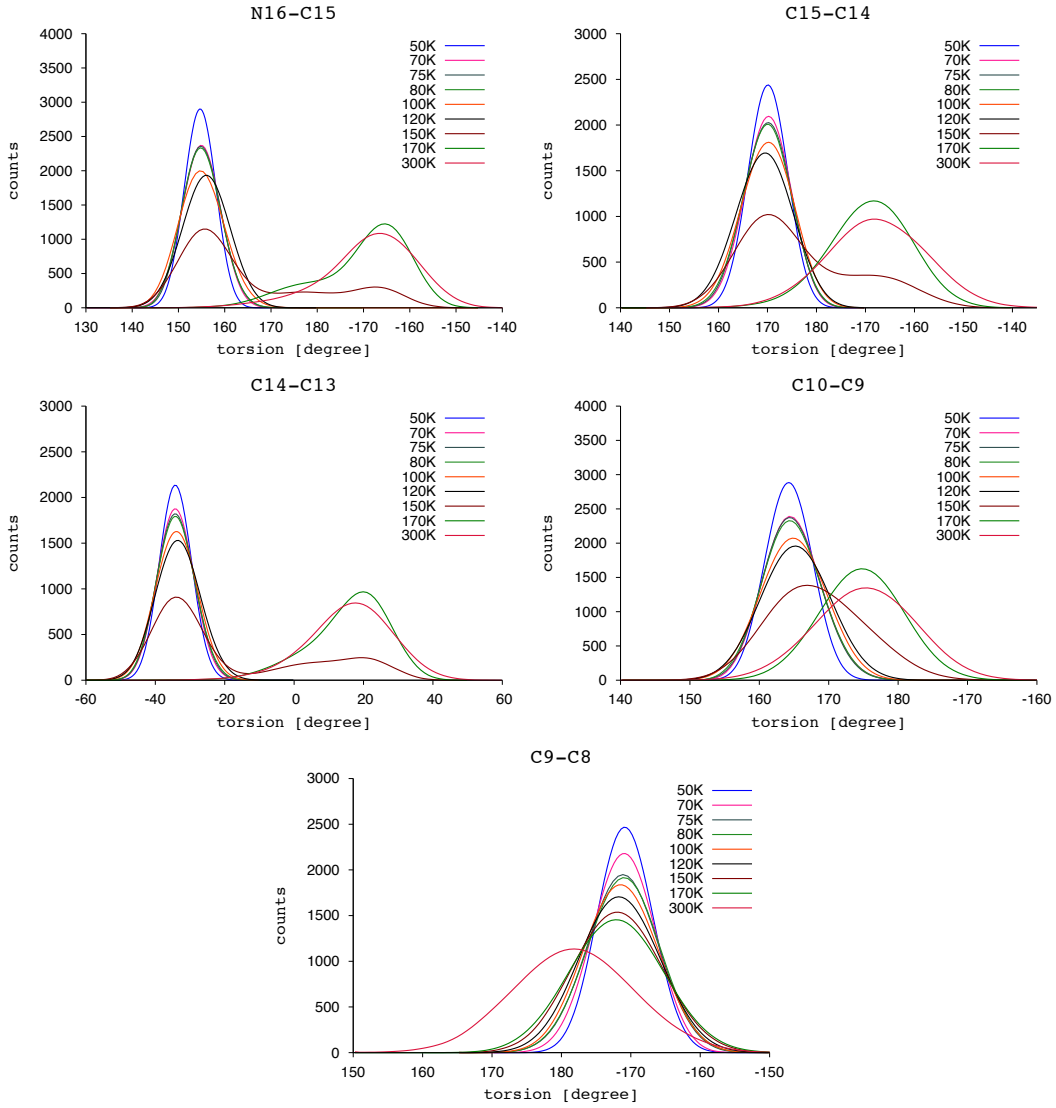


Figure 3.1 – Histograms of the representative dihedrals along the polyene chain of the retinal for different temperatures.

At 150 K the retinal stays twisted for 3.5 ns with a structure equal to the lower temperature structure ensemble. After 3.5 ns the retinal relaxes to a more nearly planar geometry and a Schiff base orientation towards the cytoplasmic side. The time series of the three important dihedrals, N16=C15, C15-C14 and C14=C13 (Fig. 3.2) highlight, that the retinal is fluctuating between a planar configuration and a slightly twisted geometry, which leads to a Schiff base orientation towards Thr89. The histograms of the C14=C13 and N16=C15 dihedrals have their second maximum at 20° and -167°, respectively, which correspond to the slightly twisted geometry. It has to be mentioned, that the simulation is obviously

not converged and the relations of the two maxima within these histograms are hard to interpret in a meaningful way. The simulation at 170 K shows a very fast relaxation of the

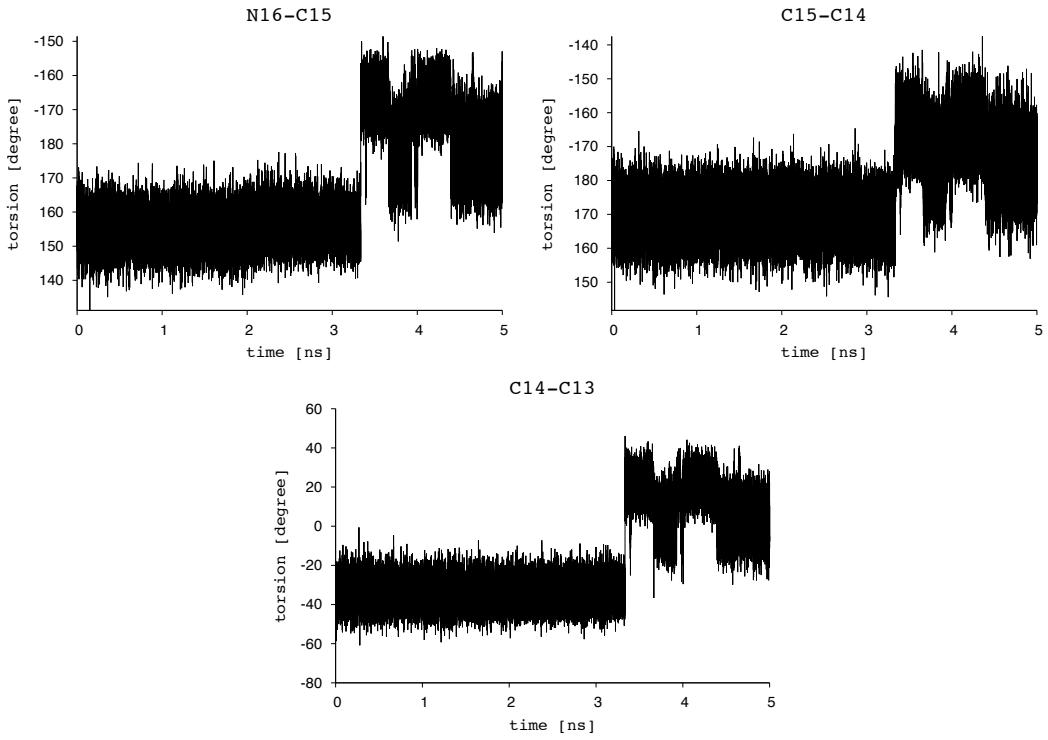


Figure 3.2 – Time dependence of the three dihedrals influencing the retinal twisting at 150 K

retinal within a few picoseconds. During the first 3 ns the retinal geometry switches between the planar and the twisted configuration like in the 150 K simulation (Fig. 3.3). The histograms for the dihedrals show maxima, which depict the preference of the twisting towards Thr89 (Fig. 3.1). Up to 170 K the water cluster at the extracellular side is stable during the simulation time of 5 ns. This changes during the simulation at 300 K. The retinal shows the same relaxation of the retinal with the difference that the retinal remains in the slightly twisted geometry. As already denoted, the water cluster undergoes a change in disposition. W402 which is hydrogen bonded to the Schiff base in the bR ground state and kept the position during the low temperature simulations, moves towards the extracellular side and adopts the position of W401. W401 also moves further to the extracellular side (Fig. 3.4c). To evaluate the derived geometries under aspect of vibrational and visual spectra, the different geometries were optimized. The re-optimized structures of the simulation below 150 K are very similar to the optimized X-ray structure with a RMSD of the binding pocket of 0.15 Å, which can be referred to the equilibration during the MD (In the following termed K Model I; see Fig. 3.4a). The two structures obtained during the 150 K and 170 K simulations, i.e. planar retinal and slightly twisted retinal (Fig. 3.4b+d) coincide into one minimum, which corresponds to a slightly twisted retinal. This implies that the planar geometry is a metastable configuration, which descends into the twisted retinal ge-

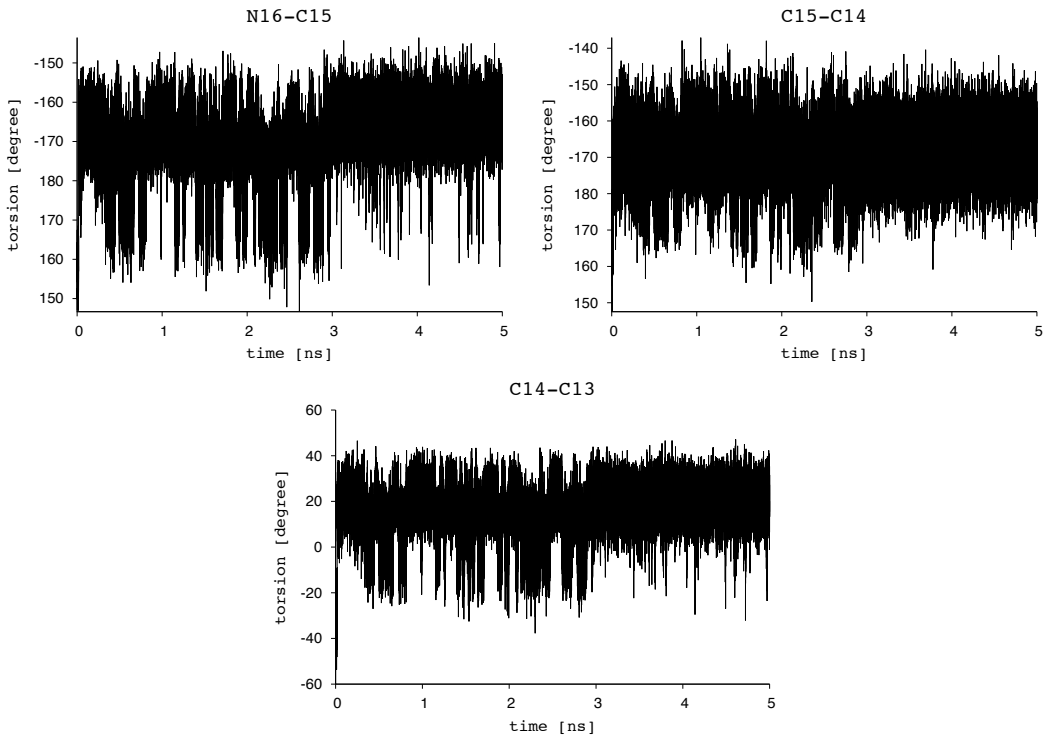


Figure 3.3 – Time dependence of the three dihedrals influencing the retinal twisting at 170 K

ometry during optimization (in the following termed K Model II; see Fig. 3.4b). The fourth structure, obtained during the 300 K simulation, shows no significant changes during optimization. The retinal keeps the twist towards Thr89 and the three water molecules of the water cluster stay at their new positions. In the following termed K Model III; see Fig. 3.4c.

Excitation energies

The calculated excitation energy of a structure and the comparison to the experimental absorption maxima can be a tool to validate this structure. In case of the K state it is somewhat different, since the experimental results contradict, ranging from spectrally silent transitions[45] within the K state to three spectrally distinguishable substates.[46] In Tab. 3.1 the calculated values for our three derived models and experimental data are listed.

Since Model I, which was extracted from the low temperature (below 150 K) molecular dynamics, resembles the cryotrapped X-ray structure, the experimental excitation energy should also be reproduced by our model. The chromophore in Model I has a vertical excitation energy of 1.92 eV. The experimental value for low temperature (77 K) is 1.96 eV. Thus, the experimental absorption maximum is reproduced within the error of our method. The assignment of Model II and III to an experimental value is much more complicated. At first sight, Model III resembles the K_L [46] or KL [25] value. However, these absorption maxima were observed after 250 ns or 150 ns, respectively. Assuming that DFTB3 reproduces a rate limiting barrier within the correct order, it is unlikely that our Model III matches with this

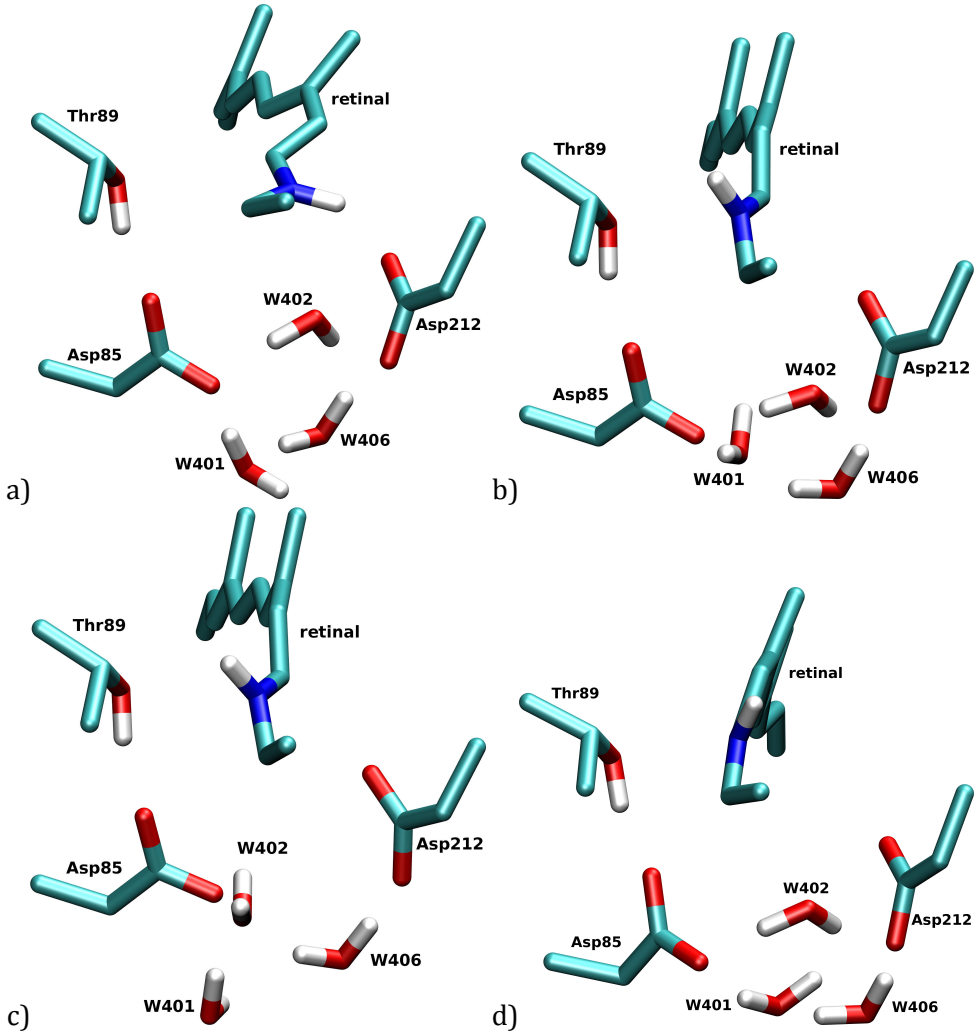


Figure 3.4 – The K state models obtained by MD: a) K model I; b) K model II; c) K model III; d) metastable structure with a planar retinal. Obtained during 150 and 170 K MDs

Table 3.1 – Comparison of calculated and experimentally derived excitation energies (in eV and nm)

MD-models	K Model I	1.92 (645)
	K Model II	2.16 (574)
	K Model II	2.12 (584)
Iwasa[176]	K_{LT}	1.96 (630)
Shichida[25]	K	2.03 (610)
	KL	2.08 (596)
Dioumaev[46]	K_E	2.06 (600)
	K_E	2.10 (590)

late K substate. Since the proposed K of Shichida *et al.* and K_E of Dioumaev and coworkers are only slightly red-shifted compared to KL and K_L, respectively, the excitation energy of

Model III matches also with K or K_E. Model II which was observed during the 150 K and 170 K simulations has an excitation energy of 2.16 eV, which does not fit with the experimental values. Since this structure is also not observed during the 300K simulation, this might imply that this structure does not occur at room temperature.

Vibrational spectra of the retinal

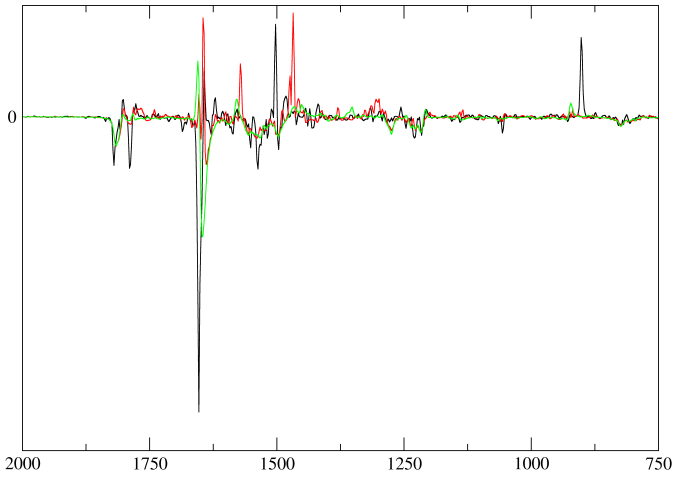


Figure 3.5 – The calculated vibrational spectra of the retinal for the three K state models. Black: K-bR-I (80K); Red: K-bR-II (135K); Green: K-bR-III (135K)

The same three models were used for the vibrational spectra. Experimental difference spectra were derived at 80 K and 135 K.[38,177] Therefore, we obtained a K-bR spectrum at 80 K (K-bR-I) with our Model I as a starting point for the simulations. The k-bR difference spectrum at 135 K were calculated for Model II and Model III. Model I was used for a K-bR spectrum at 80 K (K-bR-I). The two other models are used for K-bR spectra at 135 K (K-bR-II and III).

HOOP-Mode

The experimental K-bR (80 K) spectrum shows a strong positive peak at 960 cm^{-1}, which decreases by 50% due to heating to 135 K. Additonally, a new positive peak occurs at 985 cm^{-1} . These peaks are assigned to the HOOP-Mode of the retinal. Using our approach, we also observe a strong positive peak within our K-bR-I spectrum (Fig. 3.6). Though, it is present at 902 cm^{-1} and vanishes completely within the two K-bR (135K) spectra. We can also reproduce the arising peak at 135 K for both spectra. In case of K-bR-II the new peak consists of two peaks at 918 and 924 cm^{-1}. The two peaks are incorporated in the K-bR-III spectrum by a peak at 922 cm^{-1}. The absolute wavenumbers of our calculated HOOP modes

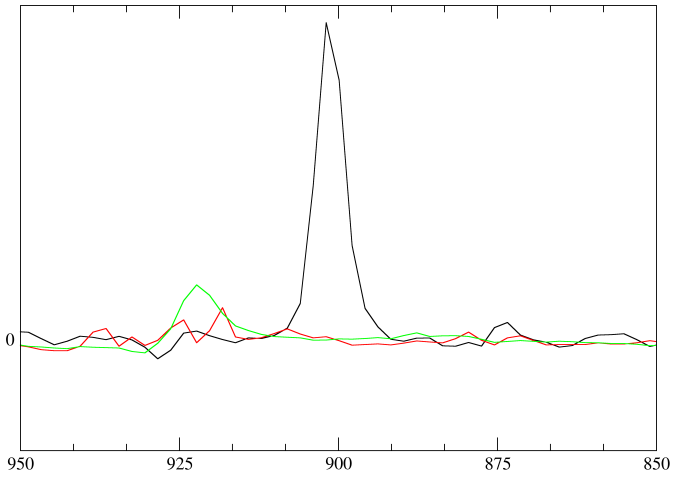


Figure 3.6 – The calculated HOOP-mode for the three K state models. Black: K-bR-I (80K); Red: K-bR-II (135K); Green: K-bR-III (135K)

are low compared to the experiment. Nevertheless, the shift between the two peaks is near to the experimental results. The difference in intensity change might occur from the fact that we describe a single state while the experiment has a mixture of states.

Chromophore vibrations

The experimental K-bR spectra show several negative peaks (1255, 1216, 1202 and 1010 cm^{-1}) which have a decreased intensity due to heating to 135 K. These wavenumbers were assigned to chromophore vibrations. Within our calculations we can observe the vibrations between 1050 and 1246 cm^{-1} (Fig. 3.7). There are two negative peaks at 1055 and 1064 cm^{-1} which can be assigned to vibration within the β-ionone ring and decrease like in experiment. Additional three peaks at 1215, 1229 and 1246 cm^{-1} occur that correspond to in-plane H/CH3 vibrations along the chromophore. The differences between K-bR-II and K-bR-III are quite small and within the error of the method.

C=C and C=N stretching

Using DFTB3 for the calculation of the chromophore stretching frequencies leads to a strong coupling between the C=C and C=N stretching vibrations.[178] In contrast to the experiment we cannot observe distinct peaks for the two different stretching modes. In experiment a negative peak at 1530 cm^{-1} and a positive peak at 1515 cm^{-1} was observed for the C=C stretching. While the C=N stretching shows a negative band at 1640 cm^{-1} and a positive peak at 1609 cm^{-1}. The calculated peaks around 1650 cm^{-1} can be assigned to the coupled vibrations.

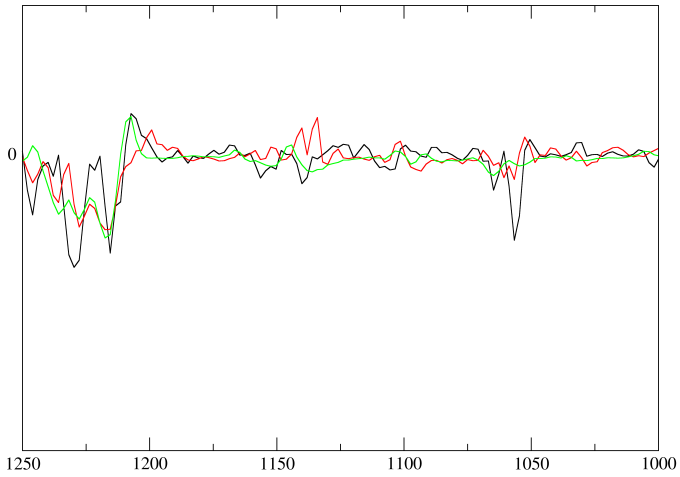


Figure 3.7 – The calculated chromophore vibrations for the three K state models. Black: K-bR-I (80K); Red: K-bR-II (135K); Green: K-bR-III (135K)

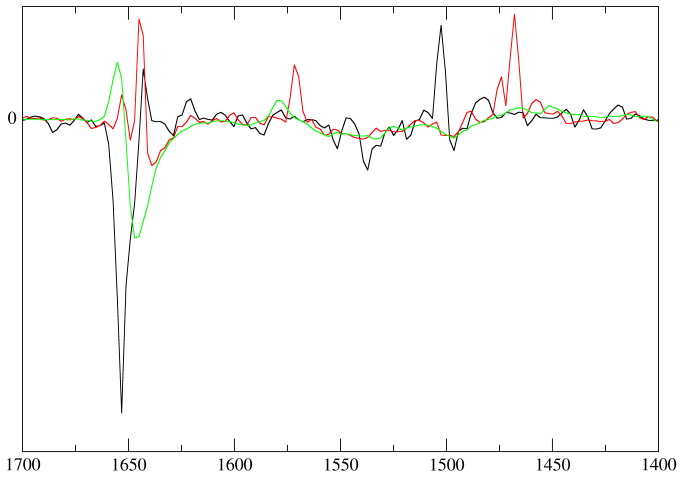


Figure 3.8 – The calculated C=C and C=N stretching spectra for the three K state models. Black: K-bR-I (80K); Red: K-bR-II (135K); Green: K-bR-III (135K)

3.1.3 Discussion

Several experiments suggest different numbers of K substates that are spectroscopically distinguishable. This ranges from three substates at 90 K,[37] over one low temperature low K state[176] plus two at room temperature[25,46] to only one spectroscopically distinguishable K state.[45]. Furthermore, FTIR spectroscopy[38,41,177] was used to identify changes in

the retinal configuration at low temperature. During our simulation we could not observe more than one K state in the low temperature region (< 150 K). One reason is of course the short simulation time of 5 ns, but also the simulations at high temperatures which results in an enhanced sampling gave no indication to additional intermediates with an absorption maximum between 625 and 645 nm. A reason for this result could be the bias which is introduced by using one X-ray structure as starting point. Nevertheless, the calculated excitation energy of our derived model (Model I) for the low temperature K state is in very good agreement with the experimental value of Iwasa *et al.*[176] and matches perfectly with one of the three low temperature K substates.[37] Vibrational spectra experiments suggest that the retinal undergoes a configurational changes caused by a temperature change from 80 K to 135 K. The main differences were the changes in the HOOP-mode, chromophore in-plane vibrations and C=C/C=N stretching vibrations. Our calculated difference spectra based on our Models I-III could reproduce the experimental spectra, and therefore support the experimental presumption that the retinal undergoes a relaxation. Our simulation at higher temperatures (> 150K) depicted a very fast transition from the twisted retinal to a nearly planar retinal with a cytoplasmic orientation. This supports the time-resolved Resonance Raman spectroscopy experiment that, suggested a very fast relaxation of the retinal (3 ps) at room temperature. UV/Vis experiments[25,46] suggest two K intermediates at room temperature, which are blue-shifted compared to the low temperature K state. In case of ref. 25 an early K and a late KL with absorption maxima of 2.03 eV and 2.08 eV, respectively. In ref. 46 a K_E (2.06 eV) and K_L (2.10 eV) were proposed. Our two models lead to excitation energies which are also blue-shifted compared to K_{LT}, but Model II which was derived by 170 K simulation is much too blue. Dioumaev and coworkers[46] proposed that the K_E is formed after 50 ps while the formation of the K_L intermediate needs 250 ns. Therefore, one can suppose that our Model III, which shows a good agreement with the K_E, resembles this early K state. The proposed late K state cannot be observed by our method, since simulations with this setup are only feasible for simulation times of up to 10 ns.

3.2 L state[1]

Despite the importance of the L state for the first proton transfer from the Schiff base to Asp85, the structural properties of this intermediate are still under debate. In detail, which geometry has the binding pocket, the orientation of the Schiff base and location of water molecules in the vicinity of the retinal. Additionally the thermal dependency of these structural factors is unclear. Since the L state is distinguishable by UV/Vis spectroscopy, we tried to understand which of the proposed models reproduces the absorption maxima measured in experiment. Therefore, we performed QM/MM molecular dynamics at room temperature and calculated the excitation energies of derived structures to give new insight into the L state at the atomistic resolution and use these to confine the possible models for the L state.

3.2.1 Methods and parameters

Protein models

We performed three independent sets of simulations on the crystal structures from ref. 51, 52 and 54 (PDB codes: 1vjm, 1ucq, 2ntw). Due to the discussed reasons in the introduction the models from ref. 47 and 53 are neglected (PDB code: 1e0p and 1o0a). In case of the model of Kouyama *et al.*,[52] the Thr89 N-C_α-C_β-C_γ dihedral was changed to a *gauche*-configuration, compared to the *trans*-geometry in the published structure, because previous computations showed that the *gauche*-configuration is energetically favored.[56]

Molecular dynamics simulation

For each L state model we performed a energy optimization followed by heating, equilibration and production run. The heating of the system was done in 20 steps, allowing for 1 ps of equilibration at each step. 300 ps of equilibration and 2.5 ns of production runs were performed using the Nose-Hoover[172,173] thermostat with an integration time step of 1.0 fs.

Quantum mechanical region

The molecular dynamics simulation have been perfomed with the same setup like for the K state calculations (section 3.1). In case of the excitation energy calculations we extended our setup. To investigate the impact of the protein environment and the size of the QM region on the excitation energies, we performed additional sets of computations with the retinal in gas phase. The according coordinates were extracted from the QM/MM trajectory using QM6. Thus, one can separate the impact of steric from the electrostatic influences on the excitation energy of the chromophore. We also assessed a smaller QM region, termed QM1, where only the retinal and the Lys216 sidechain are described on a quantum dynamical level, while the rest of the protein is included at the MM level.

[1] Is in part reproduced from Wolter, T.; Welke, K.; Phatak, P.; Bondar, A.-N. and Elstner, M.; *Phys. Chem. Chem. Phys.* **2013**, 15, 12582--12590.

3.2.2 Results

Structural features of the QM/MM optimized L state models

In this part we focus on the aspects of the geometry of the active site that are potentially important determinants of the excitation energy. Other structural aspects have been discussed already in the past.[30,56,179] Due to simplification, the QM/MM optimized structures will be denoted in the following way: L model A, B and C derived from the crystal structures from ref. 54, 51 and 52, respectively.

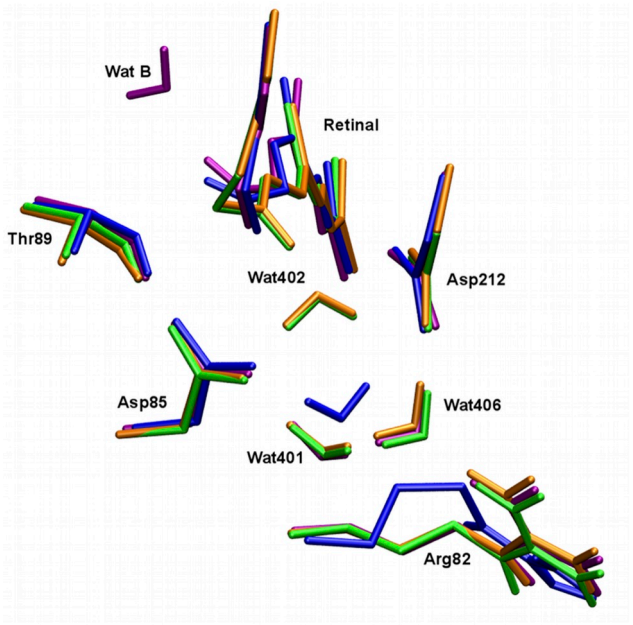


Figure 3.9 – Comparison of the QM/MM-geometry optimized L models A (green), B (blue), C (purple) and the bR resting state (orange). Note that upon geometry optimization of L model B, the retinal Schiff base segment becomes slightly twisted towards Thr89. L model A becomes almost identical with the bR resting state.

L model A

As noted before, the retinal of ref. 54 has a twisted all-*trans* comformation. Upon the geometry optimization, the retinal relaxes towards a planar all-*trans* (Fig. 3.10a). The responsible dihedral angle around the C13=C14 double bond increases from -107.9° in the crystal, to -150.7° after the optimization. All dihedrals along the polyene chain for all crystal structures and corresponding optimized structures are tabled in the Appendix (Tab. A.1). Due to the relaxation of the retinal the distance between the Schiff base nitrogen and the water 402 oxygen is slightly reduced by 0.1 Å to 2.8 Å. Moreover, the binding pocket of L model A and the optimized structure of the resting state of bR[180] are very similar. The sidechain of Arg82 moves towards the proton release group at the extracellular side of the protein.

L model B

Within the crystal structure[51] the retinal was depicted to be planar along the polyene chain. Upon geometry optimization the retinal becomes a twist towards Thr89 (Fig. 3.10b). The twist is mainly localized around the N16=C15, C15-C14 and C14=C13 bonds by 20.0°, 24.4° and 21.0°, respectively. The distance between Thr89 and Asp85 is shortened by 1.7 Å to a gap of 2.8 Å, which is quite similar to the K state (2.77 Å). FTIR studies[36] suggest that the strength of the hydrogen bond between Thr89 and Asp85 persists during transformation from K to L. Former theoretical studies[181] on the L state crystal structure of ref. 47 showed the same behavior in terms of retinal twist and formation of the hydrogen bond between Thr89 and Asp85. These independent results, suggest that a planar retinal with a cytoplasmic orientation may be less favorable.

L model C

During geometry optimization the overall starting structure of ref. 52 is preserved (Fig. 3.10c). The main changes are a shortened hydrogen bond between water B and the Schiff base by 0.3 Å and a stronger twist towards Thr89. This twist is localized at the C14=C13 and C13-C12 bonds whose dihedral angle change from 10.3° to 22.6° and 177.3° to 164.4°, respectively.

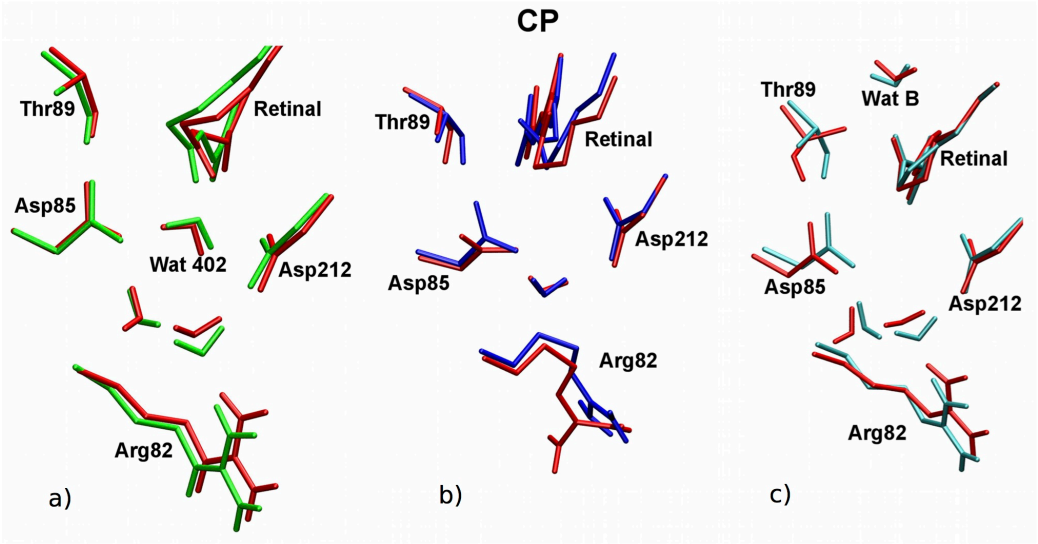


Figure 3.10 – Comparison of the starting crystal structures and the QM/MM-optimized L-state geometries. (a) L model A (green), (b) L model B (blue) and (c) L model C (cyan). The crystal structures are colored red. CP: cytoplasmic side.

QM/MM MD simulations

As mentioned earlier, crystals for the different bR photocycle intermediates are cryotrapped at certain temperatures. In case of the L intermediate temperatures from 150 to 170 K were used to purify the intermediate. This might lead to models that may not resemble the actual structure, that is functional under physiological conditions. Therefore, we pre-

formed MD simulations at room temperature to inspect the thermal stability of the models derived by QM/MM optimization and allow for a better description with conditions closer to the native environment of bR. It has to be noted, that due to the limited computational time the structural changes derived by molecular dynamics are still biased by the starting X-ray structure, because large configurational changes that are connected to large energy barriers may not be captured within the nanosecond timescale.

In case of the L models A and B, there are no notable changes during the 2.5 ns simulation at 300 K. The binding pocket including the retinal keeps the overall geometry like in the optimized model, except for small thermal fluctuations. Thus, the derived models seem to be stable at 300 K in respect to the relatively short simulation time.

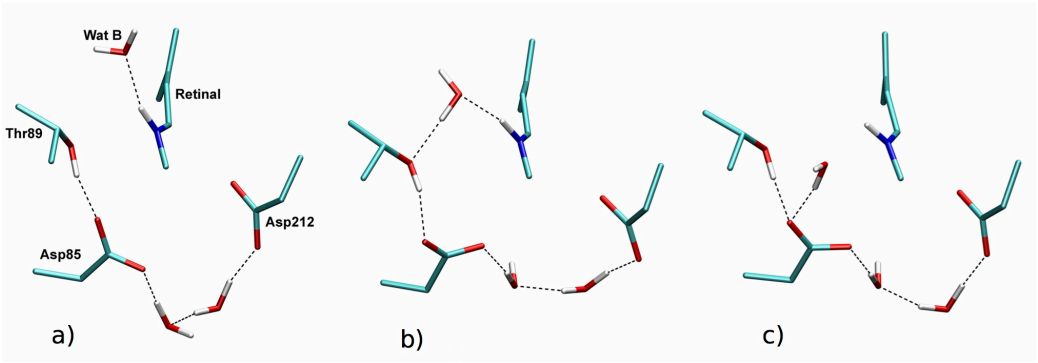


Figure 3.11 – Coordinate snapshots illustrating Schiff base water interactions sampled for L model C. (a) L model C after QM/MM optimization showing Thr89 in the *gauche* configuration, and water B as a hydrogen-bonding partner for the Schiff base. (b) Snapshot from the QM/MM MD at 300 K. The retinal is twisted, and water B bridges the Schiff base and Thr89. (c) After ~1.8 ns, the dynamics converges to a protein configuration where water B bridges the twisted retinal Schiff base and Asp85. We denote this bridged structure as L model D. The coordinate snapshots depicted in panels a-c are similar to those sampled in previous QM/MM MD and reaction path computations.[56]

On the contrary, L model C shows rearrangements in the vicinity of the retinal. Early during the equilibration of Model C, water B reorients and forms an additional hydrogen bond to Thr89 (Fig. 3.11b). This stabilizes transiently the chromophore twist towards Thr89. Within approximately 1.8 ns, retinal and water B undergo reorientations leading to geometry in which water B does not build a hydrogen bonded network with the Schiff base and Thr89, but with the Schiff base and Asp85 (Figure 3.11c). A very similar geometry was also obtained by former theoretical studies.[56] This structure will be termed in the following L Model D

Excitation energies

As already discussed earlier, excitation energies are quite sensitive to the active site structure. Configurational changes and the presence/absence of single water molecules in the active site can have some measurable influence, as discussed in previous work.[129,137--141]

Table 3.2 – Connection between structural parameters and the excitation energy of the chromophore in the gas phase (in eV and nm)

	bR	L model A	L model B	L model C	L model D
ΔE	1.86 (667)	1.86 (667)	1.88 (659)	1.92 (646)	1.93 (642)
planarity[a]	10.01	9.43	11.07	9.65	7.05
max. rotation	18.6	19.3	20.9	22.6	21.4
BLA[c] (in pm)	5.39	5.41	6.39	6.71	6.08

[a] Average deviation of dihedral angles from planarity

[b] max. rotation is around C13=C14 in all cases

[c] Bond length alternation = difference between averaged C-C single bonds and C=C double bonds

In comparison to our studies regarding the K state, we tried to elaborate these factors for the different L state models to get a better understanding which structural features have impact on the excitation energy.

Impact of the retinal geometry on the vertical excitation

There are two structural main determinants of the excitation energy of the retinal. The first important factor is the bond length alternation (BLA), which describes the ratio of the length of single and double bonds along the polyene chain. For example, increasing the BLA leads to a blue shift of the excitation energy due to less conjugated double bond. The second major factor is the deviation from planarity of the π-conjugated system. A strong twist around the C13=C14 bond leads to a red shift, because the rotation destabilizes the electronic ground state (S_0) and stabilizes the first excited state (S_1). To investigate the impact of these factors, we isolated the chromophore under vacuum by removing the protein, but keeping the retinal geometry like in our derived models.

The data summarized in Table 3.2 show that the gas phase excitation energies vary by 0.07 eV, indicating that the difference in excitation energies results from the interaction with the protein environment. The major structural determinant of the variation is the BLA of the retinal polyene chain. The deviation from planarity is similar for all models assessed here (Table 3.2). The similar values of the excitation energies computed for L model A and the bR ground state further supports our proposal that the structural model from ref. 54 is a twisted all-*trans* structure incompatible with L. The excitation energy of the retinal from L model B is slightly higher than that of the bR ground state. The increased BLA has a larger impact on the excitation energy than the slightly stronger twist of the chromophore, which leads to a 0.02 eV blue-shift compared to the bR resting state.
The blue-shift of 0.06 eV of the L model C with respect to the bR resting state can be explained by the larger BLA in L model C as compared to L model A and L model B. The marginal difference between L models C and D is caused by the compensation of two opposing effects: a more twisted retinal L model C (around C13=C14), and a smaller BLA of

the retinal in L model D.

Impact of the protein environment and size of the QM-region

Table 3.3 – The retinal bond length alternation, the protein opsin shift, and the difference dipole moments measured from the QM/MM optimized structures

L model	BLA	protein shift (eV)
A	5.41	0.47
B	6.39	0.5
C	6.71	0.63
D	6.08	0.49

The effects that are included by adding the protein, described by point charges (QM1), are (a) the stabilization of the S_0 state relative to the S_1 by the negative charged counterions, since the positive charge is shifted from the Schiff base towards the β-ionone ring during the S_0-S_1 excitation and (b) the remainder of the protein which polarizes the QM-region. This leads to a blue shift of the excitation energy induced by the protein polarization from 0.47 eV in case of model A to 0.63 eV for model C (Table 3.3). All of these values are too blue compared to the experimental value. This artifact arises from the neglect of dispersion interaction and polarization of the protein environment, if describing the electrostatic protein environment by fixed point charges.

Increasing the size of the QM region from QM1 to QM6 leads to two additional effects: (a) the polarization of the counterions and (b) a small charge transfer from the Schiff base towards the counterions and the water, which leads to a additional stabilization of the S_0 state.[141] For bR and L model A, in which the Schiff base is oriented towards the extracellular side, the protein blue-shifts (0.08 eV and 0.05 eV, QM6-QM1 in Table 3.5) are comparable. L models B, C and D, in which the Schiff base is oriented towards the cytoplasm, the protein blue-shifts range from 0.11 eV for L models B and D, to 0.23 eV for model C. These relatively large blue-shifts observed for these models are rather surprising, because the polarization of the counterions is small. Similar to previous studies,[180] we observe a correlation between the BLA and the response of the excitation energy to the electric field induced by the point charges of the protein (Table 3.3).

Furthermore, we have investigated the impact of the Arg82 configuration on the excitation energy, because in L model B the sidechain of Arg82 adopts a different orientation compared to the other models. This was achieved by mutation into a glycine, which effectively removes the charges of the arginine sidechain. In general, the replacement has only a small effect on the excitation energies (Table 3.4), which was also found for the bR resting state previously.[141,180]

In previous work[182] it was shown that water near the chromophore can have an impact on the excitation energies. On that account, we analyzed the impact of the assumed

Table 3.4 – Perturbation analysis: Influence of Arg82 and water molecules water 402 and water B on the excitation energy of the retinal chromophore

	Schiff base orient.	Orient. of Arg82[a]	R82G-Shift[b]	wat402-shift[c]	wat402-shift[d]
bR	extracellular	9.98	-0.03	-0.08	-0.16
L model A	extracellular	10.27	-0.01	-0.08	-0.19
L model B	cytoplasmic	11.22	-0.03	-	-
L model C	cytoplasmic	10.82	± 0	-0.03	-0.12
L model D	cytoplasmic	10.47	+0.02	-0.06	-0.13

[a] Distance between chromophore (N_{16}) and Arg82 (C_Z)

[b] Effect of mutation from arginine to glycine on excitation energy (QM6)

[c] Effect of deleting water402/water B on excitation energy (QM1)

[d] Effect of deleting water402/water B on excitation energy (QM6)

important water 402 and water B. Unlike Arg82, water 402 and water B have larger effects on the excitation energies computed for all models except L model C. The change in the excitation energy upon removing water 402 ranges from -0.08 eV in case of bR and L model A, to -0.03 eV in case of L model C. Note that neither water 402 nor water B is present in the L model B. Water B becomes an important determinant of the excitation energy when it moves to the bridging position between the retinal Schiff base and Asp85 (L model D). Removing the point charges of water B in L model D leads to a red-shift of 0.06 eV.

When the computations are performed with the larger QM6 region, removing water 402 from L model A leads to a significant red-shift of 0.19 eV. Water402 in bR and L model A , where it is located between the Schiff base and the Asp85/Asp212 counterions, has a slightly larger influence than water B in L models C and D, where it is positioned on the cytoplasmic side of the Schiff base. The shifts reflect the stronger polarization of the water molecule by the protein environment in the former case.

Impact of the electronic polarization

The calculation of excitation energies using fixed point charges to describe the protein environment of the chromophore leads to blue shift compared to the experimental value. The description can be improved by using a polarizable force field to describe the rest of the protein. Using our polarizable model, which includes the effect of mutual instantaneous polarization of the protein environment and the QM region results in red shifts of the excitation energies (Table 3.5). In case of an extracellular-oriented Schiff base, the polarization-induced red shift is 0.22 eV for bR and 0.23 eV for L model A, respectively. The red-shifts obtained with L models B, C and D, where the retinal is cytoplasmic-oriented are larger (-0.23 to -0.34 eV). Therefore, the large protein-induced blue-shifts are counterbalanced by polarization-induced red-shifts.

Table 3.5 – Effect of different QM fragments and polar.h on excitation energy (in eV and nm)

	bR	L-Model A	L-Model B	L-Model C	L-Model D
QM1	2.32 (534)	2.33 (532)	2.38 (521)	2.55 (486)	2.42 (512)
QM6	2.40 (517)	2.38 (521)	2.49 (498)	2.75 (451)	2.52 (492)
QM6+ polar.h	2.18 (568)	2.15 (577)	2.15 (577)	2.41 (514)	2.25 (551)
exp[a]	2.18 (568)		2.28 (544)		

[a]Absorption maxima at room temperature [23]

3.2.3 Discussion

In this work, we managed to give new insights into a very controversial topic within the bR universe. We tested three of five available X-ray structures, while we discarded the other two, because of the reasons discussed in the introduction. Using QM/MM simulations including molecular dynamics and excitation energy calculation, we evaluated the compatibility of the different models with the experimentally observed absorption maxima. To achieve this goal a reliable method for excitation energies is needed. We have used a intensively benchmarked method (see section 2.3.4), which shows reliable shifts for retinal models. Therefore, we are confident, that we can apply the method to the L intermediate and expect reliable estimates of the excitation energy.

The optimized L-model A, which is based on the X-ray of ref. 54 (pdb code: 2ntw), shows a significant structural overlap with the crystal structure of the bR resting state. As mentioned in the introduction, the retinal has a rather twisted all-*trans* than a 13-*cis* configuration. This estimate is approved during QM/MM optimization, which leads to an all-*trans* retinal. The following evaluation of the spectral properties of this structure showed that the excitation energy of L-model A is very similar to bR. Moreover, the evaluation of the different structural properties affecting the excitation energy showed a high similarity of L-model A and the bR ground state.

L-model B has a complete different active side. Nevertheless, the calculated excitation energy is exact the one of L-model A. This illustrates the problem of excitation energies. Although they are very sensitive to the structure of the active side and even the reorientation of a water molecules, they condense many structural properties into one value, which can lead to the same excitation energy for two different structures. Therefore, one has to keep in mind that a wrong structure can give the right excitation energy. However, a wrong excitation energy is a strong argument to question the corresponding structure.

The calculated excitation energy of L-model C is much too high compared to experiment. Moreover, L-model C shows a thermal instability which leads to the relocation of water B into a new position (L-model D). This was already observed by independent calculations.[56]

The relocation of water B leads to an absorption maximum shift in close agreement with experiments which show a range of 541 to 550 nm (2.29 to 2.25 eV) depending on temperature and spectral reconstruction.[37,183].
In summary, it is doubtful that L-Model A reflects the L intermediate, since the active side geometry resemble the bR ground state just like the calculated excitation energy. Also L-model B shows a red shift compared to the bR ground state, which is hard to align with the blue-shifted L intermediate. In contrast to L-model B, model C shows a blue shift compared to bR that is 0.13 eV too high. Furthermore, this model seems to be unstable in physiological conditions. L-model D is the only model in our study, which leads to an excitation energy comparable to experiment. It is possible that this is a false positive result. Nevertheless, this structure was also proposed by minimum energy pathway calculations as a possible pre-proton-transfer structure.[56] Summarizing these results, our obtained L-Model D is an eligible candidate for the pre-proton transfer structure at physiological conditions.

3.3 Conclusion

In this chapter, we could give new insights concerning binding pocket properties of the early photocycle intermediates. The main drawback of the experimental studies is the lack of direct observations at the atomistic level. There are X-ray structures of the different intermediates available, but with contradicting features. Experiments like FTIR or RR can obtain dynamical properties, but have to use models to assign their spectra to the structural changes in the protein. We tried to close the gap between these informations by QM/MM simulations. We derived structural models of the K and L state, based on available crystal structures. These structures were evaluated by calculating their vertical excitation energy and vibrational spectra to compare it with the experimental absorption maxima and FTIR spectra, respectively. Thereby, we can sort out proposed structures, which cannot reproduce the experimental absorption shift or produce the wrong vibration spectra. Nevertheless, a positive evaluation of a structural model is not a clear evidence for correctness, e.g. the correct excitation energy can be obtain by a wrong structure, since all structural parameters are condensed in value.

In case of the L-state, we doubt all evaluated cryotrapped X-ray structures to represent the L state at physiological temperatures. None of them could reproduce the experimental absorption shift. However, we could obtain a new structural model, based on the crystal structure of Kouyama *et al.*[52], which was able to reproduce this shift. Furthermore, this derived structure was already proposed by an independent theoretical study.[56]

In case of the K intermediate, we could support the model of different K states. As proposed by FTIR and Resonance Raman spectroscopy, a fast relaxation of the retinal occurs at room temperature on the picosecond scale. Based on our derived models we could reproduce these vibrational spectra and backing their assumption of a fast relaxation of the reti-

nal, followed by a increased twist afterwards. Furthermore, our models could reproduce the experimental absorption shifts in the error range of this method. Therefore, we are confident that our results contribute to the understanding of the temperature-dependent K substates.

Chapter 4

Dynamical properties of the ligand binding domain of iGluR2[1]

4.1 Evaluation of known ligands

The general understanding of the glutamate receptor, especially the ligand binding domain (LBD), is essential for the prediction of binding processes and the design of new ligands. Therefore, we performed long time scale molecular dynamics (MD) based on several crystal structures with bound cofactors, ranging from agonists to antagonists. Moreover, free energy calculations were performed to uncover the energetics of the cleft opening and ligand binding/unbinding.

4.1.1 Methods

All MD simulations are based on published X-ray structures of the iGluR2 ligand binding domain with different ligands bound to the protein. For these benchmark calculations only the monomer of the LBD was considered to ensure computational efficiency. The used cofactors are AMPA (pdb code: 1ftm),[72] glutamate (1ftj),[72] 2-BnTetAMPA (2p2a),[74] iodo-willardiine (1mqg),[73] DNQX (1ftl)[72] and the APO state (1fto).[72] Protein models were completed by automatic model building tools, embedded in cubic periodic boxes, solvated with ca. 30,000 TIP3P[184] water molecules and neutralized by adding chloride ions. The protein is described by the AMBER99SB force field[185] and the ligands were parametrized according to the GAFF force field[186] using the Antechamber module of Amber Tools 11. All systems were optimized by 500 steps of steepest descent minimization, followed by 500 ps of temperature and volume equilibration to 300 K and average system density of 0.98 g/ml. During equilibration the protein structure was restrained by harmonic forces of 1000 kJ mol^{-1} nm^{-2} on the non-hydrogen atoms. The temperature and the pressure are controlled by the Nose-Hoover thermostat[172,173] and the Parinello-Rahman[187] barostat, respectively. The sequentially computed free molecular dynamics simulations have a length of 1 μs. All

[1] Is in part reproduced with permission of Wolter, T; Steinbrecher, T and Elstner M., *PLOS One*, **2013**, 8, 1--13.

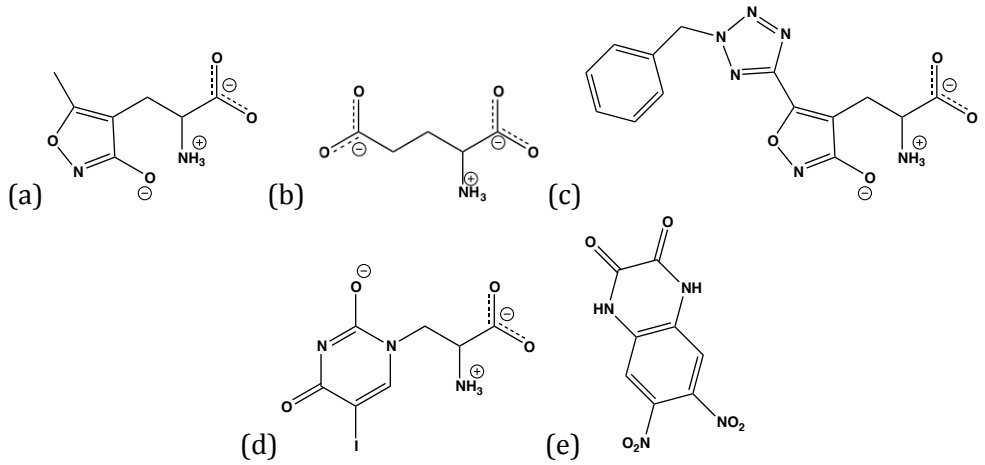


Figure 4.1 – Chemical structures of AMPA (a), glutamate (b), 2-BnTetAMPA (c), iodowillardiine (d) and DNQX (e)

simulations were performed using the Gromacs simulation package version 4.5.x.[188] The free energy calculation were conducted using the umbrella sampling (US) and Metadynamics approach. The Metadynamics calculations were performed using the PLUMED plugin 1.3 for gromacs.[189]

4.1.2 Results

Structural models

In the following the basic binding motives within the binding pocket for all used ligands are described. The description is based on the geometry optimized structures. As shown in Fig. 4.2, all ligands build a salt bridge between the negatively charged α-carboxyl group and the negatively charged sidechain of Arg485 (domain 1), which is therefore called the main anchor for the ligands. In case of DNQX the α-carboxyl group is mimicked by two carbonyl groups. Furthermore, the backbone NH of Thr480 (domain 1) provides a hydrogen bond to the same carboxyl group in all cases. The binding pockets of AMPA shows an additional hydrogen bond from the backbone NH of Ser654 (domain 2). The positively charged α-amino group of all ligands is salt-bridged to the carboxylate group of Glu705 (domain 2). This is not preserved for DNQX, where no amino group is present. Additional hydrogen bonds to the α-amino group are provided from the backbone carbonyl oxygen of Pro478 (domain 1). The main differences are located in the binding motives of the γ-carboxyl group, or the groups that mimic this feature of glutamate, and the domain 2. Except DNQX, which shows no interaction with Ser654 and Thr655, all other ligands build hydrogen bonds from the γ-group to the backbone NH and the hydroxyl group of Thr655 (domain2). It is to mention that the interaction between the backbone NH and the negatively charged oxygen of AMPA is bridged by a water molecule (W4). This interfacial water is a unique feature of AMPA. Additionally, the γ-group is stabilized by a water bridged hydrogen bonding to Thr686. Also the co-crystallized interfacial water molecules differ in number and to a small amount in

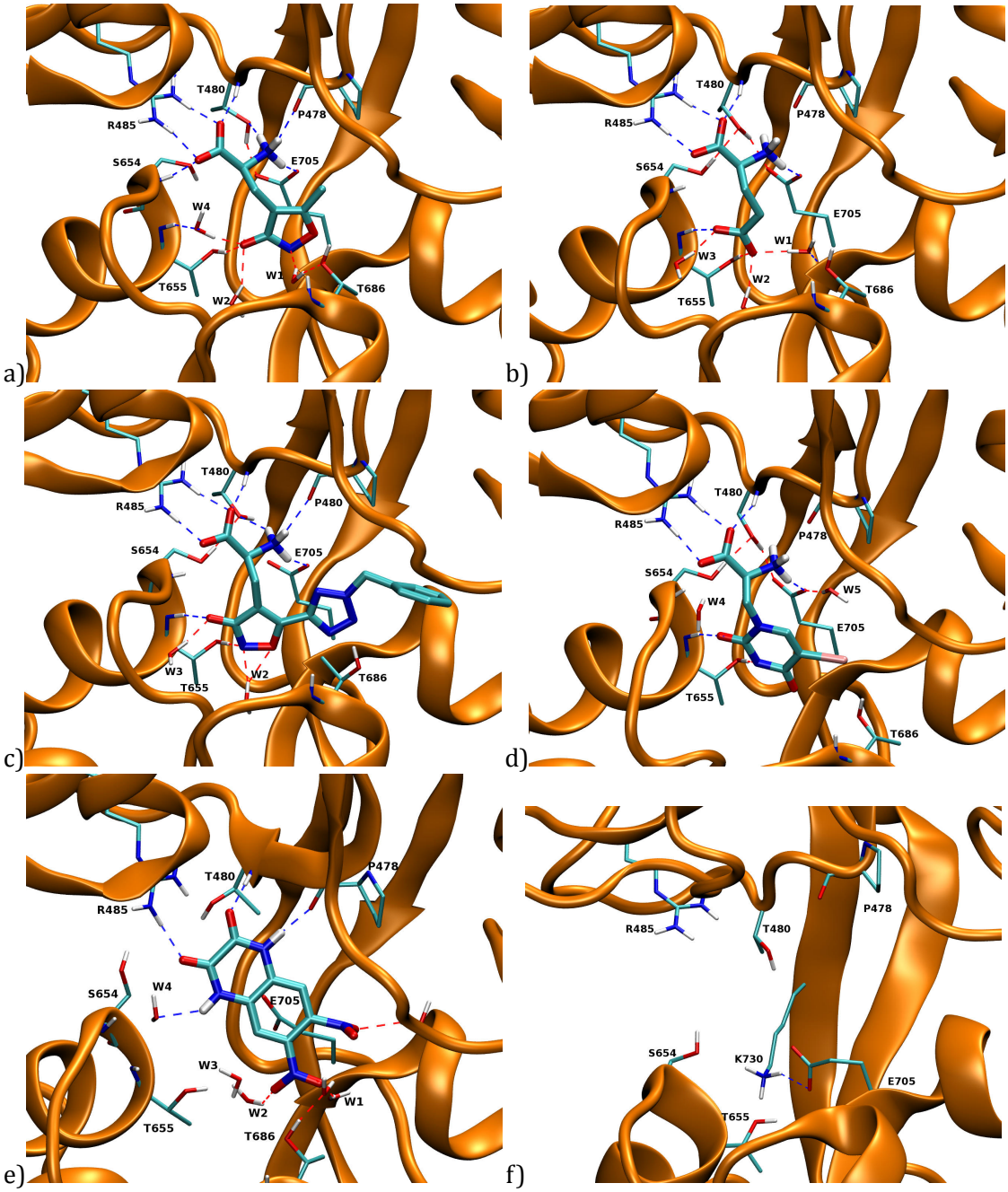


Figure 4.2 – Showing the binding pocket and the binding motives for the different ligands. a) AMPA; b) glutamate; c) 2-BnTetAMPA; d) iodo-willardiine; e) DNQX; f) APO form

location. For a more detailed comparison of the X-ray structures, see ref. 190.

AMPA

During one microsecond of MD simulation, no major changes occur. The ligand binding pocket stays closed and the overall shape of the protein is preserved, which is illustrated by the RMSD of all backbone atoms (>0.2 nm; see Fig. B.1a). The RMSD of the backbone of domain 1 is around 0.175 nm, while the value for the domain 2 is nearly all the time

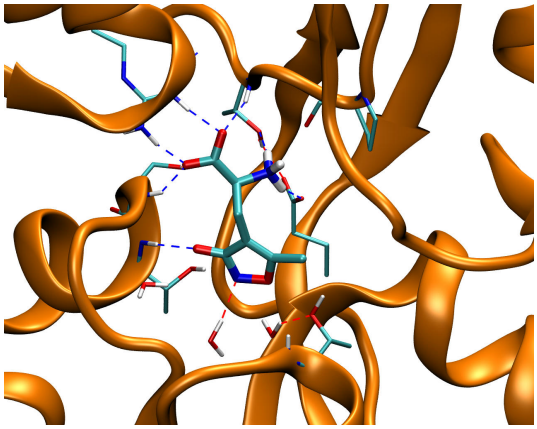


Figure 4.3 – The rearranged binding pocket with a bound AMPA after water 4 moved out of the pocket.

below 0.15 nm. This implies a high stability of the tertiary structure of the two domains. The ligand stays very stable within the pocket and shows no tendency to rearrange (RMSD < 0.15 nm). The main binding motives are all preserved. It has to be mentioned that water 4 (see Fig. 4.2 a) cannot be stabilized in this position and moves out of the binding pocket within 1 ns. Thus, the ligand and binding pocket reorient slightly, which results in a direct hydrogen bond between AMPA and the backbone NH of Thr655 (see. Fig. 4.3). Accordingly, the more sophisticated water models TIP4P[184] and TIP5P[191] were tested, with respect to the location of water 4. In both cases water 4 shows the same behavior like the TIP3P water molecule. A possible explanation for the water instability could be, that this water is highly polarized between the two groups. Since the tested water models are designed for unpolarizable force fields, the point charges on this water molecule is not adjusted for this purpose and result in wrong binding energies for this water molecule. To test this hypothesis, one has to examine polarizable forcefields or use QM/MM. However, the latter approach is too computational demanding to reach the needed simulation times.

Glutamate

The MD simulation of the LBD monomer with a glutamate in the binding pocket shows severe changes during the 1 μs. Up to 600 ns the ligand binding domain and the bound ligand keep their overall structure. After 600 ns the ligand binding domain opens. While the binding motives were retained to domain 1, the γ-carboxyl of the glutamate is hydrogen bonded to Ser654 and a intra-domain hydrogen bond between Glu705 and Thr655 is built (Fig. 4.4a). After additional 10 ns the ligand rotates out of pocket and the γ-carboxyl is surrounded by water molecules (Fig. 4.4b). After approximately 680 ns the ligand moves into the bulk water. The ligand movs back into the pocket after 760 ns and binds to domain 1, as observed before. The γ-carboxyl group is hydrogen bonded to Ser654 and Lys730, where the latter has built a salt bridge to Glu705 (Fig. 4.4c). This salt bridge was also observed in the crystal structure of the APO state and was suspected to stabilize the open

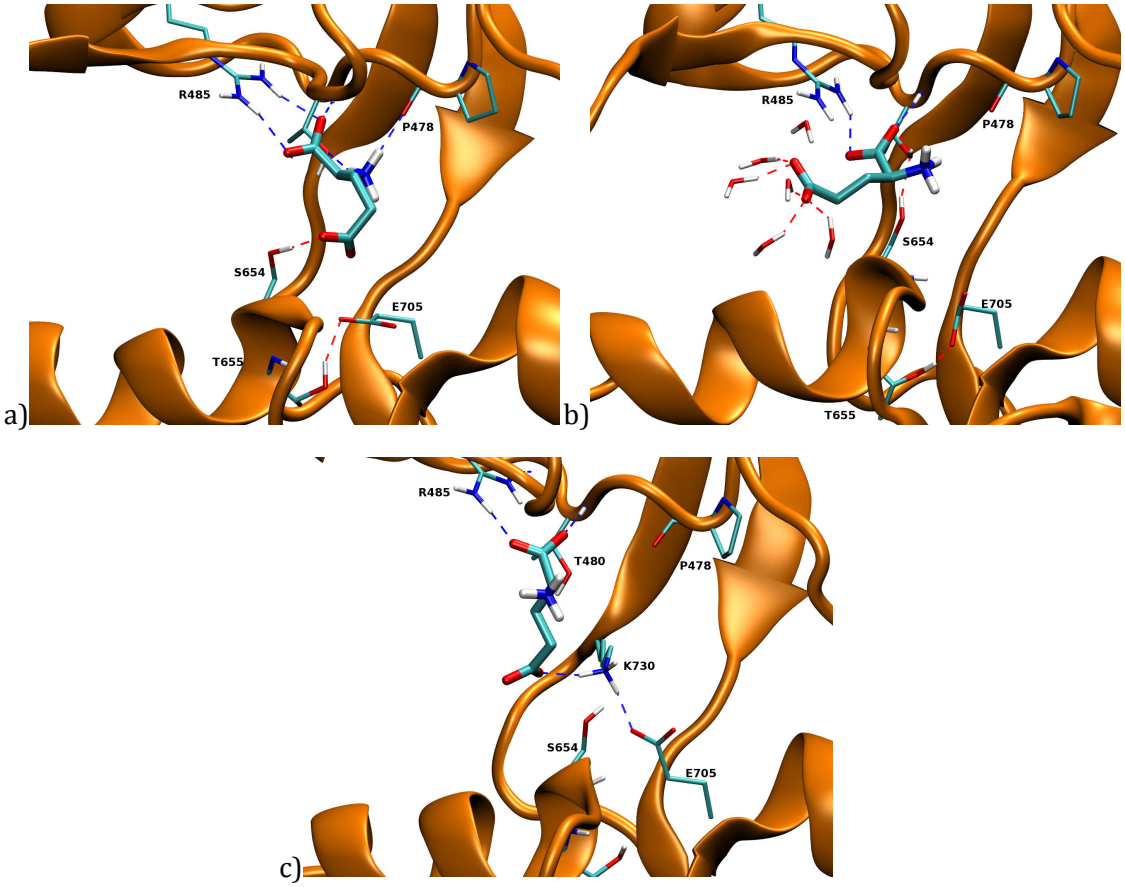


Figure 4.4 – Three snapshots along the MD, describing the unbinding/binding process

form of the LBD. After 920 ns the ligand moved again into the bulk water and stays there for the rest of the simulation. While the RMSDs of domain 1 and 2 are not influenced by the unbinding process, the RMSD of the whole backbone shows a strong correlation with the unbinding of the ligand (Fig. B.1b). This finding suggests that the tertiary structures of the two domains are very rigid and do not change due to the LBD opening.

2-BnTetAMPA and Iodo-willardiine

The structures of the LBD with a bound 2-BnTetAMPA or iodo-willardiine (IW) shows a high stability during the 1 μs. The RMSD of the ligands is around 0.1 nm, while the RMSDs of the backbone atoms is below 0.2 nm (Fig. B.1c+d). In both cases on can observe peaks of the RMSD of domain 2, which can be related to loop fluctuations.

DNQX

The simulation of the LBD with the bound antagonist DNQX depict a more flexible ligand binding domain. During the 1 μs the system fluctuates between two regions of the conformational space which is also indicated by the RMSD of the ligand and the protein backbone of the whole protein (Fig. B.2a). The first region samples the starting structure, while in the second region the LBD is more opened, comparable to the APO state. Thus, the hydro-

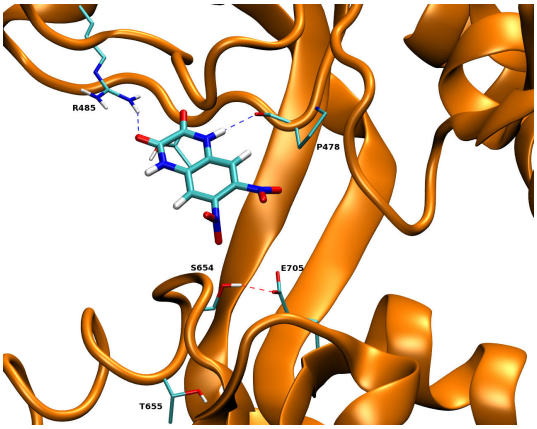


Figure 4.5 – The binding pocket with a bound DNQX during MD. The ligand shows no interaction with domain 2

gen bonds between the ligand and domain 2 are not preserved. Moreover, the hydrogen bond between Arg485 and the two carbonyl groups of DNQX is much more flexible than the salt bridge in the case of the other investigated ligands, but stays intact over the whole simulation time (Fig. 4.5).

APO

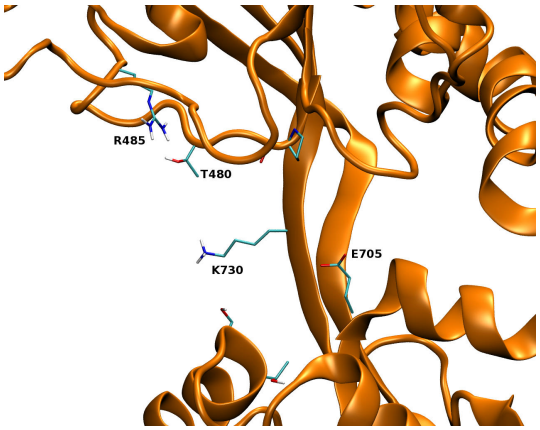


Figure 4.6 – An extreme of the LBD opening along the MD

As expected, the simulation of the APO state depicts the highest flexibility of the LBD, since the interaction between the two domains that are bridged by a ligand are nonexistent. This is reflected in a high RMSD of the protein backbone up to 0.5 nm (Fig. B.2b). Furthermore, the RMSD of the backbone of domain 1 increased steadily, which can be explained by the loops accommodating the binding positions for the ligands. These loops seem to be stabilized by a bound ligand. Even though the RMSD value does not exceed 0.3 nm. It has to be mentioned that the salt bridge between Glu705 and Lys730, which was proposed to stabilize the APO state,[72] is only stable for the first 200 ns . After that the positively charged sidechain of Lys730 orients into the bulk water and stays there for the rest

of the simulation time (Fig. 4.6).

4.1.3 Structural Properties

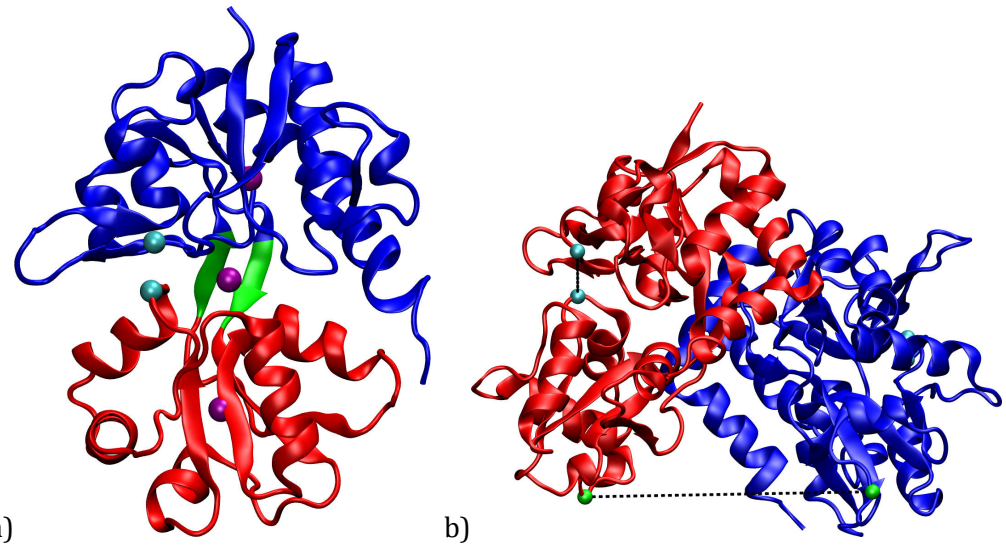


Figure 4.7 – a) Separation of the LBD into a domain1 (blue), domain 2 (red) and a hinge region (green). The purple spheres depict the COMs for calculating the angle. The cyan spheres depict the COM of G451 and S651. b) LBD dimer with different colored monomers. The cyan spheres depict the COM of G451 and S651 within a monomer. The green spheres depict the COM of P632.

To get a deeper understanding of the motions and structures of the ligand binding domain, several structural properties are measured along the trajectories. To describe the clamshell motion of the LBD, we monitored two properties: The angle between the center of mass (COM) of domain 1 (defined as residues 394-495), hinge region (residues 496-499, 729-731) and domain 2 (residues 732-771) and the distance between the center of mass of the G451 and S652 backbone (see Figure 4.7a). The influence of these two reaction coordinates on the opening of the channel is estimated by the distance between the two linker peptides P632 that are inserted into the LBD to replace the transmembrane sequence of the receptor. Since only the monomer is simulated and not the dimer, the second monomer is projected on the monomer trajectories in such a way that the two domains 1 have the same alignment like in the crystal structure (see Fig. 4.7b).

The general idea is that the increase of the angle or distance between Gly451 and Ser652 leads to a decrease of the Pro632 distance. In Fig. 4.8 these two distances and the angle are plotted for all simulations. One can see that the distance between Gly451 and Ser652 is proportional to the angle between the two domains. As expected, both are anti-proportional to the Pro632 distance. This anti-proportionality is more distinct between the two distances. In case of iodo-willardiine the fluctuations of the angle are not represented in the Pro632 distance. In general, the angle between the two domain shows a higher noise level, which might be explained by the flexible loops that also contribute to

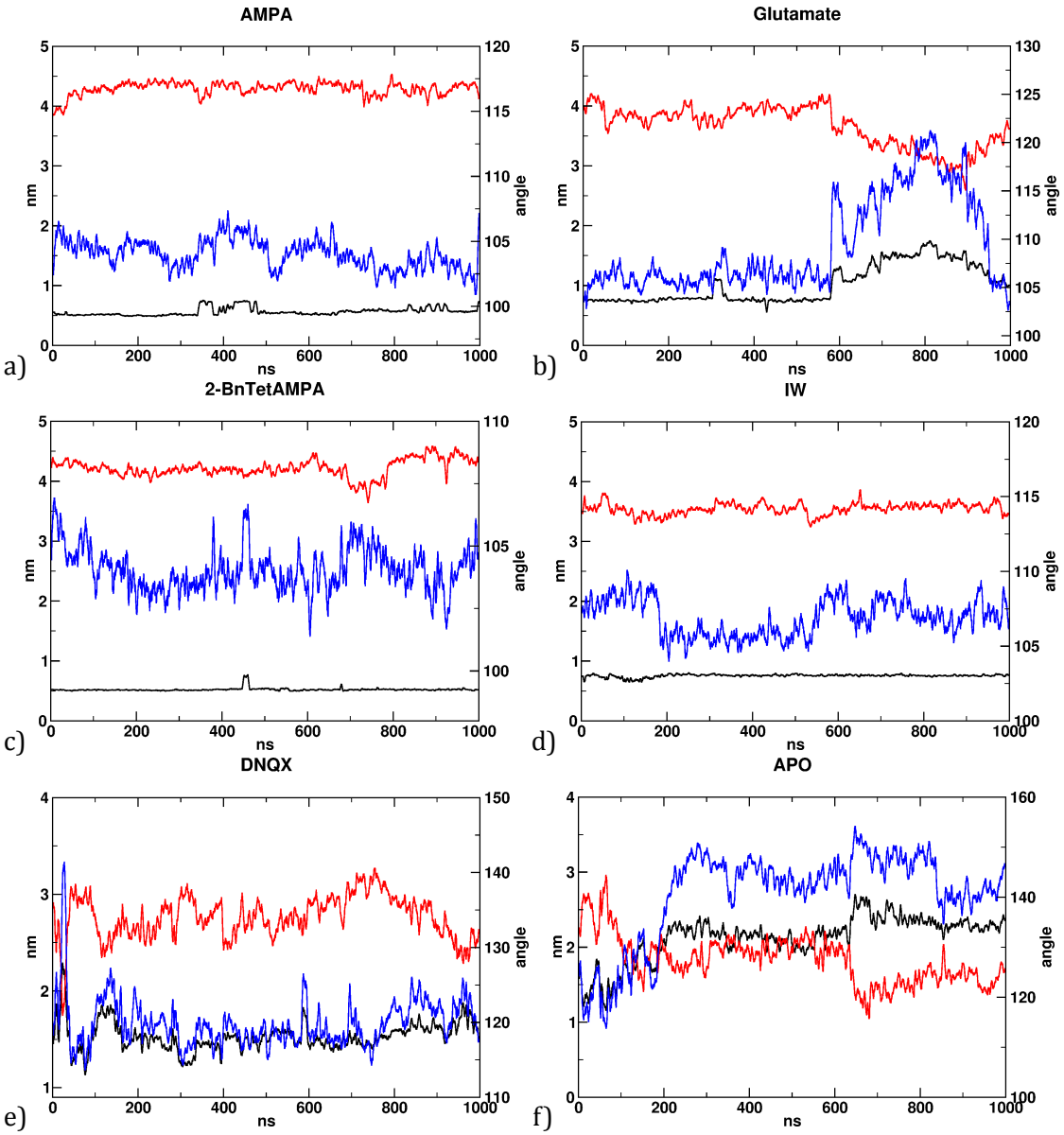


Figure 4.8 – The structural parameters along the trajectories. Black: Distance between G451 and S651; Red: Distance between both P632; Blue: Angle between the two domains.

the center of mass. This is further supported by the plots of the P632 distance against the Gly451-Ser654 distance and the angle (see Fig. 4.9). The linear dependency between the two distances is more distinct than for the angle and the distance between the two Pro632. Linear regression leads to a coefficient of determination R^2 of 0.916 and 0.856 for Gly451-Ser651/Pro632 and angle/Pro632, respectively. It is quite surprising, that the single distance at the mouth of the clamshell seems to be a better description of the ion channel opening than the angle between the domains. Since the angle seems to be a better match for the proposed clamshell motion, based on the available crystal structures, the idea occurs that the clamshell motion is not the only motion that triggers the channel opening. These additional motions might be also monitored by the distance between Gly451 and

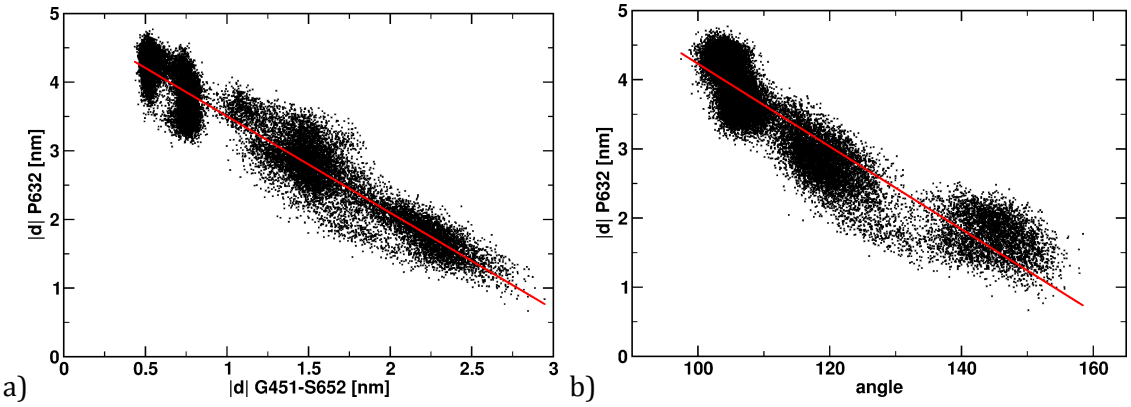


Figure 4.9 – Correlation between the different structural parameters

Ser651.

4.1.4 Principal Component Analysis

In the last section, the analysis of the structural properties along the trajectories raise the idea that the clamshell motion does not represent the whole motion of domain 2 in respect to domain 1. This was also suggested by experiments proposing that a completely closed LBD can lead to partial agonism, i.e. the ion channel is not opened to full extend.[79--82] Furthermore, rigid body principal component analysis on published X-ray structures depicted additional motions of domain 2 in respect to domain 1.[83] To expose additional motions within our molecular dynamics we performed principal component analyses (PCA). This technique removes the fast degrees of freedom of the protein and depicts the global and correlated motions of the protein. For reviews on PCA applied to proteins, see refs. 192,193. The PCAs were performed on the last 800 ns of each of the MD simulations. The first 200 ns were discarded due to equilibration of the system. During the analysis only the backbone atoms were considered. Already the first six eigenvectors contain about 80 % of the fluctuations. We tested these eigenvectors for their cosine content, since those which resemble cosines with the number of period equal to the half of their eigenvalue index, might not describe relevant motions, but random diffusion on a flat part of the free-energy landscape.[194] None of the 36 (the first six of every MD) eigenvectors have a mentionable cosine content. The overlap of the six eigenvectors between the different MDs are shown in Table 4.1. An overlap of one indicates a perfect match of two sets of eigenvectors, while an overlap of zero indicate two orthogonal set of eigenvectors. In most of the shown cases the overlap is between 0.3 and 0.4. Only the overlaps between the APO state, DNQX and glutamate are about 0.5. This might be explained by considering the protein to be in different free energy landscape regions with a differently shaped surface, which leads to different relevant motions. In case of a bound glutamate, DNQX or the APO state, the high flexibility of the protein accompanied with LBD opening and closing. The regions are connected or have a similar shape. Additionally, the last 800 ns of every MD were merged into one

Table 4.1 – Overlap of the different sets of eigenvectors

	AMPA	2-BnTetAMPA	glutamate	IW	DNQX	APO	merged
AMPA	1	0.338	0.385	0.397	0.392	0.381	0.348
2-BnTetAMPA	0.338	1	0.285	0.325	0.321	0.313	0.322
glutamate	0.385	0.285	1	0.341	0.476	0.461	0.460
IW	0.397	0.325	0.341	1	0.316	0.302	0.299
DNQX	0.392	0.321	0.476	0.316	1	0.514	0.501
APO	0.381	0.313	0.461	0.302	0.514	1	0.835
merged	0.248	0.322	0.460	0.299	0.501	0.835	1

big data set and we performed a PCA, which should give a good impression of the opening/closing process of the LBD. The calculated overlap with the single MDs is also shown in Table 4.1. First of all, the overlap between the APO simulation and the merged data set is quite big (0.835), which indicates that the principal motions of the APO state dynamics already captures most the opening/closing motions. This is further supported by the very similar overlaps between the APO state and the merged data set with the other simulations. The first eigenvector of the merged data set describes the already proposed clamshell motion. In Fig. 4.10a the six MDs are projected on the first eigenvector. Comparing Fig. 4.10a

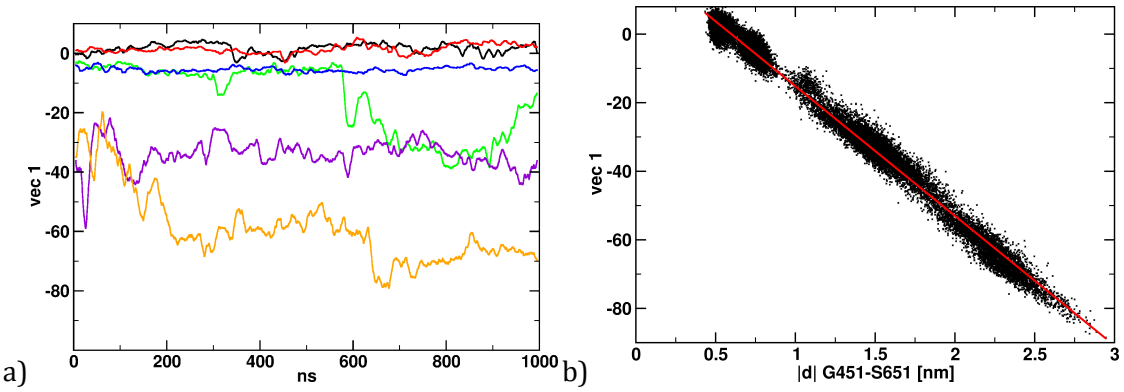


Figure 4.10 – a) Projection of the MDs on the first eigenvector. Black (AMPA), red (2-BnTetAMPA), green (glutamate), blue (IW), purple (DNQX), yellow (APO). b) Correlation between the G451-S651 distance and the motion along the first eigenvector.

and 4.8, one can see the anti-proportionality of the Gly451-Ser651 distance and the projection on the first eigenvector. The coefficient of determination for these two parameters is R^2=0.988 (see Fig. 4.10b). The projection of the trajectories along the second and third eigenvector is less significant. As expected the fluctuations along these eigenvectors is much smaller. Additionally, there is a connection between the quality of the ligand, i.e. a bound agonist like AMPA, reduce the fluctuations along the eigenvectors.

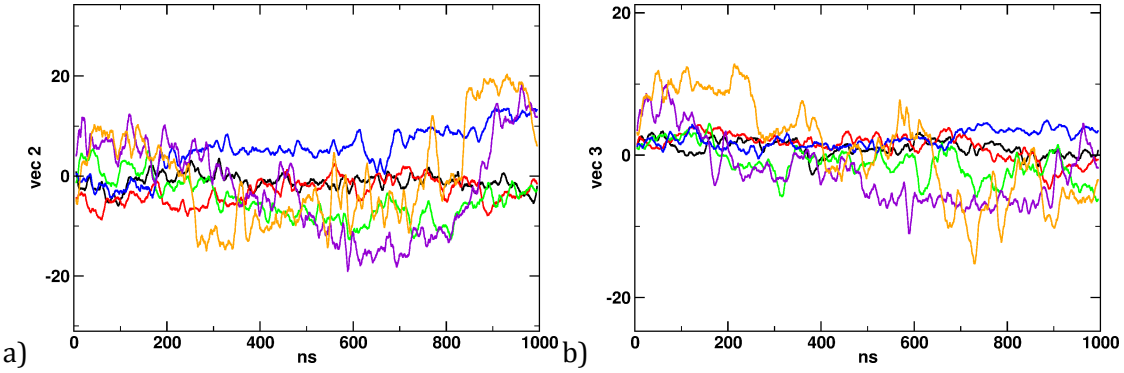


Figure 4.11 – a) Projection of the MDs on the second eigenvector. b) Projection of the MDs on the third eigenvector. Black (AMPA), red (2-BnTetAMPA), green (glutamate), blue (IW), purple (DNQX), yellow (APO).

Rigid body PCA

In order to get principal motions that only describe the LBD opening and closing and neglect all intra-domain motions, we performed also a rigid body PCA, as in a previous study.[83] Therefor, we used the simulations and removed afterwards the internal motions of the two domains. This was achieved by superimposing their optimized structures onto the corresponding residues of every MD step. For all protein-ligand complexes as well as the apo-protein, three dominant eigenvectors termed clamshell, twisting and rocking motion, in accordance to ref. 83, were obtained (see Fig. 4.12, arrow representation). These three derived eigenvectors will be used in section 4.2 to calculate the energetics of the LBD opening and closing. It is noteworthy, that the first eigenvector of both PCA approaches is very similar, while the second and third are slightly different, which is caused by the flexibility within the domains using the standard MD steps for the PCA. Nevertheless, also these eigenvectors show a high similarity. This implies that the introduced simplification of the rigid body approach is valid.

4.1.5 LBD Dimer

Additionally to the monomer simulations, we performed MD simulations of the LBD dimer. This was done to check our data received by the monomer simulations. Especially the distance between the Pro632 of the simulated monomer and the projected one may leads to artifacts, because interaction between both LBDs were neglected, which might have an influence on the behavior of the LBD. Due to computational limitations we performed only simulation for AMPA, 2-BnTetAMPA, DNQX and the APO state. In case of a bound ligand, both LBDs contain the same cofactor. Also the simulation time was reduced to 300 ns in respect to the computational cost.

In comparison to the monomer simulation, no essential differences in dynamics could be observed. The interaction between the LBD and the ligands are assumably comparable. The distances between Gly451 and Ser651 are like in the monomer simulations except for

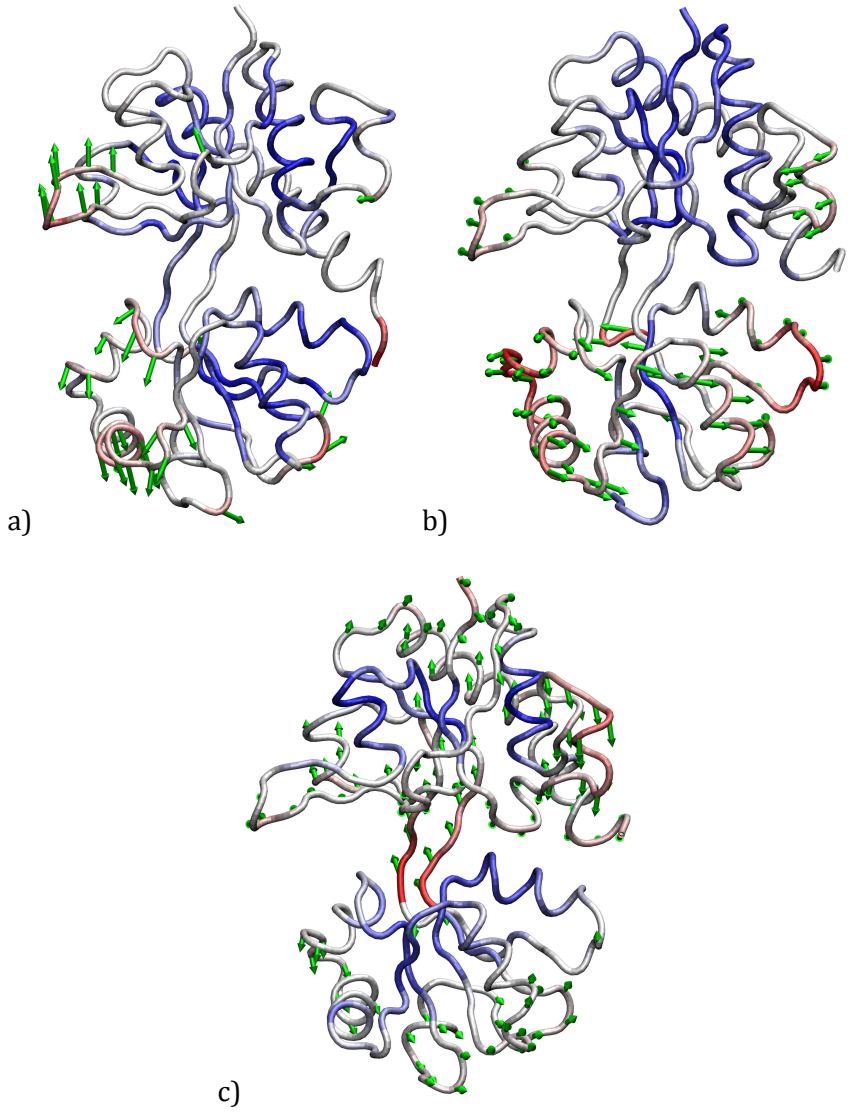


Figure 4.12 – An arrow representation of the first three principle motions. a) clamshell motion; b) twisting motion; c) rocking motion. For better visualization movies of these motion are downloadable in the supporting materials of ref. 195

the APO form of the protein (see Fig. 4.13). The G451-S651 distances in both monomers of the dimer simulation are notably smaller than in the monomer simulation, because the wide opening of the LDB monomer would lead to a collision of the two domain 2 in the dimer simulation. Additionally, the RMSD between the monomer and dimer simulation have been calculated (see Table 4.2). The small RMSD values for AMPA, BnTetAMPA and DNQX shows that the overall structure does not suffer from simulating only a monomer. In case of the APO form, the high RMSD can be explained by the huge LBD opening during the monomer simulation. A more severe difference is the obviously reduced correlation (R^2=0.896) between the G451-S651 distance and the P632 distance. The explanation for this reduced dependency is not structural changes within the single monomers, but in the

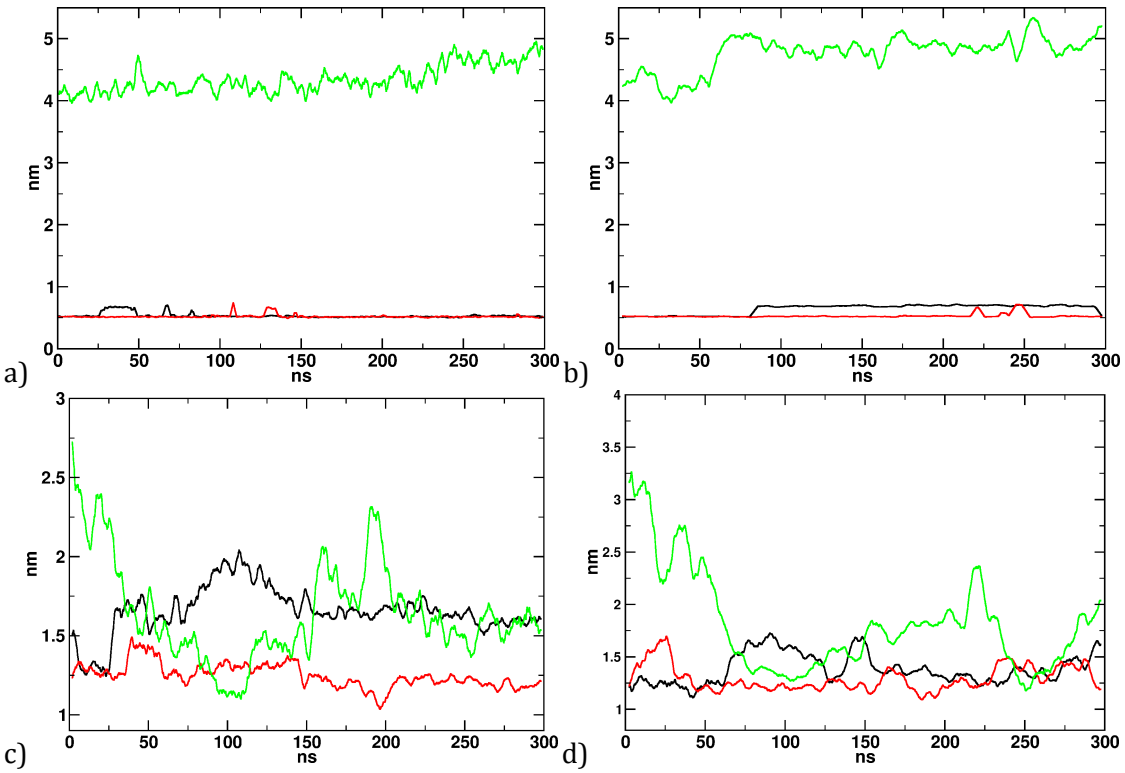


Figure 4.13 – Strcutural properties along the dimer trajectories. a) AMPA; b) 2-BnTetAMPA; c) DNQX; d) APO. Black curve: G451-S651 distance monomer A; Red curve: G451-S651 distance monomer B; Green: P632 distance.

Table 4.2 – RMSD between the monomer and dimer simulation. For calculation of the RMSD the end structures of the MD simulation have been used. All values in nm

	Dimer chain A	Dimer chain B
AMPA monomer	0.19	0.18
BnTetAMPA monomer	0.08	0.10
DNQX monomer	0.19	0.17
APO monomer	0.39	0.35

motion of the monomers to each other. A PCA of the dimer dynamics depict two mayor principal modes, which cover already over 80 % of all fluctuations. The corresponding motions are located only between the monomers and not within. The first principal mode describes a shearing motion. The second mode describes the motion of a clothespin. A schematic picture of both motion are shown in Fig. 4.14. Thus, an analyze of the correlation between the interdomain motions within a LBD monomer and the channel opening is hard to achieve.

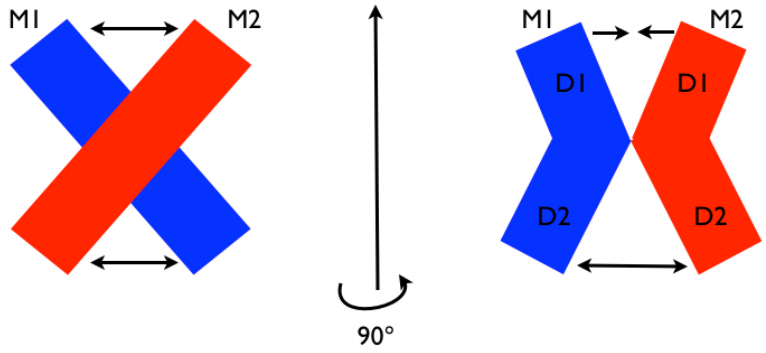


Figure 4.14 – A schematic representation of the first two principal motions of the LBD dimers. Left: The shearing motion; Right: clothespin motion.

4.1.6 Dicussion

During our extensively long MDs of one microsecond, we could observe different behaviors of the ligands and the binding domain, as well as the interaction between the two. While AMPA and 2-BnTetAMPA show very strong binding to the protein, glutamate shows a surprising unbinding of the agonist after 600 ns, which is not unrealistic at this time scales, but have not been observed in a free MD simulation before. Although, AMPA exposes strong interaction with the binding pocket, it was not possible to stabilize the interfacial water molecule that bridges the interaction between the ligand and the backbone NH. This might be caused by missing polarization within the standard force fields, which should be tested by the usage of a polarizable force field or a QM/MM approach. Nevertheless, only small rearrangements occurred and the overall structure was kept. The only partial agonist we have tested, iodo-willardiine, also preserves all interactions with the protein and the fluctuations of the binding pocket are negligible. This picture changes for the antagonist DNQX and the APO form of the LBD. Besides the more opened cleft of the LBD, also the fluctuations are getting stronger, which are caused by the reduced or completely neglected interactions within the binding pocket. The RMSDs present a picture of two fairly rigid domains which move relatively to each other. The occurrence of these motions is also depicted by the RMSD of the whole protein, but not the type of motion. Hence, for a deeper understanding of ligand induced motions in the ligand binding domain, one has to define reaction coordinates, which describes the opening and other motions of the LBD in an appropriate way. During our study we found the distance between Gly451 and Ser651 to be a very good description of the clamshell motion which was depicted by principal component analysis. Furthermore, this distance shows a high correlation with the distance between the linker peptides Pro632, which replace the transmembrane sequences of the protein, building the ion channel. These findings support the experimental results, that suggest the clamshell motion to be the motion which determines the ion channel opening. Nevertheless, we also observed deviation from this simple picture. For example glutamate and iodo-willardiine, induces the same clamshell closure, but the distance between the linkers differs by 5 Å.

Also the PCA distinguished two additional principal motions which might influence the ion channel opening. In case of glutamate and IW, no connection between the second or third eigenvector and the P632 distance could be observed. These findings are based on monomer simulation, which neglects the interaction within the LBD-dimer. An obvious result of this simplification is the excessive opening of the APO state during the MD simulation. The wide opening would lead to a clash of the two monomers. To detect less distinct biases induced by this simplification, we performed MD simulation of the dimer. The comparison of the overall structures depicted no unphysical distortion within the monomer simulation. However, a reduced correlation between the Gly451-Ser651 distance and the P632 distance was observed. This is not caused by different motions or structural features within the monomers, but by two predominant motions. These principal motions show the picture of two fairly rigid monomer moving relatively to each other (see Fig. 4.14). It is disputable if these motions can be observed in the native receptor. First, the LBDs are connected to the ion channel and a huge N-terminal domain, which should reduce these motions. Second, the 4 monomers are not arranged in parallel, but are interlaced, which further suppresses these motions. Since the introduced errors by simulating a monomer are limited to the wide opened form of the LBD, we are confident that our monomer simplification is valid in the range of normal LBD opening.

4.2 Energetics of the LBD

The energetics of the LBD opening are important for two reasons: (i) the free energy profile of the opening reveals how tight the cleft is closed and depicts conformational subspaces along the opening/closing path; (ii) the free energy difference between the open and closed state with a bound ligand is the first step to calculate ΔG_{bind} of the corresponding ligand, which is the decisive property in the development of new ligands. The free energy profiles of the transition from closed to opened state were calculated by using 1-D Umbrella Sampling and 3-D Metadynamics. Due to the extremely high computational cost, only the evaluation of AMPA, 2-BnTetAMPA and the APO-state could be achieved.

4.2.1 Opening/Closing transition via Umbrella Sampling

Due to the computational demand of Umbrella Sampling only one reaction coordinate was chosen to describe the cleft opening. In the sections 4.1.3 and 4.1.4 we established that the distance between Gly451 and Ser651 is a valid reaction coordinate to describe this domain motion. Hence, we sampled this reaction coordinate from 0.5 nm to 1.4 nm, which includes the distances of the X-ray structures of the AMPA-bound state (0.54 nm) and the APO-state (1.3 nm). The starting conformations were generated by a pulling simulation, with a very small pulling speed (10^{-5}nm/ps) to avoid unphysical distortions. Based on these starting geometries, we used 20 US windows spaced equidistant at 0.05 nm intervals, with 500 kcal mol^{-1} nm^{-2} biasing force constants. For every window a MD simulation of 300 ns length was

conducted. For analysis, the first 20 ns in every window are neglected due to equilibration of the system. Convergence was determined by histogram overlap and by batch averaging over 50 ns simulation intervals.

APO form

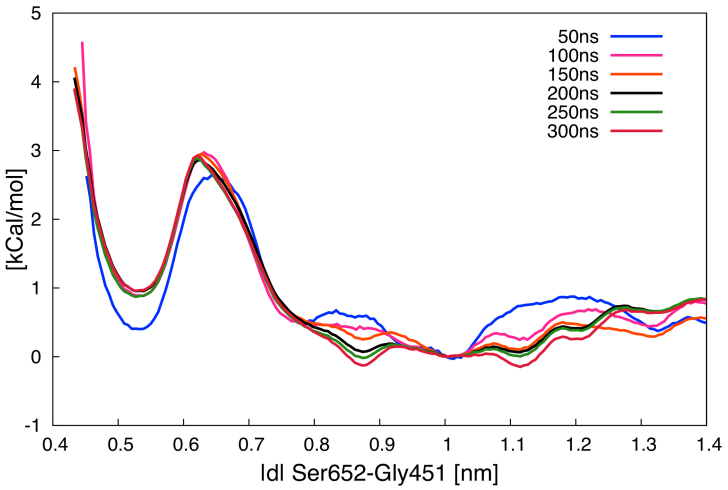


Figure 4.15 – PMF for the opening of the LBD in the APO form

For the opening process of the ligand free binding domain, a clear local energy minimum is located at 0.53 nm (+1.10 kcal/mol), which corresponds to the closed form of the receptor (see Fig. 4.15). To escape this minimum a barrier of about 2 kcal/mol has to be overcome, which leads to a broad free energy plateau between 1.0 nm and 1.4 nm instead of a clear minimum. The free energy profile depicts that the open form is more stable by ca. 1 kcal/mol, indicating that the open form is the preferred conformation of the free receptor. The shape of the derived potential of mean force (PMF) remains consistent after 150 ns, but the energy plateau at the open state configuration is subject to significant noise and slow convergence.

AMPA

The PMF for the opening of the AMPA complex shows a broad energy minimum at 0.51 nm. A second local minimum is separated by an energy barrier of ca. 5 kcal/mol. The second minimum is more shallow, is located at 0.75 nm and is 3.09 kcal/mol higher than the first minimum. The second minimum leads to a flattened PMF from 1.0 nm to 1.4 nm and includes a barely noticeable shallow minimum between 1.1 and 1.2 nm. This third minimum is identified as the open state. As one can see in Fig. 4.16, extensive sampling is needed for convergence. Especially the region around the open state needs a lot of sampling, since the PMF curve after 150 ns still show significant higher energies. The energy differences generated by the last 50 ns are small enough to consider the calculation to be converged. This leads to a free energy to open the AMPA complex of +6.18 kcal/mol. To assign these three PMF minima to receptor configuration, RMSDs between the minima and different

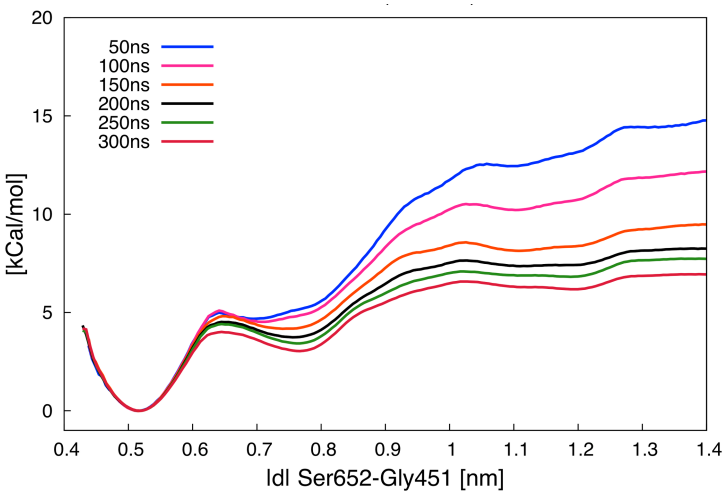


Figure 4.16 – PMF for the opening of the LBD with a bound AMPA

optimized X-ray structures were calculated. The RMSDs shown in Table 4.3 indicate that the first, second and third minimum corresponds to, respectively, the closed, semi-closed and open form of the LBD.

Table 4.3 – RMSD-values of receptor conformations and observed PMF minima.The X-ray crystal structures used for the closed, half-closed and open states had pdb identifiers 1FTM/2P2A, 1MQG and 1FTO, respectively. A clear correspondence of the observed minima and the receptor conformational states is found.

	AMPA			BTA		
Receptor	1st min	2nd min	3rd min	1st min	2nd min	3rd min
closed	0.11	0.14	0.23	0.13	0.15	0.29
half-closed	0.20	0.13	0.20	0.18	0.13	0.26
open	0.31	0.26	0.17	0.29	0.25	0.15

2-BnTetAMPA

The PMF of the 2-BnTetAMPA complex shows also a clear minimum at 0.53 nm, which corresponds to the closed state. This is also connected to a second minimum at 0.76 nm, which represents the semi-closed state. Compared the AMPA complex the energy barrier between the minima is much lower (2 kcal/mol) and also the height of the second minimum is decreased (1.35 kcal/mol) in comparison to the AMPA complex. Beyond the second minimum, the PMF rises significantly, until a plateau around 1.12 nm (+8.69 kcal/mol) is reached. Certainly, this third minimum, which corresponds to the open state, is hard to distinguish. This problem might be eased by longer simulation times, which could also lead to lower free energy for the open state. Nevertheless, the consecutive curves justifies the assumed convergence.

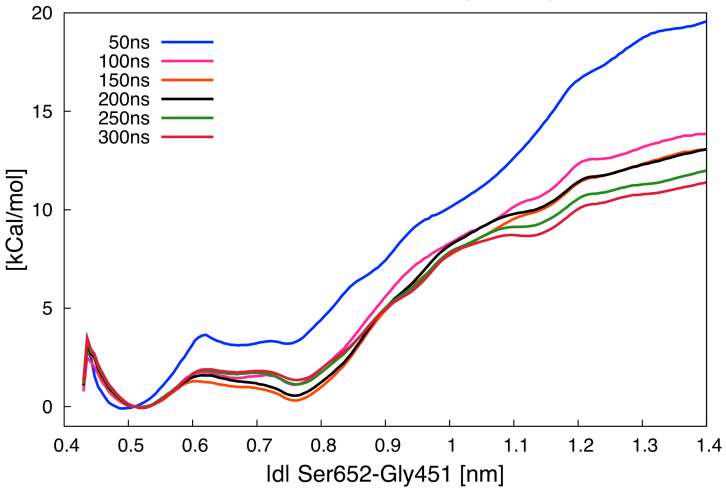


Figure 4.17 – PMF for the opening of the LBD with a bound 2-BnTetAMPA

4.2.2 Opening/Closing transition via Metadynamics

In the last section the extreme computational cost of Umbrella Sampling was shown (6μs simulation time per PMF). An alternative for this expensive method might be Metadynamics. Therefore, free energy surfaces of the LBD opening for the AMPA and 2-BnTetAMPA bound complexes have been conducted. An advantage of this method is the possibility to capture more than one dimension, which is in general also feasible using Umbrella Sampling, but only with high computational cost. As reaction coordinates, called collective variable in Metadynamics (CV), the three dominant eigenvectors from the rigid body PCA (see 4.1.4) were used. From initial analysis of 50 ns simulation with a biasing function built up by Gaussian hills with a height of 0.5 kcal/mol which were added every picosecond, significant convergence problem occurred. The complex explored the phase space along the three eigenvectors, but the system never reached the initial state which is indispensable for the convergence of a Metadynamics calculation. Moreover, we monitored strong unphysical distortions in the protein structure. This leads to the assumption, that the biasing potential was built up to quickly and that the system could not equilibrate accordingly to the changed potential surface. Hence, the protocol to build the biasing potential was refined, by adding smaller hills (0.2 kcal/mol) only every 10 ps for 200 ns simulation. After 200 ns the size of the Gaussians was reduced to 0.05 kcal/mol to smooth the resulting free energy surface and reduce the error of the method. For both complexes a 500 ns Metadynamics simulation were conducted. In case of 2-BnTetAMPA, the simulation was extended by 100 ns to ensure convergence. The free energy surfaces depict that the clamshell and the twisting mode are important for opening of both complexes. The third collective variable is less important for the AMPA complex since all minima and saddle point lay in approximately the same range of CV 3. In contrast, for 2-BnTetAMPA, the rocking motion is also an important part of the opening motion. The potential energy surface (PES) along the

three eigenvectors depicts a two-state system with minima that correspond to the open and closed state of the ligand binding domain. The free energy difference to open the LBD is +4.30 kcal/mol and 4.86 kcal/mol for AMPA and 2-BnTetAMPA, respectively. This result is in reasonable agreement with the performed Umbrella Samplings for the same process. The convergence of the simulations was assured by the fact that several transitions be-

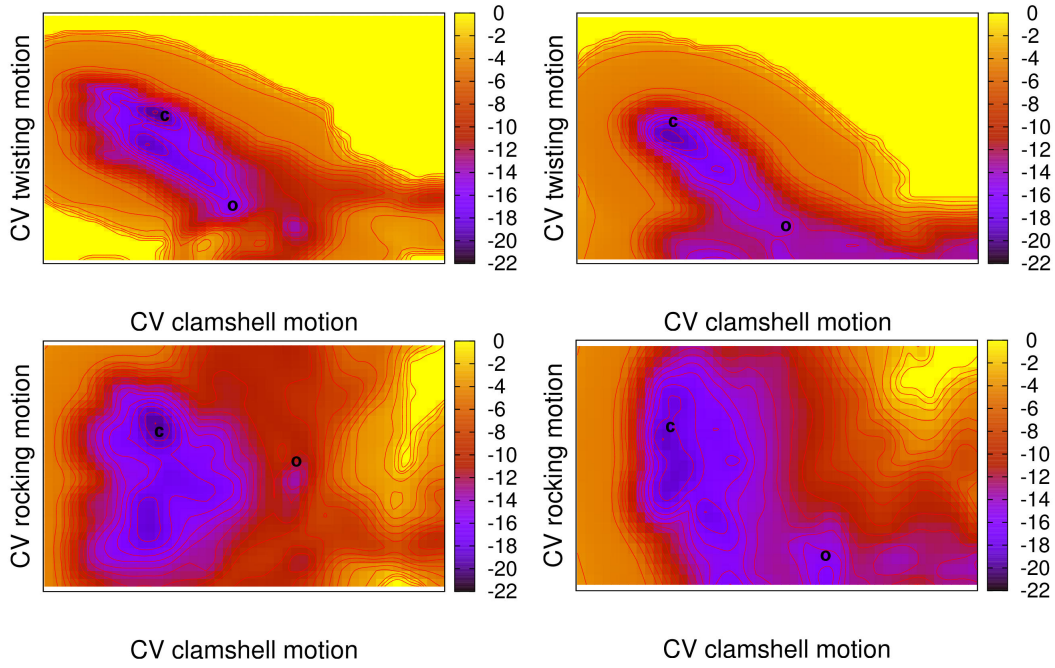


Figure 4.18 – Potential energy surface projected onto the first, second and third collective variables for the opening/closing transition of the AMPA (top) and 2-BnTetAMPA (bottom) complex

tween the two states occurred. This results leads to the conclusion that Metadynamics is capable to determine the free energy profile of the LBD with a comparable accuracy to Umbrella Sampling, though with a much lower computational cost.

4.2.3 Ligand binding/dissociation via Umbrella Sampling

The second step to describe the ligand binding to the LBD involves the dissociation of the ligand from the open form of the receptor. This was conducted only by Umbrella Sampling. Convergence of Metadynamics requires several transition between both states. This could not be achieved, because after the dissociation of the ligand, it stayed in the bulk water for several hundreds of nano second and did not found back into the binding pocket. The reaction coordinate for the dissociation was defined as a center-of-mass distance between the ligand and a group of amino acids within the binding site. Specifically, the backbone atoms of residues 399, 448-452, 462-464,476-481 and 705 were selected, resulting in a center-of-mass that ensures that the ligand is pushed out of the binding site when moving along the reaction coordinate. The ligand binding domain was aligned along the z-axis and

the reaction coordinate distance was measured in the xy-plane, bisecting the LBD between domain 1 and 2. The planar projection of the ligand position was simply used to define the position along the reaction coordinate. The ligand was fully free to move in all three directions, but the biasing potential acted only at the xy-projection of the ligand. The LBD rotation and translation were removed every 10 MD steps. We used 45 US windows along the reaction coordinate, spaced 0.075 nm apart from a starting distance of 0 nm to 3.5 nm. The biasing potential differed from 600 to 1000 kcal mol^{-1} nm^{-2}.

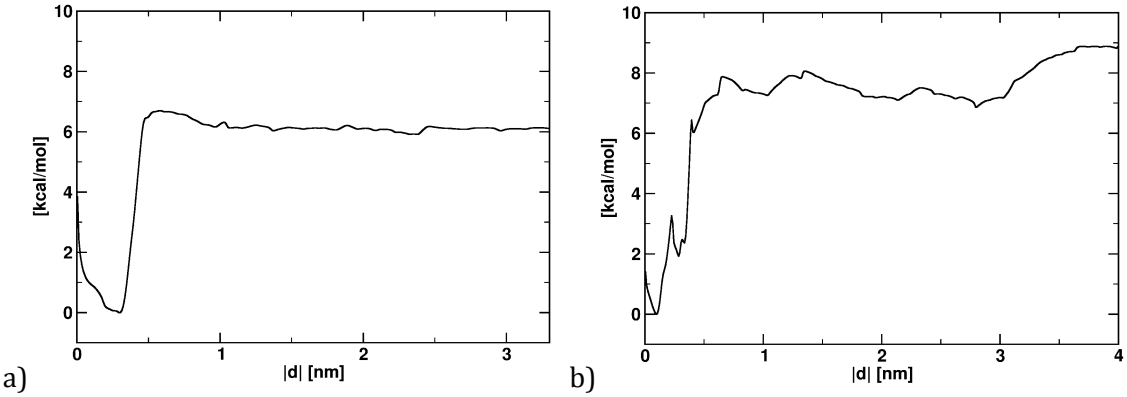


Figure 4.19 – PMF free energy curves for the removal of the ligand from the LBD. Left: AMPA; Right: 2-BnTetAMPA

The free energy curves for the dissociation into the bulk water show fairly comparable profiles for both ligands with a minimum at short distances and a high energy barrier to overcome before the ligands dissociate from the binding pocket. The difference in the positions of the minima is caused by different center-of-mass positions for both ligands. The barrier occurs early in the unbinding process and leads directly into a nearly flat energy surface. This indicates that the LBD surface does not funnel the ligands to their binding position, instead the ligands have to diffuse randomly into the LBD to reach a position close to the minimum energy binding mode. Both profiles depict only few subminima like in the opening process. During dissociation of 2-BnTetAMPA a substate is suggested by a minima at 0.4 nm, which could be due to random noise. In general, the PES of 2-BnTetAMPA is rougher, including considerable fluctuations at higher distances. The resulting free energy costs to dissociate AMPA and 2-BnTetAMPA from the open form of the LBD are 6.10 and 8.84 kcal/mol, respectively. Combining the data from both substeps derived by Umbrella Sampling, we obtain a total binding free energy of -12.3 and -17.5 kcal/mol for AMPA and BnTetAMPA, respectively.

4.2.4 Discussion

We could show, that it is possible to calculate the energetics of the ligand binding domain opening and the total binding free energy. Nevertheless, this is connected with extensive simulation time. In case of the Umbrella Sampling calculations for the opening/closing of

the LBD, more than 200 ns per US window were needed to get a converged PMF and even after 300 ns small changes in the free energy profiles are obtained. Thus, even after a total simulation time of 6 μs perfect convergence is not reached. Despite the convergence problems, three substates have been detected for the opening of the LBD. A closed state, which is most stable and corresponds to the configuration found in the X-ray structures of agonist bound LBDs. A semi-closed state, that occurs only for ligand bound complexes and resemble the opening of partial agonists and an open state which is most favorable in the APO form. Moreover, the free energy profile for the APO form shows no barrier for the wide opening of the LBD. This is consistent with the simulation of the APO monomer in section 4.1.2. Considering the dimer would probably raise an energy barrier at wider distances, due to sterical clashes.

The comparison of both ligands, present a stronger stabilization of the closed state by AMPA. In case of 2-BnTetAMPA, the closed and semi-closed states are comparable in energy, but still the closed form is preferred. The dissociation of the ligands leads to comparable profiles for both ligands. Though, the relocation of 2-BnTetAMPA into bulk water requires slightly more energy, which can be explained by the additional hydrophobic phenylring. The Metadynamics approach yielded in slightly smaller free energy differences for the LBD opening, which is more distinct for BnTetAMPA. The might be induced by the higher flexibility of the ligand, which increases the error of the methods. In general, Metadynamics is an appealing method to calculate free energies in biomolecules, due to the highly increased sampling efficiency with comparable results. Nevertheless, Umbrella Sampling is a much more robust method, since the outcome of Metadynamics simulations seem to be parameter dependent. The choice of different reaction coordinates might lead to different free energy curves for the same process. Moreover, the disregard of reaction coordinates can lead to overestimated barriers, which are not corrected by longer simulation times.[196] Thus, the usage of Metadynamics requires a good conception of the investigated process. We used a two-step model to calculate the binding free energy of the ligands, which is a realistic description of the process, because the LBD remains open after the dissociation of the ligand and no additional substep of closing after unbinding is necessary. The step of unbinding could not be obtained by the Metadynamics approach. This is caused by the convergence criteria of the method, which requires several transitions between the two states. In this case, bound and unbound state of the ligand. This leads to severe convergence problems, because the ligand stays in the bulk water after dissociation. The repeated binding of the ligand could only be achieved by additional restraint, which would influence the free energy differences of the process. Combining the free energy differences of the two steps leads to binding free energy for AMPA of -12.28 kcal/mol or -10.4 kcal/mol using Umbrella Sampling or Metadynamics for the first substep. This is an good agreement with experimental value of -10.8 kcal/mol. The calculated binding free energy of 2-BnTetAMPA is 17.53 kcal/mol and 13.7 kcal/mol using Umbrella Sampling and Metadynamics, respectively, for the opening of the LBD. Both values are much to high compared to the estimated

binding energy (about 8 kcal/mol), which is based on experimental affinity value in the low micromolar range.[81]

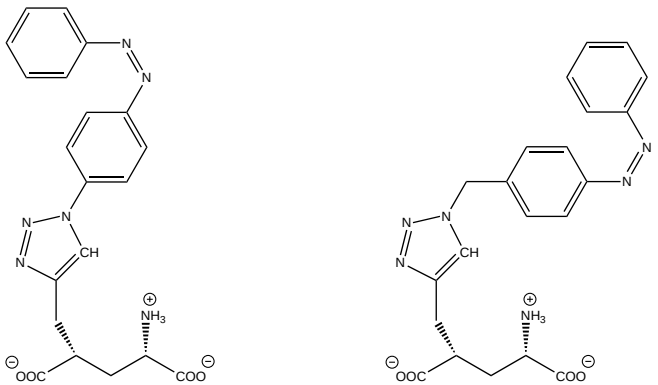
In summary, the calculation of absolute free energy differences for such a complicated system is highly demanding. The sampling efficiency is even more decreased by big and flexible ligands, while the error is increased. Thus, this approach to estimate the quality of potential ligands for such a ligand binding domain is disputable. Nevertheless, the calculated free energy profiles can help to understand the binding process and the principal motions of the LBD due to binding of a ligand.

Chapter 5

Development of photoswitchable ligands targeting iGluR2

5.1 Introduction

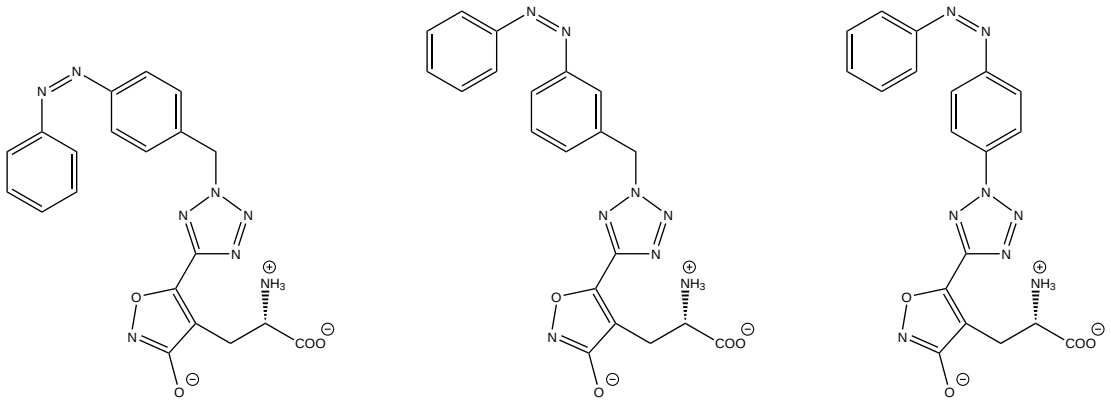
As mentioned in the introduction (section 1.2.2), glutamate receptors have been focused by the field of optogenetics. Since these receptors are not sensitive to light in their native form, a different approach has to be used. In this work we focused on the photochromic ligands (PCL), which are not permanently attached to the target protein. In the beginning of this work, several possible ligands were proposed, which are based on glutamate (Fig.5.1) or AMPA (Fig. 5.2). The photosensitive azobenzene is attached to the glutamate/AMPA-basis via a triazole or tetrazole-ring, which is adopted from the known agonist 2-BnTetAMPA.



(a) ATG (b) BnATG

Figure 5.1 – Glutamate-based ligands

Neither experimental evidence about activity of the ligands nor crystal structures were available. Thus, we used the widely established docking approach to get a first idea of possible binding motives.



(a) ATA-1 (b) ATA-2 (c) ATAMPA

Figure 5.2 – AMPA-based ligands

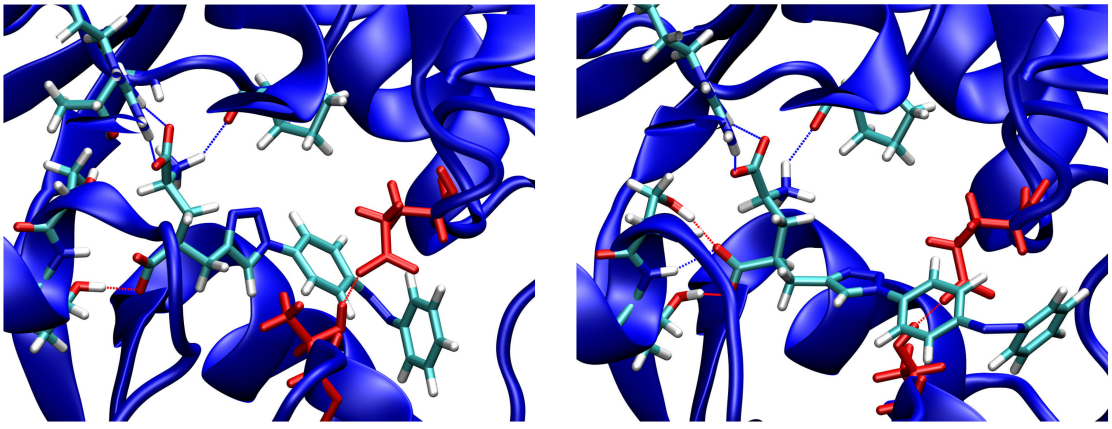
5.2 Docking

As a starting point, we considered AutoDock VINA[197] (version 1.1.1) and an evaluation version of Flexscreen[198,199] (version 20081022) and tested their reliability for this system. Therefor we evaluated the capability of reconstructing the binding motives of available crystal structures of iGluR2. For AutoDock VINA the standard parameters have been used, except the exhaustiveness that was increased from 8 to 30. For the cascadic sampling approach[200] of Flexscreen we performed 30 runs, where the number of trajectories are reduced from 500 over 5 to a number of 2. The corresponding cascadic steps have an increasing length of 5000, 30000 and 75000 steps, respectively. The temperature was decreased form 4 K to 2.5 K which leads to a more localised search in the final run. For all three cascades a timestep of 300 ps was used. The used test set includes six crystal structures of the LBD of iGluR2, which range from fully open over several degrees of partial closing to a fully closed state: 1FTM[72] (AMPA), 2P2A[74] (2-BnTetAMPA), 1FTJ[72] (glutamate), 1FTK[72] (Kainate), 1MQG[73] (IW) and 1FTL[72] (DNQX). In general, both approaches predict the correct binding motives for all ligands with a RMSD below 1.0 Å. The antagonist DNQX is the only exception, where Flexscreen presumes an inversion of the ligand compared to the reference structure. A further trend is the increase of the RMSDs for wider opened LBDs and more flexible ligands. This is quite presumably, because of the increased number of degrees of freedom, which hinder the search. These results show that the used programs are capable of predicting the right binding motives for different degrees of cleft opening. Since both programs create comparable results, we used only AutoDock VINA for further studies to reduce the computational cost. In the next step, the docking protocol was extended by the additional usage of flexible sidechains within the binding pocket. Thus, each of our photoswitchable ligand was docked using a rigid receptor binding site R as well as using two sets of flexible side chains. For the first set Flex-1, amino acid residue side chains located close to the natural binding position of glutamate in the X-ray crystal

Table 5.1 – RMSD-values of docked structure compared to the X-ray structure

	VINA	Flexscreen
AMPA	0.14	0.16
2-BnTetAMPA	0.33	0.28
glutamate	0.20	0.18
IW	0.340	0.45
kainate	0.59	0.62
DNQX	1.09	5.61

structure 1FTJ were chosen, namely Arg485, Glu402, Met708, Leu650, Thr686, Thr655, Thr480, Tyr405, Tyr450. For a second set Flex-2, side chains close to the terminal benzyl ring of 2-BnTetAMPA in the X-ray crystal structure 2P2A were selected as flexible: Ile712, Met708, Ser403, Tyr405, Thr686, Glu402, Trp767 and Tyr711. To mimic the LDB closing promoted by ligand binding, we used the six X-ray structures, which have been used for the benchmark calculation plus the crystal structure of the APO form (1FTO).[72]
After neglecting all proposed binding poses, which could not preserve the main binding motives known from co-crystallized ligands, the docking results could be divided into two groups. The first group is docked into a position where the azobenzene residue is located behind the hydrogen bond of Glu402 and Thr686 (termed as POS1). The second group includes binding poses, where the azobenzene is located in the front of that hydrogen bond (termed as POS2). This circumstance is illustrated in Fig. 5.3, using *cis*-ATG as an example for both groups. While POS1 is mainly obtained by using the 2P2A protein structure, POS2



(a) POS1 (b) POS2

Figure 5.3 – Comparison of the two possible ligand positions. Left side: POS1; right side: POS2. Glu402 and Thr686 are depicted in red for clarification.

is favored by the more opened LBD of 1FTK (see Table 5.2). Like expected, one can observe more positive binding poses by using flexible sidechains within the binding pocket.

Interestingly, no binding poses for the widely opened LBD of DNQX and the APO form was found. This might be due to the increased number of degrees of freedom, which handicaps the search for binding poses. On the other hand, only few binding poses were found for the closed forms of AMPA and glutamate. This can be explained with a too narrow binding pocket for such big molecules. Since, the idea of the docking approach is to exclude molecules as ligands for a target protein out of huge databases, docking programs are designed to be very fast. Thus, the accuracy of the method suffers from this need and tends to produce false positive results. Additionally, the huge domain motion of the LBD due to binding is very demanding for the docking approach. Using only seven different structures of the protein to mimic this behavior might be insufficient and bias the results. Therefore, the best structures, with respect to the binding energy, were used as starting structures for further studies, e.g. MD simulations.

Table 5.2 – Binding poses, conducted by three different setups.

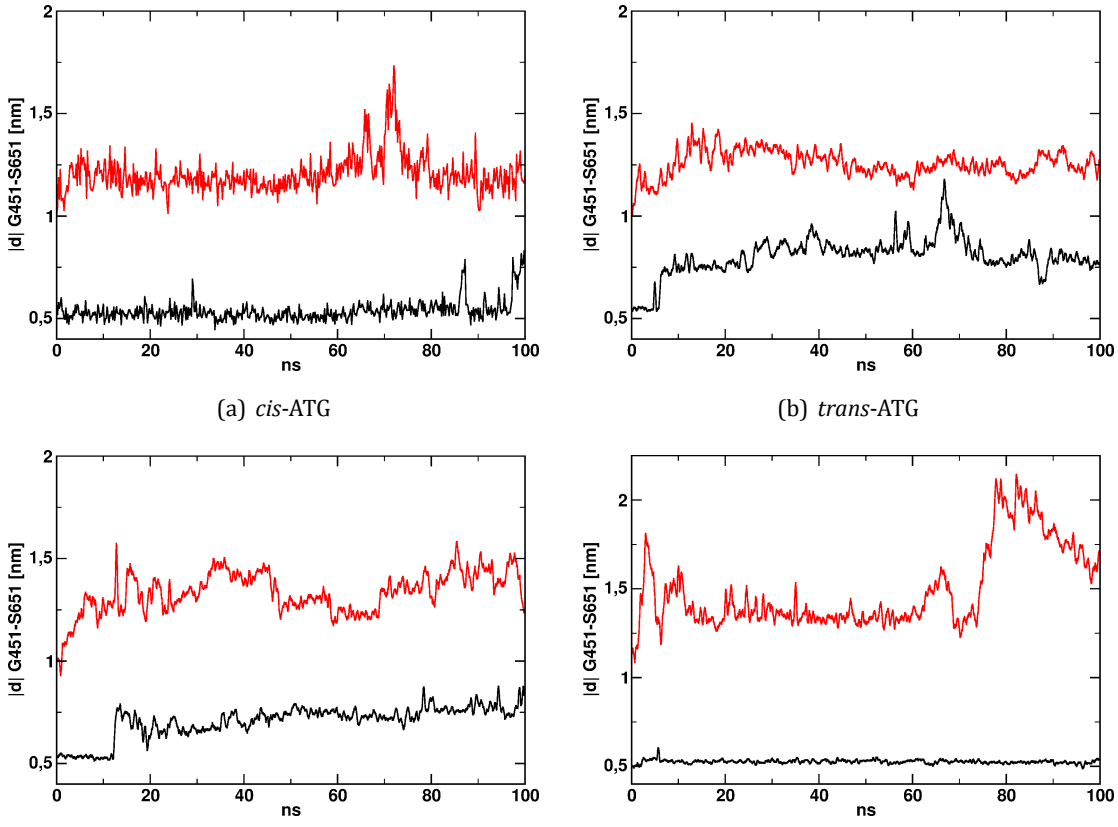
	RIGID		FLEX-1		FLEX-2	
Ligand	POS-1	POS-2	POS-1	POS-2	POS-1	POS-2
cis-ATG	x	-9.2 (1FTK)	**-8.1 (2P2A)**	-8.6 (1FTK)	-6.1 (2P2A)	**-10.9 (1FTK)**
trans-ATG	x	-9.2 (1FTK)	x	-9.2 (1FTK) -8.5 (2P2A)	**-6.0 (2P2A)**	**-10.2 (1FTK)** -6.1 (2P2A)
cis-BnATG	x	-9.6 (1FTK)	x	x	**-5.3 (2P2A)**	**-10.5 (1FTK)**
trans-BnATG	-6.2 (2P2A)	**-9.5 (1FTK)**	-9.0 (2P2A) -7.5 (1FTM) -6.9 (1FTJ)	-6.4 (1FTK)	**-11.5 (2P2A)** -11.0 (1FTM) -10.1 (1FTJ)	-8.8 (1FTK)
cis-ATA-1	x	x	**-8.3 (2P2A)**	**-7.4 (1FTK)**	x	x
trans-ATA-1	-7.4 (2P2A)	x	-10.4 (1FTM) -9.4 (2P2A)	x	**-11.5 (2P2A)**	x
cis-ATAMPA	x	x	x	x	x	x
trans-ATAMPA	x	x	**-8.2 (2P2A)**	-8.5 (1FTK)	x	**-8.8 (1FTK)**
cis-ATA-2	x	x	x	x	x	**-9.0 (1FTK)**
trans-ATA-2	-8.2 (2P2A)	x	x	x	**-10.8 (2P2A)**	x
cis-ATA-3	x	x	x	x	x	x
trans-ATA-3	-8.7 (2P2A)	x	x	x	**-12.6 (2P2A)**	**-10.5 (1FTK)**

x) No plausible binding

bold) Used for future work

5.3 MD simulations

MD simulations were performed to get a first impression of the stability of the binding poses. The general MD setup is the same like in section 4.1.1. Though, the simulation time is reduced to 100 ns due to the limited computer power at this stage of the thesis. Despite the relatively short simulation time, a general trend can be observed. All ligands in both configurations stay in the binding pocket, when their starting structure was POS1. When starting in POS2, all ligands except *cis*-ATA-2, rotate out of the binding pocket and interrupt the primary interactions with domain 2. This loss induces an opening of the LBD, which is also described by the increased distance between Gly451 and Ser651 (Fig. 5.4 and 5.5).

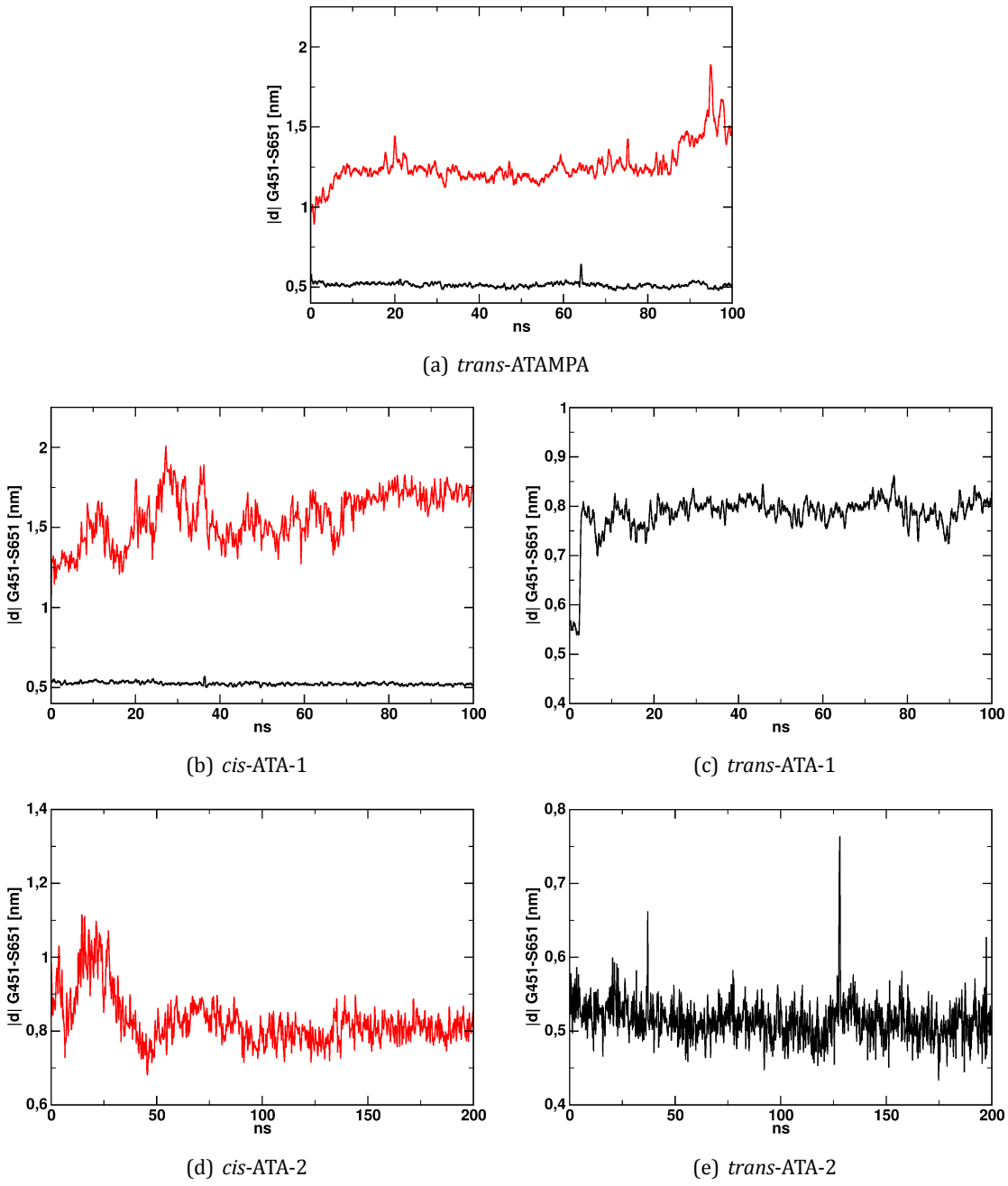


(a) *cis*-ATG (b) *trans*-ATG

(c) *cis*-BnATG (d) *trans*-BnATG

Figure 5.4 – Plots of the G451-S651 distance along the trajectories of the glutamate-based ligands. Black curve: Position 1; Red Curve: Position 2

At this point of our project our collaborators found experimental evidences, that except ATA-2, none of the five proposed ligands activate iGluR2. Thus, we discarded these ligands from further investigations. In case of ATA-2, the experiment proposed an activity in the *trans*-configuration. The activity of ATA-2 was observed in the *trans*-configuration.



(a) *trans*-ATAMPA

(b) *cis*-ATA-1

(c) *trans*-ATA-1

(d) *cis*-ATA-2

(e) *trans*-ATA-2

Figure 5.5 – Plots of the G451-S651 distance along the trajectories of the AMPA-based ligands. Black curve: Position 1; Red Curve: Position 2

5.4 ATA-2

Based on the docking results and the MD simulations of the corresponding structures, a potential mechanism was drawn up. The docking calculation could only obtain a POS1 binding pose for the *trans*-configuration and a POS2 pose for the *cis*-configuration. Furthermore, the conducted MDs showed a high stability for POS1 and high instability for POS2. Since *cis*-ATA-2 was the only ligand which showed a stable complex with the LBD, when located in position 2, we extended the simulation to 200 ns for both configurations

(Fig. 5.5d+e). Nonetheless, both configurations stayed stable in their corresponding position. But in consideration of the expected conformational changes, the statistics out of a 200 ns simulation are quite poor. The proposed change in position in accordance to the configuration change of the ligand was tested by switching it from *trans* to *cis* located in position 1.

The proper description of the photoinduced isomerization of azobenzene requires high *ab initio* methods, which are computational very demanding. Thus, we conducted forced switching simulations using an additional harmonic biasing potential of 320 kJ/mol deg^2 with a minimum corresponding to the cis-position for the azo-group N-N dihedral angle to represent the light-induced excited state potential energy function.[201] This potential was kept active for a short 500 fs length MD simulation.

After the isomerization to the *cis*-configuration, we conducted a 200 ns length MD simulation to monitor the reaction of the system according to the new ligand conformation. During this time scale, the ligand underwent no change in orientation and no reaction of the LBD was observed. As already stated before, the statistics one obtain out of one 200 ns simulations are not representative. Thus, we changed our simulation protocol to free energy calculations to capture the photoswitch mechanism.

5.4.1 Free Energy of the photoswitch mechanism

We defined three reaction coordinates to monitor the energetics of the photoswitch mechanism.

- opening of the LBD; ligand in POS1
- transition of the ligand from POS1 to POS2
- opening of the LBD; ligand in POS2

The used parameters for the opening of the LBD are taken from section 4.2.1. The reaction coordinate for the transition from POS1 to POS2 was projected on the distance between the center of mass of the azobenzene and residues in the binding pocket. These were chosen in such a way, that the azobenzene is pushed out of the pocket towards POS2 (Ser403, Pro404, Tyr405, Thr707, Tyr 711, Ile712).

Opening of the LBD (POS1)

The obtained free energy profiles for the two configurations of the ligand are very similar (Fig. 5.6). Moreover, they show the same properties like the benchmark calculations for AMPA and 2-BnTetAMPA (see section 4.2.1). For both configuration a broad minimum is at 0.53 nm like for 2-BnTetAMPA. This is connected to a second minimum (+2.1 kcal/mol). These two minima are seperated by an energy barrier of about 3.5 kcal/mol. The minimum, which corresponds to the open form, is located at 1.0 nm and has a height of about 6.6 kcal/mol. This minimum is only observed for the *cis*-configuration. The higher energies for *trans* in this region are due to convergence problem (see convergence plot in the Appendix,

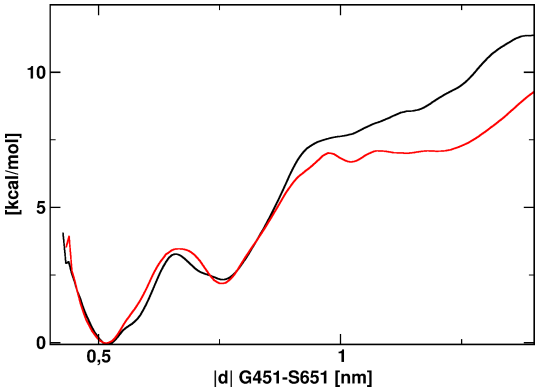


Figure 5.6 – Free energy profiles for the opening of the LBD with a bound ATA-2 in *trans* (black) and *cis*-configuration (red) in position 1

Fig. C.1), which have been also observed for 2-BnTetAMPA. In summary, the differences in the energy profiles give no evidence for a different behavior of the two configurations.

Transition from POS1 to POS2

Compared to the opening of the LBD, the two free energy profiles for the transition between the two positions have a significantly different shape (Fig. 5.7). The *trans*-configuration leads to a profile with three distinct minima. The broad minimum at 0.37 nm corresponds to POS1 (Fig. C.3a). The second minimum at 0.9 nm is 2.6 kcal/mol higher than POS1. It corresponds to an intermediate state between POS1 and POS2. The azobenzene is located between Glu402 and Thr686 and prohibits the hydrogen bond between the two residues (Fig. C.3b). The third minimum at 1.2 nm (+7.0 kcal/mol) corresponds to POS2. The azobenzene is located in front of two amino acid residues. Though, the amino acids do not form a hydrogen bond during the simulation (Fig. C.3c).

The PMF for the transition of the *cis*-ligand has a completely different shape (Fig. 5.7). Like for the *trans*-ligand, a minimum around 0.37 nm can be observed, which corresponds to POS1 (Fig. C.4a). But the profile within the minimum is rugged and cross over a small barrier of 1.1 kcal/mol to a second minimum at 0.57 nm (+0.65 kcal/mol), which is an intermediate state as observed in the *trans*-configuration (Fig. C.4b). After crossing a further barrier of 3.7 kcal/mol, the PMF shows a third minimum a 1.06 nm (+0.26 kcal/mol), which correspond to POS2 (Fig. C.4c). The differences in location of the minima and the rugged surface for the *cis*-configuration can be explained by the definition of the reaction coordinate. The center of mass of the azobenzene is different for the two conformations.

Despite the slight differences for the reaction coordinate, a significantly different PMF can be observed. The barrier to interrupt the hydrogen bond is negligible for the *cis*-ligand and can be overcome easily at room temperature. Also the rate-limiting barrier for the transition is much smaller for *cis*. But the most significant difference is the free energy differences between the two end positions. While in case of the *trans*-azobenzene, the free energy is increased by 7 kcal/mol, the difference for the *cis*-configuration is negligible and

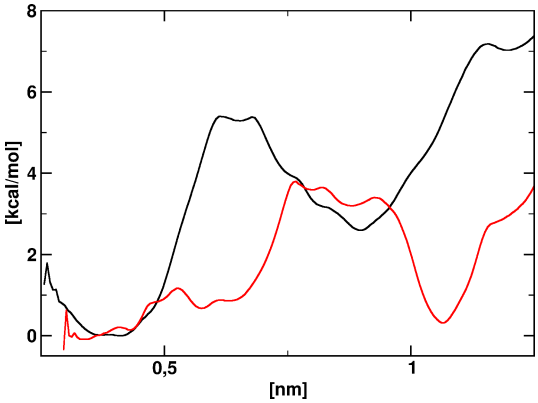


Figure 5.7 – Free energy profiles for the transition from POS1 to POS2 of ATA-2 in *trans* (black) and *cis* (red)

within the error of the method.

This depicts a possible first step within the photoswitch mechanism. While the *trans*-ligand stays stable in POS1, the *cis*-ligand can overcome the barrier at room temperature and swap between POS1 and POS2.

Opening of the LBD (POS2)

The two configurations of the ligand show different tendencies for their orientation. Thus, the impact of the POS2-location on the opening of LBD could be a further hint to unravel the photoswitch mechanism. The PMFs for the two configurations are again significantly different. The *trans*-configuration in POS2 leads to a PMF that is similar to the *trans*-configuration in POS1. Though, there are slight changes. The first minimum is shifted to 0.57 nm, according to a more opened LBD. The second minimum is at 0.75 with a height of 1.9 kcal/mol. After the second minimum the PMF rises significantly and no additional minimum is observed. This might be due to convergence problems, even if the consecutive profiles do not depict notable changes (Fig. C.5a).

The *cis*-ligand in POS2 leads in a shallow PMF. The first minimum is at 0.47 nm, but the

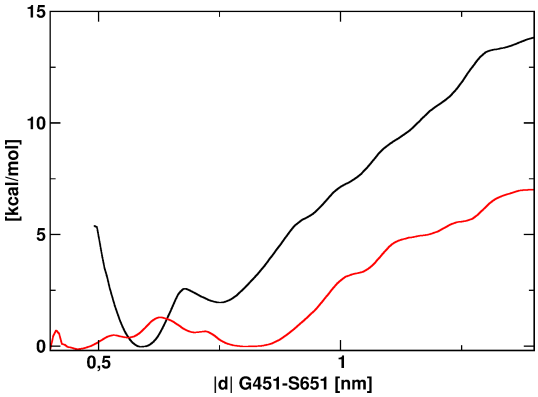


Figure 5.8 – Free energy profiles for the opening of the LBD with a bound ATA-2 in *trans* (black) and *cis*-configuration (red) in position 2

system has only to overcome a barrier of 1.2 kcal/mol to descend into a broad minimum at 0.81 nm, which has nearly the same free energy like the first minimum (+0.1 kcal/mol). There is no third minimum but a saddle point at 1.04 nm (3.3 kcal/mol), according to the open state.

In summary, the *cis*-ligand shows a clear tendency to populate the POS2 state, which is quite unlikely for the *trans*-configuration. Moreover, the *cis*-ligand in POS2 leads to very flat PMF for the opening of the LBD. The closed and semi-closed state are energetically equally favorable and the transition to the open state requires only 3 kcal/mol. These results, suggest an intriguing mechanism for the photoswitching. Though, the reaction of the ion channel has not been investigated.

At this point of the project, experiments of our collaborators gave inconclusive results about the activity of ATA-2 targeting iGluR2.[202] The same experiments, indicate a high activity for ATA-3. Thus, our focus was set to this modified ligand.

5.5 ATA-3

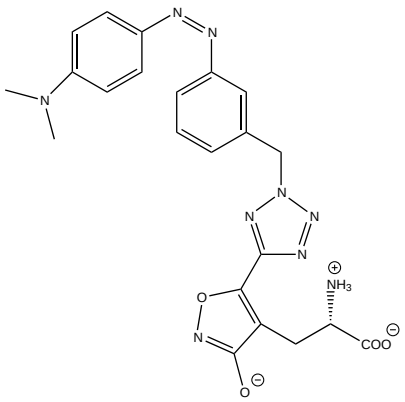


Figure 5.9 – Chemical structure of *cis*-ATA-3

The ATA-3 ligand is an advancement of ATA-2. The introduced modification is the dimethylamino substituent at the para-position of the azobenzene. The docking calculation of the ligand proposed no reasonable structure for the *cis*-configuration, but POS1 and POS2 for the *trans*-geometry.

As a starting point, we conducted 5 independent MD simulations for both models, suggested by docking calculations. After a short (500 ps) position restraint equilibration we performed 500 ns of free MD simulation to equilibrate the system. For the ligand in position 1, stable structures were found for both the receptor structure and the position of the bound ligand. RMSD values for the protein were ca. 0.15 nm after equilibration and rose slightly to ca. 0.2 nm at the end of the simulations (Fig. C.6a). In good agreement with these small RMSD values, no major conformational changes could be observed in any part of the LBD (Fig. 5.10a). The hydrogen bond between Glu402 and Thr686 remained intact during all simulations (Fig. 5.10b). Likewise, the ligand kept its starting position. Simulating

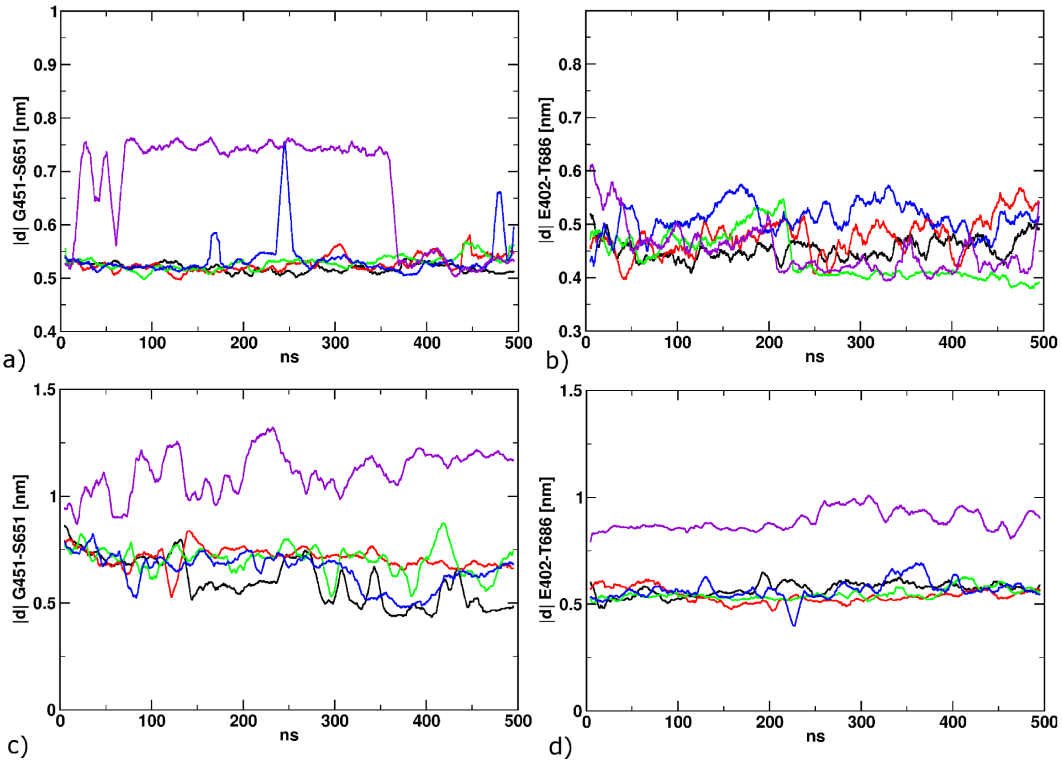


Figure 5.10 – Two structural properties along the equilibration MD for POS1 (a+b) and POS2 (c+d). The distance between Gly451 and Ser651 describes the opening of the LBD (a+c). The distance between E402 and T686 maps the formation of a hydrogen between the two residues (b+d).

the second model, with ATA-3 in POS2, depicts a reduced stability of the complex. For the initial complex based on the X-ray structure 1FTK, the ligand dissociates from the binding side spontaneously after several hundred nanoseconds and the LBD opens, comparable to the APO form. To further investigate the possibility of a ligand binding mode similar to position 2, we manually transferred the docked structure of *trans*-ATA-3 in POS2 from the 1FTK binding site to an analogous position in the binding site of the protein structure 2P2A and adjusted all side chains that possess sterical clashes with the ligand. This model was subject to a cascade of restrained equilibration MDs with decreasing restraining forces. (i) 500 ps with forces of 1000 kJ mol^{-1} nm^{-2} on the backbone atoms and the heavy atoms of the ligand; (ii) 10 ns with 100 kJ mol^{-1} nm^{-2} and (iii) 20 ns with 10 kJ mol^{-1} nm^{-2}. After the equilibration procedure, five independent simulations were perfomed as before. In four of these, the complex structure remained stable as for POS1. During the fifth simulation (purple curve)(Fig. 5.10c+d) a partial unbinding of the ligand was observed, which is accompanied by opening of the LBD.

5.5.1 Photoswitching simulations

The photoswitching of the ligands was conducted by an classical potential, as described in section 5.3, for all simulations. Afterwards, we performed free MD simulations of 900 ns length to monitor the response of the LBD according to the new ligand geometry.

ATA-3 in position 2

In all four cases, the isomerization to the *cis*-geometry does not generate any notable reaction of the protein. The hydrogen bond between Glu402 and Thr686 stays stable during the all simulations and the ligand kept its position (Fig. 5.11(c+d)). Thus, we neglected position 2 as a possible location for the *trans*-azobenzene for further investigations.

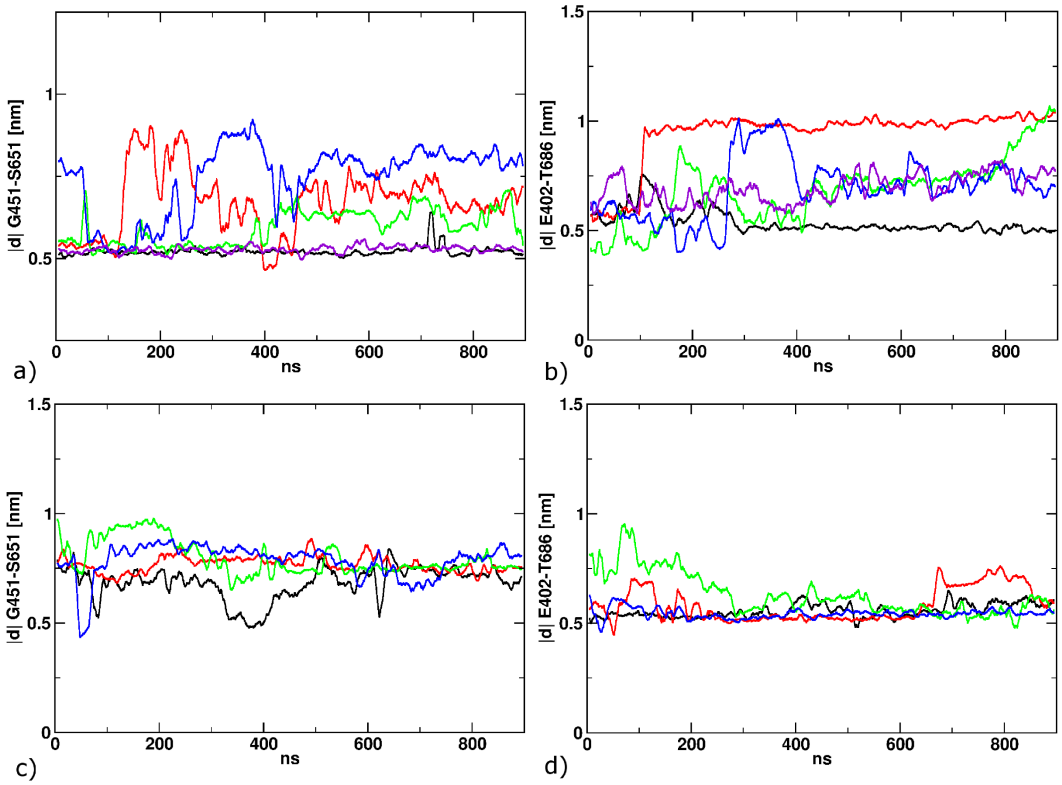


Figure 5.11 – Two structural properties after the isomerization to *cis* for POS1 (a+b) and POS2 (c+d). The distance between Gly451 and Ser651 describes the opening of the LBD (a+c). The distance between E402 and T686 maps the formation of a hydrogen between the two residues (b+d).

ATA-3 in position 1

The five simulations conducted after switching the ligand in position 1, depict a different picture compared to position 2. In one of the cases, the ligand stays in POS1 and no reaction of the protein is observed (Fig.5.11a+b, black curve). In a different simulation (purple curve) the ligand also stays in POS1, but interrupts the hydrogen bond between Glu402 and Thr686 (Fig.5.11b). Nevertheless, no opening of the LBD can be observed (Fig.5.11a). In the other three cases, a reorientation of the ligand can be observed, accompanied by a

distortion of the Glu402-Thr686 hydrogen bond. This required time for this reorientation differs from 100 ns (Fig.5.12a; red curve) to 800 ns (green curve.) Comparing the graphs 5.12(a) and 5.11(b), one can observe a clear dependency between the reorientation of the ligand and the perturbed hydrogen bond. Furthermore, the distant between the azobenzene and the pocket depict two stable states, an intermediate state at a distance around 0.75 nm, and POS2 at 1.05 nm. In all three cases, the ligand gets to the intermediate position, but only in one simulation (Fig.5.12a, blue curve) the ligand ends up in position 2. This indicates a notable second barrier between the intermediate and POS2, since in one of the simulation the system rested 800 ns in the intermediate state. The experiments de-

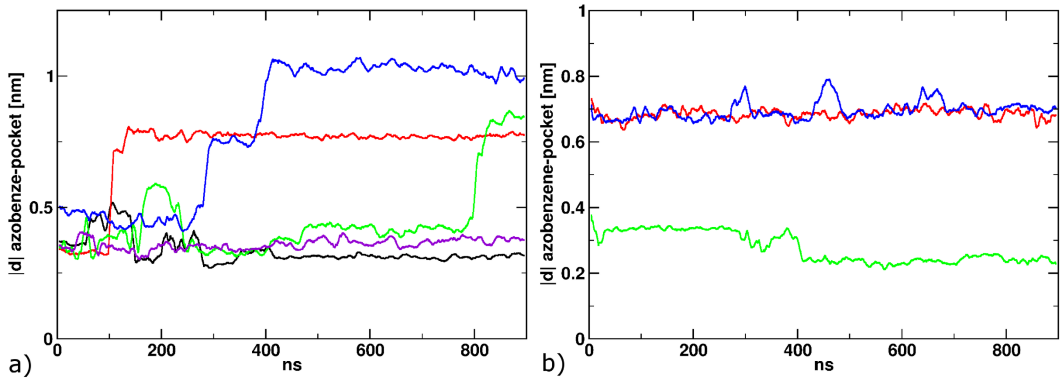


Figure 5.12 – The distance between the azobenzene and the binding pocket, that describes the translocation from position 1 to position 2. a) 900 ns after the isomerization to *cis*; b) 900ns after back-isomerization to *trans*.

pict a reversible mechanism, e.g. one can switch forth and back. Thus, the azobenzene was switched back to trans. Due to computational efficacy, only the three simulations with a new ligand position were used. The photoswitching was again obtained by a classical potential. After the *cis*-to-*trans*-isomerization, MD simulations of 900 ns length were conducted. In one of the three cases, this led in a direct reorientation into position 1 (Fig. 5.12b; green curve) and the Glu402-Thr686 hydrogen is formed again (Fig. 5.13b). The ligands in the other two simulations stayed in the intermediate state (red curve) or move from POS2 into the intermediate state (blue curve). In both cases, the Glu402-Thr686 hydrogen bond cannot be formed, due to the perturbing azobenzene (Fig. 5.13b). Thus, also a notable barrier between the intermediate and POS1 is indicated for the ligand in its *trans*-geometry. Nevertheless, based on these results, we can propose the first part of the photoswitch mechanism for ATA-3 in iGluR2. In a *trans*-conformation, a clear tendency to locate in position 1 is observable. According to the photoswitch to the *cis*-configuration, we could observe a breaking of an interdomain hydrogen bond and the reversible relocation of the ligand. It has to be mentioned, that the reorientation, has no direct correlation to the clamshell motion. Though, a change in the G451-S651 distance can be observed after reorientation of the ligand (Fig. 5.11a and 5.13a).

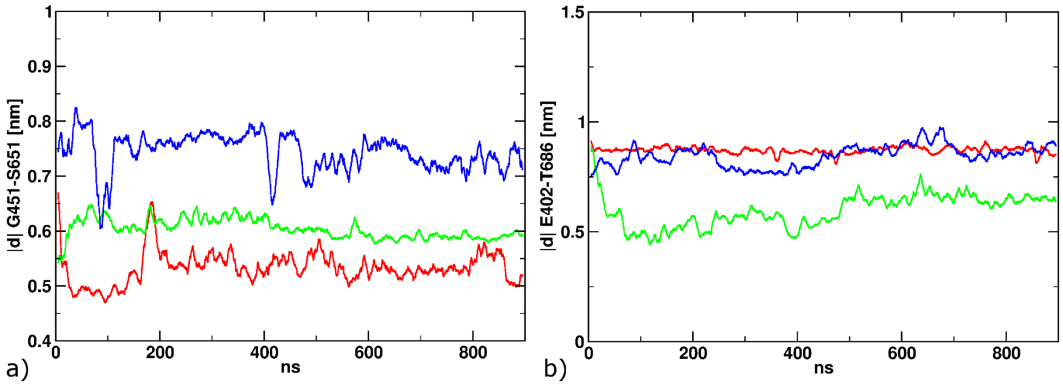


Figure 5.13 – a) The distance between Gly451 and Ser651 after the back-isomerization to *trans*; b) The distance between E402 and T686 after back-isomerization

Umbrella Sampling

Even several MD simulations with hundreds of nanosecond length, are not able to depict a clear picture of the light-induced conformational changes in the LBD. To get a more quantified picture of the first step of the proposed mechanism, we conducted Umbrella Samplings with the relocation of the ligand from POS1 to POS2 as the reaction coordinate, as done before for ATA-2. The general setup is adopted from section 5.4.1. The starting geometries were created by pulling the azobenzene from POS1 to POS2 within 100 ns.
The PMF for the transition from position 1 to position 2 for *trans*-ATA-3 depict a broad free energy minimum around 0.375 nm, which corresponds to position 1 (Fig.5.14). This leads into three subsequent shallow minima at 0.68 nm (+5.9 kcal/mol), 0.92 nm (+11.1kcal/mol) and 1.13 nm (+12.7 kcal/mol). The latter corresponds to POS2, while the second minimum (0.68 nm) describes the intermediate observed during the free MDs. The depicted minimum at 0.92 nm was not represented during our MD simulations. The general trend,

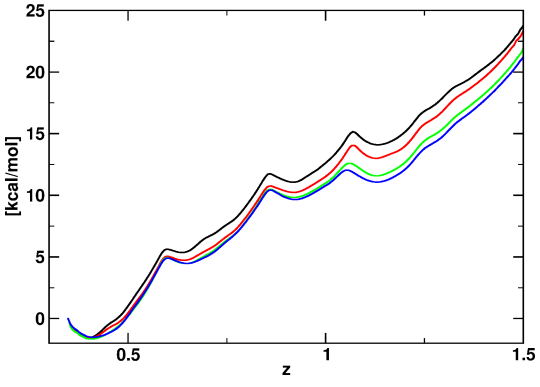


Figure 5.14 – Free energy profile for the transition of *trans*-ATA-3 from POS1 to POS2

depicting a energetically preferred position 1 for *trans*-ATA-3, is consistent with our photoswitching MDs. Nevertheless, the PMF is highly disputable, because the free energy barriers for the relocation from position 2 to 1 are all below 1 kcal/mol. This is expected for the transition from POS2 to the intermediate state, which was observed within nanosec-

onds during free MD sinulations. But the transition from the intermediate to position 1 was only achieved in one of three simulations of 900 ns length. Furthermore, the PMF is not converged after 300 ns simulation time for every window. During the last 50 ns the free energy curve changed 0.7 kcal/mol in the region of position 2 (Fig. 5.14).

The PMF for the same transition with *cis*-ATA-3 is conflicts with our observation during the free MD simulations (Fig. 5.15). Between 0.3 nm and 0.75 nm, one can see a very broad energy plateau with two minima at 0.375 and 0.55 nm, which are separated by a barrier of 0.75 kcal/mol. These minima correspond to position 1 and an intermediate, where the azobenzene is located between Glu402 and Thr686. This broad energy plateau leads to an saddle point at 0.93 nm (+7.1 kcal/mol) and a minimum at 1.07 (+8.3 kcal/mol) that resembles position 2. This strong increase of the free energy due to the transition to POS2 stands in contrast to the free MD simulations, which suggest a lower free energy for position 2 than for position 1. Moreover, the PMF is not converged within the 300 ns. In

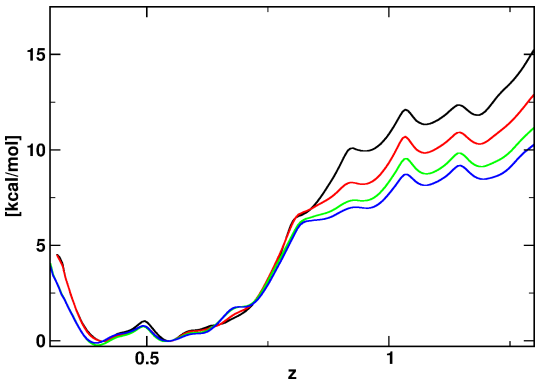


Figure 5.15 – Free energy profile for the transition of *cis*-ATA-3 from POS1 to POS2

the region around 1 nm the energies changed about 0.8 kcal/mol within the last 50 ns. It is known, that the Umbrella Sampling method suffers from the occurrence of hysteresis, e.g. the free energy profile differs between an A-to-B and a B-to-A transition. To estimate the effect of hysteresis for this reaction coordinate, we performed an Umbrella Sampling with the same reaction coordinate, but starting from position 2 and pulling towards position 1.

The resulting PMF, depict a strong hysteresis for this reaction coordinate. Position 2 is energetically favored over position 1 (Fig. 5.16). Moreover, the PMF has no minimum around 0.375 nm, but a steep increase in free energy. This might be produced by the fact that the hydrogen bond between Glu402 and Thr686, which is normally present in POS1, is not formed spontaneously during the US simulations. In general, the PMF is very flat and could not explain the long simulation times, which are needed to monitor the transition from POS1 to POS2. The contradicting PMFs raise the question, if an one-dimensional reaction coordinate, like we have selected, is capable to describe the energetics of such a transition.

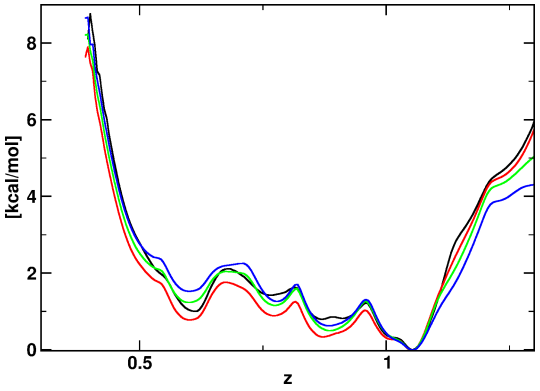


Figure 5.16 – Free energy profile for the transition of *trans*-ATA-3 from POS2 to POS1

Protein reaction

Despite the conflicting results concerning the energetics of the photoinduced changes in the binding pocket, the proposed mechanism could be valid. Therefore, one has to connect the observed changes in the binding pocket with the interdomain motions and investigate the influence on the ion channel opening. For the analysis we only considered the three simulations, where a changes in the binding pocket occurred. The Gly451-Ser651 and Pro632 distances, which showed a nearly perfect correlation for the non-switchable ligands (section 4.1.3), seem to have no connection anymore (Fig.5.17a+b), which is also supported by the negligible correlation (R^2=0.11). However, the distance between the azobenzene and the pocket, depict a dependency to the linker distance (Fig.5.17b+c). If the azobenzene moves out of the pocket, the distance between the prolines is decreased. An alignment of two structures with a ligand in position 1 and 2, depict the effect of the ligand relocation (Fig. 5.18). Due to the new position of the azobenzene, the ligand pushes down helix H, which leads to a decrease in the linker distance. In the same time the G451-S651 stays distance unaffected. Thus, we monitored the distance between the center of mass of helix H and helix B, which is a first guess for the motion of helix H (Fig.5.17d). The general shape of the distance between the helices and between the ligand and the pocket is similar, but the distance between the helices includes a lot of noise, which makes a hard to compare. The correlation between the helix-distance and the Pro632 distance is given by R^2=0.47, which is far from being a perfect correlation, but it is an improvement compared to the Gly451-Ser651 distance. In summary, the switching between the configuration of ATA-3 leads to relocation of the ligand, which not leads to the classical clamshell motion. Nevertheless, we could observe a closing of our virtual ion channel. This seems to be driven by a push down of helix H.

5.6 Discussion

We could show, that the theoretical prediction of ligand activity is a difficult task. For the investigated system this task is complicated in two ways. First, the LBD undergoes a struc-

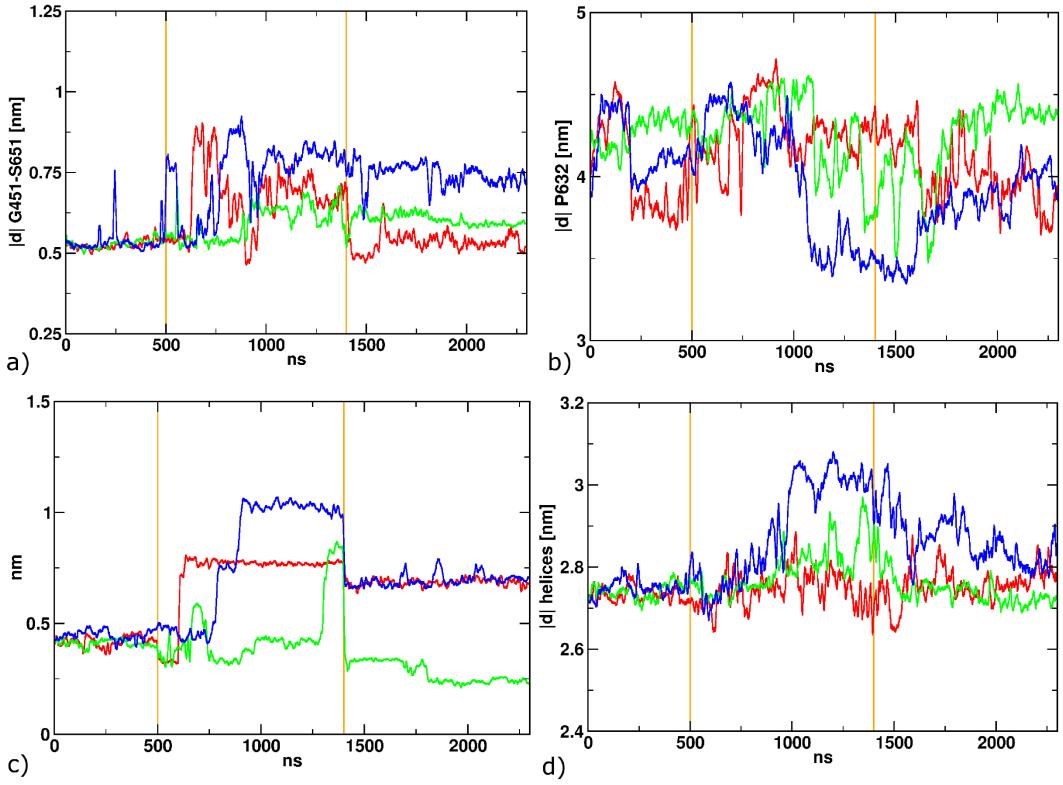


Figure 5.17 – Different structural properties along the three consecutive MDs. The vertical orange lines depict the isomerization. a) G451-S651 distance; b) P632 distance; c) Distance between the azobenzene and the pocket; d) distance between helix H and helix B

tural change due to the binding/unbinding process. Second, the investigated ligands are photoswitchable. The change in geometry does not effect the part of the molecules, that binds to the native protein residues. Therefore, the geometry of the azobenzene residue has to be the basis for the mechanism of switching between activity and inactivity. Because of the lack of crystal structures of bound ligands, we used the docking approach to obtain reasonable starting structures. The resulting two binding poses gave a first idea how the photoswitch could work. While position 1 was invariably obtained by docking into closed LBDs, position 2 was mainly achieved by using semi-closed crystal models. The different activity between the ligand configurations might be achieved by relocation of the ligand. For most of the ligands, both configurations could be docked in both positions. The resulting necessity of ranking this positions by energy is problematic, because the relocation of the azobenzene leads to small changes in the interaction energy, that are overlain by the energy contributions of the main binding motives. Furthermore, docking suffers from the tendency to give false positive results, which make is necessary to check these results by more accurate methods. We used molecular dynamic simulation methods to validate the docked structures. During the MD simulations, most of the ligand in position 2 showed a high instability within 100 ns, which was accompanied with an opening of the LBD. The ligand in position 1 stayed there during the simulation. Due to the limited simulation time,

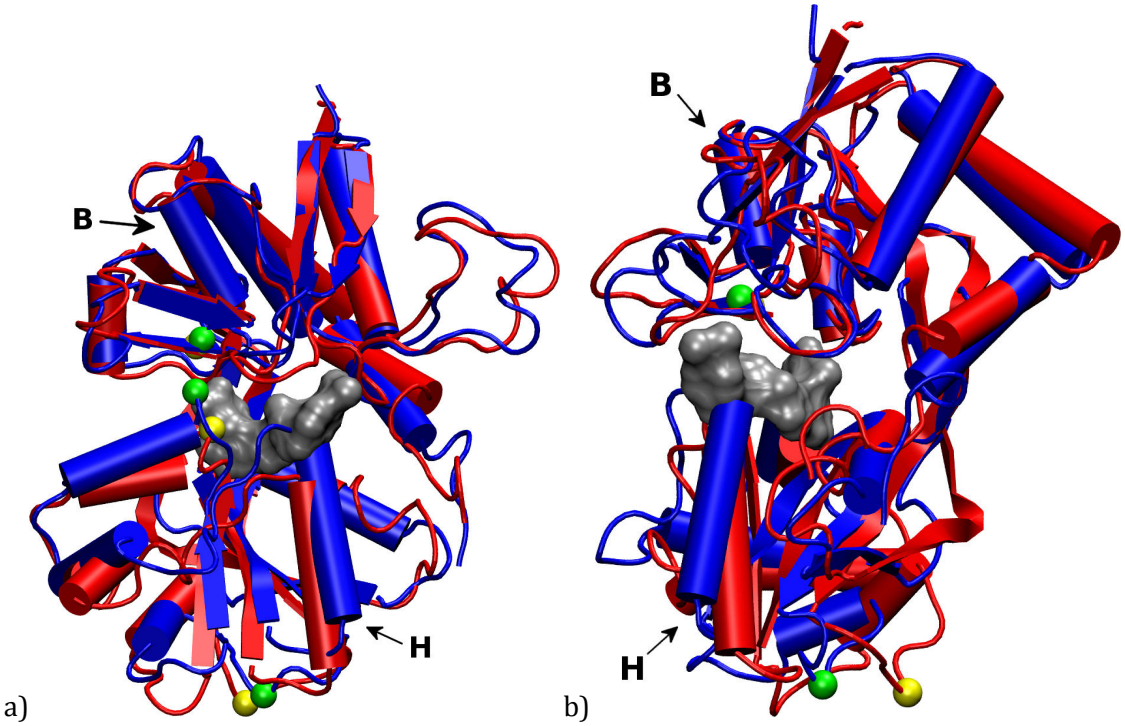


Figure 5.18 – Illustration of the push down of helix H. The blue protein structure is taken after equilibration run. The red protein structure is caused by *cis*-ATA-3 (grey bulk) in position 2. a) Front view; b) Side view

one can only make statements about instability of a structure. A stable structure might only be obtained as an artifact, caused by poor sampling. Nevertheless, the instability of position 2 supported the starting idea for the switching mechanism.

ATA-2 was the first ligand, which was suggested by experiment to show switchable activity for iGluR2. Due to the sampling problem of standard MD simulations, we used Umbrella Sampling to check our proposed mechanism. The obtained free energy profiles, showed a clear tendency of relocation of the azobenzene according to the photoisomerization and we could present the first part of a photoswitch mechanism for ATA-2. However, a huge amount of computational time was needed for these free energy curves (18 μs simulation time in total) and convergence was not achieved for all PMFs.

Because of contradicting experimental results concerning ATA-2, we focused subsequently on ATA-3, that showed distinct results. During the long MD simulation (5 times 2.4 μs), that included switching of the azobenzene from *trans* to *cis* and back, we could observe a reversible relocation of the azobenzene. While the *trans*-ligand stays in position 1, it moves spontaneously towards position 2 and breaks a hydrogen bond between the two domains of the LBD when it is isomerized to *cis*. After the back-isomerization the ligand moves back towards the starting position. It has to be mentioned, that only in one of five MDs, the full motion from position 1 to 2 and back could be monitored. In the other cases,

only a part or no motion along this reaction coordinate could be observed. This might be due to the limited simulation time. Although, we simulated 900 ns after each isomerization, it is not enough to capture the complete reaction of the system. In experiment, the fastest decay of the current due to the photoinduced isomerization from *trans* to *cis* was 47.2 $\pm$ 7.7 ms. Thus, the complete reaction of the system is far from being captured by theoretical methods. The attempt to support our results by Umbrella Sampling was prohibited by the occurrence of hysteresis and convergence problems. These problems are probably caused by the relatively simple reaction coordinate, which does not account for the motions of the protein accompanied by the relocation of the ligand. Despite the sampling problems, we investigated the reaction of the protein due to the ligand reorientation. Our results suggest that not the clamshell motion, which normally causes the ion channel opening/closing, is triggered, but a motion of only one helix. The push down of the helix causes a reduction of the linker distance, which represents the ion channel. Also for the protein reaction, one has to be aware of the possibility, that other motion and transitions might occur on a larger time scale, which we cannot capture with present computers and algorithms. The usage of enhanced sampling methods, which do not require a knowledge of the reaction coordinate, e.g. replica exchange,[203] could give additional insights.

5.7 Conclusion

During our studies concerning the LBD of iGluR2, we could elucidate several aspects of this protein. The mechanism of LBD opening and closing was mainly based on different crystal structures, with different degrees of opening. Based on these models a simple clamshell motion was derived.[77] Based on our long MD simulations (1μs), we derived additional motions, e.g. a twisting and rocking motion, which might contribute to the opening of the LBD. This was supported by our free energy calculations, which depict that, depending on the bound ligand, different principal motions are getting important for the opening. In addition to the nature of the LBD opening, we could observe a spontaneously opening of the LBD with a bound agonist, which was followed by an unbinding of the ligand. On the contrary, the MD simulations of partial agonists and antagonist, showed a stable behavior during the 1 ms. This supports the assumption, that the closing of the LBD induced by ligand-binding and the binding strength do not have to be connected. Though, one has to keep in mind the poor statistics of this observation.

We evaluated the possibility to calculate the free binding enthalpies for such a complex system, by a relatively simple two-step model. During our proof of principle with two agonists, we obtained severe convergence problems using an 1-D Umbrella Sampling, that increased the computational demand to a not justifiable amount (18 μs of simulation time). Thus, we tested the promising new method Metadynamics to obtain free energy profiles. We could show, that this method can predict comparable results. However, this method showed to be less robust in comparison to Umbrella Sampling and showed severe param-

eter dependence.

Based on this knowledge, we tried to elucidate the possible application of photoswitchable ligands for iGluR2. We could propose the first step of a photoswitch mechanism for the ligands ATA-2 and ATA-3, which invalidate the first ideas of fast dissociation of the ligand after switching to the inactive configuration. The exposure of the complete mechanism up to the ion channel opening and closing is beyond our current scope, since the experiment depicts time scales in the millisecond range, which is not even reachable with enhanced sampling techniques. Thus, the design of new ligands, only based on theoretical work, is far from being capable for such a complex system. Nevertheless, the insight from this and future work, can help in a more rational design of new ligands.

Chapter 6

Summary

In this thesis, we have investigated phenomena in proteins, that are induced by photoabsorption. In more detail, we have focused on conformational changes which occur after the protein was exposed to light of a certain wavelengths. Therefor, we have focused on two different systems. Our first investigated protein was the transmembrane proton pump bacteriorhodopsin (bR). This protein is one of the best investigated protein, including studies from X-ray crystallography over vibrational spectroscopy to theoretical studies. Due to the close-meshed grip of experimental data and conformational changes that are spatially limited to the binding pocket, we could apply higher-level methods, e.g. QM/MM MD simulations and high *ab initio* calculations to give new insights into some of the last open questions concerning the photocycle of bR.

Our studies on the bacteriorhodopsin K and L state could connect the available structural data of these intermediates with spectroscopic data, like FTIR and UV/Vis. We could show, that the published crystal structures contradict with other findings and that these cryotrapped structures do not have to represent the actual situation at room temperature. Especially, the ability to reproduce the shifts of absorption maxima could be used to rule out obtained models. Summing up our and former studies, we could propose structural models for the K and L state at physiological conditions.

The second investigated system is the ligand binding domain of iGluR2. This protein, which is not light-sensitive in the native state, is muss less studied. The main insight are based on crystal structures with different cofactors, leading to different degrees of closing. The structural changes, that are induced by the binding of a ligand, involves the motion of a complete protein domain. Thus, the considered time scale to observe the important reactions of the protein leads to the necessity of classical description of the system. Moreover, the achievement of predictive results is complicated. Nevertheless, we were able to recover principal motions within our long time MD, that correspond to the schematic model based on the X-ray structures. However, our free energy calculations of the LBD motions depict additional dynamics that play a role for the closing of the clamshell.

Equipped with the obtained knowledge about the system, we could derive a mechanism for

photoswitchable ligands, which turn the ionotropic receptor into a light-sensitive system, i.e. one can control the incorporated ion channel by light. Already the elucidation of this first step, which involves a reorientation of the ligand within the binding pocket and subsequent domain motions, was already computational demanding. Still, we could present an effect of this induced motion, which is not comparable to the standard LBD opening dynamics, on the ion channel opening. Moreover, we could show that our proposed mechanism is reversible, like preconditioned by the experimental findings.

To depict the whole mechanism, that might include subsequent domain motions and ligand unbinding, we have to evaluate other method which increases the sampling efficiency, e.g. replica exchange or coarse grain approaches. Furthermore, we have to test the effect of simulating the isolated LBD, i.e. the consideration of the whole receptor in a lipid bilayer. This also raises the need of a more efficient way to describe the system, but is still accurate enough to map the complex mechanism of photoswitching.

Bibliography

[1] J. L. Spudich, C. S. Yang and K. H. Jung, *Annu. Rev. Cell Dev. Biol*, 2000, **16**, 365--392.

[2] D. Man, W. Wang, G. Sabehi, L. Aravind and A. F. Post, *The EMBO Journal*, 2003, **22**, 1725--1731.

[3] U. Haupts, J. Tittor, E. Bamberg and D. Oesterhelt, *Biochemistry*, 1997, **36**, 2--7.

[4] J. L. Spudich, *Trends in microbiology*, 2006, **14**, 480--487.

[5] D. Oesterhelt and W. Stoeckenius, *Nat New Biol*, 1971, **233**, 149--152.

[6] D. Oesterhelt and W. Stoeckenius, *Proc. Nat. Acad. Sci. USA*, 1973, **70**, 2853--2857.

[7] H. Michel and D. Oesterhelt, *Febs Lett*, 1976, **65**, 175--178.

[8] A. E. Blaurock and W. Stoeckenius, *Nat New Biol*, 1971, **233**, 152--155.

[9] P. Unwin and R. Henderson, *Journal of Molecular Biology*, 1975, **94**, 425--440.

[10] G. S. Harbison, S. O. Smith, J. A. Pardoen, C. Winkel, J. Lugtenburg, J. Herzfeld, R. Mathies and R. G. Griffin, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 1706--1709.

[11] S. O. Smith, A. B. Myers, J. A. Pardoen, C. Winkel, P. P. J. Mulder, J. Lugtenburg and R. Mathies, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 2055--2059.

[12] R. H. Lozier, W. Niederberger, M. Ottolenghi, G. Sivorinovsky and W. Stoeckenius, *Energetics and Structure of Halophilic Microorganisms. SR Caplan and M. Ginzburg, editors. Elsevier/North-Holland Biomedical Press, New York*, 1978, 123--141.

[13] H. H. Luecke, B. B. Schobert, H. T. H. Richter, J. P. J. Cartailler and J. K. J. Lanyi, *Journal of Molecular Biology*, 1999, **291**, 13--13.

[14] H. Belrhali, P. Nollert, A. Royant and C. Menzel, *Structure*, 1999, **7**, 909--917.

[15] L. Zimanyi, G. Varo, M. Chang, B. Ni, R. Needleman and J. K. Lanyi, *Biochemistry*, 1992, **31**, 8535--8543.

[16] K. Ludmann, C. Gergely, A. Dér and G. Váró, *Biophysical Journal*, 1998, **75**, 3120--3126.

[17] T. L. Brack and G. H. Atkinson, *Journal of Molecular Structure*, 1989, **214**, 289--303.

[18] T. Kouyama, R. A. Bogomolni and W. Stoeckenius, *Biophysical Journal*, 1985, **48**, 201--208.

[19] J. K. Lanyi, *Annu. Rev. Physiol.*, 2004, **66**, 665--688.

[20] K. Rothschild, *Journal of bioenergetics and biomembranes*, 1992, **24**, 147--167.

[21] W. Zinth, M. C. Nuss, H. J. Polland, M. A. Franz and W. Kaiser, *Spectroscopy of Biological Molecules*, 1985, 325--331.

[22] M. A. Franz, W. Zinth, W. Kaiser, E. Kölling and D. Oesterhelt, *Biophysical Journal*, 1986, **49**, 651--662.

[23] R. H. Lozier, R. A. Bogomolni and W. Stoeckenius, *Biophysical Journal*, 1975, **15**, 955.

[24] J. F. Nagle, L. A. Parodi and R. H. Lozier, *Biophysical Journal*, 1982, **38**, 161--174.

[25] Y. Shichida, S. Matuoka and Y. Hidaka, *Biochimica et Biophysica*, 1983, 240--246.

[26] S. J. Milder and D. S. Kliger, *Biophysical Journal*, 1988, **53**, 465--468.

[27] A. Lewis, J. Spoonhower, R. A. Bogomolni, R. H. Lozier and W. Stoeckenius, *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 4462--4466.

[28] J. B. Ames, S. Fodor, R. Gebhard and J. Raap, *Biochemistry*, 1989, **28**, 3681--3687.

[29] K. Gerwert, G. Souvignier and B. Hess, *Proc. Natl. Acad. Sci. USA*, 1990, **87**, 9774--9778.

[30] A. Bondar, M. Elstner, S. Suhai and J. Smith, *Structure*, 2004, **12**, 1281--1288.

[31] Y. Matsui, K. Sakai, M. Murakami and Y. Shiro, *Journal of Molecular Biology*, 2002, **324**, 469--481.

[32] E. M. Landau, K. Edman, P. Nollert, A. Royant, H. Belrhali, E. Pebay-Peyroula, J. Hajdu and R. Neutze, *Nature*, 1999, **401**, 822--826.

[33] B. Schobert, J. Cupp-Vickery, V. Hornak, S. Smith and J. Lanyi, *Journal of Molecular Biology*, 2002, **321**, 715--726.

[34] P. A. Bullough and R. Henderson, *Journal of Molecular Biology*, 1999, **286**, 1663--1671.

[35] H. Kandori, N. Kinoshita, Y. Yamazaki, A. Maeda, Y. Shichida, R. Needleman, J. K. Lanyi, M. Bizounok, J. Herzfeld, J. Raap and J. Lugtenburg, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 4643--4653.

[36] H. Kandori, Y. Yamazaki, Y. Shichida, J. Raap, J. Lugtenburg, M. Belenky and J. Herzfeld, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 1571--1576.

[37] S. P. Balashov, N. V. Karneyeva, F. F. Litvin and T. G. Ebrey, *Photochemistry and photobiology*, 1991, **54**, 949--953.

[38] K. J. K. Rothschild, P. P. Roepe and J. J. Gillespie, *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1985, **808**, 140--148.

[39] S. J. Doig, P. J. Reid and R. A. Mathies, *J. Phys. Chem*, 1991, **95**, 6372--6379.

[40] R. Diller, M. Iannone, B. R. Cowen, S. Maiti, R. A. Bogomolni and R. M. Hochstrasser, *Biochemistry*, 1992, **31**, 5567--5572.

[41] A. K. Dioumaev and M. S. Braiman, *J. Phys. Chem. B*, 1997, **101**, 1655--1662.

[42] W. Hage, M. Kim, H. Frei and R. A. Mathies, *J. Phys. Chem*, 1996, **100**, 16026--16033.

[43] F. Siebert, *Laser Chem*, 1999, **19**, 179--185.

[44] J. Sasaki, T. Yuzawa, H. Kandori, A. Maeda and H. Hamaguchi, *Biophysical Journal*, 1995, **68**, 2073--2080.

[45] N. Yamamoto, T. Ebbesen and H. Ohtani, *Chemical Physics Letters*, 1994, **228**, 61--65.

[46] A. Dioumaev and J. Lanyi, *J. Phys. Chem. B*, 2009, **113**, 16643--16653.

[47] E. M. Landau, A. Royant, K. Edman, T. Ursby, E. Pebay-Peyroula and R. Neutze, *Nature*, 2000, **406**, 645--648.

[48] S. P. Balashov and T. G. Ebrey, *Photochemistry and photobiology*, 2001, **73**, 453--462.

[49] J. Lanyi, *Current Opinion in Structural Biology*, 2001, **11**, 415--419.

[50] R. Neutze, E. Pebay-Peyroula, K. Edman, A. Royant, J. Navarro and E. M. Landau, *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 2002, **1565**, 144--167.

[51] K. Edman, A. Royant, G. Larsson, F. Jacobson, T. Taylor, D. van der Spoel, E. Landau, E. Pebay-Peyroula and R. Neutze, *Journal of Biological Chemistry*, 2004, **279**, 2147--2158.

[52] T. Kouyama, T. Nishikawa, T. Tokuhisa and H. Okumura, *Journal of Molecular Biology*, 2004, **335**, 531--546.

[53] J. Lanyi and B. Schobert, *Journal of Molecular Biology*, 2003, **328**, 439--450.

[54] J. K. Lanyi and B. Schobert, *Journal of Molecular Biology*, 2007, **365**, 1379--1392.

[55] V. S. Bajaj, M. L. Mak-Jurkauskas, M. Belenky, J. Herzfeld and R. G. Griffin, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 9244--9249.

[56] A.-N. Bondar, J. Baudry, S. Suhai, S. Fischer and J. C. Smith, *J. Phys. Chem. B*, 2008, **112**, 14729--14741.

[57] S. Fodor, J. B. Ames and R. Gebhard, *Biochemistry*, 1988, **27**, 7097--7101.

[58] C. Rödig, I. Chizhov, O. Weidlich and F. Siebert, *Biophysical Journal*, 1999, **76**, 2687--2701.

[59] V. Lórenz-Fonfría, Y. Furutani and H. Kandori, *Biochemistry*, 2008, **47**, 4071--4081.

[60] R. Dingledine, K. Borges, D. Bowie and S. F. Traynelis, *Pharmacological Reviews*, 1999, **51**, 7--62.

[61] F. Asztely and B. Gustafsson, *Mol Neurobiol*, 1996, **12**, 1--11.

[62] D. Bowie, *CNS & neurological disorders drug targets*, 2008, **7**, 129.

[63] A. C. Foster and J. A. Kemp, *Curr Opin Pharmacol*, 2006, **6**, 7--17.

[64] J. C. Watkins and R. H. Evans, *Annual review of Pharmacol. Toxicol.*, 1981, **21**, 165--204.

[65] M. L. Mayer and G. L. Westbrook, *Progress in neurobiology*, 1987, **28**, 197--276.

[66] S. Schorge and D. Colquhoun, *The Journal of neuroscience*, 2003, **23**, 1151--1158.

[67] M. Mayer, *Annu. Rev. Physiol.*, 2004, **66**, 161--181.

[68] M. L. Mayer, *Current opinion in neurobiology*, 2005, **15**, 282--288.

[69] A. Kuusinen, R. Abele, D. R. Madden and K. Keinänen, *Journal of Biological Chemistry*, 1999, **274**, 28937--28943.

[70] A. Pasternack, S. K. Coleman, A. Jouppila, D. G. Mottershead, M. Lindfors, M. Pasternack and K. Keinänen, *Journal of Biological Chemistry*, 2002, **277**, 49662--49667.

[71] K. Keinanen, *Journal of Biological Chemistry*, 1996, **271**, 15527--15532.

[72] N. Armstrong and E. Gouaux, *Neuron*, 2000, **28**, 165--181.

[73] R. Jin, T. Banke, M. Mayer, S. Traynelis and E. Gouaux, *Nat Neurosci*, 2003, **6**, 803--810.

[74] S. Vogensen, K. Frydenvang, J. Greenwood, G. Postorino, B. Nielsen, D. Pickering, B. Ebert, U. Bølcho, J. Egebjerg and M. Gajhede, *J. Med. Chem*, 2007, **50**, 2408--2414.

[75] M. Sheng and S. H. Lee, *Neuroscience research*, 2003, **46**, 127--134.

[76] J. M. Henley, *Neuroscience research*, 2003, **45**, 243--154.

[77] I. Mano and Y. Lamed, *Journal of Biological Chemistry*, 1996, **271**, 15299--15302.

[78] N. Armstrong, Y. Sun and G. Chen, *Nature*, 1998, **395**, 913--917.

[79] A. Inanobe, H. Furukawa and E. Gouaux, *Neuron*, 2005, **47**, 71--84.

[80] W. Zhang, Y. Cho, E. Lolis and J. R. Howe, *Journal of Neuroscience*, 2008, **28**, 932--943.

[81] K. Frydenvang, L. L. Lash, P. Naur, P. A. Postila, D. S. Pickering, C. M. Smith, M. Gajhede, M. Sasaki, R. Sakai, O. T. Pentikäinen, G. T. Swanson and J. S. Kastrup, *Journal of Biological Chemistry*, 2009, **284**, 14219--14229.

[82] M. Holm, M. Lunn, S. Traynelis, J. Kastrup and J. Egebjerg, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 12053--12058.

[83] E. J. Bjerrum and P. C. Biggin, *Proteins*, 2008, **72**, 434--446.

[84] N. Armstrong, M. Mayer and E. Gouaux, *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 5736--5741.

[85] G. Ramanoudjame, M. Du, K. A. Mankiewicz and V. Jayaraman, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 10473--10478.

[86] E. S. Boyden, F. Zhang, E. Bamberg, G. Nagel and K. Deisseroth, *Nat Neurosci*, 2005, **8**, 1263--1268.

[87] G. Nagel, M. Brauner, J. F. Liewald and N. Adeishvili, *Current Biology*, 2005, **15**, 2279--2284.

[88] A. Bi, J. Cui, Y. P. Ma, E. Olshevskaya, M. Pu and A. M. Dizhoor, *Neuron*, 2006, **50**, 23--33.

[89] T. Ishizuka, M. Kakuda, R. Araki and H. Yawo, *Neuroscience research*, 2006, **54**, 84--94.

[90] F. Zhang, L.-P. Wang, M. Brauner, J. F. Liewald, K. Kay, N. Watzke, P. G. Wood, E. Bamberg, G. Nagel, A. Gottschalk and K. Deisseroth, *Nature*, 2007, **446**, 633--639.

[91] R. Y. Hwang, L. Zhong, Y. Xu, T. Johnson and F. Zhang, *Current Biology*, 2007, **17**, 2105--2116.

[92] D. Bellmann, A. Richardt, R. Freyberger, N. Nuwal, M. Schwärzel, A. Fiala and K. Störtkuhl, *Frontiers in Behavioral Neuroscience*, 2010, **4**, 1--10.

[93] A. B. Arrenberg, F. Del Bene and H. Baier, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 17968--17973.

[94] A. B. Arrenberg, D. Y. R. Stainier, H. Baier and J. Huisken, *Science*, 2010, **330**, 971--974.

[95] S. B. G. Abbott, R. L. Stornetta, C. S. Socolovsky, G. H. West and P. G. Guyenet, *The Journal of Physiology*, 2009, **587**, 5613--5631.

[96] S. B. G. Abbott, R. L. Stornetta, M. G. Fortuna, S. D. Depuy, G. H. West, T. E. Harris and P. G. Guyenet, *The Journal of Neuroscience*, 2009, **29**, 5806--5819.

[97] W. J. Alilain, X. Li, K. P. Horn, R. Dhingra, T. E. Dick, S. Herlitze and J. Silver, *The Journal of Neuroscience*, 2008, **28**, 11862--11870.

[98] L. Fenno, O. Yizhar and K. Deisseroth, *Annu. Rev. Neurosci.*, 2011, **34**, 389--412.

[99] F. Zhang, J. Vierock, O. Yizhar, L. E. Fenno and S. Tsunoda, *Cell*, 2011, **147**, 1446--1457.

[100] K. Deisseroth, *Nature Methods*, 2010, **8**, 26--29.

[101] G. C. R. Ellis-Davies, *Nature Methods*, 2007, **4**, 619--628.

[102] M. Volgraf, P. Gorostiza, S. Szobota, M. Helix, E. Isacoff and D. Trauner, *J. Am. Chem. Soc*, 2007, **129**, 260--261.

[103] M. Banghart, K. Borges, E. Isacoff, D. Trauner and R. H. Kramer, *Nat Neurosci*, 2004, **7**, 1381--1386.

[104] M. R. Banghart, A. Mourot, D. L. Fortin, J. Z. Yao, R. H. Kramer and D. Trauner, *Angew Chem Int Edit*, 2009, **48**, 9097--9101.

[105] M. Volgraf, P. Gorostiza, R. Numano, R. H. Kramer, E. Y. Isacoff and D. Trauner, *Nat Chem Biol*, 2005, **2**, 47--52.

[106] Z. R. Abrams, A. Warrier, D. Trauner and X. Zhang, *Frontiers in Neural Circuits*, 2010, **4**, 1--13.

[107] A. R. Dinner, G. M. Blackburn and M. Karplus, *Nature*, 2001, **413**, 752--755.

[108] A. Aksimentiev, I. A. Balabin, R. H. Fillingame and K. Schulten, *Biophysical Journal*, 2004, **86**, 1332--1344.

[109] R. A. Böckmann and H. Grubmüller, *Nat. Struct Biol.*, 2002, **9**, 198--202.

[110] B. Nagar, O. Hantschel, M. A. Young, K. Scheffzek, D. Veach, W. Bornmann, B. Clarkson, G. Superti-Furga and J. Kuriyan, *Cell*, 2003, **112**, 859--871.

[111] D. Schiff and L. Verlet, *Physical Review*, 1967, **160**, 208--218.

[112] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Dover, 1982.

[113] F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, 2013.

[114] P. Hohenberg, *Physical Review*, 1964, **136**, B864--B871.

[115] M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457--484.

[116] W. Kohn and L. J. Sham, *Physical Review*, 1965, **140**, A1133--A1138.

[117] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865--3868.

[118] C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158.

[119] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372.

[120] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260--7268.

[121] M. Gaus, Q. Cui and M. Elstner, *Journal of Chemical Theory and Computation*, 2011, **7**, 931--948.

[122] Y. Yang, H. Yu, D. York, Q. Cui and M. Elstner, *Journal of Physical Chemistry A*, 2007, **111**, 10861--10873.

[123] M. Gaus, Q. Cui and M. Elstner, *Wiley Interdisciplinary Reviews*, 2013, **00**, 1--13.

[124] K. W. Sattelmeyer, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. A*, 2006, **110**, 13551--13559.

[125] N. Otte, M. Scholten and W. Thiel, *J. Phys. Chem. A*, 2007, **111**, 5751--5755.

[126] J. J. P. Stewart, *J Mol Model*, 2007, **13**, 1173--1213.

[127] W. Weber and W. Thiel, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 2000, **103**, 495--506.

[128] E. Tajkhorshid, T. Frauenheim, S. Suhai and M. Elstner, *Chemical Physics*, 2002, **277**, 91--103.

[129] M. Wanko, M. Hoffmann, P. Strodel, A. Koslowski, W. Thiel, F. Neese, T. Frauenheim and M. Elstner, *Journal of Physical Chemistry B*, 2005, **109**, 3606--3615.

[130] A.-N. Bondar, S. Suhai, S. Fischer, J. C. Smith and M. Elstner, *J Struct Biol*, 2007, **157**, 454--469.

[131] F. Neese, *J. Chem. Phys.*, 2003, **119**, 9428.

[132] B. J. Mogensen and S. Rettrup, *International journal of quantum chemistr*, 1992, **44**, 1045--1056.

[133] J. Miralles, O. Castell, R. Caballol and J. P. Malrieu, *Chemical Physics*, 1993, **172**, 33--43.

[134] G. Hirsch, P. J. Bruna, R. J. Buenker and S. D. Peyerimhoff, *Chemical Physics*, 1980, **45**, 335--347.

[135] P. J. Bruna, S. D. Peyerimhoff and R. J. Buenker, *Chemical Physics Letters*, 1977, **52**, 442--448.

[136] A. Schafer, H. Horn and R. Ahlrichs, *Journal of Chemical Physics*, 1992, **97**, 2571--2577.

[137] M. Wanko, M. Hoffmann, T. Frauenheim and M. Elstner, *J. Phys. Chem. B*, 2008, **112**, 11462--11467.

[138] M. Wanko, M. Hoffmann, J. Fraehmcke, T. Frauenheim and M. Elstner, *Journal of Physical Chemistry B*, 2008, **112**, 11468--11478.

[139] J. S. Fraehmcke, M. Wanko, P. Phatak, M. A. Mroginski and M. Elstner, *J. Phys. Chem. B*, 2010, **114**, 11338--11352.

[140] P. Phatak, J. S. Fraehmcke, M. Wanko, M. Hoffmann, P. Strodel, J. C. Smith, S. Suhai, A.-N. Bondar and M. Elstner, *J. Am. Chem. Soc*, 2009, **131**, 7064--7078.

[141] K. Welke, J. S. Fraehmcke, H. C. Watanabe, P. Hegemann and M. Elstner, *J. Phys. Chem. B*, 2011, **115**, 15119--15128.

[142] A. Warshel and M. Levitt, *Journal of Molecular Biology*, 1976, **103**, 227--249.

[143] D. Riccardi, G. Li and Q. Cui, *J. Phys. Chem. B*, 2004, **108**, 6467--6478.

[144] S. Dapprich, I. Komáromi, K. S. Byun and K. Morokuma, *Journal of Molecular Structure*, 1999, **461**, 1--21.

[145] H. Lin and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 3992--4004.

[146] P. H. König, M. Hoffmann and T. Frauenheim, *The Journal of Physical Chemistry B*, 2005, **105**, 9082--9095.

[147] P. Sherwood, *Hybrid quantum mechanics/molecular mechanics approaches*, Clrc daresbury laboratory technical report, 2000.

[148] G. M. Torrie and J. P. Valleau, *Journal of Computational Physics*, 1977, **23**, 187--199.

[149] J. G. Kirkwood, *J. Chem. Phys.*, 1935, **3**, 300--313.

[150] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562.

[151] D. Bouzida and R. Swendsen, *Journal of Computational Chemistry*, 1992, **13**, 1011--1021.

[152] G. Bussi, A. Laio and M. Parrinello, *Phys. Rev. Lett.*, 2006, **96**, 090601.

[153] E. A. Carter, G. Ciccotti, J. T. Hynes and R. Kapral, *Chemical Physics Letters*, 1989, **156**, 472--477.

[154] M. Sprik and G. Ciccotti, *J. Chem. Phys.*, 1998, **109**, 7737.

[155] P. A. Bash, U. C. Singh, R. Langridge and P. A. Kollman, *Science*, 1987, **236**, 564--568.

[156] A. Laio, A. Rodriguez-Fortea, F. Gervasio, M. Ceccarelli and M. Parrinello, *J. Phys. Chem. B*, 2005, **109**, 6714--6721.

[157] H. A. Carlson and J. A. McCammon, *Mol Pharmacol*, 2000, **57**, 213--218.

[158] C. B-Rao, J. Subramanian and S. D. Sharma, *Drug Discovery Today*, 2009, **14**, 394--400.

[159] A. R. Leach, *Journal of Molecular Biology*, 1994, **235**, 345--356.

[160] R. Knegtel, I. D. Kuntz and C. M. Oshiro, *Journal of Molecular Biology*, 1997, **266**, 424--440.

[161] G. M. Morris, D. S. Goodsell and R. S. Halliday, *Journal of Computational Chemistry*, 1998, **19**, 1639--1662.

[162] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins*, 2003, **52**, 609--623.

[163] H. J. Böhm, *J Comput Aided Mol Des*, 1992, **6**, 593--606.

[164] D. Sitkoff, K. A. Sharp and B. Honig, *J. Phys. Chem*, 1994, **98**, 1978--1988.

[165] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187--217.

[166] L. S. Brown, G. Varo, M. Hatanaka, J. Sasaki, H. Kandori, A. Maeda, N. Friedman, M. Sheves, R. Needleman and J. K. Lanyi, *Biochemistry*, 1995, **34**, 12903--12911.

[167] G. n. Metz, F. Siebert and M. Engelhard, *Febs Lett*, 1992, **303**, 237--241.

[168] P. Phatak, N. Ghosh, H. Yu, Q. Cui and M. Elstner, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 19672--19677.

[169] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586--3616.

[170] U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718--730.

[171] A. R. Dinner, X. Lopez and M. Karplus, *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 2003, **109**, 118--124.

[172] S. Nosé, *Molecular Physics*, 1984, **52**, 255--268.

[173] W. Hoover, *Physical Review A*, 1985, **31**, 1695--1697.

[174] D. A. McQuarrie, *Statistical thermodynamics*, University Science Books, 1973.

[175] H. Yu and Q. Cui, *Journal of Chemical Physics*, 2007, **127**, 234504.

[176] T. Iwasa, F. Tokunaga and T. Yoshizawa, *Biophys. Struct. Mechanism*, 1980, **6**, 253--270.

[177] A. K. Dioumaev, J. M. Wang and J. K. Lanyi, *J. Phys. Chem. B*, 2010, **114**, 2920--2931.

[178] M. Gaus, C.-P. Chou, H. Witek and M. Elstner, *J. Phys. Chem. A*, 2009, **113**, 11866--11881.

[179] A.-N. Bondar, S. Fischer, J. C. Smith, M. Elstner and S. Suhai, *J. Am. Chem. Soc*, 2004, **126**, 14668--14677.

[180] M. Hoffmann, M. Wanko, P. Strodel, P. König, T. Frauenheim, K. Schulten, W. Thiel, E. Tajkhorshid and M. Elstner, *J. Am. Chem. Soc*, 2006, **128**, 10808--10818.

[181] A. N. Bondar, M. Elstner, S. Suhai, S. Fischer and J. C. Smith, *Phase Transitions*, 2005, **78**, 5--9.

[182] A. Strambi, P. B. Coto, N. Ferré and M. Olivucci, *Theor Chem Acc*, 2007, **118**, 185--191.

[183] R. W. Hendler, *J. Phys. Chem. B*, 2005, **109**, 16515--16528.

[184] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926--935.

[185] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712--725.

[186] J. Wang, R. Wolf, J. Caldwell, P. Kollman and D. Case, *J. Comput. Chem.*, 2004, **25**, 1157--1174.

[187] M. Parrinello and A. RAHMAN, *J Appl Phys*, 1981, **52**, 7182--7190.

[188] B. Hess, C. Kutzner and D. van der Spoel, *Journal of Chemical Theory and Computation*, 2008, **4**, 435--447.

[189] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput Phys Commun*, 2009, **180**, 1961--1972.

[190] K. Frydenvang, U. Madsen and J. Kastrup, *Neuropharmacology*, 2011, **60**, 135--150.

[191] M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 2000, **112**, 8910--8922.

[192] S. Hayward and N. Go, *Annual Review of Physical Chemistry*, 1995, **46**, 223--250.

[193] A. Kitao and N. Go, *Current Opinion in Structural Biology*, 1999, **9**, 164--169.

[194] B. Hess, *Phys Rev E*, 2000, **62**, 8438--8448.

[195] T. Wolter, T. Steinbrecher and M. Elstner, *PLoS ONE*, 2013, **8**, e58774.

[196] A. Laio and F. L. Gervasio, *Rep Prog Phys*, 2008, **71**, 126601.

[197] O. Trott and A. J. Olson, *J. Comput. Chem.*, 2009, 455--461.

[198] B. Fischer, K. Fukuzawa and W. Wenzel, *Proteins*, 2007, **70**, 1264--1273.

[199] B. Fischer, S. Basili, H. Merlitz and W. Wenzel, *Proteins*, 2007, **68**, 195--204.

[200] H. Merlitz, T. Herges and W. Wenzel, *J. Comput. Chem.*, 2004, **25**, 1568--1575.

[201] P. Nguyen and G. Stock, *Chemical Physics*, 2006, **323**, 36--44.

[202] P. Stawski, M. Sumser and D. Trauner, *Angew Chem Int Edit*, 2012, **51**, 5748--5751.

[203] Y. Sugita and Y. Okamoto, *Chemical Physics Letters*, 1999, **314**, 141--151.

Appendix A

bR

A.1 L state

Table A.1 – Dihedrals of the retinal along the polyene chain within different L structures and the bR resting state (1C3W)

	1c3w-opt	1ucq-xray	1ucq-opt	Model-D	1vjm-xray	1vjm-opt	2ntw-xray	2ntw-opt
CE-NZ	-102.5	113.2	133.9	123.8	139.0	120.6	-126.6	-100.3
NZ=C15	-164.5	-148.7	-164.3	-169.4	-179.2	-159.9	146.4	-165.9
C15-C14	171.5	-162.4	-163.0	-169.9	-179.3	-155.6	175.8	171.2
C14=C13	-151.4	10.3	22.6	21.4	-1.0	20.9	-107.9	-150.7
C13-C12	175.1	177.3	164.4	175.1	176.6	162.9	174.4	174.8
C12=C11	-172.4	176.4	-174.7	-168.9	179.9	170.3	-171.1	-172.5
C11-C10	-171.7	-178.4	-179.5	-178.4	177.9	-179.6	-170.1	-172.3
C10=C9	175.5	171.0	173.8	-175.7	177.3	175.3	-173.5	175.9
C9-C8	-166.9	-173.9	-171.0	-175.9	179.6	-168.7	-173.0	-168.4
C8=C7	172.3	172.7	173.5	178.2	177.1	172.9	169.3	172.6
C7-C6	169.0	177.3	172.5	173.5	179.5	175.2	172.1	172.6
C6=C5	179.6	174.6	179.8	-178.9	176.2	-179.7	166.9	179.4

Appendix B

Evaluation of known ligands

B.1 MD simulation of known ligands

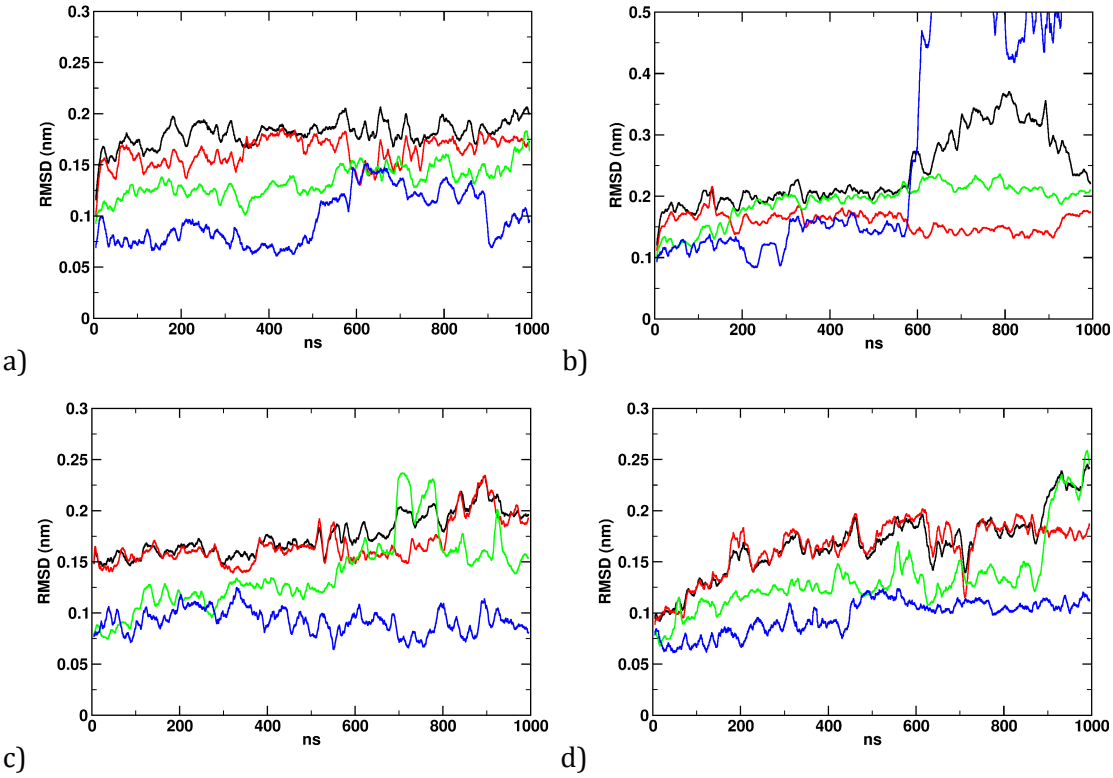
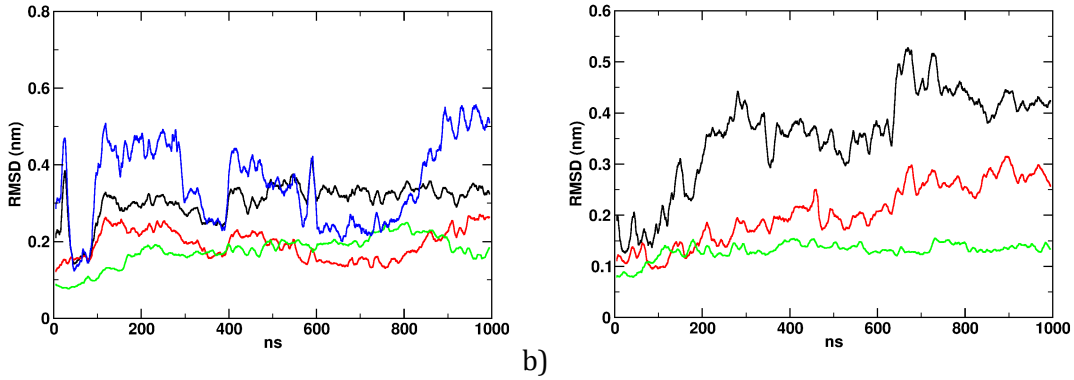


Figure B.1 – The RMSDs along the trajectories. Black: Backbone of the whole LBD; Red: Backbone of domain 1; Green: Backbone of domain 2; Blue: Non-hydrogen atoms of the ligand. a) AMPA; b) glutamate; c) 2-BnTetAMPA; d) iodo-willardiine



a) b)

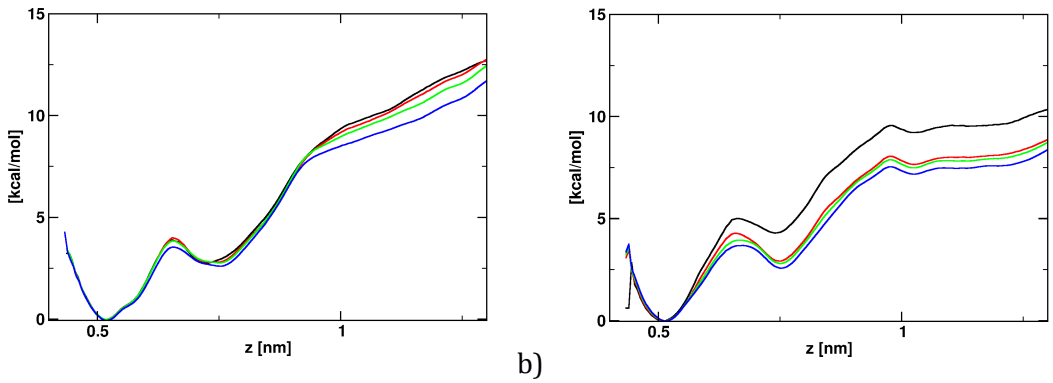
Figure B.2 – The RMSDs along the trajectories. Black: Backbone of the whole LBD; Red: Backbone of domain 1; Green: Backbone of domain 2; Blue: Non-hydrogen atoms of the ligand. a) DNQX; b) APO

Appendix C

Development of photo-switchable ligands targeting iGluR2

C.1 Photo-switch mechanism of ATA-2

C.1.1 LBD opening with bound ATA-2 in POS1



a) b)

Figure C.1 – The convergence of the PMFs with proceeding simulation time. Black: 150 ns in each window; Red: 200 ns; Green: 250 ns; Blue: 300ns.a) *trans*-ATA-2; b) *cis*-ATA-2

C.1.2 Translocation of ATA-2 from POS1 to POS2

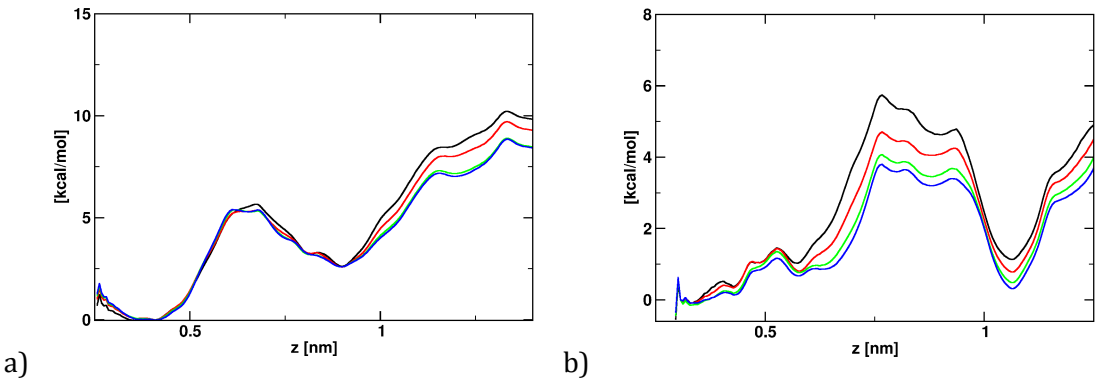


Figure C.2 – The convergence of the PMFs with proceeding simulation time. Black: 150 ns in each window; Red: 200 ns; Green: 250 ns; Blue: 300ns. a) *trans*-ATA-2; b) *cis*-ATA-2

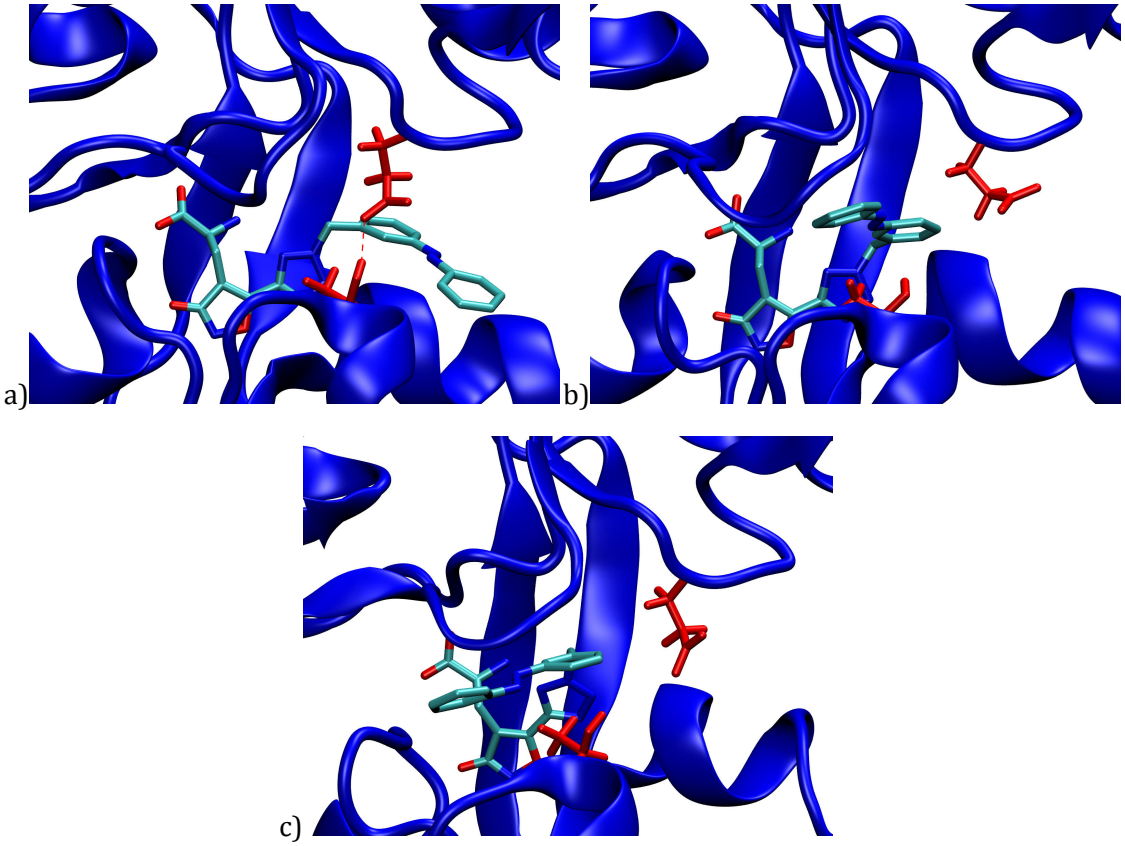


Figure C.3 – The different positions of *trans*-ATA-2 along the translocation from POS1 to POS2. a) POS1; b) intermediate state; c) POS2

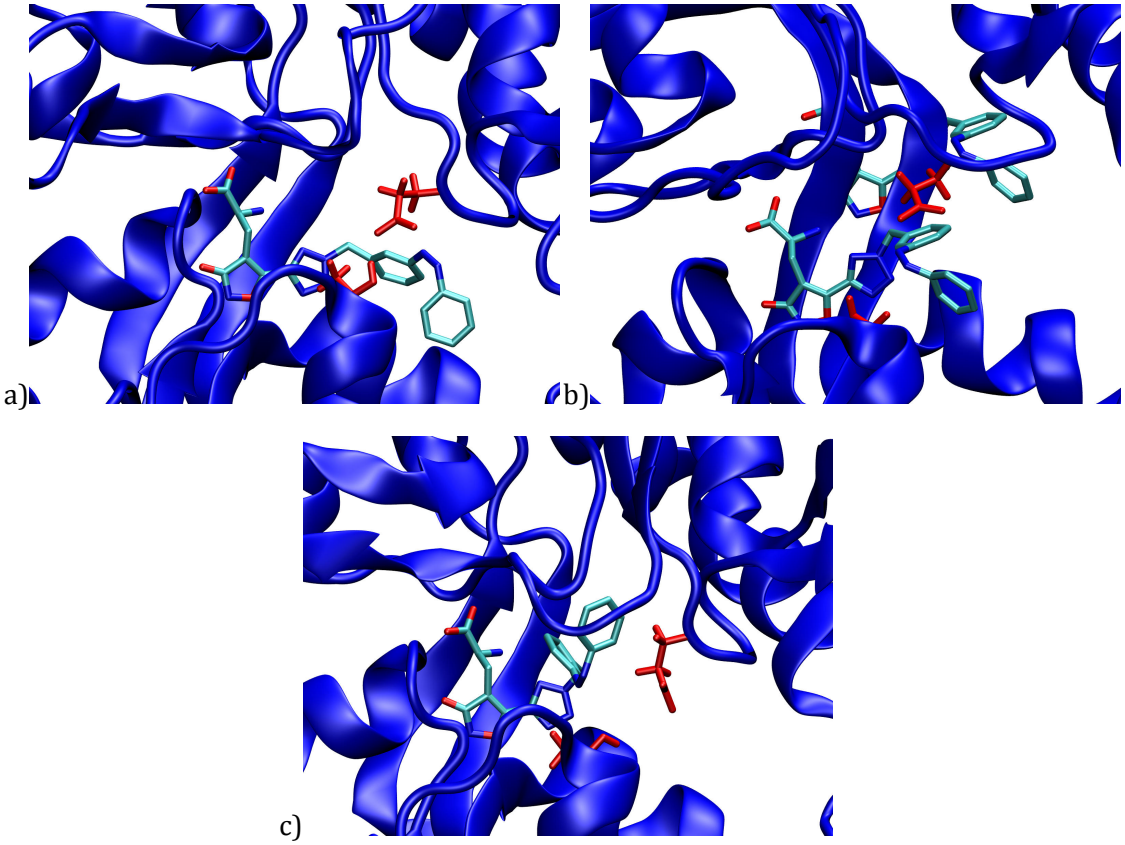


Figure C.4 – The different positions of *cis*-ATA-2 along the translocation from POS1 to POS2. a) POS1; b) intermediate state; c) POS2

C.1.3 LBD opening with bound ATA-2 in POS2

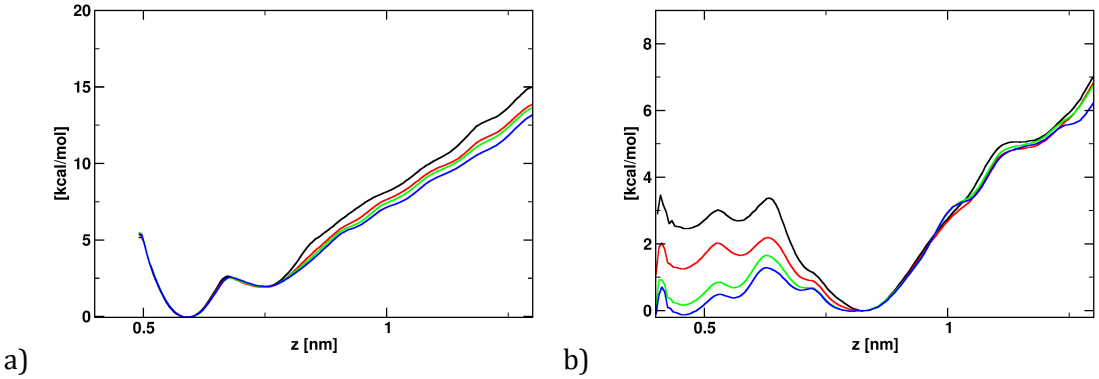
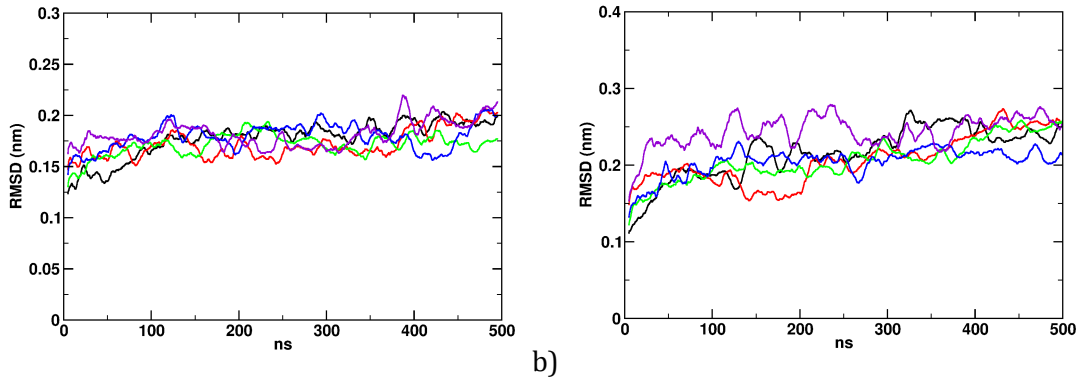


Figure C.5 – The convergence of the PMFs with proceeding simulation time. Black: 150 ns in each window; Red: 200 ns; Green: 250 ns; Blue: 300ns.a) *trans*-ATA-2; b) *cis*-ATA-2

C.2 Photo-switch mechanism of ATA-3

C.2.1 LBD opening with bound ATA-2 in POS2



a) b)

Figure C.6 – RMSD of the backbone atoms along the five trajectories. a) position 1; b) position 2