Theoretical Investigation of Kainate Receptor GluK2 and Channelrhodopsin-2: Structure and Mechanism

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von
Yanan Guo
aus
Shandong, China

KIT-Dekan: Prof. Dr. Willem M. Klopper
Referent: Prof. Dr. Marcus Elstner
Korreferent: Priv. Doz. Dr. Andreas-Neil Unterreiner
Tag der mündlichen Prüfung: 10.02.2017

Declaration

I hereby declare that the thesis is an original work. I have written this work independently and no other than the specified sources and aids are used.

Yanan Guo
Karlsruhe Institute of Technology
02.12.2016

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Dr. Marcus Elstner, who provided me the chance to access the interesting topic of “optogenetics” and the state-of-the-art computational techniques. I appreciate all his contributions of time, ideas and funding which are vital for the completion of my PhD.

The group of Prof. Dr. Dirk Trauner provided the crystal structure of the photoswitchable ligand *gluazo* bound to the kainate receptor, prompting my first research project. Dr. Tino Wolter introduced me to the fantastic protein–ionotropic glutamate receptor. He shared with me selflessly his experience in model building and GROMACS. Dr. Tomáš Kubař provided fruitful discussion in the study of kainate receptor. I benefited a lot from him in the refinement of the manuscript. He is always patient and enthusiastic in our discussions and helps me to solve the technical problems during my PhD pursuit. He maintains the group cluster, providing hardware and software assurance for the smooth progress of my calculations.

Dr. Hiroshi C. Watanabe and Dr. Marco Marazzi introduced me to the field of rhodopsins, which opened for me the door to switch to the study of channelrhodopsin-2. The PhD students Franziska E. Beyle and Beatrix M. Bold provided valuable suggestions and productive discussions. Prof. Peter Hegemann, an expert of rhodopsins, also gave useful suggestions in result analysis. The group of Prof. Dr. Walter Thiel provided technical support for the semi-empirical orthogonal method-2, allowing the analysis of the absorption spectrum. Dr. Kai Welke gave me great help in the computing of vibrational frequency and IR spectrum. Thilo Mast instructed me in the application of metadynamics for the free energy calculation. Dr. Igor Schapiro and Dr. Marco Marazzi are the mentors guiding me to the computational study of photochemistry.

I want to express my gratefulness to Dr. Tomáš Kubař, Dr. Jissy Akkarapattiakal Kuriappan, Dr. Natacha Gillet, Franziska E. Beyle, Beatrix M. Bold, Violetta Schneider, Julian Kranz and Marina Putzu for their contributions to the refinement of the thesis.

I want to thank our secretary Sabine Holthoff, who gave me great help for my stay in Germany. I would like thank all the other members of the group, the joy and enthusiasm they have make the group an appealing unity.

I want to thank the China Scholarship Council for supporting my life in Germany, and the Deutsche Forschungsgemeinschaft for providing financial support for my PhD study.

At last but not the least, I want to thank my loving, supportive, encouraging family. Their love is always the backing force supporting me through tough times in my life, no matter where I am and what I do (of course legally), and it will accompany me forever.

Yanan Guo
Karlsruhe Institute of Technology
02.12.2016

Zusammenfassung

Diese Arbeit befasst sich mit der Aufklärung der Struktur und des Mechanismus zweier verschiedener Proteine: der Kainat-Rezeptor Untereinheit (GluK2) und des Kanalrhodopsins-2 (ChR2). GluK2 ist ein ligandengesteuerter ionotroper Glutamat Rezeptor. Die Öffnung und Schließung des Kanals wird durch einen gebundenen photoschaltbaren Liganden gesteuert. ChR2 ist ein lichtgesteuerter Ionenkanal, dessen Öffnung durch direkte Lichteinstrahlung gesteuert wird. Beide Proteine sind wichtige optogenetische Werkzeuge. Da diese Proteine eine Aktivierung von neuronalen Zellen oder einen Signaltransfer zwischen den Synapsen durch Licht ermöglichen. Somit können neuronale Zellaktivitäten im Gewebe aufgezeigt und kontrolliert werden. Diese, als Optogenetik bezeichnete Technik, hat im Studium von neuronalen Krankheiten eine vielversprechende Anwendung gefunden.

Zu Beginn dieser Arbeit wurden die Modelle der Proteine gebaut. Danach wurden Multiskalenmethoden in Kombination mit klassichen molekular Mechanik und quantenmechanischen Methoden angewendet um die Beziehung zwischen Struktur und Funktion der Proteine zu untersuchen.

Der photochrome Ligand *gluazo* ist ein Azobenzen Derivat. Er verändert die Größe der Kanalöffnung des Ionenkanals GluK2 durch Änderung der molekularen Konfiguration bei Lichteinstrahlung. Die Veränderung der Proteinstruktur durch die Isomerisierung des *gluazo* ist bislang nicht gänzlich verstanden. Deswegen wurden zusätzliche Bindungsmodelle des *gluazo* (abweichend von der Kristallstruktur) mittels “Simulated Annealing” Simulationen untersucht. Die Bindungspräferenz wurde anhand von Freien Energie Berechnungen ermittelt. In der Kristallstruktur bindet das *trans-gluazo* mit dem Protein. Um den GluK2 Komplex mit gebundenem *cis-gluazo* zu erhalten und um gleichzeitig rechenzeitintensive Berechnungsmethoden zu vermeiden, wurde eine Methode verwendet bei der unter Verwendung eines “forced-switching” Protokolls die Struktur von den *trans* in den *cis* Zustand erzwungen wird. Diese Methode erfüllt alle Voraussetzungen um Modelle zu erhalten, damit das Verhalten des Proteins mit verschiedenen Liganden erforscht werden kann. Umfangreiche MD-Simulationen ergaben, dass *trans-gluazo* und *cis-gluazo* unterschiedlich interagieren. Dies führt zu unterschiedlichem Verhalten der beiden Liganden mit dem Protein: *trans-gluazo* ist ein Partialagonist, während *cis-gluazo* eine geringere Wirksamkeit als *trans-gluazo* hat.

Obwohl ChR2 und weitere verschiedene Varianten schon weitgehend in den Neurowissenschaften angewandt werden, ist die Erforschung ihres Mechanismus seit langem Gegenstand der Forschung. In dieser Arbeit wird die ChR2-C128T Mutante untersucht. Diese Mutante durchläuft einen langsameren Photocyclus. Die Lebensdauer des Zustandes, in welchem der Ionenkanal geöffnet ist, wird durch die Mutation deutlich verlängert. Ausgehend vom Homologiemodel der Mutante, wurden MD Simulationen mit einem klassischen Kraftfeld und QM/MM MD Simulationen durchgeführt. Dabei sind unterschiedliche Strukturmotive untersucht und analysiert worden. In den Simulationen weist das Aktive Zentrum von ChR2-C128T eine flexiblere Struktur auf, als das Aktive Zentrum des gut untersuchten Bacteriorhodopsin (bR). All diese Strukturmotive sind hoch gradig korreliert. Durch diese Komplexizität wird die Erforschung von ChR2 massgeblich erschwert. Zusätzlich wurde die Rationalität der erhaltenen Strukturmodelle durch Berechnung der Anregungsenergie geprüft.

Der Einfluss der Umgebung des Chromophors auf die Isomerisierung von ChR2 wurde mittels CASPT2//CASSCF/MM (CASPT2: Complete Active Space Second Order Perturbation Theory, CASSCF: Complete Active Space Self Consitent Field) untersucht. Der Isomerisierungspfad von drei verschiedenen Strukturmotiven mit einem jeweils anderen Aktiven Zentrum wurde berechnet. Die Studie zeigt, dass nur eine der drei Strukturmotive eine Isomerisierung durch eine Konische Überschneidung propagiert und zu einem strahlungslosen Zerfall führt. Das erhaltene Isomerisierungsprodukt wurde durch Berechnung des Infrarot-Spektrums analysiert und mit experimentellen Ergebnissen verglichen.

In dieser Arbeit wird die Komplexität dieser licht-sensitiven Proteine verdeutlicht, indem viele unterschiedliche aufwendige Methoden notwendig sind, um Erkenntnisse über die Sturktur und den Mechanismus zu erhalten.

Abstract

This work focuses on the structures and mechanisms of functions of two different proteins: the kainate receptor (GluK2) and the channelrhodopsin-2 (ChR2). GluK2 is a ligand-gated ionotropic glutamate receptor. Its channel opening and closing behaviors are controlled by the bound photoswitchable ligand. ChR2 is a light-gated ion channel, the opening of which is triggered directly by light. They are able to activate neuron cells or transfer signals between synapses upon illumination, and thus, control and monitor neuron cell activities in living tissues. This technique is called optogenetics. It has promising applications in studying neuronal diseases. The work starts with the structural model building of the proteins. The structure-function relationships of the proteins were investigated with multi-scale techniques, combining classic mechanical (MM) and quantum mechanical (QM) methods.

The photochromic ligand *gluazo* is an azobenzene derivative. It alters the opening extent of the ion channel of GluK2 by changing the molecular configuration upon illumination. The protein response to *gluazo* isomerization is not fully understood. Additional *gluazo* binding mode, different from that in the crystal structure, was searched through the simulated annealing simulations. The preference of available binding modes was evaluated by free energy calculations. In the crystal structure the protein binds with the *trans-gluazo*. In order to obtain the complex of GluK2 binding with *cis-gluazo* and avoid the computational expensive high level quantum mechanical method, a forced switching protocol was employed. This method fulfills the current requirement by providing models for studying protein response to different ligand conformations. Extensive molecular dynamics (MD) simulations reveal that *trans-gluazo* and *cis-gluazo* have different interactions with the protein. This leads to the different efficiencies of the two ligands: *trans-gluazo* is a partial agonist, while *cis-gluazo* has a relatively lower efficacy than *trans-gluazo*.

Although ChR2 and its various variants are widely applied in neuroscience, the investigation of their mechanisms of functions remains to be a hot topic. This study is about ChR2-C128T, which is a ChR2 mutant with slow photocycle kinetics. The mutation enlongates the life time of the open state. Starting from the mutant homology model, classical force field and combined QM/MM MD simulations were performed. Several structural motifs were investigated and analyzed. The active site of ChR2-C128T was found to be much more flexible than that of the well investigated bacteriorhodopisn (BR). All these motifs are highly correlated, which increases the complexity for studying ChR2. The rationality of the current models and samplings are proven by excitation energy calculations.

The impact of the surrounding environment on the chromophore isomerization of ChR2 was investigated using the CASPT2//CASSCF/MM method (CASPT2: complete active space second order perturbation theory, CASSCF: complete active space self consistent field). The isomerization energy pathways of three models with different active site structural motifs were calculated. The study shows that, only one of the models can propagate an isomerization through conical intersection in a radiationless decay. The received isomerization product was analyzed by computing the infared spectrum that was then compared with the experimental results.

The study shows the complexity of these light reactive proteins, while many different high level computational techniques are necessary to obtain findings about the structures and mechanisms.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

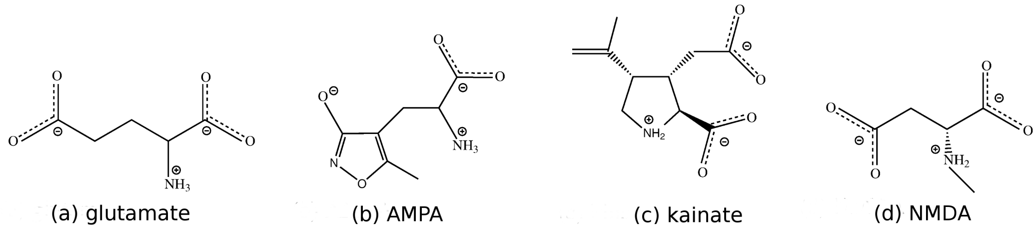
Optogenetics is a newly and rapidly developing, interdisciplinary field of research since 2002 [1]. It combines genetical and optical technologies to study neuronal networks on various spatial and temporal scales. It arises from the idea of using naturally occurring or chemically modified molecules which are reactive to light in order to excite or inhibit specific neurons. Optogenetic methods have promising applications in studying symptoms and circuits relevant to diseases such as blindness, narcolepsy, autism, Parkinson's disease. Since optogenetics is a young discipline, many of its tools are currently under investigation. The well known conventional optogenetic tools are the light-gated protein switches, i.e. photoreceptors, comprising of channelrhodopsins, halorhodopsins, sensory rhodopsins etc. Besides, the ionotropic receptors, G-protein coupled receptors are also potent optogenetic tools. They function through synthetic on-off switches or tuning elements (e.g ligands) that are attached to the signaling protein of interest, allowing the control of protein behavior under light pulse at certain wavelength. In order to improve and extend the applications of optogenetic methods, the development of optogenetic tools is essential, which further requires a thorough understanding of the structures, functions, and signaling mechanisms of current tools at the molecular level. This thesis studies two popular optogenetic tools, ionotropic glutamate receptor and channelrhodopsin.

1.1 Ligand-Gated Ionotropic Glutamate Receptor

Glutamate (Figure 1.1(a)) is the most important excitatory neurotransmitter for normal brain function. The synaptic actions of glutamate are regulated by glutamate receptors that predominantly distribute on central neuronal synapses. The family of glutamate receptors can be classified into two groups based on the mechanism by which the protein generates the excitatory postsynaptic current (EPSC) [2]: (i) ionotropic glutamate receptors (iGluRs), which form the ion channel pore upon the binding of proper ligand, also regarded as ligand-gated ion channels; (ii) metabotropic receptors (mGluRs), which modulate ion channels through a signaling cascade that involves G proteins, also referred to as G-protein coupled receptors.

The iGluRs are essential for the fast synaptic transmission between neurons. Three subfamilies of iGluRs have been identified, and they are named after their specific agonists (Figure 1.1): (i) α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid (AMPA); (ii) (2S,3S,4S)-3-(carboxymethyl)-4-prop-1-en-2-ylpyrrolidine-2-carboxylic acid (kainate); (iii) N-methyl-D-aspartate (NMDA). All of the iGluRs are nonselective cation channels, allowing the passage of K^+ and Na^+ to produce EPSC. Especially, NMDA receptors (NMDARs) are permeable to Ca^{2+}[3], which is proposed relating to long-term potentiation (LTP) and long-term depression (LTD) of the synapse.

Despite all iGluRs produce EPSC, the speed and duration of the current, and the physiological functions are distinct for each subfamily [4]. AMPA receptors (AMPARs) produce a fast increase and decrease of the electrical signal. They are responsible for fast excitatory synaptic signalling in the brain. Additionally, they are involved in the modulation of synaptic plasticity which is considered to be crucial to the learning and memory ability of mammalians. NMDARs show a voltage-dependent block of the channels, which could be relieved by depolarization of the membrane. The openning of NMDA receptors relies on the EPSC produced by AMPA receptors [5] . A slow EPSC generated by Kainate receptors (KARs) was ever reported [6]. Similar to AM-



(a) glutamate (b) AMPA (c) kainate (d) NMDA

Figure 1.1: Chemical structures of (a) glutamate; (b) AMPA; (c) kainate; (d) NMDA.

PARs, KARs mediate excitatory synaptic signals. Besides, KARs are involved in modulating the presynaptic release of neurotransmitters, and thus, in regulating the strength of synaptic connections. KARs have become the potential therapeutic targets in the treatment of pain and epilepsy [7].

Structrually, iGluRs are composed of four subunits, which assembles into dimer of dimers [8]. The tetrameric assembly can be heteromeric or homomeric. The multiple subunits of each subfamily are shown in Table 1.1. Each subunit has a typical modular construction (see Figure 1.2), consisting of an amino-terminal domain (NTD), a ligand-binding domain (LBD), a transmembrane domain (TMD), and a carboxyl-terminal domain (CTD). NTD and LBD are extracelluar domains, while CTD is an intracelluar domain [9].

The functions of NTD are still unclear. It might be involved in processes of subunit assembly, allosteric regulation and receptor trafficking [10–12]. The viewpoint that NTD could bind ligands is supported by the sequence homology and structural similarity between NTD of iGluRs and the LBD of group I mGluR (mGluR1a), and also a group of soluble bacterial periplasmatic amino acid binding proteins [8]. However, studies on subunits that lack the entire NTD reveal that these truncated subunits could assemble into receptors in which the core functions are preserved [13].

The soluble LBD is highly conserved among the different iGluR subfamilies. All LBD structures adopt a clamshell-like conformation, consisting of two half domains D1 and D2, which are mainly composed of the S1 and the S2 residue segment, respectively. The binding site for the ligand is located within the cleft between D1 and D2. Previous studies validate that for AMPARs, KARs and NMDARs, the ligand binding sites of LBD crystal structures resemble the binding sites in intact receptors [14–17]. Spectroscopy studies of AMPARs in the presence of antagonist also show that the structures of the isolated GluA2 LBD and the LBD in the full-length receptor are similar [18].

The TMDs partially build the channel pore in the iGluRs tetramer. It includes three helices (TM1, TM2, TM3) and one re-entrant pore loop. In all iGluRs, the conserved TMD is connected to the LBD through short linker residues. As shown in Figure 1.2(a), in KAR subunit GluK2, K544 and P667 connect D2, TM1 and TM2, respectively. In the X-ray structure of GluK2 LBD, S1 and S2 are usually fused through a hydrophilic linker between K544 and P667 [19, 20].

The CTD is the most diverse domain in terms of amino acid sequence. It varies in sequence and in length among iGluRs subunits. The CTD might be involved in membrane targeting, stabilization and post-translational modifications [8].

Table 1.1: Subunits of ionotropic glutamate receptors.

Receptors	AMPARs	NMDARs	KARs
Subunits	GluA1	NR1	GluK1
	GluA2	NR2a	GluK2
	GluA3	NR2b	GluK3
	GluA4	NR2c	GluK4
		NR2d	GluK5

1.1.1 Mechanism of iGluRs Channel Gating

The gating of iGluRs channel involves three sequential steps: 1) binding of the agonist to the LBD (docking of the agonist); 2) cleft closure of the LBD, preventing the dissociation of the agonist (locking of the agonist);

3) a conformational change of the ion channel. The UV and IR spectroscopy measurements show that the ligand initially forms a contact with the D1 domain of LBD, followed by a slower D2 transition to form further ligand-protein and D1–D2 interactions [21, 22] . In the tetrameric structure of iGluRs, the LBDs are arranged as back-to-back dimer interface on the D1 domain (see Figure 1.2(b)), which constrains D1 movement. In contrast, the D2 domain could move relatively freely. This is interesting considering the fact that D2 domain is connected to the ion channel-forming transmembrane helices through the linker segments on the bottom of D2. It makes the transition from LBD motion to the ion channel motion possible. Studies have shown that the linker distance increases in LBD dimer upon agonist binding. Thus, the binding of agonist leads to cleft closure of LBD, which involves movement of D2, while D1 remains relatively fixed. The D2 transition further leads to the rearrangement of transmembrane helices and drives the channel opening (see Figure 1.2(b) [23, 24].

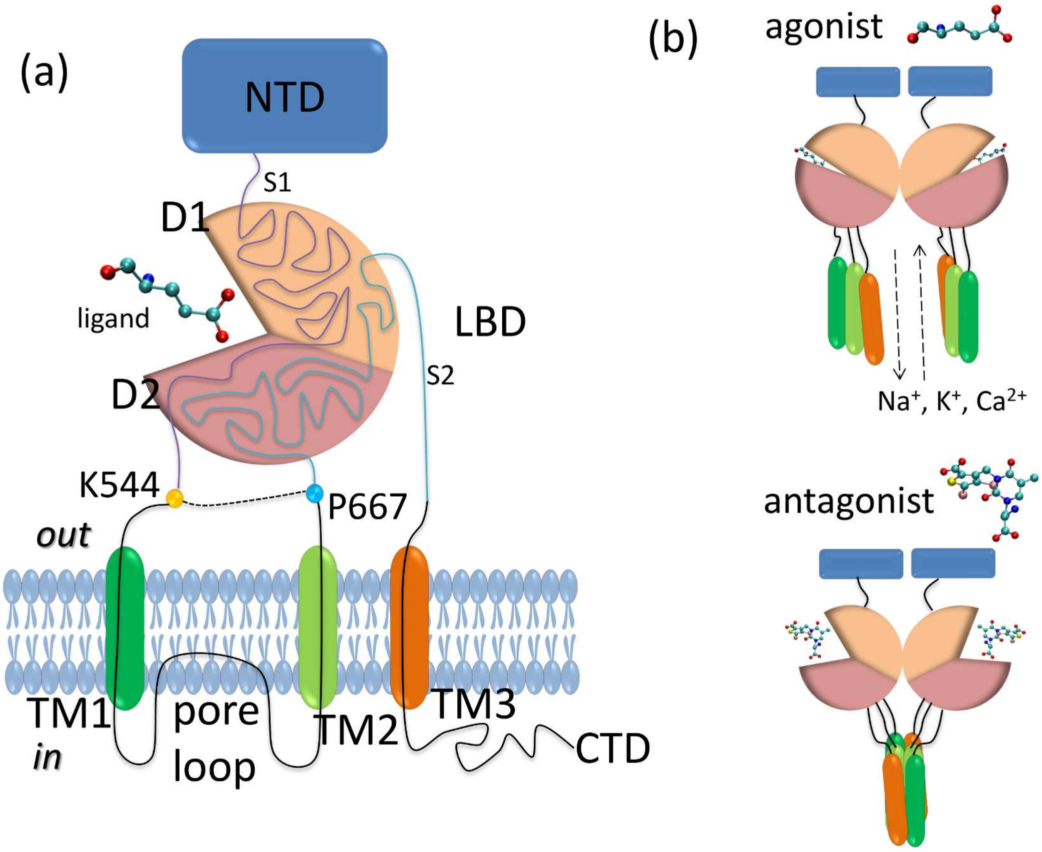


Figure 1.2: (a) Schematic illustration of the Gluk2 subunit topology. NTD: amino acid domain; LBD: ligand binding domain; TM1–3: transmembrane domain helices; CTD: intracellular domain. The D1 domain of LBD is mainly composed of the S1 segment (purple), and D2 domain of the S2 segment (blue). K544 (orange ball), P667 (blue ball) connect D2, TM1 and TM2, respectively. The dotted line depicts the hydrophilic linker connecting S1 and S2. (b) Schematic mechanism of ligand-binding-triggered channel opening. The binding of agonist leads to the full open of the channel; the receptor does not react to the binding of antagonist.

The cleft closure hypothesis is widely accepted by the researchers, which argues that the extent of LBD cleft closure is correlated with the ligand efficacy, i.e. the agonist leads to a larger cleft closure extent than the partial agonist. The cleft remains in an open state upon the binding of the antagonist [17, 19, 24, 25]. However, exceptions were found for NMDARs based on the studies of NR1 subunit. The partial agonists, 1-aminocyclopropane-1-carboxylic acid and 1-aminocyclobutane-1-carboxylic acid [26] lead to similar extent of LBD cleft closure as the agonist glycine [27].

Extensive efforts have been made to study the correlation between LBD cleft closure extent and channel conductance, which leads to a popular argument that the channel activation level is dependent on the LBD closure extent. More in detail, the large conformational change of LBD (i.e. large closure extent) induced by agonists results in the full channel opening; in contrast, the stabilized open conformation of LBD by antagonists precludes the opening of the channel [4, 17].

1.1.2 Photochromic Ligands

The use of photochromic ligands (PCLs) introduces the intrinsically light insensitive iGluRs to optogenetics. PCL was firsltly reported in the late 1960s as a photoisomerizable inactivator of chymotrypsin [28]. PCLs refer to a group of compounds which noncovalently bind to the target protein and change the configuration reversibly under illumination. This change in shape results in an alteration in the efficiency of the ligand, e.g. from agonist to antagonist or vise versa (Figure 1.3). PCLs have been investigated for various classes of target proteins, inluding ligand-gated ion channels [29, 30], enzymes [31], and receptors [32]. The most commonly used PCLs are derived from azobenzene due to its distinct configurations in *trans* and *cis* form. For instance, the compound studied in this thesis, named *gluazo* (Figure 1.4) is one of the synthesized PCLs based on azobenzene [30].

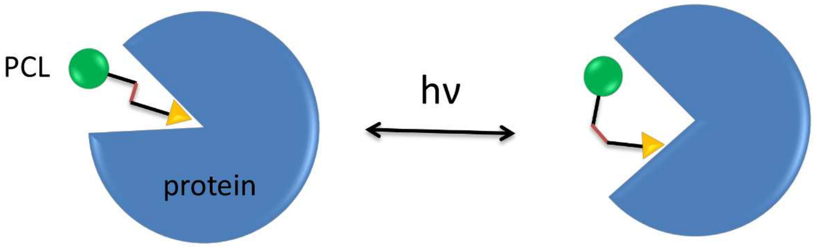


Figure 1.3: Illustration of PCL functions mechanism. The molecular configuration of PCL changes with photoisomerization, altering the ligand-protein interactions, and thus, results in different protein conformational reaction.

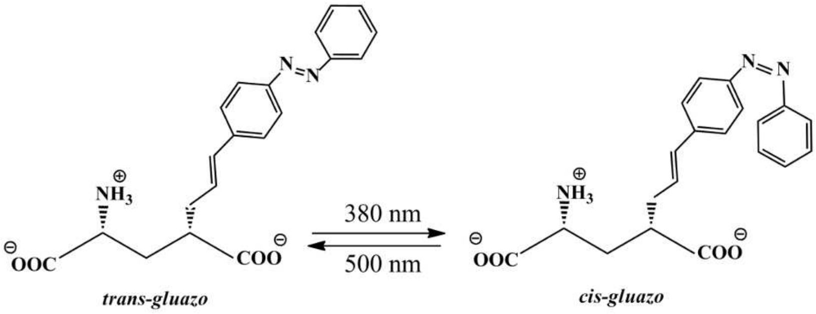


Figure 1.4: The photo-isomerization of zwitterionic *gluazo*. Figure is reproduced from ref. [33] with permission.

1.2 Rhodopsins

The rhodopsins are photochemically reactive membrane proteins containing the retinal (vitamin-A aldehyde, Figure 1.5) as their functional chromophore [34]. These proteins share some general structrual characteristics, i.e. they consist of seven transmembrane α-helices which are arranged to form an internal pocket, in which the retinal is covalently bound to the ε-carbon atom of a lysine in the seventh helix (or helix 7) through a protonated Schiff base (Figure 1.5). The light-triggered transfer of Schiff base proton is essential for the functions of rhodopsins.

Two types of rhodopsins are defined considering three different structural aspects: (i) primary amino acid sequences; (ii) helices arrangement; (iii) isomeric configuration and conformation of β-ionone ring or polyene chain of the retinal. Type I rhodopsins, also called microbial rhodopsins, are mainly found in prokaryotes such as archaea, bacteria, and now also in some eukaryotes microbes like fungi. Bacteriorhodopsin, halorhodopsin, sensory rhodopsins and channelrhodopsins all belong to this type. They perform diverse functions mainly within the membrane, as ion pumps, phototaxis receptors and ion channels. Type II rhodopsins, also called animal rhodopsins, are all found in higher eukaryotes so far. They are photosensitive receptor proteins, and

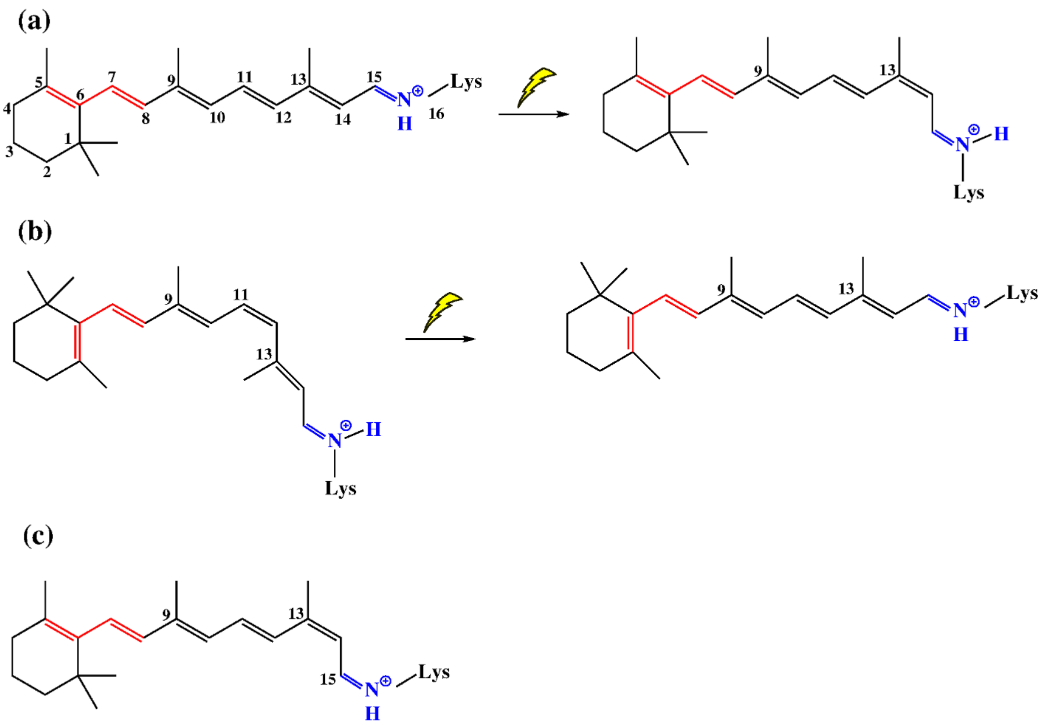


Figure 1.5: Structure of the retinal. (a)The all-*trans* retinal in type I rhodopsins, and its photoisomerization to 13-*cis* configuration under illumination. (b) The 11-*cis* retinal in type II rhodopsins, and its photoisomerization to all-*trans* configuration under illumination. (c) The 13-*cis*-15-*syn* retinal possibly coexist with all-*trans* retinal in type I rhodopsins. The Schiff base is emphasized in blue color, and the ring/polyene chain conformation is emphasized in red color.

half of the protein mass are hydrophilic loops which bind with G proteins, receptor kinases, and other signaling proteins in the cytoplasmic side [35–37]. Most of the amino acid residues within the retinal binding pocket are conserved among each type of rhodopsins or replaced by similar residues. However, they are quite different between type I and type II rhodopsins [37–42]. Besides, the electron density maps resolving helices arrangements show very different patterns between type I and type II proteins [38, 43–47].

Another difference between the two types of protein lies in the configuration and conformation of the retinal. In type I rhodopsins, the functional retinal configuration is all-*trans*, which is isomerized to 13-*cis* after photoexcitation (Figure 1.5(a)). The retinal stays attached to the apoprotein in the subsequent functional process undergoing the Schiff base deprotonation, reprotonation, the retinal reisomerization etc. and finally the protein goes back to the dark state. This procedure could be repeated and is commonly referred to as a photocycle. In type II rhodopsins, the dark state retinal configuration is 11-*cis*. It is converted to an all-*trans* conformer during the photoisomerization (Figure 1.5(b)). After that, the retinal is hydrolyzed from the apoprotein, which then leads to a different functional mechanism from type I [48]. The conformation of the β-ionone ring with respect to the polyene chain of the retinal is also different between type I and type II rhodopsins, which could be depicted through the conformation around the C6-C7 single bond. As shown in Figure 1.5, the retinal adopts the more planar ring/chain conformation (6-*s*-*trans*) in type I protein, while it is the more twisted 6-*s*-*cis* conformation in type II protein.

Bacteriorhodopsin (BR) is the most extensively studied microbial rhodopsin. The well understood details of its structure and functions can be found in literature [38, 49–53]. It is often used as the reference in the study of other microbial rhodopsins, which is also the case in the present study focusing on Channelrhodopsin-2.

1.3 Channelrhodopsins – Light-Gated Cation Channels

Channelrhoposins (ChRs) are a subfamily of microbial rhodopsins, which are found in algae. The first channelrhodopsins were discovered from the green algea *Chlamydomonoas reinhardtii*, which are named as Channelrhodopsin-1 (ChR1) and Channelrhodopsin-2 (ChR2) [54, 55]. Subsequently, a number of ChR genes are identified in various algae, such as VChR1 and VChR2 which are from the colonial algae *Volvox carteri* [56, 57]; MCbR from *Mesostigma viride* [58]; PsCbR from marine algae *Platymonas subcordiformis* [59] and CaCbR1 from *Chlamydomonas augustae* [60].

Both ChR1 and ChR2 are directly light-gated cation-selective channels, which means that the channels open rapidly after absorption of photon(s) without the involvement of transducer or secondary messenger. subsequently, certain cations are conducted [54, 55]. ChR1 has a high selectivity for proton, and it functions as a passive (from low concentration to high concentration) light-gated proton channel [54]. ChR2 shows passive permeability to both monovalent cations (Na^+, K^+ etc.) and divalent cations (Ca^{2+} etc.), and its photocurrents have an inverse relationship with the atomic radius of the cations [55]. The absorption maximum of ChR1 is at ca. 510 nm, and the protein mediates cell phototaxis at high-intensity light [61]. ChR2 absorbs maximally at ca. 470 nm and it mediates cell phototaxis at low-intensity light [61].

ChR2 is currently the best investigated ChR because of its superior expression in host cells, and of the pH-independent retinal absorption in the range pH 4-9 [62]. It was also the first ChR applied to activate neurons under illumination [63–65], and since then it has been developed as a predominant optogenetic tool.

1.3.1 Structure of Channelrhodopsins

Helices Arrangement

At present, an X-ray structure is available only for a ChR1/ChR2 chimera (C1C2) (PDB code: 3UG9). C1C2 is a hybrid composed of helices 1-5 from ChR1 and helices 6-7 from ChR2 [66]. A homologous model of ChR2 N-terminal membrane-spanning domain, using C1C2 as template is shown in Figure 1.6. Experimental studies have revealed that the seven-transmembrane-helices structure is sufficient for studying the channel activity [54, 55]. The electron microscopic structure reveals that ChR2 exists in dimers, where the third and the fourth transmembrane helices (H3 and H4) comprise the dimer interface (Figure 1.6(g)) [67].

The Retinal

The retinal is bound to a lysine residue (K257 in ChR2) which is conserved in all microbial rhodopsins through a protonated retinal Schiff base ($RSBH^+$) in dark state (Figure 1.6(b)). The viewpoint of protonated Schiff base is verified by the down-shifted fingerprint resonance Raman frequency of C=N stretching in H_2O/D_2O exchange [68]. The absorption property of the retinal is basically determined by several structural factors of the active site: (i) the retinal conformation, including the planarity of the conjugated polyene chain of the retinal and the bond length alternation (BLA), (ii) the hydrogen bonding interactions of $RSBH^+$ with surrounding residues or water molecules, (iii) the electrostatic interaction of $RSBH^+$ with charged residues. In ChR2 dark state, the retinal shows a global absorption maximum (λ_{max}) at 473 nm (2.62 eV) [69], which is ca. 90 nm blue shifted with respect to BR (560 nm, 2.21 eV) [70, 71]. Additionally, different from the single-peak absorption spectrum of BR, ChR2 has two absorption sub-peaks at 442 nm (2.81 eV) and 414 nm (3.00 eV). Lórenz-Fonfría et al. [72] reported that the absorption peaks of ChR2 arise purely from the $S_0 \rightarrow S_1$ electronic transition. This might indicate that a single ChR2 active site may produce the three-peak absorption spectrum due to various initial vibrational levels. On the other hand, it is noteworthy that, the characteristic single-peak spectrum of BR was measured from a structure with a typical active site of the so-called pentagonal cluster, being composed of three water molecules and one oxygen from each of D85 (E123 in ChR2) and D212 (D253 in ChR2) [38, 73], see Figure 1.7. Therefore, multiple active site structures might exist in ChR2 dark-adapted state, inspired from the (one structure)–(one absorption peak) relation in BR.

The retinal configuration in the dark state remains under debate. The resonance Raman spectroscopy measured at 293 K shows a C=C stretching (ethylenic mode) at 1550 cm^{-1} for the all-*trans* retinal, and at 1557 cm^{-1} for the 13-*cis*,15-*syn* retinal; the fingerprint band of C-C stretching at 1200 cm^{-1} for the all-*trans* retinal, and at 1183 cm^{-1} for the 13-*cis*,15-*syn* retinal [68]. The FT-IR difference spectroscopy at room temperature detected two negative bands at ca. 1552 cm^{-1} and 1558 cm^{-1}, which were assigned to the ethylenic vibrations of all-*trans* and 13-*cis*,15-*syn* retinal in the dark state by the authors [74]. Controversially, the solid state NMR study at 100 K suggests a pure all-*trans* retinal configuration [75].

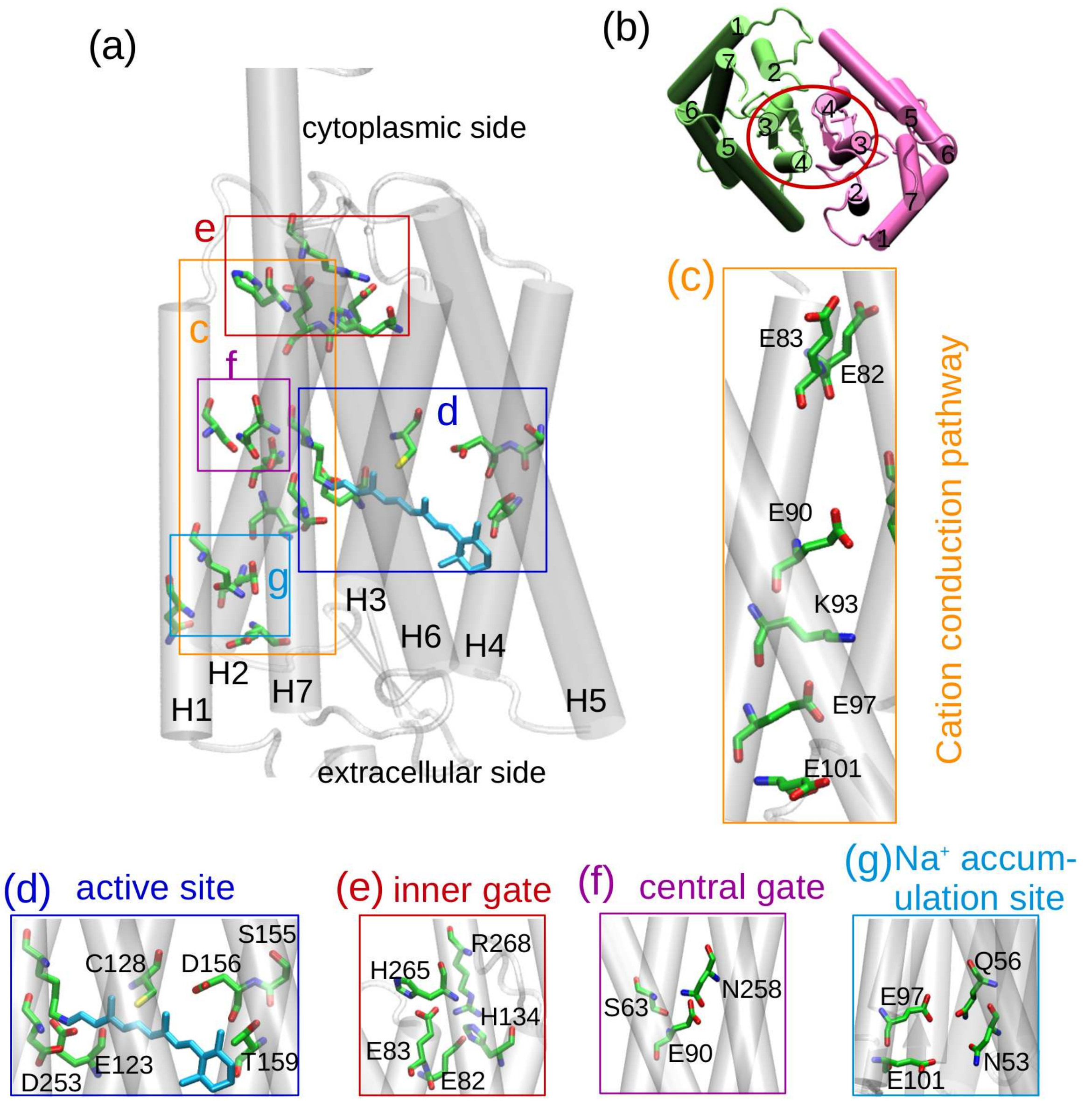


Figure 1.6: Seven-helices homologous model of ChR2 using a C1C2 X-ray crystal structure as template. (a) Overall monomer structure with framed residues of interest; (d) active site and retinal binding pocket; (c) ion permeation pathway (channel pore); (e) inner gate; (f) central gate; (g) Na^+ accumulation site; (b) overall dimer structure, the monomer-monomer interface region are labeled by red circle. The retinal is represented in cyan.

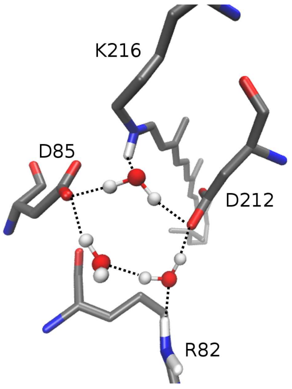


Figure 1.7: The bacteriorhodopsin active site, which is composed of three water molecules and one oxygen from each of D85 and D212.

Active Site and Retinal Binding Pocket

In ChR2 (C1C2), E123 (E162) and D253 (D292) (homologous to D85 and D212 in BR) form the potential counterion complex of $RSBH^+$. In C1C2 X-ray structure, the side chain carboxyl oxygens of E162 and D292 are close to the retinal Shiff base nitrogen with the distance of 3.4 Å and 3.0 Å, respectively. Empirical pK_a calculations [66] suggest that in C1C2, as well as in ChR1, E123 is neutral (at least in the crystal condition) and D253 is negatively charged, while in ChR2 both E123 and D253 are negatively charged. This argument is supported by spectroscopic measurements [68, 76–78] and theoretical simulations [79, 80]. Besides, in ChR2 the negatively charged E123 side chain prefers to form intrahelical hydrogen bond with T127 hydroxyl group [80].

Early studies proposed that E123 contributes to the counterion of $RSBH^+$, and serves as the primary acceptor of $RSBH^+$ proton [81, 82]. However, recently Heberle et al. [83] concluded on the basis of time-resolved step-scan Fourier transform infrared (FTIR) spectroscopy that the proton acceptor in ChR2 is D253, therefore proposing this residue as a prominent $RSBH^+$ counterion; whereas Kuhne et al. [78] showed by time-resolved FTIR that E123 and D253 are protonated, simultaneously with $RSBH^+$ deprotonation, indicating both residues might be the primary $RSBH^+$ proton acceptors.

Water might also play an important role in the ChR active site. In the C1C2 chimera only one water molecule (w619) was resolved near the $RSBH^+$, at a too large distance (4.4 Å) to be considered as direct proton acceptor of the retinal [66]. Nevertheless, Watanabe et al. [80] suggest more internal water molecules existing through the three-dimensional reference interaction site model (3D-RISM) calculation [84, 85]. The classical molecular mechanics (MM) simulations predicted more water molecules to enter the ChR2 active site than in BR [80]. The combined quantum mechanics/molecular mechanics (QM/MM) simulations sampled a direct hydrogen bond between the $RSBH^+$ and a water molecule [80], consistent with the down-shifted resonance Raman spectra of the $RSBH^+$ C=N vibration by H_2O/D_2O exchange [68, 69].

Quite a few engineered ChRs have been explored in order to expand its biological applications. Point mutant of D253E leads to a threefold slower decay of the P_2^{390} state, and correspondingly, retards channel closure. The D253N mutant does not exibit channel behavior [83]. The ChR2-E123T accelerated (ChETA) variants, a category including E123T, E123A, E123Q and combination mutants leads to smaller photocurrent amplitude compare with the wild type (WT) protein. However, they speed up the channel kinetics, and are widely used for stimulating of fast neuronal action [76, 86].

Besides, the residue pair C128 (T90 in BR) and D156 (D115 in BR), locating on helices 3 and 4 respectively (Figure 1.6 (b)), are suggested to be involved in the photoreaction. Individual mutation of either C128 or D156 leads to strongly delayed closing of the channel with respect to the WT [82, 83]. Double mutant of both residues stabilized the photocycle in the gating state, with little detectable return signal to the dark state after 30 min [87].

The hydrogen bond interaction between D156 and C128 is normally referred to as the DC gate, and it is relevant to channel gating and photocycle kinetics [88]. Currently, the structure of the DC gate is still controversial. The FTIR difference spectra from Nack et al. [89] showed that D156 should be protonated and hydrogen bonded in the dark state with the COOH vibrational frequency (νCOOH) at 1737 cm^{-1}. The same group proposed a direct hydrogen bond between the side chains of C128 and D156, whereas C128 serves as the donor [89]. This structural motif is similar to the homologous residues T90 and D115 in BR, where the hydroxyl oxygen of T90 side chain is the hydrogen bond acceptor of the carboxyl group of D115 [90, 91]. However, the structural basis for this viewpoint could not be verified by the C1C2 X-ray model, where the thiol group of C167 (C128 in ChR2) points away from D195 (D156 in ChR2) [66]. Relatively, theoretical studies reported the possibility of a water molecule bridging C128 and D156 in the WT [79, 92]. The calculated νCOOH of D156 was 1740 cm^{-1}, close to the experimental value of 1737 cm^{-1}. Especially, in C128T mutant, a direct hydrogen bond between T128 and D156 is suggested with D156 as the donor [92]. The νCOOH of D156 was calculated to be 1743 cm^{-1}, which is in good agreement with the experimental value of 1747 cm^{-1} [89].

The residue T159 is also in the vicinity of the retinal, and it forms intrahelical hydrogen bond with S155 (Figure 1.6 (b)). Mutation of T159 to cysteine, i.e. T159C mutant, results in a dramatic enhancement of photocurrent amplitudes but slows down channel kinetics in contrast to the WT. The double mutant E123T/T159C combines the rapid channel kinetics with large photocurrents, and thus, it becomes an excellent candidate for stimulating cells deep within neuronal tissue or with low light intensity [86, 93]. Moreover, a fluorescence microscopy study reveals that the substitution of T159 with aromatic amino acids (T159Y/F/W) improves the resistance of the protein to degradation in the absence of the retinal. Thus, the residue T159 is important for studying the stabilization of the apoprotein [94].

Cation Conduction Pathway

The pore formed by H1, H2, H3 and H7 was proposed to be the cation conduction pathway of ChRs [66], considering the great number of charged residues on the first two helices, especially on H2, compared to other helices. These charged residues have an inward orientation due to the nature of α-helix. The inward oriented charged groups create a hydrophilic environment that facilitates the formation of a water channel connecting the cytoplasmic and extracellular bulk phase [80]. In ChR2, these charged residues are E82, E83, E90, K93, E97, and E101 (Figure 1.6 (c)). The individual replacement of E82, E83, E97, and E101 by alanine or glutamine, leads to a significant decrease of the photocurrent amplitude with respect to WT, while the E90A/Q mutant does not change the photocurrent [95]. Consistently, the E90 is supposed to deprotonate in the early photocycle state and flips downward to facilitate the channel opening [78]. Furthermore, molecular dynamics (MD) simulations reported two Na^+ accumulation sites, which might provide an insight into the potential ion binding sites in the protein. One site is comprised of N53, Q56, E97 and E101 on the extracellular side (Figure 1.6 (f)); the other is close to E82, E83, H134, H265 and R268, which is on the cytoplasmic side and is often referred to as the inner gate (Figure 1.6 (d)) [80].

Mutagenesis studies have shown that several positions are crucial for cation selectivity, including Q56, S63, N258, and H134 [86, 96]. Residues S63, N258 together with E90 form the central gate (Figure 1.6 (e)) that is thought to block the channel [97]. Additionally, the inner gate might be also involved in keeping the channel internally closed through hydrogen bonds [98].

1.3.2 Photocycle Models of Channelrhodopsins

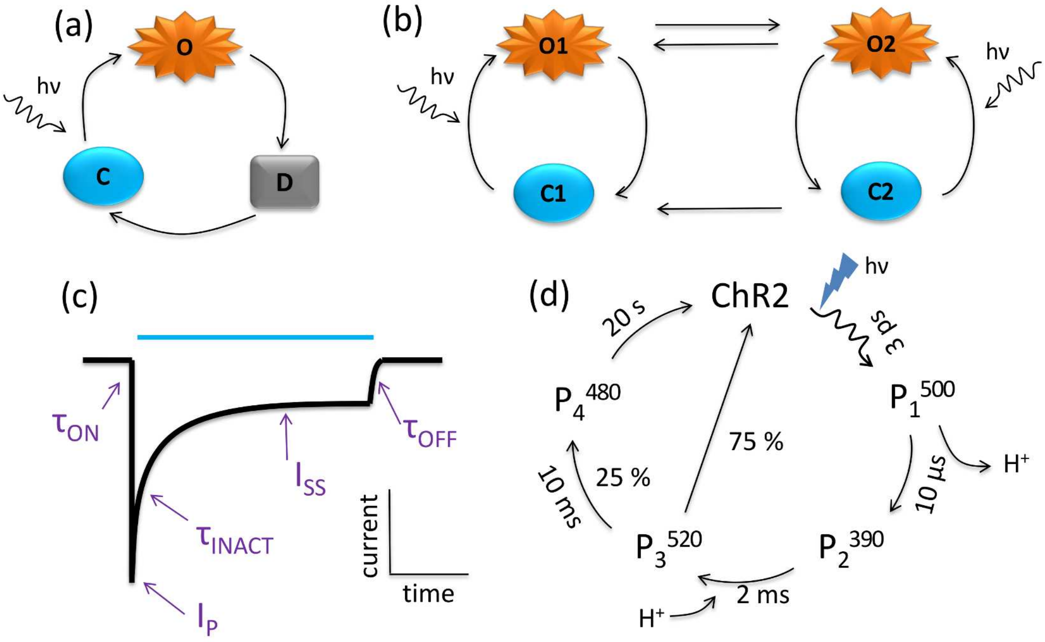


Figure 1.8: Schematic photocycle models and photocurrent of ChR2. (a) The three-state model [99], C: closed state, O: open state, D: desensitized state. The ultrafast transition to the excited state (C→C*) is eliminated from the model. (b) The four-state model [100], C1 and C2: closed states, O1 and O2: open states. (c) The photocurrent under continuous illumination (the blue bar) [100]. I_P: peak current, I_{SS}: stationary state current, τ_{ON}: activation time constant, τ_{INACT}: inactivation/desensitization time constant, τ_{OFF}: decay time constant. (d) The unidirectional photocycle proposed under single turnover condition [83], comprising four intermediate states (P_1^{500}, P_2^{390}, P_3^{520}, P_4^{480}). The photocycle is split at the level of P_3^{520} state: 75 % molecules relax directly to original ground state, 25 % molecules relax to P_4^{480} state.

The study of the photocycle of ChR2 has been one of the main interests in the optogenetic field since its first discovery [55]. There are currently two major models: a three-state model and a four-state model, as shown in Figure 1.8. The three-state model was initially proposed by Nagel et al. [55]: the closed/sensitized state (C) undergoes ultrafast transition (< 1 ns) to an excited state (C*), followed by a slower reaction to the open state (O) with a time constant of < 1 ms; the O state is closed to a desensitized state (D) (The state where the ion channel is closed, but the molecule is not ready for photoexcitation yet) with a time constant of 0.2 ms at pH=7.3; state C recoveries from state D slowly within 2 s. With such time constants, the three-state model predicts a much smaller stationary photocurrent than experiment and is not able to reproduce the characteristic peak-plateau photocurrent of ChR2 (Figure 1.8 (c)). Besides, the study shows different recovery rate (D→C, τ_{OFF}) under constant illumination from that in dark. To account for these observations, the authors postulated a four-state model with two closed states and two open states (one with large cation conductance, the other with reduced cation conductance).

After that, Hegemann et al. [101] examined several photocycle models in more detail for ChR1, and suggested a four-state two-cycle model (Figure 1.8 (b)) which provides a good reference for the subsequent investigation of ChRs photocycle. In this model, the open states O1 and O2 are reached by photoexcitation of closed states, respectively; O1 and O2 thermally decay back to C1 and C2, respectively; O1 and O2 are connected through reversible thermal transition; while C2 irreverisbly transits to C1. The stationary state of the photocurrent (I_{SS} in Figure 1.8 (c)) is explained by the relatively lower quantum efficiency of the C2→O2 transition.

Later, Nikolic et al. [99] modified the three-state model and found the main limitation of it: this model does not provide sufficient parameters to capture all the essential features of ChR2 photocurrent, especially, the proposed D→C biexponential decay. However, Stefanescu et al. [102] reported opposite conclusion with the prerequisite of nonequilibrium initial distribution of ChR2 molecules across the three states in dark adapted condition. The refinement of the four-state model was also reported by Nikolic et al. [99], Stefanescu et al. [102], Williams et al. [100] using different mathematical expressions to improve the parameters to best fit the measured photocurrents. Ernst et al. [103] even proposed a hybrid model, combining the four-state model and the three-state model.

Recent studies [77, 81, 83] of the spectra properties of ChR2 by using short laser pulses to activate the channel, proposed an unidirectional photocycle as shown in Figure 1.8 (d). The early intermediate state P_1^{500} ($\lambda_{max} \approx 500$ nm) arises after the retinal isomerization in picosecond range. Subsequently, the RSBH$^+$ deprotonates resulting in the blue shifted P_2^{390} state ($\lambda_{max} \approx 390$ nm) with half life time $t_{1/2} \approx 10$ us. The Schiff base gets reprotonated with $t_{1/2} \approx 2$ ms and generates the late red shifted P_3^{520} state ($\lambda_{max} \approx 520$ nm). The decay of P_2^{390} and the subsequent formation of P_3^{520} are concurrent with the arise of the photocurrent [104]. This suggests that the proton release of ChR2 is temporally coupled to Schiff base reprotonation, rather than to Schiff base deprotonation as in BR.

Nack et al. [104] and Lórenz-Fonfría et al. [83] reported a two-phase recovery from P_3^{520} to the sensitized ground state based on time resolved FTIR studies. This was explained by branching ChR2 photocycle at the P_3^{520} intermediate state (Figure 1.8 (d)): ca. 75 % of ChR2 molecules directly relax from P_3^{520} to the initial ground state; the remaining 25 % molecules lead to the arise of the P_4^{480} state, then slowly relax to the ground state with $t_{1/2} \approx 20$ s, which kinetics is pH-sensitive.

Spectroscopy studies on D156A and C128T variants reported that the former mutant presents a long-lived P_2^{390} state, while the later mainly a long-lived P_3^{520} state; both mutants retarded the open state and displayed similar photocurrent intensity [105, 106]. This supports the viewpoint that the open state of ChR2 consists of P_2^{390} and P_3^{520} intermediates, or at least some substates of them. Possibly, we might assign them to O1 and O2 states of the four-state model. The spectrum of P_4^{480} state is almost identical to the ground state [81, 83], and it is assumed corresponding to the desensitized state of the three-state or of the four-state model [81]. Moreover, the C1 state is usually thought representing the dark adapted state.

The viewpoint that more intermediate states should exist in ChR2 photocycle is firstly derived from the study of Bamann et al. [81]. Heberle group [107] recently identified two P_1^{500} substates (P_{1a}^{500} and P_{1b}^{500}) through FTIR technique. It is possible that more substates would be identified in the future with the appearance of more sophisticated measurement techniques and with the increase of our knowledge to ChR2.

The 13-*cis*,15-*syn* Photocycle

The photocycle of ChR is triggered by retinal isomerization around the C13=C14 bond. Since the existence of the 13-*cis*,15-*syn* retinal in the dark state is still controversial as discussed above, the contribution of the photoreaction of this isomer to the ChR2 photocycle remains elusive. Whether the 13-*cis*,15-*syn* photocycle is functionally relevant becomes another open question. In BR, the IR difference spectrum displays a band pair 1186(-)/1179(+) cm^{-1} indicating 13*cis*,15-*syn* retinal photoreaction [108]. The negative band is from the C10-C11 stretching of 13-*cis*,15-*syn* retinal [109], while the positive band is a characteristic for the formation of all-*trans* retinal arising from 13-*cis* retinal [108]. In ChR2 WT, similar bands at 1183(-)/1177(+) cm^{-1} are observed [62, 89]; and in ChR2 E123T mutant, an intense positive band at 1175^{-1} appears [107]. This reveals that the 13-*cis*,15-*syn* retinal contributes to the photoreaction of ChR2. The step-scan FTIR study of Lórenz-Fonfría et al. [107] shows that the 13-cis,15-*syn* photocycle recovers in 1 ms, which is faster than most of the proton transfer reactions, indicating that the 13-*cis* photocycle bears minor functional relevance for ChR2.

1.3.3 Structural Information of ChR2 Photointermediates

The IR difference spectroscopy has become one popular technique for studying ChR2 structural changes during the photocycle, where the negative bands represent the dark state vibrations which are missing in the respective intermediate states [81, 83, 88, 105, 107]. The amide I vibrations (1700–1620 cm^{-1}) which are from the peptide backbone C=O stretching mode (νC=O), thus reflect the conformational information of the protein backbone. Besides, the amide II vibrations, which is a combination of the C-N stretching (νC-N) and the N-H in-plane bending mode (νN-H) of the peptide bond, can also be referred to in studying protein structural changes. However, in rhodopsins difference spectra, the protein amide II region is often obscured by the retinal ethylenic C=C vibration (νC=C) [110]. The glutamate acid, aspartic acid side chain νC=O appear in the range 1700–1800 cm^{-1}.

Preparation for Cation Permeation

The P_1^{500} state of ChR2 could be trapped at 80 K according to the study of other retinal proteins [62], due to the insufficient thermal energy for overcoming the transition barrier to the next state [88]. The FTIR difference spectrum at 80 K [62, 88], and the time-resolved IR difference spectrum of P_1^{500} [83] show a strong negative band at 1664 cm^{-1}, indicating large conformational changes of the protein backbone in the early photocycle (early microseconds). The MD simulations of C1C2 reported that the steric conflict between the 13-methyl group of 13-*cis*,15-*syn* retinal and W262 (W223 in ChR2) caused the movement of H6 and H7 [111]. Considering the fact that C1C2 and ChR2 share the same H6 and H7, we might assume that the increasing retinal–protein steric interaction after retinal isomerization is (part of) the driving force leading to large protein backbone conformational changes.

A differential band 1741(+)/1735(-) cm^{-1} observed at 80 K [88], was explained by the weakening of the hydrogen bond of D156 carboxylic group [89]. The retinal is in van der Waals contact with the residue C128 that is hydrogen bonded to D156 as discussed above. It is thus possible that the retinal isomerization contributes to the weakening of D156-C128 interaction. The broad positive band at 1760 cm^{-1} assigned to D156 side chain νC=O in P_1^{500} state measured through time-resolved IR, suggesting an even weaker hydrogen bond of D156 with C128 relative to the previous observation at 80 K [83].

Another differential band at 1728(+)/1717(-) persisting in P_1^{500}, P_2^{390} and P_3^{520} states disappeared in the E90A variant. Thus, this band is assigned to the νC=O of protonated E90 side chain [83].

Onset of Cation Permeation

It is proposed that the formation of a water-filled pore (formed by H1, H2, H3, H7 as discussed above) inside the protein is the prerequisite for cation permeation [83]. The appearance of the band pair at 1663(-)/1631(+) cm^{-1} from P_1^{500} state, is assigned to the elongation of β-hairpins (1640–1625 cm^{-1}) from extracellular side loops (1662 cm^{-1}) [83]. These conformational changes are part of the reasons promoting the hydration of helices. The rising band pairs in amide I and II regions, at 1663(-)/1650(+) cm^{-1} and 1540(-)/1561(+) cm^{-1}, respectively, indicate the hydration (Readers who are interested in the detailed interpretation are guided to ref. [83]). Both bands temporally correlated well with the photocurrent, all reaching the maximum intensity at ca. 1 ms [81, 83].

Consistently, Gerwert group [78] performed MD simulations of ChR2 and sampled a water chain filling the pore formed by H1, H2, H3, H7. In this study, the authors argued that before channel opening, the protonated E90 hydrogen bonds to N258 that connects H2 and H7, preventing water influx. The residue

E90 gets deprotonated in P_3^{390} state and subsequently flips downward to the extracellular side, forming a salt-bridge with K93. The breakage of E90–N258 hydrogen bonds facilitates the formation of the water chain [78, 112]. This is controversial with the time-resolved FTIR study of Heberle group [83] who reported that the deprotonation of E90 only happens during the lifetime of the P_4^{480} state.

Channel Closure

As discussed above, Heberle group [83] suggests that the channel closes through two pathways from P_3^{390} state: $P_3^{390} \rightarrow$ dark state, and $P_3^{390} \rightarrow P_4^{480} \rightarrow$ dark state. The P_3^{390} state differs from the P_4^{480} state in at least the protonation state of D156, thus the DC gate hydrogen bond strength, and the protonation state of E90. In the P_4^{480} state, the DC gate is supposed recovered to the dark state configuration; E90 is deprotonated, resulting in a stronger interaction with N258 and contributeS to the channel closure, which is conflict with the view of Gerwert group [78]. Besides, E90 might also contribute to the pH-sensitive $P_4^{480} \rightarrow$ dark state recovery kinetics.

Conformational Changes of the Retinal

The C-C and C=C stretching modes of the retinal are sensitive to the *trans* or *cis* conformation [88]. The rising of the positive band at 1191 cm^{-1} after illumination leads to the conclusion that the retinal photoisomerization is from all-*trans*,15-*anti* to 13-*cis*,15-*anti* conformation [88]. The positive band at 986 cm^{-1} of the P_1^{500} state IR difference spectrum is assigned to the hydrogen-out-of-plane (HOOP) vibration of the retinal, indicating a nonplanar retinal conformation or a more twisted retinal than in the dark state [83]. The stretching band of C14-C15 bond is observed in both P_1^{500} and P_2^{390} states at 1173 cm^{-1}, while it is shifted in the P_3^{520} state. This reflects similar retinal conformations in P_1^{500} and P_2^{390} states, which are different from that in P_3^{520} state [83]. The retinal ethylenic vibration bands are almost canceled in the P_4^{480} state IR difference spectrum, which implies that the retinal in this state has already recovered [83].

Chapter 2

Theory

2.1 Molecular Mechanics Methods

In the study of biological systems, experimental techniques, such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, provide an average, static view of the molecules. However, this is far from sufficient for understanding the biological activity, considering the fact that biomolecules are live entities which continuously interact among themselves and with the environment. Molecular dynamics (MD) simulation with molecular mechanics (MM) force fields is a supplementary to experiment for studying the equilibrium and dynamic properties of complex systems.

In molecular mechanics methods, the "building blocks" are entire atoms, i.e. electrons are not considered as individual particles. This means that bonding information must be provided explicitly, rather than being the result of solving the electronic Schrödinger equation. Moreover, the quantum effects of the nuclear motion are also neglected, which means that the nuclei behave like classical particles. Therefore, the dynamics of atoms can be described by Newton's second law of motion

$$\vec{F_i} = m_i \, \vec{a}_i \tag{2.1}$$

$\vec{F_i}$ is the force acting on atom i, m$_i$ and $\vec{a}_i$ are its mass and acceleration, respectively. The force can be calculated as (2.2)

$$\vec{F_i} = -\frac{\partial V(\vec{r}_1, \vec{r}_2, \cdots, \vec{r}_{N-1}, \vec{r}_N)}{\partial \vec{r}_i} \tag{2.2}$$

V is the total potential energy, which is a function of all nuclei coordinates. N is the total number of atoms, and $\vec{r}_i$ is the coordinate of atom i.

Deriving the total potential energy is one of the main tasks in molecular mechanics. In biomolecular force fields, the potential energy includes the bonded and non-bonded parts. The bonded potential is expressed as

$$V_{bond} = V_{str} + V_{bend} + V_{tors} \tag{2.3}$$

V_{str} is the potential for stretching a bond between two atoms. V_{bend} represents the potential of bending an angle. V_{tors} is the torsional potential for rotation around a bond. In force fields aimed at macromolecules, V_{str} and V_{bend} are described by a harmonic potential, while a Fourier series is applied to the torsions

$$\begin{aligned} V_{bond} = & \sum_{i,j} \frac{1}{2} \, k_b^{ij} \, (r_{ij} - r_{ij}^0)^2 \; + \sum_{i,j,k} \frac{1}{2} \, k_a^{ijk} \, (\theta_{\,ijk} - \theta^{\,0}_{\,ijk})^2 \\ & + \sum_{n_{i,j,k,l}} V_{n_{ijkl}} \; \cdot \; [1 + \cos(n_{ijkl} \phi_{\,ijkl} - \gamma_{\,n_{ijkl}})] \end{aligned} \tag{2.4}$$

where k_b^{ij} and k_a^{ijk} are force constants; r_{ij} is the distance between two covalently bound atoms i and j; $\theta_{\,ijk}$ is the angle defined by three atoms i, j and k; r_{ij}^0 and $\theta^{\,0}_{\,ijk}$ represent the equilibrium values; $\phi_{\,ijkl}$ is the torsion angle defined by atoms i, j, k, and l; $V_{n_{ijkl}}$ is the height of the torsional barrier; n_{ijkl} gives the periodicity and

$\gamma_{\ n_{ijkl}}$ is the phase offset.

The non-bonded potential describes the interaction between atoms which are not linked by covalent bonds. This interaction can be divided into Van der Waals interaction and electrostatic interaction. The Van der Waals interaction describes the attraction between atoms at long range , or the repulsion at short range. The attraction, also called dispersion, is due to the induced-dipole–induced-dipole interaction which is generated by the motion of electrons. This attraction varies as the inverse sixth power of the distance between two fragments (or atoms). It is impossible to derive theoretically the functional form of the repulsive interaction. Nevertheless, there are some characteristics of this interaction, i.e. it approaches towards zero as the distance between two fragments (R_{ij}) goes to infinity, and it approaches zero faster than the R_{ij}^{-6} term as the energy goes towards zero from below.

A popular potential which fulfills the general requirements of both attractive and repulsive interactions is the *Lennard-Jones* (LJ) potential, in which the repulsion is given by the R^{-12} dependence

$$V_{LJ} = \sum_{i}^{N-1} \sum_{j>i}^{N} \varepsilon \left[\left(\frac{R_{ij}^0}{R_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^0}{R_{ij}} \right)^{6} \right] \tag{2.5}$$

N is the total number of atoms, R_{ij}^0 the equilibrium distance at which the energy is minimal, ε the depth of the minimum.

The other term of the non-bonded potential describes the interaction between positively charged parts and negatively charged parts of the molecule. This is the electrostatic interaction and is calculated by the Coulomb potential

$$V_{Coul} = \sum_{i}^{N-1} \sum_{j>i}^{N} \frac{Q_i\, Q_j}{4\, \pi\, \epsilon\, R_{ij}} \tag{2.6}$$

Q is the point charge of the fragment, ϵ the dielectric constant of the surrounding medium, R_{ij} the distance between the charges.

Notably, many force fields in use differ from each other in three main aspects: (i) The functional form of each energy term. (ii) The number of cross terms included. Cross terms are applied to describe interactions that affect others. For instance, a strongly bent water molecule tends to stretch its O–H bonds, which is accounted for through a stretch–bend cross term. Other cross terms, such as bend–bend, stretch–stretch, stretch–torsion, bend–torsion etc., may be included. (iii) The information used for fitting parameters. For instance, in the force fields which are designed primarily to treat large systems (proteins or nucleic acids), the functional forms are chosen as simple as possible. Then, harmonic functions are used for V_{str} and V_{bend}, the Lennard-Jones potential for Van der Waals interaction and no cross terms are included. The Amber99SB force field for the glutamate receptor and the Charmm36 force field for ChRs in the present thesis belong to this type. On the other hand, some force fields are designed to deal with small to medium sized molecules at a high degree of accuracy. These force fields use at least cubic or quartic order expansions to describe V_{str} and V_{bend} and possibly an exponential type potential for repulsive Van der Waals interactions. Besides, a number of cross terms are included. The MM*X* type force fields are in this category.

With the forces acting on the atoms calculated, one can obtain the trajectory describing the positions and velocities of the molecule over a time period by numerically integrating Newton's equation. There are several algorithms to do this integration, including the Euler method, the Verlet method and some more advanced methods, the details of which can be found in ref. [113].

In the present study, the Verlet type algorithm [114] was applied. Its characteristic feature is that the positions and velocities are evaluated at interleaved time points. In more detail, assuming starting at time t_0, the velocities after the first half time step $\vec{v}(t_0 + \frac{1}{2}\,\Delta\, t)$ are established through the so-called "self-starting" scheme (such as Euler, Mid-point or Runge-Kutta-4) in the first iteration. Then, in the second iteration, the positions after the first time step are calculated.

$$\vec{r}\,(t_0 + \Delta\, t) = \vec{r}(t_0) + \vec{v}\,(t_0 + \frac{1}{2}\,\Delta\, t)\ \cdot\ \Delta\, t \tag{2.7}$$

Equation (2.7) can be verified by expanding in a Taylor series and truncating after the second order

$$\begin{aligned}\vec{r}\,(t_0+\Delta\,t) &= \vec{r}(t_0)+\frac{d\vec{r}(t_0)}{dt}\;\cdot\;\Delta\,t+\frac{1}{2}\frac{d^2\vec{r}(t_0)}{dt^2}\;\cdot\;\Delta\,t^2\\ &= \vec{r}(t_0)+\vec{v}\,(t_0)\;\cdot\;\Delta\,t+\frac{1}{2}\,\vec{a}(t_0)\;\cdot\;\Delta\,t^2\\ &= \vec{r}(t_0)+(\vec{v}\,(t_0)+\vec{a}(t_0)\;\cdot\;\frac{1}{2}\,\Delta\,t)\;\cdot\;\Delta\,t\\ &= \vec{r}(t_0)+\vec{v}\,(t_0+\frac{1}{2}\,\Delta\,t)\;\cdot\;\Delta\,t\end{aligned}$$

In the same way, $\vec{v}(t_0+\frac{3}{2}\,\Delta\,t)$, $\vec{r}(t_0+2\,\Delta\,t)$, $\vec{v}(t_0+\frac{5}{2}\,\Delta\,t)\cdots$ can be calculated.

The starting positions $\vec{r}(t_0)$ can be obtained from spectroscopic data, like X-ray or NMR. The starting velocities $\vec{v}(t_0)$ are normally assigned according to a Maxwell-Boltzmann distribution [115]. Notably, the length of the time step $\Delta\,t$ is a crucial parameter in the simulation. With a too small time step, the phase space is explored very slowly, and thus normally results in a trajectory that covers only a limited fraction of the phase space. On the other hand, with a too large time step, instabilities might arise in the integration because of high energy overlaps between atoms. There are no explicit rules for calculating the most appropriate time step for a molecular dynamics simulation. A rule of thumb recommends the time step to be approximately one-tenth the time of the shortest period of motion when simulating flexible molecules. Practically, this leads to a step size of 1 fs for biological molecules.

The conventional MD simulations sample the microcanonical ensemble, i.e. the NVE ensemble [113]. This ensemble represents an isolated system, which exchanges neither energy (in the form of heat or work) nor matter (particles) with the environment. In this case, the number of particles N, the volume V and the total energy of the system E are constant. The temperature fluctuates with time, which is not the experimental situation. In an usual experimental setup, the system is in thermodynamic equilibrium with the environment. Thus, it has the same temperature as the environment. The temperature is a crucial parameter determining whether a certain part of the phase space is reachable during the MD. Unlike the microcanonical ensemble, the canonical ensemble [113] allows for the exchange of energy, keeping the number of particles, volume and temperature constant (NVT ensemble) or keeping the pressure, temperature and number of particles constant (NPT ensemble).

Some simulation algorithms have been developed to control the temperature, which are called thermostat. The Berendsen [116], Andersen [113], and Nosé–Hoover [117, 118] thermostats are the three most widely used constant-temperature schemes. In the Berendsen thermostat, the system is coupled to an infinite thermal bath with the target temperature. By scaling of velocities of all the particles, the temperature of the system fluctuates around the target temperature. The Andersen thermostat works differently from the former one. In this method, some particles are randomly selected to undergo stochastic collisions with the imaginary particles of a heat bath. For these selected particles, their new velocities are assigned from a Maxwell-Boltzmann distribution corresponding to the desired temperature. All other particles remain unaffected by the collisions. The Nosé–Hoover thermostat is more complicated mathematically and conceptionally than the other two methods. The heat bath is regarded as a part of the system and is assigned an additional coordinate and an effective "mass". Because this approach achieves a realistic canonical ensemble, it has been commonly used as one of the most accurate methods for constant-temperature MD simulations.

Most experiments are performed at constant pressure rather than at constant volume. Similar to temperature, there are algorithms, called barostat, to maintain the pressure at a constant value. The barostats vary the volume of the simulation box in order to adjust the pressure. This is done either directly (Berendsen barostat [119]) or via an extra coordinate (Parrinello-Rahman barostat [120]).

2.2 Quantum Mechanics Methods

The combined MM/MD methods can provide good insights into the native activity of biomolecules. However, it can not describe the processes involving electron motions. In order to treat the electrons and the nuclei explicitly, one has to go to quantum mechanics [121–123].

Solving the Schrödinger equation (here focusing on the time-independent, i.e the stationary Schrödinger equation, eq. 2.8) is the most essential and also the most difficult task in quantum mechanics.

$$\hat{H}\ \Psi = E\Psi \tag{2.8}$$

The Hamiltonian $\hat{H}$ of the system can be written as the sum of kinetic and potential energy operators of the electrons and nuclei

$$\hat{H} = \hat{T}_e\ +\ \hat{T}_n\ +\ \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \tag{2.9}$$

$\hat{T}_e$ is the kinetic energy of the electrons, $\hat{T}_n$ the kinetic energy of the nuclei, $\hat{V}_{ne}$ the interaction between electrons and nuclei, $\hat{V}_{ee}$ the electron–electron interaction and $\hat{V}_{nn}$ the nuclear–nuclear interaction.

Because the electron is less heavier and moves much faster than the nuclear, the nuclei can be assumed to be static and the electrons then travel in the electronic field formed by the nuclei. This is the so-called Born-Oppenheimer approximation, which is applied to simplify $\hat{H}$ by separating the motion of the nuclei from the electrons. This leads to the Hamiltonian $\hat{H}$ with the following form

$$\begin{aligned}\hat{H}(\vec{R},\vec{r}) &= \hat{T}_e\ +\ \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}\\ &= -\frac{1}{2}\sum_{i}^{N_{elec}} \nabla_i^2\ -\ \sum_{a}^{N_{nuclei}}\sum_{i}^{N_{elec}} \frac{Z_a}{|\vec{R}_a\ - \vec{r}_i|}\\ &+ \sum_{i}^{N_{elec}}\sum_{j>i}^{N_{elec}} \frac{1}{|\vec{r}_i\ - \vec{r}_j|} + \sum_{a}^{N_{nuclei}}\sum_{b>a}^{N_{nuclei}} \frac{Z_a\, Z_b}{|\vec{R}_a\ - \vec{R}_b|}\end{aligned} \tag{2.10}$$

in which $\vec{r}_i$, $\vec{r}_j$ are the positions of electrons i, j, respectively; $\vec{R}_a$, $\vec{R}_b$ are the coordinates of nuclei a, b, respectively; Z_a, Z_b are the charges of nuclei a, b, respectively; N_{elec} is the total number of electrons and N_{nuclei} is the total number of nuclei. The kinetic energy of the nuclei ($\hat{T}_n$) is neglected, and the classical nuclear-nuclear repulsion ($\hat{V}_{nn}$) turns to be a constant. Thus, $\hat{H}$ is determined by the number of electrons in the system, the positions and charges of all nuclei.

The eigenfunctions Ψ of $\hat{H}$ are the wave functions characterizing the motion of the particles. The various properties of the molecule are derived from the wave functions. In many calculations, only the wave function with the lowest energy, i.e. the ground state Ψ_0, and a few low-lying excited states are considered. The variational principle (eq. 2.11) is regarded as a recipe for finding the best approximation of the wave function.

$$E_{trial} = \frac{\langle\ \Psi_{trial}|\hat{H}|\Psi_{trial}\ \rangle}{\langle\ \Psi_{trial}|\Psi_{trial}\ \rangle}\ \geq\ E_0 \tag{2.11}$$

It states that any guessed trial wave function (Ψ_{trial}) has an energy as the expectation value of $\hat{H}$, and this energy is higher than or equal to the true ground state energy E_0.

2.2.1 The Hartree-Fock Method

Electrons are indistinguishable, i.e. exchanging any pair of electrons, the distribution of the electron density remains the same. Keep in mind that the electron density is the square of the wavefunction. According to this, the wavefunction must remains unchanged when two electrons are exchanged. Otherwise, it must change the sign. As a matter of fact, for electrons the wavefunction is required to change sign, which is the so-called *antisymmetry principle*. In the Hartree-Fock scheme, the simplest approximation to the many-electron wavefuction is to take an antisymmetrized product of one-electron wavefunctions, which can be achieved through the *Slater determinant* Φ_{SD}

$$\Psi\ \approx\ \Phi_{SD} = \frac{1}{\sqrt{N!}}\{\chi_1\,\chi_2\ \cdots\ \chi_i\ \cdots\ \chi_N\} \tag{2.12}$$

N is the total number of electrons. χ_i are one-electron functions, also known as *spin orbitals*, which are given as the product of a spatial orbital and one of the two spin functions (α, β).

Since the functional form of the wavefunction has been chosen as eq. 2.12, the next step is to find the best

Slater determinant, i.e. the best set of $\{\chi_i\}$ to minimize the expectation value of the Hamiltonian under the constraint of orthonormality

$$\begin{aligned} < \chi_i|\chi_j > &= \delta_{ij} \\ \delta_{ij} &= 1, i = j \\ \delta_{ij} &= 0, i \neq j \end{aligned} \tag{2.13}$$

This can be interpreted as that the spatial overlap between the orbitals of two electrons occupying the same space is one, while the overlap between two electrons occupying different space is zero. Applying this constraint throughout the variation of spin orbitals introduces the *Lagrangian multipliers* ε_i, which gives rise to the Hartree-Fock equations

$$\hat{f}_i \chi_i = \varepsilon_i \; \chi_i, i \in \mathrm{N} \tag{2.14}$$

N is the total number of electrons and ε_i has the physical interpretation of orbital energy (Koopmann's theorem provides the interpretation of orbital energy, the details of which can be found in any quantum chemistry book). The Fock operator $\hat{f}_i$ is defined as

$$\hat{f}_i = \underbrace{-\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{|\vec{R}_A - \vec{r}_i|}}_{\hat{h}_i} + \hat{V}_i^{HF} \tag{2.15}$$

The first two terms form the one electron operator $\hat{h}_i$, which characterizes the kinetic energy of an electron and the attraction to all the nuclei. The third term is the Hartree-Fock potential, which is composed of two components

$$\hat{V}_i^{HF} = \sum_j^N (\hat{J}_j - \hat{K}_j) \tag{2.16}$$

where $\hat{J}_j$ is the Coulomb operator, describing the potential of an electron at position $\vec{r}_i$ due to the average charge distribution of other electrons in spin orbital χ_j.

$$\hat{J}_j = \int d\vec{r}_j \frac{|\chi_j(\vec{r}_j)|^2}{|\vec{r}_i - \vec{r}_j|} \tag{2.17}$$

In eq. 2.16, $\hat{K}_j$ is the exchange operator, entirely due to the antisysmmetry of the *Slater determinant*, and has no classical interpretation.

$$\hat{K}_j = \int d^3\vec{r}_j \frac{\chi_j(\vec{r}_i)\chi_j^*(\vec{r}_j)\chi_i(\vec{r}_j)}{|\vec{r}_i - \vec{r}_j|} \tag{2.18}$$

Since the $\hat{f}_i$ is an effective one-electron operator, $\sum^N \hat{f}_i$ will characterize a system of N electrons that do not interact with each other, but experience an average repulsion potential due to the remaining $N-1$ electrons. This system is the so-called non-interacting system, of which the electron correlation is neglected. The *Slater determinant* is the exact wavefunction of the non-interacting system.

$$\hat{H}_{HF}\Phi_{SD} \;=\; E_{HF}\Phi_{SD} \;\Leftrightarrow\; \sum_i^N \hat{f}_i \; \Phi_{SD} \;=\; \sum_i^N \varepsilon_i \; \Phi_{SD} \tag{2.19}$$

Note that $\hat{f}_i$ depends through the $\hat{V}_i^{HF}$ (eqs. 2.17, 2.18) on the spin orbitals χ_i, which are derived from their own effective potentials. Therefore, the Hartree-Fock function needs to be solved iteratively until the orbitals are self-consistent (SCF), i.e. the input and output orbitals differ by less than a predetermined threshold.

In practical calculations, we normally do not directly solve the Hartree-Fock equations. The most often used strategy is to write the spin orbital as a linear combination of atomic orbitals (LCAO)

$$\chi_i = \sum_\nu^{N_{basis}} c_{\nu i} \; \phi_\nu \tag{2.20}$$

in which the atomic orbitals ϕ_ν are commonly called basis functions. A set of basis functions used in the calculations is called a basis set. There are two popular types of basis sets, the atom centered Slater type orbitals (STOs): $\phi_\nu \propto exp\left(-\zeta\left|\vec{r}\right|\right)$, and the atom centered Gaussian type orbitals (GTOs): $\phi_\nu \propto exp\left(-\alpha\left|\vec{r}\right|^2\right)$. It is common to replace the STOs by GTOs, especially when the atomic orbitals are centered on different atoms which makes the two-, three- and four-center integrals very difficult to evaluate.

Applying the LCAO ansatz to the Hartree-Fock equation, one gets

$$\hat{f}_i \sum_{\nu}^{N_{basis}} c_{\nu i}\phi_\nu = \varepsilon_i \sum_{\nu}^{N_{basis}} c_{\nu i}\phi_\nu, i \in \mathrm{N} \tag{2.21}$$

Multiplying from the left by a specific basis function and integrating over the whole space yields the Roothaan-Hall equation.

$$\sum_{\nu}^{N_{basis}} c_{\nu i} \int \phi_\mu \, \hat{f}_i \, \phi_\nu d\tau = \varepsilon_i \sum_{\nu}^{N_{basis}} c_{\nu i} \int \phi_\mu \, \phi_\nu d\tau \tag{2.22}$$

which can be conveniently written in the matrix equation form

$$\mathbf{F}\,\mathbf{C} = \mathbf{S}\,\mathbf{C}\,\mathbf{E} \tag{2.23}$$

where the elements of matrix **C** are the coefficients $c_{\nu i}$; **E** is a diagonal matrix with the elements of the orbital energies ε_i and **S** is the overlap matrix between basis functions ϕ_μ and ϕ_ν: $S_{\mu\nu} = \langle\phi_\mu|\phi_\nu\rangle$. The elements of the Fock matrix **F** can be obtained as the sum of the one-electron integrals (eq. 2.15, the first two terms on the right side), and the density matrix **P** times the two-electron integrals. They are normally written in the following shorthand notation form:

$$\begin{aligned} F_{\mu\nu} &= \langle\, \mu|\hat{f}_i|\nu\,\rangle = \, h_{\mu\nu} + \sum_{\lambda\sigma}^{N_{basis}} P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)] \\ P_{\lambda\sigma} &= \sum_{i}^{occ} c_{\lambda i} c_{\sigma i} \end{aligned} \tag{2.24}$$

The procedure of solving the Roothaan-Hall equation starts from a guess of MO coefficients. Then follows the construction of the **F** matrix and the diagonaliztion of it. The newly obtained coefficients are then used to calculate a new **F** matrix. This procedure is continued in an iterative way, until the set of coefficients used for constructing the **F** matrix is equal to (within a certain threshold in the actual calculation) that resulting from the diagonalization of **F** . Finally the spin orbitals and the orbital energies ε_i are obtained, and then one can get the total energy through the following relation.

$$E_{total} = \sum_{i}^{N} \varepsilon_i \, - \sum_{i,j}^{occ}(2J_{ij} - K_{ij}) \tag{2.25}$$

In this iterative procedure the assembly of the two-electron integrals is the most time consuming part. Considering the fact that the two-electron integrals may involve up to four different basis function indices $(\mu, \nu, \lambda, \sigma)$, the effort dealing with this part grows with the fourth power of the number of basis functions N_{basis}^4.

2.2.2 Post Hartree-Fock Electron Correlation Methods

As discussed above, in the Hartree-Fock method, the electron-electron interaction is replaced by the average interaction. Thus, the instantaneous position of an electron is not affected by the presence of a neighbouring electron. However, in fact, the motions of electrons are correlated and they tend to ‘avoid’ each other more than as suggested by the Hartree-Fock theory, and this results in a lower energy. The inclusion of such electron correlation effects are necessary for the cases where electron correlation is crucial, e.g. in the treatment of excitation problems. Next, the main methods taking care of the electron correlation will be introduced.

Configuration Interaction

The configuration interaction (CI) approach involves the Hartree-Fock excited states in the description of electronic states, i.e. the total wavefunction is regarded as a linear combination of the ground and excited states wavefunctions

$$\Psi^{CI} = c^0\Psi^0_{HF} + \sum_S c^S_{HF}\Psi^S_{HF} + \sum_D c^D_{HF}\Psi^D_{HF} + \cdots \tag{2.26}$$

where the expansion coefficients (c^0, c^S, c^D etc.) are obtained by minimizing the energy of the system using the variational approach. Ψ^S, Ψ^D etc. (also referred to *Configurational Sate Functions* (CSF)) are the wavefunctions representing the singly, doubly etc. excited electronic configurations. They are obtained from the Hartree-Fock calculation and held fixed (i.e. $\Psi^S = \sum_\nu c^S_\nu\phi_\nu$, in which the coefficients of the basis functions ϕ^i_ν are not optimized again). The number of excited wavefunctions N_{excite} is factorially dependent on the number of electrons N and the number of basis functions M_{basis}: $N_{excite} = M_{basis}!/N!(M_{basis} - N)!$. Thus the *full CI* (including all possible excited configurations, thus it recovers 100 % of the electron correlation for a given basis set) is only applicable for very small systems. In practical calculations, N_{excite} in the CI expansion (eq. 2.26) is usually restricted to reach a computational feasible model size. Including only the single excitation gives the *CI with Singles* (CIS), however, which does not give any improvement over the HF ground state energy. This is due to the *Brillouin's theorem*, under which all the matrix elements between the HF reference wavefunction and the singly excited wavefunctions are zero. See below the typical CI matrix.

$$\begin{array}{c|ccccccc}
 & \Psi^0_{HF} & \Psi^S_{HF} & \Psi^D_{HF} & \Psi^T_{HF} & \Psi^Q_{HF} & \Psi^5_{HF} & \cdots \\
\hline
\Psi^0_{HF} & E^0_{HF} & 0 & \mathcal{X} & 0 & 0 & 0 & 0 \\
\Psi^S_{HF} & 0 & E^S_{HF} & \mathcal{X} & \mathcal{X} & 0 & 0 & 0 \\
\Psi^D_{HF} & \mathcal{X} & \mathcal{X} & E^D_{HF} & \mathcal{X} & \mathcal{X} & 0 & 0 \\
\Psi^T_{HF} & 0 & \mathcal{X} & \mathcal{X} & E^T_{HF} & \mathcal{X} & \mathcal{X} & 0 \\
\Psi^Q_{HF} & 0 & 0 & \mathcal{X} & \mathcal{X} & E^Q_{HF} & \mathcal{X} & \mathcal{X} \\
\Psi^5_{HF} & 0 & 0 & 0 & \mathcal{X} & \mathcal{X} & E^5_{HF} & \mathcal{X} \\
\vdots & 0 & 0 & 0 & 0 & \mathcal{X} & \mathcal{X} & \ddots
\end{array}$$

Since only doubly excited wavefunctions have non-zero matrix elements with the HF wavefunction, the *CI with doubles* (CID) is the lowest CI level that gives an improvement over the HF result. Including both singly and doubly excited states will result in the CISD method, in which the singly excited wavefunction indirectly correlates to the HF wavefunction through the doubly excited state. This method is generally applicable for plenty of systems. The higher CI levels may also include the triply or/and quadruply excited state, resulting in the CISDT or CISDTQ model. They are rarely considered because of the large number of CSFs.

As mentioned above, in conventional CI methods, only the expansion coefficients before the determinants (c^0, c^S, c^D etc. eq. 2.26) are optimized by variational principle, but the determinants themselves are not optimized. To get a lower energy (or better wavefunction), the coefficients before the basis functions should also be permitted to vary, which leads to the *Multi-Configuration Self-Consistent Field* (MCSCF) method.

$$\Psi^{MCSCF} = \sum_\nu c^0_{MCSCF}\Psi^0_{HF} + \sum_S c^S_{MCSCF}\sum_\nu a^S_\nu\phi_\nu + \sum_D c^D_{MCSCF}\sum_\nu a^D_\nu\phi_\nu + \cdots \tag{2.27}$$

The main problem of MCSCF theory is: which electronic configuration should be included for the property of interest? One of the most attractive technique is the *Complete Active Space Self-Consistent Field* (CASSCF) method [124]. In the CASSCF method, a large number of configurations can be included by partitioning the molecular orbitals (MOs) into *active* and *inactive* spaces. The MOs belonging to the inactive space are always doubly occupied or empty in all configurations. The rest MOs belonging to the active space, allows all possible excitation states (i.e. full CI, see Figure 2.1(a)). The *Restricted Active Space Self-Consistent Field* (RASSCF) method is a variation of CASSCF, in which the active space is further divided into three parts: RAS1, RAS2 and RAS3. As illustrated in Figure 2.1(b), a limited number of electrons are allowed to be excited from RAS1 to RAS3, while RAS2 has a full CI. The CASSCF (or RASSCF) method tends to give an unbalanced description of the problem, especially if only part of the valence electrons are included in the active space. Thus, the selection of the active space is crucial, which needs to be done manually and requires the user to have some insights into the problem. In order to get more accurate wavefunctions than with the MCSCF method, one could use the MCSCF wavefunction as the reference instead of an HF type wavefunction (single determinant), see eq. 2.27. This gives the *Multi-Reference Configuration Interaction* (MRCI) method.

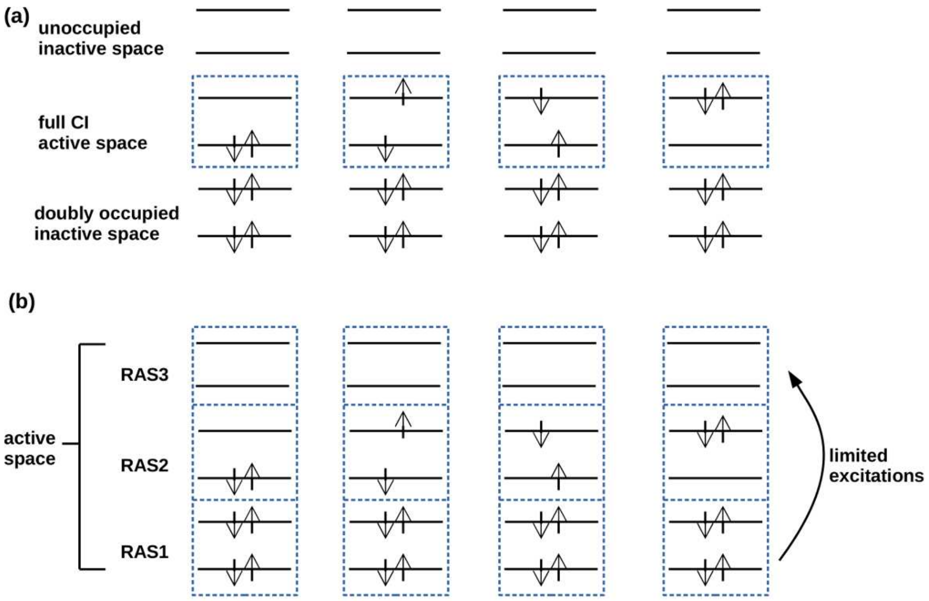


Figure 2.1: Illustration of CAS and RAS orbitals. (a) (2,2)CAS: the active space includes two electrons in two orbitals, while the inactive space includes two doubly occupied and two empty orbitals. (b) RAS: the active space of (6,6)CAS is further divided into RAS1 (two lower doubly occupied orbitals), RAS2 and RAS3 (two higher empty orbitals) The limited excitations are allowed from RAS1 to RAS3, while in RAS2 the full CI is allowed.

$$\Psi^{MRCI} = c^0_{MRCI}\Psi^0_{MCSCF} + \sum_S c^S_{MRCI}\Psi^S_{MCSCF} + \sum_D c^D_{MRCI}\Psi^D_{MCSCF} + \cdots \tag{2.28}$$

No truncated CI methods are size consistent, i.e. the energy of an N fragments system is not equal to N times the energy of one fragment. This deficiency makes the generally applicable CISD recover a smaller and smaller percentage of the correlation energy as the number of electrons increases. The lack of size deficiency can be corrected by the most widely known *Davidson correction*, which adds the approximated quadruples to the standard CISD, denoted as CISDQ.

$$\begin{aligned} E_{CISDTQ} \approx\ E_{CISDQ} =\ E_{CISD} + \Delta\ E_Q \\ \Delta\ E_Q = (1 - a_0^2)\Delta\ E_{CISD} \end{aligned} \tag{2.29}$$

Many-Body Perturbation Theory

The idea of perturbation theory is to utilize an already solved problem to deal with the target complicated one. The premise is that the solved problem should have close relation to the target problem. As a result, the Hamilton operator of the target system ($\hat{H}$) is expressed as two components, a reference $\hat{H}_0$ and a perturbation $\hat{H}^{'}$ (the difference between $\hat{H}$ and $\hat{H}_0$).

$$\hat{H} = \hat{H}_0 + \lambda\hat{H}^{'} \tag{2.30}$$

where the parameter λ determines the strength of the perturbation. It varies continuously from 0 to 1, leading to the continuous change of the eigenfunctions Ψ and eigenvalues E of $\hat{H}$. Ψ and E can be written as a Taylor series

$$\begin{aligned} \Psi = \Psi^0 + \lambda\Psi^1 + \lambda^2\Psi^2 + \cdots \\ E = E^0 + \lambda\, E^1 + \lambda^2\, E^2 + \cdots \end{aligned} \tag{2.31}$$

in which Ψ^0 and E^0, corresponding to $\lambda = 0$, are the *zeroth-order* or *unperturbed* wavefunction and energy, respectively. E^1, E^2, etc. are the first order, second order correction etc. to the energy.

In the most popular *Møller-Plesset* (MP) perturbation theory, the unperturbed Hamiltonian $\hat{H}_0$ is treated as the sum of Fock operators (eq. 2.15):

$$\hat{H}_0 = \sum_{i=1}^{N_{elec}} \hat{f}_i = \sum_{i=1}^{N_{elec}} \hat{h}_i + \sum_{j=1}^{N_{elec}} (\hat{J}_i + \hat{K}_i) \tag{2.32}$$

Then the eigenvalue of $\hat{H}_0$, i.e. the *zeroth-order* energy E^0 is equivalent to the sum of Fock operator eigenvalues ε_i (which are the energies of occupied orbitals).

$$E^0 = \sum_{i=1}^{occ} \varepsilon_i \tag{2.33}$$

In order to calculate the higher-order wavefunctions and energies, one needs to know the exact form of the perturbation operator that is defined as: $\hat{H}^{'} = \hat{H} - \hat{H}_0$. Keeping this in mind, the expression of $\hat{H}$ can be written as the sum of nuclear-electron attraction terms and electron-electron repulsion terms.

$$\hat{H} = \sum_{i=1}^{N_{elec}} \hat{h}_i + \sum_{i=1}^{N_{elec}} \sum_{j>i}^{N_{elec}} \frac{1}{|\vec{r_i} - \vec{r_j}|} \tag{2.34}$$

Subtracting eq. 2.32 from eq. 2.34, one obtains

$$\hat{H}^{'} = \sum_{i=1}^{N_{elec}} \sum_{j>i}^{N_{elec}} \frac{1}{|\vec{r_i} - \vec{r_j}|} - \sum_{j=1}^{N_{elec}} (\hat{J}_i + \hat{K}_i) \tag{2.35}$$

Then the first-order energy correction is calculated as the average of $\hat{H}^{'}$ over the zeroth-order Hartree-Fock wavefunction.

$$\begin{aligned} E^1 &= \langle \Psi^0 | \hat{H}^{'} | \Psi^0 \rangle \\ &= -\frac{1}{2} \sum_{i,j}^{N_{elec}} \frac{1}{|\vec{r_i} - \vec{r_j}|} [(ii|jj) - (ij|ij)] \end{aligned} \tag{2.36}$$

Here, ii, jj, ij in $(|)$ are the shorthand notations of the products of two molecular orbitals with indices i and j.

Combining eq. 2.32 and eq. 2.35, we can find that the first-order corrected energy is just the Hartree-Fock energy. Thus, at least the second-order correction is necessary to recover the electron correlation energy, which is the widely used MP2 theory. The second-order energy correction can be generated by replacing two occupied orbitals i and j with two virtual orbitals a and b (this can also be stated as "The second-order energy correction can be generated by promoting two electrons from occupied orbitals i and j to two virtual orbitals a and b")

$$E^2 = \sum_{j>i}^{occ} \sum_{b>a}^{vir} \frac{\langle \Psi^0 | \hat{H}^{'} | \Psi_{ij}^{ab} \rangle \langle \Psi_{ij}^{ab} | \hat{H}^{'} | \Psi^0 \rangle}{E^0 - E_{ij}^{ab}} \tag{2.37}$$

Addition of MP2 type corrections to the CASSCF zeroth-order of wavefunction results in the CASMP2 or CASPT2 method [125]. It is commonly used for excitation calculations. Inclusion of higher-order energy corrections gives the MP3, MP4 etc. levels of theories. Although they can cover more electron correlation energy then MP2, the much higher computational cost makes them less popular than MP2.

Spectroscopy-Oriented Configuration Interaction

The SORCI method [126] is an MRCI variant that is designed to calculate the energy differences between electronic states (i.e. the vertical excitation energies), which then predict the optical spectra. There are several steps required to perform the SORCI calculations. In this procedure several approximations are involved, which results in energy differences with acceptable accuracy. However, these approximations are also the reasons leading to the main drawbacks of this method.

In the first step, the starting one-electron orbitals are obtained. These orbitals can be any type of Hartree-Fock, MCSCF or Kohn-Sham. The Kohn-Sham type orbitals are preferred for systems containing transition state elements, while the Hartree-Fock type is commonly used for the other systems. In the later case it is advantageous to get the improved virtual orbitals within the $N_{elec} - 1$ electrons field. In practical calculations, the orbitals need to be selected manually and the very high and low lying orbitals are usually discarded.

The initial reference space defined by the user can be a CAS, RAS or an arbitrary list of configurations. Nevertheless, the most important configurations for all states (or roots) of interest should be contained in the active space. After the diagonalization of the Hamiltonian matrix based on the initial reference space and spin symmetry adapted CSFs, only those configurations that have a higher coefficient than the predetermined threshold T_{pre} are selected. These configurations construct the final reference space S, the re-diagonalization of which generates the 0th order wavefunction for each root of interest i.

$$\Psi_i^0 = \sum_{\mu \in S} c_{\mu i}^0 \Phi_\mu \tag{2.38}$$

The next step is an MP2 calculation for each state based on a perturbing space R that is formed by configurations from the first-order interacting space. SORCI carries the hypothesis that the second-order correction is only accurate for the weakly-interacting perturbers. Thus, the configurations $\{\Phi_\mu \in R\}$ are divided into $R^{'}$ (strongly-interacting subspace) and $R^{''}$ (weakly-interacting subspace) by comparing the second-order correction energy (eq. 2.37) with a threshold T_{sel} as follows

$$\begin{aligned} E_i^2 &\geq T_{sel}, \Phi_\mu \in R^{'} \\ E_i^2 &< T_{sel}, \Phi_\mu \in R^{''} \end{aligned} \tag{2.39}$$

In this procedure, in order to simplify the CI calculation, the perturbing configurations $\{\Phi_\mu \in R\}$ are reduced to the “relevant” configurations. This is achieved through the *difference dedicated CI* (DDCI) approach, i.e. omitting Φ_μ, whose contributions approximately cancel upon taking energy differences. Consequently, the most numerous class of configurations featuring double excitations from the internal space (doubly occupied in all references) to the external space (empty in all references) are neglected.

Subsequently, in the variation step, the Hamiltonian matrix is diagonalized in the spaces S and $R^{'}$ to generate the energy E_i and wavefunction Ψ_i^a for each state of interest. The energy E_i is further corrected with the Davidson type correction and the MP2 correction provided by configurations $\{\Phi_\mu \in R^{''}\}$, which finally gives

$$E_i = E_i^a + E_i^{Davi} + E_i^{R^{''}} \tag{2.40}$$

The contribution from configurations $\{\Phi_\mu \in R^{''}\}$, $E_i^{R^{''}}$, can be expressed in two ways. One is the normal way by promoting double excitations in $R^{''}$ space, see eq. 2.37. If the relative coefficients of the configurations $\{\Phi_\mu \in R^{''}\}$ are strongly influenced by $R^{'}$, a more accurate variant is accessible.

$$E_i^{R^{''}} = E_i^2(R) - E_i^2(R^{'}) \tag{2.41}$$

Next, the diagonalization of the averaged reduced density matrix for each state of interest yields the approximated average natural orbitals (AANOs) and their occupation numbers. The AANOs can be used as the new set of one-electron basis, which is then reduced by discarding the orbitals with occupation numbers smaller than a threshold T_{nat}. Furthermore, if the absolute value of 2 minus the occupation number of one AANO is less than T_{nat}, then this AANO is frozen in the subsequent calculation. Repeating the whole procedure once again in the basis of the AANOs, one obtains the final energies and wavefunctions.

Semi-Empirical Methods

The most time consuming part of a Hartree-Fock calculation is the computation of integrals for construction of the Fock matrix (eq. 2.24). The semi-empirical configuration interaction methods are aimed at reducing the computational cost by neglecting or approximating some of these integrals. There are multiple initial approximations to achieve this purpose: (i) only the valence electrons are considered; (ii) the core electrons are subsumed into the nuclei by reducing the nuclear charge or introducing functions to model the combined repulsion due to the nuclei and core electrons; (iii) only a minimum basis set comprised of Slater type exponential orbitals is applied for the valence electrons; (iv) the overlap matrix S in the Roothaan-Hall equation (eq. 2.23) is set equal to the identity matrix.

The *zero-differential overlap* (ZDO) approximation is the central assumption of many semi-empirical methods. In this approximation, the overlap between pairs of different atomic orbitals is set to zero for all volume elements.

$$\int \phi_\mu \phi_\nu d\nu = 0 \tag{2.42}$$

If the two atomic orbitals ϕ_μ and ϕ_ν are located on the same atom, then the differential overlap is called monoatomic differential overlap. If ϕ_μ and ϕ_ν are located on different atoms, i.e. $\mu \neq \nu$, then it is the diatomic differential overlap. Under the ZDO approximation, all three- and four-center two-electron integrals are set to zero. The remaining integrals are made into parameters and their values are assigned based on *ab initio* calculations, or by fitting to experimental data. The number of neglected integrals and the way of parameterization result in various semi-empirical methods, such as complete neglect of differential overlap (CNDO), intermediate neglect of differential overlap (INDO), neglect of diatomic differential overlap (NDDO), and some derivatives based on these three models.

The NDDO model is the most accurate of the above three mentioned traditional integral approximations. It only neglects the differential overlap between atomic orbitals on different atoms. Therefore, all the two-electron, two-center integrals $(\mu\nu|\lambda\sigma)$ with μ and ν on the same atom A, λ and σ on the same atom B are included. Then the following Fock matrix elements are obtained.

$$\begin{aligned}
F_{\mu\mu} &= H_{\mu\mu}^{core} + \sum_{\lambda\in A}\sum_{\sigma\in A}[P_{\lambda\sigma}(\mu\mu|\lambda\sigma) - \frac{1}{2}P_{\lambda\sigma}(\mu\lambda|\mu\sigma)] + \sum_{B\neq A}\sum_{\lambda\in B}\sum_{\sigma\in B} P_{\lambda\sigma}(\mu\mu|\lambda\sigma) \\
F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_{\lambda\in A}\sum_{\sigma\in A}[P_{\lambda\sigma}(\mu\nu|\lambda\sigma) - \frac{1}{2}P_{\lambda\sigma}(\mu\lambda|\nu\sigma)] \\
&\quad + \sum_{B\neq A}\sum_{\lambda\in B}\sum_{\sigma\in B} P_{\lambda\sigma}(\mu\nu|\lambda\sigma), \quad (\mu,\ \nu \in \mathrm{A}) \\
F_{\mu\nu} &= H_{\mu\nu}^{core} - \frac{1}{2}\sum_{\lambda\in B}\sum_{\sigma\in A} P_{\lambda\sigma}(\mu\sigma|\nu\lambda), \quad (\mu \in \mathrm{A}, \nu \in \mathrm{B})
\end{aligned} \tag{2.43}$$

where $F_{\mu\mu}$ are the diagonal elements, and $F_{\mu\nu}$ are the off-diagonal elements.

The introduction of the ZDO approximation results in the neglect of the orthogonalization effects in the established semi-empirical methods. Applying the orthogonalization correction to NDDO gives the OMx methods, among which the OM2 (orthogonalization correction model 2) method in combination with the graphic unitary group approach (GUGA, which is a specific MRCI method) is used to calculate the vertical excitation energies in the present study.

2.2.3 Density Functional Theory

The wavefunction is the central object of quantum chemistry. The N-electron wavefunction depends on $4N$ variables ($3N$ spatial plus N spin variables for each of the N electrons), which leads to an unmanageable size when dealing with systems containing many atoms. Great efforts have been made to find a simpler quantity to replace the wavefunction as the central variable. The electron density $\rho(\vec{r})$ is the popular target which depends only on three spatial variables. The first attempts to use $\rho(\vec{r})$ to describe the molecular system date back to the early work of Thomas and Fermi (Details of this subject can be found in the rich review literatures, such as [121–123].). However at that time, people still did not know whether computing the energy as a functional of $\rho(\vec{r})$ is physically justified. Decades later, the Hohenberg-Kohn theorem proved that it is indeed physically justified, i.e. $\rho(\vec{r})$ does uniquely determine the Hamilton operator, and therefore all properties of the system. Additionally, the variational principle holds for the density functional [122]. The DFT energy functional might be expressed as follows:

$$\begin{aligned} E_{total} &= E[\rho] + E_{nn} \\ &= T[\rho] + E_{ne}[\rho] + J[\rho] + E_{ncl}[\rho] + E_{nn} \end{aligned} \tag{2.44}$$

The total energy of the system is the sum of DFT energy $E[\rho]$ and the energy E_{nn} due to nuclei-nuclei interaction. Because E_{nn} only depends on the nuclear positions and charges, the rest of this section will focus on the complicated part $E[\rho]$. As shown in eq. 2.44, the DFT energy includes the electrons' kinetic energy T, the external potential energy due to the electron-nuclei attraction E_{ne}, the classical coulomb repulsion J and the non-classical contribution E_{ncl} (self-interaction correction, exchange and Coulomb correlation). Among these terms only the functional form of J is known: $\frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2$. In order to construct the functionals for the other terms, Kohn and Sham introduced an auxiliary *non-interacting system* with the electron density ρ_{aux} equal to that of the real interacting system (ρ_{real}).

$$\rho_{aux} = \sum_i |\varphi_i|^2 \; = \; \rho_{real} \tag{2.45}$$

The spin orbitals φ_i (also called Kohn-Sham orbitals, KS orbitals) of this auxiliary system are determined by the Kohn-Sham equation

$$\hat{f}^{KS}\varphi_i = \varepsilon_i\varphi_i \tag{2.46}$$

Here $\hat{f}^{KS}$ is the one-electron Kohn-Sham operator defined as

$$\hat{f}^{KS} = -\frac{1}{2}\nabla_i^2 + \hat{V}^{KS}(\vec{r}_i) \tag{2.47}$$

$\hat{V}^{KS}(\vec{r}_i)$ is the Kohn-Sham potential. It is essential to construct an effective $\hat{V}^{KS}(\vec{r}_i)$ such that the condition of eq. 2.45 is fulfilled. With this in mind, the energy of the real interacting system is written as follows

$$\begin{aligned} E[\rho] &= T_{aux}[\rho] + E_{ne}[\rho] + J[\rho] + E_{XC}[\rho] \\ &= T[\rho] + \int V_{ne}\rho(\vec{r})d\vec{r} + \frac{1}{2}\int\int \frac{\rho(\vec{r}_i)\rho(\vec{r}_j)}{|\vec{r}_i - \vec{r}_j|}d\vec{r}_1 d\vec{r}_2 \; + \; E_{XC}[\rho] \\ &= -\frac{1}{2}\sum_i \langle\varphi_i|\nabla^2|\varphi_i\rangle - \sum_i \int \sum_A \frac{Z_A}{|\vec{r}_i - \vec{R}_A|}|\varphi_i|^2 d\vec{r}_i \\ &+ \frac{1}{2}\sum_{i<j}\int\int |\varphi_i|^2 \frac{1}{|\vec{r}_i - \vec{r}_j|}|\varphi_j|^2 d\vec{r}_i \, d\vec{r}_j + E_{XC}[\rho] \end{aligned} \tag{2.48}$$

T_{aux} is the exact kinetic energy of the auxiliary *non-interacting system.* E_{ne} is due to the electron-nuclei interactions. J is the classical electron-electron Coulomb repulsion. The remainder E_{XC} is the so-called *exchange-correlation energy*, for which no explicit functional form is known. It is defined as

$$E_{XC}[\rho] = (T_{real}[\rho] - T_{aux}[\rho]) + E_{ncl}[\rho] \tag{2.49}$$

Namely it includes the non-classical electrostatic contributions E_{ncl}, and the remaining kinetic energy which is not covered by T_{aux}.

After the application of the variational principle to eq. 2.48 with respect to independent variations in the orbitals under the constraint of eq. 2.13 (For the details of the derivation see ref. [121].), one gets the following equation

$$\left(-\frac{1}{2}\nabla^2 + \Big[\int \frac{\rho(\vec{r}_j)}{|\vec{r}_i - \vec{r}_j|}d\vec{r}_j - \sum_A \frac{Z_A}{\vec{r}_i - \vec{R}_A} + V_{XC}(\vec{r}_i) \Big] \right)\varphi_i = \varepsilon_i\varphi_i \tag{2.50}$$

Comparing eq. 2.50 and eq. 2.46, one finds that the effective $\hat{V}^{KS}(\vec{r}_i)$ needed for getting the correct orbitals of the auxiliary *non-interacting system* to fulfill eq. 2.26 is

$$\begin{aligned} V^{KS}(\vec{r}_i) &= \int \frac{\rho(\vec{r}_j)}{|\vec{r}_i - \vec{r}_j|} d\vec{r}_j - \sum_A \frac{Z_A}{|\vec{r}_i - \vec{R}_A|} + V_{XC}(\vec{r}_i) \\ &= V_{coulomb} + V_{ne} + V_{XC} \end{aligned} \tag{2.51}$$

Assuming an explicit form of V_{XC} is known, by solving the one-electron equations (eq. 2.46) one obtains the KS orbitals φ_i describing the auxiliary *non-interacting system.* The density of the real interacting system is then given by φ_i through eq. 2.45. Inserting this density into the energy expression (eq. 2.48) yields the ground state energy.

In real applications, the users have to introduce approximation of the unknown exchange-correlation functionals $E_{XC}[\rho]$ [127]. The simple local density approximation (LDA) is only dependent on the electron density. This approximation is mainly employed by solid-state physicists, because it works well for uniform electron density distributions, such as found in metals. However, its performance in molecules is usually poor. The generalized gradient approximation (GGA) is an extension to LDA, which depends on the electron density and its gradient $\nabla\rho(\vec{r})$ at a particular point $\vec{r}$ to account for the non-homogeneity of the true electron density. For instance, PW91, PBE, BLYP, BPW91 are several popular GGA functionals. The fact that the exchange contributions are significantly larger than the corresponding correlation effects motivates us to include the exact Hartree-Fock exchange energy. This leads to so called hybrid functionals, e.g. the very popular B3LYP functional, which contains 25% of HF exchange. Current DFT still has many limitations, which can be traced to the delocalization error and the static correlation error of the approximations commonly used for $E_{XC}[\rho]$. These two types of errors lead to failures for many predicted properties, such as an underestimation of the chemical reaction barriers, the materials band gaps, or they make DFT fail to describe degenerate or near-degenerate states [128].

2.2.4 Density-Functional Tight-Binding

The tight-binding theory originates from solid-state physics, describing the properties of electrons which are tightly bound to the atoms to which they belong. The DFTB method [129–132] is an approximation derived from density functional theory. In this method, the electron density is regarded as the sum of a reference density ρ^0 and a density fluctuation $\Delta\rho$ (eq. 2.52). ρ^0 is calculated from a superposition of neutral atomic densities ρ_a^0. $\Delta\rho$ is the deviation of ρ^0 with respect to the DFT ground state density.

$$\rho(\vec{r}) \equiv \rho = \rho^0 + \Delta\rho = \sum_a \rho_a^0 + \Delta\rho \tag{2.52}$$

Then the exchange-correlation contribution $E_{XC}[\rho^0 + \Delta\rho^0]$ can be expanded in a Taylor series as

$$\begin{aligned} E_{XC}[\rho^0 + \Delta\rho] = E_{XC}[\rho^0] &+ \int \left[\frac{\delta E_{XC}[\rho]}{\delta\rho}\right]_{\rho^0} \Delta\rho \, d\vec{r} \\ &+ \frac{1}{2} \int\int \left[\frac{\delta^2 E_{XC}[\rho]}{\delta\rho\delta\rho'}\right]_{\rho^0,\rho^{0\prime}} \Delta\rho\Delta\rho' \, d\vec{r}d\vec{r}' \\ &+ \frac{1}{6} \int\int\int \left[\frac{\delta^3 \, E_{XC}[\rho]}{\delta\rho\delta\rho'\delta\rho''}\right]_{\rho^0,\rho^{0\prime},\rho^{0\prime\prime}} \Delta\rho\Delta\rho'\Delta\rho'' \, d\vec{r}d\vec{r}'d\vec{r}'' \; + \; \cdots \end{aligned} \tag{2.53}$$

Inserting eq. 2.52 and eq. 2.53 into eq. 2.48, and with the addition of the nucleus-nucleus repulsion energy E_{nn}, the total energy is written as

$$\begin{aligned} E^{DFTB}[\rho^0+\Delta\rho] = -\frac{1}{2}\int\int \frac{\rho^0\rho^{0'}}{|\vec{r}-\vec{r^6}|} - \int V_{XC}[\rho^0]d\vec{r} + E_{XC}[\rho^0] + E_{nn} \\ + \sum_{i}^{occ} n_i \left\langle \varphi_i \middle| -\frac{\nabla^2}{2} + V_{ne} + \int \frac{\rho^{0'}}{|\vec{r}-\vec{r'}|}d\vec{r'} + V_{XC}[\rho^0] \middle| \varphi_i \right\rangle \\ + \frac{1}{2}\int\int \left(\frac{1}{|\vec{r}-\vec{r'}|} + \left.\frac{\delta^2 E_{XC}[\rho]}{\delta\rho\delta\rho'}\right|_{\rho^0,\rho^{0'}} \right) \Delta\rho\Delta\rho' d\vec{r}d\vec{r'} \\ + \frac{1}{6}\int\int\int \left.\frac{\delta^3 E_{XC}[\rho]}{\delta\rho\delta\rho'\delta\rho''}\right|_{\rho^0,\rho^{0'},\rho^{0''}} \Delta\rho\Delta\rho'\Delta\rho'' d\vec{r}d\vec{r'}d\vec{r''} + \cdots \end{aligned} \tag{2.54}$$

in which n_i is the occupation number of the ith spin orbital. The standard DFTB (non-self-consistent) neglects the second and higher order terms of the Talyor expansion. If the second order term is included, the SCC-DFTB (*Self-consistent charge density functional tight-binding*, also called DFTB2) method is obtained, while including the third order term results in the DFTB3 method.

The first line of eq. 2.54 is the repulsive energy contribution (E_{rep}), including the double-counting terms, the exchange-correlation contributions and the nucleus-nucleus repulsion. In DFTB this term is usually approximated as a sum of pairwise potentials.

$$E^0 = E^{rep} = \frac{1}{2}\sum_{ab,a\neq b} V_{ab}^{rep}[\rho_a^0,\rho_b^0,r_{ab}] \tag{2.55}$$

The second line of eq. 2.54 is the energy contribution from an atomic orbital Hamiltonian that depends only on the reference density. So this term can also be written as

$$E^1 = E^{H^0} = \sum_i^{occ}\langle\varphi_i|\hat{H}^0|\varphi_i\rangle = \sum_i^{occ}\sum_{\mu\nu} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0 \tag{2.56}$$

In order to calculate the reference density, and the elements of the Hamiltonian and the overlap matrix, the contracted Slater-type basis function is used for constructing the Kohn-Sham orbitals.

$$\varphi_i = \sum_{\zeta,n,l,m} a_{n\zeta} \cdot r^{l+n} \cdot e^{-\zeta\, r} \cdot Y_{lm} \tag{2.57}$$

Y_{lm} are the spherical harmonics. ζ determines the number of primitive orbitals, which is 3 for hydrogen, 4 for carbon, nitrogen and oxygen, and 5 for halogens. A minimal basis set only including the valence orbitals is used. Then the eigenvalues of the Kohn-Sham orbitals serve as the diagonal elements of the Hamiltonian matrix. The off-diagonal elements are subjected to a two-center approximation. Because all matrix elements depend merely on interatomic distances, they only need to be pre-calculated and stored for each orbital type pair. This fact together with the using of minimal valence basis set make DFTB much more computationally efficient than full DFT.

Standard DFTB does not deal well with charge rearrangement. DFTB2 made some improvements at this point by including the second order term (the third line of eq. 2.54), which is further approximated in terms of the density fluctuations $\delta\rho$. In more details, $\delta\rho$ is written as a superposition of atomic contributions

$$\delta\rho = \sum_a \delta\rho_a \tag{2.58}$$

in which the atomic-like density fluctuations $\delta\rho_a$ are expanded as a Slater-type exponential charge density, centered at the nuclei at position $\vec{R}_a$ (i.e. only the monopole term is kept).

$$\delta\rho_a \approx \Delta\, q_a \frac{\tau_a^3}{8\pi} e^{-\tau_a|\vec{r}-\vec{R}_a|} \tag{2.59}$$

$\Delta\, q_a$ is the net charge of atom a, defined as the Mulliken charge. The interaction between the fluctuating charge densities is described by a function γ_{ab}, which have two main properties. For large interatomic distance $\vec{r}_{ab}$, it describes a pure Coulomb interaction of the two partial charges $\Delta\, q_a$ and $\Delta\, q_b$. For vanishing interatomic distance, i.e. $a = b$, γ_{aa} describes the on-site electron-electron self-repulsion, which is characterized by the Hubbard parameter U_a. U_a is twice the chemical hardness of an atom, and has an relationship with the parameter τ_a that relates to the covalent radius: $\tau_a = \frac{16}{5} U_a$. The second-order energy can be written as

$$E^2 = \frac{1}{2} \sum_{ab} \Delta\, q_a \Delta\, q_b \gamma_{ab} \tag{2.60}$$

There is a limitation of the second order approxiration: the chemical hardness of an atom (U_a) is constant, i.e. the size of the atom is the same for all charge states. In fact, the more charge is accumulated around an atom, the larger is the effective size and vice versa. The deviation is most commonly seen between hydrogen and the first-row elements. In DFTB3 a default modified γ^h_{ab} function is proposed to solve the above problem

$$\gamma^h_{ab} = \frac{1}{|\vec{R}_a - \vec{R}_b|} - S(|\vec{R}_a - \vec{R}_b|, U_a, U_b) \cdot\ h(|\vec{R}_a - \vec{R}_b|, U_a, U_b) \tag{2.61}$$

The factor h equals 1, if neither atom a or b is hydrogen. If a or/and b is hydrogen, h adopts the following expression

$$h\ = exp\left[-\left(\frac{U_a + U_b}{2}\right)^{\zeta} (|\vec{R}_a - \vec{R}_b|)^2 \right] \tag{2.62}$$

The same approximation as for E^2 is applied for E^3, which can be expressed as the variation of the γ function

$$E^3 = \frac{1}{3} \sum_{ab} (\Delta\, q_a)^2 \Delta\, q_b \Gamma_{ab} \tag{2.63}$$

Here the matrix Γ is the derivative of the γ function with respect to charge (eq. 2.64), yielding the desired chemical behavior for charged systems through the Hubbard derivative $U^d_a = \partial U_a / \partial q_a$.

$$\begin{aligned}
\Gamma_{ab} &= \left.\frac{\partial \gamma_{ab}}{\partial q_a}\right|_{q^0_a} = \left.\frac{\partial \gamma_{ab} \partial U_a}{\partial U_a \partial q_a}\right|_{q^0_a} \quad (a \neq\ b) \\
\Gamma_{ba} &= \left.\frac{\partial \gamma_{ab}}{\partial q_b}\right|_{q^0_b} = \left.\frac{\partial \gamma_{ab} \partial U_b}{\partial U_b \partial q_b}\right|_{q^0_b} \quad (a \neq\ b) \\
\Gamma_{aa} &= \left.\frac{\partial \gamma_{ab}}{\partial q_a}\right|_{q^0_a} = \left.\frac{1}{2}\frac{\partial \gamma_{aa} \partial U_a}{\partial U_a \partial q_a}\right|_{q^0_a} \quad (a = b)
\end{aligned} \tag{2.64}$$

With all these approximations, the total energy of DFTB3 is results as

$$E^{DFTB3} = \sum_i^{occ} \sum_{\mu\nu} n_i c_{\mu\ i} c_{\nu\ i} H^0_{\mu\nu} + \frac{1}{2} \sum_{ab} \Delta\, q_a \Delta\, q_b \gamma^h_{ab} + \frac{1}{3} \sum_{ab} (\Delta\, q_a)^2 \Delta\, q_b \Gamma_{ab} + \frac{1}{2} \sum_{ab, a\neq\ b} V^{rep}_{ab} \tag{2.65}$$

Variation of this expression with respect to the spin orbital coefficients gives the DFTB Kohn-Sham equations

$$\sum_{\nu} c_{\nu i} (H_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0 \quad (\nu \in b, \forall\, a :\ \mu \in a)$$

$$\mathbf{HC} = \mathbf{SCE}$$

$$H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu\nu} \sum_c \Delta\, q_c \left(\frac{1}{2}(\gamma^h_{ac} + \gamma^h_{bc}) + \frac{1}{3}(\Delta\, q_a \Gamma_{ac} + \Delta\, q_b \Gamma_{bc}) + \frac{\Delta\, q_c}{6}(\Gamma_{ca} + \Gamma_{cb}) \right) \tag{2.66}$$

in which $S_{\mu\nu}$ is the overlap matrix. The elements of **H** depend on the Mulliken charges, which in turn depend on the spin orbital coefficients. Therefore, the equations have to be solved iteratively, until the charges are self-consistent.

2.3 Combined QM/MM Methods

The molecular mechanics methods discussed in section 2.1 are capable to describe a large system with up to tens or hundreds of thousands of atoms, which is quite appreciable in simulations of biomolecules. However, these force field methods fail to characterize the electronic structure and its changes, such as by charge transfer, electronic excitation and bond breaking and formation. On the other hand, the various quantum mechanics methods are typically developed to describe the above processes involving electronic structure changes, but they are normally restricted to small systems of up to a few hundreds of atoms. Thus, schemes combining the two techniques, the so called QM/MM methods [133–135], become necessary approaches to study large biomolecular systems.

In QM/MM methods, the system is partitioned into two parts which are then described at different levels of theory. As shown in Figure 2.2, the active center, where reactions are taking place, is subsumed into the QM region that is treated quantum mechanically. The surroundings, consisting of the protein, membrane and solvent, are the MM region which is described by a force field.

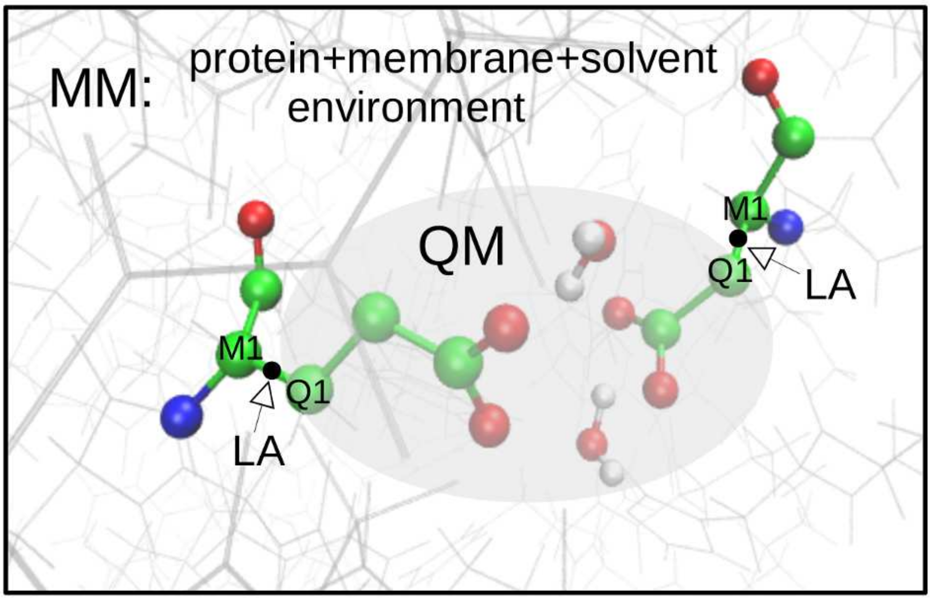


Figure 2.2: Illustration of QM/MM technique.

2.3.1 The QM/MM Energy

The total energy of the system can be expressed in two ways: subtractive or additive. In the subtractive scheme, three independent calculations are needed: (i) an MM calculation on the whole system ($E_{MM}(QM+MM)$); (ii) a QM calculation on the QM region ($E_{QM}(QM)$); (iii) an MM calculation on the QM region ($E_{MM}(QM)$). The total energy of the entire system is given by

$$E^{sub}_{QM/MM}(QM+MM) = E_{MM}(QM+MM) + E_{QM}(QM) - E_{MM}(QM) \tag{2.67}$$

where no explicit QM-MM coupling term is involved. This makes the scheme simple and straightforward but with the price of introducing some drawbacks. The QM-MM coupling is treated at the classical MM level, requiring an accurate set of MM parameters for both MM and QM regions. This has two negative consequences: (1) obtaining such accurate MM parameters for the QM region is much more difficult than for the MM region; (2) the description of the electrostatic interaction between MM and QM regions becomes erroneous, when the QM charge distribution changes in the reaction (which is the common case). Another drawback is the neglect of the polarization effects of the MM point charges on the QM region, which can be a big problem in the study of reactions involving charge transfer.

A more accurate and popular alternative to the subtractive scheme is the so-called additive scheme, in which the total energy is expressed as

$$E^{add}_{QM/MM}(QM + MM) = E_{MM}(MM) + E_{QM}(QM) + E_{QM-MM}(QM, MM) \tag{2.68}$$

There is an explicit QM-MM coupling term $E_{QM-MM}(QM, MM)$, while the MM and QM regions are handled molecular mechanically and quantum mechanically, respectively.

The essential task in the development of QM/MM algorithms is the description of the coupling between QM and MM regions. The QM-MM interaction is composed of the bonded (stretching, bending, torsion) and non-bonded (electrostatic interaction and van der Waals interacion) components, among which the electrostatic interaction is the key. Two techniques are available to deal with this interaction. One is the mechanical embedding, describing the interaction at the MM level. This method suffers the same drawbacks also occurring in the subtractive QM/MM scheme. The other method is the electric embedding, in which the QM-MM electrostatic interaction is treated by including certain one-electron terms in the QM Hamiltonian and the polarization of QM by MM will be accounted for automatically. The improvement compared to the mechanical embedding is payed for by increased computational cost and a more complicated implementation. To further improve the result, one needs to include the mutual polarization between QM and MM, i.e. the polarized QM electric field has to act back on the MM region, and then the polarized MM charges act back on the QM region till the self-consistency of the charge is reached.

The non-bonded Van der Waals interaction and the bonded interactions between MM and QM regions are handled at the MM level, which is independent of the QM/MM scheme (subtractive and additive). The MM parameters applied for the QM atoms are the key factor influencing the accuracy of these interactions. In fact, the atom types of some QM atoms can change along the reaction path, which means that one should switch the parameters during a dynamics calculation to correctly describe the interactions. However, this is usually unpractical. Alternatively, one can chose a large QM region to push the QM-MM boundary as far away from the reaction center as possible in compromise with the computational power. In principle, a large QM region reduces the deviation of bonded interactions due to the parameter choices. As to the Van der Waals interactions, the use of a large QM region does not make that much sense as for the bonded interactions. Fortunately, the introduced error is very small and does not lead to a serous problem due to the short range character of van der Waals interaction.

2.3.2 QM/MM Boundary Treatment

When the QM-MM boundary passes through covalent or ionic bonds, special care is required in the treatment of the boundary. Two approaches are available for such treatment: (i) the link atom approach, in which the "frontier atom" (Q1 in Figure 2.2) is saturated by a link atom (LA hereafter) that is usually taken to be a hydrogen atom; (ii) the localized orbitals approach, i.e. a set of localized bond orbitals are used to represent the bonds connecting QM and MM atoms (the Q1-M1 bond in Figure 2.2). The link atom method is more attractive and is widely used. Before we talk about where to put LA, firstly we need decide where to cut the bond. In principle, the atoms or groups involved in a reaction, such as bond breaking or formation,or charge transfer should be far away from the boundary. The conjugated or polarized bonds should be excluded from being cut. The bond to be cut most often is a C-C single bond.

The introduction of one LA will result in three additional degrees of freedom. To avoid this, the LA is usually put between the QM frontier atom and the MM boundary atom (i.e. on the Q1-M1 bond in Figure 2.2). The following constraint is commonly applied to define the exact position of an LA

$$C_{LA} = R_{Q1-LA}/R_{Q1-M1}, \quad C_{LA} \approx 0.71 \tag{2.69}$$

which results in a very short LA-M1 distance. This further leads to the overpolarization of the electron density at the QM-MM boundary. Various solutions have been attempted to minimize or even avoid this overpolarization effect. The simplest approach is to set the atomic charges of MM atoms nearby Q1 to zero, which often leads to the change of MM net charge and further generates artifacts in the final energy. An improved method is to delete the charges of the whole group containing the M1 atom. In this way the MM net charge is unchanged. But this method is only applicable to systems where the group containing M1 is neutral. In the case that a neutral group is not available, the so called Shift scheme finds its stage. In this method, the M1 charge is shifted evenly onto the atoms connecting to it, leading to a change of dipole moments between M1 and the atoms accepting charges from M1. In order to compensate such a dipole moment change, additional point charge pairs with opposite dipole moments are placed in the vicinity of these atoms.

2.4 Free Energy Calculation

The quantity of usable energy of a system is called free energy. Several free energy functions can be defined according to system criteria. For processes involving a system at constant pressure and temperature, the Gibbs free energy (G) is the best choice, and it finds its stage among chemists and biochemists. The Helmholtz free energy (A) is for systems at constant volume and temperature, which has great utility in physics. Next G will be highlighted.

In biological processes, the free energy difference (ΔG) reflects the reaction tendencies of molecular systems, such as the tendency of a ligand to bind to a protein and the tendency of protein conformation transitions (protein folding). The development of methodologies to calculate ΔG using molecular theory has become an important area. In general, one can obtain ΔG through standard MD in an NPT ensemble.

$$\Delta G = G(\xi_B) - G(\xi_A) = -\frac{1}{\beta}\ln\frac{P(\xi_B)}{P(\xi_A)} \tag{2.70}$$

where $G(\xi)$ is also called potential of mean force (PMF), $\beta = 1/(k_BT)$, k_B being the Boltzmann's constant, T being the absolute temperature, ξ standing for the reaction coordinate, ξ_B and ξ_A being the two states of interest (normally they are the minima along the reaction coordinate), $P(\xi)$ being the probability distribution of the conformation at ξ, which represents the occurrence frequency of this conformation relative to all other conformations.

$$P(\xi) = \frac{\int \delta(\xi^{'} - \xi)\exp(-\beta\hat{H}(x^{'}))dx^{'}}{\int \exp(-\beta\hat{H}(x^{'}))dx^{'}} \tag{2.71}$$

$\hat{H}$ (eq. 2.10) is the Hamiltonian of the conformation at $\xi^{'}$. Previous study has proved the neglectibility of the kinetic energy term in the Hamiltonian [136]. Thus, eq. 2.71 can be rewritten as

$$P(\xi) = \frac{\int \delta(\xi^{'} - \xi)\exp(-\beta V(x^{'}))dx^{'}}{\int \exp(-\beta V(x^{'}))dx^{'}} \tag{2.72}$$

where V(x$^{'}$) is the potential energy of the conformation at $\xi^{'}$, which is dependent on the selected force field as discussed in previous sections.

Sampling Issue of Standard MD

This may show in two ways:
(i) Timescale problem: One can utilize the MD simulation results only if the run is long enough for the system to cover the configuration space. In MD simulations, in order to appropriately resolve the fastest dynamics of the system (e.g. the bond stretching, bond bending), the integration time step of the equations of motions is on the order of femtoseconds. However, usually, the dynamics of interesting biological processes span a wider time range to microseconds or even to seconds. To explore these events, a quite long sampling timescale is required. Although larger and larger machines are designed, still they are far away from the requirement to approach the realistic time range. Currently, the accessible simulation time scale for classical molecular dynamics is in the range of hundreds of nanoseconds, during which most phenomena of interest become *rare events*.
(ii) Trapping in a local minimum: It is common to encounter biological processes of interest characterized by free energy barriers, which can be up to tens of kilocalories. In such a case, only if activated by the rare

thermal fluctuations, the system can overcome the barrier. But in order to obtain sufficient statistics of the rare fluctuations, an impractical amount of simulation time is required (this again comes back to the timescale problem). Consequently, in practical situations, the sampling is trapped in a local minimum and misses the global minimum along the reaction coordinate. As a result, one gets the wrong idea about the studied process. To overcome these difficulties several methods have been developed, including Umbrella Sampling [137, 138], metadynamics [139, 140], thermodynamic integration [141, 142] etc. The next two sections will introduce the basic theories behind the first two methodologies.

2.4.1 Umbrella Sampling

This method was used to evaluate the ligand binding and protein conformational change free energy properties in Chapter 3. In this method, the energy expression (energy surface) of the system is modified in order to reduce the free energy barrier by adding a bias potential $w(\xi)$.

$$V^b(x) = V^u(x) + w(\xi) \tag{2.73}$$

The superscript 'b' denotes biased quantities, 'u' unbiased quantities. The effect of the bias potential is to get sampling as uniform as possible along the whole reaction coordinate. Thus, the optimal choice would be $w(\xi) = -G(\xi)$, with which $w(\xi)$ completely compensate the free energy surface and therefore the simulation could randomly sample the whole reaction coordinate. Unfortunately, $-G(\xi)$ is unknown in prior, but rather what we aim to calculate. The popular alternative is to split the ξ of interest into a series of windows (Figure 2.3(a)). In each window a harmonic function as the bias potential is applied to keep the system close to the reference conformation ξ_i^{ref}

$$w_i(\xi) = \frac{k}{2}(\xi - \xi_i^{ref})^2 \tag{2.74}$$

where i denotes the index of the umbrella sampling windows,and k the force constant as in eq. 2.4.

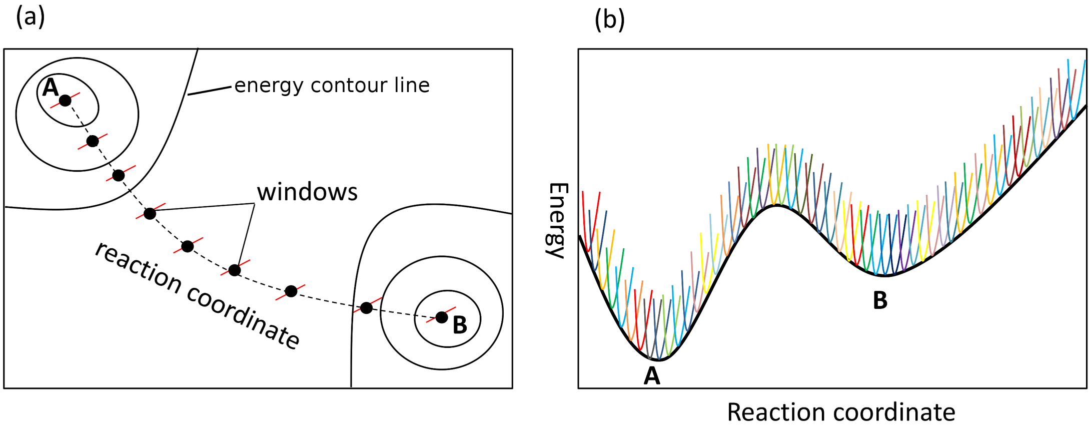


Figure 2.3: Schematic illustration of umbrella sampling method. (a) Splitting of the reaction coordinate (dashed line) between states A and B into a series of windows. (b) The contribution of the free energy $G_i(\xi)$ of each window (thin lines) to the global free energy (underlying black thick line).

The biased probability distribution of each window can be expressed as

$$P_i^b(\xi) = \frac{\int \exp(-\beta(V(x') + w_i(\xi')))\delta(\xi' - \xi)dx'}{\int \exp(-\beta(V(x') + w_i(\xi')))dx'} \tag{2.75}$$

$$= \exp(-\beta w_i(\xi)) \times \frac{\int \exp(-\beta V(x'))\delta(\xi' - \xi)dx'}{\int \exp(-\beta(V(x') + w_i(\xi')))dx'} \tag{2.76}$$

The unbiased probability distribution of each window in eq. 2.72 can be rewritten as

$$P_i^u(\xi) = P_i^b(\xi)\exp(\beta w_i(\xi)) \times \frac{\int exp(-\beta(V(x^{'}) + w_i(\xi^{'})))dx^{'}}{\int \exp(-\beta V(x^{'}))dx^{'}} \tag{2.77}$$

$$= P_i^b(\xi)\exp(\beta w_i(\xi)) \times \frac{\int \exp(-\beta V(x^{'})) \cdot \exp(-\beta w_i(\xi^{'}))dx^{'}}{\int \exp(-\beta V(x^{'}))dx^{'}} \tag{2.78}$$

$$= P_i^b(\xi)\exp(\beta w_i(\xi)) \times \int \exp(-\beta w_i(\xi^{'}))dx^{'} \tag{2.79}$$

$$= P_i^b(\xi)\exp(\beta w_i(\xi))\langle \exp(-\beta w_i(\xi))\rangle \tag{2.80}$$

Here a quantity F_i is defined as:

$$F_i = -\frac{1}{\beta}\ln\langle \exp(-\beta w_i(\xi))\rangle \tag{2.81}$$

Then the PMF of each window ($G_i(\xi)$) can be calculated from the following equation

$$G_i(\xi) = -\frac{1}{\beta}\ln P_i^b(\xi) - w_i(\xi) + F_i \tag{2.82}$$

Since $w_i(\xi)$ is harmonic, $G_i(\xi)$ from eq. 2.82 is also in the harmonic form. Therefore, combining the minima of $G_i(\xi)$ of each window one gets the global PMF $G(\xi)$ as shown in Figure 2.3(b).

In eq. 2.77 P_i^b can be obtained from the biased MD simulations. $w_i(\xi)$ is provided analytically. F_i can be estimated using various methods, among which the weighted histogram analysis method (WHAM) is commonly employed. The readers are guided to the details of WHAM in the literature [143, 144].

2.4.2 Metadynamics

This method was applied to study the residue E123 side chain conformational change free energy properties in Chapter 5.

The Algorithm

Metadynamics is powerful in reconstructing the free energy, exploring new reaction pathways and accelerating rare events of the system in standard MD at the classical or quantum level. As for umbrella sampling, in this method, a set of reaction coordinates need to be defined first of all, also referred to as collective variables (CVs). The probability distribution of the CVs is defined as

$$P(s) = \frac{\exp(-\beta\ F(s))}{\int \exp(-\beta\ F(s))ds} \tag{2.83}$$

where s denotes the values of the CVs. The free energy of the system is expressed as

$$F(s) = -\frac{1}{\beta}\ln\left(\int \exp(-\beta\ U(x))\delta(s - s(x))dx\right) \tag{2.84}$$

Here, x is the microscopic coordinate (e.g. the atomic position, electronic coordinate), $s(x)$ the function of the coordinates, and $U(x)$ the potential energy function of the system.

In metadynamics, the sampling is facilitated by introduction of a history-dependent bias potential that acts on the CVs. This potential is a sum of Gaussians deposited along a coordinate trajectory in the CVs space, which can be expressed as

$$V_G(s,t) = \omega \sum_{\substack{t^{'}=\tau_G,2\tau_G,\ldots \\ t^{'}<t}} \exp\left(-\sum_{i=1}^{d}\frac{(s_i(x) - s_i(x(t^{'})))^2}{2\delta_{s_i}^2}\right) \tag{2.85}$$

where τ_G is the frequency at which the Gaussians are deposited, δ_{s_i} the width of the Gaussian for the *i*th CV (s_i), ω the height of the Gaussian, and d the total number of CVs.

The effect of the metadynamics bias potential is to push the system out of local minima into visiting new regions of the CVs space. As shown in Figure 2.4, the black thick line in the lower panel represents a free energy surface carrying three minima. Let us assume that the starting structure is the local minimum B. In a metadynamics simulation, Gaussians are deposited with time, which results in the growth of the underlying bias potential (thin lines). At time t_2 the system escapes from minimum B by overcoming the lowest barrier on the left side and falls into the local minimum A, whereby Gaussians are deposited again. Until the accumulated bias potential completely fills the underlying free energy basin at time t_4, the system is able to come out of A. Then it diffuses between minima A and B with the continued deposition of Gaussians, till at time t_5 the highest saddle point on the right side is passed, and the system accesses minimum C. At time t_8, minimum C is also completely filled by the bias potential. After that the system randomly diffuses in the CVs space. The deposited bias potentials V_G at different times are depicted in the upper panel. Here the basic assumption of metadynamics is introduced that after a sufficient long time simulation, $V_G(s,t)$ converges to the minus of the free energy as a function of the CVs

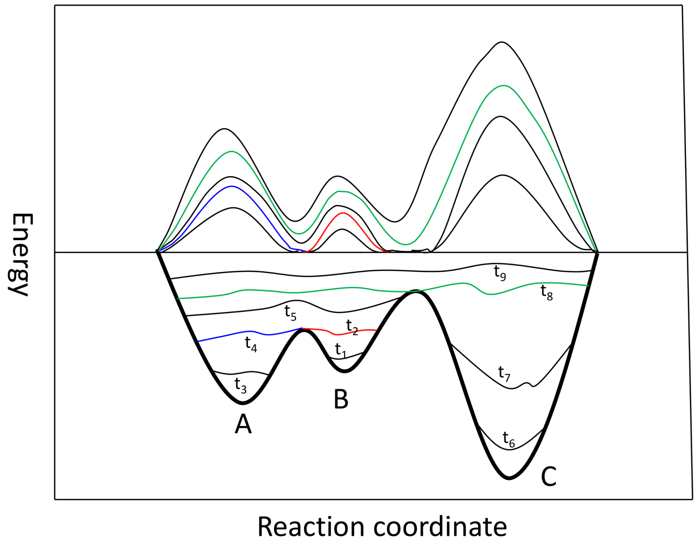


Figure 2.4: Schematic illustration of metadynamics method. Upper panel: Time evolution of deposited bias potential V_G. Red line: V_G at which minimum B is filled; blue line: V_G at which minimum A is also filled; green line: V_G at which the entire profile is filled. Lower panel: The progressive filling of the underlying free energy surface (the black thick line) through the deposition of the Gaussions along the CVs trajectory. The sum of V_G and the underlying free energy is displayed at different time (thin lines).

$$V_G(s, t \to \infty) = -F(s) + C \tag{2.86}$$

where C is an irrelevant constant. Note that this assumption does not derive from any ordinary thermodynamic identity, but was postulated heuristically.

Notably, in standard metadynamics, the height of deposited Gaussians is constant for the entire course of a simulation. This possibly results in the overfilling of the free energy surface and pushes the system to high energy region of the CVs space. Consequently, the estimated free energy from the bias potential oscillates around the real value. A solution to this problem is provided by well-tempered metadynamics [145], in which the heights of the Gaussians are decreased with the proceeding of the simulation as follows

$$\omega(t) = \omega_0 \exp\left(- \frac{V_G(s,t)}{k_B \Delta T} \right) \tag{2.87}$$

where ω_0 is the initial Gaussian height, and ΔT an input parameter with the dimension of a temperature. In a practical simulation, ΔT is determined by the "biasfactor" term γ

$$\gamma = \frac{T + \Delta T}{T} \tag{2.88}$$

T is the system temperature. The free energy is then obtained through

$$V_G(s, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(s) + C \tag{2.89}$$

The Collective Variables

The CVs can be any explicit function of the microscopic coordinates of the system, such as distance, angle, dihedral etc. The choice of the CVs can have significant influence on the reliability of metadynamics. The meaningful CVs should fullfil three criteria: (i) Discriminate clearly the initial and the final state, and include all the relevant intermediate states. (ii) Describe all the slow events of interest. (iii) Be few in number to save computation time. Choosing a proper set of CVs requires a thorough understanding of the investigated system, physically and chemically. Normally, there is no *priori* recipe for finding the right CVs. It is rather a trial and error procedure.

2.5 Vibrational Spectra Calculation

Vibrational spectroscopy is a widely used tool to study the structures and dynamics of chemical systems. Infrared (IR) and Raman spectroscopies are two of the most important experimental techniques. The corresponding vibrational spectra are accessible through quantum mechanical methodologies.

2.5.1 Normal Mode Analysis

In the NMA approach, the information of the vibrational modes and vibrational frequencies are obtained from the diagonalization of the Hessian matrix **H** that is the second order derivatives of the energy with respect to the atomic displacements.

$$H = \frac{\partial^2 E}{\partial \Delta \tilde{\vec{R}}_i \partial \Delta \tilde{\vec{R}}_j} \tag{2.90}$$

where $\Delta \tilde{\vec{R}}_i$ and $\Delta \tilde{\vec{R}}_j$ are the mass weighted nuclear displacements of atoms i and j with respect to the equilibrium positions. The use of the mass weighted coordinates simplifies the calulations.

$$\tilde{\vec{R}}_i = \sqrt{m_i}\, \vec{R}_i \tag{2.91}$$

where m_i is the mass of atom i.

The eigenvalues of **H** are proportional to the squared vibrational frequencies. The eigenvectors of **H** are the mass weighted atomic displacements that describe the vibrational modes. In cases where analytical second order derivatives are not available, numerical differentiation of the forces has to be used. However, this numerical method becomes troublesome for large systems.

The spectra intensities are calculated independently through the first order derivative of the molecular dipole moment (for IR) or polarizability (for Raman) with respect to the normal mode. The shape of the vibrational band is not accessible in the NMA approach due to the calculation at a termperature of 0 K.

2.5.2 Fourier Transform of Time-Correlation Functions

The FTTCF approach is based on Fermi's golden rule and the assumption that the linear response theory is valid. Then the IR and Raman spectra are expressed as

$$\text{IR}: \quad I(\omega) \propto \int_{-\infty}^{\infty} \langle \mu\,(t) \cdot \mu\,(0) \rangle\, e^{-i\omega t} dt \tag{2.92}$$

$$\text{Raman}: \quad I(\omega) \propto \int_{-\infty}^{\infty} \langle \alpha(t) \cdot \alpha(0) \rangle\, e^{-i\omega t} dt \tag{2.93}$$

where $\mu\,(t)$ and $\alpha\,(t)$ are the time series of molecular dipole moments and polarizabilities, respectively. They can be obtained through classical, quantum mechanical, or combined QM/MM dynamics simulations. Applying a Fourier transform on eq. 2.92 and eq. 2.93, one gets the complete IR or Raman spectrum including frequency, intensity and band shape information. A quantum correction factor for the spectral intensity is usually used to account for classical effects

$$Q_{HC} = \frac{\beta\hbar\omega}{1 - \exp(-\beta\hbar\omega)} \tag{2.94}$$

where $\beta = 1/k_B T$, k_B is the Boltzmann's constant and ω is the angular frequency that is proportional to the vibrational frequency ($\omega = 2\pi\nu$). Since FTTCF requires the computation of an ensemble of structures which is usually computationally very demanding for a single spectrum, this method is less popular then NMA in the case where the band shape is not required.

2.5.3 Principal Component Analysis

The PCA approach has been successfully used to extract the collective variables (modes) representing the main motions that contribute to the dynamics of biomolecules, like proteins. Generally, in this method, the statistical analysis is performed on the mass weighted atomic displacements to build the covariance matrix. The diagonalization of the covariance matrix yields the eigenvalues and eigenvectors carrying the information about vibrational frequencies and vibrational modes.This method requires long molecular dynamics sampling. It is only suited for calculating the slow vibrations.

2.6 Photochemistry

Photochemistry studies the physical changes and chemical reactions resulting from interactions between matter and light [146]. The energy of both matter and light is quantised, i.e. only certain specific energies are allowed. Due to the inherent quantum mechanical nature, the matter-light interaction is a complex phenomenon which is still not fully understood and is not easily described. Next, some concepts in this field and some theoretical methods for studying molecular photoisomerization will be introduced.

Light is composed of photons which carry energy. If the energy of a photon equals the energy difference between two electronic states of a molecule, the photon will be absorbed and triggers the excitation of an electron from a lower energy level to a higher energy level. The electronic wavefunction is changed due to this electronic transition, which brings the nuclei onto a different potential energy surface. This electronic transition process is associated with a transition dipole moment (μ). The efficiency of electronic transitions is usually evaluated by Fermi's Golden rule [147]

$$k_{i\rightarrow f} = \frac{2\pi}{\hbar} |\langle \Psi_f | \hat{\mu} | \Psi_i \rangle|^2 \rho_f \tag{2.95}$$

where Ψ_i is the initial state wavefunction, Ψ_f the final state wavefunction, and ρ_f the final state density.

In the excited state, there are electron(s) occupying anti-bonding orbital(s). Thus, the molecule is metastable and will relax back to the ground state sooner or later. In principle, the molecule can relax radiatively or radiationlessly (see Figure 2.5). The radiative relaxation includes fluorescence and phosphorescence. The radiationless relaxation normally refers to the decay at a conical intersection (CI, the point where the excited state and the ground state energetically degenerated, i.e. the two states have the same energy at conical intersection) [148, 149]. Which of these processes occurs depends on the structure of the molecule and also on the nature of the transition [146].

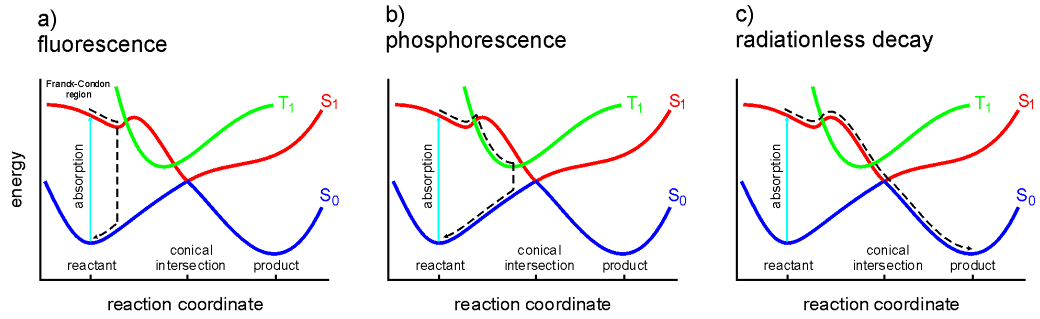


Figure 2.5: Excited state relaxation mechanisms. Photo absorption (cyan vertical arrow) promotes the molecule to the exited state S_1. The exited molecule could relax back to the ground state (S_0) through (a) fluorescence, (b) phosphorescence, or (c) decay at a conical intersection. In (a) and (b), the relaxation to the ground state is accompanied with photon emission. Moreover, in (b) the molecule undergoes a singlet-triplet transition $S_1 \rightarrow T_1$ before the photon emission. In contrast, the relaxation in (c) is radiationless. Figure is adopted from ref. [150] with permission.

Fluorescence

As shown in Figure 2.5 (a), the molecule is excited according to the Frank-Condon principle (The cyan vertical arrow). The electrons move much faster than the much heavier nuclei, and the transition from one electronic state to another is very rapid. Thus, the nuclei are assumed to be fixed during the instantaneous electronic transition [146]. After that, due to the thermal relaxation it is trapped in one minimum of the excited state potential energy surface. The only way in which the system can relax back to the ground state is to emit a photon carrying the amount of energy that equals the gap between S_1 and S_0. The efficiency of fluorescence can also be described through Fermi's Golden rule in eq. 2.95. Note that fluorescence happens between states of the same spin multiplicity.

$$m_s = \pm\frac{1}{2} \tag{2.96}$$

$$S = \sum m_s \tag{2.97}$$

$$\text{spin multiplicity } = 2S + 1 \tag{2.98}$$

where m_s is the spin quantum number denoting the electron spin up ($m_s = +\frac{1}{2}$, $\uparrow$) and spin down ($m_s = -\frac{1}{2}$, $\downarrow$). The spin multiplicity 1, 2, 3 etc. corresponds to singlet (S), doublet (D), triplet (T) electronic configurations etc., respectively.

Phosphorescence

Normally, the total electron spin is not changed in the electronic transition procedure, e.g. the $S \rightarrow T$ transition is forbidden. However, this does not strictly hold because of the coupling between the spin magnetic moment with the orbital magnetic moment generated by electron motions (spin-orbit coupling) [146]. This is especially true for systems containing heavy atoms. Therefore, when the singlet and triplet energy surfaces are very close energetically or even intersect, it is possible that the electron flips its spin and turns the molecule to a triplet state (this process is referred to as intersystem crossing), as shown in Figure 2.5 (b). Then the molecule relaxes to a minimum on the T_1 potential energy surface, from where it relaxes back to the ground state S_0 by emitting a photon.

Radiationless Decay At Conical Intersection

The occurrence of radiationless decay is because of the breakdown of the Born-Oppenheimer approximation at conical intersection. Normally, the potential energy surface is obtained within the adiabatic or Born-Oppenheimer approximation, within which the vibrational effect of the nuclei is neglected due to the much larger electronic energy gap than the nuclear vibrational energy gap. However, in a conical and near conical intersection, the vibrational gap becomes energetically comparable to the energy difference between electronic states. Therefore, the electronic and nuclear wavefunctions might interfere, which leads to the phenomenon that slight changes in nuclear positions result in transitions between different electronic states. This phenomenon is referred to as non-adiabatic coupling [151].

Figure 2.5 (c) illustrates the decay at a conical intersection. Depending on the location and the topology of the intersection, the system can either restore the initial configuration (photophysical process) or reach a chemically different species (photochemical process, dashed line in Figure 2.5 (c)). This mechanism plays an important role in chemistry and biology. For example, the *trans*$\rightarrow$*cis* isomerization or vise versa of *gluazo* and retinal under proper illumination is a photochemical reaction through this mechanism. The quantum yield of a photochemical reaction is the ratio between the number of product molecules and the number of photons absorbed.

A detailed description of non-adiabatic molecular dynamics simulations using the *diabatic surface hopping* algorithm to study the decay at conical intersection can be found in the refs. [150, 152, 153]. In this method, the ground and excited states are calculated with a high level quantum method, such as CASSCF or CASPT2. Thus, it is computationally very expensive. In the present study of *gluazo* photoisomerization, we employed a much simpler and cheaper non-equilibrium molecule dynamics strategy [154, 155]. This method could adequately fulfill our purposes of obtaining the *cis-gluazo* and investigating the protein responses to the isomerization of *gluazo*.

Previous high-level ab initio calculations support that the dihedral around the N=N double bond is the isomerization coordinate of azobenzene [156–158]. In this study, a cosine function (eq. 2.99) is applied to model a potential energy curve connecting the excited (S_1) *trans* state and the ground (S_0) *cis* state of *gluazo* (dashed line in Figure 2.6).

$$V(\theta) = \frac{K}{2}(1 + \cos\theta), \quad K = 320\,\text{kJ/mol} \tag{2.99}$$

where θ is the dihedral around the N=N double bond of *gluazo*.

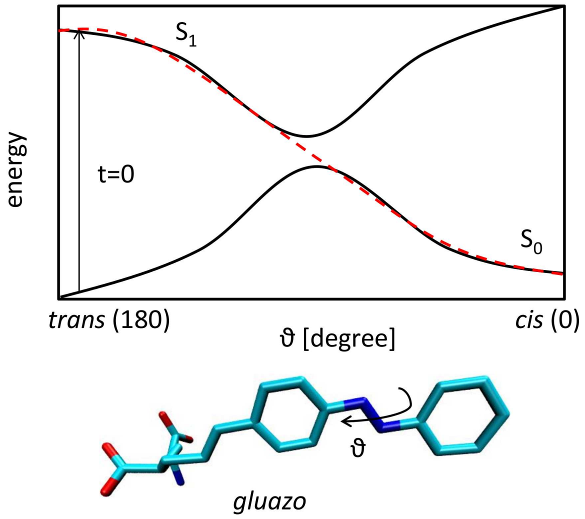


Figure 2.6: Illustration of *gluazo* photoisomerization. The solid lines denote the adiabatic potentials of the ground state (S_0) and excited state (S_1) as a function of the dihedral (θ) around the N=N double bond of *gluazo* (shown by the arrow lower in the figure). The dashed line represents the *trans-cis* photoisomerization potential, which connects the excited *trans* state and the ground *cis* state of *gluazo*.

The photoexcitation process under an ultrashort laser pulse is mimed through instantly switching from the ground state potential energy surface to the excited potential energy surface, as shown by the vertical arrow in

Figure 2.6. After that *gluazo* isomerizes along the designed potential energy curve within 100 fs (Figure 3.7). The N=N torsional potential is switched back to the ground state after 500 fs, and a standard MD simulation of 500 ns is performed to study the protein responses.

Chapter 3

Molecular Dynamics Investigation of Kainate Receptor GluK2

Reproduced from ref. [33] with permission from the PLoS ONE Owner Societies. ("Guo, Y.; Wolter, T.; Kubař, T.; Sumser, M.; Trauner, D,; Elstner, M. "Molecular dynamics investigation of *gluazo*, a photo-switchable ligand for the glutamate receptor GluK2." *PLoS ONE*, 2015, 10, e0135388."

The Trauner lab has recently introduced the concept of photopharmacology and, more specifically, the application of PCLs to edow native receptors with sensitivity towards light. Among many successfully targeted receptors, PCLs acting on iGluRs were developed [159, 160]. PCLs can be switched reversibly between two conformations (*tran*/*cis*), thereby altering their affinity towards their target receptor. In general, these molecules contain a light-sensitive azobenzene group attached to the respective ligand.

In a previous work, the effect of the PCL ATA-3 on AMPA receptors was studied [161]. Currently, a crystal structure of *gluazo* bound to the LBD GluK2 is available [20]. This makes it possible for the first time to simulate the mechanism of structural changes in the target receptor induced by the isomerization of *gluazo*. Upon illumination of UV light, the structure of *gluazo* changes from an extended and flat in the *trans* conformation to a bulky geometry in the *cis* configuration, see Figure 1.4. Following on the previous studies of GluA2 [161, 162], molecular dynamics, an ideal tool for investigation of the structure and dynamics of biomolecular systems, was employed to probe the mechanism of GluK2 response to ligand photoisomerization.

3.1 Model Building

The starting point for the current study is the X-ray structure of *trans*-*gluazo* bound to the GluK2 LBD monomer (referred to as GluK2-*trans* further), PDB ID 4H8I [20]. The amino acid residues that were not resolved in the experiment were added using Modeller v9.13 [163]. These were single residues at the termini of the peptide chain. The complete structure of GluK2-*trans* complex is shown in Figure 3.1. The complex was embedded in a periodic rectangular box sized 7.6 nm×7.1 nm×6.8 nm, solvated with ca. 10,000 TIP3P water molecules [164] and neutralized with sodium cations [165]. Amber99SB force field [166] was used to describe the protein. Because Amber99SB force field does not include parameters for ligands in the current study, all the ligands were parametrized through quantum chemical calculations.

Following the previous work on GluA2 [161], an LBD monomer rather than a dimer or even higher oligomer was simulated. One reason for this choice is to save computational resources because simulations of LBD monomer are already expensive. The other one is to avoid any sampling issues stemming from the

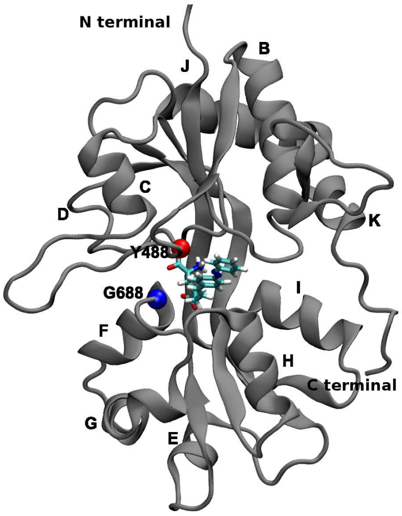


Figure 3.1: Crystal structure of GluK2 LBD domain bound with *trans-gluazo*. Cartoon mode–receptor, red ball–Y488, blue ball–G688, tubes colored by atom–ligand. The main helices of the protein are labeled.

extremely weak motions of the monomers relative to each other. The choice of simulating a monomer was assessed in ref. [161] (supporting information), in which the internal dynamics of the LBD was shown to be the same in the monomer and in the dimer. Nevertheless, there were differences between two models: (i) The clamshell was able to open wider in a free monomer due to the lack of inter-subunit contacts, which restrict the clamshell opening in an LBD dimer. (ii) The response of the channel (or the linker segment replacing the channel) induced by the motion of the clamshell was smaller in an LBD dimer than in the monomer, because of the interfering dynamics of the monomer–monomer interface. Notably, the most important qualitative results, i.e the identification of interactions governing the ligand binding and the characterization of protein response to *gluazo* isomers, will not be affected.

3.2 Simulation Details

An equilibration procedure was performed prior to the production simulations, to remove the inappropriate atomic conflicts. It started with a steepest descent minimization of maximum 2000 steps, during which the protein and the solvent were relaxed. The maximum force tolerance was set to < 100 N.

The minimized structure was then subjected to a two-phase MD equilibration. The first phase was performed under NVT ensemble for 500 ps: the initial atomic velocites were generated on the basis of Maxwell–Boltzmann distribution [115] at 300 K; the temperature was maintained at 300 K using the Nosé–Hoover thermostat [117, 118] with τ_t=0.2 ps; the harmonic position restraints were applied to the heavy atoms of the protein and the ligand with the force constant of 1000 kJ·mol^{-1} ·nm^{-2}. Subsequently, the system was proceeded to the NPT equilibration with the same position restraints as in the first phase for another 500 ps, involing an isotropic Parrinello–Rahman barostat [120] to maintain the pressure at 1 bar with a relaxation time constant, τ_p=2.0 ps and a compressibility parameter of β=4.5$\times$10^{-5} bar^{-1} for the x,y-direction and β=4.5$\times$10^{-5} for the z-direction. After that, the force constant imposed on the solute was reduced gradually from 1000 to 100 kJ·mol^{-1} ·nm^{-2} in 10 ns, and then to 10 kJ·mol^{-1} ·nm^{-2} in 20 ns before the restraints were abandoned altogether.

As for the production runs, five independent simulations of 500 ns with different initial velocities were conducted. The equation of motion was integrated through the leap-frog algorithm [114] with a time step of 2 fs. Covalent bonds were constrained using the LINCS algorithm [167]. Regarding the treatment of non-bonded interactions, the long-range point-charge electrostatics were evaluated using the particle-mesh Ewald method [168], while a cut-off of 1.0 nm was used for the van der Waals interactions. The non-bonded interaction pair-list were updated every 10 fs, applying a cut-off of 1.0 nm. All the simulations were performed with the GROMACS package, version 4.6.1 [169, 170].

3.3 Results

3.3.1 Structure of the Complex with *trans-gluazo*

Conformation Searching

Taking into account of the considerable physical size of *gluazo*, its position in the LBD of Gluk2 may not be limited to the one that occurs in the X-ray structure. In order to get an insight into other possibilities of *gluazo* binding mode, we performed several independent simulated annealing simulations. Firstly the system is heated up rapidly from a low temperature at which the properties of the system is interested in, to a high temperature. Then it is cooled down slowly back to the previous low temperature. During cooling, the studied system is allowed to randomly explore the conformational space. Simulated annealing is efficient in studying protein conformational fluctuations, and it has been successfully used by some researchers. For instance, Cordes et al. [171] studied the structure of Vpu (which is HIV-1 auxiliary protein) through simulated annealing simulation, finding five conformations for the homo-pentameric Vpu bundles. In Mayewski's research [172], the author utilized simulated annealing to search for the global minimum of tested proteins successfully. In the present study, this technique was used to search for the possible conformations besides the X-ray structure. The structure after geometry optimization was used as starting structure. For the data collecting period, the complex was directly cooled down from 450 K, 445 K, 440 K, 435 K, 430 K, respectively, to 300 K in the first 100 ns, and then was kept at 300 K for another 100 ns (see Figure 3.2).

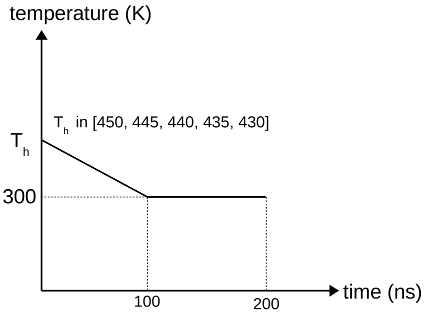


Figure 3.2: The simulated annealing scheme used in the current study.

The X-ray structure resolves the conformation that *trans-gluazo* is in front of the "E440-N721" gate (reffered as EN-gate next, Figure 3.3a). The residues E440 and N721 locate close to each other, opposite at D1 domain and D2 domain, respectively. The hydrogen bond (HBond) between E440 and N721 contributes to this EN-gate.

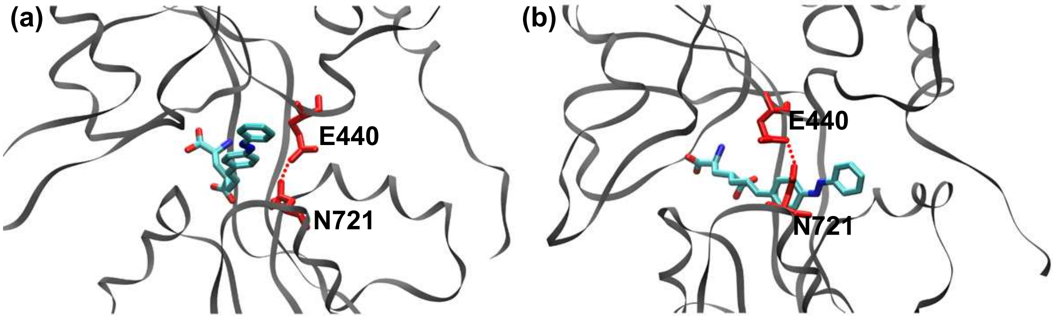


Figure 3.3: Two ligand binding modes of *trans-gluazo*. (a) Position 1 (P1): *gluazo* is in front of the EN-gate. (b) Position 2 (P2): *gluazo* is behind the EN-gate. Red dotted line–EN-gate, i.e. the hydrogen bond between E440 and N721. E440 and N721 are located close to each other, in D1 and D2 domain of LBD, respectively.

In simulated annealing simulations, the protein was crashed in the cooling procedure from 450 K to 300 K, which suggests that 450 K is too high for GluK2. Cooling from 445 K, 440 K and 430 K, the ligands stays at the initial position. In contrast, cooling from 435 K, a new ligand position with respect to the X-ray structure was observed, i.e. *gluazo* is behind the EN-gate (Figure 3.3 b). We define the ligand position in front of the EN-gate as position 1 (P1), and that behind the EN-gate as position 2 (P2). Then the X-ray structure is further defined as GluK2-*trans*-P1, the newly found complex as GluK2-*trans*-P2.

Free Energy Study of Ligand Binding Modes

Free energy calculation was performed using umbrella sampling method, in order to study the free energy properties of the two ligand binding modes. The GluK2-*trans*-P1 complex was used as the starting point. The reaction coordinate was defined as the center of mass (COM) distance between the azobenzene group (without H atoms) and a group of protein amino acid residues (111–115 and 191–205, without H atoms), see Figure 3.4. The reaction coordinate distance from 1.75 nm to 0.80 nm was sampled to guarantee that the ligand is moved from P1 to P2. Twenty one US windows were obtained, spaced by an equidistance of 0.05 nm. The initial structure for US simulations in every window was obtained through a pulling simulation, in which the pulling force was performed on the azobenzene group. A very small pulling speed of 10^{-5} nm ps^{-1} was used to avoid unphysical distortions within the protein structure. A harmonic biasing potential with a force constant of ca. 500 kcal·mol^{-1}·nm^{-2} was applied in each window. A 300 ns length MD simulation was performed for each window.

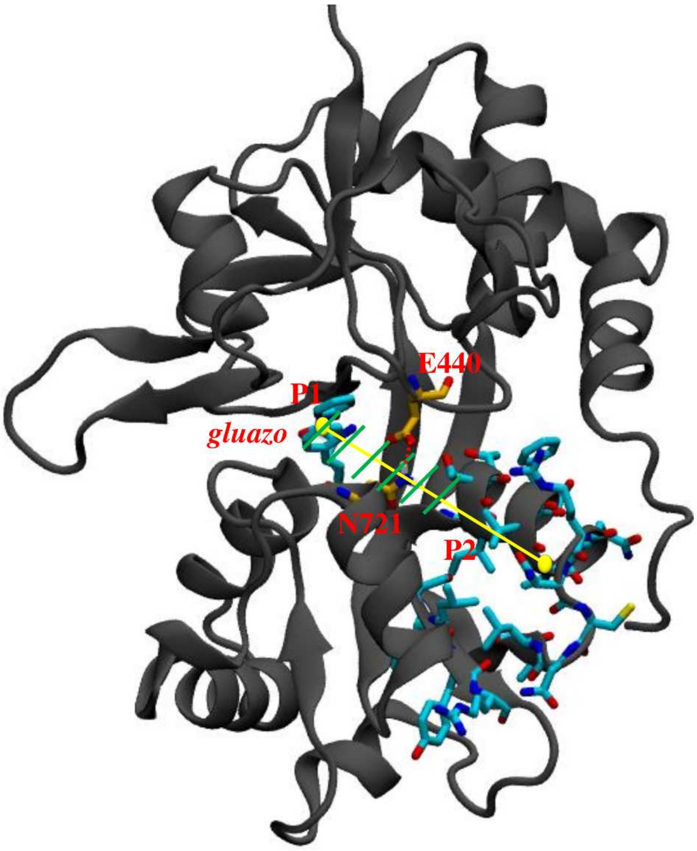


Figure 3.4: The illustration of umbrella simulation coordinate for studying ligand binding modes free energy properties. The COM defining the reaction coordinate are represented by yellow balls. The ligand will move from P1 to P2 by crossing the EN-gate. The simulation windows are depicted using the green lines.

The quality of sampling and convergence of simulations was assessed by the overlap of histograms of each window (Figure 3.5), and by the evaluation of free energies with gradually increasing simulation time (Figure 3.6). The potential mean force (PMF) changes slightly after 200 ns, indicating the convergence of the simulation.

The PMF curve after 300 ns shows an energy barrier of ca. 14.4 kcal·mol^{-1} from P1 to P2, revealing that the binding mode of *trans-gluazo* at P1 is more stable than at P2. Therefore, the present study focuses on the binding mode at P1.

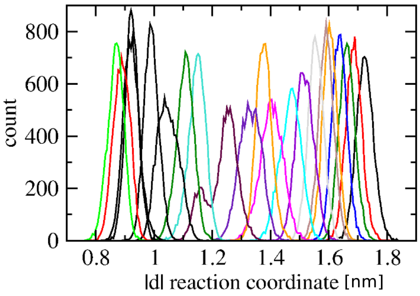


Figure 3.5: Histograms from the umbrella sampling simulations moving *trans-gluazo* from P1 to P2.

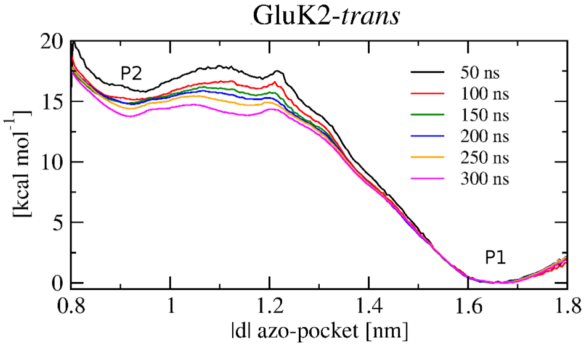


Figure 3.6: Free energy profiles for moving *trans-gluazo* from P1 to P2 obtained from umbrella sampling simulations. Each of the different curves was obtained from a series of umbrella sampling simulations with the time scale specified in the legend. Each free energy curve is shifted such that its value in the respective global minimum vanishes.

3.3.2 Isomerization of *gluazo*

A forced switching simulation as described in Chaptor 2 was performed on the GluK2-*gluazo* complex to investigate the protein response to the ligand photoisomerization. This method has been applied in previous research of GluA2 successfully [161].

The final structures from the extended equilibration were used as the starting points for the switching simulations. The C-N=N-C dihedral angle twisted from 180° to ca. 0° rapidly, accomplishing the isomerization within 100 fs. This is in accordance with the previous simulation reports of azobenzene isomerization occurring within 250 fs [154, 155, 157, 158]. The C-N=N-C dihedral after 100 fs was -3.0°, 7.8°, -4.2°, 7.1° and -5.4°, respectively, in the five independent trajectories, and it fluctuated around 0° afterwards (Figure 3.7).

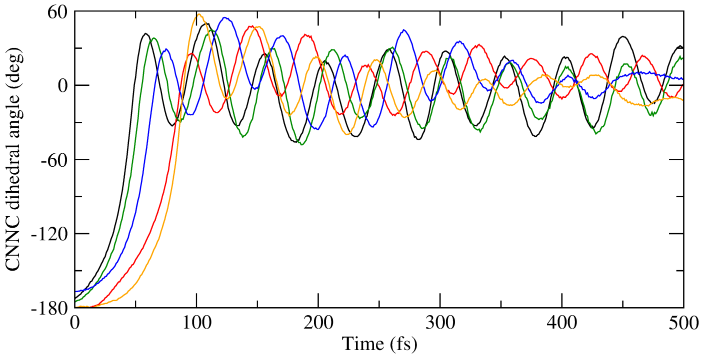


Figure 3.7: Variation of the dihedral angle around N=N during the ismerization of *trans-gluazo*. Data from five independent simulations are shown (depicted using different colors). The additional restraining potential was placed on the dihedral from $t = 0$ fs.

To investigate the ligand isomerization effects on the protein conformation during the short time scale, the root mean square deviation (RMSD) of the protein backbone relative to the starting structure was calculated (see Figure 3.8). In all of the five simulations, very small RMSD of ca. 0.02 nm was observed during the first 100 fs. This indicates that the isomerization of *gluazo* does not lead to immediate structural response of the protein. The same result was observed in the previous study on GluA2 [161]. The RMSD increases further to ca. 0.04 nm at 500 fs, which indicates that while the structure of the binding pocket and its closest vicinity adjusts to the isomerization of *gluazo* rapidly, the overall protein conformation remains rather intact on the sub-picosecond time scale (see Figure 3.9). Any large-scale conformational transitions, which lead to the opening of the channel eventually, will take place more slowly, on a time scale reaching hundreds of nanoseconds possibly.

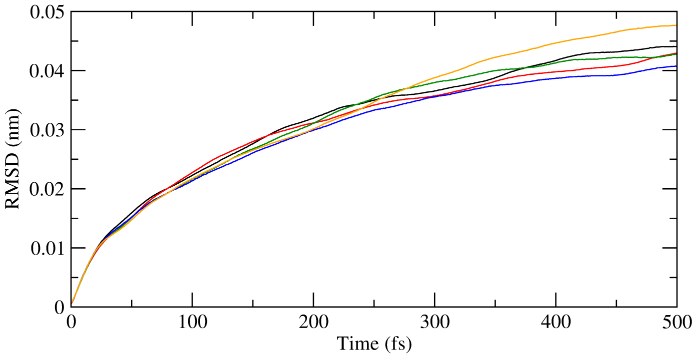


Figure 3.8: RMSD of the protein backbone relative to the starting structure in the course of the isomerization simulations of *trans-gluazo*.

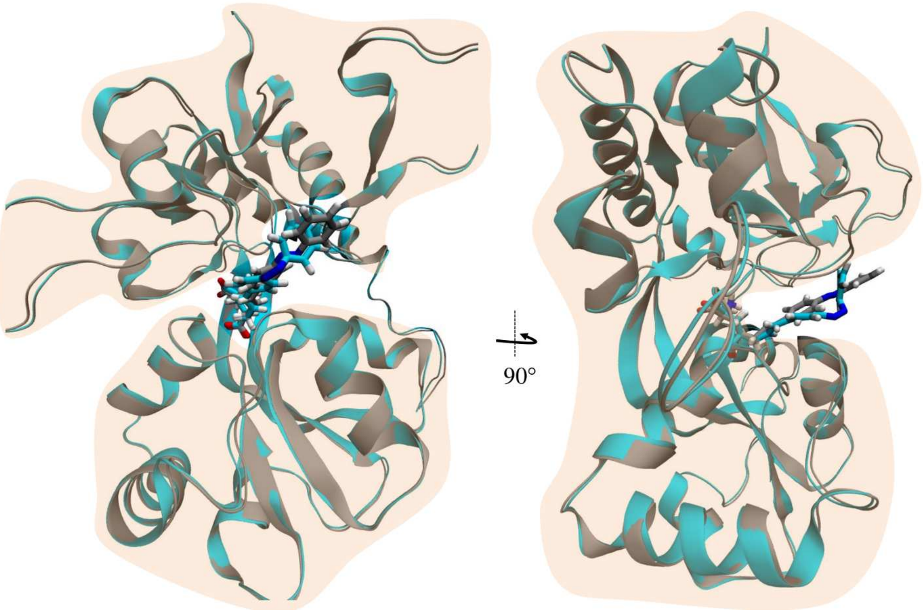


Figure 3.9: GluK2-*trans* and GluK2-*cis* complexes. The protein (ribbon) and ligand (stick) of GluK2-*trans* is colored in silver, while GluK2-*cis* is colored in cyan. The GluK2-*cis* structure is the last frame from the isomerization simulation. Left: front view. Right: side view.

The isomerization simulations provided a complex of *cis-gluazo* bound to GluK2 LBD, which will be used and referred to as GluK2-*cis* in the production MD simulations.

3.3.3 Protein Conformation Changes Upon Ligand Binding

To assess the effect of different ligand configurations on the structure and dynamics of the protein, five independent MD simulations of 500 ns each were performed on GluK2-*trans* and GluK2-*cis* complexes. Prior to this extended simulation, the GluK2-*cis* complex was subjected to the same equilibration scheme as that for GluK2-*trans*.

The RMSD of protein backbone with respect to the starting structure along each trajectory is shown in Figure 3.10 for both complexes. The value increases gradually to 0.20–0.25 nm at 50 ns in all of the simulations. In the further course of several simulations, a large increase of the RMSD values was observed. An additional analysis showed that this corresponds to the LBD clamshell wide-opening state (see below). This might have been a consequence of the simulation setup involving a LBD monomer. Recall that such a wide opening of the clamshell is not possible in a LBD oligomer due to the sterical clashes of the D2 domains of the different monomers. Thus the wide-open conformation may be over-stabilized in the current simulations artificially, while the closed and moderately open (as with a partial agonist) conformations are described correctly. Therefore, in the following discussions, the simulations where wide-open conformations occurred will not be considered, and the closed and moderately open conformations will be concentrated on.

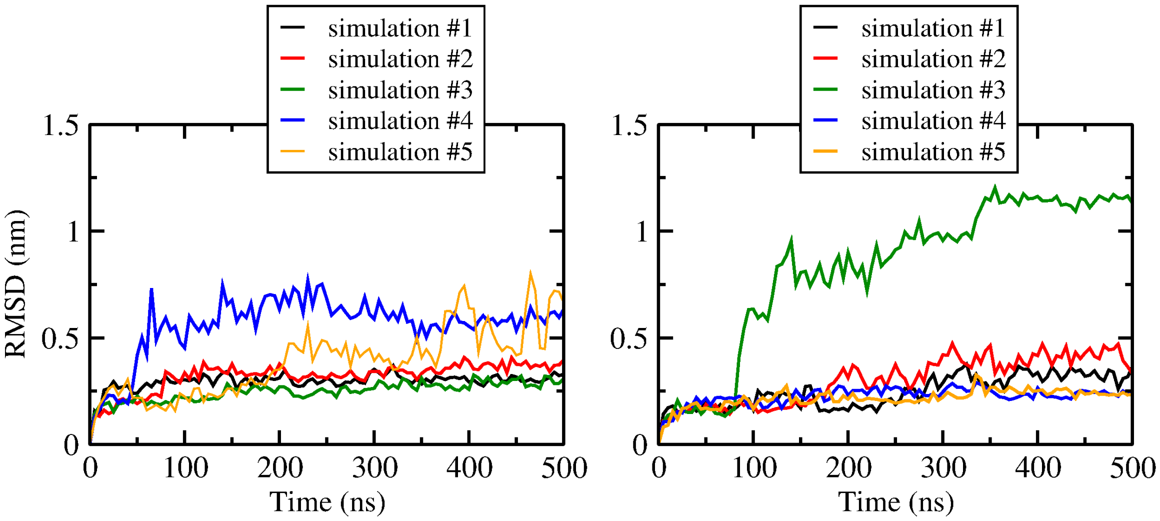


Figure 3.10: RMSD of the protein backbone with respect to the starting structure in the course of 500 ns MD simulations. Left: GluK2-*trans* complex; right: GluK2-*cis* complex. The black, red, green, blue, orange curves correspond to the simulations #1, #2, #3, #4, #5, respectively.

GluK2-*trans* Complex

Among the simulations of GluK2-*trans* complex, two trajectories show a distinct D1–D2 clamshell opening depicted through a steep rise of protein backbone RMSD (Figure 3.10 left, simulations #4 and #5). Another two simulations with comparable RMSD values (Figure 3.10 left, trajectories #1 and #2) show quite different protein stability. That is, trajectory #2 shows the clamshell opening, while trajectory #1 does not. This suggests that RMSD may not always be a good measurement of the conformational change of the protein, especially considering that the protein features flexible loops.

In GluA2, the COM distance between residues G451 (corresponds to G489 in GluK2) and S652 (D687 in GluK2) had been proposed as a good descriptor of the closing/opening extent of the LBD clamshell [161, 162]. In the present work, the COM distance of residues Y488 and G688 is considered to describbe the clamshell opening and closing movement, motivated by the large distance of these residues from the hinge region of the clamshell. Also, it can be inferred from the structure of the complex (Figure 3.1) that Y488 and G688 are located near the termini of loops connecting secondary structure elements, which will reduce the bias caused by the inherent flexibility of the loops.

The Y488–G688 distance is 0.63 nm in the closed state of GluK2 (PDB ID 1S50, bound with glutamate), and it is 0.82 nm in the partially closed state (PDB ID 2XXT, bound with kainate). Since no X-ray structure of the GluK2 *apo*-protein or a GluK2-antagonist complex is available, the complex of GluK1 bound with the antagonist UBP-315 (PDB ID 2QS1) is considered as the reference for the open state, which features a Y505–G704 COM distance of 1.15 nm (Y505 and G704 corresponds to Y488 and G688 in GluK2, respectively). The COM distance of Y488–G688 shown in Figure 3.11 captures the clamshell opening and closing in GluK2-*trans* well, with the average values of 0.97 nm, 1.19 nm, 0.75 nm, 1.65 nm and 1.25 nm for the five respective simulations #1 through #5. It is inferred that the simulations #1 and #3 (Figure 3.11 left) provided a reasonable description of the moderately open state, while the remaining three trajectories suffered from an artificially over-stabilized open state of the clamshell. In most of the following structural considerations, we will concentrate on the trajectories #1 and #3.

GluK2-*cis* Complex

For GluK2-*cis* complex, the protein structure collapses in the simulation #3 after 100 ns (Figure 3.11 right), with a large Y488-G688 distance above 2.5 nm. In the simulation #2, *cis-gluazo* dissociates after 260 ns, accompanied by the opening of the clamshell which is depicted by an increase of the Y488–G688 distance to more than 1.5 nm. The three remaining simulations #1, #4 and #5 show average Y488–G688 distance of 1.04 nm, 0.90 nm and 0.89 nm, respectively.

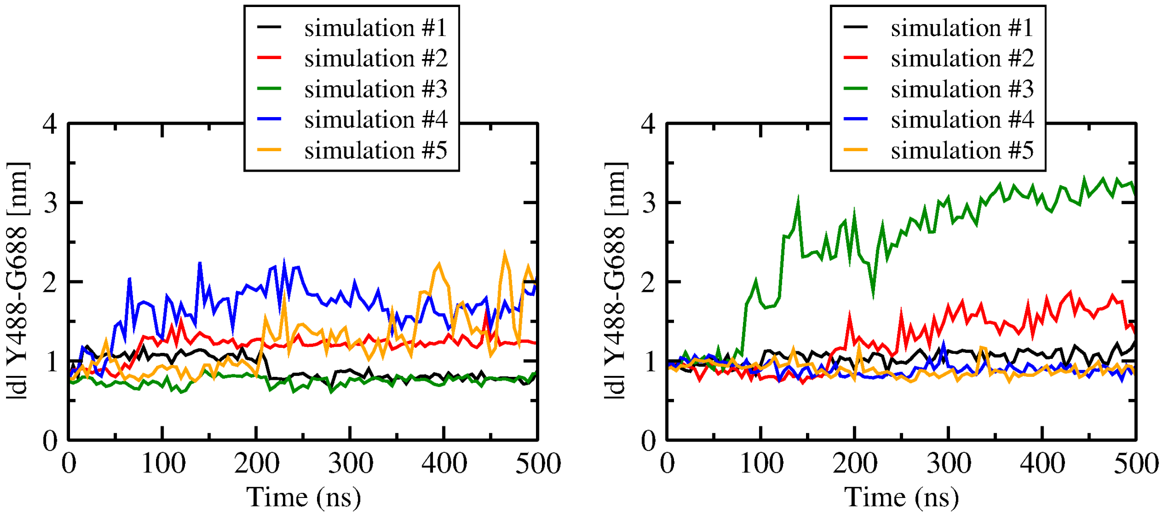


Figure 3.11: The LBD clamshell opening extent in the course of 500 ns MD simulations, depicted by the COM distance between Y488 and G688. Left: GluK2-*trans* complex; right: GluK2-*cis* complex. The black, red, green, blue, orange curves correspond to the simulations #1, #2, #3, #4, #5, respectively.

Disregarding the simulations with the wide-open clamshell, it is found that the *trans-gluazo* bound complex shows LBD clamshell conformation between closed and moderately closed states in terms of the clamshell opening extent. Relatively, the *cis-gluazo* bound LBD conformation is between moderately closed and open states.

Some Coincidences with Experiment

The ligand initiated iGluRs works through a steps procedure involving (i) the docking of the ligand and (ii) the locking of the LBD clamshell. The ligand docks to the LBD through the formation of HBonds, salt bridges or noncovalent interactions mainly with D1 domain; the interactions with the hinge region residues might also be involved. In the case of agonist, after docking, it interacts with D2 domain through noncovalent interactions. Thus, D2 domain moves towards D1 domain, as a result, the agonist is locked in the LBD clamshell. The strength of the ligand–D2 domain interactions formed in the locking step has significant effect on clamshell opening extent [4]. Previous study reported that D1 domain remains relatively fixed during the opening and closing motion of the clamshell due to the restriction from the D1–D1 interface in the intact receptor [8]. Notably, early spectroscopic work suggested that studies of the LBD clamshell motion should focus on the ligand–D2 interactions because they are involved in the locking step, which presumably leads to the clamshell closure [4, 21].

In order to get an insight into the details of LBD clamshell motion, the HBond patterns between *gluazo* and the protein of GluK2-*trans* and GluK2-*cis* complexes for all five simulations were evaluated, see Figure 3.12 and Figure 3.13, respectively. The protein residues involved are P516, A518, R523 on D1 domain, and A689, T690, E738 on D2 domain.

For GluK2-*trans* complex, in simulations #2, #4 and #5 where the clamshell is widely opened (from ca. 100 ns in #2; from ca. 50 ns in #4; from ca. 200 ns in #5; Figure 3.9 left and Figure 3.10 left), the ligand–D1 interactions are preserved even after the clamshell has opened (Figure 3.12 left). On the other hand, the ligand–D2 interactions are lost unpon the opening of the clamshell (Figure 3.12 right). This observation supports the viewpoint that the ligand mainly binds to the D1 domain in the open state of LBD clamshell [8].

The same observations are found for GluK2-*cis* complex (Figure 3.13). That is, in simulation #3, *gluazo* only interacts with D1 domain after the wide-opening of the clamshell from ca. 100 ns. In simulation #2, interactions of *gluazo* with D1 domain and D2 domain are lost upon its dissociation at ca. 100 ns. It is worth to mention that, trajectories #1, #4, #5 show average Y488–G688 COM distance of 1.04 nm, 0.90 nm, 0.89 nm, respectively, i.e. the LBD clamshell in trajectory #1 is more opened than in trajectories #4 and #5. Correspondingly, in trajectory #1 *gluazo* interacts with D2 domain through residue A689; while in trajectory #4 and #5, two residues A689 and T690 are involved. This is in good agreement with the view that the clamshell closure extent is sensitive to the strength of ligand–D2 domain interaction: the stronger is the ligand–D2 interaction, the larger is the LBD closure extent [4, 21].

3.3.4 Comparison of Ligand–Protein and Intradomain Interactions between GluK2-*trans* vs. GluK2-*cis* Complexes

Figure 3.14 and Figure 3.15 illustrate the ligand–protein interactions for GluK2-*trans* and GluK2-*cis* complexes, respectively. For comparison, five independent simulations of 500 ns each were performed for the GluK2-*glutamate* complex (PDB ID 1S50), and the corresponding ligand–protein interactions are shown in Figure 3.16. Note that, the 2D structure-diagrams of ligand–protein complexes are generated using the Poseview algrithm [173–175]. Five types of ligand–protein interactions are considered by this algrithm, i.e. hydrogen bonds, π-π stacking, π-cation interactions, metal interactions and hydrophobic contact. The details about the calculating criteria for each interaction type could be referred to in literatures [173–175].

In GluK2-*glutamate* complex, the agonist *glutamate* tightly bind to LBD domain through residues Y488, P516, A518, R523 of D1 and residues A689, T690, E738 of D2, see Figure 3.16. The benzene ring of Y488 forms π-cation interaction with the amine group of *glutamate*; the carbonyl group of P516 forms HBond with the amine group of *glutamate*; the backbone NH group of A518 froms HBond with the α-carboxyl group of *glutamate*; the guanidinium group of A523 forms salt bridge with the α-carboxyl group of *glutamate*. The backbone NH group of A689 forms HBond with the α-carboxyl group of *glutamate*; the backbone NH group and the side chain hydroxyl group of T690 forms HBond with the side chain carboxyl group (γ-carboxyl group) of *glutamate*; the γ-carboxyl group of E738 forms salt bridge with the amine group of *glutamate*.

In GluK2-*trans* and GluK2-*cis* complexes, the interactions between the glutamate moiety of *gluazo* and residues P516, A518, R523 on D1 domain are conserved. These interactions are unaffected by the configurations of *gluazo*, implying that they might be the anchor interactions maintaining the GluK2-*gluazo* complex. Previous study reported that R523 is conserved in all of the glutamate receptors and it serves as the primary anchor for the α-carboxyl group of ligands bound to GluA2 [17].

In terms of the ligand–D2 interactions, the two complexes show some differences. In GluK2-*trans* complex, the interactions between the glutamate moiety of *gluazo* and residues A689, T690, E738 are the same as in GluK2-*glutamate* complex. Relatively, in GluK2-*cis* complex, the HBond acceptor for the backbone NH group of A689 is the γ-carboxyl group of *cis-gluazo* (Hereafter the HBond is referred as (A689)NH$\cdots\gamma$-carboxyl(*cis-gluazo*)). Simulation #1 of GluK2-*trans* samples the HBonds (A689)NH$\cdots\alpha$-carboxyl(*trans-gluazo*) and (A689)NH$\cdots\gamma$-carboxyl(*trans-gluazo*) nearly equally, while mainly the (A689)NH$\cdots\alpha$-carboxyl(*trans-gluazo*) HBond occurs in simulation #3 (see Figure 3.12). As for GluK2-*cis*, all the three simulations #1, #4, #5 are dominated by the (A689)NH$\cdots\gamma$-carboxyl(*cis-gluazo*) HBond (see Figure 3.13). The importance of A689 for ligand binding affinity was proved through mutation study [176]. Therefore, it is possible that the difference of the HBond of A689 with *trans-* and *cis-gluazo* leads to distinct binding affinity of the two ligands, which eventually results in different LBD clamshell closing extent. Take this senario: The α-carboxyl group and the amine group of the glutamate moiety of *trans-* and *cis*-gluazo have the same interactions with D1 domain through P516, A518, and R523. Thus, the observed interaction of the γ-carboxyl of *cis-gluazo* with A689, indicates that helix F carrying A689 has to have shifted downwards notably. This change of the ligand-A689 interaction pattern partly explains the limited clamshell closure in GluK2-*cis* complex.

Moreover, the adoption of (A689)NH$\cdots\gamma$-carboxyl(*cis-gluazo*) HBond in GluK2-*cis* complex leads to the loss of the HBond between the side chain hydroxyl group of T690 and the side chain carboxyl group of *cis-gluazo*. This further weakens the ligand–D2 interaction.

Additionally, the HBond between E738 side chain and the amine group of *gluazo* (glutamate), which occurs in GluK2-*trans* complex (GluK2-*glutamate* complex), is almost lost in GluK2-*cis* complex. The im-

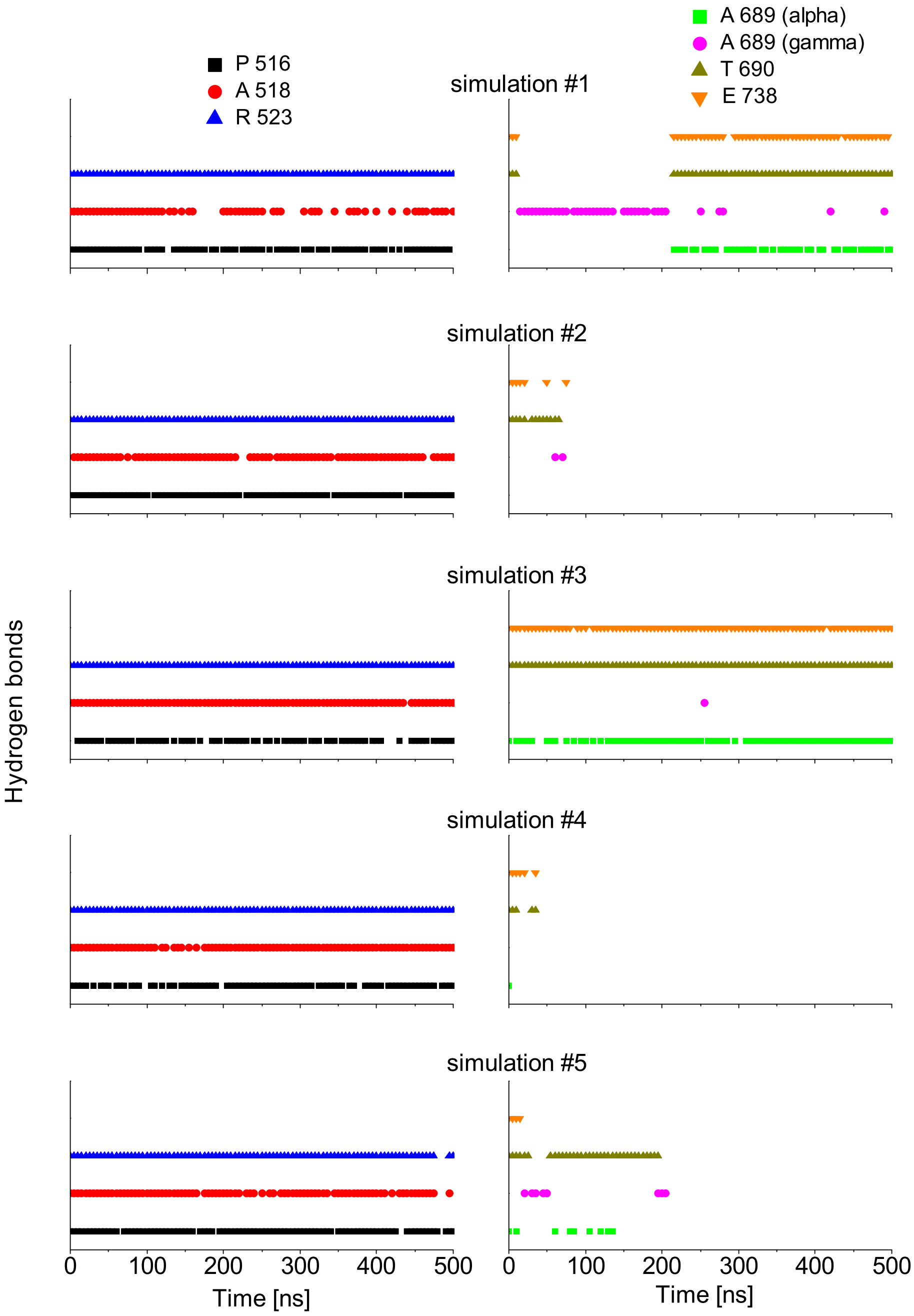


Figure 3.12: Hydrogen bond patterns in GluK2-*trans* complex in the course of 500 ns MD simulations. P516: HBonds between the α-amino acid group of *gluazo* and the backbone carbonyl group of P516; A518: HBonds between the α-carboxyl group of *gluazo* and the backbone NH group of A518; R523: HBonds between the α-carboxyl group of *gluazo* and the guanidinium group of R523; A689 (alpha): HBonds between the α-carboxyl group of *gluazo* and the backbone NH group of A689; A689 (gamma): HBonds between the γ-carboxyl group of *gluazo* and the backbone NH group of A689; T690: HBonds between the γ-carboxyl group of *gluazo* and the backbone NH group or the side chain OH group of T690; E738: HBonds between the α-amino acid group of *gluazo* and the side chain carboxyl group of E738. Data from simulations #1 through #5 from top to bottom.

portance of the corresponding HBond in GluA4 was reported in an earlier experiment, in which ^{3}H-AMPA can not bind to the receptor after E706Q mutation (E706 corresponds to E783 of GluK2) [177]. And later, this HBond was detected to be involved in LBD clamshell closure upon ligand binding [4]. Moreover, the previous vibrational spectroscopy study of GluA2 shows that the interactions of the α-amine group of the ligand with the protein plays a large role in controlling the extent of LBD cleft closure. That is, the stronger of this interaction, the larger of the cleft closure extent [178]. It is inferred that, the disapperance of the (*glu-*

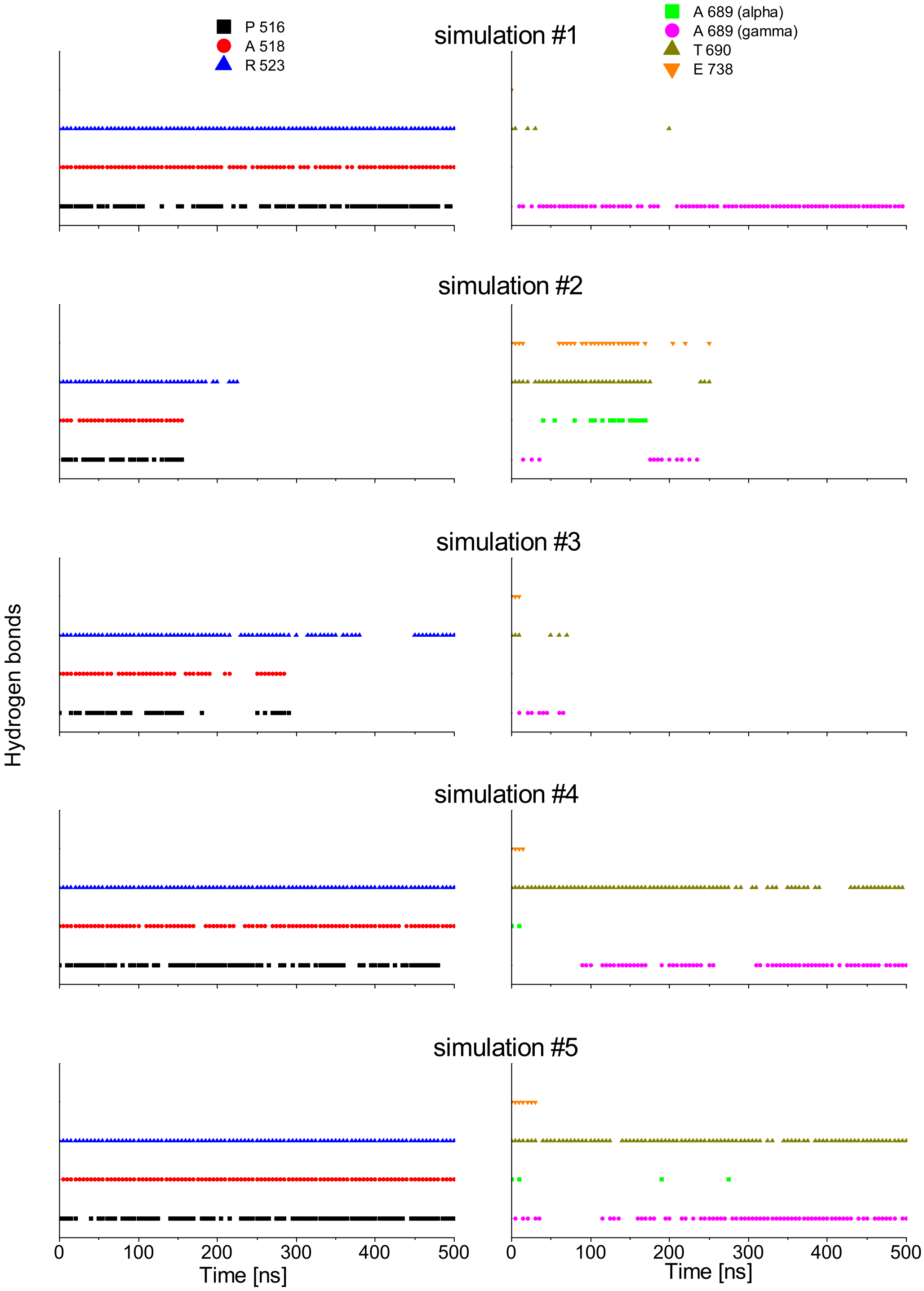


Figure 3.13: Hydrogen bond patterns in GluK2-*cis* complex in the course of 500 ns MD simulations. See the caption of Figure 3.11 for the legend.

azo)$NH_3^+ \cdots {}^-OOC$(E738) salt bridge is one of the reasons for the wider clamshell opening in the GluK2-*cis* complex.

In order to find out the reason for the weakened ligand–D2 interaction of *cis-gluazo* with respect to *trans-gluazo*, focus need be payed on the azobenzene moiety, because it is the only different part between the two ligands. The azobenzene moiety of *gluazo* forms hydrophobic contacts with the protein. This interaction is affected by *gluazo* configuration. In GluK2-*trans* complex, residues I438, Y488 on D1 domain are involved. In GluK2-*cis* complex, residues I438, G486, Y488 on D1 domain and S720 on D2 domain are involved. Note that, in GluK2-*trans* complex, Y488 has hydrophobic interaction not only with the azobenzene moiety, but also

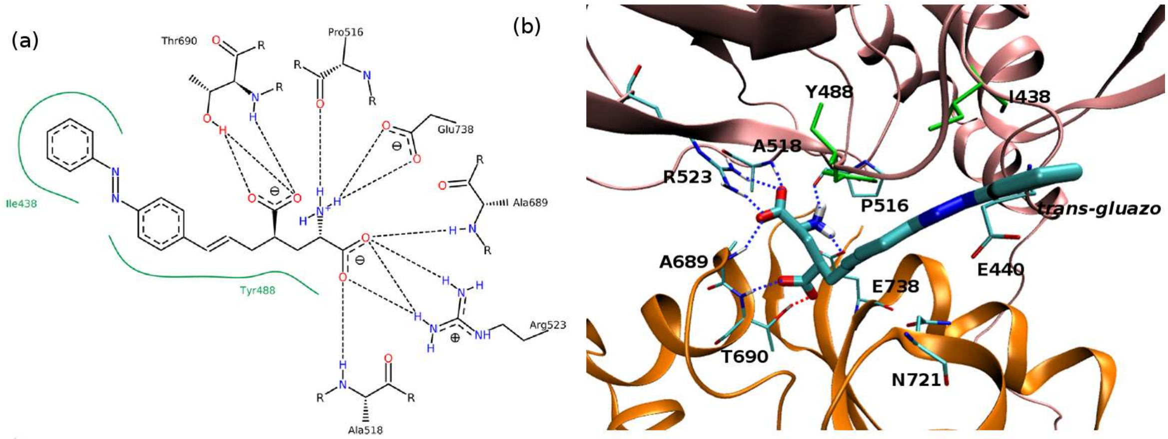


Figure 3.14: Ligand–protein interactions in the GluK2-*trans* complex. (a) 2D diagram of the ligand–protein interactions. Hydrogen bonds are represented by black dashed lines, hydrophobic contacts by green contour lines. (b) 3D diagram of the ligand–protein interactions. The protein is represented using cartoon mode, and D1 domain is in pink color, D2 domain in orange color. The ligand and interacting residues are shown in stick mode colored by atom. Residues engaged in hydrophobic interactions with the ligand are emphasized in green color. The Hydrogen bonds are depicted by blue and red dotted line.

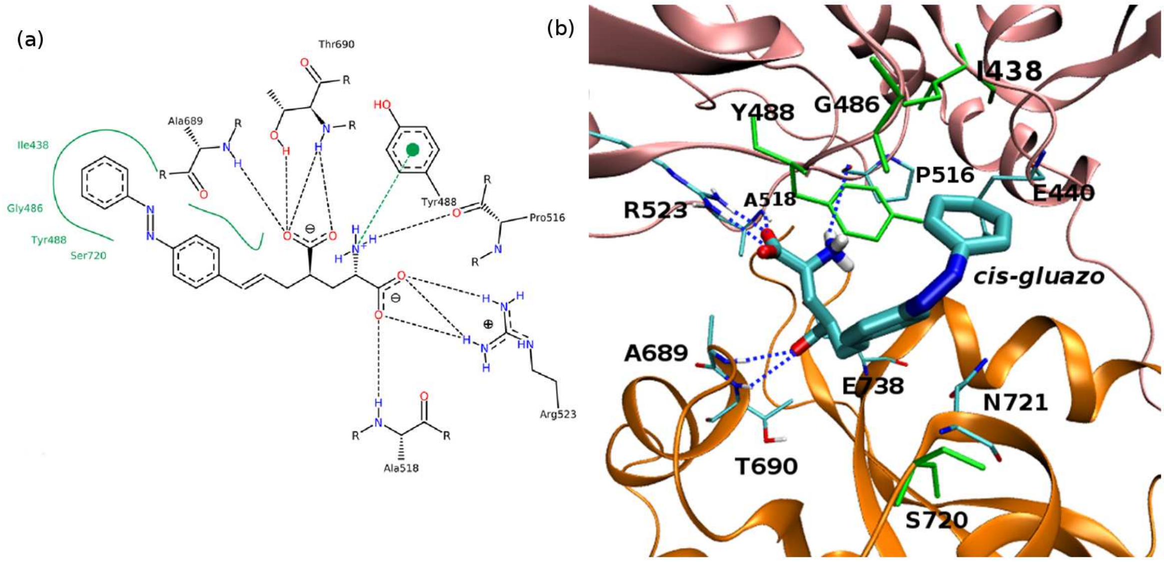


Figure 3.15: Ligand–protein interactions in the GluK2-*cis* complex. (a) 2D diagram of the ligand–protein interactions. See the caption of Figure 3.14 for legend, additionally, the green dashed line in the current figure represents the π-cation interaction. Note that the PoseView program cannot distinguish different configurations of *gluazo*, thus, *cis-gluazo* is displayed in the *trans* configuration. (b) 3D diagram of the ligand–protein interactions. See the caption of Figure 3.14 for legend.

with the backbone carbon chain of the *glutamate* moiety of *trans-gluazo*. Relatively, in GluK2-*cis* complex, all the four involved protein residues contacts only with the azobenzene moiety of *cis-gluazo*. This leads to the following ligand–D1 interaction senario (see Figure 3.17): *trans-gluazo* is bound to D1 domain through three anchors–(i) HBonds and salt bridge with protein through α-carboxyl group and amine group of the *glutamate* moiety, (ii) hydrophobic contacts with protein through the backbone carbon chain of the *glutamate* moiety, (iii) hdrophobic contacts with protein through the azobenzene moiety. On the other hand, *cis-gluazo* is bound to D1 domain through two anchors–(i) HBonds and salt bridge with protein through α-carboxyl group and amine group of the *glutamate* moiety, (ii) hdrophobic contacts with protein through the azobenzene moiety. That is, *cis-gluazo* is bound to D1 domain through two end anchor points, and the middle part (the backbone carbon chain of *glutamate* moiety) is suspended. As a result, *cis-gluazo* is more curved in the middle part than *trans-gluazo*. This further leads to the weakening of ligand–D2 interactions in GluK2-*cis* complex as

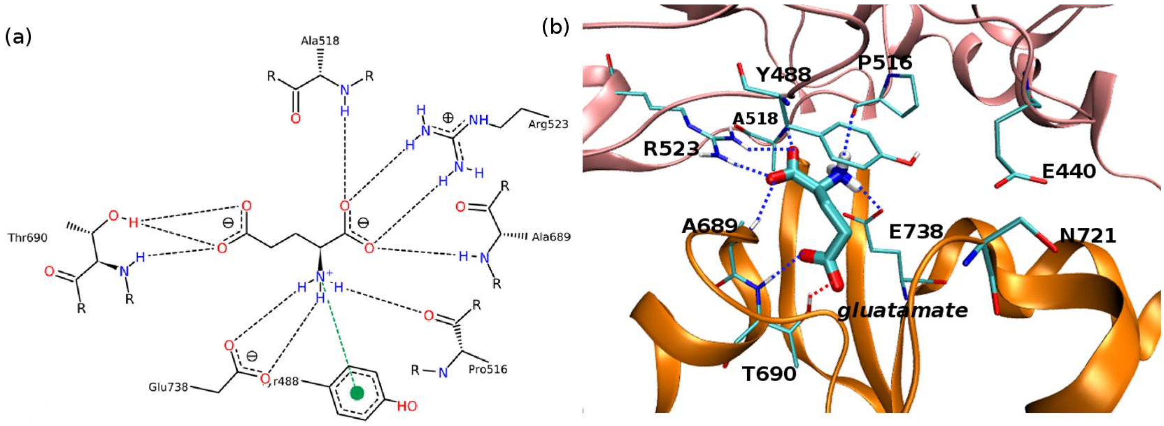


Figure 3.16: Ligand–protein interactions in the GluK2-*glutamate* complex. (a) 2D diagram of the ligand–protein interactions. See the caption of Figure 3.14 for legend. (b) 3D diagram of the ligand–protein interactions. See the caption of Figure 3.14 for legend.

discussed above.

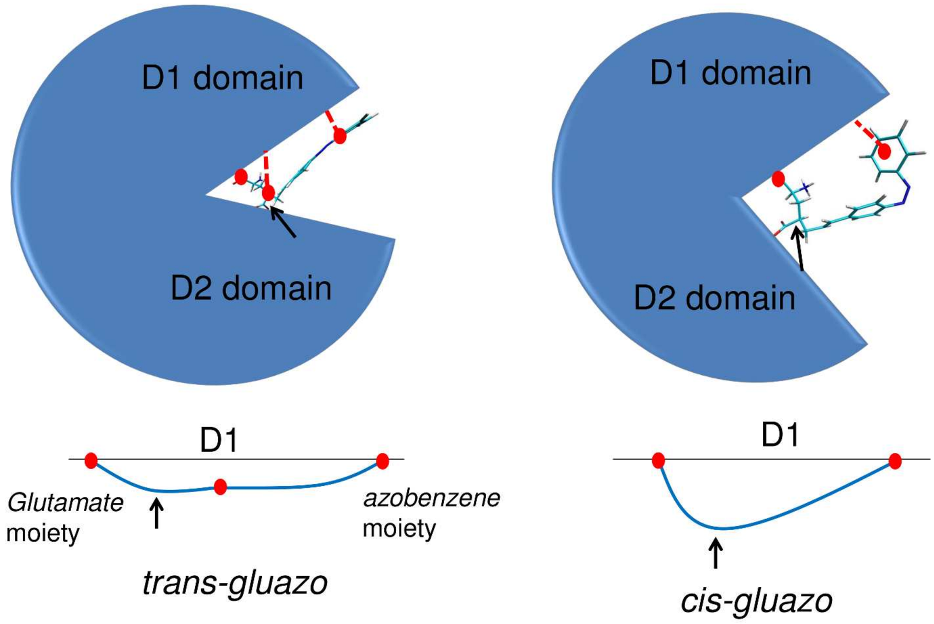


Figure 3.17: Schematic illustration of ligand–D1 domain interactions for *trans*- and *cis*-gluazo. The red cycles and dashed lines represent the anchor interactions. The curved middle part of *gluazo* is point out by arrows.

Except for the ligand–domain interactions, the D1–D2 intradomain interaction is also crucial for determining the opening/closing extent of the LBD clamshell. The HBond between E440 of D1 domain and N721 of D2 domain is interpreted as a proxy for the D1–D2 intradomain interaction. The occurrence frequency of the E440–N721 HBond in GluK2-*trans* and GluK2-*cis* complexes is listed in Table 3.1, in the form of percentages of the simulation time. The GluK2-*glutamate* complex is included as a reference, and ca. 80 % of the simulation sampled the E440–N721 hydrogen bond. The simulations #1 and #3 of GluK2-*trans* complex, in which the clamshell closure extent is intermediate between the closed state and the open state, show slightly lower occurrences of the E440–N721 HBond than the GluK2-*glutamate* complex. The simulations #1, #4 and #5 of GluK2-*cis* show rather low occurrence of the E440–N721 HBond. Therefore, the intradomain interactions are rather hindered by *cis-gluazo*, when compare to *trans-gluazo*. This obstruction destabilizes the closed conformation of the LBD clamshell [176].

Table 3.1: The occurrence frequency of the E440–N721 hydrogen bond in the five independent simulations of 500 ns for each of the complexes GluK2-*trans*, GluK2-*cis* and GluK2-*glutamate*, expressed as the percentage of the total simulation time.

	GluK2-*trans*	GluK2-*cis*	GluK2-*glutamate*
1	76%	15%	82%
2	40%	21%	90%
3	65%	4%	72%
4	5%	47%	83%
5	19%	12%	74%

3.3.5 Reaction of the Linker Segments in the Dimer

The current MD simulations were performed on the LBD monomer. In order to extrapolate the results to the ion channel reaction, virtual MD trajectories involving a dimer of LBD domain were constructed. In this procedure, one monomer from the X-ray structure (PDB ID 4H8I) was used as a template, and the MD trajectory was aligned to the D1 domain of the template monomer. This yielded a dimer composed of one dynamic monomer and one static monomer (see Figure 3.18). This made it possible to describe, albeit crudely, the orientation of TMD helices. In the LBD crystal structure as well as in our simulations however, the TMD (residues 545 through 666) is replaced by a linker segment GT (Glycine, Threonine) located at the N-terminus of helix E. The GT linker connects the residues K544 and P667, whereas K544 is linked to the transmembrane helix M1, and P667 is linked to helix M3 in the full-length receptor. Upon the binding of an agonist, the (GT)–(GT) COM distance increases, and this results in a reorientation of TMD helices and eventually the opening of the ion channel. Therefore, the COM distance of the linker segements in a dimer is regarded as the proxy for describing the opening extent of the ion channel [179].

In the current study, the COM distance of the four-residue segment K544-G-T-P667 (referred to linker–linker distance hereafter) is used to evaluate the ion channel opening. The correlation between the Y488-G488 distance and the linker–linker distance, i.e. the correlation between the LBD clamshell closure extent and the ion channel opening extent, was quantified as shown in Figure 3.18. Unlike the structural analyses, which focus on simulations sampling closed or moderately closed states, this correlation analysis is based on simulations in which the clamshell opens wide. The correlation coefficient is -0.86 for simulation #5 with largely open clamshell of GluK2-*trans* complex, and it is -0.89 for simulation #2 of GluK2-*cis* complex Such a high correlation demonstrates how the closing of the LBD clamshell translates into force that controls the structure of the ion channel.

Note that the decrease of the linker–linker COM distance in response to the opening of the clamshell is overestimated in the simulations of LBD monomer. There are two reasons for this. As discussed above, the lack of sterical clashes with the D2 domain on an interacting monomer, the clamshell is free to open too wide, leading to an overestimated shortening of the linker–linker COM distance. In addition, the interface between individual monomers in a dimer exhibits a certain flexibility as reported in ref. [180]. In the constructed trajectories, the presence of the static monomer restricts the monomer–monomer interface flexibility. This means that the current linker–linker COM distance is predicted on the basis of a relatively rigid dimer, which overestimates the linker response. Aiming merely at the qualitative comparison of *gluazo* isomers, the current dimer setup is considered reasonable.

3.3.6 Free Energy Study of LBD Clamshell Motion

The LBD clamshell conformational changes in GluK2-*trans* and GluK2-*cis* complexes were investigated by means of umbrella sampling simulation, to alleviate the issue of insufficient sampling in free MD simulations [161, 162]. The structure resulting from the extended equilibration procedure was used as a starting point of the simulations for each of the two complexes. These two starting structures possess a similar protein–ligand interaction pattern. The reaction coordinate was defined as the COM distance between Y488 and G688. An interval from 0.5 nm to 1.4 nm employing 19 or 21 windows, for GluK2-*trans* and GluK2-*cis* respectively,

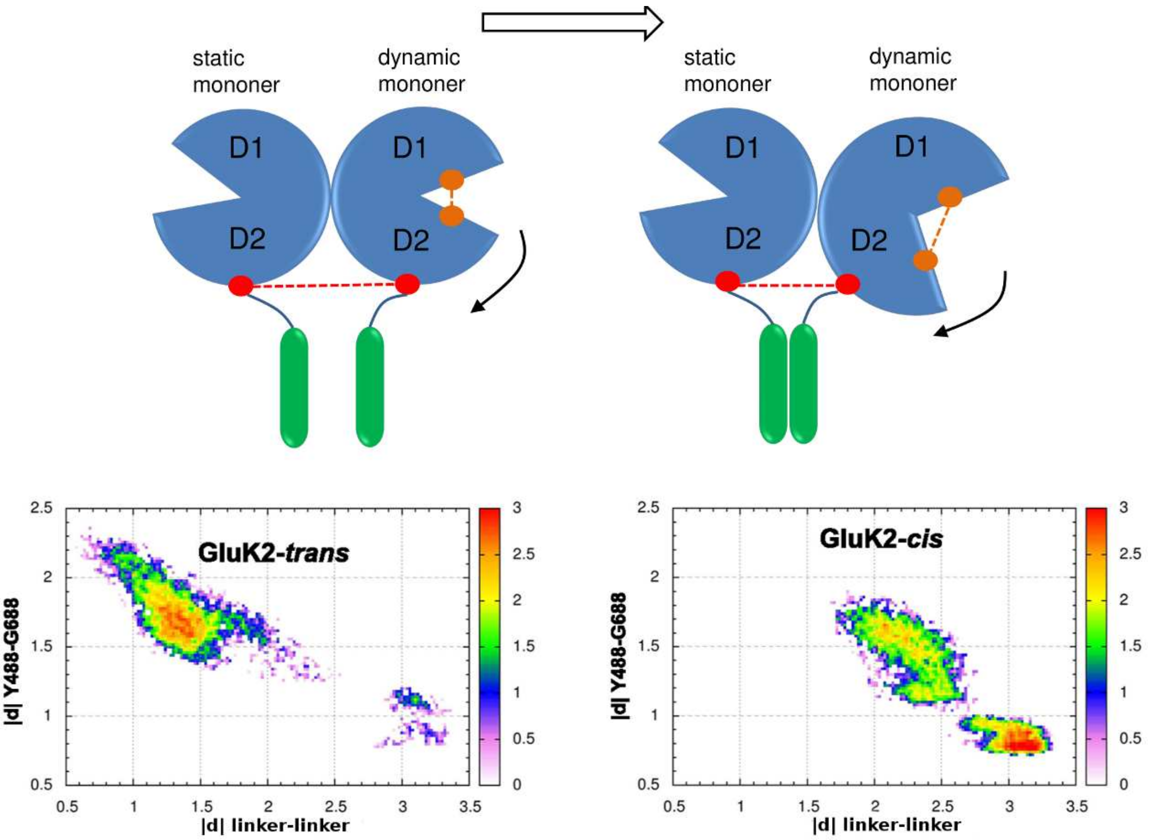


Figure 3.18: Upper: Schematic illustration of the impact of the LBD clamshell conformational changes on ion channel. The brown spheres and dashed lines depict the LBD clamshell conformational changes. The red spheres and dashed lines depict the linker–linker COM distance changes in the dimer. The green cylinders represent the ion channel in the full-length receptor. The line arrows show the moving direction of D2 domain. Lower: correlation diagram between the linker–linker COM distance (nm) and the Y488–G688 COM distance (nm) for GluK2-*trans* and GluK2-*cis* complexes. Probability density is coded by color in logarithmic scale, arbitrary unit. Note that the data were obtained from one MD simulation of each complex, and thus, the data cannot reproduce the equilibrium probabilities due to insufficient sampling. The purpose of the diagram is merely to illustrate the good linear correlation between the two quantities on a broad interval of clamshell opening extent.

was spanned to fully describe the clamshell motion. The initial structure for US simulations in every window was obtained through a pulling simulation, in which the Y488–G688 COM distance increases gradually. The pulling rate, the force constant and simulation time in each window is the same as for studying the ligand binding modes.

The well overlapped histograms of each window, and the free energies with gradually increasing simulation time of two complexes are shown in Figure 3.19 and Figure 3.20, respectively. The GluK2-*cis* complex has a smaller free energy change after 200 ns comparing to the GluK2-*trans* complex, indicating a better convergence of the former complex.

The PMF of clamshell opening of GluK2-*trans* complex shows a minimum at 0.76 nm (Figure 3.20 left), corresponding to the previously obtained equilibrated structure. The Y488–G688 COM distance of 0.76 nm is shorter than that in the partially open state (0.82 nm, PDB ID 2XXT, bound with kainate), but it is still longer than the value in closed state (0.63 nm, PDB ID 1S50, bound with glutamate). Besides, the PMF profile at 300 nm exhibits a small barrier of ca. 1.5 kcal·mol^{-1} to reach the open state. This implies *trans-gluazo* to be a partial agonist, which is in agreement with the experimental study by Reiter et al. [20] Correspondingly, the opening of the clamshell was observed in three out of five standard MD simulations #2, #4 and #5 (Figure 3.11 left). Especially, simulation #2 shows the structure corresponding to the open state minimum at 1.20 nm. The energy profile is flat between 1.20 nm and 1.35 nm, thus, there is almost no restriction to open the clamshell wide. Note that the wide open state is likely over-stabilized due to our simulation setup featuring a monomer with no interfacial restriction, which would be however present in a LBD oligomer.

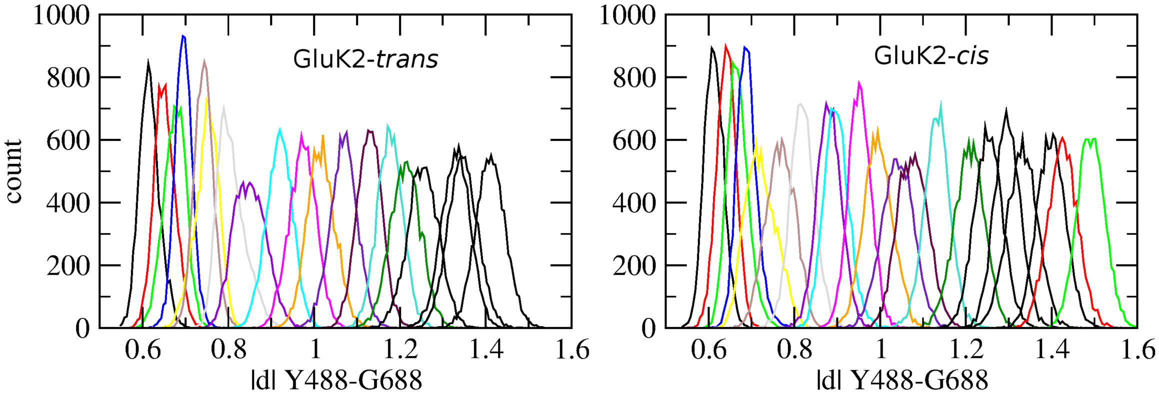


Figure 3.19: Histograms from the umbrella sampling simulations of LBD clamshell motion.

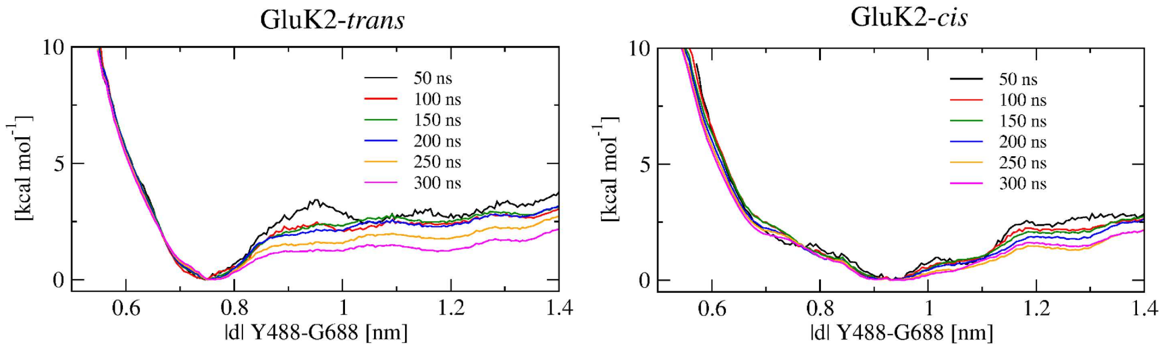


Figure 3.20: Free energy profiles of the LBD clamshell motion of GluK2-*trans* and GluK2-*cis* complexes. Each of the different curves was obtained from a series of umbrella sampling simulations with the time scale specified in the legend. Each free energy curve is shifted such that its value in the respective global minimum vanishes.

In GluK2-*cis* complex, the PMF features a free energy basin with a minimum around 0.93 nm, again corresponding to the previously obtained equilibrated structure. This minimum is separated from the open state at 1.28 nm by a barrier of 1.6 kcal·mol^{-1} (Figure 3.20 right). In agreement with this low barrier, opening of the clamshell was also observed in one of the standard MD simulations #2 (Figure 3.11 right). Based on the simulations, the free energy barrier to open the clamshell is comparable for GluK2-*trans* (1.5 kcal·mol^{-1}) and GluK2-*cis* (1.6 kcal·mol^{-1}).

It is worth to mention that, it is necessary to break three HBonds and one salt bridge in order to open the clamshell, judging from the structure of the starting molecular model. Taking this into accout, the obtained free energy barrier of ca. 1.5 kcal·mol^{-1} appears rather small. However, rather than a simple breaking of the interactions between the ligand and the D2 domain, the opening of the clamshell is a more complicated process that involves the formation of alternative interactions as well. For illustration, the ligand–D2 interactions changes during the pulling simulation should be taken into account. The HBond between the backbone NH group of A689 and *gluazo* switches from α-carboxyl to γ-carboxyl fo *gluazo* at 19–24 ns, and it remains at γ-carboxyl until this HBond breaks at 40 ns, simultaneously with the HBond between T690-NH and γ-carboxyl of *gluazo*. The salt bridge between the amine group of *gluazo* and the γ-carboxyl of E738 breaks and reforms at 21–24 ns, and it breaks completely later. Finally, the HBond between the hydroxyl group of T690 and the γ-carboxyl of *gluazo* breaks at 41 ns, after which the clamshell approches the open state.

3.4 Discussion and Conclusion

The mechanism of response of GluK2 receptor to photoswitchable *gluazo* is investigated in the present study. The *cis-gluazo* bound complex was obtained with the application of a forced switching protocol. Technically, the complete photoreaction of *gluazo* could be simulated with an appropriate QM/MM model, however, this is unnecessary in this study. Because our gist is focused on the protein response to different ligand configurations, not on the ligand isomerization itself.

The protein exhibits no ultrafast structural transition, rather the local structure in the vicinity of the ligand accommodates to the the isomerization product. A complete protein response follows on a hundreds of ns time scale.

The LBD clamshell closure extent is quantified as

$$\text{LBD closure extent} = \frac{|d|_{\text{open}} - |d|}{|d|_{\text{open}} - |d|_{\text{closed}}} \tag{3.1}$$

where $|d|$ is the current Y488–G688 COM distance, $|d|_{\text{open}} = 1.15$ nm for the antagonist bound GluK1 (PDB ID 2QS1), and $|d|_{\text{closed}} = 0.63$ nm for the agonist-bound state (with glutamate, PDB ID 1S50). The average LBD closure extent is 62 % in GluK2-*trans* complex (calculated from simulations #1 and #2), 40 % in GluK2-*cis* complex (calculated from simulations #2, #4 and #5). And it is 50 % in a partial agonist bound complex (PDB ID 2XXT). This means that, *trans-gluazo* leads to a slightly more closed LBD clamshell of GluK2 than partial agonist, *cis-gluazo* to a slightly more open LBD clamshell. Mayer et al. [19] reported that the LBD clamshell closure extent is in positive correlation with the ligand efficacy. According to this, *cis-gluazo* has a lower efficancy than *trans-gluazo*, which is in agreement with the reported lower currents intensity evoked by *cis-gluazo* relative to *trans-gluazo* in GluK2 [30].

The fundamental reasons for distinct efficacy of the two *gluazo* isomers is the difference in the ligand–protein and intradomain interactions. The *trans-gluazo* binds to D1 domain through three anchor points, which distribute at the two ends and the middle part of the ligand; relatively, *cis-gluazo* binds to D1 only through anchor points at the two ends of the ligand. As a result, *cis-gluazo* is more curved in the middle part than its *trans* counterpart. This further leads to the weakening of ligand–D2 interactions in GluK2-*cis* complex: the (*gluazo*)$\text{NH}_3^+ \cdots {}^-$OOC(E738) salt bridge and the (T690)OH$\cdots\gamma$-$^-$OOC(*gluazo*) HBond are lost comparing to the GluK2-*trans* complex. Besides, *cis-gluazo* has a stronger hinderance to the interaction between E440 (on D1 domain) and N721 (on D2 domain). This further destabilizes the closed conformation of the LBD clamshell, reducing the efficacy of *cis-gluazo*.

The current MD simulations study contributes to the understanding of protein response to different *gluazo* isomers. The *trans* isomer of *gluazo* is a partial agonist, while the *cis* isomer has a relatively lower efficacy. The interaction between *gluazo* and D2 domain plays a key role in determining LBD clamshell closure extent. The *gluazo*-D2 interaction is strongly influenced by the *gluazo*-D1 interaction.

Chapter 4

Molecular Dynamics Study of Channelrhodopsin-2 C128T Mutant

Previous studies of ChR2 mutants at amino acid position 128, C128X (X = threonine, alanine, serine), show 200- to 1000-fold extended lifetime of the conducting state compared to ChR2 wild type. As a result, cells expressing ChR2-C128X depolarize readily at quite low light intensities [82]. Because of the above properties, ChR2-C128X mutants are very useful for applications in neuroscience. For instance, Schoenenberger et al. used these mutants for characterization of pyramidal cells in hippocampal slice cultures, to reduce the photon exposure needed for the generation of spike trains [181].

Especially, in ChR2-C128T mutant, the open state(P_1^{520}) is accumulated for several seconds upon illumination. By contrast, in ChR2 wild type the open state exists in a much shorter time range [82, 106]. The mutation of C128 by T128 in ChR2 increases the light sensitivity of the protein on one hand. This reduces the photocurrent, slows down the photocycle kinetics of the protein on the other hand. Therefore, ChR2-C128T has more complex photocycle with respect to ChR2 wild type. The slow kinetics of ChR2-C128T allows the investigation of the channel opening mechanism at the molecular level.

In order to fully understand the complex function of ChR2-C128T, more attention should be devoted to its function related structures: (i) The active site, mainly comprises the retinal, the two potential counterions E123 and D253, and the water molecules in the vicinity. The hydrogen bonding network within this region might affect the protein absorption spectrum. The retinal isomerization pathway and the primary proton acceptor of $RSBH^+$ may also have relation with the active site. (ii) The DT gate, formed by interactions connecting helices 3 and 4, mainly referring to the hydrogen bond between T128 and D156. It has a great importance in controlling the channel-closing kinetics. (iii) The cation conduction pathway, which refers to the pore formed by helices 1-3 and 7. Along this pathway, S63 and N258 together with E90 form the so-called central gate. It is proposed to be involved in keeping the channel internally closed before light exposure through hydrogen bonding network.

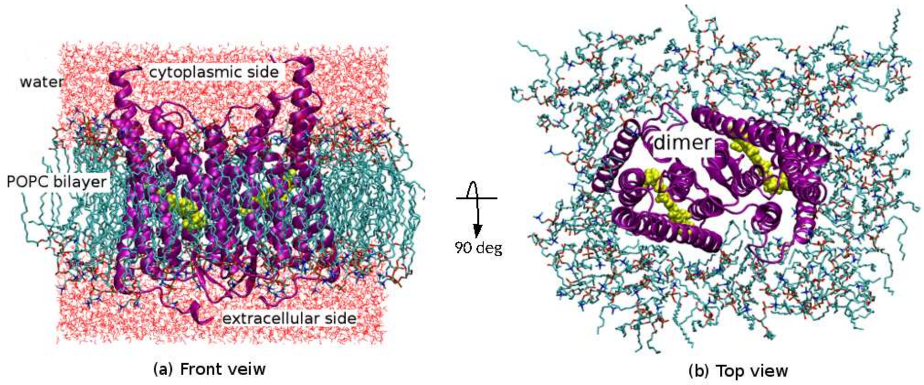
Similar to bacteriorhodopsin, ChRs undergo a photocycle initiated by the light-induced isomerization of the retinal. Up to current, the molecular details involved in the photocycle are still under debate, hence several open questions are arised. For instance, which residue serves as the primary acceptor of $RSBH^+$ proton, E123 or D253, or might be both [78, 81–83]? What triggers the channel opening, and what is the role of the central gate in this step? How the protein helices rearrange to form an opened channel pore [78]? How is the Schiff base reprotonated, and what is the proton donor and the proton transfer pathway? To answer all these questions, firstly a thorough understanding of the dark state protein structure is essential, especially of the three structural properties mentioned above. Indeed, such an knowledge would be helpful for the development of engineered ChRs, in order to expand their biological applications.

In this chapter, MD simulations based on a classical force field were performed to get the structural information of ChR-C128T.

4.1 Rhodopsin Models

A detailed ChR2 structure is urgently needed, in order to elucidate its functional mechanism at the atomic level as well as for further engineering and function. Watanabe et al. built ChR2 homology models based on the *Anabaena* sensory rhodopsin crystal structure and the C1C2 crystal structure, which identify remarkable structural information, including RSBH$^+$ orientation, water distribution near RSBH$^+$, intrahelical hydrogen bond, DC gate structure, helices conformation [79, 80].

In the current study, the ChR2 homology model from Watanabe based on the C1C2 crystal structure as described in the previous study [80] was adopted. The C1C2 X-ray structure (PDB ID 3UG9) [66] bound with all-*trans* retinal is used as the template. In the first step, the homologous model of ChR2-WT was constructed. The resulting monomer (see Figure 1.6) was used as an initial structure to build a ChR2 dimer, embedded in the POPC (1-Palmitoyl–2-oleoylphosphatidylcholine) bilayer as lipid membrane, and water molecules as explicit bulk solvent (see Figure 4.1).

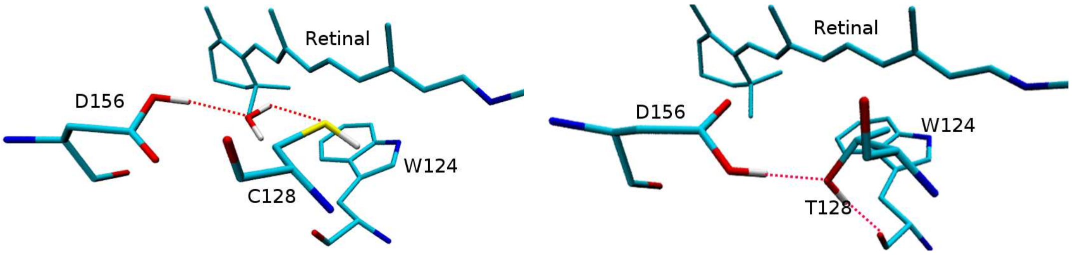


(a) Front veiw

(b) Top view

Figure 4.1: Model of ChR2: The protein dimer (purple) is inserted in a POPC lipid bilayer, which is surrounded by water molecules. The retinal chromophore (yellow) covalently links to K257 side chains. (a) Front view; (b)Top view.

The point mutation of C128T was simulated using the MMTSB toolset [182]. Previous MD simulations and vibrational frequency calculations proposed a ChR2-C128T mutant structure without any water molecule between T128 and D156, while the side chain of T128 forms an intrahelical hydrogen bond with the W124 backbone carbonyl oxygen atom [92]. To build the ChR2-C128T model, the water molecule between T128 and D156 was deleted. The side chains of T128 and D156 were adjusted to fulfill the above proposed ChR2-C128T mutant characteristics, i.e. the side chain of T128 forms a direct hydrogen bond to the side chain of D156. At the same time the side chain of T128 interacts with the W124 backbone carbonyl oxygen through an intrahelical hydrogen bond. The comparison of the DC/DT gate between ChR2-WT and ChR2-C128T is shown in Figure 4.2.



(a) ChR2-WT

(b) ChR2-C128T mutant

Figure 4.2: The DC and DT gate structures of (a) ChR2-WT; (b) ChR2-C128T mutant, respectively.

The starting active site structure was constructed according to previous theoretical and experimental findings. The counterions E123 and D253 are deprotonated as discussed in Chapter 1. The side chain of deprotonated E123 is adjusted to from a direct hydrogen bond with T127 (see Figure 4.2). Besides, the remaining titrable residues are assigned with standard protonation states, except E90 and D156 which are suggested to be protonated in the dark state [79, 83, 112].

As discussed in Chapter 1, the arrangement of water molecules in ChR2 active site is still controversial. In the present study, several water molecules near the RSBH$^+$ were identified by utilizing the DOWSER package, which locates protein internal cavities and assesses the hydrophilicity of these cavities through the calculation of the interaction energy of a water molecule with the surrounding atoms [183]. One water molecule was adjusted to form a direct hydrogen bond with the RSBH$^+$ based on previous study [80]. Note that, the side chain of E123 and the water molecule were manually rotated for each monomer, therefore, the two monomers are not guaranteed to have exactly the same starting active site structure (see Figure 4.3).

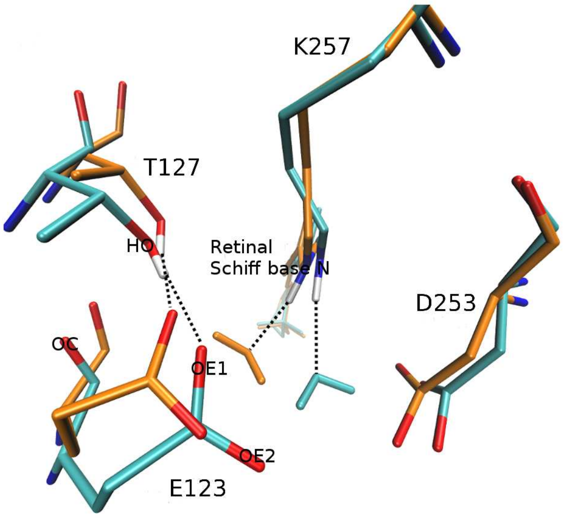


Figure 4.3: Initial structures for MM equilibration of the proposed ChR2-C128T model: hydrogen bonds (dotted lines) are set between E123 and T127 side chains, and between a water molecule and the RSBH$^+$. The two structures are shown in backbone colors of cyan and orange, respectively.

The complete model consists of 492 amino acids (246 amino acids per monomer), 107 POPC molecules, 6880 water molecules. The simulation box size is 8.22 nm × 6.74 nm × 7.45 nm.

4.2 Molecular Dynamics Simulation

4.2.1 MM Equilibration

The built models were subjected to MM equilibration. The first step was the minimization of the potential energy, which is achieved through the steepest descent method with 1000 steps. The maximum force tolerance was set as 100 Newton. Then the equilibration of the solvent was followed under NVT ensemble for 20 ps. The initial atomic velocities were generated on the basis of Maxwell–Boltzmann distribution [115] at 300 K, applying the Brendsen thermostat [119]. During this step the heavy atoms of the protein and lipid membrane were restrained to their initial positions by harmonic potentials with the force constant of 1000 kJ mol^{-1} nm^{-1}. This is an usual procedure to keep the protein intact during the equilibration of the solvent. Subsequently, the overall system was proceeded into equilibration (restrains on protein and lipid heavy atoms are removed) for 10 ns under NPT ensemble, at a pressure of 1 bar maintained using the Parrinello–Rahman barostat [184] method. A relaxation time constant of τ_p=5.0 ps and a compressibility of β=4.5$\times$10^{-5} for the x,y-direction and β=4.5$\times$10^{-5} for the z-direction were used. The temperature of 300 K was controlled by the Nosé–Hoover thermostat [117, 118] with τ_t=0.2 ps. The H-bond scheme for constraints as implemented in the LINCS algorithm [167] within the Gromacs package was employed. The time step was 2 fs.

MM force fields show severe shortcomings in the description of strongly hydrogen bonded networks. This has been demonstrated in particular for the BR active site. Using standard force fields, the water molecule binding to the RSBH$^+$ NH moiety is lost and leaves the active site, i.e. the pentagonal HBond network is broken. Alternatively, a salt bridge active site structure, RSBH$^+$–D85, is sampled [185–187]. A similar finding has been reported for ChR by Watanabe et al. [79, 80]. If the water molecule is lost during MM equilibration, it will not be restored in a limited simulation time scale. Therefore to avoid a bias due to the 10 ns MM equilibration on the following production MD starting structures, some restraints and constraint were applied to the system. More in detail, the positions of the oxygen atom of the water molecule, of the nitrogen and hydrogen atoms of the RSBH^+, were harmonically restrained with a force constant of 10000 $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$. Additionally, the hydrogen bond length between the oxygen atom of E123 side chain and the hydrogen atom of T127 side chain was constrained to a length of around 0.18 nm. These position restraints and the distance constraint were held during the whole equilibration procedure. In order to avoid accidental sampling bias, four independent equilibration simulations with different initial velocities were performed.

4.2.2 Production MD

The four structures from the NPT equilibrations were used as the starting points of the production MD simulations, time scale of 500 ns. All position restraints and the distance constraint were removed. All simulations were performed with the GROMACS package [169, 170]. The Charmm36 classical force field [188] was applied to describe the MM region, including parameters for the lipid membrane [189]. The TIP3P model [164] was used to describe water molecules.

4.3 Results and Discussion

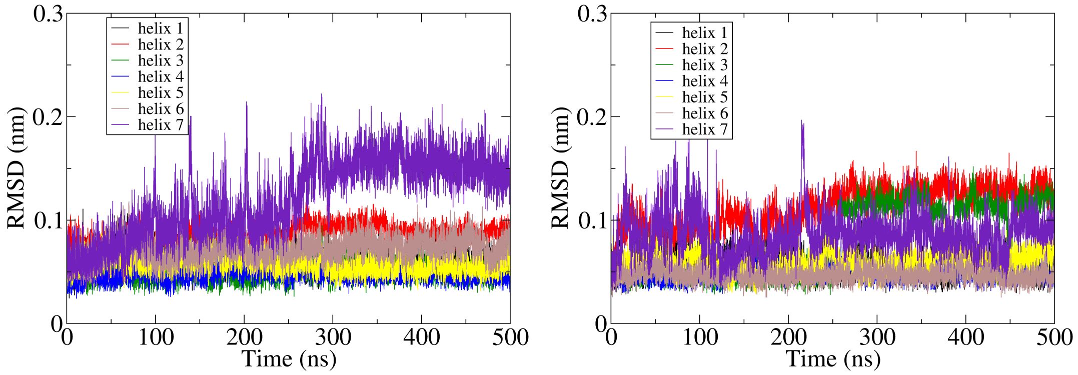
The following results concern all structural elements of ChR2-C128T mutant based on the MM simulations. In order to eliminate protein drift in the simulation box, the MD trajectories are aligned to the starting structures of the production simulations based on C_α atoms. The processed trajectories are then used for analysis.

4.3.1 Protein Conformation

The overall protein conformation change on the large time scale reflects the reliability of the current model. The root mean square deviation (RMSD) of the protein in the trajectory compared to a reference structure reflects the stability of the protein structure during the dynamics. In the current study the backbone RMSD of the seven helices with respect to the starting structure were computed, using both monomers for analysis, as shown in Figure 4.4 (The results of the other three trajectories are shown in appendix, Figure A.1). Helices 1 and 3–5 of monomer 1 show RMSD values of ca. 0.05 nm. Helices 2 and 6 show slightly larger values of ca. 0.08 nm. Helix 7 is more flexible than other helices, depicted by the pronounced width of the RMSD fluctuations, especially after 250 ns the RMSD rise up to nearly 0.2 nm. Helices 1 and 4–6 of monomer 2 show RMSD values of ca. 0.05 nm. Helix 2 shows slightly larger RMSD of ca. 0.1 nm. The RMSD of helix 3 rises from ca. 0.05 nm to ca. 0.1 nm at 250 ns. Helix 7 in monomer 2 also shows wider RMSD fluctuations than other helices. Although the helices show small structural fluctuations to various extent, the superimposition of the last snapshot from the 500 ns simulation with the starting structure shows almost no helices displacement. This means that the helices conformations of ChR2-C128T dark state are similar to the C1C2 crystal structure.

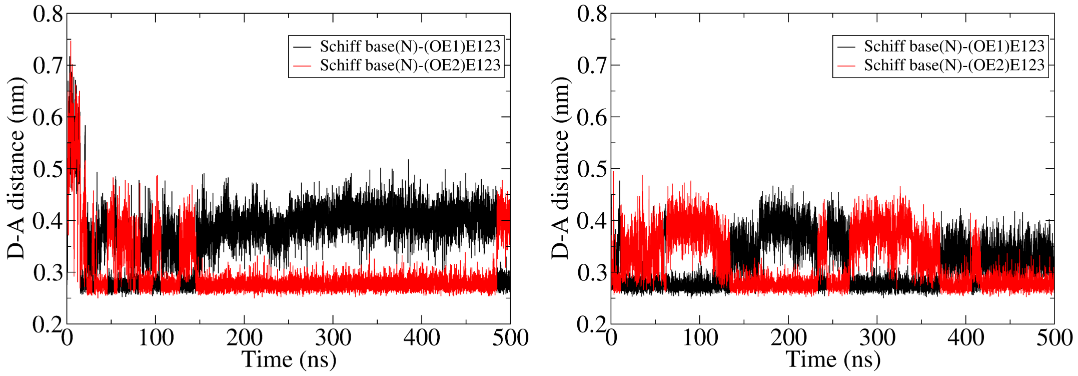
4.3.2 Active Site

The current MD simulations mainly sampled a direct salt bridge between the retinal Schiff base and the negatively charged E123, which is depicted by the Schiff base(N)-(OOC)E123 distance within 4 Å (see Figure 4.5 and appendix, Figure A.2) [190]. The water molecule that is in hydrogen-bonding contact to the Schiff base in the initial structure (see Figure 4.3) leaves the retinal binding pocket and does not rebind. This has been discussed before and is related to the deficiency of classical force fields in the description of hydrogen bonding interactions between charged residues and water molecules [80]. Note that in one trajectory in the first 320 ns (appendix, Figure A.2e) the E123 side chain flips downward to the extracellular side, which causes the breaking of the salt bridge. In the rest of the simulation time, E123 side chain flips upward and the salt bridge recovers. A water density, which depicts the possibility of water molecules occurrence, close to negatively charged E123 and D253 is observed (see Figure 4.6 and Figures A.3-A.5), cosistent with the study by Watanabe et al. for ChR2-WT [80]. This indicates that ChR2 might possess a similar water bridged active site structure as in BR.



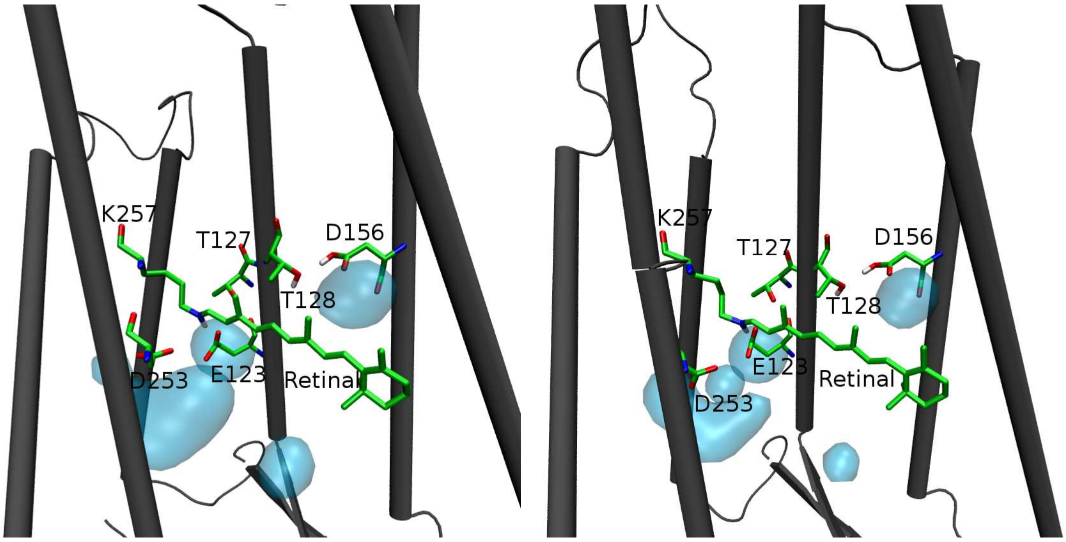
(a) monomer 1 **(b)** monomer 2

Figure 4.4: The backbone RMSD of seven helices with respect to the starting structures of two monomers. (a) monomer 1; (b) monomer 2.



(a) monomer 1 **(b)** monomer 2

Figure 4.5: The proton donor–acceptor distance between retinal Schiff base and E123 side chain. (a) monomer 1; (b) monomer 2. See the atomic labels in Figure 4.2



(a) monomer 1 **(b)** monomer 2

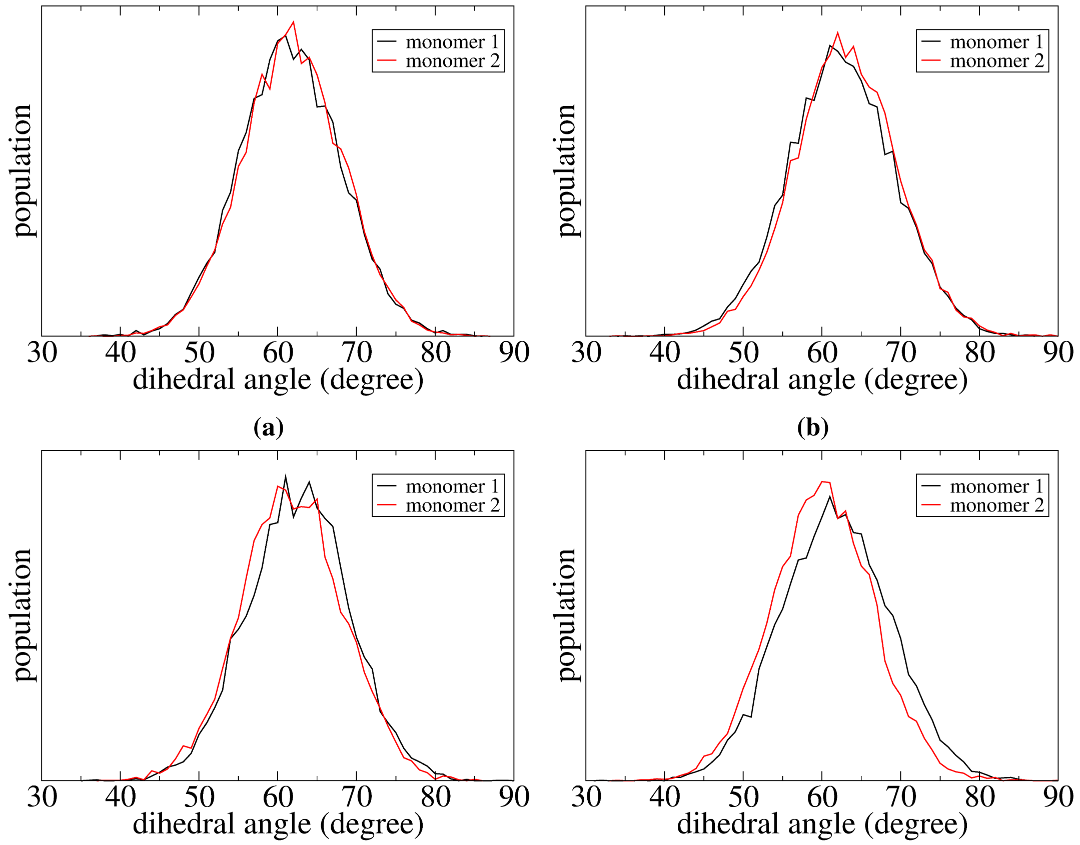
Figure 4.6: The internal water distribution: (a) monomer 1; (b) monomer 2

4.3.3 DT Gate

The MD trajectories show a stable DT gate structure, i.e. the direct hydrogen bond between D156 and T128 side chains, where D156 serves as the proton donor (Figure 4.3). This structure is similar to the homologous residues T90 (T128 in ChR2-C128T) and D115 (D156 in ChR2-C128T) of BR, where the hydroxyl oxygen of T90 side chain is the hydrogen bond acceptor of the carboxyl group of D115 [90, 91].

The N-C$_\alpha$-C$_\beta$-O$_\gamma$ dihedral angle distributions of T128 are displayed in Figure 4.7, which shows that T128 adopts the *gauche(-)* conformation with a dihderal angle of ca. 60°. One hypothesis is that the *gauche(-)/gauche(+)* conformation of threonine facilitates the formation of the intrahelical hydrogen bond with the *i*-4 residue (the residue which is four residues apart from T128) [191]. Consistently, the current MD simulations sample a stable intrahelical hydrogen bond between T128 side chain and W124 carbonyl oxygen (see Figure 4.2(b)). This promotes the conclusion that T128 cannot be the hydrogen bond donor in the DT gate of ChR2-C128T mutant. This conflicts with the experimental study of Nack et al. [89], which shows T128 as the hydrogen bond donor. However, the current result agrees well with the QM/MM study by Welke et al. [92].

The current simulations suggest a direct DT gate instead of a water bridged DT gate. Constrastly, Nack et al proposed a direct DC (in ChR2-WT) and DT (in ChR2-C128T mutant) gate structure, i.e. a direct hydrogen bond between the side chains of D156 and C128/T128. Moreover, Nack et al. proposed C128/T128 as the donor [89], this suggests D156 serves as the donor. Two observations are in favor of the DT model from the current study: (i) in the C1C2 X-ray structure, the thiol group of C167 (C128 in ChR2) points away from D195 (D156 in ChR2); (ii) the extended photocycle kinetics observed in ChR2 upon C128T mutation can hardly be explained by the same DC and DT gate structure.



(a)

(b)

(c)

(d)

Figure 4.7: The N-C$_\alpha$-C$_\beta$-O$_\gamma$ dihedral angle distribution of T128 from four trajectories.

4.3.4 Helix 3 Bending

In ChR2, helix 3 is bent due to the helix breaker P129 and the T127-E123 interaction as reported by Watanabe et al. [80]. The helix 3 bending angle is defined in the following way: (i) computation of the center of mass (COM) of residues 119–123 backbone atoms (COM 1), residues 124–128 backbone atoms (COM 2), residues 130–134 backbone atoms (COM 3) and residues 135–139 backbone atoms (COM 4); (ii) computation of vector 1–from COM 1 to COM 2, and vector 1–from COM 3 to COM 4; (iii) computation of the bending angle–the angle θ between vector 1 and vector 2 (see Figure 4.8).

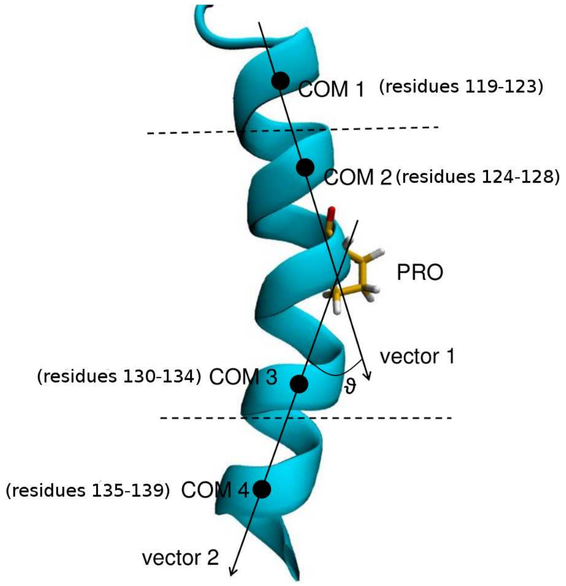


Figure 4.8: Schematic illustration of helix bending angle θ. COM 1: the center of mass of residues 119–123 backbone atoms; COM 2: the center of mass of residues 124–128 backbone atoms; COM 3: the center of mass of residues 130–134 backbone atoms; COM 4: the center of mass of residues 135–139 backbone atoms. Vector 1 is defined by COM 1 and COM 2, vector 2 by COM 3 and COM 4. The angle θ between vector 1 and vector 2 is the helix bending angle.

The helix 3 bending angles of one trajectory on the course of 500 ns are shown in Figure 4.9. (The results of the other three trajectories are shown in appendix, Figure A.6.) The average bending angles calculated every 500 ps are shown in green and blue lines for monomer 1 and monomer 2, respectively. The average angles fluctuate slightly around the starting values (ca. 27°) in both monomers.

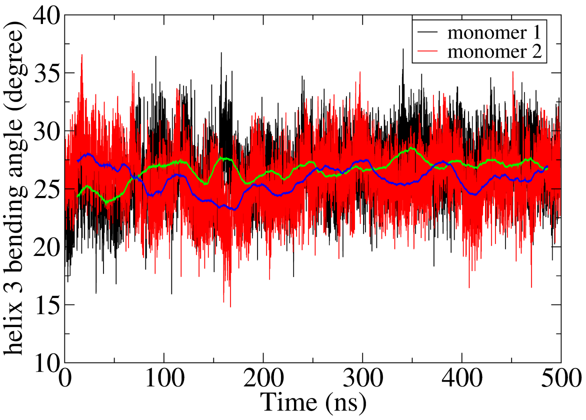
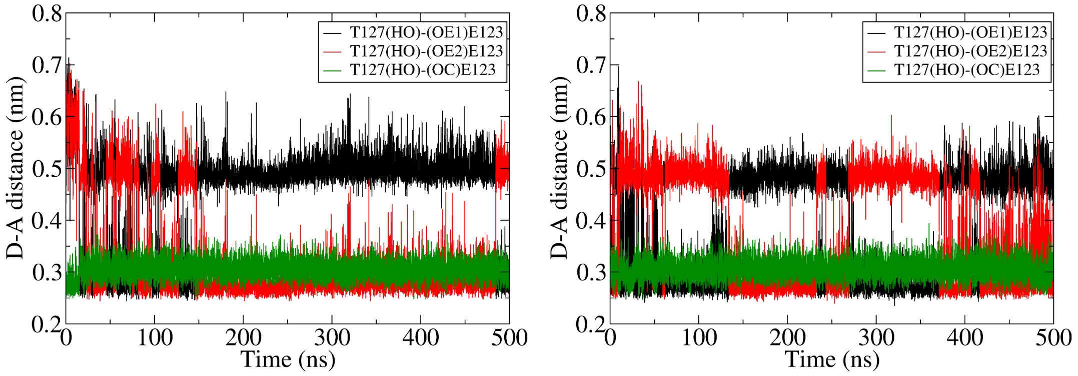
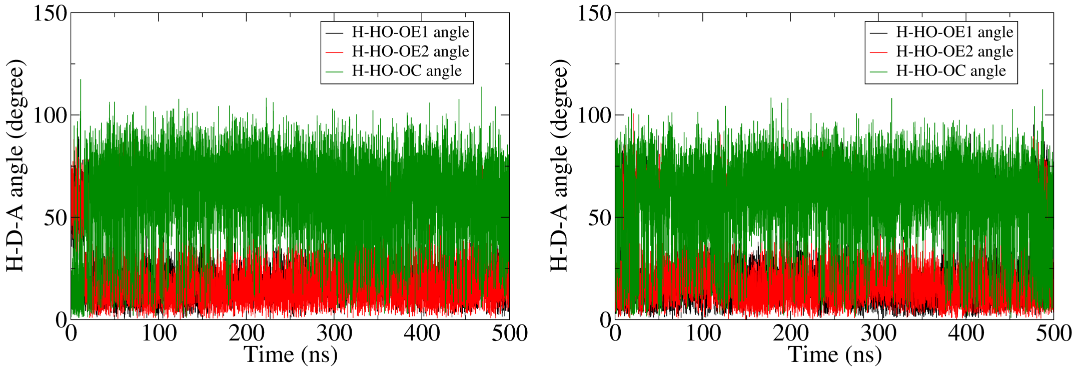


Figure 4.9: Helix 3 bending angle of two monomers. The average values calculated every 500 ps are shown in green and blue lines for monomer 1 and monomer 2, respectively.

The T127 hydroxyl group forms a hydrogen bond with the E123 side chain or backbone. The hydrogen bond is formed between T127 and E123, if the following two proposed criteria are fulfilled: (i) the donor-acceptor (D-A) distance is smaller than 0.35 nm; (ii) the frontier hydrogen-donor-acceptor (H-D-A) angle is smaller than 30°. The interchanges between these two interactions in one trajectory are shown in Figure 4.10, depicted by D-A distances and H-D-A angles. (The results of the other three trajectories are shown in appendix, Figures A.7–A.9.) In three out of four trajectories, most of the simulation time samples the hydrogen bond between T127 and E123 side chains in both monomers. In the remaining trajectory, monomer 1 shows that T127 forms a hydrogen bond with E123 backbone in the first 320 ns, and then switches mainly to E123 side chain. In monomer 2, the T127–E123 side chain hydrogen bond is also dominant (Figure A.9).



(a) monomer 1: D-A distance **(b)** monomer 2: D-A distance



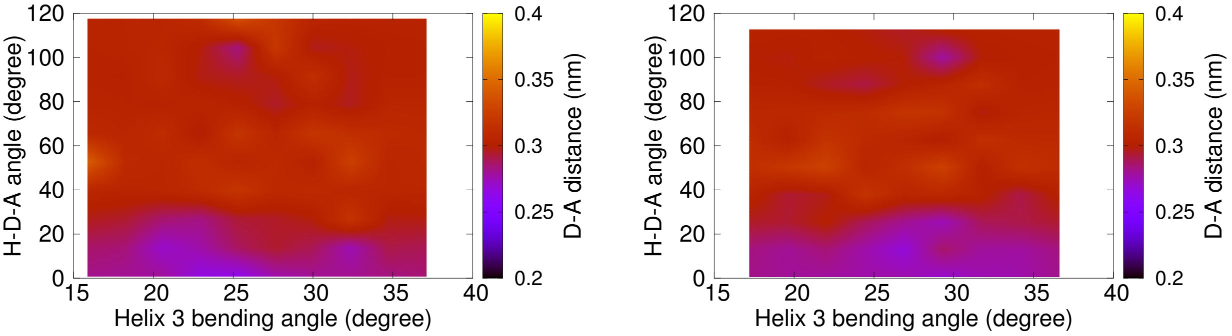
(c) monomer 1: H-D-A angle **(d)** monomer 2: H-D-A angle

Figure 4.10: T127–E123 hydrogen bonding interactions depicted by the donor-acceptor (D-A) distances and frontier H-donor-acceptor (H-D-A) angles. When the D-A distance is smaller than 0.35 nm, as well as the H-D-A angle is smaller than 30°, a hydrogen bond is formed. See atomic labels in Figure 4.2.

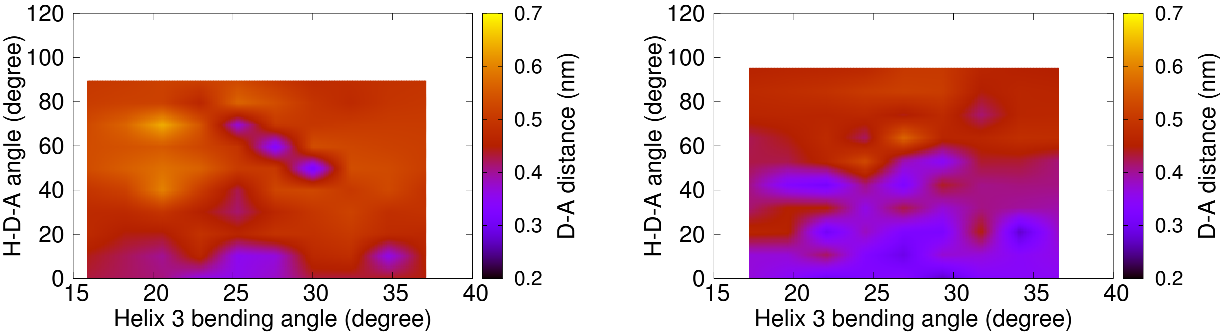
The correlation maps between helix 3 bending angle and T127–E123 hydrogen bonding interactions are shown in Figure 4.11 and Figures A.10–A.12. The left vertical axis is the H-D-A angle, whilst the D-A distance is represented according to the color bar. The horizontal axis shows the helix 3 bending angles. In all the maps the hydrogen bonding area covers almost the same helix 3 bending angle range. This means that the two T127–E123 hydrogen bonds corresponds to almost the same helix 3 bending angle range. This implies that the T127-(backbone)E123 hydrogen bond and the T127-(OOC)E123 hydrogen bond have no different influence on helix 3 bending. However, this does not support the assumption by Watanabe et al. [80], which argues that the T127-(OOC)E123 hydrogen bonding interaction reduces the helix bending with respect to the intrahelical hydrogen bond. Besides, in this former study, the helix bending is reduced during the 300 ns simulation time. The current extended simulations of 2 μs show, by a statistical analysis, that the helix 3 conformations are similar to the C1C2 crystal structure, hence, denoting an overall stability and a correct description of the structural models.

4.3.5 Central Gate

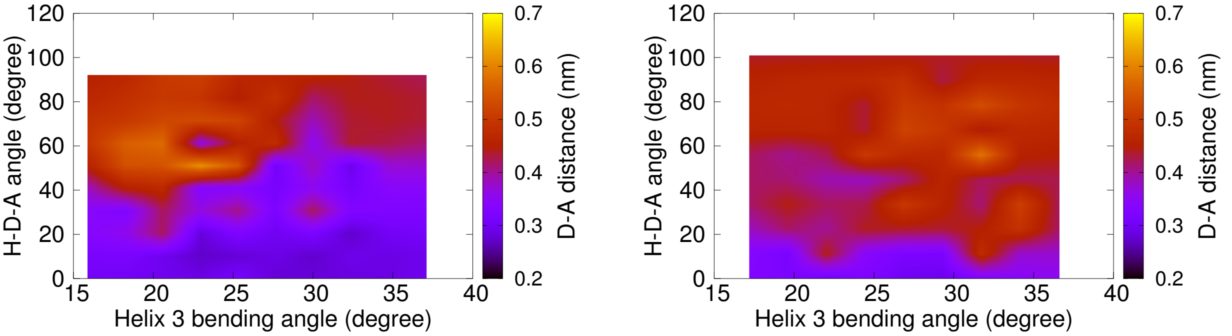
The central gate involving residues S63, N258 and E90 keeps the channel close before light exposure. A well studied central gate structure facilitates the understanding of the channel opening mechanism. The current MD simulations show that the E90 side chain is flexible. Three representative conformations are shown in



(a) monomer 1: T127-(backbone)E123 **(b)** monomer 2: T127-(backbone)E123



(c) monomer 1: T127-(OE1)E123 **(d)** monomer 2: T127-(OE1)E123



(e) monomer 1: T127-(OE2)E123 **(f)** monomer 2: T127-(OE2)E123

Figure 4.11: Correlation between helix 3 bending and T127–E123 interactions. (a) Interaction between T127 and E123 backbone in monomer 1; (b) Interaction between T127 and E123 backbone in monomer 2; (c,e) Interaction between T127 and E123 side chain in monomer 1; (d,f) Interaction between T127 and E123 side chain in monomer 2. The horizontal axis is the helix C bending angle; the vertical axis is the H-D-A angle; the color bar represent the D-A distance.

Figure 4.12. The backbone dihedral around C_β-C_γ of E90 is ca. ±180° in the conformations shown in Figure 4.12a and Figure 4.12b, respectively. In the case of Figure 4.12a, the hydroxyl group of E90(COOH) forms a hydrogen bond with the side chain carbonyl oxygen of N258, whilst the carbonyl oxygen of E90(COOH) forms a hydrogen bond with the side chain amine group of N258. Thus, helix 2 and helix 7 are connected through two hydrogen bonds. The residues S63 is not involved in the central gate hydrogen bonding network. However, because of the short atomic distances between S63 and E90/N258, the Van der Waals interactions [192] might contribute to the connection between helix 1 and helices 2 and 7.

On the other hand, in the case of Figure 4.12b, the hydrogen bond between the hydroxyl group of E90(COOH) and the side chain carbonyl oxygen of N258 is conserved. The downward orientation of the E90(COOH) carbonyl makes it too far away from the side chain amine group of N258 to form the hydrogen bond. Alternatively, the N258 side chain amine forms a hydrogen bond with the hydroxyl group of S63. In the

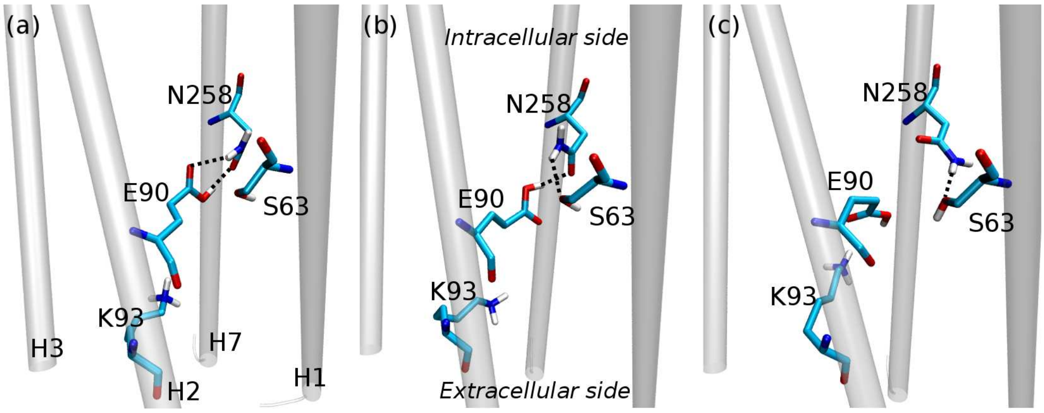


Figure 4.12: Three possible central gate structures. (a) Two hydrogen bonds between E90 and N258; (b) one hydrogen bond between E90 and N258, and one hydrogen bond between N258 and S63; (c) one hydrogen bond between N258 and S63.

trajectories, intermediate E90 conformations between the above two cases, i.e. the C_β-C_γ dihedral is between -180° and +180°, are observed (see Figure 4.13).

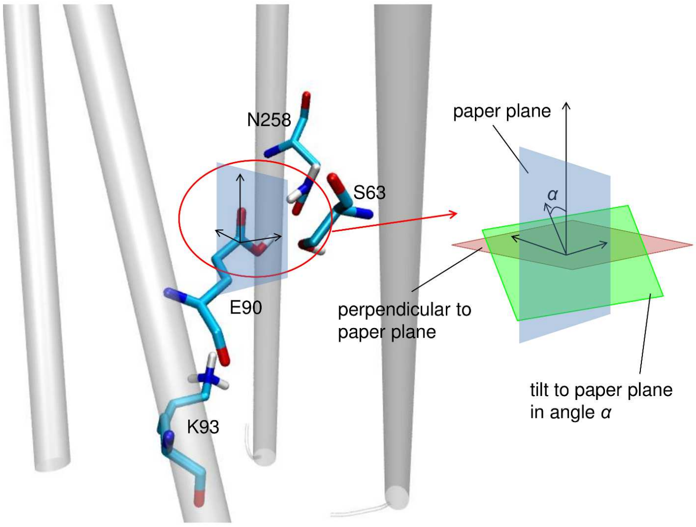


Figure 4.13: The plane formed by E90(COOH) atoms could be parallel with (C_β-C_γ dihedral is $\pm 180°$) or perpendicular to (C_β-C_γ dihedral is $\pm 90°$) the paper plane, or tilt to the paper plane forming the angle α (α is between -180° and +180°).

Interestingly, three out of four trajectories show the flip down of the E90 side chain towards the extracellular side forming a third possible central gate structure (Figure 4.12c). In this case the backbone dihedral angle around C_β-C_γ is ca. 65°. The hydrogen bond is formed between N258 side chain amine and S63 hydroxyl group. Additionally, some atomic distances between E90 and N258/S63 are within the Van der Waals interaction distance [192].

The initial ChR2-C128T model built using the C1C2 X-ray structure as the template, has the following central gate structure: one hydrogen bond between E90(COOH) and N258 side chain carbonyl, one hydrogen bond between N258 side chain amine and S63 hydroxyl. After MM equilibration, a central gate structure with two hydrogen bonds between E90(COOH) and N258 side chain is obtained. Using these two structures as the production MD starting points, a new downward oriented E90 side chain conformation was observed. The first appearance of this conformation happens after 150 ns for three out of four trajectories. As a result, there can be two, one or even no hydrogen bond between E90 and N258 side chains. In contrast, Kuhne et al. [78] reported a stable double hydrogen bond between E90 and N258, which keeps the side chains of E90 and N258 in a fixed conformation. Nevertheless in this earlier study the time scale of each simulation is only 100 ns, while an extended simulation time might be required to observe flexible E90 conformations, as performed in the current study. The multiple central gate structures obtained in the current MD simulations, however, do not conflict with the channel activation mechanism proposed by Kuhne et al. [78].

4.4 Conclusions

In this study, the long-standing function-related structural questions concerning the ChR2-C128T mutant were explored through classical MD simulations. The protein was found to have a similar conformation compared to the C1C2 crystal structure. The stable protein conformation (in terms of α-helices arrangement) in the extensive MD simulations allowed the analysis of the structural details.

The $\mathrm{RSBH}^{+}\cdots{}^{-}\mathrm{O(E123)}$ salt bridge was found to be the leading hydrogen-bonding pattern in the active site, which is ascribed to the classical force field. A stably preserved direct hydrogen bond between T128 and D156 with D156 as the proton donor was sampled. The intra- and inter-helical structural fluctuations generate flexibility in the central gate structure, ensuring in any case the presence of hydrogen bonds locking the channel in the same defined region of the ChR2-C128T mutant.

These results expand the structural database of ChR2-C128T and suggest improvements in the claim of relationships between ChR2 structure and optogenetical functions. Considering the growing scientific interest related to this protein, further efforts are needed to illustrate its mechanistic details, finally leading to the elucidation of the photocycle.

Chapter 5

Theoretical Study of Channelrhodopsin-2 C128T Mutant Dark-Adapted State Active Site

Reproduced in part from ref. [193] with permission from the Royal Society of Chemistry. ("Guo, Y.; Beyle, F. E.; Bold, B. M.; Watanabe, H. C.; Kolslowski, A.; Thiel, W.; Hegemann, P.; Marazzi, M.; Elstner, M. "Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant." *Chem. Sci.*, 2016, 7, 3879." Please see the right and permission of this paper though the following hyperlink: `http://pubs.rsc.org/en/content/requestpermission?msid=c6sc00468g`)

The detailed characterization of the active site of ChRs (including protonation states of E123, D253; water molecules arrangement; $RSBH^+$ HBond patterns) is of great importance for the following reasons: (i) The active site structure significantly affect retinal isomerization upon light absorption; (ii) The proton acceptor of $RSBH^+$ in the further photocycle is intimately related to the dark state active site structure; (iii) The active site is in direct connection with the cation conduction pathway (Figure 1.6), thus, it is important for cation conductivity and selectivity.

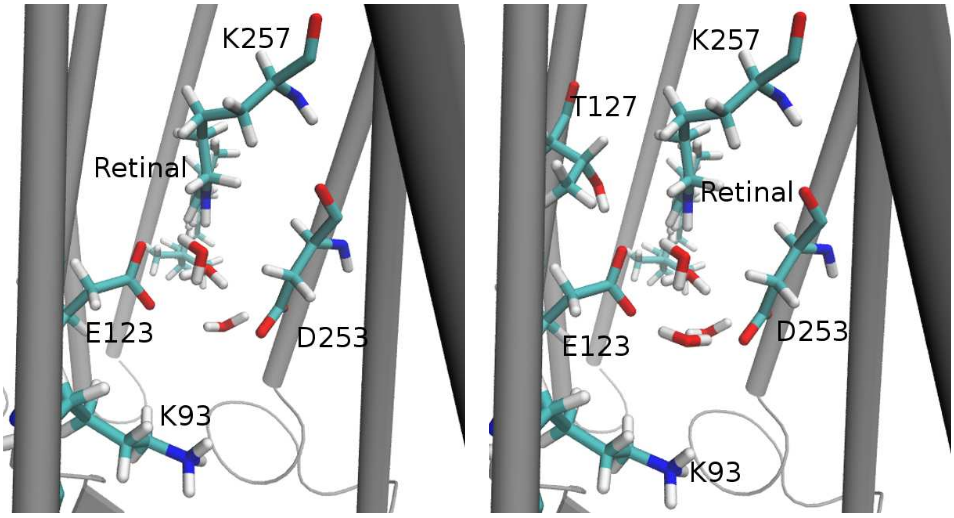
In the current chapter the active site of ChR2-C128T mutant was studied using combined QM/MM method.

5.1 Computational Details

As shown in Chapter 4, the proposed HBond betwen $RSBH^+$ and water molecule is not preserved in the MM/MD simulations [79, 80]. Besides, no large protein conformational change is observed with respect to the structures after NPT equilibration; the direct DT gate is stable. Considering the prominent role of protein-bound water molecules in microbial rhodopsins [194], in order to examine the active site structure in more detail, QM/MM MD simulations were performed using the NPT equilibrated structure as the starting point.

5.1.1 QM/MM Setup

The QM/MM models were built with QM regions (S-QM: 106 atoms, 0 charge) including the retinal chromophore and the covalently linked lysine (K257) side chain, K93, E123, and D253 side chains and three water molecules in the vicinity of the $RSBH^+$, see Figure 5.1. The remainder of the protein was treated as the MM region. The QM/MM boundary was set between the C_α and C_β atoms of the chosen QM residues. The valence of the QM fragment is saturated with a hydrogen link atom [133].



(a) small QM region **(b)** large QM region

Figure 5.1: QM regions defined for the current study. (a) Small QM region, S-QM, including the retinal chromophore and the covalently linked lysine (K257) side chain, K93, E123, and D253 side chains and three water molecules in the vicinity of the RSBH$^+$; (b) Large QM region, L-QM, additionally including T127 and one more water molecule with respect to S-QM.

A gradual release of the restraints was planned as shown in Table 5.1, finally leading to a QM/MM MD production run, where only the positions of the oxygen atoms of the QM water molecules are restrained with force constant of 500 KJ·mol^{-1} ·nm^{-1}, to retain the QM description of the active site. Otherwise, QM and MM water molecules would interchange, leading to an unbalanced description and QM convergence problems.

The production simulations sampled two E123 side chain conformations as shown in Figure 5.2 on the 2 ns simulation time, i.e. E123 side chain orients towards cytoplasmic side or orients downwards to the extracellular side. The downshift of E123 leads to the broken of the hydrogen bond between E123 and T127 side chains. Note that one MM water molecule stays stable in the active site and forms the water cluster together with the other three QM water molecules. Considering the limit of MM force field in describing the strong hydrogen bonded structures in proteins, a larger QM region (L-QM: 128 atoms, 0 charge) including additionally T127 and one more water molecule near the RSBH$^+$ relative to S-QM was considered, see Figure 5.1. The selection of a large QM region was considered necessary to obtain converged structures, as it was shown to be crucial in previous rhodopsin studies focusing on the active site [79, 80]. To get a statistical insight, 6 snapshots from the 2 ns trajectory of each simulation were randomly selected as the starting points for the final production QM/MM MD simulations.

The results shown in this chapter are all based on the simulations with L-QM.

It needs to be mentioned that, because the DT gate and the active site is correlated, principally it is crucial to include the DT gate into the QM description. However, in fact it is not feasible to include the DT gate. One either would have to use a quite large QM region, thereby reducing the amount of sampling drastically; or one would have to set up two separate QM regions, which is not a favorable choice to us. The DT gate does not have to be included into the QM region as long as it is intact, due to its long distance from the Schiff base. The current MM/MD simulations sample a single direct HBond between T128 and D156, as discussed in a previous theoretical study [92], so does the following QM/MM production simulations. Thus, the DT gate is excluded out of the QM region in all QM/MM simulations.

13-*cis*,15-*syn* Retinal

Considering the resonance Raman and IR spectroscopy studies which indicate the additional presence of the 13-*cis*,15-*syn* retinal configuration in ChR2 dark state [68, 107], the corresponding models for ChR2-C128T were built. More in detail, the 13-*cis*,15-*syn* retinal from BR (PDB code:1XOS [195]) was aligned to the all-*trans* retinal, and then the all-*trans* retinal was replaced by the aligned 13-*cis*,15-*syn* retinal. The RSBH$^+$

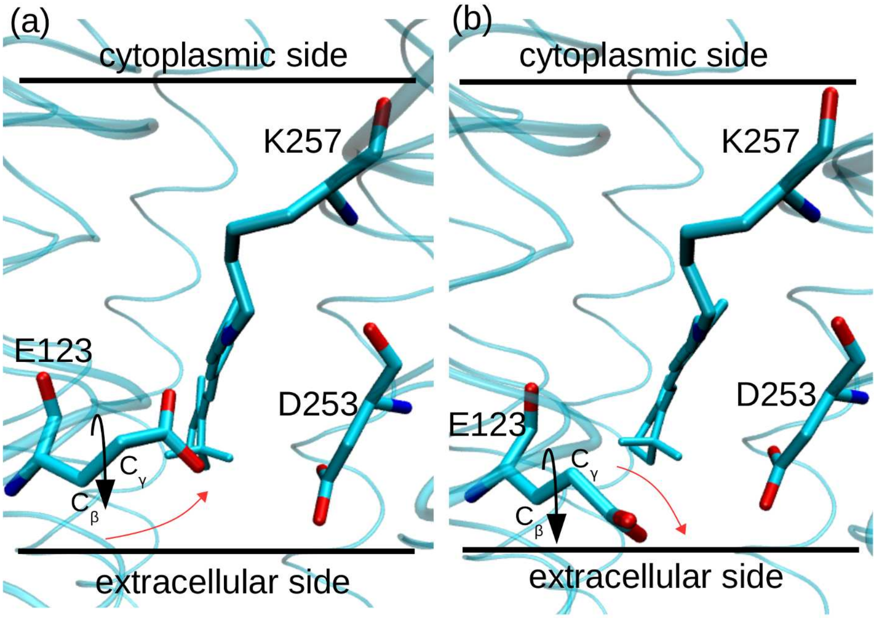


Figure 5.2: The (a) upward and (b) downward E123 side chain conformations. The black arrows depict the dihedral around the C_β-C_γ bond of E123. The red arrows depict the E123 side chain orientation.

has similar position in the all-*trans* retinal bound complex and the 13-*cis*,15-*syn* retinal bound complex. The snapshots used for aligning were randomly selected from the all-*trans* retinal trajectories with L-QM. After replacing with *cis* retinal, the protein and water environment were equilibrated energy minimization with position restraints (force constant: 1000 $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-1}$) on the retinal, K93, E123, T127, D253 and four water molecules involved in the active site hydrogen bonding patterns, to remove the inappropriate atomic conflicts. The minimization was performed at the MM level of theory using the steepest descent method. The obtained structures were used as starting points for the production QM/MM MD simulations, for which the L-QM was applied.

For comparison purposes, a QM/MM trajectory of 1 ns was performed for a BR model, with the initial structure taken from the previous study [196]. In this case, the QM region includes the retinal chromophore (in all-*trans* configuration), the K219, D85 and D212, R82 side chains, and the three water molecules forming the typical BR pentagonal cluster active site (Figure 1.7). QM/MM MD simulations in the earlier publications do not sample other ground state local minima besides this structure. Thus, the amount of sampling to cover the ground state structural phase-space is much reduced with respect to ChR2-C128T, which shows the large structural heterogeneity (see the result section). A sufficient sampling must assure that the trajectories visit all the local minima sufficiently, therefore, the BR active site structure is suitably sampled within 1 ns.

All simulations were performed with the GROMACS package [169, 170]. The Charmm36 classical force field [188] was applied to describe the MM region, including parameters for the lipid membrane [189]. The TIP3P model [164] was used to describe water molecules. The hybrid QM/MM simulations are carried out appling the DFTB3/3ob method [130–132] (third order self-consistent-charge Density-Functional-Tight-Binding method) to the QM region, as recently implemented in the GROMACS package [197]. DFTB3/3ob has been shown to model hydrogen bonded networks with similar accuracy as full DFT calculations using medium sized basis sets [130], which is essential for a reliable modeling of the extended hydrogen bonded structures in ChR2. Since trajectories on the ns time scale were required to study the formation and stability of hydrogen bonds, the application of DFTB/MM was preferred to the computationally more expensive DFT/MM. Indeed, DFTB/MM has been successfully applied to retinal proteins in the last years, in particular showing how infrared and absorption spectroscopy are sensitive to the active site structure [92, 162, 198].

Table 5.1: The setup of QM/MM simulation protocol.

Steps	Simulation time (ps)	Position restraint-1[a]	Position restraint-2[b]	Distance constraint[c]
1	200	5000	500	Yes
2	200	3000	500	Yes
3	200	2000	500	Yes
4	200	1000	500	Yes
5	200	500	500	Yes
6	200	0	500	No
7	2000	0	500	No

[a] The force constant of the position restraint on the H atom of $RSBH^+$ (unit: KJ mol^{-1} nm^{-1}).
[b] The force constant of the position restraint on the oxygen atoms of the QM water molecules (unit: KJ mol^{-1} nm^{-1}).
[c] Whether apply the distance constraint on E123–T127 HBond.

5.1.2 Excited-State Calculations

A high accuracy and extensive description of the active site by the DFTB3 method was considered fundamental to obtain reliable ground state structures, followed by a calculation of the electronic excited-state energies. The reproduction of the experimental absorption spectrum (excitation energies) is one characterization of a good active site structure. Indeed, it was shown that the remaining binding pocket (*i.e.* the other amino acids surrounding the retinal chromophore apart from E123 and D253) is responsible only for a minor hypsochromic (*i.e.* blue) shift when studying color tuning properties, and the influence of the rest of the protein is not crucial [196]. Therefore, the region outside the active site could be safely treated by classical force fields.

Two different strategies are applied in order to evaluate the absorption spectrum properties.

SORCI Calculation

Twelve geometries randomly selected from the QM/MM trajectories with S-QM region were optimized at the DFTB3/Charmm level of theory with the CHARMM37b1 suite of program [199], to ensure that harmonic fluctuations of the structures around the energy minima are neglected. For each minimized structure, a single point calculation of the electronic excitation energy was performed with the Spectroscopy-ORiented multireference Configuration Interaction (SORCI) method [126], as implemented in the ORCA program package [200].

In the QM/MM setup for the calculation, the QM region is defined including the retinal, the side chains of K257, E123 and D253, and the water molecules in the vicinity of $RSBH^+$ (see Table 5.2). In addition, for comparison purpose, the excitation energy of BR minimized structure was calculated with the same SORCI setup, and the typical pentagonal structure was treated as the QM region. Since as previously discussed [201], the BR active site has a well-defined dark state minimum, the geometry optimization of different trajectory snapshots leads to the same minimum energy structure, only one BR snapshot was calculated for the comparison in the current study.

A complete active space of 12 electrons in 12 orbitals (i.e. the 6 highest occupied molecular orbitals and the 6 lowest unoccupied molecular orbitals) was selected, including 6-π and 6-π * orbitals of the retinal chromophore. The starting orbitals were from a RHF (Restricted Hartree-Fock) [202, 203] calculation, and the split valence basis set together with one set of polarization functions on heavy atoms taken from the TurboMole library (def2-SV(P)) [204, 205] was applied. Neese et al. [126] reported that the SORCI excitation energies are fairly stable to variations of the basis set. In most cases, the results obtained with the small SV(P) basis were within ca. 0.1 eV difference compared to the more extended aug-cc-pVTZ basis set. Moreover, a much larger computational cost is expected for large molecules, like retinal. In addition, applying the same basis set as in former rhodopsin studies makes it possible to compare results. Therefore, the def2-SV(P) basis set was applied.

The three thresholds to enhance computational efficiency were set as follows. The $T_{prediag}$, giving the configuration state functions built into the reference space, was set to 10^{-3}. The selection threshold T_{sel}, which divides the first-order interacting space into strongly perturbing and weakly perturbing subspaces, was set to 10^{-6} E_h (E_h=627.51 kcal mol^{-1}). The number of approximated natural orbitals that enter the final CI calculation is limited by the threshold T_{nat} which was set to 10^{-6}. The excited state energy was converged within 0.1 eV using these threshold values according to the previous study on the retinal model [206]. Additionally, a comparison with the established CASPT2 method [125] illustrates that SORCI is a reliable method to describe excitation energies of retinal proteins [206].

The current SORCI calculations show a major difference in the electronic excited-state description between ChR2-C128T and BR. For both proteins, the optically bright electronic transition is a $^1(\pi,\pi^*)$ transition (singlet $\pi \rightarrow \pi^*$ transition), but in BR it is always a $S_0 \rightarrow S_1$ vertical excitation, while in ChR2 it corresponds to a $S_0 \rightarrow S_1$ or $S_0 \rightarrow S_2$ vertical excitation. In ChR2-C128T, for one of the sampled structures (Table 5.3, index 9), S_1 is described by a double $^1(\pi,\pi^*)$ transition (i.e. a transition involving the promotion of two electrons from bonding to antibonding π orbitals) associated with a low oscillator strength. This means that the electronic transition of interest is within the ground state and two low-lying excited states. Therefore, at least three roots should be included for the SORCI calculation of ChR2. The inversion between the S_1 and S_2 electronic states might be due to that BR and ChR2-C128T have different active site structure. Furthermore, it may lead to different description of the retinal photoisomerization with respect to BR. Following in the analysis, the optically bright electronic transition is always taken into account.

To evaluate the dependence of the excitation energy on the number of calculated exitation states, the single point energy calculation was performed with the number of roots varies from three to six, i.e. scales from two excited states to five excited states. The single excitation energies and corresponding oscillator strength for two geometries are listed in Table 5.4. The increase of the number of calculated excitation states does not have much influence on the single excitation energy, which does not vary with the geometries. Since the inclusion of more roots will unnecessarily increase the computational time, three roots were calculated in the current study.

Table 5.2: QM/MM setups for SORCI and OM2/MRCI calculations.

	QM region
QM-5	retinal+ (K257 + E123 + D253 side chains) + H_2O*
QM-1	retinal + K257 side chain**
QM-4	retinal+ (K257 + E123 side chains)**
QM-3	retinal+ (K257 + D253 side chains)**
QM-2	retinal+ K257 side chains + H_2O**

* The setup for SORCI and OM2/MRCI calculations.
** The setup for OM2/MRCI calculations.

OM2/MRCI Calculation

In order to evaluate the dynamical effect on the absorption spectrum, the semi-emprical OM2/MRCI (Orthogonalization corrected Model 2/Multi-Reference Configuration Interaction) method [207, 208] was applied on a statistical number of QM/MM trajectory snapshots. For comparison, the same strategy was applied to the BR model, calculating the OM2/MRCI excitation energy of 1000 snapshots generated by the QM/MM trajectory. The OM2/MRCI method was extensively tested [209] and successfully applied to BR, sensory rhodopsin II and bovine rhoposin [210]. The main advantages offered by OM2/MRCI with respect to *ab initio* methods as SORCI and CASPT2 is the possibility to select a larger active space at a considerable fraction of the computational cost, see Table 5.5 for the calibration. In the present study the active space of 20 electrons in 20 orbitals is used. The reference configurations comprised the closed-shell ground state configuration and four configurations generated by excitations from the two highest occupied molecular orbital (π, $\pi - 1$) to the two lowest unoccupied molecular orbitals (π^*, $\pi + 1^*$), i.e. the single excitations $\pi \rightarrow \pi^*$ and $\pi - 1 \rightarrow \pi^* + 1$ as well as the double excitations (π, π)$\rightarrow$(π^*, π^*) and (π, $\pi - 1$)$\rightarrow$(π^*, $\pi + 1$). All single and double excitations from the reference configurations were included in the OM2/MRCI calculations. On the other hand, the main drawbacks of OM2 are an overestimation of the excitation energy and the possible underestimation of the charge transfer in the excited state. In order to set the optimal conditions, such drawbacks were taken into account by calculating with OM2/MRCI method the absorption energy and oscillator strength of the same 12 ChR2-C128T geometries, which were previously optimized for SORCI calculations.

Table 5.3: The orbital transitions and oscillator strength (f) of twelve randomly selected geometries.

index	electronic states	orbital transition	f	index	electronic states	orbital transition	f
	S_0	GS[a]			S_0	GS	
1	S_1	π,π^*[b]	1.96	7	S_1	π,π^*	1.86
	S_2	$2\pi,2\pi^*$[c]	0.02		S_2	$2\pi,2\pi^*$	0.03
	S_0	GS			S_0	GS	
2	S_1	π,π^*	1.85	8	S_1	π,π^*	1.94
	S_2	$2\pi,2\pi^*$	0.02		S_2	$2\pi,2\pi^*$	0.02
	S_0	GS			S_0	GS	
3	S_1	π,π^*	1.93	9	S_1	$2\pi,2\pi^*$	0.01
	S_2	$2\pi,2\pi^*$	0.02		S_2	π,π^*	1.90
	S_0	GS			S_0	GS	
4	S_1	π,π^*	1.89	10	S_1	π,π^*	2.14
	S_2	$2\pi,2\pi^*$	0.02		S_2	$2\pi,2\pi^*$	0.002
	S_0	GS			S_0	GS	
5	S_1	π,π^*	1.92	11	S_1	π,π^*	1.87
	S_2	$2\pi,2\pi^*$	0.03		S_2	$2\pi,2\pi^*$	0.03
	S_0	GS			S_0	GS	
6	S_1	π,π^*	1.84	12	S_1	π,π^*	1.91
	S_2	$2\pi,2\pi^*$	0.02		S_2	$2\pi,2\pi^*$	0.02

[a] GS: ground state.
[b] π,π^*: single $\pi \rightarrow \pi^*$ transition .
[c] $2\pi,2\pi^*$: double $\pi \rightarrow \pi^*$ transition .

Table 5.4: The dependence of single excitation energy on the calculated number of excitation states.

N_{roots}[1]	E[2] eV (nm)	f_{osc}[3]	N_{roots}	E eV (nm)	f_{osc}
3	2.61 (476)	1.83	3	2.98 (416)	1.98
4	2.64 (469)	1.84	4	3.01 (412)	2.05
5	2.69 (461)	1.86	5	3.00 (413)	2.03
6	2.59 (478)	1.79	6	3.00 (412)	1.89

[1] N_{roots}, the number of roots.
[2] Single excitation energy.
[3] The oscillator strength.

Comparison of SORCI and OM2/MRCI

The excitation energies from SORCI and OM2/MRCI calculations with the QM region of QM-5 (see Table 5.2) show a linear correlation as shown in Figure 5.3. The correlation coefficient of ca. 0.84 indicates that OM2/MRCI is applicable for the current system. The corresponding single-point excitation energies listed in Table 5.6 depicts that a systematic hypsochromic shift of ca. 0.30 eV$\sim$0.40eV can be applied to quantitatively reproduce the SORCI excitation energy. The comparison between the simulated absorption spectra is shown in Figure 5.4. The relative intensity of each histogram is calculated as the normalized count of geometries in a defined energy window (0.05 eV), multiplied by the average oscillator strength within the same energy window.

The bond length alternation (BLA) of the retinal polyene chain, that is the difference between averaged single and double bond lengths, is directly correlated to the C=C stretch frequency and thus the absorption maximum [196]. Both the SORCI and OM2/MRCI levels of theory recover the correlation between BLA and excitation energy as shown in Figure 5.5.

OM2/MRCI Calculation Setup

There are two main effects by describing protein environment as point charges: (i) considering that the positive charge is transferred from the Schiff base towards the β-ionone ring during the S_0-S_1 excitation, the negative charged counterions in vicinity of the Schiff base stabilize the S_0 state relative to the S_1; (ii) the remainder of the protein polarize the QM region, which leads to a blue shift of the excitation energy.

In order to establish the optimal setup to be used for the calculation of the absorption spectra, the impact of protein environment and the size of the QM region (Table 5.2: QM-1, QM-2, QM-3, QM-4, QM-5) were calibrated in prior. As shown in Figure 5.6, the presence of MM point charges (left panel) results in higher excitation energies with respect to the calculations without MM point charges (right panel) for all QM regions. The absorption maximum calculated from five QM regions are similar (ca. 3.2 eV) in the presence of MM point charges. For the sake of saving computational cost, the smallest QM region, i.e. only including the retinal and the connected K257 side chain, was used for all the OM2/MRCI calculations.

For more benchmark calculation on ChR2-WT, please see our publication [193], Supporting Information.

Table 5.5: OM2 calibration calculations of the active space size, excitation level and the calculation time.[*]

active space	excitation level	calculation time	$E_{S1}{}^{a}$(eV)	$E_{S2}{}^{b}$energy (eV)
(12,12)	2	$\sim$ 14 s	2.41	3.32
	3	$\sim$ 3 min	2.41	3.32
	4	$\sim$ 25 min	2.41	3.32
	5	$\sim$ 40 min	2.41	3.32
	6	$\sim$ 60 min	2.41	3.32
(14,14)	2	$\sim$ 1 min	2.40	3.30
	3	$\sim$ 20 min	2.40	3.30
	4	$\sim$ 170 min	2.40	3.31
	5	$\sim$ 630 min	2.40	3.31
(16,16)	2	$\sim$ 3 min	2.39	3.29
	3	$\sim$ 130 min	2.39	3.29
	4	$\sim$ 1070 min	2.39	3.29
(18,18)	2	$\sim$ 7 min	2.36	3.25
	3	$\sim$ 130 min	2.36	3.25
(20,20)	2	$\sim$ 20 min	2.35	3.23
	3	$\sim$ 1020 min	2.35	3.23
(22,22)	2	$\sim$ 30 min	2.32	3.20
(24,24)	2	$\sim$ 50 min	2.32	3.18
(26,26)	2	$\sim$ 90 min	2.32	3.17
(28,28)	2	$\sim$ 160 min	2.28	3.12
(30,30)	2	$\sim$ 250 min	2.27	3.12
(32,32)	2	$\sim$ 400 min	2.27	3.1

[*] All the calculations are based on the retinal in vacuum.

[a] The first excitated state with one electron being excited from bonding to antibonding orbital.

[b] The second excitated state with two electrons being excited from bonding to antibonding orbitals.

Table 5.6: Comparison of the single-point excitation energy from SORCI and OM2/MRCI calculations (unit: eV).*

SORCI	OM2/MRCI	$\Delta E_{(SORCI-OM2)}$	SORCI	OM2/MRCI	$\Delta E_{(SORCI-OM2)}$
2.95	3.24	-0.29	2.74	3.17	-0.43
2.85	3.15	-0.30	2.83	3.17	-0.34
2.64	3.02	-0.38	2.86	3.20	-0.34
2.87	3.22	-0.35	2.96	3.26	-0.30
2.76	3.12	-0.36	2.79	3.13	-0.34
2.87	3.22	-0.35	2.78	3.14	-0.36

[*] The QM region includes the retinal chromophore and the covalently linked K257 side chain, E123, D253 side chains and the water molecules in vicinity of the PSB.

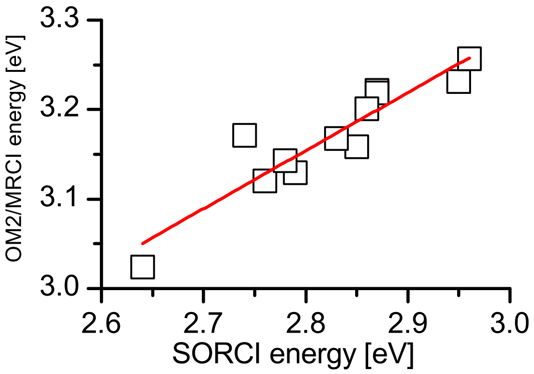
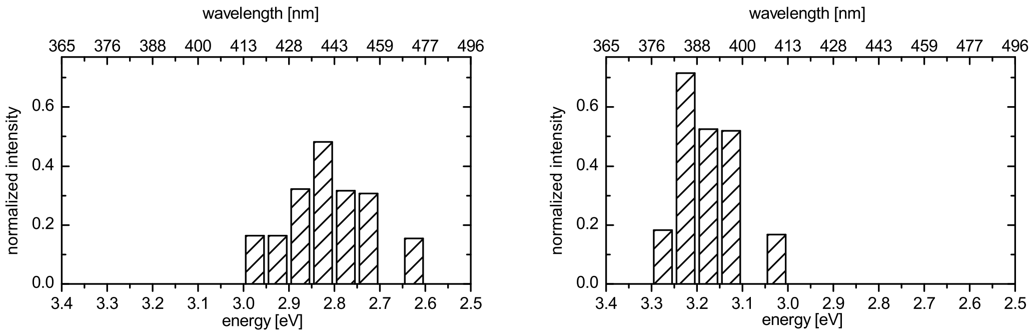


Figure 5.3: Cross-correlation function between SORCI and OM2/MRCI levels of theory. The excitation energies refer to the optically-bright electronic transitions.



(a) SORCI **(b)** OM2/MRCI

Figure 5.4: The comparison of the simulated absorption spectrum. (a) SORCI calculation; (b) OM2/MRCI calculation.

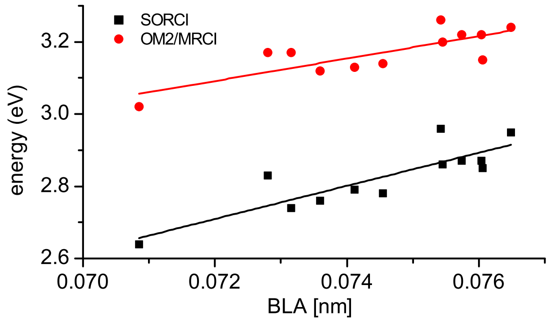
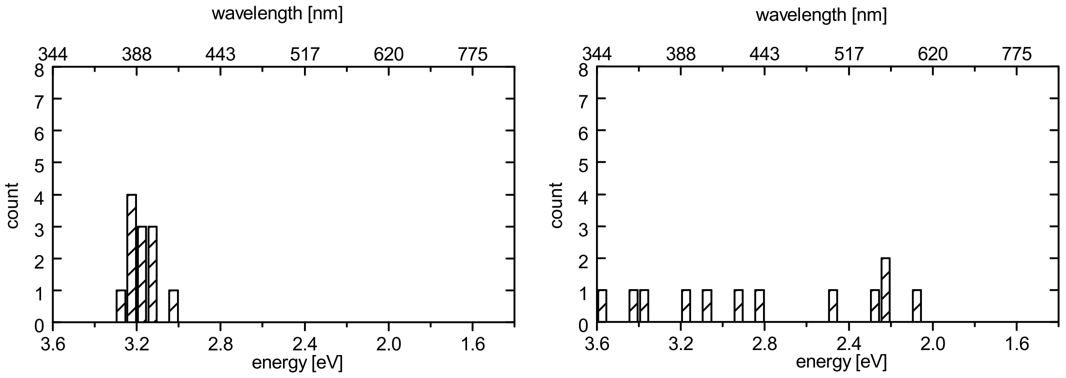
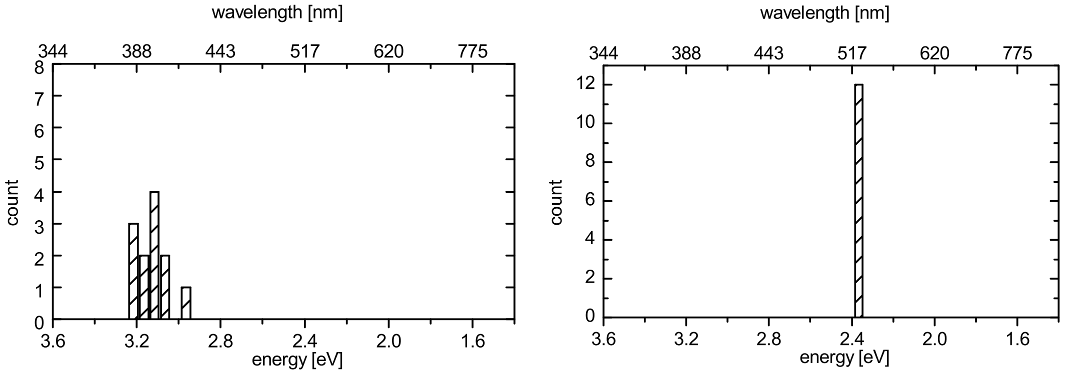


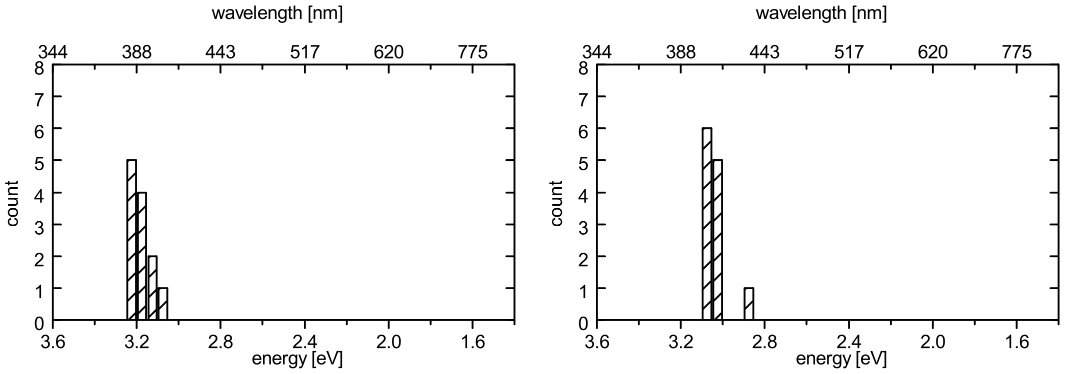
Figure 5.5: Qulitative linear dependence of the excitation energy, as a function of the retinal bond length alternation (BLA).



(a) QM-1 **(b)** QM-1



(c) QM-2 **(d)** QM-2



(e) QM-3 **(f)** QM-3

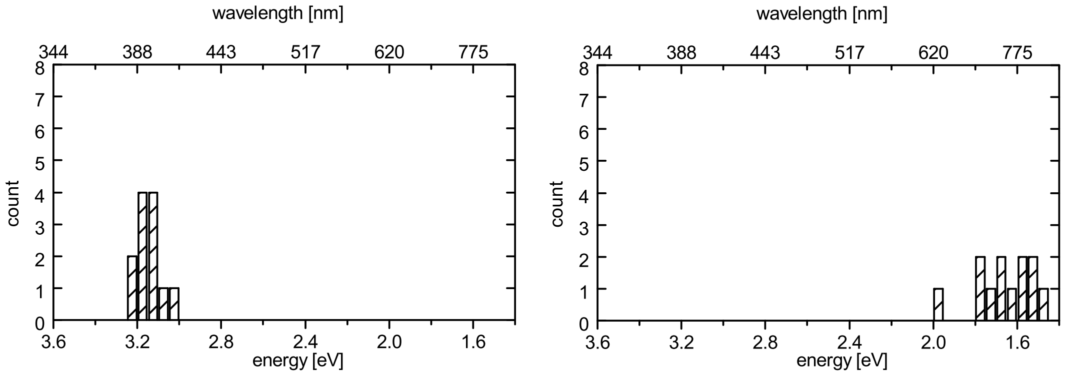
5.2 Results and Discussion

5.2.1 The Chromophore Geometry

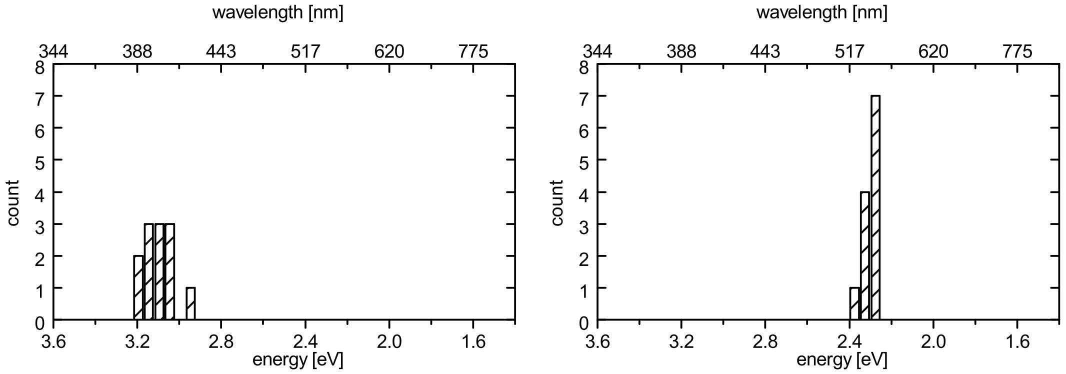
The Whole Chromophore

The retinal conjugated polyene chain bond length (Figure 5.7), curvature (κ_i, Figure 5.8) and helicity (η_i, Figure 5.9) were characterized. The way of calculating κ_i and η_i are adopted from ref. [187] as shown below in equations (5.1–5.2), respectively.

$$\kappa_i = (-1)^i(\theta_i - 120°) \tag{5.1}$$



(g) QM-4 **(h)** QM-4



(i) QM-5 **(j)** QM-5

Figure 5.6: OM2/MRCI absorption spectra calculated from different QM/MM setups as shown in Table 4.2. The left panel is with MM charge, the right panel is without MM charge. The QM regions (see Table 5.2 for details) are listed below the graphs. The vertical axis scale of (d) is 0 to 12, while in the other graphs it is 0 to 8 in order to have a good vision of the histograms.

where i is the index of the carbon atom C$_i$. This equation describes the deviation of θ_i (the C$_{i-1}$-C$_i$-X$_{i+1}$ angle) from 120° (X $\in$ C, N).

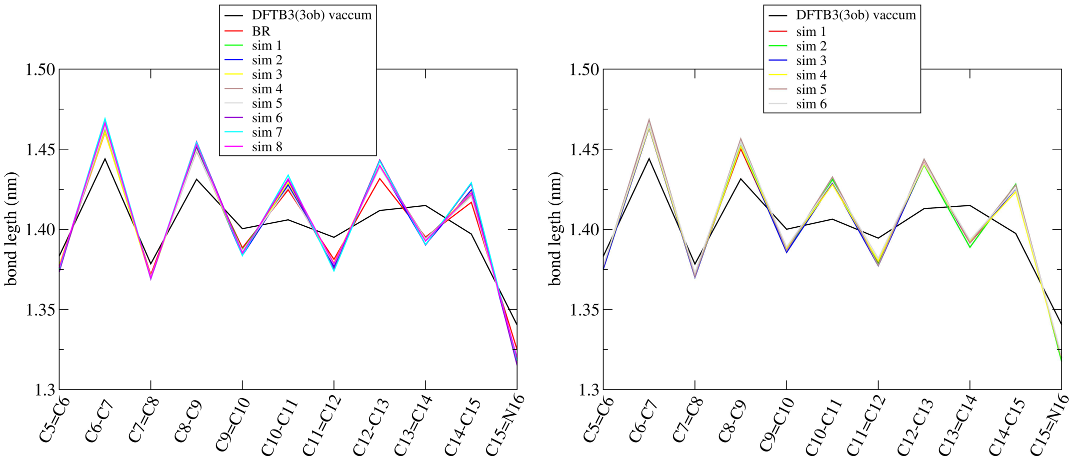
$$\eta_i = (-1)^i(\varphi_i - 180°) \tag{5.2}$$

describes the deviation of φ_i (the dihedral around the C$_i$-X$_{i+1}$ bond) from 180°, reflecting the planarity of the retinal polyene chain.

For comparison purpose, all-*trans*,15-*anti* and 13-*cis*,15-*syn* retinal isomers were optimized in vaccum at DFTB3(3ob) level of theory using DFTB+ program. The retinal models are shown in Figure 5.10, the C$_\beta$ atom of connected lysine is saturated using hydrogen atoms. In Figure 5.7 it can be seen that, the presence of the protein–ChR2 and BR, reduces the retinal polyene chain conjugation, especially in the region C9–C15. The polyene chain bond lengths of two isomers are very similar in ChR2 and BR.

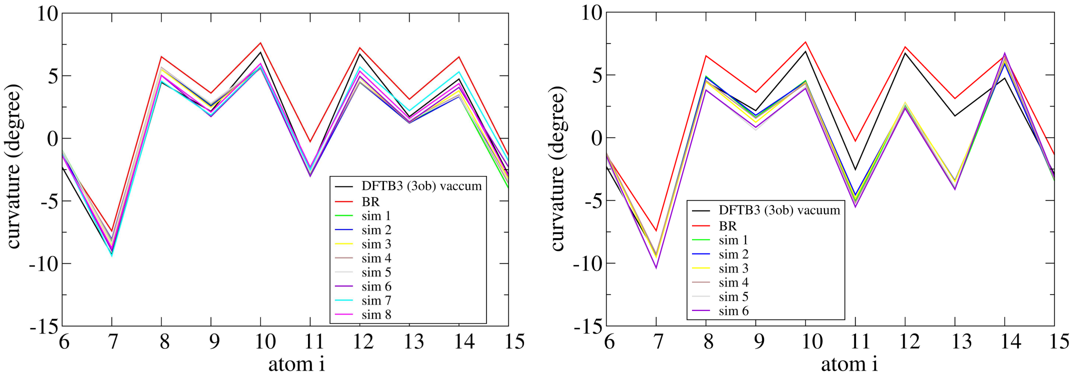
The curvature of the polyene chain is mainly due to sterical interactions of the two bulky methyl groups at C9 and C13 with surrounding hydrogen atoms. As a result, the local curvatures in regions C8–C10 and C12–C14 are specifically large. ChR2 and BR protein environment does not have much influence on retinal curvature. The curvature plots shown in Figure 5.8 depicts the local mirror symmetry of the retinal polyene chain defined by a plane vertically crossing the polyene chain at C11 (see Figure 5.10, in which the projection of the mirror plane is represented by the dashed red line). The 13-*cis*,15-*syn* retinal shows a larger absolute curvature value at C13 with respect to all-*trans*,15-*anti* isomer, which is protein environment independent. This is possibly due to the increased hydrogen–hydrogen sterical interaction between C12-H and C15-H in the *cis* isomer (Figure 5.10).

The helicity of the two isolated retinal isomers vanishes almost everywhere (Figure 5.9), revealing that the molecules have nearly planar π electron system. The ca. 15° torsion around C6–C7 single bond is caused by sterical interactions between the two methyl groups at C1 and the hydrogen atom at C8 (see Figure 5.10). The embedding of the retinal into ChR2-C128T and BR leads to twists around almost all polyene chain bonds, which are nearly independent of the protein pocket. For the *trans* isomer, each bond is twisted 5°–10°. Interestingly, the twists are alternative, i.e. in the order of clockwise (CW)–counter clockwise (CCW)–CW-CCW$\cdots$ (or CCW–CW–CCW–CW$\cdots$, depending on how to define the CW/CCW direction in the system). Consequently, the twist of the whole polyene chain is almost compensated, resulting in a planar π electron system as the isolated molecule. This viewpoint is quantitatively verified by the averaged torsional angles around C13=C14, C14-C15, C15=N, all of which are very close to 180° as shown in Table 5.7. The twist of the *cis* isomer polyene chain from C6–C11 is also almost compensated. However, there is a twist of 10°–15° around each bond from C12–N, and all these twists are in the same direction. This indicates that the *cis* retinal isomer is not planar in ChR2-C128T or BR dark state.



(a) all-*trans*,15-*anti* **(b)** 13-*cis*,15-*syn*

Figure 5.7: The averaged retinal polyene chain bond length: (a) all-*trans* retinal from eight trajectories, 1 ns of each; (b) 13-*cis*,15-*syn* retinal from six trajectories, 1 ns of each. The bond lengths of the retinal optimized at DFTB(3ob) level of theory, and the averaged retinal bond lengths from BR trajectories are also shown.

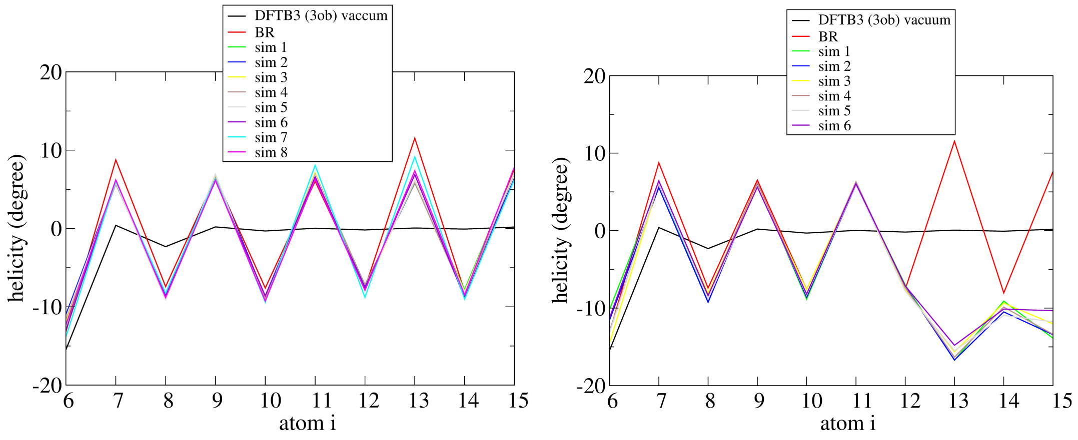


(a) all-*trans*,15-*anti* **(b)** 13-*cis*,15-*syn*

Figure 5.8: The averaged retinal polyene chain curvature: (a) all-*trans* retinal from eight trajectories, 1 ns of each; (b) 13-*cis*,15-*syn* retinal from six trajectories, 1 ns of each. The polyene chain curvature of the retinal optimized at DFTB(3ob) level of theory, and the averaged retinal polyene chain curvature from BR trajectories are also shown.

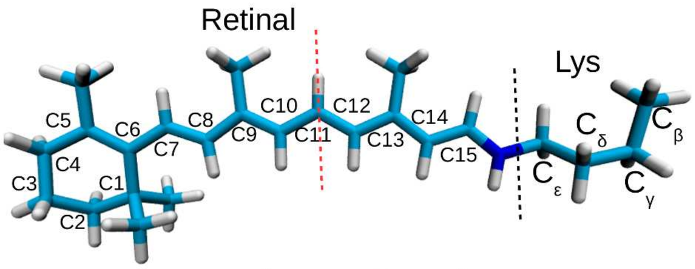
The Polyene Chain Close to NH^+

In order to get an insight into the chromophore conformation, previous studies investigated the H-C14-C15-H (HCCH) torsional angle and the C14-C15 bond [75, 211, 212]. As expected, the DFTB3(3ob) method gives completely planar *trans* and *cis* retinal in vaccum, see Table 5.7. In the current DFTB3(3ob)/MM simulations

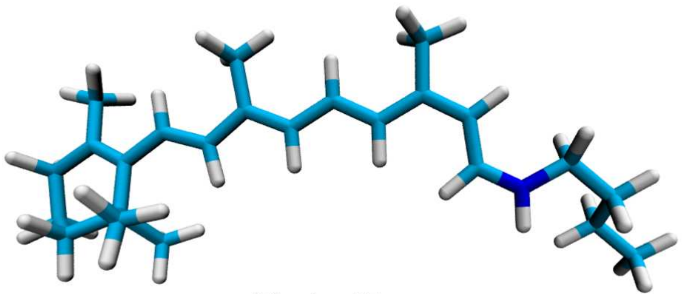


(a) all-*trans*,15-*anti* **(b)** 13-*cis*,15-*syn*

Figure 5.9: The averaged retinal polyene chain helicity: (a) all-*trans* retinal from eight trajectories, 1 ns of each; (b) 13-*cis*,15-*syn* retinal from six trajectories, 1 ns of each. The polyene chain helicity of the retinal optimized at DFTB(3ob) level of theory, and the averaged retinal polyene chain helicity from BR trajectories are also shown.



all-*trans*, 15-*anti*



13-*cis*, 15-*syn*

Figure 5.10: Two retinal isomer models in vaccum for DFTB3(3ob) optimization. The C_β of lysine is saturated using hydrogen atoms.

an averaged HCCH torsional angle of 179° of the all-*trans*,15-*anti* retinal in ChR2-C128T was obtained. The averaged C-C13=C14-C and C-C15=N-C torsional angles of 179° and 184°, respectively, further indicates that the all-*trans* retinal hydrogen-out-of-plane (HOOP) twist is negligible in ChR2-C128T dark state.

This result is controversial with that of ChR2-WT by Becker-Baldus et al. [75]. The authors measured a HCCH angle of 158°±2° using the solid-state magic-angle spinning (MAS) NMR enhanced by dynamic nuclear polarization (DNP) technique (DNP-enhanced MAS NMR technique) at around 100 K. This angle implies a significant HOOP twist, and it is conserved (156°±4°) in the P_1^{500} intermediate. This would lead to a scenario that no HOOP band is observable in the IR difference spectrum.

However, in the time-resolved IR difference spectra of ChR2-WT and ChR2-E123T at room temperature by Heberle group, HOOP vibration positive bands in P_1^{500} intermediate at 986 cm^{-1} and 988 cm^{-1}, respectively, are observed [83, 107]. There are two possible explanations for the appearance of the HOOP band: (i) The retinal is planar in dark state, twisted in P_1^{500}; (ii) The retinal is already twisted in dark state, and the twist extent is increased in P_1^{500}.

As we can see, the DNP-enhanced MAS NMR study and the IR study of ChR2 are completely conflict. The discrepancy might lies in the reliability of the techniques on ChR2. All in all, there is no clear picture of retinal conformation in ChR2 dark state.

Relatively, the current simulations of ChR2-C128T bound with 13-*cis* retinal sampled an averaged HCCH angle of 200°, which means the retinal is 20° twisted. This torsion agrees with the solid-state NMR study results on ChR2-WT (158°), BR (164°), and green proteorhodopsin (GPR, 161°), see Table 5.7. However, all these NMR studies were about all-*trans* retinal.

As can be known from Figure 5.7 and Table 5.7, the DFTB3(3ob) method overestimates the polyene chain bond conjugation, especially in the region close to NH^+. Although the introduction of protein point charges could alleviate this problem, we do not know to which extent it could be. This study shows averaged C14-C15 bond length of 1.42 Å, the same as the NMR value for GPR. However it is shorter than the ideal single C–C bond length of around 1.46 Å, as found in the crystal structures (see Table 5.7).

Table 5.7: Averaged C13=C14, C14-C15, C15=N bond length (Unit: Å); averaged torsion around C13=C14, C15=N (unit: degree); averaged H-C14-C15-H torsion (unit: degree) of the retinal in ChR2-C128T and BR. The time scale of each simulation is 1 ns. The values of the retinal optimized in vaccum at DFTB3(3ob) level of theory, and of the retinal in BR, C1C2, green proteorhodopsin (GPR) crystal structures, and some experimental values are also listed.

		bond length			torsional angle[a]		
	retinal isomer	C13=C14	C14-C15	C15=N	C13=C14	C14-C15	C15=N
ChR2-C128T	all-*tans*	1.39	1.42	1.32	179	179	184
	13-*cis*	1.39	1.42	1.32	15	200	12
BR	all-*trans*	1.39	1.42	1.32	189	184	184
isolated retinal	all-*trans*	1.41	1.40	1.34	180	180	180
	13-*cis*	1.41	1.40	1.34	0	180	0
BR (PDB code 1C3W)	all-*trans*	1.36	1.46	1.34	203	-	197
BR (PDB code 1X0S)	13-*cis*	1.37	1.48	1.42	2	-	4
C1C2 (PDB code 3UG9)	all-*trans*	1.34	1.46	1.47	180	-	167
GPR (PDB code 4JQ6)	all-*trans*	1.35	1.47	1.35	180	-	188(200)[b]
ChR2 EXP [75]	all-*trans*	-	1.51	-	-	158	-
BR EXP [211]	all-*trans*	-	-	-		164	-
GPR EXP [212]	all-*trans*	-	1.42	-	-	161	-

[a] C13=C14: C12-C3=C14-C15 torsional angle; C14-C15: H-C14-C15-H torsional angle; C15=N: C14-C15=N-C_{ε} torsional angle.

[b] Value in the parentheses is the C15=N torsional angle of the retinal in another monomer.

5.2.2 Active Site Structural Motifs

Unlike the classical force field which prefers a salt-bridge between $RSBH^+$ and negatively charged E123 side chain, the DFTB3/MM dynamics simulations sample heterogeneous active sites for both all-*trans*, 15-*anti* and 13-*cis*,15-*syn* retinals. The heterogeneity could be interpreted by (i) the coexistence of three $RSBH^+$ hydrogen bonding patterns in a rather flexible protein environment (Figure 5.11), and (ii) the coexistence of two E123 side chain conformations (Figure 5.2).

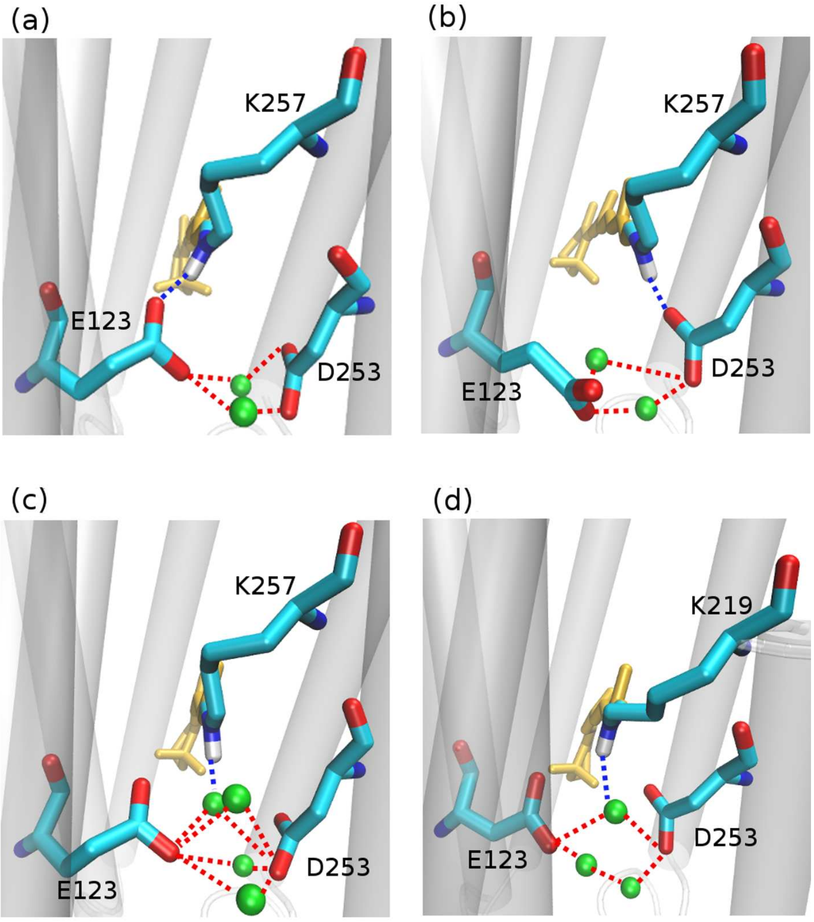


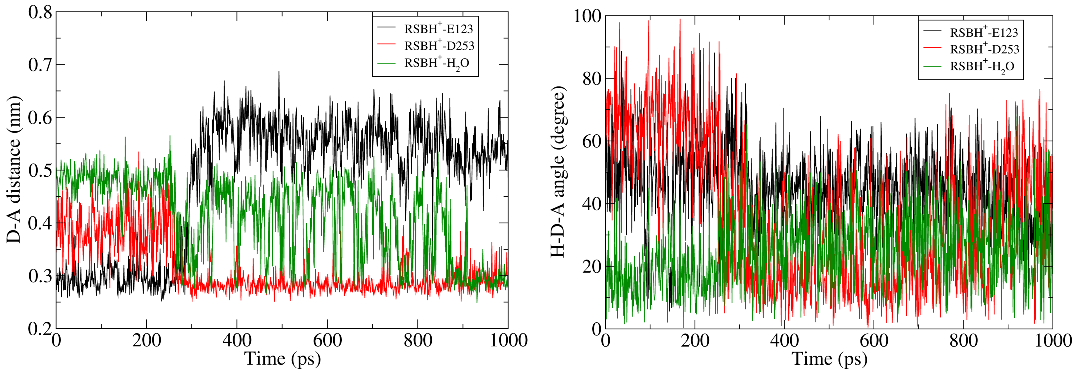
Figure 5.11: The hydrogen bonding patterns sampled in the present study: (a) -RSBH$^+\cdots{}^-$O-(E123), (b) -RSBH$^+\cdots{}^-$O-(D253), and (c) -RSBH$^+\cdots$OH$_2$. (d) Active site of BR: a pentaganal cluster is formed between the RSBH$^+$, D85, D212, and three water molecules. The dashed lines depict hydrogen bonds, and the water molecules are shown as green balls.

RSBH$^+$ Hydrogen Bonding Patterns

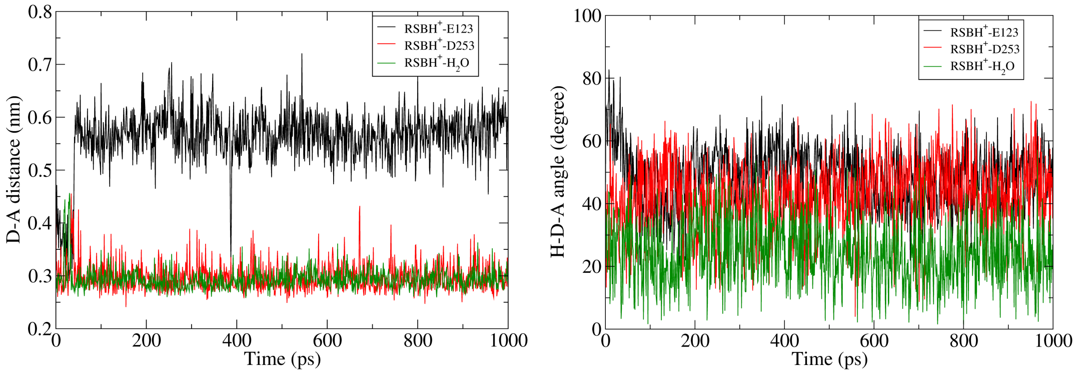
The same criteria as in Chapter 4 were adopted to assign a HBond: (i) the proton D-A distance is smaller than 0.35 nm; (ii) the frontier H-D-A angle is smaller than 30°. Three hydrogen bonding patterns were sampled in ChR2-C128T mutant, indicating a direct interaction between the RSBH$^+$ and negatively charged E123 or D253 side chains (Figure 5.11 (a), (b)) or between RSBH$^+$ and a water molecule (Figure 5.11 (c)). The three distinct hydrogen bonding motifs are defined as -RSBH$^+\cdots{}^-$O-(E123), -RSBH$^+\cdots{}^-$O-(D253), and -RSBH$^+\cdots$OH$_2$. Transitions between any two of the three hydrogen bonding patterns are observed in the trajectories. This is depicted by the temporal evolution of the D-A distances and H-D-A angles as shown in Figure 5.12. The results of the remaining trajectories are shown in Appendix A, Figure A.2 and Figure A.3.

According to the present simulations, the -RSBH$^+\cdots$OH$_2$ motif occupies 67% of the 14 ns simulation time. This motif dominates over RSBH$^+\cdots{}^-$O-(E123) (14%) and -RSBH$^+\cdots{}^-$O-(D253) (19%) motifs, indicating that a water molecule could play the role of prominent hydrogen bond partner of RSBH$^+$.

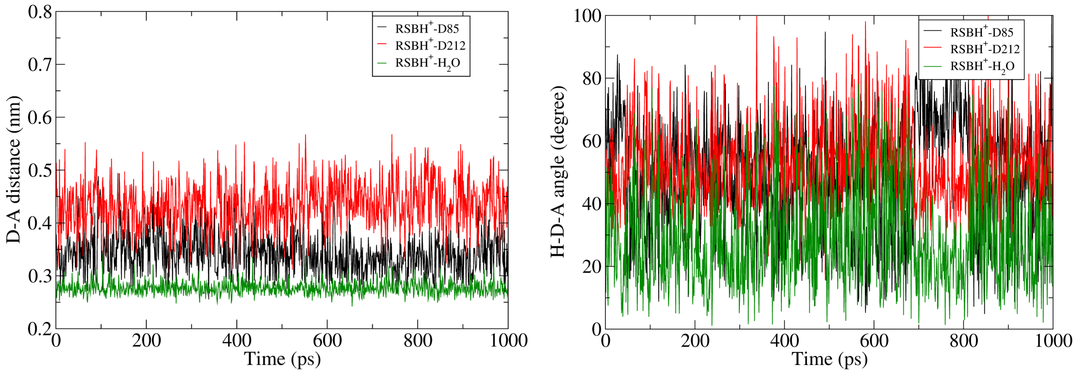
A comparison with BR is informative, especially with regard to the water arrangement within the active site. Dark state BR active site contains three water molecules forming a pentagonal cluster with the D85 and D212 side chain, where one water molecule is involved in a stable hydrogen bond with the RSBH$^+$ (Figure 5.11 (d)). This setup is well preserved in our QM/MM MD simulation (Figure 5.12 (e, f)). On the other hand, in the case of ChR2-C128T mutant, all three sampled structural motifs feature two water molecules bridging E123 and D253 side chains (Figure 5.11 (a–c)), and two additional water molecules are involved (i.e. four water molecules in total) when switching to the -RSBH$^+\cdots$OH$_2$ motif (Figure 5.11 (c)). In contrast to BR pentagonal cluster, the structure involving two water molecules was defined as the ChR2 half-barrel, the



(a) ChR2-C128T: all-*trans*,15-*anti* **(b)** ChR2-C128T: all-*trans*,15-*anti*



(c) ChR2-C128T: 13-*cis*,15-*syn* **(d)** ChR2-C128T: 13-*cis*,15-*syn*



(e) BR: all-*trans*,15-*anti* **(f)** BR: all-*trans*,15-*anti*

Figure 5.12: Stability of hydrogen bonding patterns: (a, c, e) the hydrogen bond D-A distance of ChR2-C128T bound with all-*trans*,15-*anti* retinal, 13-*cis*,15-*syn* retinal, and of BR with all-*trans*,15-*anti* retinal, respectively; (b, d, f) the hydrogen bond H-D-A angle of ChR2-C128T bound with all-*trans*,15-*anti* retinal, 13-*cis*,15-*syn* retinal, and of BR with all-*trans*,15-*anti* retinal, respectively.

structure involving four water molecules as the ChR2 barrel.

E123 Side Chain Conformations

The QM/MM simulations sample two E123 side chain conformations: (i) towards the cytoplasmic side (E123-upward), or (ii) towards the extracellular side (E123-downward) (Figure 5.2). This could be characterized by the dihedral angle around the E123 C_β-C_γ bond: a dihedral above 240° corresponds to the E123-upward conformation, while below 240° corresponds to the E123-downward conformation. About 81% of the 14 ns simulations sampled the E123-downward conformation.

In order to evaluate the sampling, the free energy profiles for ground-state transitions between the two E123 conformations were calculated at the MM level through well-tempered metadynamics method [139, 140, 145]. The dihedral angle around the C_β-C_γ bond was used as the collective variable, i.e. the reaction coordinate. The parameters were set as follows: Gaussian height of 0.2 kcal/mol, Gaussian width of 0.2 radian, Gaussian depositing frequency of 2 ps (1000 time steps), bias factor of 4, and temperature of 300 K. The simulation time scale is 100 ns. The convergence of the simulations was evaluated by comparing the free energy as a function of the collective variable at different times, as shown in Figure 5.13. The similar profile obtained after 85 ns indicates a good convergence of the simulations.

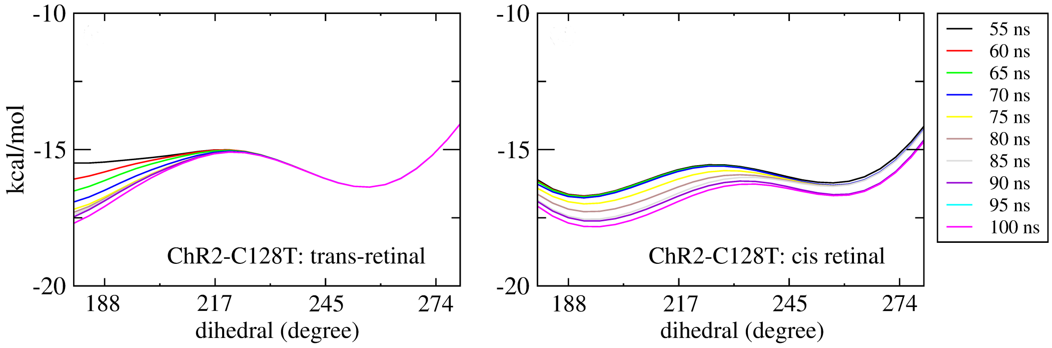


Figure 5.13: The free energy as a function of the dihedral around E123 C_β-C_γ bond, every 5 ns along the 100 ns well-tempered metadynamics simulation.

The obtained free energy profiles are shown in Figure 5.14. As can be seen that, in the complex bound with all-*trans*,15-*anti* retinal, the calculated free energy barrier from E123-upward to E123-downward is ca. 1.2 kcal/mol, while it is ca. two times higher for the reverse transition. In the case of 13-*cis*,15-*syn* retinal, the energy barrier is lower with respect to the all-*trans*,15-*anti* retinal bound complex. However, the barrier is still ca. two times higher from E123-downward to E123-upward than that of the reverse transition. This is quite consistent with our QM/MM simulations, where ca. 81% of the sampled structures correspond to the E123-downward conformation.

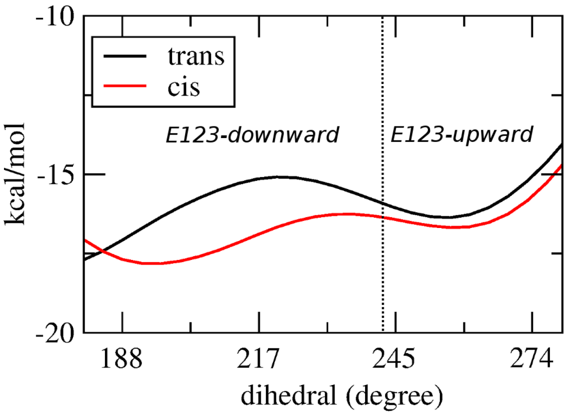


Figure 5.14: The free energy as a function of the dihedral around E123 C_β-C_γ bond. Black line: ChR2-C128T bound with all-*trans*,15-*anti* retinal, red line: ChR2-C128T bound with 13-*cis*,15-*syn* retinal. The dihedral angle block corresponding to E123-upward and E123-downward conformations are separated by a dotted line.

In conclusion, the ChR2-C128T active site is quite heterogenous compared with the rigid BR active site. The flexible water molecules and E123 side chain conformations result in three RSBH$^+$ hydrogen bonding patterns, which might contribute to the characteristic absorption spectrum of the protein. Such high flexibility determines that both E123 and D253 are potential RSBH$^+$ proton acceptors in the photocycle.

5.2.3 Absorption Spectra

The experimental absorption spectrum of dark adapted state ChR2-C128T is characterized by one distinct maximum at 480 nm (2.58 eV) and one small shoulder at ca. 450 nm (2.75 eV) [106]. In contrast, ChR2-WT spectrum shows three major peaks at 473 nm (2.62 eV), 442 nm (2.80 eV) and 414 nm (2.99 eV) [69]. The well studied BR with exclusively all-*tans*,15-*anti* retinal has a spectrum constituted by a single peak centered at ca. 570 nm (2.17 eV) [70, 71]. Therefore, all three types of rhodopsins absorb in visible range, with a considerable hypsochromic shift in ChR2-WT and ChR2-C128T relative to BR. The C128T mutation leads to some effects on the absorption spectrum.

As discussed in the above section, ChR2-C128T has active site structural motifs of -RSBH$^+\cdots{}^-$O-(E123) and -RSBH$^+\cdots{}^-$O-(D253), in which the anionic side chains of E123 and D253 could stabilize the electronic ground state S_0 and enlarge the electronic energy gap with respect to BR. This has been suggested to be the origin of the hypsochromic shift of ChR2 relative to BR [66]. However, the former QM/MM study reported that the overall electrostatic interaction of the retinal with the protein environment is responsible for a large part of the shift, while the contribution of the hydrogen bonding network is moderate [196].

The single peak property of BR absorption spectrum was explained by its single active site structure, the three-water pentagonal cluster, representing the global ground state minimum. The harmonic oscillations of the global minimum due to thermal effects leads to homogeneous spectral broadening [162]. On the other hand, ChR2-C128T active site shows several local minima of low energy barriers, which could be one explanation of the inhomogeneous broadening of the absorption spectrum.

In the current study, the electronic excitation energies of twelve ChR2-C128T geometries were calculated using the SORCI method. The result is shown in Figure 5.4 (Please see the result of the 50 ChR2-WT geometries in our paper [193], calculated by Franziska E. Beyle.), from which it can be seen that the minimized geometries do not collapse into a global ground state energy minimum, but basically maintain the initial structural motif. The resulting spectrum is centered at 2.80 eV and 2.95 eV and is 0.36 eV broad. As a comparison, we calculated the excitation energy of an optimized BR snapshot from the QM/MM trajectory. The value of 2.35 eV reveals a bathochromic shift of 0.45 eV with respect to ChR2-C128T, in good agreement with the experimental shift of 0.41 eV.

The SORCI absorption maxima (λ_{max}) of ChR2-C128T and BR are ca. 0.25 eV hypsochromic shifted, comparing with the experimental spectra. This could be explained in terms of the applied methodology: the electric field due to the classical MM point charges surrounding the QM region tends to be overestimated, which could be (at least partially) moderated by using a polarizable force field, especially considering the presence of of ionized groups in the protein interior surrounding the retinal [213]. In case of BR, a bathochromic shift of 0.2–0.3 eV is expected by including polarization effects [75, 206, 214, 215]. Nevertheless, the considerably higher computational cost required by polarizable force fields compared to classical ones makes this option unpractical for the present study, which is focused on extensive simulations of the different ChR2 structural motifs.

Since ChR2-C128T active site is inherently flexible, an exhaustive structure ensemble is necessary in order to properly describe the structural dynamical effects on the absorption spectrum. Figure 5.15 shows the OM2/MRCI results based on 14 000 snapshots (8000 snapshots of all-*trans*,15-*anti* retinal, and 6000 of 13-*cis*,15-*syn* retinal) extracted from QM/MM simulations. The results of BR from 1000 snapshots are also shown. During data analysis we should keep in mind that a hypsochromic shift of ca. 0.3–0.4 eV has to be considered when passing from SORCI to OM2/MRCI.

ChR2-C128T bound only with all-*trans*,15-*anti* retinal shows a hypsochromic shift of 0.47 eV with respect to BR. This value is quite close to that of the experiment (0.41 eV) within the limits of the OM2/MRCI accuracy. The 13-*cis*,15-*syn* retinal and the all-*trans*,15-*anti* retinal bound complexes have very similar λ_{max}. This indicates that both retinal configurations could equally be responsible of the absorption spectral shape. Hence, the assumption that the two retinal configurations coexist in ChR2 dark state cannot be verified by examining the aborption spectrum.

In order to get an insight into the composition of the absorption sepctrum, the overall histograms of the all-*trans*,15-*anti* retinals bound complex, of the 13-*cis*,15-*syn* retinals bound complex, and of the mixed all-*tans*,15-*anti* and 13-*cis*,15-*syn* retinals bound complex were deconvoluted (Figure 5.16). This allows the analysis of the contribution of the three RSBH$^+$ hydrogen bonding patterns and the two E123 side chain conformations to the spectrum.

The -RSBH$^+\cdots$OH$_2$ and the -RSBH$^+\cdots{}^-$O-(D253) patterns give almost the same λ_{max}, and are responsible for the long wavelength part of the spectrum regardless of the retinal configuration (Figure 5.16 (a, c, e)). The -RSBH$^+\cdots{}^-$O-(E123) pattern gives rise to a hypsochromic shift of ca. 0.10 eV with respect to the -RSBH$^+\cdots$OH$_2$ or the -RSBH$^+\cdots{}^-$O-(D253) pattern. In the experimental spectrum, the small shoulder is 0.17 eV hypsochromic shift relative to the λ_{max}. The calculated results are comparable to the experiment taking into account the accuracy of the applied methodology. Therefore, the three hydrogen bonding patterns might be the structural origin determining the absorption spectrum shape of ChR2-C128T. Note that the his-

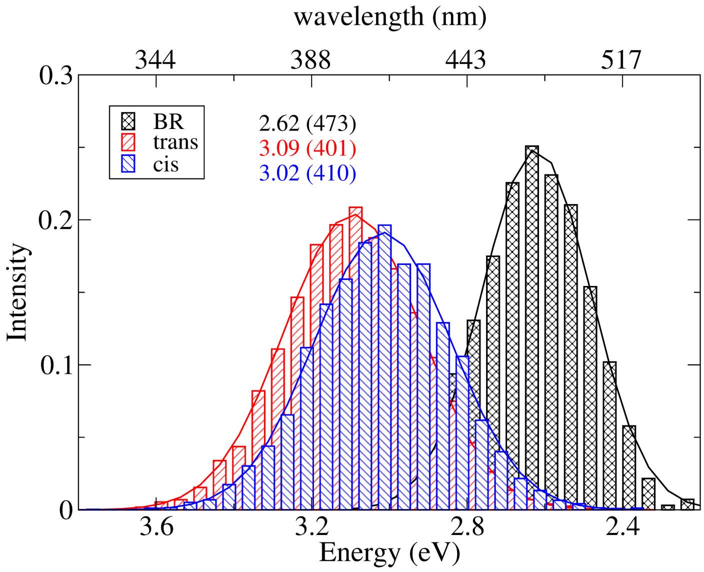


Figure 5.15: Comparison of simulated absorption spectra of ChR2-C128T and BR. The histograms are based on OM2/MRCI calculations of geometries from QM/MM trajectories. Gaussian functions are used to convolute the spectra and the corresponding maxima are listed in the same color as the histograms in eV (nm). The fitted parameters are listed in Appendix A, Table A.2. Notation: *"trans"* refers to snapshots bound with all-*tans*,15-*anti* retinal; *"cis"* refers to snapshots bound with 13-*cis*,15-*syn* retinal.

togram of the -RSBH$^+ \cdots {}^-$O-(E123) pattern of the 13-*cis*,15-*syn* retinals bound protein is not shown in Figure 5.16 (e), because only 3 out of 6000 snapshots show this structural motif.

From the perspective of the E123 side chain conformation, E123-upward shows a λ_{max} hypsochromic shift of ca. 0.1 eV with respect to E123-downward, comparable with the experimental value (0.17 eV). This suggests that the E123 side chain conformation might also be the key structural reason for explaining the absorption spectrum.

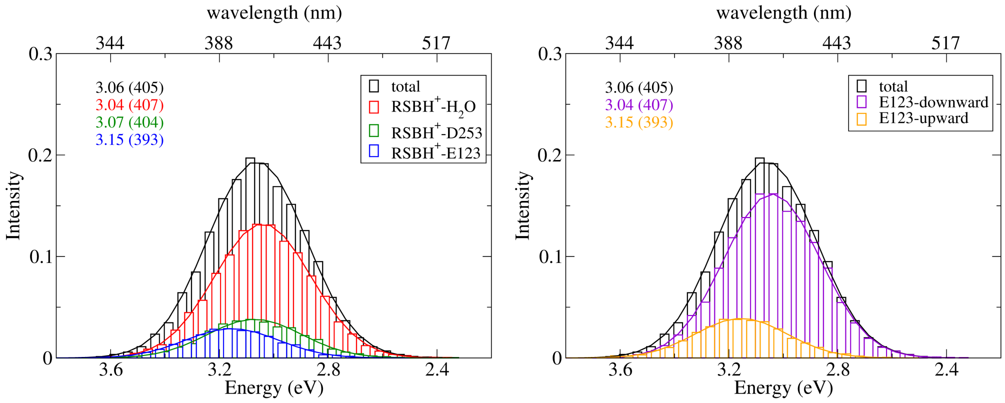
5.2.4 Concluding Remarks

Our QM/MM MD simulations suggest a heterogeneous active site structure, consisting of several local minima with low barriers being sampled at room temperature on a nanosecond time scale. The heterogeneous active site explains the one maximum–one shoulder nature of ChR2-C128T absorption spectrum, which is independent of the retinal configuration. Not only the distinct RSBH$^+$ hydrogen bonding patterns, but also the different E123 side chain conformations contribute to the structure of the spectrum.

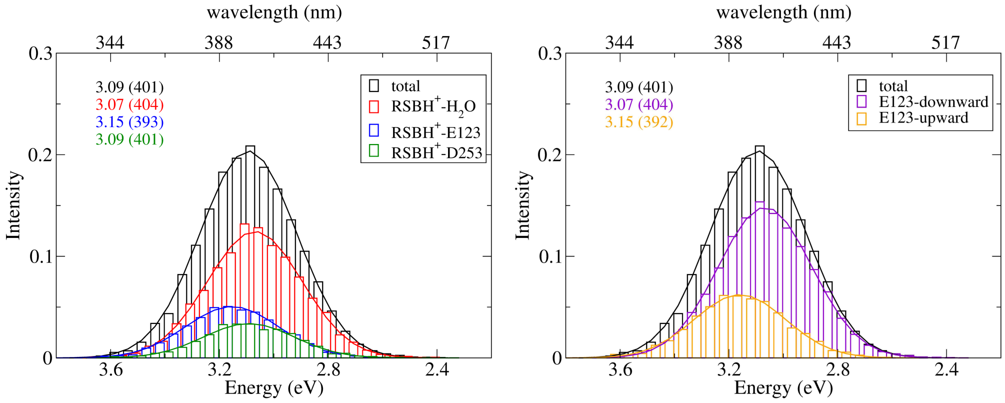
The applied computational methodology has been used frequently in recent years to compute absorption energies for various retinal proteins with a good success rate. This suggests that the effect of different retinal conformations and protein structures on computed excited states energies is well reproduced. Since the excitation energies of BR and ChR2-C128T agree quite well with experiment, we conclude that the sampled structural models in our simulations represent the ChR2-C128T structure probed in experiment with sufficient accuracy.

The 13-*cis*,15-*syn* retinals bound complex show the same structural motifs as the all-*trans*,15-*anti* retinals bound complex. The *cis*-retinal does not change the absorption spectrum with respect to the *trans*-retinal. This implies that the 13-*cis*,15-*syn* retinals cannot be excluded out from the ChR2 dark state. However, this point of view has now been clarified experimentally [75]. As to the other controversial question, i.e the retinal conformation, our DFTB3/MM simulations sampled a planar all-*trans*,15-*anti* retinal, but a twisted 13-*cis*,15-*syn* retinal with an averaged HCCH angle of 200°.

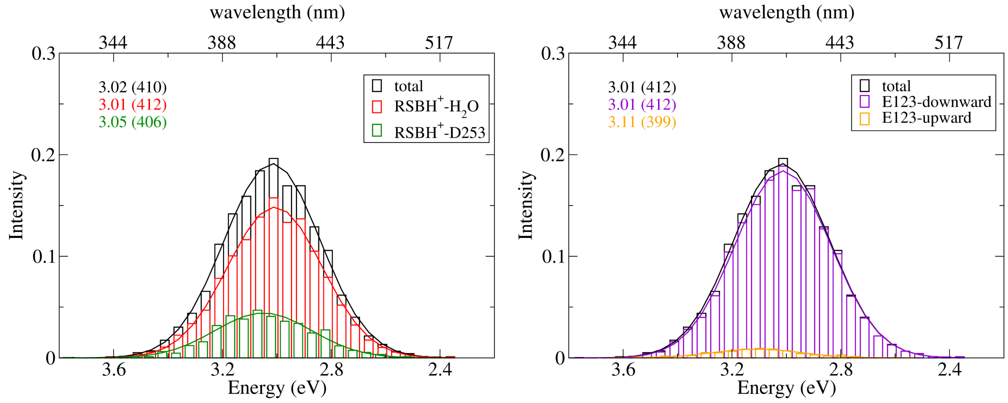
In the current study we present a very flexible active site using quantum mechanical methods. Estimating from the relative sampling amount of the structural motifs, there is a free energy difference within a few kcal mol^{-1} between different motifs. As a consequence, very small inaccuracy in the quantum chemical descriptions can shift the sampling amount of the individual states. The required accuracy is definitely higher



(a) mixed all-*trans*,15-*anti* & 13-*cis*,15-*syn* **(b)** mixed all-*trans*,15-*anti* & 13-*cis*,15-*syn*



(c) all-*trans*,15-*anti* **(d)** all-*trans*,15-*anti*



(e) 13-*cis*,15-*syn* **(f)** 13-*cis*,15-*syn*

Figure 5.16: A deconvolution based on the assignment of structural motifs is proposed for ChR2-C128T with mixed all-*trans*,15-*anti* and 13-*cis*,15-*syn* retinals (a, b), with all-*trans*,15-*anti* retinals (c, d), and with 13-*cis*,15-*syn* retinals (e, f). The contributions of the three RSHB$^+$ hydrogen bonding patterns (a, c, e) and the two E123 side chain conformations (b, d, f) are depicted. The histograms are based on OM2/MRCI calculations of geometries from QM/MM trajectories. Gaussian functions are used to convolute the spectra and the corresponding maxima are listed in the same color as the histograms in eV (nm). The fitted parameters are listed in Appendix A, Table A.2. Here, "total" denotes the histogram from all the snapshots. "RSBH$^+$-E123"corresponds to the -RSBH$^+\cdots{}^-$O-(E123) hydrogen-bonding pattern; "RSBH$^+$D253" corresponds to the -RSBH$^+\cdots{}^-$O-(D253) hydrogen-bonding pattern; "RSBH$^+$-H$_2$O" corresponds to the -RSBH$^+\cdots$OH$_2$ hydrogen-bonding pattern. In (e) the histogram of -RSBH$^+\cdots{}^-$O-(E123) is not shown, because only 3 out of 6000 snapshots show this motif.

than the chemical accuracy of 1 kcal mol^{-1}, which is hard to achieve for standard electronic structure methods. Hence, the picture we present here has a qualitative meaning, rather than quantitative. Such highly flexible active sites pose a clear challenge to computational methods, and the results should not be over-interpreted.

Chapter 6

Chromophore Photoisomerization in Channelrhodopsin-2

The photoreaction of ChRs is triggered by the light induced isomerization of its chromophore—retinal (Figure 1.6), bound to Lys257 in ChR2 via a Schiff base linkage. The retinal isomerization mechanism in ChR2 remains an unresolved problem.

A series of papers reported the retinal protein photodynamics, covering bacteriorhodopsin (bR) [216–218], vision rhodopsin (Rh) [219, 220], *Anabaena* sensory rhodopsin (ASR) [221], and channelrhodopsin (ChR) [77, 222]. These investigations deliver preliminary mechanistic scenarios of retinal photoisomerization in different proteins. Warshel and co-workers [223–225] proposed a space saving "aborted bicycle pedal" mechanism in vision Rh, involving a concerted rotation of two double bonds in opposite directions. One rotation is completed while the other one is aborted after conical intersection. A more recent study by Frutos et al. [226] proposes an "asynchronous bicycle pedal" isomerization mechanism in bovine Rh, in which the two double bond rotations are not concerted. In the "Hula-Twist" mechanism by Liu et al. [227, 228], the translocation of a single C-H unit is accompanied by a simultaneous rotation of two connected single and double bonds. This is thought to be a volume-conserving motion in the protein cavity. Strambi et al. [229] and Schapiro et al. [221] reported a complete (360°) unidirectional rotation of retinal in ASR. Altoè et al. [230] proposed the "double bicycle pedal" mechanism in bR, which is characterized by the concurrent rotation of three adjacent double bonds C11=C12, C13=C14, C15=N in alternating directions. The twisting around C13=C14 is compensated by the sum of the opposite twisting around C11=C12 and C15=N. The rest part of the retinal remains nearly intact.

However, studies of ChR photochemistry are still much lack. Our former study reveals a flexible active site of ChR2, which is much more complicated than the rigid pentagonal active site of BR [193]. This means that the mechanism of photochemical reactions in ChR2 might not be directly related from that in bR or in other Rh, because the environment surrounding the retinal significantly affects its photochemical properties. The Garavelli group [231] used *ab initio* CASPT2//CASSCF methods to investigate the effects of an acetate counterion on the photochemistry of various reduced retinal models. The authors found that the positions/orientations of the counterion affect the ground and excited electronic states stability, photoisomerization rates and stereospecificity. Virshup et al. [232] studied the full retinal chromophore *cis*→*trans* isomerization using semiempirical floating occupation molecular orbital configuration interaction (FOMO-CI) method. In gas phase and methanol solution environment, only a small amount of the 11-*cis* isomer are isomerized to all-*trans* and various di-*cis* isomers. By contrast, when the retinal is in rhodopsin environment, more than half of the product is all-*trans* isomer, and no di-*cis* isomer is observed.

The present work considers the ChR2 active site structures with various Schiff base hydrogen bonding patterns [193]. A combined QM/MM strategy is applied, where the QM calculations use an *ab initio* CASPT2//CASSCF scheme. The CASPT2 method includes the dynamical electron correlation, which is necessary for a correct description of the geometrical and electronic structures of retinal in the protein environment. The CASPT2//CASSCF/MM method provides balanced description of ground and excited states, leading to reliable reaction pathways. This scheme has been successfully used for studying the photochemistry of vision

Rh and ASR [221, 229, 233, 234].

The existence of the 13-*cis*,15-*syn* retinal in the ChR2 dark state is still controversial [68, 75, 107], and the contribution of the photoreaction of this isomer to the ChR2 photocycle remains elusive. Whether the 13-*cis*,15-*syn* retinal photocycle is functionally relevant becomes another open question. In BR, the IR difference spectrum resolves the band pair 1186(-)/1179(+) indicating the photoreaction of the 13-*cis*,15-*syn* isomer [108]. In ChR2-WT, similar bands at 1183(-)/1177(+) are observed [62, 68]. However, Lórenz-Fonfría et al. [107] reports through step-scan FTIR that the 13-*cis*,15-*syn* photocycle recovers on a time scale much faster than proton transfer reactions, implying that the photocycle of the 13-*cis*,15-*syn* retinal is functionally irrelevant in ChR2. In the present study only the photoisomerization of the all-*trans*,15-*anti* retinal isomer is considered.

6.1 Computational Methods

The Protein Model

Three models with different protonated Schiff base (RSBH$^+$) hydrogen bonding patterns were built for ChR2-WT bound with all-*trans* retinal according to our previous study [193]. The three models are: (i) RSBH-E123 – RSBH$^+$ forms hydrogen bond with E123 side chain; (ii) RSBH-D253 – RSBH$^+$ forms hydrogen bond with D253 side chain; (iii) RSBH-H$_2$O – RSBH$^+$ forms hydrogen bond with one water molecule. Each model includes one monomer of the protein and the water molecules within 6 Å of the retinal. The other monomer of the protein, the solvent, and the lipid bilayers are not considered. This is for the sake of the convenience to set up the simulations, and of saving computational time.

The RSBH$^+$ counterions E123 and D263 are deprotonated; the remaining titrable residues are assigned standard protonation states as in the Charmm 22 force field, except E90 and D156 which are set to be protonated. Three chloride ions were added to keep the system neutral.

The QM/MM Scheme

In the present study, the QM/MM methodology with hydrogen-link atom scheme was used to describe the retinal photoisomerization in ChR2. The Hamiltonian describing the QM/MM scheme is as follows:

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} - \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{q_j}{r_{ij}} + \sum_{i=1}^{N}\sum_{j=1}^{m}\frac{Z_i q_j}{R_{ij}} + E_{QM/MM}^{vdW} + E_{QM/MM}^{bonded} \tag{6.1}$$

where $\hat{H}_{QM}$ describes the QM segment, $\hat{H}_{MM}$ the MM segment. The remaining terms represent the interactions between the QM and MM segments: (1) The third term–the electrostatic interactions between QM electrons and MM point charges (q_j); *i* is the index of the electron, *n* the total number of electrons, *j* the index of the MM point charge, *m* the total number of MM point charges, r_{ij} the electron–MM point charge distance. (2) The fourth term–the electrostatic interactions between QM nuclei and MM point charges (q_j); *N* is the total number of atoms, Z_i nucleus charge, R_{ij} nucleus–MM point charge distance. (3) The fifth term–short range van der Waals interactions. (4) The last term–the additionally parametrized potentials required to describe QM/MM boundary geometry correctly.

The QM/MM boundary was placed at the N-C$_\varepsilon$ bond connecting the K257 side chain and the retinal, as shown in Figure 6.1. This choice ensures that the whole retinal is included in the QM region, and the QM segment has a moderate size. The rest of the system was treated as the MM region and was described by the Charmm22 force field.

The QM wavefunction is polarized by the MM point charges. In contrast, the MM point charges remain constant during the calculation. The frontier QM atom N is saturated with a hydrogen link atom (LAH). The link atom is fixed at a distance of 1 Å from N and kept along the N-C$_\varepsilon$ bond throughout all calculations. The LAH is treated as part of the system and it may interact with the MM point charges in order to describe the frontier correctly.

In the Charmm22 force field, each amino acid residue has an integral net charge. For instance, the lysine has a net charge of +1 representing a protonated residue. In the current QM/MM framework, the atomic charges of K257 are changed (Table 6.1) to avoid an over-polarization of the QM segment by the MM segment.

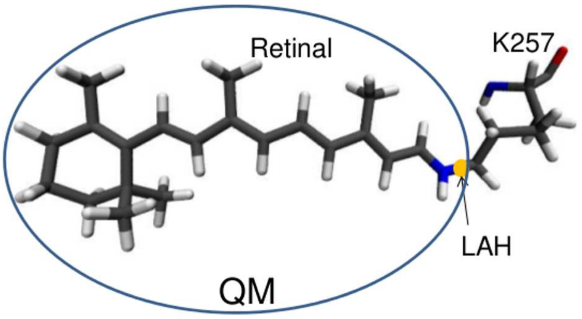


Figure 6.1: The QM/MM partitioning in the current calculations. The QM/MM boundary was placed at the N-C_ε bond connecting the K257 side chain and the retinal. The frontier QM atom N is saturated with a hydrogen link atom (LAH).

Table 6.1: The values of the reparametrized QM/MM point charges for K257.

Atom	N	H_N	C_α	H_α	$C_{carbonyl}$	$O_{carbonyl}$	C_β	H_β	H_β
Charge	-0.33876	0.28384	-0.23086	0.15174	0.74324	-0.58026	-0.00026	0.04524	0.04524
Atom	C_γ	H_γ	H_γ	C_δ	H_δ	H_δ	C_ε	H_ε	H_ε
Charge	-0.10193	0.03699	0.03699	0.02698	0.03472	0.03472	0.0000	0.0000	0.0000

The point charges of C_ε and two H_ε atoms are set to zero. The initial charges are redistributed equally on the remaining MM atoms of K257.

The QM calculations were performed at the CASPT2//CASSCF/6-31G* level of theory (i.e. geometry optimization at the CASSCF level, energy evaluation at the CASPT2 level). The active space comprises the full π-system of the retinal (i.e. 12 electrons in 12 orbitals). During the optimization, the microiteration scheme [235] was applied, under which atomic positions of relaxed MM atoms were minimized for each new QM geometry. The geometry optimization and single point calculations were carried out with the Molcas 8.0 [235] and Tinker 6.3.2 [236] programs.

The computational scheme is: (*i*) The initial structures from previous DFTB3/MM trajectories were optimized in the ground state at the B3LYP and the state-average CASSCF (SA-CASSCF) levels, respectively. The purpose of this step is to obtain a good starting structure for the excited state calculations. The resulting structures correspond to the Frank-Condon (FC) geometries. (*ii*) Perform relaxed scan in the S_1 excited state along the C12-C13=C14-C15 torsion angle. The dihedral angle was varied in steps of 5° and kept constraint in each step. The SA-CASSCF wavefunction including two roots with equal weights was used. (*iii*) Calculate CASPT2 vertical excitation energies for the CASSCF geometries from steps (*i*) and (*ii*). The energetics was determined based on both single-state (i.e. CASPT2 calculation) and multi-state (i.e. MS-CASPT2 calculation) wavefunctions.

Calculation of IR Spectrum

The IR spectrum is calculated using the Fourier transform of time-correlation functions (FTTCF) method [237]. The CHARMM37b1 software [199] interfaced with the SCC-DFTB program [130–132] is applied for the calculations. Firstly, the system is geometry optimized using the Conjugate Gradient (CONJ) minimizer until a gradient threshold of 1×10^{-3} au is reached, and subsequently, using the Newton–Raphson (ABNR) minimizer with a gradient threshold of 1×10^{-5} au. Then the system is subjected to 100 independent heating MD simulations to generate slightly different starting coordinates for the production runs. Each simulation is 5 ps with the time step of 0.5 fs. The resulting structures are proceeded to production MD simulations of 20 ps of each with the time step of 0.5 fs. During the production run, the dipole moments needed to generate IR spectrum are collected at each time step. Subsequently, the Fourier transformation of the dipole moment autocorrelation is computed to get the spectrum. 100 independent MD simulations are performed, thus, the resulting spectrum is the average of the 100 separate spectra. in order to obtain a reliable intensity pattern.

6.2 Results and Discussion

6.2.1 Optimization of Ground State Geometries

In order to get an optimal structure before starting the excited state calculation, the three models were optimized in the ground state. A previous study showed that B3LYP functional is reliable for ground state equilibrium geometries of protonated retinal models in gas phase. Relatively, the CASSCF method does not result in that good ground state geometries as the B3LYP functional [238]. For the present models, the optimizations at both B3LYP and CASSCF levels are performed. The protein was frozen, and the water molecules within 6 Å of the QM atoms were relaxed during the optimization. The orientation of the retinal NH^+ moiety of the RSBH-E123 model and the RSBH-H_2O model changed slightly with respect to the initial structure (see Figure 6.2). The water molecules rearranged slightly as well. Nevertheless, the characteristic active site hydrogen bonding patterns were preserved for both models.

In the case of the RSBH-D253 model (Figure 6.2 (b)), the orientation of the retinal NH^+ moiety flipped from the D253 side chain towards one water molecule. Correspondingly, the hydrogen bond was rearranged from -RSBH$^+ \cdots$ $^-$(O)-D253 to -RSBH$^+ \cdots$ OH_2. In order to study all the three protein environments effects on retinal isomerization, the initial structure with -RSBH$^+ \cdots$ $^-$(O)-D253 HBond is directly used for the excited state calculations.

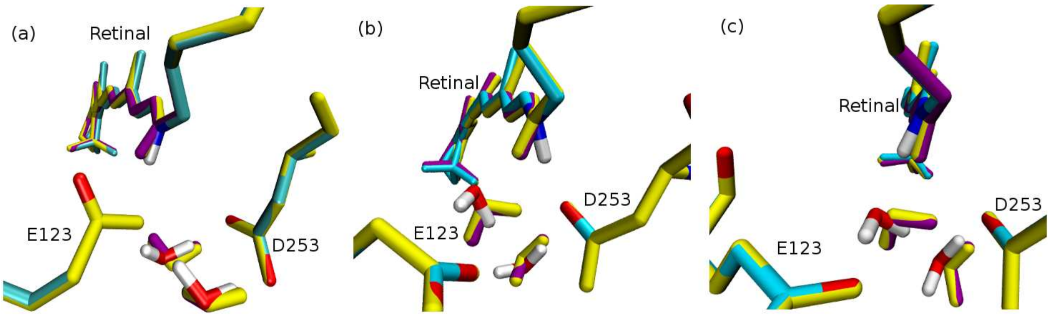
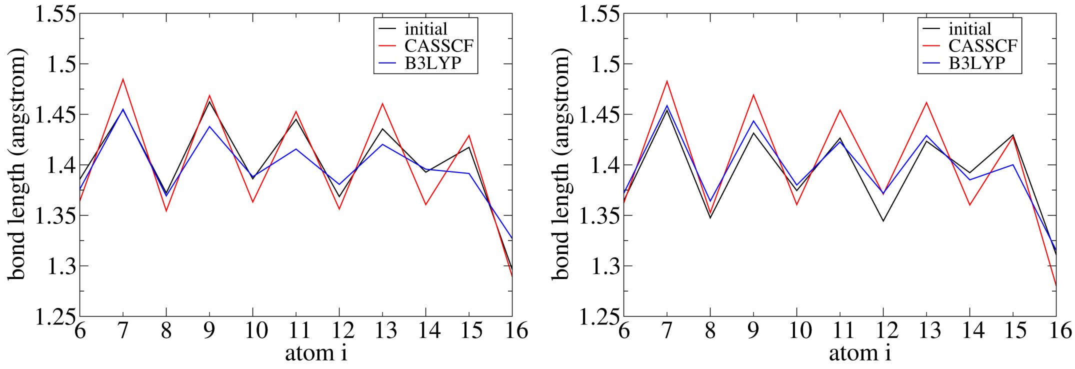


Figure 6.2: Superposition of the ground state CASSCF geometry (yellow), B3LYP geometry (purple) and the initial geometry (cyan) of ChR2-WT. (a) RSBH-E123 model, (b) RSBH-D253 model, (c) RSBH-H_2O model. The protein was frozen, and the water molecules within 6 Å of the QM atoms were relaxed during the optimization.

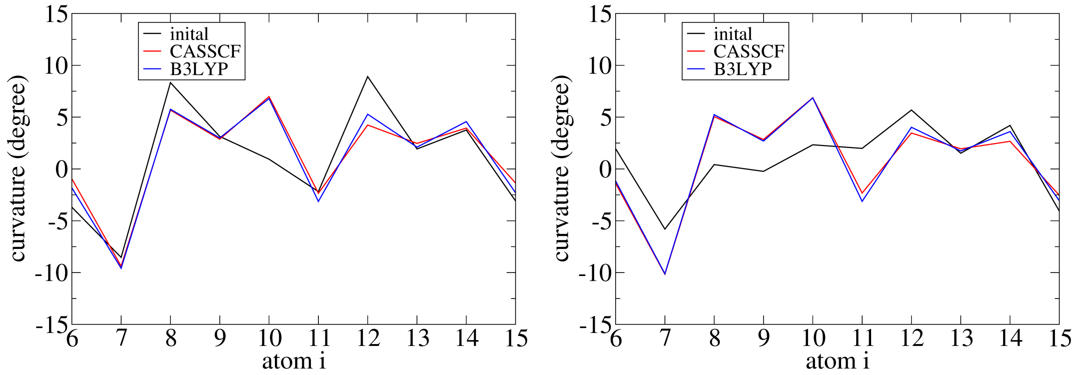
To further compare the B3LYP and CASSCF geometries, the bond length (Figure 6.3), curvature (Figure 6.4) and helicity (Figure 6.5) of the conjugated polyene chain of the retinal for the RSBH-E123 and RSBH-H_2O models were characterized. The curvature and the helicity were obtained with equations 5.1 and 5.2 in Chapter 5.

The bond length alternation of the B3LYP geometry is slightly smaller than that of the CASSCF geometry. This means that B3LYP provides a stronger retinal polyene chain conjugation than CASSCF. No significant changes in the backbone curvature and planarity are observed between B3LYP and CASSCF geometries. In conclusion, the B3LYP and CASSCF methods result in quite similar retinal backbone geometries for the RSBH-E123 and RSBH-H_2O models.



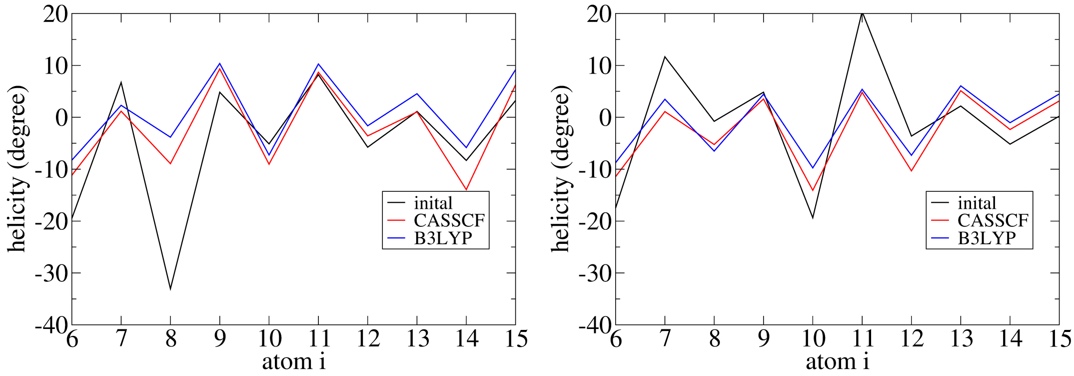
(a) ChR2-WT: RSBH-E123 model **(b)** ChR2-WT: RSBH-H_2O model

Figure 6.3: The retinal polyene chain bond length of the CASSCF, B3LYP, and initial geometries.



(a) ChR2-WT: RSBH-E123 model **(b)** ChR2-WT: RSBH-H_2O model

Figure 6.4: The retinal polyene chain curvature of the CASSCF, B3LYP, and initial geometries.



(a) ChR2-WT: RSBH-E123 model **(b)** ChR2-WT: RSBH-H_2O model

Figure 6.5: The retinal polyene chain helicity of the CASSCF, B3LYP, and initial geometries.

6.2.2 Calibration of Computational Parameters and Model Setup

IPEA Shift

In CASPT2 calculations, the use of the ionization potential–electron affinity (IPEA) shift parameter underestimates excitation energies by modifying diagonal elements of the zeroth-order Hamiltonian [239]. Previous studies on retinal chromophore analogues of visual rhodopsin in vacuo showed that for the CASSCF-optimized geometries, CASPT2/6-31G* methodology with IPEA shift = 0.0 au results in better excitation energies than the standard value of IPEA shift = 0.25 au [238, 240]. This is because the choice of IPEA shift = 0.0 au for the CASPT2//CASSCF/6-31* scheme leads to a significant cancellation of errors: CASSCF geometry results in blue-shifted excitation energy with respect to CASPT2 or B3LYP geometry; a medium-sized basis set like 6-31G* results in blue-shifted excitation energy with respect to larger basis sets; IPEA shift = 0.0 au results in red-shifted excitation energy relative to the standard IPEA shift. In the current study, the performance of IPEA shift = 0.0 au and IPEA shift = 0.25 au were evaluated for the retinal in the protein, in which three roots were included and the level shift (LS) parameter is 0.1 au. As shown in Table 6.2, IPEA shift = 0.0 au results in absorption maximum (λ_{max}) closer to the experimental value (λ_{max} = 2.62 eV) than IPEA shift = 0.25 au. This implies that for the current protein models, the inclusion of MM point charges does not alter the choice of IPEA shift = 0.0 au for the CASPT2//CASSCF/6-31* scheme . Therefore, in this thesis, CASPT2 calculations without IPEA shift parameter are applied to obtain the retinal isomerizaiton energy pathways of ChR2-WT.

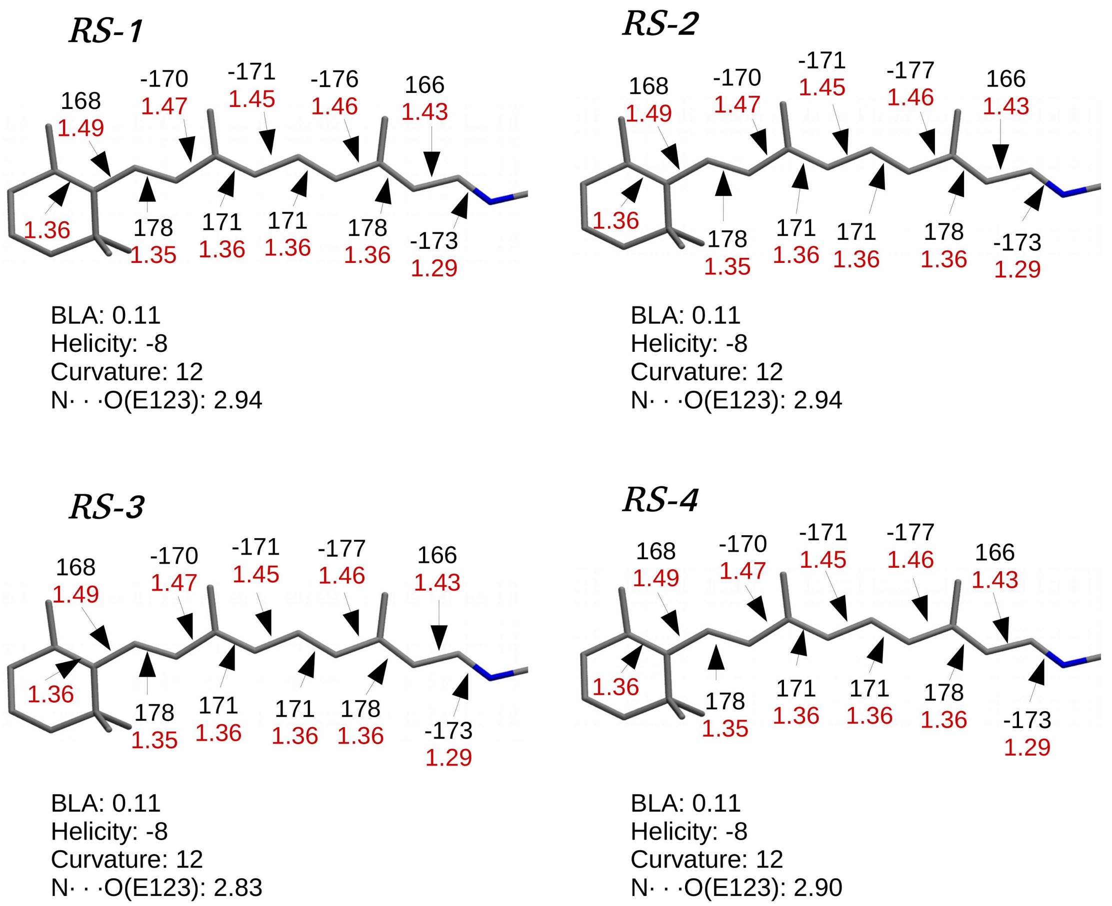
Table 6.2: CASPT2 and MS-CASPT2 vertical excitation energies, and the corresponding CASSCF energies for CASSCF FC geometries of RSBH-E123, RSBH-H_2O and RSBH-D253 models. The geometries were optimized in the ground state in which the protein was frozen, and the QM atoms (i.e the retinal) and water molecules within 6 Å of the QM atoms were relaxed. The energies of three roots based on IPEA shift = 0.0 au and based on IPEA shift = 0.25 au (values in the parentheses) are listed. The medium-sized basis set 6-31* was applied, LS = 0.1 au. The experimental absorption maximum is 2.62 eV (473 nm). Energy unit: eV.

Structure	State	E_{CASPT2}	$E_{MSCASPT2}$	E_{CASSCF}
RBSH-E123	S_0	0	0	0
	S_1	2.57 (3.05)	2.59 (2.93)	3.86 (3.91)
	S_2	3.47 (3.84)	3.84 (4.18)	4.45 (4.46)
RBSH-H_2O	S_0	0	0	0
	S_1	3.05 (3.52)	2.72 (3.16)	4.16 (4.22)
	S_2	3.26 (3.70)	4.05 (4.35)	4.70 (4.74)
RBSH-D253	S_0	0	0	0
	S_1	3.06 (3.53)	2.75 (3.16)	4.22 (4.22)
	S_2	3.33 (3.70)	4.13 (4.35)	4.75 (4.75)

Relaxed Scan Schemes

In order to determine the influence of relaxation of the retinal surrounding environment on spectroscopic properties and isomerization energy pathway, four relaxed scan schemes were employed for the RSBH-E123 model. *RS-1*: The QM atoms and water molecules within 6 Å of the QM atoms are free to relax. *RS-2*: The QM atoms, water molecules within 6 Å of the QM atoms and the side chain of K257 are free to relax. *RS-3*: The QM atoms, water molecules within 6 Å of the QM atoms and the side chains of K257, E123, D253 are free to relax. *RS-4*: The QM atoms, water molecules within 6 Å of the QM region,the side chains of K257, E123, D253 and of residues forming the retinal binding pocket (V86, E90, W124, T127, C128, I131, T159, P178, L182, W223, P227, P230, M255, S256, C259, W260) are free to relax. The rest of the protein is frozen.

The CASSCF FC geometries resulting from the four optimization schemes are shown in Figure 6.6. The retinal geometry is little affected by the relaxation of the binding pocket residues. The relaxation of K257 does not change the interaction of $RSBH^+$ with its surroundings with respect to the structure obtained by freezing the whole protein. The further relaxation of E123 and D253, or the whole binding pocket results in slightly stronger interaction between $RSBH^+$ and the carboxyl of E123, depicted by the decreased shortest distance between the nitrogen of $RSBH^+$ and the carboxyl oxygen of E123. Therefore, the positive charge of $RSBH^+$ is more strongly stabilized in geometries from *RS-3* and *RS-4* than those from *RS-1* and *RS-2*. Consequently, λ_{max} of *RS-3* and *RS-4* geometries are higher than those of*RS-1* and *RS-2* geometries (see Table 6.3).



RS-1
BLA: 0.11
Helicity: -8
Curvature: 12
N· · ·O(E123): 2.94

RS-2
BLA: 0.11
Helicity: -8
Curvature: 12
N· · ·O(E123): 2.94

RS-3
BLA: 0.11
Helicity: -8
Curvature: 12
N· · ·O(E123): 2.83

RS-4
BLA: 0.11
Helicity: -8
Curvature: 12
N· · ·O(E123): 2.90

Figure 6.6: The CASSCF FC geometries based on the four optimization schemes. The bond lengths (Å, red numbers), dihedral angles (degrees, black numbers), BLA vaules (Å), helicity (degrees) and curvature (degrees) of the retinal polyene chain, and the shortest distance (Å) between nitrogen of RSBH$^+$ and carboxyl oxygen of E123 (N· · · O(E123)). The equations of computing the total helicity and curvature of the polyene chain are referred to ref. [187].

Table 6.3: The vertical excitation energies of three roots of the FC geometry based on the four relaxed scan schemes. The medium-sized basis set 6-31* was applied, IPEA = 0.0 au, LS = 0.1 au. Unit of dihedral angles: degree. Unit of energy: eV. Experimental λ_{max} = 2.62 eV.

	dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MSCASPT2}$	E_{CASSCF}
RS-1	FC(178)	S_0	S_0	0	0	0
			S_1	2.57	2.59	3.86
			S_2	3.47	3.84	4.45
RS-2	FC(178)	S_0	S_0	0	0	0
			S_1	2.57	2.59	3.86
			S_2	3.47	3.84	4.45
RS-3	FC(178)	S_0	S_0	0	0	0
			S_1	2.88	2.65	4.09
			S_2	3.33	4.04	4.62
RS-4	FC(178)	S_0	S_0	0	0	0
			S_1	2.93	2.67	4.10
			S_2	3.32	4.03	4.64

The CASPT2, MSCASPT2 and CASSCF energy pathways along the C12-C13=C14-C15 torsion angle based on the four relaxed scan schemes are shown in Figures 6.7–6.9 (Please see the corresponding energy values in Appendix C Tables C.1–C.8). The schemes *RS-1* and *RS-2* result in almost the same pathways.

Except for the absorption dark state S_2 energy profile, the overall shape of the pathways of *RS-3* and *RS-4* are almost the same as those of *RS-1* and *RS-2*. This reveals that the retinal isomerization property will not be altered by an additional relaxation of the retinal binding pocket using the following computational parameters: basis set = 6-31G*, IPEA = 0.0 au, LS = 0.1 au, number of roots = 3.

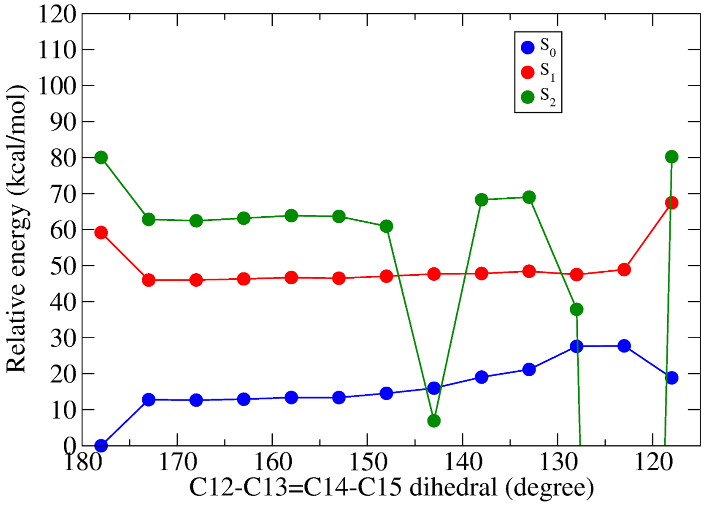
Intruder State Problem

Notably, for all the four relaxed scan schemes, the CASPT2 pathways show the intruder state problem, i.e. the cross of different electronic state energies at the region far from conical intersection. In order to eliminate this problem, the excitation energies were calculated using a large LS parameter, which adds a shift to the external part of the zeroth order Hamiltonian [241, 242].

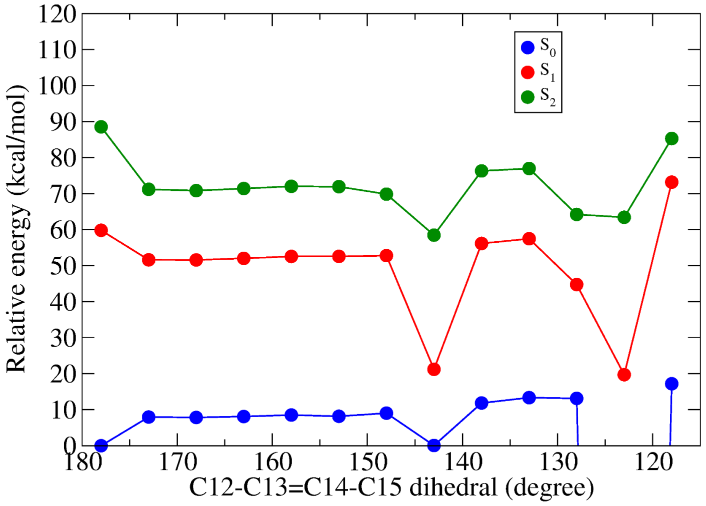
To determine the appropriate LS parameter, calibration calculations were performed with various LS values for one geometry of the RSBH-E123 model based on the relaxed scan scheme *RS-1* (see Table 6.4). The excited states have the same reference weight of the CASSCF wavefunctions as the ground state using LS values larger than 0.1 au, and correspondingly, the intruder state problem is removed. In the current study, the LS of 0.2 au is applied. The computed excitation energies along the isomerization coordinate are listed in Appendix C Table C.9. As is shown, the S_1 and S_2 states are far away from each other in energy (larger than 15 kcal/mol). Additionally, the S_1–S_2 coupling is very weak, indicated by the small $S_1 \rightarrow S_2$ oscillator strength. This means that the exclusion of S_2 state from the calculations will not affect the obtained results. Therefore, for the eventual analysis, the excitation energies were determined over two roots with equal weights, using LS = 0.2 au.

Table 6.4: Excitation energies (eV) and reference weights of the CASSCF wavefunctions for different LS values (au). The calculations are performed on geometry of the RSBH-E123 model with C12-C13=C14-C15 dihedral of 143°. This geometry was obtained from the relaxed scan scheme *RS-1*. Basis set = 6-31G*, IPEA = 0.0 au, number of roots = 3.

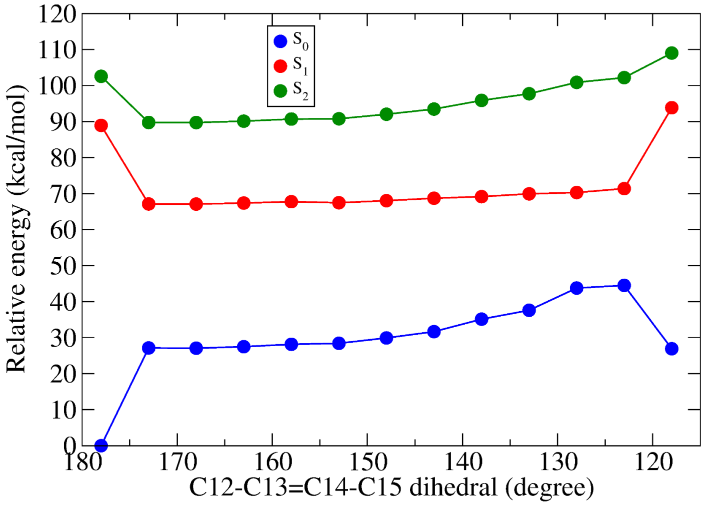
	E_{CASPT2} (eV)				
LS	E_{S_1}	E_{S_2}	weight GS	weight S_1	weight S_2
0.1	2.07	0.30	0.54	0.54	0.35
0.2	2.32	3.23	0.56	0.56	0.56
0.3	2.68	3.62	0.58	0.58	0.58
0.4	3.13	4.09	0.59	0.59	0.59
0.5	3.65	4.62	0.61	0.61	0.61



(a) CASPT2

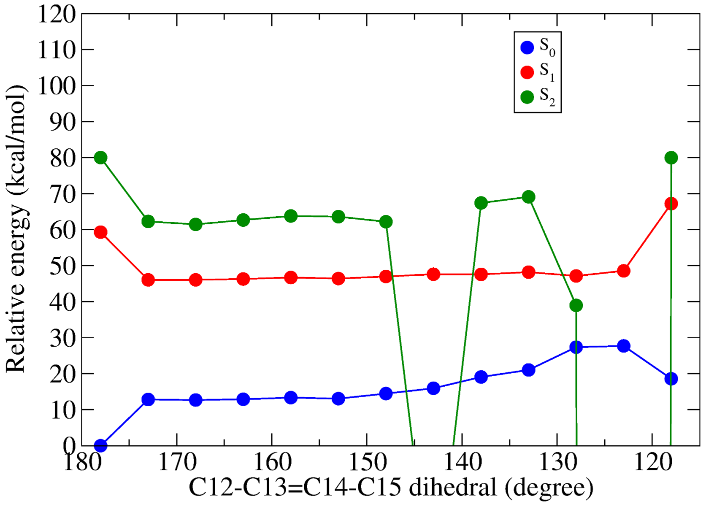


(b) MSCASPT2

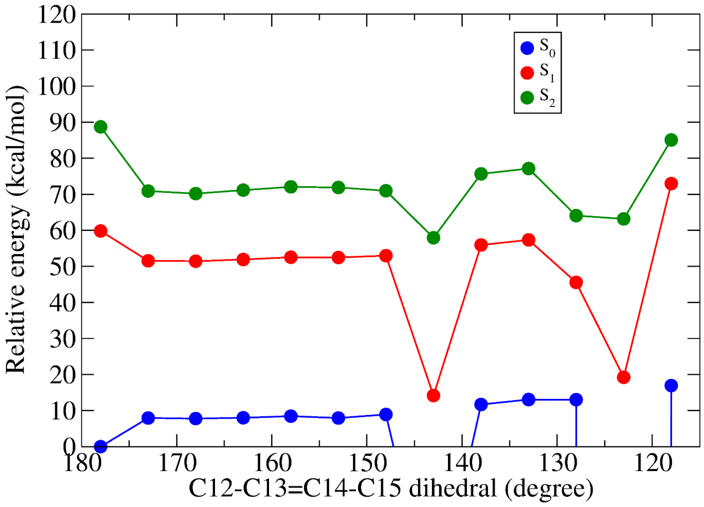


(c) CASSCF

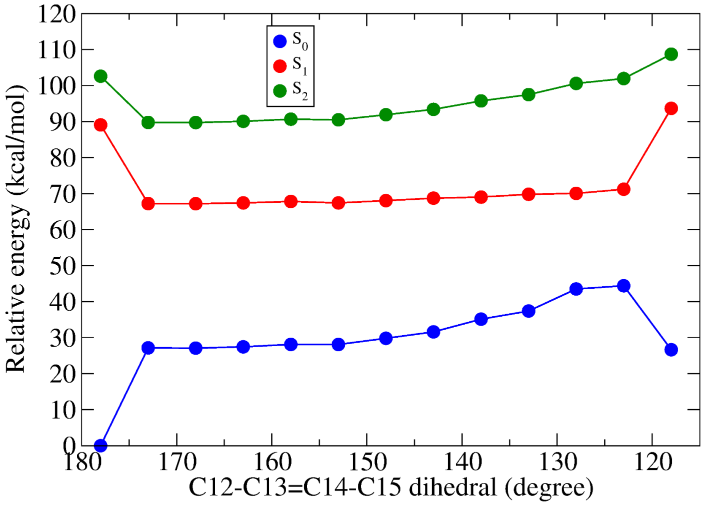
Figure 6.7: The CASPT2, MSCASPT2 and CASSCF energy pathways of RSBH-E123 model based on the relaxed scan scheme *RS-1* (see main text for the details).



(a) CASPT2

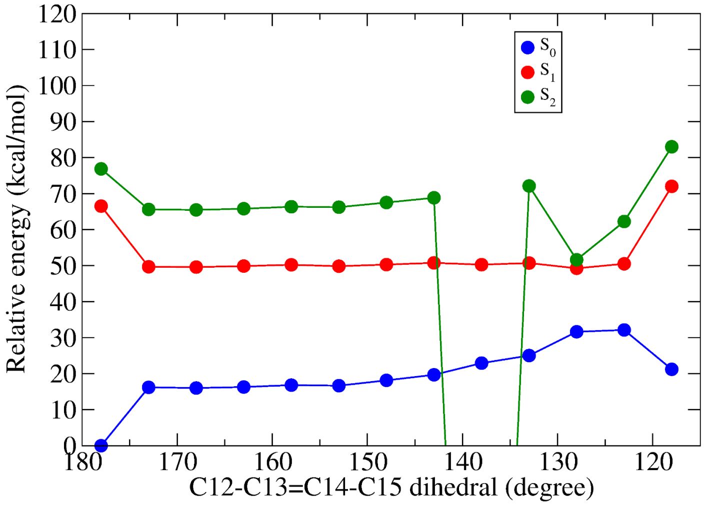


(b) MSCASPT2

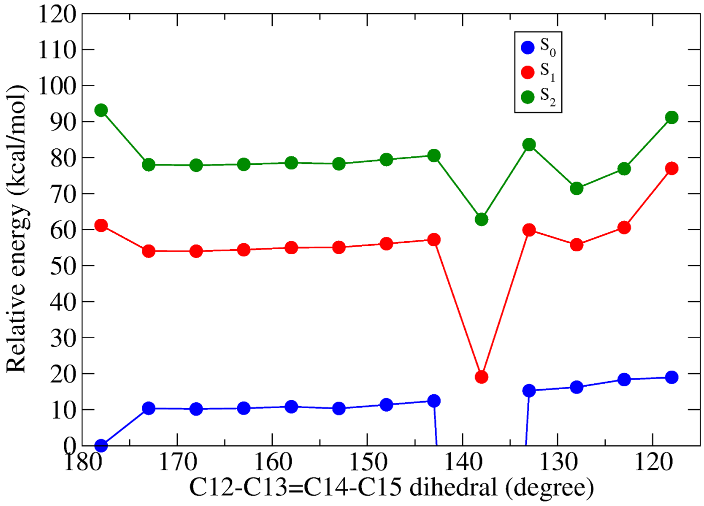


(c) CASSCF

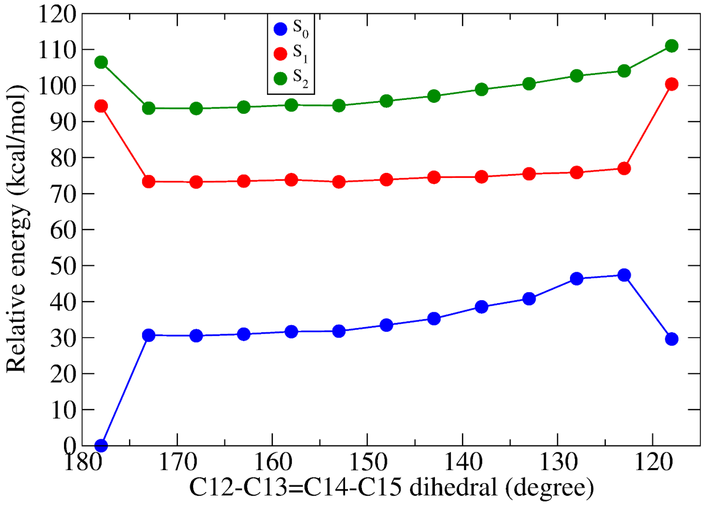
Figure 6.8: The CASPT2, MSCASPT2 and CASSCF energy pathways of RSBH-E123 model based on the relaxed scan scheme *RS-2* (see main text for the details).



(a) CASPT2

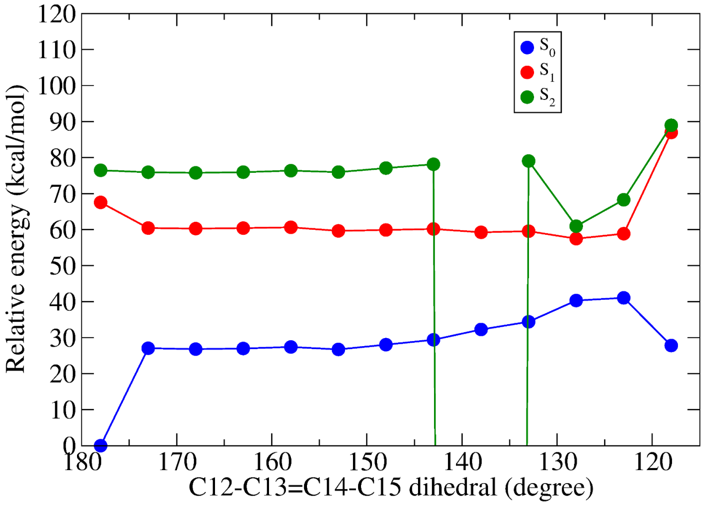


(b) MSCASPT2

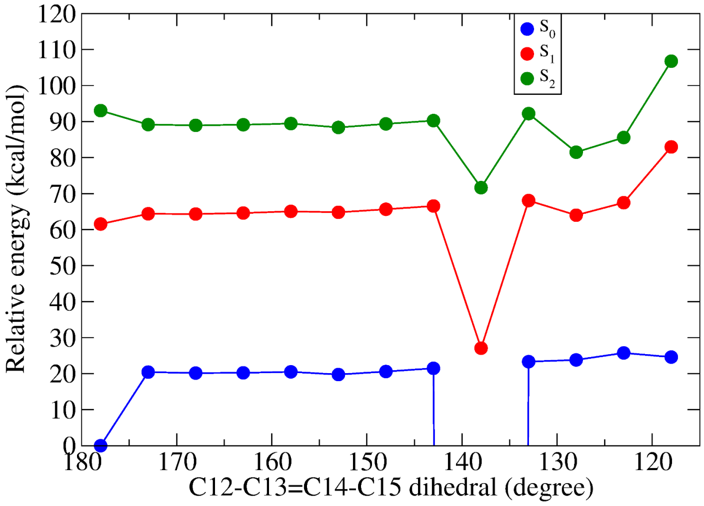


(c) CASSCF

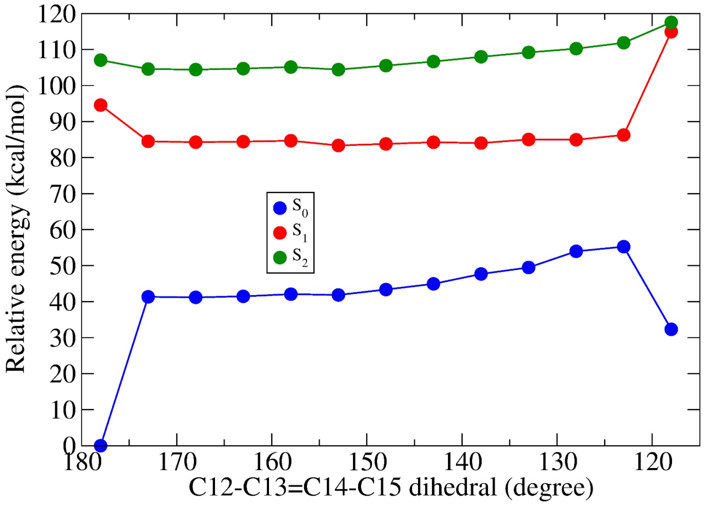
Figure 6.9: The CASPT2, MSCASPT2 and CASSCF energy pathways of RSBH-E123 model based on the relaxed scan scheme *RS-3* (see main text for the details).



(a) CASPT2



(b) MSCASPT2



(c) CASSCF

Figure 6.10: The CASPT2, MSCASPT2 and CASSCF energy pathways of RSBH-E123 model based on the relaxed scan scheme *RS-4* (see main text for the details).

6.2.3 Chromophore Isomerization

The retinal isomerization was studied using two different sets of computational parameters: (i) Relaxed scan scheme: *RS-1*; excitation energy calculations: basis set = 6-31G*, IPEA = 0.0 au, LS = 0.2 au, number of roots = 2. (ii) Relaxed scan scheme: *RS-4*, using the same excitation energy calculations parameters as (i). Then the impact of the relaxation of the retinal binding pocket on the isomerization properties will be demonstrated for the above excitation energy calculation parameters. Since the CASPT2 method gives better description at the region close to the CI than MSCASPT2, the following discussions are based on CASPT2 calculations.

Properties of S_1 Energy Profile

Relaxed Scan *RS-1*

The FC geometry of the RSBH-E123 model has a torsion of 178° about the -C13=C14- double bond. Starting from the FC, unconstrained geometry optimization on S_1 was performed, which results in a minimum with C12-C13=C14-C15 torsion of 171°. Further excited state optimization leads to another minimum. The two minima are referred to as *minim-1* and *minim-2* hereafter, respectively (see Figure 6.11). Previous studies of bovine rhodopsin reported two different types of S_1 energy minima [243, 244]. The first minimum (the so called locally excited state, LE state) is close to the FC point in terms of geometry and electronic structure. The second minimum (charge transfer state, CT state) is characterized by a complete bond length inversion in the middle part of the retinal polyene chain, by larger twist around the isomerization double bond, by different electronic structure with respect to the first minimum. The retinal geometries of FC, *minim-1* and *minim-2* of the RSBH-E123 model are shown in Figure 6.12 (a). The structure of *minim-1* differs largely from that of FC point: the bond lengths of the middle part of the polyene chain (segment C8 – C14) are completely inverted in *minim-1*, depicted by a negative BLA. This indicates that the *minim-1* is more similar to the CT state than to the LE state. The *minim-2* is 0.7 kcal/mol above *minim-1* (Table 6.5), separated from *minim-1* by a negligible barrier of 0.8 kcal/mol. The bond lengths of the retinal at *minim-1* and *minim-2* are almost the same. The transition from *minim-1* to *minim-2* involves noticeable -C11=C12- and -C15=N- bond twists in the opposite direction to the -C13=C14- bond twist, during which the retinal from C11 to the β-ionone ring part stays almost intact. As is shown in Figure 6.11, *minim-1* and *minim-2* belong to a flat region of the S_1 path, where the increase of the torsion around the -C13=C14- bond (from 171° to 153°) hardly affects the S_1 energy.

The unconstrained geometry optimization of the RSBH-H_2O model leads to *minim-1* with -C13=C14- torsion of 170°. It is separated from the more twisted *minim-2* (150° twisted around -C13=C14- bond) by an energy barrier of 2.2 kcal/mol, which is ca. twice larger than that of the RSBH-E123 model. The *minim-2* is 2 kcal/mol above *minim-1*. The S_1 relaxation of the RSBH-D253 model results in one minimum, spanning a flat region from 179° to 164° along the -C13=C14- torsion, where the profile energetically increases by 1.1 kcal/mol.

Upon leaving the S_1 minimum, the retinal photoisomerization reaction proceeds towards the CI, where the S_1 and S_0 energy surfaces are degenerate. In RSBH-H_2O and RSBH-D253 models, the CI point is separated from the S_1 minimum by a transition state (TS) (see Figure 6.11). The energy barrier is 8.9 kcal/mol and 12.5 kcal/mol for the RSBH-H_2O and the RSBH-D253 models, respectively (see Table 6.5). In the RSBH-E123 model, the CI point is 4.9 kcal/mol above the S_1 minimum without transition state. The very fast retinal isomerization in ChR2, within 400 fs [72], implies a barrierless S_1 path or a S_1 path with a small energy barrier. Apparently, the energy barriers of the three models, especially those of RSBH-H_2O and RSBH-D253 models are too high to lead to the very fast isomerization.

Table 6.5: CASPT2//CASSCF/MM relative energies of characteristic structures along the isomerization coordinate. The relative energies are calculated with respect to the ground state energy of the FC geometry. Basis set = 6-31G*, IPEA = 0.0 au, LS = 0.2 au, number of roots = 2, relaxed scan scheme = *RS-1*. Unit: kcal/mol.

Model	electronic state	FC	minim-1	minim-2	TS	CI
RSBH-E123	S_0	0.0	10.6	12.9		44.5
	S_1	59.5	47.1	47.8		52.7

RSBH-H_2O	S_0	0.0	9.8	13.7	30.3	57.4
	S_1	78.1	57.8	59.8	68.7	59.8
RSBH-D253	S_0	0.0	9.0		32.4	54.7
	S_1	79.2	56.7		69.2	55.6

Relaxed Scan *RS-4*

The relaxation of the residues surrounding the retinal results in a gently sloped S_1 energy profile of the RSBH-E123 model (see Figure 6.13). Upon leaving the FC point, from 171° to 128° along the -C13=C14- torsion, the energy is reduced by ca. 2 kcal/mol gradually. After that, a small barrier of ca. 1.4 kcal/mol occurs, separating the steep profile, which leads to the CI point.

Relatively, the S_1 paths of the RSBH-H_2O and RSBH-D253 models are weakly affected by the relaxation of the retinal binding pocket (see Figure 6.13). The high energy barriers of 11.2 kcal/mol and 12.4 kcal/mol for the RSBH-H_2O and the RSBH-D253 models, respectively, make these two paths inefficient. In the following sections, the discussions will be presented based on energy pathways from relaxed scan *RS-4*.

Conical Intersection

The CI structures and positions of the three models are hardly affected by the relaxation of the protein. The CI structure of the RSBH-E123 model displays a high twist around the -C13=C14- bond (118°), which is 60° more twisted with respect to the FC structure (see Figure 6.12). In the RSBH-H_2O model, the torsion around the -C13=C14- bond of the CI structure is 105°. Relatively, in the RSBH-D253 model the torsion is 119°. In the latter two models, the retinal of the CI is 70° more twisted about the central -C13=C14- bond relative to the FC geometry (see Figure 6.12). This means that the S_1/S_0 crossing occurs earlier in the RSBH-E123 model than in the RSBH-H_2O and RSBH-D253 models. Cembran et al. [231] investigated the counterion (acetate) effect on the photoisomerization using the simplified retinal models (without β-ionone ring and methyl groups). The calculations show that when the counterion is placed close to the Schiff base, the S_1/S_0 crossing occurs later compared to the counterion position close to the middle part or to the C-tail (the tail opposite to the Schiff base) of the retinal model. It seems that the current calculations lead to conclusions that are the opposite of those from the former study. However, some factors should be taken into account in the comparison: in the previous study (i) the retinal model is simplified; (ii) the protein environment effect is not considered. These two factors might affect the results largely.

Notably, when passing from FC to CI, the extent of the total twist around the -C13=C14- bond are the same for the RSBH-H_2O and RSBH-D253 models. Checking the S_1 relaxed geometries of RSBH-D253 model, it is found that at the early stage of the relaxed scan the RSBH$^+$ NH switched from D253 side chain to one water molecule, resulting in a structure similar to the RSBH-H_2O model. Therefore, it is reasonable that similar S_1 profiles were obatined for the two models: both show a transition state; both have a high energy barrier to overcome to reach the CI; both show a 70° twist around the -C13=C14- bond from FC to CI. In order to get a better understanding of the retinal photoisomerization of the RSBH-D253 model, additional starting structures should be considered.

The CI structure is located ca. 27 kcal/mol below the FC point in the RSBH-E123 model, and this energy difference is ca. 18 kcal/mol in the RSBH-H_2O model, and ca. 23 kcal/mol in the RSBH-D253 model (see Table 6.6). Although the FC–S_1 minimum excess energy for all three models is large enough in principle to overcome the energy barrier on the S_1 reaction pathway, the barrier as large as 11.2 kcal/mol and 12.5 kcal/mol for the RSBH-H_2O and the RSBH-D253 models, respectively, is still too high. Considering the S_1 structural similarity between the RSBH-H_2O and RSBH-D253 models as discussed above, one conclusion might be obtained: the retinal photoisomerization in the RSBH-H_2O model is inefficient for generating the *cis*-retinal.

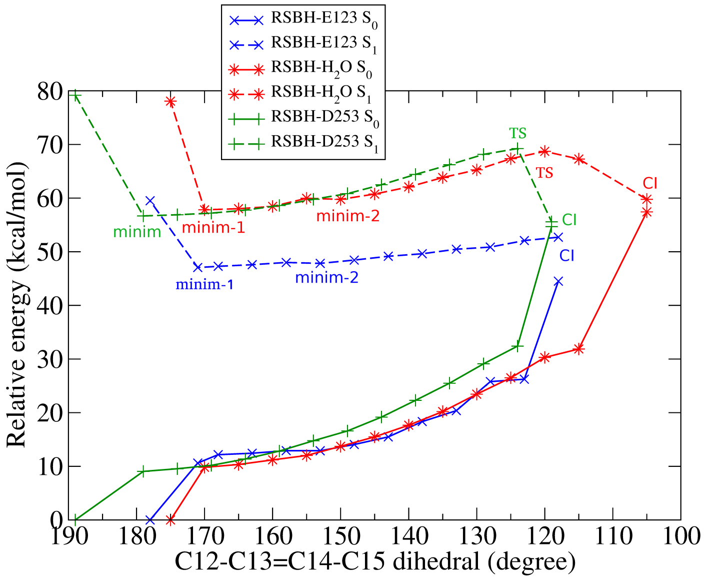


Figure 6.11: The retinal isomerization pathways of RSBH-E123, RSBH-H$_2$O and RSBH-D253 models based on relaxed scan scheme *RS-1*.

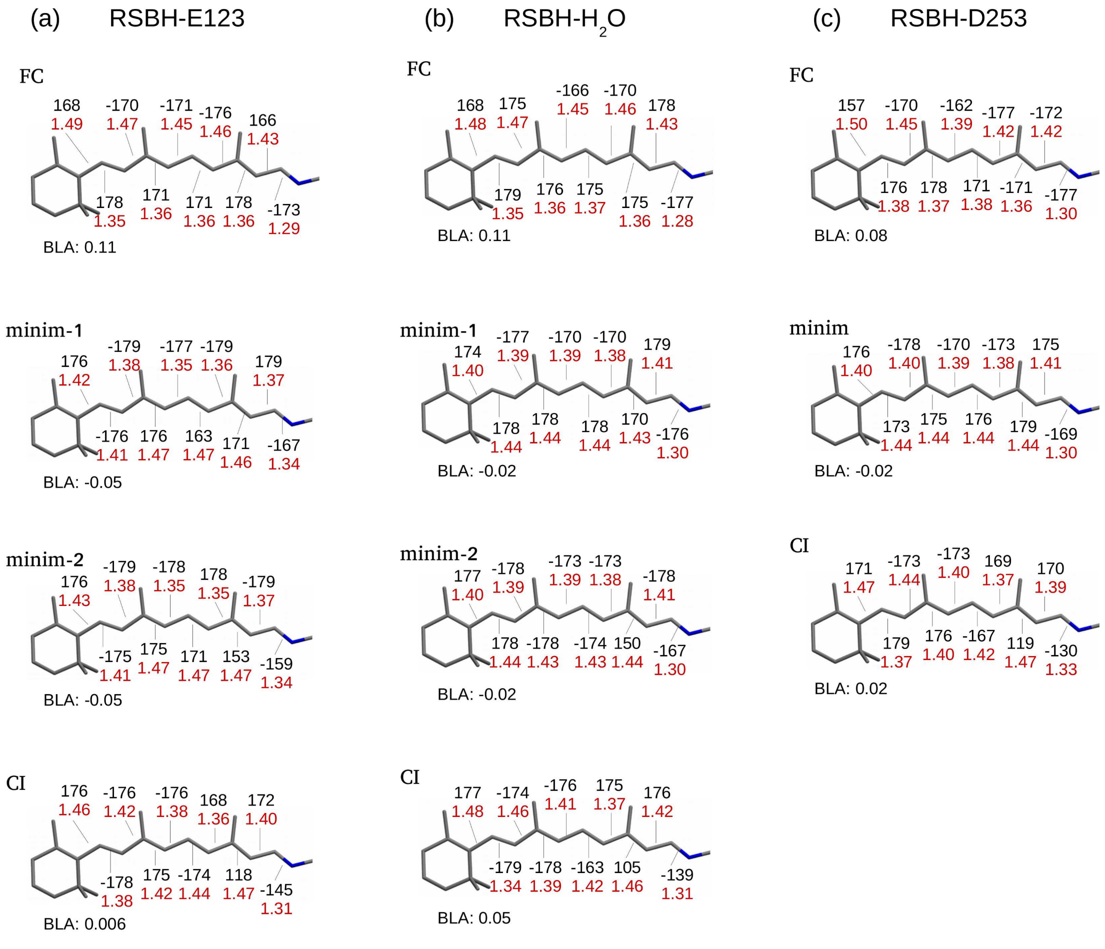


Figure 6.12: The retinal geometrical parameters, including the bond lengths (Å, red), dihedral angles (degree, black), ployene chain BLA (Å), of the stationary points along the S$_1$ energy profile of (a) RSBH-E123 model, (b) RSBH-H$_2$O model, (c) RSBH-D253 model.

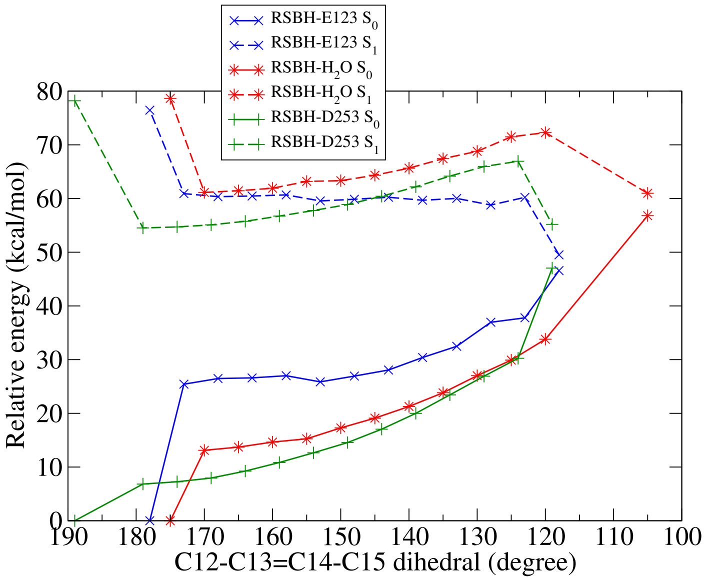


Figure 6.13: The retinal isomerization pathways of RSBH-E123, RSBH-H_2O and RSBH-D253 models based on the relaxed scan scheme *RS-4*.

Table 6.6: CASPT2//CASSCF/MM relative energies of characteristic structures along the isomerization coordinate. The relative energies are calculated with respect to the ground state energy of the FC geometry. Basis set = 6-31G*, IPEA = 0.0 au, LS = 0.2 au, number of roots = 2, relaxed scan scheme = *RS-4*. Unit: kcal/mol.

Model	electronic state	FC	minim	TS	CI
RSBH-E123	S_0	0.0	37.0	37.8	46.6
	S_1	76.4	58.8	60.2	49.5
RSBH-H_2O	S_0	0.0	13.1	33.8	56.8
	S_1	78.6	61.1	72.3	60.9
RSBH-D253	S_0	0.0	6.8	30.3	47.0
	S_1	78.2	54.5	66.9	55.1

Isomerization Motion

The complete chromophore isomerization involves three consecutive steps, which are characterized by different structural changes, see Figure 6.14. The first step is characterized by the inversion of polyene chain bond lengths, which takes places immediately after light absorption. This is in agreement with the reports on BR and Rh [230, 245–247]. The following step is the rotation about the -C13=C14- double bond accompanied by small rotations about two nearby double bonds -C11=C12- and -C15=N-, allowing the system to reach the CI. This rotational motion is the so-called “aborted double bicycle-pedal” mechanism (i.e. the rotation around -C13=C14- is complete, while it is aborted around -C11=C12- and -C15=N- after CI), involving a tiny pyramidalization at C14 in ChR2. This motion is reported for BR by Altoè [230], in which the tiny pyramidalizations locate at C12 and C13 conflicting with Hayashi’s study at C14 [248]. The structural deformation due to the pyramidalization leads to the increase of S_0 energy, reducing the S_1–S_0 energy gap, which enhances the transition probability [248]. Finally, the isomerized retinal decays to the ground state through the CI, reaching the P_1 state in the photocycle. Consequently, the double bonds are reconstructed.

The three adjacent rotations around double bonds -C11=C12-, -C13=C14-, -C15=N- proceed in alternating directions. This is consistent with BR and Rh [230]. The twisting extent of the -C11=C12- and -C15=N- bonds are smaller than that experienced by -C13=C14- rotation. It is interesting that the combined rotation around -C11=C12- (+23° for RSBH-E123, +19° for RSBH-H_2O, +17° for RSBH-D253) and -C15=N- (+22° for RSBH-E123, +37° for RSBH-H_2O, +39° for RSBH-D253) almost fully compensates for the rotation around C13=C14 (-53° for RSBH-E123, -65° for RSBH-H_2O, -60° for RSBH-D253) at the twisted CI. The β-ionone ring part of the retinal only shows very small geometrical change along the isomerization coordinate (Figure 6.14 (b)), further allowing the minimum volume isomerization.

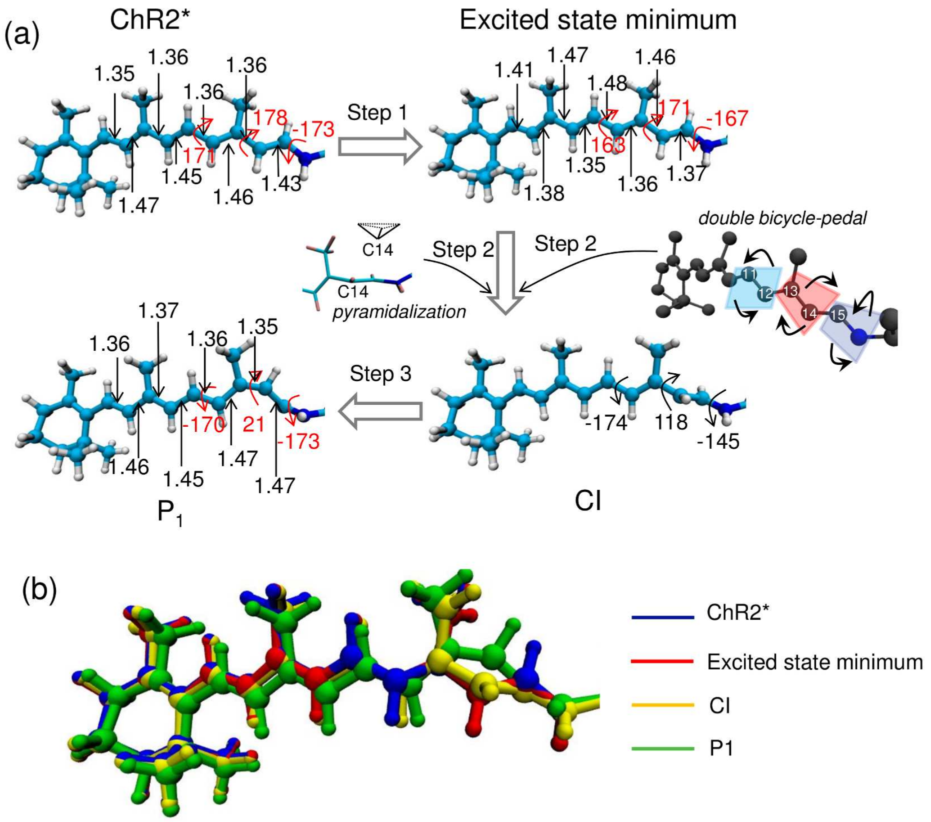


Figure 6.14: Retinal isomerization mechanism in ChR2 illustrated using the RSBH-E123 model. (a) The three consecutive steps of retinal isomerization and the corresponding representative geometries. (b) Superposition of different geometries along the isomerization coordinate. All the distances are in Å, dihedrals in degree.

6.2.4 Early P_1 Intermediate

P_1 state is the first intermediate of ChR2 photocycle that appears after the all-*trans*→13-*cis* photoisomerization of the chromophore. The fact that the first proton transfer during the photocycle occurs in the $P_1 \rightarrow P_2$ transition emphasizes the importance of P_1 state [72].

IR studies at 80 K suggest significant alterations of the protein secondary structure in the early stage of the photocycle [62, 88]. The recent early IR difference spectra of ChR2-WT at room temperature resolved similar protein conformational changes as the previous studies at 80 K [107]. The authors predicted at least two P_1 substates with avsimilar retinal conformation and electrostatic environment but differing in the structure of the apoprotein. These structural changes of the apoprotein rise with a half-life time of 400 ns. The HOOP (Hydrogen Out Of Plane) vibration band [83] at 988 cm^{-1} suggests a twisted retinal conformation in ChR2 early photocycle intermediates, which is in analogy to the positive HOOP band at 983 cm^{-1} of bR and between 995 and 965 cm^{-1} of several microbial rhodopsins.

Geometry optimization on the ground state starting from the CI point results in the early P_1 intermediate. As discussed above, because of the high excited state energy barrier, the RSBH-H_2O and RSBH-D253 pathways are not taken into account. The ground state relaxation involves a further twist around the central -C13=C14- bond. The geometry resulting from the RSBH-E123 pathway has C12-C13=C14-C15 dihedral of 21°. Moreover, it is ca. 20° twisted around the -C15=N- bond and ca. 10° twisted around the -C11=C12- bond.

To evaluate the retinal geometry of P_1 state, the FTIR difference spectrum was computed using the FTTCF method. The spectrum (Figure 6.15) shows vibrational bands at 894, 938, 1158, 1201, 1300 cm^{-1}. The Raman spectroscopy study by Heberle group [222] reports the characteristic 13-*cis* retinal C-C stretching band at 1157, 1196 cm^{-1}, and the CH_3 rocking band at 1301 cm^{-1}. The present computed spectrum has very similar bands, verfying that the 13-*cis* retinal was obtained. Additionally, the computed bands at 894, 938 cm^{-1} could be assigned to the HOOP vibration, indicating a highly twisted retinal.

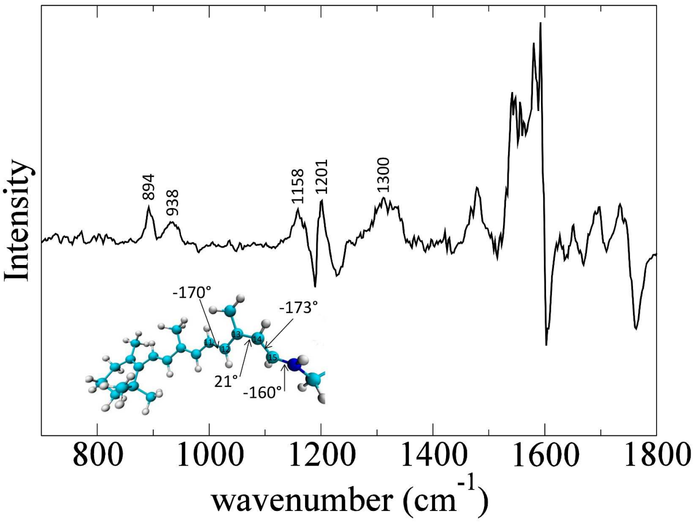


Figure 6.15: The computed FTIR difference spectrum of the P_1 geometry from the RSBH-E123 pathway.

6.3 Conclusions

The CASPT2//CASSCF/MM method was used to study the chromophore isomerization properties of ChR2. For the current models, the following computational parameters should be employed to obtain rational isomerization pathways: relaxation of the retinal, the water molecules, the residues surrounding the retinal in the relaxed scan optimization; basis set = 6-31G*, IPEA = 0.0 au, LS = 0.2 au for the 2-root CASPT2 calculations. Multiple environments surrounding the RSBH$^+$ are considered. The environment (protein+water) surrounding the RSBH$^+$ has significant impact on the retinal isomerization pathway. Two types isomerization pathways were summerized as shown in Figure 6.16 based on the current study. In the case of the negatively charged counterion forming strong salt bridge with RSBH$^+$, the retinal approaches the twisted 13-*cis* conformer through CI radiationless decay. Before reaching the CI, the S_1 path propagates with small steepness (Figure 6.16 (a)). The IR difference spectrum of the obtained retinal comprises the reported experimental vibrational bands. The retinal is ca. 20° twisted around 13=C14, C15=N bonds and ca. 10° twisted around the C11=C12 bond. When the RSBH$^+$ forms hydrogen bond with a water molecule, there is a high barrier on the S_1 path separating the FC active region from the CI reactive region (Figure 6.16 (b)). This pathway is inefficient because of the unrealistic long time scale needed to overcome such a high energy barrier.

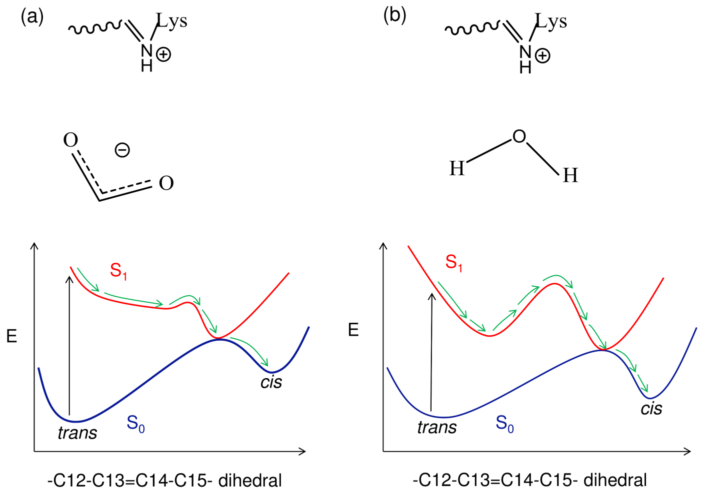


Figure 6.16: Two models and the corresponding retinal isomerization pathways.

Chapter 7

Conclusions and Outlook

Experimental and theoretical research of protein prototypes, including GluRs and ChRs, has promoted the understanding of light-activated ion channels. This provides guidelines for designing new versatile optogenetic tools. A promising application of those proteins would be the specific activation of defined cells in the brain, which contributes to the studies of neuron diseases.

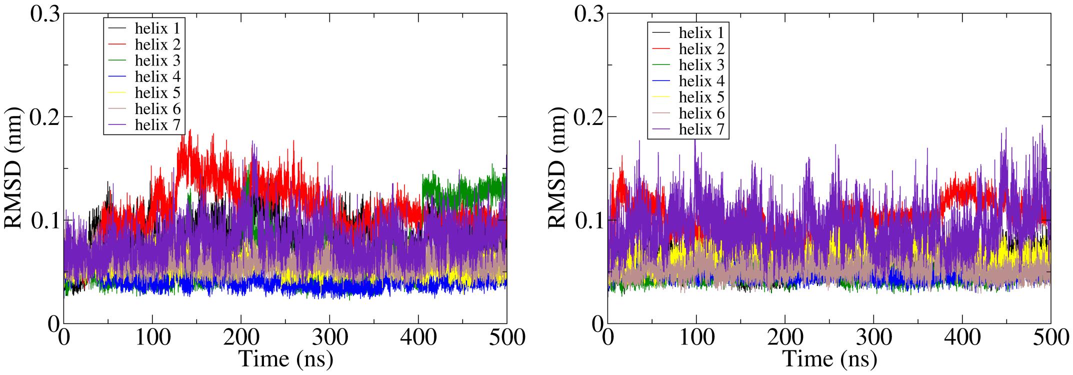
In this thesis, the structure–function relationships of GluK2 and ChR2 were investigated using state-of-the-art computational methods. In Chapter 3, the response of the GluK2 receptor to the binding of different *gluazo* isomers was investigated. Extensive molecular dynamics simulations show that *trans-gluazo* is more efficient than *cis-gluazo* in promoting the opening of the channel. The atomistic mechanism of this phenomenon was interpreted by the different ligand–protein interactions of the GluK2-*trans-gluazo* complex and the GluK2-*cis-gluazo* complex. The *trans-gluazo* binds to the upper clamshell of the ligand binding domain through three anchor points, which are located at the two ends and the middle part of the ligand molecule. By contrast, *cis-gluazo* binds to the upper domain only through two anchor points, located at the two ends of the ligand. Consequently, *cis-gluazo* is more curved in the middle part than *trans-gluazo* upon binding to the ligand binding domain. This further leads to the weakening of the interactions between *gluazo* and the lower clamshell of the ligand binding domain. The hydrogen bond of *gluazo* with T690, and the salt bridge with E738 determine the interactions of *gluazo* with the lower clamshell of the ligand binding domain. These findings may guide the design of photochromic ligands that interact with the ligand binding domains of glutamate receptors.

In the computational study of ChR2, the principal problem is the lack of crystal structure. In Chapters 4 and 5, a ChR2-C128T mutant homology model was used to explore the function-related structural questions. Classical MD simulations showed that the conformation of the dark state of ChR2-C128T is similar to the C1C2 crystal structure. Taking into account the shortcomings of classical force fields in describing strong hydrogen bonded networks, the QM/MM method was applied to explore the active site structure. At the theory level of DFTB3, three hydrogen bonding networks between RSBH$^+$ and the environment were sampled. This means that ChR2-C128T has a more complex active site than the rigid BR active site. The heterogeneous active site could explain the absorption spectrum of ChR2-C128T, i.e. the peaks of the spectrum may be assigned to the various active site hydrogen bonding patterns. This hypothesis is independent of the retinal configuration (all-*trans*,15-*anti* vs. 13-*cis*,15-*syn*). Thus, the current simulations do not support any side in the conflict of the retinal configuration (whether 13-*cis*,15-*syn* retinal exists in the dark state). Considering the growing scientific interest related to this protein, further efforts are needed to illustrate its mechanistic details, finally leading to the elucidation of the photocycle.

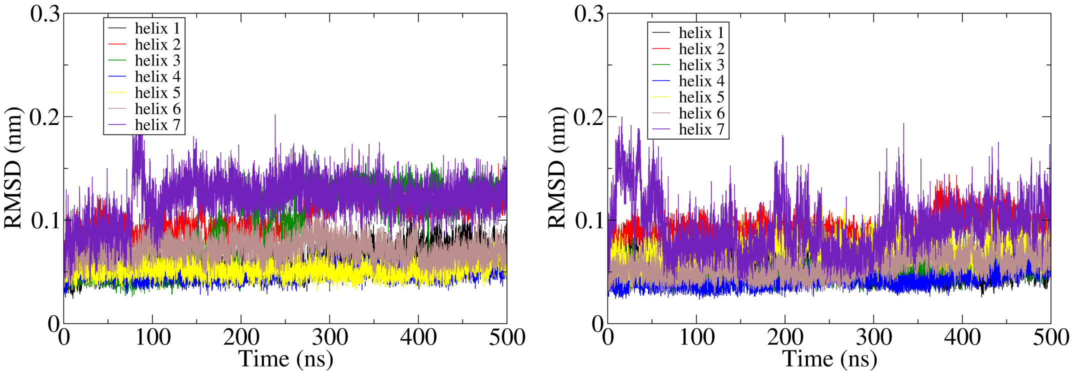
In Chapter 6, the first step of ChR2-WT photocycle, i.e. the retinal photoisomerization, was investigated employing the CASPT2//CASSCF/MM method. The influences of three active site motifs on the retinal isomerization were explored. Only one of the three active site structures can promote efficient isomerization. In future work, additional RSBH-D253 structures should be considered to span the full scope of the environmental effects on retinal isomerization, because of the S_1 structural coincidence of RSBH-D253 with RSBH-H_2O along the isomerization coordinate. Although further experimental validations of the 13-*cis*,15-*syn* photocycle of ChR2 are needed, computational studies of the isomerization 13-*cis*,15-*syn*$\rightarrow$ all-*trans*,15-*syn* are encouraged.

Appendix A

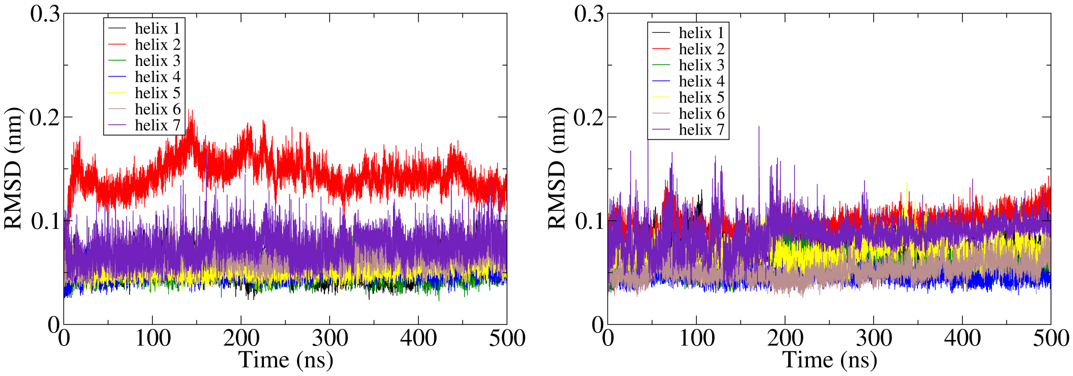
Appendices of Chapter 4



(a) trajectory 1: monomer 1 **(b)** trajectory 1: monomer 2

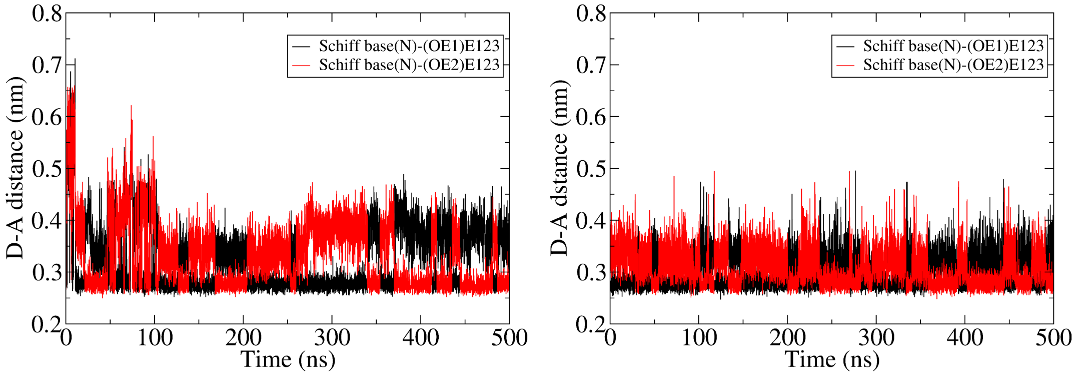


(c) trajectory 2: monomer 1 **(d)** trajectory 2: monomer 2

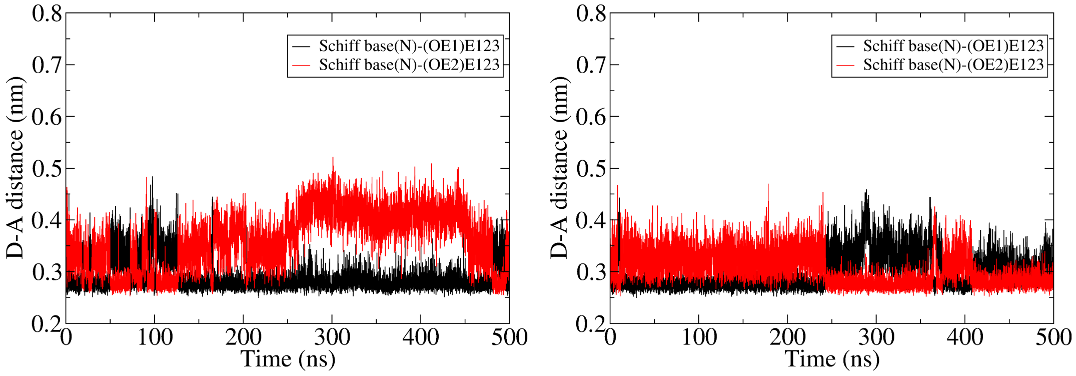


(e) trajectory 3: monomer 1 **(f)** trajectory 3: monomer 2

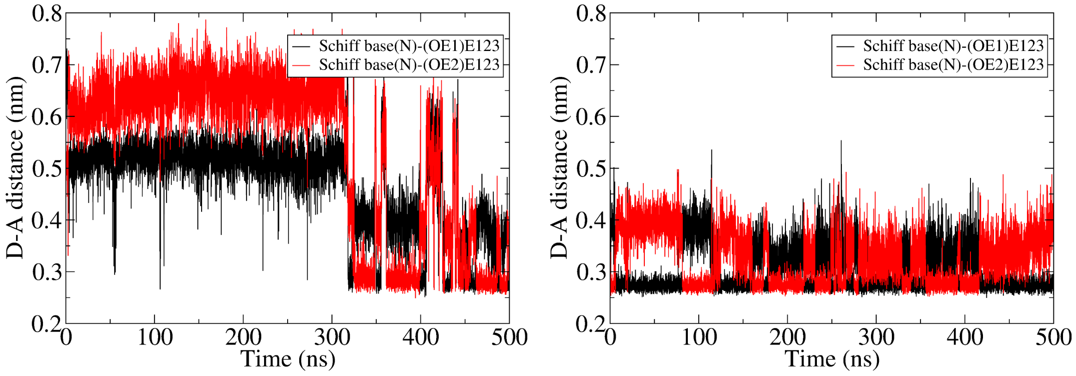
Figure A.1: The backbone RMSD of seven helices with respect to the starting structures of two monomers of three trajectories.



(a) trajectory 1: monomer 1 **(b)** trajectory 1: monomer 2

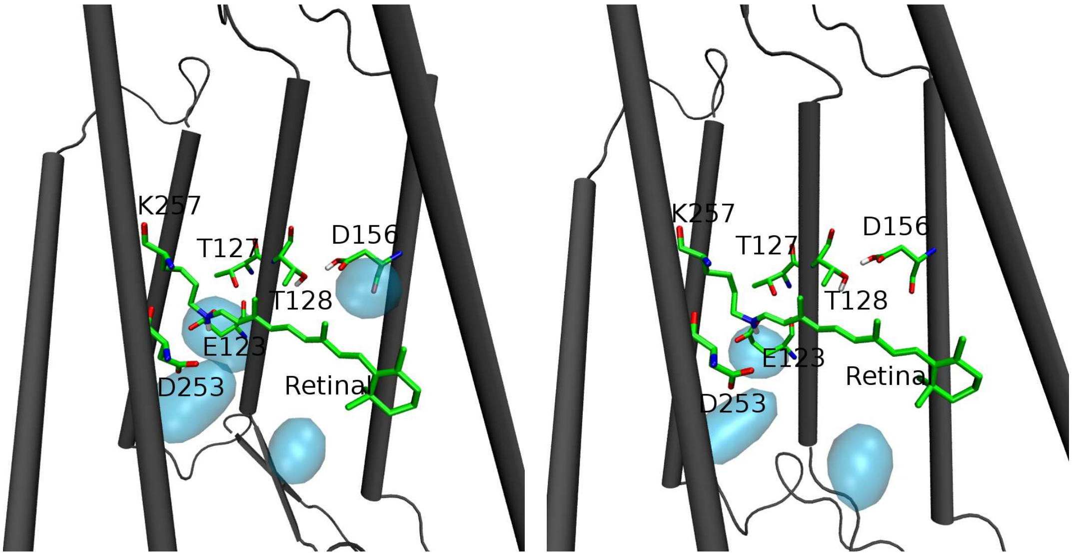


(c) trajectory 2: monomer 1 **(d)** trajectory 2: monomer 2



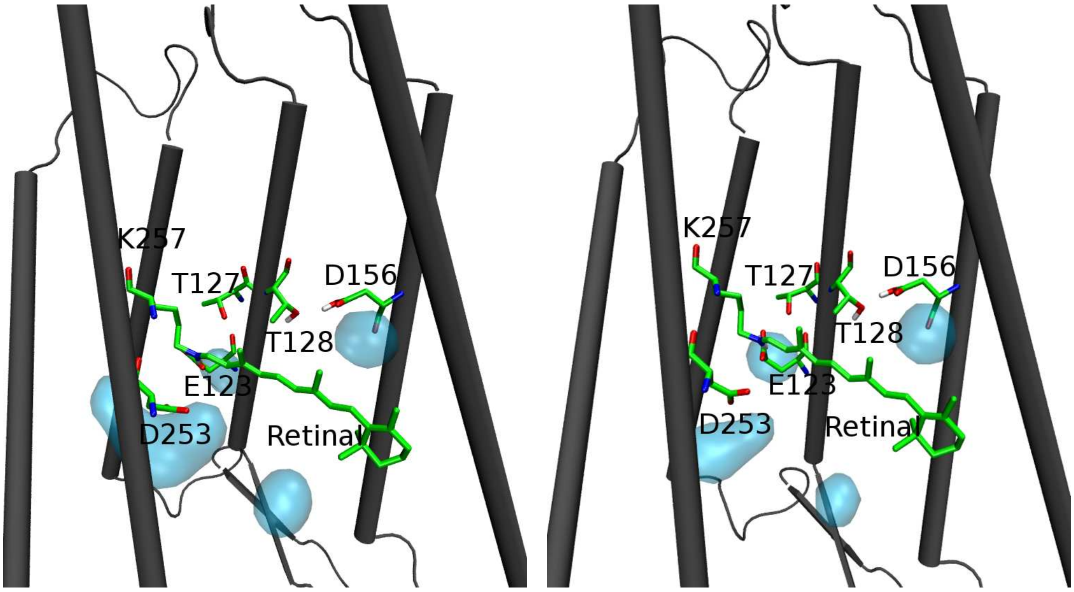
(e) trajectory 3: monomer 1 **(f)** trajectory 3: monomer 2

Figure A.2: The proton donor–acceptor distance between retinal Schiff base and E123 side chain of three trajectories. See atomic labels in Figure 4.2.



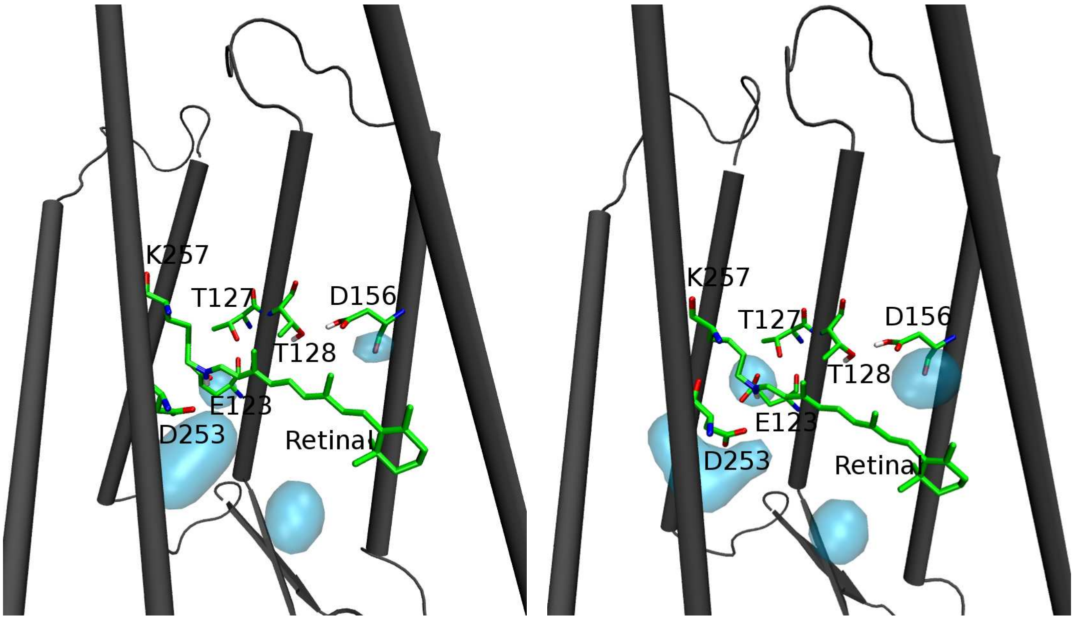
(**a**) monomer 1 (**b**) monomer 2

Figure A.3: The internal water distribution: (a) monomer 1; (b) monomer 2.



(**a**) monomer 1 (**b**) monomer 2

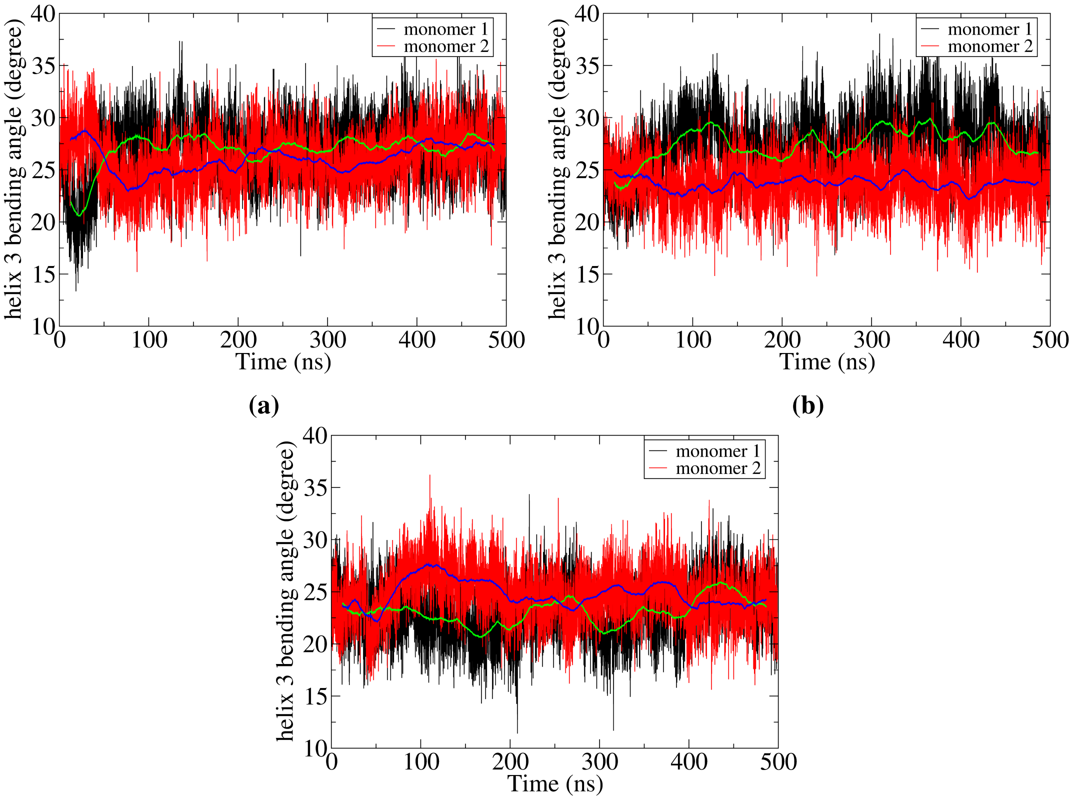
Figure A.4: The internal water distribution: (a) monomer 1; (b) monomer 2.



(a) monomer 1

(b) monomer 2

Figure A.5: The internal water distribution: (a) monomer 1; (b) monomer 2.

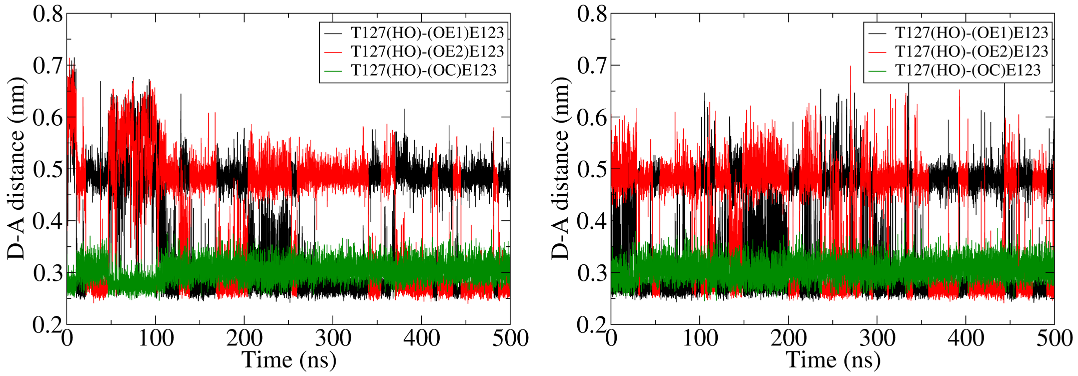


(a)

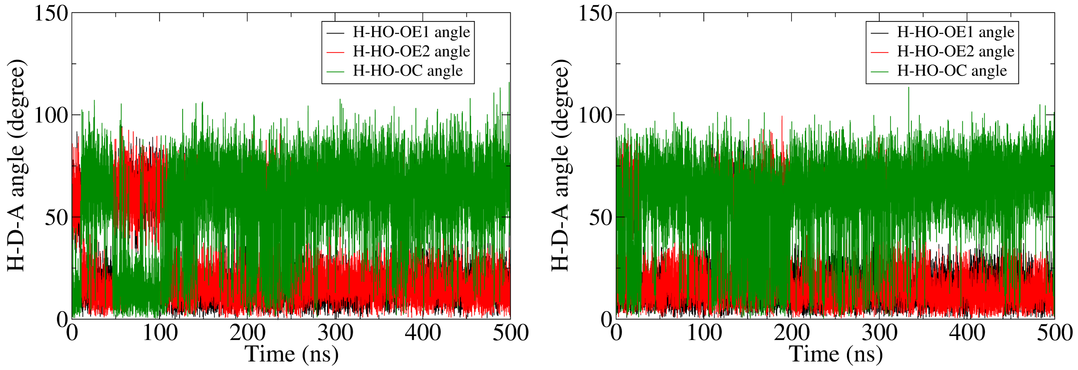
(b)

(c)

Figure A.6: Helix 3 bending angles of two monomers for three trajectories. The average values calculated every 500 ps are shown in green and blue lines for monomer 1 and monomer 2, respectively.

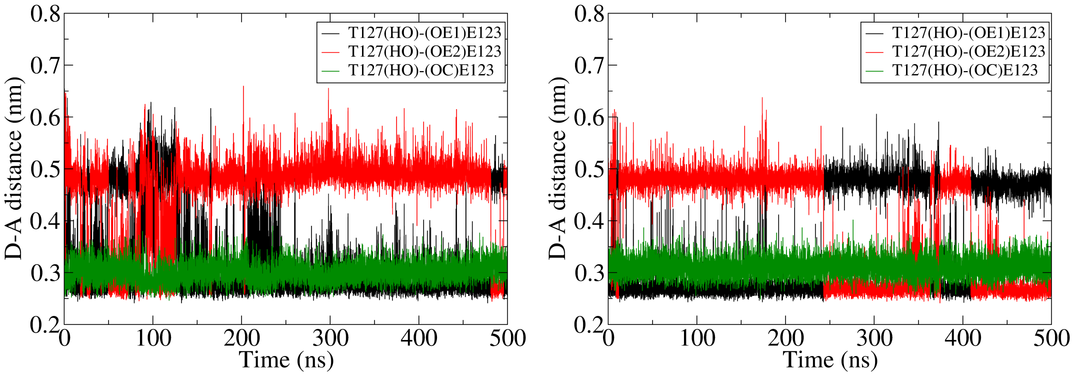


(a) monomer 1: D-A distance **(b)** monomer 2: D-A distance

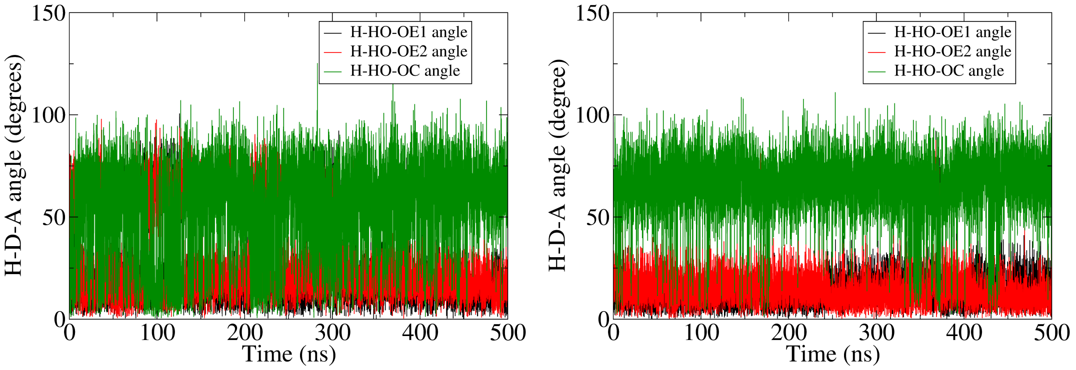


(c) monomer 1: H-D-A angle **(d)** monomer 2: H-D-A angle

Figure A.7: T127–E123 hydrogen bonding interactions depicted by the donor-acceptor (D-A) distances and the frontier H-donor-acceptor (H-D-A) angles. See main text for the criteria of defining a hydrogen bond. See atomic labels in Figure 4.2.

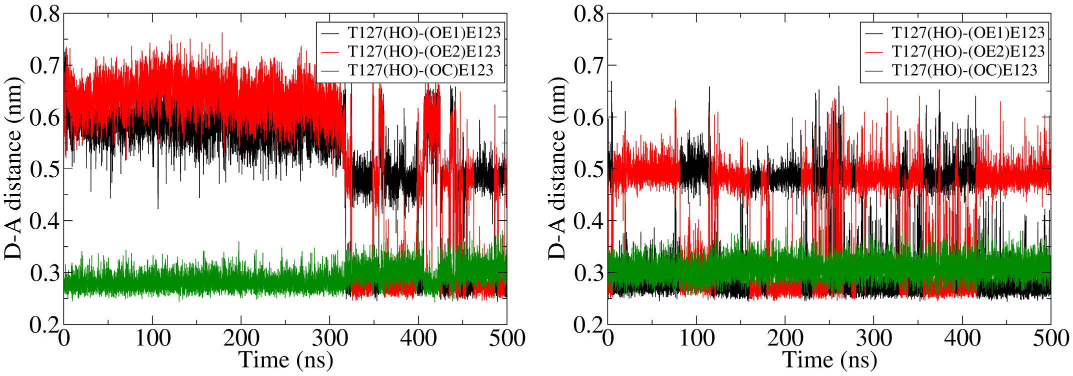


(a) monomer 1: D-A distance **(b)** monomer 2: D-A distance



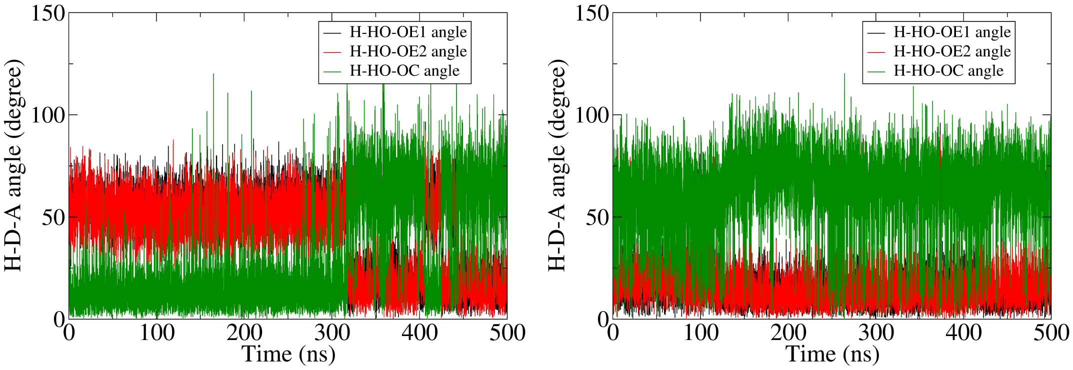
(c) monomer 1: H-D-A angle **(d)** monomer 2: H-D-A angle

Figure A.8: T127–E123 hydrogen bonding interactions depicted by the donor-acceptor (D-A) distances and frontier H-donor-acceptor (H-D-A) angles. See main text for the criteria of defining a hydrogen bond. See atomic labels in Figure 4.2.



(a) monomer 1: D-A distance

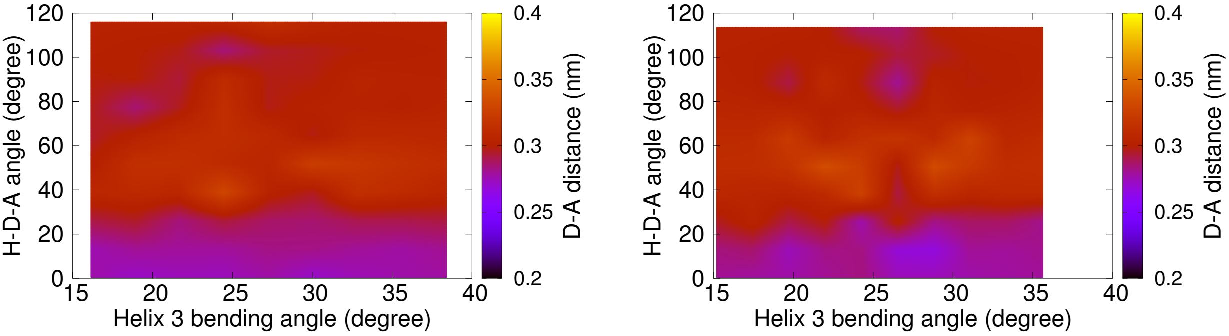
(b) monomer 2: D-A distance



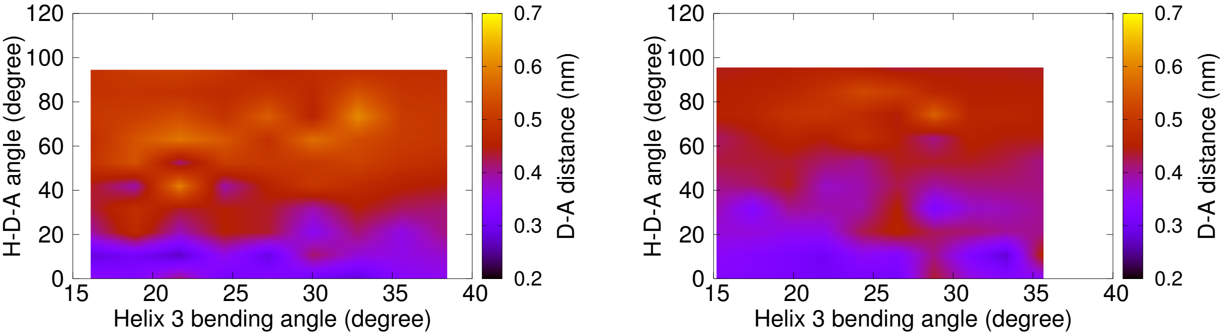
(c) monomer 1: H-D-A angle

(d) monomer 2: H-D-A angle

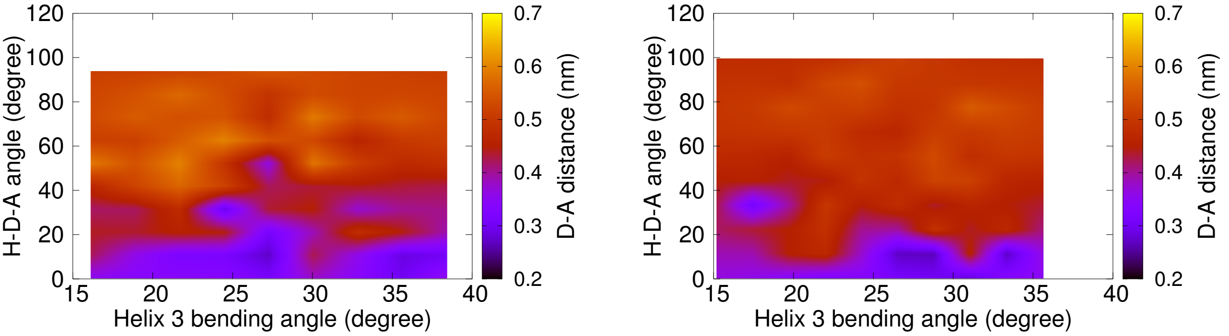
Figure A.9: T127–E123 hydrogen bonding interactions depicted by the donor-acceptor (D-A) distances and frontier H-donor-acceptor (H-D-A) angles. See main text for the criteria of defining a hydrogen bond. See atomic labels in Figure 4.2.



(a) monomer 1: T127-(backbone)E123 **(b)** monomer 2: T127-(backbone)E123

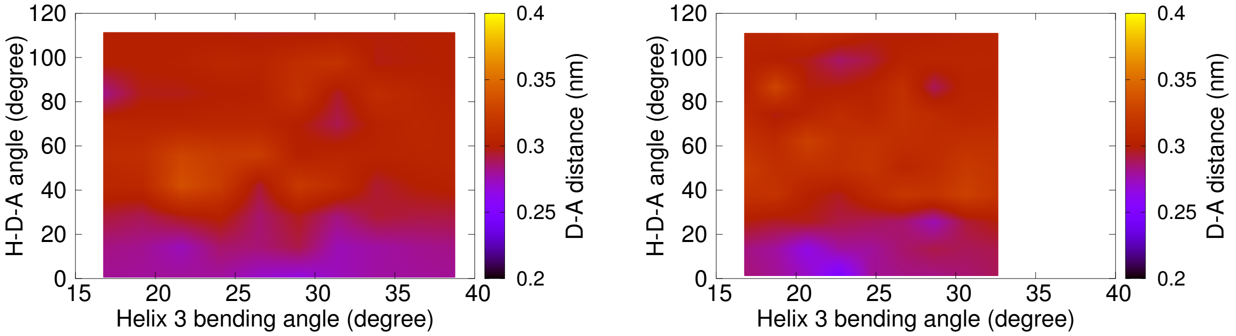


(c) monomer 1: T127-(OE1)E123 **(d)** monomer 2: T127-(OE1)E123

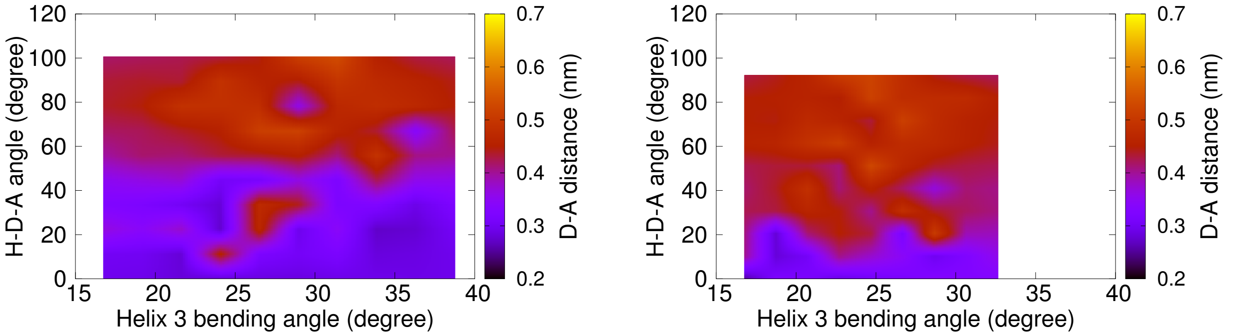


(e) monomer 1: T127-(OE2)E123 **(f)** monomer 2: T127-(OE2)E123

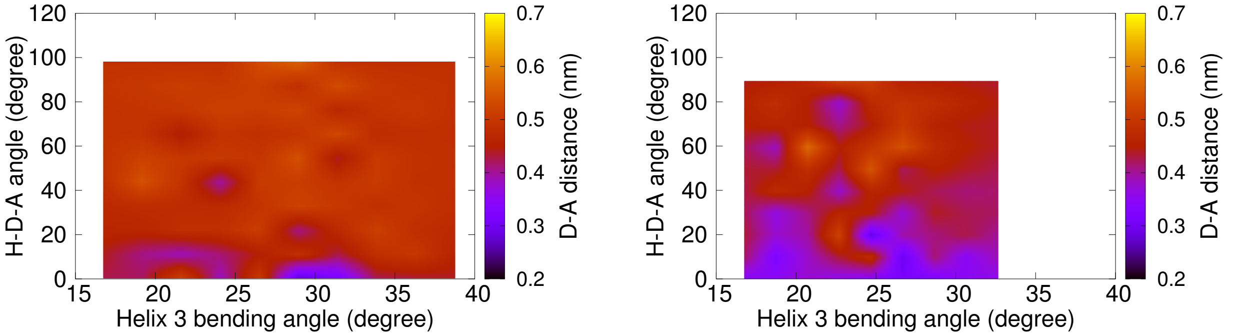
Figure A.10: Correlation between helix 3 bending and T127–E123 interactions. (a) Interaction between T127 and E123 backbone in monomer 1; (b) Interaction between T127 and E123 backbone in monomer 2; (c,e) Interaction between T127 and E123 side chain in monomer 1; (d,f) Interaction between T127 and E123 side chain in monomer 2. The horizontal axis is the helix C bending angle; the vertical axis is the H-D-A angle; the color bar represent the D-A distance.



(a) monomer 1: T127-(backbone)E123 **(b)** monomer 2: T127-(backbone)E123

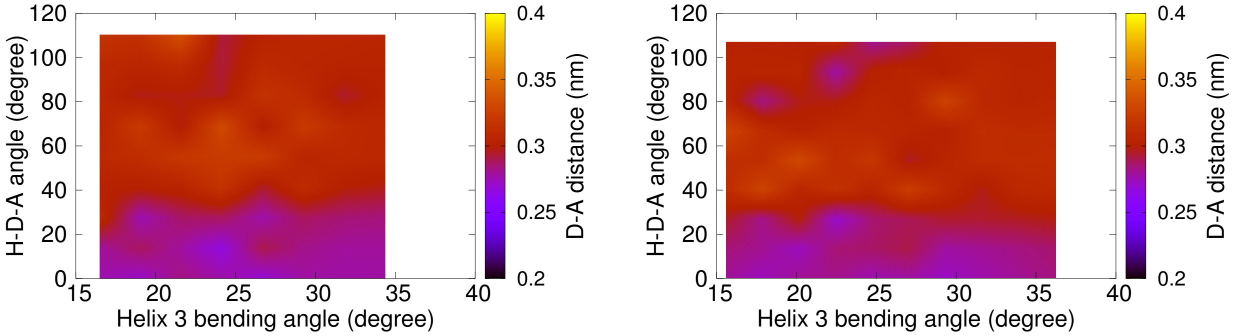


(c) monomer 1: T127-(OE1)E123 **(d)** monomer 2: T127-(OE1)E123



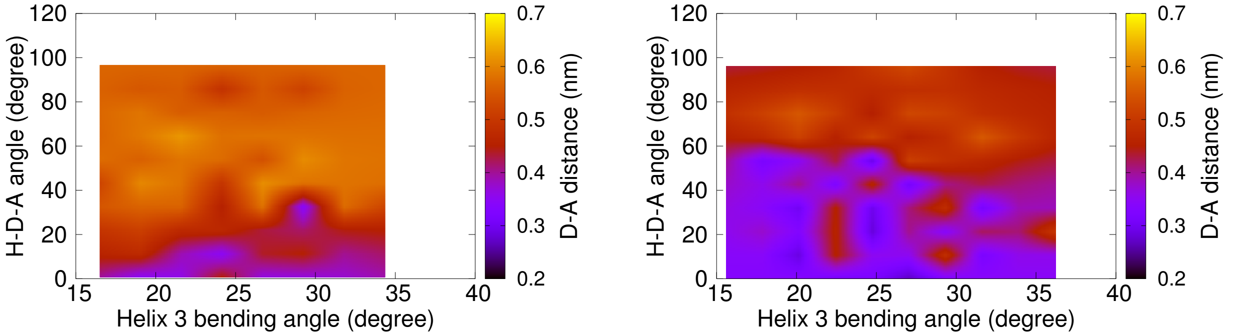
(e) monomer 1: T127-(OE2)E123 **(f)** monomer 2: T127-(OE2)E123

Figure A.11: Correlation between helix 3 bending and T127–E123 interactions. (a) Interaction between T127 and E123 backbone in monomer 1; (b) Interaction between T127 and E123 backbone in monomer 2; (c,e) Interaction between T127 and E123 side chain in monomer 1; (d,f) Interaction between T127 and E123 side chain in monomer 2. The horizontal axis is the helix C bending angle; the vertical axis is the H-D-A angle; the color bar represent the D-A distance.



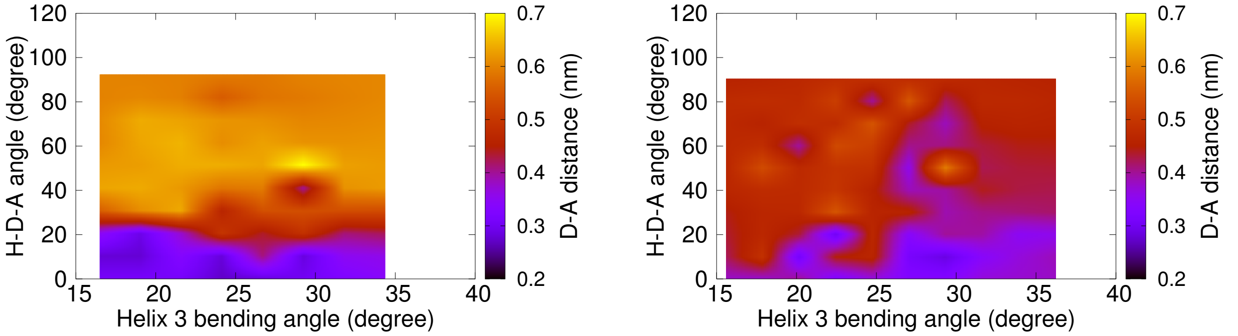
(a) monomer 1: T127-(backbone)E123

(b) monomer 2: T127-(backbone)E123



(c) monomer 1: T127-(OE1)E123

(d) monomer 2: T127-(OE1)E123



(e) monomer 1: T127-(OE2)E123

(f) monomer 2: T127-(OE2)E123

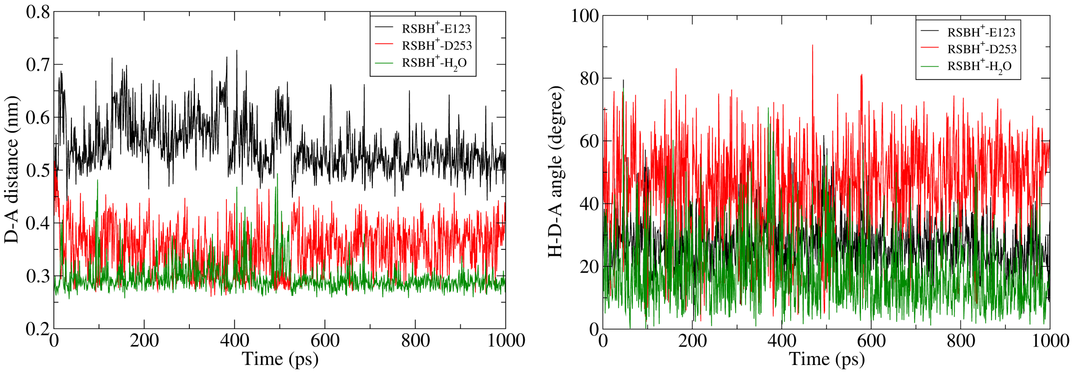
Figure A.12: Correlation between helix 3 bending and T127–E123 interactions. (a) Interaction between T127 and E123 backbone in monomer 1; (b) Interaction between T127 and E123 backbone in monomer 2; (c,e) Interaction between T127 and E123 side chain in monomer 1; (d,f) Interaction between T127 and E123 side chain in monomer 2. The horizontal axis is the helix C bending angle; the vertical axis is the H-D-A angle; the color bar represent the D-A distance.

Appendix B

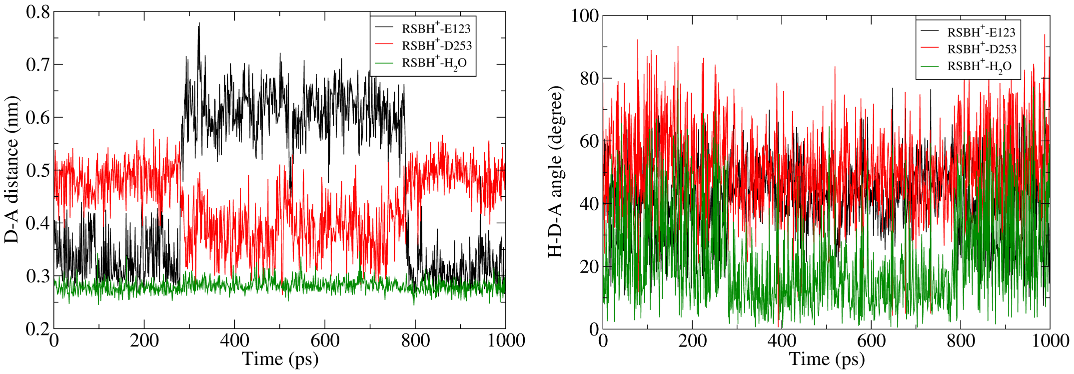
Appendices of Chapter 5

Table B.1: Excitation energies for the five QM regions shown in Table 4.2 with and without MM charges.

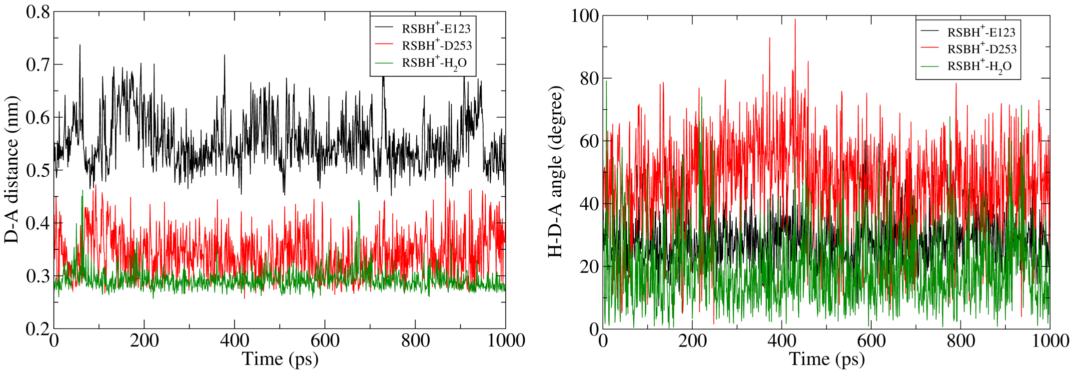
QM-1					
MM charges	vacuum	ΔE	MM charges	vacuum	ΔE
3.16	2.35	0.81	3.11	2.32	0.79
3.11	2.33	0.78	3.12	2.32	0.80
2.98	2.32	0.66	3.17	2.34	0.83
3.19	2.33	0.86	3.19	2.34	0.85
3.08	2.33	0.75	3.08	2.33	0.75
3.19	2.33	0.86	3.10	2.35	0.75
QM-2					
MM charges	vacuum	ΔE	MM charges	vacuum	ΔE
3.16	2.33	0.83	3.10	2.29	0.81
3.05	2.38	0.67	3.10	2.28	0.82
2.93	2.26	0.67	3.13	2.27	0.86
3.19	2.28	0.91	3.19	2.33	0.86
3.04	2.32	0.72	3.05	2.26	0.79
3.15	2.32	0.83	3.07	2.29	0.78
QM-3					
MM charges	vacuum	ΔE	MM charges	vacuum	ΔE
3.22	3.08	0.14	3.16	3.04	0.12
3.19	2.89	0.30	3.17	3.02	0.15
3.05	3.01	0.04	3.22	3.08	0.14
3.23	3.01	0.21	3.24	3.08	0.16
3.14	3.04	0.10	3.14	3.08	0.06
3.24	3.06	0.18	3.16	3.09	0.07
QM-4					
MM charges	vacuum	ΔE	MM charges	vacuum	ΔE
3.23	2.21	1.02	3.17	2.09	1.08
3.15	3.06	0.09	3.16	2.28	0.88
3.02	3.18	-0.16	3.20	2.39	0.81
3.22	2.49	0.73	3.25	2.20	1.05
3.12	2.80	0.32	3.12	3.56	-0.44
3.21	2.90	0.31	3.14	3.44	-0.30
QM-5					
MM charges	vacuum	ΔE	MM charges	vacuum	ΔE
3.18	1.70	1.48	3.14	1.66	1.48
3.15	1.45	1.70	3.14	1.78	1.36
3.02	1.59	1.43	3.20	1.79	1.41
3.18	1.98	1.20	3.20	1.64	1.56
3.09	1.51	1.58	3.11	1.53	1.58
3.19	1.67	1.52	3.12	1.56	1.56



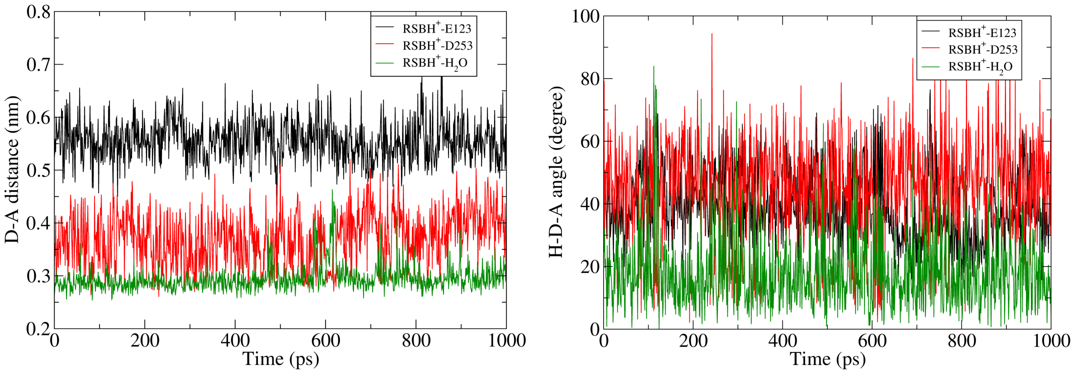
(a) simulation 1 **(b)** simulation 1



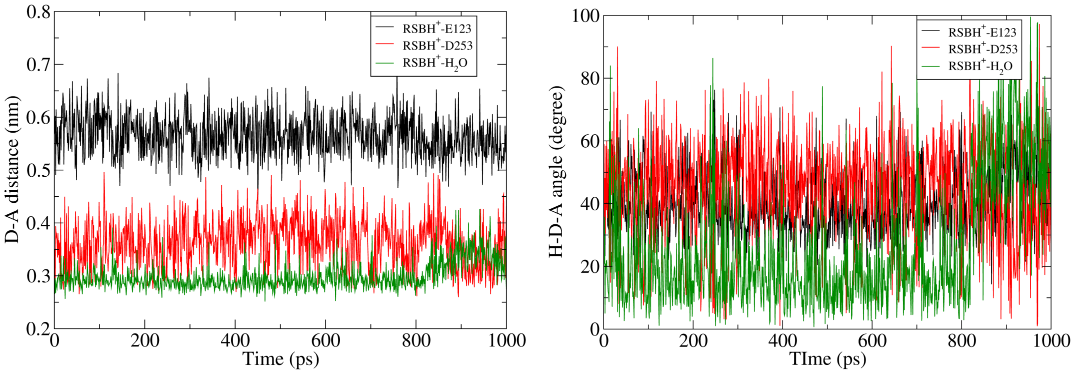
(c) simulation 2 **(d)** simulation 2



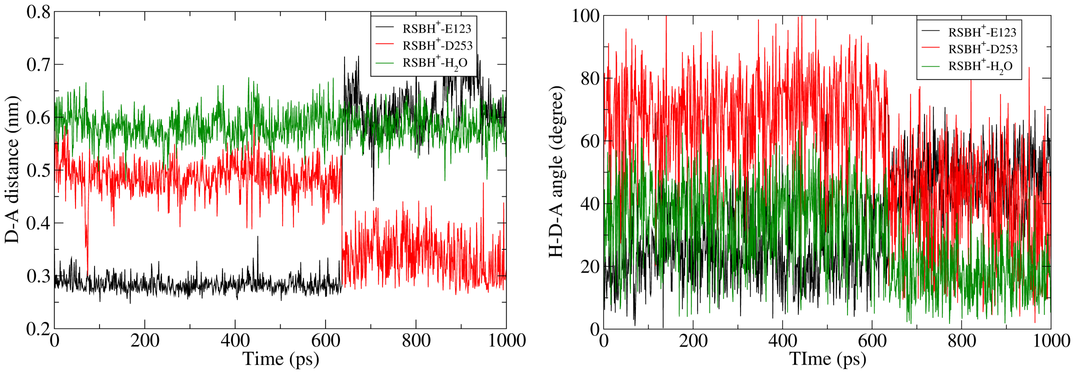
(e) simulation 3 **(f)** simulation 3



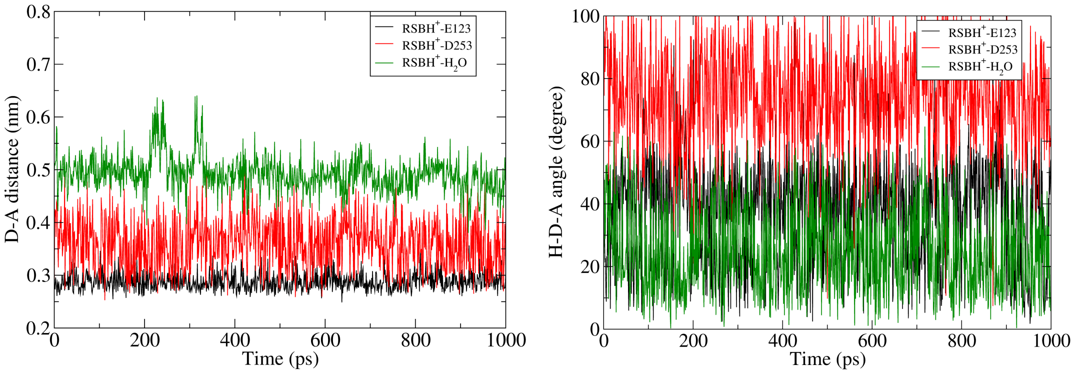
(g) simulation 4 **(h)** simulation 4



(i) simulation 5 **(j)**

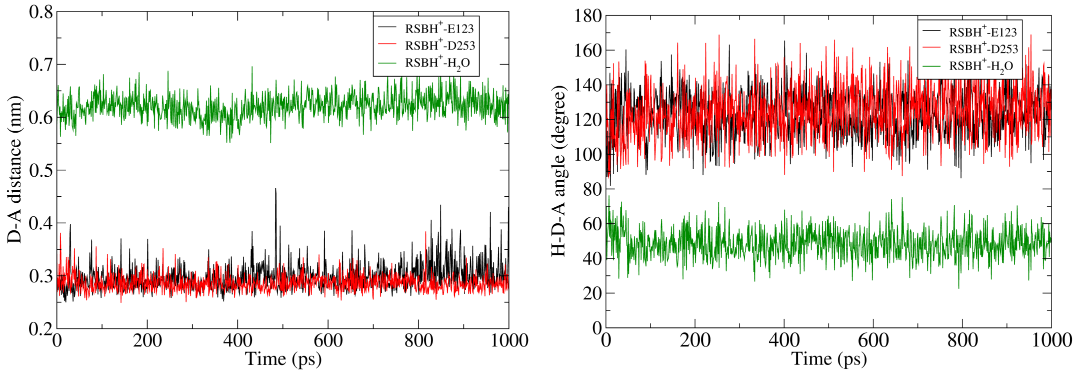


(k) simulation 6 **(l)** simulation 6

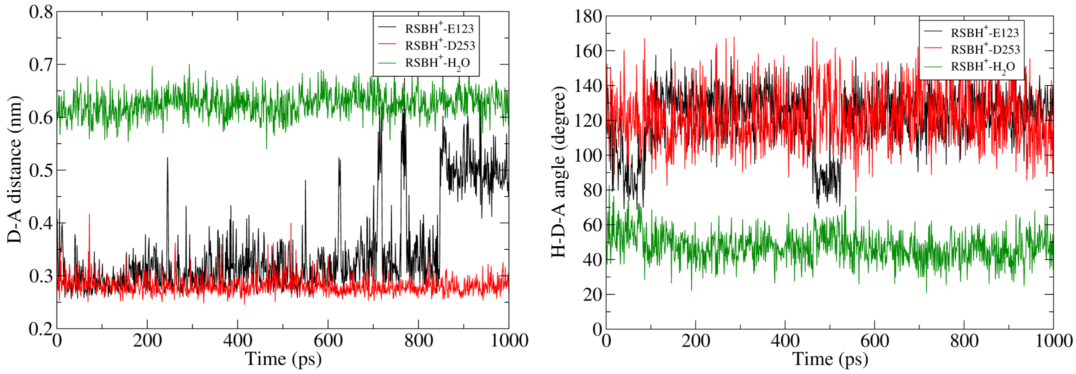


(m) simulation 7 **(n)** simulation 7

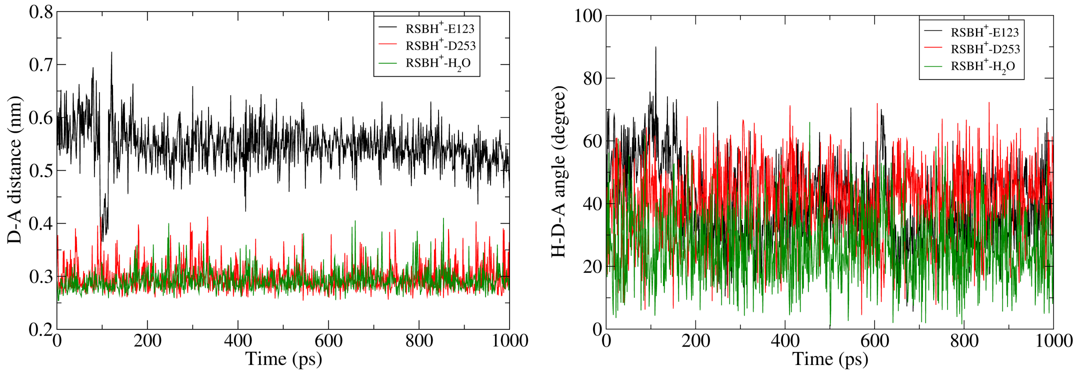
Figure B.1: Stability of hydrogen bonding patterns in ChR2-C128T bound with all-*trans*,15-*anti* retinal in seven trajectories, 1 ns of each. Please see the result of the other trajectory in the main text.



(a) simulation 1 **(b)** simulation 1



(c) simulation 2 **(d)** simulation 2



(e) simulation 3 **(f)** simulation 3

Table B.2: Fitting parameters used to convolute the OM2/MRCI histograms into Gaussian functions of Figures 5.15, 5.16 in the main text.

Gaussian function: $y=-A_0*\exp(A_2*(A1-x)^2)$

		A_0	A_1	A_2	Fitting parameter R^2	
BR	*trans*	total	-0.248719	2.62184	-25.5929	0.997833
ChR2-C128T	*trans*+*cis*[a]	total[d]	-0.19388	3.06225	-15.2799	0.99131
		E123[e]	-0.0288588	3.15394	-17.058	0.997762
		D253[f]	-0.0380906	3.07066	-16.6173	0.994485
		H_2O[g]	-0.132155	3.04065	-15.8689	0.999090
		upward[h]	-0.039007	3.15495	-17.5993	0.998072
		downward[i]	-0.161623	3.04139	-16.138	0.998690
	trans[b]	total	-0.203821	3.09375	-16.0137	0.999303
		E123	-0.0505065	3.15373	-17.0181	0.997741
		D253	-0.0337882	3.09147	-17.8271	0.992998
		H_2O	-0.124503	3.07004	-16.5484	0.997336
		upward	-0.0622742	3.15965	-17.6607	0.997861
		downward	-0.148302	3.06711	-16.9397	0.999051
	cis[c]	total	-0.191343	3.01701	-16.109	0.997703
		D253	-0.0442035	3.04954	-15.7172	0.985149
		H_2O	-0.148484	3.00725	-16.6555	0.997210
		upward	-0.00855356	3.10777	-18.9443	0.978048
		downward	-0.184176	3.01285	-16.2734	0.997784

[a] *trans*+*cis*: snapshots with mixed all-*trans*,15-*anti* and 13-*cis*,15-*syn* retinals.

[b] *trans*: snapshots with all-*trans*,15-*anti* retinals.

[c] *cis*: snapshots with 13-*cis*,15-*syn* retinals.

[d] total: all snapshots.

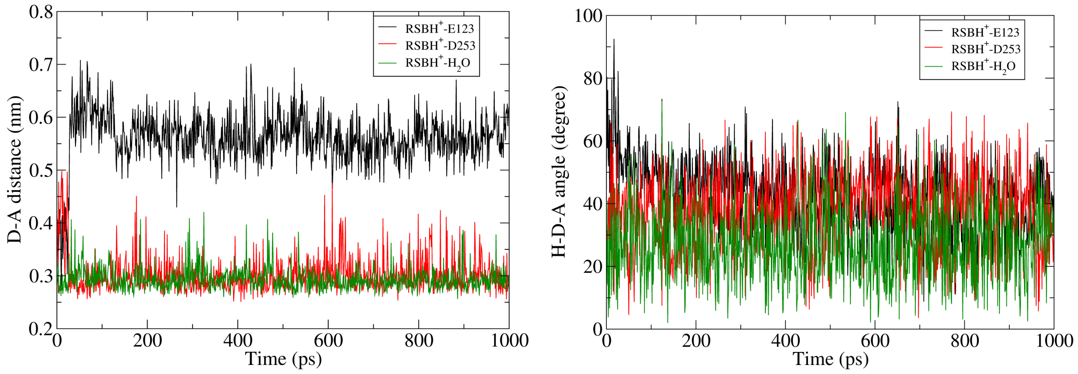
[e] E123: snapshots with -RSBH$^+\cdots{}^-$O-(E123) hydrogen bond.

[f] D253: snapshots with -RSBH$^+\cdots{}^-$O-(D253) hydrogen bond.

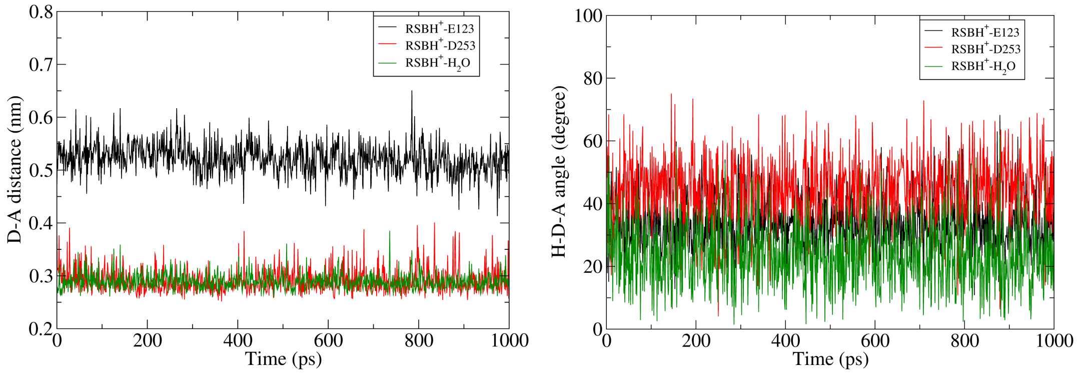
[g] H_2O: snapshots with -RSBH$^+\cdots OH_2$ hydrogen bond.

[h] upward: snapshots with E123-upward conformation.

[i] downward: snapshots with E123-downward conformation.



(g) simulation 4 **(h)** simulation 4



(i) simulation 5 **(j)** simulation 5

Figure B.2: Stability of hydrogen bonding patterns in ChR2-C128T bound with 13-*cis*,15-*syn* retinal in five trajectories, 1 ns of each. Please see the result of the other trajectory in the main text.

Appendix C

Appendices of Chapter 6

Table C.1: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-1* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au. Unit of dihedral angles: degree. Unit of energy: Hartree.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
FC(178)	S_0	S_0	-832.43214724	-832.43699735	-829.826855
		S_1	-832.33784254	-832.34169973	-829.685145
		S_2	-832.30462097	-832.29591368	-829.663438
173	S_1	S_0	-832.41176365	-832.42430900	-829.783619
		S_1	-832.35884712	-832.35475789	-829.719888
		S_2	-832.33203224	-832.32357611	-829.683861
168	S_1	S_0	-832.41198137	-832.42450451	-829.783706
		S_1	-832.35882230	-832.35481015	-829.719944
		S_2	-832.33265516	-832.32414417	-829.683901
163	S_1	S_0	-832.41157093	-832.42409073	-829.783056
		S_1	-832.35835014	-832.35412091	-829.719493
		S_2	-832.33149741	-832.32320684	-829.683268
158	S_1	S_0	-832.41082174	-832.42340973	-829.781999
		S_1	-832.35773489	-832.35325300	-829.718879
		S_2	-832.33036273	-832.32225663	-829.682331
153	S_1	S_0	-832.41085102	-832.42398728	-829.781548
		S_1	-832.35804984	-832.35324518	-829.719352
		S_2	-832.33071651	-832.32238492	-829.682191
148	S_1	S_0	-832.40898281	-832.42260446	-829.779170
		S_1	-832.35716483	-832.35290138	-829.718455
		S_2	-832.33505127	-832.32569308	-829.680217
143	S_1	S_0	-832.40668917	-832.43691368	-829.776384
		S_1	-832.35617320	-832.40321358	-829.717354
		S_2	-832.42111745	-832.34385255	-829.677901
138	S_1	S_0	-832.40179521	-832.41812842	-829.770880

		S_1	-832.35596323	-832.34753635	-829.716651
		S_2	-832.32334253	-832.31543619	-829.674077
		S_0	-832.39841749	-832.41575127	-829.766943
133	S_1	S_1	-832.35496198	-832.34543297	-829.715424
		S_2	-832.32217960	-832.31437483	-829.671124
		S_0	-832.38820107	-832.41606100	-829.757119
128	S_1	S_1	-832.35645899	-832.36570667	-829.714842
		S_2	-832.37179911	-832.33469150	-829.666105
		S_0	-832.38796653	-833.19767444	-829.755933
123	S_1	S_1	-832.35422896	-832.40558632	-829.713104
		S_2	-833.19702335	-832.33595809	-829.664009
		S_0	-832.40212231	-832.40960520	-829.783972
118	S_0	S_1	-832.32468766	-832.32039625	-829.677299
		S_2	-832.30428666	-832.30109519	-829.653144

Table C.2: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-1* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G$*$. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Units of energy: Kcal/mol. The relative energies are calculated with respect to FC(S_0). CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
		S_0	0	0	0
FC(178)	S_0	S_1	59.18	59.80	88.92
		S_2	80.02	88.53	102.55
		S_0	12.79	7.96	27.13
173	S_1	S_1	46.00	51.61	67.12
		S_2	62.82	71.17	89.73
		S_0	12.65	7.84	27.08
168	S_1	S_1	46.01	51.57	67.09
		S_2	62.43	70.82	89.70
		S_0	12.91	8.10	27.48
163	S_1	S_1	46.31	52.01	67.37
		S_2	63.16	71.40	90.10
		S_0	13.38	8.53	28.15
158	S_1	S_1	46.69	52.55	67.76
		S_2	63.87	72.00	90.69
		S_0	13.36	8.16	28.43
153	S_1	S_1	46.50	52.56	67.46
		S_2	63.65	71.92	90.78
		S_0	14.54	9.03	29.92
148	S_1	S_1	47.05	52.77	68.02
		S_2	60.93	69.84	92.02

143	S$_1$	S$_0$	15.98	0.05	31.67
		S$_1$	47.67	21.20	68.71
		S$_2$	6.92	58.45	93.47
138	S$_1$	S$_0$	19.05	11.84	35.12
		S$_1$	47.81	56.14	69.15
		S$_2$	68.28	76.28	95.87
133	S$_1$	S$_0$	21.17	13.33	37.60
		S$_1$	48.43	57.46	69.92
		S$_2$	69.01	76.95	97.72
128	S$_1$	S$_0$	27.58	13.14	43.76
		S$_1$	47.50	44.74	70.29
		S$_2$	37.87	64.20	100.87
123	S$_1$	S$_0$	27.72	-477.33	44.50
		S$_1$	48.89	19.71	71.38
		S$_2$	-479.97	63.40	102.19
118	S$_0$	S$_0$	18.84	17.19	26.91
		S$_1$	67.43	73.17	93.85
		S$_2$	80.23	85.28	109.01

Table C.3: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-2* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Unit of energy: Hartree. CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
FC(178)	S$_0$	S$_0$	-832.43984763	-832.44475039	-829.834611
		S$_1$	-832.34537493	-832.34943412	-829.692662
		S$_2$	-832.31237433	-832.30341238	-829.671128
173	S$_1$	S$_0$	-832.41939958	-832.43209498	-829.791341
		S$_1$	-832.36648229	-832.36260074	-829.727487
		S$_2$	-832.34061245	-832.33179860	-829.691590
168	S$_1$	S$_0$	-832.41963545	-832.43234673	-829.791440
		S$_1$	-832.36647559	-832.36279485	-829.727552
		S$_2$	-832.34193149	-832.33290095	-829.691645
163	S$_1$	S$_0$	-832.41930499	-832.43198094	-829.790874
		S$_1$	-832.36607583	-832.36202195	-829.727183
		S$_2$	-832.34000576	-832.33138368	-829.691104
158	S$_1$	S$_0$	-832.41855874	-832.43126996	-829.789821
		S$_1$	-832.36544443	-832.36103254	-829.726555
		S$_2$	-832.33821610	-832.32991677	-829.690156
153	S$_1$	S$_0$	-832.41903380	-832.43210278	-829.789823

		S_1	-832.36590547	-832.36115933	-829.727201
		S_2	-832.33851798	-832.33019514	-829.690423
		S_0	-832.41677944	-832.43053802	-829.787117
148	S_1	S_1	-832.36499431	-832.36034069	-829.726193
		S_2	-832.34076373	-832.33165877	-829.688175
		S_0	-832.41443393	-832.51997666	-829.784254
143	S_1	S_1	-832.36398955	-832.42219779	-829.725085
		S_2	-832.51618999	-832.35243902	-829.685792
		S_0	-832.40944614	-832.42612596	-829.778627
138	S_1	S_1	-832.36402552	-832.35560528	-829.724626
		S_2	-832.33245282	-832.32419324	-829.682058
		S_0	-832.40634095	-832.42392047	-829.775030
133	S_1	S_1	-832.36304183	-832.35336352	-829.723379
		S_2	-832.32975728	-832.32185607	-829.679288
		S_0	-832.39625896	-832.42397976	-829.765266
128	S_1	S_1	-832.36473551	-832.37215102	-829.722987
		S_2	-832.37779117	-832.34265486	-829.674337
		S_0	-832.39568770	-842.23832071	-829.763853
123	S_1	S_1	-832.36249300	-832.41409562	-829.721154
		S_2	-842.23830074	-832.34406510	-829.672182
		S_0	-832.41024554	-832.41776683	-829.792159
118	S_0	S_1	-832.33277027	-832.32848701	-829.685365
		S_2	-832.31243441	-832.30919638	-829.661407

Table C.4: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-2* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Units of energy: kcal/mol. The relative energies are calculated with respect to FC(S_0). CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
		S_0	0	0	0
FC(178)	S_0	S_1	59.28	59.81	89.07
		S_2	79.99	88.69	102.59
		S_0	12.83	7.94	27.15
173	S_1	S_1	46.04	51.55	67.22
		S_2	62.27	70.88	89.75
		S_0	12.68	7.78	27.09
168	S_1	S_1	46.04	51.43	67.18
		S_2	61.44	70.19	89.71
		S_0	12.89	8.01	27.45
163	S_1	S_1	46.29	51.91	67.41
		S_2	62.65	71.14	90.05

		S_0	13.36	8.46	28.11
158	S_1	S_1	46.69	52.53	67.81
		S_2	63.77	72.06	90.65
		S_0	13.06	7.94	28.10
153	S_1	S_1	46.40	52.45	67.40
		S_2	63.59	71.88	90.48
		S_0	14.48	8.92	29.80
148	S_1	S_1	46.97	52.97	68.03
		S_2	62.18	70.97	91.89
		S_0	15.95	-47.21	31.60
143	S_1	S_1	47.60	14.15	68.73
		S_2	-47.91	57.93	93.39
		S_0	19.08	11.69	35.13
138	S_1	S_1	47.58	55.94	69.02
		S_2	67.39	75.65	95.73
		S_0	21.03	13.07	37.39
133	S_1	S_1	48.20	57.35	69.80
		S_2	69.08	77.12	97.47
		S_0	27.35	13.03	43.51
128	S_1	S_1	47.13	45.56	70.05
		S_2	38.94	64.07	100.57
		S_0	27.71	-6145.55	44.40
123	S_1	S_1	48.54	19.24	71.20
		S_2	-6148.62	63.18	101.93
		S_0	18.58	16.93	26.64
118	S_0	S_1	67.19	72.96	93.65
		S_2	79.95	85.06	108.69

Table C.5: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-3* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Unit of energy: Hartree. CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
		S_0	-832.70167455	-832.70749315	-830.099203
FC(178)	S_0	S_1	-832.59568319	-832.61000963	-829.948926
		S_2	-832.57920277	-832.55905772	-829.929538
		S_0	-832.67586417	-832.69093162	-830.050370
173	S_1	S_1	-832.62250287	-832.62140727	-829.982339
		S_2	-832.59714141	-832.58316956	-829.949872
		S_0	-832.67616277	-832.69127400	-830.050535
168	S_1				

		S_1	-832.62262860	-832.62144344	-829.982545
		S_2	-832.59735001	-832.58342394	-829.950011
		S_0	-832.67571665	-832.69092291	-830.049871
163	S_1	S_1	-832.62220701	-832.62080671	-829.982150
		S_2	-832.59682919	-832.58302323	-829.949419
		S_0	-832.67487531	-832.69026103	-830.048739
158	S_1	S_1	-832.62163652	-832.61986859	-829.981553
		S_2	-832.59595125	-832.58233347	-829.948487
		S_0	-832.67513172	-832.69100844	-830.048529
153	S_1	S_1	-832.62225689	-832.61974697	-829.982456
		S_2	-832.59614207	-832.58277527	-829.948731
		S_0	-832.67276900	-832.68939179	-830.045851
148	S_1	S_1	-832.62155712	-832.61813570	-829.981498
		S_2	-832.59408714	-832.58088576	-829.946672
		S_0	-832.67029908	-832.68764722	-830.042991
143	S_1	S_1	-832.62078602	-832.61634062	-829.980436
		S_2	-832.59199049	-832.57908775	-829.944505
		S_0	-832.66514068	-833.04598296	-830.037765
138	S_1	S_1	-832.62158072	-832.67709617	-829.980232
		S_2	-833.04372323	-832.60736551	-829.941564
		S_0	-832.66178554	-832.68314253	-830.034192
133	S_1	S_1	-832.62087270	-832.61203676	-829.978928
		S_2	-832.58676713	-832.57424608	-829.939073
		S_0	-832.65125904	-832.68159586	-830.025293
128	S_1	S_1	-832.62317826	-832.61859983	-829.978315
		S_2	-832.61942215	-832.59366377	-829.935549
		S_0	-832.65049096	-832.67824203	-830.023694
123	S_1	S_1	-832.62118043	-832.61093511	-829.976513
		S_2	-832.60248557	-832.58497982	-829.933365
		S_0	-832.66786968	-832.67721693	-830.052016
118	S_0	S_1	-832.58694829	-832.58480403	-829.939235
		S_2	-832.56942983	-832.56222685	-829.922276

Table C.6: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-3* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Units of energy: kcal/mol. The relative energies are calculated with respect to FC(S_0). CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
		S_0	0	0	0
FC(178)	S_0	S_1	66.51	61.17	94.30
		S_2	76.85	93.14	106.47

173	S_1	S_0	12.20	10.39	30.64
		S_1	49.68	54.02	73.33
		S_2	65.60	78.01	93.71
168	S_1	S_0	16.01	10.18	30.54
		S_1	49.60	54.00	73.20
		S_2	65.46	77.85	93.62
163	S_1	S_0	16.29	10.40	30.96
		S_1	49.87	54.40	73.45
		S_2	65.79	78.11	93.99
158	S_1	S_0	16.82	10.81	31.67
		S_1	50.22	54.99	73.83
		S_2	66.34	78.54	94.58
153	S_1	S_0	16.66	10.34	31.80
		S_1	49.84	55.06	73.26
		S_2	66.22	78.26	94.42
148	S_1	S_0	18.14	11.36	33.48
		S_1	56.07	52.97	73.86
		S_2	79.45	70.97	95.71
143	S_1	S_0	19.69	12.45	35.27
		S_1	50.76	57.20	74.53
		S_2	68.83	80.58	97.07
138	S_1	S_0	22.93	-212.41	38.55
		S_1	50.26	19.07	74.66
		S_2	-214.64	62.83	98.92
133	S_1	S_0	25.03	15.28	40.80
		S_1	50.70	59.90	75.47
		S_2	72.11	83.61	100.48
128	S_1	S_0	31.64	16.25	46.38
		S_1	49.26	55.78	75.86
		S_2	51.61	71.43	102.69
123	S_1	S_0	32.12	18.36	47.38
		S_1	50.51	60.59	76.99
		S_2	62.24	76.88	104.07
118	S_0	S_0	21.21	19.00	29.61
		S_1	71.99	76.99	100.38
		S_2	82.98	91.16	111.02

Table C.7: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-4* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Unit of energy: Hartree. CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
FC(178)	S_0	S_0	-832.69432500	-832.69994420	-830.092078
		S_1	-832.58668187	-832.60185962	-829.941398
		S_2	-832.57247493	-832.55167798	-829.921502
173	S_1	S_0	-832.65122414	-832.66736761	-830.026210
		S_1	-832.59801816	-832.59732990	-829.957425
		S_2	-832.57333242	-832.55787720	-829.925438
168	S_1	S_0	-832.65160516	-832.66781869	-830.026457
		S_1	-832.59827639	-832.59744656	-829.957792
		S_2	-832.57360821	-832.55822451	-829.925669
163	S_1	S_0	-832.65132339	-832.66770022	-830.025985
		S_1	-832.59804228	-832.59701402	-829.957561
		S_2	-832.57329832	-832.55794975	-829.925252
158	S_1	S_0	-832.65062945	-832.66725845	-830.025025
		S_1	-832.59768662	-832.59627828	-829.957152
		S_2	-832.57264623	-832.55742556	-829.924550
153	S_1	S_0	-832.65173676	-832.66846763	-830.025394
		S_1	-832.59927654	-832.59667436	-829.959248
		S_2	-832.57327496	-832.55914626	-829.925677
148	S_1	S_0	-832.64962713	-832.66712710	-830.022985
		S_1	-832.59886702	-832.59532742	-829.958568
		S_2	-832.57150628	-832.55754591	-829.923921
143	S_1	S_0	-832.64747573	-832.66572078	-830.020468
		S_1	-832.59844419	-832.59388913	-829.957823
		S_2	-832.56979942	-832.55610943	-829.922123
138	S_1	S_0	-832.64291639	-836.14633099	-830.016105
		S_1	-832.59989774	-832.65680499	-829.958197
		S_2	-836.14610459	-832.58578275	-829.920033
133	S_1	S_0	-832.63950598	-832.66275162	-830.013226
		S_1	-832.59940786	-832.59150015	-829.956591
		S_2	-832.56836984	-832.55303191	-829.918081
128	S_1	S_0	-832.63009439	-832.66199059	-830.006079
		S_1	-832.60272335	-832.59795020	-829.956718
		S_2	-832.59721990	-832.57009684	-829.916406
123	S_1	S_0	-832.62889534	-832.65892018	-830.004039
		S_1	-832.60052842	-832.59241911	-829.954610
		S_2	-832.58550070	-832.56358517	-829.913832

118	S_0	S_0	-832.65000816	-832.66073190	-830.040587
		S_1	-832.55573690	-832.56775695	-829.908940
		S_2	-832.55257442	-832.52983064	-829.904772

Table C.8: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-4* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Units of energy: kcal/mol. The relative energies are calculated with respect to FC(S_0). CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
FC(178)	S_0	S_0	0	0	0
		S_1	67.55	61.55	94.55
		S_2	76.46	93.04	107.04
173	S_1	S_0	27.05	20.44	41.33
		S_1	60.43	64.39	84.50
		S_2	75.92	89.15	104.57
168	S_1	S_0	26.81	20.16	41.18
		S_1	60.27	64.32	84.27
		S_2	75.75	88.93	104.42
163	S_1	S_0	26.98	20.23	41.47
		S_1	60.42	64.59	84.41
		S_2	75.95	89.10	104.69
158	S_1	S_0	27.42	20.51	42.08
		S_1	60.64	65.05	84.67
		S_2	76.35	89.43	105.13
153	S_1	S_0	26.72	19.75	41.84
		S_1	59.64	64.80	83.35
		S_2	75.96	88.35	104.42
148	S_1	S_0	28.05	20.59	43.36
		S_1	59.90	65.65	83.78
		S_2	77.07	89.36	105.52
143	S_1	S_0	29.40	21.48	44.94
		S_1	60.17	66.55	84.25
		S_2	78.14	90.26	106.65
138	S_1	S_0	32.26	-2162.64	47.67
		S_1	59.25	27.07	84.01
		S_2	-2166.02	71.64	107.96
133	S_1	S_0	34.40	23.34	49.48
		S_1	59.56	68.05	85.02
		S_2	79.04	92.19	109.19
128	S_1	S_0	40.31	23.82	53.97

		S_1	57.48	64.00	84.94
		S_2	60.93	81.48	110.24
		S_0	41.06	25.74	55.24
123	S_1	S_1	58.86	67.47	86.26
		S_2	68.29	85.57	111.85
		S_0	27.81	24.61	32.31
118	S_0	S_1	86.96	82.95	114.92
		S_2	88.95	106.75	117.54

Table C.9: CASPT2 and CASSCF vertical excitation energies of geometries based on the relaxed scan scheme *RS-1* (see main text for the details) of the ChR2-WT RSBH-E123 model. Basis set: 6-31G*. IPEA shift = 0.0 au, LS = 0.2 au. Unit of dihedral angles: degree. Units of energy: kcal/mol. The relative energies are calculated with respect to FC(S_0). CASSCF wavefunctions were obatined from state-average calculations over three roots with equal weights.

dihedral -C13=C14-	opt. state	electronic states	E_{CASPT2}	$E_{MS-CASPT2}$	E_{CASSCF}
		S_0	0	0	0
FC(178)	S_0	S_1	59.78	60.35	88.92
		S_2	80.80	88.15	102.55
		S_0	13.01	8.97	27.13
173	S_1	S_1	46.45	51.75	67.12
		S_2	65.94	72.59	89.73
		S_0	12.87	8.85	27.08
168	S_1	S_1	46.47	51.77	67.09
		S_2	65.85	72.48	89.70
		S_0	13.13	9.10	27.48
163	S_1	S_1	46.77	52.13	67.37
		S_2	66.17	72.75	90.10
		S_0	13.60	9.53	28.15
158	S_1	S_1	47.16	52.63	67.76
		S_2	66.68	73.20	90.69
		S_0	13.59	9.23	28.43
153	S_1	S_1	46.97	52.64	67.46
		S_2	66.66	73.27	90.78
		S_0	14.76	10.27	29.92
148	S_1	S_1	47.53	53.51	68.02
		S_2	67.86	74.30	92.02
		S_0	16.21	11.45	31.67
143	S_1	S_1	48.16	54.52	68.71
		S_2	69.29	75.61	93.47
		S_0	19.30	13.20	35.12
138	S_1	S_1	48.31	55.99	69.15
		S_2	72.01	78.33	95.87

133	S_1	S_0	21.43	14.82	37.60
		S_1	48.95	57.34	69.92
		S_2	74.02	80.16	97.72
128	S_1	S_0	27.86	17.89	43.76
		S_1	48.02	60.12	70.29
		S_2	79.60	85.40	100.87
123	S_1	S_0	28.01	18.75	44.50
		S_1	49.42	60.94	71.38
		S_2	80.15	85.81	102.19
118	S_0	S_0	20.31	19.07	26.82
		S_1	70.18	76.05	97.97
		S_2	88.89	92.18	112.25

Abbreviations

AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid
AMPARs	AMPA Receptors
BLA	Bond Length Alternation
BR	Bacterhodopsin
CASPT2	Complete Active Space Second Order Perturbation Theory
CASSCF	Complete Active Space Self Consistent Field
ChR1	Channelrhodopsin-1
ChR2	Channelrhodopsin-2
C1C2	Channelrhodopsin-1/Channelrhodopsin-2 chimera
CI	Conical Intersection
DFT	Density Functional Theory
DFTB3	Extended self-consistent charge density functional tight-binding
FC	Frank Condon
GluK2	Subtype of Kainate Receptor
HOOP	Hydrogen-Out-Of-Plane
iGluRs	Ionotropic Glutamate Receptors
IR	Infared Radiation
IPEA	Ionization potential - Electron Affinity
KAR	Kainate Receptors
LBD	Ligand Binding Domain
LS	Level Shift
MD	Molecular Dynamics
MM	Molecular Mechanics
MSCASPT2	multi-state CASPT2
NMDA	N-methyl-D-aspartate
NMDARs	NMDA Receptors
NTD	amino-terminal domain
PCL	Photochromic Ligand
QM	Quantum Mechanics
RAS	Restricted Active Space
RSBH$^+$	protonated retinal Schiff base
SA-CASSCF	state-average CASSCF
TMD	Transmembrane Domain

Publications and Copyright

List of Publications

1. Guo, Y.; Beyle, F. E.; Bold, B. M.; Watanabe, H. C.; Koslowski, A.; Thiel, W.; Hegemann, P.; Marazzi, M.; Elstner, M. “Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant.” *Chem. Sci.*, 2016, 7, 3879.

2. Guo, Y.; Wolter, T.; Kubař, T.; Sumser, M.; Trauner, D,; Elstner, M. “Molecular dynamics investigation of *gluazo*, a photo-switchable ligand for the glutamate receptor GluK2.” *PLoS ONE*, 2015, 10, e0135388.

3. Vogt, A.; Guo, Y.; Tsunoda, S. P.; Kateriya, S.; Elstner, M.; Hegemann, P. “Conversion of a light-driven proton pump into a light-gated ion channel.” *Sci. Rep.*, 2015, 5, 16450.

Copyright

⇒ Chapter 3 is reproduced with permission from Guo, Y.; Wolter, T.; Kubář, T.; Sumser, M.; Trauner, D,; Elstner, M. “Molecular dynamics investigation of *gluazo*, a photo-switchable ligand for the glutamate receptor GluK2.” *PLoS ONE*, 2015, 10, e0135388.

⇒ Chapter 5 is reproduced in part with permission from Guo, Y.; Beyle, F. E.; Bold, B. M.; Watanabe, H. C.; Koslowski, A.; Thiel, W.; Hegemann, P.; Marazzi, M.; Elstner, M. “Active site structure and absorption spectrum of channelrhodopsin-2 wild-type and C128T mutant.” *Chem. Sci.*, 2016, 7, 3879. This work was a coorperation with Franziska E. Beyle, where both authors contributed equally. The work was shared as follows: Franziska E. Beyle conducted the simulations and part of the analysis of ChR2 WT, while Yanan Guo performed the simulations and analysis of ChR2-C128T mutant and part of the analysis of ChR2 WT.

Curriculum Vitae

Name: Yanan Guo

Nationality: People's Republic of China

10.2013 - 02.2017	PhD Student of Prof. Dr. Marcus Elstner in Institute of Physical Chemistry, Karlsruhe Institute of Technology, Karlsruhe, Germany
09.2010 - 05.2013	Master Student of Prof. Dr. Xiong Lu in School of Materials Science and Engineering, Southwest Jiaotong University, Chengdu, China
09.2006 - 06.2010	Bachelor Student in School of Materials Science and Engineering, Southwest Jiaotong University, Chengdu, China
09.2003 - 06.2006	Highschool, Heze, China

Bibliography

[1] Hegemann, P.; Sigrist, S. *Optogenetics*; Freie Universität Berlin: Berlin, Germany, 2013.

[2] Palmada, M.; Centelles, J. J. *Front. Biosci.* **1998**, *3*, d701–718.

[3] Crowder, J. M.; Croucher, M. J.; Bradford, H. F.; Collins, J. F. *J. Neurochem.* **1987**, *48*, 1917–1924.

[4] Madden, D. R. *Nat. Rev. Neurosci.* **2002**, *3*, 91–101.

[5] Johnson, J.; Ascher, P. *Biophys J.* **1990**, *57*, 1085–1090.

[6] Castillo, P. E.; Malenka, R. C.; Nicoll, R. A. *Nature* **1997**, *388*, 182–186.

[7] Lerma, J.; Paternain, A. V.; Rodríguez-Moreno, A.; López-García, J. C. *Physiol. Rev.* **2001**, *81*, 971–998.

[8] Traynelis, S. F.; Wollmuth, L. P.; McBain, C. J.; Menniti, F. S.; Vance, K. M.; Ogden, K. K.; Hansen, K. B.; Yuan, H.; Myers, S. J.; Dingledine, R. *Pharmacol. Rev.* **2010**, *62*, 405–496.

[9] Hollmann, M.; Maron, C.; Heinemann, S. *Neuron* **1994**, *13*, 1331 – 1343.

[10] Hansen, K. B.; Furukawa, H.; Traynelis, S. F. *Mol. Pharmacol.* **2010**, *78*, 535–549.

[11] Möykkynen, T.; Coleman, S. K.; Semenov, A.; Keinänen, K. *J. Biol. Chem.* **2014**, *289*, 13197–13205.

[12] Kuusinen, A.; Abele, R.; Madden, D. R.; Keinänen, K. *J. Biol. Chem.* **1999**, *274*, 28937–28943.

[13] Pasternack, A.; Coleman, S. K.; Jouppila, A.; Mottershead, D. G.; Lindfors, M.; Pasternack, M.; Keinänen, K. *J. Biol. Chem.* **2002**, *277*, 49662–49667.

[14] Kuusinen, A.; Arvola, M.; Keinänen, K. *EMBO J.* **1995**, *14*, 6327–6332.

[15] Keinänen, K.; Jouppila, A.; Kuusinen, A. *Biochem. J.* **1998**, *330*, 1461–1467.

[16] Ivanovic, A.; Reiländer, H.; Laube, B.; Kuhse, J. *J. Biol. Chem.* **1998**, *273*, 19933–19937.

[17] Armstrong, N.; Gouaux, E. *Neuron* **2000**, *28*, 165 – 181.

[18] Deming, D.; Cheng, Q.; Jayaraman, V. *J. Biol. Chem.* **2003**, *278*, 17589–17592.

[19] Mayer, M. L. *Neuron* **2005**, *45*, 539–552.

[20] Reiter, A.; Skerra, A.; Trauner, D.; Schiefner, A. *Biochemistry* **2013**, *52*, 8972–8974.

[21] Abele, R.; Keinänen, K.; Madden, D. R. *J. Biol. Chem.* **2000**, *275*, 21355–21363.

[22] Chen, H.-S. V.; Lipton, S. A. *J. Pharmacol. Exp. Ther.* **2005**, *314*, 961–971.

[23] Jin, R.; Gouaux, E. *Biochemistry* **2003**, *42*, 5201–5213.

[24] Jin, R.; Banke, T. G.; Mayer, M. L.; Traynelis, S. F.; Gouaux, E. *Nat. Neurosci.* **2003**, *6*, 803–810.

[25] Armstrong, N.; Mayer, M.; Gouaux, E. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5736–5741.

[26] Priestley, T.; Laughton, P.; Myers, J.; Le Bourdellés, B.; Kerby, J.; Whiting, P. J. *Mol. Pharmacol.* **1995**, *48*, 841–848.

[27] Inanobe, A.; Furukawa, H.; Gouaux, E. *Neuron* **2005**, *47*, 71–84.

[28] Kaufman, H.; Vratsanos, S. M.; Erlanger, B. F. *Science* **1968**, *162*, 1487–1489.

[29] Bartels, E.; Wassermann, N. H.; Erlanger, B. F. *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 1820–1823.

[30] Volgraf, M.; Gorostiza, P.; Szobota, S.; Helix, M. R.; Isacoff, E. Y.; Trauner, D. *J. Am. Chem. Soc.* **2006**, *129*, 260–261.

[31] Fujita, D.; Murai, M.; Nishioka, T.; Miyoshi, H. *Biochemistry* **2006**, *45*, 6581–6586.

[32] Nargeot, J.; Lester, H. A.; Birdsall, N. J.; Stockton, J.; Wassermann, N. H.; Erlanger, B. F. *J. Gen. Physiol.* **1982**, *79*, 657–678.

[33] Guo, Y.; Wolter, T.; Kubař, T.; Sumser, M.; Trauner, D.; Marcus, E. *PLoS ONE* **2015**, *10*, e0135399.

[34] Spudich, J. L.; Yang, C.-S.; Jung, K.-H.; Spudich, E. N. *Annu. Rev. Cell Dev. Biol.* **2000**, *16*, 365–392.

[35] Khorana, H. G. *Proc. Natl. Acad. Sci.* **1993**, *90*, 1166–1171.

[36] Helmreich, E. J.; Hofmann, K.-P. *Biochim. Biophys. Acta, Rev. Biomembranes* **1996**, *1286*, 285 – 322.

[37] Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Trong, I. L.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000**, *289*, 739–745.

[38] Luecke, H.; Schobert, B.; Richter, H.-T.; Cartailler, J.-P.; Lanyi, J. K. *J. Mol. Biol.* **1999**, *291*, 899 – 911.

[39] Bieszke, J. A.; Braun, E. L.; Bean, L. E.; Kang, S.; Natvig, D. O.; Borkovich, K. A. *Proc. Natl. Acad. Sci.* **1999**, *96*, 8034–8039.

[40] Sasaki, J.; Brown, L.; Chon, Y.; Kandori, H.; Maeda, A.; Needleman, R.; Lanyi, J. *Science* **1995**, *269*, 73–75.

[41] Baldwin, J. M.; Schertler, G. F.; Unger, V. M. *J. Mol. Biol.* **1997**, *272*, 144 – 164.

[42] Shieh, T.; Han, M.; Sakmar, T. P.; Smith, S. O. *J. Mol. Biol.* **1997**, *269*, 373 – 384.

[43] Royant, A.; Nollert, P.; Edman, K.; Neutze, R.; Landau, E. M.; Pebay-Peyroula, E.; Navarro, J. *Proc. Natl. Acad. Sci.* **2001**, *98*, 10131–10136.

[44] Kolbe, M.; Besir, H.; Essen, L.-O.; Oesterhelt, D. *Science* **2000**, *288*, 1390–1396.

[45] Schertler, G. F.; Hargrave, P. A. *Proc. Natl. Acad. Sci.* **1995**, *92*, 11578–11582.

[46] Davies, A.; Schertler, G. F.; Gowen, B. E.; Saibil, H. R. *J. Struct. Biol.* **1996**, *117*, 36–44.

[47] Krebs, A.; Villa, C.; Edwards, P. C.; Schertler, G. F. *J. Mol. Biol.* **1998**, *282*, 991 – 1003.

[48] McBee, J. K.; Palczewski, K.; Baehr, W.; Pepperberg, D. R. *Prog. Retin. Eye Res.* **2001**, *20*, 469 – 529.

[49] Garczarek, F.; Gerwert, K. *Nature* **2006**, *439*, 109–112.

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

[51] Bondar, A.-N.; Fischer, S.; Suhai, S.; Smith, J. C. *J. Phys. Chem. B* **2005**, *109*, 14786–14788.

[52] Smith, S. O.; Pardoen, J. A.; Mulder, P. P. J.; Curry, B.; Lugtenburg, J.; Mathies, R. *Biochemistry* **1983**, *22*, 6141–6148.

[53] Phatak, P. V. Investigation of proton transfer pathways in bacteriorhodopsin with multi-length-scale simulations. Ph.D. thesis, TU Braunschweig, 2009.

[54] Nagel, G.; Ollig, D.; Fuhrmann, M.; Kateriya, S.; Musti, A. M.; Bamberg, E.; Hegemann, P. *Science* **2002**, *296*, 2395–2398.

[55] Nagel, G.; Szellas, T.; Huhn, W.; Kateriya, S.; Adeishvili, N.; Berthold, P.; Ollig, D.; Hegemann, P.; Bamberg, E. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13940–13945.

[56] Zhang, F.; Prigge, M.; Beyrière, F.; Tsunoda, S. P.; Mattis, J.; Yizhar, O.; Hegemann, P.; Deisseroth, K. *Nat. Neurosci.* **2008**, *11*, 631–633.

[57] Kianianmomeni, A.; Stehfest, K.; Nematollahi, G.; Hegemann, P.; Hallmann, A. *J. Plant. Physiol.* **2009**, *151*, 347–366.

[58] Govorunova, E. G.; Spudich, E. N.; Lane, C. E.; Sineshchekov, O. A.; Spudich, J. L. *mBio* **2011**, *2*, 00115–11.

[59] Govorunova, E. G.; Sineshchekov, O. A.; Li, H.; Janz, R.; Spudich, J. L. *J. Biol. Chem.* **2013**, *288*, 29911–29922.

[60] Ogren, J. I.; Mamaev, S.; Russano, D.; Li, H.; Spudich, J. L.; Rothschild, K. J. *Biochemistry* **2014**, *53*, 3961–3970.

[61] Sineshchekov, O. A.; Jung, K.-H.; Spudich, J. L. *Proc. Natl. Acad. Sci.* **2002**, *99*, 8689–8694.

[62] Ritter, E.; Stehfest, K.; Berndt, A.; Hegemann, P.; Bartl, F. J. *J. Biol. Chem.* **2008**, *283*, 35033–35041.

[63] Boyden, E. S.; Zhang, F.; Bamberg, E.; Nagel, G.; Deisseroth, K. *Neuroscience* **2005**, *8*, 1263–1268.

[64] Li, X.; Gutierrez, D. V.; Hanson, M. G.; Han, J.; Mark, M. D.; Chiel, H.; Hegemann, P.; Landmesser, L. T.; Herlitze, S. *Proc. Natl. Acad. Sci.* **2005**, *102*, 17816–17821.

[65] Nagel, G.; Brauner, M.; Liewald, J. F.; Adeishvili, N.; Bamberg, E.; Gottschalk, A. *Curr. Biol.* **2005**, *15*, 2279 – 2284.

[66] Kato, H. E.; Zhang, F.; Yizhar, O.; Ramakrishnan, C.; Nishizawa, T.; Hirata, K.; Ito, J.; Aita, Y.; Tsukazaki, T.; Hayashi, S.; Hegemann, P.; Maturana, A.; Ishitani, R.; Deisseroth, K.; Nureki, O. *Nature* **2012**, *482*, 369–374.

[67] Müller, M.; Bamann, C.; Bamberg, E.; Kühlbrandt, W. *J. Mol. Biol.* **2011**, *414*, 86–95.

[68] Nack, M.; Radu, I.; Bamann, C.; Bamberg, E.; Heberle, J. *FEBS Lett.* **2009**, *583*, 3676–3680.

[69] Bruun, S. B.; Stoeppler, D.; Keidel, A.; Kuhlmann, U.; Luck, M.; Diehl, A.; Geiger, M.-A.; Woodmansee, D.; Trauner, D.; Hegemann, P.; Oschkinat, H.; Hildebrandt, P.; Stehfest, K. *Biochemistry* **2015**, *54*, 5389–5400.

[70] Duñach, M.; Marti, T.; Khorana, H. G.; Rothschild, K. J. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9873–9877.

[71] Schertler, G. F.; Lozier, R.; Oesterhelt, D. *EMBO J.* **1991**, *10*, 2353–2361.

[72] Lórenz-Fonfría, V. A.; Heberle, J. *Biochim. Biophys. Acta, Bioenerg.* **2014**, *1837*, 626 – 642.

[73] Baer, M.; Mathias, G.; Kuo, I.-F. W.; Tobias, D. J.; Mundy, C. J.; Marx, D. *ChemPhysChem* **2008**, *9*, 2703–2707.

[74] Lórenz-Fonfría, V. A.; Schultz, B.-J.; Resler, T.; Schlesinger, R.; Bamann, C.; Bamberg, E.; Heberle, J. *J. Am. Chem. Soc.* **2015**, *137*, 1850–1861.

[75] Becker-Baldus, J.; Bamann, C.; Saxena, K.; Gustmann, H.; Brown, L. J.; Brown, R. C. D.; Reiter, C.; Bamberg, E.; Wachtveitl, J.; Schwalbe, H.; Glaubitz, C. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9896–9901.

[76] Gunaydin, L. A.; Yizhar, O.; Berndt, A.; Sohal, V. S.; Deisseroth, K.; Hegemann, P. *Nat. Neurosci.* **2010**, *13*, 387–392.

[77] Verhoefen, M.-K.; Bamann, C.; Blöcher, R.; Förster, U.; Bamberg, E.; Wachtveitl, J. *ChemPhysChem* **2010**, *11*, 3113–3122.

[78] Kuhne, J.; Eisenhauer, K.; Ritter, E.; Hegemann, P.; Gerwert, K.; Bartl, F. *Angew. Chem. Int. Ed.* **2015**, *54*, 4953–4957.

[79] Watanabe, H. C.; Welke, K.; Schneider, F.; Tsunoda, S.; Zhang, F.; Deisseroth, K.; Hegemann, P.; Elstner, M. *J. Biol. Chem.* **2012**, *287*, 7456–7466.

[80] Watanabe, H. C.; Welke, K.; Sindhikara, D. J.; Hegemann, P.; Elstner, M. *J. Mol. Biol.* **2013**, *425*, 1795–1814.

[81] Bamann, C.; Kirsch, T.; Nagel, G.; Bamberg, E. *J. Mol. Biol.* **2008**, *375*, 686–694.

[82] Berndt, A.; Yizhar, O.; Gunaydin, L. A.; Hegemann, P.; Deisseroth, K. *Nat. Neurosci.* **2009**, *12*, 229–234.

[83] Lórenz-Fonfría, V. A.; Resler, T.; Krause, N.; Nack, M.; Gossing, M.; Fischer von Mollard, G.; Bamann, C.; Bamberg, E.; Schlesinger, R.; Heberle, J. *Proc. Natl. Acad. Sci.* **2013**, *110*, E1273–E1281.

[84] Sindhikara, D. J.; Yoshida, N.; Hirata, F. *J. Comput. Chem.* **2012**, *33*, 1536–1543.

[85] Sindhikara, D. J.; Hirata, F. *J. Phys. Chem. B* **2013**, *117*, 6718–6723.

[86] Berndt, A.; Schoenenberger, P.; Mattis, J.; Tye, K. M.; Deisseroth, K.; Hegemann, P.; Oertner, T. G. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7595–7600.

[87] Yizhar, O. et al. *Nature* **2011**, *477*, 171–178.

[88] Radu, I.; Bamann, C.; Nack, M.; Nagel, G.; Bamberg, E.; Heberle, J. *J. Am. Chem. Soc.* **2009**, *131*, 7313–7319.

[89] Nack, M.; Radu, I.; Gossing, M.; Bamann, C.; Bamberg, E.; von Mollard, G. F.; Heberle, J. *Photochem. Photobiol. Sci.* **2010**, *9*, 194–198.

[90] Perálvarez-Marín, A.; Márquez, M.; Bourdelande, J.-L.; Querol, E.; Padrós, E. *J. Biol. Chem.* **2004**, *279*, 16403–16409.

[91] Joh, N. H.; Min, A.; Faham, S.; Whitelegge, J. P.; Yang, D.; Woods, V. L.; Bowie, J. U. *Nature* **2008**, *453*, 1266–1270.

[92] Welke, K.; Watanabe, H. C.; Wolter, T.; Gaus, M.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6651–6659.

[93] Prigge, M.; Schneider, F.; Tsunoda, S. P.; Shilyansky, C.; Wietek, J.; Deisseroth, K.; Hegemann, P. *J. Biol. Chem.* **2012**, *287*, 31804–31812.

[94] Ullrich, S.; Gueta, R.; Nagel, G. *J. Biol. Chem.* **2013**, *394*, 271–280.

[95] Sugiyama, Y.; Wang, H.; Hikima, T.; Sato, M.; Kuroda, J.; Takahashi, T.; Ishizuka, T.; Yawo, H. *Photochem. Photobiol. Sci.* **2009**, *8*, 328–336.

[96] Plazzo, A. P.; De Franceschi, N.; Da Broi, F.; Zonta, F.; Sanasi, M. F.; Filippini, F.; Mongillo, M. *J. Biol. Chem.* **2012**, *287*, 4818–4825.

[97] Grote, M.; Engelhard, M.; Hegemann, P. *Biochim. Biophys. Acta - Bioenerg.* **2014**, *1837*, 533 – 545.

[98] Schneider, F.; Grimm, C.; Hegemann, P. *Annu. Rev. Biophys.* **2015**, *44*, 167–186.

[99] Nikolic, K.; Grossman, N.; Grubb, M. S.; Burrone, J.; Toumazou, C.; Degenaar, P. *Photochem. Photobiol.* **2009**, *85*, 400–411.

[100] Williams, J. C.; Xu, J.; Lu, Z.; Klimas, A.; Chen, X.; Ambrosi, C. M.; Cohen, I. S.; Entcheva, E. *PLoS Comput. Biol.* **2013**, *9*, e1003220.

[101] Hegemann, P.; Ehlenbeck, S.; Gradmann, D. *Biophys. J.* **2005**, *89*, 3911–3918.

[102] Stefanescu, R. A.; Shivakeshavan, R.; Khargonekar, P. P.; Talathi, S. S. *Bull. Math. Biol.* **2013**, *75*, 2208–2240.

[103] Ernst, O. P.; Murcia, P. A. S.; Daldrop, P.; Tsunoda, S. P.; Kateriya, S.; Hegemann, P. *J. Biol. Chem.* **2008**, *283*, 1637–1643.

[104] Nack, M.; Radu, I.; Schultz, B.-J.; Resler, T.; Schlesinger, R.; Bondar, A.-N.; del Val, C.; Abbruzzetti, S.; Viappiani, C.; Bamann, C.; Bamberg, E.; Heberle, J. *FEBS Lett.* **2012**, *586*, 1344–1348.

[105] Bamann, C.; Gueta, R.; Kleinlogel, S.; Nagel, G.; Ernst, B. *Biochemistry* **2010**, *49*, 267–278.

[106] Stehfest, K.; Ritter, E.; Berndt, A.; Bartl, F.; Hegemann, P. *J. Mol. Biol.* **2010**, *398*, 690–702.

[107] Lórenz-Fonfría, V. A.; Schultz, B.-J.; Resler, T.; Schlesinger, R.; Bamann, C.; Bamberg, E.; Heberle, J. *J. Am. Chem. Soc.* **2015**, *137*, 1850–1861.

[108] Mizuide, N.; Shibata, M.; Friedman, N.; Sheves, M.; Belenky, M.; Herzfeld, J.; Kandori, H. *Biochemistry* **2006**, *45*, 10674–10681.

[109] Smith, S. O.; Pardoen, J. A.; Lugtenburg, J.; Mathies, R. A. *J. Phys. Chem.* **1987**, *91*, 804–819.

[110] Maeda, A. *Isr. J. Chem.* **1995**, *35*, 387–400.

[111] Takemoto, M.; Kato, H. E.; Koyama, M.; Ito, J.; Kamiy, M.; Hayashi, S.; Maturana, A. D.; Deisseroth, K.; Ishitani, R.; Nureki, O. *PLoS One* **2015**, *10*, e0131094.

[112] Eisenhauer, K.; Kuhne, J.; Ritter, E.; Berndt, A.; Wolf, S.; Freier, E.; Hegemann, P.; Gerwert, K. *J. Biol. Chem.* **2012**, *287*, 6904–6911.

[113] Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*, 2nd ed.; Elsevier: United States of America, 2002.

[114] Van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Simul.* **1988**, *1*, 173–185.

[115] Zhigilei, L. V.; Garrison, B. J. *Appl. Phys. Lett.* **1997**, *71*, 551–553.

[116] Lemak, A. S.; Balabaev, N. K. *Mol. Simul.* **1994**, *13*, 177–187.

[117] Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

[118] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

[119] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

[120] Nosé, S.; Klein, M. L. *Mol. Phys.* **1983**, *50*, 1055–1076.

[121] Parr, R. G.; Yang, W. *Density functional theory of atoms and molecules*, 2nd ed.; Oxford University Press, Inc.: New York, United Sates of America, 1989.

[122] Koch, W.; Holthausen, M. C. *A chemist's guide to density functional theory*, 2nd ed.; John Wiley and Sons, Ltd.: Weinheim, Germany, 2001.

[123] Jensen, F. *Introduction to computational chemistry*, 2nd ed.; John Wiley and Sons, Ltd.: University of Southern Denmark, Odense, Denmark, 2007.

[124] Malmqvist, P.-A. k.; Roos, B. O. *Chem. Phys. Lett.* **1989**, *155*, 189–194.

[125] Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

[126] Neese, F. *J. Chem. Phys.* **2003**, *119*, 9428–9443.

[127] Demichelis, R.; Civalleri, B.; Ferrabone, M.; Dovesi, R. *Int. J. Quantum Chem.* **2010**, *110*, 406–415.

[128] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792–794.

[129] Elstner, M.; S., S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[130] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

[131] Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

[132] Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2014**, *10*, 1518–1537.

[133] Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, *117*, 185–199.

[134] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[135] Brunk, E.; Rothlisberger, U. *Chem. Rev.* **2015**, *115*, 6217–6263.

[136] Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.

[137] Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187 – 199.

[138] Kästner, J. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932–942.

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

[140] Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843.

[141] Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.

[142] Sprik, M.; Ciccotti, G. *J. Chem. Phys.* **1998**, *109*, 7737–7744.

[143] Shankar, K.; Djamal, B.; Robert H., S.; Peter A., K.; John M., R. *J. Comupt. Chem.* **1992**, *13*, 1011–1021.

[144] Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40 – 57.

[145] Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

[146] Wardle, B. *Principles and applications of photochemistry*, 1st ed.; John Wiley and Sons, Ltd.: Manchester Metropolitan University, Manchester, UK, 2009.

[147] Dirac, P. A. M. *Proc. R. Soc. London, Ser. A* **1927**, *114*, 243–265.

[148] Worth, G. A.; Cederbaum, L. S. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158.

[149] Polli, D.; Altoè, P.; Weingart, O.; Spillane, K. M.; Manzoni, C.; Brida, D.; Tomasello, G.; Orlandi, G.; Kukura, P.; Mathies, R. A.; Garavelli, M.; Cerullo, G. *Nature* **2010**, *467*, 440–443.

[150] Groenhof, G. Understanding light-induced conformational changes in molecular systems from first principles. Ph.D. thesis, University of Groningen, Groningen, 2005.

[151] Baer, M. *Phys. Rep.* **2002**, *358*, 75 – 142.

[152] Groenhof, G.; Bouxn-Cademartory, M.; Hess, B.; de Visser, S. P.; Berendsen, H. J. C.; Olivucci, M.; Mark, A. E.; Robb, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 4228–4233.

[153] Schäfer, L. Photoactivated processes in condensed phase studied by molecular dynamics simulations. Ph.D. thesis, Technischen Universität Carolo-Wilhelmina zu Braunschweig, 2007.

[154] Nguyen, P. H.; Stock, G. *Chem. Phys.* **2006**, *323*, 36–44.

[155] Nguyen, P. H.; Gorbunov, R. D.; Stock, G. *Biophys. J.* **2006**, *91*, 1224–1234.

[156] Ishikawa, T.; Noro, T.; Shoda, T. *J. Chem. Phys.* **2001**, *115*, 7503–7512.

[157] Gagliardi, L.; Orlandi, G.; Bernardi, F.; Cembran, A.; Garavelli, M. *Theor. Chem. Acc.* **2004**, *111*, 363–372.

[158] Cembran, A.; Fernando, B.; Marco, G.; Gagliardi, L.; Giorgio, O. *J. Am. Chem. Soc.* **2004**, *126*, 3234–3243.

[159] Volgraf, M.; Gorostiza, P.; Numano, R.; Kramer, R. H.; Isacoff, E. Y.; Trauner, D. *Nat. Chem. Biol.* **2006**, *2*, 47–52.

[160] Stawski, P.; Sumser, M.; Trauner, D. *Angew. Chem. Int. Ed.* **2012**, *51*, 5748–5751.

[161] Wolter, T.; Steinbrecher, T.; Trauner, D.; Elstner, M. *PLoS ONE* **2014**, *9*, e92716.

[162] Wolter, T.; Steinbrecher, T.; Elstner, M. *PLoS ONE* **2013**, *8*, e58774.

[163] Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.

[164] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

[165] Åqvist, J. *J. Phys. Chem.* **1990**, *94*, 8021–8024.

[166] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 712–725.

[167] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[168] Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.

[169] Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

[170] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[171] Cordes, F. S.; Kukol, A.; R., F. L.; Arkin, I. T.; Sansom, M. S. P.; Fischer, W. B. *Biochim. Biophys. Acta, Biomembranes* **2001**, *1512*, 291–298.

[172] Mayewski, S. *Proteins Struct. Funct. Bioinf.* **2005**, *59*, 152–169.

[173] Stierand, K.; Maaß, P.; Rarey, M. *Bioinformatics* **2006**, *22*, 1710–1716.

[174] Stierand, K.; Rarey, M. *ChemMedChem* **2007**, *2*, 853–860.

[175] Stierand, K.; Rarey, M. *Med. Chem. Lett.* **2010**, *1*, 540–545.

[176] Swanson, G. T.; Gereau IV, R. W.; Green, T.; Heinemann, S. F. *Neuron* **1997**, *19*, 913–926.

[177] Lämpinen, M.; Pentikainen, O.; Johnson, M. S.; K., K. *EMBO* **1988**, *17*, 4704–4711.

[178] Mankiewicz, K.; Rambhadran, A.; Wathen, L.; Jayaraman, V. *Biochemistry* **2008**, *47*, 398–404.

[179] Bjerrum, E. J.; Biggin, P. C. *Proteins Struct. Funct. Bioinf.* **2008**, *72*, 434–446.

[180] Nayeem, N.; Mayans, O.; Green, T. *J. Neurosci.* **2011**, *31*, 2916–2924.

[181] Schoenenberger, P.; Gerosa, D.; Oertner, T. O. *PLOSONE* **2009**, *4*, e8185.

[182] Feig, M.; Karanicolas, J.; Brooks, C. r. *J. Mol. Graph. Model.* **2004**, *22*, 377–395.

[183] Zhang, L.; Hermans, J. *Proteins Struct. Funct. Bioinf.* **1996**, *24*, 433–438.

[184] Parrinello, M.; Rahmann, A. *J. Chem. Phys.* **1982**, *76*, 2662–2666.

[185] Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2002**, *83*, 1281–1297.

[186] Kandt, C.; Schlitter, J.; Gerwert, K. *Biophys. J.* **2004**, *86*, 705–717.

[187] Babitzki, G.; Denschlag, R.; Tavan, P. *J. Phys. Chem. B* **2009**, *113*, 10483–10495.

[188] Huang, J.; Mackerell, A. D. J. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

[189] Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D. J.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

[190] Kumar, S.; Nussinov, R. *ChemBioChem* **2002**, *3*, 604–617.

[191] Ballesteros, J. A.; Deupi, X.; Olivella, M.; Haaksma, E. E.; Pardo, L. *Biophys. J.* **2000**, *79*, 2754 – 2760.

[192] Steiner, T.; Desiraju, G. R. *Chem. Commun.* **1998**, 891–892.

[193] Guo, Y.; Beyle, F. E.; Bold, B. M.; Watanabe, H. C.; Koslowski, A.; Thiel, W.; Hegemann, P.; Marazzi, M.; Elstner, M. *Chem. Sci.* **2016**, *7*, 3879–3891.

[194] Gerwert, K.; Freier, E.; Wolf, S. *Biochim. Biophys. Acta, Bioenerg.* **2014**, *1837*, 606–613.

[195] Nishikawa, T.; Murakami, M.; Kouyama, T. *J. Mol. Biol.* **2005**, *352*, 319–328.

[196] Welke, K.; Frähmcke, J. S.; Watanabe, H. C.; Hegemann, P.; Elstner, M. *J. Phys. Chem. B* **2011**, *115*, 15119–15128.

[197] Kubař, T.; Welke, K.; Groenhof, G. *J. Comput. Chem.* **2015**, *36*, 1978–1989.

[198] Wolter, T.; Elstner, M.; Fischer, S.; Smith, J. C.; Bondar, A.-N. *J. Phys. Chem. B* **2015**, *119*, 2229–2240.

[199] Brooks, B. R. et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[200] Neese, F. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

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

[202] Sperber, G. *Int. J. Quantum Chem.* **1971**, *5*, 189–214.

[203] Paldus, J.; Chin, E. *Int. J. Quantum Chem.* **1983**, *24*, 373–394.

[204] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[205] Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

[206] Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

[207] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

[208] Koslowski, A.; Beck, M. E.; Thiel, W. *J. Comput. Chem.* **2003**, *24*, 714–726.

[209] Silva-Junior, M. R.; Thiel, W. *J. Chem. Theory Comput.* **2010**, *6*, 1546–1564.

[210] Frähmcke, J. S.; Wanko, M.; Phatak, P.; Mroginski, M. A.; Elstner, M. *J. Phys. Chem. B* **2010**, *114*, 11338–11352.

[211] Lansing, J. C.; Hohwy, M.; Jaroniec, C. P.; Creemers, A. F. L.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **2002**, *41*, 431–438.

[212] Mao, J.; Do, N.-N.; Scholz, F.; Reggie, L.; Mehler, M.; Lakatos, A.; Ong, Y.-S.; Ullrich, S. J.; Brown, L. J.; Brown, R. C. D.; Becker-Baldus, J.; Wachtveitl, J.; Glaubitz, C. *J. Am. Chem. Soc.* **2014**, *136*, 17578–17590.

[213] Warshel, A.; Kato, M.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

[214] Warshel, A.; Chu, Z. T. *J. Phys. Chem. B* **2001**, *105*, 9857–9871.

[215] Houjou, H.; Inoue, Y.; Sakurai, M. *J. Phys. Chem. B* **2001**, *105*, 867–879.

[216] Mathies, R.; Brito Cruz, C.; Pollard, W.; Shank, C. *Science* **1988**, *240*, 777–779.

[217] Kobayashi, T.; Saito, T.; Ohtani, H. *Nature* **2001**, *414*, 531–534.

[218] Herbst, J.; Heyne, K.; Diller, R. *Science* **2002**, *297*, 822–825.

[219] Schoenlein, R.; Peteanu, L.; Mathies, R.; Shank, C. *Science* **1991**, *254*, 412–415.

[220] Wang, Q.; Schoenlein, R.; Peteanu, L.; Mathies, R.; Shank, C. *Science* **1994**, *266*, 422–424.

[221] Schapiro, I.; Ruhman, S. *Biochim. Biophys. Acta, Bioenergetics* **2014**, *1837*, 589–597.

[222] Schnedermann, C.; Muders, V.; Ehrenberg, D.; Schlesinger, R.; Kukura, P.; Heberle, J. *J. Am. Chem. Soc.* **2016**, *138*, 4757–4762.

[223] Warshel, A. *Nature* **1976**, *260*, 679–683.

[224] Arieh, W.; Natalia, B. *J. Am. Chem. Soc.* **1982**, *104*, 1469–1476.

[225] Warshel, A.; Chu, Z.; Hwang, J.-K. *Chem. Phys.* **1991**, *158*, 303–314.

[226] Frutos, L. M.; Andruniów, T.; Santoro, F.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci.* **2007**, *104*, 7764–7769.

[227] Liu, R. S.; Asato, A. E. *Proc. Natl. Acad. Sci.* **1985**, *82*, 259–263.

[228] Liu, R. S. H. *Acc. Chem. Res.* **2001**, *34*, 555–562.

[229] Strambi, A.; Durbeej, B.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21322–21326.

[230] Altoè, P.; Cembran, A.; Olivucci, M.; Garavelli, M. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20172–20177.

[231] Cembran, A.; Bernardi, F.; Olivucci, M.; Garavelli, M. *J. Am. Chem. Soc.* **2004**, *126*, 16018–16037.

[232] Virshup, A. M.; Punwong, C.; Pogorelov, T. V.; Lindquist, B. A.; Ko, C.; Martínez, T. J. *J. Phys. Chem. B* **2009**, *113*, 3280–3291.

[233] Ferré, N.; Olivucci, M. *J. Am. Chem. Soc.* **2003**, *125*, 6868–6869.

[234] Andruniów, T.; Ferré, N.; Olivucci, M. *Proc. Natl. Acad. Sci.* **2004**, *101*, 17908–17913.

[235] Aquilante, F. et al. *J. Comput. Chem.* **2016**, *37*, 506–541.

[236] Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016–1024.

[237] Kaminski, S.; Gaus, M.; Phatak, P.; von Stetten, D.; Elstner, M.; Mroginski, M. A. *J. Chem. Theory Comput.* **2010**, *6*, 1240–1255.

[238] Walczak, E.; Szefczyk, B.; Andruniów, T. *J. Chem. Theory Comput.* **2013**, *9*, 4915–4927.

[239] Ghigo, G.; Roos, B. O.; Malmqvist, P.-A. *Chem. Phys. Lett.* **2004**, *396*, 142–149.

[240] Gozem, S.; Melaccio, F.; Lindh, R.; Krylov, A. I.; Granovsky, A. A.; Angeli, C.; Olivucci, M. *J. Chem. Theory Comput.* **2013**, *9*, 4495–4506.

[241] Roos, B. O.; Andersson, K. *Chem. Phys. Lett.* **1995**, *245*, 215–223.

[242] Forsberg, N.; Malmqvist, P.-A. *Chem. Phys. Letters* **1997**, *274*, 196–204.

[243] Walczak, E.; Andruniów, T. *Phys. Chem. Chem. Phys.* **2015**, *17*, 17169–17181.

[244] Laricheva, E. N.; Gozem, S.; Rinaldl, S.; Melaccio, F.; Valentini, A.; Olivucci, M. *J. Chem. Theory Comput.* **2012**, *8*, 2559–2563.

[245] Bonačić-Koutecký, V.; Schöffel, K.; Michl, J. *Theor. Chim. Acta* **1987**, *72*, 459–474.

[246] Garavelli, M.; Bernardi, F.; Olivucci, M.; Vreven, T.; Klein, S.; Celani, P.; A. Robb, M. *Faraday Discuss.* **1998**, *110*, 51–70.

[247] Tomasello, G.; Olaso-González, G.; Altoè, P.; Stenta, M.; Serrano-Andrès, L.; Merchán, M.; Orlandi, G.; Bottoni, A.; Garavelli, M. *J. Am. Chem. Soc.* **2009**, *131*, 5172–5186.

[248] Hayashi, S.; Tajkhorshid, E.; Schulten, K. *Biophys. J.* **2003**, *85*, 1440–1449.