学位論文

Molecular dynamics simulations of eukaryotic membrane channel and transporter

(真核生物由来膜チャネル・膜輸送体の
分子動力学シミュレーションによる解析)

平成29年12月 博士(理学) 申請

東京大学大学院理学系研究科
生物科学専攻
武本 瑞貴

Abstract

Cells and organelles are separated from the external environment by the biological membrane, and importing or exporting the materials across the membrane, called membrane transport, is essential for the survival of the cell. This membrane transport is carried out by some groups of membrane proteins, called membrane channels and transporters. They are divided into some subgroups according to the direction of the transport and the energy source. By focusing on the requirements that membrane proteins of each group must satisfy at the minimum, mechanisms in common within these groups can be considered. To understand these mechanisms of channels and transporters, not only their static crystal structures, but their dynamics are also important. All-atom molecular dynamics (MD) simulation is a suitable method to explore these problems. However, at the same time, the conventional all-atom MD simulation has a problem that it cannot observe the event in a long, physiologically relevant timescale due to the fundamental limits of the computation. This problem is solved by employing the enhanced sampling techniques. In chapter 1, the general mechanisms for channels and transporters and the conformational sampling methods in MD simulation are summarized. In this study, the author focused on two proteins, that is a light-gated channel, channelrhodopsin (ChR), and an antiporter, triose-phosphate/phosphate translocator (TPT), to unravel their mechanisms in atomic detail, using conventional MD simulation and enhanced sampling MD simulations.

In chapter 2, the MD simulations of channelrhodopsin (ChR) were performed. ChR is a light-gated cation channel that responds to blue light. Since ChR can be readily expressed in specific neurons to precisely control their activities by light, it has become a powerful tool in neuroscience. Although the recently solved crystal structure of a chimeric ChR, C1C2, provided the structural basis for ChR, our understanding of the molecular mechanism of ChR still remains limited. Here electrophysiological analyses and conventional all-atom MD simulations are performed, to investigate the importance of the intracellular and central constrictions of the ion conducting pore observed in the crystal structure of C1C2. The electrophysiological analysis revealed that two glutamate residues, Glu122 and

Glu129, in the intracellular and central constrictions, respectively, should be deprotonated in the photocycle. The simulation results suggested that the deprotonation of Glu129 in the central constriction leads to ion leakage in the ground state, and implied that the protonation of Glu129 is important for preventing ion leakage in the ground state. Moreover, the 13-*cis* retinal bound; *i.e.*, activated C1C2 model was designed, and MD simulations were performed to investigate the conformational changes in the early stage of the photocycle. These simulations suggested that retinal photoisomerization induces the conformational change toward channel opening, including the movements of TM6, TM7, and TM2. These insights into the dynamics of the ground states and the early photocycle stages enhance our understanding of the channel function of ChR.

In chapter 3, the author performed the MD simulation of triose-phosphate/phosphate translocator (TPT). Secondary active membrane transporters translocate their substrates using the electrochemical potentials of other chemical compounds while undergoing large-scale conformational changes. Despite extensive structural studies of transporters, the atomic details of the transport mechanism, such as the coupling mechanism of the local and global conformational changes, still remain elusive. Here, a series of all-atom molecular dynamics simulations of the TPT was performed, which exports organic phosphates, such as triose-phosphate, in the chloroplast stroma in a strict counter exchange with inorganic phosphate (P_i). Biased sampling methods, including string method and umbrella sampling, successfully reproduced the conformational changes between the inward- and outward-facing states, along with the substrate binding. The free energy landscape of this entire TPT transition pathway demonstrated the alternating access and substrate translocation mechanisms. The results revealed that P_i is relayed by positively charged residues along the transition pathway. Furthermore, the conserved Glu207 functions as a “molecular switch”, linking the local substrate binding and the global conformational transition. The results provide quantitative and atomic-detailed insights into the energy coupling mechanism of secondary active antiporters.

Contents

List of Figures

List of Tables

Abbreviations

NPT	isothermal-isobaric ensemble
NVT	isothermal (canonical) ensemble
13-*cis*R	13-*cis* retinal
3-PGA	3-phosphoglyceric acids
ATR	all-*trans* retinal
BEUS	bias-exchange umbrella sampling
BPTI	bovine pancreatic trypsin inhibitor
BR	bacteriorhodopsin
C1C2	chimeric construct of *C. reinhardtii* ChR1 and ChR2
ChR	channelrhodopsin
COM	center of mass
CV	collective variable
$\mathrm{D_{min}}$	minimum distances
DEER	double electron-electron resonance
DMT	drug metabolite transporter

EGTA	ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid
FEP	free energy perturbation
FTIR	fourier transform infrared
GPT	glucose 6-phosphate/phosphate translocator
GsGPT	*Galdieria sulphuraria* GPT
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IF	inward-facing
IQR	interquartile range
LFEP	lowest free energy pathway
MD	molecular dynamics
MFS	major facilitator superfamily
Occ	occluded
OF	outward-facing
P_i	inorganic phosphate
pathCV	path collective variable
pathCV(ζ)	path collective variable ζ
pathCV(s)	path collective variable s

PC1	the first principal component of protein
PCA	principal component analysis
PDB	protein data bank
PHSM	*post-hoc* string method
PME	particle mesh Ewald
PMF	potential of mean force
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PPT	phospho*enol*pyruvate/phosphate translocator
pPT	plastidic phosphate tranlocator
RBF	radial basis function
REMD	replica-exchange molecular dynamics
RMSD	root-mean-square deviation
s.d.	standard deviation
s.e.m.	standard error of mean
SEC-MALS	size exclusion chromatography-multi angle light scattering
SMD	steered molecular dynamics
SMR	small drug resistance
SMwST	string method with swarms of trajectories
SnYddG	*Starkeya novella* YddG
SRII	sensory rhodopsin II

TI	thermodynamic integration
TM	transmembrane
TP	triose phosphates
TPT	triose-phosphate/phosphate translocator
WHAM	weighted histogram analysis method
WT	wild type
XPT	xylulose 5-phosphate/phosphate translocator

Table of amino acid abbreviations

3-letter abbreviation	1-letter abbreviation	Full name
Ala	A	alanine
Cys	C	cystein
Asp	D	aspartic acid
Glu	E	glutamic acid
Phe	F	phenylalanine
Gly	G	glycine
His	H	histidine
Ile	I	isoleucine
Lys	K	lysine
Leu	L	leucine
Met	M	methionine
Asn	N	asparagine
Pro	P	proline
Gln	Q	glutamine
Arg	R	arginine
Ser	S	serine
Thr	T	threonine
Val	V	valine
Trp	W	tryptophan
Tyr	Y	tyrosine

1 General Introduction

1.1 Basics of membrane channels and transporters

1.1.1 Membrane transport

Cells and organelles are separated from the external environment by the lipid bilayer, which is mainly composed of phospholipids. Uptake of compounds essential for cells or organelles and elimination of unnecessary compounds are important for cell growth. However, free diffusion of the hydrophilic compounds and ions is so slow that cells and organelles have the mechanism that selectively facilitates the diffusion of the compounds and ions, called membrane transport.

Membrane transport is classified into 2 classes by its transport direction. The first class is passive transport (Fig. 1.1; left and center). The passive transport facilitates the "downhill" translocation of substrates, in which the substrates diffuse across the membrane along their electrochemical potential. Thus, the passive transport system itself does not need the external energy source for transporting. This passive transport includes membrane channel and some types of membrane transporter (passive transporter) molecules. The other class of membrane transport is active transport (Fig. 1.2; right). This system translocates substrates against their electrochemical potential, which means the active transport system requires an external energy source, such as ATP hydrolysis energy. The other types of transporter molecules (active transporters) are classified into this active transport system.

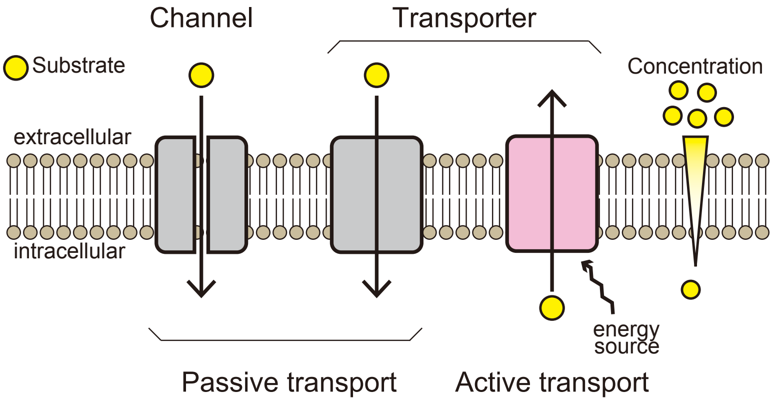


Figure 1.1: Schematic diagram for the classification of membrane transport.

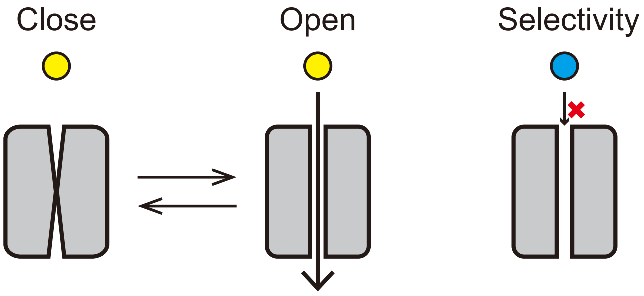


Figure 1.2: Schematic representation of the membrane channel. Open/close regulation (conformational change) and the substrate selectivity are the required mechanism for membrane channels.

1.1.2 General mechanisms of membrane channels

Membrane channels function as the passive transport system in which the substrates (in most case, inorganic ions) are conducted along their electrochemical potential. In the transport process of channels, the number of transported substrate molecules is not regulated. Although some channels constantly open the gate and function as a leaky channel, conformational regulation is required for most membrane channels. In the non-conducting, closed state, a leakage of the substrates must be blocked. Therefore, the conformational transition mechanism required for the membrane channels is the transition between two states, opening, and closing of substrate conducting pathway (Fig. 1.2; left).

Another requirement for the membrane channels is substrate selectivity (Fig. 1.2; right). Some channels conduct ions non-selectively while others conduct ions with high specificity.

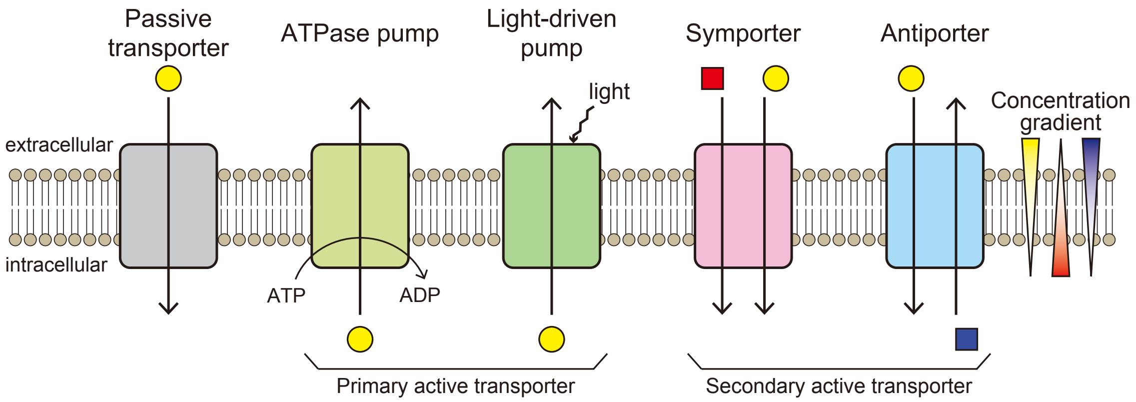


Figure 1.3: Classification of membrane transporters. The gradient for each substrate is shown in gradation color bar.

These two mechanisms, open/close regulation, and substrate selectivity are universally required for all membrane channel, and understanding these mechanisms is essential.

1.1.3 Classification of membrane transporters

Membrane transporters include both types of passive and active transport, called passive transporters and active transporter, respectively. (Fig. 1.3).

Passive transporters, also called permease, conducts the “down-hill” transport of substrates. The difference between channels and passive transporters is their stoichiometry: in the passive transporters, a fixed number of substrates are transported upon a single conformational change cycle.

In contrast, active transporters conduct the “up-hill” transport of the substrates, which requires an extra energy source. Active transporters are further divided into two classes according to the energy source: primary active transporters and secondary active transporters. The primary active transporters utilize light, redox or ATP hydrolysis energy, for up-hill translocating of substrate (Fig. 1.3; light green and green). The secondary active transporters catalyze a transport of compounds against their electrochemical potential, using the electrochemical potential of another compound (co-substrate) created by the primary active transporters (Fig. 1.3; magenta and cyan).

The secondary active transporters include two classes according to the transport direction of substrate and co-substrate. The first group is called symporter, which translocates

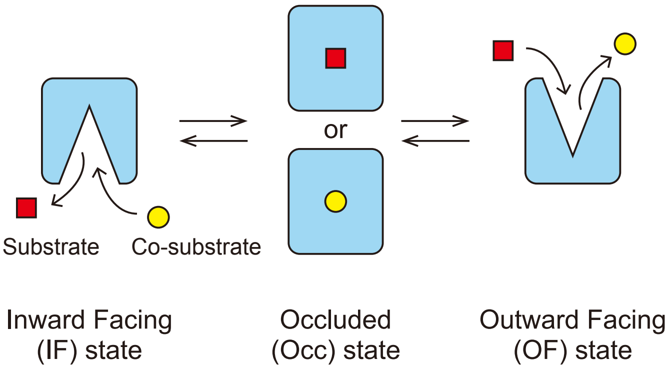


Figure 1.4: Conceptual model of alternating access mechanism. An example of antiporter is shown.

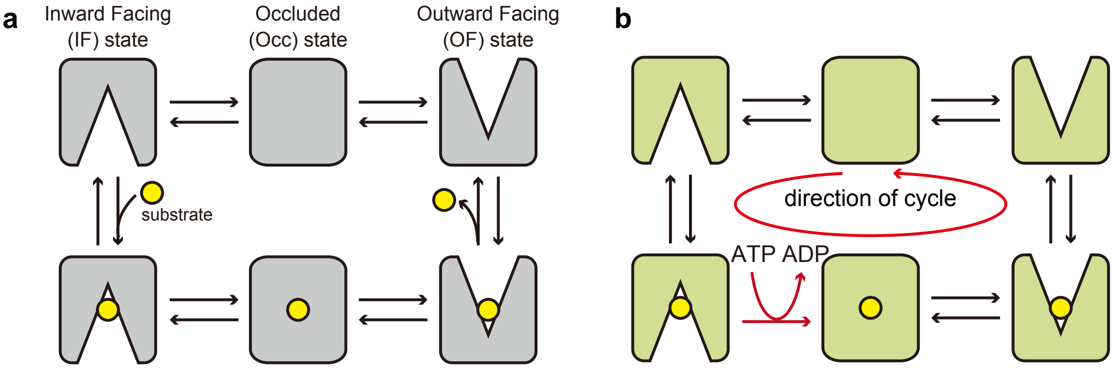


Figure 1.5: Simplified mechanisms of the transport cycle of (a) passive membrane transporter and (b) primary active transporter (ATPase pump). Note that the binding/unbinding process of ATP and ADP is omitted in the panel b.

substrate and co-substrate in the same direction (Fig. 1.3; magenta). The other group is called antiporter, which translocates substrate and co-substrate in the opposite direction (Fig. 1.3; cyan) In both cases, the co-substrates is transported along their electrochemical potential while the substrates are transported against their electrochemical potential. These active transporters achieve the uptake of compounds essential for cells or organelles or the elimination of unnecessary compounds.

1.1.4 General mechanisms of transporters

Generally, all membrane transporters achieve the transport of substrates by undergoing conformational transitions between the inward-facing (IF) state, the outward-facing (OF) state, and their intermediate, the occluded (Occ) state (Fig. 1.4). In this conformational

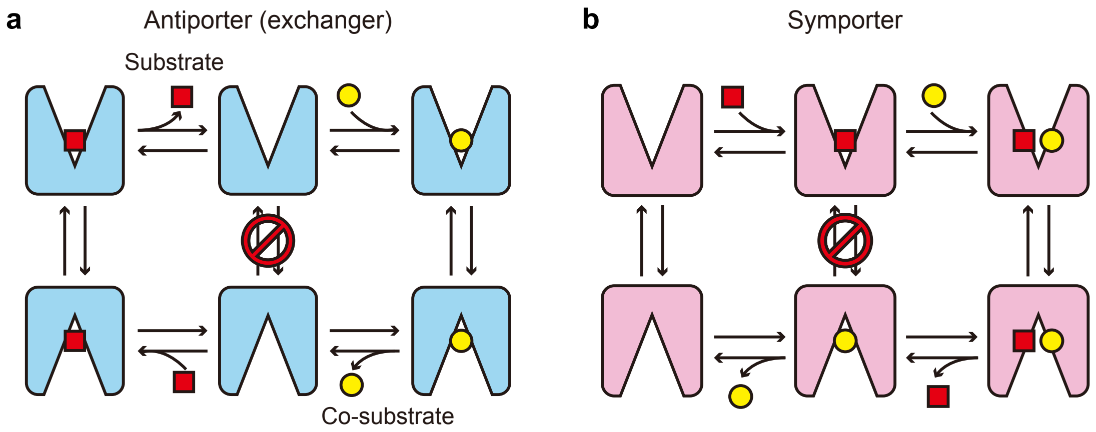


Figure 1.6: Simplefied transport cycle model of secondary active transporters in (a) antiporters and (b) symporters.

change scheme, substrate binding site, which usually locates at the center of the transmembrane domain, is exposed to intracellular or extracellular side alternately. This mechanism is called the alternating access mechanism [1, 2] and universally required for all transporters, since opening the gates on both sides of the membrane would permit the free diffusion of the substrate across the membrane, which means the transporter molecule would function as a channel. Thus, we can illustrate the conformational change as a cycle reaction like Fig. 1.5a.

In the case of the passive transporter, there is no restriction to the conformational transition and all state transitions shown in Fig. 1.5a are regarded as reversible processes at equilibration. Thus, the conformational transition between IF and OF is permitted regardless of whether the substrate is bound or not in the passive transporters.

In the primary active transporter, at least one transition of the equilibrium cycle is an irreversible process (an example of ATPase pump is shown Fig. 1.5b). This irreversibility is guaranteed by the extra energy source, such as ATP hydrolysis or light reception. This mechanism drives the cycle single direction, resulting in the up-hill transport of substrate. In this case, the coupling between ATP hydrolysis and substrate binding is important; otherwise, wasteful ATP hydrolysis would take place.

In the secondary active transporters, the conformational change of antiporters and symporters must be strictly coupled with substrate binding/unbinding. In the case of antiporters, the conformational transition between IF and OF is permitted when and only when the substrate or co-substrate is bound, and the transition in *apo* state is prohibited (Fig. 1.6a). On the other hand, in the case of symporters, the conformational transition between IF and OF is permitted when both substrates are bound or when the binding site is not occupied. And the conformational transition is prohibited when either of the substrate or co-substrate is bound to the binding site (Fig. 1.6b). These coupling mechanisms between conformational transition and substrate binding, in addition to the alternating access mechanism, are a necessary and sufficient condition the secondary active transporters. And understanding these coupling mechanisms is essential to elucidate the fundamental mechanism of the secondary active transporter.

1.2 Basics of molecular dynamics (MD) simulation

1.2.1 Conventional MD simulation

Structural information of proteins is essential to understand their functions and mechanisms. Most of the structures of proteins are obtained by X-ray crystallography since it provides the high-resolution atomic structure for a wide range of molecular weight protein. However, the protein structure of X-ray crystallography is the static snapshot which is averaged over all protein molecules in the crystal, although the actual protein structure in solution is fluctuating around the energetically stable structure. Moreover, as discussed in section 1.1, many proteins undergo large conformational change, which usually cannot be observed in a single crystal structure.

Molecular dynamics (MD) simulation enables to observe the dynamics of biological molecules by numerically solving the equation of motion for all atoms including protein, water, lipids, and ions. Since the first MD simulation of biological molecule concerning the bovine pancreatic trypsin inhibitor (BPTI) was performed [3], MD simulations have been used to study the dynamics and conformational transitions of various proteins [4]. While only a several pico-seconds of simulation was performed in the first MD simulation of BPTI, recent evolution in computer technology facilitates the MD simulation with large biological molecules in long timescale (~ ns-μs).

1.2.2 Difficulty in free energy calculation by MD simulation

In theory, one can reconstruct the free energy landscape along the important degree of freedom from the trajectories of MD simulation. The free energy difference ΔG_{AB} between given states A and B are related to the probability of the system to be found in the state A (p_{A}) and in the state B (p_{B}):

$$\Delta G_{\mathrm{AB}} = -\beta^{-1} \ln \frac{p_{\mathrm{A}}}{p_{\mathrm{B}}} = -\beta^{-1} \ln \frac{\int_{\mathrm{B}} e^{-\beta U(\boldsymbol{x})} d\boldsymbol{x}}{\int_{\mathrm{A}} e^{-\beta U(\boldsymbol{x})} d\boldsymbol{x}} \tag{1.1}$$

where $\beta = 1/k_B T$ is inverse temperature of system, $U(\boldsymbol{x})$ is the internal energy of the system, and the integral $\int_{\mathrm{A}} d\boldsymbol{x}$ run over all the configurations that make up state A. To accurately estimate this probability, the sufficient number of sampling of the state transition between A and B is required. In practice, however, sampling the sufficient number of the state transition in finite computational time scale is difficult when the high free energy barrier exists between the state A and B. A recently developed specialized hardware for MD simulation, called Anton [5] enables to perform several hundreds of micro-seconds simulation, but this micro-seconds time scale is still short to sufficiently sample the state transition of the biological molecules; although a free energy minimum state is well sampled, the other free energy minima and the high free energy region such as transition state are not sampled.

1.2.3 Enhanced sampling methods in MD simulation

To solve the problem in calculating the free energy difference of protein conformation sampling, a lot of conformational sampling techniques called enhanced sampling methods are developed [6]. These methods can be grouped into following 4 categories.

The first strategy is exploiting the thermal fluctuations by setting the higher temperature of the system than the normal simulation. This includes the replica-exchange molecular dynamics (REMD) (also called parallel tempering) algorithm [7]. In REMD simula-

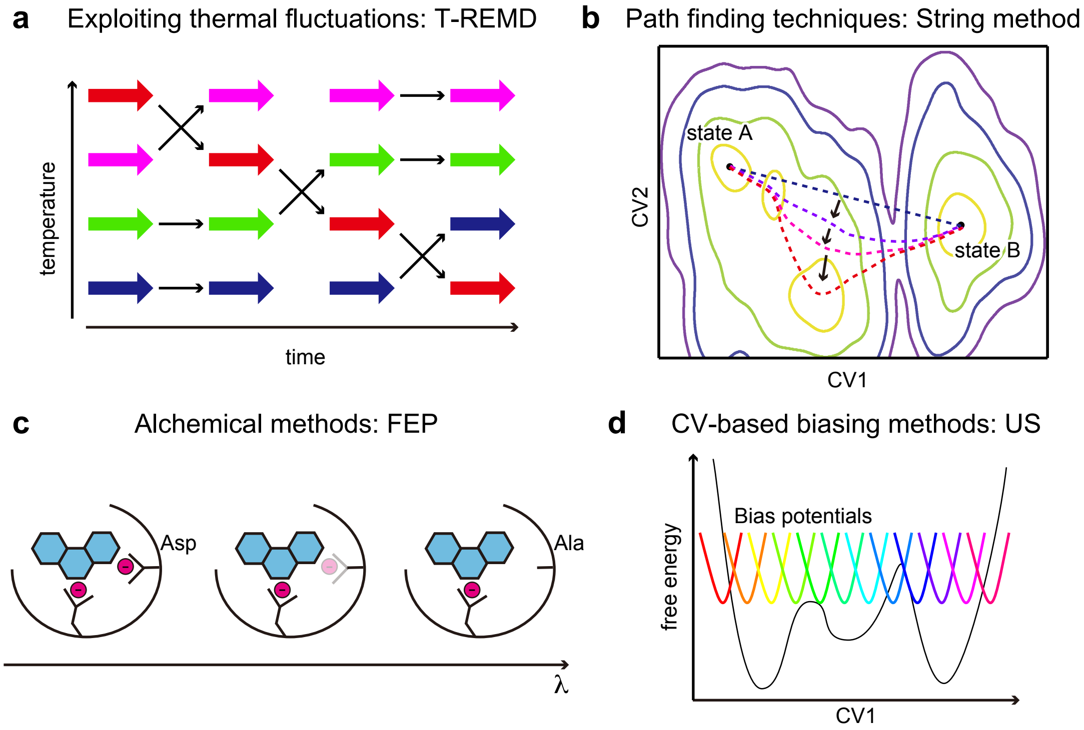


Figure 1.7: Schematic representations of enhanced sampling techniques in MD simulation. CV means collective variable. (a) The example of temperature-replica exchange MD (T-REMD), which is the strategy that exploits thermal fluctuations. The replicas with difference temperatures are exchanged to enhance the conformational sampling. (b) The example of string method, which is one of the transition path sampling techniques. The initial transition path (blue dashed line) is sequentially updated until the transition path is converged to the minimum free energy path (red dashed line). (c) Schematic representation of the alchemical method. In this example Asp residue in a ligand binding pocket in a protein gradually changes to Ala residue with coupling parameter λ. The free energy difference will be calculated by the free energy perturbation (FEP) or thermodynamic integration (TI) framework, to obtain the difference of binding energy. (d) The example of umbrella sampling (US) is shown. In this method, quadratic biasing potential (shown in rainbow parabolas; called windows) are spaced along a CV to enhance the sampling of the high free energy regions. The simulation results from independent windows are then post-processed by weighted histogram analysis method (WHAM) to obtain the free energy landscape (black line).

tion, multiple replicas with different temperatures are simulated in parallel, and exchanges between neighboring replicas are attempted periodically and accepted or rejected by the Metropolis criterion (Fig. 1.7a). These exchanges result in a random walk of temperature space, which enables more effective conformational sampling than the normal constant temperature simulation.

The second strategy is called path-finding technique. This strategy exploits the knowledge about the initial and final states of the conformational change and tries to find the optimal transition pathway between them, that is minimum free energy path. The representative algorithm is string method [8], which searches the minimum free energy path by iteratively updating the transition path (Fig. 1.7b). This method requires a measure of the progression of the system along the path.

If you are not interested in the transition state between two states and interested only in the free energy difference, you can use the so-called alchemical methods. In this framework, state A gradually disappear from the system, and state B gradually appears in the system, with the coupling parameter (usually called λ; Fig. 1.7c). By using the free energy perturbation (FEP) or thermodynamic integration (TI) frameworks, we can obtain the free energy difference between state A and state B. Note that this method is usually employed for calculating the free energy differences between the small changes, such as binding energy difference by a point mutation, and not used for calculating the free energy in large conformational change.

The last strategy is a biased sampling protocol, including umbrella sampling [9]. In this strategy, sampling is performed only along a pre-defined reaction coordinate (usually implemented by collective variable (CV)), which focuses on a limited region of the entire configuration space (Fig. 1.7d). A bias potential is used to enhance the sampling along the selected degrees of freedom, and after the simulation, the bias will be removed (or re-weighted) by an appropriate method.

In addition to these methods, the combination of these biased sampling methods with the replica-exchange approach is also developed, such as Hamiltonian-REMD [10] (H-REMD; also called bias-exchange umbrella sampling (BEUS) [11]). In this approach, the simulation of many replicas with a bias for different CVs are performed, and they are periodically exchanged under the Metropolis criterion. This method can speed-up the convergence of the free energy landscape.

1.3 Overview of this study

In this paper, the author performed MD simulations to reveal the molecular mechanisms of a membrane channel and a membrane transporter.

In chapter 2, a series of non-biased MD simulation for a light-gated cation channel, called channelrhodopsin, is performed. The author performed the MD simulations with protonated or deprotonated states for the titratable residues in the ion conducting pathway and revealed that deprotonation of a glutamate residue on ion conducting pathway causes an ion leakage in the ground state. The author further performed the MD simulation of the photo-activated, *i.e.*, 13-*cis* retinal bound state, and revealed the conformational change of the first step of channel opening.

In chapter 3, the author performed the molecular mechanism of triose-phosphate/phosphate translocator (TPT), which exchanges the product of photosynthesis (triose-phosphate) with inorganic phosphate (P_i) on the inner membrane of the chloroplast. By employing the enhanced sampling methods, the entire transport cycle of TPT is reconstructed from one P_i bound crystal structure. Based on these reconstructed structures, the fundamental mechanisms as an exchanger including substrate binding mechanism, alternating access mechanism and antiporter mechanism as described in section 1.1.4 are discussed.

2 Molecular Dynamics Simulation of Channelrhodopsin

2.1 Introduction

2.1.1 Channelrhodopsin (ChR)

Most organisms perceive light-carried information by using rhodopsin family proteins, which are covalently bound to a retinal chromophore. During evolution, rhodopsin family proteins have acquired divergent functions, as light sensors, ion pumps, and ion channels. Channelrhodopsin (ChR) is the only channel-type rhodopsin family protein identified thus far. ChR was originally isolated from *Chlamydomonas reinhardtii*, and was characterized as a light-gated cation channel in 2002 [12]. Upon excitation by blue light, ChR permeates several monovalent and divalent cations, including H^+, Na^+, K^+, and Ca^{2+}. As the net inward flow of cations under physiological ionic conditions depolarizes cell membranes, ChR has become not only an attractive target of biophysical studies, but also a powerful tool for the neuroscience field (optogenetics). Since ChR was first expressed in mammalian neurons in 2005 [13, 14], the number of studies using ChR has rapidly increased [15]. The integration of ChR and fiber-optic devices allows the control of neural activity with high spatial and temporal precision, even within systems such as freely moving mammals. Thus, ChR is now recognized as a powerful tool for modulating neural activity, in order to elucidate the neural circuit foundations of normal and pathological behaviors [15, 16]. However, despite the rapid progress in optogenetics, our understanding of the mechanism by which

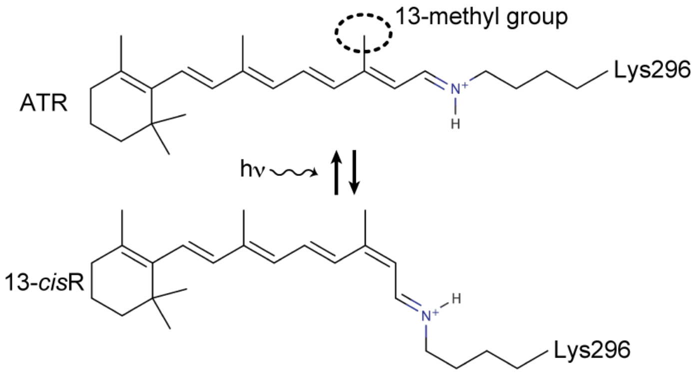


Figure 2.1: Photoisomerization of retinal molecule in microbial rhodopsin. Chemical structures of ATR and 13-*cis*R are shown.

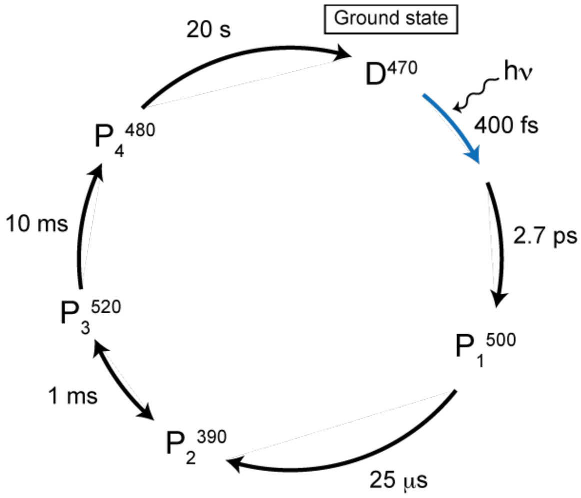


Figure 2.2: Simplified model of the ChR photocycle. The representative time scale is shown in each state transition.

ChR permeates cations in response to light is still limited.

Previous electrophysiological analyses of ChR revealed that its N-terminal extracellular domain (N-domain) and seven-transmembrane (TM) domain are sufficient for the channel function, while its large C-terminal intracellular domain is dispensable [12, 17]. In the center of the TM domain, an all-*trans* retinal (ATR) (Fig. 2.1) is covalently bound to Lys296 (Lys257 in ChR2) on TM7, thus forming the protonated Schiff base, and the positive charge of the Schiff base is stabilized by two counterions, Glu162 and Asp292 (Glu123 and Asp253 in ChR2, respectively). Upon receiving blue light, ATR isomerizes to 13-*cis* retinal (13-*cis*R) (Fig. 2.1), which triggers a series of chemical reactions and conformational changes, called the photocycle (Fig. 2.2). Electrophysiological and spectroscopic experiments revealed that the ChR photocycle comprises the early photoproducts, P_1^{500} and

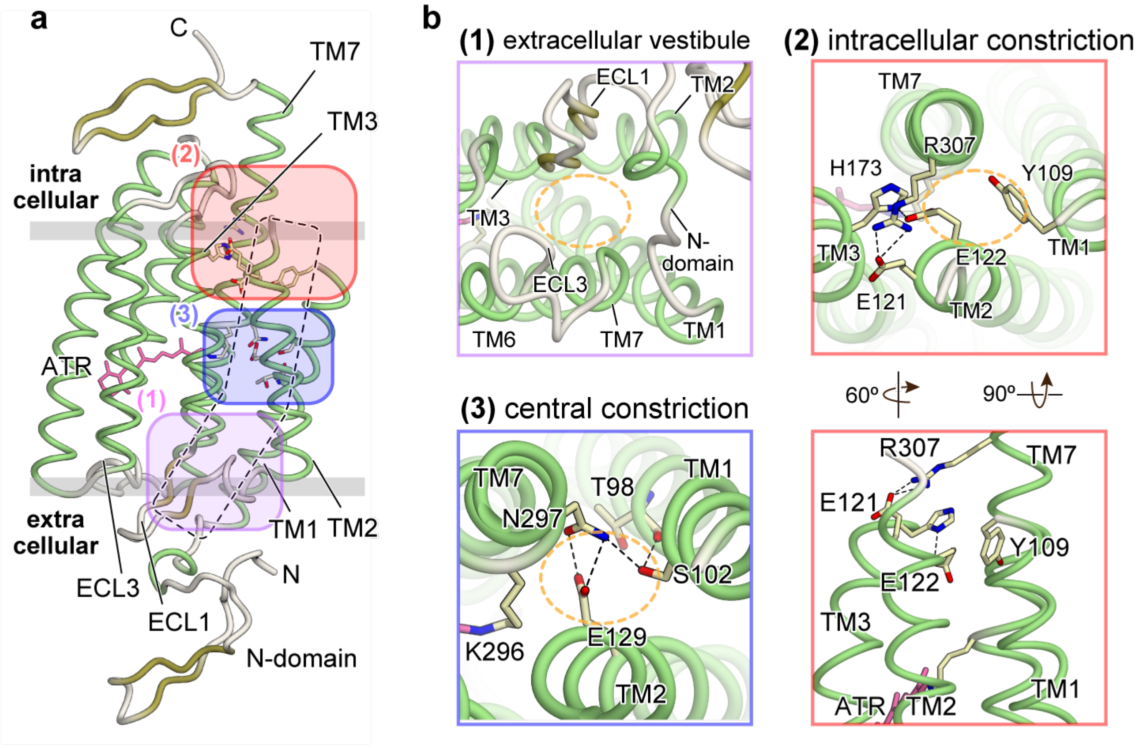


Figure 2.3: The crystal structure of the channelrhodopsin (ChR). The two constrictions in the crystal structure of C1C2 (PDB ID 3UG9) and the electrophysiological analysis of the constrictions. (a) Overall structure of C1C2, viewed parallel to the membrane with the three key regions highlighted (magenta, blue, and red). The dashed area represents the putative ion-conducting pathway. (b) Magnified views of the highlighted regions in (a). Black dashed lines are hydrogen bonds, and orange dashed circles represent the putative conducting pathway.

P_2^{390}, the conducting state, P_3^{520}, the late intermediate state, P_4^{480}, and other intermediates in the branched side pathways [18–23].

2.1.2 Crystal structure of ChR

In addition to these electrophysiological and spectroscopic analyses, the 2.3 Å resolution structure of the chimeric construct of *C. reinhardtii* ChR1 and ChR2 (C1C2) provided detailed insights into the architecture of ChR, including its cation-conducting pathway [17]. ChR forms a dimer in this crystal structure, and the electronegative pore in each monomer, formed by TMs 1, 2, 3, and 7, probably functions as the cation-conducting pathway, consistent with the previous computational and electrophysiological studies [17, 24–30] (Fig. 2.3a). The crystal structure revealed that an extracellular vestibule, formed by the N-domain and extracellular loops 1 and 3, expands to a diameter of about 8 Å, and suggested that this passage would allow water molecules to move from the extracellular side to the middle of the pathway [17] (Fig. 2.3b, panel (1)). In contrast to the extracellular half, the intracellular half of the pathway is occluded at two constrictions, revealing a closed state. This is consistent with the fact that C1C2 was crystallized and harvested in the dark. These two constrictions, called the intracellular and central constrictions (corresponding to Fig. 2.3a, panel (2) and (3) respectively), contain titratable residues, including Glu122 and Glu129, which are highly conserved in the ChR family, but not in other rhodopsin family members. These residues link the TM helices by hydrogen bonds and a salt bridge to occlude the conducting pathway, and these interactions are considered to be disrupted in the conducting state. Although the crystal structure of C1C2 identified certain residues that may be important for channel opening, it remains unknown how the residues in these constrictions prevent ion leakage in the ground state, and how the retinal isomerization induces the conformational change towards channel opening.

2.1.3 Overview of this study

In this study, the functional roles of Glu122 and Glu129 in the two constrictions were investigated by electrophysiological analyses and MD simulations. The results revealed that Glu129 in the central constriction is important for preventing ion leakage in the ground state. In addition, the conformational change upon retinal isomerization, which is the very first stage in the photocycle, was revealed by performing the MD simulation with 13-*cis* retinal-bound C1C2.

2.2 Materials and methods

2.2.1 Simulation system setup

The atomic coordinates of the crystal structure of ChR, C1C2 (protein data bank (PDB) ID 3UG9 [17]) were retrieved from the PDB. Two disordered regions (Gly110-Thr117 and

Asn327-Glu332) were modeled with the program Modeller [31]. The protonation state of each glutamate or aspartate was determined based on the pattern of hydrogen bonds observed in the crystal structure, as well as the results of the PROPKA calculation [32], except for Glu163, which was reported to be deprotonated in the ground state [33]. The lipid bilayer composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules was generated using the VMD plugin [34]. The full simulation system included C1C2, POPC, water molecules, and chloride and sodium ions. The system size was 128×128×128 Å, and contained about 190,000 atoms.

All of the water molecules observed in the crystal structure were kept. The missing hydrogen atoms were built with the program VMD [34]. An explicit solvent periodic boundary system was prepared. The net charge of the simulation systems was neutralized through the addition of chloride and sodium ions upto the concentration of 150 mM NaCl. The molecular topologies and parameters from Charmm22/CMAP force field [35, 36] were used for proteins, and from the Charmm36 force field [37] were used for the lipid molecules. The topology and force field parameters for 13-*cis*R were the same as those for ATR, which are also compatible with 13-*cis*R (designed by the collaborators, Dr. Motoshi Kamiya (RIKEN) and Dr. Shigehiko Hayashi (Kyoto University)), developed based on the previously reported values [38–41].

2.2.2 Production run

Molecular dynamics simulations were performed with the program NAMD 2.8 [42]. The systems were first energy minimized for 1,000 steps with fixed positions of the non-hydrogen atoms, and then for another 1,000 steps with 10 kcal mol^{-1} Å^{-2} restraints for the non-hydrogen atoms, except for the lipid molecules within 5.0 Å from the proteins. Next, equilibration was performed for 0.01 ns under isothermal (canonical) ensemble (*NVT*) conditions, with 10 kcal mol^{-1} Å^{-2} restraints for the non-hydrogen atoms of the protein. Finally, equilibration was performed for 0.5 ns under isothermal-isobaric ensemble (*NPT*) conditions with the 1.0 kcal mol^{-1} Å^{-2} restraints. Production process for each protonation state was performed in 150 ns as shown in Table 2.1. In the equilibration and production processes, the pressure and temperature were set to 1.0 atm and 300 K, respectively. Constant temperature was maintained by using Langevin dynamics. Constant pressure was maintained by using the Langevin piston Nosé-Hoover method [43]. Long-range electrostatic interactions were calculated by using the particle mesh Ewald (PME) method [44]. Each simulation was performed at least twice with different initial velocities, and similar results are obtained from these trajectories.

2.2.3 Modeling of 13-*cis* retinal bound C1C2

The coordinates of Lys-13-*cis*R were obtained from the K-intermediate of bacteriorhodopsin (PDB ID: 1IXF [45]). The main chain atoms (*i.e.*, the C, O, N, and CA atoms) and the β-ionone ring of this Lys-13-*cis*R molecule were superimposed on those of Lys296-ATR of C1C2, and then the Lys296-ATR of C1C2 was replaced with the Lys-13-*cis*R. After this replacement, 1,000 steps of energy minimization were performed with fixed positions of the non-hydrogen atoms. Subsequently, another 1,000 steps of energy minimization were performed with 10 kcal mol^{-1} restraints for the non-hydrogen atoms, except for the lipid molecules within 5.0 Å from the protein. The equilibration and production runs were performed in the same manner as described above.

2.2.4 Data analysis

For each obtained trajectories, the minimum distance between a pair of residues, the root-mean-square deviation (RMSD) values for the atom groups of interest, and the correlation coefficients between all C$_\alpha$ atoms were calculated. The calculations were performed by the QScript implemented in cuemol (`http://www.cuemol.org/`) or a python package for analyzing the MD trajectories, MDAnalysis [46]

The RMSD was calculated using following equation:

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i}^{N}\left|\boldsymbol{r}_i^t - \boldsymbol{r}_i^{ref}\right|^2} \tag{2.1}$$

where $\boldsymbol{r}_i^t$ is the coordinate of i-th atom at time t, $\boldsymbol{r}_i^{ref}$ is the coordinate of i-th atom of reference structure, and N is the total number of atom. The RMSD calculations were performed after superimposing each trajectory frame on TM helices 1-7 of the crystal structure.

The correlation coefficient between i-th and j-th C_α atoms (C_{ij}) was calculated using following equation:

$$C_{ij} = \frac{\left\langle (\boldsymbol{r}_i - \bar{\boldsymbol{r}}_i) \cdot (\boldsymbol{r}_j - \bar{\boldsymbol{r}}_j) \right\rangle}{\sqrt{\left\langle |\boldsymbol{r}_i - \bar{\boldsymbol{r}}_i|^2 \right\rangle \left\langle |\boldsymbol{r}_j - \bar{\boldsymbol{r}}_j|^2 \right\rangle}} \tag{2.2}$$

where $\boldsymbol{r}_i$ is the coordinate of i-th C_α atom, $\langle \boldsymbol{x} \rangle$ is the time average of vector $\boldsymbol{x}$, and $\bar{\boldsymbol{r}}_i$ is the averaged coordinate of i-th C_α atom. The correlation coefficient calculations were performed after superimposing each trajectory frame on TM helices 3-5 of the crystal structure.

2.2.5 Electrophysiological analysis

The electrophysiological analysis of C1C2 wild type (WT) and mutants were performed in collaboration with Dr. Ito (King's college London) and Dr. Maturana (Nagoya university).

For electrophysiology, human embryonic kidney (HEK)293 cells were cultured on polylysine-coated, glass bottom culture dishes (Matsunami), and were transfected with 0.1 mg of a plasmid construct encoding the GFP-tagged C1C2 WT or the GFP-tagged C1C2 mutants. At 24-30 h after transfection, the cells were placed in bath medium, containing 140 mM NaCl, 1 mM $CaCl_2$, 2 mM $MgCl_2$, 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.4 with NaOH), and 5 mM glucose, under an inverted microscope (Olympus IX71). A borosilicate patch pipette (Harvard Apparatus), with a resistance of about 5-8 MΩ, was filled with 140 mM KCl, 5 mM ethylene glycol-bis(2-aminoethylether)-*N,N,N',N'*-tetraacetic acid (EGTA), 2 mM $MgCl_2$, and 10 mM HEPES (pH 7.2 with KOH). The C1C2 currents were recorded in both the voltage-clamp mode and whole-cell configuration. The cells were held at a membrane potential of -80 mV, and were depolarized by 10 mV voltage steps of 1.8 s up to 170 mV. The light-dependent currents were activated 200 ms after the depolarization step, with 465 nm light (1.5 mW mm^{-2}) for 1,000 ms, elicited by a high power LED illumination system (LEX2-B, Brainvision) connected to an A/D converter (Digidata 1440, Axon CNS, Molecular Devices), and controlled by the pClamp10 software (Axon CNS, Molecular Devices). Currents were measured using an Axopatch 200B amplifier (Axon CNS, Molecular Devices), filtered at 2 kHz and sampled at 5 kHz, using a Digidata 1440A digitizer (Axon CNS, Molecular Devices) controlled by the pClamp10 software (Axon CNS, Molecular Devices).

To estimate the membrane expression levels of C1C2 and its mutants, the ratio of the membrane and cytosolic fluorescence values was determined. The cells were transfected with 0.1 mg WT C1C2 or C1C2 mutants, using Fugene 6, and cultured for 30 h. They were then washed with PBS, fixed with 4% paraformaldehyde in PBS for 20 min at room temperature (20 °C), and washed again with PBS before microscopy observations. GFP fluorescence was observed with a laser confocal microscope (FV1000 Olympus).

2.3 Results

2.3.1 Deprotonation of Glu122 and Glu129 is important for channel activity

The crystal structure of C1C2 revealed the central and intracellular constrictions in the pore [17]. In the intracellular constriction, the salt bridge between Glu121 and Arg307 links TM2 to TM7, and the hydrogen bond between Glu122 and His173 links TM2 to TM3 (Fig. 2.3b panel (2)). In the central constriction, the hydrogen bond between Glu129 and Asn297 links TM2 to TM7 and occludes the pore (Fig. 2.3b panel (3)). In this crystal structure, hydrogen bonds are formed between Glu122 and His173, and Glu129 and Asn297, suggesting the protonation of the side chains of Glu122 and Glu129 [17]. To clarify the functional importance of the protonation states of these glutamate residues, the E122Q and E129Q mutants were prepared, expressed in HEK293 cells, and their photocurrents in response to 465-nm light pulses were recorded. The results revealed that the E122Q and E129Q single mutants each showed a significant decrease in the photocurrent,

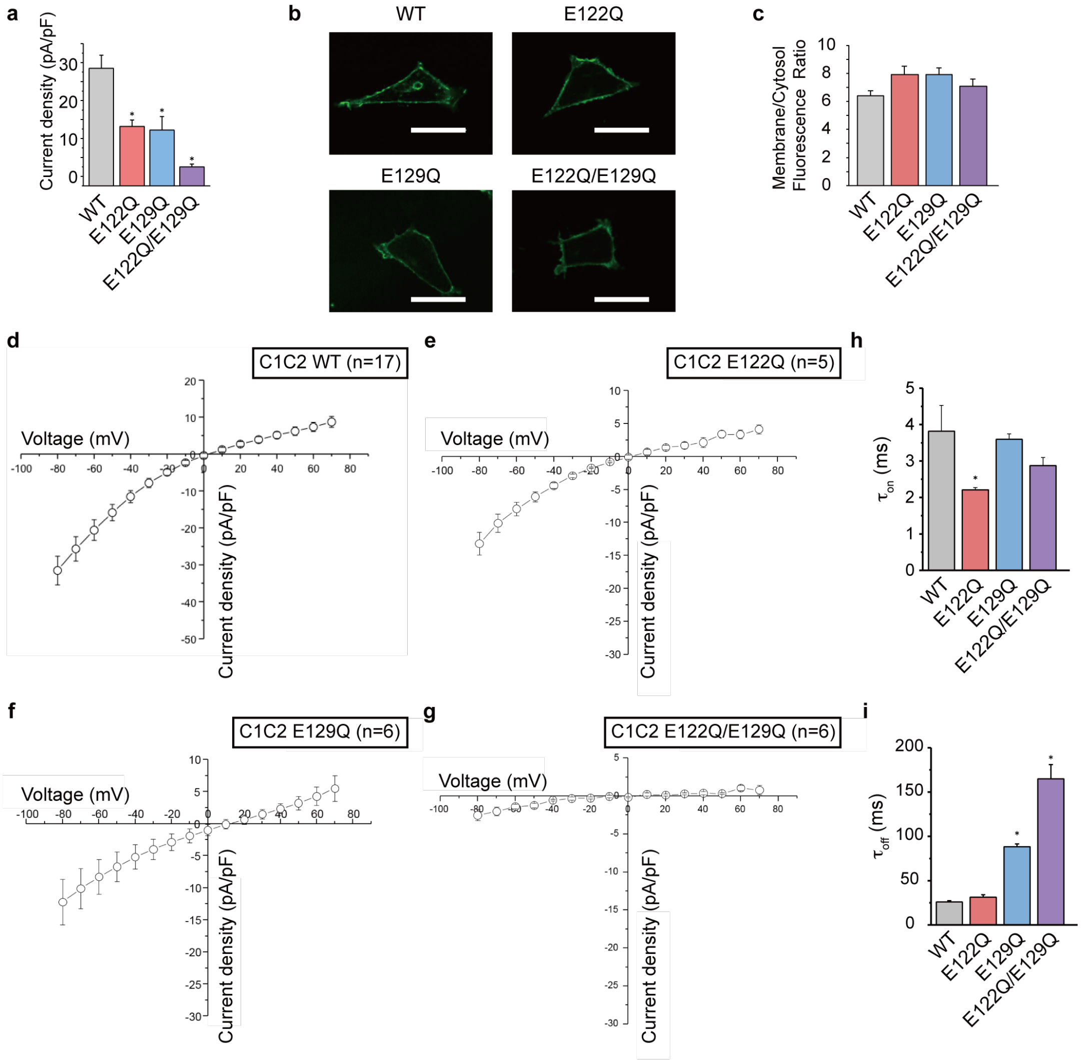


Figure 2.4: Electrophysiological analysis of C1C2. (a) The peak amplitudes of the photocurrents, normalized by the cell's input capacitance. (b) Conforcal images of representative HEK293 cells expressing the C1C2 WT and its mutants. Scale bar represents 30 μm. (c) The expression level of each C1C2 variant measured by the membrane/cytosol ratio of GFP fluorescence. (d-g) The current-voltage (I-V) relation curves for each mutant. (h, i) The kinetic parameters for each mutant, (h) opening rates (τ_{on}) and (i) closing rates (τ_{off}). The error bars represent standard error of mean (s.e.m.) of 3 experiments (n = 5-17 cells). * $p < 0.05$.

This analysis was performed in collaboration with Assoc. prof. Andrés D. Maturana (Nagoya University) and Dr. Jumpei Ito (King's College London).

Table 2.1: Simulation set list performed in this study. “p” represents that the residue(s) is protonated, and “Δp” represents that the residue(s) is deprotonated.

name	Retinal	E122	E129	time
ATR-E122p-E129p	ATR	p	p	150 ns
ATR-E122Δp-E129Δp	ATR	Δp	Δp	150 ns
ATR-E122p-E129Δp	ATR	p	Δp	150 ns
ATR-E122Δp-E129p	ATR	Δp	p	150 ns
13-*cis*R-E122Δp-E129p	13-*cis*R	Δp	p	150 ns

and that the double mutation E122Q/E129Q almost completely abolished the photocurrent (Fig. 2.4a and d-g), despite their stable membrane expression (Fig. 2.4 b, c). Given that the glutamine mutation can mimic the protonated glutamic acid, these results suggested that Glu122 and Glu129 (Glu83 and Glu90 in ChR2) are deprotonated during the photocycle. These findings are consistent with previous spectroscopic studies showing that Glu90 in ChR2 is protonated in the dark/closed state, and its protonation state changes during the photocycle [24, 47]. In addition, the calculation of the kinetic parameters, τ_{on} and τ_{off}, showed that the E122Q and E129Q mutations affected the τ_{on} and τ_{off} values, respectively (Fig. 2.4h, i).

2.3.2 Deprotonation of both Glu122 and Glu129 results in proton leakage in the ground state

To investigate the effects of the protonation states of Glu122 and Glu129 on the ChR structure, atomic models of C1C2 with different protonation states were created based on the crystal structure (PDB entry 3UG9 [17]), and performed all-atom MD simulations in the presence of a POPC lipid bilayer.

First, the ATR-E122p-E129p simulation was performed (Table 2.1), which has the same protonation state as the crystal structure. In this simulation, the overall structure was stable, and no large structural changes were observed (Fig. 2.5a; gray). The interactions in both the central and intracellular constrictions were also preserved during the 150 ns simulation. The salt bridge between Glu121 and Arg307 in the intracellular constriction, and the

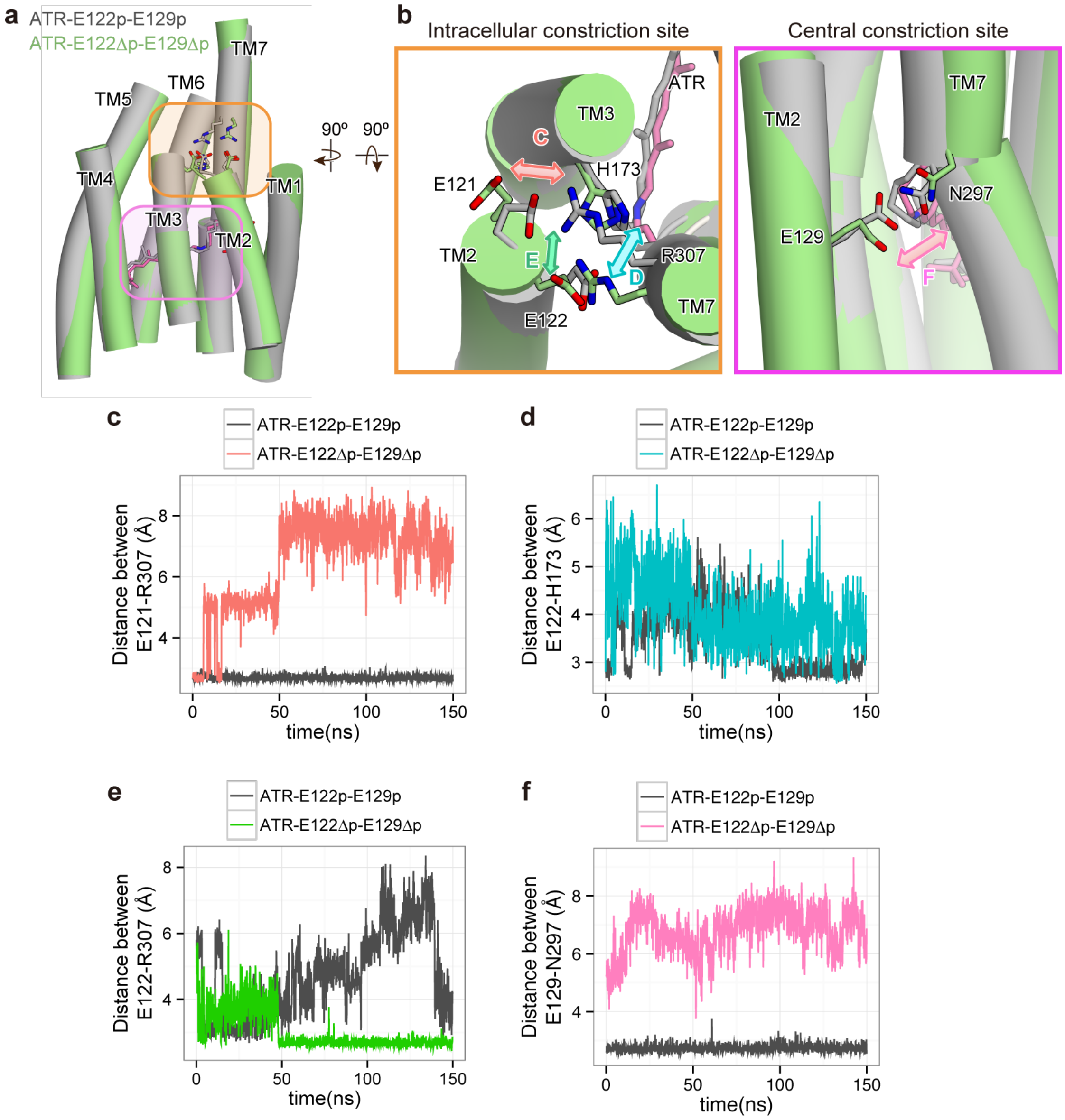


Figure 2.5: Structural comparison of the intracellular constrictions between the ATR-E122p-E129p and ATR-E122Δp-E129Δp simulations. (a) Overall structures of snapshots from the last frame of both simulations. Key residues are highlighted in orange and magenta. (b) Magnified view of the intracellular and central constrictions (left and right panels, respectively). Double arrows indicate the possible motions of Glu121-Arg307 (red), Glu122-His173 (cyan), Glu122-Arg307 (green) and Glu129-Asn297 (magenta). (C-F) Distances between (C) Glu121-Arg307, (D) Glu122-His173, (E) Glu122-Arg307, and (F) Glu129-Asn297.

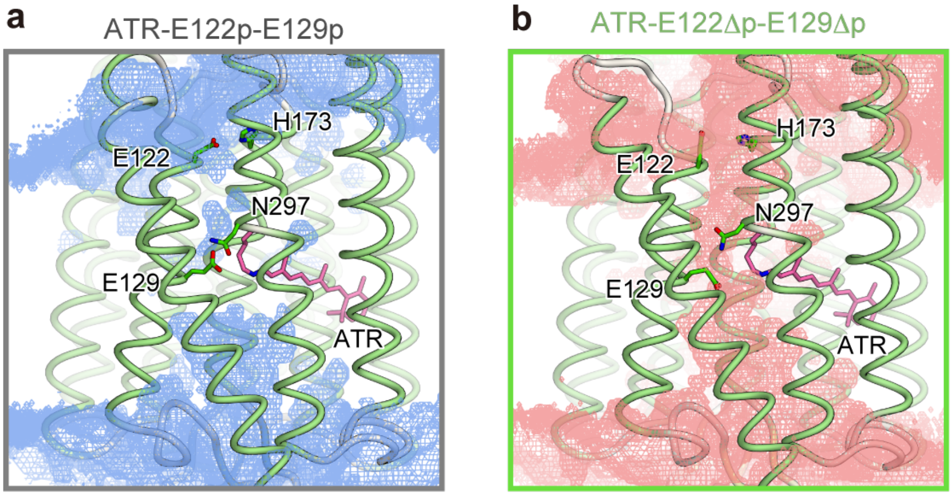


Figure 2.6: Distribution of water molecules in the (a) ATR-E122p-E129p and (b) ATR-E122Δp-E129Δp simulations. The distribution maps are contoured at the probability density of 0.0015 molecules $Å^{-3}$ ns^{-1}. The time-averaged structure of the protein over 150 ns is shown.

hydrogen bond between Glu129 and Asn297 in the central constriction, were stable (Fig. 2.5c, f; gray line). The hydrogen bond between Glu122 and His173 in the intracellular constriction was disrupted and reformed again during the 150 ns simulation (Fig. 2.5d). As a result, a small number of water molecules entered through the intracellular constriction (Fig. 2.6a). The extracellular side of the channel pathway, especially the extracellular vestibule observed in the crystal structure, was filled with water molecules (Fig. 2.6a). However, no water channel was formed between the intra- and extracellular sides of the membrane, due to the stable hydrogen-bonds of the central constriction (Glu129-Asn297).

Next, the results of the ATR-E122p-E129p and ATR-E122Δp-E129Δp simulations were compared (Table 2.1), to examine the effects of the protonation states of Glu122 and Glu129. In the ATR-E122Δp-E129Δp simulation, the overall structure was stable (Fig. 2.5a; green), but several changes in the intracellular and central constrictions were observed. In the intracellular constriction, the salt bridge between Glu121 and Arg307 was disrupted in the 150 ns simulation (Fig. 2.5c; red line), and the hydrogen bond between Glu122 and His173 was also disrupted, as a result of the deprotonation of Glu122 (Fig. 2.5d; cyan line). Instead, the deprotonated Glu122 formed a stable salt bridge with the side chain of Arg307 (Fig. 2.5e; green line). In the central constriction, the hydrogen bond

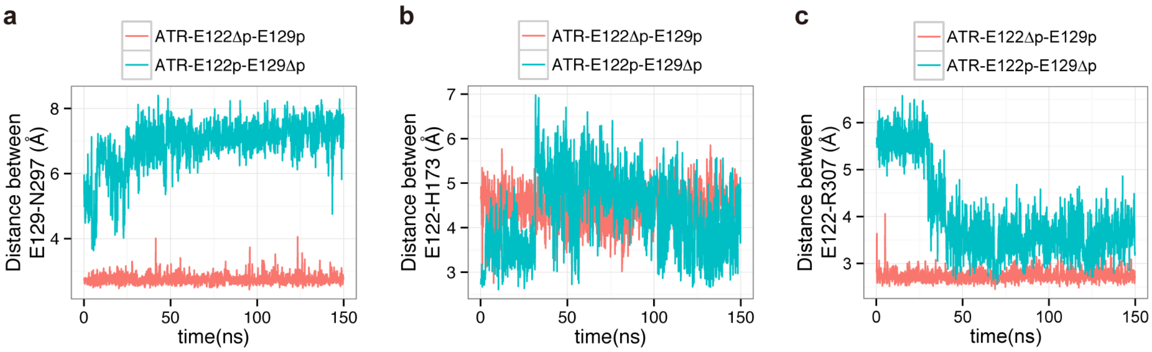


Figure 2.7: Distances between (a) Glu129-Asn297, (b) Glu122-His173 and (c) Glu122-Arg307 in the ATR-E122p-E129Δp and ATR-E122Δp-E129p simulations.

between Glu129 and Asn297 was disrupted upon the Glu129 deprotonation (Fig. 2.5f; magenta line). The water distribution in the ATR-E122Δp-E129Δp simulation revealed that water molecules from both the intracellular and extracellular entrances filled the channel pathway, which resulted in the formation of the water-mediated hydrogen-bond network connecting the intra- and extracellular sides of the membrane (Fig. 2.6b). The formation of this water-mediated hydrogen-bond network in the ground state may result in the leakage of protons across the membrane, without channel activation. Therefore, the protonation of Glu122 and/or Glu129 is important for preventing ion leakage in the ground state.

2.3.3 Glu129 protonation is important for preventing proton leakage

To investigate the effect of the deprotonation of each glutamate residue, we performed simulations in which either Glu122 or Glu129 was deprotonated (ATR-E122p-E129Δp and ATR-E122Δp-E129p ; Table 2.1). In these simulations, similar structural changes to those in the ATR-E122Δp-E129Δp simulation were observed: the Glu129-Asn297 hydrogen bond was disrupted upon Glu129 deprotonation in the ATR-E122p-E129Δp simulation (Fig. 2.7a; cyan line), while the Glu122-His173 hydrogen bond was disrupted and the salt bridge between Glu122 and Arg307 was formed upon Glu122 deprotonation in the ATR-E122Δp-E129p simulation (Fig. 2.7b, c; red line). However, the water distributions revealed that no water channels were formed in these simulation (Fig. 2.8), suggesting that

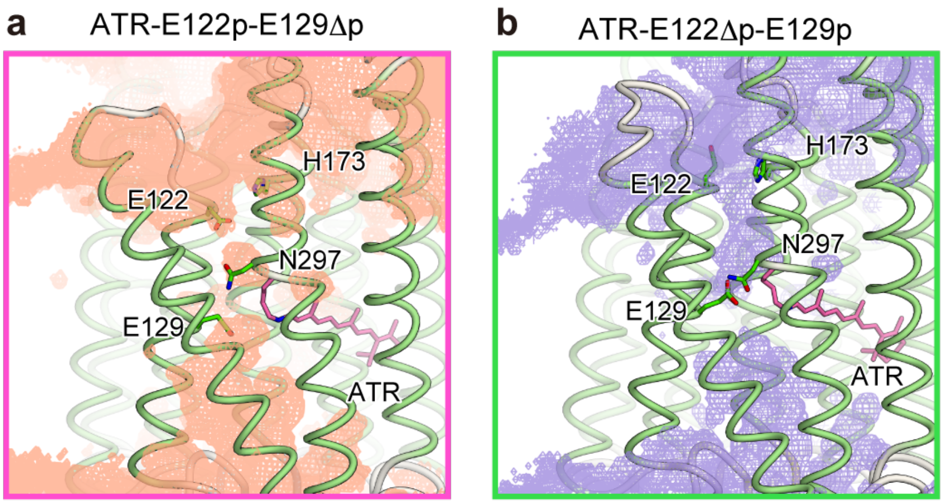


Figure 2.8: Distributions of water molecules in the ATR-E122p-E129Δp and ATR-E122Δp-E129p simulations. The distribution map is contoured at the probability density of 0.0015 molecules $Å^{-3}$ ns^{-1}. The time-averaged structure of the protein over 150 ns is shown.

ion or proton leakage did not occur. In the ATR-E122p-E129Δp simulation, the water permeation was disturbed by the interaction between Glu122 and His173, and a small number of water molecules entered through these constrictions, similar to the ATR-E122p-E129p simulation. In the ATR-E122Δp-E129p simulation, the hydrogen bond between Glu129 and Asn297 blocked the water permeation. In contrast to the case of the Glu122-His173 interaction, no water molecules passed through the channel pathway, because of the stable Glu129-Asn297 hydrogen bond. Therefore, we hypothesized that either Glu122 or Glu129 is protonated in the ground state, to prevent the water channel formation and proton leakage.

A recent structural analysis of the C1C2 variant under neutral-pH conditions suggested that Glu122 is deprotonated in the ground state [48]. In this C1C2 variant, although the overall structure is essentially the same as that of the previous C1C2 structure obtained under low-pH conditions, the hydrogen bond between Glu122 and His173 is absent, and instead Glu122 forms a salt bridge with Arg307. This result suggested that Glu122 is deprotonated and interacts with Arg307, rather than His173, under neutral-pH conditions, as observed in the ATR-E122Δp-E129Δp and ATR-E122Δp-E129p simulations (Fig. 2.5d, e and Fig. 2.7b, c). Therefore, the author concluded that Glu122 is deprotonated in the

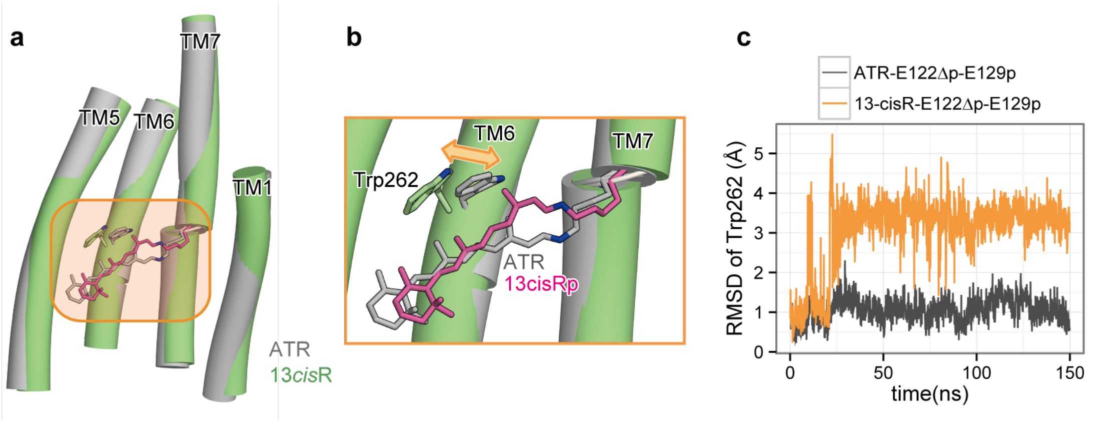


Figure 2.9: The conformational change in Trp262 upon retinal isomerization. (a) Structural comparison between the snapshots from the ATR-bound (grey) and 13-*cis*R-bound (green) simulations. (b) Magnified view of retinal and Trp262, from the orange-highlighted region in the left panel. Double arrows indicate the possible motions of Trp262. (c) The RMSD values of the Trp262 atoms, relative to those of the crystal structure.

ground state, and that the protonated Glu129 and its hydrogen bond with Asn297 are important for preventing ion leakage across the membrane.

2.3.4 ATR isomerization to 13-*cis*R results in the movement of the cytoplasmic regions of TM6 and TM7

The initial event of the photocycle, upon the absorption of blue light, is the isomerization of all-*trans* retinal (ATR) to 13-*cis* retinal (13-*cis*R), which occurs on a femto-second time scale (Fig. 2.2) [19,23]. To investigate this initial conformational change in the photocycle, the MD simulation of C1C2 with 13-*cis*R was performed. The initial structure was modeled by simply replacing the ATR moiety of the crystal structure with 13-*cis*R, and subsequent energy minimization. We assumed that Glu122 was deprotonated in the ground state, considering the results of the ATR-E122p-E129Δp and ATR-E122Δp-E129p simulations, as described above. In this 13-*cis*R bound simulation (13-*cis*R-E122Δp-E129p ; Table 2.1), conformational changes in the TM domains were observed. The 13-methyl group of 13-*cis*R shifted toward the cytoplasmic side and pushed out the indole ring of Trp262 (Trp223 in ChR2) on TM6 (Fig. 2.9). The local steric conflict of the 13-methyl group with Trp262 caused the subsequent movement of the cytoplasmic half of TM6 (Fig. 2.11a-c). This movement was consistent with previous studies of other microbial rhodopsins, such

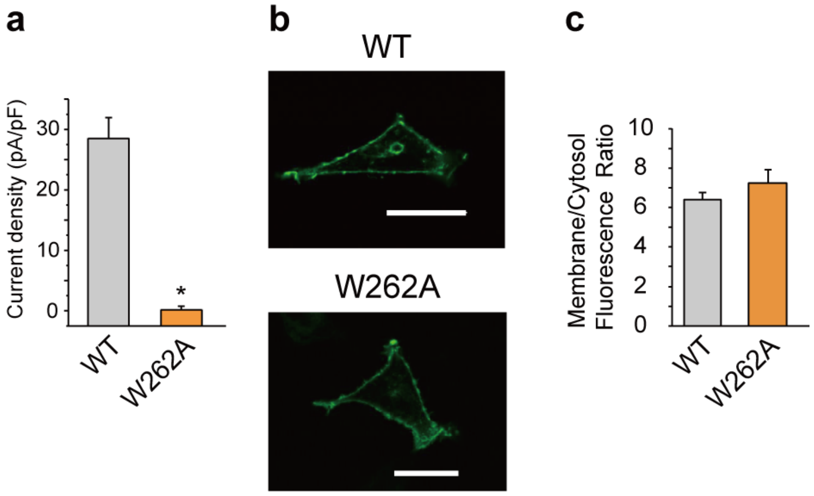


Figure 2.10: Electrophysiological analysis for W262A mutant. (a) The peak amplitudes of the photocurrents, normalized by the cell's input capacitance. (b) Conforcal images of representative HEK293 cells expressing the C1C2 WT and W262A mutants. Scale bar represents 30 μm. (c) The expression level of W262A mutant measured by the membrane/-cytosol ratio of GFP fluorescence. The error bars represent s.e.m. of 3 experiments (n = 5-17 cells). * $p < 0.05$.

This analysis was performed in collaboration with Assoc. prof. Andrés D. Maturana (Nagoya University) and Dr. Jumpei Ito (King's College London).

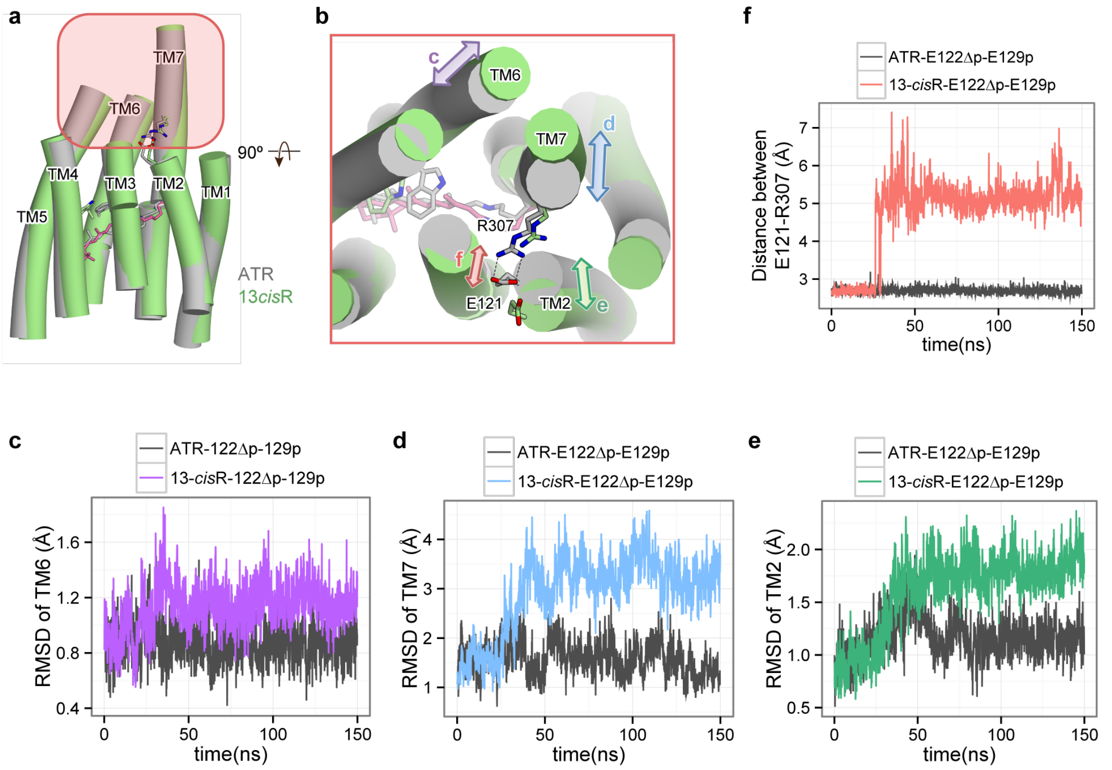


Figure 2.11: The movements of TM helices upon retinal isomerization. (a) Structural comparison between the snapshots from the ATR-bound (grey) and 13-*cis*R-bound (green) simulations. (b) Magnified cytoplasmic view of the red-highlighted region in the left panel. (c-e) The RMSD values of (c) TM6, (d) TM7 and (e) TM2, compared between the ATR-bound and 13-cisR-bound forms. (F) Distance between Glu121-Arg07 in the intracellular constriction.

as bacteriorhodopsin (BR) and sensory rhodopsin II (SRII), which indicated that the retinal isomerization and the steric collision between the 13-methyl group of 13-*cis*R and the tryptophan residue on TM6 (Trp182 in BR, Trp171 in SRII and Trp262 in C1C2) cause the movements of TM6 and TM7 [38–40]. To verify the functional importance of Trp262, the photocurrents of the W262A mutant of C1C2 were measured in HEK293 cells, revealing that this mutant completely abolished the photocurrent, despite its robust membrane expression (Fig. 2.10). These results suggested that the presence of a bulky side chain adjacent to the 13-methyl group of 13-*cis*R is important for triggering the channel opening.

In addition to the movement of TM6, the outward movement of the cytoplasmic half of TM7 was observed in the 13-*cis*R-E122Δp-E129p simulation (Fig. 2.11b, d). To visualize the correlated motions of TM6 and TM7, the correlation coefficient between each C_α atom in the TM domains was calculated. The correlation matrix showed that the cytoplasmic

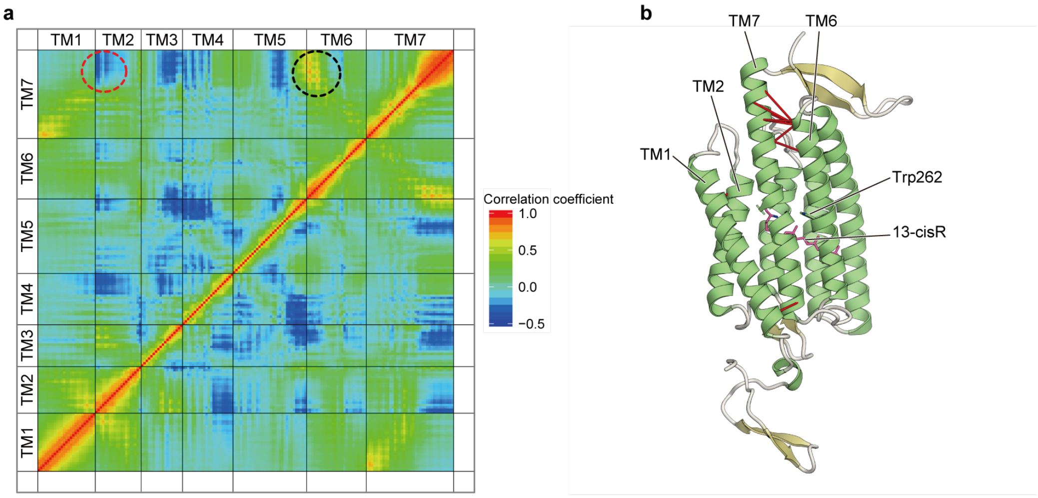


Figure 2.12: Correlation analysis for the 13-*cis*R-E122Δp-E129p simulation. (a) The matrix of correlation coefficients for the pairs of C_α atoms. (b) Mapping of the correlation coefficients to the structure. The black dashed circle represents the pair of C_α atoms with a correlation coefficient greater than 0.7. The red dashed circle represents the pair of C_α atoms in TM2 and TM7 that has a negative correlation coefficient.

halves of TM6 and TM7 have a strong correlation, with a coefficient greater than 0.7 (Fig. 2.12; black dashed circle in panel a). Moreover, this correlation matrix illustrates that the cytoplasmic half of TM2 has a negative correlation coefficient with TM7 (Fig. 2.12a; red dashed circle), indicating that the cytoplasmic halves of TM2 and TM7 move in opposite directions (Fig. 2.12b, e). These movements of TM2 and TM7 facilitate the disruption of the salt bridge between Glu121 and Arg307 in the intracellular constriction (Fig. 2.11f). Taken together, the results of the 13-*cis*R-E122Δp-E129p simulation suggested that retinal isomerization induces the series of conformational changes of Trp262, followed by those of TM6, TM7, and TM2, toward channel opening. This is consistent with previous results obtained by cryo-electron microscopy and double electron-electron resonance (DEER) spectroscopy, which suggested the movements of TM2, 6 and 7 in ChR2 during the photocycle, based on projection difference maps [49–51].

2.4 Discussion

2.4.1 Decrease of photocurrent in E122Q mutant

The electrophysiological analysis in Fig. 2.4 revealed the decreased photocurrent in the E122Q mutant (Fig. 2.4a). Given the structural similarity between glutamine and protonated glutamic acid, a transient interaction between Gln122 and His173 may occur in this mutant, as observed in the ATR-E122p-E129p and ATR-E122p-E129Δp simulations (Fig. 2.5d and Fig. 2.7b). In these Glu122 protonated simulations, the interaction between Glu122 and His173 restricted the entrance of water molecules through the intracellular constriction (Fig. 2.6a; Fig. 2.8a). Therefore, in the E122Q mutant, a similar interaction between Gln122 and His173 might disturb the water and ion permeation at the intracellular constriction (Fig. 2.6a), resulting in the decreased photocurrent of the E122Q mutant.

2.4.2 Deprotonation of Glu129

The electrophysiological analysis in Fig. 2.4 also revealed the decreased photocurrent and increased τ_{off} parameter of the E129Q mutant (Fig. 2.4a, i), suggesting the importance of Glu129 deprotonation during the photocycle. Consistent with the proposal in this study, other studies showed that Glu129 is deprotonated and negatively charged in the conducting state, which is important for the ion selectivity [24, 52–54]. However, the timing of the Glu129 deprotonation in the photocycle still remains controversial. Notably, some studies showed that Glu129 (Glu90 in ChR2) is deprotonated in the early stage of the photocycle of ChR2 [24,52], while others indicated that it is deprotonated in the desensitized stage of the photocycle in both ChR2 and C1C2 [53,54]. The results of the current electrophysiological analysis of the E129Q mutant are consistent with the latter results, since the mutation affected τ_{off}, rather than τ_{on} (Fig. 2.4h, i). It should be noted here that a recent study suggested that the timing of the Glu129 (Glu90 in ChR2) deprotonation is different between the ChR2 and ChR1/ChR2 chimeras, including C1C2 [54].

In addition, a previous study demonstrated that the replacement of Glu90 in ChR2 with lysine or arginine converted ChR2 into a chloride-selective channel [55]. Another report also revealed that the replacement of several residues in C1C2, including Glu129, converted it into an anion-selective channel [56]. Moreover, a recent crystal structure of ChR2 in the closed state revealed that Glu90 (Glu129 in C1C2) does not formed the hydrogen bond with Asn258 (Asn297 in C1C2) [57] (PDB ID: 6EID, 6EIG). In these ChR proteins, different hydrogen bond and/or salt bridge pairs from Glu129 and Asn297 might prevent ion leakage in the ground state.

Further spectroscopic and crystallographic analyses for the protonation state of Glu129 is required.

2.4.3 Photoactivation mechanism

The 13-*cis*R-E122Δp-E129p simulation results suggested that the isomerization of retinal induces the conformational change of Trp262, which is followed by the subsequent movements of TM6, TM7 and TM2 (Figs. 2.9, 2.11). Given that the P_1^{500} state in the photocycle is observed ~50 ns after the retinal isomerization [19, 21] and that large conformational changes of the protein backbone were observed in this P_1^{500} state by fourier transform infrared (FTIR) spectroscopy [19, 58], the structural changes, including the movements of TM6, TM7 and TM2, observed in the 13-*cis*R-E122Δp-E129p simulation are likely to represent the structural transition from the ground state to the P_1^{500} state. Moreover, this P_1^{500} state is considered to be a non-conducting state [18–23], which is consistent with the observations in this simulation: the channel pore radii on both the cytoplasmic and extracellular sides were still comparable to those of the ATR-E122Δp-E129p simulation, and thus too narrow to permeate monovalent or divalent cations.

The events which occur after retinal isomerization (*i.e.*, P_1^{500}) involve the proton transfer between the Schiff base of the retinal and the surrounding protein side chains, which finally results in the ion-conducting state. An investigation of the subsequent events is beyond the scope of the present study, since the time scales of those events are several tens of microseconds to milli-seconds [19, 21, 58], and thus exceed the limitations of the conventional MD simulation.

Nevertheless, the present study provides mechanistic insights into the nature of the two constrictions and the early response of the protein conformations to the isomerization of the retinal chromophore, and will serve as a framework for future investigations of the formation of the ion-conducting state.

3 Molecular Dynamics Simulation of Triose-phosphate/phosphate Translocator

3.1 Introduction

3.1.1 Plastidic phosphate tranlocator (pPT) family and triose-phosphate/phosphate translocator (TPT)

The plastidic phosphate tranlocator (pPT) family is one of the membrane transporter groups that exchanges phosphorylated C3, C5 and C6 carbon sugars with inorganic phosphate (P_i) on the inner membrane of the plastid in plant and alga cells [59,60]. The pPT family is further divided into four subfamilies, triose-phosphate/phosphate translocator (TPT), glucose 6-phosphate/phosphate translocator (GPT), xylulose 5-phosphate/phosphate translocator (XPT) and phospho*enol*pyruvate/phosphate translocator (PPT), according to their substrate specificities for triose-phosphate, glucose-6-phosphate, xylulose-5-phosphate and phospho*enol*pyruvate, respectively [59, 61–63]. The TPT family in green plants exports triose phosphates (TP) and 3-phosphoglyceric acids (3-PGA) produced by photosynthesis and imports P_i into stroma [59, 64, 65] (Fig. 3.1). Thus, TPT contributes to efficient carbon fixation and plant growth [66–71]. Each pPT subfamily catalyzes the strict 1:1 counter exchange of phosphorylated carbon compounds with P_i to ensure phosphate homeostasis between the stroma and the cytosol [72, 73]. In addition, they can also catalyze (nonpro-

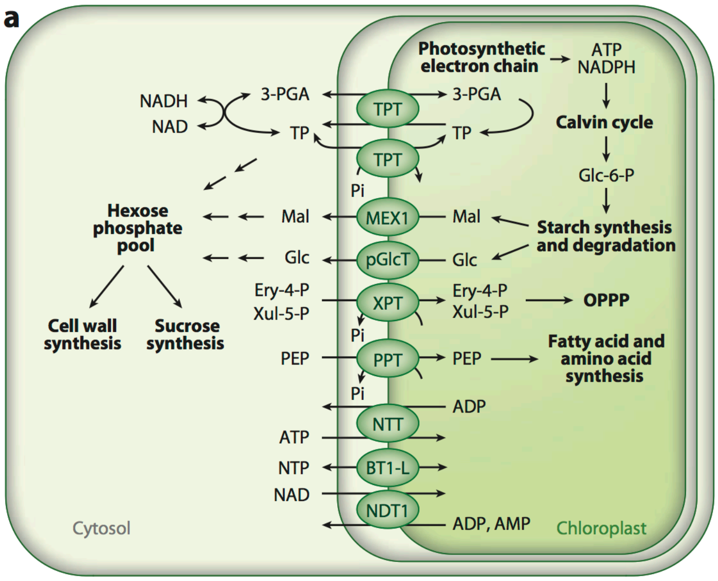


Figure 3.1: Overview of the metabolite transport pathways in chloroplast. Figure cited from [59] Fig.1 panel a.

ductive) P_i/P_i homo-exchange in both chloroplast and reconstituted system [74, 75].

3.1.2 Crystal structure of *Galdieria sulphuraria* GPT (GsGPT)

Recently, our group reported the crystal structure of *Galdieria sulphuraria* GPT (GsGPT) from the thermophilic red alga *Galdieria sulphuraria* [76], which is functionally similar to TPT, rather than GPT, despite its name based on the sequence similarity [75].

Although the size exclusion chromatography-multi angle light scattering (SEC-MALS) analysis shows that the purified GsGPT is monomeric in solution [76], the crystal structure of GsGPT is sovled in a dimer form (Fig. 3.2a). Each protomer is composed of 10 α-helical TM helices (Fig. 3.2b), and adopts the drug metabolite transporter (DMT) superfamily [77] fold, similar to that of the recently reported bacterial metabolite transporter, *Starkeya novella* YddG (SnYddG) [78], in which the N-terminal half (TM1-5) is related to the C-terminal half (TM6-10) by two-fold pseudo-symmetry.

Our research group obtained the two crystal structures bound with different substrates, P_i and 3-PGA, and the overall conformation is quite simillar (Fig. 3.2c). In both crystal structures, the phosphate moieties of the substrates are recognized by the well-conserved basic residues, Lys204, Lys362, and Arg363 (Fig. 3.2b, c:right panel). This central sub-

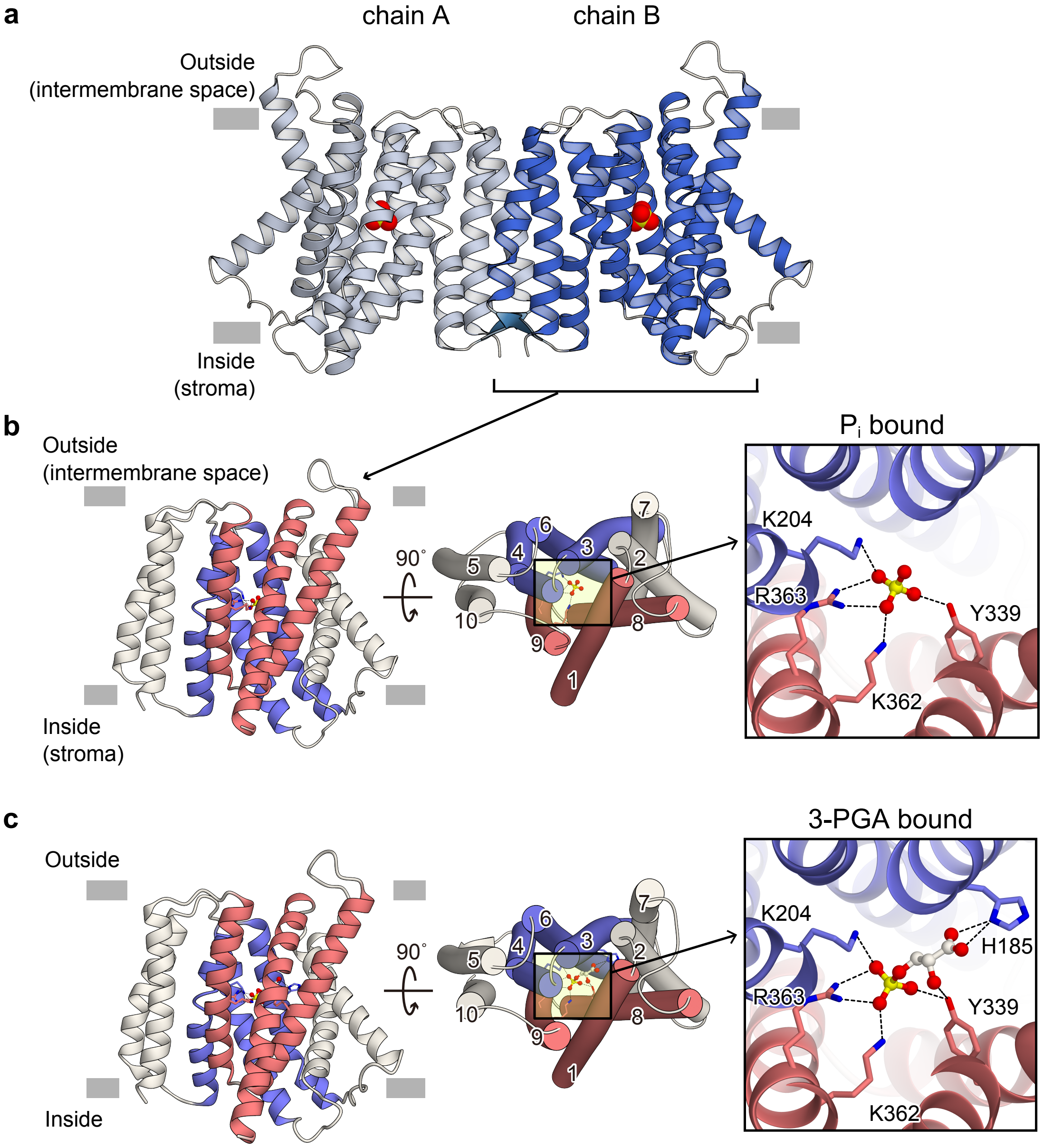


Figure 3.2: The crystal structure of *Galdieria sulphuraria* GPT (GsGPT). (a) The crystal structure of GsGPT in dimeric form. (b, c) The monomeric structure of GsGPT in the (b) P_i-bound state (PDB ID: 5Y78) and (c) 3-PGA-bound state(PDB ID: 5Y79), viewed from the plane of the membrane (left panel) and from the intermembrane space (center panel). Right panel shows the close-up view of the central binding site in the P_i-bound state. Key residues involved in substrate binding are shown in stick models. Dotted lines represent polar interactions.

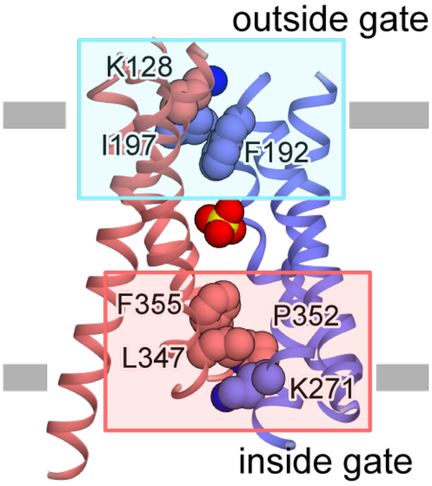


Figure 3.3: Inside and outside gates of GsGPT. The inside and outside gates are highlighted within magenta and cyan rectangles, respectively. The substrate and gate-forming residues are shown with space-filling models. TM2, 5, 7 and 10 are not shown.

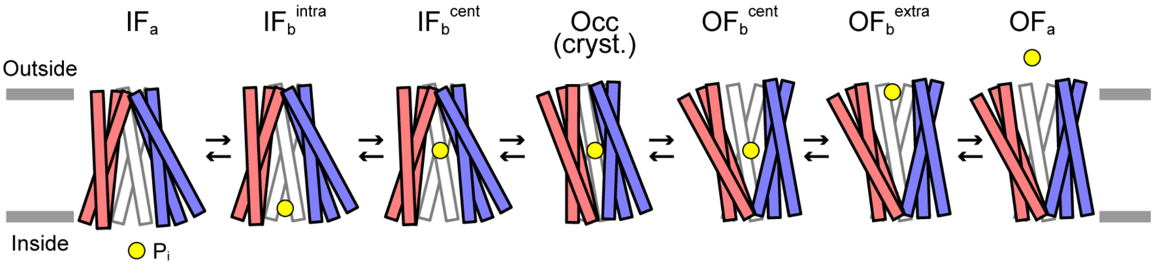


Figure 3.4: Schematic representation of states, IF$_a$ (IF-*apo*), IF$_b^{intra}$ (IF-bound at intracellular gate), IF$_b^{cent}$ (IF-bound at central binding site), Occ, OF$_b^{cent}$ (OF-bound at central binding site), OF$_b^{extra}$ (OF-bound at extracellular gate) and OF$_a$ (OF-*apo*), involved in the currently studied transport cycle. P$_i$ is represented by the yellow circle. The TM segments are colored in the same manner as in panel a. TM2 and 7 are not shown.

strate binding site is occluded from the solvent by hydrophobic residues, referred to as the inside and outside gates (Fig. 3.3). These observations indicated that the crystal structure of GsGPT adopts a substrate-bound Occ state. The structural comparison with SnYddG in the OF state revealed the basis of the transport mechanism, in which the opening and closing of these gates regulate the IF and OF conformational changes [76]. However, the detailed mechanisms of the conformational changes and the substrate binding/releasing by the DMT superfamily have remained elusive.

3.1.3 Overview of this study

In this study, the author investigated the dynamics of the GsGPT using the MD simulations.

First, the dynamics of GsGPT in the dimeric form were examined in the explicit solvent and lipid bilayer. The author revealed that the removal of the P_i from the crystal structure led to the spontaneous conformational transition to IF or OF states, while the substrate bound form is stable in Occ state within 100 ns. A steered molecular dynamics (SMD) simulation [79] revealed that the conformational change of each GsGPT protomer occurs independently. The author concluded that one protomer is sufficient for the transport activity.

Next, by combining the biased sampling techniques described in section 1.2.2, the author achieved the *in silico* reconstruction of a feasible transport cycle for GsGPT monomer (Fig. 3.4), and observed the substrate binding and releasing events for the first time in the DMT superfamily transporters. Comprehensive structural analyses revealed the strict 1:1 exchange mechanism of GsGPT in atomic detail. An unexpected common feature of the conformational transition mechanisms with the well-studied major facilitator superfamily (MFS) transporters [80] was found, despite their completely different protein folding. Along with the computational simulations, the functional analysis of the mutants provided experimental evidence supporting the MD simulation observations. Taken together, the author revealed the atomistic molecular mechanism of GsGPT, which involves the structural coupling between the local conformational change induced by substrate binding and the global conformational transition of the transporter.

3.2 Materials and methods

3.2.1 Simulation system setup for GsGPT dimer

The simulation system included GsGPT dimer in P_i bound form (PDB ID: 5Y78), POPC, TIP3P water and 150 mM NaCl. The disordered side chain in the TPT protein was modelled by COOT [81]. To embed the protein to POPC bilayer, the protocol described by Javanainen was used [82]. In short, the bilayer with 446 POPC molecules was prepared with VMD [34]. The POPC membrane patch and the GsGPT dimer were placed in the same simulation box, and placed next to the lipid bilayer. The restraints for the all protein

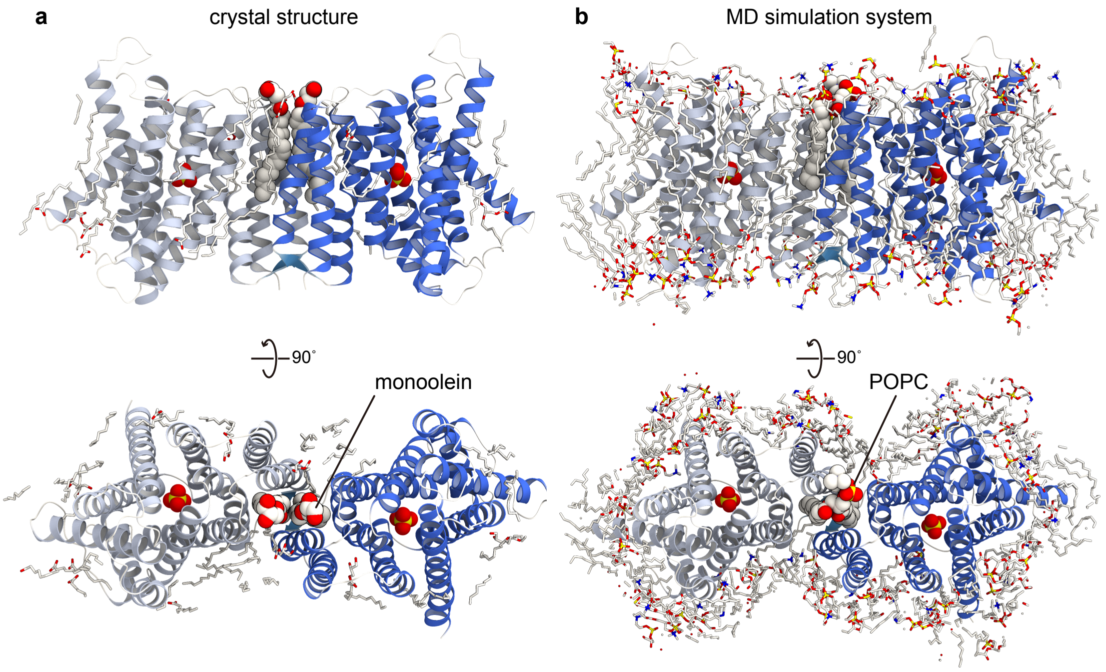


Figure 3.5: Lipid molecules in (a) crystal structure and (b) MD simulation system. (a) All monoolein molecules visible in crystal structure and (b) POPC molecules around 8 Å of protein are shown in stick model. The lipid monolecules existing in the protein dimerization interface and P_i are shown in CPK model.

atoms with the force constant of 10,000 kJ mol^{-1} nm^{-2} were included. The Z-direction restraints for the nitrogen atoms of the lipid headgroup and the terminal groups of the lipid tails with the force constant of 100 kJ mol^{-1} nm^{-2} are also included. The system was minimized, and then simulated for 1 ns with a large lateral pressure using the semi-isotropic Berendsen barostat [83] with 1,000 bar and 1 bar for X-Y direction and Z-direction, respectively. Then, the system was relaxed for 10 ns under the 1 bar pressure.

After embedding into the membrane, one POPC molecule was placed in the GsGPT dimerization interface corresponding to the two monoolein molecules in the crystal structure (Fig. 3.5). Finally, the periodic boundary system, including 136,668 (P_i-bound) and 136,652 (without P_i) atoms, with the size of 90.7 × 147.9 × 100.0 Å was prepared. The net charge of the solute was neutralized with sodium and chloride ions. The molecular topologies and force field parameters from CHARMM36 [84] was used.

3.2.2 Production run of GsGPT dimer simulation

MD simulations were performed by the program Gromacs 5.0.5 [85]. First, energy minimization was performed using steepest descent with cut-off of 1000.0 kJ mol^{-1} nm^{-1}. Next, the random velocity is assigned according to a Maxwell distribution at temperature of 310 K to each atom, and performed a equilibration run (eq1) for 100 ps in the *NVT* (310 K, $90.7 \times 147.9 \times 100.0$ Å volume). Finally, a equilibration run (eq2) was performed for 1000 ps in the *NPT* (310 K, 1 bar). The position of non-hydrogen atoms in the protein and phosphate was restrained with force constant of 1000 kJ mol^{-1} nm^{-2} in the minimization and the equilibration runs. Production runs were performed for 100 ns in the *NPT* ensemble (310 K, 1 bar). The same simulations were performed twice with different initial velocities.

Constant temperature was maintained by using V-rescaling [86] with a time constant of 0.1 ps in eq1, while Nosé-Hoover thermostat [87, 88] with a time constant of 0.5 ps is used in eq2 and the production runs. Pressure was controlled with semiisotropic coupling to Parrinello-Rahman [89] barostat with a time constant of 5.0 ps and with a compressibility of $4.5\mathrm{x}10^{-5}$ bar^{-1}. LINCS algorithm [90] was used to constrain all hydrogen-containing bonds to enable a 2 fs timestep. Long range electrostatic interactions were calculated with the PME method [44].

3.2.3 Steered molecular dynamics (SMD) for GsGPT dimer

Steered molecular dynamics (SMD) [79] simulation was performed with the center of mass (COM) pulling option of Gromacs software. The results of production run suggested that the conformational change from Occ to OF and IF can be induced by the two distances, $\mathrm{D_{out}}$ and $\mathrm{D_{in}}$: where $\mathrm{D_{out}}$ was defined as the distance between the C_α mass centers of the two residue groups, 126-130 and 191-200, and $\mathrm{D_{in}}$ was defined as the distance between the C_α mass centers of the two residue groups, 269-273 and 347-355 (Fig. 3.6; center, right panel).

COM pulling is performed by the moving center harmonic restraint with a constant velocity of 0.00005 nm/ps (= 0.5 Å/ns), and with a force constant of 5,000 kJ mol^{-1} nm^{-2}

(= 11.96 kcal mol^{-1} Å^{-2}). The pulling force is applied for 20 ns (=10 Å moving of D_{out} or D_{in}), followed by the 80 ns relaxation simulation without the pulling force. The same pulling simulations were performed D_{in} and D_{out} for each protomer. In total, 4 patterns of pulling simulation were performed. The same simulation sets were performed twice with different initial velocities, and obtained similar results.

3.2.4 Simulation system setup for GsGPT monomer

All MD simulations with GsGPT monomer were performed with NAMD 2.9-2.11 [42], along with its Colvars Module [91]. The simulation system was 96 × 96 × 100 Å^3, and contained GsGPT monomer (PDB ID: 5Y78; chain B), a molecule of P_i, a POPC bilayer, 150 mM NaCl, and TIP3P water molecules. All of the water molecules that were visible in the crystal structure were kept. The missing hydrogen atoms were added by the psfgen plugin of VMD [34]. The net charge of the system was neutralized by adding 150 mM NaCl. The molecular topology and force field parameters from Charmm36 were used [84, 92]. The systems were first energy minimized for 1,000 steps with fixed positions of the non-hydrogen atoms, and then for another 1,000 steps with 10 kcal mol^{-1} Å^{-2} restraints for the non-hydrogen atoms, except for the lipid molecules within 5.0 Å from the protein. Next, equilibrations were performed for 0.1 ns with *NVT*, with 10 kcal mol^{-1} Å^{-2} restraints for the heavy atoms of the protein. Finally, equilibration was performed for 1.0 ns with *NPT* with the 1.0 kcal mol^{-1} Å^{-2} restraints for all C_α atoms of the protein, followed by a 100 ns equilibration without the restraints. The equilibrated system was further used for the next simulation, as indicated in Table 3.1 and below. For all simulations in Table 3.1, the SHAKE algorithm was used to constrain all hydrogen-containing bonds to enable a 2 fs timestep. Constant temperature was maintained with Langevin dynamics at 310 K, and constant pressure was maintained with a Nosé-Hoover Langevin piston [43] at 1 bar. Electrostatic interactions were calculated by the PME method [44].

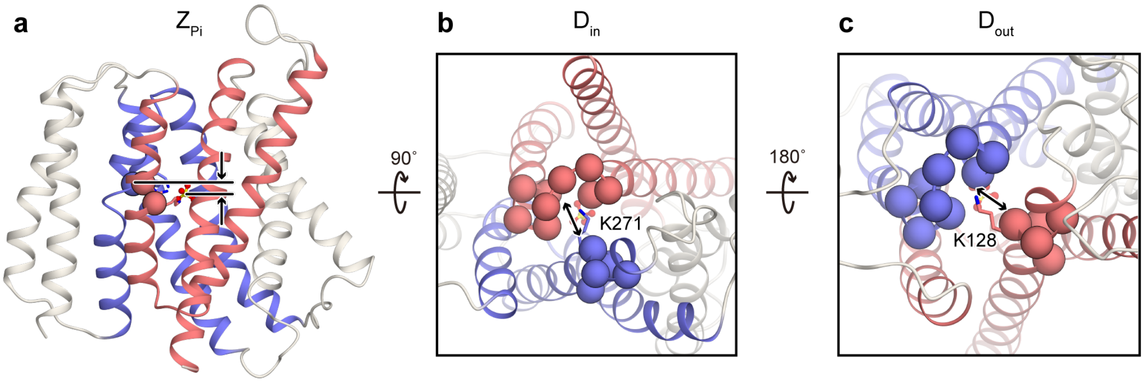


Figure 3.6: The definitions of collective variables (CVs). (a) Definition of Z_{Pi} . P_i is shown with a ball and stick model, and the C_α atoms of Lys204, Lys362 and Arg363 used for the reference points are shown with spheres. (b, c) Definitions of D_{in} and D_{out} . The C_α atoms used for the calculation are shown as spheres.

3.2.5 Collective variables (CVs)

In order to induce the conformational change from the Occ state crystal structure, the CVs of ΔD and Z_{Pi} were used. ΔD was defined as follows: $\Delta D = D_{\mathrm{out}} - D_{\mathrm{in}}$, where D_{out} and D_{in} is defined in section 3.2.3 (Fig. 3.6). Z_{Pi} was used to specify the position of the substrate P_i, and defined as the projection of the distance vector between the phosphorus atom of P_i and the reference point onto the membrane normal axis (Fig. 3.6; left panel):

$$Z_{\mathrm{Pi}} = (\boldsymbol{r}_{\mathrm{P}} - \boldsymbol{r}_{ref}) \cdot \boldsymbol{e}_z \tag{3.1}$$

where $\boldsymbol{r}_{\mathrm{P}}$ is the position of the phosphorus atom of P_i, $\boldsymbol{r}_{ref}$ is the reference point, and $\boldsymbol{e}_z$ is the cell basis vector along z-axis. The reference point of Z_{Pi} was defined as the center of mass of the C_α atoms of the P_i binding residues in the crystal structure, *i.e.*, K204, K362 and R363 (Fig. 3.6; left panel). In all of the simulations, P_i was also restrained within a cylinder with a 10 Å radius, which was aligned to the membrane normal axis and centered to the mass center of the protein, using a half-harmonic potential with a spring constant of 10 kcal mol^{-1} Å^{-2}.

To sample the configurations for the entire transport cycle of GsGPT, we used the following strategy. Generally, it is difficult to sample sufficiently in an n-dimensional CV space, $(z_1, z_2, \dots, z_n) = \boldsymbol{z}$, due to the computational cost. A practical solution is to perform

Table 3.1: Simulation set list performed in this study

Sim No.	Transition	Method	Collective variable	# of replicas × sim. time
1	cryst. $\rightarrow$ $\mathrm{IF_b}^{\mathrm{cent}}$	SMD	ΔD	10 ns
2	$\mathrm{IF_b}^{\mathrm{cent}}$ $\rightarrow$ Occ $\rightarrow$ $\mathrm{OF_b}^{\mathrm{cent}}$	SMD	ΔD	20 ns
3	$\mathrm{IF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{IF_b}^{\mathrm{intra}}$	SMD	$\mathrm{Z_{Pi}}$	10 ns
4	$\mathrm{OF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{OF_b}^{\mathrm{extra}}$	SMD	$\mathrm{Z_{Pi}}$	10 ns
5	$\mathrm{IF_b}^{\mathrm{intra}}$ $\leftrightarrow$ Occ $\leftrightarrow$ $\mathrm{OF_b}^{\mathrm{extra}}$	SMwST	ΔD / $\mathrm{Z_{Pi}}$	40 × 1.2 ns = 48 ns
6	$\mathrm{IF_b}^{\mathrm{intra}}$ $\leftrightarrow$ Occ $\leftrightarrow$ $\mathrm{OF_b}^{\mathrm{extra}}$	BEUS	pathCV(s)	20 × 20 ns = 400 ns
7	$\mathrm{IF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{IF_a}$	SMD	$\mathrm{Z_{Pi}}$	25 ns
8	$\mathrm{IF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{IF_a}$	BEUS	$\mathrm{Z_{Pi}}$	16 × 20 ns = 320 ns
9	$\mathrm{OF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{OF_a}$	SMD	$\mathrm{Z_{Pi}}$	25 ns
10	$\mathrm{OF_b}^{\mathrm{cent}}$ $\rightarrow$ $\mathrm{OF_a}$	BEUS	$\mathrm{Z_{Pi}}$	16 × 20 ns = 320 ns
11	$\mathrm{IF_a}$ $\leftrightarrow$ Occ $\leftrightarrow$ $\mathrm{OF_a}$	SMwST	ΔD / $\mathrm{Z_{Pi}}$	40 × 5.5 ns = 220 ns
12	$\mathrm{IF_a}$ $\leftrightarrow$ Occ $\leftrightarrow$ $\mathrm{OF_a}$	BEUS	pathCV(s)	40 × 50 ns = 2000 ns
			total	~ 3.4 μs

the sampling only around a particular pathway (or string), $s(z)$. In this study, the minimum free energy path of the transport pathway is explored using the string method with swarms of trajectories (SMwST) [93] in a two-dimensional CV space, (ΔD , $\mathrm{Z_{Pi}}$). To achieve the conformational sampling along the refined transition path by the bias-exchange umbrella sampling (BEUS) [10, 11, 94] (simulation sets 6 and 12 in Table 3.1), the so-called path collective variable (pathCV) s and ζ [95] were used:

$$s(z) = \frac{1}{M-1}\frac{\sum_{i=1}^{M}(i-1)e^{-\lambda(\mathbf{z}-\mathbf{z}_i)^2}}{\sum_{i=1}^{M}e^{-\lambda(\mathbf{z}-\mathbf{z}_i)^2}} \tag{3.2}$$

$$\zeta(z) = -\frac{1}{\lambda}\ln\left(\sum_{i=1}^{M}e^{-\lambda(\mathbf{z}-\mathbf{z}_i)^2}\right) \tag{3.3}$$

where λ is the inverse square of the average RMSD between images, M is the number of images, z is a vector for the current values of the CV, and z_i is the CV vector for the i-th image. In this paper, z was represented as a two-dimensional vector (ΔD , $\mathrm{Z_{Pi}}$).

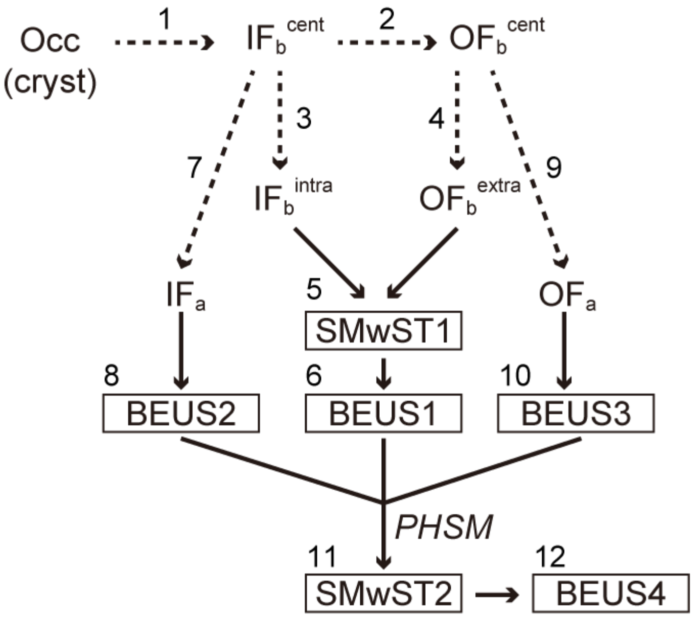


Figure 3.7: Graphical representation of the iterative scheme used for designing the simulations shown in Table3.1. The number corresponds to the simulation numbering in Table3.1.

3.2.6 Sampling protocol to reconstruct the transport cycle of GsGPT

To reconstruct the entire transport cycle of GsGPT, the author employed an iterative sampling approach. The details of all simulation sets are shown in Table 3.1. The relations of each simulation sets are shown in Fig. 3.7.

At first, to generate the IF_b^{cent} and OF_b^{cent} states from the Occ state crystal structure, SMD simulations using the CV ΔD were performed (simulation sets 1 and 2). These simulations were performed by moving the center harmonic restraint with a force constant of 10 kcal mol^{-1} $Å^{-2}$ and using the ΔD ranges from 0 to -10 Å and 0 to 10 Å, to generate IF_b^{cent} and OF_b^{cent} , respectively. The final structures from simulation sets 1 and 2 were used as the initial structures of the simulation sets 3 and 4, respectively, in which the conformational transitions to IF_b^{intra} and OF_b^{extra} were induced using the CV Z_{Pi} within the ranges from 0 to -15 Å and 0 to 15 Å, respectively.

As the $IF_b^{intra} \leftrightarrow OF_b^{extra}$ transition was expected to undergo complicated conformational changes, including large scale substrate translocation and a significant local conformational change of the protein, we first refined the $IF_b^{intra} \leftrightarrow OF_b^{extra}$ transition with string method with swarms of trajectories (SMwST), prior to reconstructing the entire conformational transition, $IF_a \leftrightarrow Occ \leftrightarrow OF_a$. The final frames of these simulation sets were used to generate the initial string of the $IF_b^{intra} \leftrightarrow OF_b^{extra}$ transition. The initial $IF_b^{intra} \leftrightarrow OF_b^{extra}$ transition string was defined with 40 images: 10 frames from simulation set 3 (Z_{Pi} = 0, -1.5, -3.0, . . . , -15 Å), 10 frames from simulation set 4 (Z_{Pi} = 0, 1.5, 3.0, . . . , 15 Å) and 20 frames from simulation set 2 (ΔD = -10, -9, ⋯, 10 Å). This initial string was refined with SMwST in the two-dimensional (ΔD , Z_{Pi}) CV space (simulation set 5). The selected images were minimized with 10,000 steps, followed by 1 ns of equilibration in the NPT ensemble with ΔD and Z_{Pi} restraints to the initial values. Twenty swarms of 0.1 ps unrestrained simulation were followed by 20 ps of restrained simulation, and the force constants for both ΔD and Z_{Pi} were 1.0 kcal mol^{-1} Å^{-2}. The string was converged after 234 iterations. The last 20 strings were averaged and used as the $IF_b^{intra} \leftrightarrow OF_b^{extra}$ transition path in simulation set 6. To perform the free energy calculation with BEUS along the refined transition string, the path collective variables (pathCVs) were used (defined by equation 3.2 and 3.3) with the λ value of 1.0. The 20 windows for the collective variable s = $0.025, 0.075, \ldots, 0.975$ (increased by 0.05) were used for BEUS with a harmonic force constant of 200 kcal mol^{-1} Å^{-2}. The collective variable path collective variable ζ (pathCV(ζ)) was restrained to 0.0 with a harmonic force constant of 200 kcal mol^{-1} Å^{-2}. The replica exchange was attempted between neighboring replicas with an interval of 2.0 ps for each unique pair. The atomic coordinates were sampled every 2.0 ps.

Next, to induce the conformational transition to the apo structure, the final frames of simulation sets 1 and 2 were further employed in the SMD simulation, using the CV Z_{Pi} (simulation sets 7 and 9) with the ranges from 0 to -45 Å and from 0 to 45 Å, respectively. BEUS with 16 windows along Z_{Pi} for the transition from IF_b^{cent} to IF_a was performed with window centers (and force constants) of -2.0, -4.5, -6.5, -8.0, -9.0, -10.5, -12.0, -14.0, -15.5, -18.0, -20.5, -22.0, -24.0, -28.0, -33.0 and -40.0 Å (0.2, 0.5, 1.0, 1.0, 1.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.1, 0.1, 0.05 and 0.01 kcal mol^{-1} Å^{-2}) (simulation set 8). BEUS with 16 windows along Z_{Pi} for the transition from OF_b^{cent} to OF_a was also performed with window centers and force constants of 0, 2.5, 4.0, 6.0, 8.5, 10.5, 12.5, 14.0, 16.0, 18.0, 20.0, 22.0, 25.0, 29.0, 33.0 and 40Å (0.5, 1.0, 1.0, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.1, 0.1, 0.05 and 0.05 kcal mol^{-1} Å^{-2}) (simulation set 10).

The structures and free energy information obtained from simulation sets 6, 8 and 10 were utilized for the PHSM [94] to generate the initial string for the transition $IF_a \leftrightarrow OF_a$. The *post-hoc* string method (PHSM) was performed as described previously [94]. First, the initial string was generated by a non-parametric version of the lowest free energy pathway (LFEP) [94, 96]. The following 37 metrics were used for the non-parametric LFEP and PHSM algorithm: (i) Z_{Pi} , (ii) first 8 C_α -based principal components, (iii) D_{in} and D_{out} , (iv) four atomic distances of K128-F192O, K128-V195O, K271-L347O and K271-V350O, (v) minimum distances (D_{min}) between P_i and pore-lining polar residues: N124, K128, H201, K204, R266, K271, Y339, K362 and R363, (vi) D_{min} values of internal salt bridges E207-K204, E207-K362 and E207-R363, and (vii) χ_1 angles for all 10 residues listed in (v) and (vi). A distance was defined in this 37-dimensional CV space of (i) to (vii), using a diagonal metric matrix with the values 1/Å^2, 0.04/Å^2, 1/Å^2, 1/Å^2, 0.01/Å^2, 1/Å^2 and 0.0025/(1°)^2 for the elements in (i) to (vii), respectively. These weights were roughly chosen based on the relative variance of the CVs. The initial string generated by the non-parametric LFEP was used for the PHSM algorithm with 400 image centers. The tube thickness $\varepsilon = 0.5$ was used in the PHSM algorithm, since it gives the smoothest pathway as compared with other results based on $\varepsilon = 0.1, 0.2, \ldots, 1.0$.

The string of the $IF_a \leftrightarrow OF_a$ transition generated by the PHSM algorithm was further refined with SMwST (simulation set 11) in the two-dimensional CV space (ΔD , Z_{Pi}). The 40 images were selected from the string refined with the PHSM algorithm at even intervals. The selected images from PHSM were minimized and equilibrated in the same manner as the simulation set 5. In this analysis, 20 swarms of 5 ps unrestrained simulation were performed, followed by 10 ps of restrained simulation with force constants of 10.0 kcal mol^{-1} for each CV. The string was converged after 50 iterations. The final 10 strings were averaged, and used for the final BEUS sampling.

The final BEUS sampling along the refined $IF_a \leftrightarrow OF_a$ transition string (simulation set 12) was performed using the path collective variable s (pathCV(s)) and pathCV(ζ) with

$\lambda = 0.1$ (equations (3.2), (3.3)). The 40 images corresponding to the image points of the averaged string were selected and equilibrated for 0.1 ns with the CVs (ΔD , Z_{Pi}) restrained. In total, 40 windows, $s = 0.015, 0.035, \ldots, 0.99$ (increased by 0.025) for the pathCV(s) were used for BEUS, with a force constant of 500 kcal mol^{-1}. The pathCV(ζ) is restrained to the region pathCV(ζ) < 1.0 with half-harmonic restraints with a force constant of 100 kcal mol^{-1}. The exchange was attempted between the neighboring replicas with an interval of 10 ps for each unique pair. The sampling was performed for 50 ns, and the coordinates for all atoms were sampled every 2 ps, to obtain 1,000,000 structures (= 40 replicas x 50 ns / 2 ps) in total.

3.2.7 Data analysis

Softwares

Data analysis was performed using MDAnalysis [46], mdtraj [97], and HOLE [98], along with *in-house* python and C++ codes. All molecular graphics were illustrated with CueMol2 (`http://www.cuemol.org/`). Plot graphics were generated with ggplot2 [99], seaborn [100] and matplotlib [101]. The C++ codes for the non-parametric LFEP and PHSM analyses are provided from Dr. Mahmoud Moradi (University of Arkansas).

RMSD

The RMSD was calculated using the equation 2.1. In this chapter, the superimposition to the reference structure was performed with all C_α atoms of GsGPT if not specified.

Principal component analysis (PCA)

The PCA was performed on the all C_α atoms of GsGPT.

The coordinates of the trajectory were superimposed on the averaged structure of simulation set 12 using all C_α atoms of GsGPT. The covariance matrix of the atomic fluctuations

$\mathbf{C} = \{C_{ij}\}$ was calculated as following:

$$C_{ij} = \langle (\boldsymbol{r}_i - \langle \boldsymbol{r}_i \rangle) \cdot (\boldsymbol{r}_j - \langle \boldsymbol{r}_j \rangle) \rangle \tag{3.4}$$

where $\boldsymbol{r}_i, \boldsymbol{r}_j$ are the Cartesian coordinates of the i-th and j-th atom, and $\langle \cdot \rangle$ represents the average over the trajectory. The covariance matrix can be decomposed as

$$\mathbf{C} = \mathbf{P} \boldsymbol{\Delta} \mathbf{P}^{\top} \tag{3.5}$$

where $\boldsymbol{\Delta}$ is the diagonal matrix whose diagonal elements are the eigenvalues, and $\mathbf{P}$ is the eigenvectors which represents the principal components and is used for projection of the protein movements.

The number of hydrogen bonds

The calculation of the number of hydrogen bonds was performed by the following equation:

$$\sum_{i \in group} \frac{1 - (d_i/d_0)^n}{1 - (d_i/d_0)^m} \tag{3.6}$$

where d_i is the distance between two hydrogen-bonding atoms, d_0 is a cutoff distance, and n and m are integers that determine the behavior of the switching function. The sum is taken over all of the hydrogen bonding atom pairs. In this study, $d_0 = 3.3$ Å, $n = 6$ and $m = 8$ were used.

3.2.8 Free energy calculations

To calculate the potential of mean force (PMF) from the BEUS simulations along the pathCV(s) (simulation set 12), the C implementation of the weighted histogram analysis method (WHAM) algorithm [102] (Grossfield, Alan, “WHAM: the weighted histogram analysis method”, version 2.0.9: `http://membrane.urmc.rochester.edu/content/`

`wham`) was used. In short, the following equation was solved iteratively:

$$p_i = \frac{\sum_{k=1}^{N_{sim}} B_k}{\sum_{m=1}^{N_{sim}} N_m \exp[\beta F_m - \beta U_m(q_i)]} \tag{3.7}$$

$$F_i = -\beta^{-1} \ln \left[\sum_{j=1}^{N_{bin}} p_j \exp[-\beta U_j(q_i)] \right] \tag{3.8}$$

where p_i and F_i are the probability and free energy for the i-th bin, respectively, N_{sim} is the number of replicas, N_{bin} is the number of bins, B_i is counts in i-th bin, N_i is the number of snapshots of i-th replica, q_i is the coordinate fo i-th bin, and $U_i(q_j)$ is the bias potential,

$$U_i(q_j) = \frac{1}{2} k_i (q_j - q_i^{center})^2 \tag{3.9}$$

where q_i^{center} is the center value of the harmonic potential for i-th window.

To perform the *post-hoc* string method analysis, the weights for each structure w^t were required [94]. They were calculated by the following generalized WHAM equation [94, 103]:

$$w^t = \left(\sum_j T_j \exp\left[-\beta \left(U_j(\zeta^t) - F_j \right) \right] \right)^{-1} \tag{3.10}$$

$$\exp[-\beta F_j] = \sum_t w^t \exp\left[-\beta U_j(\zeta^t) \right] \tag{3.11}$$

where $\beta = 1/k_B T$ is the inverse temperature, T_j is the number of samples collected for image j, F_j is the PMF for the j-th umbrella window, ζ^t is the CV at time t, and $U_j(x)$ is the j-th umbrella potential as defined by equation 3.9.

The projection of the PMF from a CV space to another one, $\boldsymbol{\xi}$, was performed with the following equation [94]:

$$G(\boldsymbol{\xi}) = -\frac{1}{\beta} \ln \left[\sum_t w^t \mathrm{K} \left(\boldsymbol{\xi}(x^t), \boldsymbol{\xi} \right) \right] \tag{3.12}$$

where K(., .) is a kernel function. In this study, the radial basis function (RBF) kernel was used as defined following:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left[-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{2\sigma}\right] \tag{3.13}$$

In this study, $\sigma = 1.0$ was used.

3.2.9 Liposeme transport assay

The transport assays using proteo-liposome are performed in collaboration with Mr. Yongchan (The University of Tokyo).

The liposome assays were performed as previously described [76]. In brief, the liposomes reconstituted with yeast membranes expressing WT or mutant GsGPT were preloaded with 30 mM NaH_2PO_4, and were mixed with an equal volume of extra-liposomal solution containing 1 mM $[^{32}P]-NaH_2PO_4$ (0.1 mCi ml^{-1}). The reaction was terminated by passing the liposomes through AG-1 X8 resin. The total amounts of incorporated P_i were measured at 30 min. The control experiment was performed with membranes from yeast cells harboring the empty vector. The mean count value of the control experiment was subtracted from each count value, and the values shown in the figures were normalized with the mean value of the WT. The expression of each mutant was verified by a western-blot analysis, using an anti-His-tag polyclonal antibody (code PM032; MBL).

3.3 Results

3.3.1 The removal of P_i leads to spontaneous gate opening

First of all, the author performed the MD simulation of the dimeric GsGPT in explicit water and POPC membrane.

In the simulation with initial structure of P_i bound crystal structure, significant conformational change was not observed, and GsGPT is stable over 100 ns (Fig. 3.8a, b) In contrast, in the simulation with initial structure in which the P_i molecule was removed from

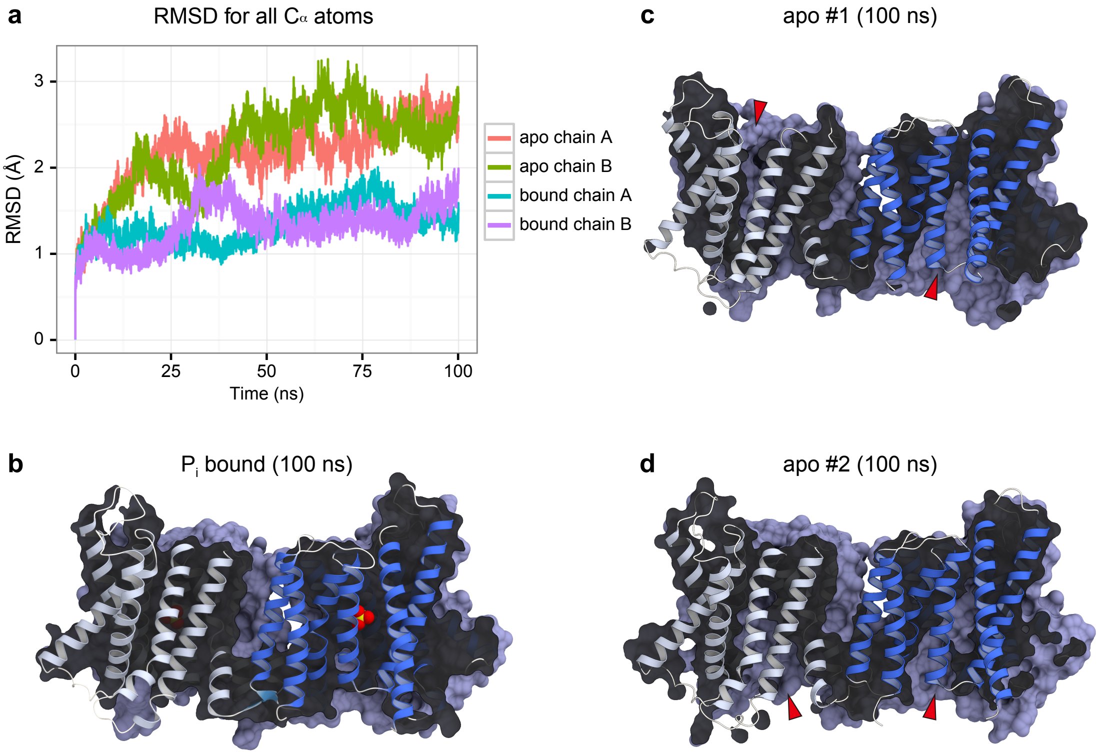


Figure 3.8: The results of the conventional MD simulation of GsGPT in dimeric state. (a) The time course of the RMSD value for the all C_α atoms. The reference structure is the crystal structure. (b-d) Surface section representation of the final snapshot of each simulation. The apo#1 and apo#2 in panel (c) and (d) represent the first and second trial of the simulation, respectively.

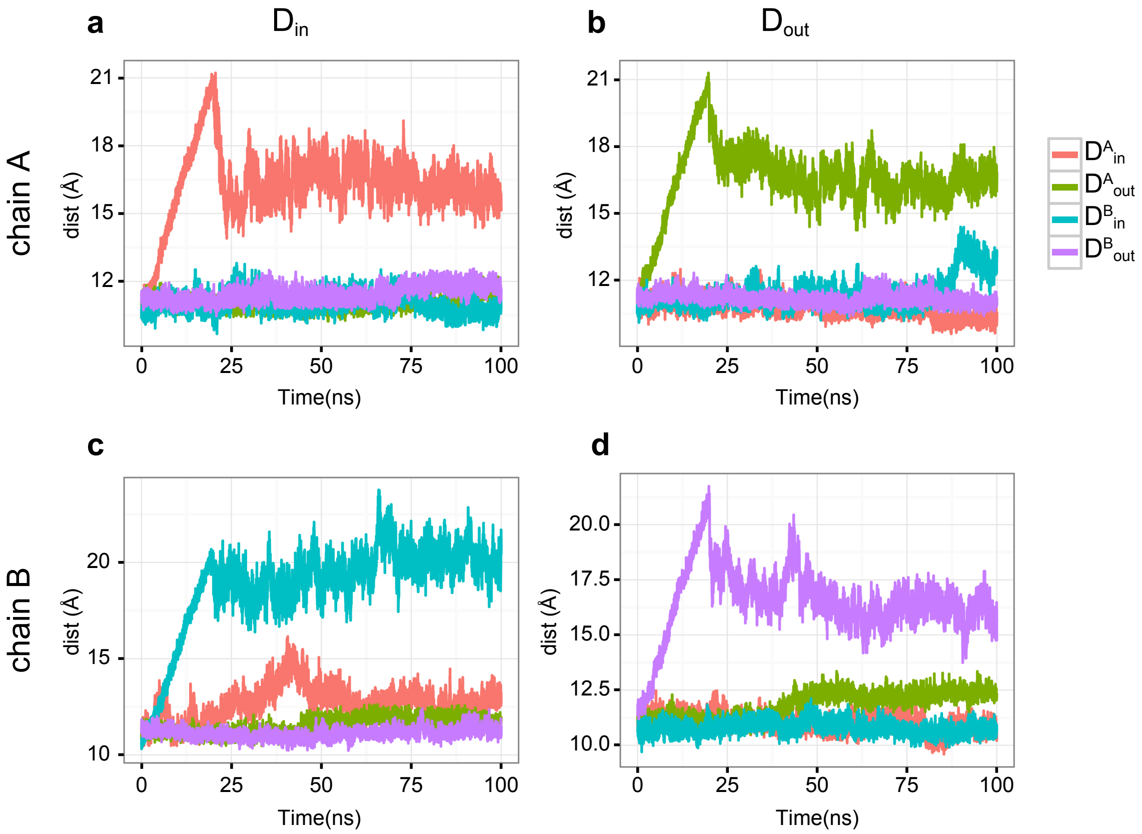


Figure 3.9: The results for SMD for each (a) D^A_{in} , (b) D^A_{out} (c) D^B_{in} and (d) D^B_{out} .

the crystal structure, (*i.e.* started from Occ-*apo* state), a rapid and large conformational change within 10 ns was observed (Fig. 3.8a; red and green lines). The molecular surface representation shows one protomer of GsGPT underwent the conformational change from Occ to OF, and the other underwent the conformational change from Occ to IF (Fig. 3.8c). Another independent trial of this simulation with re-assigned random initial velocity resulted in the different conformational change: the both protomer underwent the conformational change to IF (Fig. 3.8d). This result suggests that the conformational change of each protomer occurs independently.

3.3.2 The movement of each protomer is independent

To confirm the independency of the conformational change of each protomer, the author performed a steered molecular dynamics (SMD) simulation which induces the conformation of each protomer to inward or outward conformation. Various trials revealed that two distance parameters, D_{in} and D_{out} , were sufficient to induce the transition to the IF and

OF states from the Occ state crystal structure (Fig. 3.10), where D_{in} and D_{out} represent the center of mass distances between the C_α atoms of the inside and outside gate residues, respectively (see section 3.2.5 and Fig. 3.6 for a detailed definition). Then, one variable from four variables that is D_{in} and D_{out} for each protomer, D^A_{in} , D^B_{in} , D^A_{out} and D^B_{out} , is selected, and applied the force with constant velocity moving center harmonic restraints within 20 ns. Then, relaxation simulation without the force during 80 ns is performed. In these four SMD simulations, a significant change is observed only in the controlled variable, and any change is not observed the others (Fig. 3.9). These result suggest that the structural change of each gate does not affect that of the others, and thus the conformational change of each protomer is independent.

Although GsGPT forms dimer in the crystal structure, it is suggested that the monomer is sufficient to the conformational change required for the transport activity. Thus, hereinafter, the author performed MD simulations of GsGPT using one protomer, for computational efficiency.

3.3.3 Reconstruction of the GsGPT transport cycle

From the MD simulations above, the author obtained the stable atomic structures of IF-*apo* and OF-*apo* states, which are the both end-point structure of the transport cycle shown in Fig. 3.4. As summarized in the section 1.2.3, a feasible structural transition pathway between these end-points can be found by employing the path sampling methods. These sampling methods require a set of reaction coordinates which can describe the transition path of GsGPT. Thus, the author first tried to find these reaction coordinates (or, collective variables (CVs)). The best solution was using D_{in} and D_{out} again, and revealed that these variables were sufficient to induce the transition in monomer GsGPT to the IF and OF states from the Occ state crystal structure (Fig. 3.10) Moreover, it was revealed that the one-dimensional CV, $\Delta D = D_{out} - D_{in}$, can also be used to induce the conformational change from IF to OF (Fig. 3.10e, f). For simplicity, the author used this ΔD to describe the conformational space of GsGPT. Another reaction coordinate for the transition cycle is the position of P_i along the membrane normal axis, referred to as Z_{Pi} (Fig. 3.6; left

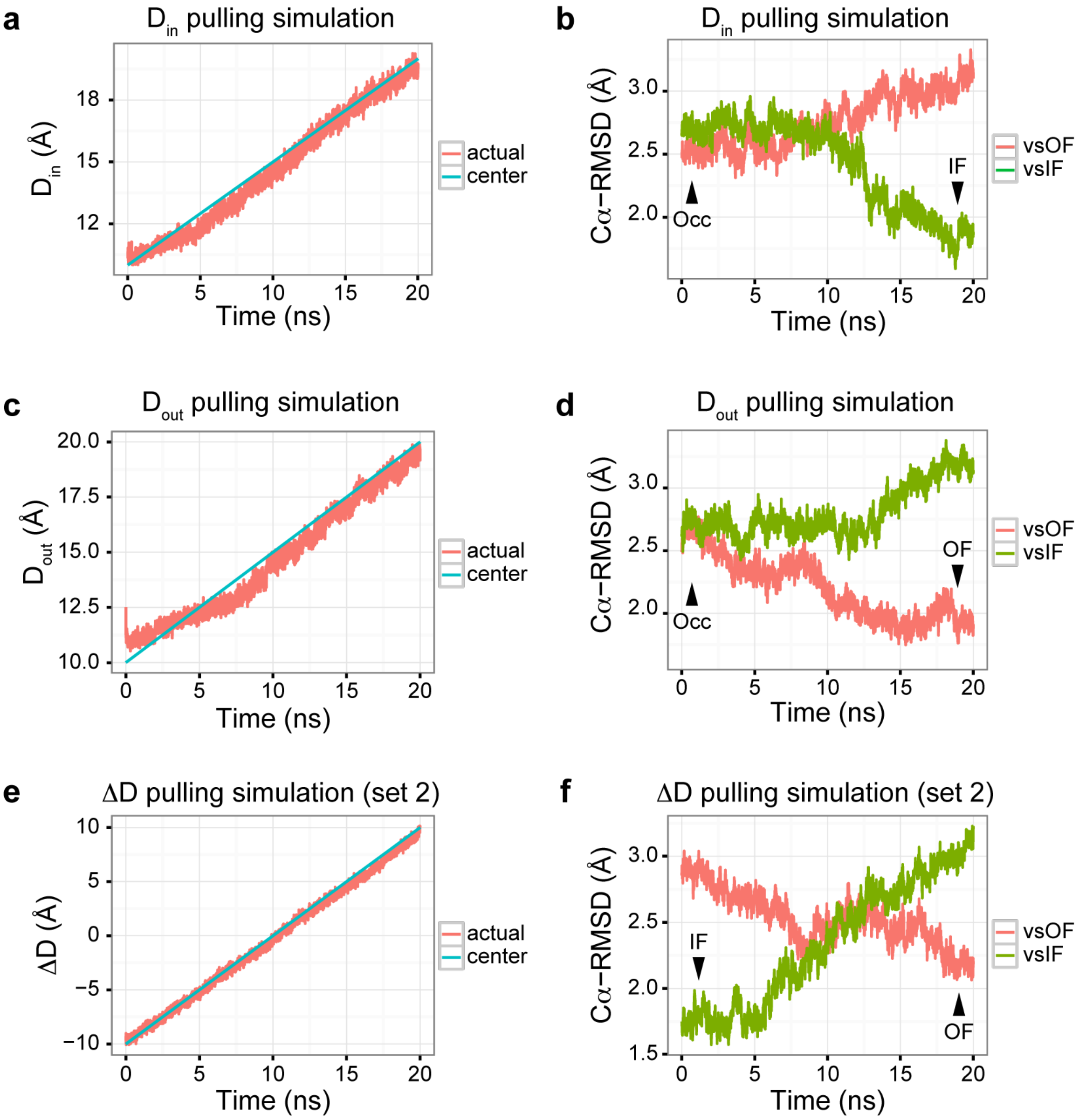


Figure 3.10: Time courses of the CVs and RMSD values in SMD simulation of GsGPT monomer. (a, c, e) Time courses of the CVs, (a) D_{in} , (c) D_{out} and (e) ΔD , in the SMD simulation of each variable. The blue line represents the center of the harmonic potential, and the red line represents the actual value of the CV. (b, d, f) Time courses of the RMSD values of all C_α atoms relative to the OF (red) and IF (green) conformations.

panel). Finally, the author performed 12 sets of biased samplings in the two-dimensional CV space, ΔD and Z_{Pi} .

The reconstruction of the transport cycle is performed by the following steps (Fig. 3.7e): first of all, the SMD simulation was performed to induce the transitions from Occ to IF_b^{cent} and OF_b^{cent} (IF- and OF-bound at the central binding site, respectively) (Fig. 3.4), using the CV ΔD (simulation sets 1 and 2 in Table 3.1). The subsequent SMD simulation for Z_{Pi} was performed to generate the structures in which P_i is bound near the gate, IF_b^{intra} (IF-bound at intracellular gate) and OF_b^{extra} (OF-bound at extracellular gate) (simulation sets 3 and 4). To search the minimum free energy pathway of the $IF_b^{intra} \leftrightarrow$ Occ $\leftrightarrow OF_b^{extra}$ transition in the ΔD - Z_{Pi} two-dimensional space, the SMwST [93] algorithm was employed (simulation set 5), followed by the free energy calculation with a set of BEUS [10, 11, 94] along the refined transition string (simulation set 6) using “path collective variables” (pathCV(s) and pathCV(ζ); equations (3.2) and (3.3)) [95]. Next, to obtain the *apo* states of GsGPT, IF_a and OF_a in Fig. 3.4, the P_i was pulled out from the IF_b^{cent} and OF_b^{cent} states to the bulk solvent region (simulation sets 7 and 9), and performed free energy calculations along the Z_{Pi} variable (simulation sets 8 and 10). The results of these three sets of free energy calculations were combined with the recently developed PHSM [94] to estimate the minimum free energy pathway for the full transition cycle, $IF_a \leftrightarrow$ Occ $\leftrightarrow OF_a$. The extracted transition pathway was further refined with SMwST (simulation set 11). Finally, the free energy calculation for the entire transport cycle (simulation set 12) was performed along this optimized transition pathway, using pathCV(s) and pathCV(ζ). The author obtained the entire transport cycle of GsGPT with 1,000,000 configurations, and calculated the free energy along the optimum transition pathway using the WHAM [102] (Fig. 3.11). The following structural analyses were performed with the structures obtained from this simulation set 12.

3.3.4 Overall conformational change of GsGPT

To investigate the overall conformational change of GsGPT, representative structures corresponding to the free energy basins of the IF_a and OF_a states (Fig. 3.11) were compared

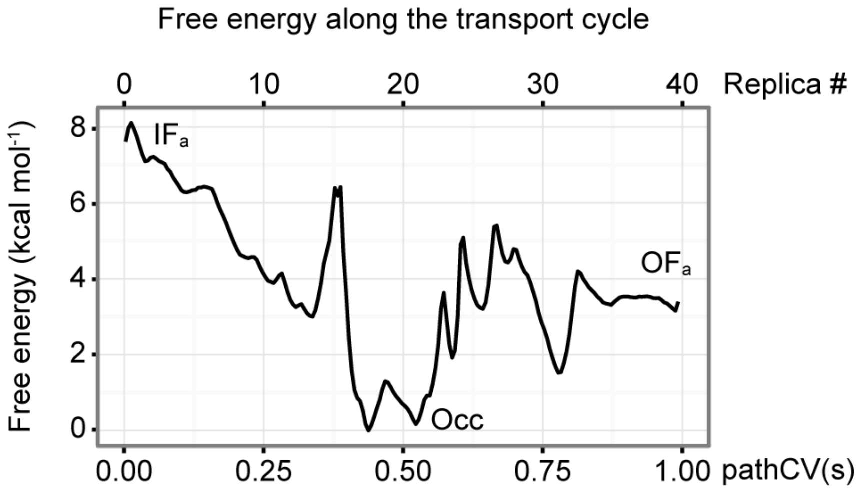


Figure 3.11: Free energy profile along the simulated transport cycle shown in Fig.3.4, based on the final simulation (simulation set 12 in Table 3.1). pathCV(s) represents path collective variable s (equation (3.2)).

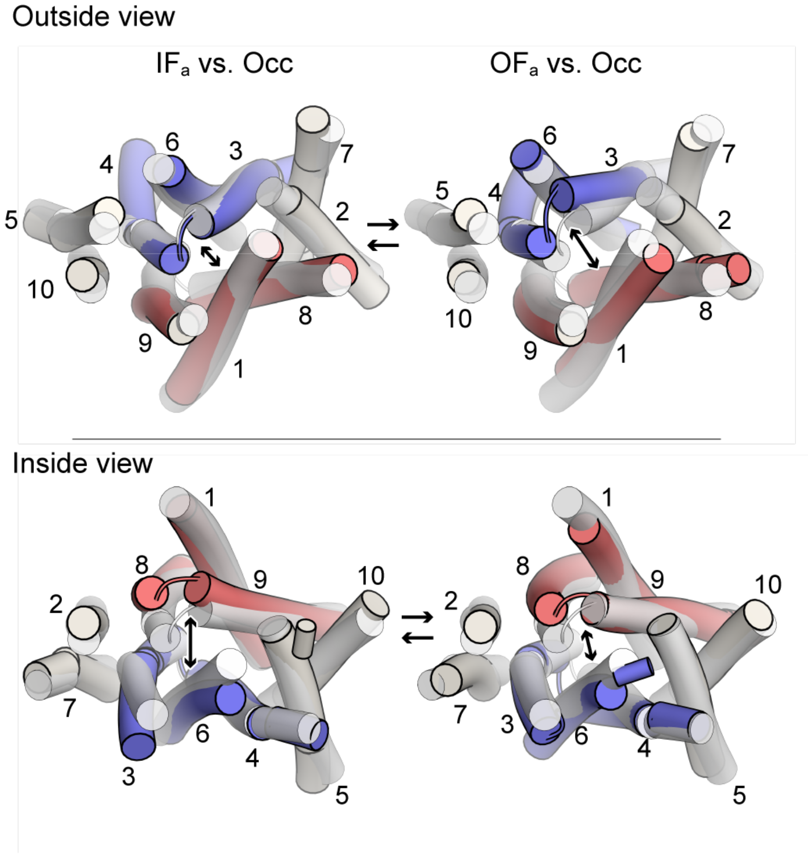


Figure 3.12: Outside (upper) and inside (lower) views of representative structures in the OF_a and IF_a states, shown in cylinder representations. Each structure was superimposed onto the Occ state structure, shown with a transparent cylinder. Bundle1 and bundle2 are colored red and blue, respectively, and the other helices are colored white. The number represents the numbering of the TM helices. The narrow cylinders represent the 3_{10} helix region.

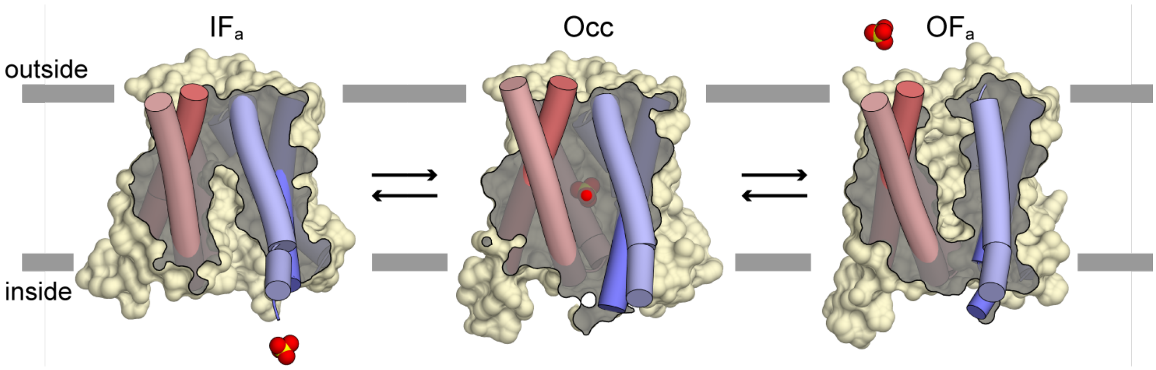


Figure 3.13: Cross-sections of surface representations for the representative structures from the IF$_a$, Occ and OF$_a$ states. P$_i$ is shown in a CPK model.

with that of the Occ state (Fig. 3.12). The major differences between the IF and OF states were the outward-gate distance between the hairpin of TM3-4 and TM1 (Fig. 3.12;upper), and the inward-gate distance between the hairpin of TM8-9 and TM6 (Fig. 3.12;lower). The superimposition of the IF$_a$ and OF$_a$ states onto the Occ state revealed that the TM bundles composed of TM1, 8 and 9 (bundle1) and TM6, 3 and 4 (bundle2) are directly involved in the overall conformational change. In contrast, the other helices (TM2, 5, 7 and 10) do not participate in the open/closed conformational changes, and function as a scaffold for the conformational changes of bundle1 and bundle2. Hereafter, we refer to bundle1 and bundle2 (*i.e.*, TM1, 3, 4, 6, 8, and 9) as the core domain.

The molecular surfaces for the representative structures illustrate that the substrate binding site is exposed to opposite sides of the membrane in an alternating fashion, known as the alternating access mechanism [1, 2] (Fig. 3.13). To confirm whether all of the structures obtained from the simulation satisfy the alternating access mechanism, pore radius calculations were performed along the transport pathway by HOLE [98] for all structures obtained from the simulation (for representative results, Fig. 3.14). The outward and inward minimum radii were defined as the minimum radii of the pore in the regions of $-20 < z < -7$ and $7 < z < 20$, respectively, where z represents the coordinate along the pore pathway ($z = 0$ corresponds to the central binding site, and $z > 0$ corresponds to the outward direction). The plot of the outward vs. inward minimum radii illustrates that the conformational transition between the IF and OF states always visits the Occ-state region, and none of the structures visit the region where the substrate binding site is exposed to both

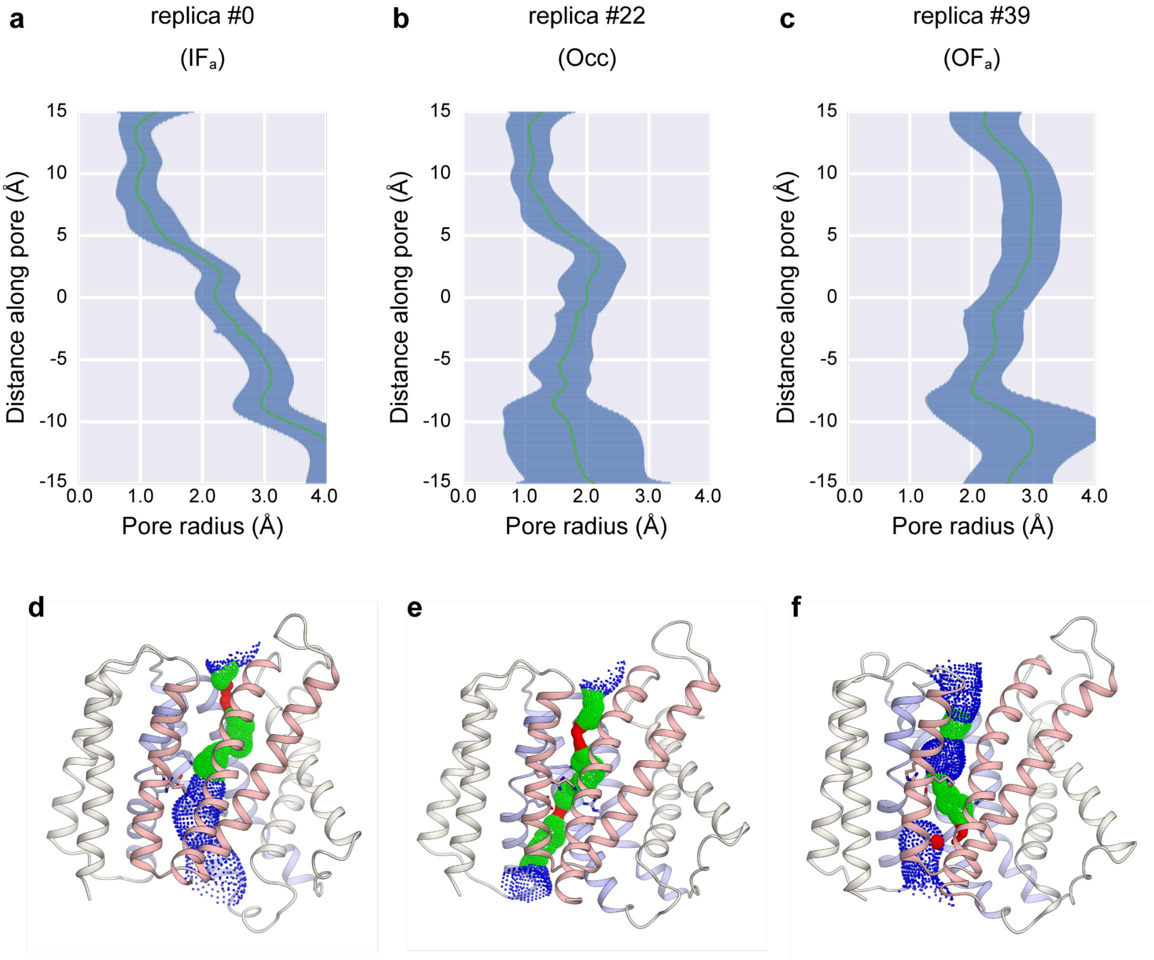


Figure 3.14: Representative results for the HOLE analysis. (a-c) Minimum radius along the path through the central binding site for (a) replica# 0 (IF), (b) replica# 22 (Occ) and (c) replica# 39 (OF). The green line represents the mean value and the blue line represents the mean value and standard deviation (s.d.) for each distance along the pore. (d-f) Representative structures of the (d) IF, (e) Occ and (f) OF states. The red region represents $r < 0.6$ Å, the green region represents $0.6 < r < 1.2$ Å and the blue region represents $r > 1.2$ Å.

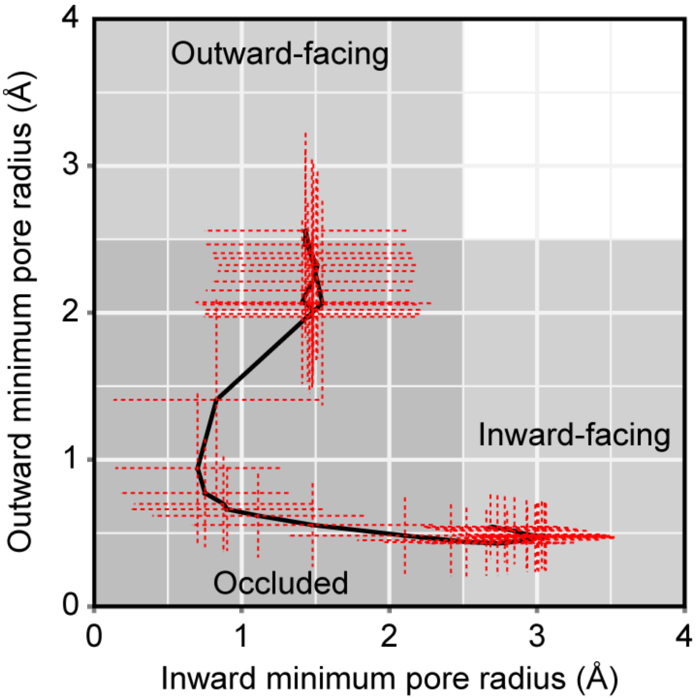


Figure 3.15: Pore radii at the outside and inside gates. The mean values of the outside and inside radii in each replica were plotted. The error bars represent the s.d. in each replica.

the inward- and outward-side solvent (Fig. 3.15). This result indicates that the obtained transition pathway satisfies the alternating access, suggesting the validity of the simulated transition pathway.

Next, to reveal the mechanism of the overall conformational change, we investigated the internal conformational change within the core domain helices, bundle1 and bundle2. Bundle1 and bundle2 in the IF_a conformation both superimpose well onto those in the OF_a conformation, respectively (Fig. 3.16a, b), suggesting the semi-rigid body movement of the core domains along the transport cycle. In fact, the distributions of the RMSD values of the bundle1 and bundle2 C_α atoms against the crystal structure illustrated that the RMSD values for only bundle1 or bundle2 remained lower than that of bundle1+2 over the entire transport cycle (Fig. 3.16c). These results suggest that the conformational transition between IF and OF is governed by the semi-rigid body tilting motions of bundle1 and bundle2, rather than the internal bending and straightening motion of the TM helices. To confirm this mechanism, the free energy was projected onto the two-dimensional space consisting of the the first principal component of protein (PC1) of the C_α atoms and the inter-helical angle between TM1 and TM6 (Fig. 3.18e). The PC1 correlates well with the

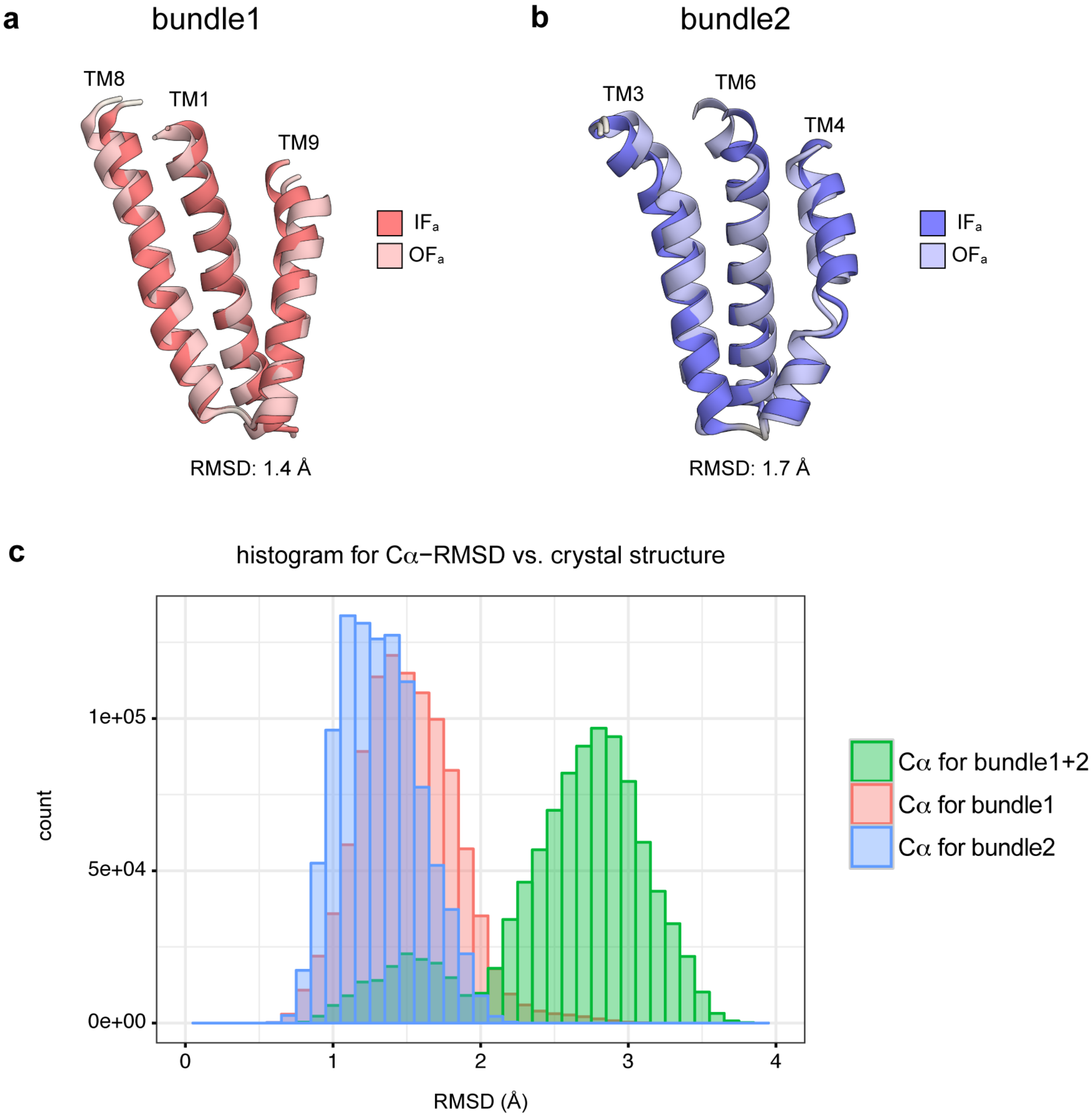
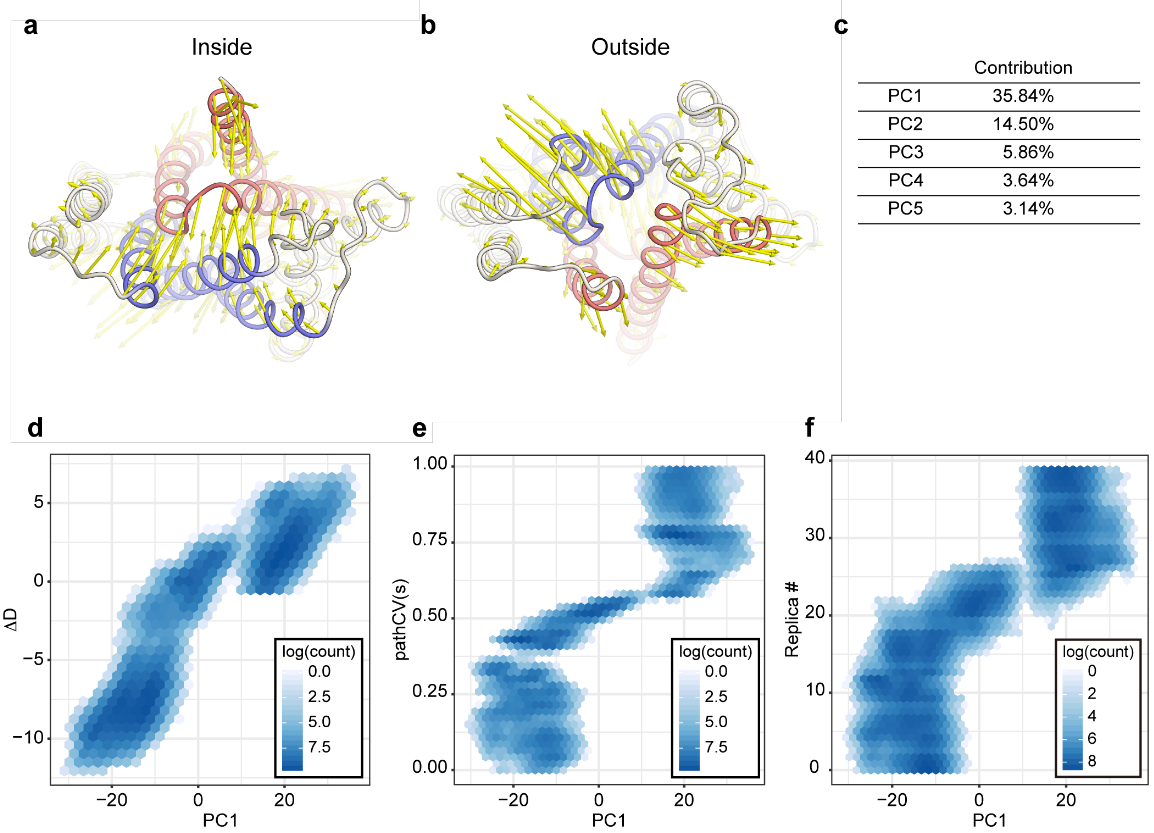


Figure 3.16: (a, b) Representative conformations of (a) bundle1 and (b) bundle2 from the IF and OF conformations. The root-mean-square deviation (RMSD) values of the C_α atoms for each structure are shown below. (c) The histogram of RMSD values of C_α atoms in bundles1+2 (green), bundle1 only (red) and bundle2 only (blue) values relative to the crystal structure. The RMSD values were calculated for the α-helix region of each bundle.



	Contribution
PC1	35.84%
PC2	14.50%
PC3	5.86%
PC4	3.64%
PC5	3.14%

Figure 3.17: (a, b) Vector representation of the eigenvector of the first principal component of protein (PC1) of GsGPT movement, calculated by principal component analysis (PCA). (c) The contributions of PC1-5 for the overall deviations. (d-f) Correlations of the PC1 vs. (d) ΔD , (e) the number of replicas and (f) the pathCV(s) shown in a 2D histogram. The count value is shown in a log scale.

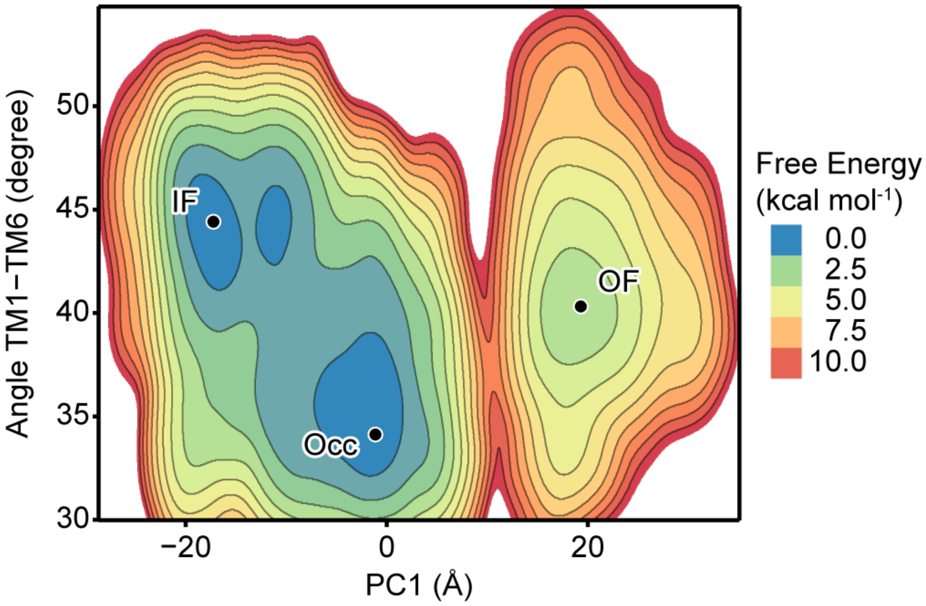


Figure 3.18: Free energy landscape in terms of (PC1, TM1-TM6 angle) space. PC1 represents the first principal component of all C_α atoms, and the inter-helical angle between TM1 and TM6 was calculated using the roll axis of the helix (obtained from the principal axis component analysis for all C_α atoms).

ΔD , pathCV(s), and replica number (Fig. 3.17), and represents the global conformational transition as well as the ΔD . The inter-helical angle between TM1 and TM6 represents the inter-domain angle between bundle1 and bundle2. This free energy landscape shows three distinct free energy basins corresponding to the IF, Occ and OF states, suggesting that the tilting motions of bundle1 and bundle2 play a major role in the overall conformational change.

These results are in contrast with the previously suggested mechanism hypothesized from the static crystal structures of SnYddG and GsGPT, in which the bending motion of the TM helices governs the conformational transition [76]. In the previous report of the bacterial DMT superfamily transporter, SnYddG, the bending and straightening of the core bundle, especially TM3-4 and TM8-9, were suggested to regulate the opening and closing of the inside and outside gates [78]. However, the present results suggest that the tilting motion of the semi-rigid core domain, which resembles the rocker-switch mechanism proposed for the MFS transporters [2], is sufficient to explain the conformational transition of GsGPT. Thus, the alternating access mechanism of the DMT superfamily is unexpectedly similar to that of the MFS transporters, despite their completely different TM topologies.

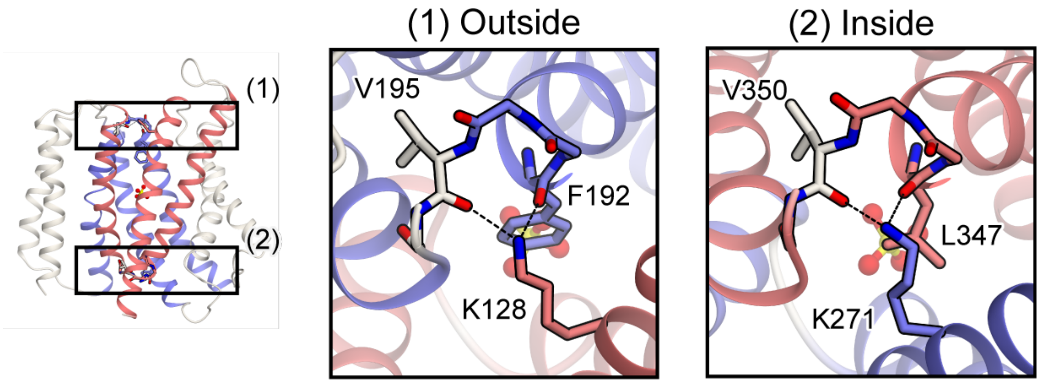


Figure 3.19: Overall structure of GsGPT (upper) and close-up views of the outside (center) and inside (lower) gates. The dotted line represents a hydrogen bond between a Lys residue and a main chain carbonyl atom.

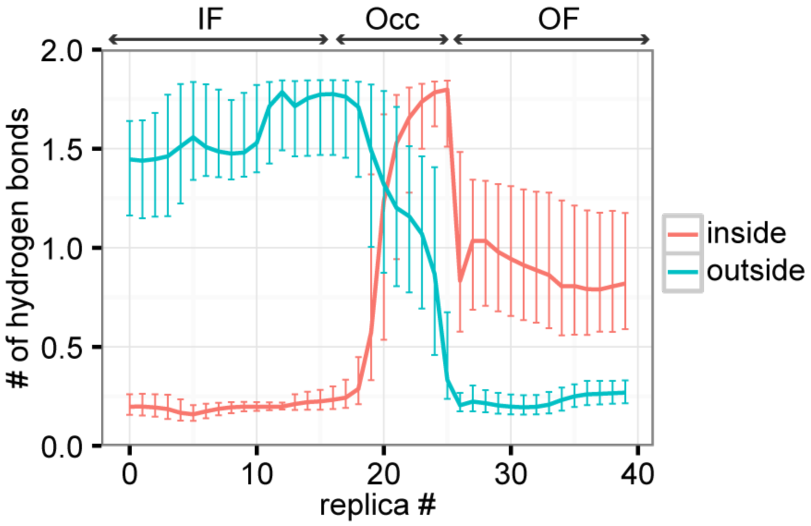


Figure 3.20: The number of hydrogen bonds in the inside and outside gates (magenta and cyan, respectively), shown in panel a. The median value in each replica is plotted. Error bars represent the interquartile range (IQR).

3.3.5 Hydrogen bonds at the inside and outside gates

In the crystal structure of GsGPT, the well conserved Lys271 and Lys128 residues form hydrogen bonds with the main chain carbonyl atoms on the opposite bundle, and lock the inside and outside gates, respectively (Fig. 3.19). In the previous analysis based on the structural comparison with SnYddG, the disruption and formation of these hydrogen bonds were suggested to regulate the gate opening and closing [76]. However, in the present simulation, the two hydrogen bonds of Lys128 shown in Fig. 3.20 were not always maintained, even in the IF state (Fig. 3.19; cyan), and the two hydrogen bonds of Lys271 were also not always maintained, even in the OF state (Fig. 3.20; magenta). Furthermore, the

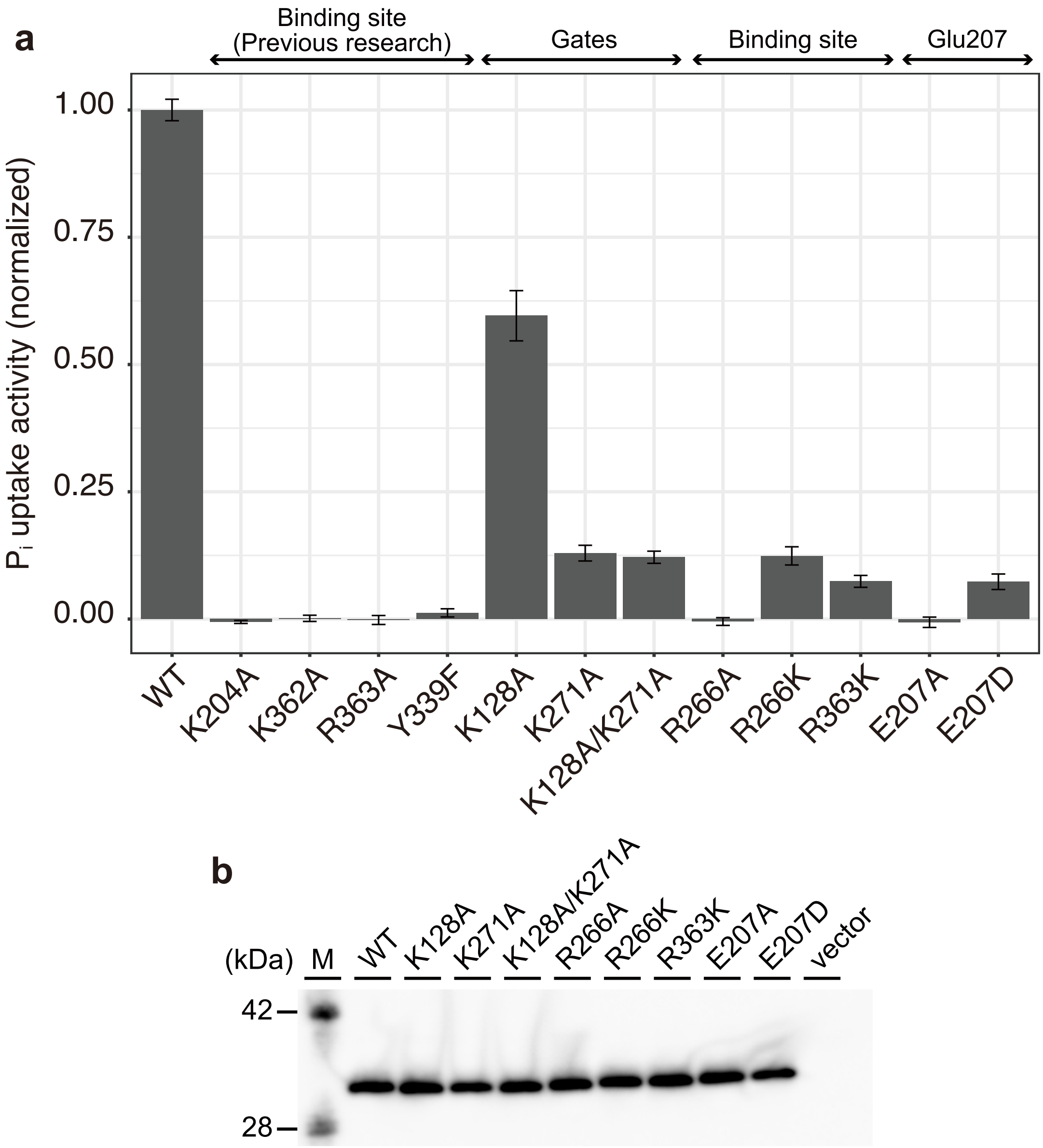


Figure 3.21: (a) Liposome-based mutational analysis for GsGPT mutants. The levels of [^{32}P]-P_i uptake by GsGPT mutants were normalized with the mean value of the wild-type. The error bars represent s.e.m. (n=3-9). The values of K204A, K362A and R363A are from the previous study [76]. (b) Western blotting analysis for WT and each GsGPT mutant, confirming the comparable expression levels. “M” represents the marker lane.

liposome-based transport assays of K271A, K128A and their double mutant revealed that all of these mutants still retain the P_i/ P_i homo-exchange activity, as compared with the previously reported mutants in the substrate binding site, K204A, K362A and R363A [76] (Fig. 3.21). These results suggest that the hydrogen bond formation by Lys271 or Lys128 is not crucial for the conformational change. These investigations are consistent with the previous functional analysis of the GsGPT homologue, the apicoplast phosphate translocator of *Toxoplasma gondii* (TgAPT), which revealed that the K67A mutation of TgAPT, corresponding to the K128A mutation of GsGPT, retained 10-20% activity [104]. The current simulation suggests that Lys128 and Lys271 play another role in the conformational transition of GsGPT, rather than the regulation of gate opening and closing by hydrogen bond formation.

3.3.6 Substrate binding and translocation mechanism

To gain detailed insight into the mechanism for substrate binding and translocation, the free energy information was projected onto the PC1- Z_{Pi} plane, and compared the representative structures of the free energy basins (Fig. 3.22). The translocation of P_i from the outside (intermembrane side) to the inside (stroma side) can be described by the following reaction steps. (i) At first, P_i in the outside solution is captured by the positively charged residues, and passed to Lys128 (Fig. 3.22a and Fig. 3.23a). (ii) In addition to Lys128, P_i is captured by Lys204, one of the central substrate binding residues (Fig. 3.22b and Fig. 3.23b). (iii) The salt bridge between P_i and Lys128 is disrupted, and then P_i binds to Lys362 and Arg363 (Fig. 3.22c and Fig. 3.23c, d). (iv) The hydrogen bond between P_i and Tyr339 on TM9 is formed (Fig. 3.22d and Fig. 3.23e), which enables the outward halves of the core domains to approach each other, and the conformational change from the OF to Occ states is facilitated. (Fig. 3.22d). This configuration is the most stable state in the simulation (Fig. 3.11 and Fig. 3.22), consistent with the fact that the two crystal structures of GsGPT were obtained in the substrate-bound Occ state. Note that two distinct configurations were degenerate in this Occ state in terms of the binding modes of Arg266: in one state, Arg266 binds to P_i with a water-mediated indirect interaction, as observed in the crystal structure

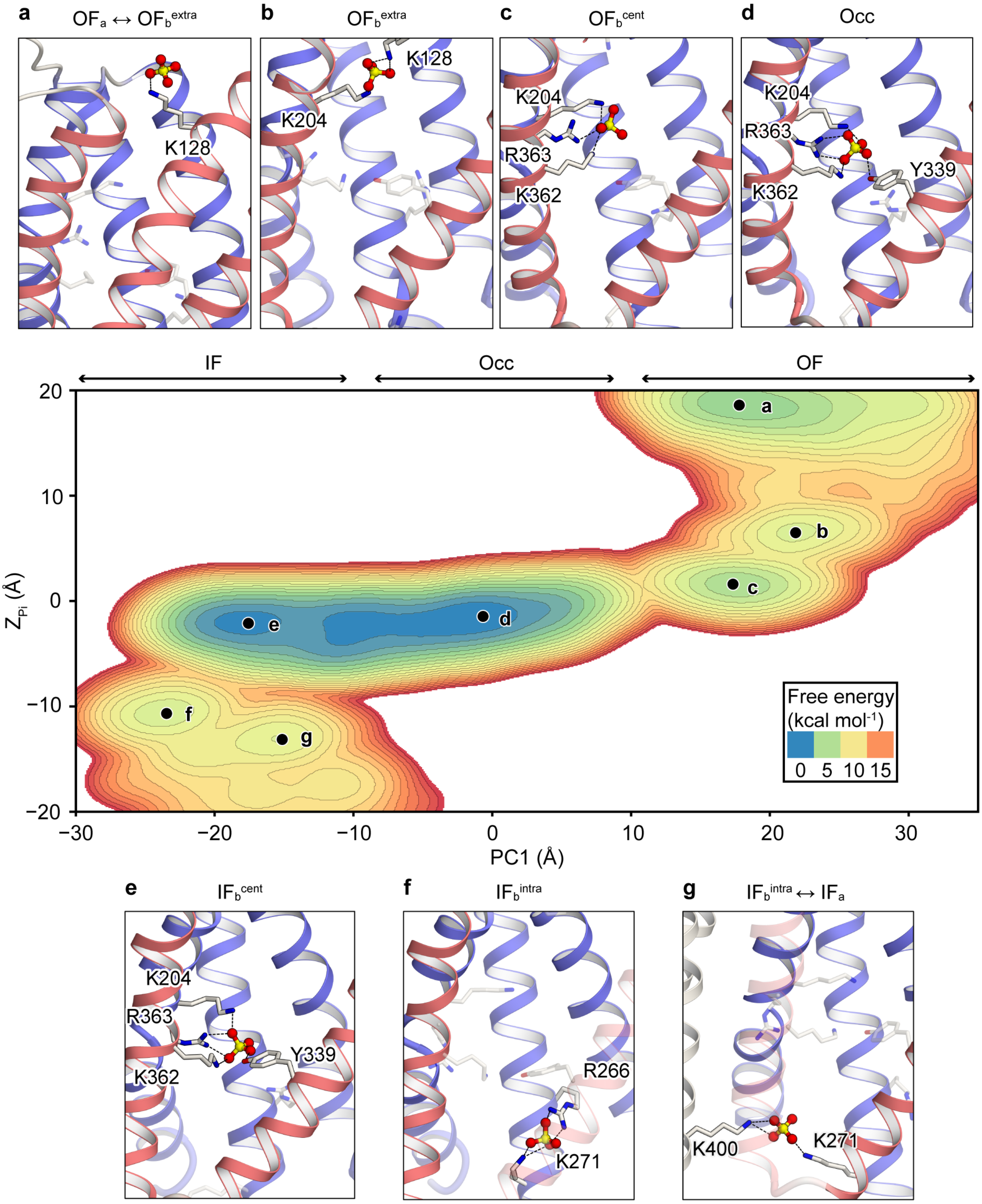


Figure 3.22: Free energy landscape in the (PC1, $Z_{\mathrm{P_i}}$) space is shown for the regions in which $\mathrm{P_i}$ is bound to GsGPT. (a-g) Snapshots of the binding site conformation corresponding to each free energy basin. The residues that directly interact with $\mathrm{P_i}$ in the transport cycle are shown with stick models. Dotted lines represent strong interactions in which the distance between $\mathrm{P_i}$ and each residue is less than 3.0 Å. TM1 is not shown, except in panel a.

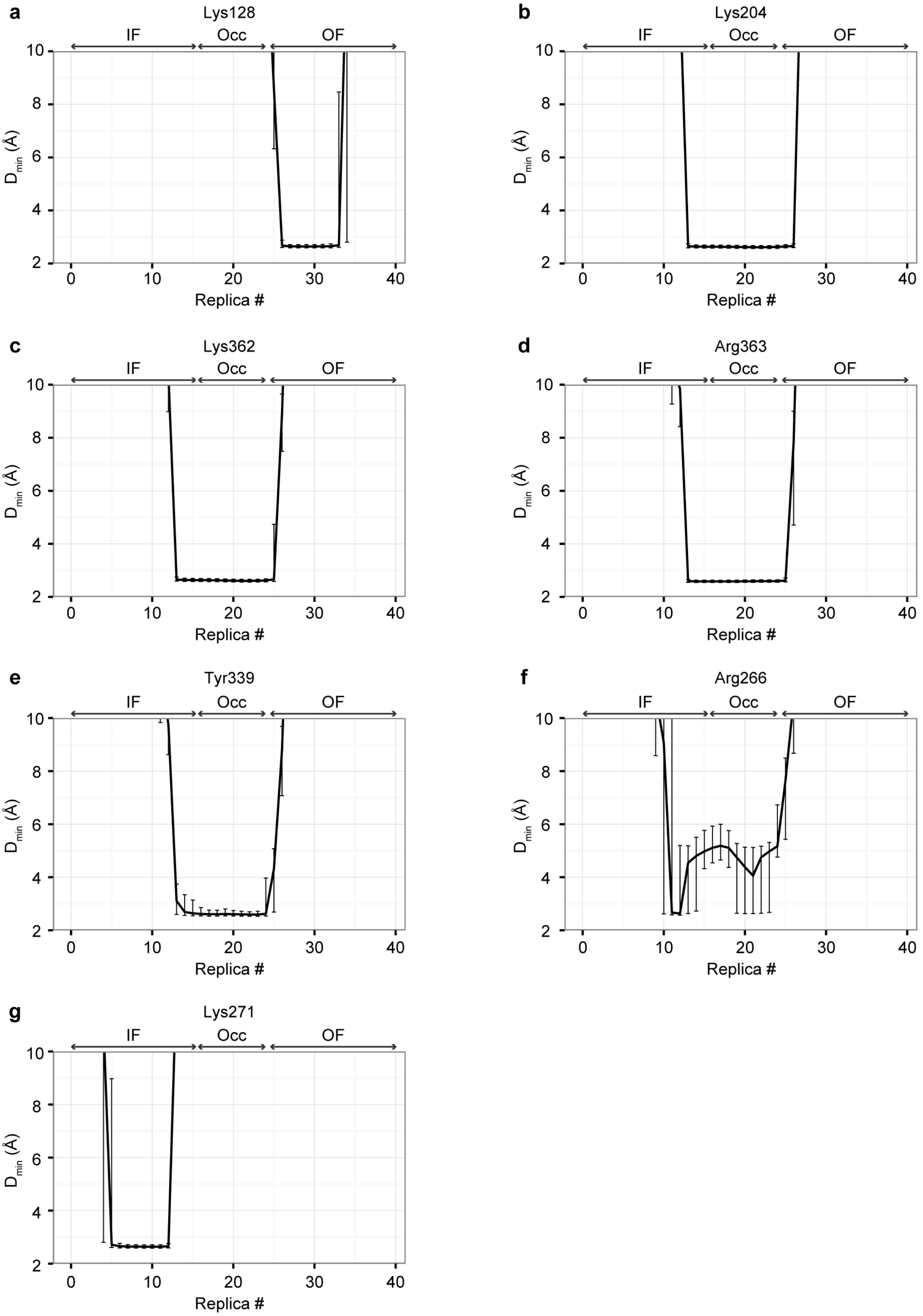


Figure 3.23: The minimum distances between P_i and the P_i interacting residues, (a) Lys128, (b) Lys204, (c) Lys362, (d) Arg363, (e) Tyr339, (f) Arg266 and (g) Lys271. The median value within each replica is plotted, and the error bars represent the interquartile range (IQR).

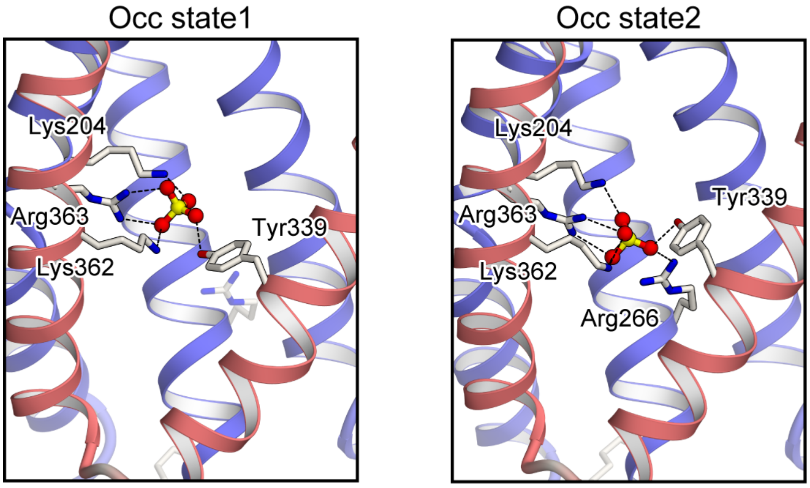


Figure 3.24: (a, b) Close-up views of the substrate binding site for two conformations corresponding to the free energy basins shown in Fig. 3.22d. TM1 is not shown.

(Fig. 3.24a); in the other, Arg266 directly forms a salt bridge with P_i (Fig. 3.24b). These two states can be discriminated by the slight difference of Z_{Pi} , and this difference might determine the direction of the conformational transition from the Occ state to the IF or OF state. (v) The inside gate is disrupted, and the conformational change from the Occ to IF state occurs (Fig. 3.22e). (vi) P_i is dissociated from the central binding site and bound to both Arg266 and Lys271 (Fig. 3.22f and Fig. 3.23f, g). (vii) The salt bridge between P_i and Arg266 is disrupted, and P_i is released to the inside solution mediated by Lys271 and other basic residues located on the inward surface, *e.g.* Lys400 (Fig. 3.22g). Each free energy basin corresponds to a distinct substrate binding mode, and the transport mechanism can be explained as the relaying of P_i by these pore lining, basic residues (except for Tyr339). The free energy barriers between the P_i binding modes correspond to the energy required for disrupting the salt bridge between P_i and these basic residues.

The D_{min} plot between P_i and these pore lining residues shows that the residues in the central binding site, Lys204, Tyr339, Lys362 and Arg363, interact strongly with the substrate in all of the OF, Occ and IF states (Fig. 3.23b-e). This observation is consistent with the previous mutational analysis, in which the mutations of these residues, K204A, Y339F, K362A and R363A, abolished the P_i/P_i homo-exchange activity [76]. It is also interesting to note that Lys271 and Lys128 in the inside and outside gates strongly interact with P_i in the transport cycle. Given that the hydrogen bond formation of Lys128 and Lys271 is

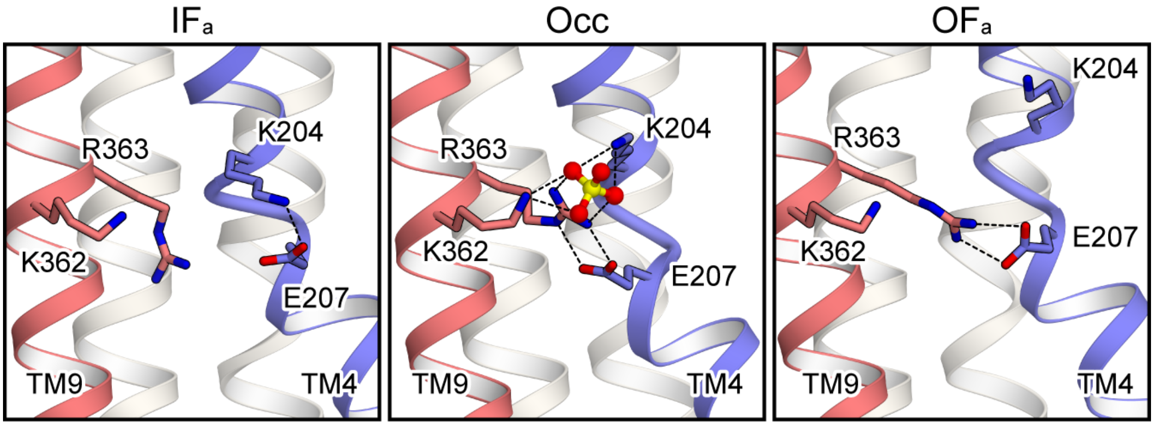


Figure 3.25: Close-up view of the substrate binding site for representative structures of IF, Occ and OF. Dotted lines represent strong salt bridge interactions, in which the distance between each atom is less than 3.5 Å.

dispensable for the conformational transition (section 3.3.5), the decreases of the transport activity in the K128A, K271A and K128A/K271A mutants may be due to the loss of these interactions. Thus, the author conclude that Lys128 and Lys271 primarily function as the initial binding sites for the substrates, rather than in the inside and outside gate formation.

To further support this relaying mechanism of P_i, additional mutational analyses using the liposome-based assay were performed (Fig. 3.21). The result revealed that the R266A mutant also completely abolishes the P_i exchanging activity, showing the importance of Arg266 for the transport mechanism. In addition, the conservative mutants, R266K and R363K, exhibited significantly decreased activities (Fig. 3.21), suggesting that the shapes of these side chains are also important for the recognition of the tetrahedral arrangement of the substrate.

3.3.7 Coupling mechanism of substrate binding and protein conformational change

To reveal the conformational coupling mechanism of GsGPT, the author focused on Glu207 on TM4, which is the only conserved acidic residue in the positively-charged substrate binding pocket (Fig. 3.25). Calculations of the minimum distances between the neighboring basic residues, Lys204 and Arg363, revealed that Glu207 exchanges its salt bridge partner during the transport cycle (Fig. 3.25 and Fig. 3.26). In the IF_a state, Glu207

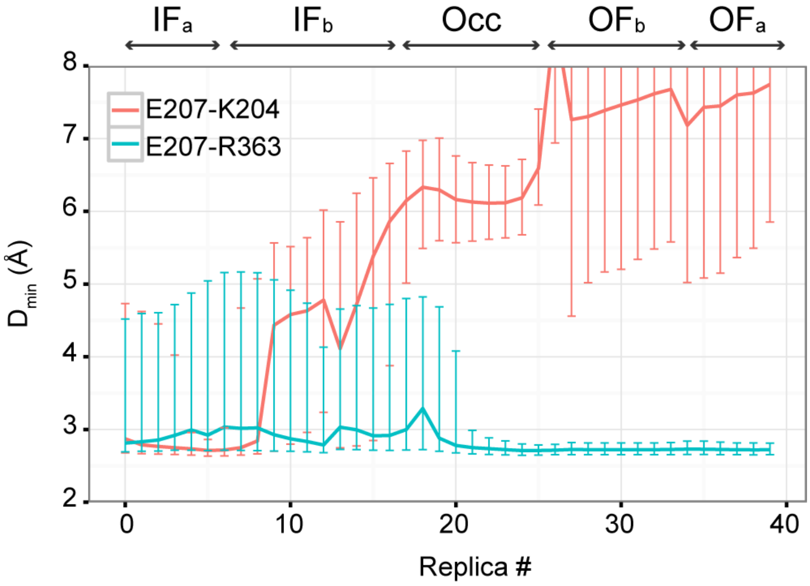


Figure 3.26: Minimum distance between Glu207 and Lys204 (magenta) and between Glu207 and Arg363 (cyan). The median value of the distance in each replica is plotted, and the error bars represent the IQR.

mainly forms an intra-helical salt bridge with Lys204 (Fig. 3.26; magenta line). This intra-helical salt bridge fixes the side chain of Lys204 to the Glu207 side, thereby causing the electrostatic repulsion between Lys204 and Arg363 to prevent the conformational transition to Occ and stabilize the IF conformation (Fig. 3.25; left). By contrast, in the $\mathrm{OF_a}$ state, Glu207 forms a stable inter-helical salt bridge with Arg363 on TM9 (Fig. 3.26; cyan line). This inter-helical salt bridge between Glu207 and Arg363 clamps the inward halves of TM4 and TM9 together (Fig. 3.25; right), thereby fixing the arrangement of bundle1 and bundle2 to the OF conformation. $\mathrm{P_i}$ binding to these basic residues weakens the salt bridges between the basic residues and Glu207, which may facilitate the exchange of the salt bridge partner of Glu207 (Fig. 3.25; center).

This salt-bridge rearrangement involving Glu207 leads to a plausible mechanism of the conformational coupling, which can explain how the local conformational change upon the substrate binding is amplified to the global motions of bundle1 and bundle2. In this mechanism, Glu207 plays a role as a “switch” that turns on the global tilting motions of TM4 and TM9 upon substrate binding. Since TM4 and TM9 constitute the (semi-)rigid bodies of bundle1 and bundle2, respectively, these tilting motions of TM4 and TM9 directly result in the global IF/OF conformational change of GsGPT (Fig. 3.27). This mechanism is in

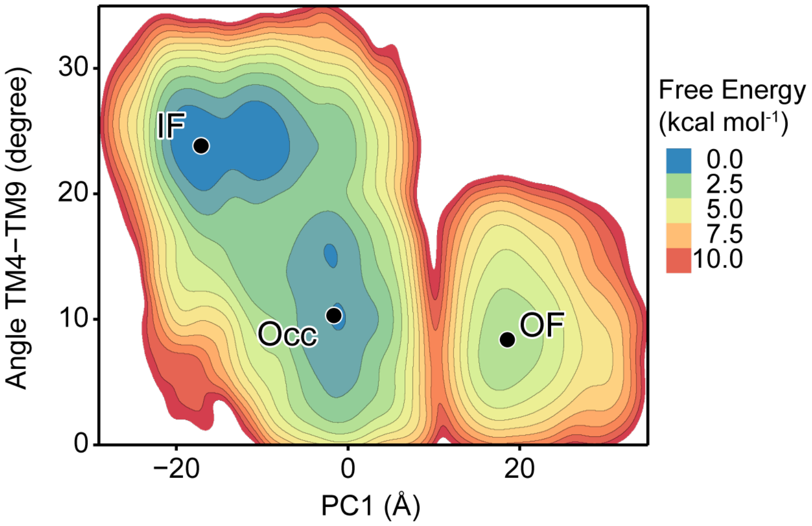


Figure 3.27: Free energy landscape in terms of (PC1, TM4-TM9 angle) space. PC1 represents the first principal component of all C_α atoms, and the inter-helical angle between TM4 and TM9 was calculated using the roll axis of the helix (obtained from the principal axis component analysis for the C_α atoms).

contrast to the previous working model based on the static crystal structure of GsGPT, in which the electrostatic repulsion by the positively charged residues in the substrate binding pocket is simply neutralized by the binding of the negatively-charged substrate, thereby enabling the global IF/OF conformational change [76]. To further support the importance of Glu207, the liposome-based assays of Glu207 mutants were performed. The results revealed that the E207A mutation abolishes the P_i/P_i homo-exchange activity (Fig. 3.21), in spite of its stable membrane expression (Fig. 3.21b). Moreover, the conservative mutation, E207D, significantly reduces the exchange activity (Fig. 3.21), probably due to the insufficient side chain length to form the salt bridge network (Fig. 3.25). Taken together, in addition to the previous mechanism based on the electrostatic repulsion and neutralization, the switch mechanism by Glu207 plays a pivotal role in the coupling mechanism of the local and global conformational changes, thus allowing the strict 1:1 exchange of the substrates by GsGPT.

3.4 Discussion

3.4.1 Summary of this chapter

The author performed MD simulation of GsGPT in both dimeric and monomeric state. In the dimeric state of GsGPT, the spontaneous conformational change upon the substrate removal from the crystal structure was observed. It was indicated that the conformational change of each protomer of GsGPT is independent each other. In the monomeric state of GsGPT, the entire transport cycle of GsGPT was successfully reconstructed from a single crystal structure, by combining string method and umbrella sampling. The analysis of all configurations obtained from the simulation revealed the fundamental antiport mechanism of GsGPT in atomic detail.

3.4.2 Instability of Occ-*apo* state

In the dimeric state simulation, the spontaneous conformational change upon the substrate removal from the crystal structure was observed (Fig. 3.8). This result shows that Occ-*apo* state is unstable and this state rapidly transitions to $\mathrm{IF_a}$ or $\mathrm{OF_a}$ states. This conformational transition is due to the nature of an exchanger; Occ-*apo* state must not exist in exchanger as described in the section 1.1.4. The author also attempted to reconstruct the structural transition in *apo* state, but the conformational change between $\mathrm{IF_a}$ and $\mathrm{OF_a}$ cannot be induced by the SMD simulation using $\Delta \mathrm{D}$, due to high free energy barrier between these states. This barrier guarantees the strict antiport of GsGPT, and this molecular mechanism is revealed to be the switching by Glu207 residue in the current study.

3.4.3 Generality of the transport mechanism in the pPT family

In this simulations, a highly symmetric substrate, $\mathrm{P_i}$, is used to simplify the CV used in the simulations. However, we can gain insight into other pPT family protein substrates that have more complicated structures, such as 3-PGA. Although pPT family proteins transport various types of phosphosugars, depending on their subtypes [59, 76], all of the substrates

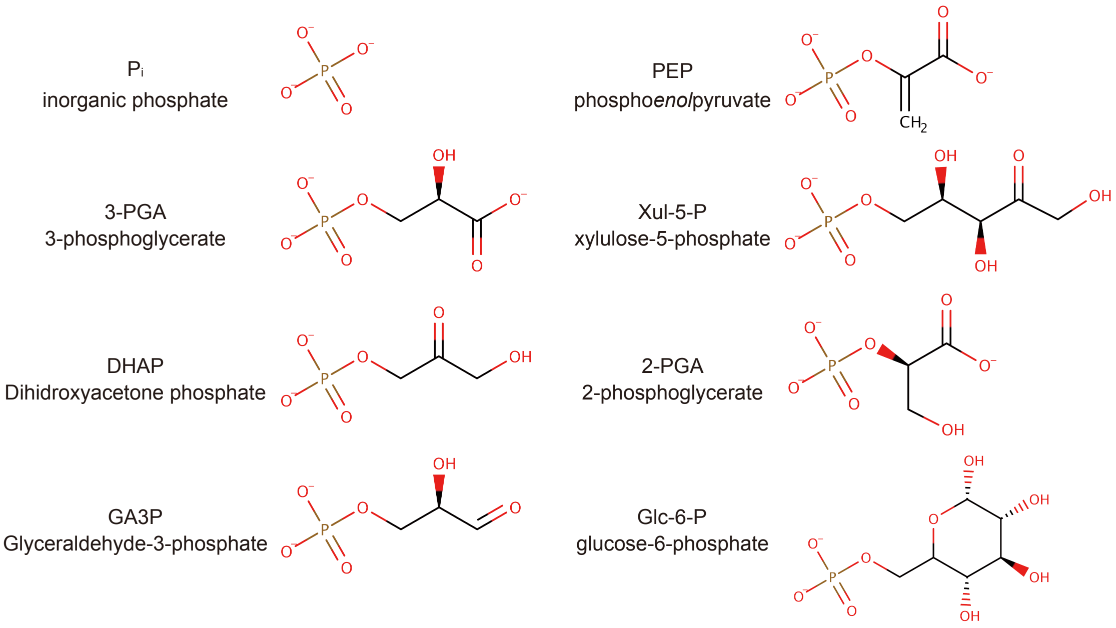


Figure 3.28: Substrates of plastidic phosphate tranlocator (pPT) families.

have a common structural feature: one oxygen atom of P_i is replaced with a sugar moiety (Fig. 3.28). In the current simulation, the substrate binding pocket in all of the binding modes can accommodate a compound in which one oxygen atom of P_i is replaced with a sugar moiety (Fig. 3.22a-g), as observed in the previous 3-PGA bound crystal structure [76]. In addition, the previous study revealed that the recognition and discrimination of the sugar moiety of the substrates is achieved by the side chains located on the opposite side of the P_i binding residues, such as His185 (Fig. 3.2b), which are not involved in the currently proposed transport mechanisms. The conformational transition is regulated by the binding of the phosphate moiety, and the binding manner of the phosphate moiety is the same in both the P_i and 3-PGA bound crystal structures [76] (Fig. 3.2). Therefore, it is highly likely that the transport mechanism of other substrates by GsGPT is similar to that of P_i observed in the current simulations. Moreover, given that the P_i binding residues observed in current simulation (Lys128, Lys204, Arg266, Lys271, Tyr339, Lys362 and Arg363) are well conserved among the pPT family members [76], and that the free energy landscape is governed almost entirely by the binding mode of P_i to these conserved residues, all of the transporters belonging to the pPT family may have similar free energy landscapes to that of GsGPT (Fig. 3.22). Thus, the results can be generalized to the transport mechanism for all pPT family transporters and substrates.

3.4.4 Comparison of mechanism with other transporter proteins

The previous research on the bacterial DMT superfamily protein, SnYddG, revealed that the best-characterized small drug resistance (SMR) family protein, *E. coli* EmrE [105,106], shares a similar TM topology with SnYddG, suggesting an evolutional relationship between SMR and other DMT proteins [78]. EmrE forms a dimer of four TM segments, and exchanges protons and cationic drugs [105, 106]. The structural comparison of EmrE with SnYddG and GsGPT revealed that EmrE lacks the TM helices corresponding to TM2 and TM7 in SnYddG and GsGPT [78]. The current simulation showed that these TM helices work as a scaffold, and are not directly involved in the conformational transition. Thus, the transport mechanism of EmrE can also be explained as the rocker-switch movement of a pair of three TM bundles (TM1, 2 and 3 of each protomer), corresponding to bundle1 and bundle2 of GsGPT (Fig. 3.13).

It is worth comparing the conformational regulation mechanism of GsGPT with that of GlpT(PDB ID:1PW4), which belongs to MFS and functions as a glycerol-3-phosphate/P_i exchanger [107, 108]. GlpT has a different evolutionary origin and protein folding from those of GsGPT, but is functionally similar to GsGPT. The previous biochemical and MD simulation studies [94, 109] revealed that the inter- and intra-domain salt bridges between two basic (R45 and K46) and two acidic (D274 and E299) residues stabilize the different conformations during the transport cycle, and substrate binding weakens these interactions, allowing the conformational transition. In the case of GsGPT, the current simulations revealed that the inter- and intra-helical salt bridges involving Glu207, which connect bundle1 and bundle2, stabilize the IF and OF conformations (Fig. 3.25), and these interactions are weakened by substrate binding. Furthermore, the overall conformational change is described by the rocker-switch motions of bundle1 and bundle2 (Fig. 3.13), which are widely observed in MFS transporters, including GlpT [107, 108]. Therefore, the conformational change mechanism by the rocker-switch motion and the conformational regulation mechanism by the salt bridge formation and disruption in GlpT and GsGPT are quite similar to each other. These similarities between GlpT and GsGPT over protein superfamily, in terms of both functions and mechanisms, may represent an example of the convergent evolution of the transporter mechanisms.

4 Concluding Remarks

In this study, MD simulations of a membrane channel, ChR, and an exchanger, TPT, to reveal the fundamental mechanisms required for channels and transporters in atomic detail.

In chapter 2, Glu129 is revealed to be important for preventing the proton leakage in the ground state, which is one of the basic mechanisms of membrane channels. This mechanism was revealed by observing the distribution of water molecules in the Glu129 deprotonated simulation, and could not be observed without MD simulation. The modeling and simulation with 13-*cis*R bound ChR revealed the conformational change at the early stage of channel opening. However, the channel opening enough to conduct the solvated cations cannot be observed in current simulations. Lots of experimental results show that it takes tens of microseconds to transition from ground state to conducting state (Fig. 2.2), which exceeds the timescales of the conventional MD simulation, and some enhanced sampling methods or simulation started with the initial structure of experimental structure in an intermediate state will be required. To perform these simulations, some experimental information of the conducting state of ChR are required, but these experimental results are still controversial. Further structural study for the intermediates of ChR is needed.

In chapter 3, the author performed the MD simulation of GsGPT. The combination of enhanced sampling methods enables to predict the feasible conformations from one crystal structure, and the feasibility of the sampled state transition can be evaluated by the reweighted free energy landscapes. The two fundamental mechanism as an exchanger have unraveled in atomic detail; the alternating access mechanism and the coupling mechanism between substrate binding and conformational change. The reconstructed transport cycle of GsGPT enables to easily compare the mechanisms of GsGPT with other transporters (section 3.4.4) and revealed the similar mechanisms between the evolutionary unrelated transporters. The reconstruction of functional reaction cycle (transport cycle, enzymatic cycle, etc) is important not only in a technical point but also for discussing or comparing the mechanism of the protein.

Secondary active transporters are attractive targets for the structural biology in the point of the strict coupling mechanisms between substrate binding and permitted conformational changes. GsGPT is a example of antiporters, in which the state transition in *apo* state is prohibited. In the case of symporters, however, the required mechanism seems to be more complicated than that of antiporters, as pointed in section 1.1.4: the state transition in *apo* and both substrate-bound state is permitted. To my knowledge, the enhanced sampling simulation for a symporter over its entire transport cycle is not performed. Similar approach as the current study will reveal the coupling mechanism between substrate binding and conformational change regulation for symporters in atomic detail, although there would be a difficulty in performing the substrate binding and releasing simulation since the co-substrate of the most symporters are proton.

MD simulation is a powerful tool to investigate the structural dynamics of proteins, which cannot be obtained from a single crystal structural analysis. A number of novel approaches for conformational sampling have been established along with the development of the computer technology. Along this evolution in MD simulation, recent innovation in structural analysis methods such as cryo-electron microscopy enables to reveal the novel structure of membrane protein. The combination of these structural analysis methods and MD simulations will reveal the novel dynamics and mechanisms of membrane proteins.

References

[1] O Jardetzky. Simple allosteric model for membrane pumps. *Nature*, 211(5052):969–970, 1966.

[2] David Drew and Olga Boudker. Shared Molecular Mechanisms of Membrane Transporters. *Annu. Rev. Biochem.*, 85:543–72, 2016.

[3] J. Andrew McCammon, Bruce R. Gelin, and Martin Karplus. Dynamics of folded proteins. *Nature*, 267(5612):585–590, jun 1977.

[4] Martin Karplus and Ja Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.*, 9(9):646–652, sep 2002.

[5] David E. Shaw, J.P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC14 Int. Conf. High Perform. Comput. Networking, Storage Anal.*, volume 2015-Janua, pages 41–53. IEEE, nov 2014.

[6] Tyler J. Harpole, Lucie Delemotte, B B A Biomembranes, Tyler J. Harpole, and Lucie Delemotte. Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. *Biochim. Biophys. Acta - Biomembr.*, (July):0–1, nov 2017.

[7] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, nov 1999.

[8] Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125(2):024106, 2006.

[9] G.M. M. Torrie and J.P. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187–199, feb 1977.

[10] Yuji Sugita, Akio Kitao, and Yuko Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113(15):6042, 2000.

[11] Mahmoud Moradi and Emad Tajkhorshid. Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems. *J. Chem. Theory Comput.*, 10(7):2866–2880, jul 2014.

[12] Georg Nagel, Doris Ollig, Markus Fuhrmann, Suneel Kateriya, Anna Maria Musti, Ernst Bamberg, and Peter Hegemann. Channelrhodopsin-1: a light-gated proton channel in green algae. *Science*, 296(5577):2395–2398, 2002.

[13] Edward S Boyden, Feng Zhang, Ernst Bamberg, Georg Nagel, and Karl Deisseroth. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.*, 8(9):1263–1268, sep 2005.

[14] Toru Ishizuka, Masaaki Kakuda, Rikita Araki, and Hiromu Yawo. Kinetic evaluation of photosensitivity in genetically engineered neurons expressing green algae light-gated channels. *Neurosci. Res.*, 54(2):85–94, 2006.

[15] Karl Deisseroth. Optogenetics. *Nat. Methods*, 8(1):26–29, 2011.

[16] Xue Han. In vivo application of optogenetics for neural circuit analysis, 2012.

[17] Hideaki E Kato, Feng Zhang, Ofer Yizhar, Charu Ramakrishnan, Tomohiro Nishizawa, Kunio Hirata, Jumpei Ito, Yusuke Aita, Tomoya Tsukazaki, Shigehiko Hayashi, Peter Hegemann, Andrés D Maturana, Ryuichiro Ishitani, Karl Deisseroth, and Osamu Nureki. Crystal structure of the channelrhodopsin light-gated cation channel. *Nature*, 482(7385):369–374, feb 2012.

[18] Christian Bamann, Taryn Kirsch, Georg Nagel, and Ernst Bamberg. Spectral Characteristics of the Photocycle of Channelrhodopsin-2 and Its Implication for Channel Function. *J. Mol. Biol.*, 375(3):686–694, 2008.

[19] Eglof Ritter, Katja Stehfest, Andre Berndt, Peter Hegemann, and Franz J Bartl. Monitoring light-induced structural changes of channelrhodopsin-2 by UV-visible and Fourier transform infrared spectroscopy. *J. Biol. Chem.*, 283(50):35033–35041, dec 2008.

[20] Christian Bamann, Ronnie Gueta, Sonja Kleinlogel, Georg Nagel, and Ernst Bamberg. Structural guidance of the photocycle of channelrhodopsin-2 by an interhelical hydrogen bond. *Biochemistry*, 49(2):267–278, 2010.

[21] Katja Stehfest and Peter Hegemann. Evolution of the channelrhodopsin photocycle model, 2010.

[22] Eglof Ritter, Patrick Piwowarski, Peter Hegemann, and Franz J. Bartl. Light-dark adaptation of channelrhodopsin C128T mutant. *J. Biol. Chem.*, 288(15):10451–10458, apr 2013.

[23] Katja Stehfest, Eglof Ritter, André Berndt, Franz Bartl, and Peter Hegemann. The branched photocycle of the slow-cycling channelrhodopsin-2 mutant C128T. *J. Mol. Biol.*, 398(5):690–702, 2010.

[24] Kirstin Eisenhauer, Jens Kuhne, Eglof Ritter, André Berndt, Steffen Wolf, Erik Freier, Franz Bartl, Peter Hegemann, and Klaus Gerwert. In channelrhodopsin-2 Glu-90 is crucial for ion selectivity and is deprotonated during the photocycle. *J. Biol. Chem.*, 287(9):6904–11, feb 2012.

[25] Anna Pia Plazzo, Nicola De Franceschi, Francesca Da Broi, Francesco Zonta, Maria Federica Sanasi, Francesco Filippini, and Marco Mongillo. Bioinformatic and mutational analysis of channelrhodopsin-2 protein cation-conducting pathway. *J. Biol. Chem.*, 287(7):4818–4825, 2012.

[26] Hiroshi C. Watanabe, Kai Welke, Daniel J. Sindhikara, Peter Hegemann, and Marcus Elstner. Towards an understanding of channelrhodopsin function: Simulations lead to novel insights of the channel mechanism. *J. Mol. Biol.*, 425(10):1795–1814, 2013.

[27] Saki Tanimoto, Yuka Sugiyama, Tetsuo Takahashi, Toru Ishizuka, and Hiromu Yawo. Involvement of glutamate 97 in ion influx through photo-activated channelrhodopsin-2. *Neurosci. Res.*, 75(1):13–22, jan 2013.

[28] Olga Gaiko and Robert E. Dempski. Transmembrane domain three contributes to the ion conductance pathway of channelrhodopsin-2. *Biophys. J.*, 104(6):1230–1237, 2013.

[29] Maria Müller, Christian Bamann, Ernst Bamberg, and Werner Kühlbrandt. Projection structure of channelrhodopsin-2 at 6 Å resolution by electron crystallography. *J. Mol. Biol.*, 414(1):86–95, 2011.

[30] Jan Hoffmann, Lubica Aslimovska, Christian Bamann, Clemens Glaubitz, Ernst Bamberg, and Bernd Brutschy. Studying the stoichiometries of membrane proteins by mass spectrometry: microbial rhodopsins and a potassium ion channel. *Phys. Chem. Chem. Phys.*, 12(14):3480–3485, apr 2010.

[31] a. Sali, L. Potterton, F. Yuan, H. Van Vlijmen, and M. Karplus. Evaluation of

comparative protein modeling by MODELLER. *Proteins Struct. Funct. Genet.*, 23(3):318–326, 1995.

[32] Delphine C. Bas, David M. Rogers, and Jan H. Jensen. Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins Struct. Funct. Genet.*, 73(April):765–783, 2008.

[33] Shota Ito, Hideaki E Kato, Reiya Taniguchi, Tatsuya Iwata, Osamu Nureki, and Hideki Kandori. Water-containing hydrogen-bonding network in the active center of channelrhodopsin. *J. Am. Chem. Soc.*, 136(9):3475–3482, mar 2014.

[34] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38, 1996.

[35] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins † . *J. Phys. Chem. B*, 102(18):3586–3616, apr 1998.

[36] Alexander D. Mackerell, Michael Feig, and Charles L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25(11):1400–1415, aug 2004.

[37] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and

M. Karplus. CHARMM: The biomolecular simulation program. *J. Comput. Chem.*, 30(10):1545–1614, 2009.

[38] Richard Neutze, Eva Pebay-Peyroula, Karl Edman, Antoine Royant, Javier Navarro, and Ehud M. Landau. Bacteriorhodopsin: A high-resolution structural view of vectorial proton transport, 2002.

[39] S Subramaniam, M Gerstein, D Oesterhelt, and R Henderson. Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.*, 12(1):1–8, jan 1993.

[40] S Subramaniam and R Henderson. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature*, 406(6796):653–657, 2000.

[41] Ivan Gushchin, Anastasia Reshetnyak, Valentin Borshchevskiy, Andrii Ishchenko, Ekaterina Round, Sergei Grudinin, Martin Engelhard, Georg Büldt, and Valentin Gordeliy. Active state of sensory rhodopsin II: Structural determinants for signal transfer and proton pumping. *J. Mol. Biol.*, 412(4):591–600, sep 2011.

[42] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–802, dec 2005.

[43] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 103(11):4613, 1995.

[44] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089, 1993.

[45] Yasuhiro Matsui, Keisuke Sakai, Midori Murakami, Yoshitsugu Shiro, Shin ichi Adachi, Hideo Okumura, and Tsutomu Kouyama. Specific damage induced by

X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.*, 324(3):469–481, nov 2002.

[46] Naveen Michaud-Agrawal, Elizabeth J Denning, Thomas B Woolf, and Oliver Beckstein. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, 32(10):2319–27, jul 2011.

[47] Víctor A Lórenz-Fonfría, Tom Resler, Nils Krause, Melanie Nack, Michael Gossing, Gabriele Fischer von Mollard, Christian Bamann, Ernst Bamberg, Ramona Schlesinger, and Joachim Heberle. Transient protonation changes in channelrhodopsin-2 and their relevance to channel gating. *Proc. Natl. Acad. Sci. U. S. A.*, 110(14):E1273–81, 2013.

[48] Hideaki E. Kato, Motoshi Kamiya, Seiya Sugo, Jumpei Ito, Reiya Taniguchi, Ayaka Orito, Kunio Hirata, Ayumu Inutsuka, Akihiro Yamanaka, Andrés D. Maturana, Ryuichiro Ishitani, Yuki Sudo, Shigehiko Hayashi, and Osamu Nureki. Atomistic design of microbial opsin-based blue-shifted optogenetics tools. *Nat. Commun.*, 6(May):7177, may 2015.

[49] Maria Müller, Christian Bamann, Ernst Bamberg, and Werner Kühlbrandt. Light-Induced Helix Movements in Channelrhodopsin-2. *J. Mol. Biol.*, 427(2):341–349, 2015.

[50] Thomas Sattig, Christian Rickert, Ernst Bamberg, Heinz-Jürgen Jürgen Steinhoff, and Christian Bamann. Light-induced movement of the transmembrane Helix B in Channelrhodopsin-2. *Angew. Chemie - Int. Ed.*, 52(37):9705–9708, sep 2013.

[51] Nils Krause, Christopher Engelhard, Joachim Heberle, Ramona Schlesinger, and Robert Bittl. Structural differences between the closed and open states of channelrhodopsin-2 as observed by EPR spectroscopy. *FEBS Lett.*, 587(20):3309–3313, oct 2013.

[52] Jens Kuhne, Kirstin Eisenhauer, Eglof Ritter, Peter Hegemann, Klaus Gerwert, and

Franz Bartl. Early Formation of the Ion-Conducting Pore in Channelrhodopsin-2. *Angew. Chemie Int. Ed.*, 54(16):4953–4957, 2015.

[53] Karelia Ruffert, Bettina Himmel, Deepti Lall, Christian Bamann, Ernst Bamberg, Heinrich Betz, and Volker Eulenburg. Glutamate residue 90 in the predicted transmembrane domain 2 is crucial for cation flux through channelrhodopsin 2. *Biochem. Biophys. Res. Commun.*, 410(4):737–743, jul 2011.

[54] Asumi Inaguma, Hisao Tsukamoto, Hideaki E. Kato, Tetsunari Kimura, Toru Ishizuka, Satomi Oishi, Hiromu Yawo, Osamu Nureki, and Yuji Furutani. Chimeras of Channelrhodopsin-1 and -2 from Chlamydomonas reinhardtii Exhibit Distinctive Light-induced Structural Changes from Channelrhodopsin-2. *J. Biol. Chem.*, page jbc.M115.642256, 2015.

[55] Jonas Wietek, J Simon Wiegert, Nona Adeishvili, Franziska Schneider, Hiroshi Watanabe, Satoshi P Tsunoda, Arend Vogt, Marcus Elstner, Thomas G Oertner, and Peter Hegemann. Conversion of channelrhodopsin into a light-gated chloride channel. *Science*, 344(6182):409–12, 2014.

[56] Andre Berndt, Soo Yeun Lee, Charu Ramakrishnan, and Karl Deisseroth. Structure-guided transformation of channelrhodopsin into a light-activated chloride channel. *Science*, 344(6182):420–4, 2014.

[57] Oleksandr Volkov, Kirill Kovalev, Vitaly Polovinkin, Valentin Borshchevskiy, Christian Bamann, Roman Astashkin, Egor Marin, Alexander Popov, Taras Balandin, Dieter Willbold, Georg Büldt, Ernst Bamberg, and Valentin Gordeliy. Structural insights into ion conduction by channelrhodopsin 2. *Science (80-.).*, 358(6366):eaan8862, nov 2017.

[58] Ionela Radu, Christian Bamann, Melanie Nack, Georg Nagel, Ernst Bamberg, and Joachim Heberle. Conformational changes of channelrhodopsin-2. *J. Am. Chem. Soc.*, 131(21):7313–7319, 2009.

[59] Andreas P M Weber and Nicole Linka. Connecting the plastid: transporters of the plastid envelope and their role in linking plastidial with cytosolic metabolism. *Annu. Rev. Plant Biol.*, 62:53–77, 2011.

[60] Silke Knappe, Ulf-Ingo Flügge, and Karsten Fischer. Analysis of the plastidic phosphate translocator gene family in Arabidopsis and identification of new phosphate translocator-homologous transporters, classified by their putative substrate-binding site. *Plant Physiol.*, 131(3):1178–1190, 2003.

[61] Birgit Kammerer, Karsten Fischer, Bettina Hilpert, Sabine Schubert, Michael Gutensohn, Andreas Weber, and Ulf-ingo I Flügge. Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. *Plant Cell*, 10(1):105–17, jan 1998.

[62] Michael Eicks, Verónica Maurino, Silke Knappe, Ulf-Ingo Flügge, and Karsten Fischer. The plastidic pentose phosphate translocator represents a link between the cytosolic and the plastidic pentose phosphate pathways in plants. *Plant Physiol.*, 128(2):512–22, feb 2002.

[63] K. Fischer, B Kammerer, M Gutensohn, B Arbinger, A Weber, R E Häusler, and U I Flügge. A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. *Plant Cell*, 9(3):453–62, mar 1997.

[64] U I I Flügge, Karsten Fischer, Armin Gross, Walter Sebald, Fritz Lottspeich, and Christoph Eckerskorn. The triose phosphate-3-phosphoglycerate-phosphate translocator from spinach chloroplasts: nucleotide sequence of a full-length cDNA clone and import of the in vitro synthesized precursor protein into chloroplasts. *EMBO J.*, 8(1):39–46, jan 1989.

[65] Ulf-Ingo Flügge, Andreas Weber, Karsten Fischer, Fritz Lottspeich, Christoph Eckerskorn, Karin Waegemann, and Jürgen Soll. The major chloroplast envelope

polypeptide is the phosphate translocator and not the protein import receptor. *Nature*, 353(6342):364–367, sep 1991.

[66] Jorg W Riesmeier, U I Flügge, Burkhard Schulz, Dieter Heineke, H W Heldt, Lothar Willmitzer, and Wolf B Frommer. Antisense repression of the chloroplast triose phosphate translocator affects carbon partitioning in transgenic potato plants. *Proc. Natl. Acad. Sci. U. S. A.*, 90(13):6160–4, jul 1993.

[67] Dieter Heineke, Anne Kruse, Ulf-Ingo Ingo Flügge, WolfB. B. Frommer, JörgW. W. Riesmeier, Lothar Willmitzer, and HansW. W. Heldt. Effect of antisense repression of the chloroplast triose-phosphate translocator on photosynthetic metabolism in transgenic potato plants. *Planta*, 193(2):174–180, mar 1994.

[68] R E Häusler, N H Schlieben, and U I Flügge. Control of carbon partitioning and photosynthesis by the triose phosphate/phosphate translocator in transgenic tobacco plants (Nicotiana tabacum). II. Assessment of control coefficients of the triose phosphate/phosphate translocator. *Planta*, 210(3):383–90, feb 2000.

[69] R E Häusler, N H Schlieben, P Nicolay, K L Fischer, K L Fischer, and U I Flügge. Control of carbon partitioning and photosynthesis by the triose phosphate/phosphate translocator in transgenic tobacco plants (Nicotiana tabacum L.). I. Comparative physiological analysis of tobacco plants with antisense repression and overexpression of t. *Planta*, 210(3):371–82, feb 2000.

[70] Andrea Hattenbach, B. Muller-Rober, G. Nast, and Dieter Heineke. Antisense Repression of Both ADP-Glucose Pyrophosphorylase and Triose Phosphate Translocator Modifies Carbohydrate Partitioning in Potato Leaves. *Plant Physiol.*, 115(2):471–475, oct 1997.

[71] Anja Schneider, Rainer E. Häusler, Uner Kolukisaoglu, Reinhard Kunze, Eric van der Graaff, Rainer Schwacke, Elisabetta Catoni, Marcelo Desimone, and Ulf-Ingo Flügge. An Arabidopsis thaliana knock-out mutant of the chloroplast triose

phosphate/phosphate translocator is severely compromised only when starch synthesis, but not starch mobilisation is abolished. *Plant J.*, 32(5):685–99, dec 2002.

[72] Andreas P.M. Weber, Rainer Schwacke, and Ulf-Ingo Flügge. Solute Transporters of the Plastid Envelope Membrane. *Annu. Rev. Plant Biol.*, 56(1):133–164, 2005.

[73] Rainer Fliege, Ulf Ingo Flügge, Karl Werdan, and Hans W. Heldt. Specific transport of inorganic phosphate, 3-phosphoglycerate and triosephosphates across the inner membrane of the envelope in spinach chloroplasts. *BBA - Bioenerg.*, 502(2):232–247, may 1978.

[74] Ulf-Ingo Flugge. PHOSPHATE TRANSLOCATORS IN PLASTIDS. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, 50:27–45, jun 1999.

[75] Marc Linka, Aziz Jamai, and Andreas P M Weber. Functional characterization of the plastidic phosphate translocator gene family from the thermo-acidophilic red alga Galdieria sulphuraria reveals specific adaptations of primary carbon partitioning in green plants and red algae. *Plant Physiol.*, 148(3):1487–96, 2008.

[76] Yongchan Lee, Tomohiro Nishizawa, Mizuki Takemoto, Kaoru Kumazaki, Keitaro Yamashita, Kunio Hirata, Ayumi Minoda, Satoru Nagatoishi, Kouhei Tsumoto, Ryuichiro Ishitani, and Osamu Nureki. Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat. Plants*, 3(10):825–832, oct 2017.

[77] Donald L Jack, Nelson M Yang, and Milton H. Saier. The drug/metabolite transporter superfamily. *Eur. J. Biochem.*, 268(13):3620–3639, jul 2001.

[78] Hirotoshi Tsuchiya, Shintaro Doki, Mizuki Takemoto, Tatsuya Ikuta, Takashi Higuchi, Keita Fukui, Yoshihiro Usuda, Eri Tabuchi, Satoru Nagatoishi, Kouhei Tsumoto, Tomohiro Nishizawa, Koichi Ito, Naoshi Dohmae, Ryuichiro Ishitani, and Osamu Nureki. Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature*, 534(7607):417–420, may 2016.

[79] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72(4):1568–1581, apr 1997.

[80] Esben M Quistgaard, Christian Löw, Fatma Guettou, and Pär Nordlund. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Publ. Gr.*, pages 1–10, 2016.

[81] P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(4):486–501, apr 2010.

[82] Matti Javanainen. Universal method for embedding proteins into complex lipid bilayers for molecular dynamics simulations. *J. Chem. Theory Comput.*, 10(6):2577–2582, 2014.

[83] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, oct 1984.

[84] Jeffery B Klauda, Richard M Venable, J Alfredo Freites, Joseph W. O'Connor, Douglas J Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D. MacKerell, and Richard W Pastor. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B*, 114(23):7830–7843, jun 2010.

[85] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. GRGMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[86] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1), 2007.

[87] Shuichi Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.

[88] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[89] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190, 1981.

[90] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18(12):1463–1472, 1997.

[91] Giacomo Fiorin, Michael L. Klein, and Jérôme Hénin. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.*, 111(22-23):3345–3362, 2013.

[92] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E M Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ 1 and χ 2 Dihedral Angles. *J. Chem. Theory Comput.*, 8(9):3257–3273, 2012.

[93] Albert C. Pan, Deniz Sezer, and Benoît Roux. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B*, 112(11):3432–3440, 2008.

[94] Mahmoud Moradi, Giray Enkavi, and Emad Tajkhorshid. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nat. Commun.*, 6:8393, 2015.

[95] Davide Branduardi, Francesco Luigi Gervasio, and Michele Parrinello. From A to B in free energy space. *J. Chem. Phys.*, 126(5):054103, feb 2007.

[96] Bernd Ensing, Alessandro Laio, Michele Parrinello, and Michael L. Klein. A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J. Phys. Chem. B*, 109(14):6676–6687, 2005.

[97] Robert T. McGibbon, Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee-Ping Wang, Thomas J. Lane, and Vijay S. Pande. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.*, 109(8):1528–1532, 2015.

[98] Oliver S. Smart, Joseph G. Neduvelil, Xiaonan Wang, B. A. Wallace, and Mark S P Sansom. HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.*, 14(6):354–360, 1996.

[99] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[100] Michael Waskom, Olga Botvinnik, Drewokane, Paul Hobson, David, Yaroslav Halchenko, Saulius Lukauskas, John B Cole, Jordi Warmenhoven, Julian de Ruiter, Stephan Hoyer, Jake Vanderplas, Santi Villalba, Gero Kunter, Eric Quintero, Marcel Martin, Alistair Miles, Kyle Meyer, Tom Augspurger, Tal Yarkoni, Pete Bachant, Mike Williams, Constantine Evans, Clark Fitzgerald, Brian, Daniel Wehner, Gregory Hitz, Erik Ziegler, Adel Qalieh, and Antony Lee. seaborn: v0.7.1 (June 2016), 2016.

[101] John D. Hunter. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, 9(3):90–95, 2007.

[102] Shankar Kumar, John M Rosenberg, Djamal Bouzida, Robert H Swendsen, and Peter A Kollman. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13(8):1011–1021, oct 1992.

[103] Christian Bartels. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.*, 331(5-6):446–454, 2000.

[104] Carrie F. Brooks, Hanne Johnsen, Giel G. van Dooren, Mani Muthalagi, San San

Lin, Wolfgang Bohne, Karsten Fischer, and Boris Striepen. The Toxoplasma Apicoplast Phosphate Translocator Links Cytosolic and Apicoplast Metabolism and Is Essential for Parasite Survival. *Cell Host Microbe*, 7(1):62–73, 2010.

[105] Y.-J. Chen, Owen Pornillos, Samantha Lieu, Che Ma, Andy P Chen, and Geoffrey Chang. X-ray structure of EmrE supports dual topology model. *Proc. Natl. Acad. Sci.*, 104(48):18999–19004, nov 2007.

[106] Katherine Henzler-Wildman. Analyzing conformational changes in the transport cycle of EmrE. *Curr. Opin. Struct. Biol.*, 22(1):38–43, 2012.

[107] Yafei Huang, M Joanne Lemieux, Jinmei Song, Manfred Auer, and Da-Neng Wang. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science*, 301(5633):616–20, aug 2003.

[108] Christopher J. Law, Peter C. Maloney, and Da-Neng Wang. Ins and Outs of Major Facilitator Superfamily Antiporters. *Annu. Rev. Microbiol.*, 62(1):289–305, 2008.

[109] Christopher J. Law, Jonas Almqvist, Adam Bernstein, Regina M. Goetz, Yafei Huang, Celine Soudant, Aatto Laaksonen, Sven Hovmöller, and Da Neng Wang. Salt-bridge Dynamics Control Substrate-induced Conformational Change in the Membrane Transporter GlpT. *J. Mol. Biol.*, 378(4):826–837, 2008.

Original papers

1. Kumazaki, K., Chiba, S., Takemoto, M., Furukawa, A., Nishiyama, K., Sugano, Y., Mori, T., Dohmae, N., Hirata, K., Nakada-Nakura, Y., Maturana, A.D., Tanaka, Y., Mori, H., Sugita, Y., Arisaka, F., Ito, K., Ishitani, R., Tsukazaki, T., and Nureki, O. (2014). Structural basis of Sec-independent membrane protein insertion by YidC. *Nature* **509**, 516-520.

2. Kato, H.E., Inoue, K., Abe-Yoshizumi, R., Kato, Y., Ono, H., Konno, M., Hososhima, S., Ishizuka, T., Hoque, M.R., Kunitomo, H., Ito, J., Yoshizawa, S., Yamashita, K., Takemoto, M., Nishizawa, T., Taniguchi, R., Kogure, K., Maturana, A.D., Iino, Y., Yawo, H., Ishitani, R., Kandori, H., and Nureki, O.. (2015). Structural basis for Na^+ transport mechanism by a light-driven Na^+ pump. *Nature* **521**, 48-53.

3. Takemoto, M., Kato, H.E., Koyama, M., Ito, J., Kamiya, M., Hayashi, S., Maturana, A.D., Deisseroth, K., Ishitani, R., and Nureki, O. (2015). Molecular Dynamics of Channelrhodopsin at the Early Stages of Channel Opening. *PLOS ONE* **10**, e0131094.

4. Tanaka, Y., Sugano, Y., Takemoto, M., Mori, T., Furukawa, A., Kusakizako, T., Kumazaki, K., Kashima, A., Ishitani, R., Sugita, Y., Nureki, O., and Tsukazaki, T. (2015). Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. *Cell Rep.* **13**, 1561-1568.

5. Tsuchiya, H., Doki, S., Takemoto, M., Ikuta, T., Higuchi, T., Fukui, K., Usuda, Y., Tabuchi, E., Nagatoishi, S., Tsumoto, K., Nishizawa, T., Ito, K., Dohmae, N.,

Ishitani, R., and Nureki, O. (2016). Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature* **534**, 417-420.

6. Kasuya, G., Fujiwara, Y., Takemoto, M., Dohmae, N., Nakada-Nakura, Y., Ishitani, R., Hattori, M., and Nureki, O. (2016). Structural Insights into Divalent Cation Modulations of ATP-Gated P2X Receptor Channels. *Cell Rep.* **14**, 932-944.

7. Kasuya, G., Hiraizumi, M., Maturana, A.D., Kumazaki, K., Fujiwara, Y., Liu, K., Nakada-Nakura, Y., Iwata, S., Tsukada, K., Komori, T., Uemura, S., Goto, Y., Nakane, T., Takemoto, M., Kato, H.E., Yamashita, K., Wada, M., Ito, K., Ishitani, R., Hattori, M., and Nureki, O. (2016). Crystal structures of the TRIC trimeric intracellular cation channel orthologues. *Cell Res.* **26**, 1288-1301.

8. Kasuya, G., Yamaura, T., Ma, X.-B., Nakamura, R., Takemoto, M., Nagumo, H., Tanaka, E., Dohmae, N., Nakane, T., Yu, Y., Ishitani, R., Matsuzaki, O., Hattori, M., and Nureki, O. (2017). Structural insights into the competitive inhibition of the ATP-gated P2X receptor channel. *Nat. Commun.* **8**, 876.

9. Lee, Y., Nishizawa, T., Takemoto, M., Kumazaki, K., Yamashita, K., Hirata, K., Minoda, A., Nagatoishi, S., Tsumoto, K., Ishitani, R., and Nureki, O. (2017). Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat. Plants* **3**, 825-832.

10. Takemoto, M., Lee, Y., Ishitani, R., and Nureki, O. (2017). Free energy landscape for the entire transport cycle of triose-phosphate/phosphate translocator. *bioRxiv*

Acknowledgements

First, I would like to thank Professor Osamu Nureki for his supervision and encouragement during the bachelor's, master's and Ph.D. courses. Besides Prof. Nureki, I have been assisted by many members of the Nureki laboratory. In particular, I would like to thank Assoc. prof. Ryuichiro Ishitani for guidance and assistance with performing simulations and writing papers, and Mr. Yongchan Lee for performing the liposome transport assay of GsGPT and discussing for the results of MD simulations. I also would like to thank Dr. Tomohiro Nishizawa, Dr. Takanori Nakane, Dr. Michio Koyama and Dr. Hideaki Kato for technical advice and experimental support, and Ms. Rieko Yamazaki for secretarial assistance.

This study was performed in collaboration with many scientists. I would like to thank Assoc. prof. Andrés D. Maturana (Nagoya University) and Dr. Jumpei Ito (King's College London) for the collaboration with the electrophysiological analysis of channelrhodopsin, Prof. Shigehiko Hayashi (Kyoto University) and Dr. Motoshi Kamiya (RIKEN) for the collaboration with the development of the topology and force field of the all-*trans* and 13-*cis* retinal and Prof. Karl Deisseroth for helpful comments on the MD simulation study of channelrhodopsin. I also thank Dr. Mahmoud Moradi (University of Arkansas) for kindly providing the *post-hoc* string method program and Dr. Yuji Sugita (RIKEN) for technical advice throughout all of the simulations.

The simulations of channelrhodopsin (chapter 2) were conducted using the RIKEN Integrated Cluster of Clusters (RICC) facility, the Fujitsu PRIMEHPC FX10 System (Oakleaf-FX) in the Information Technology Center, The University of Tokyo, and the K computer provided by the RIKEN Advanced Institute for Computational Science, through the HPCI System Research Project (Project ID: hp120309). The simulations of triose-phosphate/phosphate translocator (chapter 3) were partially performed on HOKUSAI Great-Wave at the RIKEN Advanced Center for Computing and Communication, the mini-K supercomputer system at the SACLA facility, and the NIG supercomputer at ROIS National Institute of Genetics. Without these computational resources, the MD simulations in this study could not be performed.

During the bachelor's, master's and Ph.D. courses, I enjoyed many exciting experiences with the members of the Nureki laboratory. I would like to thank all members of the Nureki laboratory for stimulating discussion for the research. In particular, I thank Mr. Reiya Taniguchi and Mr. Go Kasuya as supportive labmates since the undergraduate course.

Finally, I would like to thank my family for continuous support throughout my primary education, the bachelor's, master's and Ph.D. courses, and my life.

Mizuki Takemoto