

Investigating the Function of a Pentameric Ligand-Gated Ion Channel with Molecular Simulation



Marc André Dämgen
Wolfson College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2019

Acknowledgements

Firstly, I would like to express my gratitude to my supervisor Professor Phil Biggin for his advise and support. I was given an unusual degree of autonomy, which allowed me to be guided by my scientific curiosity and explore a diverse array of academic opportunities during my PhD, which certainly aided my growth as a scientist.

I consider myself fortunate for the opportunity to carry out my graduate research in the SBCB with its extremely talented members from so many diverse backgrounds. This thesis has greatly benefited from inspiration and interaction with these exceptional colleagues. My thanks goes to all past and present members, and more than a few deserve explicit mentioning. Lucy Ladefoged, Maria Musgaard and Ole Andersen were always available for questions when I started out in the field of biomolecular simulation. Some of the analysis performed in Chapter 3 on the open state of the glycine receptor was aided by helpful discussions with Shanlin Rao, Hannah Smith and Gianni Klesse, who has developed the wonderful software CHAP. Research on protein-lipids interactions with coarse-grained MD simulations by members of the Sansom group has inspired the work in Chapter 4 and I owe my thanks to Wanling Song and Anna Duncan for constructive feedback during the project and comments on the chapter. Further, I would like thank Matteo Aldeghi, Irfan Alibi and Joe Bluck for discussions around free energy calculations which added value to Chapter 5. Irfan deserves a special thanks for comments on Chapter 5, for dealing with IT-related issues in the SBCB as well as his general helpful attitude. Last but not least, I would like to thank Nicholas Michelarakis and Teresa Paramo for a lot of fun in and outside of the lab. I would also like to express my gratitude towards our colleagues at UCL for a fruitful collaboration and stimulating discussions, particularly Lucia Sivilotti and Remigijus Lape.

This thesis would not have been possible without the financial support from the Oxford-Wolfson-Marriott-Graduate Scholarship in Biochemistry, the German National Academic Foundation (Studienstiftung des deutschen Volkes), the Goodger and Schorstein Research Scholarship in Medical Sciences, and the Vice Chancellors' Fund and they are all gratefully acknowledged.

My time in Oxford has been enriched by the companionship of great friends. Cami has been one of the first people I have met in Oxford and is the best friend one could ask for. Javi became also an incredible friend. I will never forget the

magical moments of playing music together and even having had some stage time with Felipe and Andrius during the open mic nights organized by Pablo, who all have become dear friends whom I shall miss. A special thanks goes to Claire and Rahel for their friendship. I would also like to thank my friends outside of Oxford, particularly Marco, Ortwin, Marc, Imke and Sofia. Words cannot express how thankful I am for the time spent with my girlfriend Emily and for all her support.

Finally, I would like to thank my family, particularly my mother Gudrun and my sister Yvonne for their love, support and encouragement.

In grateful memory of my father Michael.

Abstract

Pentameric ligand-gated ion channels (pLGICs) are key players in mediating fast neurotransmission. Upon neurotransmitter binding, they undergo conformational changes that lead to the opening of the channel pore, allowing ions to flow in or out of the cell. They are involved in a range of neurological diseases which makes them important drug targets. The glycine receptor (GlyR) is a chloride-selective member of this receptor family, which is an excellent model system, because structures in key functional states are available and it has been extensively characterized experimentally. This thesis therefore seeks to investigate the function of the GlyR as a representative of the pLGIC superfamily with molecular dynamics (MD) simulations with four focal points.

1. Recent work has questioned the physiological relevance of the open state structure of the GlyR, since it rapidly collapses in MD simulations. I show that the collapse can be avoided by a careful equilibration protocol that reconciles the more problematic regions of the original density map and gives a stable open state that shows frequent selective chloride permeation. The stability can be explained by the 9' leucine gate residues being locked into so far unreported hydrophobic pockets that are functionally conserved across pLGICs.

2. pLGICs are known to be sensitive to their lipid environment, yet it is still poorly understood how lipids modulate receptor function. By embedding this receptor in a realistic model of a neuronal membrane and performing coarse-grained molecular dynamics simulations, I find that the lipid environment as well as the lipid-protein interaction strength are dependent on the receptor state, suggesting that lipids play a crucial role in stabilizing either the active or inactive state. Cholesterol can bind to the binding site of the positive allosteric modulator ivermectin, indicating that intrinsic membrane lipids might have positive modulatory function.

3. A molecular level of understanding of partial agonism in pLGICs has remained elusive thus far. Using atomistic MD simulations, I characterize the binding modes and determine the binding free energies of the full agonist glycine and several partial agonists. All agonists stabilize loop C, which covers the binding site like a lid, in a closed conformation, but the interactions in the binding site are less optimal for partial agonists than for glycine.

4. The atomistic details of the gating mechanism of pLGICs are still not fully understood. Via a combination of targeted MD and Markov state modelling, I reconstruct the equilibrium distribution and underlying kinetics of the gating process. The analysis reveals that the closed-to-open transition involves an intermediate shut state.

Contents

List of Figures	xi
List of Tables	xiii
List of Abbreviations	xv
1 Introduction	1
1.1 Neuronal Activity	2
1.2 Synaptic Transmission	3
1.2.1 Electrical Synapses	4
1.2.2 Chemical Synapses	4
1.2.3 Signal Integration	6
1.3 Pentameric Ligand-Gated Ion Channels	7
1.3.1 Eukaryotic pLGICs	8
1.3.2 Prokaryotic pLGICs	11
1.3.3 Conserved Overall Architecture	11
1.3.4 The Activation Mechanism - Insights from Single Channel Recordings	20
1.3.5 Partial Agonism	23
1.3.6 Allosteric Modulation	25
1.4 Thesis Organization and Motivation	26
2 Theory and Methods	29
2.1 The Physical Basis for Molecular Simulation	30
2.1.1 Key Concepts of Classical Mechanics	30
2.1.2 Key Concepts of Statistical Mechanics	36
2.2 Molecular Dynamics	38
2.2.1 Integration of the Equations of Motion	39
2.2.2 Potential Energy Models (Force Fields)	42
2.2.3 Periodic Boundary Conditions and Box Shapes	48
2.2.4 Common Speed up Techniques	50
2.2.5 Controlling Temperature and Pressure: Thermostats and Barostats	55

2.2.6	The Practical Workflow of a MD Simulation	59
2.3	Free Energy Calculations	61
2.3.1	The Alchemical Cycle	63
2.3.2	Bennett's Acceptance Ratio	66
2.3.3	Multistate Bennett's Acceptance Ratio	66
2.4	Markov State Modelling	67
2.4.1	Definition of a Markov State Model	67
2.4.2	Necessary Properties of the Underlying Dynamical Process .	67
2.4.3	Time Evolution of a Markov State Model	68
2.4.4	Analyzing Markov State Models	70
3	A Refined Open State of the Glycine Receptor	71
3.1	Introduction	72
3.2	Methods	75
3.2.1	Homology Modelling	75
3.2.2	System Setup	75
3.2.3	Simulation Protocol	76
3.3	Results	80
3.3.1	The MD-Refined Structure is Physically Open, Hydrated and Selectively Permeant to Chloride Ions	80
3.3.2	The Open Conformation is Stabilized by the Leucine Gate Residues Filling a Hydrophobic Void at the Subunit Interface	85
3.3.3	The Cryo-EM Based Structure Collapses due to a Sub-Optimal Orientation of the Leucine Gate Residues	89
3.4	Discussion	91
3.5	Conclusions	99
4	Protein-Lipid Interactions in a Neuronal Membrane	101
4.1	Introduction	102
4.2	Methods	106
4.2.1	Homology Modelling	106
4.2.2	System Set Up	107
4.2.3	Simulation Protocol	108
4.2.4	Analysis and Visualization	108
4.3	Results	108
4.3.1	The Lipid Environment is Dependent on the Receptor State	108
4.3.2	The Duration of Protein-Lipid Contacts is Dependent on the Receptor State	112
4.3.3	Membrane Lipids Can Interact with the Extracellular Domain of the Receptor	115

4.3.4	Cholesterol Can Bind to the Binding Site of the Positive Allosteric Modulator Ivermectin	123
4.3.5	The Binding Mode of the Anionic Lipid DOPS in GlyR Differs from the PIP2 Binding Mode in the GABA _A R	126
4.4	Discussion	128
4.5	Conclusions	132
5	Agonist Binding in the Glycine Receptor	135
5.1	Introduction	135
5.2	Methods	138
5.2.1	Standard MD Simulations	138
5.2.2	Free Energy Calculations	143
5.3	Results	145
5.3.1	Binding Modes of Agonists	145
5.3.2	Agonist Influence on Receptor Dynamics	160
5.3.3	Loop C Dynamics	161
5.3.4	Binding Free Energies	167
5.4	Discussion	171
5.5	Conclusions	175
6	Towards an Atomistic Gating Mechanism	179
6.1	Introduction	180
6.2	Methods	182
6.2.1	Targeted MD	182
6.2.2	Unbiased MD along the Gating Path	183
6.2.3	MSM Estimation and Validation	184
6.2.4	Analysis	186
6.3	Results	187
6.3.1	Targeted MD Generates an Initial Guess for a Gating Path .	187
6.3.2	Insights into the Gating Mechanism from Markov State Modelling	192
6.4	Discussion	196
6.5	Conclusions	198
7	Conclusions	199
7.1	A Refined Open State of the Glycine Receptor	200
7.1.1	Contributions	200
7.1.2	Impact	201
7.1.3	The MD-refined Open State in Light of New Structures . . .	202
7.2	Protein-Lipid Interactions in a Neuronal Membrane	203

7.2.1	Contributions	203
7.2.2	Impact	204
7.3	Agonist Binding in the Glycine Receptor	204
7.3.1	Contributions	204
7.3.2	Impact	205
7.4	Towards an Atomistic Gating Mechanism	206
7.4.1	Contributions	206
7.4.2	Impact	207
References		209

List of Figures

1.1	Schematic overview of a neuron	3
1.2	Schematic overview of a chemical synapse	5
1.3	Signal integration	7
1.4	Overall architecture of the pLGIC superfamily	12
1.5	Overall pLGIC architecture exemplified by the GlyR	13
1.6	Orthosteric binding site	14
1.7	Glycine binding mode in PDB: 5TIN	15
1.8	Ion channel pore	17
1.9	ECD-TMD interface	19
1.10	Flip mechanism	21
1.11	Partial agonism	24
1.12	Allosteric sites	25
2.1	Typical functional form of a force field	43
2.2	Periodic boundary conditions	49
2.3	Optimal box geometry	50
2.4	Alchemical cycle	64
3.1	Simulation system: GlyR in POPC membrane	76
3.2	Pore restraints	78
3.3	Pore profile of GlyR structures	81
3.4	Pore profiles from three independent simulations with a stable open state	82
3.5	Ion permeation in simulations with a stable open state	84
3.6	Critical changes in pore geometry	86
3.7	Functional conservation of hydrophobic pockets in pLGIC superfamily	87
3.8	Details of the hydrophobic pocket and orientation of T6'	88
3.9	Structural features of the collapsed state	90
3.10	Pore profile of a hydrophobic collapsed state	91
3.11	Pore profiles from three independent simulations with a collapsed state	92
3.12	Ion permeation in simulations with a stable open state	93
3.13	L9' gate overlay with PDB: 3EAM	94

3.14	L9' gate overlay with PDB: 6HIN	96
3.15	Cryo-EM density map of PDB: 3JAE	97
3.16	RMSF near P-2' gate	98
4.1	CG simulation system: GlyR in neuronal membrane	105
4.2	2-d lipid densities	109
4.3	3-d lipid densities	111
4.4	Protein-lipid contact times per residue colour-coded on receptor . .	113
4.5	Protein-lipid contact times per residue bar plot	114
4.6	Phospholipid interactions at the ECD-TMD interface	116
4.7	Protein-lipid contact times at the ECD-TMD interface	118
4.8	Comparison of phospholipid density with experimental structures .	121
4.9	Overlay of cholesterol density with ivermectin	124
4.10	Comparison of DOPS density with PIP2 in PDB: 6I53	127
4.11	Aromatic residues at the M4-M1/M3 interface of different pLGICs .	130
5.1	Agonist molecules of the GlyR	137
5.2	Simulation system: ECD with positional restraints	139
5.3	Binding mode of glycine	146
5.4	Densities of characteristic agonist atoms	151
5.5	Probability densities of selected key distances in the orthosteric binding site	153
5.6	Binding mode of D-alanine	156
5.7	Binding mode of sarcosine	157
5.8	Binding mode of β -alanine	159
5.9	Binding mode of taurine	160
5.10	C $_{\alpha}$ -RMSF of the ECD from apo and holo simulations	162
5.11	Degree of loop C closure and its dynamic flexibility from apo and holo simulations	164
5.12	Mechanism of loop C opening/closing	166
5.13	Partial charges in the GAFF2 and Horn force fields	176
6.1	Dimensionality reduction with TICA	185
6.2	MSM construction	185
6.3	Chapman-Kolmogorow test	186
6.4	Illustration of gating CVs	187
6.5	Normalized gating CVs for all force constants	189
6.6	RMSD and gating CVs for force constant $\kappa=200,000 \text{ kJmol}^{-1}\text{nm}^{-2}$	191
6.7	Free energy surface	193
6.8	MSM eigenvectors	194
6.9	A 3-state gating model	195

List of Tables

4.1	Composition of a human neuronal model membrane	107
5.1	Overview of unbinding events at the different subunit interfaces . .	147
5.2	Key distances to characterize the binding mode in the orthosteric binding site	154
5.3	Results of free energy calculations	168
5.4	Breakdown of free energy calculations	169
6.1	Relative occupancies of closed, intermediate-shut (flip) and open state	196

List of Abbreviations

AA	...	All-atom
AChBP	...	Acetylcholine binding protein
BAR	...	Bennett’s acceptance ratio
CG	...	Coarse-grained
COG	...	Centre of geometry
Cryo-EM	...	Cryogenic electron microscopy
CHOL	...	Cholesterol
CV	...	Collective variable
DOPS	...	1,2-dioleoyl-sn-glycero-3-phospho-L-serine
ECD	...	Extracellular domain
EC₅₀	...	Half maximal effective concentration
EPSP	...	Excitatory postsynaptic potential
ELIC	...	GABA/cysteamine-gated channel from <i>Erwinia chrysanthemi</i>
GABA	...	γ -aminobutyric acid
GABA_AR	...	γ -aminobutyric acid type A receptor
GLIC	...	Proton-gated ion channel from <i>Gloeobacter violaceus</i>
GluCl	...	Glutamate-gated chloride channel
GlyR	...	Glycine receptor
GPCR	...	G-protein coupled receptor
ICD	...	Intracellular domain
InsP3	...	Inositol-1,4,5-trisphosphate
IPSP	...	Inhibitory postsynaptic potential
ITS	...	Implied timescale
LGIC	...	Ligand-gated ion channel
LJ	...	Lennard-Jones

MD	Molecular dynamics
MBAR	Multiple Bennett's acceptance ratio
MM-PBSA	Molecular-mechanics Poisson-Boltzmann surface area
MSM	Markov state model
nAChR	Nicotinic acetylcholine receptor
NMR	Nuclear magnetic resonance
PA	Phosphatidic acid
PC	Phosphatidylcholine
PCCA++	Perron cluster cluster analysis
PE	Phosphatidylethanolamine
PIP2	Phosphatidylinositol 4,5-bisphosphate
pLGIC	Pentameric ligand-gated ion channel
PME	Particle mesh Ewald
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PS	Phosphatidylserine
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SM	Sphingomyelin
TICA	Time-lagged independent component analysis
THDOC	Allotetrahydrodeoxycorticosterone
TMD	Transmembrane domain
UA	United-atom
vdW	Van der Waals
WT	Wild type
ZAC	Zinc-activated channel
1-d, 2-d, 3-d	One-dimensional, two-dimensional, three-dimensional
5-HT₃	5-hydroxytryptamine (serotonin)
5-HT₃R	5-hydroxytryptamine (serotonin) type 3 receptor
(+)	Principal subunit
(-)	Complementary subunit

These are some of the things that hydrogen atoms do, given fifteen billion years of cosmic evolution.

— Carl Sagan

1

Introduction

Contents

1.1	Neuronal Activity	2
1.2	Synaptic Transmission	3
1.2.1	Electrical Synapses	4
1.2.2	Chemical Synapses	4
1.2.3	Signal Integration	6
1.3	Pentameric Ligand-Gated Ion Channels	7
1.3.1	Eukaryotic pLGICs	8
1.3.2	Prokaryotic pLGICs	11
1.3.3	Conserved Overall Architecture	11
1.3.4	The Activation Mechanism - Insights from Single Channel Recordings	20
1.3.5	Partial Agonism	23
1.3.6	Allosteric Modulation	25
1.4	Thesis Organization and Motivation	26

The flow of electrical signals through our nervous circuits allows us to perceive the outside world, to walk and speak, to learn and think, and, ultimately, gives rise to our mind. The ultimate goal of neuroscience is to understand in molecular detail how all these functions are carried out by our nervous system. While we are still far away from this level of understanding, there has been significant progress in elucidating the molecular mechanisms that govern our nervous system. This thesis contributes a small piece of the puzzle towards this molecular understanding.

This introduction aims to provide the relevant general background to contextualize the results of this thesis. In doing so, I will begin at the cellular level of neuronal activity and synaptic transmission¹ and then provide an overview of the class of pentameric ligand-gated ion channels (pLGICs), which mediate most of the fast synaptic transmission. Zooming in at a molecular level, I will present the overall architecture of this ion channel family, review insights about the activation mechanism from electrophysiological single channel recordings, introduce the phenomenon of partial agonism and briefly cover allosteric modulation sites. While in this thesis, molecular dynamics simulations solely of the glycine receptor (GlyR) are performed as a prototypic representative of the pLGIC superfamily, often comparisons between other members of the pLGIC family are drawn, as their structure and function closely resemble each other. Therefore, a rather broad perspective was chosen to cover this receptor class as a whole. The introduction closes with an overview of the thesis organization and its contributions to the field.

1.1 Neuronal Activity

Important units of the nervous system are nerve cells (neurons) and glial cells (glia). Neurons are the core functional units which create electrical signals and react to them, while glia play a supportive role for the neurons. Neurons are highly specialized cells and typical structural features are depicted in Figure 1.1.

How do neurons carry out their function? Receiving and processing incoming stimuli is typically facilitated by the dendrites and the cell body (soma). Incoming signals received by the dendrites move passively to the soma until they eventually reach the axon hillock. This is a specialized structure forming a junction between the soma and the axon. It is here where all incoming signals are integrated and where it is determined if the neuron generates an electrical signal in response to them. If the sum of all incoming signals is above a certain threshold, then an action potential is fired, which is electrically conducted down the axon. Towards the end, the axon branches out into axon terminals which form connections to other neurons

¹These two sections are based on the textbooks [1, 2].

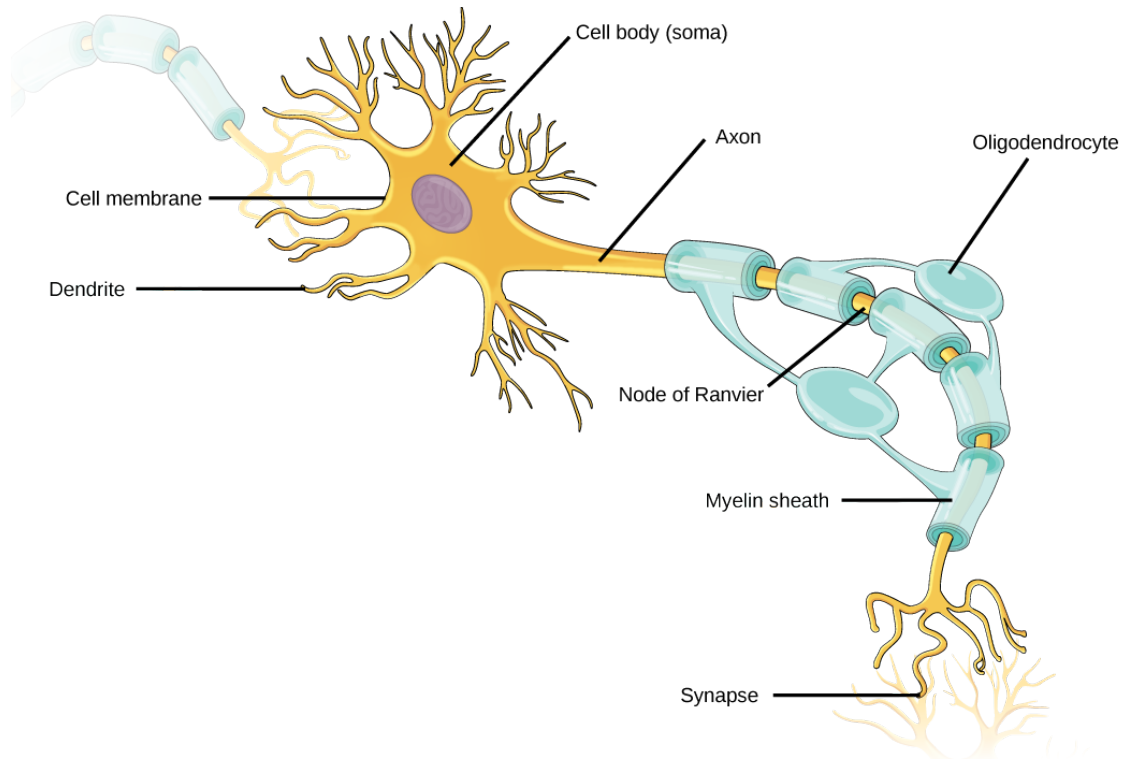


Figure 1.1: Schematic overview of a neuron. Figure reproduced from [1] under a Creative Commons Attribution 4.0 International License.

or target cells, to which the signal is passed on. Neuron-to-neuron connections are called synapses and the process of how nerve signals are communicated across synapses will be described in more detail in the following section.

In this context it is important to keep in mind that neuronal activity results not from a single neuron alone, but from the interplay of a myriad of neurons being in contact with each other. Each neuron engages with other neurons via many synaptic connections, for instance neocortical neurons have on average 7,000 synaptic connections [3]. Considering the human brain is estimated to have around 100 billion neurons [4], the resulting total number of synaptic connections is an astronomically large number.

1.2 Synaptic Transmission

Synapses are the communication interfaces between neurons. Here, an incoming action potential from a sending (presynaptic) neuron leads to the transmission of

a signal to a receiving (postsynaptic) neuron, which contributes to the likelihood that the postsynaptic neuron might fire another action potential. This process is called synaptic transmission and can either be electrical or chemical.

1.2.1 Electrical Synapses

In an electrical synapse, there is a direct physical connection of the presynaptic and postsynaptic membranes via integral channel proteins forming gap junctions. A gap junction consists of two hemichannels, which connect across the intercellular space. In vertebrates, these hemichannels are formed by connexin proteins. Each hemichannel is composed of six connexin subunits, each made up of four transmembrane helices [5]. The action potential is nothing but a current carried by directed ion flow. Ions can pass the gap junctions and thus, the electrical signal is transferred directly from the presynaptic to the postsynaptic neuron. Electrical synapses are less common than chemical synapses.

1.2.2 Chemical Synapses

Chemical synapses communicate an incoming electrical nerve signal via translating it into a chemical signal and then reconvert it into an electrical signal. The overall mechanism can be subdivided into several steps as illustrated in Figure 1.2.

Mechanism of Chemical Synaptic Transmission

When an action potential arrives at an axon terminal, this translates at a molecular level into the membrane being depolarized. This triggers the opening of voltage-gated calcium channels. Driven by a concentration gradient, calcium ions stream passively into the cell where they initiate a signalling cascade that ultimately leads to synaptic vesicles fusing with the membrane. These vesicles are filled with neurotransmitters which are released into the synaptic cleft, where they passively diffuse. These neurotransmitters can bind to receptors located in the postsynaptic membrane. Ligand-gated ion channels (LGICs) (or ionotropic receptors) are an important class of such receptors. Upon neurotransmitter binding they undergo

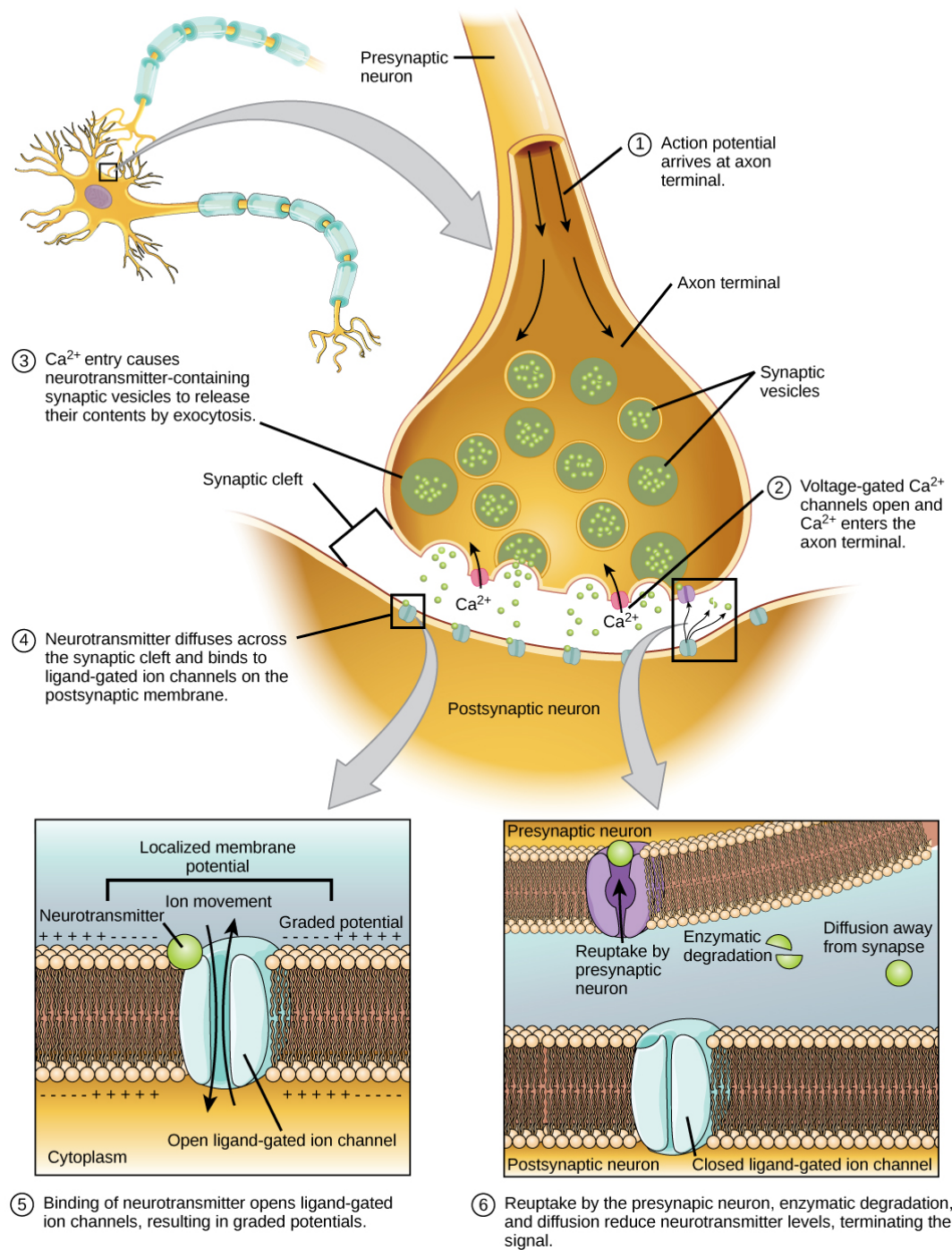


Figure 1.2: Schematic overview of a chemical synapse. Figure reproduced from [1] under a Creative Commons Attribution 4.0 International License.

conformational changes that lead to the opening of an ion channel pore. Due to the membrane potential, ions flow into the postsynaptic neuron which produces a localized change in the membrane potential.

It should be noted that another important class of postsynaptic receptors are metabotropic receptors. These belong to the family of G-protein coupled receptors (GPCRs) and mediate slower synaptic transmission by triggering a signalling cascade on the inside of the pre- or postsynaptic membrane which ultimately alters the behaviour of the pre- or postsynaptic neuron.

Excitatory and Inhibitory Synapses

LGICs are highly selective for either cations or anions. Cation (Na^+ , K^+ , Ca^{2+}) inflow makes the inside of the cell more positive and thus causes the postsynaptic membrane to depolarize. This is called an excitatory postsynaptic potential (EPSP), as it makes the postsynaptic neuron more likely to fire an action potential. Inflow of anions (mostly Cl^-), on the other hand, results in a hyperpolarization of the postsynaptic membrane. Such an inhibitory postsynaptic potential (IPSP) makes the generation of an action potential more difficult. Usually a neuron releases only one type of neurotransmitter, e.g. those releasing acetylcholine are called ‘cholinergic’, those releasing glycine ‘glycinergic’ etc.

Signal Termination

For correct synaptic function, the signal must be terminated, so that the postsynaptic membrane can be reset to its normal resting potential, such that it can react to new arriving signals. To this end, the synaptic cleft must be cleared of neurotransmitters, which occurs in several ways. Neurotransmitters can simply diffuse away, they can be actively broken down by enzymes or removed via reuptake by the presynaptic neuron.

1.2.3 Signal Integration

A neuron can receive excitatory (EPSPs) as well as inhibitory inputs (IPSPs) from thousands of synaptic connections. All of these signals are summed up at the axon hillock and it is the overall net change of the postsynaptic membrane potential that

decides if a new action potential is fired. This process cannot only take signals arriving instantaneously from different regions (spatial integration) into account, but also process signals received within a short time frame (temporal integration) and is illustrated by a simplified example in Figure 1.3.

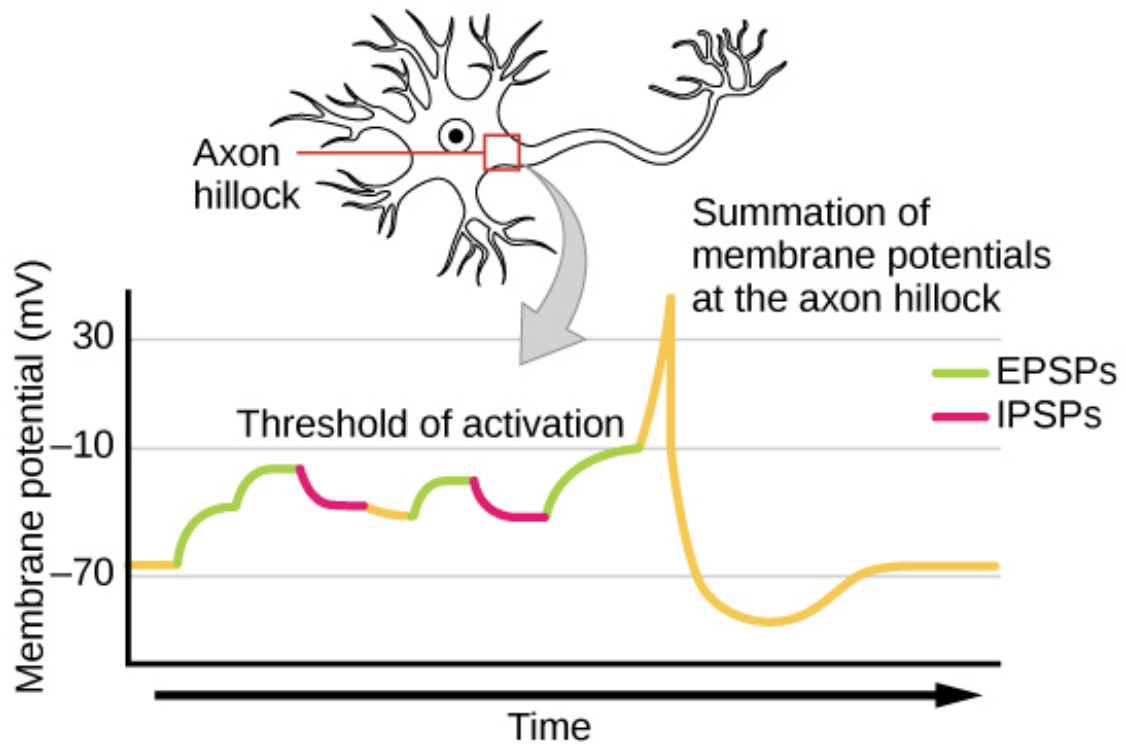


Figure 1.3: A simplified example how incoming signals can lead to the creation of an action potential at the axon hillock. Figure reproduced from [1] under a Creative Commons Attribution 4.0 International License.

1.3 Pentameric Ligand-Gated Ion Channels

The superfamily of ligand-gated ion channels includes the subfamily of pentameric ligand-gated ion channels (pLGICs), the tetrameric ionotropic glutamate receptors and the trimeric P2X receptors [6]. In this thesis the sole focus lies on the pLGIC family, also called the Cys-loop family or Pro-loop family. The overview presented here is mainly based on the two most recent reviews [7, 8] as well as [6] with additional sources cited explicitly.

Members of the pLGIC family are pervasively found in all neurons and their primary function is to mediate neuronal communication.

They are located in the postsynaptic membrane, where neurotransmitter binding leads to conformational changes that open a channel pore permitting passive ion passage which leads to cell hyperpolarization or depolarization.

Their dysfunction is associated with a wide range of neurological diseases such as epilepsy, Parkinson's, Alzheimer's or schizophrenia, and thus they constitute pivotal drug targets, see e.g. [9–11].

In mammals, the pLGIC superfamily consists of the cation-selective, excitatory nicotinic acetylcholine receptors (nAChRs), 5-hydroxytryptamine (serotonin) type 3 receptors (5-HT₃Rs) and a single zinc-activated channel (ZAC) as well as the anion-selective inhibitory γ -aminobutyric acid type A receptors (GABA_ARs) and glycine receptors (GlyRs). Moreover, the invertebrate glutamate-gated chloride channel (GluCl) from *Caenorhabditis elegans* as well as the prokaryotic proton-gated ion channel from *Gloeobacter violaceus* (GLIC) and the prokaryotic GABA/cysteamine-gated channel from *Erwinia chrysanthemi* (ELIC) further add to the diversity of pLGICs.

All members consist of five subunits (hence the name pentameric) and share a similar overall architecture that translates into a similar functional mechanism.

1.3.1 Eukaryotic pLGICs

Nicotinic Acetylcholine Receptor

The nicotinic acetylcholine receptor (nAChR) is the best-studied representative of all pLGICs. Acetylcholine is its natural neurotransmitter and nAChRs are non-specific cation channels and as such their effect is excitatory.

17 known subunits ($\alpha 1 - \alpha 10$, $\beta 1 - \beta 4$, γ , δ , ϵ) allow for immense combinatorial possibilities to form pentameric subtypes.

nAChRs are ubiquitously expressed across the nervous system and apart from neuron-to-neuron synapses (mostly $\alpha 4\beta 2$ heteropentamers and $\alpha 7$ homopentamers [12]) play an important role at neuromuscular junctions (predominantly the $\alpha 1_2\beta 1\delta\epsilon$ adult muscle-type) [6].

Certain mutations in neuronal nAChR result in nocturnal frontal lobe epilepsy [13, 14], while mutations in muscle-type nAChR can cause congenital myasthenic

syndrome [14, 15]. Additionally, dysfunction of nAChRs has been linked to schizophrenia [16], epilepsy [13] and there is strong evidence that nAChRs are implicated in Parkinson's and Alzheimer's disease [17, 18]

Consequently, nAChRs are important drug targets. Carbamazepine is an antiepileptic drug that acts as an open channel blocker of nAChRs [19, 20]. Since nAChRs are responsible for nicotine and tobacco addiction [21], there is also a huge market for smoking cessation drugs, such as varenicline[22].

5-hydroxytryptamine (serotonin) Type 3 Receptor

Besides the nAChR, the other major excitatory (i.e. cation-selective) pLGIC is the 5-hydroxytryptamine type 3 receptor (5-HT₃R). As its name indicates, it is activated by the neurotransmitter 5-hydroxytryptamine (5-HT₃), also known as serotonin. Note that serotonin also activates other receptors that belong to the GPCR superfamily and are termed 5-HT_x, $x = \{1, 2, 4, 5, 6, 7\}$.

While five known subunits termed A–E exist, 5-HT₃Rs are primarily expressed as A homopentamers or AB heteropentamers. 5-HT₃Rs are located in the central and peripheral nervous systems [23, 24].

They play roles in cognition, regulate the vomiting reflex, alter gastrointestinal motility and are involved in anxiety control, the reward system and pain processing. Accordingly, they are targeted by drugs for the treatment of nausea and vomiting, gastrointestinal disorders, irritable bowel syndrome and anxiety, among others [25–27].

Zinc Activated Receptor

In addition to the nAChR and the 5-HT₃R, another excitatory cationic zinc-activated channel (ZAC) exists [28], but it remains the least understood member of the pLGIC superfamily and recently it was shown that copper and protons are also agonists of ZAC [29].

γ -aminobutyric Acid Type A Receptor

In contrast to the excitatory cationic receptors, the anion-selective γ -aminobutyric acid type A receptor (GABA_AR) is the major inhibitory receptor and as such is responsible for most of the fast inhibitory neurotransmission in the adult nervous system. Note that in contrast to the type A receptors, the so called GABA_BRs are metabotropic receptors belonging to the GPCR superfamily.

19 known subunits, named $\alpha 1$ - $\alpha 6$, $\beta 1$ - $\beta 3$, $\gamma 1$ - $\gamma 3$, δ , ϵ , θ , π and $\rho 1$ - $\rho 3$, give rise to a large diversity of receptor subtypes. The canonical assembly consists of two α subunits, two β subunits and a fifth subunit [30, 31]. Recent work has identified the $\beta 3$ subunit as necessary and sufficient for hippocampal inhibitory transmission [32].

The GABAergic system is associated with neurodevelopmental disorders, such as fragile X syndrome, Rett syndrome and Dravet syndrome [33] as well as schizophrenia [16] and certain mutations are responsible for several epileptic syndromes [34]. Moreover, GABAergic inhibitory pathways mediate in part how ethanol alters neuronal activity [35]. In terms of existing drugs, GABA_ARs are targeted by benzodiazepines which produce sedative, anxiolytic, muscle relaxant and anticonvulsant effects by positive modulation [36].

Glycine Receptor

Apart from GABA_ARs, inhibitory synaptic transmission is also mediated by the anionic glycine receptors (GlyRs) which are activated by the amino acid glycine.

While five subunits are known ($\alpha 1$ - $\alpha 4$ and β), in humans the gene for α_4 is a pseudo-gene, reducing the number of GlyR subunit types to four [37]. The predominant form in adults is the heteromeric $\alpha_1\beta$ GlyR [38], but GlyRs also assemble as functional $\alpha 1$ homopentamers. The exact subunit stoichiometry of the predominant $\alpha 1\beta$ form is not totally confirmed. Experimental results allow for a $2\alpha : 3\beta$ or a $1\alpha : 4\beta$ stoichiometry, where newer atomic force microscopy results indicate an invariant composition of $2\alpha 1 : 3\beta$ with a $\beta - \alpha_1 - \beta - \alpha_1 - \beta$ subunit arrangement [39].

GlyRs are mostly expressed in the brain stem and spinal cord and are involved in motor coordination, sensory processing, including vision and audition, and pain perception [38, 40–42]. Just like GABA_ARs, they are also involved in the effects that ethanol exerts on the nervous system [43]. Furthermore, certain mutations result in a rare genetic disorder, called hyperekplexia or startle disease, because patients suffer from an exaggerated startle reflex in response to acoustic or tactile stimuli [44, 45]. While to date most drugs targeting GlyRs combat chronic pain, there is great potential for drug development addressing temporal lobe epilepsy, autism, breathing disorders, hyperekplexia, motor neuron disease and alcoholism [46, 47].

Glutamate-gated Chloride Channel

Closely related to GlyRs, but only found in protostome invertebrate phyla is the glutamate-gated chloride channel (GluCl) [48]. GluCls control locomotion and feeding and mediate sensory inputs into behaviour. Parasitic nematodes and arthropods can be killed by drugs from the macrocyclic lactone family of anthelmintics and pesticides which specifically target GluCls [48].

1.3.2 Prokaryotic pLGICs

In 2004 prokaryotic homologues of pLGICs were identified by Tasneem et al. [49], adding even more diversity to the pLGIC family. The GABA/cysteamine-gated channel from *Erwinia chrysanthemi* (ELIC) as well as the proton-gated ion channel from *Gloeobacter violaceus* (GLIC) are both cation-selective. While their physiological role is unclear, they have provided the first high resolution structures of the pLGIC superfamily [50] and put the evolution of pLGICs into a new perspective [51, 52].

1.3.3 Conserved Overall Architecture

Over the last decade, many structures of pLGICs have been solved in different functional states, both with X-ray crystallography and cryogenic electron microscopy (cryo-EM), revealing that the overall architecture is well conserved among all members of the superfamily (see Figure 1.4).

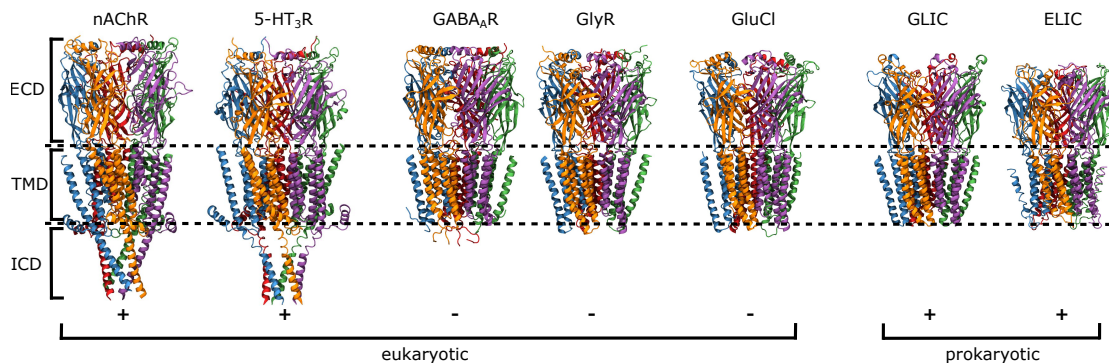


Figure 1.4: Overall architecture of the pLGIC superfamily. pLGICs consist of five subunits (coloured differently) that arrange pseudo-symmetrically around the ion conducting pore. The approximate location of the lipid headgroups of the membrane is indicated by dashed lines. + and - signs denote whether the channels are selectively permeable for cations or anions, respectively. Eukaryotic pLGICs consist of an extracellular domain (ECD), a transmembrane domain (TMD) and an intracellular domain (ICD), which is only partially resolved in cationic channel structures and removed or not resolved in the structures of the anionic channels. Prokaryotic channels do not have an intracellular domain. PDB codes of the structures shown are: nAChR (6PV7 [53]), 5-HT₃R (6DG8 [54]), GABA_AR (6I53 [55]), GlyR (5TIN [56]), GluCl (4TNW [57]), GLIC (6HZW [58]) and ELIC (2VL0 [59]).

The structures show that pLGICs assemble as pentamers with a C5-pseudo-symmetry axis along the centre of the ion channel pore. Figure 1.5 shows more detail of the overall structure of a pLGIC, exemplified by the GlyR.

The extracellular domain (ECD), mostly organized in a β -sandwich fold, contains the neurotransmitter binding sites (orthosteric binding sites), which are located at the subunit interfaces (see Figure 1.5b). The number of orthosteric sites depends on the receptor and its subunit composition, for instance, while the homomeric $\alpha 1$ GlyR has five orthosteric sites, the muscle $\alpha 1_2\beta 1\delta\epsilon$ nAChR only has two at the $\alpha 1\delta$ and $\alpha 1\epsilon$ interfaces. The ECD of eukaryotic pLGICs also has a conserved disulfide bridge, forming a characteristic Cys-loop and has been the origin for the name ‘Cys-loop receptors’. However, since the prokaryotic channels lack this disulfide bridge, this term has been superseded by the term pLGICs.

The transmembrane domain (TMD) hosts the ion channel pore. Each subunit consists of four sequential α -helices (M1-M4). According to Barrantes [61], these helices can be subdivided into three regions that form concentric rings (see Figure 1.5c) from the central pore towards the protein periphery. The inner ring consists of the

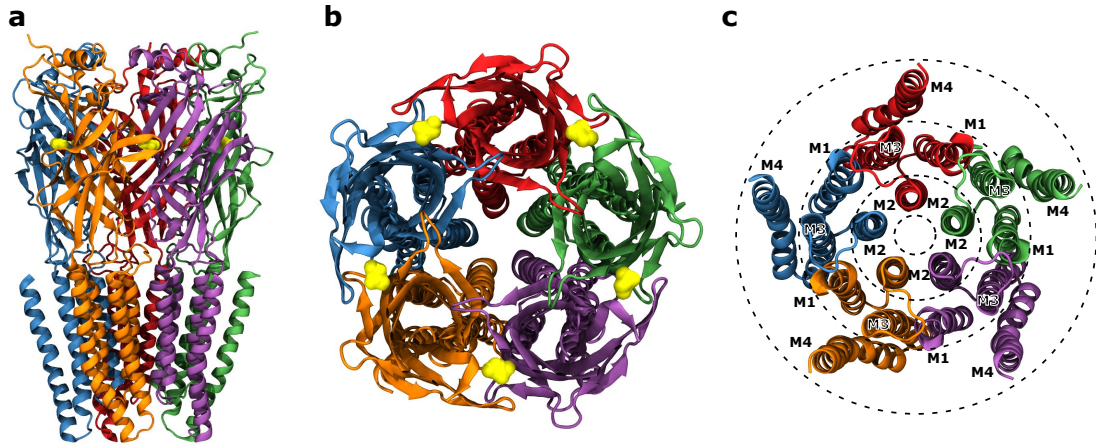


Figure 1.5: Overall pLGIC architecture exemplified by the GlyR. Each pLGIC has five subunits (coloured differently) that assemble pseudo-symmetrically around the ion channel pore as symmetry axis. Neurotransmitters (yellow) bind to conserved orthosteric sites at subunit interfaces of the ECD, but the number of binding sites depends on the receptor type and its subunit composition. (a) Side view (PDB: 5TIN [56]). (b) View from the extracellular towards the intracellular side with the upper half of the ECD removed to reveal the view of the neurotransmitter binding sites (PDB: 5TIN [56]). (c) The TMD can be subdivided into three regions (concentric dashed rings) from the inner pore towards the outer membrane facing surface (same perspective as in (b)) (PDB: 5CFB [60]). Each subunit consists of four consecutive transmembrane helices M1-M4.

M2 helices of all subunits which line the ion-conducting pore. The middle ring is formed by the M1 and M3 helices of all subunits whose surface is partly in contact with the plasma membrane. The outer ring is made up of the M4 helices of all subunits which are most exposed to the membrane environment.

Eukaryotic pLGICs also have an intracellular domain (ICD) with large diversity, both in length (80-200 residues) and sequence. The ICD is a mostly unstructured loop region connecting the TMD helices M3 and M4. Due to its unstructured nature, it is usually replaced by short peptide linkers to facilitate structure determination.

Interestingly, molluscan glial cells naturally express a truncated version of nAChRs. It only consists of the ECD and is called acetylcholine binding protein (AChBP). It plays a modulatory function in synaptic transmission [62] and since its structure had been solved before the full length receptor structures, was an important structure template for the ECD of the superfamily.

The Neurotransmitter Binding Site

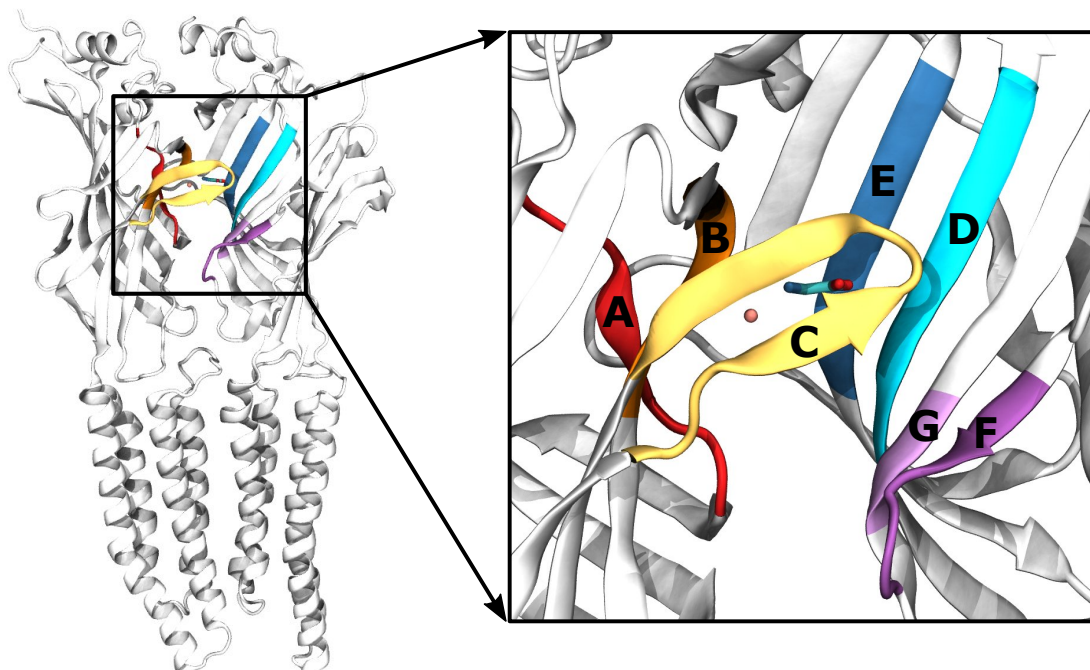


Figure 1.6: Location of the orthosteric binding site at the interface of two adjacent subunits in pLGICs exemplified by the GlyR (PDB: 5TIN). The regions forming the orthosteric site are traditionally called loops A-F with loops A-C forming the principal face and loops D-G the complementary face of the orthosteric site. Here, the neurotransmitter glycine as well as a stable water molecule in the binding pocket (rose sphere) are shown explicitly as an example and to give an orientation where the neurotransmitter molecule is located in the orthosteric site.

The orthosteric binding site of pLGICs is located at the interface of adjacent subunits of the ECD (see Figure 1.6). It is formed by regions that are traditionally called loops A-C belonging to the principal (+) subunit and loops D-G (which are actually β -strands) being part of the complementary (-) subunit. Several pLGIC structures could be resolved with orthosteric ligands bound. A common pharmacophore in all neurotransmitters is a positively charged ammonium group which can bind to the principal face of the orthosteric site, mainly via cation- π interactions. The interactions with the complementary subunit are more diverse, depending on ligand and different members (as well as subtypes) of the pLGIC class. Reviews that go more in depth for the entire superfamily are [63–65].

Since only agonist binding to the GlyR will be studied in this thesis, here I only review the binding mode of the endogenous neurotransmitter glycine to the GlyR. In the currently highest resolution structure (2.61 Å) of the human $\alpha 3$ GlyR (PDB: 5TIN) [56], density for glycine was for the first time discernible (see Figure 1.7). The zwitterionic glycine molecule faces the principal subunit with its positively charged ammonium group, while its negatively charged carboxylate moiety points towards the complementary subunit. The ammonium end is stabilized by a cation- π interaction with (+)F207 and an electrostatic interaction with (+)E157 which is bridged by a stable water molecule in the binding pocket as well as an H-bond with the backbone carbonyl oxygen of (+)F159. The carboxylate end forms H-bonds with (+)T204 and (-)S129 and forms a saltbridge with (-)R65 which simultaneously acts as a bidentate H-bond donor.

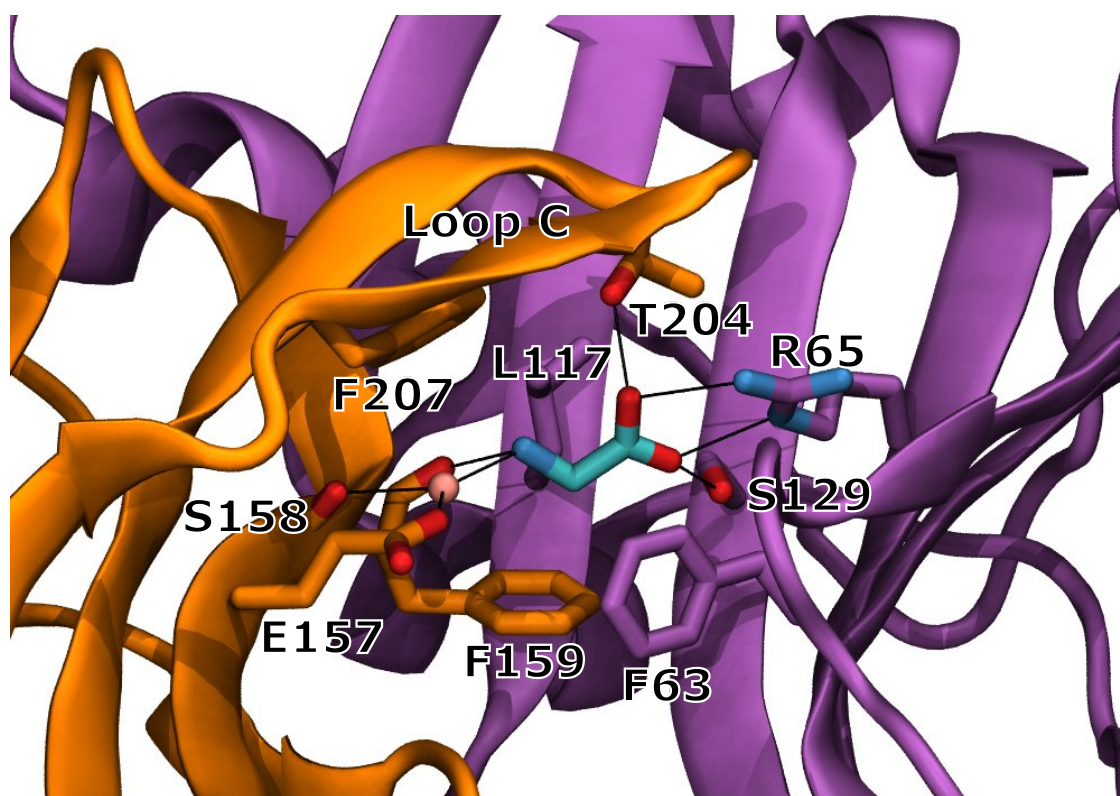


Figure 1.7: Binding mode of the neurotransmitter glycine in the orthosteric site of GlyR (PDB: 5TIN). View from the membrane towards the extracellular matrix. The principal subunit is coloured in orange, the complementary subunit in purple. Important residues for binding are shown explicitly.

This glycine binding mode, including the stable water molecule in the binding site,

has been previously predicted by docking and MD-refinement in homology models of the human $\alpha 1$ GlyR based on a GluCl template (PDB: 3RIF [66]) by Yu et al. [67] as well as based on a zebrafish $\alpha 1$ GlyR template (PDB: 3JAE [68]) by the author [69].

Note that loop C covers the orthosteric site in a lid-like fashion and can thus shield it from bulk water which protects the ligand from unbinding. It seems a common feature of all pLGICs that loop C is dynamically flexible in the apo state, while ligand binding goes hand in hand with loop C capping the binding site and being more rigid (see the review by Nys et al. [63] and references therein for experimental evidence as well as, e.g. [70–72] for evidence from MD simulations). The distance change of C_α atoms in loop C between the ligand bound and apo state can vary up to 8 Å [73].

The Ion Channel Pore

The TMD hosts the ion channel pore which is lined by the M2 helices of all five subunits (see Figure 1.8). This creates a layered organization of pentagonal rings of pore-lining residues that are highly-conserved. Hence, for easier comparison, a prime numbering scheme is used for residues of the M2 helices for all pLGICs.

There are two main constriction points. One is located in the centre of the pore at the 9' position and formed by a ring of hydrophobic residues. Leucine is highly conserved at the 9' position (with the exception of 5-HT₃R B subunit and GLIC, where the 9'L is replaced by 9'V or 9'I, respectively), which is why it is commonly referred to as the leucine gate. The other constriction is at the intracellular end of the TMD, where the -2'/-1' position faces the pore lumen. Cationic pLGICs typically have a glutamate at the -1' position, while in anionic channels the constriction is usually formed by an alanine or proline at the -2' position.

From the structural data of pLGICs in different conformational states, the dual gating hypothesis has emerged [76]. In the resting gate both the 9' and -2'/-1' constriction points form a gate, while in the active state both these gates are open. In the desensitized state the 9' gate is open, while the -2'/-1' gate is closed.

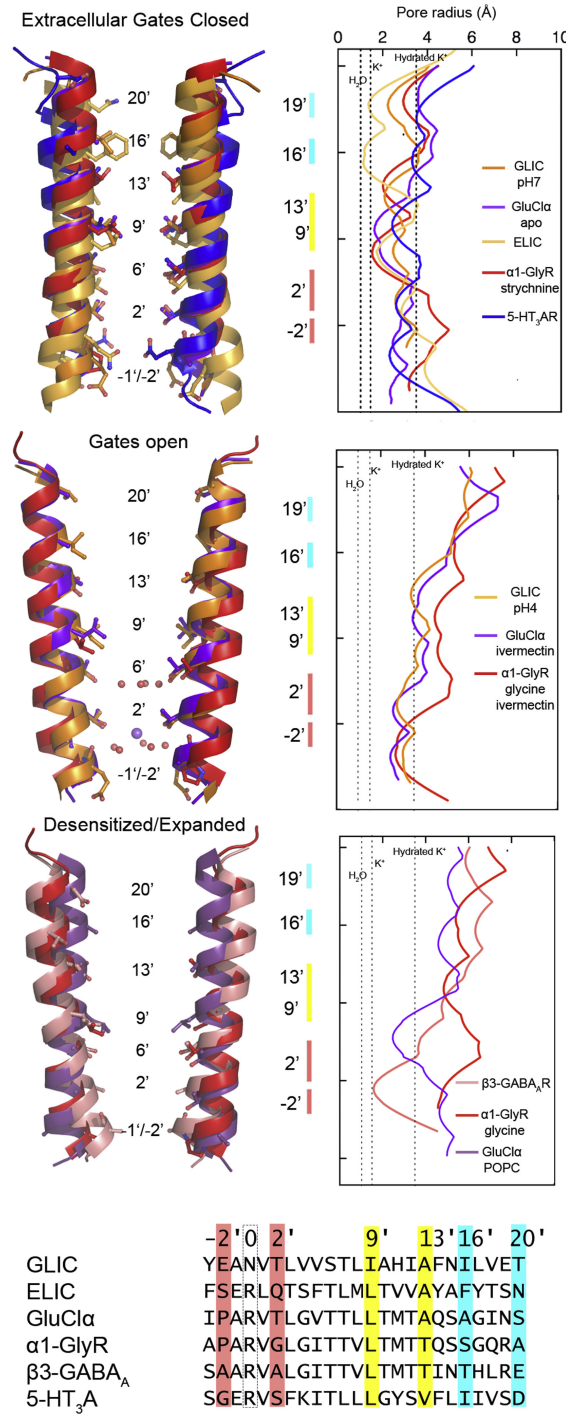


Figure 1.8: The ion channel pore formed by the M2 helices of the TMD in different conformational states on the left with the corresponding pore radii determined with Hole [74, 75] on the right. The radii of water, desolvated K⁺ as well as K⁺ with a hydration shell are indicated by dashed lines. Top: Closed/resting states represented by GLIC, GluClα, ELIC, α1-GlyR and 5-HT₃R structures. Middle: Open/Active States represented by GLIC, GluClα and α1-GlyR structures. Bottom: Expanded/Desensitized states from β3-GABA_AR, α1-GlyR and GluClα. A sequence alignment of the pore-lining M2 helices is placed below. Figure reproduced with permission from Elsevier Neuron, [7], <https://doi.org/10.1016/j.neuron.2016.03.032> (Copyright Elsevier Neuron 2016).

For the constriction at the 9' position, the concept of hydrophobic gating [77, 78] plays an important role. Hydrophobic gating occurs in narrow hydrophobic pores where the energetically favoured expulsion of water prohibits ion permeation. Beckstein and Sansom [79] have studied in detail the influence of radius, hydrophobicity as well as flexibility of the pore on the degree of dewetting and found that a radius of less than ca. 4-5 Å in hydrophobic regions can lead to local dehydration and thus be a prohibitive energetic obstacle for ion permeation.

While there has been some controversy about the precise origins of selectivity for cations or anions in pLGICs, there is now consensus that it is mainly governed by the intracellular end of the TMD pore [9, 80–85]. An extensive mutagenesis study by Cymes and Grosman [85] has revealed a complex hierarchic model for charge selectivity with the first turn of the M2 helix (where the -2/-1' gate is located) dominating permeation control, but other ionized residues along the permeation pathway (both in the TMD as well as the ECD) also contribute in an ancillary manner.

The ECD-TMD Interface as Communication Hub for Intrinsic Signal Transduction

The interface between the ECD and TMD forms the communication hub for conformational change induced by agonist binding between the ECD and the TMD (see Figure 1.9), which carries the channel pore. The ECD is covalently linked to the TMD via the end of the β_{10} -strand (pre-M1) with the M1 helix of the TMD. In addition, several well-conserved loops engage in non-bonded interactions between these two domains. The β_1 - β_2 loop (loop 2) of the ECD can interact with the part of the M2-M3 loop that is close to the M2 helix of the TMD. The characteristic Cys loop (β_6 - β_7 loop or loop 7) in the ECD can form contacts with the part of the M2-M3 loop that is close to the M3 helix, with the top of the M3 helix as well as with the extracellular extension of the M4 helix. The β_8 - β_9 loop of the ECD can form contacts with the M2-M3 loop of an adjacent subunit.

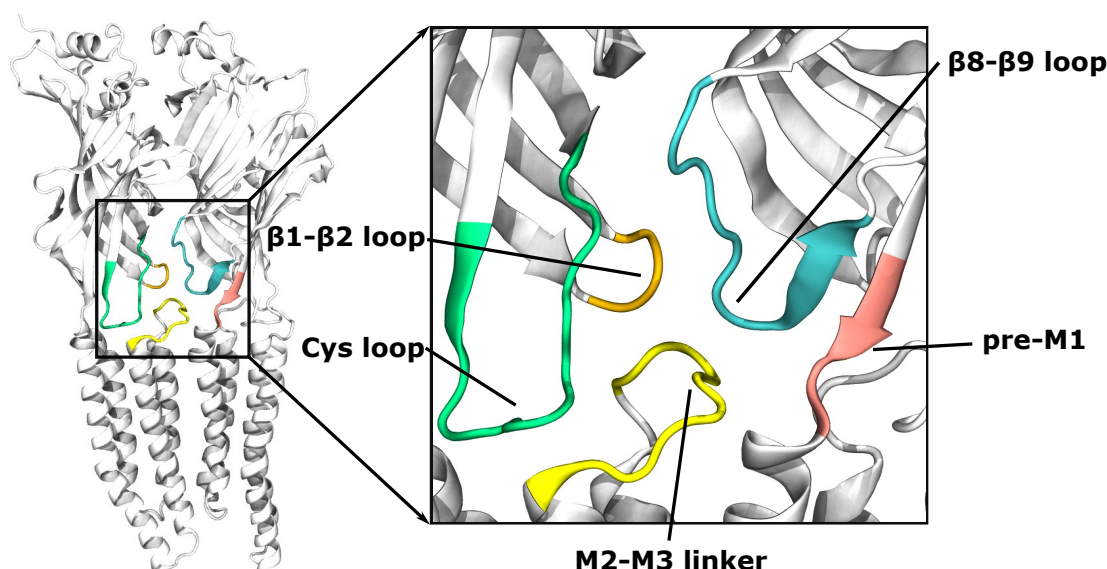


Figure 1.9: The ECD-TMD interface is the communication platform for signal transduction between the neurotransmitter binding ECD and the TMD carrying the channel pore (here illustrated by the GlyR structure PDB: 5TIN).

The Intracellular Domain

The cytoplasm facing ICD formed by the M3-M4 loop is only present in eukaryotic pLGICs and varies strongly in length and sequence. Since it is usually replaced by short peptide linkers to facilitate structure determination, very little is known about its structure. At high resolution it has only been partially resolved so far in structures of the 5-HT₃R [54, 86] and the nAChR [53]. In these structures, a loop connecting the M3 helix with an α -helix oriented almost in parallel to the membrane plane (MX) is visible as well as a long α -helical intracellular extension of the M4 helix (MA stretch).

The ICD plays an important role for the interaction with intracellular proteins. For instance, it facilitates clustering through direct interactions with rapsyn for muscle nAChRs [87] or via gephyrin [88] for GABA_ARs and GlyRs. GlyRs can be modulated via the ICD through interactions with the G protein $\beta\gamma$ dimer [89] (see [90] for a detailed review of the role of the ICD for GlyRs). From studies on cationic pLGICs, the picture has emerged that ions exit the ICD through

lateral fenestrations [53, 91–94].

1.3.4 The Activation Mechanism - Insights from Single Channel Recordings

A longstanding question in the LGIC field is to understand the activation mechanism: How does agonist binding translate into channel opening? Based on studies of the endplate nAChR, in 1957 del Castillo and Katz were the first to postulate a mechanism that separates agonist binding and channel opening into two separate steps [95]:



They proposed that binding of the agonist (A) to a receptor (R) leads to a complex (AR) which is still inactive (channel shut) and only after this complex is formed a conformational change occurs resulting in the active (channel open) state AR^* . The first binding step is characterized by the equilibrium dissociation constant K_A and the second shut-open transition is characterized by the equilibrium constant E defined as the ratio of the rate constants of opening over shutting.

Since then mechanistic insight into the activation mechanism through the unique insights offered by single channel electrophysiological recordings has come a long way [96–99]. For the GlyR, the nAChR, and the 5-HT₃R, it could be demonstrated that the activation mechanism after ligand binding, rather than going directly from shut to open as proposed by del Castillo and Katz, involves one or more shut intermediate states [100–109]. Different terms have been used to describe these mechanisms including shut intermediate states, such as 'flip', 'prime' or 'catch-and-hold' (see [110] for a short overview article).

Since in this thesis the focus lies on the GlyR, only the flip model is reviewed here, which is the state of the art for the activation mechanism derived from single channel recordings. It was proposed for the first time in 2004 by Burzomato et al. [100] for both the $\alpha 1$ GlyR homopentamer as well as the $\alpha 1\beta$ GlyR heteropentamer and is shown in Figure 1.10.

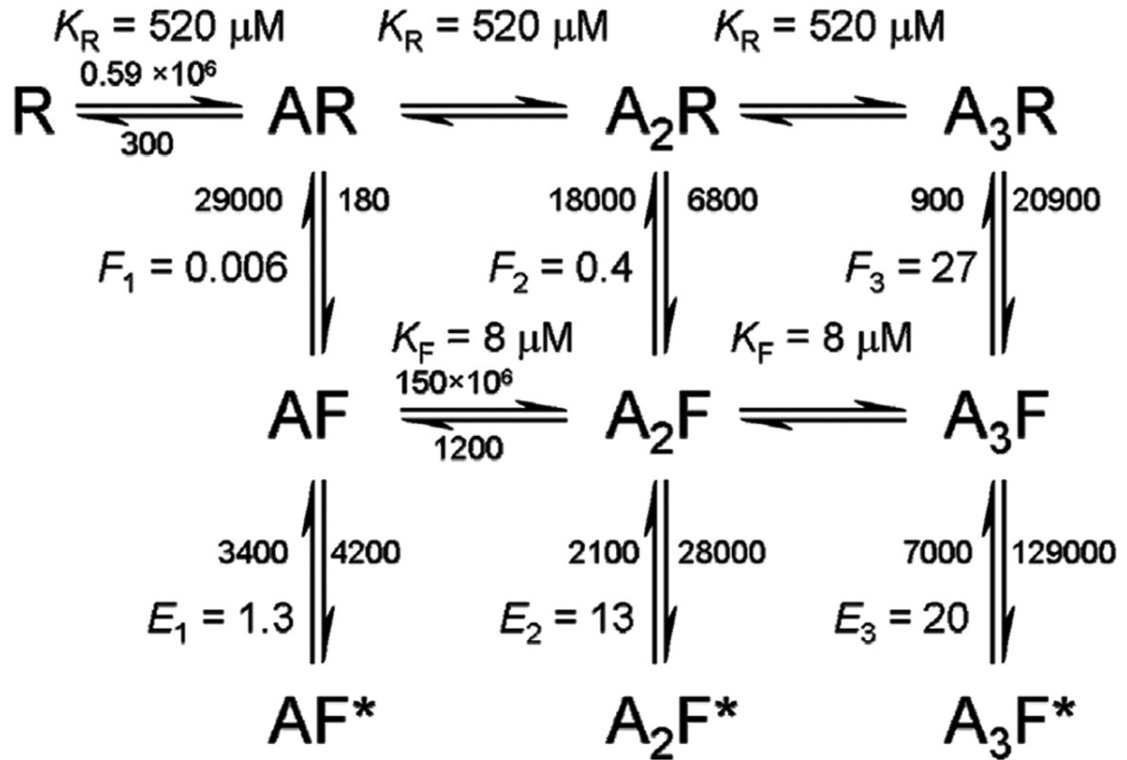


Figure 1.10: The ‘flip’ mechanism proposed by Burzomato et al. [100] to fit the single-channel activity of $\alpha 1\beta$ GlyR in response to its natural agonist glycine. R denotes the resting (inactive) conformation (top row), F the shut intermediate (‘flipped’) conformation (middle row) and F the active (open) channel conformation (bottom row). The equilibrium dissociation constant of agonist from the resting states is K_R and from the flipped states is K_F . The equilibrium constants for the resting-to-flipped transition are denoted by F , while the equilibrium constants for the conformational change from the flipped to the open state are denoted by E . Numbers next to arrows represent the rate constants (units are s^{-1} or $\text{s}^{-1} \text{M}^{-1}$ as appropriate). Figure reproduced from [100], <https://doi.org/10.1523/JNEUROSCI.3424-04.2004> (Copyright 2004 Society for Neuroscience).

Since the GlyR has five orthosteric binding sites, it can bind up to five glycine molecules. While one, two or three bound glycine molecules lead to openings with measurable frequency, it could not be determined if more glycine molecules can bind. Three glycine molecules are sufficient to reach the channel's maximum open probability. So even if more than three glycine molecules are bound, they cannot further increase the channel open probability [98]. The flip mechanism therefore includes four shut states with zero (R), one (AR), two (A_2R), three or more agonist molecules bound (A_3R) (top row). Each of the liganded states can change conformation to an open state (AF^* , A_2F^* and A_3F^* , bottom row). The novel feature of the flip model is that it postulates intermediate states (AF , A_2F and A_3F , middle row) along this transition which are still shut, the so called 'flipped' states.

Note that per construction the rate constants connecting the different liganded resting states (top row) are the same (and thus K_R), i.e. the proposed mechanism assumes no cooperativity of binding. The same is postulated for the flipped states (middle row).

In this concrete example of the $\alpha 1\beta$ GlyR, the agonist affinity is increased in the flipped state relative to the resting state which was also found for the $\alpha 1$ [100] and $\alpha 2$ [103] GlyR homopentamers. More generally, increased agonist affinity to the intermediate shut states relative to the resting states seems to be a common feature also in other members of the pLGIC family [105, 107, 109], suggesting that the orthosteric binding site of the intermediate state is structurally more similar to the active state than to the resting state.

The theoretical postulation of intermediate shut states raises the question about their physical nature. What do these intermediate states look like structurally? Mukhtasimova et al. [105] proposed that in the nAChR 'priming' corresponds to the capping of loop C over the binding site. They substantiated their hypothesis by showing that forcing loop C into a capped position via an engineered disulfide bridge had a similar effect as adding agonist. However, for GlyRs it is unlikely that loop C closing alone corresponds to flipping. With loop C covering the binding site, one would expect that agonist association would be slower, but the opposite is

observed for GlyRs: the agonist association rate constant for the flipped state is increased compared to the resting state [100, 101, 103]. As more complex collective variables involved in the gating process become apparent through insights from molecular dynamics simulations [111], it is likely that the flipping transition is a more complex conformational change that precedes channel opening.

The flip mechanism presented here has emerged from over 30 different putative tested mechanisms as the one which can explain the single-channel recording data best (maximum-likelihood) [100]. However, the current precision limits of single-channel recordings do not allow the complete estimation of rate constants for more complex mechanisms such as the primed model [105]. The flip model is probably still an oversimplification [110].

1.3.5 Partial Agonism

Partial agonists bind and activate a receptor, but can only elicit a partial response relative to a full agonist, even at saturating concentrations. For ligand-gated ion channels the 'response' can be defined as the maximum open probability, i.e. the proportion of time a receptor is open at saturating agonist concentrations.

To illustrate this, consider the following example (see Figure 1.11a). The natural agonist glycine, when applied at a concentration that saturates the binding sites, can elicit a maximum probability of 96 % for the ion channel to be open. Taurine on the other hand, binds at the same binding pocket as glycine, yet can only yield a maximum open probability of 54 %, even at saturating concentrations [101].

In the framework of the flip model, partial agonism can be explained by splitting agonist efficacy into two steps: an initial transition to the shut flipped state (flipping), followed by the actual channel opening. The partial agonist taurine is less efficient in making the channel enter the flipped state than the full agonist glycine. Once in the flipped state, the ability to open the channel is quite similar. Thus, with the partial agonist bound, the channel remains overall more time in the resting (closed) state, which in turn leads to a lower maximum open probability [101].

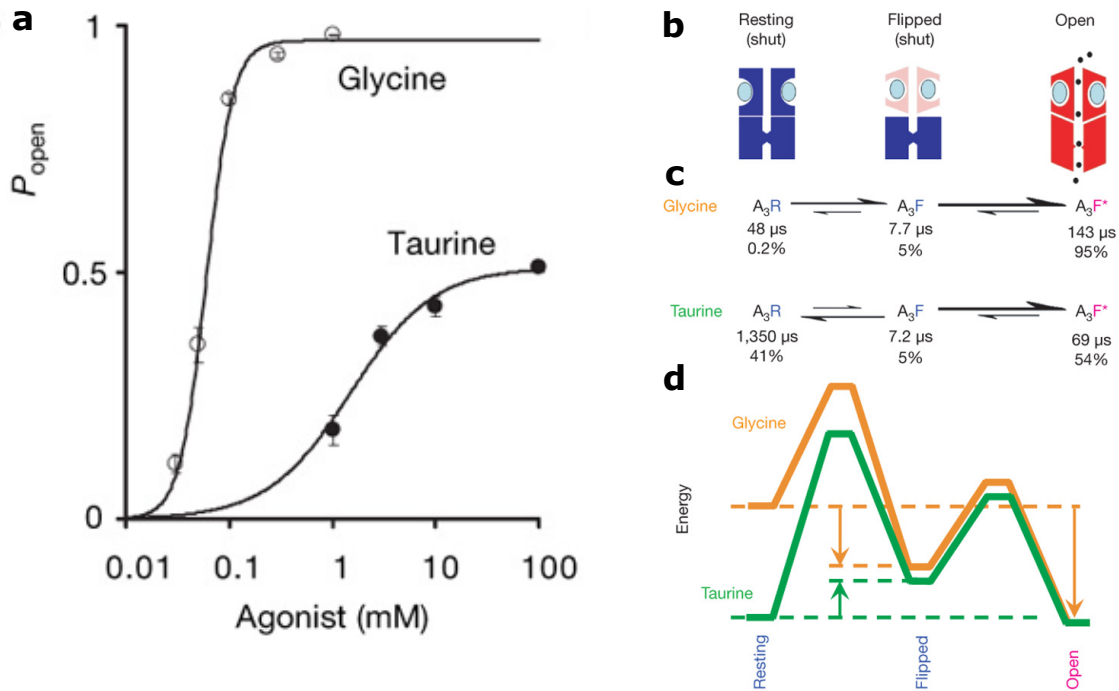


Figure 1.11: Partial agonism illustrated for glycine (full agonist) vs. taurine (partial agonist). (a) Open probability vs. agonist concentration. (b) Schematic representation of the the resting, flipped and open receptor. (c) Mean life time of each state and the equilibrium proportion of time spent in the resting, flipped and open state for glycine and taurine, respectively. The size of the arrows indicate the relative rates for the transitions between the different states. (d) Schematic illustration of the energies corresponding to the different states for glycine (orange) and taurine (green). Glycine can elicit a maximum open probability of 96 %, while taurine can only achieve 54 %. The right panel (b-d) provides an explanation for this in terms of the flip model. The rates connecting the flipped and open state are similar for both agonists. The main difference is that the full agonist glycine is much more efficient in favouring the transition from the resting to the flipped state than the partial agonist taurine. Figures reprinted by permission from Springer Nature, [101], <https://doi.org/10.1038/nature07139> (Copyright Springer Nature 2008).

1.3.6 Allosteric Modulation

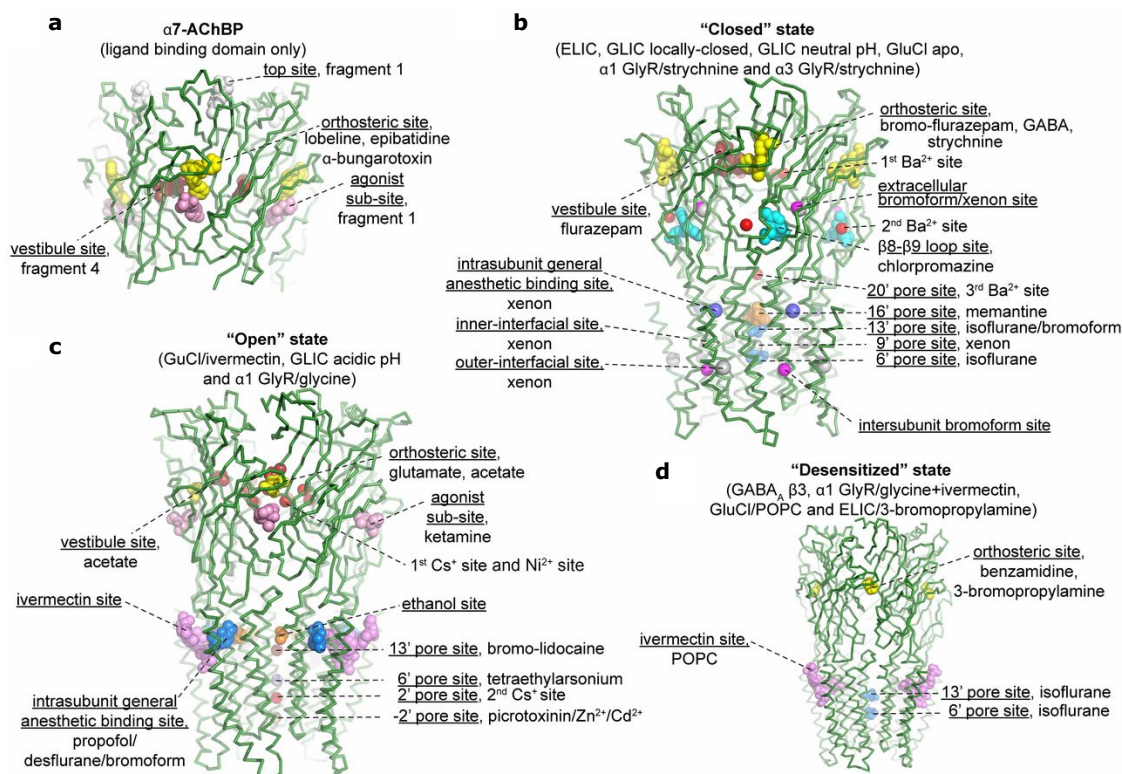


Figure 1.12: Overview of allosteric modulators bound at various binding sites to (a) AChBP and different conformational states of pLGICs ((b) closed, (c) open, (d) desensitized). The receptor backbone is shown in ribbon representation, while allosteric modulators are highlighted in van der Waals representation. The same colour scheme is used for modulators occupying overlapping sites in different states. Figure reproduced from [112], <https://doi.org/10.1073/pnas.1603101113>.

In addition to neurotransmitter activation, pLGICs can be modulated by a diverse set of both endogenous compounds, including membrane lipids, as well as drug molecules that bind at various locations other than the orthosteric site. Figure 1.12 provides an overview of allosteric sites in pLGICs in different conformational states. A detailed coverage of all allosteric modulators is beyond the scope of this introduction and it is referred to [8, 112–122] as excellent sources of information. Here, I only want to highlight the ivermectin binding site, which is located at a cleft formed by adjacent subunits in the upper half of the TMD in the open and desensitized conformations and was structurally resolved bound to the GlyR [68, 123], where it acts as a positive allosteric modulator.

1.4 Thesis Organization and Motivation

Chapter 2 provides the theoretical background for the computational methods used in this thesis.

Chapters 3-6 cover the major results of this thesis. The study presented in Chapter 3 was motivated by an ongoing dispute about the physiological relevance of the open state structure of the GlyR solved by cryo-EM in 2015. The structure has been rendered questionable by several independent groups because it is unstable and collapses in MD simulations. In this work, I resolve this controversy by developing a simulation protocol for atomistic MD simulations through which a stable open state of the GlyR can be obtained and demonstrating that its stability can be explained in terms of rearrangements in the gating regions of the receptor.

In Chapter 4, I explore the protein-lipid interactions of the GlyR as a representative of the pLGIC class in a neuronal membrane model. The motivation stems from the fact that pLGICs are known to be sensitive to their lipid environment and there is conclusive experimental evidence that certain lipid types such as cholesterol and anionic lipids are crucial for nAChR function. Yet, it is still poorly understood where lipid binding occurs, what the precise interactions are and how this modulates the receptor's behaviour. For the first time, I have therefore embedded the GlyR in a realistic model of a neuronal membrane and performed coarse-grained molecular dynamics simulations. A detailed map of the immediate lipid environment could be drawn which shows that it is dependent on the conformational state of the receptor, suggesting that lipids play an important role in stabilizing either the active or inactive state. Moreover, from a detailed analysis of selected lipid interaction hot spots, hypotheses are generated how certain lipid types might modulate receptor function.

Chapter 5 investigates how the natural neurotransmitter glycine and a number of partial agonists bind to the orthosteric binding site of the GlyR. Several chemically similar agonists are known to bind at the same binding site of the GlyR, but vary greatly in their ability to keep the channel open. While the binding mode of glycine to the GlyR could be resolved by X-ray crystallography, structural data of partial agonists bound to the GlyR has remained elusive thus far. In this study I set out to

explore possible structural explanations at an atomistic level of resolution for this phenomenon of partial agonism. Using atomistic MD simulations, I characterize the binding modes of various agonists and analyze their influence on receptor dynamics. Moreover, for the first time, the affinity of these compounds are calculated via alchemical free energy calculations and compared to affinity estimates from single channel recordings. From the insights obtained, hypotheses about the structural foundation of partial agonism in GlyR are formulated and interpretations derived from previous experimental studies are challenged.

Chapter 6 seeks to obtain insights into the gating mechanism via a combination of targeted MD and Markov state modelling. While the number of available structures of pLGICs in different functional states has largely increased over the last few years, the atomistic details of the full gating mechanism are still not fully understood. In order to obtain a more detailed insight of the structural intermediates connecting the closed and open state as well as their kinetic connections, I construct a Markov state model based on an accumulated simulation time of 85.5 μ s of unbiased atomistic MD simulations seeded along the gating pathway obtained by targeted MD. The resulting equilibrium distribution of states along the gating pathway as well as the underlying kinetics reveals metastable states and the main kinetic bottlenecks of the gating process.

Finally, in Chapter 7 the results and impact of this thesis are summarized in more detail.

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.

— Paul Dirac

2

Theory and Methods

Contents

2.1	The Physical Basis for Molecular Simulation	30
2.1.1	Key Concepts of Classical Mechanics	30
2.1.2	Key Concepts of Statistical Mechanics	36
2.2	Molecular Dynamics	38
2.2.1	Integration of the Equations of Motion	39
2.2.2	Potential Energy Models (Force Fields)	42
2.2.3	Periodic Boundary Conditions and Box Shapes	48
2.2.4	Common Speed up Techniques	50
2.2.5	Controlling Temperature and Pressure: Thermostats and Barostats	55
2.2.6	The Practical Workflow of a MD Simulation	59
2.3	Free Energy Calculations	61
2.3.1	The Alchemical Cycle	63
2.3.2	Bennett’s Acceptance Ratio	66
2.3.3	Multistate Bennett’s Acceptance Ratio	66
2.4	Markov State Modelling	67
2.4.1	Definition of a Markov State Model	67
2.4.2	Necessary Properties of the Underlying Dynamical Process	67
2.4.3	Time Evolution of a Markov State Model	68
2.4.4	Analyzing Markov State Models	70

This chapter provides the theoretical background of the simulation methods used in this thesis. A comprehensive review is beyond the scope of this chapter. For an in-depth discussion of all methods, refer to the standard textbooks in the field, such

as the ones by Tuckerman [124], Allen and Tildesley [125], Frenkel and Smit [126], Berendsen [127] and Leach [128], which are the core sources for this chapter. Here, the focus is on a high-level description of the key concepts of molecular dynamics (MD) simulation in a way that makes clear why and how MD can characterize physical observables and processes at the molecular level. In order to address the 'why', the first section briefly summarizes the physical basis for molecular simulation by revisiting the key concepts of classical¹ and statistical mechanics². The 'how' is addressed in the second section on molecular dynamics simulations³. A third section covers the advanced technique of free energy calculations⁴. Finally, a fourth section briefly introduces the key concepts of Markov state modelling⁵.

2.1 The Physical Basis for Molecular Simulation

In biomolecular simulation, we want to study a system of interest, e.g. a protein in its natural environment (surrounded by bulk water, ions, eventually also embedded in a membrane). We want to know: How will the system evolve over time? And how can we calculate observables from this information that we could compare to experimental observables? This section provides the physical background as to how these questions can be answered.

2.1.1 Key Concepts of Classical Mechanics

Given some initial conditions and some dynamic law, how will a system evolve over time? Answering this question is the job of classical mechanics. This section introduces some of the key concepts of classical mechanics that form the basis of classical molecular simulation, such as the *state of a system*, a *trajectory*, *phase space*, the *classical equations of motion* as well as the intrinsic physical properties of classical mechanics, such as *determinism*, *time-reversibility* as well as *energy and phase space conservation*.

¹Key sources: [124], [129] and [130]

²Key sources: [124], [131] and [132]

³Key sources: [124], [125], [126], [127], [128] and [133]

⁴Key sources: [124], [134], [135], [136] and [137]

⁵Key sources: [138] and [139]

While classical mechanics was originally developed to explain the motion of the planets and stars as well as other large objects, it turns out that it is an astonishingly good approximation for the molecular world, which is best described by quantum mechanics.

The State of a System

In classical mechanics, given a system of N particles (e.g. atoms) in 3d-space ($\mathbb{R}^3$), all information about the state of the system is contained in the positions $r_1, \dots, r_N \in \mathbb{R}^3$ and velocities $v_1, \dots, v_N \in \mathbb{R}^3$ of all particles. For simplicity of notation, often only r is written for $r_1, \dots, r_N$ and equally v for $v_1, \dots, v_N$.

An example for a system in biomolecular simulation would be a protein in its natural environment (surrounded by bulk water, ions, eventually also embedded in a membrane) and its state would be characterized completely by knowing the positions and velocities of all its particles (usually described at the level of atoms).

The State Space

The set of all possible positions and velocities is called the *state space* ($\mathbb{R}^{6N}$), which is $6N$ -dimensional. Each state of the system, corresponds to a single point in this state space. Sometimes, momenta $p_i = m_i v_i$ (m_i mass of particle i) are used instead of velocities, in which case the state space is called *phase space* and will be denoted by Γ in equations. The phase space can be subdivided into the *configuration space*, which only contains the positions, and the *momentum space*, which only contains the momenta.

To illustrate these abstract terms, consider a protein in solution. It can be unfolded or folded with a plethora of configurations along the folding pathway. Each of these configurations (including the positions of the water particles) correspond to a single point in configuration space. If for each of these configurations, one would also consider the momenta of all particles (points in momentum space), then one would have a complete description of the states of the system (points in phase space).

A Trajectory

As the system evolves in time t , the positions as well as the velocities (or momenta) change. One can picture this as if one would follow the system's evolution in state space with a pencil, connecting an infinite number of states with a line. This 'history' $r_1(t), \dots, r_N(t), v_1(t), \dots, v_N(t)$ of the system is called a *trajectory*.

Going back to the example of protein folding, the pathway from the unfolded state to the folded state (which includes the positions and momenta of all particles, including the solvent at all instants in time) is a trajectory of the folding process.

The Equations of Motion

So how do we obtain such a trajectory? What is the dynamic rule that governs the movement of all particles? This rule was published by Newton in 1687 and laid the foundation for classical mechanics. It relates the force $F \in \mathbb{R}^3$ that acts on a particle to the acceleration $a = \dot{v} = \ddot{r} \in \mathbb{R}^3$ it will experience due to the force where its mass m is the proportionality constant:

$$F = ma. \quad (2.1)$$

While Newton did not know about our modern concept of potential energy, there is a link between the potential energy $U(r)$ and the force acting on a particle i , given by

$$F_i(r) = -\nabla_{r_i} U(r) = -\frac{\partial U(r)}{\partial r_i} \quad (2.2)$$

With equation (2.2), *Newton's equations of motion* write

$$m_i \ddot{r}_i = -\nabla_{r_i} U(r) \quad \forall i = 1, \dots, N, \quad (2.3)$$

yielding a set of $3N$ coupled second order differential equations. This means that if we can calculate the potential energy of a system, then we can predict how it will evolve over time. For now, we will just assume that we can always do so and explain in more detail in section 2.2.2 how the potential energy is calculated in practice.

If we go back to our protein folding example, we now have a dynamic rule that tells us how the atoms will move over time and all we need to know to

calculate the folding pathway is a recipe of how to calculate the potential energy depending on the positions of all particles.

An alternative formulation of classical mechanics was introduced by Hamilton in 1833. While Newton used Cartesian coordinates to describe the positions $r_1, \dots, r_n = x_1, y_1, z_1, \dots, x_N, y_N, z_N$ and velocities $v_1, \dots, v_n = v_{x_1}, v_{y_1}, v_{z_1}, \dots, v_{x_N}, v_{y_N}, v_{z_N}$, Hamilton used the more general framework of so-called generalized coordinates $q_1, \dots, q_n$ for positions and $p_1, \dots, p_n$ for momenta ($n = 3N$). The generalized coordinates are usually chosen such that for a specific system the solution of the equations of motion becomes easier. Here, it is sufficient to just think of these generalized coordinates as the positions and velocities in Cartesian space. Again, for brevity one usually writes q for $q_1, \dots, q_n$ and p for $p_1, \dots, p_n$. The total energy of a system is described by the *Hamiltonian* as the sum of the kinetic energy T and the potential energy U :

$$H(q, p, t) = T + U = \sum_{i=1}^n \frac{p_i^2}{2m_i} + U(q, p, t). \quad (2.4)$$

In the following, only the case of time-independent potentials, so-called *conservative* potentials $U(q, p)$ will be considered, such that the Hamiltonian is also time-independent ($H(q, p)$). Hamilton's equations of motion write $\forall i = 1, \dots, n$:

$$\begin{aligned} \dot{p}_i &= - \frac{\partial H}{\partial q_i}, \\ \dot{q}_i &= \frac{\partial H}{\partial p_i}. \end{aligned} \quad (2.5)$$

This is a set of $6N$ coupled first order differential equations. It can be shown that Hamilton's formulation of classical mechanics is equivalent to Newton's. Both formulations have their own advantages, as some insights are easier obtained from one or the other. It is therefore useful to be familiar with both of them.

Key Properties of Classical Mechanics

Classical mechanics has some intrinsic properties which any simulation method, that aims to generate dynamics of a classical system, should satisfy.

First, classical mechanics is *deterministic*, i.e. given a state (all positions and velocities) of a system, it is uniquely determined how the system will evolve in the

future. This is a direct consequence of the equations of motion which have a unique solution, given an initial set of all coordinates and velocities (*initial conditions*).

Second, classical mechanics is *time-reversible*, i.e. a system's trajectory is the same if propagated forwards or backwards in time. Under a reversal of the direction of time $t \rightarrow -t$, the positions and momenta transform according to $q_i \rightarrow q_i$ and $p_i = m_i \frac{dq_i}{dt} \rightarrow m_i \frac{dq_i}{d(-t)} = -p_i$. Thus, Hamilton's equations of motion transform according to

$$\begin{aligned} \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} &\rightarrow \frac{d(-p_i)}{d(-t)} = -\frac{\partial H}{\partial q_i} \Leftrightarrow \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \\ \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} &\rightarrow \frac{dq_i}{d(-t)} = \frac{\partial H}{\partial(-p_i)} \Leftrightarrow \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}. \end{aligned}$$

As can be seen, Hamilton's equations of motion remain the same when reversing the direction of time.

Third, in classical mechanics the total energy is conserved. To prove this, recall that the total energy is given by the Hamiltonian $H(q, p)$ and conservation of energy means that it does not change over time, i.e. $\frac{dH(q,p)}{dt} = 0$. Writing out the time derivative of a conservative Hamiltonian yields:

$$\begin{aligned} \frac{dH(q(t), p(t))}{dt} &\stackrel{\text{chain rule}}{=} \sum_{i=1}^{3N} \left(\frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right) \\ &\stackrel{\text{Hamilton's equations of motion}}{=} \sum_{i=1}^{3N} \left(\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \\ &\stackrel{\text{Schwarz's theorem}}{=} 0 \quad \square. \end{aligned}$$

Note that this proof did not make use of a specific functional form of the Hamiltonian, it was only assumed that it has continuous partial second derivatives and that Hamilton's equations of motion hold.

Fourth, in classical mechanics *phase space volume is conserved* with time. This is one of the central theorems of classical mechanics and is commonly known as *Liouville's theorem*. What does it mean? Recall that each point in phase space corresponds to a state of the system. Each of these points could be the starting point of a trajectory which evolves through the phase space according to the equations of motion. One can picture how one could track this time evolution by following it

with a pencil and obtain a curve that represents the trajectory. Now imagine you select an infinite number of points in phase space and track their evolution with an infinite number of pencils. One can think of this infinite number of points that move through phase space according to the equations of motion as the flow of a fluid, the *phase space fluid*. Liouville's theorem means that this fluid cannot be compressed or decompressed over time. We will see in the next section 2.1.2 that a consequence of this theorem is that observables obtained as time averages from molecular dynamics simulations, are time-independent, as would be expected for an equilibrium observable. The proof of Liouville's theorem is straightforward if one makes an analogy to fluid dynamics. Imagine a regular fluid, composed of particles in 3d-space. The flow of such a fluid can be described by a velocity field at each point in 3d-space $v(x, y, z) = (v_x, v_y, v_z) \in \mathbb{R}^3$. Let us assume that the flow does not depend on time, i.e. it is *stationary*. Imagine an infinitesimally small cube at an arbitrary position within the fluid. The amount of fluid particles in this cube is constant over time, if the same amount of particles that enters the cube also leaves the cube per time increment. This is true if the overall flow velocity through the walls of the cube is constant or, mathematically speaking

$$\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = 0 \Leftrightarrow \nabla \cdot v = 0,$$

where $\nabla \cdot v$ is called the divergence. Now let us go back to the fluid flow in phase space. Here, we do not look at particles that move in 3d-space, but we look at system states that move through the 6N-dimensional phase space. The corresponding velocity field of the phase space fluid is thus 6N-dimensional $v = \dot{q}_1, \dots, \dot{q}_{3N}, \dot{p}_1, \dots, \dot{p}_{3N}$. The generalization of the concept of divergence to higher dimensions is straightforward and for the 6N-dimensional phase space fluid vector field given by

$$\nabla \cdot v = \sum_{i=1}^{3N} \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right).$$

Insertion of Hamilton's equations of motion yields:

$$\nabla \cdot v = \sum_{i=1}^{3N} \left(\frac{\partial}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \stackrel{\text{Schwarz's theorem}}{=} 0 \quad \square.$$

Thus, the phase space fluid is neither incompressible nor decompressible, i.e. phase space volume is conserved over time.

2.1.2 Key Concepts of Statistical Mechanics

Having answered our original question of how to evolve a biomolecular system over time, it remains to clarify how we can calculate observables of interest from this information that we could compare to experimental observables. The answer to this question is provided by the physical theory of statistical mechanics. Statistical mechanics forms the bridge between the microscopic description of individual atoms and macroscopic observables such as temperature, pressure, entropy, free energy etc. If one knows the physical laws that govern the microscopic world, how atoms interact and move, the apparatus of statistical mechanics allows to derive macroscopic properties of the system in a rigorous way. Thus, it provides the microscopic basis for thermodynamics and, moreover, enables the calculation of additional properties not dealt with in thermodynamics, such as structural properties via probability measures or dynamical properties (e.g. rate constants or spectra) via time correlation functions.

The Ensemble Concept

The foundation of statistical mechanics is the idea of an *ensemble*, first introduced by Gibbs. An ensemble is a virtual collection of a very large number of systems with the same microscopic interactions, which can be described by a Hamiltonian, and the same set of macroscopic properties, such as number and type of molecules, volume or total energy. All members of the ensemble are replica of the same system and identical from a thermodynamic perspective, as they all share the same set of macroscopic properties, however, microscopically, they are not necessarily identical, as many different microscopic states can give rise to the same macroscopic state. It is assumed that each possible microscopic state is represented an equal number of times in the ensemble. This is a postulate of statistical mechanics and called the *principle of equal a priori probabilities*. An observable which can be expressed as a function of the microscopic degrees of freedom is then defined as the ensemble average of this function over all members of the ensemble.

Ensemble Averages

Within the framework of classical ensemble theory, every macroscopic observable A can be obtained using a microscopic function of the coordinates and momenta of all particles $a(q, p)$. Here, we assume that the system is in thermodynamic equilibrium, such that the probability of finding a system in a certain state does not change with time. In thermodynamic equilibrium, the connection between A and $a(q, p)$ is given by an ensemble average

$$A = \langle a(q, p) \rangle := \int_{\Gamma} a(q, p) f(q, p) dq dp. \quad (2.6)$$

The probability density $f(q, p)$ can be written as a function F of the Hamiltonian $H(q, p)$ which is normalized via a normalization factor $1/Z$:

$$f(q, p) = \frac{1}{Z} F(H(q, p)), \quad (2.7)$$

where

$$Z = \int_{\Gamma} F(H(q, p)) dq dp \quad (2.8)$$

is the so-called *partition function*, which is a measure of the number of accessible states in the phase space of a given ensemble. Equations (2.6) and (2.8) represent the central results of equilibrium statistical mechanics, because they determine all observable quantities (including all thermodynamic properties). The specific form of the probability density $f(q, p)$ and thus the partition function depends on the type of ensemble, i.e. the macroscopic variables used to define the ensemble. The main job of statistical mechanics is to derive this probability measure for the different types of ensembles. However, solving the integral (2.6) analytically to obtain an observable of interest is only feasible for very simple systems. For biomolecular systems, it is necessary to use approximative methods, such as molecular dynamics simulations in order to calculate the observables of interest.

Ergodic Hypothesis

The *ergodic hypothesis* is the central assumption that allows the calculation of equilibrium observables from molecular dynamics simulations. This hypothesis states that the ensemble average of an observable can be replaced by the time average over an ergodic trajectory in the limit of infinite sampling:

$$\langle a(q, p) \rangle := \frac{1}{Z} \int_{\Gamma} a(q, p) F(H(q, p)) \, dq dp = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T a(q(t), p(t)) dt =: \bar{a}. \quad (2.9)$$

With an MD simulation an estimate of this time average $\bar{a}$ of an observable A of interest can be calculated. The MD-trajectory consists of the positions and momenta at discrete time steps, resulting in a discrete number of M simulation frames. Due to Liouville's theorem, all time frames contribute equally to the overall time average, such that it is obtained via

$$\bar{a} \approx \frac{1}{M} \sum_{i=1}^M a(q_i, p_i). \quad (2.10)$$

In the limit of $M \rightarrow \infty$, this approximation becomes exact, provided the system is ergodic.

2.2 Molecular Dynamics

The calculation of the integrals to obtain ensemble averages (section 2.1.2, equation (2.6)) for a general many-body system with non-linear interactions is analytically intractable. Therefore, approximate methods must be applied in practice.

Molecular dynamics is such a method that has the advantage that both equilibrium averages as well as the system's dynamics can be obtained simultaneously. In classical molecular dynamics, the nuclei are treated as classical point particles that evolve according to the classical equations of motion (2.3) under the influence of a conservative potential that describes the microscopic interactions.

Classical molecular dynamics can be derived through several approximations from non-relativistic quantum mechanics as outlined in Chapter 2 of the textbook by Marx and Hutter [140]. One of the key approximations in this context is the

Born-Oppenheimer approximation, which assumes that the motion of the nuclei and the electrons can be treated separately. This approximation is based on the large mass difference between nuclei and electrons, such that one can assume that the electrons follow the nuclei instantaneously (adiabatically). In the Born-Oppenheimer approximation the kinetic energy of the nuclei is neglected and their positions enter the Schrödinger equation parametrically. This allows solving the electronic Schrödinger equation separately for any given configuration of clamped nuclei. If one solves the electronic Schrödinger equation for many configurations of fixed nuclear positions, one obtains the corresponding energy values that lie on the so-called Born-Oppenheimer potential energy surface U^{BO} . The second main approximation is to treat the nuclei as classical point particles that move according to the classical equations of motion under the effective potential given by the Born–Oppenheimer potential energy surface. Typically the electrons are assumed to be in the ground state and thus the energy of the Born-Oppenheimer potential energy surface is the ground-state solution U_0^{BO} of the time-independent electronic Schrödinger equation with clamped nuclei.

For intuitive reasons, the theory of MD is explained here in terms of atoms as point particles. However, the principles of MD also apply in a more general sense to point particles that carry a mass and somehow interact with each other through a Hamiltonian. They are therefore also applicable to coarser models of molecular matter, where groups of atoms can be represented by a single particle.

2.2.1 Integration of the Equations of Motion

Solving the equations of motion equation (2.3) analytically is only possible for very simple systems, e.g. the harmonic oscillator. Systems that are interesting for biomolecular simulation, however, are far too complex and require an approximated numerical solution. To this end, the time evolution of the system is discretized into very small discrete time steps Δt , such that one obtains a trajectory consisting of discrete positions and velocities. The simplest way to obtain a numerical integration scheme is to use Taylor expansions.

Verlet Algorithm

The Verlet algorithm can be derived by Taylor expansions of the positions forward and backward in time. For an arbitrary particle i , we obtain

$$r_i(t + \Delta t) = r_i(t) + \dot{r}_i(t)\Delta t + \frac{1}{2}\ddot{r}_i(t)\Delta t^2 + \frac{1}{3!}\dddot{r}_i(t)\Delta t^3 + \mathcal{O}(\Delta t^4) \quad (2.11)$$

$$\begin{aligned} r_i(t - \Delta t) &= r_i(t) + \dot{r}_i(t)(-\Delta t) + \frac{1}{2}\ddot{r}_i(t)(-\Delta t)^2 + \frac{1}{3!}\dddot{r}_i(t)(-\Delta t)^3 + \mathcal{O}(\Delta t^4) \\ &= r_i(t) - \dot{r}_i(t)\Delta t + \frac{1}{2}\ddot{r}_i(t)\Delta t^2 - \frac{1}{3!}\dddot{r}_i(t)\Delta t^3 + \mathcal{O}(\Delta t^4). \end{aligned} \quad (2.12)$$

Adding these two expansions cancels out all odd-powered terms, resulting in

$$r_i(t + \Delta t) + r_i(t - \Delta t) = 2r_i(t) + \ddot{r}_i(t)\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.13)$$

$$\Leftrightarrow r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \ddot{r}_i(t)\Delta t^2 + \mathcal{O}(\Delta t^4). \quad (2.14)$$

Since $a_i = \ddot{r}_i$ and using Newton's second law $F = ma$, we can rewrite

$$r_i(t + \Delta t) = 2r_i(t) - r_i(t - \Delta t) + \frac{F_i(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4). \quad (2.15)$$

Neglecting the terms of $\mathcal{O}(\Delta t^4)$ yields a numerical approximation to the true propagation through configuration space, called the Verlet algorithm. Note that this integration scheme can propagate the positions without the need of velocities $v_i = \dot{r}_i$. However, the velocity for any particle can always be obtained retrospectively via

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \quad (2.16)$$

or more accurately via

$$v_i(t) = \frac{r_i(t + \Delta t) - r_i(t - \Delta t)}{2\Delta t} + \frac{F_i(t - \Delta t) - F_i(t + \Delta t)}{12} + \mathcal{O}(\Delta t^3). \quad (2.17)$$

The Verlet algorithm can also be derived via the powerful Trotter factorization scheme [124] and thus fulfils the important properties inherent to classical mechanics and which therefore any numerical integration scheme should have:

1. determinism,

2. time reversibility,
3. energy conservation,
4. conservation of phase space volume (Liouville's theorem).

Moreover, it is also advantageous from a computational perspective, because per time step it only requires one force evaluation which is what dominates the computational effort.

Leap-frog Algorithm

The leap-frog algorithm is a variant of the Verlet-algorithm that is commonly implemented in most MD engines, such as GROMACS [141], and was used for all standard MD simulations in this thesis. In fact it can be shown that it is equivalent to the Verlet algorithm and thus preserves all desirable symmetries of classical mechanics. In this scheme, the velocities and positions of all particles $1, \dots, i, \dots, N$ are evolved via the equations:

$$v_i \left(t + \frac{\Delta t}{2} \right) = v_i \left(t - \frac{\Delta t}{2} \right) + \frac{F_i(t)}{m_i} \Delta t, \quad (2.18)$$

$$r_i(t + \Delta t) = r_i(t) + v_i \left(t + \frac{\Delta t}{2} \right). \quad (2.19)$$

Given the initial positions and velocities, this scheme first propagates the velocities and then the positions. This generates a trajectory in which the positions and velocities are not calculated at the same points in time, but separated by a half time-step. One can visualize this propagation scheme as if the new position propagation leaps a full time step over the already calculated velocity half a time step in front of the old position and then the new velocity leaps a full time step over the previously calculated new position half a time step in front of the old velocity and so forth and so on, giving rise to the name of the scheme.

Initializing the Velocities

As we have seen in the previous section, the leap-frog algorithm requires a set of both the initial positions and the velocities in order to be able to generate a trajectory. The positions are obtained as described in section 2.2.6. But how are the initial velocities obtained? We know that atomic velocities at a macroscopic level correspond to temperature and that the link between the macroscopic temperature and the microscopic velocities is given by the Maxwell-Boltzmann distribution, which for a given temperature T describes the probability density $p(v)$ of finding a particle with mass m at a certain velocity v :

$$p(v) = \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{mv^2}{2k_B T} \right). \quad (2.20)$$

Hence, any desired temperature can be imposed on the simulation system by initiating the atomic velocities according to a Maxwell-Boltzmann distribution, and indeed this is how it is done in practice.

2.2.2 Potential Energy Models (Force Fields)

The potential energy of the electronic ground state of the Born-Oppenheimer potential energy surface U_0^{BO} can be calculated approximately with explicit quantum mechanical electronic structure methods on the fly during a MD simulation. These methods are referred to as *ab-initio* MD with Car-Parrinello MD, Born-Oppenheimer MD or Ehrenfest MD as different flavours. For large systems, however, such direct methods are not efficient enough to reach relevant time scales and simpler models are required.

To this end, the global potential energy surface U_0^{BO} is approximated in terms of a truncated expansion of many-body contributions

$$U_0^{BO} \approx \sum_{i=1}^N U^{(1)}(r_i) + \sum_{i<j}^N U^{(2)}(r_i, r_j) + \sum_{i<j<k}^N U^{(3)}(r_i, r_j, r_k) + \dots \quad (2.21)$$

This potential energy model is only dependent on the positions of the nuclei as degrees of freedom. Different functional forms exist and the associated parameters for specific atom types are determined semi-empirically by both *ab initio* quantum

mechanical calculations as well as from fitting to experimental data. The combination of a functional form for the potential energy together with the associated set of parameters is commonly referred to as a *force field*. Based on the expansion in equation (2.21), force fields can be divided into two classes. *Pairwise additive force fields* only consider terms up to two-body interactions, while *many-body force fields* also include higher order terms.

Pairwise Additive Fixed-Charge Force Fields

Many of the force fields in use today are *pairwise additive fixed charge force fields*. All force fields used in this thesis belong to this class. Their functional form is illustrated in Figure 2.1. The contributions to the potential energy $U(r)$ can be divided into

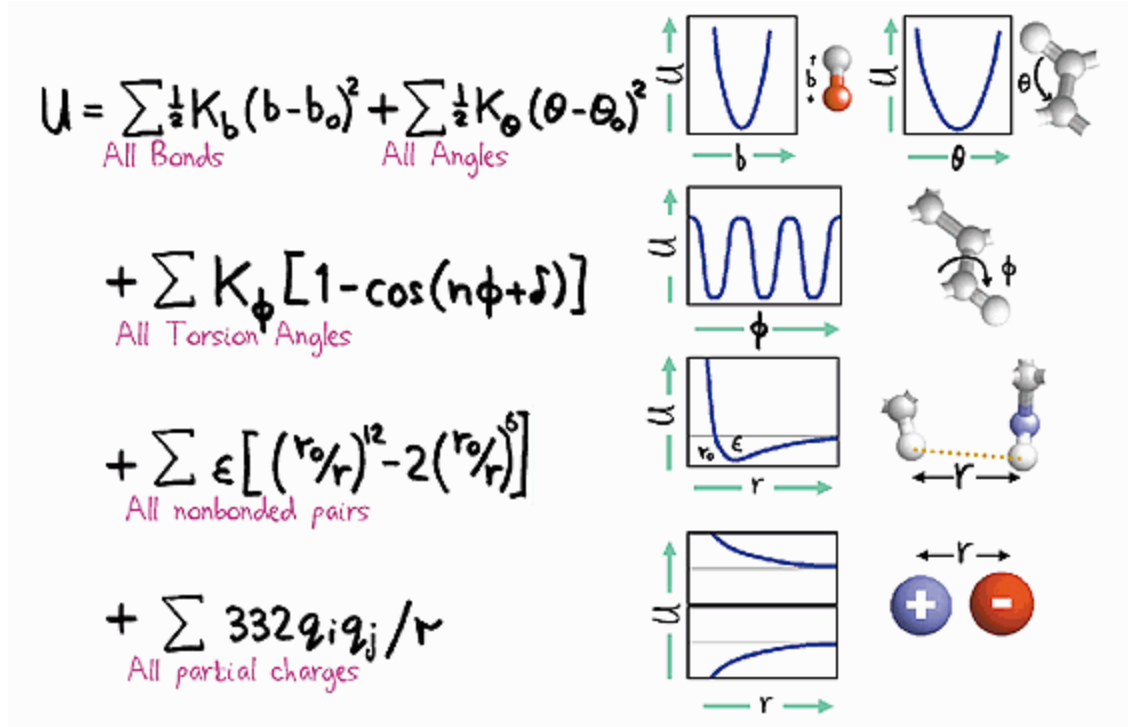


Figure 2.1: Typical functional form of a force field, modelling the potential energy as a sum of pairwise additive interactions. Figure reprinted by permission from Springer Nature, Nature Structural Biology, [142], <https://doi.org/10.1038/87545>, Copyright Springer Nature 2001.

bonded $U_{\text{bonded}}(r)$ and non-bonded interactions $U_{\text{non-bonded}}(r)$, such that

$$U(r) = U_{\text{bonded}}(r) + U_{\text{non-bonded}}(r). \quad (2.22)$$

Three types of bonded interactions are typically accounted for, namely bonds, angles and dihedrals:

1. A covalent *bond* between two atoms with positions r_i, r_j can be approximated by a harmonic potential

$$U_{\text{bond}}(r_i, r_j) = \frac{1}{2}k_{b_{ij}}(r_{ij} - r_{0_{ij}})^2, \quad (2.23)$$

where $r_{ij} = |r_i - r_j|$ is the distance between the two atoms, $r_{0_{ij}}$ is the equilibrium bond length and $k_{b_{ij}}$ is the force constant which determines the rigidity of the bond. Since the bond length typically does not deviate much from its equilibrium value, this second order Taylor expansion is usually a sufficient description. Note, however, that due to this functional form, a bond cannot break. Including higher order terms of the Taylor expansion to describe the bond improves accuracy, but comes with additional computational cost. Another alternative to describe bond stretching is the more realistic Morse potential

$$U_{\text{bond}}^{\text{Morse}}(r_i, r_j) = D_{ij} \left(1 - \exp(-\beta_{ij}(r_{ij} - r_{0_{ij}}))\right)^2. \quad (2.24)$$

2. A covalent *bond angle* between three atoms with positions r_i, r_j, r_k can also be reasonably well described by a harmonic angular potential

$$U_{\text{angle}}(r_i, r_j, r_k) = \frac{1}{2}k_{\theta}(\theta - \theta_0)^2, \quad (2.25)$$

where

$$\theta = \arccos\left(\frac{r_{ij} \cdot r_{jk}}{|r_{ij}r_{jk}|}\right) \quad (2.26)$$

is the angle between the three atoms, θ_0 the equilibrium angle and k_{θ} is the force constant determining how rigid the angle bending is.

3. A covalent *dihedral angle* ϕ between four atoms with positions r_i, r_j, r_k, r_l can be modelled via a periodic function, such as

$$U_{\text{dihedral}}(r_i, r_j, r_k, r_l) = k_{\phi}(1 + \cos(n\phi - \phi_0)), \quad (2.27)$$

where

$$\phi = \arccos \left(\frac{r_{ij} \times r_{kj}}{|r_{ij} \times r_{kj}|} \cdot \frac{r_{jk} \times r_{lk}}{|r_{jk} \times r_{lk}|} \right) \quad (2.28)$$

is the dihedral angle between the normals of the two planes spanned by r_i, r_j, r_k and r_j, r_k, r_l , ϕ_0 is the reference dihedral angle and k_ϕ a constant that determines the magnitude of the energy separation along the dihedral angle. Note that this functional form creates only equal energy minima, such that for example the trans and the two gauche conformations are the same. In order to correct for this, an additional interaction between atoms i and l is introduced, which is called the 1-4 interaction. As an alternative to a 1-4 interaction, one could also use a set of periodic functions with different periodicity or a set of powers of higher order cosine functions, such as the Ryckaert-Bellmans potential:

$$U_{\text{dihedral}}^{\text{RB}}(r_i, r_j, r_k, r_l) = \sum_{n=0}^5 C_n \cos^n \phi. \quad (2.29)$$

In order to enforce planarity of planar groups, such as aromatic rings, or stereochemistry via preventing enantiomers from flipping to their mirror image, *improper dihedrals* are introduced, which act on four atoms via a harmonic restraining potential

$$U_{\text{improper}}(r_i, r_j, r_k, r_l) = \frac{1}{2} k_\phi (\phi - \phi_0)^2. \quad (2.30)$$

All bonded interactions listed above are simply added up, resulting in the total potential energy from bonded interactions:

$$U_{\text{bonded}}(r) = \sum_{\text{bonds}} U_{\text{bond}} + \sum_{\text{angles}} U_{\text{angle}} + \sum_{\text{dihedrals}} U_{\text{dihedral}} + \sum_{\text{improper dihedrals}} U_{\text{improper}}. \quad (2.31)$$

The parameters in the functional forms can in principle all be derived from quantum mechanical calculations. Typically the energies are calculated for a series of molecular configurations for a wide range of atomic species and a potential of the functional forms above is then fitted onto these discrete energies by optimizing the parameters. However, some of these parameters are often inferred from experimental

data, for instance equilibrium bond length and angles from structural data and force constants from spectroscopic data.

As for the non-bonded interactions it should be stressed that they also act intramolecularly with some exclusions. Pairs that are already involved in the bonded interactions of bonds (1-2 interactions) or angles (1-3 interactions) are always excluded from the non-bonded interactions. Pairs involved in 1-4 interactions of dihedrals are either excluded or used in a modified form, depending on the specific force field. Two types of non-bonded interactions are important at a molecular level, namely the electrostatic interactions described by the Coulomb law and the van der Waals interactions which can be described by Lennard-Jones potentials.

1. In fixed-charge force fields, each atom i is assigned a partial charge q_i and the pairwise *Coulomb interactions* due to these charges can be exactly described by Coulomb's law:

$$U_{\text{Coulomb}}(r_i, r_j) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \quad (2.32)$$

where ϵ_0 is the permittivity in vacuum. While the functional form of Coulomb's law is exact, the approximative nature of fixed-charge force fields lies in the use of fixed point charges centred on the atoms. In reality, charge density can be pictured as a cloud that is delocalized over an entire molecule and the influence of the electrostatic environment can induce changes in this charge density, which is called *polarization*. While fixed-charge force fields can include the average polarization, they cannot account for explicit polarization effects due to a changing electrostatic environment. The partial charges are derived from ab initio quantum mechanical calculations which give the true electrostatic potential for a given molecule structure (usually the minimum energy structure obtained from a geometry optimization using a quantum mechanical level of theory). Via a fitting procedure the optimal point charges centred on the atomic positions can be found that best reproduce the overall electrostatic potential. However, the true charge density changes depending on the particular configuration of the molecule and it is thus challenging to

also account for the accessible configuration space at a relevant temperature with fixed-charge models.

2. *Van der Waals interactions* are modelled with Lennard-Jones potentials

$$U_{\text{LJ}}(r_i, r_j) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (2.33)$$

where the r^{-12} term describes short-range repulsion and the r^{-6} term the longer range dispersion interactions. The graph of this function can be seen in Figure 2.1 and its minimum is located at $r_{ij}^{\text{min}} = 2^{1/6}\sigma_{ij}$ with well depth $-\epsilon_{ij}$. Thus, the parameter σ_{ij} controls the position of the minimum and the parameter ϵ_{ij} its depth. Their values are calculated from the single-atom parameters involved in the pairwise interaction, which are defined for each atom type in a force field via the Lorentz-Berthelot combination rules

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \quad (2.34)$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}. \quad (2.35)$$

There are many ways to determine the Lennard-Jones parameters. Usually they are obtained from iterative fitting procedures with the goal to reproduce thermodynamic properties of experimental data such as liquid densities and enthalpies of vapourization.

The total potential energy from non-bonded interactions is in theory obtained as the sum over all pairwise non-bonded interactions (except the exclusions):

$$U_{\text{non-bonded}}(r) = \sum_{\text{pairs}} U_{\text{Coulomb}} + \sum_{\text{pairs}} U_{\text{LJ}}. \quad (2.36)$$

However, in practice, the long-range non-bonded interactions are treated in an approximative manner to reduce computational costs. Further details are presented in section 2.2.4.

Having discussed the functional form of pairwise additive fixed-charge force fields in detail, I want to give a brief overview of popular force fields of this class used in practice in the field of biomolecular simulations. They provide parameters for the

most relevant biochemical species and typically include amino acids, ions and water and often extensions to lipids, nucleic acids and carbohydrates are available. These force fields can be further subdivided by their level of resolution into *all-atom* (AA), *united-atom* (UA) and *coarse-grained* (CG) force fields. The most popular all-atom force fields are the *Assisted Model Building and Energy Refinement* (AMBER) [143–145] and *Chemistry at Harvard Macromolecular Mechanics* (CHARMM) [146] force fields. United-atom force fields incorporate some hydrogen atoms into the heavy atom that they are bound to. Usually these are aliphatic carbon atoms and the resulting CH₂ or CH₃ groups are treated as a single particle in the united-atom force field. Notable examples of this subclass are the *Groningen Molecular Simulation* (GROMOS) [147] and the *Optimized Potential for Liquid Simulations united-atom* (OPLS-UA) [148] force fields. Coarse-grained force fields represent even more atoms as a single particle and the most widely used representative of this class is the *MARTINI* force field [149]. For more detailed information, the reader is referred to reviews discussing force fields in more depth [150–156] and to several benchmark studies comparing different force fields [157–164].

Since many force fields have been developed along different routes based on different principles and data, different force fields cannot simply be mixed and matched. When combining force fields, it is important that they have been parameterized consistently. In this thesis for example, for AA simulations the AMBER99SB-ILDN force field [143] was used for the protein and ions, Slipids [165–167] for lipids, GAFF [168] (version 2) for ligands, parameters by Horn [169] for zwitterionic amino acids and the TIP3P [170] model for water. These force fields have been developed in a consistent way and they can thus be combined. For CG simulations, the MARTINI force field [149] was used which can describe all biomolecular species simulated in this thesis.

2.2.3 Periodic Boundary Conditions and Box Shapes

The computational cost of MD simulations limits the size of systems that can be studied. However, simulating an isolated system of condensed matter in vacuum

would introduce artefacts due to surface effects. This is why periodic boundary conditions as illustrated in Figure 2.2 are used. The simulation box is replicated in all three spacial dimensions to form an infinite lattice. If a molecule leaves the central box, one of its images enters the central box through the opposite boundary.

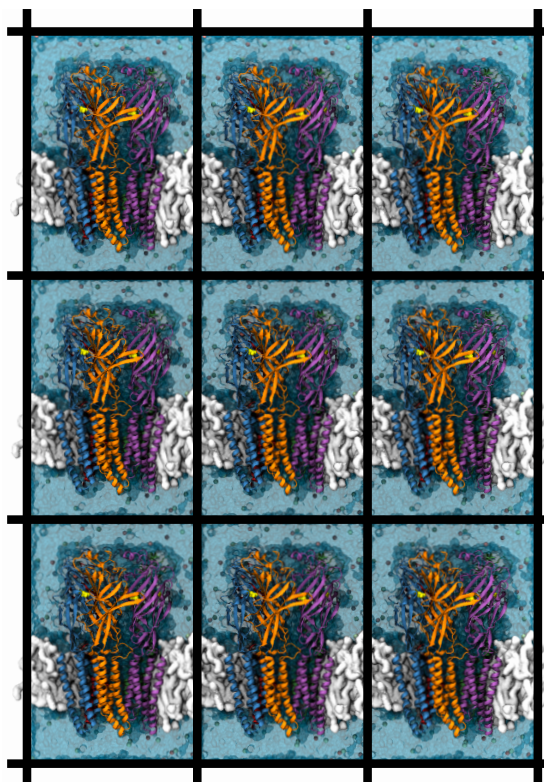


Figure 2.2: Periodic boundary conditions illustrated on a simulation box containing an ion channel in a membrane solvated in water and ions.

When simulating a protein such as in Figure 2.2, it is important to allow for sufficient buffer between the protein and the periodic boundaries to avoid artefacts from the protein interacting non-negligably with one of its periodic images and thus itself. Typically a buffer of 10-15 Å is considered sufficient, depending on the flexibility of the protein and the force field settings.

Note that the unit cell does not necessarily need to be rectangular, but can have any triclinic shape. Since the computational cost of a classical MD simulation scales with the numbers of particles N as $\mathcal{O}(N \ln(N))$, the most efficient box shape is the one that minimizes the box volume. For a solute of arbitrary shape that can freely rotate and translate, the rhombic dodecahedron is the most efficient box

size. For an ion channel with a cylindric shape in a membrane that can only rotate around its pore axis which is perpendicular to the membrane plane, as illustrated in Figure 2.2, in fact the best box geometry is a hexagonal prism (see Figure 2.3).

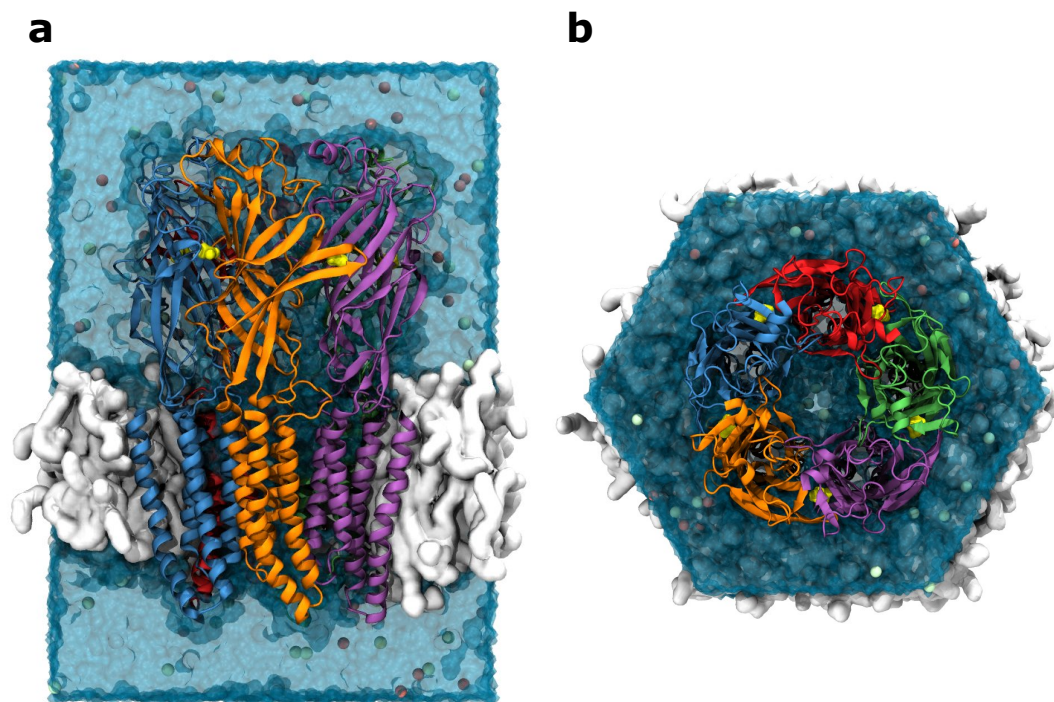


Figure 2.3: A hexagonal prism as optimal box geometry illustrated on a simulation box containing a cylindrically shaped ion channel in a membrane solvated in water and ions. (a) View parallel to the membrane, (b) view from the extracellular towards the intracellular matrix.

The simulation box shown in Figure 2.3 also illustrates that due to the periodic boundary conditions, no membrane potential via different ion concentrations can be simulated, as nothing prohibits ions swapping between the extracellular and intracellular side through the periodic boundaries. However, this could be circumvented by simulating two oppositely oriented membranes, each containing a membrane protein [171], which yields two separate aqueous compartments even with periodic boundaries in all three spatial dimensions.

2.2.4 Common Speed up Techniques

In this section, I discuss common speed up techniques to reduce the computational cost of MD simulations.

Treatment of Long-range Interactions

The calculation of the non-bonded interactions for all N particles in a pairwise manner according to equations (2.33) and (2.32) scales with $\mathcal{O}(N^2)$ and is thus the computationally most expensive part of the overall calculation of the potential energy. To reduce the computational cost, a cut-off distance can be used beyond which the non-bonded interactions are either neglected or are dealt with in a computationally less demanding approximative way.

1. The long range (dispersion) part of the van der Waals interactions according to equation (2.33) has a r^{-6} distance dependence and thus decays fairly rapidly to zero as the distance between particles increases. Therefore, a cut-off scheme can be used within good approximation, where the interactions are calculated explicitly within the cut-off distance r_c and a tail correction is used to account for this finite truncation.

If we truncate a potential $U(r)$ at a cutoff-distance r_c , the tail correction in 3d space can be estimated via

$$U_{\text{tail}} = \frac{N^2}{2V} \int_{r_c}^{\infty} U(r) g(r) 4\pi r^2 dr. \quad (2.37)$$

where V is the system's volume and $g(r)$ the particle-particle radial distribution function [126]. Note that $\lim_{r \rightarrow \infty} g(r) = 1$, and $g(r)$ can be approximated to be 1 for $r > r_c$. Insertion of the Lennard-Jones potential $\sum_{i < j}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$ for $U(r)$ then gives the dispersion correction to the potential energy:

$$U_{\text{tail}}^{LJ} = -\frac{8\pi N^2}{3V} \sum_{i < j}^N \frac{\epsilon_{ij} \sigma_{ij}^6}{r_c^3} \left(\frac{\sigma_{ij}^6}{r_c^6} - 1 \right). \quad (2.38)$$

In simulations with constant pressure, a similar correction is applied to the pressure (see [125, 127, 150] for more details).

Since an abrupt cut-off of the van der Waals interactions is necessarily artificial, different strategies have been developed to deal with this discontinuity. A popular approach is to apply a switching function that scales the interactions

smoothly to zero within a small range near the cut-off distance and a more detailed discussion can be found in [125–127].

2. Compared to the long range dispersion interactions, the long range Coulomb interactions decay much slower with r^{-1} (cf. equation (2.32)) and therefore, applying a cut-off distance beyond which the electrostatic interactions are neglected introduces a large error [172]. A simple tail correction cannot be used, because the integral in equation (2.37) diverges if $n \geq 3$ in $U(r) \sim r^{-n}$ [173].

One of the main approaches to deal with this issue is the so-called *Ewald summation*. Assuming charge neutrality, the total Coulomb energy under periodic boundary conditions is given by

$$U_{\text{Coulomb}}(r) = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum'_{m \in \mathbb{Z}^3} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{r_{ij} + mL}, \quad (2.39)$$

where L is the box size and the outer sum runs over all replica of the unit cell, where the prime indicates that $i = j$ is omitted for $m = 0$. However, the sum in equation (2.39) is only conditionally convergent and thus not suitable for direct numerical evaluation. The basic idea of Ewald's ansatz is to express equation (2.39) as quickly converging sums via formally replacing the $1/r$ interaction via

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}. \quad (2.40)$$

A usual choice for $f(r)$ is the error function $\text{erfc}(\alpha r) := \frac{1}{\sqrt{\pi}} \int_{\alpha r}^{\infty} e^{-t^2} dt$, so that the point charges are replaced by Gaussian charge distributions such that the total charge distribution is not modified. A detailed derivation can be found in [127]. The result is a decomposition of equation (2.39):

$$U_{\text{Coulomb}} = U_{\text{real}} + U_{\text{reciprocal}} + U_{\text{self}} + U_{\text{dipole}}, \quad (2.41)$$

where U_{real} is a short-range component that quickly converges in real space, $U_{\text{reciprocal}}$ a long-range component that quickly converges in reciprocal space,

U_{self} the self energy and U_{dipole} the dipole correction [174]. The complexity of the Ewald summation is $\mathcal{O}(N^{3/2})$.

A further improvement to $\mathcal{O}(N \ln(N))$ can be achieved with the *particle mesh Ewald* (PME) method [175]. Here, the principal idea is to approximate the reciprocal part on a mesh by solving Poisson's equation of a smooth charge distribution on a grid using fast Fourier transforms. The smooth PME method is a further development that makes use of efficient B-spline interpolation of the structure factors [176].

A computationally much cheaper, but less accurate, alternative to lattice based methods is the *reaction field* approach, introduced by Barker and Watts [177] and later generalized for ionic systems [178]. Within a cut-off distance the electrostatic interactions are calculated explicitly via Coulomb's law (equation (2.32)), beyond it the environment is treated in a mean-field approximation with a constant dielectric constant.

In practice, when choosing the cut-off distances for the non-bonded interactions, it is important to keep in mind that the force fields themselves have been parameterized with a certain long-range treatment and the resulting parameters are thus optimized for these specifications. For example, the AMBER99SB-ILDN force field used for the protein in AA simulations in this thesis was parameterized with a 10 Å cut-off for the van der Waals interactions, using dispersion correction to correct for this finite truncation, and also 10 Å for the short-range contribution of the Coulomb interactions, using PME for the long-range part.

Neighbour Lists

Since the non-bonded interactions are only calculated explicitly within a certain cut-off as discussed in the previous section 2.2.4, it is necessary to determine for each particle which particles are within that cut-off distance which scales with $\mathcal{O}(N^2)$ if all possible pairs are checked. *Neighbour lists* are an efficient way to reduce the computational cost associated with pair searches. The idea is to keep

a list of all neighbours within the cut-off distance plus a buffer for each particle which is only updated at certain intervals, which typically range between 10-20 integration steps. For the non-bonded force calculation of each particle, only the particles in the neighbour list need to be considered and if the distance is smaller than the cut-off, they are included in the force calculation. The buffer ensures that particles that are in close proximity of the cut-off distance can be taken into account if they move within the cut-off distance in between list updates. Automatic list update algorithms also exist (see [125] for more details).

Holonomic Constraints

For stable integration the size of the time step is limited by the fastest time scale in the system. For atomistic systems, these are the bond vibrations, particularly of bonds involving the light hydrogen atoms and limits the time step to 1 fs. Removing these fastest motions by constraining the bonds involving hydrogens to their equilibrium length allows for a time step of 2 fs and thus doubling computational efficiency. Sometimes, all bonds are constrained, which enables a time step of ca. 2.5 fs. Such bond constraints are common practice, since bond vibrations themselves are of little interest in understanding biomolecular systems. Constraining bond angles is not recommended, as this may lead to artefacts in the conformational dynamics [179].

Several algorithms to implement such holonomic constraints exist, whose underlying principle is to reset bonds to their equilibrium length after the positions have been updated [127]. SHAKE [180] was designed for the Verlet algorithm, while RATTLE [181] was developed for the velocity Verlet version [125]. Nowadays, the LINCS algorithm [182, 183] is very popular due to its stability and robustness [127]. For small molecules, the constraint equations can be solved analytically via the SETTLE algorithm [184], which is commonly used to constrain water molecules. For details on how to consider constraints in the equations of motion and the algorithmic details of implementation it is referred to [124, 125].

2.2.5 Controlling Temperature and Pressure: Thermostats and Barostats

MD simulations as presented so far with a constant number of particles N in a fixed volume V and constant total energy E produces states representative of a *microcanonical* (NVE) ensemble. However, biomolecular processes occur under constant temperature (T) and pressure (P). In order to mimic these natural conditions and also to be able to compare simulation results to experimental data, simulations should sample from an *isothermic-isobaric* (NPT) ensemble. To this end, the simulation system is coupled to a thermostat and a barostat. While a wide range of these have been developed (see [133] for an overview with further references for details), here I only review the ones used in this thesis.

Thermostats

In classical statistical mechanics the *equipartition theorem* relates the temperature T of an N -particle system in thermodynamic equilibrium to its average kinetic energy via

$$\left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle = \frac{3}{2} N k_B T. \quad (2.42)$$

This result does not depend on the particular form of the conservative potential energy function (force field) [185]. The *instantaneous temperature* T_{inst} is thus defined as

$$T_{\text{inst}} = \frac{1}{3Nk_B} \sum_{i=1}^N m_i v_i^2 \quad (2.43)$$

and its average is equal to the equilibrium temperature. If constraints, such as fixed bonds, are placed on the system and the overall translation is removed during the simulation, the degrees of freedom are reduced to $N_f = N - N_C - 3$, where N_C is the number of independent internal constraints [125]. T_{inst} can thus be readily calculated at any velocity update time step. The goal of a thermostat is to modify the velocities v_i such that on average a constant temperature is maintained and moreover to generate the correct temperature fluctuations of the ensemble.

The idea behind the *Berendsen thermostat* [186] is to rescale the velocities such that the instantaneous temperature T_{inst} of the system decays to the target temperature T_{target} as:

$$\frac{dT_{\text{inst}}}{dt} = \frac{T_{\text{target}} - T_{\text{inst}}}{\tau}, \quad (2.44)$$

which causes the system's instantaneous temperature to decay exponentially to the target temperature with a time constant τ [127]. The specific implementation of this idea is to rescale the velocities with a scaling factor λ given by

$$\lambda^2 = 1 + \frac{\Delta t}{\tau} \left(\frac{T_{\text{target}}}{T_{\text{inst}}} - 1 \right). \quad (2.45)$$

With the coupling constant τ , the strength of the temperature coupling can be controlled. However, the Berendsen thermostat does not produce the distribution of a canonical (NVT) ensemble [125].

A popular thermostat that does produce the correct canonical ensemble is the *Bussi thermostat* [187], which combines velocity rescaling with a stochastic approach. It is a modification of the velocity rescaling method, where the velocities are scaled by a factor $\sqrt{\bar{K}/K}$, where K is the instantaneous kinetic energy and $\bar{K}$ is the target kinetic energy. Instead of using a fixed target kinetic energy $\bar{K}$, the key idea of the Bussi thermostat is to select the target value K_t stochastically, such that is drawn from the canonical equilibrium distribution of the kinetic energy. This can be achieved as follows. At a coupling step, the kinetic energy K can be calculated and then propagated using an auxiliary continuous stochastic dynamics

$$dK = (\bar{K} - K) \frac{dt}{\tau} + 2 \sqrt{\frac{K \bar{K}}{N_f}} \frac{dW}{\sqrt{\tau}}, \quad (2.46)$$

where N_f is the number of degrees of freedom, $\bar{K}$ is the target kinetic energy, dW denotes a Wiener process and the coupling constant τ determines the time scale of the thermostat. Note that the first term in equation (2.46) is equivalent to the Berendsen thermostat and thus causing an exponential relaxation to the target kinetic energy (and thus the target temperature). The second term is a noise term

that ensures the correct fluctuations. The velocities can then be rescaled so as to enforce this new value of kinetic energy K_t via

$$\alpha = \sqrt{\frac{K_t}{K}}. \quad (2.47)$$

In practice, the Bussi thermostat is nowadays commonly used in standard MD simulations, since it produces the correct ensemble and only minimally affects the system's dynamics [188, 189].

Another stochastic approach of controlling the temperature corresponds to *Langevin dynamics*. In this framework the microcanonical equations of motion are supplemented with Brownian dynamics, mimicking an implicit solvent via inclusion of viscosity and random collisions [133]:

$$\dot{q}_i = \frac{p_i(t)}{m_i}, \quad (2.48)$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i} + g_i(t) \quad \forall i = 1, \dots, N_f, \quad (2.49)$$

where

$$g_i(t) = -\gamma p_i(t) + \sigma \dot{W}_i \quad (2.50)$$

is the additional term that is added to Hamilton's equations of motion. $-\gamma p_i(t)$ corresponds to a frictional force, where γ is the friction coefficient. $\sigma \dot{W}_i$ mimics random collisions via a vector $W_i(t)$ of N_f independent Wiener noises as if an implicit solvent was present. The parameter σ controls the strength of the random forces and is linked to γ via $\sigma = \sqrt{2m_i\gamma k_B T}$. Details of the algorithmic implementation can be found in [125]. In alchemical free energy calculations (see section 2.3), Langevin dynamics is used for temperature control, because it generates the correct distribution of kinetic energies for decoupled states [190], which is important when decoupling the ligand during the alchemical cycle [191].

Barostats

The instantaneous pressure of a classical N -particle system can be calculated efficiently during an MD simulation using the *virial theorem* as formulated in [192]:

$$P_{\text{inst}} = \frac{1}{3V} \left(\sum_{i=1}^N m_i v_i^2 - \sum_{i=1}^N r_i \nabla_{r_i} U(r) \right). \quad (2.51)$$

Since the force $-\nabla_{r_i}U(r)$ acting on each particle is evaluated at every time step, the evaluation of the pressure is computationally inexpensive. Similar to thermostats, the aim of a barostat is to maintain an average constant pressure with the correct pressure fluctuations. As can be seen from equation (2.51), this can be achieved by modifying the volume of the simulation box as well as the particle coordinates.

The *Berendsen barostat* [186] follows an analogous approach as the Berendsen thermostat. The pressure is regulated according to the first order differential equation

$$\frac{dP_{\text{inst}}}{dt} = \frac{P_{\text{target}} - P_{\text{inst}}}{\tau}, \quad (2.52)$$

which causes the system's instantaneous pressure to decay exponentially to the target pressure with a time constant τ . This is achieved by scaling the coordinates and volume according to

$$r' = \chi r \quad (2.53)$$

$$V' = \chi^3 V, \quad (2.54)$$

where the scaling factor χ is given by

$$\chi^3 = 1 - \beta_T \frac{\Delta t}{\tau} (P_{\text{target}} - P_{\text{inst}}). \quad (2.55)$$

For this, the isothermal compressibility $\beta_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T$ is assumed to be known, but as it only enters as a product with $\Delta t/\tau$, its value does not need to be known precisely [127]. In biomolecular simulations, it is common practice to use the compressibility of water under standard conditions ($4.5 \cdot 10^{-5} \text{ bar}^{-1}$ [193]). The Berendsen barostat is well suited to establish a target pressure quickly and is therefore often used in the equilibration phase. However, it does not generate a known ensemble [127] and therefore is not recommended for production runs.

The *Parrinello-Rahman barostat* [194] on the other hand is well suited to produce the correct ensemble and it is therefore recommended for production runs. In this method additional degrees of freedom are added to the Hamiltonian such that a modified set of equations of motion is obtained. The equations involved are

somewhat more complex and it is referred to the original paper for their derivation [194]. The Parrinello-Rahman barostat has the advantage that the vectors spanning the simulation cell can be scaled independently.

If the simulation system is homogenous, the barostat is applied in an isotropic fashion, i.e. all three dimensions are treated equally. For simulations containing a lipid bilayer, however, the barostat should be applied semi-isotropically, such that the directions which span the membrane plane are scaled independently from the direction orthogonal to the membrane plane [195].

2.2.6 The Practical Workflow of a MD Simulation

The practical workflow of a MD simulation consists of the following steps.

System Setup

The goal of the system setup from an abstract point of view is to provide all particle positions as the configurational part of the initial conditions to be able to evolve the system according to the classical equations of motion. In biomolecular simulation, one tries to model the system that is to be studied as realistically as possible based on experimental information available. Here, I exemplify the modelling procedure of a system to study a membrane protein. The protein structure is obtained from structural data from cryo-EM, X-ray crystallography or NMR. If no structural data for the organism of choice is available, homology modelling is used to generate a model based on an available template structure. Typically some additional computational modelling is necessary to model in eventually missing loops or not fully resolved side chains in the structure and to add hydrogen atoms. Here it is a modeller's choice how to protonate titratable residues or the histidine tautomers. A standard approach is to use protonation states at physiological pH and choose histidine tautomers such that H-bond networks are optimized. The protein is then embedded in a lipid bilayer and water as well as ions are added at a physiological concentration. Adding ions also serves the purpose of creating overall charge neutrality of the system which is important if PME is used for the

long-range treatment of the electrostatic interactions. To reduce the computational cost of the MD simulation, the box geometry should be chosen such that the system size is as small as possible as discussed in section 2.2.3.

Energy Minimization

The system's potential energy $U(r)$ is usually very high after the initial system setup, particularly when some particles clash. If an MD simulation was started right away, this would lead to high forces that would cause numerical instability of the integration scheme and the simulation might crash. Therefore, an *energy minimization* is used to relax the system into a local energy minimum, e.g via the steepest descent method. In fact, it is only necessary to ensure that the maximum force falls below a certain tolerance and usually the energy minimization is terminated once this threshold is reached.

Equilibration

The purpose of the *equilibration* is to establish the system at its target temperature and pressure as well as to let all system components such as the lipid bilayer, solvent and protein relax due to their interactions. Usually this process is split up into two phases. First the temperature is equilibrated in the NVT-ensemble. The velocity initialization via the Maxwell-Boltzmann distribution might create very large velocities on a few particles which are quickly equilibrated due to energy exchange between all particles and the action of a thermostat. Since the initial placement of solvent molecules in a grid-like fashion exerts large artificial forces on the protein, the protein heavy atoms are kept close to their original positions with harmonic restraints. In the second equilibration phase, a barostat is coupled to the system such that the target pressure can be established in the NPT-ensemble. During this phase the position restraints on the protein are typically released gradually.

Production Run

Once the system is relaxed and the target temperature and pressure are established, the system's dynamics samples from the correct NPT-ensemble. All analysis is based on this part of the simulation, called the *production run*.

2.3 Free Energy Calculations

Free energy is the work needed to carry out a thermodynamic transformation of a system from one state to another. It is a useful concept, because it determines whether a process occurs spontaneously or not. Any spontaneous process minimizes the free energy until it reaches its minimum at thermodynamic equilibrium. The rate with which a state is reached from another state depends on the free energy barrier separating the two states. In the NVT-ensemble the free energy is called the *Helmholtz free energy* A and is linked to the partition function via

$$A(N, V, T) = -k_B T \ln Z_{NVT}, \quad (2.56)$$

where for a classical N -particle system

$$Z_{NVT} = C_N \int_{\Gamma} \exp \left(\frac{H(q, p)}{k_B T} \right) dq dp \quad (2.57)$$

is the partition function of the NVT ensemble with prefactor

$$C_N = \frac{1}{h^{3N} N_1! N_2! \dots} \quad (2.58)$$

for a multicomponent system with N_1 particles of type 1, N_2 particles of type 2,..., and h is the Planck constant. In the NPT-ensemble the free energy is called the *Gibbs free energy* G and given analogously as

$$G(N, P, T) = -k_B T \ln Z_{NPT}, \quad (2.59)$$

where for a classical N -particle system

$$Z_{NPT} = I_N \int_V \int_{\Gamma} \exp \left(\frac{H(q, p) + PV}{k_B T} \right) dq dp dV \quad (2.60)$$

is the partition function of the NPT ensemble with coefficient

$$I_N = \frac{1}{V_0 C_N} \quad (2.61)$$

for a multicomponent system as above with an additional reference volume V_0 to make the overall expression dimensionless.

The Helmholtz free energy difference between two states 1 and 0 expressed in terms of equation (2.56) writes

$$\Delta A_{1,0} = A_1 - A_0 = -k_B T \ln \left(\frac{\int_{\Gamma_1} \exp \left(\frac{H_1(q,p)}{k_B T} \right) dq dp}{\int_{\Gamma_0} \exp \left(\frac{H_0(q,p)}{k_B T} \right) dq dp} \right). \quad (2.62)$$

Analogously, the Gibbs free energy difference between two states 1 and 0 is obtained via equation (2.59) as

$$\Delta G_{1,0} = G_1 - G_0 = -k_B T \ln \left(\frac{\int_{V_1} \int_{\Gamma_1} \exp \left(\frac{H_1(q,p) + PV}{k_B T} \right) dq dp dV}{\int_{V_0} \int_{\Gamma_0} \exp \left(\frac{H_0(q,p) + PV}{k_B T} \right) dq dp dV} \right). \quad (2.63)$$

It is worth pointing out, that the free energy change due to changes in the average volume are small at physiological pressures [136]. Thus, in biomolecular contexts, the PV component to the free energy can be neglected as long as the average volume corresponding to a certain pressure is enforced in simulations by a barostat. Consequently within this approximation, the Gibbs free energy difference can be replaced by the Helmholtz free energy difference.

In the context of reversible ligand binding to a receptor



the *affinity* is commonly described by the equilibrium *dissociation constant*

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}} = \frac{[L][R]}{[LR]}, \quad (2.65)$$

where $[X]$ denotes the molar concentration of species X at equilibrium. Under isothermal-isobaric conditions the Gibbs free energy of binding $\Delta G_{\text{binding}}$ is linked to the equilibrium dissociation constant K_d via

$$\Delta G_{\text{binding}} = -RT \ln \left(\frac{K_d}{c^0} \right), \quad (2.66)$$

where the standard state concentration c^0 of 1 mol/L makes the argument of the logarithm dimensionless and R is the ideal gas constant ($8.314 \text{ JK}^{-1}\text{mol}^{-1}$) which is related to the Boltzmann constant k_B by Avagadro's number N_A via $R=N_A k_B$.

While there are several ways to calculate $\Delta G_{\text{binding}}$ computationally, *alchemical free energy calculations* have recently been demonstrated to be very accurate to determine binding free energies of ligands to bromodomains [196, 197]. This approach has been used in this thesis to calculate the $\Delta G_{\text{binding}}$ of various agonists to the GlyR and this section provides a brief theoretical background of this method.

2.3.1 The Alchemical Cycle

Alchemical free energy calculations take advantage of the fact that the free energy is a state function and as such does not depend on the path taken to arrive at a specific state. More specifically, the ligand-binding process can be expressed in terms of a non-physical pathway (therefore called alchemical) which can be computed much more efficiently than simulating over a long time to sample sufficient binding and unbinding events such that $\Delta G_{\text{binding}}$ could be estimated via equation (2.66). The alchemical cycle typically employed in ligand-receptor binding [135] is visualized in Figure 2.4 and explained in the figure caption. $\Delta G_{\text{binding}}$ is related to the alchemical transitions via

$$\Delta G_{\text{binding}} = G_{\text{elec+vdw}}^{\text{solv}} + G_{\text{restr}}^{\text{solv}} + \Delta G_{\text{elec+vdw}}^{\text{prot}} + \Delta G_{\text{restr}}^{\text{prot}} \quad . \quad (2.67)$$

The transformations between different states are achieved by scaling the Hamiltonian with a parameter $\lambda \in [0, 1]$, where $\lambda = 0$ yields the system's Hamiltonian at the start of the transition and $\lambda = 1$ the system's Hamiltonian at the end of the transition. In practice, this alchemical transition pathway is discretized into a finite amount of λ values. For each of these λ values the simulation system is simulated in the NPT-ensemble for a limited time in order to create a sufficient amount of uncorrelated samples from the NPT-ensemble. The free energy change along the transition can then be calculated as described in the following sections 2.3.2 and 2.3.3.

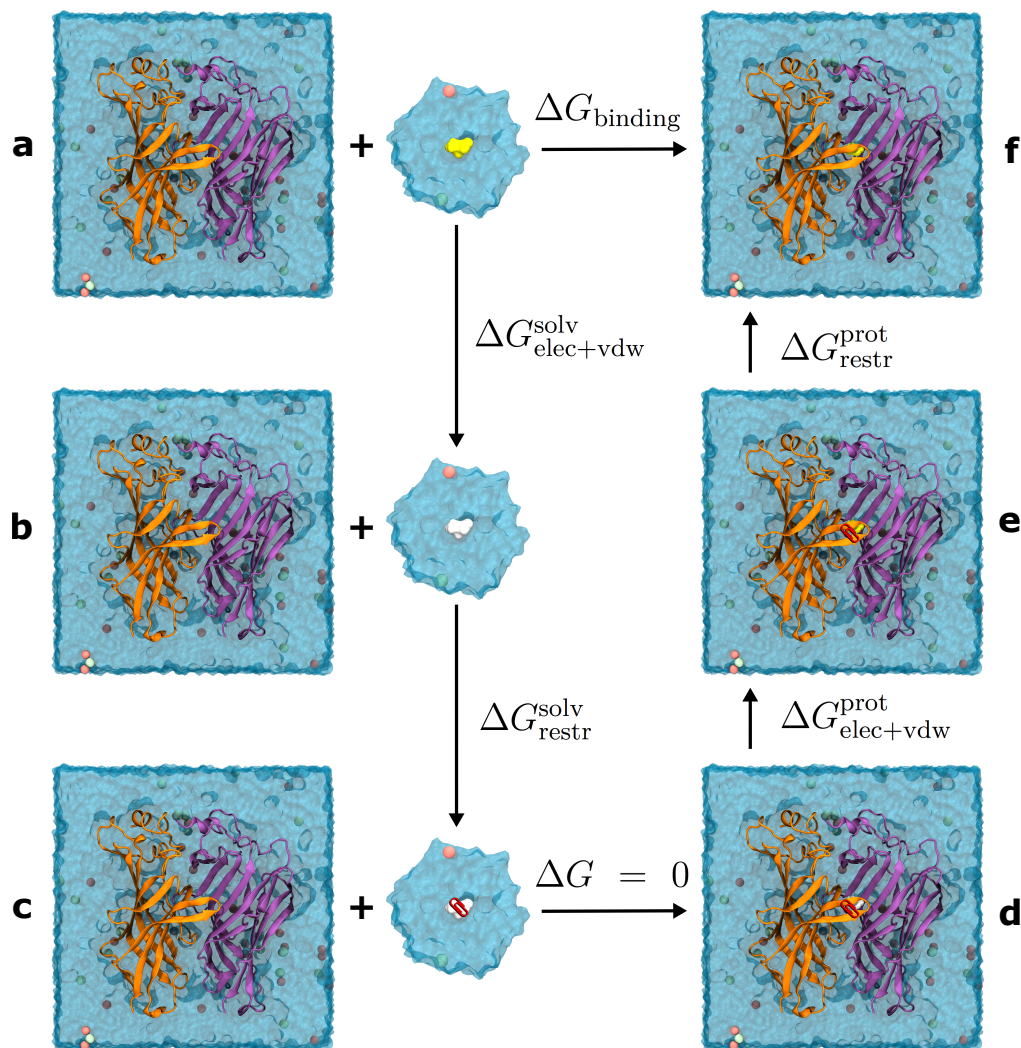


Figure 2.4: Thermodynamic cycle to determine the binding free energy of a ligand to a receptor, exemplified via two adjacent subunits of the ECD of the GlyR as receptor and zwitterionic glycine as ligand. The binding free energy of the ligand is the free energy difference between the ligand bound to its receptor (f) and the receptor plus the ligand not interacting with each other (a). Since the free energy is a state function, i.e. it does not depend on the path taken between the two end states, it can be calculated in a computationally efficient manner via a non-physical (alchemical) pathway as illustrated here. Turning off the non-bonded (electrostatic and van der Waals) interactions between the ligand and the solution yields a non-interacting solute (b, indicated by white colour) and the free energy difference between states (b) and (a), $\Delta G_{\text{elec+vdw}}^{\text{solv}}$, corresponds to the free energy of desolvation of the ligand. The ligand is then restrained (c) and the corresponding $\Delta G_{\text{restr}}^{\text{solv}}$ can be calculated analytically as demonstrated by Boresch et al. [198]. Since the ligand is not interacting with its environment, state (c) and state (d) have no difference in free energy. The nonbonded interactions between the ligand and its environment are then turned on again (e, indicated by yellow colour), while it is still restrained, corresponding to $\Delta G_{\text{elec+vdw}}^{\text{prot}}$. The restraints are necessary in the (d) to (e) transition to keep the not fully interacting ligand in its position in the binding pocket and prevent its dissociation from the binding site. Finally, the restraints are turned off ($\Delta G_{\text{restr}}^{\text{prot}}$) to obtain the endpoint of the cycle, the receptor-ligand complex (f).

In the cycle depicted in Figure 2.4, the ligand is restrained to prevent it from leaving the binding site when its interactions with the environment are scaled down. While different types of restraints can be employed, the set of harmonic restraints proposed by Boresch et al. [198] is particularly convenient. First, the restraints fix the ligand in a certain orientation relative to the receptor which reduces the simulation time needed to achieve sufficient sampling. Second, they allow for an analytic calculation of $G_{\text{restr}}^{\text{solv}}$ and thus there is no need for explicit MD simulations for the corresponding transition. The Boresch restraints act on a distance, two angles and three dihedrals between three ligand and three receptor atoms as specified in [198]. The Boresch formula to calculate $G_{\text{restr}}^{\text{solv}}$ is given by

$$G_{\text{restr}}^{\text{solv}} \approx A_{\text{restr}}^{\text{solv}} = -RT \ln \left(\frac{8\pi^2 V^0 \sqrt{K_r K_{\theta_A} K_{\theta_B} K_{\phi_A} K_{\phi_B} K_{\phi_C}}}{r^2 \sin \theta_A \sin \theta_B (2\pi k_B T)^3} \right), \quad (2.68)$$

where r is the restrained distance length, θ_A and θ_B are the restrained angles, ϕ_A , ϕ_B and ϕ_C are the restrained dihedrals, the K_x ($x \in \{r_0, \theta_A, \theta_B, \phi_A, \phi_B, \phi_C\}$) denote the corresponding force constants used in the harmonic restraints and V^0 is the standard volume of $1660 \text{ \AA}^3$, corresponding to the standard concentration of 1 mol/L.

Equation (2.68) has been derived without considering symmetry. The presence of symmetry in the protein P, ligand L or the protein-ligand complex PL leads to an additional symmetry contribution of

$$G_{\text{sym}} \approx A_{\text{sym}} = -RT \ln \left(\frac{\sigma_{PL}}{\sigma_P \sigma_L} \right), \quad (2.69)$$

where σ_{PL} , σ_P , σ_L denote the symmetry number of the symmetry group of the protein ligand complex, the protein, and the ligand, respectively [198, 199].

Assuming the ligand orientation prescribed by the Boresch restraints is the dominant binding mode, the result obtained for G_{binding} is physically rigorous and the only sources of error are a systematic error due to the limited accuracy of the force field and the random error because of the finite discretization of the transitions as well as finite sampling of each λ window. This makes this method superior to other computationally less expensive approaches to estimate G_{binding}

such as docking scores or implicit solvent end point methods such as molecular mechanics Poisson-Boltzmann surface area (MM-PBSA).

Several methods exist to calculate the free energy change along the transitions and can be divided into two classes: thermodynamic integration [200] and free energy perturbation methods [201]. In the following, the *Bennett Acceptance Ratio* (BAR) [202] (section 2.3.2) and its multistate extension *Multistate BAR* (MBAR) [203] (section 2.3.3) are explained in more detail, as the latter was used in this thesis, because it makes the most efficient use of data when samples from more than one state are available.

2.3.2 Bennett's Acceptance Ratio

Bennett's acceptance ratio [202] is a method for estimating the free energy difference between two different equilibrium states described by the Hamiltonians H_0 and H_1 . Bennett's starting point was the equation

$$\Delta A_{1,0} = k_B T \ln \left(\frac{\langle a(q, p) \exp(-\frac{H_0 - H_1}{k_B T}) \rangle_1}{\langle a(q, p) \exp(-\frac{H_1 - H_0}{k_B T}) \rangle_0} \right), \quad (2.70)$$

which holds for any function $a(q, p)$ [137]. Using variational calculus to find a choice of $a(q, p)$ that minimizes the variance of the free energy, he obtained an implicit function of $\Delta A_{1,0}$:

$$\sum_{i=1}^{n_1} \frac{1}{1 + \frac{n_1}{n_0} \exp\left(\frac{H_1 - H_0 - \Delta A_{1,0}}{k_B T}\right)} = \sum_{j=1}^{n_0} \frac{1}{1 + \frac{n_0}{n_1} \exp\left(\frac{H_0 - H_1 + \Delta A_{1,0}}{k_B T}\right)}, \quad (2.71)$$

where n_1 and n_0 are the number of uncorrelated samples from each state. Equation (2.71) can be solved numerically to yield an estimate for $\Delta A_{1,0}$. Shirts demonstrated that this is the maximum likelihood estimator of the measured data [204], which means that it estimates $\Delta A_{1,0}$ in a statistically optimal way.

2.3.3 Multistate Bennett's Acceptance Ratio

The multistate Bennett's acceptance ratio (MBAR) is a multistate extension of BAR [203]. In this ansatz, a series of $K \times K$ weighting functions $\alpha_{ij}(q)$ are derived to minimize the uncertainties in free energy differences between all K states considered

simultaneously. In addition, it provides a direct way to estimate the uncertainty of the calculated free energies. In the limit of only two states, MBAR reduces to BAR.

2.4 Markov State Modelling

2.4.1 Definition of a Markov State Model

A Markov state model (MSM) is an approximation of the molecular kinetics of a system. It consists of two things: A discretization of the continuous phase space (usually only configuration space) into discrete states $1, \dots, n$ and a Markovian transition matrix $T \in \mathbb{R}^{n \times n}$, whose elements t_{ij} define the conditional probability, given that the system is in state i , after a lag time τ it will be in state j . The transition matrix T thus defines the evolution of the system over time in a probabilistic way.

MSMs have been used in the context of biomolecular simulations because they can reconcile simulation information from many independent individual trajectories in order to rigorously reconstruct the equilibrium distribution and kinetics of a system whose time scales may be far beyond those of the individual trajectories.

2.4.2 Necessary Properties of the Underlying Dynamical Process

The underlying dynamical process (in the context of MD-simulations, this is a set of MD trajectories) from which a MSM is constructed is assumed to have the following properties:

1. The dynamical process is a *Markov process*, i.e. in time-continuous dynamics, the instantaneous change of the system state only depends on the current state of the system. Since classical mechanics is deterministic, this condition is fulfilled for classical MD.
2. The dynamical process is *time-homogeneous*, i.e. the transition probability from any subset of the state space to another (or the same) subset of the phase

space does not change over time. Because in classical mechanics Liouville's theorem holds, classical MD satisfies this condition.

3. The dynamical process is *ergodic*, i.e. the ergodic hypothesis holds. While this is difficult to prove, in practice for sufficiently complex systems, such as biomolecular systems, this is always assumed to be true.
4. The dynamical process fulfills the condition of *detailed balance*, i.e. in equilibrium the fraction of systems transitioning from a subset x to another (or the same) subset y of state space per time is the same as the fraction of systems transitioning from y to x . Detailed balance is a more general concept than time-reversibility and is generally expected to be fulfilled in equilibrium MD.

If a discrete MSM is constructed from an underlying process that satisfies these properties, the resulting transition matrix is a time-homogeneous discrete Markov chain on a finite state space, with a transition matrix T that is primitive (irreducible and aperiodic) and reversible.

2.4.3 Time Evolution of a Markov State Model

The dynamics in the discretized state space are completely determined by the transition matrix T . In contrast to classical mechanics, however, the time evolution is not deterministic, but probabilistic. Given an initial distribution of states $\mu_{(0)} \in \mathbb{R}^n$, the distribution after one lag time τ is obtained by matrix multiplication with the transition matrix via:

$$\mu_{(1)}^T = \mu_{(0)}^T \cdot T \quad (2.72)$$

and thus the distribution at any multiple k of the lag time τ is given by

$$\mu_{(k)}^T = \mu_{(0)}^T \cdot T^k, \quad (2.73)$$

i.e. the complete future of the system (the discretized state space flow) is determined by iterative multiplication of the transition matrix T .

In order to see what the key processes are that govern this time evolution, we can make use of the fact that a primitive and reversible transition matrix can always be diagonalized, i.e. there exists a change of basis (given by an invertible matrix $S \in \mathbb{R}^{n \times n}$), such that

$$D = S^{-1}TS, \quad (2.74)$$

where

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \quad (2.75)$$

is a diagonal matrix with eigenvalues $\lambda_1 = 1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n$. Note that the columns of the matrix S are the corresponding eigenvectors. The state distribution at any multiple k of the lag time τ can then be expressed as

$$\mu_{(k)}^T = \mu_{(0)}^T \cdot T^k = \mu_{(0)}^T \cdot \overbrace{S^{-1}D \underbrace{SS^{-1}}_I DS \dots S^{-1}D \underbrace{SS^{-1}}_I DS}^{k \text{ times}} = \mu_{(0)}^T \cdot S^{-1}D^k S, \quad (2.76)$$

where I is the identity matrix. Since

$$D^k = \begin{pmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{pmatrix}, \quad (2.77)$$

it is easy to see that the dynamics are governed by the eigenvalues $\lambda_1, \dots, \lambda_n$ of the transition matrix T and the corresponding eigenvectors forming the columns of the matrix S . The overall dynamics result from a superposition of these n individual processes. The eigenvectors contain the changes in probability for each discrete state i and thus correspond to the modes of probability flow between the system's discrete states. The corresponding eigenvalues λ_i express the speed of the corresponding probability flow. Hence, the diagonalization of the transition matrix T can be viewed as a principal component analysis of the system's dynamics. The eigenvalues can be seen as the weights of the processes and all processes with eigenvalue $\lambda_i < 1$ decay exponentially with the timescale t_i of the decay process

$$t_i = -\frac{\tau}{\ln \lambda_i}. \quad (2.78)$$

In the long term limit of $k \rightarrow \infty$ only the process with eigenvalue $\lambda_1 = 1$ remains. This process is in fact stationary and corresponds to the stationary (equilibrium) distribution of the system. In fact, it can be shown that for a transition matrix with the properties outlined in section 2.4.2 any starting distribution converges to a unique stationary (i.e. equilibrium) distribution μ_* :

$$\lim_{k \rightarrow \infty} \mu_{(k)}^T = \lim_{k \rightarrow \infty} \mu_{(0)}^T \cdot T^k = \mu_*, \quad (2.79)$$

where μ_* corresponds to the first eigenvector. Since $\lambda_1 = 1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_n$, the process corresponding to the second eigenvalue and eigenvector is the slowest process, followed by the third eigenvalue/vector pair being the second slowest process and so on.

2.4.4 Analyzing Markov State Models

In practice, a MSM consists of several hundred to thousands of microstates which makes it difficult to interpret. In order to construct more comprehensible models of the system's dynamics, a coarse grained MSM can be constructed by merging many microstates into only a few macrostates. To this end, typically advanced variants of the Perron Cluster Cluster Analysis (PCCA) are used. The coarser macrostates correspond to metastable states of the system. In order to map out the flux between these metastable states (i.e. the kinetics), transition path theory is employed. The details of these techniques are beyond the scope of this text and it is referred to the excellent textbooks [138, 139] for further details with regard to all aspects of MSMs. The author has written a review contribution that illustrates the typical workflow in the context of biomolecular simulation [205].

3

A Refined Open State of the Glycine Receptor

Contents

3.1	Introduction	72
3.2	Methods	75
3.2.1	Homology Modelling	75
3.2.2	System Setup	75
3.2.3	Simulation Protocol	76
3.3	Results	80
3.3.1	The MD-Refined Structure is Physically Open, Hydrated and Selectively Permeant to Chloride Ions	80
3.3.2	The Open Conformation is Stabilized by the Leucine Gate Residues Filling a Hydrophobic Void at the Subunit Interface	85
3.3.3	The Cryo-EM Based Structure Collapses due to a Sub-Optimal Orientation of the Leucine Gate Residues	89
3.4	Discussion	91
3.5	Conclusions	99

What truly represents the physiological open state of pLGICs has been subject to debate for several years. This dispute was triggered by an open state structure of the GlyR solved by cryo-EM in 2015 [68], which has been rendered questionable by several independent groups because it is unstable and collapses in MD simulations. In this chapter, I resolve this controversy by developing a simulation protocol

for atomistic MD simulations through which a stable open state of the GlyR can be obtained and demonstrating that its stability can be explained in terms of rearrangements in the gating regions of the receptor. Much of this chapter is based upon the following publication: M.A. Dämgen and P.C. Biggin. ‘A Refined Open State of the Glycine Receptor Obtained via Molecular Dynamics Simulations.’ *Structure* 2020, 28, 1-10.

3.1 Introduction

Between 2015 and 2017 various structures of the GlyR with agonist, antagonist and modulators bound have been resolved with cryoelectron microscopy (cryo-EM) [68] as well as X-ray crystallography [56, 60, 123] and have been annotated to closed, open and desensitized states.

From these structures and those of other members of the superfamily alongside a large amount of mutagenesis work (see [206] and references therein), it became apparent that there are two main constriction points in the channel pore, namely a ring of five hydrophobic residues (usually leucine) at the 9' position of the pore-lining M2 helices and a ring of five proline residues at the -2' position, which I hereafter refer to as L9' and -P2' gate, respectively. The L9' gate, located in the middle of the TMD closes the channel in the antagonist (strychnine) bound structures (PDB: 5CFB and 3JAD [68]). The P-2' gate is near the intracellular mouth of the channel pore and gates the channel in the desensitized-like structures (PDB: 5TIN, 5TIO [56], 5VDH, 5VDI [123] and 3JAF [68]).

Among these structures, the glycine bound cryo-EM structure 3JAE [68] is of particular interest due to its unusually wide-open pore conformation relative to other open-like structures of the pLGIC superfamily, particularly at the -2' position. The structure 4HFI [207] of the proton-activated bacterial *Gloeobacter violaceus* (GLIC) at pH=4 and the 3RIF structure [66] of the invertebrate glutamate-gated chloride channel (GluCl) from *Caenorhabditis elegans* were classified as open-like pLGIC structures. The distance from the pore centre to the backbone C_α

at the -2' position is about 2 Å smaller for these two structures than for the wide-open 3JAE structure [208].

Another factor that adds to the uncertainty of the -2' conformation is that the electron density map of the wide open GlyR is locally extremely poor in this region [68]. This is an indicator of a higher dynamic variability in this region. Moreover, the final step of fitting an atomic model into a density map with standard software packages can result in folds in which some regions might not be stable under physiological conditions. One reason for this is the potential energy model of standard software packages that usually only accounts for the protein in vacuum, while neglecting the full natural environment, i.e. the explicit presence of bulk water as well as a lipid bilayer for membrane proteins. Another drawback is that entropic effects at a physiologically relevant temperature are not accounted for. MD simulations provide a more sophisticated model, including the ability to account for the natural environment and entropic effects. Yet, they are much more costly, both in regard to the time necessary to set these simulations up as well as in regard to actual computation time.

MD simulations have provided useful insight into the functional annotation of states to ion channel structures which is difficult to infer from structural information alone. An important concept in this context is hydrophobic gating [77, 78]. Hydrophobic gating occurs in narrow hydrophobic pores where the energetically favoured expulsion of water prohibits ion permeation. This means that a pore whose dimensions are theoretically large enough to allow for ion permeation is not necessarily ion-permeant due to hydrophobic effects. Pore hydration has been suggested as a proxy for ion-permeability [209] and used to annotate ion channels as functionally open or closed.

Gonzalez-Gutierrez et al. [208] and Trick et al. [209] independently annotated the wide-open 3JAE GlyR structure as open. Furthermore, Gonzalez-Gutierrez et al. found (via simulations employing a reduced model with restraints applied to backbone atoms) that the 3JAE wide-open pore to have a four times higher single channel conductance than the experimental value (although it should be noted that

the simulations were performed at 37 °C without the extracellular and intracellular domain present, while experiments were conducted at room temperature with the full receptor). Moreover, blocking the channel pore with picrotoxin in their simulations still resulted in ion conduction whereas no currents were observed experimentally. This substantiates the notion that the wide-open 3JAE atomic model has an overly dilated pore. In these simulations, artificial restraints were applied to the protein backbone, such that it cannot move freely and stays very close to the original cryo-EM model. Interestingly, if no such restraints are applied, the structure undergoes a so-called hydrophobic collapse to a distinct state with a significantly smaller pore radius [210]. This behaviour has also been observed in simulations of other open (but not 'wide-open') structures of the pLGIC family [70, 211, 212]. They all have in common that the hydrophobic collapse is initiated via the non-polar interaction of the 9' gate residues.

I hypothesized that the wide-open atomic model 3JAE that has been fitted into the cryo-EM density map may not correspond to an energetically stable representation of the open state under physiological conditions. However, there might be other models in agreement with the cryo-EM density map that correspond to a stable open state. In order to find alternative open state structures, I developed a careful equilibration protocol where the protein is given maximum freedom to explore other open state configurations while preventing the pore from collapsing. Then all restraints on the pore are released and the structure found through this careful equilibration procedure stays open with a hydrated pore that allows selectively for frequent chloride ion permeation. The model provides alternative conformations of the leucine side-chains that are not discernible in the original density map. Moreover, I provide a structural explanation as to why the structure found via MD simulations is a stable open state, whereas the wide-open structure modelled into the cryo-EM density map undergoes a hydrophobic collapse.

3.2 Methods

3.2.1 Homology Modelling

A human $\alpha 1$ GlyR model was constructed from the open-state structure of *Danio rerio* $\alpha 1$ GlyR (glycine-bound open state; PDB: 3JAE [68]) as previously reported [69]. Briefly, sequences were aligned using Muscle [213] and adjusted manually. MODELLER version 9.16 [214] was used to generate 100 models and from of the intersection of the top 10 of both molecular PDF and DOPE score [215], the model with the best QMEAN score [216] was chosen for simulation. The final model consisted of residues P10-R309 linked via an AGT-tripeptide to K386-K419. In order to avoid artificial charges, the termini of the model were capped with N-terminal acetyl and C-terminal acetamide with Maestro (Schrödinger Release 2013-2: Maestro, version 9.5, Schrödinger, LLC, New York, NY, 2013). Hydrogens were added to the protein with the pdb2gmX tool of GROMACS 5.1 [141], with standard protonation states at pH=7.4.

3.2.2 System Setup

The structure for the natural neurotransmitter glycine was taken from the ZINC ligand database [217] and positioned into all 5 orthosteric binding sites according to its position in the currently highest resolution GlyR structure (PDB: 5TIN [56]), 2.6 Å resolution (note that glycine was bound to the receptor but no density discernible in PDB: 3JAE [68], 3.9 Å resolution). A stable water molecule in the binding pocket which is resolved in PDB: 5TIN and shown to be stable in MD-simulations was also included [67, 69].

Simulations were set up using a coarse-grained approach with the MARTINI force field version 2.2 using ‘martinize’ [149] to coarse-grain the protein model. The systems were setup with the programme ‘insane’ [218], using a hexagonal prism as the periodic simulation box. The protein was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer, solvated in water and sodium and chloride ions were added to neutralize the net charge and simulate a physiological

concentration of 150 mM. GROMACS 2018 [141] was used as molecular dynamics engine for all simulations. After energy minimization, the system was equilibrated at 310 K and 1 bar for 10 ns. Simulation details were based on the recommended new-rf.mdp parameters [219] with appropriate adjustments. The final frame was then converted to all-atom resolution via the programme ‘backward’ [220]. The backmapped receptor was replaced by the original all-atom model with the glycine ligands and stable water molecules bound, resulting in a total of 136,334 atoms and is visualized in Figure 3.1.

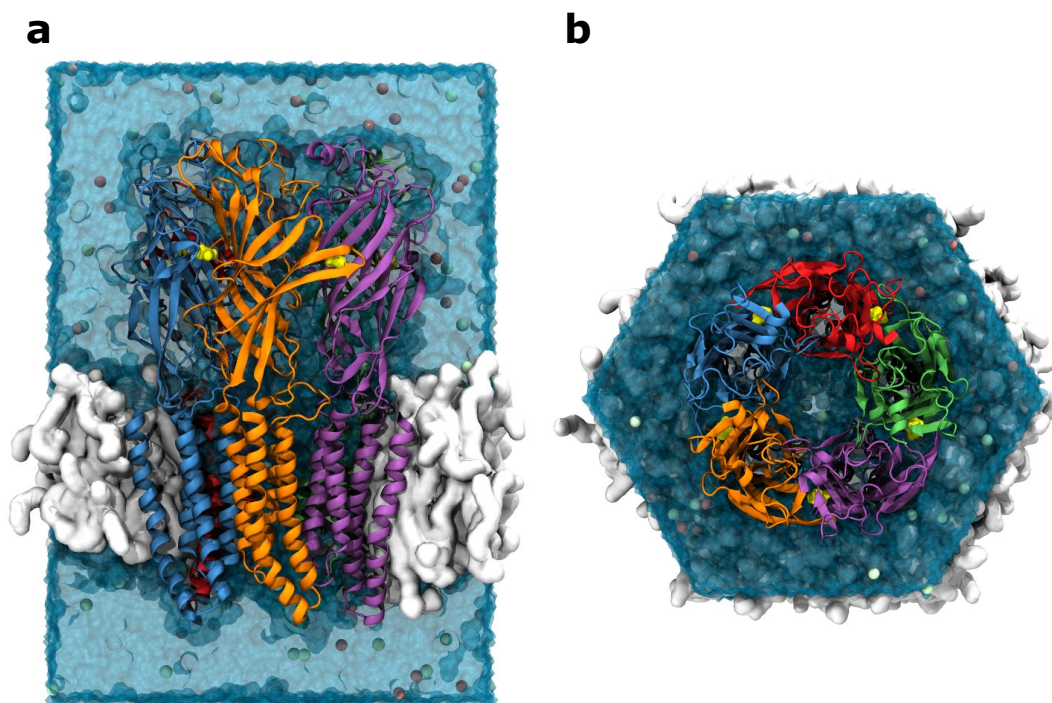


Figure 3.1: Simulation system with a hexagonal prism as box geometry. The GlyR is embedded in a lipid membrane and solvated in 150 mM NaCl solution. (a) View parallel to the membrane, (b) view from the extracellular towards the intracellular matrix.

3.2.3 Simulation Protocol

All following steps were performed with GROMACS 2018 [141]. Periodic boundary conditions in all three spatial dimensions were applied to mimic bulk conditions. Each system was energy-minimized using the steepest descent algorithm until the maximum force fell below a tolerance of $1000 \text{ kJmol}^{-1}\text{nm}^{-1}$. The velocities were

initiated according to a Maxwell-Boltzmann distribution at $T=310$ K and each system was equilibrated for 100 ps in the NVT-ensemble. All protein, ligand and binding pocket water heavy-atoms were position-restrained harmonically with a force constant of $1000 \text{ kJmol}^{-1}\text{nm}^{-2}$. This was followed by a series of equilibration phases in the NPT-ensemble to establish pressure at $P=1$ bar and to gradually relax the protein and membrane. For 1 ns the same position restraints as in the NVT-equilibration were applied, followed by 10 ns in which only the protein backbone as well as the ligand's C_α atom and the oxygen of the binding pocket water were restrained harmonically with a force constant of $1024 \text{ kJmol}^{-1}\text{nm}^{-2}$ in order to relax the protein side chains and to give the ligands some orientational freedom. These restraints were then reduced in a step-wise fashion via halving them every 1 ns to a final value of $64 \text{ kJmol}^{-1}\text{nm}^{-2}$ in order to gradually relax the protein backbone as well as the ligand orientations and the water in the binding pocket.

The final frame after this initial equilibration period was used as the starting point for three independent simulations of 300 ns length which were all subject to hydrophobic collapse. In order to find a stable open channel structure, I extended the initial equilibration as follows. The ion channel pore was kept open via flat-bottom cross distance restraints between the C_α atoms of pore-lining residues (pore restraints, see Figure 3.2). The flat bottom potential only acts with a linear force when the C_α atoms of pore-lining residues ($-2'$, $2'$, $6'$, $9'$, $13'$, $16'$, $20'$) of non-adjacent subunits come closer to each other than in their distance in the open-state cryo-EM structure 3JAE. This keeps the channel pore from collapsing initially and gives it enough freedom to find a stable configuration. This additional equilibration period with pore restraints followed the initial equilibration for 150 ns. The final frame was used as the starting configuration for three independent repeats, initiated with different velocities. All three repeats showed a stable open channel configuration. Note that in order to prevent possible unbinding of glycine ligands [67], flat bottom restraints on the glycine ligands were applied, so that they can move freely within a reasonable volume in the orthosteric binding site but are prevented from unbinding. These restraints act if the distance between C_α atoms of key residues in the binding

site and the C_α atom of the glycine ligand exceed a threshold distance (namely, 10.60 Å to the (+)F207 C_α , 10.53 Å to the (-)F63 C_α and 9.15 Å to the (-)S129 C_α with a force constant of 5000 kJmol⁻¹nm⁻²).

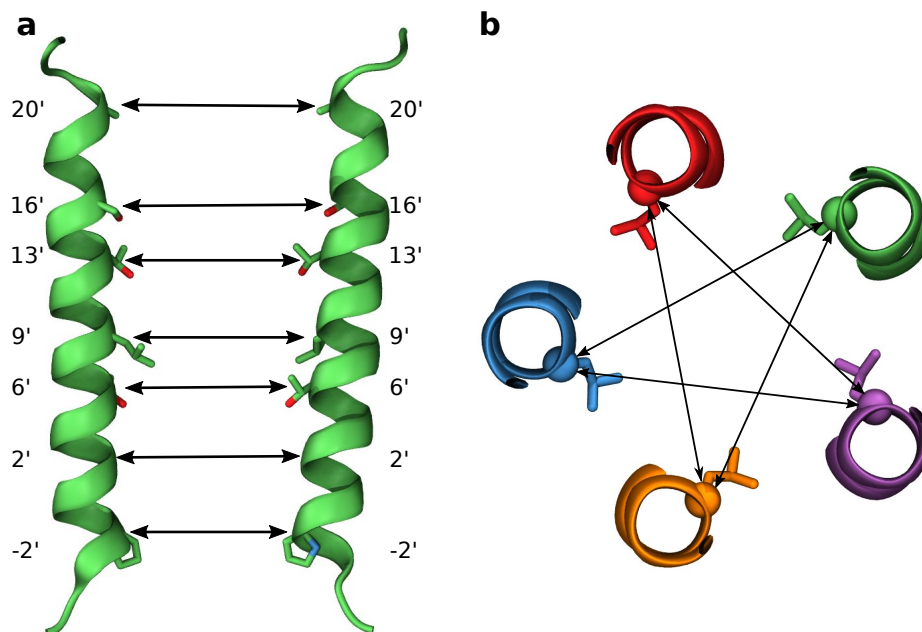


Figure 3.2: Pore Restraints. (a) Two non-adjacent pore lining M2 helices with seven pore lining residues. The double arrows indicate the pore cross distances on which the flat bottom pore restraints act. (b) Illustration of the pore cross distances at the L9' gate (view from the extracellular toward the intracellular side). In total, there are five cross distances for the 9' ring. The cross distances for the other pore lining residues are not shown for clarity, but, similarly, there are five cross distances per pore lining residue position, so that in total $7 \times 5 = 35$ cross distances are restrained. Figure reproduced from [221] under a CC BY 4.0 license.

The force fields used were AMBER99SB-ILDN [143] for protein and ions, TIP3P for water [170], Slipids [165–167] for lipids and the parameters for zwitterionic amino acids by Horn [169] for the glycine ligands.

Most of the details for the calculation of the potential energy are based on the settings used to parameterize the AMBER99SB-ILDN force field [143]. The van der Waals interactions were cut off at 10 Å and dispersion correction was applied to energy and pressure to correct for this finite truncation. Electrostatic interactions were treated using the smooth Particle Mesh Ewald method [175, 176], where the real space contribution was cut off at 10 Å and the reciprocal energy

term obtained in k-space was calculated on a grid with 1.2 Å spacing using 4th order B-splines for interpolation. The Verlet cut-off scheme was used to generate a pair-list with buffering using a tolerance of 0.005 kJ/mol/ps for pair-interactions per particle. With this scheme the cut-off distance for the short-range neighbour list is set automatically by mdrun as well as the update frequency of the neighbour list for which a minimum of 10 time steps was specified. Grid search was chosen as the algorithm to search for neighbours.

In all dynamics simulations, Hamilton's equations of motion were integrated numerically with the leap frog integration algorithm [222] which is a time-reversible and symplectic scheme that preserves the important symmetries of classical mechanics.

All protein bonds involving hydrogen were constrained with the LINCS algorithm [182] and water molecules were kept rigid via the SETTLE algorithm [184], such that a timestep of 2 fs could be employed (with the exception of the initial NVT-equilibration where a shorter timestep of 1 fs was used).

The Bussi thermostat [187] was coupled to the system to maintain the temperature constant at physiological 310 K. To ensure that weakly coupled parts of the system have the same temperature, three groups were coupled separately, namely the protein together with ligands and waters in the binding pocket, the lipid membrane, and bulk water with ions. For all groups, the same coupling constant of 0.1 ps was used.

In all simulations with constant pressure, a barostat was coupled semi-isotropically to the system to maintain a physiological pressure of 1 bar. The time constant used for pressure coupling was 1 ps and the isothermal compressibility set to $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. For the initial equilibration, the Berendsen barostat [186] was used and the Parrinello-Rahmann barostat [194] for production runs as well as the simulation period with pore restraints.

In order to avoid artificial translational drift, the centre of mass motion due to translation was removed every 100 time steps independently for the group of the protein, ligand and binding pocket water and the group of bilayer, bulk water and ions.

During production runs, the configurational part of phase space was written out every 10 ps.

Analysis

Pore profiles, volumes and hydrophobicity were calculated with CHAP [223]. The pore radius profiles are reported as mean and 1 standard deviation over 300 equally distributed data points per 300 ns simulation. Hydrophobicity was calculated according to the Wimley and White scale [224, 225]. Trajectory analysis was performed with MDAnalysis [226] and in-house scripts. The local resolution of the cryo-EM density map was estimated with RESMAP [227], the real space correlation between atomic model and map was calculated with Phenix [228]. Visualization was performed in VMD [229] and UCSF Chimera [230].

3.3 Results

3.3.1 The MD-Refined Structure is Physically Open, Hydrated and Selectively Permeant to Chloride Ions

The results of this and the following section are based on three-independent production simulations of 300 ns duration which were initiated from the final frame of 150 ns equilibration with flat bottom pore restraints. All simulations gave stable open-pore configurations, did not exhibit hydrophobic collapse and exhibited similar pore profiles.

As a first assessment of the functional state of the receptor observed in the simulations, I examine the pore radius profile of the TMD where the L9' and P-2' gates are located (Figure 3.3a). Compared to the wide open cryo-EM structure 3JAE (red), the mean radius observed in the simulation (black) is slightly narrower at the P-2' ring, but would still allow for the permeation of hydrated chloride ions (dashed cyan). It clearly is not a desensitized state with the P-2' constriction being 2 Å wider than the currently highest resolution crystal structure of a desensitized state, 5TIN (orange). This result could be reproduced in two further independent simulations (see Figure 3.4).

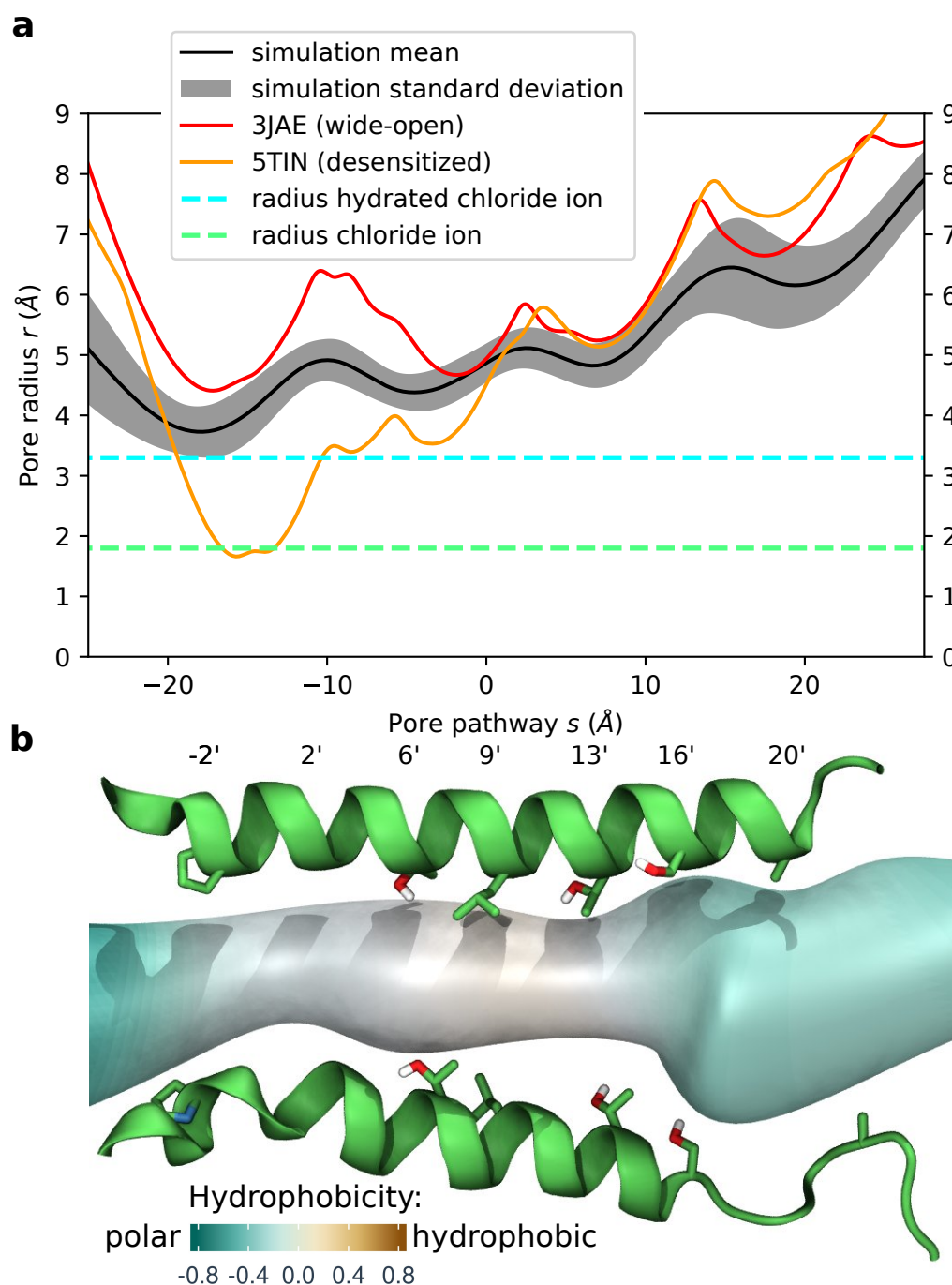


Figure 3.3: Pore profile of GlyR structures. (a) A typical pore radius profile obtained from a production run (mean, black; standard deviation, gray) alongside the profile of the cryo-EM wide-open structure (PDB: 3JAE) [68] (red) as well as the desensitized crystal structure with currently highest-resolution (PDB: 5TIN) [56] (orange). The radii of a dehydrated and a hydrated chloride ion are indicated by dashed green and cyan lines, respectively. The pore is positioned in this and subsequent figures such that the L9' ring is located at $s=0$. The GlyR structure from the simulation is physically open and theoretically allows permeation of hydrated chloride ions. (b) Corresponding pore region of a representative structure from the simulation (minimal C_α root-mean-square deviation to the average structure). M2 helices with pore lining residues (only two non-adjacent subunits shown) together with the pore volume coloured by the hydrophobicity of the surrounding based on the Wimley-White hydrophobicity scale rescaled from -1 (very hydrophilic) to 1 (very hydrophobic) [224, 225]. Figure reproduced from [221] under a CC BY 4.0 license.

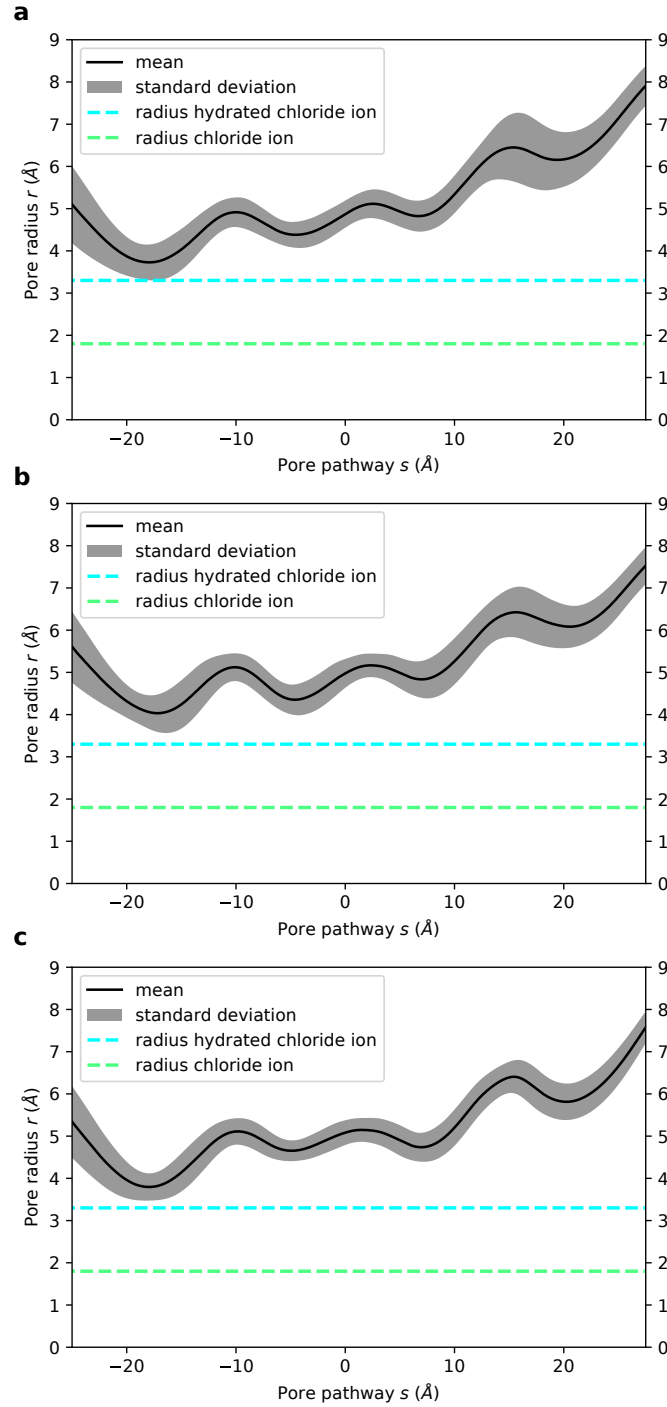


Figure 3.4: Pore profiles from three independent simulations with a stable open state. Time-averaged pore radius profiles of the transmembrane domain obtained from three independent simulations (of 300 ns length each) initiated with different velocities from the system obtained after 150 ns equilibration with pore restraints. No restraints were applied to keep the pore open. The radii of a dehydrated and a hydrated chloride ion are indicated by dashed green and cyan lines, respectively. In all three repeats, the structure from the simulation is physically open and theoretically allows permeation of hydrated chloride ions. Figure reproduced from [221] under a CC BY 4.0 license.

The corresponding pore region of a representative simulation structure (the structure with minimal C_α RMSD to the average structure) is illustrated in Figure 3.3b. The pore volume corresponding to this representative simulation snap shot is coloured by the hydrophobicity of the amino acids surrounding the pore. Hydrophobicity is highest in the L9' region which acts as a hydrophobic gate in the resting state of the receptor.

While the physical pore dimensions allow for hydrated chloride ion permeation, the concept of hydrophobic gating [77, 78] is another crucial factor to consider when assigning a functional state to an ion channel. Hydrophobic gating occurs in narrow hydrophobic pores where the energetically favoured expulsion of water prohibits ion permeation. Beckstein and Sansom [79] have studied in detail the influence of radius, hydrophobicity as well as flexibility of the pore on the degree of dewetting and found that a radius of less than ca. 4-5 Å in hydrophobic regions can lead to local dehydration and thus be a prohibitive energetic obstacle for ion permeation. Since the pore radius of the most hydrophobic region around the L9' ring in the GlyR border these dimensions, I proceed with an assessment of the hydration state of the pore as well as the explored space by ions.

The ion densities obtained over 300 ns of unrestrained simulation (Figure 3.5a-c) reveal that chloride ions can indeed penetrate the transmembrane pore and show selectivity over sodium ions, which preferentially occupy regions in the upper part of the extracellular domain. To resolve the underlying dynamics over time, I analyzed the z-coordinates of water, chloride and sodium ions inside the pore within 5 Å of the ion channel axis (Figure 3.5d-f). Chloride ions frequently penetrate the transmembrane pore region, while sodium only rarely accesses this region. In three repeats over the 300 ns each, 52, 53 and 49 permeation events occur for chloride and 1, 1 and 0 permeation events for sodium, respectively. This is in line with the experimental estimation of a ratio of 28:1 selectivity of chloride over sodium [83]. Furthermore, the whole channel pore is hydrated throughout the simulation in all three repeats, demonstrating that the observed pore radius at the hydrophobic L9' gate is above the threshold for hydrophobic gating.

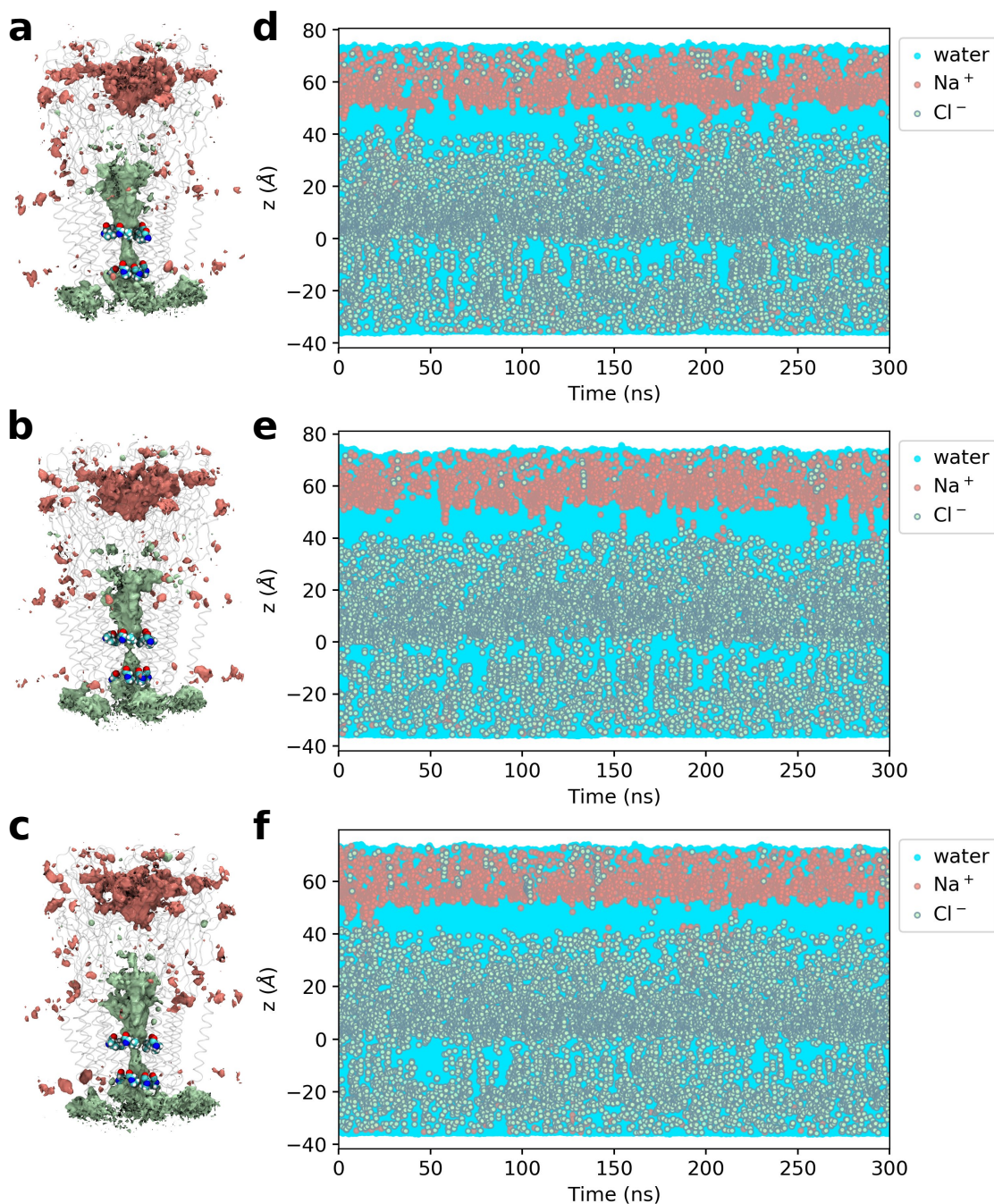


Figure 3.5: Ion densities and trajectories of water and ions within 5 Å of the channel axis for simulations with a stable open state (150 mM NaCl with no membrane potential). (a-c) Densities of chloride (green) and sodium (red) ions over 300 ns. The isosurfaces represent a density value of 0.5 particles/nm³. Transparent ribbons indicate the receptor backbone. The L9' ring and P-2' ring residues are shown in van der Waals representation. The densities prove chloride occupancy in the transmembrane pore and, moreover, show selectivity for chloride over sodium in this region. (d-f) Trajectories of water, chloride and sodium ion z-coordinates within 5 Å of the channel axis inside the pore over 300 ns (represented by blue, green and red circles, respectively). The 0 point of the z-axis is positioned at the L9' ring and the receptor structures in the left panels are aligned and scaled correspondingly to allow for spatial orientation along the z-axis. The whole channel pore is wetted throughout the simulations (dewetted regions would appear as white stretches). While chloride frequently penetrates the transmembrane pore, sodium only does so very rarely, again demonstrating the chloride selectivity for this channel. Figure reproduced from [221] under a CC BY 4.0 license.

3.3.2 The Open Conformation is Stabilized by the Leucine Gate Residues Filling a Hydrophobic Void at the Subunit Interface

Having annotated the receptor state as functionally open, the first question that arises is: Why does the structure stay open? What structural changes have occurred during the equilibration period with restraints to keep the channel open, such that after the release of these restraints the channel remains in an open ion-conductive state?

The key conformational changes during the equilibration phase with pore restraints are depicted in Figure 3.6. Taking a cross section at the level of the L9' gate reveals that at the beginning of the equilibration phase, there are substantial gaps between the subunit interfaces near the leucine residues (Figure 3.6a). During the equilibration the leucine residues that originally point slightly more towards the pore centre, rotate further away from the pore centre and bury themselves deeply into these voids. By engaging in hydrophobic interactions with the hydrophobic residues lining these hydrophobic pockets, they lock the subunit interfaces as well as the neighbouring M1 helix stronger together. With L9'=L261 being located on the M2 helix of the principal (+) subunit, the corresponding hydrophobic void is formed by I257, V260 and T264 belonging to the M2 helix of the principal (+) subunit, L255, T258, T259, T262 belonging to the M2 helix of the complementary (-) subunit as well as L233, I234 and L237 of the M1 helix of the complementary (-) subunit. Across the pLGIC superfamily, all the corresponding positions are occupied by hydrophobic residues and thus the hydrophobic pocket is functionally preserved (Figure 3.7). This suggests that the open state of the whole superfamily is characterized by the -2' hydrophobic residues (L or I/V in some cases) being buried in these hydrophobic voids.

It should be pointed out that the hydrophobic interaction between the L9' residue and the methyl-group of the T6' residue of the complementary subunit leads to the T6' hydroxyl group facing towards the centre pore (Figure 3.8). This makes the profile more hydrophilic at the 6' ring and thus facilitates pore hydration and

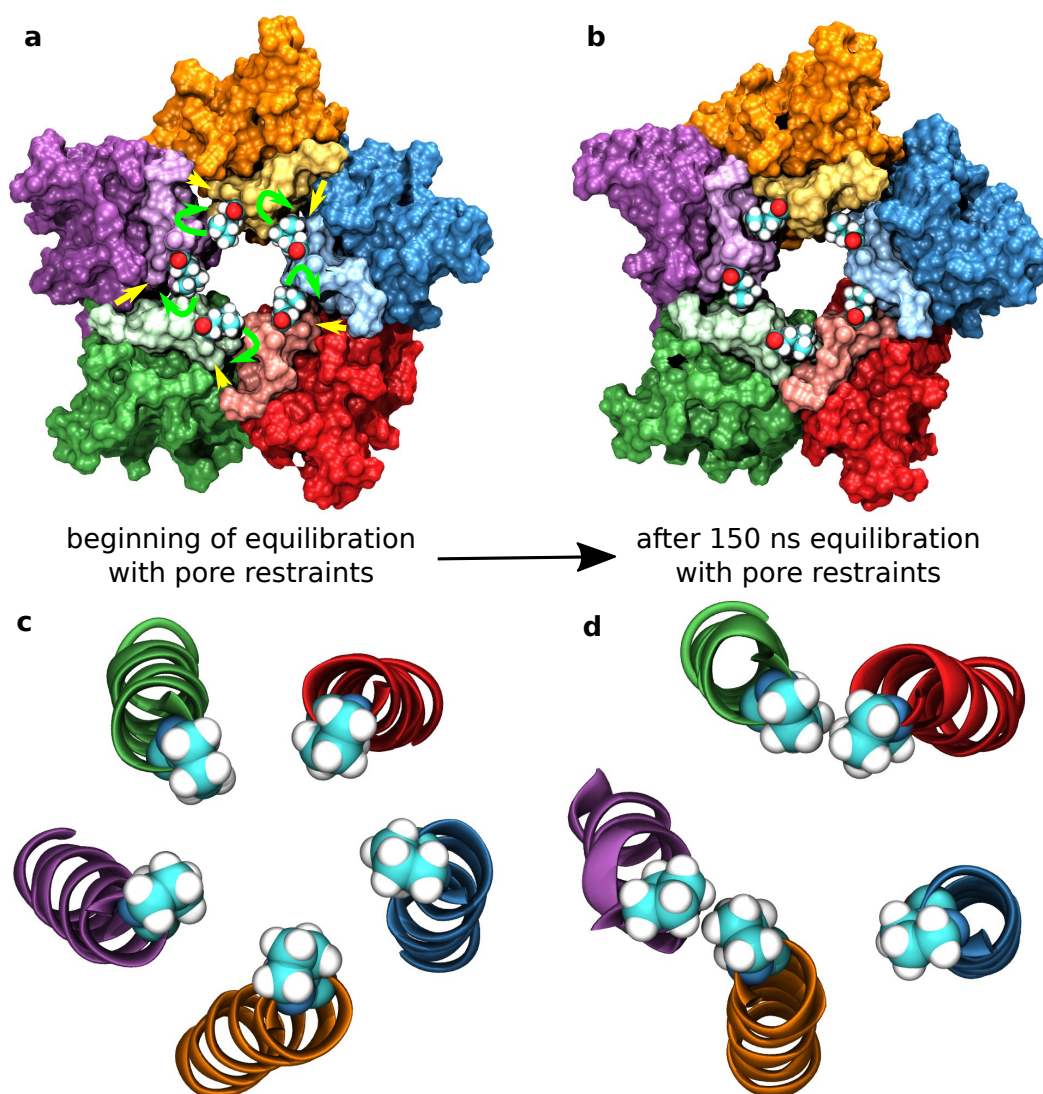


Figure 3.6: Critical changes in pore geometry. (a and b) View from the extracellular toward the intracellular matrix. The five subunits are coloured differently with residues that normally form the hydrophobic binding pocket around L9' in lighter colours. While at the beginning (a) the five hydrophobic voids at the subunit interfaces are empty and the L9' residues are orientated slightly more toward the pore center (b), after 150 ns of equilibration with pore restraints the L9' residues have buried themselves deeply into these hydrophobic pockets and make close contact with the surrounding residues in these pockets, thus firmly locking the structure into an open state conformation. (c and d) View of the M2 helices with P-2' gate shown explicitly from the intracellular toward the extracellular side. While at the beginning (c), the structure is five-fold symmetrical, after 150 ns of equilibration with pore restraints an asymmetric arrangement has been attained (d). Figure reproduced from [221] under a CC BY 4.0 license.

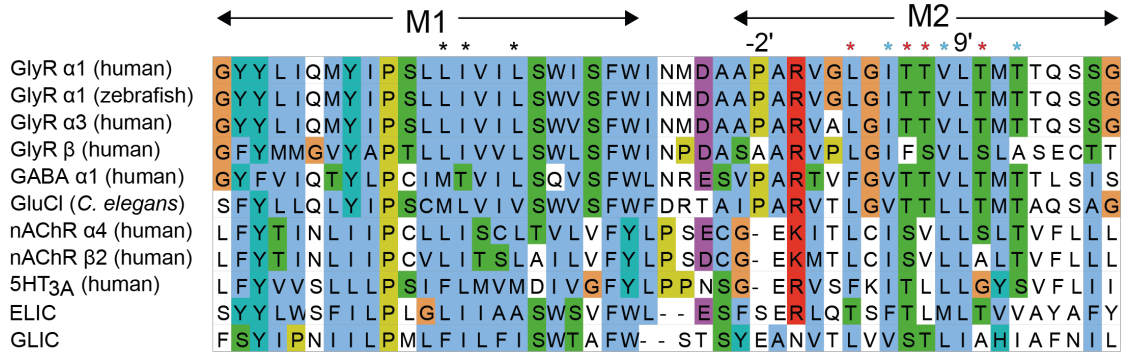


Figure 3.7: Multiple sequence alignment of pLGIC representatives. The conservation of the leucine at L9' is well described. Less well appreciated is the functional conservation of the residues that form the hydrophobic pocket for the leucine in the open state formed by I257, V260 and T264 (I5', V8' and T12' respectively and shown as cyan asterisks) belonging to the M2 helix of the principal (+) subunit, L255, T258, T259, T262 (L3', T6', T7' and T10' respectively and shown as red asterisks) belonging to the M2 helix of the complementary (-) subunit as well as L233, I234 and L237 of the M1 helix of the complementary (-) subunit, shown as black asterisks. Figure reproduced from [221] under a CC BY 4.0 license.

ion permeation. Thus, I postulate that the (+)L9'(-)T6' hydrophobic interaction plays an important role in the gating mechanism of GlyR. The T6' position is highly conserved across anionic members of the pLGIC superfamily and only occasionally replaced by a serine which lacks the additional methyl group of threonine to engage in hydrophobic interactions with L9'. For the anionic pLGICs, the (+)L/I/V9'(-)T6' interaction could thus be an additional way to fine-tune pore hydration in the gating process.

While the L9' residues lock the pore in an open conformation, the P-2' gate residues rearrange from a symmetric conformation (Figure 3.6c), which is enforced by the five-fold symmetry restraints when constructing the cryo-EM model, in an asymmetrical manner (Figure 3.6d). In this arrangement, four vicinal proline residues form two pairs via hydrophobic interactions, increasing the distances toward the non-interacting vicinal proline residues, such that the overall asymmetric conformation withstands a hydrophobic collapse.

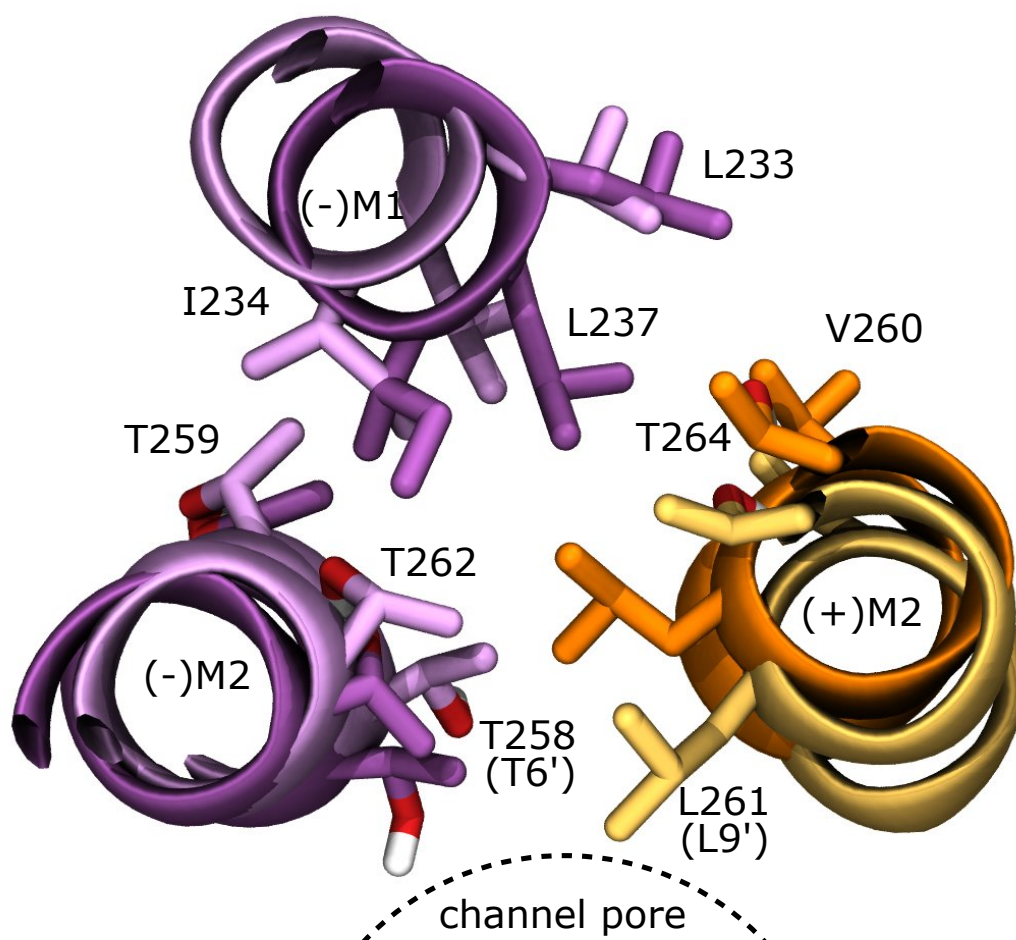


Figure 3.8: Details of the hydrophobic pocket and orientation of T6'. Horizontal slice viewed from the extracellular side with the light colours representing the structure at the beginning of the equilibration with pore restraints and dark colours the structure at the end of 150 ns equilibration with pore restraints. The principal subunit is coloured in orange, the complementary subunit in purple. Key residues that form the hydrophobic pocket are indicated. Note also the position of T6' which rotates slightly towards the central pore with its hydroxyl group facing towards the channel, thus facilitating channel hydration. Figure reproduced from [221] under a CC BY 4.0 license.

3.3.3 The Cryo-EM Based Structure Collapses due to a Sub-Optimal Orientation of the Leucine Gate Residues

The results of this section are based on three independent production runs of 300 ns length that followed the initial standard equilibration without the additional 150 ns of additional equilibration with flat bottom pore restraints.

Having shown that the key to a stable open state structure is the occupation of all five hydrophobic voids at the subunit interface by the L9' gate residues, it is easy to explain why the original cryo-EM conformation undergoes a hydrophobic collapse. In the original cryo-EM conformation (Figure 3.6a), the L9' residues are not buried in the hydrophobic voids at the subunit interface, but rather point slightly more towards the pore centre. In simulations that start from this conformation, the presence of explicit water together with protein dynamic flexibility at a physiological temperature then renders these hydrophobic voids extremely unstable. Collapse usually occurs when two subunits that form a void come closer together and push the corresponding leucine residue toward the pore center. It is also possible that a leucine residue randomly ends up in a hydrophobic void. However, the overall collapse is a chaotic process, such that at least two (but usually more) hydrophobic voids collapse with the corresponding leucine residues being pushed towards the pore centre, decreasing the pore radius significantly (Figure 3.9a). The pore lining M2 helices come closer to each other, such that the P-2' mutually experience their strong hydrophobic interactions, leading to a further collapse of the P-2' ring (Figure 3.9b).

A recent publication by Cerdan et al. [210] presents an ion-permeable state of the GlyR obtained from MD simulations. I argue that this state merely corresponds to a collapsed structure. On a structural level, it shows the key features of a collapsed state (see Figure 3.9c and d). At the L9' gate, two hydrophobic voids are collapsed and the P-2' ring is collapsed, too. The pore radius profile resembles that of the collapsed simulation with the L9' gate pore radius being slightly larger, but the P-2' gate pore radius slightly smaller. Cerdan et al. reported a single chloride ion permeation event over a total unrestrained simulation time of 500 ns from two repeats. Chloride permeation, however, is not impossible for a collapsed

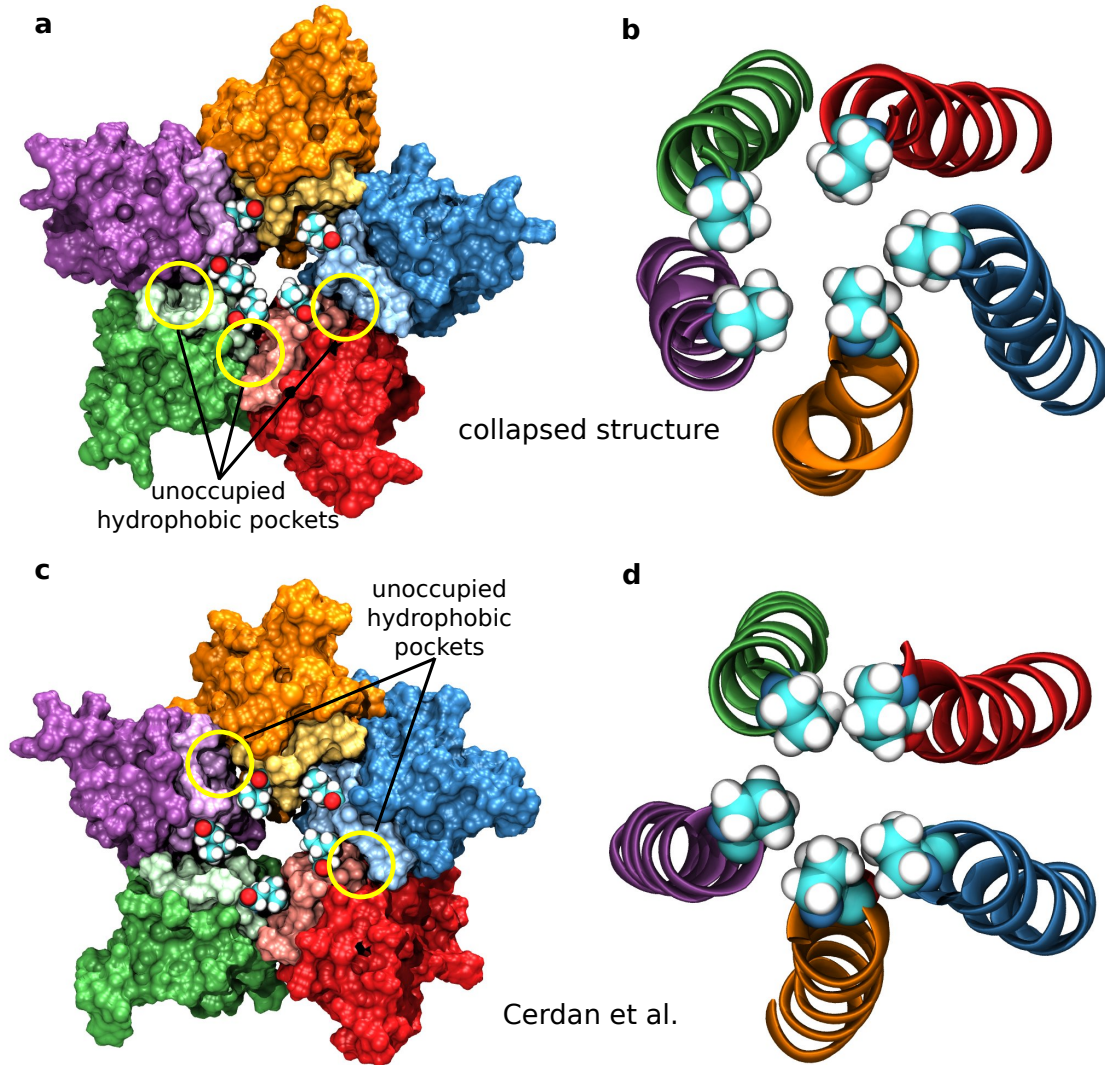


Figure 3.9: Structural features of the collapsed state. (a and b) Representative structure from a collapsed simulation. (a) Although two hydrophobic pockets end up accommodating the corresponding leucine residues, the other three hydrophobic voids undergo a hydrophobic collapse, where the two subunits come closer together and push the leucine residues toward the pore center and decrease the pore radius significantly. (b) With the pore lining M2 helices coming closer together, the P-2' residues undergo a hydrophobic collapse due to their strong hydrophobic interactions. (c and d) Representative structure from a simulation by Cerdan et al. [210]. (d) Two hydrophobic pockets fail to accommodate their corresponding L9' residue, leading to a much smaller pore radius. (d) Without all L9' gate residues locking the structure into a stable open state, the P-2' residues come so close to each other that their strong hydrophobic interactions lead to a hydrophobic collapse in the P-2' ring region as well. Figure reproduced from [221] under a CC BY 4.0 license.

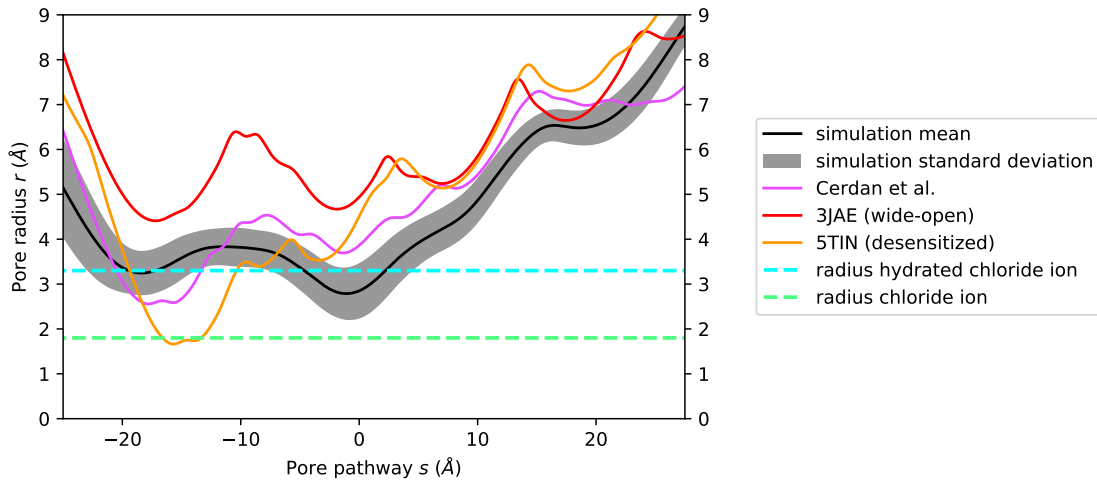


Figure 3.10: Pore profile of a hydrophobic collapsed state. Pore radius profile of a hydrophobic collapse simulation (mean, black; standard deviation, gray) alongside the profile of an MD structure obtained by Cerdan et al. [210] (purple) as well as the cryo-EM wide-open structure PDB: 3JAE (red) and the desensitized crystal structure with currently highest-resolution PDB: 5TIN (orange). The radii of a dehydrated and a hydrated chloride ion are indicated by dashed green and cyan lines, respectively. Figure reproduced from [221] under a CC BY 4.0 license.

state. Figure 3.12d-f illustrates that occasionally chloride ions can penetrate the transmembrane pore region. In three simulations with a collapsing pore of 300 ns length each, 3, 3, and 0 chloride ion permeation events occur, respectively. Nevertheless, this is more than an order of magnitude less than the 52 chloride permeation events over 300 ns observed in the simulations with a stable open state. The chloride ion density obtained from the collapsed simulations has a value below 0.5 particles/nm³ between the L9' and the P-2' gate. Hence, no green isosurface is visible in this region in Figure 3.12a-c. In light of these results, I conclude that Cerdan et al.'s GlyR structure does not correspond to a stable open state, but rather to a collapsed state.

3.4 Discussion

Simulations of open states of the pLGIC superfamily undergo a hydrophobic collapse once artificial restraints on the protein structure are released [70, 210–212]. This is why it has become common practice to apply restraints on the protein backbone

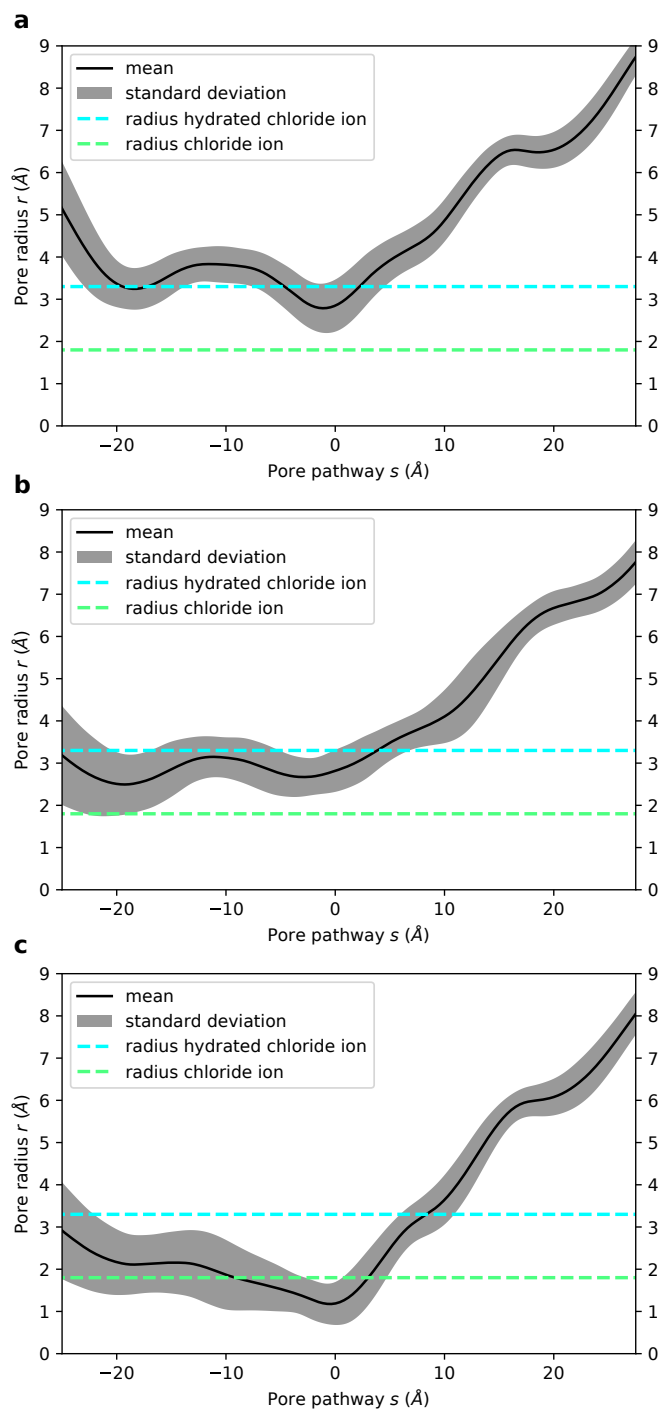


Figure 3.11: Pore profiles from three independent simulations with a collapsed state. Time-averaged pore radius profiles of the transmembrane domain obtained from three independent simulations (of 300 ns length each) initiated with different velocities from the system obtained immediately after the short initial equilibration before the structure is relaxed via the application of minimally invasive pore restraints. The radii of a dehydrated and a hydrated chloride ion are indicated by dashed green and cyan lines, respectively. The L9' and P-2' rings are the main constriction points. If at all, only partially dehydrated chloride ions could theoretically permeate through the pore. Figure reproduced from [221] under a CC BY 4.0 license.

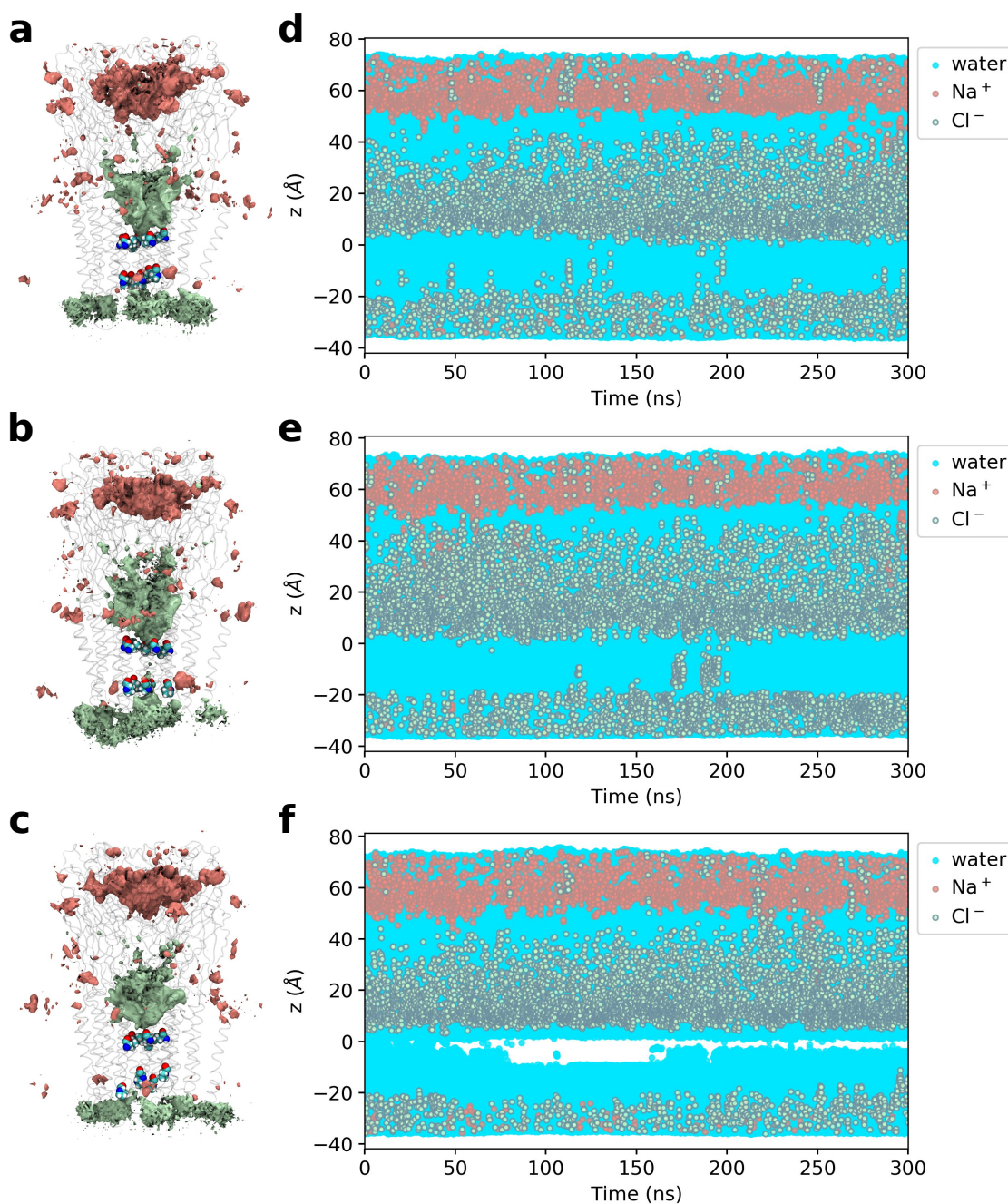


Figure 3.12: Ion densities and trajectories of water and ions within 5 Å of channel axis for simulations with a collapsed state (150 mM NaCl with no membrane potential). (a-c) Densities of chloride (green) and sodium (red) ions over 300 ns production run. The isosurfaces shown represent a density value of 0.5 particles/nm³. Transparent ribbons indicate the receptor backbone, The L9' ring and P-2' ring residues are shown in van der Waals representation. (d-f) Trajectories of water as well as chloride and sodium ion z-coordinates within 5 Å of the channel axis inside the pore over 300 ns (represented by blue, green and red circles, respectively). The 0 point of the z-axis is positioned at the L9' ring and the receptor structures in the left panels are aligned and scaled correspondingly to allow for spatial orientation along the z-axis. The white stretches in (f) demonstrate local dewetting of the pore. Chloride ions access the transmembrane pore much less frequently than in the stable open state simulations. Figure reproduced from [221] under a CC BY 4.0 license.

to artificially fix its position [54, 208, 209]. Cross-pore restraints have been used before to restrain the subsection of the channel near the hydrophobic gate [231, 232]. Interestingly, Chiodo et al. [232] observed a stable hydrated state of a homology model of the nAChR, upon the release of such restraints. However, although they observed water permeation through the channel, not a single ion permeation event was observed in the course of 800 ns (aggregated) simulation time. A similar lack of ion permeation was also reported for an open state model of the GluCl [211]. Thus, it is unclear to what extent those studies capture a genuine open state. GLIC is a prokaryotic member of the pLGIC superfamily and a structure exists which has been annotated as an apparently open state. A comparison of our open state with the 3EAM structure [233] shows that the isoleucine of GLIC at the equivalent 9' position, is pointing slightly more towards the channel pore (Figure 3.13). However, this structure is stabilized in a non-physiological way by non-polar regions of detergents molecules, which are trapped inside the pore and exclude water in this region. Thus, a direct meaningful comparison here is difficult.

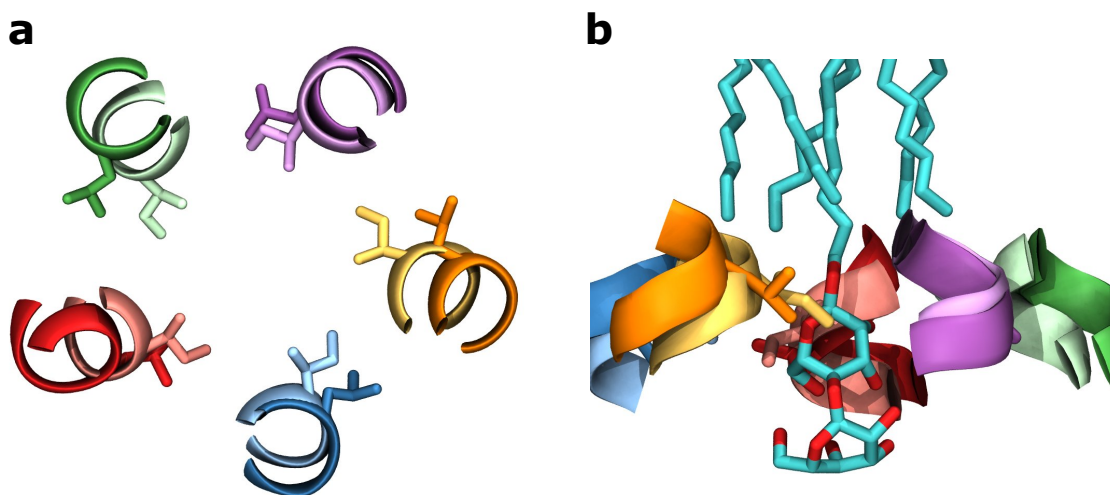


Figure 3.13: Comparison of the leucine gate between the representative open state GlyR structure from simulations and the GLIC open state structure 3EAM [233]. The open state model (dark colours) was fitted to the C_{α} atoms of L9' of the GLIC structure (light colours). (a) is a top-down view of a horizontal slice at the L9' position and highlights that the 9' isoleucine residues of GLIC are pointing slightly more towards the channel pore. (b) Side view of the same overlay, but with detergent molecules in the channel pore of the 3EAM structure shown explicitly. They provide an artificially hydrophobic environment and exclude water from this region of the channel pore.

Recently, Polovinkin et al. [234] reported that while simulations starting from one structure of the 5-HT₃R (I2, PDB: 6HIQ) gave the familiar hydrophobic collapse, simulations starting from another (F, PDB: 6HIN) allowed ions and water to flow throughout the trajectory and that wetting was correlated to rotation of L9' out of the pore lumen. The 6HIN structure shows that the L9' leucines are indeed rotated away from the pore axis (Figure 3.14), consistent with the refined model of the GlyR presented here and supporting the view that this corresponds to an open state. By forming strong hydrophobic interactions with residues of the M2 and M1 transmembrane helices of the complementary subunit, the leucines at L9' lock the pore into a metastable open conformation.

It is insightful in this context to revisit the original density map of the wide-open 3JAE structure from cryo-EM [68] paying particular attention to the areas of the key conformational changes that stabilize the open state, namely the L9' and P-2' region. Close inspection shows no density for the orientation of the L9' gate side chains (Figure 3.15a), meaning that their atomic coordinates are purely a result of computational modelling and that alternative leucine side-chain conformations are equally probable.

Inspection of the density at the P-2' gate region also shows that the map is of particularly poor resolution in this region (Figure 3.15c and d) and likely corresponds to the high mobility of this region. The data would be consistent with an ensemble of asymmetric P-2' gate conformations which, when averaged, result in a symmetric distribution of density. Consistent with this, the simulations confirm a higher mobility in this area, as measured by the root mean square fluctuation (RMSF) of C_α atoms (Figure 3.16). While typical RMSF values for C_α atoms in α -helices are around 0.7 Å, these values are significantly increased near the P-2' gate (> 1 Å). This means more conformational diversity of the intracellular pore mouth is accessible at a physiological temperature of 37 °C. What is observed in the simulations, is that during two of the simulation runs, the prolines are highly dynamic yet remain as two pairs and one lone proline. In the third simulation, a lone proline starts to interact with a pair and the simulation proceeds as an

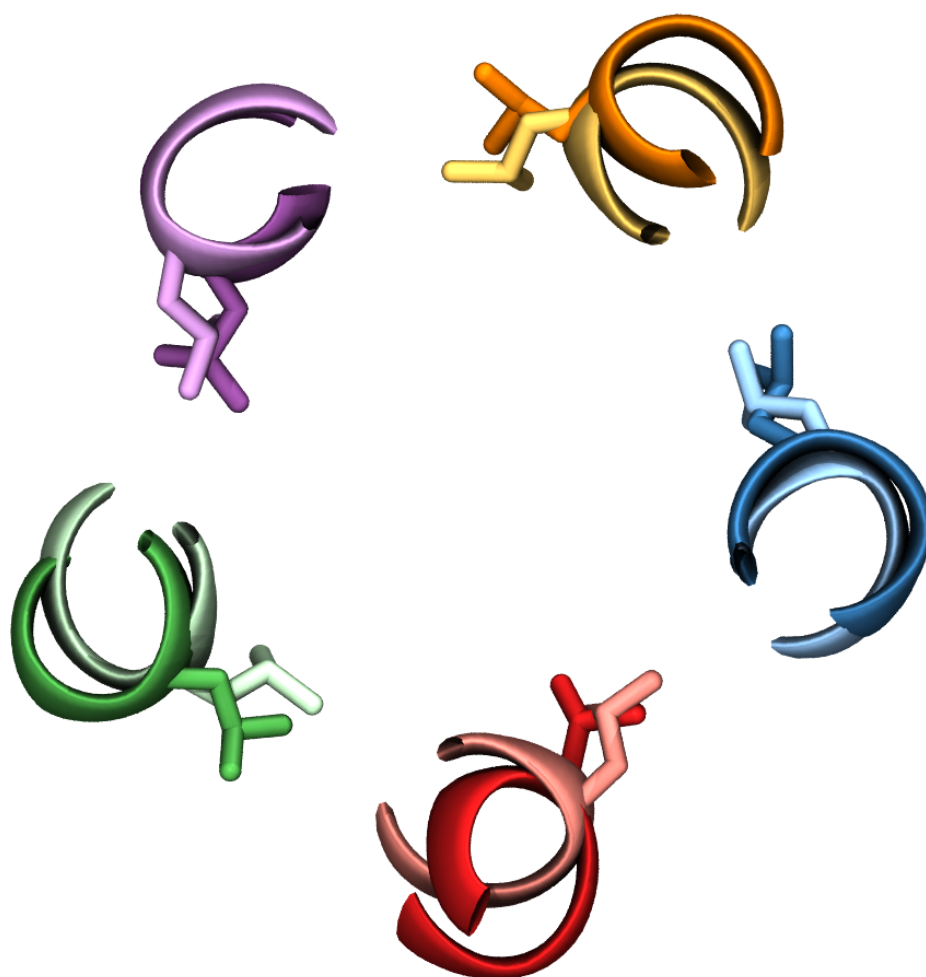


Figure 3.14: Comparison of the leucine gate between the representative open state GlyR structure from simulations and the 5-HT₃R. The open state model (dark colours) was fitted to the C_α atoms of L9' of the 5-HT₃R structure (light colours, PDB: 6HIN). Shown is a top-down view of a horizontal slice at the L9' position and highlights the similarity in position of the leucines at this position in an open state. Figure reproduced from [221] under a CC BY 4.0 license.

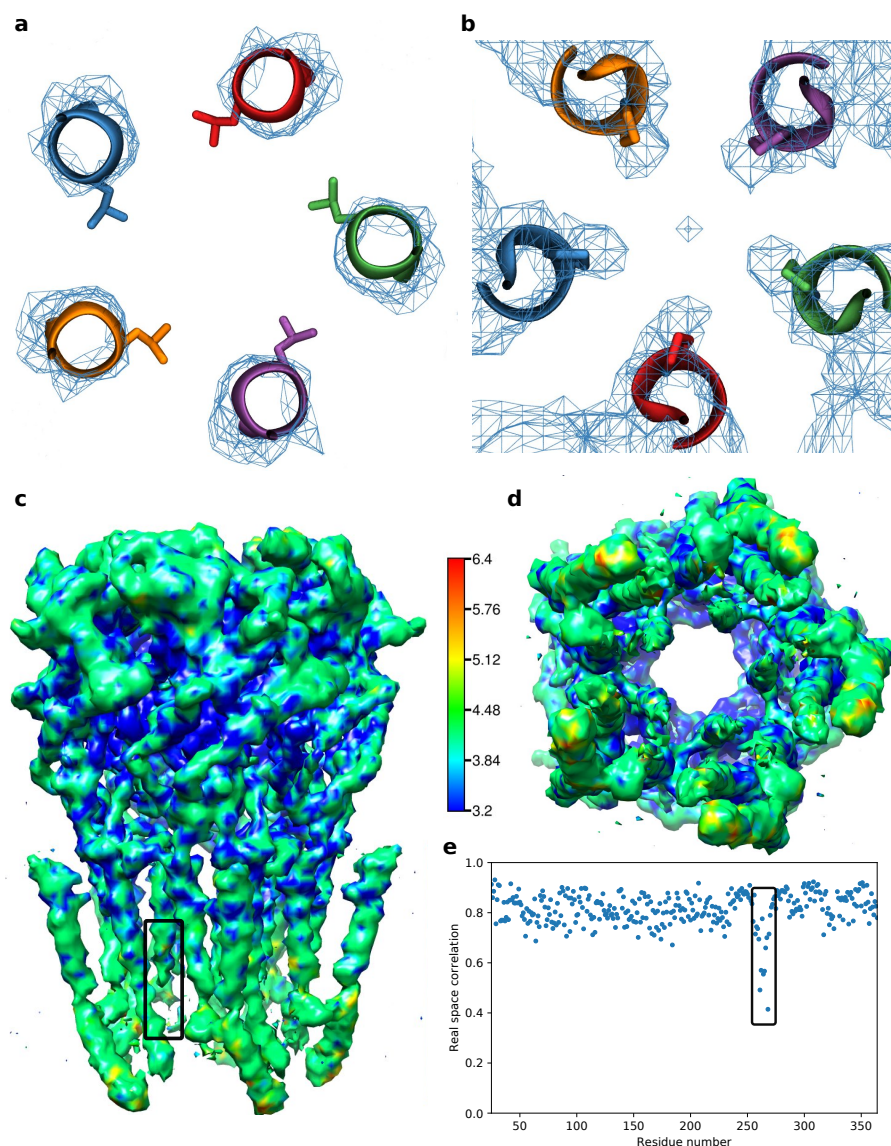


Figure 3.15: Cryo-EM density map and fitted model of PDB: 3JAE structure [68]. (a) Density of the L9' gate region, contoured at an isosurface value of 6.0 together with the fitted model structure for PDB: 3JAE (viewed from the extracellular toward the intracellular side). The density is insufficient to indicate the leucine side chain orientation, which is therefore purely a result of computational modelling. (b) Density of the P-2' gate region, contoured at an isosurface value of 5.8 together with the fitted model structure for PDB: 3JAE (viewed from intracellular toward the extracellular side). The density suggests a narrower -2' pore restriction as the fitted model. (c and d) Density map coloured according to local resolution estimated using RESMAP and visualized with Chimera: side view (c) and view from the intracellular side toward the extracellular side (d). The local resolution is particularly poor in the lower pore lining M2 helix region and its linker to the M1 helix (highlighted by the box) which suggests a higher dynamic flexibility in this region. (e) Real space correlation between the fitted atomic model and the density map calculated with PHENIX. The correlation is particularly poor in the lower half of the pore lining M2 helix and its linker to the M1 helix (highlighted by black box), indicating that the quality of the atomic model is particularly poor in this region. Figure reproduced from [221] under a CC BY 4.0 license.

arrangement of three and two prolines, thus maintaining a dynamic asymmetry. The density map in the P-2' ring region shows that there is density for proline further towards the pore centre (Figure 3.15b), suggesting that on average the pore radius might in reality be ca. 1 Å or so smaller than the unusual wide-open -2' Pro radius of the modelled atomic structure. This is an interesting discrepancy between the density map and the fitted model, as our simulations also suggest a smaller P-2' pore radius of comparable magnitude (see Figure 3.3a). Moreover, the real space correlation between the modelled structure for 3JAE and the density map is particularly meagre in this region, with R0' peaking with a very low correlation below 0.4 (Figure 3.15e). This means that the atomic model 3JAE is of poor quality in the M1-M2 linker and the lower M2 region, where the P-2' gate is located.

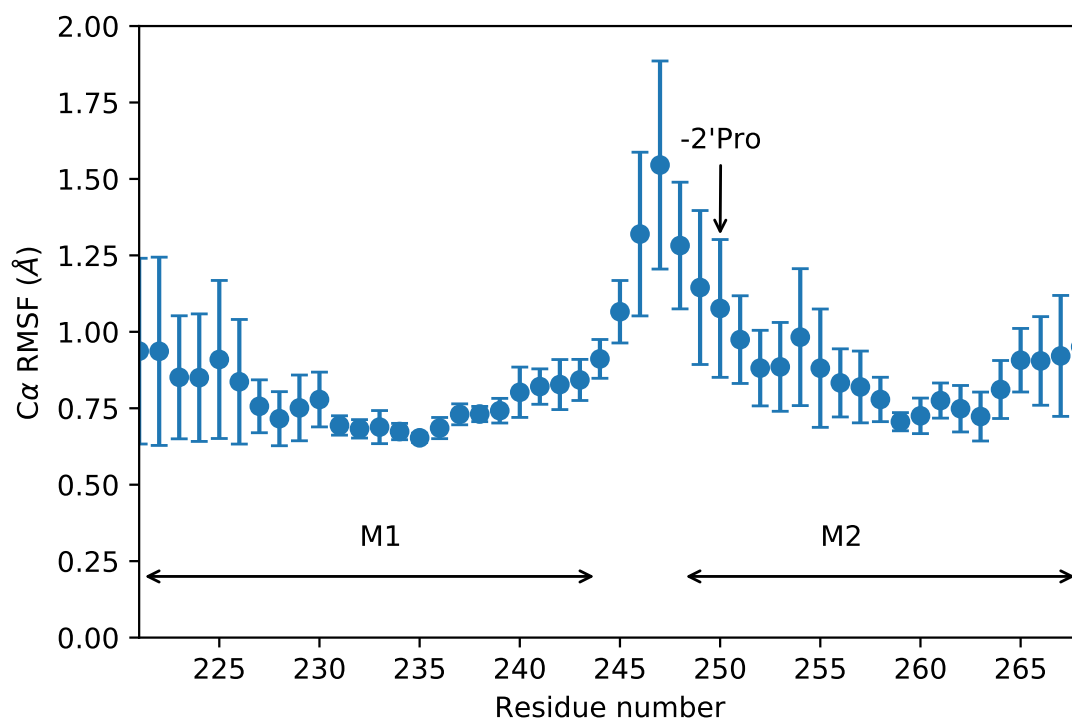


Figure 3.16: Root mean square fluctuation (RMSF) of C_{α} atoms of M1 and M2 helices averaged over all five subunits with standard deviation as error bars. The values corresponding to the region near the intracellular pore opening of the pore lining M2 helix (where the P-2' gate is located) are very high compared to typical values for α -helical C_{α} atoms of around 0.7 Å. This indicates a high conformational flexibility that is accessible at a physiological temperature of 37 °C and can explain the poor density in the cryo-EM map in this region. Figure reproduced from [221] under a CC BY 4.0 license.

In conclusion, the results presented here demonstrate the potential of MD simulations to refine structures of channel proteins derived from cryo-EM or X-ray crystallography and highlight the potential issues surrounding them. MD has the power to help resolve these problems and indeed flexible fitting methods [235], seem like a promising way forward to utilize this more directly at the model building stage. In the case of the pLGIC superfamily, the question of what truly represents the physiological open state has been subject to debate since 2015. Although it is clear that the observation of collapsed, mostly non-conducting, states are not likely representative of the true open state, it does raise the question of whether these collapsed states are in any way physiologically relevant. It may still be the case that they represent some kind of intermediate that is part of the gating process (including, for example, recovering from desensitization). If this was the case, the gating process would not be a concerted mechanism. To resolve that is likely to require substantial further work.

3.5 Conclusions

In this work, I have reported an MD-refined open and ion-conductive structure of the GlyR as a representative open state of the pLGIC superfamily. In light of the simulation results, and in revisiting the original cryo-EM density map together with the fitted atomic model 3JAE, I provide a new interpretation of the experimental data. I argue that the open state is stabilized by the L9' residues, whose orientation is not resolved in the cryo-EM density map, being buried into hydrophobic pockets at the subunit interfaces, thus locking the receptor in an open state at the 9' ring. I also suggest that the poor resolution of the cryo-EM map at the P-2' ring is the result of high dynamic flexibility in this region and the open state corresponds to an (on average symmetric) ensemble of possible asymmetric arrangements of the proline residues. Moreover, I argue that the -2' constriction is ca. 1 Å smaller in radius than the fitted atomic model 3JAE suggests, which is underpinned by the original cryo-EM density in this region as well as the MD-simulation results and would be more in line with other open structures of the pLGIC superfamily. It should be

noted that the proline at the -2' position is not conserved amongst other members of the Cys-loop family, and whether the equivalent residues behave in a similar manner remains to be seen. On the other hand, the non-polar 9' position is highly conserved across the pLGICs (L or occasionally I/V) and so are the hydrophobic pockets. Thus, the new hydrophobic lock mechanism reported here is likely to play a key role in gating for the whole superfamily.

4

Protein-Lipid Interactions in a Neuronal Membrane

Contents

4.1	Introduction	102
4.2	Methods	106
4.2.1	Homology Modelling	106
4.2.2	System Set Up	107
4.2.3	Simulation Protocol	108
4.2.4	Analysis and Visualization	108
4.3	Results	108
4.3.1	The Lipid Environment is Dependent on the Receptor State	108
4.3.2	The Duration of Protein-Lipid Contacts is Dependent on the Receptor State	112
4.3.3	Membrane Lipids Can Interact with the Extracellular Domain of the Receptor	115
4.3.4	Cholesterol Can Bind to the Binding Site of the Positive Allosteric Modulator Ivermectin	123
4.3.5	The Binding Mode of the Anionic Lipid DOPS in GlyR Differs from the PIP2 Binding Mode in the GABA _A R	126
4.4	Discussion	128
4.5	Conclusions	132

pLGICs have been shown to be sensitive to their membrane environment, yet our understanding where lipid binding occurs and how it modulates receptor function is still rudimentary. This chapter reports a study, where I have embedded

the GlyR as a prototypic member of the pLGIC superfamily in a realistic model of an in-vivo mimetic neuronal membrane and performed coarse-grained molecular dynamics simulations to probe and characterize the molecular level detail of these interactions in the inactive and active state of the receptor.

4.1 Introduction

pLGICs are modulated by their lipid environment, but exactly where lipids bind, and how they act, is poorly understood [119, 236, 237]. Via reconstitution experiments, cholesterol (CHOL) and anionic lipids have been demonstrated to be crucial for nAChR function [238–240]. In pure phosphatidylcholine (PC) or phosphatidylethanolamine (PE) membranes, nAChRs bind agonist, but remain in an unresponsive, non-conductive state (which was termed the uncoupled state). However, ternary mixtures of PC with CHOL and the anionic lipid phosphatidic acid (PA) completely restore nAChR functionality. Interestingly, binary mixtures of PC/CHOL and PC/PA alone, are not sufficient to stabilize a large pool of agonist-responsive nAChRs. Furthermore, from biophysical studies the M4 transmembrane helix has emerged as playing a role as a lipid sensor [241]. Since mutations along this helix alter gating, M4-lipid interactions are likely to influence the gating pathway [242–244]. Detailed mechanistic insight at a molecular level remains elusive, but recent experimental evidence supports the so called M4-M1/M3 lipid sensor model which hypothesizes that enhancing M4-M1/M3 interactions potentiates pLGIC function [245]. These interactions are facilitated by aromatic residues lining this interface [246], but if they are lacking, then lipids might modulate the enhancement of M4-M1/M3 interactions. Experimental evidence for this hypothesis is provided by studies of GLIC and ELIC in pure PC membranes [245]. ELIC has very few aromatic residues at the M4-M1/M3 interface, just like the nAChR, and reconstituted in pure PC membranes, is equally unresponsive upon agonist binding. GLIC, on the other hand, has more aromatic residues at the M4-M1/M3 interface and, remains functional in pure PC membranes. However, by engineering additional

aromatic residues at the M4-M1/M3 interface of ELIC, its functionality in PC membranes can be restored.

Increasingly more structural data of pLGICs with resolved lipids, neurosteroids and detergents is becoming available, providing insights into the molecular details of protein interactions with lipids or lipid-like molecules. Known binding sites for phospholipids are located at the extracellular half of the TMD on the surface of the the M1 and M4 helices as seen in a recent GABA_AR structure (PDB: 6I53 [55]) or inserted between the M4 helix and the adjacent M1 and M3 helices in several GLIC structures (e.g. PDB: 3EAM [233] and 6HZW [58]). Phospholipids can also bind at the interface of adjacent subunits at the extracellular TMD half between (+)M3 and (-)M1 as showcased by the crystal structure of GluCl (PDB: 4TNW [57]) which overlaps with the binding site of the positive allosteric modulator ivermectin in GlyR (PDB: 5VDH [123] and 3JAF [68]) and GluCl (PDB: 3RIF [66]). Interestingly, the anionic lipid phosphatidylserine (PS) is a competitor for the ivermectin binding site in GluCl [57], showcasing the competitive nature for lipid occupation on the receptor surface. Moreover, the ivermectin binding site overlaps with the interaction site of neurosteroids. Cholesteryl hemisuccinate is a detergent that is often used to replace CHOL in crystallization of membrane proteins and is bound to the ivermectin binding site in a GABA_AR chimera (PDB: 5OSA [247]), suggesting that cholesterol itself might be able to interact at the ivermectin binding site. Furthermore, the endogenous neurosteroid allotetrahydrodeoxycorticosterone (THDOC) acts as a potent positive allosteric modulator of GABA_ARs and binds in a similar location. Anionic lipids are found in the intracellular membrane leaflet and accordingly, a phosphatidylinositol 4,5-bisphosphate (PIP2) binding site was unambiguously identified in the intracellular leaflet half of the TMD next to the M3 and M4 helices of the $\alpha 1$ subunits in the GABA_AR structure (PDB: 6I53 [55]). Two CHOL molecules have recently been resolved at the intracellular leaflet TMD half near the M3 and M4 helices of a cryo-EM structure of the nAChR (PDB: 6CNJ and 6CNK [248]) but it is also thought to interact with the M4 helix of

the nAChR [245]. An comprehensive list of all current pLGIC structures with lipids resolved can be found in [237].

While insights from growing structural data are valuable, most are obtained under non-physiological conditions. For example, the artificial crystal lattice organization in crystal structures might lead to altered lipid positions as opposed in a natural membrane environment. In reality, different lipid types compete over interaction sites on the protein surface, but experimental conditions rarely account for this. For example when obtaining the crystals for GluCl (PDB: 4TNW) only 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) was added as a single lipid and bound to the ivermectin binding site [57], yet other lipid types might have a higher affinity to this site. Unnatural detergent molecules might preferentially bind to what would be actual lipid sites in a natural membrane environment. The most recent cryo-EM structures of receptors reconstituted in lipid nanodisks provide the closest mimic to a natural membrane environment, yet the influence of artificial surface tension on protein structure and dynamics is yet to be determined. Very rarely, lipid densities from experimental techniques allow for unambiguous determination of headgroups or chain length and thus, assignment of the actual lipid type. It is not uncommon that lipid density is simply modelled as a PC molecule with no justification (see for example the supplementary information of [233] or the methods section of [55]). Moreover, lipid interactions are of a highly dynamic nature [249], which poses a limit on the resolution that can be obtained for lipid densities.

Molecular dynamics (MD) simulations provide a non-invasive complementary tool to experimental techniques that allows to probe protein-lipid interactions in a natural membrane environment under physiological temperature and pressure [250]. Particularly, coarse-grained (CG) MD simulations (using, e.g. the popular MARTINI model [149]) allow probing the protein-lipid interactions at long simulation time scales which is required to rigorously sample the protein-lipid interaction space [249]. They have been successfully applied to provide structural insights into protein-lipid interactions of key functional membrane proteins, such as transporters [251], ion channels [252] and GPCRs [253], to name a few.

While it is experimentally known that pLGICs are sensitive to their lipid environment and that CHOL and anionic lipids are both essential for nAChR function, it is still the standard approach to perform simulations of this receptor class in simple bilayer models that only consist of a single lipid class, usually POPC (e.g. [54]). Few simulations have included the functionally crucial lipid types, nevertheless in very simplified membrane models of binary or tertiary composition with no asymmetry between leaflets [254, 255]. Yet, *in vivo* plasma membranes consist of many different lipid types that are distributed asymmetrically across the leaflets. However, so far no simulations of pLGICs in such a realistic membrane environment have been performed. Based on extensive lipidomics data, Ingolfsson et al. [256] have derived a coarse-grained model of a neuronal membrane with asymmetric leaflet composition which can be used as an *in vivo*-mimetic model for the natural environment of pLGICs.

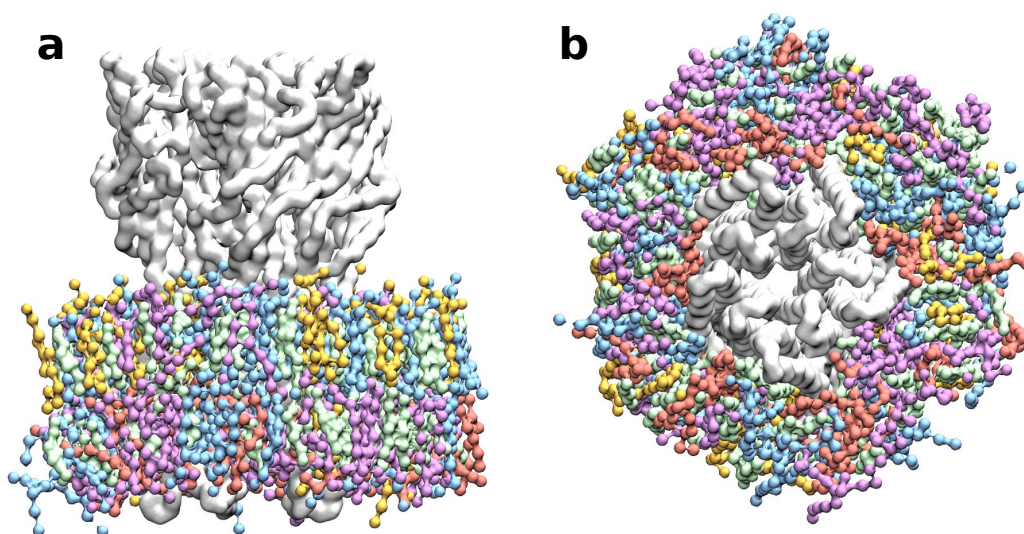


Figure 4.1: The simulation system (a) viewed in parallel to the membrane plane and (b) viewed from the intracellular side. GlyR (light grey) is embedded in a neuronal model membrane (POPC: blue, POPE: purple, SM: gold, CHOL: green, DOPS: red), water and ions are not shown for clarity.

Here, I have used molecular modelling to reconstitute a prototypical pLGIC, namely GlyR, in an *in vivo*-mimetic neuronal membrane (see Figure 4.1) and performed CG-MD simulations to probe its protein-lipid interactions. The membrane

model is a simplified version of [256] that consists of representatives for the five main lipid types found in neuronal plasma membranes: PC, PE, sphingomyelin (SM), PS, and CHOL. The reduction of the model to these five lipid species is necessary to improve the sampling of the protein-lipid interactions for each lipid type. Yet, it still contains all essential lipid types, particularly the functionally relevant CHOL and anionic PS, in a typical asymmetric composition of a neuronal membrane. Moreover, in order to examine possible state-dependent protein-lipid interactions, I have performed two separate sets of simulations, each containing a different state of a human homology model of the $\alpha 1$ GlyR. One is derived from the inactive (closed) state (PDB: 5CFB [60]) and one is derived from the active (open) state (PDB: 3JAE [68]). Lipid densities and local residence times reveal that the different conformational states recruit a different lipid environment, thus indicating that lipids may play a modulatory role in stabilizing different receptor states. Particularly, I report so far unknown phospholipid interactions at the ECD-TMD interface which are likely to play a role in stabilizing the tertiary structure in this critical region for signal transduction and might even influence the signal transduction itself. Furthermore, CHOL can bind to the binding site of the positive allosteric modulator ivermectin and thus might itself act as a membrane-intrinsic modulation molecule. Noteworthy is also that the anionic lipid DOPS can bind near the intracellular ends of the M3 and M3 helices which overlaps with the PIP2 binding site in human GABA_AR, but the details of the headgroup interactions differ. The insights gained by elucidating a detailed map of the state-dependent protein-lipid interactions with a pLGIC could be exploited for rational drug development of lipid-like receptor modulators.

4.2 Methods

4.2.1 Homology Modelling

The active state homology model of human $\alpha 1$ GlyR was obtained as described in section 3.2.1. The model in the inactive conformation was based on the human $\alpha 3$ GlyR (strychnine-bound closed state; PDB: 5CFB [60]) and obtained analogously,

with the only difference that symmetry restraints were enforced in the model building process, as the template was missing certain residues near the M3-M4 loop in some subunits, which were present in others. For consistency, the final model of the inactive state also consisted of residues P10-R309 linked via an AGT-tripeptide to K386-K419.

4.2.2 System Set Up

The homology models of GlyR were converted to coarse-grained MARTINI representations via ‘martinize’ [149]. Using the ‘insane’ script [218], each of the coarse-grained protein structures was separately embedded in an asymmetric neuronal model membrane using a hexagonal prism as the periodic simulation box, solvated in water and sodium and chloride ions were added to neutralize the net charge and simulate a physiological concentration of 150 mM (see Figure 4.1). The membrane model is based on [256] and was simplified to consist of representatives of the five most essential lipid species as specified in table 4.1. For each receptor state, 10 systems were set up with the same number of molecules per species, but with different random seeding of the lipid positions in the membrane plane to improve sampling. No constraints were imposed to maintain the asymmetric lipid distribution across leaflets, but based on previous experience only CHOL can interchange the leaflets on timescales of tens of μ s.

Table 4.1: Composition of a human neuronal model membrane (simplified to five essential lipid types based on the model by Ingolfsson et al. [256]).

lipid species	extracellular / mol %	intracellular / mol %	total / mol %
POPC	24.4	13.7	19.2
POPE	11.1	21.5	16.2
SM	19.9	3.0	11.6
CHOL	44.7	44.9	44.8
DOPS	0	16.9	8.3

4.2.3 Simulation Protocol

All CG-MD simulations were performed with the MARTINI force field version 2.2 for protein and version 2.0 for lipids [149]. GROMACS 2018 [141] was used as molecular dynamics engine for all simulations and periodic boundary conditions were applied in all three spatial dimensions. After energy minimisation, the systems were simulated at 310 K and 1 bar for 40 μ s. Simulation details were based on the recommended new-rf.mdp parameters [219] and appropriately adjusted. Harmonic restraints with a force constant of 1,000 kJmol⁻¹nm⁻² were placed on all protein particles to preserve the conformational state close to the original crystal or cryo-EM structure.

4.2.4 Analysis and Visualization

Trajectory analysis was performed on the last 20 μ s out of a total of 40 μ s simulation time using frames at a separation of $dt=1$ ns and averaged over all 10 repeats per conformational state. Software used for analysis was MDAnalysis [226], MDTraj [257] and in-house scripts, part of which are now publicly available on <https://github.com/wlsong/PyLipID>. Visualization was performed in VMD [229].

4.3 Results

4.3.1 The Lipid Environment is Dependent on the Receptor State

In order to explore the possible modulatory role of membrane lipids on GlyR activation, I analyzed the lipid distribution of each lipid type per leaflet for both the inactive and active state (Figure 4.2). Hot spots that are strongly preferred by certain lipid types can be clearly identified. A strong colour gradient between the immediate protein's surroundings and the membrane bulk indicates that the receptor recruits a lipid-type specific environment at the protein-membrane interface.

Most importantly, CHOL shows that not only the occupation intensity, but also the location of interaction hot spots, depend on the receptor state. Most noteworthy is that in the active state CHOL in the extracellular leaflet shows a

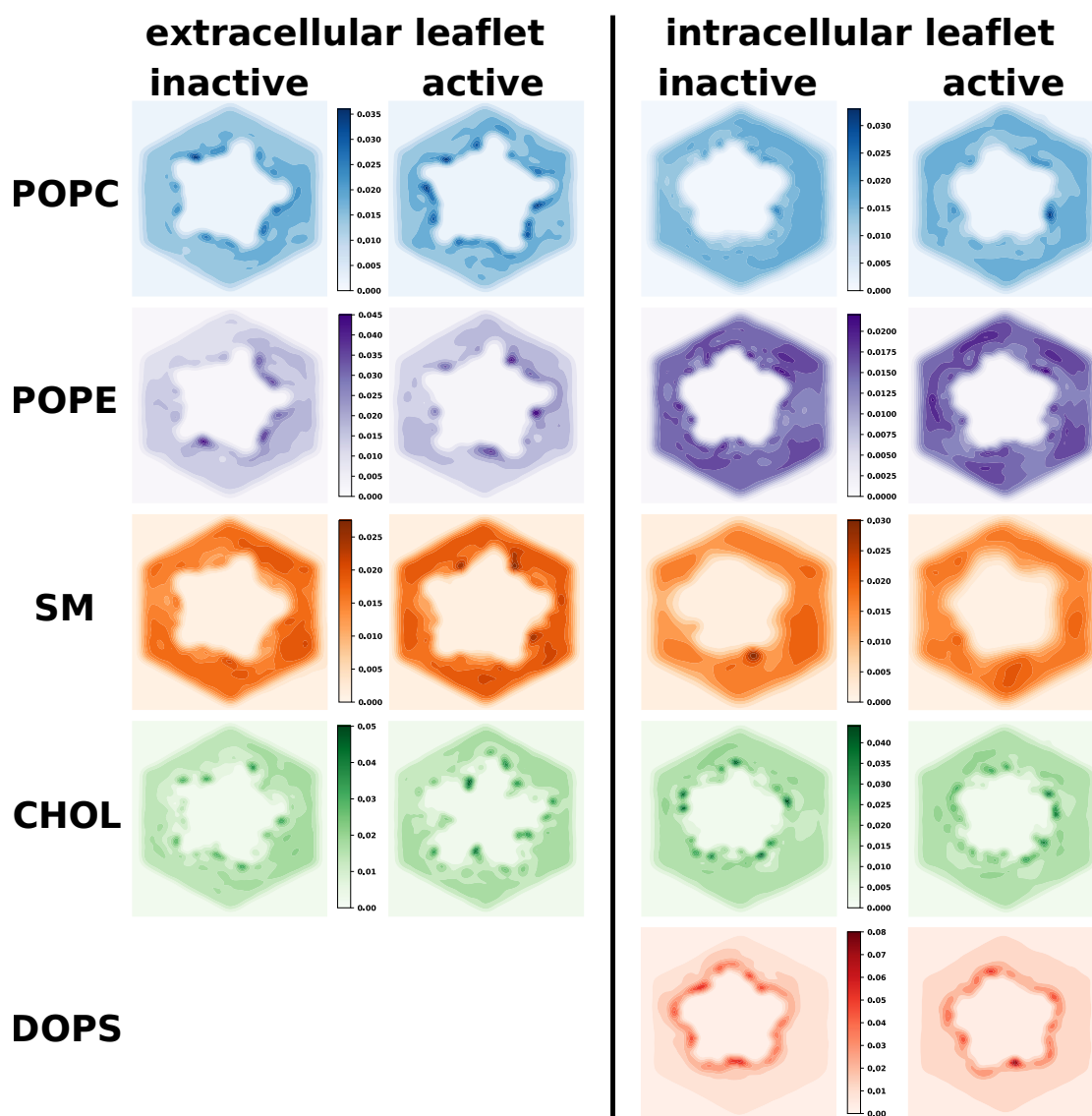


Figure 4.2: Probability densities of different lipid types around the receptor in the extracellular and intracellular leaflet for the inactive and active state. The colour scales were chosen so as to highlight the state-dependent differences per leaflet and lipid type. Darker colours represent interaction hotspots. For CHOL the location and intensity of interaction hot spots changes between the inactive and active state. POPC, POPE and SM can occupy similar sites on the receptor surface, but their intensity varies between states, suggesting a state-dependent affinity change.

strong preference to penetrate deeply into the inter-subunit interfaces. This is where the binding site of ivermectin, a positive allosteric modulator, is located and thus an important region for modulating receptor function.

At this level of resolution, the interaction hot spots of the remaining lipids (POPC, POPE, SM and DOPS) do not vary much when comparing between the inactive and the active state of the receptor. However, the relative occupation probabilities at these locations are clearly dependent on the state of the receptor, indicating that the binding affinities of these locations are sensitive to the conformational state of the receptor.

Moreover, it can be observed that POPC, POPE and SM can all occupy the same interaction hot spots, meaning that they compete over the same binding locations. The fact that these three lipid types engage in similar interactions with the receptor is not surprising, given the similarity of their headgroups: POPC and SM both have a phosphocholine headgroup, while POPE has a phosphoethanolamine headgroup which is chemically very similar, as it is still zwitterionic but less bulky, as the choline ($-\text{N}-(\text{CH}_3)_3$ group) is replaced by an ammonium ($-\text{NH}_3$) group. Note that in the coarse-grained MARTINI force field employed here, these groups only differ by a single bead (particle type), namely Q0 for $-\text{N}-(\text{CH}_3)_3$ in POPC and SM and Qd for $-\text{NH}_3$ in POPE.

If simulations were perfectly converged, one would expect a C_5 -symmetry around a five-fold symmetric receptor. While this is achieved quite well for CHOL, this is not quite the case for the other lipid types which are much less represented in the overall membrane composition and thus require longer simulation times to fully converge. It should be noted, however, that 10 repeats of 40 μs of simulation time each per receptor state well exceeds currently typical timescales of CG-MD simulations to study protein-lipid interactions which are typically in the range of 5-10 repeats of 2-10 μs [253, 258].

In order to get a more detailed picture of the protein-lipid interactions of the first lipid shell, I calculated the lipid densities in 3-d space (Figure 4.3).

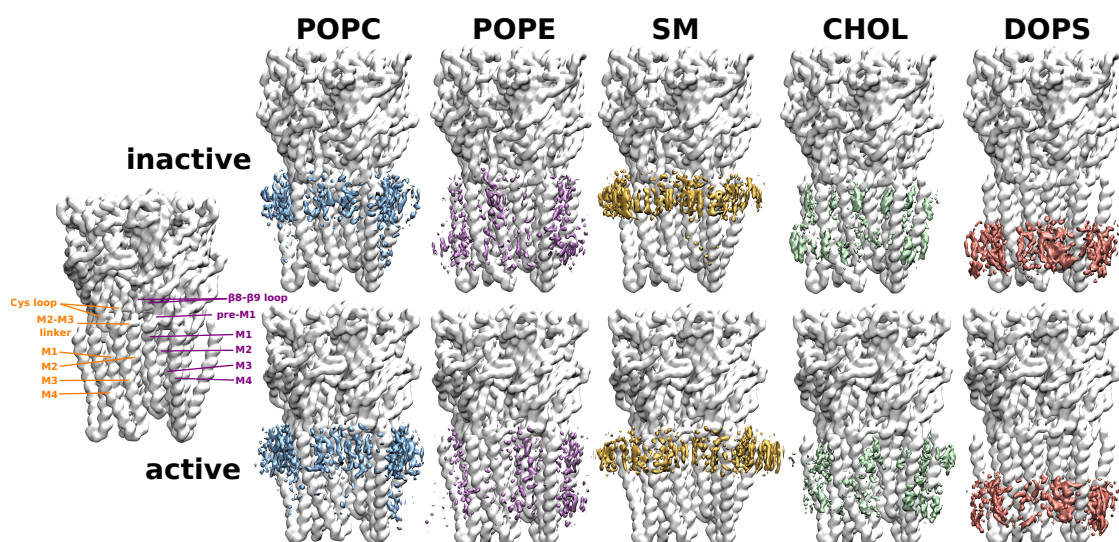


Figure 4.3: Densities of different lipid types in 3d space for the inactive and active state. For spatial orientation, the receptor structure on the left is labelled with key regions of adjacent subunits - principal (orange) and complementary subunit (purple). In order to better compare the individual lipid type differences for the two conformational states, the densities of each lipid type are shown at different isosurface values, such that they are reweighted as if the membrane consisted of equal amounts of each lipid type (the isosurface for the POPC density is shown at a value of 6.3 (6.3) molecules/nm³, POPE density at 5.4 (5.4) molecules/nm³, SM density at 3.7 (3.8) molecules/nm³, CHOL density at 15 (15) molecules/nm³ and DOPS density at 5.8 (5.6) molecules/nm³ for the inactive state (active state)). In the active state, CHOL can insert itself deep into a crevice that is opening up at the subunit interface between the (+)M3 and (-)M1 helices, where the binding site for the positive allosteric modulator ivermectin is located, indicating a possible positive modulatory role of CHOL. The density of the anionic lipid type DOPS is higher in the inactive vs. the active state, suggesting that anionic lipids may aid in stabilizing the inactive receptor state. POPC, POPE and SM can all compete for the same regions on the receptor surface, which is not surprising given their similar chemical character.

Striking state-dependent changes can be observed for CHOL. Probably most important is that transitioning from the inactive to the active state, CHOL penetrates into the subunit interface between the crevice of the (+)M3 and (-)M1 helices. Since this is where the ivermectin binding site is located, CHOL might act as an intrinsic membrane molecule with positive modulatory function in the receptor activation process. This is supported by the experimental evidence that CHOL is crucial for the activation of the nAChR, another member of the pLGIC superfamily [238–240].

Another functionally important lipid class for nAChRs are anionic lipids [13]. In the simulations, I have used DOPS which is an important anionic lipid type in neuronal membranes that only occurs in the intracellular leaflet. It is found that the DOPS density is higher in the inactive state compared to the active state, suggesting DOPS (and thus possibly in general anionic lipids) aid in stabilizing the inactive state over the active state.

Overall, the different lipid types compete with each other for the same interaction regions on the receptor surface. Particularly, the neutral phospholipids POPC, POPE and SM can all interact with the same parts of the receptor. This is not surprising, as they are chemically more similar within themselves than compared to the steroid CHOL or the anionic phospholipid DOPS (as discussed above).

4.3.2 The Duration of Protein-Lipid Contacts is Dependent on the Receptor State

As a next step towards a more detailed understanding of the local strength of the protein-lipid interactions, I analyzed the binding residues on the receptor for each specific lipid type. Binding residues would typically be identified as those residues with long contact times with lipid molecules. I therefore calculated the average duration of continuous contacts between a molecule of a given lipid species and each protein residue (using a 6 Å cut-off) as a proxy for local binding affinity. Note that this can also be seen as the local residence time for a certain lipid molecule on a specific receptor residue. For each receptor state, in addition to averaging

over all 10 simulation repeats, the data was also averaged over all five identical subunits. In Figure 4.4 I have projected this data colour-coded onto the protein structure to obtain a detailed map of the protein-lipid contact times.

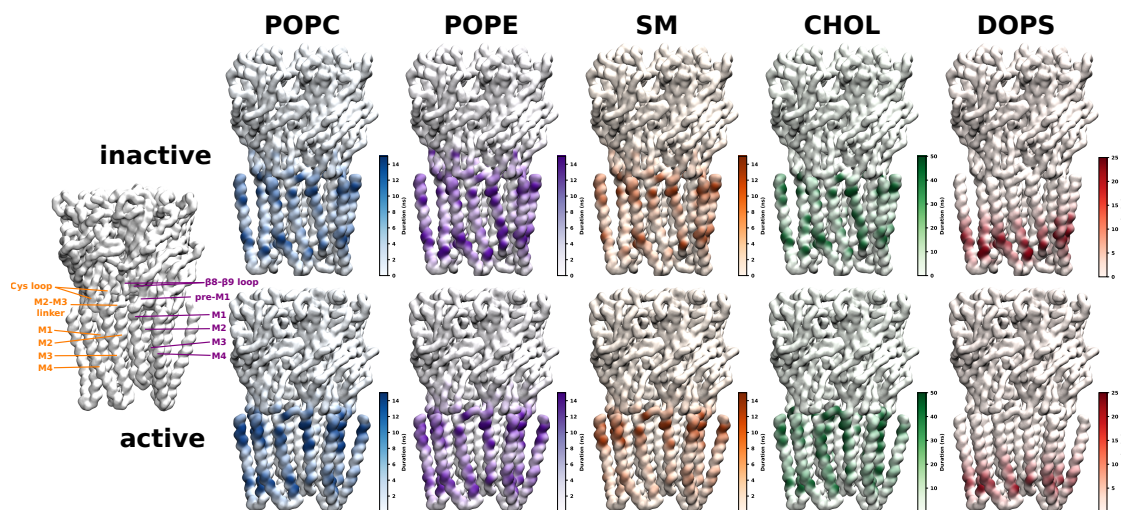


Figure 4.4: Mean duration of protein-lipid contact per residue for the inactive and active state. A cut-off of 6 Å (first lipid shell) was used to define a protein-lipid contact. For spatial orientation, the receptor structure on the left is labelled with key regions of a principal (orange) and a complementary subunit (purple). The colour scales were chosen so as to highlight the state-dependent differences per lipid type in an optimal way (the maximal contact duration can be longer than the maximum value on the colour scale, see Figure 4.5). Darker colours represent regions with longer protein-lipid contact times. The per residue-interaction times are a proxy for the local binding affinity of a certain lipid type. CHOL forms the longest-lasting contacts with the receptor, followed by the anionic lipid DOPS, and then the chemically similar phospholipids POPC, POPE, SM, which supports the experimental evidence that CHOL and anionic lipids are important for nAChR receptor function. Interestingly, all phospholipids interact with the ECD-TMD interface which is crucial for signal-transduction from the neurotransmitter binding ECD towards the TMD which contains the channel pore, thus, phospholipids could possibly modulate the conformational changes important for signal transduction at this pivotal interface.

Out of all lipid types, CHOL engages in the longest contacts with the receptor (see also Figure 4.5), indicating its most prevalent role as lipid interaction partner. The second most important lipid type in terms of interaction duration is DOPS. Finally, POPC, POPE and SM fall into a similar range of shorter contact times with the receptor. This hierarchy is in agreement with the experimental insight that CHOL and a negatively charged lipid type (such as DOPS) are essential for nAChR function.

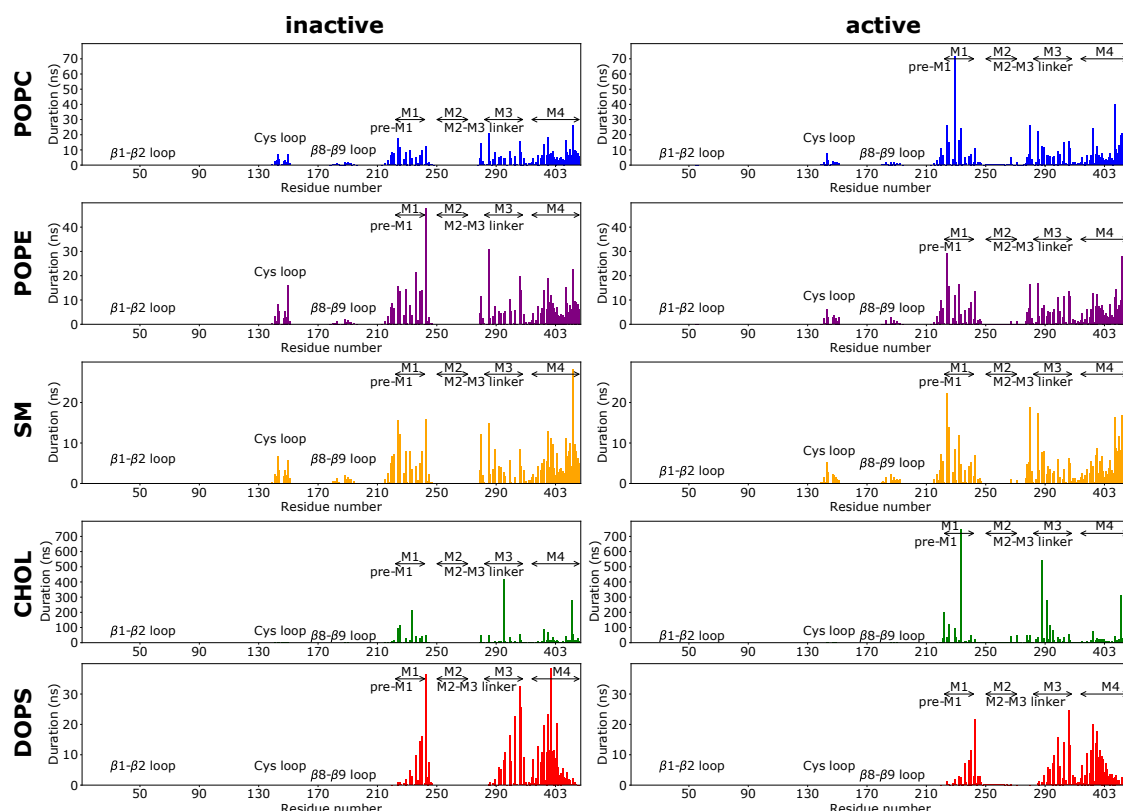


Figure 4.5: Mean duration of protein-lipid contact per residue for the inactive and active state. A cut-off of 6 Å (first lipid shell) was used to define a protein-lipid contact.

As expected due to their chemical similarity, POPC, POPE and SM have very similar contact duration maps. For these lipids, a trend towards longer interaction times in the extracellular half of the TMD is observed, when comparing the inactive with the active state. This suggests that the active state is stabilized by stronger interactions with these lipid types in the extracellular leaflet. CHOL forms much longer contacts in the active state at the subunit interface where the binding site of ivermectin, a positive allosteric modulator, is located. The average of the mean contact times of important residues in the ivermectin binding pocket¹ goes up from 6 ns in the inactive state to 15 ns in the active state. In the active state, this site is more accessible and CHOL engages in longer contacts with the TMD regions forming this allosteric site, namely the extracellular halves of the (+)M3 -helix, (+)M2 and (-)M1 helix as well as the (+)M2-M3 linker. Note that via direct interactions with

¹Important residues as shown explicitly in Figure 4.9 in the ivermectin binding pocket are D284 and A288 on the (+)M3 helix, A277, S278 and V280 on the (+)M2-M3 linker, S267 on the (+)M2 helix and I255, Q226, P230, I229 and L233 on the (-)M1 helix.

the pore-lining M2 helix (which are also formed by the positive allosteric modulator ivermectin), CHOL could directly influence the dynamics of the pore-lining M2 helices. Lastly, DOPS, which is only present in the lower leaflet, has longer contact times with the intracellular half of the TMD in the inactive vs. the active state.

Finally, I would like to highlight that the phospholipids POPC, POPE and SM can form contacts with regions of the extracellular domain, which show longer contact times in the inactive vs. the active state. This aspect will be analyzed more detail in the following section.

4.3.3 Membrane Lipids Can Interact with the Extracellular Domain of the Receptor

The ECD-TMD interface of pLGICs is of particular importance for signal transduction, as it is the communication interface between the neurotransmitter binding ECD and the TMD containing the channel pore (see left panel of Figure 4.6). Neurotransmitter binding is thought to trigger configurational changes in the $\beta 8$ - $\beta 9$ loop (=loop F), the $\beta 10$ strand that precedes the M1 helix (pre-M1), the $\beta 1$ - $\beta 2$ loop and the Cys loop (= $\beta 6$ - $\beta 7$ loop) in the ECD which form contacts with the M2-M3 linker of the TMD and thus transduce the signal into opening of the channel pore in the TMD. Note that loop C (= $\beta 9$ - $\beta 10$ loop) that covers the neurotransmitter binding site is known to close due to ligand binding and is directly linked to the $\beta 8$ - $\beta 9$ loop and the pre-M1 region.

The simulations show that phospholipids (POPC, POPE and SM) interact with the receptor at precisely this crucial communication interface. The 3-d phospholipid densities for the inactive and active state of the receptor are illustrated in the right panel of Figure 4.6. Note, however, that the side chain orientations depicted in Figure 4.6 correspond to a static structure, namely the atomistic receptor model that the coarse-grained receptor model was obtained from. Under physiological conditions, the side chain orientations would be highly flexible and dynamically adapt to the motion of an interacting lipid. The exact side chain orientations with respect to the lipid densities from this static structure should therefore not be

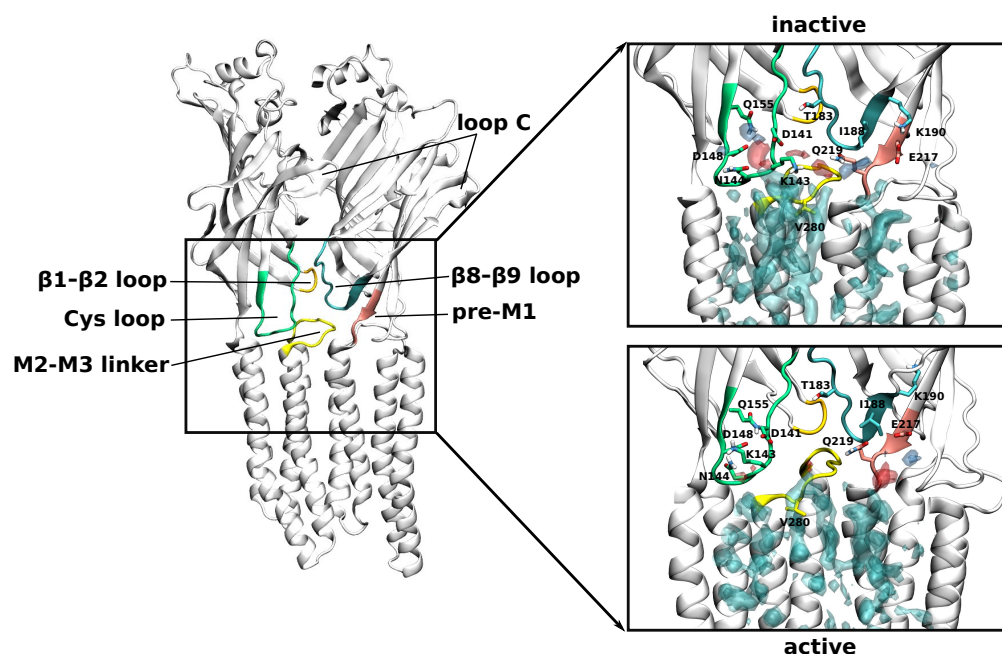


Figure 4.6: Phospholipid interactions with the ECD-TMD interface. Phospholipid densities at the communication interface between the ECD and TMD for the inactive and active state. Density arising from the positively charged headgroup beads (choline (-N-(CH₃)₃ group) or ammonium (-NH₃)) are shown in transparent blue, from the negatively charged phosphate headgroup beads (PO₄) is shown in transparent red and from the lipid tails is shown in transparent cyan. The densities corresponding to each lipid type are shown at different isosurface values, such that they are reweighted as if the upper membrane leaflet consisted of equal amounts of each lipid type (the isosurface for the POPC density is shown at a value of 8.2 (8.2) molecules/nm³, POPE density at 3.8 (3.6) molecules/nm³, SM density at 6.6 (6.8) molecules/nm³ for the inactive state (active state)). The densities were calculated from the last 20 μ s (out of 40 μ s) and averaged over 10 repeats per conformational state. In order to show the detailed atomistic interactions, the receptor is shown in an atomistic representation which formed the input to obtain the coarse-grained representation used in the simulations. In the inactive state, the phospholipid headgroups interact with regions slightly further up towards the extracellular side of the ECD. These interactions are mediated mostly via polar side chains of the Cys loop, the β 8- β 9 loop as well as the part of the β 10 strand that precedes the M1 helix (pre-M1). Note that the latter two are directly connected with loop C whose closing is key to neurotransmitter binding and thus, signal transduction towards the ECD-TMD interface.

overinterpreted. In the inactive state, the phospholipids preferentially interact with regions further up towards the extracellular side of the ECD. Particularly, the Cys loop interacts with the zwitterionic headgroups of phospholipids via the charged and polar residues D141, K143, N144, D148 and Q155. Another interaction hot spot is located more towards the complementary subunit, where the headgroups interact with Q219 and E217 from the pre-M1 region as well as T183 and K190 from the $\beta 8$ - $\beta 9$ loop. In the active state, on the other hand, the phospholipid densities are overall less far up towards the extracellular side of the ECD than in the inactive state. Moreover, the phospholipid headgroup densities are less defined near the Cys loop, indicating a higher degree of flexibility as well as a weaker interaction strength. In the active state, the phospholipid headgroup density is much more defined at the interaction hot spot mediated primarily via polar and electrostatic interactions with Q219 and E217 from the pre-M1 region of the complementary subunit, indicating more stable interactions at this site. Furthermore, a more defined lipid binding site appears in the active state at the subunit interface which is not as clearly separated in the inactive state. This density originates mostly from SM (compare also additional density spots of SM at the subunit interfaces in the extracellular leaflet in the active state in Figure 4.2) and could be due to favourable interactions between SM and CHOL which preferentially interacts with a region at the subunit interface slightly further towards the inside of the membrane and is discussed in more detail in the next section. In summary, the phospholipid density analysis at the ECD-TMD interface suggests that in the inactive state, the phospholipid headgroups form more stable interactions with the Cys loop while in the active state, they form more stable interactions with the pre-M1 region.

This is also reflected by the per residue contact times (see Figure 4.7). Overall, the contact times of phospholipids with the Cys loop are longer in the inactive state vs. the active state (Figure 4.7a), which is a proxy for a higher binding affinity of phospholipids to the Cys loop in the inactive state. Particularly, K143 emerges as the most important residue in the Cys loop. Due to its long flexible side chain with its terminal ammonium group carrying a positive electric charge, it can act

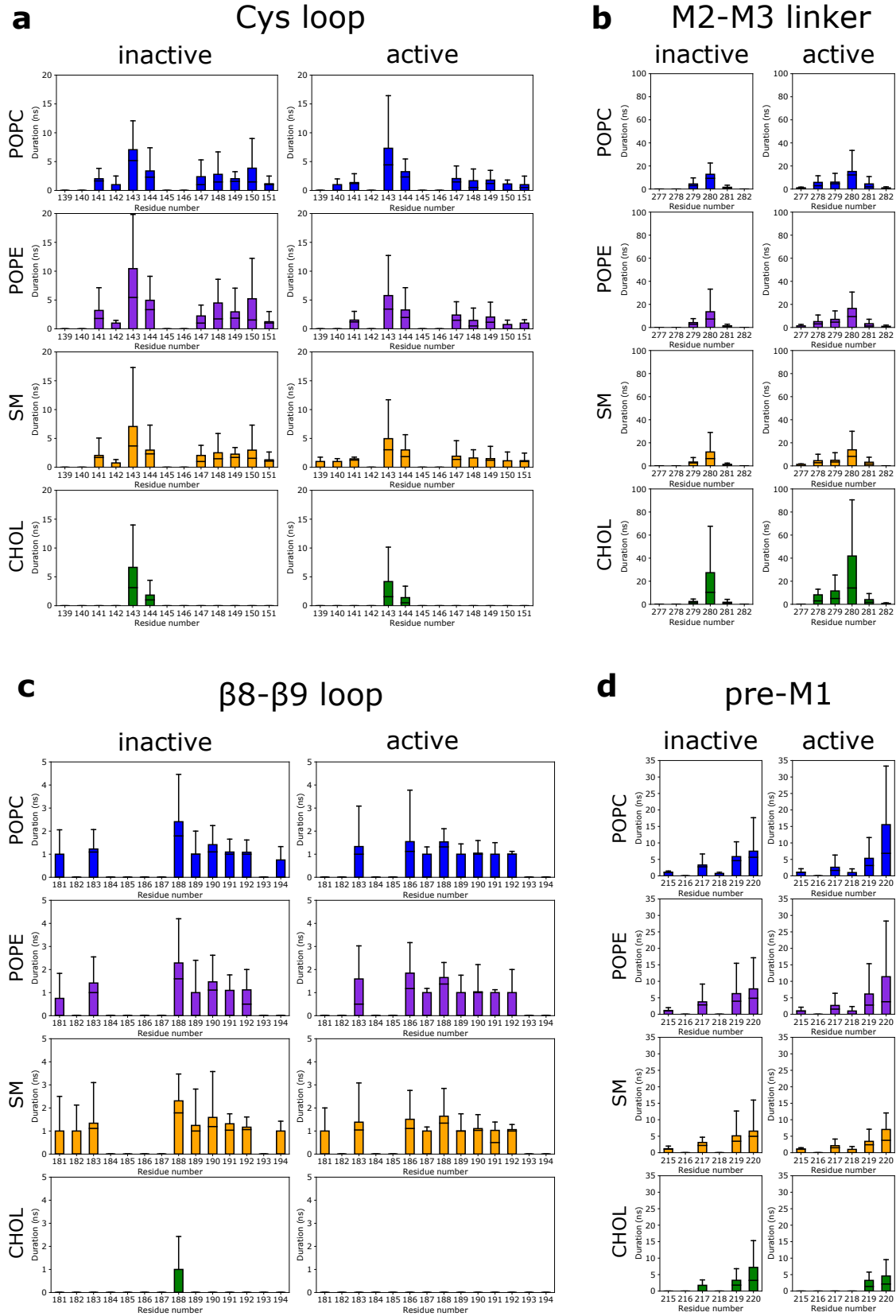


Figure 4.7: Mean duration of protein-lipid contact per residue for the inactive and active state at the ECD-TMD interface. Box plots show the distribution from 10 repeats and five subunits per conformational state for (a) the Cys loop, (b) the M2-M3 linker, (c) the $\beta 8$ - $\beta 9$ loop and (d) the pre-M1 region on the $\beta 10$ strand. Note the different y-axis scaling to better compare the differences for each of these four regions.

as a lipid recruiter, both by forming electrostatic interactions with the negatively charged phosphate groups of the phospholipid headgroups and as H-bond donor to the oxygen of the CHOL hydroxy group (cf. lipid snorkeling by lysine [259]). On the other hand, the contact times with the pre-M1 region are dominated by longer contact times of the phospholipids POPC and POPE with M220 in the active state compared to the inactive state (Figure 4.7d), indicating a higher affinity of these phospholipids to the pre-M1 region in the active state. Since methionine's side chain is rather hydrophobic, the main driving force for lipid interactions in the pre-M1 region seems to be of hydrophobic nature, in contrast to the strong polar driving forces of the lipid interactions with the Cys loop. Out of all the parts that make up the ECD-TMD interface, the M2-M3 linker shows overall the longest contact times with lipids, which is not surprising, given it is nearest to the TMD. The lipid contact intensifies in the active state compared to inactive state, both in terms of the numbers of residues in contact with lipids (residues 277-282 in active vs. 279-281 in inactive state) as well as in terms of the individual length of the per residue lipid contacts with V280 being the most dominant lipid interaction residue, suggesting that lipid binding in this region is primarily driven by hydrophobic interactions. Finally, the $\beta 8$ - $\beta 9$ loop has much shorter contact times with lipids than the other parts of the ECD-TMD interface and thus its role for lipid modulation is likely to be less important. However, short lipid contacts of a few nanoseconds length might be sufficient to lower the energy barrier for conformational changes in this loop. The main differences between the inactive and active state are that I188 contact times are longer in the inactive state vs. active state, while residues 186 and 187 only make contact with lipids in the active state. Note that no lipid contacts were formed with the $\beta 1$ - $\beta 2$ loop in neither the inactive nor active state (data not shown). This loop is located too far away from the membrane and located towards the inside of the ECD part of the ion channel pore to be accessible to membrane lipids.

While for the GlyR, no phospholipid density has been resolved in the structures available at present, structural data of other receptors of the pLGIC superfamily with phospholipid or detergent sites at the ECD-TMD interface exist for comparison

with the lipid densities observed in the simulations. Since these structures are more similar to the active state (open or desensitized channels), I compare them with the phospholipid densities obtained from the active state simulations. Three phospholipid interaction sites per subunit interface are observed in the active state (see Figure 4.8), resulting in a total of 15 phospholipid binding sites for the whole pentameric receptor. The first one is located between the M3 and M4 helices of the principal subunit with the headgroup interacting with the Cys loop, the second one is at the subunit interface between the (+)M3 and (-)M1 helices and the headgroup interacting with the (+)M2-M3 linker and a third one is between the M1 and M4 helices of the complementary subunit with the headgroup interacting with the pre-M1 region. The overall position of the first and third binding site agree well with existing structural data. The second binding site at the subunit interface is a new binding site predicted by simulation which has not been observed by structure determination methods so far.

The first phospholipid binding site identified from the simulation data coincides well with a detergent molecule site resolved in a recent GLIC crystal structure (PDB: 6HZW [58]) (see Figure 4.8a). The polar detergent headgroup forms polar interactions with the positively charged R117 and R118 side chains in the Cys loop of GLIC. Similarly, in the human $\alpha 1$ GlyR the corresponding polar side chains K143 and N144 in the Cys loop accommodate the negatively charged phosphate headgroup density originating from phospholipids via polar interactions. The hydrophobic detergent tail in the GLIC structure interacts with the membrane facing surface of the M3 and M4 helices which is also where the phospholipid tail density is located in the simulations.

The third phospholipid binding site identified by simulation is also found in the same GLIC structure as well as in a recent GABA_AR cryo-EM structure, although the structurally resolved binding modes differ slightly among each other and with regard to the density observed in the simulations. In GLIC (see Figure 4.8b), a phosphocholine molecule is bound between the M1 and the M4 helix. Note that the M4 helix in the GLIC structure is kinked slightly more away from the M1/M3

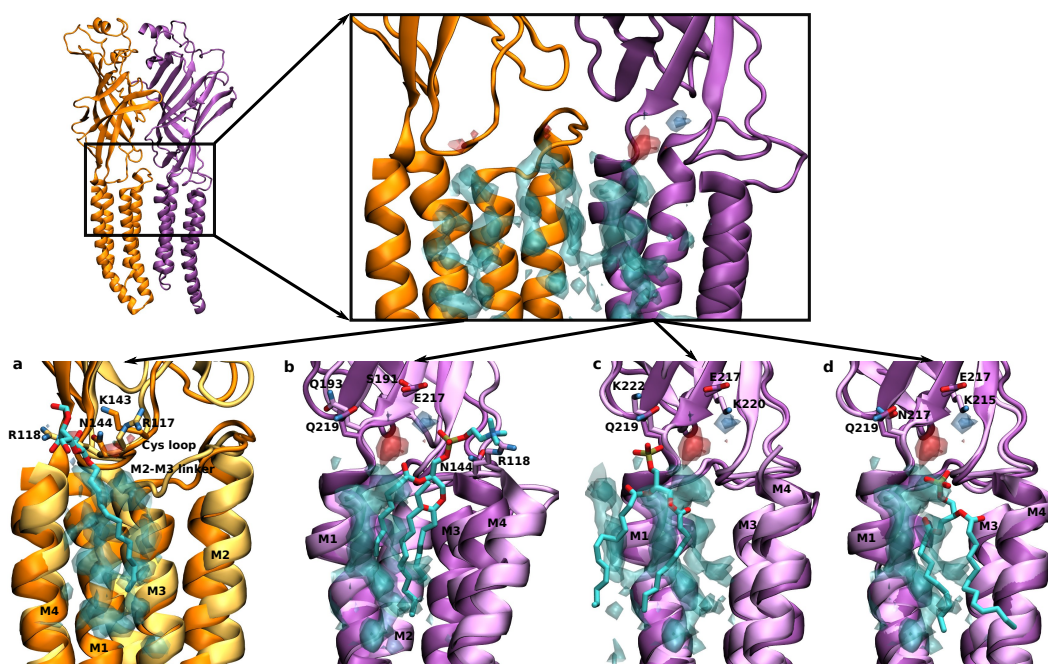


Figure 4.8: Overlay of phospholipid densities at the ECD-TMD interface with experimental structures. Overlays of the human $\alpha 1$ GlyR model (darker colours) and phospholipid density (transparent, phosphate headgroup density in red, choline/ammonium headgroup density in blue, tail density in cyan) from simulations in the active state with experimentally resolved structures of the pLGIC superfamily (lighter colours) with phospholipid or detergent bound. (a) Overlay with GLIC crystal structure (PDB: 6HZW [58]), which has a detergent molecule bound that interacts with its polar headgroup with the Cys loop, while its hydrophobic tail mainly interacts with the M3 helix. (b) Overlay with GLIC crystal structure (PDB: 6HZW [58]) which has a phosphocholine lipid bound that interacts with its polar headgroup with the pre-M1 region as well as the Cys loop, while its hydrophobic tails bridge the interface between the M4 and M1/M3 helices. (c) Overlay with the $\alpha 1$ subunit of the human GABA_AR cryo-EM structure (PDB: 6I53 [55]), which has a phospholipid bound that interacts with its polar headgroup with the pre-M1 region, while its hydrophobic tails interact mainly with the M1 helix. (d) Overlay with the $\alpha 3$ subunit of the human GABA_AR cryo-EM structure (PDB: 6I53 [55]), which has a phospholipid bound that interacts with its polar headgroup with the pre-M1 region, while its hydrophobic tails interact with both the M1 and M4 helices.

interface than in the human $\alpha 1$ GlyR, allowing the phosphocholine molecule to insert itself deeper in between the M4-M1/M3 helix interface while the phospholipid density in the simulations does not penetrate this interface but is rather located at the membrane facing surface of the M1 and M4 helices. The looser contact of the M4 helix with the M1/M3 helix interface in GLIC can be explained in terms of less aromatic residues that hold the M4 helix close to the M1/M3 helices in the human $\alpha 1$ GlyR and is covered in more detail in the discussion. Moreover, the phosphocholine insertion at this site in GLIC is further supported by an electrostatic interaction of the phosphate headgroup with R118 on the M2-M3 linker, whereas the human $\alpha 1$ GlyR has an asparagine at this position which cannot engage in such an interaction due to its much shorter side chain and lack of charge.

In the human GABA_AR cryo-EM structure (PDB: 6I53 [55]), two slightly different phospholipid binding modes are observed in this region. Figure 4.8c shows a phospholipid bound to the $\alpha 1$ subunit which is positioned slightly more towards the subunit interface, mainly interacting with the M1 helix via its lipid tails and its phosphate headgroup is stabilized via electrostatic interactions with K222 in the pre-M1 region. In Figure 4.8d an alternative mode of a phospholipid bound to the $\alpha 3$ subunit can be seen. The lipid tails are located on the membrane facing surface of the M1 and M4 helices and its phosphate headgroup can engage in electrostatic interactions with K215 in the pre-M1 region. The phospholipid modes in the human GABA_AR cryo-EM structure are similar to the phospholipid densities observed in the simulations of the active human $\alpha 1$ GlyR in the sense that the lipid tail positions are located on the membrane facing surfaces of the M1 and M4 helices, rather than penetrating the M4-M1/M3 interface as seen in the GLIC structure (Figure 4.8b). This can be explained due to a similar close connection of the M4 helix with the M1/M3 helices which is facilitated by a similarly large number of aromatic residues in the human GABA_AR as in the human $\alpha 1$ GlyR which I will expand on in the discussion.

In summary, phospholipids can interact with the ECD-TMD interface that is a crucial communication interface for signal transduction from the neurotransmitter

binding ECD towards the channel hosting TMD. For the inactive state, the phospholipid headgroups interact with regions slightly further towards the extracellular side of the ECD, especially the Cys loop, while in the active state stronger interactions with the pre-M1 region are observed. In the active state, three binding sites per subunit interface are distinguishable out of which two have previously been structurally characterized as phospholipid or detergent binding sites in the pLGIC superfamily, while one site at the subunit interface is a so far unreported new predicted site from simulations. Further work is required to determine the exact role of the phospholipids for receptor function. In atomistic simulations of the ECD alone, however, it is observed that the ECD-TMD interface tertiary structure unfolds in the absence of a lipid bilayer. Hence, one certain function of the lipids is to maintain the tertiary structure of the ECD-TMD interface intact, which is a crucial requirement for the signal transduction mechanism.

4.3.4 Cholesterol Can Bind to the Binding Site of the Positive Allosteric Modulator Ivermectin

Since CHOL has the longest interaction times of all lipid types with the receptor in the simulations and is experimentally known to be important for nAChR function, its interactions with the receptor at a key site for positive allosteric modulation by ivermectin are analyzed in more detail in this section. This binding site is located in a cleft at the interface of adjacent subunits in the extracellular leaflet half of the TMD. Ivermectin is wedged between the principal subunit and complementary subunit via the (+)M3 and (-)M1 helices and forms important contacts with the pore lining (+)M2 helix and the (+)M2-M3 linker. This site is not accessible in the inactive receptor state, but opens up in the active or desensitized states of the receptor. While CHOL preferentially interacts at the mouth of the ivermectin binding site in the inactive state, it only deeply penetrates it in the active state, as can be seen in Fig. 4.2-4.4. In the active state, CHOL density overlaps with the position of structurally resolved ivermectin and the density arising from the simulations can be assigned to two CHOL molecules.

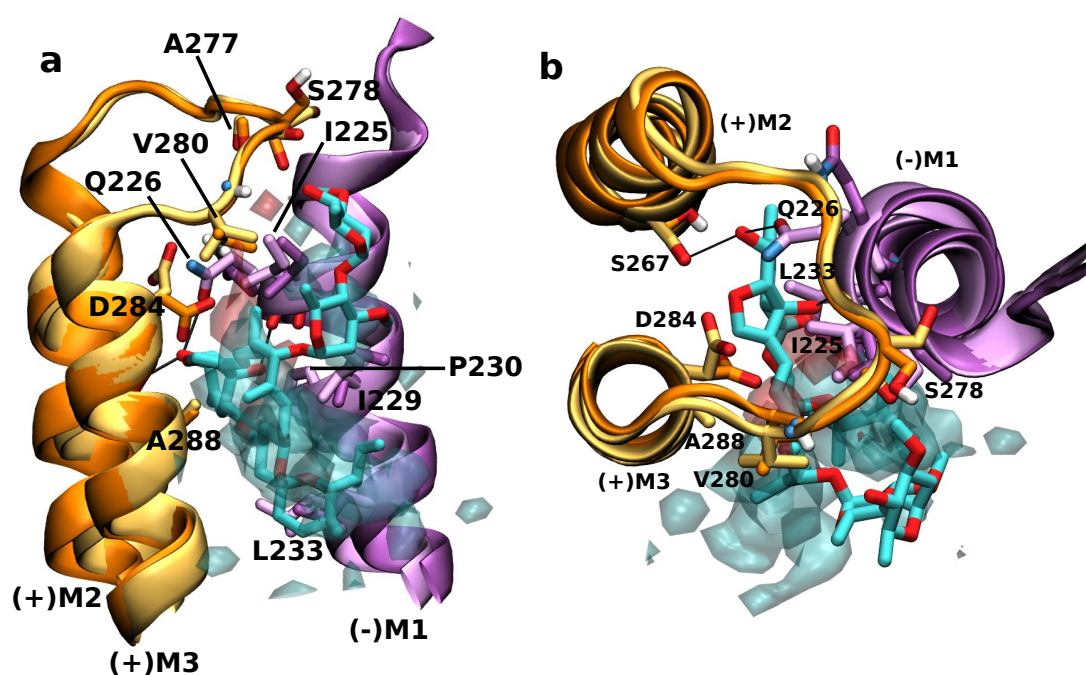


Figure 4.9: Overlay of CHOL density with positive allosteric modulator ivermectin at the subunit interface. Overlay of the human $\alpha 1$ GlyR model (darker colours) and CHOL density (transparent, hydroxy group density in red, remaining hydrophobic steroid skeleton in cyan) from simulations in the active state with the currently highest resolution crystal structure of human $\alpha 3$ GlyR (lighter colours) with ivermectin bound (PDB: 5VDH [123]). Shown is the ivermectin binding site at the interface of adjacent subunits with the (+)M2 and (+)M3 helices of the principal subunit (orange) and the (-)M1 helix of the complementary subunit (purple) forming a crevice and the (+)M2-M3 linker acting as a lid towards the extracellular side. Ivermectin and important residues for ivermectin or CHOL binding are shown explicitly and key interactions are highlighted by black lines. (a) Side view almost in parallel to the membrane and (b) view from the extracellular side towards the intracellular side.

Figure 4.9 shows an overlay of this region with CHOL density from the active state simulations and the positive allosteric modulator ivermectin resolved in the crystal structure of human $\alpha 3$ GlyR (PDB: 5VDH [123]). Ivermectin binding is driven by extensive hydrophobic interactions with the (+)M3 and (-)M1 helices as well as the (+)M2-M3 linker. More specifically, A288 from the (+)M3 helix, I225, I229, P230 and L233 from the (-)M1 helix as well as V280 from the (+)M2-M3 linker engage in various hydrophobic interactions with ivermectin. Likewise, these hydrophobic residues are in direct contact with the density arising from the hydrophobic CHOL steroid skeleton (coloured in transparent cyan). Yet, ivermectin can penetrate the cleft deeper than CHOL and form polar interactions via its benzofuran head. Specifically, via its C5 hydroxy group it forms a direct H-bond with S267 from the pore lining (+)M2 helix as well as with Q226 from the (-)M1 helix. The CHOL density overlays well with the spiroketal and disaccharide part of ivermectin. The density suggests that one CHOL molecule forms polar interactions via its hydroxy group (transparent red) with D284 from the (+)M3 helix, while a second CHOL molecule can form polar interactions via its hydroxy group with polar backbone atoms of the (+)M2-M3 linker, namely the backbone amine group of V280, the carbonyl group of S278 and the carbonyl group of A277. Interestingly, the polar atoms of ivermectin's disaccharide tail are also orientated towards this region of the (+)M2-M3 linker. The second CHOL molecule interacting with the (+)M2-M3 linker could thus mimic the disaccharide interactions from ivermectin. The importance of the hydrophobic residues lining the ivermectin binding site is backed up by mutational data [260], e.g. mutations that make the binding site narrower, such as A288F or P230W, abolish ivermectin agonist activity or lead to a significant decrease in ivermectin activation, respectively. This might also be the reason that ivermectin has no activity in the 5HT₃R, where A288 is replaced by the larger leucine. Interestingly, ivermectin is more potent in GluCl [261], where A288 is substituted by the smaller glycine residue, allowing ivermectin to penetrate ca. 1 Å deeper into the cleft as in the human $\alpha 3$ GlyR [123]. Atomistic MD simulations have previously also identified preferred CHOL binding to the ivermectin binding

site in the GABA_AR [255], however, on a sub-microsecond time scale the sampling of the highly dynamic interactions was much more limited.

In conclusion, CHOL can bind to the ivermectin binding site and mimic interactions similar to the spiroketal and disaccharide group of ivermectin via two separate CHOL molecules, but it does not penetrate the intrasubunit cleft as deeply as ivermectin which can interact with its benzofuran head directly with the pore lining M2 helix. The fact that ivermectin has higher potency in GluCl where it is buried ca. 1 Å deeper in the cleft as in the human $\alpha 3$ GlyR, suggests that potency is linked to penetration depth. This could explain why CHOL is indeed important for receptor function by supporting the expansion of the ion channel pore through moderate penetration of the ivermectin binding site, however, ivermectin itself is a more potent allosteric modulator, as it can bind deeper into the subunit interface where it makes contact with the pore lining M2 helix. Understanding the detailed atomistic interactions of CHOL in the ivermectin binding site could help in future efforts to rationally design new therapeutic agents for a wide range of neurological disorders such as chronic pain.

4.3.5 The Binding Mode of the Anionic Lipid DOPS in GlyR Differs from the PIP2 Binding Mode in the GABA_AR

Reconstitution experiments have shown that apart from CHOL, anionic lipids are important for receptor function in the nAChR [262]. Moreover, for the first time a recent cryo-EM structure of a desensitized human GABA_AR reconstituted in a lipid nanodisc has revealed the binding site of an anionic lipid, namely PIP2, to the pLGIC superfamily. In this section, the binding mode of the resolved PIP2 lipid is compared with density of the anionic lipid type DOPS from the simulations with the active GlyR structure, as the active state more closely resembles the desensitized GABA_AR.

Figure 4.10 shows the PIP2 binding site at one of the $\alpha 1$ subunits of the human $\alpha 1\beta 3\gamma 2$ GABA_AR (PDB: 6I53 [55]) overlaid with the density originating from DOPS from the active state simulations. The overall location of the anionic binding site at the intracellular end of the M4 and M3 helices is similar. However, the

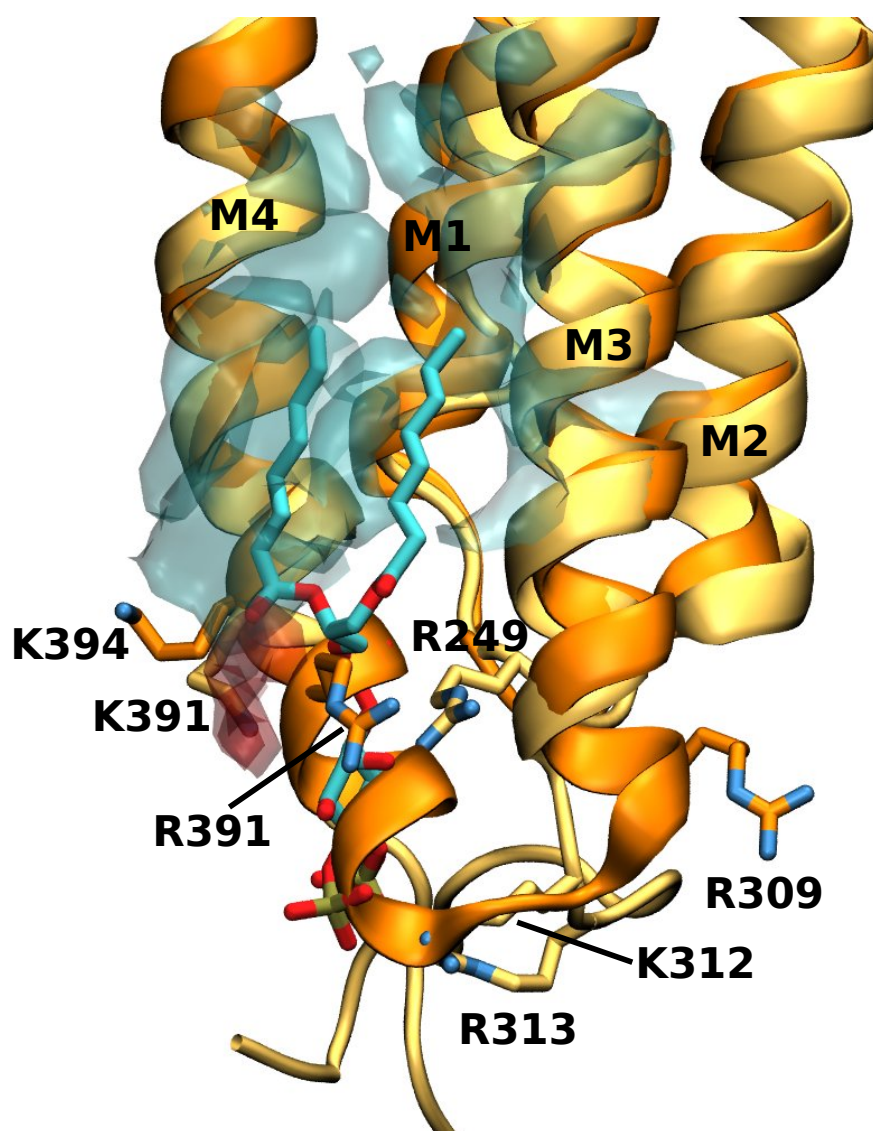


Figure 4.10: Overlay of DOPS density with PIP2 near the M4 helix. Overlay of the human $\alpha 1$ GlyR model (darker colours) and DOPS density (transparent, negatively charged headgroup density in red, lipid tail density in cyan) from simulations in the active state with the cryo-EM structure of human $\alpha 1\beta 3\gamma 2$ GABA_AR (PDB: 6I53 [55]) (lighter colours) with PIP2 bound. Shown is the PIP2 binding site at one of the $\alpha 1$ subunits. PIP2 and important residues for anionic lipid binding are shown explicitly.

binding modes differ between PIP2 and DOPS, which is not surprising due to the different headgroups of the anionic lipids as well as structural differences between the GABA_AR structure and the GlyR structure in this region. The most striking difference in the receptor structures is that the M4 helix of the GlyR structure maintains its helical secondary structure for two more turns towards the intracellular side than the M4 helix of the GABA_AR structure. As a consequence, this region sterically prohibits access of DOPS in the GlyR, which is precisely where the inositol-1,4,5-trisphosphate (InsP3) headgroup of PIP2 interacts extensively with the protein backbone of the GABA_AR. Moreover, the positively charged side chains of the GABA_AR (R249, K312, R313) that electrostatically stabilize the negatively charged InsP3 headgroup of PIP2 in this region are either not present in human α 1 GlyR or in orientations too far away to engage in such interactions (R249 and R309). In GlyR, the negatively charged phosphatidylserine headgroup of DOPS forms stable electrostatic interactions slightly further towards the extracellular direction of the M4 helix with the positively charged side chains R391 and K394. Furthermore, the InsP3 headgroup of PIP2 is overall larger and carries a three-fold negative charge, whereas the phosphoserine headgroup of DOPS is smaller and carries only a single negative charge, which is also likely to have an effect on the different binding locations for these headgroups. To sum up, human α 1 GlyR has a binding site for anionic lipids at the intracellular end of the M4 and M3 helices, which is also where PIP2 binding is observed in at the α 1 subunits of human α 1 β 3 γ 2 GABA_AR (PDB: 6I53 [55]), but the electrostatic stabilization of the negatively charged lipid headgroup occurs about two turns further in the extracellular direction of the M4 helix.

4.4 Discussion

The nAChR is the best studied member of the pLGIC superfamily with regards to protein-lipid interactions [119]. CHOL and anionic lipids have both been shown to be essential for nAChR function [238–240, 262]. nAChRs reconstituted in pure PC or PE membranes bind agonist, but do not gate open. The unresponsive state

of the nAChR has been called the uncoupled state. However, mechanistic insight into the molecular details that characterize the uncoupled vs. the responsive (or coupled) state has been elusive so far. Lipid dependency seems to be an important theme for the whole pLGIC superfamily, as also the function of GABA_ARs is dependent on CHOL levels [263].

Since current evidence suggests that lipids can modulate nAChR function via interactions with the M4 helix as well as the adjacent M1 and M3 helices, the so called M4 lipid-sensor model has emerged [245]. It hypothesizes that the uncoupled state is structurally characterized by the M4 helix being more distant to its neighbouring M1 and M3 helices and thus the extracellular post-M4 cannot form interactions with the Cys loop of the ECD. Lipids could influence the binding of the M4 helix to the adjacent M1/M3 surface and thus either favour the coupled or uncoupled state [264]. As aromatic residues lining the contact interface of the M4 and M1/M3 helices have been shown to be the main driving force of M4 helix binding to the M1/M3 surface during the folding of the GlyR [246], it has further been hypothesized that the varying number of these π - π interactions in pLGIC members is linked to the degree of dependence on the lipid environment [245]. The more aromatic residues support the M4-M1/M3 binding, the less dependent is a receptor on its lipid environment.

Figure 4.11 gives an overview of the varying degree of aromatic residues lining the M4-M1/M3 interface across representatives of the pLGIC superfamily. In the human $\alpha 1$ GlyR aromatic residues line the M4-M1/M3 interface (Figure 4.11a) in an equally distributed manner, leading to a tight binding of the M4 helix to the M1/M3 interface and thus, according to the M4 lipid-sensor model, suggesting little dependence on the lipid environment. A similarly aromatic rich interface is found in the $\beta 3$ GABA_AR (Figure 4.11b). The prokaryotic channel GLIC has only eight aromatic residues at the M4-M1/M3 interface with a striking gap at the extracellular half of the TMD (Figure 4.11c), indicating weaker binding of the upper half of the M4 helix. On the other end of the spectrum are the nAChR ($\alpha 4$ subunit shown in (Figure 4.11d)) and ELIC (Figure 4.11e) with only six and four aromatic residues, respectively, suggesting a weak binding of the M4 helix and the

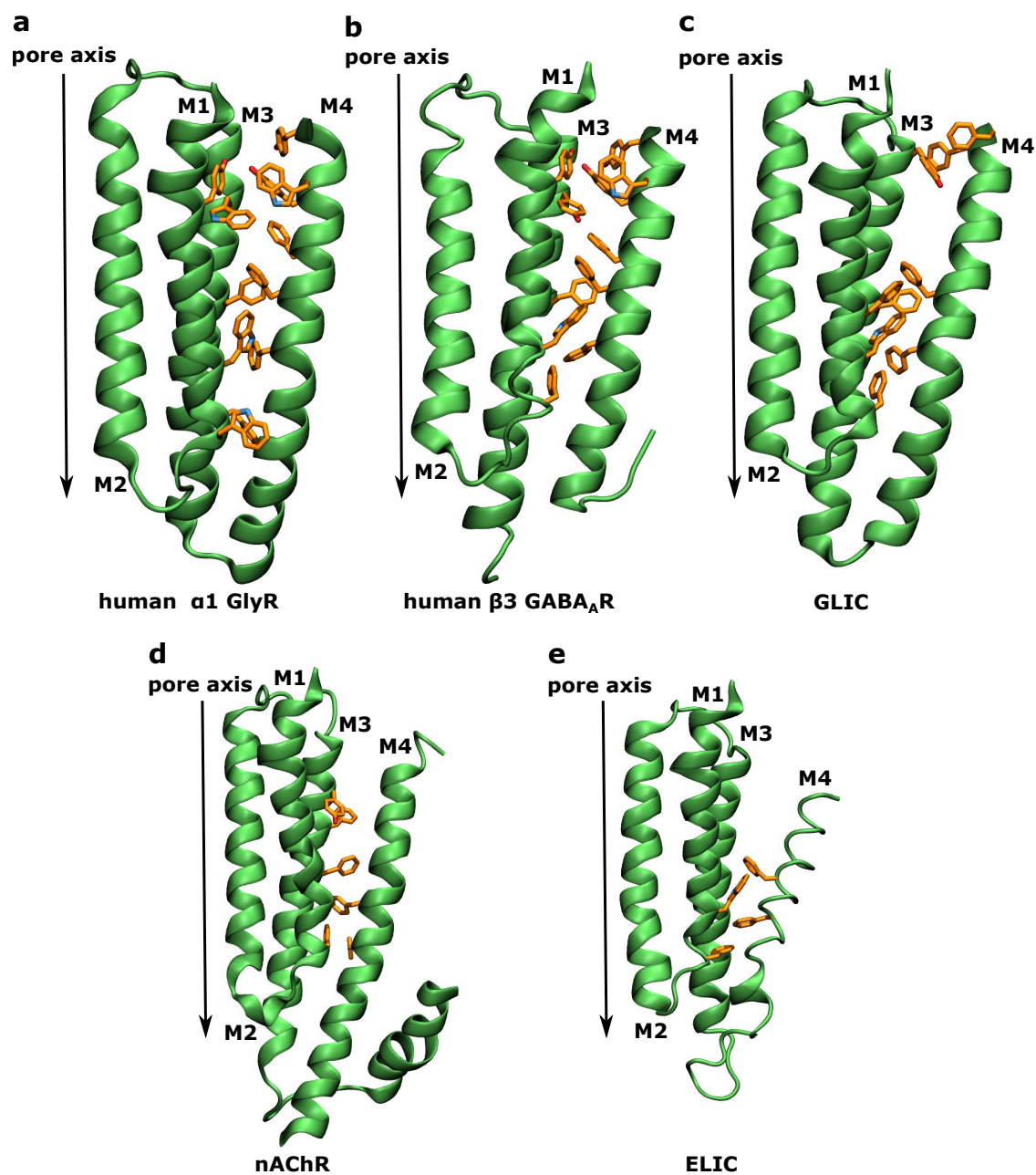


Figure 4.11: Aromatic residues at the M4-M1/M3 interface of different pLGICs. TMD region of a single subunit viewed in parallel to the membrane plane with aromatic residues at the M4-M1/M3 interface explicitly shown in orange for (a) human $\alpha 1$ GlyR (human homology model from PDB: 3JAE [68]), (b) human $\beta 3$ GABA_AR (from PDB: 6I53 [55]), (c) GLIC (from PDB: 6HZW [58]), (d) $\alpha 4$ human nAChR (from PDB: 5KXI [265]) and (e) ELIC (from PDB: 2VL0 [59]).

M4 lipid-sensor model would predict a strong dependence on the lipid environment. Indeed, the nAChR and ELIC have been demonstrated to be highly dependent on their lipid environment and do not undergo agonist-induced opening in minimal PC membranes, whereas GLIC remains functional [245]. While experimental data for GlyR and GABA_AR in this respect are not available, their larger number of aromatic contacts along the M4-M1/M3 interface than GLIC suggest that they remain functional in minimal PC membranes.

Interestingly, the analysis of aromatic contacts at the M4-M1/M3 interface explains the simulation result that phospholipids bind to the membrane facing surface of the M1 and M4 helices of GlyR (see Figure 4.8), rather than inserting in between the M4-M1/M3 interface. The plethora of aromatic contacts between the M4 helix and the M1/M3 surface binds the M4 helix strongly to the receptor in GlyR. This is also the reason why similar surface binding of phospholipids is observed in the aromatic-rich GABA_AR (see Figure 4.8c and 4.8d). For GLIC, however, the gap of aromatic residues in the extracellular half of the M4-M1/M3 interface, leads to a weaker binding of the M4 helix that allows phospholipids to insert themselves between the M4 and M1/M3 helices, as can be seen in the GLIC crystal structure with the M4 helix kinked away from the M1/M3 surface and a phospholipid inserted between this interface (see Figure 4.8b).

The protein-lipid interactions described here form the basis for further experimental studies to address their physiological functional consequences *in vivo*. From a pharmacological perspective, the detailed picture of protein-lipid interactions for the GlyR as a pLGIC representative forms the basis to guide future design principles for lipid-like drugs that can partition into the membrane and bind to allosteric sites within the TMD or even at the ECD-TMD interface. pLGIC dysfunction can lead to a range of neurological disorders such as Alzheimer's, Parkinson's or epilepsy which makes them pivotal drug targets [9, 10]. However, existing drugs suffer from side effects due to poor subtype selectivity [11], i.e. the same drug molecule does not only bind to its desired target receptor, but also to other receptor subtypes where it creates side effects. For rational drug design purposes, the neurotransmitter

binding sites (orthosteric sites) are the natural target. The problem is that they are highly conserved across receptor subtypes. Hence, the same drug molecule is likely to bind to the orthosteric site of the desired receptor subtype, but also to other receptor subtypes where it would provoke undesired side effects. Targeting allosteric (non-orthosteric) modulator sites has great potential to overcome this problem for the compelling reason that they exhibit more sequence diversity between receptor subtypes. Many allosteric sites are located at the lipid-accessible regions of the TMD. Hence, the detailed interaction map around the GlyR as a representative of the pLGIC class obtained in this study, particularly the CHOL interactions at the ivermectin binding site, forms the basis for the rational design of more selective, lipid-like drug molecules.

With the ongoing cryo-EM structural revolution, our structural knowledge of interacting lipids with pLGICs will undoubtedly increase with more high resolution structures being resolved that show electron density for lipid molecules. Since it is still difficult to identify the exact lipid type from the electron density data alone, MD simulations will be an important complementary technique in order to both identify the molecular identities of the lipid electron density as well as provide insight into the dynamic lipid interactions with the receptor [237].

4.5 Conclusions

For the first time, a representative pLGIC, namely the GlyR, was computationally reconstituted in a realistic model of a neuronal membrane and CG-MD simulations were performed to probe and characterize the molecular level detail of the protein-lipid interactions in two key functional receptor states (inactive and active). The simulations revealed that the lipid environment is dependent on the receptor state. The differences in the local residence times between the inactive and active state are a proxy for changing lipid binding affinities at their respective interaction sites that must occur during receptor activation. This suggests that the lipid environment itself might play an active role in regulating the activation of the receptor. Via analyzing the detailed molecular interactions, so far unreported

protein-lipid interactions at ECD-TMD interface could be characterized, suggesting that lipids might play a modulatory role at this communication interface in the activation process of the receptor. Two phospholipid interaction sites at the ECD-TMD interface coincide well with known phospholipid or detergent binding sites in GLIC and the GABA_AR and, in addition, a so far unreported third phospholipid binding site at the extracellular half the TMD at the subunit interface in the active state of GlyR is predicted. Furthermore, the simulations show that CHOL binds to the binding site of the positive allosteric modulator ivermectin, indicating that CHOL might have positive modulatory function. Finally, a binding site for the anionic lipid DOPS at the intracellular leaflet half of the TMD is identified. Located on the surface of the M4 and M3 helices, it roughly coincides with the location of PIP2 binding in the human GABA_AR, yet the details of the headgroup interactions are different, mainly due to differences in the receptor secondary structure in this region.

Thousands have lived without love, not one without water.

— Wystan H. Auden

5

Agonist Binding in the Glycine Receptor

Contents

5.1	Introduction	135
5.2	Methods	138
5.2.1	Standard MD Simulations	138
5.2.2	Free Energy Calculations	143
5.3	Results	145
5.3.1	Binding Modes of Agonists	145
5.3.2	Agonist Influence on Receptor Dynamics	160
5.3.3	Loop C Dynamics	161
5.3.4	Binding Free Energies	167
5.4	Discussion	171
5.5	Conclusions	175

The focus of this chapter is on the orthosteric binding site of the GlyR. Atomistic MD simulations were performed to characterize the binding modes of full and partial agonists and to analyze their influence on the receptor dynamics. Moreover, their binding free energies were determined via free energy calculations.

5.1 Introduction

The gating process of pLGICs is initiated via the binding of agonists to orthosteric binding sites at subunit interfaces of the ECD. For the GlyR, many different agonists with varying efficacy are known. To illustrate this, consider the following example.

The natural agonist glycine, when applied at a concentration that saturates the binding sites, can elicit a maximum probability of 96 % for the ion channel to be open. Taurine on the other hand, binds at the same binding pocket as glycine, yet can only yield a maximum open probability of 54 %, even at saturating concentrations [101]. Such agonists that produce a response below that of a full agonist are called partial agonists. However, our knowledge of how partial agonists stimulate differential responses is still rudimentary.

An understanding of the phenomenon of partial agonism could be exploited for rational drug design. Depending on the clinical situation, the ideal therapeutic drug should elicit a fine-tuned channel response somewhere in between that of a full agonist and a silent antagonist. Besides the GlyR, other members of the pLGIC superfamily are also potential targets for drugs with partial agonist properties, such as the nAChR for the smoking cessation drug varenicline [22].

A first step towards the understanding of partial agonism has emerged from single channel electrophysiological data, which led to the proposition of one or more short-lived intermediate shut states (flipped/primed) along the channel activation pathway [100, 102, 104]. In this model, the limited efficacy of partial agonists is explained by partial agonists being less efficient in promoting the transition to an intermediate shut state, while once the intermediate state is reached, their ability to make the channel enter the active state is similar. Agonist affinity is increased in the intermediate shut state relative to the resting state [101, 105, 107, 109], which suggests that the orthosteric binding site of the intermediate state is structurally more similar to the active state than to the resting state.

While providing a coarse explanation of partial agonism, this model gives no insight at the level of atomistic resolution, which is necessary for any rational drug design purposes. Here, I focus on the GlyR because it is an excellent model system for the pLGIC family due to a plethora of experimental and structural data in different conformational states. In a recent high resolution crystal structure of human $\alpha 3$ GlyR (PDB: 5TIN) [56], density for the natural agonist glycine was discernible and allowed for a detailed characterization of its binding mode in

agreement with predicted insights from MD simulations, including a stable water molecule in the binding pocket [67]. However, no structures of the GlyR with partial agonists bound have been resolved or simulated with MD simulations yet.

An important structural feature of the orthosteric binding site that has emerged for all pLGICs is loop C, which covers it like a lid towards the extracellular bulk phase. Structural data as well as tryptophan-fluorescence studies on other Cys-loop receptors show that in the apo state this loop is dynamically flexible, while ligand binding leads to loop C closing in on the binding site with confined flexibility (see the review by Nys et al. [63] and references therein), which is also observed in MD-simulations, e.g. [70–72].

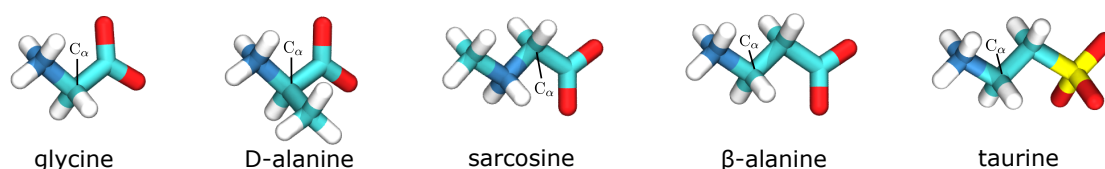


Figure 5.1: Agonist molecules of the GlyR investigated in this chapter. For the GlyR, glycine acts as a full agonist while all other ligands are partial agonists. There is no obvious correlation between the molecular properties of the agonists per se (such as size/shape or electrostatics) with the degree of full/partial agonism. Common features of all agonists are that they are zwitterionic with a positively charged ammonium end and a negatively charged carboxylate or sulfonate end. D-alanine differs from glycine by replacing its hydrogen side chain by a bulkier methyl side chain. Sarcosine derives from glycine via addition of a methyl group bound to the ammonium nitrogen. β -alanine is obtained from glycine by inserting a methylene group between the C_α carbon and carboxylate carbon. Taurine is related to β -alanine by exchanging the carboxylate group with a sulfonate group. For all agonists, the C_α atom is defined as indicated in this figure.

Considering the cost and effort associated with solving high resolution structures, computational methods provide a much cheaper and faster route to obtain structural insight into the binding modes of various agonists to the GlyR. The predictive power of MD simulations in this respect has been demonstrated by correctly characterizing the binding mode of glycine in the GlyR by Yu et al. [67] including a stable water molecule which was later confirmed by X-ray crystallography [56].

Here, I use MD simulations to study the behaviour of the full agonist glycine as well as the partial agonists D-alanine, sarcosine, β -alanine and taurine (see

Figure 5.1) in the orthosteric binding site of the human α_1 GlyR. This homopentamer was chosen as it allows for direct comparison with electrophysiological data from single channel recordings as well as extensive mutagenesis studies. The aim of this Chapter is to:

1. Structurally characterize the binding modes of various GlyR agonists.
2. analyze the agonists' effect on the receptor dynamics with a special focus on loop C.
3. Functionally annotate the residues in the orthosteric binding site and dissect their role for ligand binding as well as loop C dynamics.
4. Calculate the binding free energies of the different agonists with state of the art alchemical free energy calculations.

5.2 Methods

5.2.1 Standard MD Simulations

System Setup

Since the orthosteric binding sites are located at the subunit interfaces of the ECD, only this domain (residues P10-M220 per subunit) was extracted from the full length receptor homology model from Chapter 3 (human α_1 GlyR based on the glycine bound cryo-EM structure 3JAE). In order to avoid unnatural charges at the newly introduced artificial C-termini, they were capped with N-methyl (NME) groups. The TMD and the lipid bilayer were not included in the simulation system, as they do not play a role in agonist binding. This is a common approach for ligand-gated ion channels if the observables of interest are located in the orthosteric binding site [266–269]. Moreover, the AChBP is a naturally occurring homologue of the ECD of the nAChR and showcases that the ECD architecture of pLGICs can stably assemble in the absence of the TMD and retain its ligand binding function [270]. The advantage of this modelling choice is that the reduced system size allows for longer simulation times and thus a better sampling of relevant binding modes.

The structures for the ligands glycine, D-alanine, sarcosine, β -alanine and taurine were taken from the ZINC ligand database [217]. Zwitterionic glycine was positioned into all five orthosteric binding pockets as described in Chapter 3. The systems with the other ligands were obtained by aligning these ligands with the glycine backbone. All holo systems included a stable water molecule in the orthosteric site at the position resolved in PDB: 5TIN [56].

For each ligand type the ECD-ligand complexes as well as an apo ECD were then positioned into a simulation box of hexagonal prism shape, solvated with water and sodium and chloride ions were added to neutralize each system for a final physiological concentration of 150 mM. This yielded systems containing approximately 78,000 atoms and an example system with glycine bound is visualized in Figure 5.2 (a) and (b).

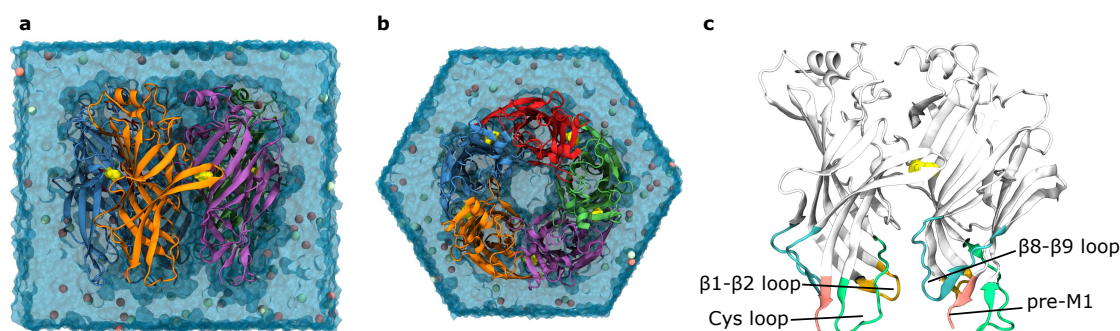


Figure 5.2: Simulation system and positional restraints. The simulation system contains the ECD of the GlyR in a hexagonal box of water with 150 mM NaCl viewed (a) in parallel to the membrane plane and (b) from the extracellular side towards the membrane. Each subunit is coloured differently and ligands are shown in yellow surface representation. (c) Positional restraints act on the $\beta 1$ - $\beta 2$ loop (gold), Cys loop (green), $\beta 8$ - $\beta 9$ loop (cyan) and the pre-M1 helix region (pink) in order to maintain the receptor's tertiary structure in this region in the absence of a TMD and a lipid bilayer. Agonist in the orthosteric binding site covered by loop C is shown in yellow surface representation. Only two subunits are shown for clarity, but these restraints were placed on all subunits.

Simulation Protocol

The AMBER99SB-ILDN force field [143] was used for protein and ions and the TIP3P model for water [170]. The ligands were parameterized using ACPYPE [271] which is a python wrapper script around ANTECHAMBER [272] in order to create

topologies and parameters in the general AMBER force field, version 2 (GAFF2) [168]. AM1-BCC charges [273] were used. In addition, the zwitterionic amino acids glycine and D-alanine were also simulated with parameters by Horn [169] which were optimized based on the AMBER99SB-ILDN force field which was used for the protein. Periodic boundary conditions in all three spatial dimensions were applied to mimic bulk conditions. All following steps were performed with GROMACS 2018 [141].

Each system was energy-minimized using the steepest descent algorithm until the maximum force fell below a tolerance of $1000 \text{ kJmol}^{-1}\text{nm}^{-1}$. The velocities were initiated according to a Maxwell-Boltzmann distribution at $T=310 \text{ K}$ and each system was equilibrated for 100 ps in the NVT-ensemble. All protein, ligand (if present) and binding pocket water (if present) heavy-atoms were position-restrained harmonically with a force constant of $1000 \text{ kJmol}^{-1}\text{nm}^{-2}$. This was followed by a series of equilibration phases in the NPT-ensemble to establish pressure at $P=1 \text{ bar}$ and to gradually relax the protein complexes. For 1 ns the same position restraints as in the NVT-equilibration were applied, followed by 5 ns in which only the protein backbone as well as the ligand's C_α atom (in holo systems) and the oxygen of the binding pocket water (in holo systems) were restrained harmonically with a force constant of $1024 \text{ kJmol}^{-1}\text{nm}^{-2}$ in order to relax the protein side chains and to give the ligands some orientational freedom. These restraints were then reduced in a step-wise fashion via halving them every 1 ns to a final value of $64 \text{ kJmol}^{-1}\text{nm}^{-2}$ in order to gradually relax the protein backbone as well as - in holo simulations - the ligand orientations and the water in the binding pocket.

Since in the absence of restraints a local loss of tertiary structure was observed in the loop regions of the ECD facing the TMD region, harmonic positional restraints ($1024 \text{ kJmol}^{-1}\text{nm}^{-2}$) were placed on these loop regions (see Figure 5.2 (c)) to maintain the structure close to the cryo-EM template containing a TMD. Specifically, the restraints acted on the backbone of residues A52-D57 of the $\beta 1$ - $\beta 2$ loop (gold), residues A137-Q150 of the Cys loop (green), residues A179-D194 of the $\beta 8$ - $\beta 9$ loop (cyan) and residues E217-M220 of the pre-M1 helix region (pink), where additionally a particularly high force constant of $10000 \text{ kJmol}^{-1}\text{nm}^{-2}$ was placed on the C_α

atom of C-terminal M220 to mimic the presence of the TMD. These loop restraints acted during the production runs as well as the equilibration phases, such that these regions were exempted from the restraint reduction in the equilibration phase described above. Due to these positional restraints in all simulation phases, the protein cannot undergo rotation or translation, such that a hexagonal prism could be chosen as the optimal simulation box type for the cylindrical ECD.

After the equilibration, production runs of 1 μ s length were carried out in the NPT-ensemble to mimic isothermal-isobaric physiological conditions. During production runs only the loop restraints described above were in place to preserve tertiary structure in the TMD-facing part of the ECD close to the cryo-EM template structure. Three independent repeats were run per system (glycine (Horn), glycine (GAFF2), D-alanine (Horn), D-alanine (GAFF2), sarcosine (GAFF2), β -alanine (GAFF2), taurine (GAFF2), apo), creating different dynamics by different initial velocity seeds after the energy minimization.

Most of the details for the calculation of the potential energy are based on the settings used to parameterize the AMBER99SB-ILDN force field [143]. The van der Waals interactions were cut off at 10 Å and dispersion correction was applied to energy and pressure to correct for this finite truncation. Electrostatic interactions were treated using the smooth Particle Mesh Ewald method [175, 176], where the real space contribution was cut off at 10 Å and the reciprocal energy term obtained in k-space was calculated on a grid with 1.2 Å spacing using 4th order B-splines for interpolation. The Verlet cut-off scheme was used to generate a pair-list with buffering using a tolerance of 0.005 kJ/mol/ps for pair-interactions per particle. With this scheme the cut-off distance for the short-range neighbour list is set automatically by mdrun as well as the update frequency of the neighbour list for which a minimum of 10 time steps was specified. Grid search was chosen as the algorithm to search for neighbours.

In all dynamics simulations, Hamilton's equations of motion were integrated numerically with the leap frog integration algorithm [222] which is a time-reversible and symplectic scheme that preserves the important symmetries of classical mechanics.

All protein bonds involving hydrogen were constrained with the LINCS algorithm [182] and water molecules were kept rigid via the SETTLE algorithm [184], such that a timestep of 2 fs could be employed (with the exception of the initial NVT-equilibration where a shorter timestep of 1 fs was used).

The Bussi thermostat [187] was coupled to the system to maintain the temperature constant at physiological 310 K. To ensure that weakly coupled parts of the system have the same temperature, in the apo systems, the protein was thermostatted separately from water and ions. In holo simulations, the protein together with ligands and waters in the binding pocket were coupled as one group to the thermostat, while the remaining system part (water and ions) was coupled separately. For all groups, the same coupling constant of 0.1 ps was used.

In all simulations with constant pressure, a barostat was coupled isotropically to the system to maintain a physiological pressure of 1 bar. The time constant used for pressure coupling was 1 ps and the isothermal compressibility set to $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. In the NPT-equilibration the Berendsen barostat [186] was used since it can establish a target pressure more rapidly than other schemes. However, just as the Berendsen thermostat, it fails to produce the correct fluctuations [127]. Hence, for production runs the Parinello-Rahmann barostat [194] was utilized, because it produces the correct fluctuations.

In order to avoid artificial translational drift, the centre of mass motion due to translation was removed every 100 time steps independently for the group of the protein (and ligand and binding pocket water if present) and the group of water with ions.

During production runs, the configurational part of phase space was written out every 10 ps.

Analysis

Trajectories were analyzed with MDAnalysis [226]. Visualization was performed in VMD [229].

5.2.2 Free Energy Calculations

System Setup

In order to calculate the binding free energies of a protein-ligand complex, containing only a single agonist molecule, the ECD-ligand complexes described in section 5.2.1 were further reduced to only contain two adjacent subunits and the ligand that is bound at the subunit interface. This protein-ligand complex was then inserted in a cuboidal box, solvated with water and neutralized via addition of NaCl to a physiological concentration of 150 mM, which resulted in total system sizes of ca. 44,000 atoms for the various ligand-receptor complexes.

For each ligand, such a system was energy minimized and equilibrated following the same protocol as described in section 5.2.1. In order to obtain an equilibrated orientation of ligand and side chain orientations in the orthosteric binding site, a 15 ns long production run followed this equilibration phase with the same settings as reported in section 5.2.1. For each ligand, the frame obtained after 15 ns production simulation was used as the input system for the protein-ligand leg of the alchemical cycle (right column of Figure 2.4).

For the leg of the alchemical cycle containing the ligand in solvent only, the ligand was centred in a dodecahedral box with a minimum distance of 12 Å to the box surface and the box was filled with water and NaCl at 150 mM. Depending on the ligand, this yielded system sizes between 1,533 and 1,860 atoms.

Simulation Protocol

All simulations for the binding free energy calculation were performed with GRO-MACS 2016.3 [141] as MD engine. The receptor-ligand leg of the alchemical cycle was subdivided into 52 discrete steps. First, the ligand restraints were added over 12 steps with non-uniformly distributed λ values (0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.75, 1.0). Then the electrostatic interactions, followed by the van der Waals interactions were turned off, each over 20 windows in a linear fashion using $\Delta\lambda = 0.05$. Taken together, this resulted in a total of 52 independent simulation windows for the receptor-ligand leg of the alchemical pathway.

For the ligand-in-solvent leg of the alchemical cycle, only the annihilation of the nonbonded interactions was explicitly simulated using the same $\Delta\lambda = 0.05$ spacing for both the Coulomb and the van der Waals interactions as employed for the receptor-ligand leg, while the free energy contribution from restraining the ligand was calculated analytically using Boresch’s formula [198]. This yielded a total of 41 independent simulation windows for the ligand-in-solvent leg of the alchemical cycle.

For each window from both legs of the cycle, the following protocol was employed. First, systems were energy minimized using the steepest descent algorithm for 10,000 steps and stopped before if the maximum force fell below $10 \text{ kJmol}^{-1}\text{nm}^{-1}$. Subsequently, Maxwell-Boltzmann-distributed particle velocities were assigned corresponding to a physiological temperature of 310 K and equilibrated in the canonical ensemble for 200 ps in the ligand-in-solvent leg or 100 ps for the ligand-receptor leg in the NVT-ensemble, followed by 1 ns NPT-equilibration to establish the pressure at 1 bar. During the energy minimization as well as all equilibration phases, harmonic position restraints were placed on the heavy atoms of the ligand (and protein if present) with a force constant of $1000 \text{ kJmol}^{-1}\text{nm}^{-2}$. After the equilibration, a production run was performed of 10 ns length with no restraints for windows of the ligand-in-solvent leg and 40 ns length with loop restraints as described in section 5.2.1 and visualized in Figure 5.2 (c) for windows of the ligand-receptor leg. For each ligand type, three independent repeats of both legs of the alchemical cycle were performed, i.e. 3×41 windows of ligand in solution simulations ($3 \times 41 \times 10 \text{ ns} = 1,230 \text{ ns}$ accumulated simulation time from production runs) and 3×52 windows of ligand-receptor simulations ($3 \times 52 \times 40 = 6,240 \text{ ns}$ accumulated simulation time from production runs).

The same force fields and simulation parameters as in section 5.2.1 have been used, except for the following alterations. Temperature was coupled using Langevin dynamics via the stochastic leap-frog integration scheme with a coupling constant of 2 ps. A soft-core potential for decoupling the van der Waals interactions was employed with an α value of 0.5 and power 1 for λ in the soft-core function. A 2 fs

time step was already used in the NVT-equilibration phase. $\frac{\partial H}{\partial \lambda}$ from which the changes in free energy are calculated was written out every 200 ps.

Analysis

The results in form of the $\frac{\partial H}{\partial \lambda}$ values obtained from the production runs of all windows of a leg of the alchemical cycle were analyzed with the alchemical analysis tool by Klimovich et al. [274]. Standard settings were used and the first 1 ns of the production run of each window was discarded as an additional equilibration period. The results reported in this Chapter are based on the multiple Bennet acceptance ratio (MBAR) which is currently the best and therefore recommended method to calculate free energy differences from $\frac{\partial H}{\partial \lambda}$ simulation data [203].

5.3 Results

5.3.1 Binding Modes of Agonists

For all ligands, several unbinding events occur at various binding pockets over the course of 1 μ s simulation time per trajectory, which are summarized in Table 5.1. Agonist unbinding has also been observed in two of three repeats of 150 ns MD simulations by Yu et al. [67] who constructed a model of human α_1 GlyR based on the template structure of the glutamate-gated chloride channel (GluCl) from *C. elegans* (42 % sequence identity). Single channel recordings of rat GlyR α_1 could not determine if more than three glycine molecules can actually bind to the receptor and it was found that three glycines are sufficient for the channel to reach its maximum open probability [98]. Hence, the observed unbinding events are consistent with both previous experimental and computational results.

The Binding Mode of the Natural Agonist Glycine

We begin the binding mode analysis with the natural agonist glycine. An atomistic-level of understanding of its interactions in the orthosteric binding site will form the cornerstone for the understanding of the interactions of all other partial agonists discussed in this Chapter. In 2017 Huang et al. [56] solved a crystal structure of the

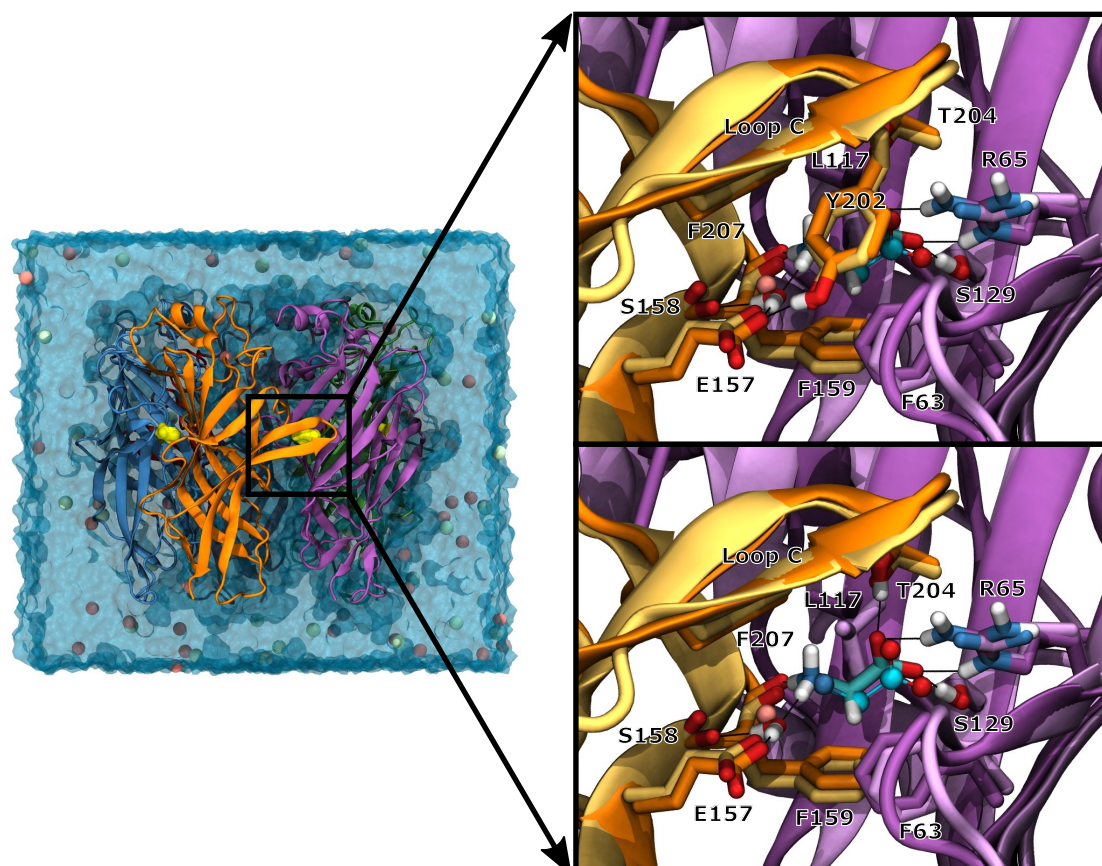


Figure 5.3: Binding mode of the natural agonist glycine in the orthosteric site of the GlyR. Overlay of the final frame of simulation repeat 1, Horn parameters, (darker colours) with the glycine-bound X-ray structure PDB: 5TIN [56] (lighter colours, the crystal structure glycine molecule is shown in ball-and-stick representation and the oxygen of a stable water molecule as a rose sphere). View from the membrane towards the extracellular matrix. H-bonds are indicated by black lines. For clarity, Y202 is not shown in the bottom panel.

Table 5.1: Overview of unbinding events at the different subunit interfaces (AB, BC, CD, DE and EA) for all holo simulations. The ✓ symbol indicates that the ligand is bound for the whole duration of the simulation, while the ✗ symbol denotes when the ligand dissociates from this binding site at some point of the production run.

ligand	AB	BC	CD	DE	EA	total unbinding events
glycine (Horn)	✓	✓	✓	✓	✓	0
	✓	✓	✓	✗	✓	1
	✓	✓	✓	✓	✓	0
glycine (GAFF2)	✗	✓	✓	✓	✗	2
	✗	✓	✓	✗	✓	2
	✗	✓	✗	✓	✗	3
D-alanine (Horn)	✓	✓	✗	✓	✓	1
	✓	✓	✓	✓	✓	0
	✓	✓	✓	✓	✓	0
D-alanine (GAFF2)	✗	✓	✓	✗	✗	3
	✓	✗	✓	✓	✓	1
	✗	✓	✗	✓	✓	2
sarcosine (GAFF2)	✗	✓	✓	✗	✓	2
	✗	✓	✓	✓	✓	1
	✗	✓	✓	✓	✗	2
β -alanine (GAFF2)	✓	✓	✓	✓	✗	1
	✗	✗	✓	✓	✗	3
	✓	✓	✗	✗	✗	3
taurine (GAFF2)	✓	✓	✓	✗	✓	1
	✓	✓	✓	✓	✓	0
	✓	✗	✓	✓	✓	1

GlyR (human $\alpha 3$ homopentamer) with the currently best resolution of 2.61 Å (PDB: 5TIN), such that density for glycine was discernible for the first time. Figure 5.3 shows an overlay of a representative binding mode obtained from simulations (last frame of repeat 1, Horn parameters) with this crystal structure. The structures were aligned on the C_α atoms of key residues in the orthosteric binding site (E157, S158, F159, Y202, T204 and F207 of the principal subunit (+) and F63, R65, L117 and S129 of the complementary subunit (-)) at the BC interface. The zwitterionic glycine molecule has two principal parts with which it interacts in the orthosteric binding site: its positively charged ammonium moiety which can also act as a H-bond donor and its negatively charged carboxylate group which also acts as H-bond acceptor. The ammonium moiety engages in three different ways with the principal subunit of the GlyR. Firstly, it forms a cation- π interaction with (+)F207 from loop C,

secondly it forms a H-bond with the backbone carbonyl oxygen of (+)F159 and thirdly, it forms an electrostatic interaction with (+)E157 which is bridged via a stable water molecule in the binding site. This water molecule itself is stabilized via an extensive hydrogen bond network: it donates one of its hydrogen atoms to form a H-bond with one of the carboxylate oxygens of (+)E157, it donates its other hydrogen atom to engage in a H-bond with the backbone carbonyl oxygen of (+)S158 and its oxygen atom accepts one of the hydrogen atoms of the ammonium group of glycine to form a third H-bond. Glycine's carboxylate group interacts with three residues: firstly, it forms a H-bond with (+)T204 from the tip of loop C, secondly it forms a saltbridge with (-)R65 and thirdly, it makes a H-bond with (-)S129. Note that the guanidinium group also acts as a bidentate H-bond donor, while the carboxylate-oxygens of glycine act as H-bond acceptors. The binding mode observed in the simulations is in excellent agreement with the binding mode in the crystal structure PDB: 5TIN, and even the presence of a stable water molecule is confirmed as well. Through the ligand-receptor interactions, the principal and complementary subunits are brought closer together, particularly the flexible loop C which covers the binding site like a lid is stabilized in a closed conformation by the presence of glycine. The latter is mainly facilitated by glycine's carboxylate group which forms both H-bonds with (+)T202 on the tip of loop C as well as (-)S129 of the complementary subunit as well as glycine's ammonium moiety forming a cation- π interaction with (+)F207 located on loop C.

All the residues discussed above have previously been identified as crucial for glycine binding in $\alpha 1$ GlyRs by experimental studies. Very early T204 has been identified by Vandenberg et al. [275] as important for glycine binding by site-directed mutagenesis. In 2005 Grudzinska et al. [276] demonstrated that alanine mutations of F63, R65, E157 and F207 increased EC_{50} values more than 1000-fold compared to the wild type. Interestingly, the R65K mutation also increases the EC_{50} by 2 orders of magnitude, which suggests that it is not only the positive charge of R65 that contributes to glycine binding, but also its ability to additionally engage in two H-bonds with glycine's carboxylate group as predicted by our simulations

and in agreement with the PDB: 5TIN binding mode (cf. Figure 5.3). Finally, in 2014, S129 was predicted to be involved in glycine binding from MD simulations of a homology model of human $\alpha 1$ GlyR by Yu et al. which was experimentally confirmed via S129A having a 9-fold increase in EC_{50} value in the same paper [67]. Moreover, this publication was the first computational study to predict a stable water molecule in the binding site which was later structurally resolved in the crystal structure PDB: 5TIN. A mutagenesis study with unnatural amino acids by Pless et al. [277] concluded that only F159 engages in a cation- π interaction, while F207 and F63 do not. In contrast to this, the PDB: 5TIN structure clearly reveals a cation- π interaction with F207 in agreement with our simulation data (cf. Figure 5.3). However, the results by Pless could also be interpreted in a different way. For example, the EC_{50} values in fact do increase for fluorinated F207 (ca. 500-fold for F2-F207 and more than 100-fold for F3-F207), indicating weaker binding due to a reduced cation- π interaction, yet the authors argue against a cation- π interaction because they do not see the same apparent correlation as found for F159. Another possibility, namely that the artificially introduced fluorine atoms to the phenyl ring might yield steric clashes which would cause a conformational reorganization that could explain a change in EC_{50} values, is not discussed at all.

Apart from these principal interactions with the ligand described above, residues in the orthosteric site also play an important role in ligand binding via shielding the inner pocket volume from the surrounding bulk water and thus prevent ligand dissociation. The hydrophobic residues (+)F207, (-)L117, (-)L127 together with the hydrophobic methylgroup of (+)T204 form a hydrophobic shield towards the extracellular matrix direction. Towards the membrane direction the aromatic residues (-)F63, (+)F159 and (+)Y202 (and less direct (+)F99 and (-)F44) hydrophobically shield the orthosteric site from bulk water. The aromatic residues (-)F63, (+)F159 and (+)Y202 also have a stabilizing effect on the ammonium moiety via cation- π interactions, although they are weaker due to the larger distance compared to (+)F207. Note how (+)Y202 is anchored in this shielding position via a H-bond between its hydroxy hydrogen and an oxygen of (+)E157. We will see in section 5.3.3

that this H-bond can break and (+)Y202 can rotate into a different conformation that opens up loop C. The importance of the hydroxy group of Y202 which can engage in the E157-Y202 H-bond is further supported by experimental mutagenesis data. Y202F and Y202L both lead to similar reductions in glycine affinity [278], which suggests that rather than the aromatic character, it is the ability of residue 202 to form a H-bond with E157, thus stabilising a closed loop C conformation, which is crucial for glycine binding.

It is remarkable how the orthosteric binding site of the GlyR has been evolutionarily optimized to accommodate a zwitterionic glycine molecule together with a water molecule via electrostatic interactions and H-bonds and at the same time shielding the ligand from bulk water via hydrophobic residues.

Comparison with Binding Modes of Partial Agonists

How does the binding mode of the full agonist glycine compare with the binding modes of partial agonists? To this end, a density analysis of key functional ligand atoms was performed and averaged over all three simulation repeats per system using frames separated by 100 ps. Since the least unbinding events occur at the BC interface (cf. Table 5.1, this interface was chosen for detailed analysis and is depicted in Figure 5.4. It reveals the preferred location of the ammonium nitrogen (blue), the carboxylate/sulfonium oxygens (red) and water oxygen (rose) in 3d space. Moreover, the ligand-specific important carbon of D-alanine's methyl side chain as well as the N-methyl carbon of sarcosine are visualized in cyan. Common features of all agonists are the orientation of the negatively charged carboxylate/sulfonium group facing the positively charged (-)R65 residue of the complementary subunit while the positively charged ammonium moiety is oriented towards the negatively charged (+)E157 residue of the principal subunit and this latter interaction is bridged by a water molecule.

While these agonist-specific densities provide an ensemble-averaged representation of the agonists' position, they cannot capture the dynamics of side chains in the orthosteric binding site, which flexibly adapt in response to the interactions with

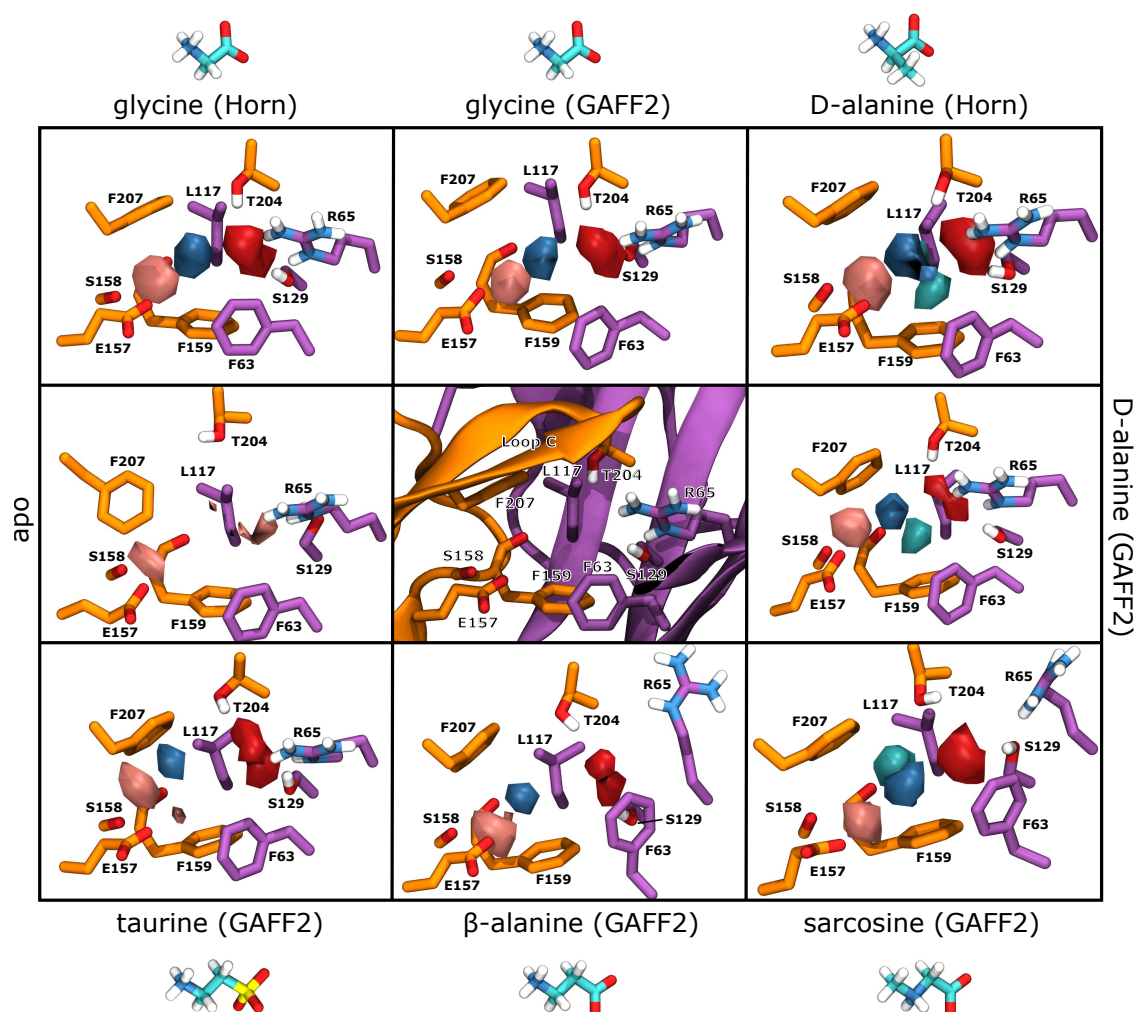


Figure 5.4: Agonist and water densities in the orthosteric binding site of the GlyR. View of the BC interface from the membrane towards the extracellular matrix. The densities arising from the agonists' nitrogen atom is coloured in blue, while density corresponding to the agonists' oxygen atoms is coloured in red. Density from water oxygens is visualized in rose. For D-alanine, density from the methyl side chain carbon is additionally viewed in cyan and for sarcosine, density from the N-methyl carbon is also additionally shown in cyan. All densities were averaged over three independent repeats. The receptor structures shown correspond to the final frame of simulation repeat 1 of each system. Common main features of the holo simulations are the agonist orientation, facing the principal subunit with its positively charged ammonium moiety and the complementary subunit with its negatively charged carboxylate or sulfonium group. In all cases, the interaction of the positive ammonium end with the negatively charged carboxylate group of E157 is bridged via a water molecule. This water binding site is also occupied in the simulations of the apo state. The density of the N-methyl carbon of sarcosine suggests that it dampens the cation- π interaction with F207. For D-alanine the overlap of densities of the ammonium nitrogen and the side chain methyl carbon suggests that these atoms can interchange places via rotation around the C_{α} - $C_{\text{carboxylate}}$ bond and thus the cation- π interaction with F207 is weaker on average.

the agonist molecule. This is why, additionally, key distances between the agonist and the residues in the binding pocket were analyzed as well. These distances were chosen based on the interactions observed for the natural agonist glycine both from structural and simulation data as discussed in section 5.3.1. In all cases non-hydrogen atoms were chosen. In order to filter out noise from rotation of the ligands' carboxylate or sulfonium group, the corresponding C or S atom were chosen and similarly to filter out noise by residue rotations, the centre of geometry (COG) of the phenyl ring of aromatic side chains, the central C of the guanidinium group of R65 and the C $_{\gamma}$ of L117 were chosen as anchor points for distance measurement. In total, 11 distances were monitored and are summarized in Table 5.2.

The corresponding time series data with frames separated by 1 ns for all three simulation repeats per system and for all 5 binding pockets was filtered to contain distance data only if a ligand was bound to a binding pocket. A ligand is considered not bound in the case that the distance between the ligand's C $_{\alpha}$ atom and the C $_{\alpha}$ of any of key residues in the orthosteric binding site was larger than a cutoff of 12 Å. Otherwise it is considered to be bound. From this filtered data, probability densities for each of the key distances were estimated using kernel density estimation and a representative selection is shown in Figure 5.5. The probability densities give us an idea about the relative occupation of different binding modes as defined by these distances and thus, their overall importance. Each local maximum corresponds to a local minimum on the free energy surface. It would be beyond the scope of this Chapter to discuss the structures corresponding to all energy minima for all agonists. The point of this analysis is to substantiate the main binding modes discussed here with a quantitative analysis and further illustrate that beyond these main modes, many more metastable binding modes exist. However, due to their low occupancy, their overall contribution is probably negligible compared to the main modes discussed here. In the following, the binding modes for each ligand will be discussed in more detail.

First, I briefly inspect the glycine-specific densities in order to provide a posthoc justification as to why the binding mode presented in 5.3.1 can be considered as

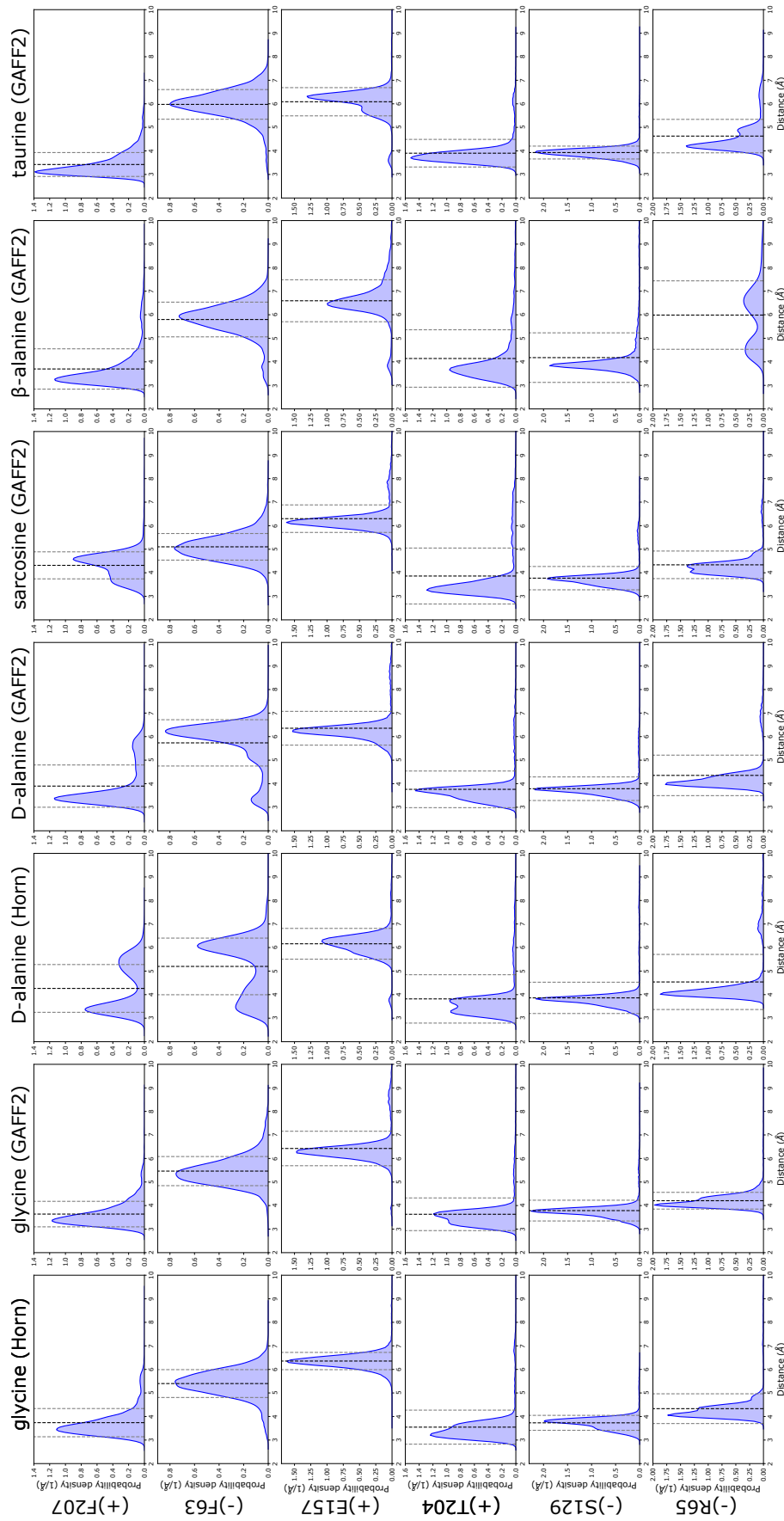


Figure 5.5: Probability densities of selected key distances in the orthosteric binding site as a finger print for binding modes of several GlyR agonists. First row: (+)F207 phenyl COG - ammonium N distance, second row: (-)F63 phenyl COG - ammonium N distance, third row: (+)E157 carboxylate C - ammonium N distance, fourth row: (+)T204 hydroxy O - carboxylate C/ sulfonium S distance, fifth row: (-)S129 hydroxy O - carboxylate C/ sulfonium S distance, sixth row: (-)R65 guanidinium C - carboxylate C/ sulfonium S distance. The mean value and one standard deviation from the mean are indicated via dashed black and grey lines, respectively.

Table 5.2: Key distances to characterize the binding mode in the orthosteric binding site.

interaction	protein residue anchor point	agonist anchor point
electrostatic interaction between the negatively charged carboxylate group of (+)E157 and the positively charged ammonium group of the ligand	(+)E157 carboxylate C	ammonium N
(water mediated) H-bond between the backbone carbonyl oxygen of (+)S158 and one of the ammonium hydrogens of the ligand	(+)S158 carbonyl O	ammonium N
cation- π interaction between the centre of geometry (COG) of (+)F159's phenyl ring and the positively charged ammonium group of the ligand	(+)F159 phenyl COG	ammonium N
H-bond between the backbone carbonyl oxygen of (+)F159 and one of the ammonium hydrogens of the ligand	(+)F159 carbonyl O	ammonium N
cation- π interaction between the centre of geometry (COG) of (+)Y202's phenyl ring and the positively charged ammonium group of the ligand	(+)Y202 phenyl COG	ammonium N
H-bond between the hydroxy hydrogen of (+)T204 and one of the oxygens of the carboxylate/sulfonium group of the ligand	(+)T204 hydroxy O	carboxylate C/ sulfonium S
cation- π interaction between the centre of geometry (COG) of (+)F207's phenyl ring and the positively charged ammonium group of the ligand	(+)F207 phenyl COG	ammonium N
cation- π interaction between the centre of geometry (COG) of (-)F63's phenyl ring and the positively charged ammonium group of the ligand	(-)F63 phenyl COG	ammonium N
electrostatic interaction between the positively charged guanidinium group of (-)R65 and the negatively charged carboxylate/sulfonium group of the ligand	(-)R65 guanidinium C	carboxylate C/ sulfonium S
hydrophobic interaction between hydrophobic side chain of (-)L117 and the C $_{\alpha}$ atom of the ligand	(-)L117 C $_{\gamma}$	C $_{\alpha}$
H-bond between the hydroxy hydrogen of (-)S129 and one of the oxygens of the carboxylate/sulfonium group of the ligand	(-)S129 hydroxy O	carboxylate C/ sulfonium S

representative. Clearly, the density for carboxylate oxygens suggest a H-bond network with (+)T204, (-)S129 and (-)R65, while the ammonium nitrogen density indicates a cation- π interaction with (+)F207, a H-bond to the backbone carbonyl oxygen of (+)F159 and an electrostatic interaction with (+)E157 bridged via water density, as captured by the representative snapshot described in Figure 5.3. Moreover, the densities obtained using the Horn parameters or the GAFF2 parameters for zwitterionic glycine are very similar. Furthermore, the mean values of the probability density distributions corresponding to these key interactions are close to the optimal values for these interactions (cf. Figure 5.5). The cation- π interaction between the ammonium group and (+)F207 has its average close to the optimal value of 3.75 Å [279]. The average (-)R65-carboxylate saltbridge is at its optimal distance between guanidinium carbon and carboxylate carbon of 4.3 Å [280].

The Binding Mode of D-alanine

The respective particle densities for D-alanine look similar to those of glycine. However, D-alanine has an additional methyl side chain which can be directed towards (-)F63, but it can also face (+)F207. The methyl side chain density overlaps with the ammonium nitrogen density, pointing to the fact that these groups can swap positions via rotation around the C_{α} - $C_{\text{carboxylate}}$ bond. Thus, it follows that averaged over time, the cation- π interaction with (+)F207 would be weaker compared to glycine due to this rotation. The two conformations of the ammonium moiety either facing (+)F207 or (-)F63 are further substantiated by the corresponding bimodal probability densities in Figure 5.5. Representative snapshots illustrating the two methyl side chain conformations are shown in Figure 5.6.

These modes reveal that the H-bond between (+)Y202 and (+)E157 is stretched when D-alanine's methyl side chain points towards (+)F159. The weakened H-bond is thus less effective in contributing towards a closed loop C conformation. This, together with the on average weakened cation- π interaction with (+)F207 shows that D-alanine is less effective in contributing to a closed loop C conformation compared to the natural agonist glycine. Hence, with D-alanine the free energy barrier for loop

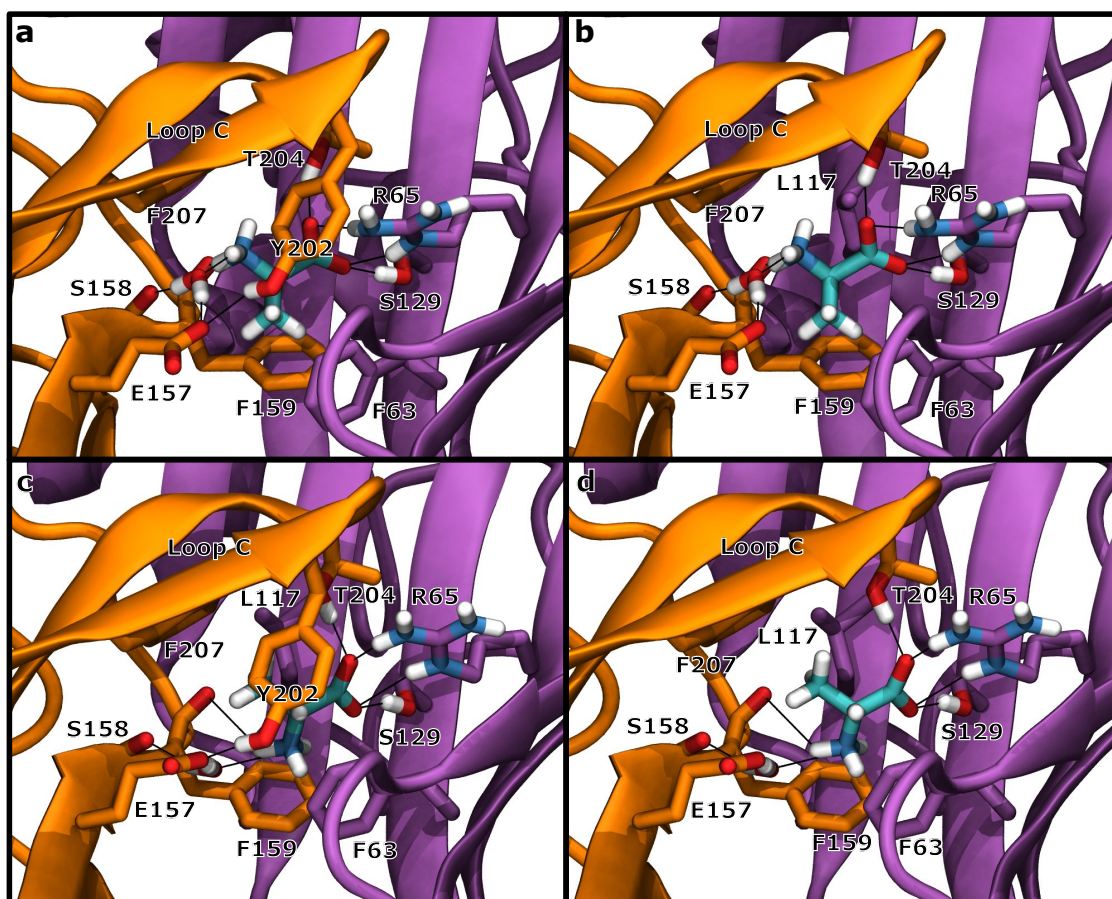


Figure 5.6: Binding modes of the partial agonist D-alanine in the orthosteric site of the GlyR. View from the membrane towards the extracellular matrix. H-bonds are indicated by black lines. The ammonium group and methyl side chain can interchange their position via rotation around the C_{α} - $C_{\text{carboxylate}}$ bond. The top row shows snapshots with the ammonium group pointing towards (+)F207 and the methyl side chain facing (+)F159, (a) with (+)Y202 and (b) without (+)Y202 shown. The bottom row shows the alternative conformation with these groups interchanged, (c) with (+)Y202 and (d) without (+)Y202. Note how in the top row conformation the H-bond between (+)Y202 and (+)E157 is elongated compared to the bottom row conformation as well as compared to the canonical binding mode of glycine.

C to spontaneously open up is lower which makes it easier for D-alanine to unbind compared to glycine. This is in line with kinetic estimates from single channel recordings which posit that the main difference between glycine and D-alanine affinity is due to a faster dissociation rate of D-alanine, while association rates are similar ([100] for glycine and unpublished data by the group of Lucia Sivilotti for D-alanine).

The Binding Mode of Sarcosine

Sarcosine differs from glycine by an additional methyl group bound to the ammonium nitrogen. Its density in Figure 5.4 reveals that while the carboxylate end forms the canonical interactions with (+)T204, (-)S129 and (-)R65 and the ammonium moiety still interacts electrostatically with (+)E157 bridged via a stable water molecule, the cation- π interaction with (+)F207 is dampened by precisely this additional N-methyl group. This also manifests itself in the (+)F207-N distance probability distribution which for sarcosine is shifted towards longer distances compared to glycine (see Figure 5.5). Figure 5.7 depicts a representative snapshot of the binding mode. Hence, sarcosine destabilizes the closed loop C conformation via blocking the cation- π interaction with (+)F207.

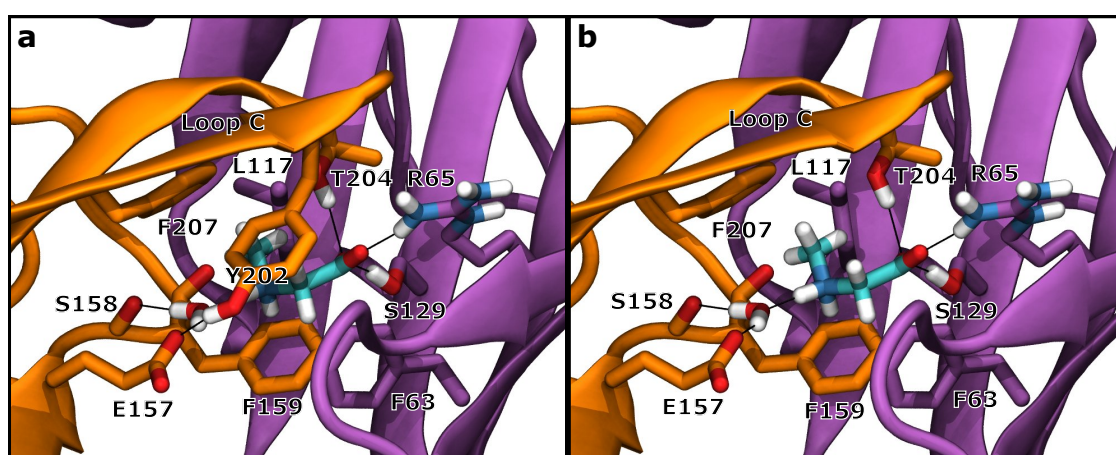


Figure 5.7: Binding mode of the partial agonist sarcosine in the orthosteric site of the GlyR. View from the membrane towards the extracellular matrix. Y202 is not shown in (b). H-bonds are indicated by black lines. The key difference to the glycine mode is that sarcosine's N-methyl group weakens the cation- π interaction with (+)F207.

The Binding Mode of β -alanine

The densities for β -alanine are similar to the ones of glycine, suggesting similar interactions. However, β -alanine has an additional methylene group inserted between the ammonium and carboxylate ends which makes it slightly longer. In order to accommodate this slightly longer molecule, one could imagine that either the ammonium end occupies the position of the stable water molecule in the binding site or that the overall position of the ligand would be shifted further towards the complementary subunit. While the former is occasionally the case as indicated by the small local maximum in the probability density corresponding to the E157 carboxylate C - N distance (see Figure 5.5), what is dominantly observed in the simulations is the latter. This leads to the position of the carboxylate end being shifted slightly further towards the complementary subunit compared to glycine. This in turn leads to a less optimal positioning for hydrogen bonding to (+)T204 and (-)S129 as can be seen from the probability densities for the respective distances which are shifted towards longer H-bond distances, cf. Figure 5.5. Most interestingly, this strongly affects the saltbridge with (-)R65. The underlying probability density is bimodal, with one local maximum near the optimal distance of 4.3 Å between the guanididinium carbon and the carboxylate carbon [280], while the other local maximum is at a much longer distance which would correspond to a broken saltbridge. Examples of binding modes with a broken and intact saltbridge are depicted in the top and bottom row of Figure 5.8, respectively.

The Binding Mode of Taurine

Taurine is similar to β -alanine, as it also has an additional methylene group compared to glycine between its charged ends. The difference to β -alanine is that the negatively charged carboxylate group is replaced by the larger negatively charged sulfonium group. As expected from this molecular similarity, a similar density distribution is observed as for β -alanine with the negative sulfonium group being positioned further towards the complementary subunit compared to glycine's carboxylate group position. Similarly as for β -alanine, this leads to weaker hydrogen

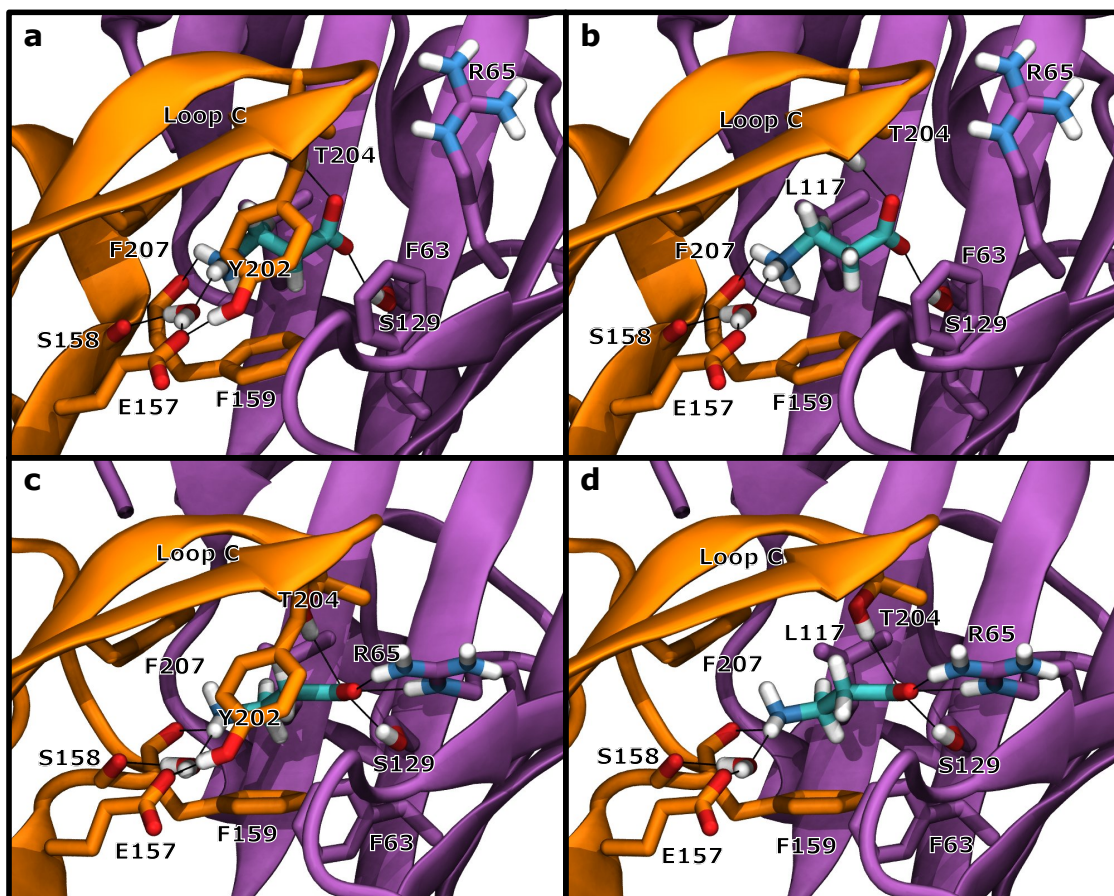


Figure 5.8: Binding mode of the partial agonist β -alanine in the orthosteric site of the GlyR. View from the membrane towards the extracellular matrix. H-bonds are indicated by black lines. Compared to glycine, the slightly longer β -alanine molecule is positioned slightly further towards the complementary subunit which leads to weaker hydrogen bonding to (+)T204 and (-)S129 and two similarly stable conformations with the saltbridge to (-)R65 broken (top row) or formed (bottom row). (b) and (d) are the same snapshots as (a) and (c), respectively, just with Y202 not shown.

bonding to (+)T204 and (-)S129 as suggested by a shift to longer distances in the corresponding distance densities (see Figure 5.5). The saltbridge with (-)R65, although weakened compared to glycine, is intact most of the time (cf. Figure 5.5). Similar to β -alanine, the long taurine molecule can also replace the water in the binding pocket by its ammonium group as indicated by the small local maximum between 3 and 4 Å of the E157 carboxylate C - N distance (see Figure 5.5). However, the relative amount of time spent in this ‘dry’ mode is small, just as for β -alanine. A representative binding mode can be seen in Figure 5.9. Note how the bulkier sulfonium group wedges loop C into a more open conformation in comparison to glycine. In the simulations it is observed that the sulfonium group rotates more frequently than the carboxylate group of other agonists. This provides an additional explanation as to why the interactions at the negatively charged end of taurine are slightly weaker. Taurine thus destabilizes a fully closed loop C via its bulkier and dynamically more flexible sulfonium group.

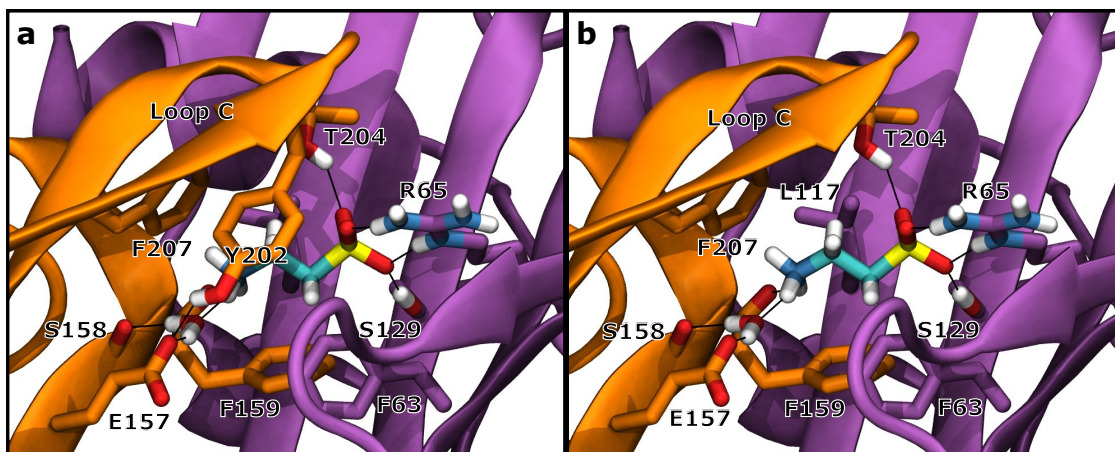


Figure 5.9: Binding mode of the partial agonist taurine in the orthosteric site of the GlyR. View from the membrane towards the extracellular matrix. H-bonds are indicated by black lines. Taurine wedges loop C into a slightly more open conformation via its bulky sulfonium group.

5.3.2 Agonist Influence on Receptor Dynamics

In order to investigate the agonists' influence on the overall receptor dynamics in the ECD, in this section I will analyze the root-mean-square fluctuation (RMSF)

of the C_α atoms, which is a measure for the degree of local dynamic flexibility. Since in the holo simulations occasional unbinding occurs for various agonists at different subunit interfaces during the course of the simulations, only simulation frames in which all five orthosteric sites have a ligand bound are included in the analysis (an equal frame spacing of 1 ns was used). The obtained data sets are then averaged over all three repeats and all five identical subunits and the standard deviation is calculated from this total of fifteen subunit C_α -RMSF profiles. As can be seen from Figure 5.10, overall, the dynamic flexibility of the ECD is quite similar for the apo as well as the holo systems. Naturally, loop regions as well as the unrestrained N-termini display a higher degree of conformational flexibility, while secondary structure elements (mostly β -sheets in the ECD) are more stable. Note that the C-termini as well as flexible loops at the ECD-TMD interface have been artificially restrained in the simulations to prevent a loss of tertiary structure due to the absence of the TMD and membrane lipids in this region and, thus, exhibit a very low C_α -RMSF, which is a simulation artefact resulting from the compromise of only simulating the ECD of the GlyR. The simulations are therefore not suitable to provide insights into the receptor dynamics in the restrained loop regions of the ECD-TMD interface. While overall, the RMSF behaviour among both the apo and the holo systems is quite similar, the loop C region (residues 195-216) is a striking exception. Hence, the most important insight from this analysis is the agonists' ability to decrease the flexibility of loop C which covers the orthosteric binding site like a lid. Thus, it can be concluded that the effect of agonists bound in the orthosteric site has a localized effect on the receptor dynamics which reduces loop C flexibility. This will be explored in more detail in the next section.

5.3.3 Loop C Dynamics

Loop C covers the orthosteric binding site in a lid-like fashion. It is implicated in regulating the ligand accessibility in pLGICs and its closure upon ligand binding is thought to be the first step in the gating mechanism [63]. In the previous section, it has emerged as a key region whose dynamics is influenced by the presence of

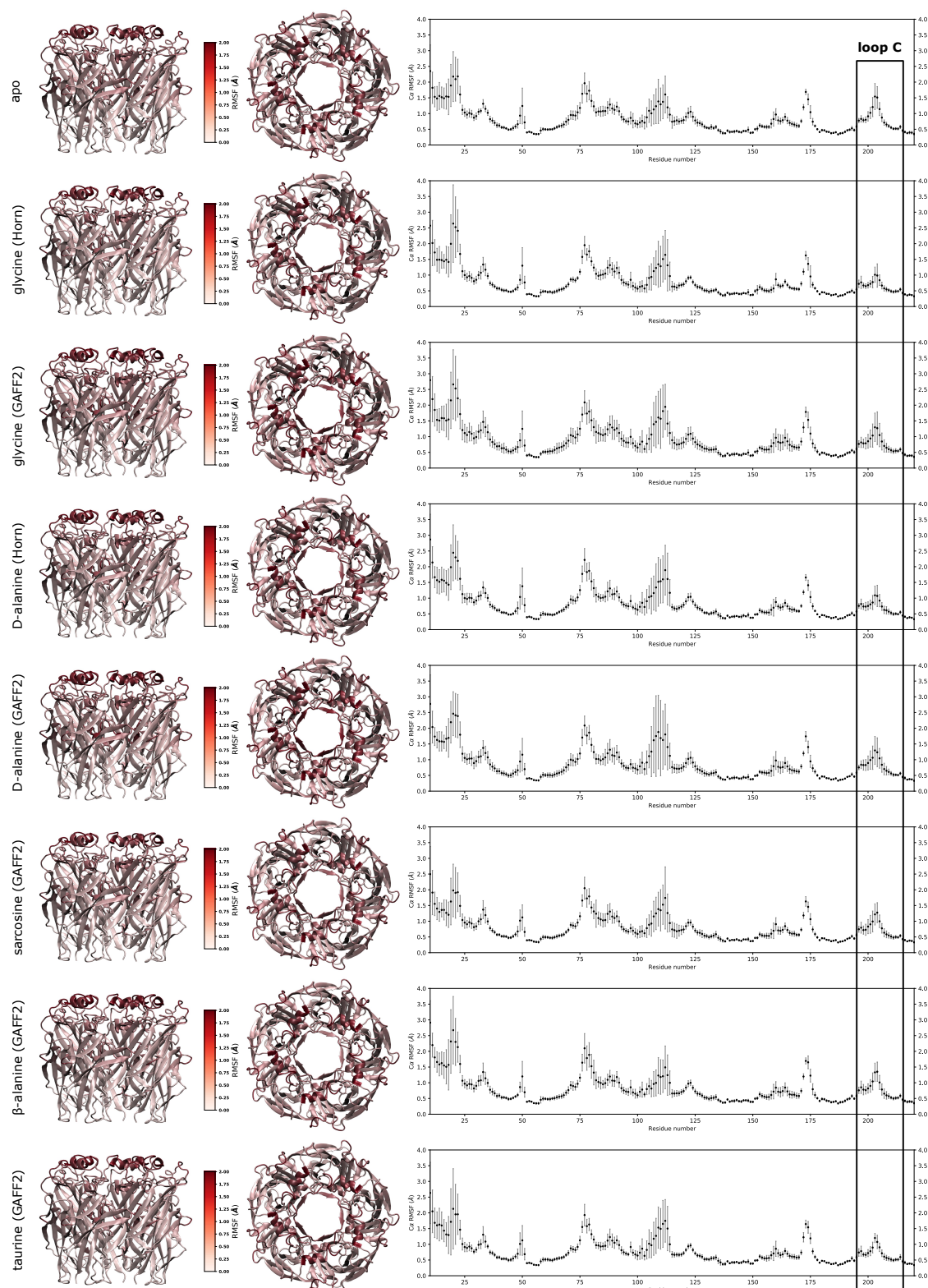


Figure 5.10: C_{α} -RMSF of the ECD from apo and holo simulations. The RMSF values were averaged over all repeats and all subunits, only considering frames with a fully liganded ECD. From left to right: C_{α} -RMSF colour-coded onto the ECD structure, view parallel towards the membrane plane and view from the extracellular towards the intracellular matrix. C_{α} -RMSF vs. residue number with standard deviation. While the global dynamic flexibility of the ECD is similar between apo and holo simulations, the presence of agonist in the orthosteric binding site has a rather localized effect on the receptor dynamics. Most striking is that agonists reduce the dynamic flexibility of loop C (residues 195-216).

agonists in the orthosteric binding site. Therefore, the dynamic behaviour of this loop will be explored in more detail in this section.

To monitor the degree of loop C closure, the distance between the C $_{\alpha}$ atom of (+)T204, positioned at the tip of loop C, and the C $_{\alpha}$ atom of (-)S129 which is located on the opposite side of the binding pocket, was chosen. Moreover, as discussed in section 5.3.1, both of these residues form H-bonds with agonists which is likely to affect loop C conformational dynamics. As previously mentioned, in the holo simulations spontaneous unbinding events occur for various agonists at different binding sites and hence, data frames (every 1 ns) from a binding pocket without a ligand bound are excluded from the distance analysis in order to only consider data from binding pockets with a ligand bound. In Figure 5.11, the probability distribution of this distance is shown in the left column and is complemented by the C $_{\alpha}$ -RMSF profile of loop C as flexibility measure in the right column for apo and holo simulations (obtained as described in section 5.3.2).

In the apo state, the probability density of the (+)T204(-)S129 C $_{\alpha}$ distance is smeared out over a range from 6-16 Å. Its mean value is at 10.2 Å and has a standard deviation of 1.7 Å. Moreover, it is bimodal with local maxima at 8.8 Å and 11.2 Å, suggesting two main conformations of loop C. In the following, the one corresponding to the shorter distance will be referred to as ‘closed’ and the one corresponding to the longer distance will be called ‘open’. On the other hand, in all holo simulations, the probability densities are shifted towards shorter distances in more confined ranges. The average distance is between 8-9 Å for glycine, D-alanine and sarcosine and between 9-10 Å for β -alanine and taurine. This shift demonstrates that the presence of agonists in the orthosteric binding site stabilizes loop C in a closed conformation. The observation, that the probability densities do not resemble a perfect Gaussian distribution, points to the fact that there are several metastable states of loop C. Each of them would be expected to be occupied in a normally distributed manner and the overall density observed is the result of a superimposition of all normal distributions from various metastable states. For example, for glycine there are two local maxima quite close to each other, indicating two main metastable

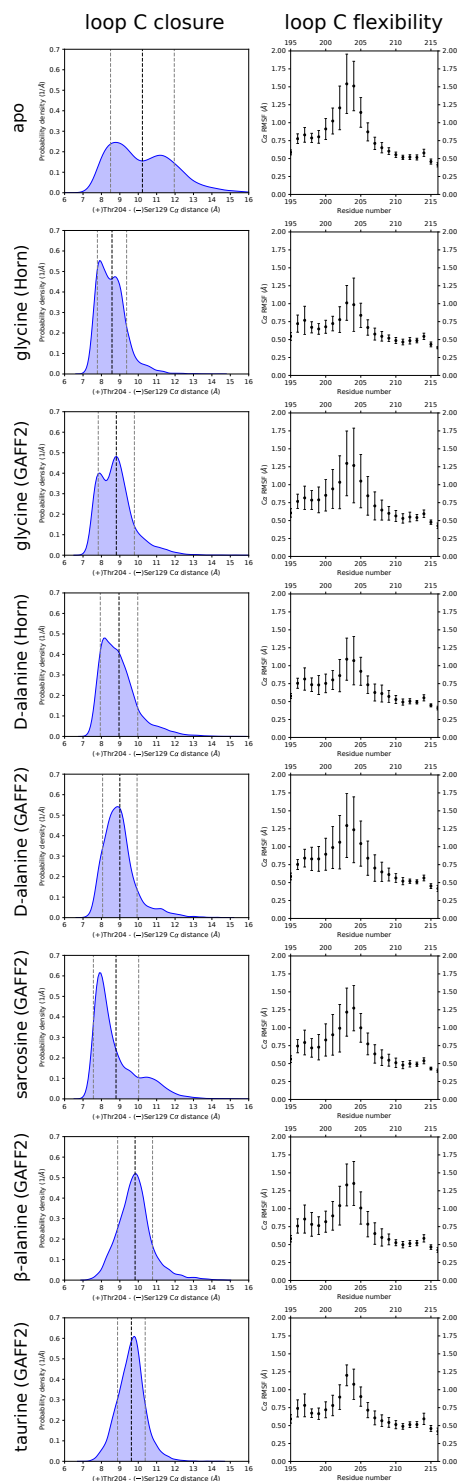


Figure 5.11: Degree of loop C closure and its dynamic flexibility from apo and holo simulations. Left column: Degree of loop C closure as evidenced by the probability density of the (+)T204(-)S129 C_{α} distance. The mean value and one standard deviation from the mean are indicated via dashed black and grey lines, respectively. While in apo simulations the density is spread out over a larger distance range, in holo simulations it is shifted towards shorter distances with a more confined range, meaning that agonists stabilize loop C in a closed conformation. Right Column: C_{α} RMSF with standard deviations of loop C residues as a measure for loop C dynamic flexibility. The presence of agonists in the orthosteric binding site reduces loop C dynamic flexibility compared to apo simulations.

states that are both on the closed end of the conformational spectrum accessible to loop C. Interestingly, both the parameters by the Horn and the GAFF2 force field capture these two local maxima, however, their relative occupation (i.e. free energy), is different between these two force fields. The partial agonists D-alanine and sarcosine seem to have one strongly favoured closed main mode, but a tail towards longer distances indicates that loop C can open up slightly more, although this is less probable (i.e. energetically less stable). The probability densities of the bulkier ligands β -alanine and taurine resemble a normal distribution the most, suggesting that both these ligands stabilize a single stable loop C conformation.

From a wider range of conformational space accessible to loop C as observed in the apo simulations and a narrower range as observed in the holo simulations, one would expect a higher dynamic flexibility of loop C in the apo simulations and more rigid loop C behaviour in the holo simulations. Indeed, this is reflected by the C_α -RMSF profile of loop C displayed in the right column of Figure 5.11. While in the apo simulations, loop C exhibits a high dynamic flexibility, it is reduced in all holo simulations. Especially, the residues on loop C which are engaged in key interactions with the agonists (Y202, T204 and F207) as discussed in section 5.3.1 contribute to less conformational flexibility, as can be seen in the reduction of the corresponding C_α -RMSF values.

The fact that loop C is more flexible in the apo state while it is more confined upon ligand binding is in line with experimental data and previous computational studies on other members of the pLGIC superfamily. From structural data as well as Trp fluorescence studies the picture has emerged that loop C can access multiple conformations in the apo state, but upon ligand binding its dynamic flexibility becomes restricted (see the review by Nys et al. [63] and references therein). In agreement with these experimental findings, various simulation studies on different Cys-loop receptors came to the same conclusion, e.g. [70–72].

In the apo simulations, while loop C is overall dynamically more flexible than in the holo simulations, particularly wide opening events were observed (in repeat 2 at the DE interface as well as in repeat 3 at the AB and BC interface) which are

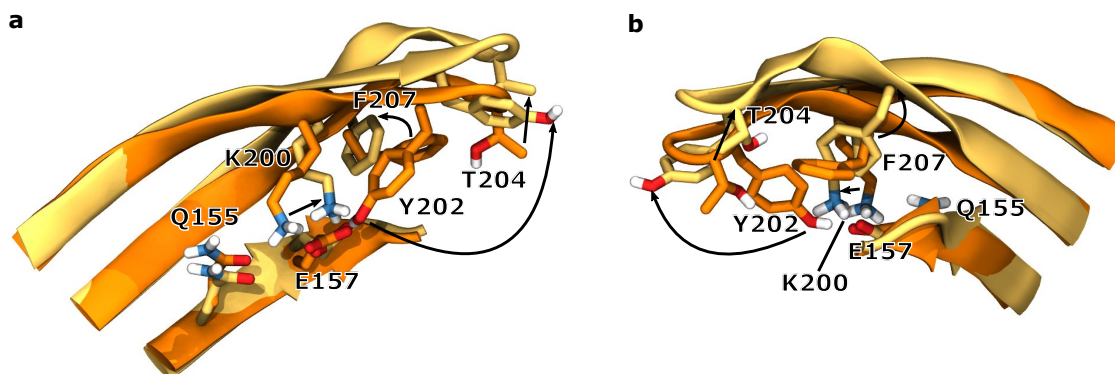


Figure 5.12: Loop C in closed (darker colours) and open conformation (lighter colours) as observed in apo simulations. Panel (a) is viewed from the membrane towards the extracellular matrix, while panel (b) illustrates the view in the opposite direction. Arrows indicate residue-specific rearrangements governing the overall conformational change. Y202 undergoes the largest conformational movement. It stabilizes the closed loop C conformation via a H-bond with E157, while in the open loop C conformation it faces the complementary subunit where it can interact with R65 through a cation- π interaction between the positively charged guanidinium group of R65 and the aromatic π -electron system of the phenyl ring of Y202.

characterized by the breaking of the H-bond between (+)Y202 and (+)E157 and a rotation of (+)Y202 towards the complementary subunit. Figure 5.12 exemplifies the configurational changes between the closed and wide open loop C conformation. In absence of an agonist molecule, the main molecular driving force that keeps loop C from fully opening up is the H-bond between the hydroxy hydrogen of (+)Y202 and one of the carboxylate oxygens of (+)E157. The spontaneous breakage of this H-bond allows (+)Y202 to rotate and face towards the complementary subunit which goes hand in hand with an overall configurational movement that opens up loop C. When (+)Y202 faces the complementary subunit, it can interact with (-)R65 through a cation- π interaction between the positively charged guanidinium group of (-)R65 and the aromatic π -electron system of the phenyl ring of (+)Y202. However, while having a stabilizing effect of the open loop C conformation, this interaction is not strictly necessary to keep loop C open. This opening of loop C is reversible and spontaneous closing was observed as well in repeat 2 at the DE interface.

The structural differences between the closed and open loop C conformation, provide us with insights concerning how agonists can stabilize the closed over the open conformation, namely via a H-bond with (+)T204, a cation- π interaction with

(+)F207 and via not interfering with the H-bond between (+)E157 and (+)Y202. As seen in section 5.3.1, while the full agonist glycine contributes in all these three ways to a closed loop C conformation, all partial agonist in some way contribute less to at least one of these three ways. As spontaneous loop C opening would facilitate unbinding of agonists, the insight from the simulations that partial agonists are less efficient in stabilizing loop C in a closed conformation could provide a structural explanation for the experimental finding that D-alanine (unpublished data by Lucia Sivilotti's group) as well as β -alanine and taurine [281] have higher estimated dissociation rates from $\alpha 1$ GlyR than the full agonist glycine.

In light of this mechanism for loop C opening/closing, the interpretation by Pless et al. [277] that Y202 cannot engage in a cation- π interaction with glycine based on the Y202F mutation leading to a 480-fold higher EC_{50} value [278] can be challenged. I propose that the effect of the Y202F mutation is due to the inability of F202 to engage in a hydrogen bond with E157 which stabilizes loop C in a closed conformation. With Y202F, loop C is much more likely to be in the open conformation depicted in lighter colours in Figure 5.12 with F202 pointing towards the complementary subunit which weakens agonist binding. However, due to Y202's ability to engage in a H-bond via its hydroxy group with E157, loop C is stabilized in a closed conformation which favours stronger agonist binding. In this conformation, the aromatic character of Y202 can contribute to a cation- π interaction, since the distance between the ammonium N and the centre of geometry of the phenyl group is between 3-6 Å (cf. Figure 5.5), which is in the energetically favourable range (with 3.75 Å being optimal for methylammonium and phenyl in aqueous solution [279]).

5.3.4 Binding Free Energies

In this section, the results of the the absolute binding free energy calculations of the agonists discussed in this Chapter are presented. They were obtained computationally via the use of an alchemical cycle as explained in section 5.2 via

$$\Delta G_{\text{binding}} = G_{\text{elec+vdw}}^{\text{solv}} + G_{\text{restr}}^{\text{solv}} + \Delta G_{\text{elec+vdw+restr}}^{\text{prot}} \quad (5.1)$$

The results are summarized in table 5.3. A more detailed breakdown of the different components along the alchemical pathway are reported in table 5.4. The first data column contains $\Delta G_{\text{restr}}^{\text{solv}}$, which corresponds to the desolvation free energy of the ligand, i.e. the higher the value, the more free energy is needed to desolvate the ligand (or vice versa, the more free energy is released when the ligand is solvated). The second data column reports $\Delta G_{\text{restr}}^{\text{solv}}$, which corresponds to the free energy of restraining the ligand's movement and was calculated analytically via Boresch's formula [198]. The third data column lists $\Delta G_{\text{elec+vdw+restr}}^{\text{prot}}$, which corresponds to the free energy of inserting an unrestrained non-interacting ligand into its binding site and then allowing it to interact and release all restraints on it. Finally, the last data column contains the binding free energy $\Delta G_{\text{binding}}$ obtained via equation (2.67). The errors reported for the individual repeats are uncertainties estimated with the MBAR method from the $\frac{\partial H}{\partial \lambda}$ data on the assumption that the phase space has been sufficiently sampled. However, the divergence in the individual values shows that this assumption cannot be justified. Therefore, averages over all three independent repeats with the corresponding standard deviation were calculated in order to provide a more realistic error estimate that takes limited sampling into account.

Table 5.3: Results of the free energy calculations for various agonists of the GlyR ($\Delta G_{\text{binding}}^{\text{computational}}$) compared with estimates from single channel recording data ($\Delta G_{\text{binding}}^{\text{experimental}}$). All values are in units of kcal/mol. * : No data fitted to a flip model available. ** : rat $\alpha 1\beta$ heteropentamer.

Ligand	$\Delta G_{\text{binding}}^{\text{computational}}$	$\Delta G_{\text{binding}}^{\text{experimental}}$
glycine (Horn)	-5.7 ± 1.6	-5.2 ± 0.1
glycine (GAFF2)	-5.0 ± 1.2	-5.2 ± 0.1
D-alanine (Horn)	-6.8 ± 1.1	-3.7 ± 0.1
D-alanine (GAFF2)	-3.3 ± 1.0	-3.7 ± 0.1
sarcosine (GAFF2)	-5.3 ± 1.1	N/A*
β -alanine (GAFF2)	-2.9 ± 0.7	N/A*
taurine (GAFF2)	-5.4 ± 1.1	$-4.3 \pm 0.3^{**}$

The binding free energy of glycine obtained with the parameters by Horn was -5.7 ± 1.6 kcal/mol which is within error of the result obtained with the GAFF2 parameters (-5.0 ± 1.2 kcal/mol). These computational results are in line with the

Table 5.4: Breakdown of free energy calculations performed in triplica for various agonists of the GlyR. All values are in units of kcal/mol. $\Delta G_{\text{restr}}^{\text{solv}}$ has been calculated via the Boresch formula and is an analytical result that does not carry an error. The uncertainties reported with $\Delta G_{\text{elec+vdw}}^{\text{solv}}$ and $\Delta G_{\text{elec+vdw+restr}}^{\text{prot}}$ result from the MBAR uncertainty estimation and the error for $\Delta G_{\text{binding}}$ is obtained via error propagation. Since the MBAR error does not include uncertainty resulting from limited sampling, averages over all 3 independent repeats together with the corresponding standard deviation are reported to provide a more realistic error estimate. Note that the ligands glycine, sarcosine, β -alanine and taurine have a symmetry number of $\sigma_L=2$ (identity and a symmetry plane). Therefore, a symmetry correction of $\Delta G_{\text{sym}} = -RT \ln \sigma_L = -0.427$ kcal/mol was added, resulting in the $\Delta G_{\text{binding}}$ values reported here.

Ligand	$\Delta G_{\text{elec+vdw}}^{\text{solv}}$	$\Delta G_{\text{restr}}^{\text{solv}}$	$\Delta G_{\text{elec+vdw+restr}}^{\text{prot}}$	$\Delta G_{\text{binding}}$
glycine (Horn)	38.728 ± 0.008	6.871	-52.583 ± 0.007	-7.411 ± 0.011
	38.735 ± 0.008	6.871	-49.526 ± 0.008	-4.347 ± 0.011
	38.752 ± 0.008	6.871	-50.552 ± 0.057	-5.356 ± 0.058
average	38.738 ± 0.012	6.871	-50.887 ± 1.556	-5.705 ± 1.561
glycine (GAFF2)	43.535 ± 0.008	6.927	-56.075 ± 0.009	-6.040 ± 0.012
	43.510 ± 0.008	6.927	-55.369 ± 0.008	-5.359 ± 0.011
	43.540 ± 0.008	6.927	-53.686 ± 0.010	-3.646 ± 0.013
average	43.528 ± 0.016	6.927	-55.043 ± 1.227	-5.015 ± 1.234
D-alanine (Horn)	36.974 ± 0.009	6.865	-49.820 ± 0.014	-5.981 ± 0.017
	37.012 ± 0.009	6.865	-51.990 ± 0.053	-8.113 ± 0.054
	37.013 ± 0.009	6.865	-50.192 ± 0.008	-6.314 ± 0.012
average	37.000 ± 0.022	6.865	-50.667 ± 1.160	-6.803 ± 1.147
D-alanine (GAFF2)	42.383 ± 0.010	6.901	-51.600 ± 0.015	-2.316 ± 0.018
	42.370 ± 0.010	6.901	-53.653 ± 0.010	-4.382 ± 0.014
	42.418 ± 0.010	6.901	-52.553 ± 0.011	-3.234 ± 0.015
average	42.390 ± 0.025	6.901	-52.602 ± 1.027	-3.311 ± 1.035
sarcosine (GAFF2)	40.249 ± 0.009	6.814	-53.173 ± 0.013	-6.537 ± 0.016
	40.249 ± 0.009	6.814	-51.680 ± 0.008	-5.044 ± 0.012
	40.235 ± 0.009	6.814	-51.076 ± 0.013	-4.454 ± 0.016
average	40.244 ± 0.008	6.814	-51.976 ± 1.079	-5.345 ± 1.074
β -alanine (GAFF2)	14.271 ± 0.008	6.859	-22.972 ± 0.014	-2.269 ± 0.016
	14.264 ± 0.008	6.859	-23.601 ± 0.007	-2.905 ± 0.011
	14.277 ± 0.008	6.859	-24.323 ± 0.005	-3.614 ± 0.009
average	14.271 ± 0.007	6.859	-23.632 ± 0.676	-2.929 ± 0.673
taurine (GAFF2)	51.000 ± 0.026	6.905	-61.775 ± 0.006	-4.297 ± 0.027
	50.863 ± 0.027	6.905	-63.784 ± 0.006	-6.443 ± 0.028
	51.078 ± 0.027	6.905	-62.908 ± 0.013	-5.352 ± 0.030
average	50.980 ± 0.109	6.905	-62.822 ± 1.007	-5.364 ± 1.073

affinity of glycine to the flipped state of the rat $\alpha 1$ GlyR of -5.2 ± 0.1 kcal/mol (estimated from single channel recording data [100]). The flipped state value was

used for comparison, as it is assumed to more closely resemble the open state than the resting state. Differences could also arise due to the difference in temperature, as experiments were conducted at ca. 21 °C while the computational results provide an estimate at physiological 36.85 °C. Differences due to the different organisms are unlikely as the sequences between rat and human $\alpha 1$ GlyR do not differ in the ECD where the orthosteric binding site is located.

For D-alanine, however, the two different force fields produce different values for the free energy that the error does not account for, namely -6.8 ± 1.1 kcal/mol (Horn) and -3.3 ± 1.0 kcal/mol. The reason for this is that the solvation of D-alanine in the GAFF2 parameterization is by 5.4 ± 0.0 kcal/mol more favourable than in the Horn parameterization, while the respective free energy difference of $\Delta G_{\text{elec+vdw+restr}}^{\text{prot}}$ is only 1.9 ± 1.6 kcal/mol. The computational binding free energy obtained with GAFF2 parameters is within error of the affinity of D-alanine to the flipped state of human $\alpha 1$ GlyR of 3.7 ± 0.1 kcal/mol (estimated from unpublished single channel recording data at T=21 °C from the Sivilotti group), suggesting that the GAFF2 parameter set might be better suited to match affinity from experiments. When comparing binding free energies with each other, the underlying force field should be consistent. Hence, only the GAFF2 binding free energies will be used in the following when comparisons are made.

The calculated binding free energy of sarcosine is -5.3 ± 1.1 kcal/mol and thus within error margins not distinguishable from glycine. Currently, no experimental affinity data for comparison exists for sarcosine.

β -alanine was ranked as the weakest binder by the binding free energy calculations with $\Delta G_{\text{binding}} = -2.9 \pm 0.7$ kcal/mol. No experimental affinity estimates from single channel recordings exist for β -alanine that include the flipped state in kinetic fitting schemes.

Finally, taurine's calculated binding free energy amounts to -5.4 ± 1.1 kcal/mol which is similar to that of sarcosine and also not distinguishable from the value for glycine within the associated uncertainty. This is in agreement within error margins with the experimental estimate from single channel recordings of -4.3 ± 0.3 kcal/mol

to the flipped state [101]. However, it should be noted that this experimental estimate results from measurements of the rat $\alpha 1\beta$ heteropentameric GlyR, which has a higher affinity of glycine to the flipped state than the homopentamer, namely -6.9 ± 0.1 kcal/mol. Hence, a direct comparison between computational and experimental results is difficult.

Overall, the calculated binding free energies have the same order of magnitude as the estimates from single channel recording data (when available) and are in reasonable agreement considering error margins. However, the standard deviations associated with the calculated binding free energies span a range between 0.7 and 1.6 kcal/mol. Considering that the differences in affinity between agonists estimated experimentally are of a similar magnitude as these standard deviations, the precision limit of the computational results makes it difficult to compare the binding free energies of agonists with each other.

5.4 Discussion

The binding mode for glycine obtained here via MD simulations is in excellent agreement with mutagenesis and structural data, including the position of a stable water molecule in the binding site. This is encouraging in terms of the predictive power for the binding modes described for the partial agonists for which no structural is available so far.

When affinity estimates from single channel electrophysiology data were available, the computed binding free energies are in line with them. However, both precision as well accuracy could be improved.

The precision could be improved by targeting the ligand-receptor leg of the alchemical cycle, as the error from the solvation free energies is negligible. This could be achieved by increasing the sampling time of each window and performing more than three independent repeats of the entire leg. However, it should be noted that the ligand-complex free energy calculations with 52 windows of 40 ns length each already are the computationally most extensive alchemical free energy calculations performed to date. For comparison, a landmark study by Aldhegi et al. [196] only

employed 42 windows of 10 ns length each, resulting in better precision than here and excellent agreement with experiment. This shows that the quality of alchemical binding free energies is strongly system-dependent.

Besides improving the precision, there are many ways to improve the accuracy (i.e. the systematic error) of the binding free energy calculations. The accuracy could be improved by using a finer discretization of the λ parameter that controls the alchemical pathway, however, the accuracy improvement for any finer discretization than $\lambda = 0.05$ spacing for both the Coulomb and van der Waals interactions is expected to be small compared to other possible ways to address the systematic error. It is well known that the accuracy of binding free energy calculations depends on the quality of the force field parameters [282–284]. From a survey of solvation free energies using different charge models, AM1-BCC [273] and RESP [285] charges emerged to be in best agreement with experimental data [283]. Both of these charge models have been used here, AM1-BCC charges in combination with the GAFF2 force field parameters [168] and RESP charges in combination with parameters by Horn [169] that are based on the AMBER99SB-ILDN force field [143] which was also used for the protein. The differences between the two charge models are illustrated in Figure 5.13 for glycine and D-alanine. The main difference between the charge models is that the distribution of partial charges at the ammonium and carboxylate groups is more extreme in the AM1-BCC model than in the RESP model, resulting in a more polar molecule. This explains the lower solvation free energies for the AM1-BCC charge model compared to the RESP model. From a theoretical perspective, RESP charges should be of higher quality than AM1-BCC charges. The van der Waals parameters can also strongly impact binding free energies [284]. However, the interactions for the zwitterionic compounds examined here with water as well as the orthosteric binding site of the GlyR are mainly dominated by electrostatic interactions and thus, optimization of the van der Waals parameters is less important than the Coulomb parameters. Bonded force field parameters, particularly torsions, can also have an influence on calculated affinities [197].

Even if all those factors were taken into account with perfect accuracy, classical fixed-charge force fields by construction cannot account for polarizability, i.e. the dynamic adaptation of the charge distribution in a molecule in response to its environment. Polarizability is known to be important in many situations in biochemistry and it is expected to be particularly important for the binding of zwitterionic molecules in the orthosteric binding site of the GlyR. Specifically, inclusion of explicit polarization could more accurately account for the cation- π interactions of the ammonium group with aromatic residues as well as account for dynamic charge rearrangements in the zwitterionic ligands in response to the electric field in the orthosteric binding site spanned by the negatively charged (+)E157 and the negatively charged (-)R65 residues [286]. The pairwise additive fixed-charge force field AMBER99SB-ILDN used in this study is limited in its ability to correctly capture cation- π interactions. The charge distribution in the phenyl ring of phenylalanine residues is such that the carbons carry mostly slightly negative partial charges (C_γ : 0.01, C_δ : -0.13, C_ϵ : -0.17 and C_ζ : -0.11), while the hydrogens carry slightly positive partial charges (H_δ : 0.13, H_ϵ : 0.14 and H_ζ : 0.13), which gives rise to an electric field that would be expected for a phenyl ring. Since the distance between the methylammonium group of glycine and the phenyl COG is close to the optimum of 3.75 Å [279] (cf. Figure 5.5), it appears that at least the distance of the cation- π interaction is well described. However, the energetics might be less accurate and moreover, the fixed partial charges cannot capture any induced polarization, e.g. due to the presence of a cation. To improve the cation- π interaction in the context of the AMBER force field, Monoux and Chipot [287] have proposed a short-range 4-12 potential to reproduce induction phenomena in an average sense. More recently in the context of the Charmm force field, Khan et al. [288] suggested a modification of the Lennard-Jones parameters to recover missing induction effects. Neglect of polarizability in general has been shown to produce inaccurate binding free energies in cases where polarization changes occur upon ligand binding [289–291]. This is also expected for zwitterionic ligand binding to the GlyR. In reality, the zwitterionic agonists would undergo induced polarization due to the electrostatic

field in the binding pocket exerted by (+)E157 and (-)R65, which would lower the free energy of the protein-ligand complex. In solution, however, there is no such external electric field and polarization contributions from polarizable water and ions could be captured equally well in a mean-field approximation by classical fixed-charged models. It is therefore likely that the force fields employed here, which do not account for induced polarization effects, systematically underestimate the absolute binding free energy. Hence, in theory, more accurate affinities could be obtained with polarizable force fields such as AMOEBA [292] or Charmm Drude [293], but the computational overhead that comes along with these more elaborate models practically limits the sampling to much shorter time scales which would introduce a larger systematic error due to poor sampling. However, with increasing computing power, this approach could become feasible in the near future.

Independent from the force field used, the choice of the Boresch restraints which keep the ligand in a pre-defined orientation in the binding pocket also affect the outcome of a free energy calculation. For a rigid ligand binding pose, Boresch restraints are well suited, however, if the ligand and residues in the binding pocket are very dynamic, they are likely to introduce a systematic error. Take D-alanine, for example, where standard MD-simulations show that rotation around the C_{α} - $C_{\text{carboxylate}}$ bond leads to a weakened cation- π interaction with (+)F207 on average. The Boresch restraints employed here do not allow for this rotation and, thus, introduce an artificial bias of the binding pose. Using various different types of Boresch restraints could lead to better accuracy. Ideally, the ligand should have maximal freedom to change orientation. This could be achieved by a different set of restraints that only confine the ligand in a fixed volume in the binding pocket. Practically, however, this would require much longer sampling times to yield convergence.

A promising approach that can overcome the shortcomings imposed by using orientational restraints on the ligand is funnel metadynamics [294]. This is an enhanced sampling method, where metadynamics is used to speed up the ligand's exploration of phase space together with a funnel-shaped restraint potential to keep

the ligand to relevant parts of the phase space while sampling a statistically sufficient amount of unbinding and rebinding events to provide a binding free energy estimate. The advantage is that the ligand is free to explore the whole binding pocket and all relevant metastable binding poses, leading to a more accurate affinity estimate. Furthermore, mechanistic insight of the binding/unbinding pathway can be obtained and kinetics (k_{on} and k_{off}) can be directly compared to experimental estimates from single channel recordings. The free energy landscape of the binding/unbinding pathways could also be exploited for rational drug design in order to customize the drug binding kinetics. In fact, the funnel metadynamics approach has already been successfully applied to GABA binding to the RDL receptor, an insect-specific member of the pLGIC superfamily [269].

Lastly, the conserved water binding site in the orthosteric binding pocket deserves a final remark. Since it is important for the binding mode of all partial agonists, as structures for these agonists appear, the water densities obtained in this study can help to distinguish density originating from ligand and this stable water, so it can be included in the modelling process. Moreover, as this water site is also occupied in the apo state, it opens up a new avenue for rational drug design which can take advantage of designing compounds with groups that can replace this stable water, leading to additional affinity gains, both from enthalpy contributions from additional interactions with the receptor as well as entropy contributions from releasing the confined water molecule to the bulk phase.

5.5 Conclusions

In conclusion, in this Chapter the binding modes of the full agonist glycine and the partial agonists D-alanine, sarcosine, β -alanine and taurine have been characterized, their influence on receptor dynamics, particularly loop C, has been analyzed and their binding free energies were determined.

From the binding mode of the natural agonist glycine, key residues in the orthosteric binding site could be functionally annotated. While glycine's negatively charged carboxylate group is stabilized by H-bonds with (+)T204 and (-)S129 as

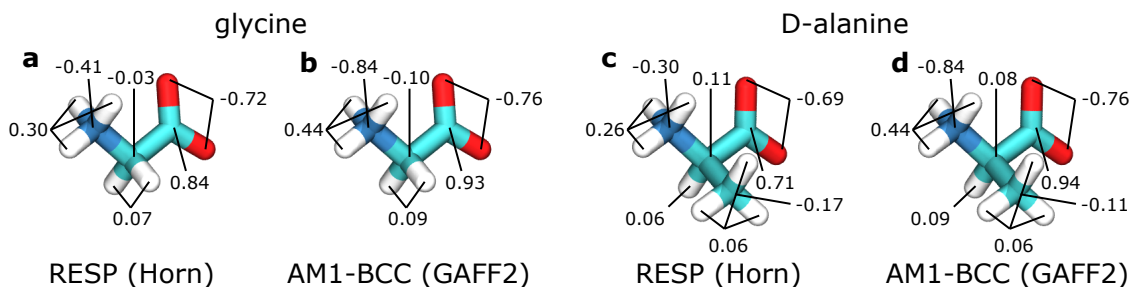


Figure 5.13: Comparison of the partial charges in the two different parameterizations used for glycine and D-alanine. In the parameter set by Horn atomic charges were obtained via the restrained electrostatic potential (RESP) [285] charge derivation strategy following the established procedure [295, 296]. The atomic charges used with the GAFF2 force field parameters were assigned using AM1-BCC charges [273]. The main difference between the two sets of partial charges is that for the AM1-BCC model, the charge distribution at the charged ammonium and carboxylate ends is more extreme according to electronegativity.

well as a saltbridge with R65 which simultaneously acts as a bidentate H-bond donor, its positively charged ammonium moiety forms a cation- π interaction with (+)F207, a H-bond with the backbone carbonyl oxygen of (+)F159 and engages in an electrostatic interaction with (+)E157 which is bridged by a stable water molecule. Hydrophobic shielding from bulk water is another important function for ligand binding carried out towards the extracellular side by (+)F207, (-)L117, (-)L127 together with the hydrophobic methylgroup of (+)T204 and towards the membrane direction by (-)F63, (+)F159 and (+)Y202 (and less direct (+)F99 and (-)F44). The H-bond between Y202 and E157 is a particularly important interaction that stabilizes loop C in a closed conformation. Other contributions to a closed loop C conformation are mediated by glycine via its interactions with (+)F207 and (+)T204. Via engaging in interactions with both the principal and complementary subunit, the presence of glycine leads to a contraction of the orthosteric binding site, thus bringing the adjacent subunits closer together. The binding mode obtained from simulations is in excellent agreement with the crystal structure PDB: 5TIN as well as experimental mutagenesis data.

The partial agonists D-alanine, sarcosine, β -alanine and taurine have very similar binding features compared to glycine. In common are the agonist orientation, facing the principal subunit with the positively charged ammonium moiety and the complementary subunit with the negatively charged carboxylate or sulfonium group.

In all cases, the interaction of the positive ammonium end with the negatively charged carboxylate group of (+)E157 is bridged via a water molecule. However, subtle differences do exist. D-alanine's bulkier side chain can weaken the Y202-E157 H-bond which is crucial for a closed loop C conformation. Moreover, D-alanine's ammonium and methyl group can rotate around the C_{α} - $C_{\text{carboxylate}}$ bond and thus the cation- π interaction with (+)F207 is weaker on average. Similarly, sarcosine's N-methyl-group dampens the cation- π interaction with (+)F207. For β -alanine, the saltbridge with (-)R65 is weakened. Taurine's bulkier sulfonium group wedges loop C slightly more open than glycine's more compact carboxylate group and its interactions are weaker, as it rotates more frequently. Overall, all partial agonists interact less optimally than glycine with the receptor and contribute less efficiently towards a closed loop C conformation. This provides a structural explanation for faster dissociation rates of partial agonists compared to glycine.

The agonists' influence on receptor dynamics is localized on loop C, stabilising it in a closed conformation and restricting its dynamic flexibility compared to the apo state, which is in agreement with experimental data as well as MD simulations obtained from other members of the pLGIC superfamily.

Lastly, for the first time absolute binding free energies for agonists to the active conformation of GlyR have been calculated computationally. The affinities obtained with GAFF2 parameters are: -5.0 ± 1.2 kcal/mol for glycine, -3.3 ± 1.0 kcal/mol for D-alanine, -5.3 ± 1.1 kcal/mol for sarcosine, -2.9 ± 0.7 kcal/mol for β -alanine and -5.4 ± 1.1 kcal/mol for taurine. They are in line with estimates for the flipped state from single channel recordings (where data are available). Since the relative affinity differences between partial agonists based on available experimental data are subtle and of similar magnitude as the precision limit of the computational results, it remains difficult to delineate the contributions of affinity to efficacy.

You can never cross the ocean unless you have the courage to lose sight of the shore.

— Christopher Columbus

6

Towards an Atomistic Gating Mechanism

Contents

6.1	Introduction	180
6.2	Methods	182
6.2.1	Targeted MD	182
6.2.2	Unbiased MD along the Gating Path	183
6.2.3	MSM Estimation and Validation	184
6.2.4	Analysis	186
6.3	Results	187
6.3.1	Targeted MD Generates an Initial Guess for a Gating Path	187
6.3.2	Insights into the Gating Mechanism from Markov State Modelling	192
6.4	Discussion	196
6.5	Conclusions	198

A long-standing goal of the pLGIC community is to understand the gating mechanism at an atomistic level of resolution. While the number of high resolution pLGIC structures in different states has greatly increased, predicted short-lived intermediates are difficult to trap and a dynamic understanding including all intermediates has remained elusive thus far. In this chapter, I strive to obtain insights into the gating mechanism of the human $\alpha 1$ GlyR using a combination of targeted MD and Markov State Modelling.

6.1 Introduction

pLGICs are allosteric machines. Upon the binding of neurotransmitters at their ECD, they undergo conformational changes that lead to the opening of the channel pore in the TMD [7]. However, a detailed atomistic mechanism of the full gating process has remained elusive thus far.

Over the last decade a large increase in solved structures of various members of the pLGIC superfamily in different states has provided an atomistic, yet stroboscopic insight into the gating process. Particularly, pLGIC members with both closed and open states resolved provide valuable endpoints for the gating transition. Several structures in the closed and open state have been solved for the prokaryotic pH-controlled GLIC [58, 207, 233, 297]. Moreover, closed and open state structures of eukaryotic members of the superfamily are available, such as for the cationic 5HT₃R [54, 234] as well as the anionic GlyR [60, 68]. However, there has been some controversy about the physiological relevance of the GlyR open state structure PDB: 3JAE, as it was unstable in MD simulations [210], but a stable open-state structure could be obtained via MD-refinement as demonstrated in Chapter 3 [221].

Through structural comparison of these end-point structures, conformational changes of the gating mechanism have been inferred (see [7] for a detailed review). There are several common main features characterizing the gating transition, which is not surprising considering the conserved architecture within the pLGIC superfamily. From the perspective of the closed-to-open transition, the ECD undergoes a quaternary anticlockwise twist relative to the TMD. Moreover, an overall contraction of the upper ECD part is observed that is commonly referred to as ‘unblooming’ in the literature. It is noteworthy that both of these global rearrangements have been predicted by normal mode analysis [298, 299]. At the TMD level, the subtleties in conformational rearrangement are slightly more diverse within the superfamily, but the key to channel pore opening is always the widening of the pore radius by movement of the M2 helices.

Compelling evidence for a similar gating mechanism across different pLGICs is further provided by genetic engineering of functional chimeras, consisting of

domains from different members of the superfamily, such as the $\alpha 7$ -nAChR(ECD)-5HT₃R(TMD+ICD) [300], $\alpha 7$ -nAChR(ECD)/ $\alpha 1$ -GlyR(TMD+ICD) [301], ELIC(ECD)/GLIC(TMD) [302] and even prokaryotic-eukaryotic combinations, such as GLIC(ECD)/ $\alpha 1$ -GlyR(TMD) [303] and ELIC(ECD)/ $\alpha 7$ -nAChR(TMD) [304].

Prokaryotic pLGICs seem to contain the minimal common core necessary for gating, which suggests that the ICD is not crucial to gating. This also seems to be true for the eukaryotic GlyR construct that lacks the ICD, for which two-electrode voltage clamp electrophysiology experiments have demonstrated that it remains functional [68].

However, the structures only provide the endpoint states of the gating transition and the actual gating pathway might involve several metastable intermediate structures not yet captured by structural methods. From electrophysiological single channel recordings, it has become apparent that there are one or more short-lived intermediate shut states (flipped/primed) along the channel activation pathway [100–109]. Yet, a structural characterization of these intermediates has remained elusive. Interestingly, an intermediate state of GLIC as well as GlyR with activated ECD and closed TMD could be trapped by cross-linking at the outer region of the pore to constrain its conformation [305, 306], however, given these artificial cross-links, it is unclear to what extent these structures might correspond to a physiologically relevant intermediate structure.

In this respect, MD simulations can be a powerful tool to detect short-lived or highly flexible intermediates which are difficult to trap experimentally, but might be crucial for the molecular understanding of the allosteric mechanism. For pLGICs they have already contributed some insight into the dynamics governing conformational changes (refer to the reviews [307, 308] and references therein). Indeed, some of the global gating movements inferred from structural data could also be observed to a limited extent in MD simulations, see e.g. [70, 212]. The main limitation, however, are the accessible time scales of standard MD simulations, for which the full gating mechanism of pLGICs on the order of several hundred μ s [100] is currently still out of reach. The most noteworthy recent computational

advances towards deciphering the full gating transition are the elastic-network Brownian dynamics simulation approach developed by Orellana et al. applied to GLIC [309] as well as the string-method with swarm trajectories method [310] applied to GLIC by Lev et al. [111, 311]. However, both methods are limited in the sense that, while they provide optimized transition pathways, they do not provide direct insight into the dynamics.

While various enhanced sampling methods, such as metadynamics [312], exist that could provide insight into the dynamics of the gating process, Markov state models (MSMs) [138, 139] provide the advantage that they do not require an a-priori knowledge about important collective variables (CVs) of the transition. This approach has already been applied successfully to reveal detailed insights into the intrinsic conformational dynamics upon the binding event in various proteins (see e.g. [313, 314] for overview articles and in particular [315] for an example of ligand modulation of activation pathways).

In this study, I set out to obtain insights into the gating mechanism of the human $\alpha 1$ GlyR as a representative for the pLGIC superfamily. To this end, targeted MD is used to generate an initial guess pathway between the resting state based on PDB: 5CFB [60] and the refined open state structure from Chapter 3 [221]. Unbiased MD simulations initiated from 285 frames along this guess pathway totalling in an accumulated simulation time of 85.5 μ s have been performed from which a fully connected MSM has been constructed, which reweights the trajectories such that the equilibrium distribution (and thus the free energy surface) as well as kinetics of the full gating pathway can be recovered.

6.2 Methods

6.2.1 Targeted MD

System Setup

The simulation system was obtained from a CG-system with the inactive receptor state obtained via the same procedure as described in section 4.2 after 20 μ s of NPT-simulation. It was reverted to an atomistic representation via the programme

‘backward’ [220]. The backmapped receptor was substituted by the original all-atom homology model of human $\alpha 1$ GlyR based on the human $\alpha 3$ GlyR (strychnine-bound closed state; PDB: 5CFB [60]). Glycine ligands and stable binding site water molecules had been inserted into the homology model via alignment with the glycine positions in PDB: 5TIN [56]. This resulted in an atomistic system with the same neuronal brain membrane composition as specified in section 4.2 and a total number of 142,729 atoms.

Simulation Protocol

The system was energy-minimized and equilibrated following the same initial equilibration protocol (before the additional phase with pore restraints) as described in section 3.2.3, only that GROMACS version 2016.4 [141] was used as MD-engine. The system was then simulated for an additional 10 ns in the NPT-ensemble with the Parrinello-Rahman barostat with no restraints other than flat bottom restraints on the ligands as described in section 3.2.3. The final frame was the starting point for targeted MD simulations. The target structure was the representative stable open state structure from Chapter 3. Using GROMACS version 2016.4 patched with plumed version 2.2.3 [316], six targeted MD simulations were performed using different force constants $\kappa=1\cdot 10^5$ kJmol⁻¹nm⁻², $2\cdot 10^5$ kJmol⁻¹nm⁻², $5\cdot 10^5$ kJmol⁻¹nm⁻², $1\cdot 10^6$ kJmol⁻¹nm⁻², $2\cdot 10^6$ kJmol⁻¹nm⁻², $6\cdot 10^5$ kJmol⁻¹nm⁻². The force constant was increased linearly from 0 to its final value over 25 ns and then kept constant for another 25 ns. It acted on all protein atoms and the RMSD was constantly measured after optimal alignment with the target structure. The remaining simulation parameters were the same as described in section 3.2.3, using flat bottom restraints on the ligands to prevent them from unbinding, but no pore restraints. Positions were written out every 10 ps.

6.2.2 Unbiased MD along the Gating Path

285 frames along the initial guess path trajectory generated by targeted MD with $\kappa=2\cdot 10^5$ kJmol⁻¹nm⁻² were extracted as starting configurations for unbiased MD

simulations. For each system seed the simulation protocol consisted of energy minimization, 100 ps NVT-equilibration and 1 ns NPT-equilibration and a 300 ns production run with the same settings as specified in section 3.2.3. This resulted in an accumulated simulation time of 85.5 μ s for MSM construction.

6.2.3 MSM Estimation and Validation

All the following steps have been performed using the pyEMMA software package [317], version 2.5.7.

Dimensionality Reduction

The natural internal degrees of freedom of proteins that capture their conformational flexibility are their backbone dihedrals. This feature has therefore been chosen as the input for further dimensionality reduction. Time-lagged independent component analysis (TICA) [318] has been employed to reduce the dimensionality of the full dihedral space onto a space where distance corresponds to the slowest kinetic time scales. A lag time of 30 ns and a dimensionality of 2 has been chosen for the TICA space construction. Figure 6.1 illustrates the distribution of the simulation data in this TICA space.

MSM Construction

The optimal choice for the number of microstates of the MSM is a compromise between reducing the discretization error by using as many microstates as possible, but at the same time obtain sufficient statistics for transitions between these microstates, which limits the number of microstates [319]. The VAMP-2 score [320] can provide a good guidance for this choice, though it is generated using MSMs that were not validated and thus should be considered a heuristic. As can be seen in Figure 6.2a, choosing 250 microstates is a good compromise between minimizing the number of microstates for better transition statistics, while the VAMP-2 score is (within error) in the saturated region. Therefore, k-means clustering was performed on the simulation data in TICA space with 250 cluster centres, resulting in a Voronoi tessellation of the TICA space, where each Voronoi cell represents a microstate

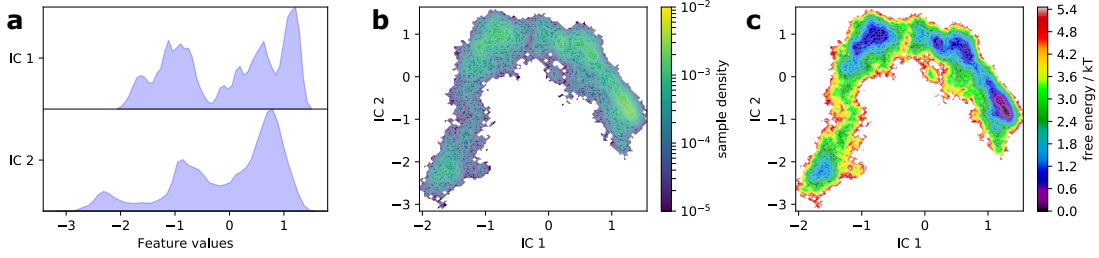


Figure 6.1: Distribution of the accumulated simulation data in TICA space (a-b) and corresponding free energy surface (c). In this TICA space, the closed starting structure used for targeted MD has coordinates $(IC1, IC2) = (1.3, -0.8)$ and the stable open target structure $(IC1, IC2) = (-1.5, -2.1)$. The analysis in section 6.3.2 will show that the transition region between the shut and open state lies in the upper part of (b) and (c) around $IC1 = 0$ and that the gating process involves a shut intermediate state, which is roughly in the region with $IC1 < 0$ and $IC2 > 0$. The TICA coordinates thus roughly align with the transition between the closed and open state ($IC1$) and between the closed state and the intermediate shut state ($IC2$) and further indicate that these are the slowest kinetic processes in the system. Note that the free energy surface does not correspond to the correct equilibrium free energy surface, as it has not been reweighted according to the equilibrium information contained in the MSM yet.

(cf. Figure 6.2b). Finally, various MSMs were estimated using different lag times and the implied time scales (ITS) are visualized in Figure 6.2c. Convergence of the slowest two ITS is reached (within error) at a lag time of 50 ns, which was therefore chosen as the lag time for the final MSM.

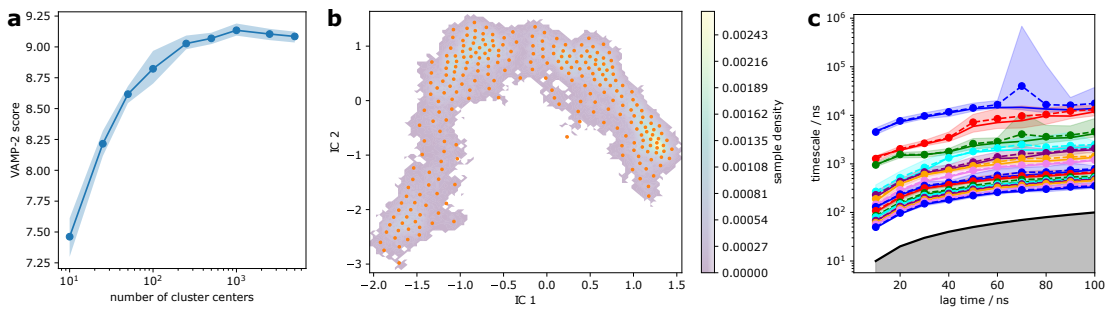


Figure 6.2: MSM construction. (a) VAMP-2 score for different numbers of microstates. (b) 250 cluster centres from k-means clustering, each defining a microstate via a Voronoi tessellation of the TICA space. (c) 15 slowest implied time scales (ITS) for MSMs estimated at various lag times. Convergence of the slowest two ITS is reached (within error) at a lag time of 50 ns. One can further see a separation of time scales between the two slowest (blue and red) and the remaining ITS, indicating that the overall dynamics of the system is governed by two slow processes separating the configuration space into three main states.

MSM Validation

The convergence of implied time scales is already a first validation test to ensure Markovianity. Additionally, a Chapman-Kolmogorow test was conducted as a second validation step, using three metastable states, as the ITS plot indicates a separation of time scales after the two slowest ITS. It demonstrates that the MSM estimated with a 50 ns lag time is consistent with the simulation data up to 200 ns Figure 6.3.

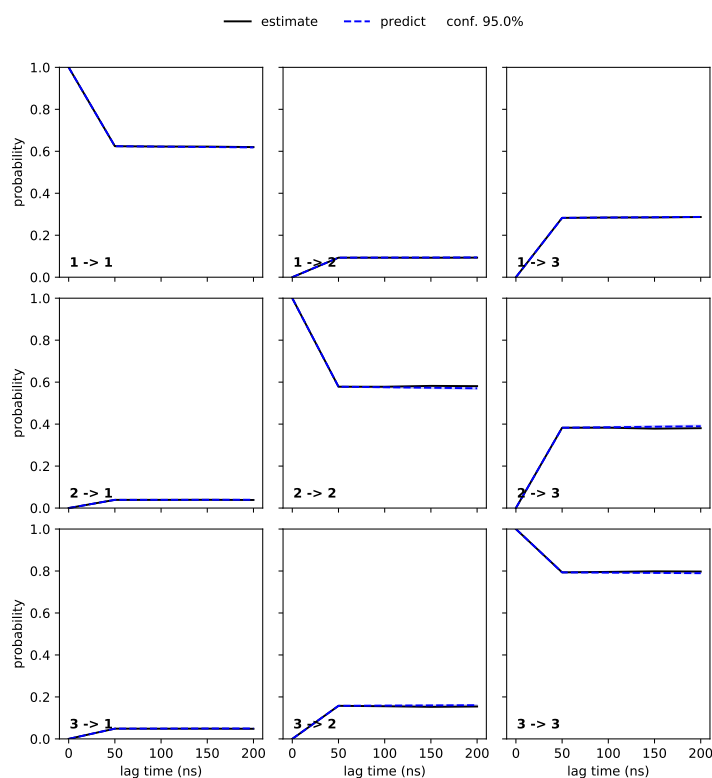


Figure 6.3: The Chapman-Kolmogorow test with three metastable states demonstrates that the MSM estimated with a 50 ns lag time is consistent with the simulation data up to 200 ns.

6.2.4 Analysis

Structural analysis was performed with MDAnalysis [226]. Visualization was performed in VMD [229]. All MSM-related analysis was performed with pyEMMA [317], version 2.5.7.

6.3 Results

6.3.1 Targeted MD Generates an Initial Guess for a Gating Path

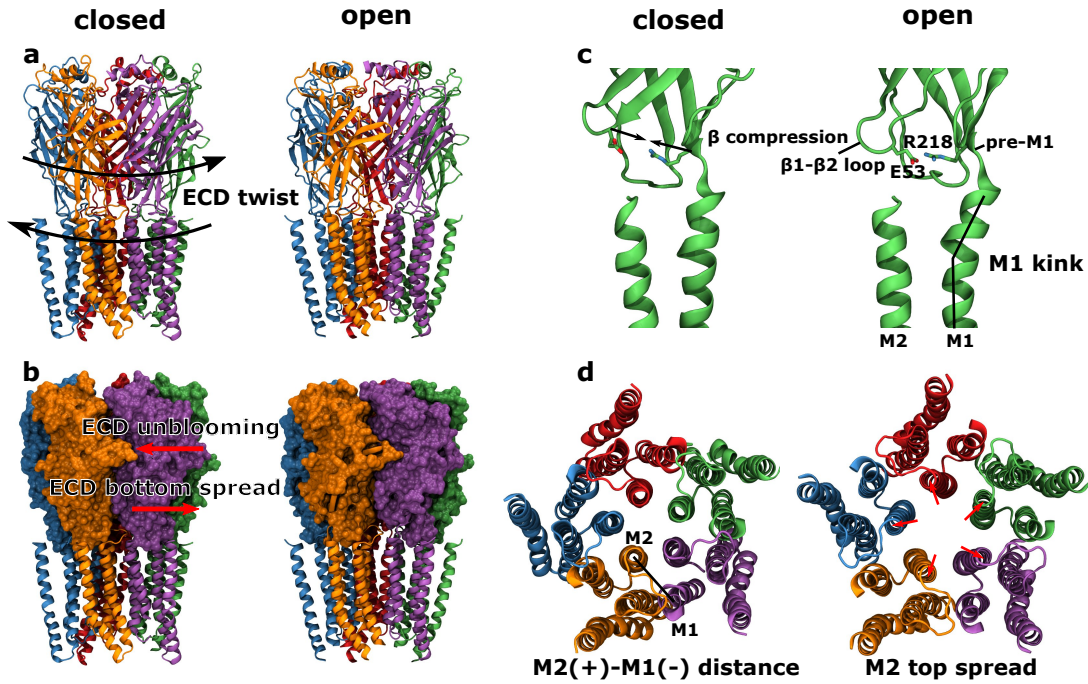


Figure 6.4: Illustration of gating CVs in the closed-to-open transition for the GlyR. (a) ECD twist, (b) ECD unblooming and ECD bottom spread, (c) β compression and M1 kink, (d) M2(+)-M1(-) distance and M2 top spread.

Lev et al. have identified collective variables (CVs) that characterize the conformational space of the full gating transition of the pLGIC GLIC [111]. Here, I use them analogously for the homology model of the human $\alpha 1$ GlyR (see Figure 6.4 for their illustration). Since Lev et al. named them from the perspective of the open-to-closed transition, I renamed some of them here, so that the names are suitable for the closed-to-open transition. Note that each point in the following list constitutes five CVs, resulting in a total of 35 CVs:

1. *ECD twist*: Dihedral defined by the ECD¹ subunit C_α centre of geometry (COG) - full ECD C_α COG - full TMD² C_α COG - TMD subunit C_α COG (Figure 6.4a).

¹Here, I define the ECD as residues P10-M220.

²Here, I define the TMD as residues G221-R309-A-G-T-K386-K419.

2. *ECD unblooming*³: Distance between the full ECD C_α COG and the C_α COG of the upper half⁴ of the ECD per subunit (Figure 6.4b).
3. *ECD bottom spread*: Distance between the full ECD C_α COG and the C_α COG of the lower half of the ECD per subunit (Figure 6.4b).
4. *β compression*⁵: Lower β -sandwich compression defined as the subunit-wise distance between the C_α COG of residues 51-55 (corresponding to residues 30-34 in GLIC) and the C_α COG of residues 216-220 (corresponding to residues 190-194 in GLIC) (Figure 6.4c).
5. *M2(+)-M1(-) distance*: Distance between M2(+) (C_α COG of residues 268-272 (corresponding to residues 241-245 in GLIC)) and M1(-) (C_α COG of residues 226-230 (corresponding to residues 200-204 in GLIC)) (Figure 6.4d).
6. *M2 top spread*: Distance between subunit C_α COG of residues 268-272 (corresponding to residues 241-245 in GLIC) and full TMD C_α COG (Figure 6.4d).
7. *M1 kink*: Angle defined by subunit C_α COG of residues 226-230 (corresponding to residues 200-204 in GLIC) - subunit C_α COG of residues 228-232 (corresponding to residues 202-206 in GLIC) - subunit C_α COG of residues 231-242 (corresponding to residues 205-216 in GLIC) (Figure 6.4c).

The first task is to assess whether the targeted MD simulations could capture the complete gating pathway from the closed to the open state. To this end, the gating CVs were calculated, averaged over all five subunits and scaled on an interval from 0 to 1 to allow relative comparison between all gating CVs. Their time-evolution is plotted in Figure 6.5 for six targeted MD-simulations with different force constants. As expected, the higher the force constant, the quicker the transition from the

³Lev et al.: ECD top spread.

⁴The exact definition of lower and upper ECD half is not provided in Lev et al.'s paper [111]. Here, I define the upper half of the ECD as follows: The ECD's midpoint in the direction of the channel axis is used to define all the residues with C_α atoms below this midpoint as belonging to the lower half and all residues with C_α atoms above this midpoint as the upper half. Specifically, the lower ECD half consists of residues 42-63, 100-107, 132-156, 170, 172-201, 209-220 and the upper ECD half consists of residues 10-41, 64-99, 108-131, 157-169, 171, 202-208.

⁵Lev et al.: β expansion.

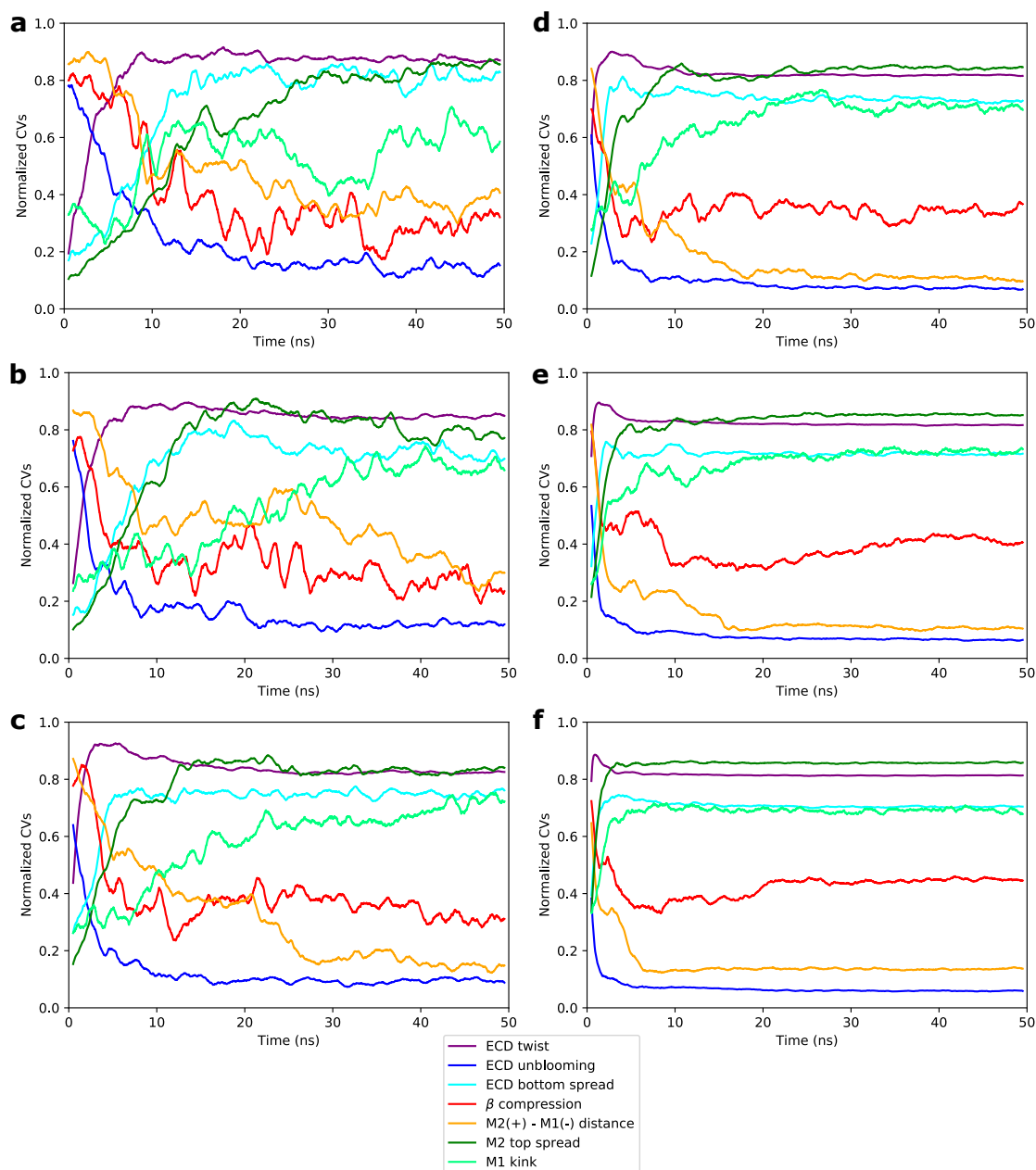


Figure 6.5: Gating CVs (averaged over all five subunits and scaled on an interval between 0 and 1) over the course of 50 ns of targeted MD with different force constants κ (running average over 100 frames). (a) $\kappa = 1 \cdot 10^5 \text{ kJmol}^{-1}\text{nm}^{-2}$, (b) $\kappa = 2 \cdot 10^5 \text{ kJmol}^{-1}\text{nm}^{-2}$, (c) $\kappa = 5 \cdot 10^5 \text{ kJmol}^{-1}\text{nm}^{-2}$, (d) $\kappa = 1 \cdot 10^6 \text{ kJmol}^{-1}\text{nm}^{-2}$, (e) $\kappa = 2 \cdot 10^6 \text{ kJmol}^{-1}\text{nm}^{-2}$, (f) $\kappa = 5 \cdot 10^6 \text{ kJmol}^{-1}\text{nm}^{-2}$.

closed to open state takes place during the 50 ns of targeted MD. The order of the different conformational changes described by the gating CVs is the same for the different force constants. First, the ECD twists alongside with ECD unblooming. Then the ECD bottom spread follows and the lower β sheets in the lower ECD contract (β compression) simultaneously. After these initial conformational changes in the ECD, the signal is transduced to the TMD. The M2(+)-M1(-) distance decreases at the same time as the upper part of the M2 helices spread out (M2 top spread). Finally, the M1 helices kink, completing the closed-to-open gating transition. The observed sequence of configurational changes in the ECD elicited by neurotransmitter binding, followed by configurational changes in the TMD is encouraging, as this is what one would expect.

In the simulations with force constants $\kappa=1 \cdot 10^6$ kJmol⁻¹nm⁻² and higher Figure 6.5d-f, the gating transition occurs rapidly within the first 8-20 ns of the simulation. Further, from these simulations with higher force constant, it can be read out to what values the CVs converge for the open state. Thus, it can be seen for example that the targeted MD with the lowest force constant $\kappa=1 \cdot 10^5$ kJmol⁻¹nm⁻² (Figure 6.5a) does not fully reach the open state characteristic values for all CVs, particularly the ECD bottom spread, the M2(+)-M1(-) distance and the M1 kink. A good initial guess pathway for the gating transition is expected to be obtained with the smallest force constant possible, so as to minimize the artificial bias when forcing the transition, while at the same time evolving the gating CVs over the full range for the closed-to-open transition. In this sense, the targeted MD with $\kappa=2 \cdot 10^5$ kJmol⁻¹nm⁻² seems to be optimal, as it is a good compromise between evolving the gating CVs over the full duration of the targeted MD and sampling a range of gating CV values that is close to the full spectrum for the closed-to-open transition.

For this particular targeted MD simulation, the time series of the C_α root mean square deviation (RMSD) to the target structure as well as the gating CVs for all five subunits/chains ($7 \times 5=35$ CVs) are depicted in Figure 6.6. From explicitly analyzing the CVs per subunit, it becomes clear that the gating transition obtained

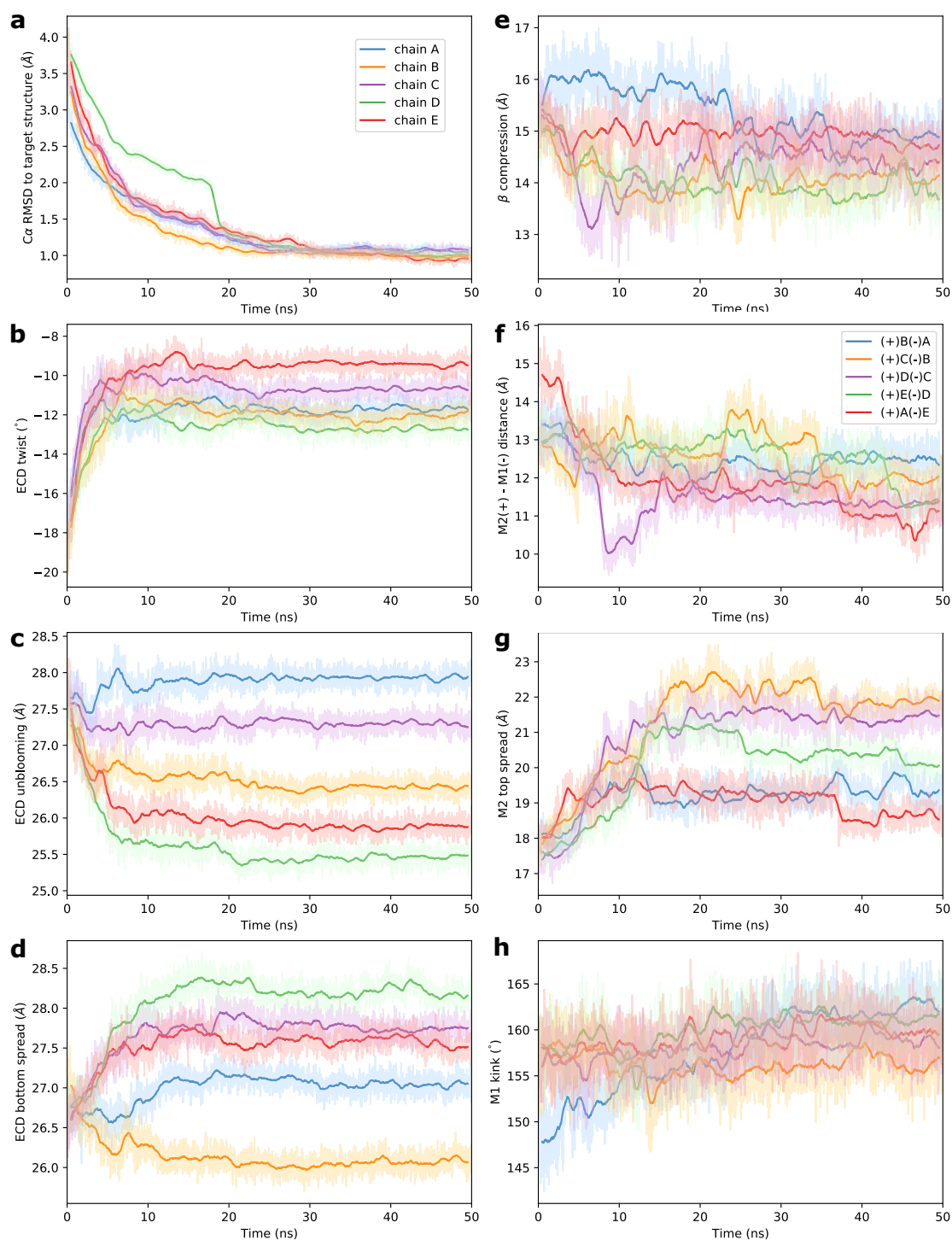


Figure 6.6: C_α RMSD to target structure and gating CVs for all five subunits over the course of 50 ns of targeted MD with force constant $\kappa=200,000 \text{ kJmol}^{-1}\text{nm}^{-2}$. (a) C_α RMSD to target structure, (b) ECD twist, (c) ECD top spread, (d) ECD bottom spread, (e) β expansion, (f) M2(+) - M1(-) distance, (g) M2 top spread, (h) M1 kink. The legends for the five subunit colours in (a) applies to all other panels, apart from panel (f), which has its own legend. Darker colours represent the running mean over 100 frames, lighter colours represent the (fluctuating) real-time values.

via targeted MD is not a concerted process. This is also what is observed for the GLIC transition obtained via the string method with swarms of trajectories by Lev et al. [111]. At the end of the simulations, the receptor structure is within ca. 1 Å C_α RMSD distance to the open state structure and visual inspection shows that it indeed has the structural features of the open state, most prominently evidenced via the outward rotation of the L9' gate residues into their hydrophobic pockets allowing for full pore hydration and thus ion permeation as described in Chapter 3. This trajectory was therefore chosen as an initial guess path for the closed-to-open transition. Unbiased MD simulations from 285 trajectory frames have been run for 300 ns each and the resulting trajectories of an accumulated simulation time of 85.5 μ s were used to build a MSM in order to obtain an unbiased equilibrium free energy surface of the gating pathway, which is described in the following section.

6.3.2 Insights into the Gating Mechanism from Markov State Modelling

While the initial guess pathway for the gating mechanism obtained by targeted MD is not exact, as the energetics are not correct, it still gives a tentative pathway that can be further refined. Here, this has been done via unbiased MD simulations initiated along the initial guess pathway, which were then reconciled via Markov state modelling. Thus, the trajectory data was reweighted in a rigorous way, such that the equilibrium distribution as well as the kinetics of the full gating pathway are recovered.

The Free Energy Surface of the Gating Pathway

Using the MSM, a reweighted free energy surface can be generated that corresponds to the equilibrium distribution (Figure 6.7). Minima correspond to metastable states. Structural analysis reveals that the part to the right of the encircled transition region corresponds to states with a closed channel, while the other half corresponds to states with an open channel pore. The closed state of the starting structure used for targeted MD has coordinates (IC1,IC2)=(1.3,-0.8). On the way from this starting point towards the transition region, the gating CVs characterizing the conformational

changes in the ECD progress (activated ECD), while the channel pore in the TMD is still shut. The global minimum corresponds to such an intermediate shut state with an activated ECD. The closed-to-open transition is structurally characterized by a fourth L9' gate residue rotating out of the pore centre into its hydrophobic pocket. The region corresponding to the open state can be structurally further subdivided into an open state where all five L9' residues are accommodated in their hydrophobic pockets or whether one of them is still oriented towards the pore centre. Surprisingly, the latter is lower in energy than the former. However, this could be an artefact due to insufficient sampling of the transition region between these two states.

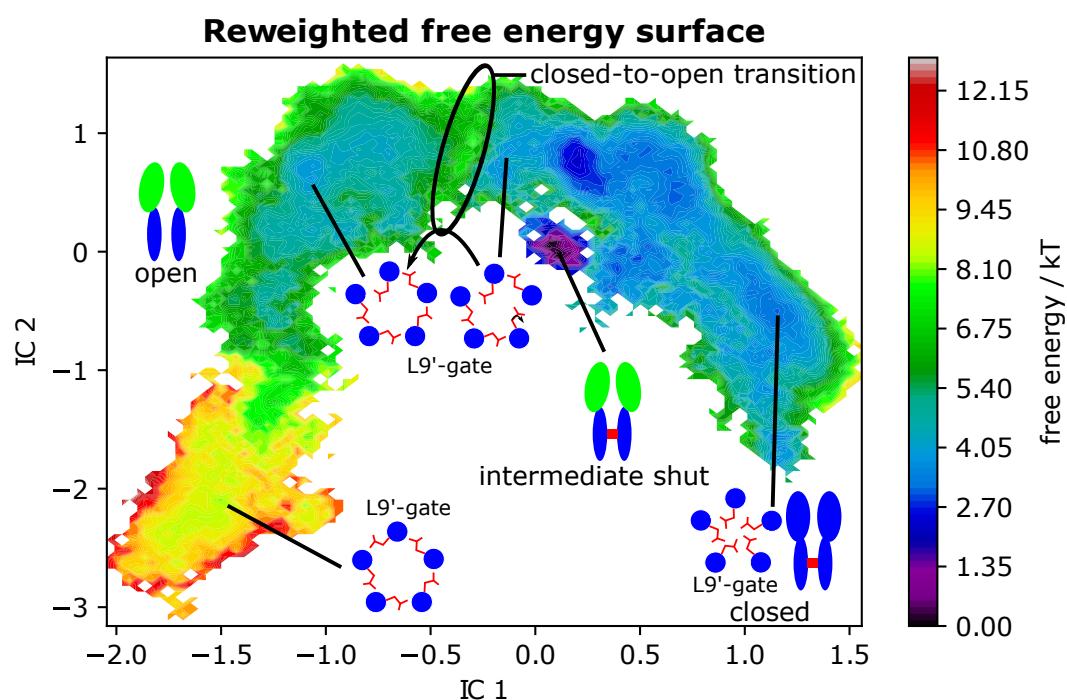


Figure 6.7: Reweighted free energy surface of the gating process in TICA space according to the MSM. Minima correspond to metastable states. The region right of the closed-to-open transition region corresponds to a closed channel, while the region left to the closed-to-open transition region corresponds to an open channel. The closed-channel states can be further subdivided into a closed state with inactivated ECD and an intermediate shut state with an activated ECD (green colour in schematic). The transition between a closed and open channel pore is structurally characterized by a fourth L9' residue rotating out of the pore into its hydrophobic pocket. The open states can be further subdivided whether all five L9' residues are rotated into their hydrophobic pockets or just four of them.

The Slowest Processes of the Gating Pathway

The information about the slowest processes in the system are encoded in the eigenvectors and corresponding eigenvalues of the transition matrix of the MSM. The eigenvalues λ_i are related to the ITS t_i via $t_i = \frac{-\tau}{\ln|\lambda_i(\tau)|}$, where τ is the lag-time of the MSM. From the inspection of the five slowest ITS (Figure 6.8a), there is a time scale separation between the first two and the third ITS, indicating the system's dynamics is governed by two slow processes between three metastable states. The corresponding eigenvectors contain the information about the probability flow through configuration space that represent these slowest processes. Their projection onto TICA space is visualized in Figure 6.8b and c. The slowest process is the transition from the closed to the open state along IC1 with an ITS of ca. $16 \pm 3 \mu\text{s}$. The 2nd slowest process is the transition between the closed state and the intermediate shut state with an ITS of ca. $7 \pm 2 \mu\text{s}$.

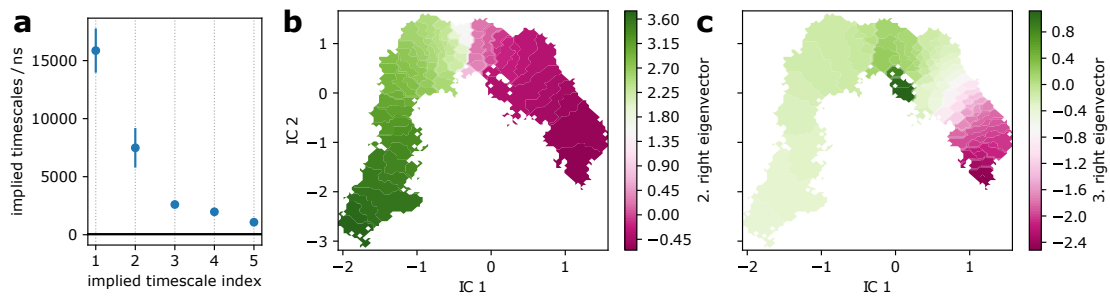


Figure 6.8: The slowest ITS as well as the 2nd and 3rd right eigenvectors of the MSM visualized in TICA space. (a) The five slowest ITS. There is clearly a separation of time scales between the first two and the other ITS. This indicates that the system's dynamics is governed by two slow processes which correspond to the 2nd and 3rd right eigenvectors of the MSM transition matrix. (b) The 2nd right eigenvector corresponds to the probability flow of the slowest process and its eigenvalue is related to the slowest ITS of ca. $16 \pm 3 \mu\text{s}$. This corresponds to the closed-to-open transition. (c) The 3rd right eigenvector corresponds to the probability flow of the 2nd slowest process and its eigenvalue is related to the 2nd slowest ITS of ca. $7 \pm 2 \mu\text{s}$. This corresponds to the transition between the closed state and the intermediate shut state.

A 3-State Gating Model

Since the kinetic analysis of the slowest processes of the gating pathway revealed a time scale separation with two slow processes between three metastable states

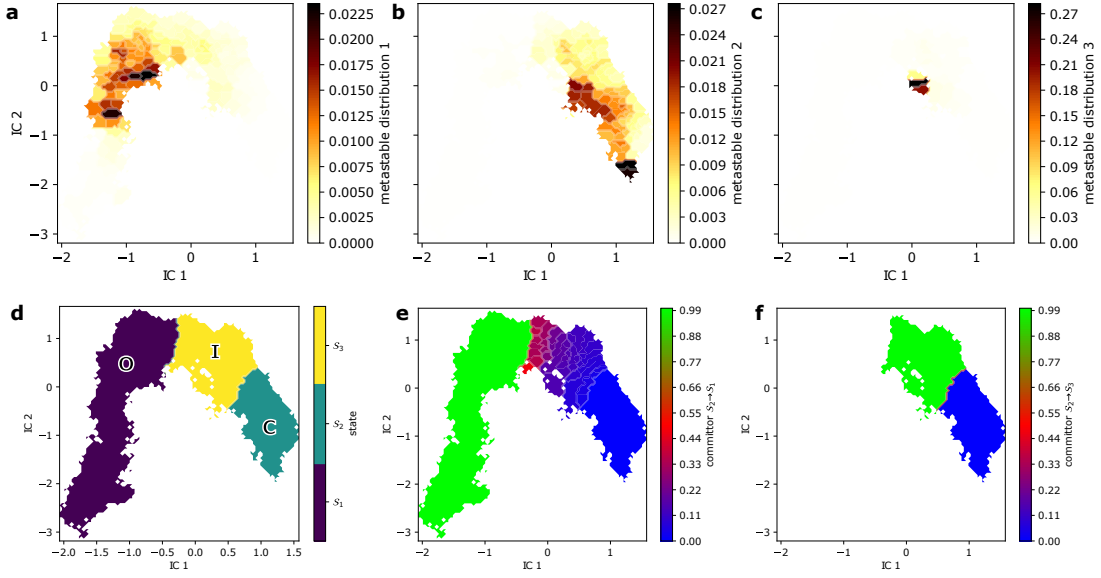


Figure 6.9: A 3-state gating model. (a-c) The result of the membership probability of the microstates using fuzzy PCCA++, given three metastable macrostates. (d) Conversion of these distributions into a crisp assignment to three metastable macrostates, resulting in a closed state (C), an intermediate shut state (I) and an open state (O). (e) The committor for the transition from the closed (C) to the open state (O) reveals the transition region around $IC1=0$ (red), which lies in the region of the intermediate shut state (I). (f) The committor for the transition from the closed (C) to the intermediate shut state (I) reveals that the committor value is constant within the metastable macrostates.

governing the system's dynamics, a fuzzy Perron cluster cluster analysis (PCCA++) was performed to compute the probability of each microstate of the MSM to belong to one of three metastable macrostates (Figure 6.9a-c). In order to obtain a crisp assignment from these distributions the argmax of each microstate's memberships was used ((Figure 6.9d).) Structurally, these three metastable states correspond to the closed (C), the intermediate shut (I) and the open state (O) (cf. Figure 6.7). This 3-state model is a coarse-grained representation of the dynamics. The stationary probabilities of the C, I and O states are 19 %, 69 % and 12 %, respectively. The mean first passage time (MFPT) of the C to O state is $93.1 \pm 72.5 \mu s$ (reverse: $28.7 \pm 14.5 \mu s$). The MFPT of the C to I state is $3.8 \pm 0.5 \mu s$ (reverse: $13.7 \pm 23.4 \mu s$). The MFPT of the I to O state is $87.1 \pm 79.7 \mu s$ (reverse: $11.3 \pm 1.8 \mu s$). Finally, a committor analysis can reveal the transition regions between the different macrostates. Given two macrostates A and B, the forward committor is defined for each microstate as the probability that being at this state, state B will be reached next, rather than

state A. The forward committor for the C and O state is depicted in Figure 6.9e. It reveals the region around $IC1=0$ as the transition region (red) for the closed-to-open transition, which lies in the region of the intermediate shut state. The forward committor for the C and I state is constant within the metastable macrostates with a sharp transition line between those states (Figure 6.9f).

6.4 Discussion

It is remarkable that the MSM predicts an intermediate shut state in the gating process, just as electrophysiological single channel recording data suggests. It is tempting to identify the intermediate shut state with the flipped state predicted from single channel recordings (cf. section 1.3.4). However, the experimentally estimated stationary occupancies for the closed, flipped and open states do not agree well with the stationary distribution of these macrostates in the coarse-grained 3-state model (see table 6.1). There are several possible explanations for this discrepancy. The

Table 6.1: Relative occupancies of closed, intermediate-shut and open state from the MSM compared with estimates fitted to the flip state model from single channel recordings [100].

State	Closed (C)	Intermediate-shut (flip) (I)	Open (O)
MSM	19 %	69 %	12 %
Single channel recordings	0.3 %	2.4 %	97.3 %

single channel recordings estimates are obtained with full length receptors, including the ICD, whereas the simulated receptor model lacks the ICD, because its structure has not been resolved thus far. Other explanations could lie in the computational modelling choices. For example the way in which macrostates were assigned based on the argmax of the membership probability. The fuzzy memberships (Figure 6.9a-c) demonstrate that the assignments of some microstates to a macrostates is not that clear. Further refinement of macrostate definitions considering structural analysis might result in better agreement with experimental data. Another important aspect is the limited quality of the MSM which can be dissected into several factors. First, it is likely that the sampling is not sufficient to reweigh the free energy surface

correctly. The quality of the MSM could greatly benefit from additional sampling of the transition regions between energy minima. Second, while best practices have been followed, there are many ways to construct a MSM and optimizing parameters such as the TICA lag time, TICA dimensions, cluster numbers and experimenting with different ways of clustering could possibly produce MSMs of better quality. Theoretically sound validation methods, such as the generalized matrix Rayleigh quotient [257] can help guiding the selection process of an optimal MSM among several models. Third, the sampled phase space is to some extent biased by the seeds used. Here, only the part of phase space can be sampled that is within reach of 300 ns unbiased MD simulation from the initial guess pathway obtained by targeted MD. Including seeds from other guess pathways as well as extending the simulation time of the individual unbiased MD simulations could reduce the risk of missing a possibly important part of phase space for the gating pathway. Finally, the accuracy of any MSM is limited by the force field. Particularly with respect to kinetics, force fields might be less accurate, as they were primarily parameterized with respect to time-independent thermodynamic properties.

Apart from improving the quality of the MSM, substantial further structural analysis is required to characterize the various metastable states corresponding to minima in the free energy surface as well as the transition states that connect them. This could then shed light into the detailed atomistic mechanism underlying the signal transduction from the ECD towards the TMD during gating. Via a flux analysis, the kinetics connecting those metastable states can be determined. To this end, it is necessary to construct a finer coarse-grained model that includes more than just three macrostates.

Future directions are ample. The accumulated simulation data could be analyzed with respect to protein-lipid interactions to see if and how lipids might play an active role in modulating the gating transition. Further, it would be interesting to extend the study to see how startle disease mutations change the free energy surface of the gating pathway. One could also analyze the effect of adding allosteric modulators, such as ivermectin. By studying systems with different partial agonists,

one could investigate the atomistic origin of partial agonism. Do partial agonists elicit entirely different gating pathways or do they just reroute the flux along the same intermediate states? Finally, the approach could be extended to other members of the pLGIC superfamily to see if and how the gating mechanisms might differ.

6.5 Conclusions

In this study, I used targeted MD to generate an initial guess pathway for the gating process of the human $\alpha 1$ GlyR. Using established gating CVs for the pLGIC superfamily, it could be demonstrated how targeted MD can indeed capture the full gating transition. In order to obtain the correct equilibrium distribution, 285 independent unbiased MD simulations were initiated from seeds along the gating pathway and reconciled via Markov state modelling. The accumulated simulation data of 85.5 μ s constitutes the largest atomistic MD study performed for pLGICs thus far. The free energy surface of the gating pathway revealed several minima corresponding to metastable states. Structural analysis in combination with PCCA revealed a coarse 3-state model of gating consisting of a closed state, an intermediate shut state and an open state, in line with the flip model from single channel recordings. Quantitative discrepancies suggest that additional sampling is needed to improve the quality of the MSM. The slowest process of the system could be assigned to the closed-to-open state transition, while the second slowest process corresponds to the transition between the closed and the intermediate shut state. The latter shows structural characteristics of an activated ECD, while the channel pore is still closed. The main free energy barrier in the closed-to-open transition seems to be the rotation of a fourth L9' residue into its hydrophobic pocket. The open state can be further subdivided in a metastable state with all five L9' residues locked into their hydrophobic pockets and another where one L9' residue is still oriented towards the pore centre. Overall, the structural analysis suggests a non-concerted mechanism for the gating process. Further structural analysis is required to elucidate the molecular details of how the signal is transduced from the ECD to the ion channel pore during the gating process.

Have no fear of perfection - you'll never reach it.

— Salvador Dali

7

Conclusions

Contents

7.1	A Refined Open State of the Glycine Receptor	200
7.1.1	Contributions	200
7.1.2	Impact	201
7.1.3	The MD-refined Open State in Light of New Structures	202
7.2	Protein-Lipid Interactions in a Neuronal Membrane	203
7.2.1	Contributions	203
7.2.2	Impact	204
7.3	Agonist Binding in the Glycine Receptor	204
7.3.1	Contributions	204
7.3.2	Impact	205
7.4	Towards an Atomistic Gating Mechanism	206
7.4.1	Contributions	206
7.4.2	Impact	207

The solution of the first full length receptor structures for the GlyR in closed, open and desensitized states at the end of 2015 [60, 68] provided the starting points for the exploration of this pLGIC via MD simulations. Moreover, from an experimental perspective the GlyR is an excellent model system. It has been extensively characterized by mutagenesis studies and is particularly amenable to single channel recordings, which can separate out the reaction steps in the activation mechanism with the kinetic landscape that connects them. In this thesis, homology models of the human $\alpha 1$ GlyR have been constructed based on the closed human

$\alpha 3$ GlyR (PDB: 5CFB [60]) and the open zebrafish $\alpha 1$ GlyR (PDB: 3JAE [68]) structures as templates and studied with MD simulations. This model system was chosen due to its relevance as a potential drug target in humans and because it allows direct comparison to extensive experimental data from single channel recordings. Ultimately, insights into GlyR function can be transferred to a certain extent to the whole pLGIC superfamily, as all members of this class share a similar overall architecture and there is growing evidence for a similar activation mechanism.

In this thesis, several aspects relevant to GlyR function were investigated which required zooming in into different regions of the GlyR structure and using a diverse set of computational simulation techniques. In Chapter 3, the attention was on the ion channel pore in order to address the question of the true physiological open state. In Chapter 4, the outer membrane accessible surface of the TMD was the focus of interest in order to characterize its interactions with its natural lipid environment in a neuronal membrane. In Chapter 5, the focal point was the orthosteric binding site for the exploration of the detailed atomistic interactions of various agonists with the receptor. In Chapter 6, structural and kinetic insights about the gating mechanism are obtained. This Chapter recapitulates the specific results of Chapters 3, 4, 5, and 6, highlights the contributions to the field and points out their potential impact.

7.1 A Refined Open State of the Glycine Receptor

7.1.1 Contributions

In Chapter 3, the so far unanswered question of the true open state structure of the GlyR has been addressed. Recent work has questioned the physiological relevance of the open state structure, since it rapidly collapses in molecular dynamics simulations.

Here, I show that the collapse can be avoided by a careful equilibration protocol that reconciles the more problematic regions of the original density map and gives a stable open state that shows frequent selective chloride permeation.

By revisiting the original experimental density map, I demonstrate that its quality in the key gating regions is not sufficient to unambiguously fit side chain

orientations into it. The simulation data suggests that the -2' proline residues attain an asymmetric configuration that can withstand a hydrophobic collapse at the -2' ring. Moreover, the -2' constriction point is slightly narrower than in the atomistic model fitted into the cryo-EM structure, which is in closer agreement with the actual underlying density map and also with the conformation of other open structures of the pLGIC superfamily. The lower region of the M2 helices is highly dynamic, suggesting that many different asymmetric conformations are accessible at physiological temperatures which would give rise to a structural ensemble which could explain the low resolution in the density map in this region. Most importantly, I argue that the side chain orientations of the 9' leucine residues, long shown to be key components of the gate, have been fitted in wrong, and that they are in fact accommodated in previously unreported hydrophobic pockets. I show that their wrong orientation has been the reason for the collapse of the open state structure in molecular simulations so far. Moreover, using a sequence alignment I reveal that the residues that constitute these hydrophobic pockets are functionally conserved across the whole superfamily of pLGICs. This suggests that locking the 9' residues in these hydrophobic pockets is a crucial part of the gating mechanism for this receptor class.

7.1.2 Impact

This study provides a crucial correction to the open state structure published in Nature in 2015 [68]. It is of importance to the whole receptor superfamily as it provides a simulation protocol that overcomes the instability of open-like structures which solves a longstanding problem in the molecular dynamics simulation community and now provides a means to study open-like structures via simulations.

A corrected atomistic description of the active state as provided here is highly relevant for future structure-based drug design for a large range of treatments for neurological diseases and addictive disorders. The success of structure-based drug design is dependent on knowing the correct structure and, given the structural conservation among the pLGICs, this refined model will be a valuable new template for the active state, and thus highly relevant to the pharmacology community.

Furthermore, these results will be of much broader interest to the whole structural biology community, because it demonstrates how current protocols of atomic model fitting to cryoelectron microscopy or X-ray crystallography density maps can lead to wrong atomic models in regions of poor electron density, advocating the use of more sophisticated molecular dynamics based flexible fitting methods already at the model building stage. This is an area that is clearly going to become increasingly prevalent with the ongoing revolution in cryo-EM.

7.1.3 The MD-refined Open State in Light of New Structures

During the write up of this thesis new GlyR open state structures were reported on BioRxiv. Kumar et al. [321] could capture a structure with an open pore geometry of the GlyR (zebrafish $\alpha 1$ homopentamer) with glycine bound in the orthosteric binding sites as well as the channel blocker picrotoxin bound in the channel pore. While the pore geometry corresponds to an ion-permeable pore, the solved structure had density of a picrotoxin molecule between the 2' and 9' position which would prevent ion permeation. However, MD simulations with picrotoxin removed showed that the pore was permeable to chloride ions. The orientation of the 9'L residues as shown in Figure 1 of [321] are rotated further away from the luminal pore into their hydrophobic pockets, as predicted by MD-refinement reported in Chapter 3. Moreover, the radius of the -2'P constriction is almost ca. 1.5 Å narrower than for the 3JAE open state from Du et al. [68], which is in better agreement with the MD-refined structure, which also predicts a ca. 1 Å narrower radius at the 2'P position. Moreover, a recent taurine bound open state GlyR structure (zebrafish $\alpha 1$ homopentamer) from Yu et al. [322] (Figure 2 therein) also supports the view that the 9'L residues are directed further away from the pore and that the -2'P constriction also appears to be narrower than the 3JAE open state structure. Thus, it appears that the MD-refined open state model was correct in predicting the 9'L residues being rotated further away from the pore

into hydrophobic pockets, while the -2'P' constriction is narrower than originally thought based on the 3JAE open state structure.

7.2 Protein-Lipid Interactions in a Neuronal Membrane

7.2.1 Contributions

In Chapter 4, the protein-lipid interactions of the GlyR in its native membrane environment are characterized.

For the first time, I have embedded this receptor in a realistic model of a neuronal membrane and performed coarse-grained molecular dynamics simulations.

I find that the lipid environment as well as the lipid-protein interaction strength are dependent on the receptor state, suggesting that lipids play a crucial role in stabilizing either the active or inactive state. By comparing the simulation results to existing structural data of the pLGIC superfamily, known lipid binding sites are confirmed, but also so far unreported new protein-lipid interactions are predicted. In particular, I characterize previously unknown phospholipid interactions at the communication interface between the extracellular and transmembrane domain that vary between the inactive and active state, suggesting that phospholipids might play a modulatory role in the activation process of the receptor. Moreover, in the active receptor state, cholesterol can bind to the binding site of the positive allosteric modulator ivermectin, indicating that cholesterol might have positive modulatory function. Thus, the hypothesis that endogenous lipids might play a modulatory role is substantiated and structural insight is provided as to how cholesterol could influence receptor function in the pLGIC class, as it was found to be crucial experimentally for the nAChR. Finally, a binding site for the anionic lipid DOPS is identified at the intracellular leaflet half of the TMD on the surface of the M4 and M3 helices which roughly coincides with the location of PIP2 binding in the human GABA_AR, yet the details of the headgroup interactions are different, mainly due to differences in the receptor's secondary structure in this region.

7.2.2 Impact

By drawing a detailed picture of the protein-lipid interactions, I elucidate the structural basis for allosteric modulation possibilities which can be exploited for the rational design of lipid-like drugs.

This has important ramifications, because existing drugs for pLGICs suffer from side effects due to poor subtype selectivity, i.e. the same drug molecule does not only bind to its desired target receptor, but also to other receptor subtypes where it creates side effects. For rational drug design purposes, the neurotransmitter binding sites (orthosteric sites) are the natural target. The problem is that they are highly conserved, i.e. the amino acids that line these sites have been preserved by evolution and are extremely similar across receptor subtypes. Hence, the same drug molecule is likely to bind to the orthosteric site of the desired receptor subtype, but also to other receptor subtypes where it would provoke undesired side effects. Targeting allosteric (non-orthosteric) modulator sites has great potential to overcome this problem for the compelling reason that they exhibit more sequence diversity between receptor subtypes.

Many allosteric sites are located at the lipid-accessible regions of the TMD. It has therefore been hypothesized that lipids themselves can act as allosteric modulators by interacting with these sites by modifying the free energy landscape between functional states. Hence, it is a prerequisite to the understanding of allosteric modulation to first characterize the protein-lipid interactions in the natural membrane environment. This forms the basis for the rational design of lipid-like drug molecules.

7.3 Agonist Binding in the Glycine Receptor

7.3.1 Contributions

Chapter 5 is concerned with the interactions of various agonists in the orthosteric binding site of the active conformation of the GlyR.

From the binding mode of the natural agonist glycine, key residues in the orthosteric binding site could be functionally annotated. The binding mode found by simulations is in excellent agreement with the structurally resolved binding mode of glycine in the PDB: 5TIN structure [56] as well as with mutagenesis data.

The partial agonists D-alanine, sarcosine, β -alanine and taurine share very similar binding features with glycine. They all face the principal subunit with the positively charged ammonium group and the complementary subunit with their negatively charged carboxylate or sulfonium group. Moreover, all binding modes include a stable water molecule in the binding pocket. However, small differences to the optimal glycine binding mode give rise to less optimal receptor interactions, such that the partial agonists contribute less efficiently towards a closed loop C conformation. This provides a structural explanation for faster dissociation rates of partial agonists compared to glycine as estimated from single channel recordings.

All agonists have a very localized effect on the receptor dynamics which is focused on loop C, which covers the binding site like a lid and thus shields it from bulk water. The closed loop C conformation is stabilized by the presence of agonists compared to a much more dynamic loop C in the apo state, in agreement with experimental data as well as MD simulations obtained from other members of the pLGIC superfamily.

Finally, binding free energies for agonists to the active conformation of GlyR have been determined via alchemical free energy calculations. Where affinity estimates are available from single channel recordings, the computational results are in line with them within error. However, the relative affinity differences between partial agonists based on available experimental data are subtle and within the accuracy limit of the computational results. Thus, it remains difficult to delineate the contributions of affinity to efficacy.

7.3.2 Impact

The detailed insight into the binding modes of partial agonists at an atomistic level of resolution contributes to the molecular level of understanding for partial agonism as long as experimentally obtained structural data remains elusive.

Indeed, such an understanding could be exploited for rational drug design. Ideally, one would like to fine-tune the response elicited by the therapeutic molecule, such that it is somewhere in between that of a full agonist and a silent antagonist, depending on the clinical situation. Apart from GlyR, other members of the pLGIC class are also potential targets for drugs with partial agonist properties, as exemplified by the smoking cessation drug varenicline which acts on the nAChR.

If GlyR structures with partial agonists will be resolved in the future, the location of the stable water binding site identified in this work could aid in the modelling process in order to distinguish if density stems from this water or from the ligand.

Furthermore, because the water site in the binding pocket is also occupied in the apo state, this opens up a compelling way to design drug molecules with groups that can replace this stable water, which would lead to additional affinity gains, both entropy-wise by releasing the confined water as well as enthalpy-wise by favourable interactions of the additional engineered group of the drug compound with the receptor.

7.4 Towards an Atomistic Gating Mechanism

7.4.1 Contributions

In Chapter 6, insights about the gating mechanism of the GlyR are obtained via a combination of targeted MD and Markov state modelling.

Targeted MD is employed to create a tentative pathway for the gating process. By using established CVs for the pLGIC superfamily as progress parameters for the gating process, it could be demonstrated that targeted MD can indeed capture the full gating transition. To refine this tentative pathway, 285 unbiased MD simulations of 300 ns length each were initiated from starting structures along the gating pathway. This resulted in an accumulated simulation time of 85.5 μ s, constituting the largest atomistic MD study performed for pLGICs thus far. From this unbiased simulation data a MSM was constructed, thus recovering the equilibrium distribution and kinetics of the gating process. Structural analysis paired with MSM-analysis techniques revealed that the system's dynamics is dominated by two slowest processes

between three metastable states, where the slowest process corresponds to the transition of the closed to the open state and the second slowest process to the transition from the closed into an intermediate shut state. This model is in qualitative agreement with the flip model predicted from single channel recording data. However, quantitative discrepancies point to the fact that the MSM is not fully converged and additional sampling is needed to improve its quality. Structurally, the intermediate shut state shows characteristics of an activated ECD with a shut TMD. The main free energy barrier in the closed-to-open transition corresponds to the rotation of a fourth L9' residue into its hydrophobic pocket. The open state can be further subdivided into a state with all five L9' residues being accommodated in their hydrophobic pockets and a state where one of them is still oriented towards the pore centre. Overall, the analysis suggests a non-concerted mechanism for the gating process. Further structural analysis will be required to elucidate the atomistic details of how the signal is transduced from the neurotransmitter binding site to the ion channel pore.

7.4.2 Impact

An understanding of the gating process at an atomistic level of resolution opens up new avenues for drug design, as intermediate states can be detected that have not been characterized by structural methods so far, which can serve as novel drug targets. Metastable structures could also help guiding the classification process when analyzing cryo-EM data and thus improve structure resolution. Furthermore, it could help rationalize why certain mutations lead to diseases such as startle disease. Due to the overall structural conservation across the pLGIC superfamily, a common gating mechanism is likely, which is further substantiated by functional chimeras consisting of domains from different members of the pLGIC. Thus, results obtained for the GlyR as a model system are probably transferable to some extent.

This study showcases that the simulation overhead associated with studying the full gating process with atomistic MD is now within reach of current state of the art supercomputing. It is thus likely that further questions involving the full gating

pathway of pLGICs can be addressed in the near future, such as how allosteric modulators influence receptor function or what gives rise to the phenomenon of partial agonism at a molecular level.

References

- [1] C. Molnar and J. Gair. *Concepts of biology: 1st Canadian Edition*. Victoria, B.C.: BCcampus. Retrieved from <https://opentextbc.ca/biology/>. 2019.
- [2] E.R. Kandel et al. *Principles of neural science*. 5th. The McGraw-Hill Companies, Inc., 2013.
- [3] B. Pakkenberg et al. “Aging and the human neocortex”. In: *Exp Gerontol* 38.1-2 (2003), pp. 95–9.
- [4] S. Herculano-Houzel. “The human brain in numbers: A linearly scaled-up primate brain”. In: *Front Hum Neurosci* 3 (2009), p. 31.
- [5] S. Maeda et al. “Structure of the connexin 26 gap junction channel at 3.5 Å resolution”. In: *Nature* 458.7238 (2009), pp. 597–602.
- [6] S. P. H. Alexander et al. “The concise guide to pharmacology 2019/20: Ion channels”. In: *Br J Pharmacol* 176 Suppl 1 (2019), S142–S228.
- [7] A. Nemecz et al. “Emerging molecular mechanisms of signal transduction in pentameric ligand-gated ion channels”. In: *Neuron* 90.3 (2016), pp. 452–70.
- [8] M. Cecchini and J. P. Changeux. “The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions allosteric modulation”. In: *Neuropharmacology* 96.Pt B (2015), pp. 137–49.
- [9] A. A. Jensen et al. “Neuronal nicotinic acetylcholine receptors: Structural revelations, target identifications, and therapeutic inspirations”. In: *J Med Chem* 48.15 (2005), pp. 4705–45.
- [10] C. Gotti, M. Zoli, and F. Clementi. “Brain nicotinic acetylcholine receptors: Native subtypes and their relevance”. In: *Trends Pharmacol Sci* 27.9 (2006), pp. 482–91.
- [11] E. B. Van Arnam and D. A. Dougherty. “Functional probes of drug-receptor interactions implicated by structural studies: Cys-loop receptors provide a fertile testing ground”. In: *J Med Chem* 57.15 (2014), pp. 6289–300.
- [12] M. Zoli, F. Pistillo, and C. Gotti. “Diversity of native nicotinic receptor subtypes in mammalian brain”. In: *Neuropharmacology* 96.Pt B (2015), pp. 302–11.
- [13] D. Bertrand. “Neuronal nicotinic acetylcholine receptors and epilepsy”. In: *Epilepsy Curr* 2.6 (2002), pp. 191–193.
- [14] O. K. Steinlein. “Ion channel mutations in neuronal diseases: A genetics perspective”. In: *Chem Rev* 112.12 (2012), pp. 6334–52.
- [15] S. M. Sine. “End-plate acetylcholine receptor: structure, mechanism, pharmacology, and disease”. In: *Physiol Rev* 92.3 (2012), pp. 1189–234.

- [16] H. Lin et al. "Cortical parvalbumin GABAergic deficits with alpha7 nicotinic acetylcholine receptor deletion: implications for schizophrenia". In: *Mol Cell Neurosci* 61 (2014), pp. 163–75.
- [17] M. R. Picciotto and M. Zoli. "Neuroprotection via nAChRs: The role of nAChRs in neurodegenerative disorders such as Alzheimer's and Parkinson's disease". In: *Front Biosci* 13 (2008), pp. 492–504.
- [18] A. S. Valles, M. V. Borroni, and F. J. Barrantes. "Targeting brain alpha7 nicotinic acetylcholine receptors in Alzheimer's disease: Rationale and current status". In: *CNS Drugs* 28.11 (2014), pp. 975–87.
- [19] F. Picard et al. "Mutated nicotinic receptors responsible for autosomal dominant nocturnal frontal lobe epilepsy are more sensitive to carbamazepine". In: *Epilepsia* 40.9 (1999), pp. 1198–209.
- [20] M. O. Ortells and G. E. Barrantes. "Molecular modelling of the interactions of carbamazepine and a nicotinic receptor involved in the autosomal dominant nocturnal frontal lobe epilepsy". In: *Br J Pharmacol* 136.6 (2002), pp. 883–95.
- [21] M. De Biasi and J. A. Dani. "Reward, addiction, withdrawal to nicotine". In: *Annu Rev Neurosci* 34 (2011), pp. 105–30.
- [22] S. Lam and P. N. Patel. "Varenicline: A selective alpha4beta2 nicotinic acetylcholine receptor partial agonist approved for smoking cessation". In: *Cardiol Rev* 15.3 (2007), pp. 154–61.
- [23] A. J. Thompson and S. C. Lummis. "5-HT₃ receptors". In: *Curr Pharm Des* 12.28 (2006), pp. 3615–30.
- [24] S. C. Lummis. "5-HT(3) receptors". In: *J Biol Chem* 287.48 (2012), pp. 40239–45.
- [25] M. Fayyaz and J. M. Lackner. "Serotonin receptor modulators in the treatment of irritable bowel syndrome". In: *Ther Clin Risk Manag* 4.1 (2008), pp. 41–8.
- [26] J. Walstab, G. Rappold, and B. Niesler. "5-HT(3) receptors: Role in disease and target of drugs". In: *Pharmacol Ther* 128.1 (2010), pp. 146–69.
- [27] T. K. Machu. "Therapeutics of 5-HT₃ receptor antagonists: Current uses and future directions". In: *Pharmacol Ther* 130.3 (2011), pp. 338–47.
- [28] P. A. Davies et al. "A novel class of ligand-gated ion channel is activated by Zn²⁺". In: *J Biol Chem* 278.2 (2003), pp. 712–7.
- [29] S. M. Trattnig et al. "Copper and protons directly activate the zinc-activated channel". In: *Biochem Pharmacol* 103 (2016), pp. 109–17.
- [30] Y. Chang et al. "Stoichiometry of a recombinant GABA_A receptor". In: *J Neurosci* 16.17 (1996), pp. 5415–24.
- [31] W. Sieghart and G. Sperk. "Subunit composition, distribution and function of GABA(A) receptor subtypes". In: *Curr Top Med Chem* 2.8 (2002), pp. 795–816.
- [32] Q. A. Nguyen and R. A. Nicoll. "The GABA_A receptor beta subunit is required for inhibitory transmission". In: *Neuron* 98.4 (2018), 718–725 e3.
- [33] S. Braat and R. F. Kooy. "The GABA_A receptor as a therapeutic target for neurodevelopmental disorders". In: *Neuron* 86.5 (2015), pp. 1119–30.

- [34] R. L. Macdonald, J. Q. Kang, and M. J. Gallagher. "Mutations in GABAA receptor subunits associated with genetic epilepsies". In: *J Physiol* 588.Pt 11 (2010), pp. 1861–9.
- [35] R. J. Howard et al. "Alcohol-binding sites in distinct brain proteins: The quest for atomic level resolution". In: *Alcohol Clin Exp Res* 35.9 (2011), pp. 1561–73.
- [36] J. M. Fritschy and P. Panzanelli. "GABAA receptors and plasticity of inhibitory neurotransmission in the central nervous system". In: *Eur J Neurosci* 39.11 (2014), pp. 1845–65.
- [37] J. Simon et al. "Analysis of the set of GABA(A) receptor genes in the human genome". In: *J Biol Chem* 279.40 (2004), pp. 41422–35.
- [38] J. W. Lynch. "Native glycine receptor subtypes and their physiological roles". In: *Neuropharmacology* 56.1 (2009), pp. 303–9.
- [39] Z. Yang et al. "Stoichiometry and subunit arrangement of alpha1beta glycine receptors as determined by atomic force microscopy". In: *Biochemistry* 51.26 (2012), pp. 5229–31.
- [40] P. Legendre. "The glycinergic inhibitory synapse". In: *Cell Mol Life Sci* 58.5-6 (2001), pp. 760–93.
- [41] J. W. Lynch. "Molecular structure and function of the glycine receptor chloride channel". In: *Physiol Rev* 84.4 (2004), pp. 1051–95.
- [42] S. Dutertre, C. M. Becker, and H. Betz. "Inhibitory glycine receptors: An update". In: *J Biol Chem* 287.48 (2012), pp. 40216–23.
- [43] D. I. Perkins et al. "Molecular targets and mechanisms for ethanol action in glycine receptors". In: *Pharmacol Ther* 127.1 (2010), pp. 53–65.
- [44] R. J. Harvey et al. "The genetics of hyperekplexia: More than startle!" In: *Trends Genet* 24.9 (2008), pp. 439–47.
- [45] A. Bode and J. W. Lynch. "The impact of human hyperekplexia mutations on glycine receptor structure and function". In: *Mol Brain* 7 (2014), p. 2.
- [46] C. F. Burgos, G. E. Yevenes, and L. G. Aguayo. "Structure and pharmacologic modulation of inhibitory glycine receptors". In: *Mol Pharmacol* 90.3 (2016), pp. 318–25.
- [47] J. W. Lynch et al. "Glycine receptor drug discovery". In: *Adv Pharmacol* 79 (2017), pp. 225–253.
- [48] A. J. Wolstenholme. "Glutamate-gated chloride channels". In: *J Biol Chem* 287.48 (2012), pp. 40232–8.
- [49] A. Tasneem et al. "Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels". In: *Genome Biol* 6.1 (2005), R4.
- [50] R. J. Hilf and R. Dutzler. "A prokaryotic perspective on pentameric ligand-gated ion channel structure". In: *Curr Opin Struct Biol* 19.4 (2009), pp. 418–24.
- [51] R. N. Beech and C. Neveu. "The evolution of pentameric ligand-gated ion-channels and the changing family of anthelmintic drug targets". In: *Parasitology* 142.2 (2015), pp. 303–17.

- [52] M. Jaiteh, A. Taly, and J. Henin. “Evolution of pentameric ligand-gated ion channels: Pro-loop receptors”. In: *PLoS One* 11.3 (2016), e0151934.
- [53] A. Gharpure et al. “Agonist selectivity and ion permeation in the alpha3beta4 ganglionic nicotinic receptor”. In: *Neuron* 104.3 (2019), 501–511 e6.
- [54] S. Basak et al. “Cryo-EM reveals two distinct serotonin-bound conformations of full-length 5-HT3A receptor”. In: *Nature* 563.7730 (2018), pp. 270–274.
- [55] D. Laverty et al. “Cryo-EM structure of the human alpha1beta3gamma2 GABAA receptor in a lipid bilayer”. In: *Nature* 565.7740 (2019), pp. 516–520.
- [56] X. Huang et al. “Crystal structures of human glycine receptor alpha3 bound to a novel class of analgesic potentiators”. In: *Nat Struct Mol Biol* 24.2 (2017), pp. 108–113.
- [57] T. Althoff et al. “X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors”. In: *Nature* 512.7514 (2014), pp. 333–7.
- [58] H. Hu et al. “Electrostatics, proton sensor, and networks governing the gating transition in GLIC, a proton-gated pentameric ion channel”. In: *Proc Natl Acad Sci U S A* 115.52 (2018), E12172–E12181.
- [59] R. J. Hilf and R. Dutzler. “X-ray structure of a prokaryotic pentameric ligand-gated ion channel”. In: *Nature* 452.7185 (2008), pp. 375–9.
- [60] X. Huang et al. “Crystal structure of human glycine receptor-alpha3 bound to antagonist strychnine”. In: *Nature* 526.7572 (2015), pp. 277–80.
- [61] F. J. Barrantes. “Phylogenetic conservation of protein-lipid motifs in pentameric ligand-gated ion channels”. In: *Biochim Biophys Acta* 1848.9 (2015), pp. 1796–805.
- [62] T. K. Sixma and A. B. Smit. “Acetylcholine binding protein (AChBP): A secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels”. In: *Annu Rev Biophys Biomol Struct* 32 (2003), pp. 311–34.
- [63] M. Nys, D. Kesters, and C. Ulens. “Structural insights into Cys-loop receptor function and ligand recognition”. In: *Biochem Pharmacol* 86.8 (2013), pp. 1042–53.
- [64] T. Lynagh and S. A. Pless. “Principles of agonist recognition in Cys-loop receptors”. In: *Front Physiol* 5 (2014), p. 160.
- [65] S. A. Pless and L. G. Sivilotti. “A tale of ligands big and small: an update on how pentameric ligand-gated ion channels interact with agonists and proteins”. In: *Curr Opin Physiol* 2 (2019), pp. 19–26.
- [66] R. E. Hibbs and E. Gouaux. “Principles of activation and permeation in an anion-selective Cys-loop receptor”. In: *Nature* 474.7349 (2011), pp. 54–60.
- [67] R. Yu et al. “Agonist and antagonist binding in human glycine receptors”. In: *Biochemistry* 53.38 (2014), pp. 6041–51.
- [68] J. Du et al. “Glycine receptor mechanism elucidated by electron cryo-microscopy”. In: *Nature* 526.7572 (2015), pp. 224–9.
- [69] F. Safar et al. “The startle disease mutation E103K impairs activation of human homomeric alpha1 glycine receptors by disrupting an intersubunit salt bridge across the agonist binding site”. In: *J Biol Chem* 292.12 (2017), pp. 5031–5042.

- [70] O. Yoluk, E. Lindahl, and M. Andersson. “Conformational gating dynamics in the GluCl anion-selective chloride channel”. In: *ACS Chem Neurosci* 6.8 (2015), pp. 1459–67.
- [71] Z. Mohammad Hosseini Naveh et al. “Conformational changes in acetylcholine binding protein investigated by temperature accelerated molecular dynamics”. In: *PLoS One* 9.2 (2014), e88555.
- [72] P. L. Mallipeddi, S. E. Pedersen, and J. M. Briggs. “Interactions of acetylcholine binding site residues contributing to nicotinic acetylcholine receptor gating: role of residues Y93, Y190, K145 and D200”. In: *J Mol Graph Model* 44 (2013), pp. 145–54.
- [73] F. Gao et al. “Agonist-mediated conformational changes in acetylcholine-binding protein revealed by simulation and intrinsic tryptophan fluorescence”. In: *J Biol Chem* 280.9 (2005), pp. 8443–51.
- [74] O. S. Smart, J. M. Goodfellow, and B. A. Wallace. “The pore dimensions of gramicidin A”. In: *Biophys J* 65.6 (1993), pp. 2455–60.
- [75] O. S. Smart et al. “HOLE: A program for the analysis of the pore dimensions of ion channel structural models”. In: *J Mol Graph* 14.6 (1996), pp. 354–60, 376.
- [76] M. Gielen and P. J. Corringer. “The dual-gate model for pentameric ligand-gated ion channels activation and desensitization”. In: *J Physiol* 596.10 (2018), pp. 1873–1902.
- [77] Mark S.P. Sansom Oliver Beckstein Philip C. Biggin. “A hydrophobic gating mechanism for nanopores”. In: *J. Phys. Chem. B* 105 (2001), pp. 12902–12905.
- [78] P. Aryal, M. S. Sansom, and S. J. Tucker. “Hydrophobic gating in ion channels”. In: *J Mol Biol* 427.1 (2015), pp. 121–30.
- [79] O. Beckstein and M. S. Sansom. “The influence of geometry, surface character, and flexibility on the permeation of ions and water through biological pores”. In: *Phys Biol* 1.1-2 (2004), pp. 42–52.
- [80] J. L. Galzi et al. “Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic”. In: *Nature* 359.6395 (1992), pp. 500–5.
- [81] P. J. Corringer et al. “Mutational analysis of the charge selectivity filter of the $\alpha 7$ nicotinic acetylcholine receptor”. In: *Neuron* 22.4 (1999), pp. 831–43.
- [82] M. J. Gunthorpe and S. C. Lummis. “Conversion of the ion selectivity of the 5-HT(3a) receptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily”. In: *J Biol Chem* 276.24 (2001), pp. 10977–83.
- [83] A. Keramidas et al. “Cation-selective mutations in the M2 domain of the inhibitory glycine receptor channel reveal determinants of ion-charge selectivity”. In: *J Gen Physiol* 119.5 (2002), pp. 393–410.
- [84] M. Sunesen et al. “Mechanism of Cl⁻ selection by a glutamate-gated chloride (GluCl) receptor revealed through mutations in the selectivity filter”. In: *J Biol Chem* 281.21 (2006), pp. 14875–81.

- [85] G. D. Cymes and C. Grosman. “Identifying the elusive link between amino acid sequence and charge selectivity in pentameric ligand-gated ion channels”. In: *Proc Natl Acad Sci U S A* 113.45 (2016), E7106–E7115.
- [86] G. Hassaine et al. “X-ray structure of the mouse serotonin 5-HT₃ receptor”. In: *Nature* 512.7514 (2014), pp. 276–81.
- [87] B. Zuber and N. Unwin. “Structure and superorganization of acetylcholine receptor-rapsyn complexes”. In: *Proc Natl Acad Sci U S A* 110.26 (2013), pp. 10622–7.
- [88] J. M. Fritschy, R. J. Harvey, and G. Schwarz. “Gephyrin: Where do we stand, where do we go?” In: *Trends Neurosci* 31.5 (2008), pp. 257–64.
- [89] G. E. Yevenes et al. “Modulation of glycine-activated ion channel function by G-protein betagamma subunits”. In: *Nat Neurosci* 6.8 (2003), pp. 819–24.
- [90] G. Langlhofer and C. Villmann. “The intracellular loop of the glycine receptor: It’s not all about the size”. In: *Front Mol Neurosci* 9 (2016), p. 41.
- [91] S. P. Kelley et al. “A cytoplasmic region determines single-channel conductance in 5-HT₃ receptors”. In: *Nature* 424.6946 (2003), pp. 321–4.
- [92] C. Song and B. Corry. “Role of acetylcholine receptor domains in ion selectivity”. In: *Biochim Biophys Acta* 1788.7 (2009), pp. 1466–73.
- [93] J. E. Carland et al. “Mutagenic analysis of the intracellular portals of the human 5-HT_{3A} receptor”. In: *J Biol Chem* 288.44 (2013), pp. 31592–601.
- [94] D. Di Maio, B. Chandramouli, and G. Brancato. “Pathways and barriers for ion translocation through the 5-HT_{3A} receptor channel”. In: *PLoS One* 10.10 (2015), e0140258.
- [95] J. Del Castillo and B. Katz. “Interaction at end-plate receptors between different choline derivatives”. In: *Proc R Soc Lond B Biol Sci* 146.924 (1957), pp. 369–81.
- [96] D. Colquhoun. “Binding, gating, affinity and efficacy: The interpretation of structure-activity relationships for agonists and of the effects of mutating receptors”. In: *Br J Pharmacol* 125.5 (1998), pp. 924–47.
- [97] David Colquhoun, Remigijus Lape, and Lucia Sivilotti. “Chapter 8 - Single ion channels”. In: *Single molecule biology*. Ed. by Alex E. Knight. New York: Academic Press, 2009, pp. 223–251.
- [98] L. G. Sivilotti. “What single-channel analysis tells us of the activation mechanism of ligand-gated channels: the case of the glycine receptor”. In: *J Physiol* 588.Pt 1 (2010), pp. 45–58.
- [99] C. Bouzat and S. M. Sine. “Nicotinic acetylcholine receptors at the single-channel level”. In: *Br J Pharmacol* 175.11 (2018), pp. 1789–1804.
- [100] V. Burzomato et al. “Single-channel behavior of heteromeric $\alpha 1\beta$ glycine receptors: An attempt to detect a conformational change before the channel opens”. In: *J Neurosci* 24.48 (2004), pp. 10924–40.
- [101] R. Lape, D. Colquhoun, and L. G. Sivilotti. “On the nature of partial agonism in the nicotinic receptor superfamily”. In: *Nature* 454.7205 (2008), pp. 722–7.
- [102] A. Auerbach. “Gating of acetylcholine receptor channels: Brownian motion across a broad transition state”. In: *Proc Natl Acad Sci U S A* 102.5 (2005), pp. 1408–12.

- [103] P. Krashia et al. “The long activations of alpha2 glycine channels can be described by a mechanism with reaction intermediates ("flip")”. In: *J Gen Physiol* 137.2 (2011), pp. 197–216.
- [104] R. Lape et al. “The alpha1K276E startle disease mutation reveals multiple intermediate states in the gating of glycine receptors”. In: *J Neurosci* 32.4 (2012), pp. 1336–52.
- [105] N. Mukhtasimova et al. “Detection and trapping of intermediate states priming nicotinic receptor channel opening”. In: *Nature* 459.7245 (2009), pp. 451–4.
- [106] N. Mukhtasimova, C. J. daCosta, and S. M. Sine. “Improved resolution of single channel dwell times reveals mechanisms of binding, priming, and gating in muscle AChR”. In: *J Gen Physiol* 148.1 (2016), pp. 43–63.
- [107] S. Jadey and A. Auerbach. “An integrated catch-and-hold mechanism activates nicotinic acetylcholine receptors”. In: *J Gen Physiol* 140.1 (2012), pp. 17–28.
- [108] P. Purohit et al. “Catch-and-hold activation of muscle acetylcholine receptors having transmitter binding site mutations”. In: *Biophys J* 107.1 (2014), pp. 88–99.
- [109] J. Corradi, F. Gumilar, and C. Bouzat. “Single-channel kinetic analysis for activation and desensitization of homomeric 5-HT(3)A receptors”. In: *Biophys J* 97.5 (2009), pp. 1335–45.
- [110] A. J. Plested. “Don’t flip out: AChRs are primed to catch and hold your attention”. In: *Biophys J* 107.1 (2014), pp. 8–9.
- [111] B. Lev et al. “String method solution of the gating pathways for a pentameric ligand-gated ion channel”. In: *Proc Natl Acad Sci U S A* 114.21 (2017), E4158–E4167.
- [112] M. Nys et al. “Allosteric binding site in a Cys-loop receptor ligand-binding domain unveiled in the crystal structure of ELIC in complex with chlorpromazine”. In: *Proc Natl Acad Sci U S A* 113.43 (2016), E6696–E6703.
- [113] P. Rucktooa, A. B. Smit, and T. K. Sixma. “Insight in nAChR subtype selectivity from AChBP crystal structures”. In: *Biochem Pharmacol* 78.7 (2009), pp. 777–87.
- [114] J. E. Baenziger and P. J. Corringer. “3D structure and allosteric modulation of the transmembrane domain of pentameric ligand-gated ion channels”. In: *Neuropharmacology* 60.1 (2011), pp. 116–25.
- [115] T. Lynagh and J. W. Lynch. “Molecular mechanisms of Cys-loop ion channel receptor modulation by ivermectin”. In: *Front Mol Neurosci* 5 (2012), p. 60.
- [116] A. K. Hamouda et al. “Photoaffinity labeling of nicotinic receptors: Diversity of drug binding sites!” In: *J Mol Neurosci* 53.3 (2014), pp. 480–6.
- [117] R. J. Howard, J. R. Trudell, and R. A. Harris. “Seeking structural specificity: Direct modulation of pentameric ligand-gated ion channels by alcohols and general anesthetics”. In: *Pharmacol Rev* 66.2 (2014), pp. 396–412.
- [118] A. Taly et al. “Allosteric regulation of pentameric ligand-gated ion channels: An emerging mechanistic perspective”. In: *Channels (Austin)* 8.4 (2014), pp. 350–60.
- [119] J. E. Baenziger et al. “Nicotinic acetylcholine receptor-lipid interactions: Mechanistic insight and biological function”. In: *Biochim Biophys Acta* 1848.9 (2015), pp. 1806–17.

- [120] L. Sauguet, A. Shahsavar, and M. Delarue. “Crystallographic studies of pharmacological sites in pentameric ligand-gated ion channels”. In: *Biochim Biophys Acta* 1850.3 (2015), pp. 511–23.
- [121] B. A. Sparling and E. F. DiMauro. “Progress in the discovery of small molecule modulators of the Cys-loop superfamily receptors”. In: *Bioorg Med Chem Lett* 27.15 (2017), pp. 3207–3218.
- [122] Z. Fourati et al. “Structural basis for a bimodal allosteric mechanism of general anesthetic modulation in pentameric ligand-gated ion channels”. In: *Cell Rep* 23.4 (2018), pp. 993–1004.
- [123] X. Huang, H. Chen, and P. L. Shaffer. “Crystal structures of human GlyR α 3 bound to ivermectin”. In: *Structure* 25.6 (2017), 945–950 e2.
- [124] Mark Tuckerman. *Statistical mechanics: Theory and molecular simulation*. Oxford: Oxford University Press, 2010.
- [125] Michael P. Allen and Dominic J. Tildesley. *Computer simulation of liquids*. 2nd. Oxford University Press, 2017.
- [126] Daan Frenkel and Berend Smit. *Understanding molecular simulation*. 2nd. Academic Press, Inc., 2001.
- [127] Herman J. C. Berendsen. *Simulating the physical world: Hierarchical modeling from quantum mechanics to fluid dynamics*. Cambridge: Cambridge University Press, 2007.
- [128] A.R. Leach. *Molecular modelling: Principles and applications*. 2nd edition. Pearson Education. Prentice Hall, 2001.
- [129] Leonard Susskind and George Hrabovsky. *The theoretical minimum: What you need to know to start doing physics*. New York: Basic Books. A Member of the Perseus Books Group, 2013.
- [130] H. Goldstein, C.P. Poole, and J.L. Safko. *Classical mechanics*. Addison Wesley, 2002.
- [131] D. Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, 1987.
- [132] Donald A. McQuarrie. *Statistical mechanics*. University Science Books, 2000.
- [133] E. Braun et al. “Best practices for foundations in molecular simulations [Article v1.0]”. In: *Living J Comput Mol Sci* 1.1 (2019).
- [134] Christophe Chipot and Andrew Pohorille. *Free energy calculations: Theory and applications in chemistry and biology*. Springer Series in Chemical Physics. Springer-Verlag Berlin Heidelberg, 2007.
- [135] Matteo Aldeghi, Joseph P. Bluck, and Philip C. Biggin. “Absolute alchemical free energy calculations for ligand binding: A beginner’s guide”. In: *Computational Drug Discovery and Design*. Ed. by Mohini Gore and Umesh B. Jagtap. New York, NY: Springer New York, 2018, pp. 199–232.
- [136] Michael R. Shirts. “Best practices in free energy calculations for drug design”. In: *Computational Drug Discovery and Design*. Ed. by Riccardo Baron. New York, NY: Springer New York, 2012, pp. 425–467.

- [137] Michael R. Shirts, David L. Mobley, and Scott P. Brown. “Free-energy calculations in structure-based drug design”. In: *Drug Design: Structure- and Ligand-Based Approaches*. Ed. by Charles H. Reynolds, Dagmar Ringe, and Jr Kenneth M. Merz. Cambridge: Cambridge University Press, 2010, pp. 61–86.
- [138] C. Schütte and M. Sarich. *Metastability and Markov state models in molecular dynamics: Modeling, analysis, algorithmic approaches*. Vol. 24. Courant Lecture Notes. American Mathematical Society, 2013.
- [139] G.R. Bowman, V.S. Pande, and F. Noé. *An introduction to Markov state models and their application to long timescale molecular simulation*. Advances in Experimental Medicine and Biology. Springer, 2014.
- [140] Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: Basic theory and advanced methods*. Cambridge: Cambridge University Press, 2009.
- [141] M.J. Abraham et al. “GROMACS: High performance molecular simulations through multi-levelparallelism from laptops to supercomputers”. In: *SoftwareX* 1 (2015), pp. 19–25.
- [142] Michael Levitt. “The birth of computational structural biology”. In: *Nature Structural Biology* 8.5 (2001), pp. 392–393.
- [143] K. Lindorff-Larsen et al. “Improved side-chain torsion potentials for the Amber ff99SB protein force field”. In: *Proteins* 78.8 (2010), pp. 1950–8.
- [144] J. A. Maier et al. “ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB”. In: *J Chem Theory Comput* 11.8 (2015), pp. 3696–713.
- [145] C. Tian et al. “ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution”. In: *J Chem Theory Comput* (2019).
- [146] R. B. Best et al. “Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles”. In: *J Chem Theory Comput* 8.9 (2012), pp. 3257–3273.
- [147] M. M. Reif, P. H. Hunenberger, and C. Oostenbrink. “New interaction parameters for charged amino acid side chains in the GROMOS force field”. In: *J Chem Theory Comput* 8.10 (2012), pp. 3705–23.
- [148] William L. Jorgensen and Carol J. Swenson. “Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides”. In: *Journal of the American Chemical Society* 107.3 (1985), pp. 569–578.
- [149] D. H. de Jong et al. “Improved parameters for the Martini coarse-grained protein force field”. In: *J Chem Theory Comput* 9.1 (2013), pp. 687–97.
- [150] S. Riniker. “Fixed-charge atomistic force fields for molecular dynamics simulations in the condensed phase: An overview”. In: *J Chem Inf Model* 58.3 (2018), pp. 565–578.
- [151] P. Dauber-Osguthorpe and A. T. Hagler. “Biomolecular force fields: Where have we been, where are we now, where do we need to go and how do we get there?” In: *J Comput Aided Mol Des* 33.2 (2019), pp. 133–203.

- [152] P. E. Lopes, O. Guvench, and Jr. MacKerell A. D. “Current status of protein force fields for molecular dynamics simulations”. In: *Methods Mol Biol* 1215 (2015), pp. 47–71.
- [153] A. P. Lyubartsev and A. L. Rabinovich. “Force field development for lipid membrane simulations”. In: *Biochim Biophys Acta* 1858.10 (2016), pp. 2483–2497.
- [154] P. S. Nerenberg and T. Head-Gordon. “New developments in force fields for biomolecular simulations”. In: *Curr Opin Struct Biol* 49 (2018), pp. 129–138.
- [155] S. J. Marrink and D. P. Tieleman. “Perspective on the Martini model”. In: *Chem Soc Rev* 42.16 (2013), pp. 6801–22.
- [156] B. M. H. Bruininks, P. C. T. Souza, and S. J. Marrink. “A practical view of the Martini force field”. In: *Methods Mol Biol* 2022 (2019), pp. 105–127.
- [157] S. Piana, K. Lindorff-Larsen, and D. E. Shaw. “How robust are protein folding simulations with respect to force field parameterization?” In: *Biophys J* 100.9 (2011), pp. L47–9.
- [158] K. Lindorff-Larsen et al. “Systematic validation of protein force fields against experimental data”. In: *PLoS One* 7.2 (2012), e32131.
- [159] K. A. Beauchamp et al. “Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements”. In: *J Chem Theory Comput* 8.4 (2012), pp. 1409–1414.
- [160] M. Paloncova et al. “Benchmarking of force fields for molecule-membrane interactions”. In: *J Chem Theory Comput* 10.9 (2014), pp. 4143–51.
- [161] F. Vitalini, F. Noe, and B. G. Keller. “Molecular dynamics simulations data of the twenty encoded amino acids in different force fields”. In: *Data Brief* 7 (2016), pp. 582–90.
- [162] D. Poger, B. Caron, and A. E. Mark. “Validating lipid force fields against experimental data: Progress, challenges and perspectives”. In: *Biochim Biophys Acta* 1858.7 Pt B (2016), pp. 1556–65.
- [163] K. Pluhackova et al. “A critical comparison of biomembrane force fields: Structure and dynamics of model DMPC, POPC, and POPE bilayers”. In: *J Phys Chem B* 120.16 (2016), pp. 3888–903.
- [164] V. Gapsys et al. “Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan”. In: *Angew Chem Int Ed Engl* 55.26 (2016), pp. 7364–8.
- [165] J. P. Jambeck and A. P. Lyubartsev. “Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids”. In: *J Phys Chem B* 116.10 (2012), pp. 3164–79.
- [166] J. P. Jambeck and A. P. Lyubartsev. “An extension and further validation of an all-atomistic force field for biological membranes”. In: *J Chem Theory Comput* 8.8 (2012), pp. 2938–48.
- [167] J. P. Jambeck and A. P. Lyubartsev. “Another piece of the membrane puzzle: Extending Slipids further”. In: *J Chem Theory Comput* 9.1 (2013), pp. 774–84.
- [168] J. Wang et al. “Development and testing of a general amber force field”. In: *J Comput Chem* 25.9 (2004), pp. 1157–74.

- [169] A. H. Horn. “A consistent force field parameter set for zwitterionic amino acid residues”. In: *J Mol Model* 20.11 (2014), p. 2478.
- [170] William L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935.
- [171] C. Kutzner et al. “Computational electrophysiology: The molecular dynamics of ion channel permeation and selectivity in atomistic detail”. In: *Biophys J* 101.4 (2011), pp. 809–17.
- [172] Michael Bergdorf, Christine Peter, and Philippe H. Hünenberger. “Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: A continuum electrostatics study”. In: *The Journal of Chemical Physics* 119.17 (2003), pp. 9129–9144.
- [173] B. A. Wells and A. L. Chaffee. “Ewald summation for molecular simulations”. In: *J Chem Theory Comput* 11.8 (2015), pp. 3684–95.
- [174] Markus Deserno and Christian Holm. “How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines”. In: *The Journal of Chemical Physics* 109.18 (1998), pp. 7678–7693.
- [175] Tom Darden, Darrin York, and Lee Pedersen. “Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems”. In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.
- [176] Ulrich Essmann et al. “A smooth particle mesh Ewald method”. In: *The Journal of Chemical Physics* 103.19 (1995), pp. 8577–8593.
- [177] J. A. Barker and R. O. Watts. “Monte Carlo studies of the dielectric properties of water-like models”. In: *Molecular Physics* 26.3 (1973), pp. 789–792.
- [178] Ilario G. Tironi et al. “A generalized reaction field method for molecular dynamics simulations”. In: *The Journal of Chemical Physics* 102.13 (1995), pp. 5451–5459.
- [179] W. F. Van Gunsteren and Martin Karplus. “Effect of constraints on the dynamics of macromolecules”. In: *Macromolecules* 15.6 (1982), pp. 1528–1544.
- [180] Jean-Paul Ryckaert, Giovanni Cicciotti, and Herman J. C. Berendsen. “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes”. In: *Journal of Computational Physics* 23.3 (1977), pp. 327–341.
- [181] Hans C. Andersen. “Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations”. In: *Journal of Computational Physics* 52.1 (1983), pp. 24–34.
- [182] Berk Hess et al. “LINCS: A linear constraint solver for molecular simulations”. In: *Journal of Computational Chemistry* 18.12 (1997), pp. 1463–1472.
- [183] Berk Hess. “P-LINCS: A parallel linear constraint solver for molecular simulation”. In: *Journal of Chemical Theory and Computation* 4.1 (2008), pp. 116–122.
- [184] Shuichi Miyamoto and Peter A. Kollman. “Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962.

- [185] Michael P. Eastwood et al. “Equipartition and the calculation of temperature in biomolecular simulations”. In: *Journal of Chemical Theory and Computation* 6.7 (2010), pp. 2045–2058.
- [186] H. J. C. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *The Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.
- [187] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical sampling through velocity rescaling”. In: *The Journal of Chemical Physics* 126.1 (2007), p. 014101.
- [188] Giovanni Bussi and Michele Parrinello. “Stochastic thermostats: Comparison of local and global schemes”. In: *Computer Physics Communications* 179.1 (2008), pp. 26–29.
- [189] Joseph E. Basconi and Michael R. Shirts. “Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations”. In: *Journal of Chemical Theory and Computation* 9.7 (2013), pp. 2887–2899.
- [190] Stephen C. Harvey, Robert K. Z. Tan, and Thomas E. Cheatham Iii. “The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition”. In: *Journal of Computational Chemistry* 19.7 (1998), pp. 726–740.
- [191] N. Goga et al. “Efficient algorithms for Langevin and DPD dynamics”. In: *Journal of Chemical Theory and Computation* 8.10 (2012), pp. 3637–3649.
- [192] C. G. Gray and K. E. Gubbins. *Theory of molecular fluids: Fundamentals*. Vol. 1. New York: Oxford University Press, 1984.
- [193] Mary Vedamuthu, Surjit Singh, and G. Wilse Robinson. “Properties of liquid water. 4. The isothermal compressibility minimum near 50 °C”. In: *The Journal of Physical Chemistry* 99.22 (1995), pp. 9263–9267.
- [194] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *Journal of Applied Physics* 52.12 (1981), pp. 7182–7190.
- [195] Peter Tieleman et al. *Molecular simulations and biomembranes: From biophysics to function*. RSC Biomolecular Sciences. The Royal Society of Chemistry, 2010.
- [196] M. Aldeghi et al. “Accurate calculation of the absolute free energy of binding for drug molecules”. In: *Chem Sci* 7.1 (2016), pp. 207–218.
- [197] M. Aldeghi et al. “Predictions of ligand selectivity from absolute binding free energy calculations”. In: *J Am Chem Soc* 139.2 (2017), pp. 946–957.
- [198] Stefan Boresch et al. “Absolute binding free energies: A quantitative approach for their calculation”. In: *The Journal of Physical Chemistry B* 107.35 (2003), pp. 9535–9551.
- [199] D. L. Mobley, J. D. Chodera, and K. A. Dill. “On the use of orientational restraints and symmetry corrections in alchemical free energy calculations”. In: *J Chem Phys* 125.8 (2006), p. 084902.
- [200] John G. Kirkwood. “Statistical Mechanics of Fluid Mixtures”. In: *The Journal of Chemical Physics* 3.5 (1935), pp. 300–313.

- [201] Robert W. Zwanzig. “High [U+2010]temperature equation of state by a perturbation method. I. Nonpolar gases”. In: *The Journal of Chemical Physics* 22.8 (1954), pp. 1420–1426.
- [202] Charles H. Bennett. “Efficient estimation of free energy differences from Monte Carlo data”. In: *Journal of Computational Physics* 22.2 (1976), pp. 245–268.
- [203] M. R. Shirts and J. D. Chodera. “Statistically optimal analysis of samples from multiple equilibrium states”. In: *J Chem Phys* 129.12 (2008), p. 124105.
- [204] M. R. Shirts et al. “Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods”. In: *Phys Rev Lett* 91.14 (2003), p. 140601.
- [205] M. A. Dämgen and P. C. Biggin. “Computational methods to examine conformational changes and ligand-binding properties: Examples in neurobiology”. In: *Neurosci Lett* 700 (2019), pp. 9–16.
- [206] A. J. Thompson, H. A. Lester, and S. C. Lummis. “The structural basis of function in Cys-loop receptors”. In: *Q Rev Biophys* 43.4 (2010), pp. 449–99.
- [207] L. Sauguet et al. “Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels”. In: *EMBO J* 32.5 (2013), pp. 728–41.
- [208] G. Gonzalez-Gutierrez et al. “Chasing the open-state structure of pentameric ligand-gated ion channels”. In: *J Gen Physiol* 149.12 (2017), pp. 1119–1138.
- [209] J. L. Trick et al. “Functional annotation of ion channel structures by molecular simulation”. In: *Structure* 24.12 (2016), pp. 2207–2216.
- [210] A. H. Cerdan, N. E. Martin, and M. Cecchini. “An ion-permeable state of the glycine receptor captured by molecular dynamics”. In: *Structure* 26.11 (2018), 1555–1562 e4.
- [211] M. H. Cheng and R. D. Coalson. “Energetics and ion permeation characteristics in a glutamate-gated chloride (GluCl) receptor channel”. In: *J Phys Chem B* 116.46 (2012), pp. 13637–43.
- [212] N. Calimet et al. “A gating mechanism of pentameric ligand-gated ion channels”. In: *Proc Natl Acad Sci U S A* 110.42 (2013), E3987–96.
- [213] R. C. Edgar. “MUSCLE: Multiple sequence alignment with high accuracy and high throughput”. In: *Nucleic Acids Res* 32.5 (2004), pp. 1792–7.
- [214] B. Webb and A. Sali. “Comparative protein structure modeling using MODELLER”. In: *Curr Protoc Bioinformatics* 47 (2014), pp. 5 6 1–32.
- [215] M. Y. Shen and A. Sali. “Statistical potential for assessment and prediction of protein structures”. In: *Protein Sci* 15.11 (2006), pp. 2507–24.
- [216] P. Benkert, M. Kunzli, and T. Schwede. “QMEAN server for protein model quality estimation”. In: *Nucleic Acids Res* 37.Web Server issue (2009), W510–4.
- [217] T. Sterling and J. J. Irwin. “ZINC 15 - ligand discovery for everyone”. In: *J Chem Inf Model* 55.11 (2015), pp. 2324–37.
- [218] T. A. Wassenaar et al. “Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations”. In: *J Chem Theory Comput* 11.5 (2015), pp. 2144–55.

- [219] D. H. de Jong et al. “Martini straight: Boosting performance using a shorter cutoff and GPUs”. In: *Computer Physics Communications* 199 (2016), pp. 1–7.
- [220] T. A. Wassenaar et al. “Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models”. In: *J Chem Theory Comput* 10.2 (2014), pp. 676–90.
- [221] Marc A. Dämgen and Philip C. Biggin. “A refined open state of the glycine receptor obtained via molecular dynamics simulations”. In: *Structure* 28 (2020), pp. 1–10.
- [222] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. Taylor & Francis, Inc., 1988, p. 564.
- [223] G. Klesse et al. “CHAP: A versatile tool for the structural and functional annotation of ion channel pores”. In: *J Mol Biol* 431.17 (2019), pp. 3353–3365.
- [224] W. C. Wimley, T. P. Creamer, and S. H. White. “Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides”. In: *Biochemistry* 35.16 (1996), pp. 5109–24.
- [225] W. C. Wimley and S. H. White. “Experimentally determined hydrophobicity scale for proteins at membrane interfaces”. In: *Nat Struct Biol* 3.10 (1996), pp. 842–8.
- [226] N. Michaud-Agrawal et al. “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations”. In: *J Comput Chem* 32.10 (2011), pp. 2319–27.
- [227] A. Kucukelbir, F. J. Sigworth, and H. D. Tagare. “Quantifying the local resolution of cryo-EM density maps”. In: *Nat Methods* 11.1 (2014), pp. 63–5.
- [228] P. D. Adams et al. “PHENIX: A comprehensive Python-based system for macromolecular structure solution”. In: *Acta Crystallogr D Biol Crystallogr* 66.Pt 2 (2010), pp. 213–21.
- [229] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. In: *J Mol Graph* 14.1 (1996), pp. 33–8, 27–8.
- [230] E. F. Pettersen et al. “UCSF Chimera - a visualization system for exploratory research and analysis”. In: *J Comput Chem* 25.13 (2004), pp. 1605–12.
- [231] D. N. LeBard et al. “General anesthetics predicted to block the GLIC pore with micromolar affinity”. In: *PLoS Comput Biol* 8.5 (2012), e1002532.
- [232] L. Chiodo et al. “A structural model of the human alpha7 nicotinic receptor in an open conformation”. In: *PLoS One* 10.7 (2015), e0133011.
- [233] N. Bocquet et al. “X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation”. In: *Nature* 457.7225 (2009), pp. 111–4.
- [234] L. Polovinkin et al. “Conformational transitions of the serotonin 5-HT3 receptor”. In: *Nature* 563.7730 (2018), pp. 275–279.
- [235] M. Igaev et al. “Automated cryo-EM structure refinement using correlation-driven molecular dynamics”. In: *Elife* 8 (2019).
- [236] J. A. Poveda et al. “Towards understanding the molecular basis of ion channel modulation by lipids: Mechanistic models and current paradigms”. In: *Biochim Biophys Acta Biomembr* 1859.9 Pt B (2017), pp. 1507–1516.

- [237] A. L. Duncan, W. Song, and M. S. P. Sansom. “Lipid-dependent regulation of ion channels and G protein-coupled receptors: Insights from structures and simulations”. In: *Annu Rev Pharmacol Toxicol* (2019).
- [238] T. M. Fong and M. G. McNamee. “Correlation between acetylcholine receptor function and structural properties of membranes”. In: *Biochemistry* 25.4 (1986), pp. 830–40.
- [239] M. Criado, H. Eibl, and F. J. Barrantes. “Effects of lipids on acetylcholine receptor. Essential need of cholesterol for maintenance of agonist-induced state transitions in lipid vesicles”. In: *Biochemistry* 21.15 (1982), pp. 3622–9.
- [240] A. K. Hamouda et al. “Assessing the lipid requirements of the Torpedo californica nicotinic acetylcholine receptor”. In: *Biochemistry* 45.13 (2006), pp. 4327–37.
- [241] C. M. Henault et al. “The role of the M4 lipid-sensor in the folding, trafficking, and allosteric modulation of nicotinic acetylcholine receptors”. In: *Neuropharmacology* 96.Pt B (2015), pp. 157–68.
- [242] Y. H. Lee et al. “Mutations in the M4 domain of Torpedo californica acetylcholine receptor dramatically alter ion channel function”. In: *Biophys J* 66.3 Pt 1 (1994), pp. 646–53.
- [243] J. A. Lasalde et al. “Tryptophan substitutions at the lipid-exposed transmembrane segment M4 of Torpedo californica acetylcholine receptor govern channel gating”. In: *Biochemistry* 35.45 (1996), pp. 14139–48.
- [244] C. Bouzat et al. “Mutations at lipid-exposed residues of the acetylcholine receptor affect its gating kinetics”. In: *Mol Pharmacol* 54.1 (1998), pp. 146–53.
- [245] C. L. Carswell, J. Sun, and J. E. Baenziger. “Intramembrane aromatic interactions influence the lipid sensitivities of pentameric ligand-gated ion channels”. In: *J Biol Chem* 290.4 (2015), pp. 2496–507.
- [246] S. Haeger et al. “An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors”. In: *Nat Struct Mol Biol* 17.1 (2010), pp. 90–8.
- [247] D. Laverty et al. “Crystal structures of a GABAA-receptor chimera reveal new endogenous neurosteroid-binding sites”. In: *Nat Struct Mol Biol* 24.11 (2017), pp. 977–985.
- [248] Jr. Walsh R. M. et al. “Structural principles of distinct assemblies of the human $\alpha 4 \beta 2$ nicotinic receptor”. In: *Nature* 557.7704 (2018), pp. 261–265.
- [249] V. Corradi et al. “Emerging diversity in lipid-protein interactions”. In: *Chem Rev* 119.9 (2019), pp. 5775–5848.
- [250] G. Hedger and M. S. P. Sansom. “Lipid interaction sites on channels, transporters and receptors: Recent insights from molecular dynamics simulations”. In: *Biochim Biophys Acta* 1858.10 (2016), pp. 2390–2400.
- [251] L. Domicевичa, H. Koldso, and P. C. Biggin. “Multiscale molecular dynamics simulations of lipid interactions with P-glycoprotein in a complex membrane”. In: *J Mol Graph Model* 80 (2018), pp. 147–156.
- [252] A. L. Duncan et al. “Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes”. In: *Sci Rep* 7.1 (2017), p. 16647.

- [253] W. Song et al. “State-dependent lipid interactions with the A2a receptor revealed by MD simulations using in vivo-mimetic membranes”. In: *Structure* 27.2 (2019), 392–403 e3.
- [254] M. H. Cheng, Y. Xu, and P. Tang. “Anionic lipid and cholesterol interactions with alpha4beta2 nAChR: Insights from MD simulations”. In: *J Phys Chem B* 113.19 (2009), pp. 6964–70.
- [255] J. Henin et al. “A predicted binding site for cholesterol on the GABAA receptor”. In: *Biophys J* 106.9 (2014), pp. 1938–49.
- [256] H. I. Ingolfsson et al. “Computational lipidomics of the neuronal plasma membrane”. In: *Biophys J* 113.10 (2017), pp. 2271–2280.
- [257] R. T. McGibbon et al. “MDTraj: A modern open library for the analysis of molecular dynamics trajectories”. In: *Biophys J* 109.8 (2015), pp. 1528–32.
- [258] L. Sharp, R. Salari, and G. Brannigan. “Boundary lipids of the nicotinic acetylcholine receptor: Spontaneous partitioning via coarse-grained molecular dynamics simulation”. In: *Biochim Biophys Acta Biomembr* 1861.4 (2019), pp. 887–896.
- [259] E. Strandberg and J. A. Killian. “Snorkeling of lysine side chains in transmembrane helices: How easy can it get?” In: *FEBS Lett* 544.1-3 (2003), pp. 69–73.
- [260] T. Lynagh et al. “Molecular determinants of ivermectin sensitivity at the glycine receptor chloride channel”. In: *J Biol Chem* 286.51 (2011), pp. 43913–24.
- [261] Q. Shan, J. L. Haddrill, and J. W. Lynch. “Ivermectin, an unconventional agonist of the glycine receptor chloride channel”. In: *J Biol Chem* 276.16 (2001), pp. 12556–64.
- [262] C. J. daCosta et al. “Lipid-protein interactions at the nicotinic acetylcholine receptor. A functional coupling between nicotinic receptors and phosphatidic acid-containing lipid bilayers”. In: *J Biol Chem* 277.1 (2002), pp. 201–8.
- [263] T. Sooksawate and M. A. Simmonds. “Effects of membrane cholesterol on the sensitivity of the GABA(A) receptor to GABA in acutely dissociated rat hippocampal neurones”. In: *Neuropharmacology* 40.2 (2001), pp. 178–84.
- [264] C. J. daCosta and J. E. Baenziger. “A lipid-dependent uncoupled conformation of the acetylcholine receptor”. In: *J Biol Chem* 284.26 (2009), pp. 17819–25.
- [265] C. L. Morales-Perez, C. M. Noviello, and R. E. Hibbs. “X-ray structure of the human alpha4beta2 nicotinic receptor”. In: *Nature* 538.7625 (2016), pp. 411–415.
- [266] C. Melis, S. C. Lummis, and C. Molteni. “Molecular dynamics simulations of GABA binding to the GABAC receptor: The role of Arg104”. In: *Biophys J* 95.9 (2008), pp. 4115–23.
- [267] J. A. Ashby et al. “GABA binding to an insect GABA receptor: A molecular dynamics and mutagenesis study”. In: *Biophys J* 103.10 (2012), pp. 2071–81.
- [268] F. Comitani et al. “Insights into the binding of GABA to the insect RDL receptor from atomistic simulations: A comparison of models”. In: *J Comput Aided Mol Des* 28.1 (2014), pp. 35–48.

- [269] F. Comitani, V. Limongelli, and C. Molteni. “The free energy landscape of GABA binding to a pentameric ligand-gated ion channel and its disruption by mutations”. In: *J Chem Theory Comput* 12.7 (2016), pp. 3398–406.
- [270] A. Shahsavar et al. “Structural studies of nicotinic acetylcholine receptors: Using acetylcholine-binding protein as a structural surrogate”. In: *Basic Clin Pharmacol Toxicol* 118.6 (2016), pp. 399–407.
- [271] A. W. Sousa da Silva and W. F. Vranken. “ACPYPE - AnteChamber PYthon Parser interfacE”. In: *BMC Res Notes* 5 (2012), p. 367.
- [272] J. Wang et al. “Automatic atom type and bond type perception in molecular mechanical calculations”. In: *J Mol Graph Model* 25.2 (2006), pp. 247–60.
- [273] A. Jakalian, D. B. Jack, and C. I. Bayly. “Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation”. In: *J Comput Chem* 23.16 (2002), pp. 1623–41.
- [274] P. V. Klimovich, M. R. Shirts, and D. L. Mobley. “Guidelines for the analysis of free energy calculations”. In: *J Comput Aided Mol Des* 29.5 (2015), pp. 397–411.
- [275] R. J. Vandenberg, C. A. Handford, and P. R. Schofield. “Distinct agonist- and antagonist-binding sites on the glycine receptor”. In: *Neuron* 9.3 (1992), pp. 491–6.
- [276] J. Grudzinska et al. “The beta subunit determines the ligand binding properties of synaptic glycine receptors”. In: *Neuron* 45.5 (2005), pp. 727–39.
- [277] S. A. Pless et al. “A cation- π interaction in the binding site of the glycine receptor is mediated by a phenylalanine residue”. In: *J Neurosci* 28.43 (2008), pp. 10937–42.
- [278] S. Rajendra et al. “The unique extracellular disulfide loop of the glycine receptor is a principal ligand binding element”. In: *EMBO J* 14.13 (1995), pp. 2987–98.
- [279] Justin P. Gollivan and Dennis A. Dougherty. “A computational study of cation– π interactions vs salt bridges in aqueous media: Implications for protein engineering”. In: *Journal of the American Chemical Society* 122.5 (2000), pp. 870–874.
- [280] S. Pylaeva, M. Brehm, and D. Sebastiani. “Salt bridge in aqueous solution: Strong structural motifs but weak enthalpic effect”. In: *Sci Rep* 8.1 (2018), p. 13626.
- [281] T. M. Lewis, P. R. Schofield, and A. M. McClellan. “Kinetic determinants of agonist action at the recombinant human glycine receptor”. In: *J Physiol* 549.Pt 2 (2003), pp. 361–74.
- [282] D. L. Mobley et al. “Predicting absolute ligand binding free energies to a simple model site”. In: *J Mol Biol* 371.4 (2007), pp. 1118–34.
- [283] D. L. Mobley et al. “Predictions of hydration free energies from all-atom molecular dynamics simulations”. In: *J Phys Chem B* 113.14 (2009), pp. 4533–7.
- [284] G. J. Rocklin, D. L. Mobley, and K. A. Dill. “Calculating the sensitivity and robustness of binding free energy calculations to force field parameters”. In: *J Chem Theory Comput* 9.7 (2013), pp. 3072–3083.

- [285] Christopher I. Bayly et al. “A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model”. In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280.
- [286] V. S. Inakollu et al. “Polarisable force fields: What do they add in biomolecular simulations?” In: *Curr Opin Struct Biol* 61 (2020), pp. 182–190.
- [287] Hervé Minoux and Christophe Chipot. “Cation– π Interactions in proteins: Can simple models provide an accurate description?” In: *Journal of the American Chemical Society* 121.44 (1999), pp. 10366–10372.
- [288] H. M. Khan et al. “Improving the force field description of tyrosine-choline cation- π interactions: QM investigation of phenol-N(Me)₄(+) interactions”. In: *J Chem Theory Comput* 12.11 (2016), pp. 5585–5595.
- [289] D. Jiao et al. “Calculation of protein-ligand binding free energy by using a polarizable potential”. In: *Proc Natl Acad Sci U S A* 105.17 (2008), pp. 6290–5.
- [290] Christopher M. Baker. “Polarizable force fields for molecular dynamics simulations of biomolecules”. In: *WIREs Computational Molecular Science* 5.2 (2015), pp. 241–254.
- [291] L. L. Duan, G. Q. Feng, and Q. G. Zhang. “Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand binding energy”. In: *Sci Rep* 6 (2016), p. 31488.
- [292] Y. Shi et al. “The polarizable atomic multipole-based AMOEBA force field for proteins”. In: *J Chem Theory Comput* 9.9 (2013), pp. 4046–4063.
- [293] P. E. Lopes et al. “Force field for peptides and proteins based on the classical drude oscillator”. In: *J Chem Theory Comput* 9.12 (2013), pp. 5430–5449.
- [294] V. Limongelli, M. Bonomi, and M. Parrinello. “Funnel metadynamics as accurate binding free-energy method”. In: *Proc Natl Acad Sci U S A* 110.16 (2013), pp. 6358–63.
- [295] F. Y. Dupradeau et al. “The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building”. In: *Phys Chem Chem Phys* 12.28 (2010), pp. 7821–39.
- [296] E. Vanquelef et al. “R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments”. In: *Nucleic Acids Res* 39.Web Server issue (2011), W511–7.
- [297] L. Sauguet et al. “Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation”. In: *Proc Natl Acad Sci U S A* 111.3 (2014), pp. 966–71.
- [298] A. Taly et al. “Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism”. In: *Biophys J* 88.6 (2005), pp. 3954–65.
- [299] E. J. Bertaccini et al. “Effect of cobratoxin binding on the normal mode vibration within acetylcholine binding protein”. In: *J Chem Inf Model* 48.4 (2008), pp. 855–60.
- [300] J. L. Eisele et al. “Chimaeric nicotinic-serotonergic receptor combines distinct ligand binding and channel specificities”. In: *Nature* 366.6454 (1993), pp. 479–83.

- [301] T. Grutter et al. “A chimera encoding the fusion of an acetylcholine-binding protein to an ion channel is stabilized in a state close to the desensitized form of ligand-gated ion channels”. In: *C R Biol* 328.3 (2005), pp. 223–34.
- [302] N. Schmandt et al. “A chimeric prokaryotic pentameric ligand-gated channel reveals distinct pathways of activation”. In: *J Gen Physiol* 146.4 (2015), pp. 323–40.
- [303] G. Duret et al. “Functional prokaryotic-eukaryotic chimera from the pentameric ligand-gated ion channel family”. In: *Proc Natl Acad Sci U S A* 108.29 (2011), pp. 12143–8.
- [304] T. S. Tillman et al. “ELIC- $\alpha 7$ nicotinic acetylcholine receptor ($\alpha 7$ nAChR) chimeras reveal a prominent role of the extracellular-transmembrane domain interface in allosteric modulation”. In: *J Biol Chem* 289.20 (2014), pp. 13851–7.
- [305] M. S. Prevost et al. “A locally closed conformation of a bacterial pentameric proton-gated ion channel”. In: *Nat Struct Mol Biol* 19.6 (2012), pp. 642–9.
- [306] M. S. Prevost et al. “Intermediate closed state for glycine receptor function revealed by cysteine cross-linking”. In: *Proc Natl Acad Sci U S A* 110.42 (2013), pp. 17113–8.
- [307] R. Salari, S. Murlidaran, and G. Brannigan. “Pentameric ligand-gated ion channels : Insights from computation”. In: *Mol Simul* 40.10-11 (2014), pp. 821–829.
- [308] A. Crnjar et al. “Mutagenesis computer experiments in pentameric ligand-gated ion channels: The role of simulation tools with different resolution”. In: *Interface Focus* 9.3 (2019), p. 20180067.
- [309] L. Orellana et al. “Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations”. In: *Nat Commun* 7 (2016), p. 12575.
- [310] A. C. Pan, D. Sezer, and B. Roux. “Finding transition pathways using the string method with swarms of trajectories”. In: *J Phys Chem B* 112.11 (2008), pp. 3432–40.
- [311] B. Lev and T. W. Allen. “Simulating ion channel activation mechanisms using swarms of trajectories”. In: *J Comput Chem* (2019).
- [312] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proc Natl Acad Sci U S A* 99.20 (2002), pp. 12562–6.
- [313] R. D. Malmstrom et al. “On the application of molecular-dynamics based Markov state models to functional proteins”. In: *J Chem Theory Comput* 10.7 (2014), pp. 2648–2657.
- [314] D. Shukla et al. “Markov state models provide insights into dynamic modulation of protein function”. In: *Acc Chem Res* 48.2 (2015), pp. 414–22.
- [315] K. J. Kohlhoff et al. “Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways”. In: *Nat Chem* 6.1 (2014), pp. 15–21.
- [316] Gareth A. Tribello et al. “PLUMED 2: New feathers for an old bird”. In: *Computer Physics Communications* 185.2 (2014), pp. 604–613.

- [317] M. K. Scherer et al. “PyEMMA 2: A software package for estimation, validation, and analysis of Markov models”. In: *J Chem Theory Comput* 11.11 (2015), pp. 5525–42.
- [318] G. Perez-Hernandez et al. “Identification of slow molecular order parameters for Markov model construction”. In: *J Chem Phys* 139.1 (2013), p. 015102.
- [319] J. H. Prinz et al. “Markov models of molecular kinetics: Generation and validation”. In: *J Chem Phys* 134.17 (2011), p. 174105.
- [320] A. Mardt et al. “VAMPnets for deep learning of molecular kinetics”. In: *Nat Commun* 9.1 (2018), p. 5.
- [321] Arvind Kumar et al. “Mechanisms of activation and desensitization of full-length glycine receptor in membranes”. In: *bioRxiv* (2019), p. 788695.
- [322] Jie Yu et al. “Mechanism of gating and partial agonist action in the glycine receptor”. In: *bioRxiv* (2019), p. 786632.