Étude de l'anesthésie générale à l'échelle atomique par modélisation d'un homologue bactérien du récepteur nicotinique humain

Benoist Laurent

Thèse de doctorat de l'université Paris Diderot (Paris 7)
École Doctorale Interdisciplinaire Européenne Frontières du Vivant

Étude de l'anesthésie générale à l'échelle atomique par modélisation d'un homologue bactérien du récepteur nicotinique humain

Présentée par

Benoist Laurent

Pour obtenir le grade de Docteur de l'université Paris Diderot (Paris 7)

Spécialité : Biochimie informatique et statistique

Thèse soutenue le 16 juin 2014 devant le jury composé de :

Dr. Annick Dejaegere	Université de Strasbourg	Rapporteur
Dr. Thomas Grutter	CNRS	Rapporteur
Pr. Anne-Claude Camproux	Université Paris Diderot (Paris 7)	Examinateur
Dr. Nicolas Férey	Université Paris Sud (Paris 11)	Examinateur
Dr. Damien Laage	CNRS	Examinateur
Dr. Pierre-François Lavallée	CNRS	Examinateur
Dr. Marc Baaden	CNRS	Directeur de thèse

À mon épouse, Lydie, mon modèle et ma source d'inspiration dans la science comme dans la vie.

À ma fille, Flore, qui a plongé ma vie dans une incertitude insoluble qu'elle balaye chaque fois qu'elle plonge son regard dans le mien.

Remerciements

Mes remerciements s'adressent tout d'abord à mon directeur de thèse, Marc Baaden, qui s'est battu avec moi pour obtenir un financement et qui m'a accompagné pendant ces trois années. Tu m'as laissé la liberté dont j'avais besoin (parfois plus !) et toujours respecté mes choix. Merci d'avoir choisi de ne jamais rien m'imposer, mais de me convaincre au terme de débats souvent interminables.

Je remercie chaleureusement les laboratoires Servier pour avoir accepté de financer ce travail pendant plus de trois ans en me laissant la plus grande liberté, et plus particulièrement Olivier Nosjean qui m'a soutenu dès le départ. Merci également à mon école doctorale, Frontières du Vivant et à la Fondation Bettencourt Schueller.

Je tiens à remercier le Professeur Philippe Derreumaux pour m'avoir accueilli au sein du Laboratoire de Biochimie Théorique.

Mes sincères remerciements aux membres du jury qui ont accepté d'évaluer mes travaux. Un merci particulier pour Nicolas Férey et Damien Laage qui ont été mes tuteurs tout au long de ma thèse et qui ont partagé avec moi leur vision et leurs idées sur ce travail.

Je remercie très chaleureusement les membres du LBT, passés et présents, qui, chacun à sa manière, contribuent à faire de ce laboratoire un lieu de travail stimulant que je quitte à regret. Merci à Samuel qui m'a apporté son expertise dans de nombreux domaines et répondu à mes questions incessantes avec patience et sarcasmes ! J'aimerais également adresser quelques mots à trois membres du LBT qui m'ont particulièrement soutenus pendant cette thèse : Fabio Sterpone, Antoine Taly et Jérôme Hénin. Merci Jérôme de m'avoir montré que même les choses les plus simples sont d'une complexité abyssale et d'avoir su me les expliquer avec sérénité et courage.

Je n'oublie pas mes collègues de pasteuriens, Frédéric Poitevin, Ludovic Sauguet, Marie Prevost, Pierre-Jean Corringer et Marc Delarue, pour m'avoir fait profiter de leur expertise insondable des pLGICs.

Merci à mes amis, avec qui je partage depuis de nombreuses années mes joies, passions, doutes et espoirs et sur qui j'ai toujours pu compter. Un mot particulier pour Paul qui, il y a maintenant huit ans, m'a poussé à reprendre mes études et sans qui aujourd'hui, j'aurais sans doute un emploi stable !

Un dernier mot pour ma famille et plus particulièrement mes parents, sans qui, pour citer un de mes illustres aînés, je ne serais pas là. Merci pour l'énergie que vous avez investi dès le plus jeune âge à m'éveiller aux curiosités de la vie, probablement malgré vous. Je ne pourrais jamais vous rendre ce que vous m'avez donné en me permettant de reprendre mes études et d'aller au bout de mes rêves.

Abbreviations

5-HTR 5-HydroxyTryptamine Receptor

ATP Adenosin Triphosphate

BAR Bennett's Acceptance Ratio

cAMP cyclic Adenosine Monophosphate

CeCILL Ce(A) C(NRS) I(NRIA) L(ogiciel) L(ibre)

CHARMM Chemistry at HARvard Molecular Mechanics

CNS Central Nervous System

CPU Central Processing Unit

CSS Cascading Style Sheets

DSF Desflurane

ECD Extracellular Domain

ELIC *Erwinia chrysanthemi* Ligand-gated Ion Channel

FEB Free Energy of Binding

FFT Fast Fourier Transforms

GA General Anesthetic

GABA γ-Aminobutiric Acid

$GABA_AR$ GABA Receptor of type A

GLIC *Gloeobacter violaceus* Ligand-gated Ion Channel

GluCl Glutamate-gated Chloride ion channel

GlyR Glycine Receptor

GNU GNU's Not Unix!

GPL General Public License

GROMACS Groningen Machine for Chemical Simulations

HTML Hypertext Markup Language

IDRIS Institut du Développement et des Ressources en Informatique Scientifique

IMD Interactive Molecular Dynamics

LC Locally Closed

MBR Bromoform

MD Molecular Dynamics

nAChR Nicotinic Acetylcholine Receptor

NAMD Not (just) Another Molecular Dynamics program

NetCDF Network Common Data Form

NMR Nuclear Magnetic Resonance

NPT Number of particles Pressure Temperature

NVT Number of particles Volume Temperature

OPEP Optimized Potential for Efficient rotein structure Prediction

PBC Periodic Boundary Conditions

PBE Poisson-Boltzmann Equation

PDB Protein Data Bank

PFL Propofol

pLGIC pentameric Ligand-Gated Ion Channel

PME Particle Mesh Ewald

PNS Peripheral Nervous System

reST reStructured Text

RMSD Root Mean Square Deviation

SAXS Small-Angle X-ray Scattering

TMD Transmembrane Domain

VMD Visual Molecular Dynamics

WT Wild-Type

Contents

List of Figures

List of Tables

A short primer about this thesis

0

Neuroreceptors are membrane proteins responsible for nervous signal transduction between neurons and every part of the body. At synaptic ends, they receive the signal from the presynaptic cell in the form of a neurotransmitter and respond with the opening of an ion channel located on the postsynaptic cell, leading to transmission of the signal across the nervous ending. Dysfunction of neuroreceptors is associated to several disorders of the central nervous system including *myasthenia gravis*, epilepsy, addiction to nicotine and alcohol and several cognitive and mental disorders such as schizophrenia, Alzheimer's and Parkinson's diseases. They are also involved in general anesthesia mechanisms.

In this work, I focused on the latter, trying to understand the open question of general anesthesia mechanisms at the atomistic scale. General anesthetics have been shown to target neuroreceptors, in particular the pentameric Ligand-Gated Ion Channel (pLGIC) family. As a threedimensional structure of an eukaryotic member of this family is particularly difficult to solve, numerous groups worldwide focus now on prokaryotic members on this family, including *Gloeobacter violaceus* Ligand-gated Ion Channel (GLIC). GLIC is a homopentameric ion channel the opening of which is controlled by pH variations. Since its first crystallization in 2009, several high resolution structures of this bacterial channel have been resolved, including mutants, open, locally closed and closed conformations, and co-crystals with modulating ions, alcohols, benzodiazepines, local and general anesthetics.

My work is primarily based on crystals of general anesthetics propofol, desflurane and bromoform bound to GLIC in open and locally closed conformations. I attempt to characterize anesthetics behavior while bound to allosteric sites in conformations close to the crystal structures thanks to *in silico* methods such as Molecular Dynamics (MD) simulations and Free Energy of Binding (FEB) calculations.

This manuscript is structured as follows.

Chapter 1, "Biological Background", briefly introduces basic notions on the nervous signal transduction in the human body, followed by a short history on the development of anesthesia. I then focus on the structure and function of human neuroreceptors before concluding with the state-of-the-art of the study of bacterial and unicellular members of the pLGIC superfamily.

In Chapter 2, entitled "Molecular Modeling: Theory And Practice", I shortly introduce the main methods I applied in the context of this project: MD simulations and free energy calculations. I then develop some of the principal recurring difficulties that have to be faced calculating MD simulations of biomolecules.

Chapter 3, "High-Performance Computing And Large Scale Data Analysis", deals with the answer I found to methodological issues I faced on a daily basis, mainly related to high-performance computing and large scale data analysis. In particular, I present the Epock software, which I implemented during my PhD project and that aims to efficiently calculate protein pocket volumes.

The next two chapters are devoted to the characterization of general anesthetics behavior bound to the GLIC channel.

Chapter 4, "Probing pLGICs with bromoform reveals many interconnected anesthetic binding sites", aims to describe the bromoform interactions with GLIC while bound to several binding sites. For this purpose, I combine three complementary simulation strategies, trying to answer three principal questions: i) are bromoform crystallographic binding sites spontaneously accessible? ii) how does bromoform dynamics evolve while bound to these sites? iii) what is bromoform's affinity for each binding site?

In the next chapter, 'Propofol & desflurane simulations provide new insights into anesthetic action at the atomic scale", I extend the problems addressed in studying bromoform to general anesthetics propofol and desflurane. I aim to check whether hypotheses I established for bromoform are verified on other general anesthetics that exhibit quite different properties compared to bromoform. I also address several additional issues such as the symmetry of anesthetic binding and the extensive characterization of pocket volume.

The work is then concluded with final remarks and a general perspective. Supplementary information is provided in the appendix.

Biological Background

1

1.1 The human nervous system, a central player in general anesthesia

Here I will briefly describe in a top-down manner the structure of relevant parts of the human nervous system in relation to general anesthesia. The nervous system is an ensemble of structures that coordinates an animal's voluntary and involuntary actions and transmits signals to different body areas. In vertebrates, it consists of two main parts, called the Central Nervous System (CNS) and the Peripheral Nervous System (PNS). The CNS consists of the brain and the spinal cord. The PNS consists mainly of nerves, which are long fibers that connect the CNS to every part of the body. The PNS also includes peripheral ganglia and the enteric nervous system, a semi-independent part of the nervous system that controls the gastrointestinal system.

1.1.1 Overall structure

The nervous system is made of two main categories of cells: neurons and glial cells.

Neurons, or nerve cells, are distinguishable from other cells in a number of ways but their most fundamental property is that they communicate with other cells through specialized intercellular adhesion sites called synapses. A typical neuron has four morphologically defined regions: the cell body, dendrites, axon and presynaptic terminals as shown in figure 1.1. Dendrites provide a highly branched, elongated surface for receiving signals. The axon conducts electrical impulses rapidly over long distances to their synaptic terminal, which releases neurotransmitter onto target cells (Kandel *et al.*, 2000). An axon can extend to different parts of the body and make thousands of synaptic contacts. A nerve is a bundle of axons.

All neural cells that lack the capacity of transmitting rapid signals in the form of an action potential (see section 1.1.3) are categorized into a broad class termed glia. In mammals, glial cells include microglia, astrocytes, Schwann cells and oligodendrocytes (figure 1.2). They ensure a wide range of functions, some of them probably unknown yet, including homeostasis maintenance, neuron support, nutrition and insulation to speed up electrical communication (Allen and Barres, 2009). Glial cell are an essential component of the nervous system and can constitute the major part of a brain: human brain has about 90 % glial cells, an elephant's brain 97 %.

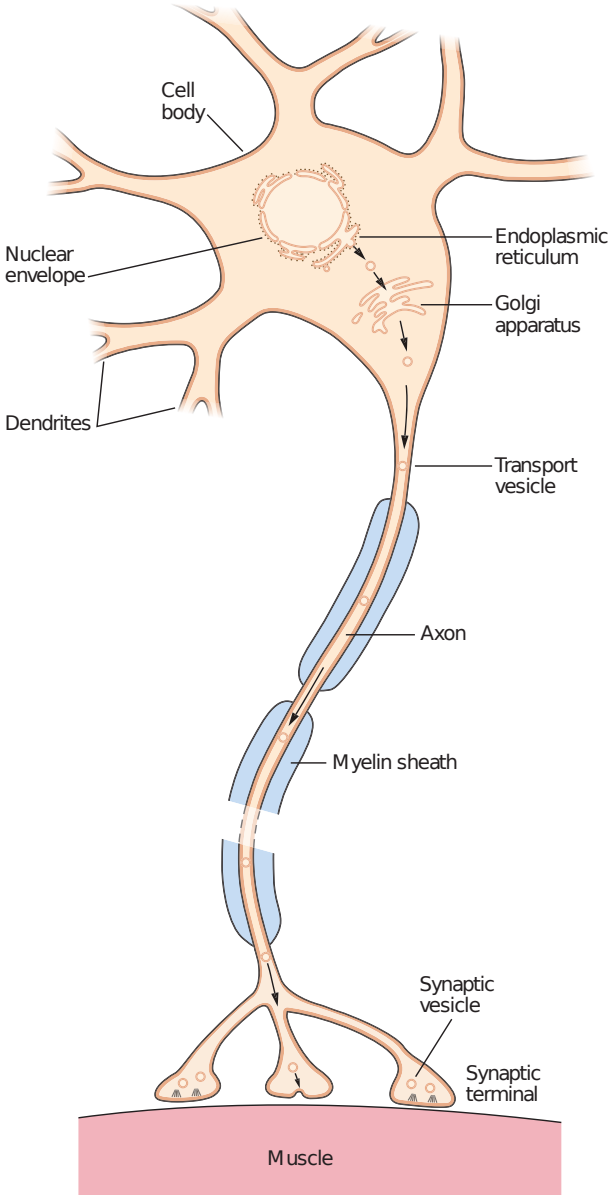

Figure 1.1 – The structure of a neuron. The cell body and nucleus of a spinal motor neuron are surrounded by a double-layered membrane, the nuclear envelope, which is continuous with the endoplasmic reticulum. The space between the two membrane layers that constitutes the nuclear envelope is continuous with the lumen of the endoplasmic reticulum. Dendrites emerge from the basal aspect of the neuron, the axon from the apical aspect. From Kandel *et al.* (2000).

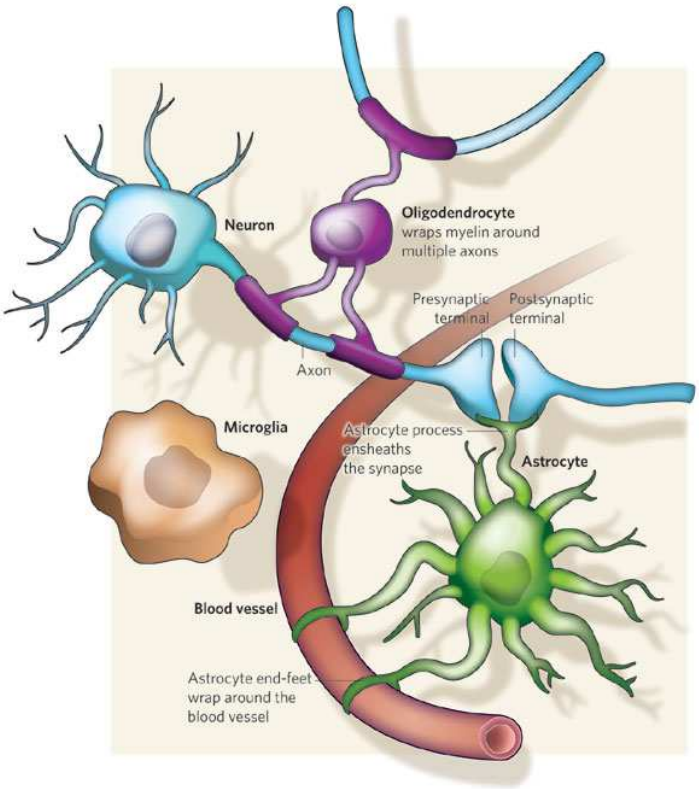

Figure 1.2 – Different types of glia interact with neurons and the surrounding blood vessels. Oligodendrocytes wrap myelin around axons to speed up neuronal transmission. Astrocytes extend processes that en-sheath blood vessels and synapses. Microglia keep the brain under surveillance for damage or infection. From Allen and Barres (2009).

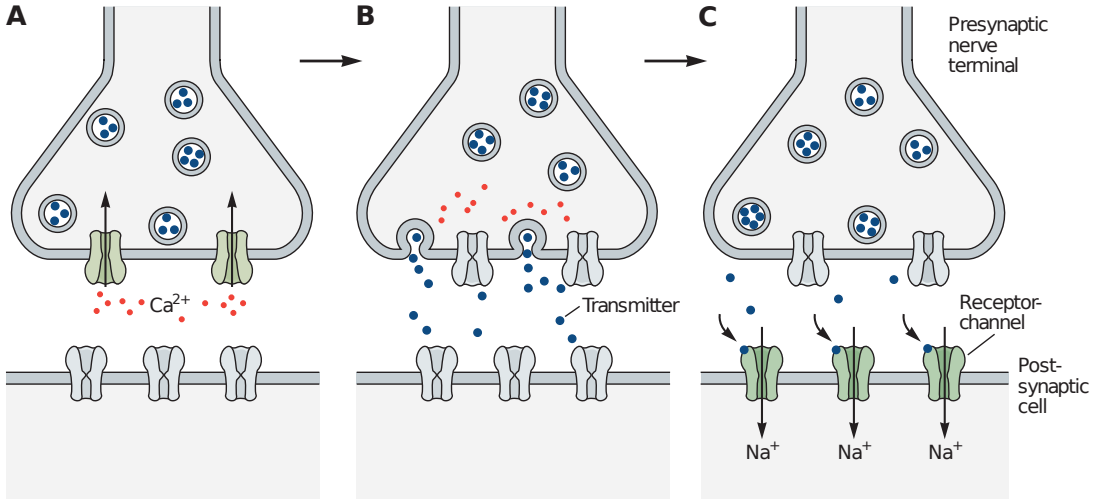

Figure 1.3 – The chemical synaptic transmission. A) An action potential arriving at the presynaptic terminus causes voltage-gated Ca^{2+} channels on the presynaptic membrane to open. B) The opening of the Ca^{2+} channels causes high concentration of intracellular Ca^{2+} which causes synaptic vesicles containing neurotransmitter molecules to fuse with the presynaptic membrane and transmitters to be liberated in the synaptic cleft. C) The released neurotransmitter molecules diffuse across the synaptic cleft and bind the neuroreceptor on the postsynaptic membrane. As ion channels open, the membrane potential of the postsynaptic cell changes. Adapted from Kandel *et al.* (2000).

1.1.2 The synapse - creating the right interconnections

The synapse is a specialized structure that mediates a functional interaction between two neurons or between a neuron and another cell type. Synapses may be of chemical as well as electrical nature. Chemical and electrical synapses differ not only by the mechanism of information transfer, but also in their morphological organization. At electrical synapses, the pre- and postsynaptic cells communicate through gap junctions, cell-to-cell pores that serve as a conduit between the cytoplasm of two cells. Consequently, the space between the two cells, called the synaptic cleft, is on the order of 2 to 4 nm wide. In contrast, in the much more common chemical synapses, the synaptic cleft is wider, on the order of 20 to 40 nm (Hormuzdi *et al.*, 2004).

Chemical synaptic transmission depends on the diffusion of a neurotransmitter across the synaptic cleft. Neurotransmitter molecules are contained in synaptic vesicles. At arrival of an electrical signal, voltage-gated Ca^{2+} channels at the presynaptic terminus open, allowing Ca^{2+} ions to enter the cell. The rise of intracellular calcium initiates synaptic vesicle fusion with the presynaptic membrane, therefore the neurotransmitter liberation in the synaptic cleft (figure 1.3). The neurotransmitter diffuses and binds to its receptor on the postsynaptic membrane, which responds by opening and letting ions pass from the extracellular environment to the cytosol. If the ion flow is adequate, it will provoke a depolarization of the postsynaptic membrane which will be transmitted through the axon of the receptor cell. Importantly, chemical synapses can amplify the signal they receive since one synaptic vesicle releases thousands of neurotransmitter molecules that can open thousands of ion channels on the target cell. A small presynaptic nerve which generates a weak electrical current can therefore depolarize a large postsynaptic cell.

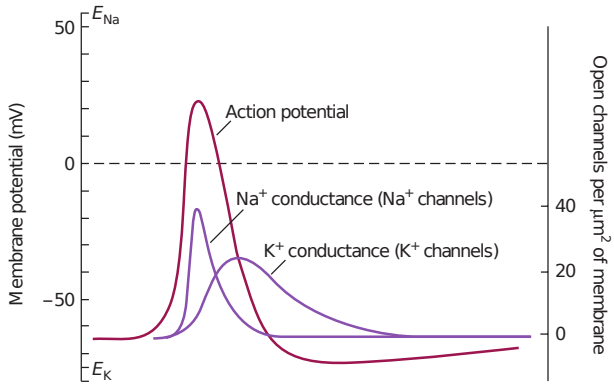

Figure 1.4 – The action potential. The sequential opening of voltage-gated Na+ and K+ channels generates the action potential. From Kandel *et al.* (2000).

1.1.3 The action potential

An enormous amount of work has been realized by Hodgkin and Huxley (1952) leading to the detailed comprehension of the sequence of events that constitute the action potential. As described in section 1.1.2, the neurotransmitters binding to their receptor lead to the opening of ion channels, therefore the entering of ions in the intracellular space. Consequently, the membrane potential changes, rising if cations enter (depolarization), declining if anions enter (hyperpolarization). If the depolarization is sufficient, *i.e.* exceeds a *threshold* value[1], voltage-gated Na^+ channels rapidly open resulting in an inward Na^+ current. This current causes further depolarization, thereby opening more Na^+ channels, resulting in a further increase of the inward current. The opening of Na^+ channels causes the rising phase of the action potential (figure 1.4). The depolarization gradually inactivates the voltage-gated Na^+ channels and opens, with some delay, voltage-gated K^+ channels, resulting in an outward K^+ flow that tends to repolarize the membrane. As these channels remain open for some time after the resting potential has been reached, this current leads to a transient shift of the membrane potential to values more negative than the resting potential (Kandel *et al.*, 2000). The combined effect of this increase in K^+ conductance combined with the inactivation of Na^+ channels underlies the *absolute refractory period*, the brief period after which an action potential cannot be triggered. As some K^+ channels begin to close and some Na^+ channels recover from inactivation, the membrane enters the *relative refractory period*, during which an action potential can be triggered by applying stronger stimuli that those normally required to reach threshold. The membrane potential returns to its resting value as all the K^+ channels finally close and the initial concentration of intracellular ions is restored by nongated and gated ion channels responsible for maintaining ion balance at rest.

1.1.4 Medical implications: general anesthesia

The development of general anesthesia

General anesthesia has several purposes:

[1]The action potential obeys the *all-or-none principle*: stimuli below the threshold do not produce an action potential while stimuli above the threshold all produce an action potential with the same amplitude. The intensity of the stimulus affects the action potential frequency.

- analgesia *i.e.* loss of response to pain,
- amnesia *i.e.* loss of memory,
- unconsciousness,
- immobility *i.e.* loss of motor reflexes,
- relaxation of skeletal muscles.

In contrast, local anesthesia does not provoke amnesia or unconsciousness. Furthermore, the effect of a local anesthesia is restricted to a small part of the body. In a strict sense, local anesthesia refer for example to a tooth or an area of the skin. Any larger region such as leg or arm is covered under the term *regional anesthesia*. Usually, a local anesthetic cannot be used as general anesthetic and conversely.

The first attempts of anesthesia presumably occurred during prehistory through the administration of herbal remedies. Opium and alcohol were used in the Antiquity as narcotic and sedative, respectively, although the question of which people are at the origin of this usage is still debated (Krikorian, 1975). Interestingly, a Chinese legend wants that a Chinese physician successfully used herbal decoctions to render patients unconscious for several days and practice surgery upon them. The exact formula he used likely disappeared at his death. For centuries, physicians used oral as well as inhaled anesthetics in therapeutics in the form of herbal mixtures, often composed in part of *Papaver somniferum*, a plant from which opium is prepared.

Several notable advances in anesthesia, local and general, were made during the 19th century. The first documented successful use of general anesthesia is generally considered as Hanoka Seishu mastectomy on 13 October 1804 (Izuo, 2004). This Japanese surgeon, after years of efforts, finally developed a formula which he named *tsusensan* composed of several plants from which, by the way, opium is not present. The same year, Friedrich Sertürner isolated morphine from opium, a molecule still commonly used as an analgesic. Along the 19th century, several uses of diethyl ether, the analgesic properties of which have been described in the 16th century, are reported for local as well as general anesthesia. In the middle 1800s, chloroform was discovered in Europe and rapidly replaced ether in Europe and Western countries, but was discarded because of its tendency to cause fatal cardiac arrhythmia. Most modern anesthetics have been developed by pharmaceutical labs, including propofol and desflurane; their action mechanism will be discussed in chapter 5. The development of most anesthetics has been primarily empirical, even sometimes fortuitous.

Hypotheses on general anesthetics action

Several theories have been formulated on general anesthetics action, always relative to the modulation of membrane proteins in the neuronal membrane. Paul Ehrlich (1854–1915) first proposed the concept of highly specific interactions between drugs and receptors, stating that a drug acts only when bound to its target (Weir, 2006). As this theory was judged hardly applicable to general anesthetics because of their chemical and structural diversity (figure 1.5), theories implying nonspecific perturbations of neurons have been formulated. In the early 20th century, HH Meyer and CE Overton independently reported a correlation between the solubility of narcotics in olive oil and their anesthetic power: the greater the

lipid solubility of the compound, the greater its anesthetic potency. This relation became known as the *Meyer-Overton correlation*, refined by Meyer's son in Meyer, 1937. Theories have then been developed mentioning that anesthetic solubilization in neuronal membranes alters the function of ion channels. However, these theories suffer several weaknesses:

- several compounds with structures similar to anesthetics and high lipid solubility do not act as anesthetics (figure 1.6A,B),
- general anesthetic ability to perturb membranes *in vitro* can be reproduced by a temperature drop to less than 1 °C, a change well within the physiological range and clearly not sufficient to induce loss of consciousness,
- some enantiomers of general anesthetics do not produce identical clinical effects, although they have the same properties in lipid bilayers, (figure 1.6C)
- there appears to be a cutoff effect above a certain molecular volume which is indicative of anesthetic agents interacting with binding site(s) of finite dimensions (Bradley *et al.*, 1984).

Finally, it has been demonstrated that a range of general anesthetics act as competitive antagonists on the firefly luciferase, a soluble protein (Franks and Lieb, 1984). Remarkably, inhibition of luciferase was directly correlated with anesthetic potency, providing persuasive evidence that general anesthetic drugs could selectively interact with proteins. However, some groups still argue that anesthetics may alter membrane properties which would play a role in anesthesia (Bahri *et al.*, 2007; Baenziger *et al.*, 2008; Hansen *et al.*, 2013).

In vitro experiments demonstrated that general anesthetics alter the function of several neurotransmitter receptors at clinically relevant concentrations (Weir, 2006) including the GABA Receptor of type A ($GABA_AR$), the Nicotinic Acetylcholine Receptor (nAChR), the Glycine Receptor (GlyR) and the 5-HydroxyTryptamine Receptor (5-HTR) of subtype 3 ($5\text{-}HT_3R$).

1.1.5 Neuroreceptors, a likely target for general anesthetics

Neuroreceptors are membrane proteins localized on the postsynaptic membrane that share a few common properties, including the binding of neurotransmitters.

A major second property is that, in contrast to molecular pumps, the flux of ions through the channel is passive. It is therefore determined not by the channel itself, but rather by the electrostatic and diffusional driving forces across the membrane. Finally, those channels are selective, which means that they allow particular types of ions to cross the membrane. Most channels are selective to only one type of ion that is usually present is the extracellular environment. Anion channels conduct only one physiological ion, chloride (Cl^-). Cation channels are selective to Na^+, K^+, or Ca^{2+}. The effect of a synaptic potential, *i.e.* excitatory or inhibitory, depends on the type of ion that permeates into the postsynaptic cell: cations, that increase the membrane potential, trigger action potentials while anions cause an inhibition of the signal. Neurons can receive signals from both excitatory and inhibitory synapses. Although some transmitters can produce excitatory as well as inhibitory potentials, most transmitters produce a single type of response because they bind the same type of channel wherever they are met in the body. Excitatory synaptic action

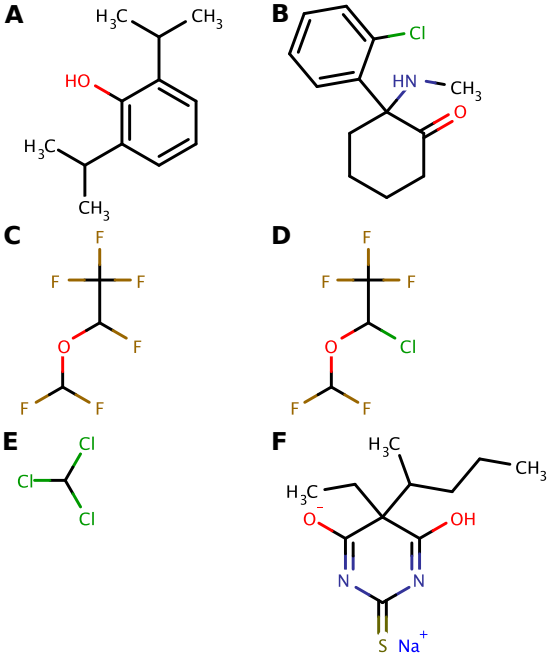

Figure 1.5 – Structure of common general anesthetics. A) Propofol. B) Ketamine. C) Desflurane. D) Isoflurane. E) Chloroform (not used clinically anymore). F) Sodium thiopental (also known as sodium pentothal).

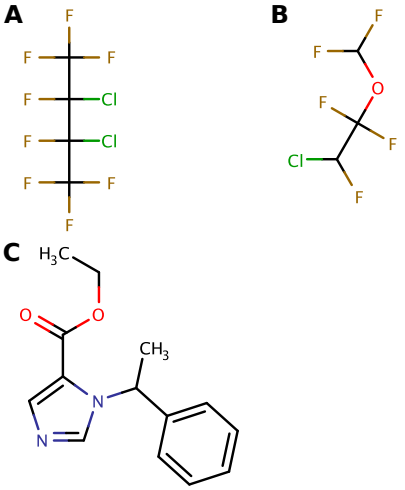

Figure 1.6 – Examples of compounds that do not obey Meyer Overton's rule. A) 2,3-dichlorooctafluorobutane has very high lipid solubility and properties similar to anesthetic but do not provoke anesthetia. B) Enflurane (isomere of isoflurane – figure 1.5C – that is 45 to 90 percent less potent than isoflurane). C) Etomidate (R etomidate is 10 times more potent that S etomidate).

Neurotransmitter	Receptor	Receptor type	Signal Type
Acetylcholine	nAChR	ionotropic	excitatory
Glutamate	iGluR	ionotropic	excitatory
ATP	P2XR	ionotropic	excitatory
Glycine	GlyR	ionotropic	inhibitory
GABA	$GABA_AR$	ionotropic	inhibitory
GABA	$GABA_BR$	metabotropic	excitatory
Serotonin*	5-HTR	metabotropic	excitatory

Table 1.1 – Example neurotransmitters and associated signal type. Examples of common neurotransmitters associated with their most common receptor type and action on the signal.
* Serotonin receptors are metabotropic receptors with the notable exception of the 5-HT_3 receptor which is a ligand-gated Na^+ and K^+ channel. 5-HT receptors are excitatory receptors with the exception of subtypes 5-HT_1 and 5-HT_5.

is usually mediated by ionotropic glutamate and nicotine receptor channels that are permeable to sodium and potassium. Inhibitory synaptic action is usually mediated by GABA and glycine receptors that are permeable to chloride, as summarized in table 1.1.

Neurotransmitters control the opening of ion channels on the postsynaptic cell either directly or indirectly, by acting on different types of receptors.

Receptors that gate ion flow indirectly, known as *metabotropic receptors*, include for example the serotonin receptor (5-HTR) and the GABA receptor of type B. They are usually made of a single subunit, at most two, that are distinct from the ion channel they regulate. Activation of these receptors stimulates the production of a second messenger, cAMP for instance, which activates a protein kinase, an enzyme that phosphorylates different substrate proteins. In many cases, the protein kinases directly phosphorylate ion channels, leading to their opening or closing (Kandel *et al.*, 2000). These several steps account for the delay in the synaptic transmission compared to direct gating.

Ligand-gated channels, or *ionotropic receptors*, gate ion flow directly by opening upon neurotransmitter binding. They are composed of three to five identical or homologous subunits symmetrically arranged around a central ionic pore. The channel displays an Extracellular Domain (ECD) where the neurotransmitter binds and a Transmembrane Domain (TMD) that forms an ion-conducting pore (see figure 1.7A-B). Besides the pentameric channels, this family includes the trimetric P2X receptors and the tetrameric glutamate receptor[2]. The pentameric superfamily comprises the nAChR, the $GABA_AR$ and the GlyR. The pLGICs are also named *Cys-loop* receptors due to the presence in the extracellular domain of a defining loop of approximately 13 residues flanked by two canonical cysteines linked by a disulfide bridge. Most signaling between neurons in the CNS involves ionotropic receptors, as well as synapses at the neuromuscular junction which involve exclusively nAChRs.

To date, the only complete structure of an animal[3] Cys-loop receptor available is that of the nAChR from *Torpedo marmorata* (figure 1.7; Unwin, 2005). It was solved at a 4 Å resolution by cryoelectron microscopy, a resolution at which a high uncertainty exists on side chain localization. Besides, this structure

[2] The glutamate channel from *Caenorhabditis elegans* (GluCl) is pentameric and chloride selective (see section 1.2.2).
[3] Apart from that of the nematode *C. elegans* (see section 1.2).

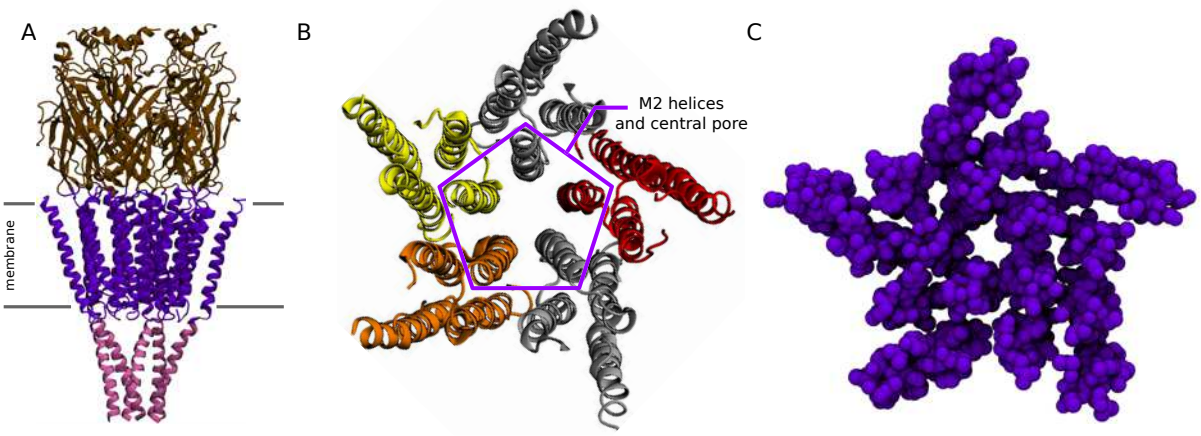

Figure 1.7 – The structure of the nicotinic receptor from *Torpedo marmorata*. A) Side view of the whole structure represented in cartoon. The extracellular, transmembrane and intracellular domains are colored in brown, purple and pink, respectively. B) Top view of the transmembrane domain colored by subunit type (α, γ, α, β, δ). Each subunit contributes four helices, the M2 helix lining the channel pore. C) View of the extracellular half of the TMD in space filling representation showing the numerous gaps induced by the low packing between helices (see also appendix A.1).

is, to date, highly controversial since important gaps exist between the TMD helices (figure 1.7C), gaps that are hardly compatible with the hypothesis according to which they are filled with water as originally assumed (Miyazawa *et al.*, 2003) because of the strong density visible at these locations in the density map used to obtain the structure of the nAChR TMD (PDB-id 1OED) and the hydrophobic nature of the residues surrounding these gaps. Brannigan *et al.* proposed that these gaps are actually filled with cholesterol, since they successfully docked cholesterol into it (Brannigan *et al.*, 2008) and since cholesterol is required for nAChR's proper function (Dalziel *et al.*, 1980). Furthermore, the lack of resolution of electron microscopy data led to the introduction of residue assignment errors in helices in the first atomic model of the TMD alone (Miyazawa *et al.*, 2003), in which residues are shifted by one helical turn from their correct position. The error became evident inspecting homologous structures (see section 1.2) and has been formally proven from direct experimental testing (Mnatsakanyan and Jansen, 2013). As the raw data from Unwin's work have not been released, further refinement is made impossible and the uncertainties concerning this model have not been dissipated.

1.2 Bacterial and invertebrate homologues to the human nicotinic receptor

1.2.1 Why study channels from bacteria?

In 2004, Tasneem *et al.* searched for distant representatives of the Cys-loop family in organisms outside the animal lineage, faced with the fact that ancestors of voltage-gated potassium and sodium channels have been identified in non-animal eukaryotes, as well as numerous prokaryotes. Interestingly, this indicates that these channels were used in other signaling contexts by a variety of organisms way before the origin of the animal nervous system (Ito *et al.*, 2004). As Tasneem *et al.* argue that the TMD is compositionally biased and tends to recover false positives in iterative sequence searches, they used only the ECD for their

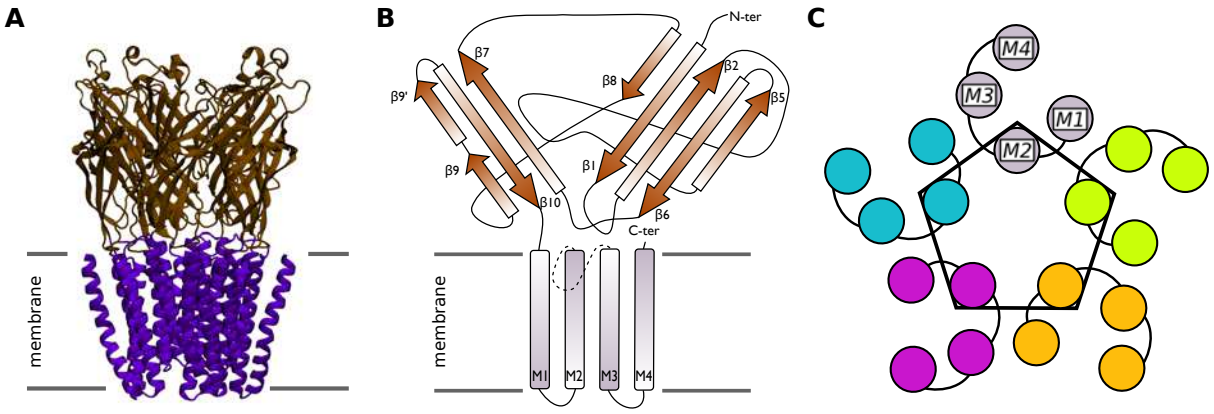

Figure 1.8 – GLIC: a bacterial homologue to the human nicotinic receptor. A) Side view of the GLIC channel. Compared to the nAChR (see figure 1.7), GLIC's structure lacks the intracellular domain, the helices at the top of the ECD and the two cysteines that border the signature loop (not visible here). B) Topology of one of GLIC's five identical subunits, which is the same as for ELIC and GluCl. C) Organination of the transmembrane domain. Each subunit (represented with different colors) contributes to four helices named M1 to M4. M2 helices line the channel pore.

queries. They used PSI-BLAST to search in all organisms with genomic sequence data available at that time, with initial queries such as the human acetylcholine receptor α7 chain or the human GABA receptor α4 chain. In addition to animal sequences, these searches recovered sequences from bacteria such as the cyanobacteria *Gloeobacter violaceus* and the γ-proteobacteria *Erwinia chrysanthemi*[4], among others.

1.2.2 A conserved general receptor organization

The work by Tasneem *et al.* turned out to be a breakthrough in the study of pLGICs. Several groups started focusing on bacterial members of the superfamily, despite the debate on the applicability of discoveries from prokaryotes to eukaryotes. Intensive efforts lead to the crystallization of two bacterial pLGICs, namely ELIC from *Erwinia chrysanthemi* (Hilf and Dutzler, 2008) and GLIC from *Gloeobacter violaceus* (Bocquet *et al.*, 2009; Hilf and Dutzler, 2009). The structure of the eukaryotic glutamate-gated chloride channel GluCl, from *Caenorhabditis elegans*, has been solved few years later in complex with the positive allosteric modulator ivermectin (Hibbs and Gouaux, 2011).

These structures show an important similarity with the nAChR although being simpler (figure 1.8). The ECD is structured in a β-sandwich fold stabilized through conserved hydrophobic residues (Corringer *et al.*, 2012) but lack the N-terminal helix and the two cysteines that border the signature loop. The TMD of each protomer is made of four helices named M1 to M4 (figure 1.8B). The M2 helices form the pore of the channel and are thus critical segments of the ion conduction pathway (figure 1.8C). In contrast to the nAChR, they do not display a cytoplasmic domain.

Despite this conserved general organization, the structure and length of loops connecting β-sheets in the ECD vary along pLGICs although they are critical for the quaternary assembly of the receptor (the sequence similarity between those pLGICs is notably low, as shown in table 1.2). The evolutionary

[4]Recent taxonomic revisions have caused the bacteria *Erwinia chrysanthemi* to be renamed *Dickeya dadantii* (Samson *et al.*, 2005).

	GLIC	ELIC	GluCl	nAChR
GLIC	–	45 (22)	36 (23)	41 (19)
ELIC	–	–	39 (21)	41 (23)
GluCl	–	–	–	42 (26)

Table 1.2 – Sequence similarities of non-human pLGICs. Sequence similarities calculated with protein-protein BLAST. Percentages of identity are given between parentheses. Sequence accession ids are Q7NDN8, P0C7B7, Q17328, and Q9UGM1 for GLIC, ELIC, GluCl, and nAChR respectively.

explanation for these differences is that these loops are believed to play a crucial role in the binding of the agonist and the transduction of the signal to the TMD. Similarly, connecting loops in the TMD play a determinant role for the channel function, such as the M2-M3 loop that actively participates to the signal transduction (Corringer *et al.*, 2012).

1.2.3 Key features of an ion channel: opening and closing

The comprehension of an ion channel's transition from open to closed state, named *gating*, is essential for the understanding of the mechanics of signal transduction. Based on normal modes analysis, Taly *et al.* suggested that nAChR gating involved a quaternary twist-motion rearrangement of the ECD and the TMD (Taly *et al.*, 2005). This analysis, later performed on the crystal structures of GLIC and ELIC, suggested that this gating mechanism was applicable to both prokaryotic channels (Bocquet *et al.*, 2009; Cheng *et al.*, 2009).

However, the detailed function of a channel's gating mechanism can hardly be understood without knowing how to activate and inactivate it. GluCl, by definition, is gated by glutamate. GLIC's natural ligand was known even before the resolution of its crystal structure: the proton, meaning that GLIC opens at acidic pH and closes at neutral pH (Bocquet *et al.*, 2007). On the other hand, ELIC's detailed investigation was slowed down by the fact that activating ligands were unknown until a list of several primary amines that include GABA were found to activate the receptor (Zimmermann and Dutzler, 2011). Notably, GLIC is a cation channel with similar permeabilities for Na^+ and K^+ (Bocquet *et al.*, 2007). Its conductance is 8 pS. At −60 mV, GLIC permeates only 3 to 4 ions per microsecond, making this process very expensive in terms of computational cost to study thanks to MD simulations (see sections 2.2 and 3.1.2).

A second prerequisite to understand atomic details of a channel's gating mechanism is to obtain the structures of both its open and resting state, the endpoints of the gating transition. Based on the calculation of the channel pore radius, one can determine whether a channel is in a conducting conformation or not, *i.e.* if ions can pass through the pore. GLIC and GluCl display very similar open conformations, while ELIC displays a closed pore.

Several attempts have been made to model the gating transition from GLIC's open state to ELIC's closed state, assuming that ELIC is a good model for GLIC's resting form (Zhu and Hummer, 2009, 2010; Nury *et al.*, 2010; Calimet *et al.*, 2013), an assumption that is still debated today considering the low sequence identity between the two proteins (see table 1.2). Furthermore, it is not clear whether ELIC's crystal structure represents the resting state as channels may adopt different closed conformations, in their

desensitized state for example (Gonzalez-Gutierrez and Grosman, 2010). However, the work by Nury *et al.* is interesting from several points of view. First, as GLIC is sensitive to pH, they induced the channel closure by modifying the protonation state of selected residues. Second, they simulated a fully atomistic model of the GLIC system for one microsecond, leading to the closure on only two over five subunits, which suggests that at least another microsecond would be required to simulate GLIC's full closure. Finally, they proposed that GLIC's closure starts by the formation of a hydrophobic gate between residues 9' and 16' in prime notation[5], with a twist motion of the top of M2 helices. These findings revealed particularly consistent with crystal structures of open and nonconductive GLIC as discussed below.

Concerning GluCl's gating, it turns out that, upon ivermectin removal, the channel closes at the hundred nanosecond timescale (Calimet *et al.*, 2013; Yoluk *et al.*, 2013). This transition is probably too swift to be biologically relevant, as a full gating event is assumed to take place on the millisecond timescale. This swiftness probably indicates a bias in the crystal structure and/or the simulation setup. However, several observations made during the transition of GluCl from open to closed state are consistent with the quaternary model proposed in the light of recent high resolution structures of GLIC.

Among all pLGICs, GLIC is currently probably the best structural model to study transitions from open to closed state since it has been crystallized in three distinct conformations, firstly at acidic pH in an open conformation (Bocquet *et al.*, 2009; Hilf and Dutzler, 2009), later mutants have been crystallized in a Locally Closed (LC) conformation (Prevost *et al.*, 2012; Gonzalez-Gutierrez *et al.*, 2013), and very recently an X-ray structure of Wild-Type (WT) GLIC in its resting state has finally been released (Sauguet *et al.*, 2014). The LC form shares most structural features with the open state but displays a closed pore as a result of the concerted bending of the extracellular part of its M2 helices. It was recently demonstrated that WT GLIC can adopt the LC form and that the open and LC forms coexist at acidic pH (Sauguet *et al.*, 2014), consistently with the assumption that the LC form can represent a late intermediate in the course of activation. These structures allowed to confirm that GLIC's gating involves a marked twist motion of the ECD, coupled with an inward tilt motion. The structure of the resting state shows that the conformation of M2 helices is remarkably different from that observed in ELIC: GLIC's pore appears closed due to a concerted bending of the upper part of its M2 helices (figure 1.9). This motion is independent of M3 helix orientation, unchanged compared to the open form. In ELIC, the M2 helix axis appears straight and M2 and M3 helices seem strongly attached to each other.

1.3 pLGICs are modulated by a variety of molecules

Over decades of research, pLGICs turned out to be modulated by numerous compounds with very different physico-chemical properties. I chose to present some of the most important related studies, sorting the compounds by the localization of the binding site.

[5] The prime notation has been introduced decades ago and aims to number residues that line the channel pore. Hence, residues with the same prime number have the same location on the M2 helix, residue 1' being located on the intracellular end of M2.

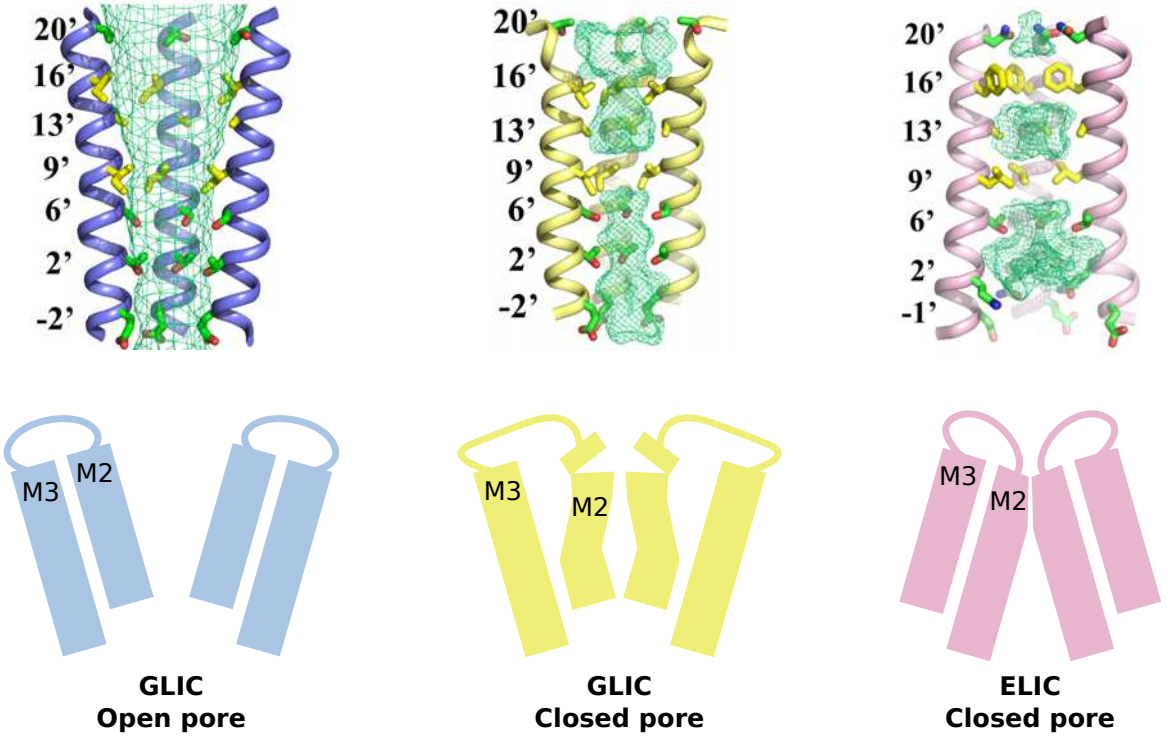

Figure 1.9 – Comparing GLIC and ELIC closed state. Top: enlarged views of the pore. The solvent-accessible region is shown by a green mesh, the side chain of the pore lining residues are shown as sticks with polar and hydrophobic residues colored in green and yellow, respectively. Bottom: schematic representation of the M2 and M3 helix relative positions. Adapted from Sauguet *et al.* (submitted).

1.3.1 Modulation through the sites in the ECD

pLGICs are modulated by benzodiazepines, a class of widely prescribed drugs that display anxiolitic, anti-convulsive and sedative properties, targeting mainly GABA receptors. They have been shown to bind ELIC's ECD in two distinct sites with associated opposite modulation effects (Spurny *et al.*, 2012):

- activation through an intrasubunit site at low concentration,
- inhibition through an intersubunit site at high concentration.

Notably, the intersubunit site matches a site previously described on $GABA_AR$ (Ramerstorfer *et al.*, 2011).

Divalent cations such as Ca^{2+} or Zn^{2+} have been suggested to play an important role in a biological context. Ca^{2+} can potentiate nAChRs (Vernino *et al.*, 1992; Mulle *et al.*, 1992) and inhibit 5-HTRs (Peters *et al.*, 1988), while Zn^{2+} can have potentiating or inhibiting effects, depending on its concentration and the type of pLGIC (Laube *et al.*, 1995; Palma *et al.*, 1998). Divalent ions have been shown to inhibit ELIC (Zimmermann *et al.*, 2012), with a binding site at the interface between two subunits.

A 2.4 Å structure of GLIC allowed to detect Br^-, Cs^+ and Rb^+ at several binding sites in the vestibule edge region of the ECD (Sauguet *et al.*, 2013b). In the same study, the authors suggest two binding sites for acetate, one overlapping the benzodiazepine binding site identified in ELIC, the second at the interface between two subunits.

A final study should be mentioned here. Pan *et al.* (2012) proposed a binding site for the General Anesthetic (GA) ketamine located at the interface between subunits in GLIC's ECD. Interestingly, this structure is, to my knowledge, the only crystal structure of a GA bound to a pLGIC extracellular domain. However, further investigation in Marc Delarue's group highlighted several structures with unattributed densities at the same location, even in absence of ketamine. The same group used these data to try to propose an inhibition mechanism of GLIC by ketamine, using the Perturbation-based Markovian Transmission model (Mowrey *et al.*, 2013a). This work has, in my opinion, two weaknesses: i) the doubts that still exist on the structure of ketamine bound to GLIC and ii) the lack of dynamic analyses to confirm the validity of the paths from ECD to TMD that would explain ketamine's action.

Importantly, most of these binding sites, especially ion ones, are still largely unexplored, as most groups focused on the modulating sites in the TMD. A summary of anesthetics and alcohol binding sites found by X-ray crystallography in pLGICs is given in table 1.3.

1.3.2 Modulation through the sites in the TMD

The TMD is the target of general and local anesthetics, alcohols and several cations.

Modulation sites for alcohols and GAs have been characterized experimentally by combining photolabelling (Hamouda *et al.*, 2013), site-directed mutagenesis and electrophysiology (Olsen *et al.*, 2013; Howard *et al.*, 2011b). Three principal binding sites for GAs and alcohols have been identified within the TMD: i) an intrasubunit pocket located within the M1-4 helix bundle, ii) an intersubunit pocket located roughly at the same height than the intrasubunit pocket and iii) a channel site located at the extracellular end of the pore, between the M2 helices. pLGICs crystal structures have been solved in complex with GAs and

	Protein	Ligand	Resolution (Å)	Reference
GAs	GLIC	propofol	3.3	Nury *et al.* (2011)
	GLIC	desflurane	3.1	Nury *et al.* (2011)
	GLIC (F238A)	bromoform	3.1	Sauguet *et al.* (2013a)
	GLIC	bromoform	2.7	Sauguet *et al.* (2013a)
	ELIC	bromoform	3.7	Spurny *et al.* (2013)
Alcohols	GLIC (F238A)	ethanol	2.8	Sauguet *et al.* (2013a)
	GLIC (F238A)	2-bromo-ethanol	3.1	Sauguet *et al.* (2013a)
Channel blockers	GLIC	bromo-lidocaine	3.5	Hilf *et al.* (2010)
	GLIC	tetra-ethyl-arsonium	3.5	Hilf *et al.* (2010)
	GLIC	tetra-methyl-arsonium	3.6	Hilf *et al.* (2010)
	GLIC	tetra-butyl-antimony	3.7	Hilf *et al.* (2010)
	GLIC	picrotoxine	3.4	Hibbs and Gouaux (2011)

Table 1.3 – Crystal structure of general anesthetics, alcohols and channel blockers bound to a member of the pLGIC family. Adapted from Sauguet *et al.* (submitted).

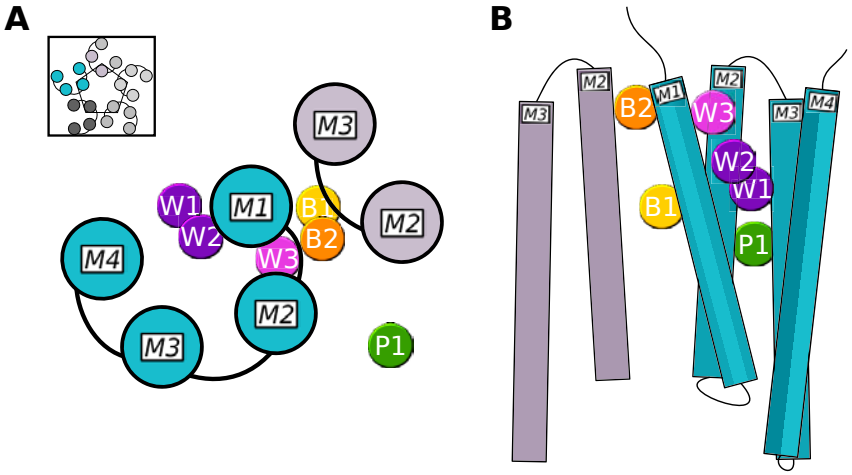

Figure 1.10 – Location of the anesthetic binding sites highlighted in pLGICs transmembrane domain. Top (A) and side (B) views of one and a half subunit transmembrane domain showing the three intrasubunit binding sites (W1 to W3), the intersubunit site (B1) present in the GlyR and in GLIC F238A mutant, Nury *et al.*'s linking tunnel (B2) and the pore site (P1).

alcohols, covering these three sites. For convenience, we introduce here a nomenclature for these binding sites that will be consistently numbered as summarized in figure 1.10 and referred to by their number.

A brominated variant of the local anesthetic lidocaine, which is known for its pore blockage properties, as well as several cations were co-crystallized with GLIC few years ago (Hilf *et al.*, 2010). Docking based data suggested GLIC's pore blockage by propofol and isoflurane with a micromolar affinity (Brannigan *et al.*, 2010; LeBard *et al.*, 2012). More recently, the co-crystal structure of ELIC with bromoform provided another experimental evidence that anesthetics can bind pLGIC's pore (Spurny *et al.*, 2013). Inhibition through the pore can be understood intuitively and two distinct mechanisms probably coexist. Open-channel blockers, such as lidocaine, are most often charged molecules that carry the same charge as the permeating ion and block the pore by mimicking ion permeation up to the point where they sterically prevent ion conduction and jam the channel in an open conformation. On the other hand, general anesthetics such as propofol, desflurane or isoflurane are mostly believed to block the pore in an allosteric fashion, either selecting or favoring, then stabilizing a closed conformation. However, from a mechanistic point of view, the hypothesis that pore blockage/closing alone is at the origin of the anesthetic-induced effects is difficult to reconcile with mutagenesis data indicating that mutations in the intrasubunit pocket region affect anesthetic action (Nury *et al.*, 2011). Binding sites in this region should therefore closely be examined, too.

Propofol, desflurane and bromoform bind the intrasubunit pocket in GLIC (Nury *et al.*, 2011; Sauguet *et al.*, 2013a; Chiara *et al.*, 2014). While propofol and desflurane binding poses are virtually the same, bromoform adopted three distinct poses in the cavity: site W1 (the W standing for *within* the subunit) overlaps the propofol and desflurane binding site; site W2 lies closer to the M1 helix and partially overlaps site W1; site W3 is deeper inserted in the cavity, between the M1 and M2 helices, at the interface between the intra- and intersubunit regions. Ethanol, 2-bromo-ethanol and bromoform were found to bind an intersubunit cavity in the structure of a GLIC ethanol sensitive variant, namely the mutant F238A. Ethanol, 2-bromo-ethanol and bromoform were shown to bind an ethanol-sensitive variant of GLIC, namely the mutant F238A, by docking in an intersubunit cavity (Sauguet *et al.*, 2013a) that will be referred to as site B1 (the B standing for *between* the subunits). This site had been previously suggested for ethanol binding to the glycine receptor by MD simulations (Murail *et al.*, 2011). GAs and alcohols binding to these sites produce opposite effects on channel function. In GLIC, the intersubunit site B1 is thought to be potentiating (Sauguet *et al.*, 2013a; Brömstrup *et al.*, 2013; Murail *et al.*, 2012; Howard *et al.*, 2011a) while the intrasubunit W1 site is inhibitory (Nury *et al.*, 2011). Another intersubunit site has been suggested for propofol binding to $GABA_AR$ by photolabelling at the ECD and the TMD interface (Yip *et al.*, 2013), that had been suggested for desflurane based on MD data (Nury *et al.*, 2011). This site will be referred to as site B2.

A founding hypothesis of my work is that anesthetics modify the equilibrium between the channel open and closed states, complying with the definition of an allosteric modulator (figure 1.11). However, it should be mentioned that some experts believe that anesthetics block ion flow by sterically obstructing the channel. A consensus exists for some anesthetics such as lidocaine which is believed to be an open channel blocker. On the other hand, some anesthetics such as propofol for example have been proven to bind to GLIC's intrasubunit pocket (Nury *et al.*, 2011) and suggested to bind the pore as well (LeBard *et al.*, 2012). Propofol is therefore believed by some to be a steric channel blocker and not an allosteric inhibitor.

The emerging picture is that modulation is the effect of competitive binding between the intersubunit

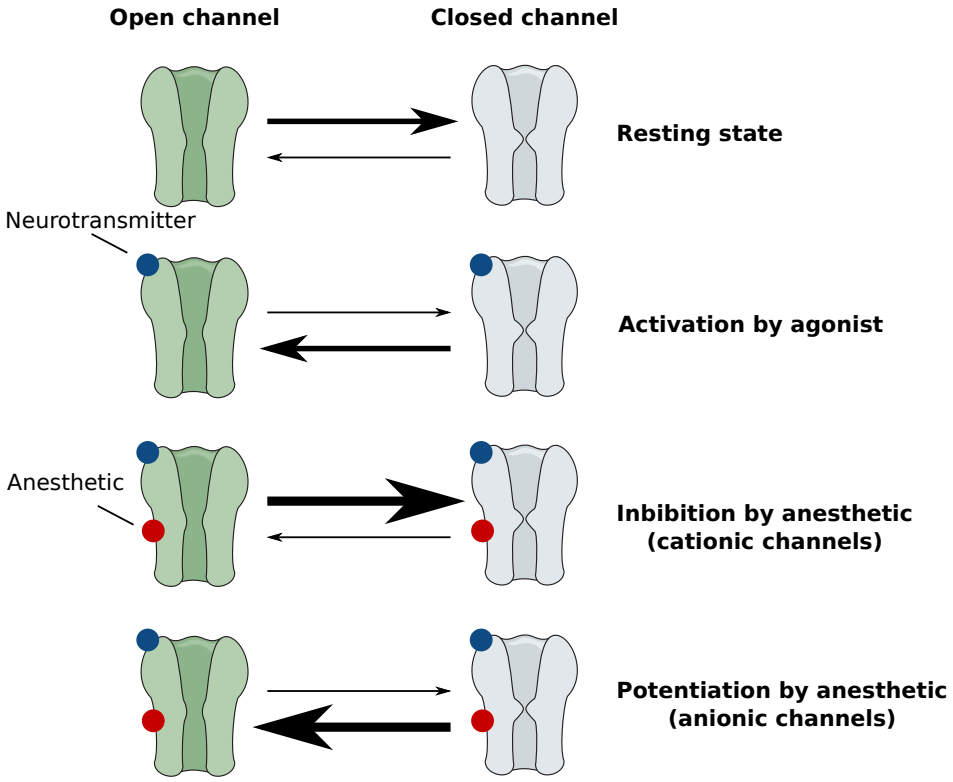

Figure 1.11 – Hypothesis of anesthetic action on the function of ionotropic channels. Neurotransmitters as well as anesthetics are believed to change the equilibrium between the channel conformations *i.e.* open and closed. The neurotransmitter modifies the equilibrium in favor of the open state. At excitatory channels, anesthetics favor the closed state, preventing cations to enter the cell therefore the increase of the membrane potential. At inhibitory channels, they favor the open state, allowing anions to enter the cell which have the effect of decreasing the membrane potential. At excitatory as well as at inhibitory channels, anesthetic action is therefore to inhibit the transmission of the action potential.

potentiating site and the intrasubunit inhibitory site, which is consistent when applied to mammalian pLGICs since GAs and alcohols potentiate inhibitory channels $GABA_A$Rs and GlyRs, while they inhibit the excitatory nAChRs. Although, despite the accumulation of crystal structures of general anesthetics bound to a member of the pLGIC family, the molecular mechanism of allosteric inhibition by anesthetics binding to intra- and intersubunit pockets is still poorly understood. Regarding the probable existence of general anesthetic binding sites in pLGICs pore, channel modulation is to be regarded as the combined effect of binding to the intra, inter, and pore regions.

1.4 Context in October 2010

In the next few lines, I attempt to provide some context for the choices I made when I started to work on the project, in October 2010.

A close collaboration was initiated in 2009 between Marc Delarue's group at Institut Pasteur and my PhD supervisor, Marc Baaden. In January 2011, Marc Delarue's group released the first structure of general anesthetics bound to GLIC (Nury *et al.*, 2011). His group co-crystallized propofol and desflurane bound to

GLIC's intrasubunit pocket.

As electrophysiology measurement showed marked differences in the modulation of several GLIC mutants by these GAs, Marc Baaden was in charge of running MD simulations aiming to understand this phenomena. These simulations yielded interesting but incomplete data on anesthetic dynamics while bound to the receptor, mainly suggesting that channel closure could be caused by the repetitive contacts between GAs and the M2 helices.

We decided to center the start of my thesis on three principal aspects: i) the extensive characterization of desflurane and propofol dynamics bound to the crystallographic site; ii) understanding why propofol inhibits more the T255A mutant than the WT GLIC; iii) understanding why desflurane has an opposite effect than propofol on this mutant, *i.e.* it is less effective on the mutant than on the WT.

In early 2013, our collaborators at Institut Pasteur came back to us with several structures of bromoform bound to GLIC, displaying among other previously unseen sites, a bromoform molecule bound to the pore of GLIC in LC conformation. As I had developed a suite of tools that allowed me to rapidly launch MD simulations on a system of interest and efficiently analyze the results, our collaborators asked us to characterize GLIC's inhibition by bromoform.

In the next two chapters, I will introduce the methods I used and developed during this project. The characterization of GLIC's inhibition by bromoform will be detailed in chapter 4. The following chapter will be devoted to the study of propofol and desflurane dynamics while bound to GLIC. Finally, I will end the present manuscript with some concluding remarks.

Molecular Modeling: Theory And Practice

2

In this chapter, I introduce the principal method I have been using in this work: molecular modeling. Molecular modeling includes a wide range of methods, from mixed experimental-theoretical to purely theoretical ones.

As an example of a so-called *mixed method*, Nuclear Magnetic Resonance (NMR) and X-ray crystallography are widely used to determine the three-dimensional structures of molecules. Both methods have a strong *in vitro* component including protein expression, purification and raw data acquisition. Since none of these techniques routinely allow to precisely determine the position of atoms from raw data, *in silico* models are used to fit atoms in the signal acquired from the machines (Trabuco *et al.*, 2008; Brünger *et al.*, 1998). Pure *in silico* methods developed during the last decades faced with the lack of *in vitro* methods able to describe molecule dynamics at the atomistic scale. MD simulations are a member of the large family of pure *in silico* methods, further including Brownian Dynamics, Normal Modes Analysis, or docking for example. These methods belong to the family of molecular mechanics methods that require the use of a force field (see section 2.1) to describe the interactions between a system of particles. Despite their empirical nature, these methods have proven their ability to reproduce data obtained from "wet lab" experiments and are used to answer questions that *in vitro* and *in vivo* procedures cannot.

In 2013, the Nobel Prize in Chemistry was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *for the development of multiscale models for complex chemical systems*, an acknowledgment that *in silico* methods are to be considered as an essential tool that can, together with *in vitro* and *in vivo* methods, address some of the most challenging questions of our time.

This chapter is divided into four sections. The first three sections briefly introduce respectively the concepts of force field, MD simulation and free energy calculations. Finally, the fourth section will be devoted to challenges mostly related to the simulation of biological systems.

2.1 Force Fields

As opposed to quantum mechanical representations that aim to describe the dual particle-like and wave-like behavior of energy and matter, molecular mechanics representations use classical mechanics to model a molecular system. Molecular mechanics methods require the use of a force field that describes the interactions between a system of particles with contributions of processes such as the stretching of bonds,

$$V_{\text{potential}} = \sum_{\text{bonds}} \frac{k_i}{2}(l_i - l_{i,0})^2 +$$

$$\sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 +$$

$$\sum_{\text{torsions}} \frac{V_n}{2}\left[1 + \cos(n\omega - \gamma)\right] +$$

$$\sum_{i=1}^{N-1}\sum_{j=i}^{N} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\mathbf{r}_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{\mathbf{r}_{ij}}\right)^{6}\right] +$$

$$\sum_{i=1}^{N-1}\sum_{j=i}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 \mathbf{r}_{ij}}$$

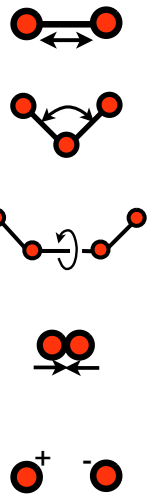

Figure 2.1 – Equation of potential energy and schematic representation of the various contributions. The total potential energy is the sum of bonded (bonds, angles, torsions) and non-bonded (steric, electrostatic) interactions.

the opening and closing of angles, the rotations about single bonds and non-bonded interactions between particles.

It should be noted that some methods do not account for all these components. As an example, in *rigid* docking bonded interactions are often ignored while classical spring network models do not explicitly account for electrostatic interactions (Tirion, 1996; Bahar *et al.*, 1997; Hinsen, 1998).

The classical form of a molecule's potential energy is shown in the illustrated equation of figure 2.1. For each component, energy penalties are associated with the deviation from equilibrium values. The resolution of these equilibrium values, even when based on experimental data, is often adjusted to fit a molecule's macroscopic properties that can be measured thanks to *in vitro* experiments. This process is called *parametrization*. For example, during the parametrization of small molecules such as anesthetics, it is common to use the density of a pure solution of this molecule and the solvation energy of this molecule in water as target properties. As another example, lipids are often parametrized in such a way that the order parameter of each carbon atom fits *in vitro* data.

As a consequence and since force fields are parametrized to fit a finite set of properties that can be different from one force field to another, it is not surprising that a given force field may perform better reproducing certain properties compared to another one (and conversely). It is then crucial to choose with precaution which force field is the most appropriate to reproduce the properties one wants to study. For example, several studies highlight the fact that the Amber03 force field overstabilizes helical structures (Cino *et al.*, 2012; Lindorff-Larsen *et al.*, 2012) while it has been suggested that the OPLS-AA force field may be biased in favor of bends over helices (Cao *et al.*, 2011; Cino *et al.*, 2012; Vamparys, 2013).

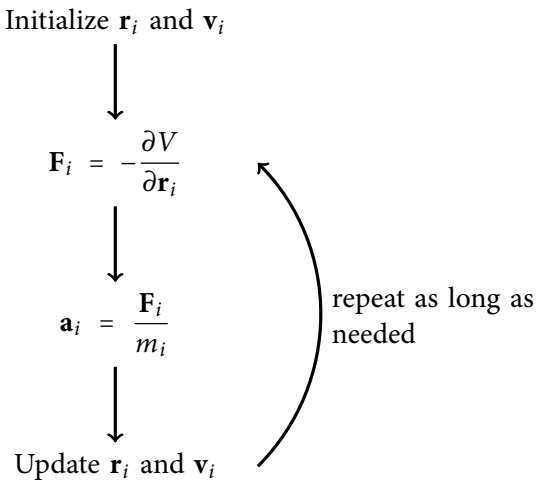

Figure 2.2 – The algorithm underpinning molecular dynamics simulations. With $\mathbf{r}_i$ the cartesian coordinates of the atom i, $\mathbf{v}_i$ its velocity, $\mathbf{F}_i$ the net force acting on it, V the potential energy applied to it, $\mathbf{a}_i$ the acceleration applied to it and m_i its mass.

2.2 Molecular Dynamics Simulations

Molecular Dynamics (MD) is a computer simulation method that integrates Newton's laws of motion to calculate successive configurations of a system, resulting in a trajectory of atom coordinates and velocities over time.

The second law of motion states that *the acceleration of a body depends directly upon the net force acting upon the body, and inversely upon the mass of the object*. Thus $\mathbf{F} = m\mathbf{a}$, where $\mathbf{F}$ is the net force acting on the object, m its mass and $\mathbf{a}$ its acceleration. Thanks to a force field, $\mathbf{F}$ can be calculated for each particle of a system. Since the particle mass is known, its acceleration can be calculated, therefore its position at the next iteration can be deduced. This process, called *integration of equations of motions* can be performed several times to obtain a trajectory for each particle in the system (figure 2.2).

The integration of motion equations can be done thanks to several integrators, each one having specific properties. Next I will develop a few examples of such integrators.

Because MD is an *in silico* method, the software used to run the calculations plays a central role in the data acquisition. During my thesis these issues proved important. In this part I will introduce some differences that exist in the main two software suites that I used for MD simulations, namely NAMD and GROMACS.

2.2.1 Equation of motion integration algorithms

This part could lead us into the deepest pits of classical physics and mathematics. I chose to present only a few key concepts in order to highlight some of the most important properties that should be considered before running an MD simulation.

There are a variety of algorithms that can be used to solve ordinary differential equations. The most basic methods are probably the *Euler method*, named after Leonard Euler who proposed it in the late 17th century, and its generalization by C. Runge and M. W. Kutta, called the *Runge-Kutta* method.

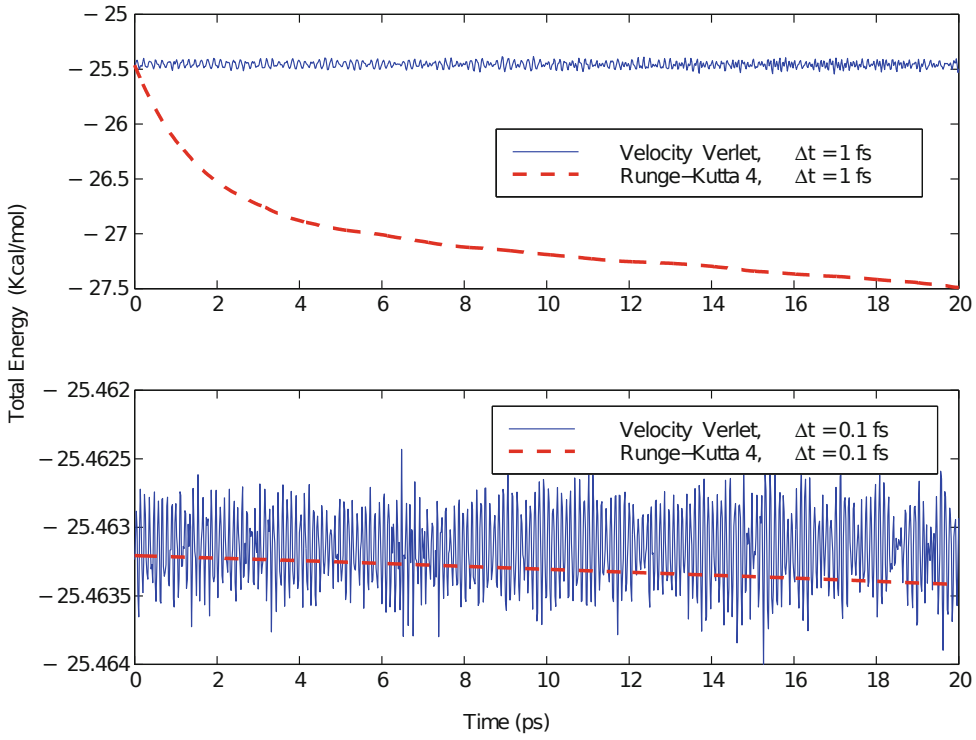

Figure 2.3 – Sympletic *vs* non-symplectic integrators. Energy evolution of a water tetramer simulated by the symplectic Verlet scheme (solid line) versus the non-symplectic Runge-Kutta integrator (dashed line) at two time steps (0.1 and 1 fs). From Schlick (2010).

These methods are quite intuitive and easy to implement, but do not feature one of the most important properties an integrator should have in MD: they are not symplectic integrators. Sympletic integrators are integrators that preserve specific properties associated with the Hamiltonian system of differential equations, including its value *i.e.* its energy (Schlick 2010, section 14.2). In practice, the total energy is not preserved exactly but the energy error remains contained over time, which is different from non-symplectic integrators which display a systematic energy drift over time. Runge-Kutta or Euler integrators are therefore not to be used in MD.

The most popular methods for integrating Newton's laws of motions derives from the Verlet method (Verlet, 1967) in which the positions are given by

$$\mathbf{r}(t+\Delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\Delta t) + \frac{\mathbf{F}}{m}\Delta t^2 \tag{2.1}$$

with Δt the time between two snapshots. The velocities are given by

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\Delta t) - \mathbf{r}(t-\Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2) \tag{2.2}$$

with $\mathcal{O}(\Delta t^2)$ being the interpolation error. NAMD implements the *velocity-Verlet* algorithm (Phillips *et al.*, 2005) while GROMACS' default MD integrator is *leap-frog* (Pronk *et al.*, 2013). GROMACS also implements a variety of integrators which are mostly variants of either leap-frog or velocity-Verlet, each one aiming to either increase accuracy or efficiency under certain circumstances.

2.2.2 Integration time step

The upper limit for the integration time step Δt in the integration scheme depends on the fastest motions in the system. For biological systems under biological temperature and pressure conditions, these motions are light-atom bond vibrations, which are on the order of 10 fs (Schlick 2010, section 14.2.3). It is usually accepted that Δt has to be one order of magnitude lower than the fastest motions in the system, *i.e.* 1 fs in this case. As the amount of CPU-time[1] a user can spend calculating an MD simulation is fixed and since the computation time of one time step is constant, it is therefore useful to be able to increase the value of Δt to speed up the simulation: computing 1,000,000 simulation steps with Δt = 1 fs will output a 1 ns long simulation trajectory, while, with Δt = 2 fs, 2 ns can be calculated in the same amount of time.

The integration time step can be increased by treating bond stretching degrees of freedom as rigid. The traditionnal algorithm to constrain bonds, implemented in NAMD, is SHAKE (Ryckaert *et al.*, 1977). The LINCS algorithm (Hess *et al.* 1997, a default in GROMACS), has proven of higher efficiency and presents better convergence properties. Both algorithms solve the same problem *i.e.* resetting coupled constraints after an unconstrained update. Interestingly, a third algorithm optimized for rigid water molecules, named SETTLE (Miyamoto and Kollman, 1992), has been implemented in both software packages and is commonly used in the simulations of biological solutions.

An alternative scheme to treat high-frequency vibrational modes is to separate the calculation of the force on a particle into two components: short-range and long-range forces, the underlying idea being that long-range forces vary more slowly than short-range forces. This technique, known as *multiple-time-stepping*, allows to use different time steps for bonded and non-bonded interactions. NAMD implements this idea separating bonded forces, Lennard-Jones and short-range electrostatic forces and finally long-range electrostatic forces in three different loops. Typical multiple time step example values would be 2 fs, 2 fs and 6 fs. Multiple-time-stepping is not yet implemented in GROMACS.

2.2.3 Non-bonded interactions under periodic boundary conditions

Periodic boundary conditions

Without specific boundary conditions, the simulation of a molecular system would take place in vacuum: particles at the system's border are surrounded by nothing. At a moderately short time scale, this would lead to serious artefacts because the whole system would diffuse in the infinite space of cartesian coordinates. The solution to this issue is to circumscribe the system in a box just large enough to encompass every particles of the system. But this would lead to other artefacts: border particles would repeatedly collide with the box, which would strongly impact their behavior. Because the size of the systems that can be currently simulated with MD is still so small, the abnormal behavior of border particles would finally impact little by little every components of the system. A solution to this problem is to periodically repeat the simulation cell, a technique known as Periodic Boundary Conditions (PBC). A particle exiting on one side of the cell enters from the opposite face with the same velocity (figure 2.4). Besides, border particles on one side of the cell interact with particles at the opposite side. This technique therefore simulates an

[1]Cost of a simulation, expressed as the amount of time for which a Central Processing Unit (CPU) was used processing instructions of a computer program (definition from wikipedia.org). For example, a calculation that was run on 128 CPUs and lasted 2 hours consumed $t = 2 \times 128 = 256$ CPU-hours.

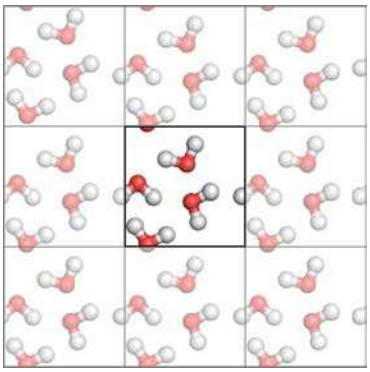

Figure 2.4 – Periodic boundary conditions. The initial cell (central) is replicated in each direction so that border particles interact with particles from the neighboring cell. Particles that exit a cell on one side enter the cell on the opposite side.

infinite system, but interactions with contributions from an infinite number of neighboring images have to be calculated, with consequences on computational efficiency.

Non-bonded interactions

The complexity of computing non-bonded interactions is $\mathcal{O}(N^2)$ where N is the number of atoms. This is because non-bonded interactions have to be computed between all pairs of atoms. Such a complexity implies a huge computational cost: Schlick (2010) estimates to roughly 2 years the CPU-time required to calculate a single nanosecond of a 10,000-atom system. Fortunately, techniques have been developed to reduce the dramatic cost of computing non-bonded interactions without decreasing the accuracy of simulations of biomolecules in solvent.

The first technique is to use a spherical cutoff scheme, which is easy to implement and cheaper than brute calculation ($\mathcal{O}(N)$). There are three categories of cutoff techniques, all of them setting the contribution of atoms remote from each other of a distance $r > b$ to 0. The difference between the techniques are their behavior when $r <= b$ (figure 2.5):

- the truncation approach does not change the value of the energy when $r <= b$, therefore abruptly sets the interaction energy to 0 when $r > b$,
- switching schemes smoothly change the energy value for $a <= r <= b$,
- shift functions alter the function for all $r < b$.

In simulations of biomolecules, spherical cutoff schemes are mostly used for representing van der Waals interactions, which are quite short-range interactions.

Electrostatic interactions are long-range interactions that play a critical role in biomolecules. The Particle Mesh Ewald (PME) approach has revolutionized biomolecular simulations, reducing the computational cost of calculating long-range interactions from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ (Darden *et al.*, 1993). This technique, which has been developed specifically to be used under PBC, employs direct calculation for short-range interactions while long-range interactions are calculated in reciprocal space thanks to Fast Fourier Transforms (FFT).

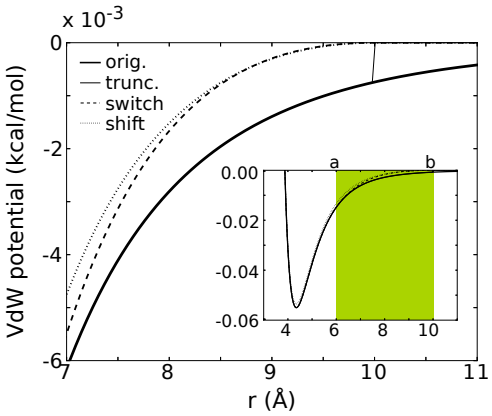

Figure 2.5 – Spherical cutoff schemes for non-bonded interactions. Van der Waals potential for various cutoff schemes with buffer region 6-10 Å. Example of a $C_\beta - C_\beta$ interaction with parameters taken from the CHARMM program. Adapted from Schlick (2010).

2.2.4 Statistical ensembles: thermostats and barostats

Biological organisms live under strict conditions of temperature and pressure. To mimic a biological environment, it is fundamental to be able to reproduce the atomic vibrational movement associated with a given temperature and pressure. It is therefore very common for simulations of biological interest to run in the NPT ensemble, *i.e.* an ensemble where the Number of particles, the Pressure and the Temperature are retained constant.

Temperature coupling

The temperature T of a system is related to its kinetic energy, which is given by:

$$E_{\text{kinetic}} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2}\nu k T \tag{2.3}$$

where ν is the number of degrees of freedom, *i.e.* 3 per atom (velocities along x, y and z axes)[2] minus the number of constraints applied to the system. For example, for a system of N atoms with fixed bond length and center of mass momentum removal, $\nu = 3N - N_{\text{bonds}} - 3$. Controlling the temperature can therefore be done by a simple scaling of the velocities at each time step. However, algorithms coupling to an external bath are often preferred because they allow the temperature to fluctuate about the desired temperature.

The Berendsen *weak-coupling* scheme (Berendsen *et al.*, 1984) is a widely used algorithm. It has been implemented in several MD software packages such as NAMD and GROMACS. However, while it performs fast temperature equilibration, it does not generate rigorous canonical averages (Leach, 2001). Different canonical algorithms are therefore often used for the production phase, such as the Andersen thermostat (Andersen, 1980) or the Nosé-Hoover scheme (Nosé, 1984; Hoover, 1985). GROMACS also implements a *velocity rescaling* method which is very similar to a Berendsen thermostat with an additional stochastic term that ensures a correct canonical ensemble (Bussi *et al.*, 2007).

[2] There are 6 degrees of freedom per atom in the system, 3 for its velocity components plus 3 for the positions. As the kinetic energy does not depend on atomic positions, here each atom accounts for 3 degrees of freedom.

Pressure coupling

The instantaneous pressure $p(t)$ is computed as the trace

$$p = \frac{1}{3} \text{Trace } \mathbf{P} \tag{2.4}$$

of the pressure tensor

$$\mathbf{P} = \frac{2}{V}\left(\frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v}_i \otimes \mathbf{v}_i - \Xi\right) \tag{2.5}$$

from the virial tensor

$$\Xi = -\frac{1}{2}\sum_{i<j} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij}. \tag{2.6}$$

Constant pressure simulations require periodic boundary conditions (see section 2.2.3). Pressure is controlled by adjusting the volume, tuning both the particle coordinates and the box vectors at each time step. The box size can be adjusted in different ways. For the simulation of membrane proteins, the box's dimension are often constrained to vary semi-isotropically *i.e.* the X and Y dimensions are adjusted identically while the Z dimension is adjusted independently.

Both NAMD and GROMACS implement the Berendsen pressure coupling. It is to be noted that, according to the GROMACS manual, *although this algorithm yield to the correct average pressure, it does not yield the exact NPT ensemble and it is not yet clear exactly what errors this approximation may yield*. GROMACS implements two alternatives: the Parrinello-Rahman approach (Parrinello and Rahman, 1981; Nosé and Klein, 1983) and the MMTK approach (Martyna *et al.*, 1996), that can be used only with specific integrators. As for NAMD, it provides constant pressure simulation using a modified Nosé-Hoover method in which Langevin dynamics is used to control fluctuations in the barostat.

In the last few sections, I intended to highlight some differences that exist between two widely used software packages for MD simulations, NAMD and GROMACS. The implicit idea behind this comparison is that it may be difficult to run simulations with the exact same setups using these programs, as they often do not implement the same algorithms on the one hand, and as even when they do, different implementations of the same algorithm may lead to numerical differences *e.g.* due to rounding, which can have dramatic consequences in numerical simulations. As an example, Marc Baaden's group has been the first one to my knowledge to raise an important issue concerning the simulations of the GLIC channel. He and Samuel Murail, noticed that in all simulations calculated with the NAMD software, water molecules escape GLIC's pore after approximately one hundred nanoseconds, which is not the case when running simulations with GROMACS in which case the pore remains hydrated on the microsecond timescale. Detailed investigations have, to date, not allowed to precisely determine the origin of these different behaviors (see also section 2.5.4).

2.3 Free energy calculations

The free energy is a highly desirable quantity to compute because it describes quantitatively the thermodynamical equilibrium. It allows especially to probe states of a system that are not accessible experimentally. As an example, several hypotheses coexist on the location of GLIC binding sites for anesthetics. Alchemical transformations constitute a precious tool to calculate free energies of binding, potentially allowing to discriminate binding sites with a lower affinity from high-affinity ones.

In this work, free energy calculations were performed using the thermodynamic integration method that allows to calculate free energy differences between two states of a system. As depicted by figure 2.6, while the free energy of binding of a ligand to a receptor cannot be directly computed, this quantity is accessed employing a thermodynamic cycle. The thermodynamic method requires independent simulations in which the Hamiltonian $\mathcal{H}$ of the system slowly changes from that describing system A to that describing system B. This modification of $\mathcal{H}$ is a function of a coupling parameter λ that ranges from 0 where $\mathcal{H}(\mathbf{p}, \mathbf{r}; 0) = \mathcal{H}^A(\mathbf{p}, \mathbf{r})$ to 1 where $\mathcal{H}(\mathbf{p}, \mathbf{r}; 1) = \mathcal{H}^B(\mathbf{p}, \mathbf{r})$, where $\mathbf{p}$ is the particle momenta and $\mathbf{r}$ is the particle cartesian coordinates.

The free energy F of a system at thermodynamical equilibrium is related to the partition function Q of the system of an NVT ensemble (constant number of particles, volume and temperature):

$$F(\lambda) = -\frac{1}{\beta} \ln Q \qquad \text{with} \quad \beta = \frac{1}{k_B T} \tag{2.7}$$

$$Q = c \iint \exp\left[-\beta \mathcal{H}(\mathbf{p}, \mathbf{r}; \lambda)\right] d\mathbf{p} d\mathbf{r} \qquad \text{with} \quad c = \frac{1}{h^{3N} N!} \tag{2.8}$$

where k_B is Boltzmann's constant, T is the temperature, N is the number of particles in the system and h is Plank's constant. The Hamiltonian $\mathcal{H}(\mathbf{p}, \mathbf{r}; \lambda)$ gives the total energy of the system in a given configuration, *i.e.* a given set of momenta and coordinates at a given value of coupling constant λ (Christ *et al.*, 2010).

Compared to the free energy F, the Gibbs free energy is related to the partition function Δ of an NPT ensemble, which, at thermodynamical equilibrium, is

$$G(\lambda) = -\frac{1}{\beta} \ln \Delta \tag{2.9}$$

$$\Delta = c \iiint \exp\left[-\beta \mathcal{H}(\mathbf{p}, \mathbf{r}; \lambda) - \beta p V\right] d\mathbf{p} d\mathbf{r} dV \tag{2.10}$$

where p and V are the pressure and the volume of the system, respectively. The difference in free energy between states A and B of the system are

$$\Delta G_{BA}(p, T) = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_{NpT;\lambda} d\lambda \tag{2.11}$$

The value of ΔG_{BA} can be calculated by running several independent simulations with different values of λ. The energy variations between states is calculated in each window using the Bennett's Acceptance Ratio (BAR) method (Bennett, 1976) and the total energy variation is $\Delta G_{BA} = \sum_i \Delta G_i$, with i the window index. Free energy of binding calculations are no exception to the ergodicity principle: the value obtained is expected to be accurate as long as the sampling is adequate. The number of steps computed in each

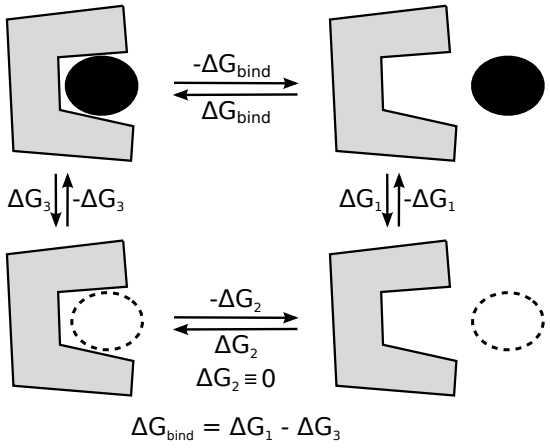

Figure 2.6 – A thermodynamic cycle for the computation of the free energy of binding of a ligand to a receptor. ΔG_{bind} can be accessed by calculating ΔG_1 and ΔG_3. Adapted from Brömstrup *et al.* (2013).

frame has therefore to be adjusted to the system (the bigger a system is, the longer a simulation has to be run to obtain the necessary sampling).

2.4 Application: bromoform force field parameterization

2.4.1 Principle

The algorithm of the simulation of a biped walking could be constructed from two steps that would repeat endlessly: 1) move left foot, 2) move right foot. Now to extract data from such a simulation as for example "how long would it take the biped to walk from Paris to Strasbourg?", parameters have to be provided to the simulation such as the biped speed and step size for example. For this very basic example, the parameter (*i.e.* the biped speed) can be measured accurately. However, it is not always the case for molecules.

As for all numerical simulations, MD simulations obey similar rules and the representation of the molecules that compose a system requires parameters such as the bond distances, the valence angles, the dihedral angles, the atom charges and van der Waals radii, *etc.* Several of these parameters can be measured experimentally with more or less accuracy. For the purposes of biological simulations, molecule force field parameters are usually adjusted so that the molecule's macroscopic properties match the experiment, as developed in section 2.1.

As it is often the case for small molecules that are not widespread in the biological medium, bromoform parameters were not available in any of the common force fields and had therefore to be computed.

2.4.2 Methods

The parameters I chose to target for bromoform were its density at 25 °C, its vaporization enthalpy and its solvation free energy, which is classic for such small molecules.

Density at 25 °C

The density of bromoform at a given temperature can be accessed thanks to simulations of pure bromoform solutions, *i.e.* composed only of bromoform molecules. As both the number of bromoform molecules in the box N and the box volume V are known, the density of the solution at equilibrium is given by

$$\rho = \frac{N \times M}{N_A \times V} \tag{2.12}$$

where N_A is Avogadro's constant (N_A =6.02 $\times$ 10^{23} mol^{-1}) and M is the molar mass ($M_{bromoform}$ = 253.73 g/mol).

To calculate bromoform density, I chose to run a single simulation of 1000 bromoform molecules in a 5 $\times$ 5 $\times$ 5 Å box for 5 nanoseconds.

Vaporization enthalpy

The vaporization enthalpy can be easily calculated from the simulation of a pure bromoform box. It corresponds to the sum of non-bonded energies divided by the number of bromoform molecules in the box. Vaporization enthalpy was extracted from the same simulation as the density.

Solvation free energy

Solvation free energy can be accessed thanks to a thermodynamical integration. A single bromoform molecule is solvated and simulated in a water box. Bromoform charges and steric interactions with water are slowly turned off. The energy difference between the initial and the final states corresponds to the solvation free energy.

A single bromoform molecule was solvated in a cubic water box containing 343 water molecules. Free solvation energy was calculated using similar setup as developed in section 2.6.3.

2.4.3 Results

The final parameters are the result of successive iterations and parameter refinements to match as well as possible the desired properties.

The parametrization results I obtained are summarized in table 2.1. Some simulation values, such as the density or the vaporization enthalpy match experimental values quite well. Notably, the value of the dipole moment, which can be calculated from the molecule geometry, is very sensitive to the set of atom charges and to the angle between Br-C-Br atoms. We therefore judge this difference of 29 % acceptable, since a difference of more than 20 % is observed for the dipole moment of TIP3P water, a well established model of water molecule. Furthermore, it is recommended that force field parameters for MD simulations of condensed phase properties overestimate the gas-phase dipole moment of a given molecule by as much as 20 % (Vanommeslaeghe *et al.*, 2009).

The van der Waals parameters of bromoform carbon and hydrogen atoms have been taken from the Amber99sb-isbn force field atom types CT and H1, respectively. Bromine van der Waals parameters come from the OPLS force field. Atom charges have been calculated thanks to the R.E.D tools (Dupradeau *et al.*, 2010). Bond lengths have been taken from Lide (2003) as well as the value of the Br-C-Br angle. The value

	Experience	Simulation	Δ (%)
Density at 25 °C	2.879 g/cm^3	2.932 g/cm^3	**1.8**
Vaporization enthalpy	46.05 kJ/mol	42.03 kJ/mol	**−9.6**
Solvation free energy	−8.9 kJ/mol	−6.2 kJ/mol	**−43.5**
Dipole moment	0.99 D	1.39 D	**28.8**

Table 2.1 – Bromoform parametrization result summary.

of the Br-C-H angle has been calculated to match a perfect tetrahedron, according to the Br-C-Br angle. The corresponding GROMACS parameter (`.itp`) file is available in appendix B.1.

2.5 Difficulties

Biological solutions are intrinsically complex and crowded mixtures. Some groups aim at modeling such systems in their full complexity, as illustrated by a snapshot of a recent simulation by McGuffee and Elcock (2010) endeavoring to understand diffusion properties of the bacterial cytoplasm (see figure 2.7). However, such studies are currently limited to simplified approaches, such as Brownian Dynamics of rigid molecules.

Focusing on the case of GLIC, I discuss in this part several challenges for the study of biological membrane protein solutions. Simplifications are necessary in constructing the model, in defining the composition of the system and in choosing the concentration of each species. Uncertainties remain since neither experiments nor calculations can resolve issues such as reliably choosing the protonation states of each of GLIC 's titratable groups. I discuss intrinsic properties of the models that remain unclear at this time such as the behavior of water in hydrophobic nanoconfinement.

This part has been largely inspired by a review article I wrote and published in Laurent *et al.* 2013 (see appendix B.2).

2.5.1 System composition

Biological solutions of interest contain several ingredients including proteins, nucleic acids, sugars, ions, water and other small molecules (alcohols, anesthetics, *etc.*). Depending on the goal a study aims to achieve, it might be necessary to include all these compounds in a model used for MD simulations.

Cells communicate with each other thanks to membrane-embedded receptors whose hydrophobic transmembrane domains are stabilized by the membrane environment. Hence, lipid bilayers or bilayer mimetics have to be included in a realistic model. Typical biological membranes are themselves complex ensembles composed of lipids, carbohydrates, proteins and cholesterol in the case of eukaryotic cell (van Meer *et al.*, 2008). The proportion of each molecular species depends on the cell type. Furthermore, different types of phospholipids are mixed in the membranes, with proportions again depending on the cell type and possible leaflet asymmetry. Matrix protein networks may further structure the membrane. To make the membrane picture more intricate, microdomains enriched in cholesterol, also called *lipid rafts*, could play a role in the cellular function by compartmentalizing specific lipids and membrane proteins. However the existence and role of such microdomains are very controversial.

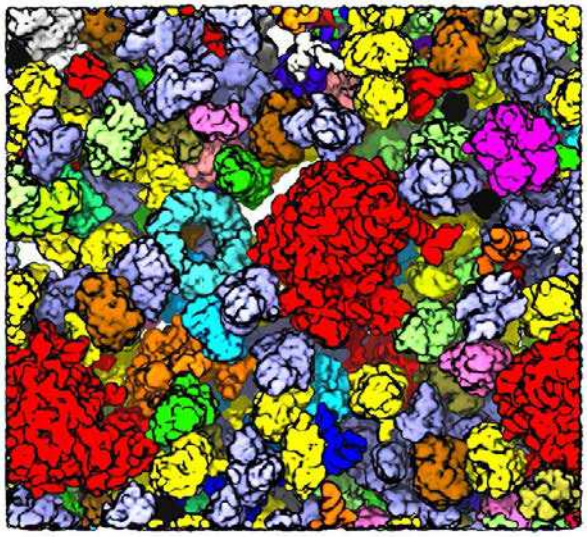

Figure 2.7 – Biological solutions are crowded mixtures. Snapshot of a bacterial cytoplasm model whose dynamics has been studied by Brownian Dynamics (courtesy of Dr. Adrian Elcock; McGuffee and Elcock 2010).

Molecule	Approx. percentage
cholesterol	29
phosphatidylcholine	28
phosphatidylethanolamine	26
phosphatidylserine	10
phosphatidylinositol	2
sphingomyelin	2
glycolipids	3

Table 2.2 – A human synaptic membrane composition. Approximative ratios as described by Wood *et al.* (1989)

Although some MD simulation studies now attempt to accurately model the complexity of *e.g.* the *Escherichia coli* bacterial membranes (Piggot *et al.*, 2011), generally speaking such a complexity is very difficult to reproduce and the composition of the membrane is often simplified by using a single type of phospholipid. As an example, the GLIC simulation system is composed of the protein embedded in a fully hydrated phosphatidylcholine membrane, whereas a typical synaptic plasma membrane composition is diversified as described in table 2.2

Unfortunately, our knowledge about lipid compositions of various organisms and cell types is still very limited, yet it is established that lipid *building blocks* have the potential to generate up to tens of thousands of different molecular species (Shevchenko and Simons, 2010). This situation may improve with recent efforts in the field of lipidomics (Alex Brown, 2012).

On the order of 150 mM NaCl or KCl ions are typically added to mimic physiologic electrolyte concentration. Sometimes higher concentrations up to 1 M are used to enhance the probability of processes such

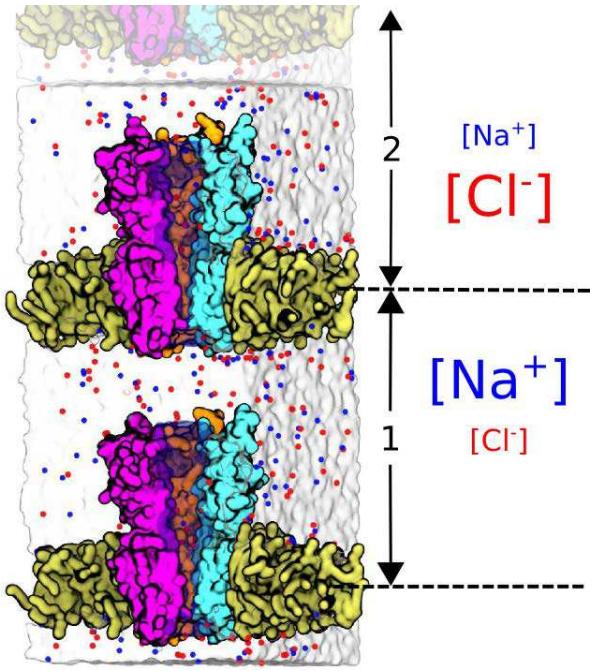

Figure 2.8 – System set-up used to study ion permeation through GLIC according to the double bilayer method. The two compartments display different ion concentrations that create an electric potential between them, driving the ions through the channel. Only three of the five protein subunits and half of the membrane are represented in order to expose the pore interior, displayed as a blue transparent surface.

as ion permeation. These processes may be driven by cross-membrane potentials which are typically in the range of −90 to −40 mV. To reproduce such a potential in the simulation may require special methods such as applying an electric field using additional forces on all charged particles (Gumbart *et al.*, 2012). Maintaining a charge imbalance between two solution compartments has initially been proposed by Sachs *et al.* (2004). This approach has subsequently been used by several groups to study the permeation of ion channels (Kutzner *et al.*, 2011; Lee *et al.*, 2008). Herrera and Pantano (2009) proposed a variation of the method where ionic motion is restricted anisotropically to one side of the system. Finally, Bostick and Berkowitz (2003) used vacuum slabs to separate the compartments of a single unit cell. A setup to study ion permeation through GLIC using the double bilayer approach is shown in figure 2.8.

2.5.2 Concentrations

Defining molecular concentrations at the microscopic level is not straight forward. Most simulation studies aim to match experimental conditions. Yet, due to limited computational resources, simulation *boxes* are typically designed as small as possible, focusing on the immediate membrane environment of the channel of interest and minimizing the bulk solvent part. This introduces a bias in the calculation of concentrations.

In many experiments involving membrane proteins, a salt concentration of about 150 mM NaCl is required to warrant an appropriate osmotic pressure. The number of Na^+ and Cl^- ions to add to a simulation system may be calculated using equation 2.13 with $C_{water} = 55$ M. The calculated number of ions is adjusted to neutralize the global charge of the system (*i.e.* protein charges), hence in this case $N_{Na} \neq N_{Cl}$. For the highly charged GLIC channel, there is an important imbalance between positive and negative

ions, in our simulations typically about 90 Cl^- and 50 Na^+ are present. These numbers correspond to a concentration of about 150 mM when calculated on the basis of the Cl^- ions or 80 mM on the basis of the Na^+ ions. If the system was much larger with an extensive bulk solvent part, this difference would progressively become negligible, tending towards $N_{Na} \approx N_{Cl}$.

$$N_{Na} = N_{Cl} = N_{water} \times C_{NaCl}/C_{water} \tag{2.13}$$

Ion concentration could in principle constitute a way to study the influence of pH by explicitly adding protons to the simulation system (Baaden *et al.*, 2001). However, this approach is impractical in the present case because a huge simulation box would be required: at pH 4.6, a box containing 2,189,589 water molecules is necessary to observe a single H^+ ion, *i.e.* 50 times the actual simulation box size we use for GLIC. This echoes the fact that the number of water molecules in simulations is generally too low to mimic dilute solutions. A higher ratio of water to (protein + membrane) would be required to properly account for bulk concentrations.

Furthermore, concentration is a dynamic property. Brannigan *et al.* (2010) reported simulations in which they flooded both GLIC and nAChR receptors with isoflurane, a general anesthetic. During the simulation an important number of the hydrophobic isoflurane molecules partition into the lipid bilayer. Hence, at the end of the simulation, the concentration of isoflurane in the solvent is significantly lower (< 10 mM) than when it started (> 100 mM). This process is depicted on figure 2.9 along with data for alcohol partitioning (Murail *et al.*, 2011).

Brannigan's study implies that a long equilibration of the system may be necessary before concentrations can be measured reliably in order to allow solute molecules to partition between aqueous and membrane phases. It may be debated whether concentrations should be calculated with respect to the water phase only, with respect to water and membrane or with respect to the entire simulation box.

2.5.3 Protonation state

Knowing the protonation state of ionizable residues is a key issue to reliably model a protein. The protonation state depends on a residue's local environment. Standard pK_a values measured in bulk cannot be applied to buried protein residues, in particular for membrane proteins with an environment that largely differs from aqueous solution. GLIC is constituted of five symmetric protomers and the location of its 81 × 5 ionizable residues is shown in figure 2.10. We may consider that equivalent residues in each subunit bear identical protonation states. This assumption leads to approximately $2^{81} = 10^{19}$ possible combinations of protonation states. Tang and coworkers suggest that the protonation state of some titratable groups may be different from one protomer to another leading to up to 10^{98} different combinations (Cheng *et al.*, 2010; Willenbring *et al.*, 2011), a figure exceeding the number of particles in the universe!

The development of methods for calculating pK_a values of titratable groups in proteins was pioneered by Tanford and Kirkwood (1957) who proposed to represent the protein as an impenetrable sphere, which allows to analytically solve the Poisson-Boltzmann Equation (PBE). The increase in computing performances has facilitated the development of many PBE solvers, including the widely used APBS software (Fogolari *et al.*, 2002; Baker *et al.*, 2001). Nielsen and coworkers showed that a Finite Difference Poisson-Boltzmann method yields better results when adding an explicit step to optimize the hydrogen bond

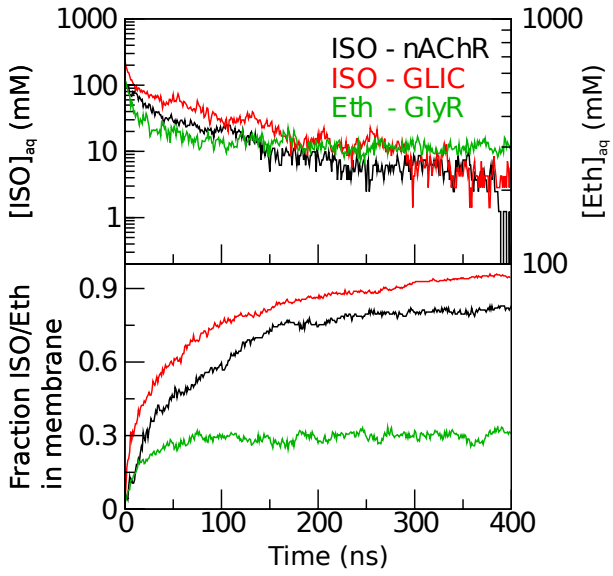

Figure 2.9 – Isoflurane and ethanol partitioning along flooding simulations. Isoflurane (black and red) and ethanol (green) partition into the membrane during the equilibration of a GLIC, a nAChR and a GlyR system, respectively. The aqueous concentration of isoflurane (top) decreases for the benefit of the fraction in the membrane (bottom). The same behavior is observed for the partitioning of ethanol during the equilibration of a GlyR system (green) but to a lesser extent. Due to its more hydrophilic properties, ethanol concentration decreased to half the initial one (≈ 300 mM) at the end of the simulation. This is in contrast to isoflurane: its concentration drops to less than 10% of the starting one (≈ 10 mM).

network (Nielsen *et al.*, 1999). Ideally, protein conformational flexibility should be taken into consideration for calculating pK_a values. Specific terms have been included in some algorithms (Alexov and Gunner, 1997) and, more recently, methods based on the λ-dynamics approach using constant pH MD and Replica Exchange Molecular Dynamics emerged (Donnini *et al.*, 2011; Williams *et al.*, 2011; Meng and Roitberg, 2010). These latter methods are currently still under development and have so far only been tested on small non-membrane peptides or proteins. PROPKA (Li *et al.*, 2005; Bas *et al.*, 2008) may be one of the most commonly used empirical approaches because it is very fast.

In order to setup simulations of the GLIC system, we assessed the results of several widely used programs and web services. These pK_a predictions yielded widely varying pK_a shifts as illustrated in figure 2.11. We settled on the use of the Yasara software (Krieger *et al.*, 2002) mixing Ewald summation and hydrogen bonding network optimization to determine if a titratable group should be protonated or not (Krieger *et al.*, 2006). The Yasara results remain in a reasonable pK_a shift range, whereas some of the other methods suggest huge shifts (figure 2.11). We applied a consensus approach, only protonating residues that were simultaneously found to change ionization state in all five subunits. Subsequent Brownian Dynamics simulations suggested that a neutral H11' residue is most compatible with ion permeation (data courtesy of Prof. Toby Allen; not shown) and the influence of this protonation state has been tested in more recent simulations. Many efforts in improving the crystallization protocol for GLIC recently lead to a higher resolution structure in which ion binding can be predicted between residue D86 and D88. This is a strong indication that these residues should not be protonated (Sauguet *et al.*, 2013b). These findings allowed us to iteratively improve our protonation state estimate for GLIC.

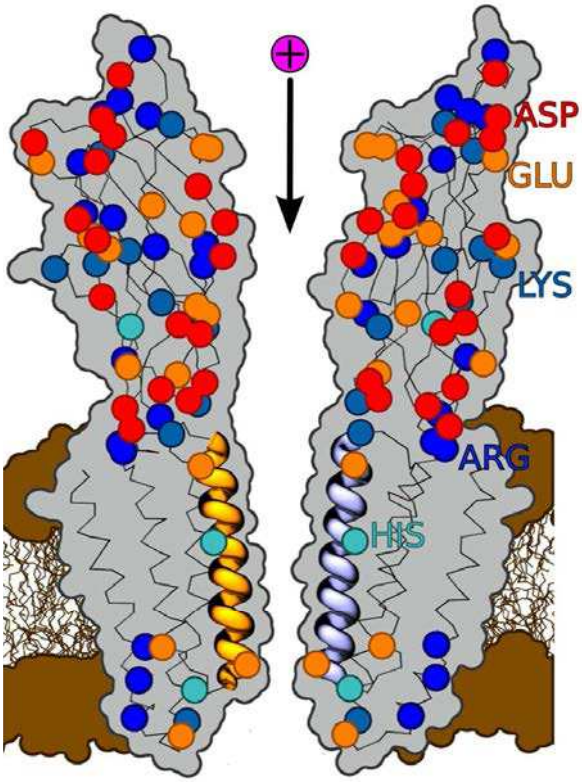

Figure 2.10 – Localization of ionizable residues shown in a cross-section of the GLIC ion channel (grey). M2 helices, in cartoon representation, line the pore through which cations (pink) cross the membrane (ochre).

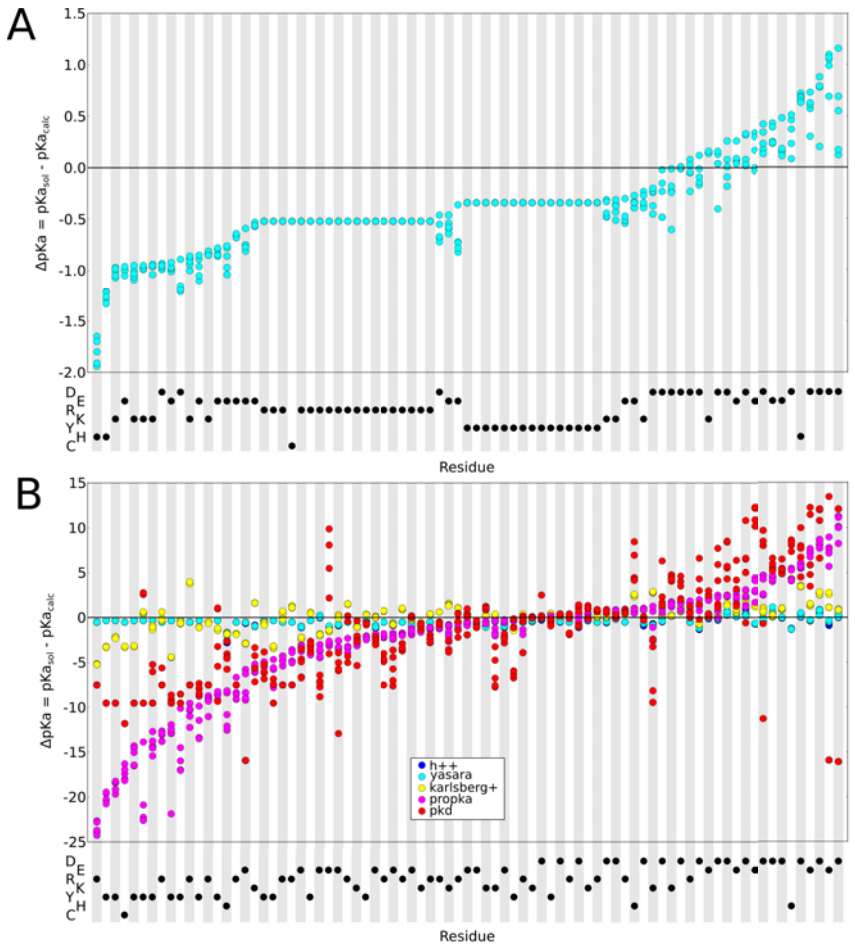

Figure 2.11 – pK_a shift predictions with respect to standard values for all ionizable residues in GLIC obtained using different software packages. Residues are ordered according to $\Delta pK_a = (pK_{a\,\text{solution}} - pK_{a\,\text{calc}})$ with respect to the Yasara software (A) or the PROPKA software (B), respectively. The bottom part of each panel indicates the amino acid type of each residue.

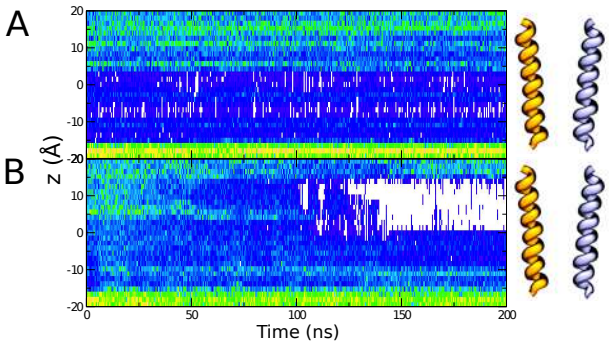

Figure 2.12 – Hydration traces of GLIC's pore during two representative simulations. Minor changes in the simulation parameters can make a noticeable difference between a fully hydrated channel (A) and a channel that dehydrates spontaneously in the upper part of the M2 helix-lined pore (B). For both simulations, the protonation states were identical (Bocquet *et al.*, 2009), the only differences were the forcefield and the MD software used (amber99 and GROMACS for simulation A, *vs.* Charmm22 and NAMD2 for simulation B).

2.5.4 Solvation in special/unusual environments

The complex shapes of proteins may feature channels and cavities providing special, potentially solvated nano-environments. Water in such hydrophobic nanoconfinement may be particularly unstable, a phenomenon known as capillary evaporation. Several groups have observed and characterized dewetting transitions in MD simulations, for example in the context of nanopores (Beckstein and Sansom, 2003; Beckstein *et al.*, 2001) or in the bacterial mechanosensitive channels MscL and MscS (Anishkin *et al.*, 2010; Anishkin and Sukharev, 2004). Roth and coworkers suggest that capillary evaporation could constitute an intrinsic property of some channels (Roth *et al.*, 2008) and may be a widespread biological mechanism. In the case of GLIC, extensive sampling lead us to observe an unexpected pore dewetting behavior (see figure 2.12A), as did several other groups (LeBard *et al.*, 2012; Willenbring *et al.*, 2011). Yet we cannot currently conclude whether GLIC belongs to a family of *bubble gated ion channels*, since ongoing studies in our lab suggest that subtle changes in the simulation parameters may prevent dewetting to occur (see figure 2.12B). Another very recent study is more affirmative (Zhu and Hummer, 2012b). It should be noted that forcefield parameters generally have not been tuned to reproduce the behavior of water in such special environments, which is in part due to the lack of experimental data.

2.5.5 Sampling, statistics, timescale

A fundamental question before starting any computational study is how to best spend the limited amount of available computing time. Strategies may vary in between two extremes: a) running many short simulations from several starting points or b) running an extended one-shot simulation. Shaw et al. recently showed that the result of the second approach matches experimental data very well, when the MD simulations are long enough (Shaw *et al.*, 2010).

In 2010, my PhD host lab studied GLIC gating in a 1 microsecond MD simulation suggesting a *domino* gating mechanism in which subunits sequentially switch from an open to a closed conformation (Nury *et al.*, 2010). Despite the large amount of computational resources (approx. 10 months of calculations on

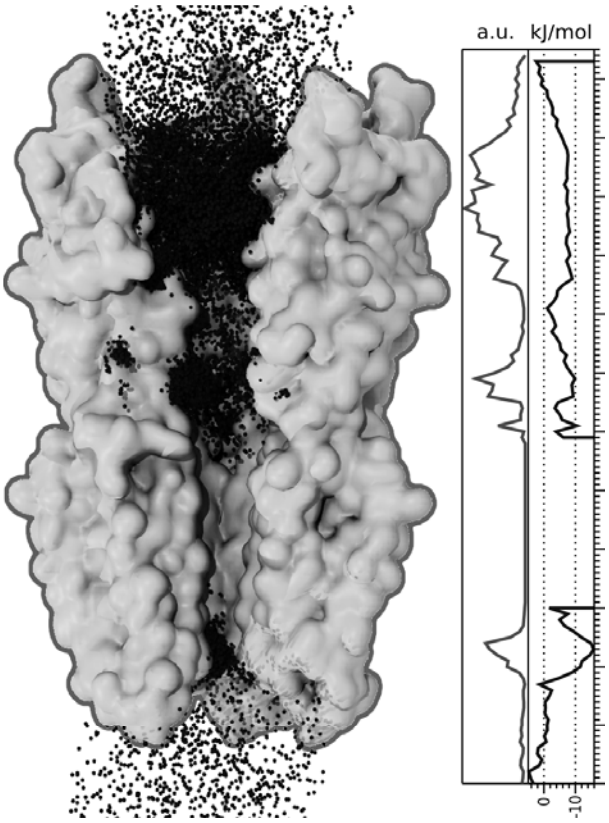

Figure 2.13 – Sodium ion occupancy and related relative Boltzmann energy accumulated during a one microsecond MD simulation. Ions are blocked in the upper part of the transmembrane domain.

a supercomputer in 2009, *i.e.* tens of years on a recent desktop machine), only two protomers had fully undergone this transition to a closed state at the end of the simulation, suggesting that a much longer simulation was required to achieve a complete gating transition in all five protomers. Longer simulations are also needed to characterize processes such as ion permeation. Since GLIC has a low conductivity of 8 ps, one should observe an estimated passage of only 3 ions per microsecond at −65 mV. A cheaper alternative is to map the affinity of ions for a certain position along the channel pore. Such a graph was determined previously from the 1 microsecond gating simulation for a non-conductive state with a central barrier (figure 2.13). It is usually admitted that a simulation should be run at least 10 times longer than the slowest timescale of interest (Zuckerman, 2011). This is often impossible since many relevant biomolecular timescales exceed 1 microsecond. Typically, the neuromuscular acetylcholine receptor's gating is expected to be in the range of 1 to 10 µs (Chakrapani and Auerbach, 2005) which implies MD simulations from 10 to 100 µs.

Running several short comparative simulations may be more appropriate for ligand binding studies, for example involving drugs and anesthetics as illustrated in the results of this PhD thesis. An advantage of such short simulations is to remain close to a well defined state, *e.g.* a crystal structure, rather than moving away from the experimentally backed conformation to some transient intermediate state. Furthermore one may reduce the number of unproductive runs where the drug may diffuse out of the binding pocket into the solvent. Many short simulations with slightly different ligand starting conformations improve statistics and sampling. We employ such an approach to study two general anesthetics, propofol and desflurane, that

have recently been co-crystallized with GLIC (Nury *et al.*, 2011). This study revealed a binding site in the upper part of the transmembrane domain of the protein. Other binding sites for general anesthetics and alcohols, including transmembrane, extracellular and pore sites have been suggested (Cheng *et al.*, 2010; Chen *et al.*, 2010; Brannigan *et al.*, 2010; Howard *et al.*, 2011a). Channel blocking by charged quaternary ammonium compounds, divalent ions and lidocaine has been shown using electrophysiology and X-ray crystallography (Hilf *et al.*, 2010), also suggesting binding sites located in the pore of the channel. These observations pose the problem of sampling from a combinatorial point of view: multiplying the number of sites by the number of ligands, then by the number of mutants one wishes to test quickly leads to an intractable required total simulation time.

Lipids are crucial for the structure and function of membrane proteins. Bilayers with complex compositions pose a particular sampling challenge (Soares and Straatsma, 2008). A misplaced lipid in a simulation setup might have consequences on the whole trajectory, in particular if it were to play a specific biological role. With a diffusion coefficient of the order of 10^{-8} cm²/s, a lipid embedded in a membrane is expected to have a mean-square displacement of 4 nm² for a one microsecond long simulation. In our GLIC simulations, convergence for this value sets on beyond 100 ns and fully stabilizes at about 500 ns. Hence, the timescale of most current studies does not allow for an extensive reorganization of lipids around membrane proteins. Parton et al. recently addressed this problem while simulating a whole vesicle, demonstrating the importance of lipid diffusion for protein aggregation (Parton *et al.*, 2011). The authors however highlight that the coarse grained models are highly simplified and inevitably approximate the nature of the protein-lipid and protein-protein interactions. de Meyer *et al.* (2010) previously suggested the role of cholesterol in protein clustering using dissipative particle dynamics Monte Carlo and a more simplified model.

At last, the problem of simulation convergence should be raised briefly. Methods for the quantification of sampling have been proposed for several decades, yet none has been widely adopted. In 2000, Berk Hess proposed a method based on principal component analysis (Hess, 2000) that has been used by other groups to evaluate the convergence of a set of MD simulations (Faraldo-Gómez *et al.*, 2004; Grossfield *et al.*, 2007). Faraldo-Gómez and coworkers focus on convergence of membrane protein simulations, and although the timescale is relatively short by today's standards, their main findings are likely still valid. Their work concludes that structured transmembrane domains converge relatively fast, even on a 10 ns timescale, but more mobile parts are under-sampled. Grossfield et al. calculated 26 independent 100 ns molecular dynamics runs of rhodopsin and found similar results (Grossfield *et al.*, 2007). To date, despite new method proposals (Grossfield and Zuckerman, 2009; Zhang *et al.*, 2010; Zhu and Hummer, 2012a), sampling quality is often tentatively assessed based on several simple criteria. A single descriptor may be monitored along a simulation until it reaches a stable value. The Root Mean Square Deviation (RMSD), which is a descriptor for molecular deformation, is a common but controversial criterion. A variation consists in stopping a simulation after a descriptor reaches an experimental reference value and remains close to it for a certain time. Another approach is to use several independent MD simulations with different starting points. When these simulations converge to a similar state, sampling is considered sufficient.

2.6 Setups and methods used in this work

In this subsection I first describe the simulation setups that form the basis for this thesis, representing a total of over 1600 nanoseconds sampling accumulated on the anesthetic-bound GLIC system. At the end I highlight a few non-standard methods used to calculate confidence intervals on these data sets, characterize the location of the anesthetic in a given binding site and determine binding pocket volume.

This work reports MD simulations of three different anesthetics bound to GLIC. Molecular models of propofol (PFL) and desflurane (DSF) bound to GLIC were built using a pre-equilibrated system of GLIC embedded in a fully hydrated lipid bilayer. GLIC's initial configuration was based on PDB ID 3EAM, in which protein conformation is virtually identical to the later released high resolution structure 4HFI (RMSD on heavy atoms is 0.4 Å), and then equilibrated without restraints for several tens of nanoseconds. GLIC's conformation at the end of the equilibration phase displayed an RMSD relative to 3EAM of 2.5 Å (calculated and fitted on the protein Cα atoms).

For bromoform, molecular models of open GLIC were built from PDB ID 4HFI (wild-type) and 4HFD (F238A). The system was equilibrated with harmonic constraints on the protein backbone for 200 ns.

Residue protonation state was assigned in the same fashion as in previous simulations (Nury *et al.*, 2011) on the basis of pKa calculations with the Yasara software (Krieger *et al.*, 2012) to represent the most probable pattern at pH 4.6, with residues E26, E35, E67, E69, E75, E82, D86, D88, E177 and E243 being protonated. All histidines were doubly protonated (unless stated otherwise). The models were inserted in a fully hydrated palmitoyl-2-oleoyl-sn-glycerol-phosphatidylcholine (POPC) lipid bilayer. The net charge of the system was neutralized with Na^+ and Cl^- counter ions. The NAMD (Phillips *et al.*, 2005) and GROMACS (Pronk *et al.*, 2013) software suites were used for short and long MD simulations, respectively.

2.6.1 Short 8 ns long MD simulations

Anesthetic initial poses

The ligand was inserted into a previously equilibrated system of GLIC embedded in a fully hydrated lipid bilayer. Bromoform (MBR) poses have been generated by randomly moving and rotating bromoform molecules around the crystallographic binding site. Propofol and desflurane poses have been generated by taking the largest clusters from a 30 ns long MD simulation of the GA bound to GLIC. Previous coordinates were calculated using the GROMACS `g_cluster` program with the gromos algorithm. The cutoff distance for the clustering has been determined empirically to fit the number of starting conformations we needed *i.e.* approximately 125.

GA molecules were assigned different conformations in each of the five GLIC subunits and in each of the 25 systems that were simulated achieving a total of 125 different poses, which maximizes anesthetic sampling in the binding pocket.

Each system was then minimized for 1000 steps and ran for 8 ns using the run parameters described below.

Anesthetic	GLIC variant	Conformation*	Binding site†	Sampling (ns)	Total sampling (ns)
MBR	WT	Open	W1	25 × 8 = 200	**1000**
MBR	WT	LC	P1	10 × 8 = 80	**80**
MBR	F238A	Open	B1	25 × 8 = 200	**1000**
DSF	WT	Open	W1	25 × 8 = 200	**1000**
DSF	WT	LC	W1	25 × 8 = 200	**1000**
DSF	T255A	Open	W1	25 × 8 = 200	**1000**
DSF	T255A	LC	W1	25 × 8 = 200	**1000**
PFL	WT	Open	W1	25 × 8 = 200	**1000**
PFL	WT	LC	W1	25 × 8 = 200	**1000**
PFL	T255A	Open	W1	25 × 8 = 200	**1000**
PFL	T255A	LC	W1	25 × 8 = 200	**1000**

Table 2.3 – Systems simulated by means of short MD simulations. Each GLIC protomer hosts a GA. Considering that each GA molecule is independent from the ones in the neighboring subunits, an MD simulation of 5 GAs bound to GLIC therefore provides 5 times the sampling.
* LC conformation is defined in section 1.2.3. †Site numbering is defined in section 1.3.2.

MD run parameters

MD simulations were performed using the CHARMM27 (MacKerell *et al.*, 1998) force field. Temperature and pressure were maintained using Langevin dynamics (Kubo *et al.*, 1992) and a Langevin Piston (Feller *et al.*, 1995), respectively, at 310 K and 1 bar. Short-range non-bonded interactions were computed using a potential switching from 8.5 to 10 Å. Long-range interactions have been treated using PME (Darden *et al.*, 1993). The same protocol has been used for each system for which short MD simulations were calculated (table 2.3).

2.6.2 Long MD simulations beyond the hundred nanoseconds timescale

The flooding simulation setup was carried out by my colleague Samuel Murail.

For long MD simulations a previously equilibrated system containing GLIC, 246 POPC lipids, 29141 water molecules, 170 Cl^- and 135 Na^+ ions (*i.e.* a total of 146,000 atoms) in an hexagonal box was used to create the system with 200 bromoform molecules. It was equilibrated for 50 ns with position constraints on GLIC C_α atoms with the 4HFI structure as a reference. Then four iterations were used to add slowly the bromoform and avoid aggregates due to its low solubility. In each iteration, 50 molecules of bromoform were added by replacing random water molecules 10 Å away of protein and 4 Å away of the membrane. The system was then minimized for 10,000 steps and equilibrated with position constraints on GLIC C_α atoms with the 4HFI structure as a reference. In the two first iterations, equilibrations were 50 ns long, and 100 ns long in the two following. In a last step bromoform molecules which were bound in the intrasubunit cavity were replaced in the water phase and the system was minimized for 10,000 steps. This equilibrated system was then used as starting point for the three flooding simulations. For simulation of F238A, the phenyl

side chain was removed manually from the system and a minimization of 10,000 steps was run. In each of the three simulations, a supplementary equilibration step was used consisting in a 10 ns equilibration with position constraints on heavy atoms, and 20 ns with position constraints on C_α atoms. Reference structures used were, for WT open, WT LC, and F238A open simulations, respectively, PDB:4HFI, the structure presented in chapter 4 and PDB:4HFD. Production runs were finally carried out for 1 µs without any constraints. Simulations were performed using GROMACS 4.6.3 using virtual interaction sites, 5 fs time steps, and all bond lengths constrained with the LINCS algorithm (Hess *et al.*, 1997). Electrostatics interactions were computed using particle mesh ewald summation at every step. A 10 Å cutoff was used for non-bonded interactions and the neighbor list was updated every 5 steps. Three baths (protein, water and ion, membrane) were coupled to a temperature of 310 K using the Bussi velocity rescaling thermostat with a time constant of τ = 0.1 ps. The x/y dimensions were scaled isotropically with a Berendsen weak barostat and the z dimension independently to reference pressures of 1 bar, τ = 5 ps and compressibility of 4.5×10^{-5} bar^{-1}. During equilibration position restraints of 1000 kJ/(mol nm) were used.

2.6.3 Free energy calculations

Thirdly, we calculated the bromoform affinity for each of the 6 binding sites in WT, F238A, open and LC variants of GLIC using alchemical free energy calculations (figure 2.14). The bromoform poses displayed in the crystal structure were used when available. The bromoform pose in intersubunit site B2 was extracted from a short MD simulation. Bromoform was inserted into a previously equilibrated system of GLIC embedded in a fully hydrated lipid bilayer. MD simulations were performed using the CHARMM36 (Huang and MacKerell, 2013) force field. The system was minimized for 10,000 steps. Two successive equilibrations with constraints on a reference structure, typically a crystal structure, were performed: 5 ns constraining protein heavy atoms and bromoform, then 20 ns constraining protein Cα atoms and bromoform. During these two equilibration steps, constraints were also applied on the dihedral angle between Y197 C-C_α-C_β-C_γ atoms ensuring an angle of 173.5° in GLIC open form and 91.8° in the locally closed conformation. These values correspond to the two principal modes of the angle distribution observed along short MD simulations. A thermodynamic cycle was then applied to calculate free energies of binding of bromoform to GLIC using a similar protocol as that described in (Brömstrup *et al.*, 2013), however as a large number of such calculations was carried out, we optimized the protocol in terms of number of windows and sampling times. Coulombic and van der Waals interactions were decoupled using a decoupling parameter λ linearly increasing from 0 to 1. Coulombic interactions were decoupled along 11 independent steps while 21 steps were necessary to decouple van der Waals interactions. At each λ-point, the system was minimized for 5000 steps, equilibrated for 10 ps in the NVT ensemble, then equilibrated for 100 ps in the NPT ensemble. Bromoform positions were harmonically constrained during these two equilibration steps with a force constant of 1000 kJ/(mol nm^2). Production simulations were run using the `sd`[3] integrator with a time step of 2 fs. During the production phase, bromoform positions were restrained using an umbrella potential with a force constant of 100 kJ/(mol nm^2). For coulombic interaction decoupling, 2 ns were carried out. For van der Waals interactions, 3 ns were carried out for the first 14 λ-points (initial $\lambda < 0.7$) and 10 ns for the remaining 7 points (initial $\lambda >= 0.7$). The same protocol was applied for decoupling bromoform in water. The calculation of the binding free energy was carried out

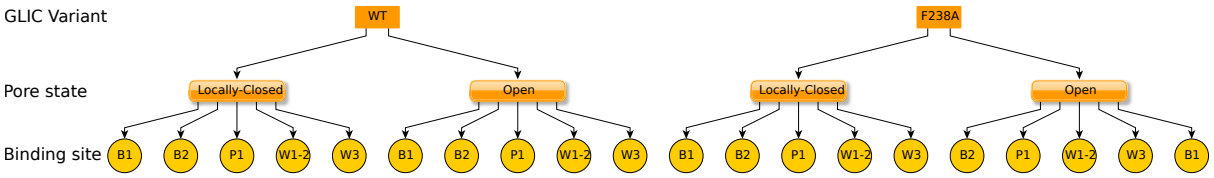

Figure 2.14 – Extensive screening of bromform's affinity for GLIC. Bromoform's free energy of binding was calculated for 5 binding sites in GLIC wild-type, mutant F238A, in both open and locally closed conformation.

using the BAR method (Bennett, 1976) as implemented in the `g_bar` program from the GROMACS suite.

2.6.4 Confidence interval on means calculation

Comparing two means extracted from MD simulations requires a robust methodology that is not well established in the field. A classical method such as the Shapiro-Wilk test is often not applicable in MD because this test, as well as most parametric tests, require the data to be normally distributed and have equal variances, which is often not the case in MD. Non parametric tests, such as the Kolmogorov-Smirnov test, are therefore a better choice but, as well as the parametric tests by the way, are biased by the number of observations: they will return significant p-values if the number of observations is important, even if the difference between the distributions is minimal. In MD, it is very common to calculate a mean on hundreds or thousands of steps of a simulation. Another approach has therefore to be used to calculate robust means with confidence intervals.

I chose to use the bootstrapping method. This method consists in calculating an estimator, typically the mean, of a distribution using a random resampling of the distribution with replacement. Numerous resamplings have to be done, to finally obtain as many estimators extracted from the slightly different subsamples extracted from the original data set. This method has several advantages, especially to calculate means with confidence intervals. First, it can be used on non normally distributed data since the ensemble of means that is calculated will most likely be normally distributed itself. Second, this method is sensitive to variance in the initial distribution which means that two sets of data centered on the same value but with different variances will yield different confidence intervals.

In my case, I chose to use 1000 resamplings.

2.6.5 Binding site occupancies

Binding sites occupancies have been computed by calculating the distance between the anesthetic and a reference position taken from relevant crystal structures. The occupancy of site B2, which is not a crystallographic site, has been calculated with respect to a position extracted from a short MD reference simulation. A site is defined *as occupied* at a time t if the distance between the center of mass of the anesthetic molecule and the reference position are within a cutoff. The cutoff value I chose is 4 Å, which is quite restrictive considering the volume of the intrasubunit pocket. It is therefore important to note that the analysis may indicate that the anesthetic does not occupy any binding site strictly speaking while

[3] The `sd` integrator implemented in GROMACS is an accurate leap-frog stochastic integrator which also acts as a thermostat.

being inside the pocket.

Occupancy maps have been calculated with Visual Molecular Dynamics (VMD)'s volmap tool, using a classical van der Waals radius, combining all frames and averaging the data.

2.6.6 Pocket volume calculation

Binding pocket volume calculations have been carried out using the Epock software (see section 3.3).

To calculate the volume of the intrasubunit pocket, only frames with anesthetic molecules closer than 4 Å from sites W1, W2 or W3 have been taken into account.

The final volume average value and corresponding confidence interval have been calculated by bootstrapping (see section 2.6.4.) considering the last 3 nanoseconds of simulation.

2.6.7 Contacts

Contacts between the anesthetic and the protein residues have been calculated with the VMD `measure contacts` procedure. A contact with a residue is counted if any atom from the anesthetic is closer than 4 Å of any atom of the residue. The number of contacts with a residue at a time t is summed over the five subunits of the protein. The final percentage of contacts between the anesthetic and a residue is defined as the sum of the number of contacts at each frame divided by the number of frames in the simulation. This percentage is therefore the probability that any of the five anesthetic molecules present in a simulation contacts the correspondind residue on any of the five GLIC chains.

High-Performance Computing And Large Scale Data Analysis

3

The founding principle of statistical physics concerns ergodicity stating that the time average of one sequence of events is the same as the ensemble average. Hence, as any statistical analysis, data obtained from MD simulations can only be trusted if numerous uncorrelated events have been observed. As discussed in section 2.5.5, there are basically two ways one can apply this principle to MD simulations: a) running an extended one-shot simulation; b) running many short simulations from several starting points. Depending on the study focus, the first, second or both methodologies may apply. For example, a study that aims to describe the process of binding of a ligand to a protein would most probably require long simulations in which the ligand is not bound to the protein at the start as opposed to the description of the interactions between a protein and a bound ligand, which would require many short simulations to avoid the ligand unbinding. Nowadays, it is very common to run one to two long simulations since, as will be justified in this chapter, it often implies less work.

A major aspect of this work has been the description of the dynamics of GAs bound to GLIC. I chose the second approach, *i.e.* the calculation of several short simulations to achieve extensive sampling of the ligand dynamics while bound to the protein.

In this part, my goal is to introduce the main technical difficulties I have been facing. The first section will be devoted to the specific hurdles related to the approach I used, while the following is related to the system's size.

3.1 Computing the simulations

3.1.1 The need for high-performance computers

As GLIC is a membrane protein, a minimum system for studying this channel at an all-atom resolution is made of several molecule types, leading to a total number of approximately 200,000 atoms (see table 3.2). The simulation of such a number of particles remains challenging and requires computational power that is often not accessible locally. Supercomputers are therefore required to produce data in a reasonable time. As an example, using 184 cores on `jade@cines.fr`[1] allows to run an MD simulation at a speed of 10 ns/24h. While computing 10 ns of simulation takes one day on `jade`, it would virtually take 46 days to

get the same amount of data on a recent desktop computer.

I focused on 12 subsystems including different GLIC mutants in different conformations with different GAs bound to it. Since 25 simulations of 8 ns each have been run for each system, I calculated the equivalent of 9600 ns, which would have been done in a total of 960 days if they had been run one after the other on a super computer, and more than 120 years on a recent desktop computer, pointing out the essential need for supercomputers in theoretical biophysics. The total amount of CPU-time[2] spent on this aspect of the project is estimated to 4,377,528 jade equivalent CPU-hours[3].

3.1.2 Optimizing the available resources

The number of CPU-hours available for a project is finite, obviously. To best spend the limited amount of available computing time, it is necessary to run several benchmarks to avoid loosing time by running a suboptimized simulation. The procedure to find the set of parameters that gives the best computation speed is well defined and should be carried out every time one starts to work on new machines or a new system since this set of parameters depends both on the topology of the system and the architecture of the machines.

The very first step to optimize a simulation is to reduce to the minimum the number of particles that compose a system. The size of the simulation box (see section 2.2.3) has therefore to be well chosen not being too large, which will increase the number of lipids (in the case of a membrane protein) and solvent molecules, but also not being too small to avoid contacts between protein periodic images or artificial structuring of the membrane.

The second step is to run several very short simulations (on the order of 100 ps) varying the number of cores used that will allow to estimate the speed a simulation is computed at (*i.e.* number of nanoseconds calculated per day), and the computing time consumed. The optimal number of processors and cores (recent processors have up to 16 cores, and, sometimes, leaving one core available yields better results) can then be defined according to the project needs. A number of processors which is a power of two usually yields better performance. However, NAMD developers advise for a maximum speed to use a number of cores proportional to the system's size. For instance, it can be deduced from figure 3.1 that the maximum speed (number of nanoseconds computed per day) is reached at 544 CPUs. It is then a waste of resources, again for this particular system on this particular machine, to run a simulation using 1024 CPUs, as no speed-up is achieved. The second element that will influence the choice of the number of cores to use is the number of nanoseconds that can be computed with a certain amount of CPU-hours, 300,000 for instance. This amount is inversely proportional to the number of cores used. It is clear on figure 3.1, that, at maximum speed (*i.e.* 512 CPUs), only 500 ns can be computed while more than 700 ns can be computed with 128 CPUs. On the other hand, the 500 ns would be obtained in X days only with 512 CPUs, whereas Y days are required with the more economical 128 CPUs for 700 ns.

Last but not least, the program options should be in their turn optimized. In NAMD, for instance,

[1] `jade` is a supercomputer located at Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier).

[2] See definition on page 27.

[3] The simulations were dispatched on `jade@cines.fr`, `turing@idris.fr`, `babel@idris.fr`, `ada@idris.fr`, `vargas@idris.fr` and `curie@tgcc.fr`. This value is the number of CPU-hours that would have been consumed if all the simulations had been carried out on `jade@cines.fr`.

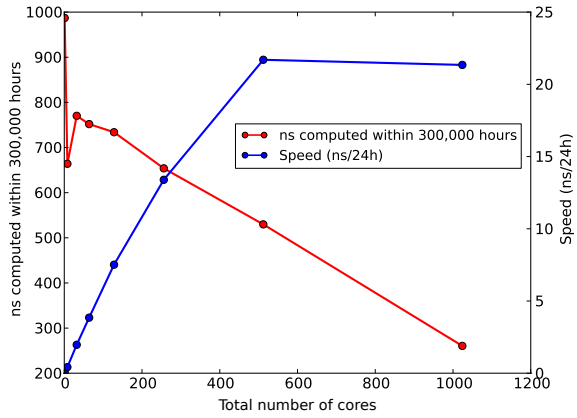

Figure 3.1 – Benchmarking the speed of a simulation on a machine. The total number of cores is the product of the number of CPUs and the number of cores on each CPU. Benchmark realized with NAMD in 2011 on `jade@cines.fr`.

dividing patches[4] in halves often leads to better results. The PME calculation (electrostatic interactions) can be optimized by defining the dimension of the grid for PME decomposition and defining the number of processors on which PME should be computed and reserve them for PME only. Several other options can be adjusted and will not be developed here.

Table 3.1 compares some French supercomputer performances. Although it may seem illogical, the fastest machine is not necessarily the one to use for all purposes. For instance, `curie@tgcc.fr`'s cores are slightly less efficient to run an MD than `ada@idris.fr`'s but, on the other hand, `curie`'s overall speed is higher than `ada`'s since it has more cores: 80,640 cores *vs* 10,624 on `ada`. This reduces the waiting time between jobs and may allow to run more jobs in parallel, depending on a given supercomputer's policy. Furthermore, more cores could be used on `curie` than on `ada` speeding-up the simulation (but increasing the cost of a simulation in CPU-hours/ns). As as second example, `turing@idris.fr` is one order of magnitude slower than every other supercomputer listed in table 3.1. However, this machine turned out to be suitable to run short MD simulations (8 ns). The specific rules of the computer centers have to be taken into account. At IDRIS, the number of jobs a user can run at the same time is limited to 3, and the number of jobs the same user can have in queue is also restricted. This rule can strongly slow down a project's proceedings when numerous jobs have to be run. To compute 20 different free energies of binding of bromoform to GLIC (see section 2.6.3), I had to run a total of 640 jobs. For the reason stated above, most of these jobs (in the limit of the available CPU time) were run on `curie` to benefit of the unrestricted number of jobs a single user can run at the same time on this machine. The remaining jobs were run with a special priority on the `ada` machine, thanks to the assistance of the IDRIS support team.

3.1.3 Data storage

The storage of these data has to be considered as a major concern, since more than 6 TB have been produced carrying out this part of the project. Besides, safety demands storing at least two copies of the data in

[4] The patch is used by NAMD as the fundamental unit of spatial decomposition.

Machine	Number of cores	Speed (ns/day)	ns/100,000 CPU hours	days/µs
turing@idris.fr	256	3.54	58	283
ada@idris.fr	256	57.66	938	17
curie@tgcc.fr	256	53.03	863	19
jade@cines.fr	256	43.61	710	23
hades@lbt.ibpc.fr	180	35.88	865	28

Table 3.1 – Comparing common supercomputers. Benchmark realized with GROMACS 4.6.3 on a system made of 146,182 particles.

different places, allowing to recover data if one copy is damaged or deleted.

The original data have been stored on gaya@idris.fr, that offers a capacity of 6.6 PB of taped storage. Copies had to be stored locally. However, the lab's current storage setup imposes limits on the recurrent backup of such large quantities of data. The solution I found is to switch from the original NAMD dcd format to the GROMACS xtc format, which has been optimized for floating point numbers compression. Hence, an xtc file is more than 3 times smaller than a dcd file and displays a negligible precision drift. It therefore turns out a very good choice for original data backup and everyday analyses. Finally, xtc files generated for this part of the project constitute a total of 1.9 TB of data.

From a more general point of view, I think that GROMACS's xtc format should always be used to store system trajectories for numerous reasons. The first reason is the impressive gain of space detailed above. The second reason is that GROMACS can perform a variety of operations on xtc files such as filtering, fitting, translating, *etc.* The third reason is that dealing with smaller files will significantly impact the time a program will require to run an analysis, even its capacity to run an analysis. As an example, the VMD software loads into memory the whole trajectory at start. Obviously, the larger the file, the longer the loading, which can become critical if the file size exceeds 1 GB.

Finally, backup software run time considerably depends on the number of files to save. Some MD software, such as OPEP (Chebaro *et al.*, 2012), store the system trajectory as Protein Data Bank (PDB) formatted files at a rate of one file per frame. One trajectory could therefore be stored as several thousands (millions) of files which will take a considerable amount of time to backup. This may even be more critical at supercomputer centers, as the number of available inodes on the file system may be limited.

Dealing with such a volume of data divided into 400 independent simulations is not trivial. Improving the efficiency and the scaling of the analysis processes turned out to be unavoidable.

3.2 Scaling and parallelization of the analysis processes

Handling several similar simulations at the same time can reveal itself time consuming and many mistakes can slip into the process if some slight changes have to be made from one simulation to the other. Since most recent studies favor the calculation of one to a handful (< 5) of long simulations upon the calculation of many short ones, no software has been developed to perform this kind of specific task.

A good knowledge of Unix tools combined with programming skills allowed me to handle 400

Molecule type	Number of molecules	Number of atoms
Protein	1555 residues (5 chains of 311 residues)	25,385
Lipids	301	41,138
Water	43,882	131,646
Ions	143	143
Total	–	198,312

Table 3.2 – Composition of a minimal GLIC system. *minimal* is to be understood as *that contains the minimum number of species to calculate an all-atom simulation.* We estimate that the total number of atoms can be further reduced by ≈ 25% by agressively optimizing the simulation box shape.

independent simulations very efficiently considering both time concerns and risk minimization. As an example, a script aiming to get each simulation ready to run on a supercomputer is shown in appendix C.1. The script creates one directory per simulation with all the materials required to run the MD simulation on a cluster. It implements the possibility of choosing on which cluster the simulations have to be ran and adapts the submission scripts in consequence. Such kind of scripts are not major progresses in the field but have to be implemented when running tens (hundreds in my case) of similar simulations that may vary by a handful parameters.

The same kind of approach had to be applied to the simulation analysis. I chose to write one specific `Makefile` for each analysis which allowed me to take advantage of the multi-threaded nature of the `make` program and to run up to 12 analyses at the same time, which became critical when analyses have to be run on hundreds of simulations. Furthermore, by writing scripts as flexible as possible, running a new analysis turned out a matter of minutes, even on hundreds of simulations.

3.3 Efficient Analysis Software Need: The Epock Software

This part is adapted from Laurent *et al.* (submitted) and specifically adresses the development of a pocket volume analysis tool in the context of handling more and more massive amounts of MD data.

Owing to recent advances in hardware and software, MD simulations enable the study of the evolution of biomolecular systems of increasing size and complexity over time. Repeatedly D.E. Shaw showed the possibility of breaking the millisecond barrier using the Anton supercomputer and the Desmond computer program (Lindorff-Larsen *et al.*, 2011). The drawback of this progress is the generation of increasingly large MD datasets (see section 3.1), with consequences for subsequent analysis. It is therefore crucial to develop improved software tools able to analyze these datasets in a reasonable time.

The volume of an internal protein pocket is of fundamental importance to ligand accessibility and mobility inside the pocket. Along years, several programs and algorithms that aim to quantify the volume of a protein cavity have been developed and, among them, only few are designed to efficiently manage dynamic data from MD. Limited performance often prohibits their use on large datasets. To tackle this issue, I developed Epock, a program that allows efficient measurement of the evolution of protein pocket volume during MD trajectories.

3.3.1 Program features

Epock is a command-line program that requires input in the form of a system topology and an MD trajectory in GROMACS xtc format (Pronk *et al.*, 2013), which can be either atomistic or coarse-grained. An Epock configuration file specifies a set of parameters for each cavity to be characterized, including a maximum encompassing region for the cavity (MER). The MER provides explicitly defined bounds for each cavity by combining simple 3D objects (spheres, cylinders and cuboids) to create complex shapes (a concept known as "solid constructive geometry"). This allows to unequivocally follow a priori determined cavities over time, whereas Epock is not intended for cavity searches. My implementation extends the method proposed by Durrant and coworkers in the POVME program (Durrant *et al.*, 2011). The Epock Tcl/Tk plugin for VMD (Humphrey *et al.*, 1996) provides an intuitive way to choose and position shapes to define the MER (see figure 3.3A-B) [5].

For each pocket, Epock calculates the space accessible to a probe, called "free space", which is the set of all grid points with a distance to protein exceeding the user-defined probe radius (typically, 1.4 Å). The number of grid points that overlap each residue is stored and can be outputted as "residue contribution". The volume of the so-called free space is then calculated at higher precision using a finer grid.

Epock outputs pore profiles by calculating the radius of the largest disc that can fit the previously detected free space along the Z axis. The results are similar to those obtained with the Hole software (Smart *et al.* 1993. see figure 3.2) [6].

Epock produces several output files, including the computed trajectory of free space over time, a feature inspired by the `trj_cavity` software (Paramo *et al.*, 2014). This trajectory is directly readable in VMD, which makes the relationship between pocket volume and protein conformation highly intuitive. Epock results for pocket volume, residue contribution and pore profile can be plotted directly in VMD using the plugin, or by running the Python scripts that are freely distributed with Epock.

3.3.2 Application: the GLIC ion channel

The *Gloeobacter violaceus* Ion Channel (GLIC) previously introduced in this PhD manuscript features numerous pockets, including a binding site for general anesthetics (Nury *et al.*, 2011). It is a challenging test case because of its size, 1555 residues, and the presence of multiple pockets that often connect to each other and/or to the central pore. The volume of a single pocket was computed over an 800-frame trajectory of the protein (25385 atoms, 75 MB) on Mac OS 10.6.8 with 2 × 2.93 GHz Quad-Core Intel Xeon processors and 8 GB 1066 MHz DDR3 memory.

Examples of Epock output are shown in figure 3.3. The chosen analysis example shows that the cavity volume dramatically decreases after *c.a.* 1500 ps, (see red curve in figure 3.3C). Epock's residue contribution analysis shows a particularly high variability for residue Y197 (see cyan curve in figure 3.3C, and figure 3.3D). Simultaneous visualization of the protein trajectory alongside the pocket free space in VMD (figure 3.3E-F) confirms that movement of the Y197 side chain is largely responsible for the volume decrease.

[5] The VMD plugin has been developed by Matthieu Chavent and Caroline Dahl from the Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry, University of Oxford, UK

[6] The pore profile feature has been developped by Tristan Cragnolini, Laboratoire de Biochimie Théorique, CNRS UPR 9080, Univ. Paris Diderot, France

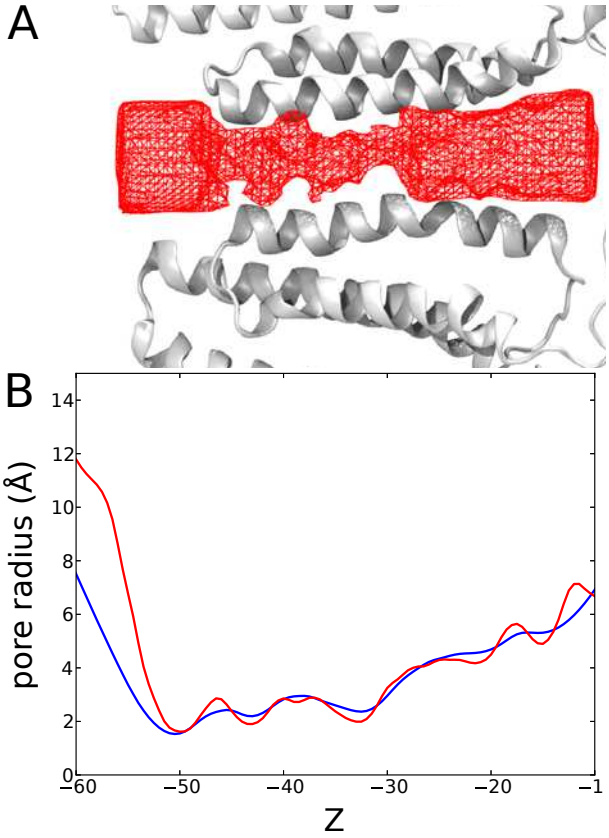

Figure 3.2 – Calculation of a pore profile with Epock. A) GLIC transmembrane domain (protein backbone is represented as white cartoon, Epock's pore surface as red wireframe). The pore profile has been calculated given a 14 Å-radius cylinder as include region and a superimposed 7 Å-radius cylinder as contiguous seed (see Epock's manual for more information). The surface has been calculated from Epock's output using the VMD Volmap tool. B) Comparison of the average profile of the GLIC pore of a 800-frame trajectory. Epock (red) and Hole (blue) results are very similar.

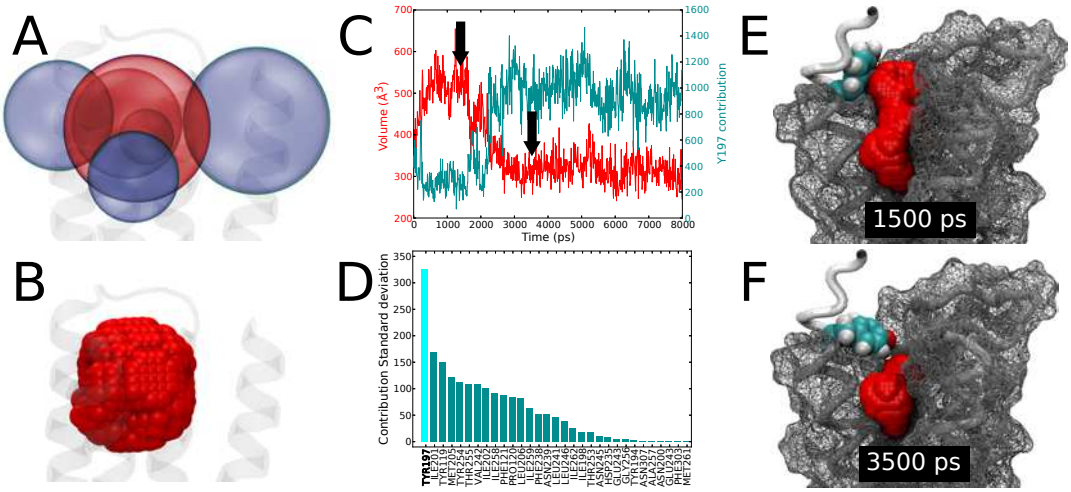

Figure 3.3 – From Epock setup to result analysis. A) Graphical interface for defining the MER using the VMD Epock plugin. The MER consists of a combination of volumes to include (spheres in red) and exclude (spheres in purple), giving rise to a custom complex geometric shape for analysis. B) Grid points that compose the MER. C) Pocket volume (red) and residue contribution of Y197 (cyan) during an MD simulation. D) Standard deviation of residue contribution ordered from highest to lowest. E-F) Protein conformation and pocket (protein surface in grey mesh, backbone as white tube, Y197 as spheres colored by atom type, pocket accessible space as red spheres) at t = 1500 ps (E) and t = 3500 ps (F).

We compared Epock execution speed to two existing programs: i) mdpocket (Schmidtke *et al.*, 2011) that uses Voronoi diagrams and has been specifically designed to calculate pocket volume for MD simulations and ii) POVME (Durrant *et al.*, 2011) which implements an algorithm similar to Epock for free space detection but with differences in the free space volume calculation. The same input grid can be given for all three programs, allowing for meaningful performance comparisons. Epock ran in 5 seconds. This is a dramatically higher speed than both mdpocket and POVME, which feature computing times on the hour timescale (5 and 3 hours, respectively). We hypothesize that POVME's execution time is largely related to its implementation in Python which is known for being slower than the corresponding C++ executable. The reason why Epock is faster than mdpocket may be due to the numerous additional analyses that mdpocket performs during a run, and that can not be disabled.

3.3.3 Making Epock public

The source code distribution

A distributed software requires protection against abusive use such as copy and distribution for commercial use. I strongly believe in open-source projects, especially for science, as well as Epock co-developers. We decided to make Epock's source code accessible to anyone so that developers could enhance the program over years or build a new program inspired by it. As it is crucial for us to assure the accessibility of the source code of any program inspired by Epock, Epock source code is distributed under the CeCILL license, a modified version (and still compatible with it) of the GNU General Public License (GPL).

To encourage developers to contribute to Epock, its source code is versioned with mercurial, a distributed source control management system. This technology allows a developer to access all past code

modifications and keep track of new code modifications while developing a new feature. If a developer wants to contribute to Epock, his changes can be *pulled* into Epock's mother repository so that the whole history of the new feature development is then accessible from it.

Epock' source code is hosted by bitbucket.org and is available at

http://bitbucket.org/epock/epock.

The software distribution

Nowadays, besides the publication of an article or application note in a scientific journal, it is crucial for a software to be visible on the Internet, so that, in the case of Epock for example, anyone looking for a *program for pocket volume measurement in molecular dynamics* can reach Epock. Part of the time I spent on the Epock project was therefore naturally devoted to the creation of a website explaining the method underlying Epock, its usage and a series of application examples.

As writing HTML code can be time consuming, I have been looking for a solution allowing me to write text files with a simplified markup language and translate them to HTML. The best solution I found is the Sphinx tool that was originally created for the new Python documentation, and, from a general point of view, is particularly adapted to the code documentation. In the case of Epock, I did not want to document the code itself but only to build a showcase allowing to download the package, access the online manual and read more about Epock.

Sphinx inputs are, besides a configuration script, text files in reST format, a markup language that allows high efficiency during the writing process since reST files are much simpler to write than HTML code. It can be guessed from figure 3.4 (that shows the reST input file and the corresponding webpage) that reST is a very powerful language that can produce very rich content: this thesis manuscript could perfectly have been written in reST and rendered into a PDF document thanks to Sphinx! A few more hours were also required to customize the page layout and, more importantly, write the CSS files for a stylized and original website.

Epock's website is hosted by bitbucket.org and is available at

http://epock.bitbucket.org.

The methodological aspect of this project has now been covered. In the next two chapters, I will focus on the results I obtained on the study of general anesthetic action at the atomic scale. The accurate characterization of binding pocket volumes did play an important role in these investigations.

3.4 BioSpring: an Augmented Spring Network Simulation Engine

As has already been discussed, MD simulations of GLIC require consequent computational power and a simulation may run for weeks, if not months. Here I will describe BioSpring, a computational tool for much faster - but also more approximate - simulations of macromolecular systems. BioSpring is not an alternative to MD simulations, but a useful complementary tool to characterize a molecular system.

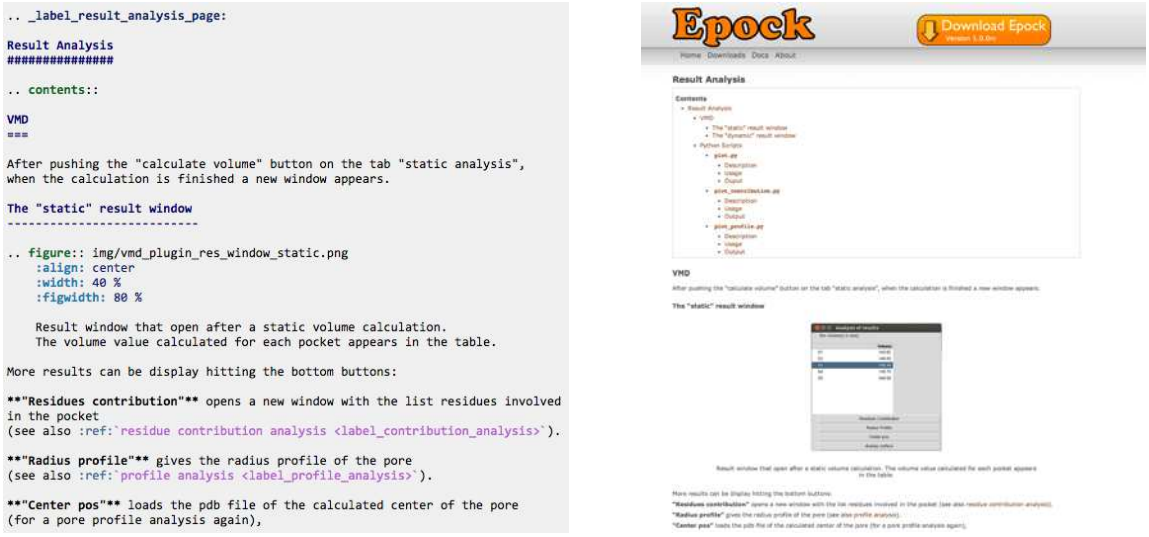

Figure 3.4 – Writing Epock's website. The reST source code (left) that, after processing by Sphinx, will produce the HTML that can be rendered in any web browser (right).

3.4.1 Principle

BioSpring is a calculation engine allowing to run molecular simulations of spring network models. Spring networks are a simplified representation in which the system structure is maintained by interactions mimicking springs: harmonic potentials are created between neighbor particles so that the more the distance d_{ij} between two particles i and j is distant from their equilibrium distance d°_{ij}, the greater is the associated energy. The equilibrium distance d°_{ij} is defined as the distance between particles i and j at $t_{\circ}$. By definition, this force tends to drive the system back to its equilibrium state: the initial configuration of the system.

Biospring has two major originalities. First, it takes into account non-bonded interactions between particles which can allow the system to reach metastable states different from its initial configuration. Second, Biospring implements an interface to Interactive Molecular Dynamics (IMD), a technique in which a user can input forces to the system manually as the simulation is going on thanks to a dedicated input device. This input device can be a simple mouse or a haptic device that allows to send back forces to the user. IMD therefore allows to very intuitively dock a ligand on a receptor or fold a protein. Another application has been demonstrated recently by Molza *et al.* (2014) in a study where the authors use a map extracted from Small-Angle X-ray Scattering (SAXS) experiments to run a targeted folding thanks to BioSpring.

3.4.2 My contribution

BioSpring development was initiated by Nicolas Férey in 2008. He implemented BioSpring's core plus several associated tools.

My initial thesis project included several BioSpring-IMD experiments on GLIC including docking of anesthetics and studies of the channel gating. The current BioSpring force-fields turned out unsuitable to reproduce hydrophobic interactions with sufficient accuracy, so anesthetic docking tries were unfruitful

as were tests on GLIC's gating. In the latter case because the spring network revealed itself too rigid to properly reproduce the motions of the M2 helices. To make these tests, I nevertheless regularly improved BioSpring in different ways I will develop in this part.

Input/Output

BioSpring input files are

- the simulation configuration file (containing time step, number of steps…),
- the system topology.

The system topology file contains the parameters for all the particles of a system including coordinates, radius, charge, *etc.* Reading and writing tabulated files can raise problems, especially for format specification. BioSpring lead developers chose to use the NetCDF format for topology files in which data are stored as arrays (figure 3.5). I improved NetCDF reader and writer classes already existing in BioSpring by performing numerous sanity checks to make sure the file format behaves as expected, a crucial step to avoid unauthorized memory access, which can turn out very difficult to debug. I also implemented methods to automatically write NetCDF binary files and implemented BioSpring support for the newest versions of the NetCDF library.

Associated tools

Initially, BioSpring conversion tools from PDB format to NetCDF format offered limited flexibility. In the context of scientific experiments using BioSpring, it is very common that the spring stiffness between two particles or two groups of particles has to be adjusted, some springs have to be removed, others have to be added, *etc.* Despite the difficulty of the task, these operations had to be done manually, which was time consuming and a potential source of errors. As an example, to add a single spring between two particles, the user had to

1. find the id number of the two particles (this part was usually done using VMD, making sure the `serial` parameter of VMD corresponds to the actual particle id in BioSpring),
2. add 1 to the dimension `spring_number`,
3. add the two particle ids to the `springs` array
4. add the appropriate stiffness for this spring to the `springsstiffness` array,
5. add the appropriate spring equilibrium distance to the `springsequilibrium` array (note that the spring stiffness and equilibrium parameters have to be inserted at the exact same position as the spring is inserted in the `springs` array).

I developed three tools named `pdb2spn`, `editspn` and `mergespn` to deal with most use cases and significantly reduce both the time spent on the topology tuning and the probability of making mistakes during the process.

```
netcdf model {
    dimensions:
        spatialdim = 3 ;
        particle_number = 2 ;
        particlename_length = 4 ;
        chainname_length = 4 ;
        resname_length = 4 ;
        springdim = 2 ;
        spring_number = 1 ;

    variables:
        float coordinates(particle_number, spatialdim) ;
            coordinates:units = "angstrom" ;
            coordinates:long_name = "Particle coordinates" ;

        float charges(particle_number) ;
            charges:long_name = "Particle charge id" ;
            charges:units = "electron" ;

        float radii(particle_number) ;
            radii:units = "A" ;
            radii:long_name = "Particle radius" ;

        float epsilon(particle_number) ;
            epsilon:units = "kJ.mol 1" ;
            epsilon:long_name = "Particle epsilon for Lennard Jones" ;

        float mass(particle_number) ;
            mass:units = "Da" ;
            mass:long_name = "Particle mass" ;

        ...

        int springs(spring_number, springdim) ;
            springs:long_name = "Spring between particle referenced by 2 particle id
s" ;

        float springsstiffness(spring_number) ;
            springsstiffness:long_name = "Spring stiffness" ;

        float springsequilibrium(spring_number) ;
            springsequilibrium:long_name = "Spring distance equilibrium" ;

    data:
        coordinates = 0, 0, 0,
                      2, 0, 0 ;
        particleids = 0, 1 ;
        particlenames = "N", "N" ;
        charges =  0.4157,  0.4157 ;
        radii = 1.824, 1.824 ;
        epsilon = 0.17, 0.17 ;
        mass = 14.01, 14.01 ;
        surfaceaccessibility = 0, 0 ;
        hydrophobicityscale = 0.112, 0.112 ;
        resnames = "VAL", "VAL" ;
        resids = 1, 2 ;
        chainnames = "A", "A" ;
        dynamicstate = 0, 1 ;
        springs = 0, 1 ;
        springsstiffness = 1 ;
        springsequilibrium = 2 ;
}
```

Figure 3.5 – The NetCDF array-oriented format. Example of a NetCDF text file for a system containing two particles. BioSpring input is the binary version of this file that can be created thanks to the `ncgen` program distributed along with the NetCDF library.

`pdb2spn` is a utility that converts a PDB formatted file to a binary NetCDF formatted file. Prior to the development of this tool, text files were generated. They had to be converted to a binary in a second time, thanks to a tool distributed with the NetCDF library. The creation of a binary file is the only new feature of `pdb2spn` but it is incorporated within a framework in which BioSpring topology files should not be edited by hand.

`editspn` allows to edit BioSpring binary topology files. It implements features such as spring creation from a cutoff, add or remove springs thanks to a selection language, modify particle positions…

`mergespn` aims to merge two BioSpring binary topology files. This is particularly useful for creating a system in which several parts have different flexibility levels. Two spring network models can be created for two distinct molecules of a system, with different spring cutoffs and/or spring stiffness thanks to two calls to `pdb2spn`. They are gathered together in a second time thanks to `mergespn` which can additionally create springs between the two structures.

The compilation process

BioSpring software is made of 98 source files representing a total exceeding 20,000 lines of C++ code. It has several dependencies such as the NetCDF library and a few more libraries. The build process can therefore not be managed by hand.

For decades, developers used the GNU build system known as *Autotools*, a suite of programs designed to generate a `configure` script for the project. This script, to be executed by the user prior to compilation, generates the `Makefile` that will produce the software targets (programs, libraries, *etc.*) by invoking the command `make` (figure 3.6). This system has proven both its robustness and difficulty of use since Autotools input files have a very particular syntax that makes writing them tough and improvement even tougher since the whole file has to be read again and understood before being modified. The CMake software was developed in this context in the early 2000s, with a main objective to simplify the writing of configuration files. The developer has to write basically a single input file and CMake generates the appropriate `Makefile`. Notably, CMake configuration variables are very easy to modify, so the build settings can be tuned very quickly. CMake is now used by thousands of developers to compile hundreds of projects including very large projects such as KDE, a Unix desktop environment, MySQL, a database management system and BioSpring!

BioSpring lead developers chose CMake as build process management system, facing the fact that the less time is spent on the compilation, the more is spent on the actual code development. My contribution to the build process has been to improve the CMake input file by adding several options to customize the build process and make the input file clearer from a general point a view. A very interesting CMake feature is the package search: CMake can search for libraries, programs or any kind of dependency a project has. I wrote several CMake search scripts that were not already included in the CMake package, such as for the NetCDF and the MDDriver libraries for example.

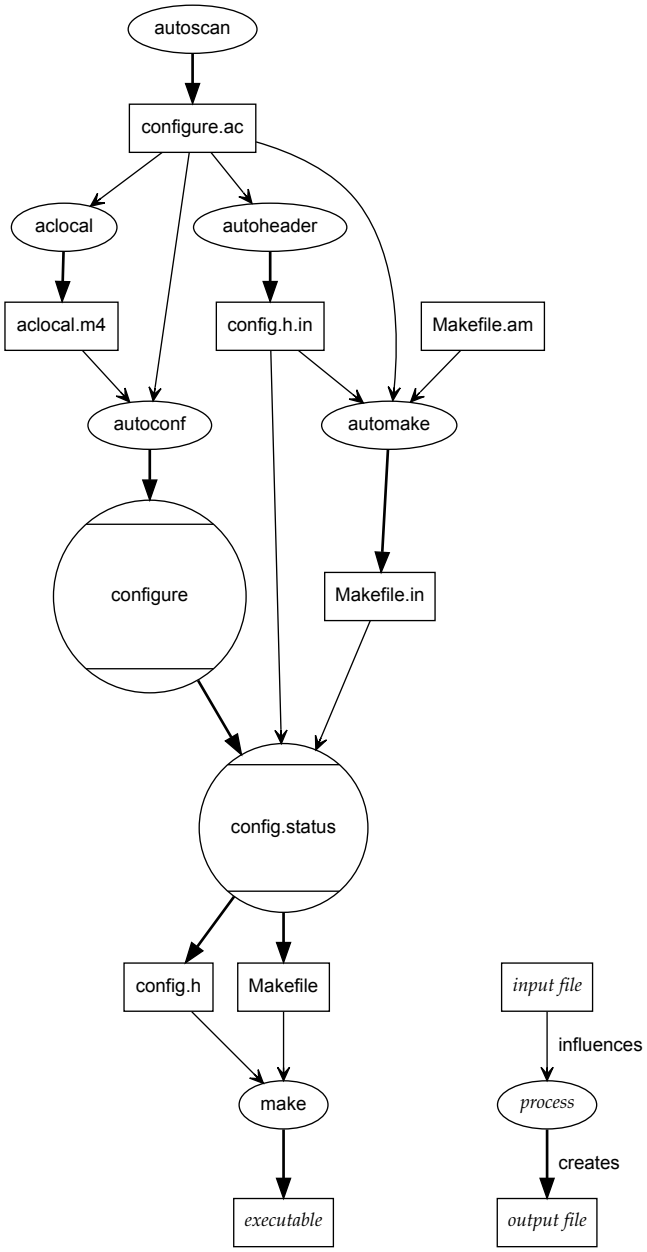

Figure 3.6 – GNU autoconf and automake process for generating makefiles. From http://www.wikipedia.org

Probing pLGICs with bromoform reveals many interconnected anesthetic binding sites

4

This chapter is devoted to the characterization of general anesthetic bromoform binding sites described in Sauguet *et al.* (2013a). I characterize the three sites and an additional pore site described in a new crystal structure of GLIC in locally closed conformation. I combine several computational approaches to address three key questions: (i) are the crystal binding sites spontaneously accessible? (ii) can bromoform travel from one site to another? (iii) what is the bromoform affinity for each binding site? Molecular dynamics simulations of flooding the receptor with bromoform recover most of the experimentally observed sites, with a modulated occupancy between the open and the locally closed conformations. Sixty short MD simulations were carried out to extensively explore the binding pockets, providing data on possible routes connecting them. These simulations furthermore highlight residues such as Y197 that may play key roles in controlling the interaction between anesthetic and receptor molecules. FEB calculations indicate significant affinity for all crystallographic binding sites in open and locally-closed conformations, in some cases modulated by pH. They support the critical role of Y197 into anesthetic binding.

The chapter is largely inspired by a scientific article currently in preparation. I am the first author of this article. Furthermore this work features contributions from Ludovic Sauguet and Marc Delarue (Institut Pasteur) who covered the X-ray crystallography part of this project, my colleague Samuel Murail who carried out the setup of microsecond timescale MD simulations as well as the main part of their analysis, and Marc Baaden who supervised the project. I was in charge of writing the article, managing the different contributions, running and analyzing all short MD simulations, running and analyzing all FEB calculations as well as partly analyzing long MD simulations.

4.1 Results

4.1.1 Bromoform-bound crystal structure of the GLIC channel in its locally-closed conformation

In order to study the properties of bromoform-binding to the GLIC receptor in its locally-closed conformation, we grew crystals of the GLIC K33C L246C variant in the presence of bromoform. GLIC K33C L246C variant is a particularly adapted model for this study as it is known to crystallize in a locally-closed

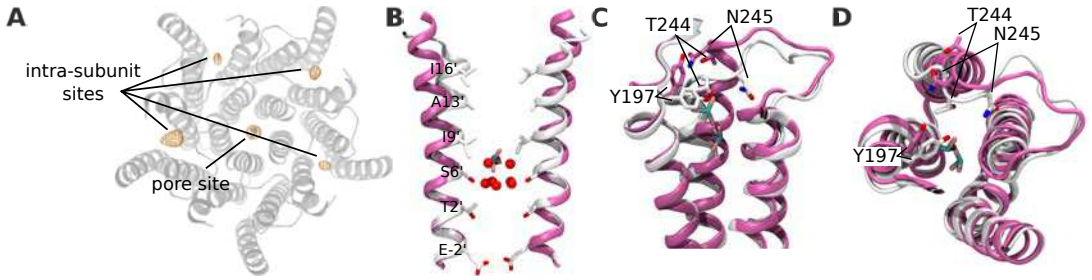

Figure 4.1 – A bromoform-bound structure of GLIC in its locally closed conformation. A) Top view of GLIC TMD. Bromoform densities, shown in ochre wireframe, are present in the intrasubunit site of four over five subunit, and in the new pore site. B) Location of bromoform (sticks) in the pore site. Two M2 helix backbones of the LC and open conformations are represented in pink and white, respectively. Water molecules detected in GLIC high resolution structure are represented as red spheres. C-D) Side and top view of one intrasubunit site in open (pink) and locally closed (white) conformations. Residues possibly responsible for decrease of site W3 accessibility are showed as sticks.

conformation but shares indistinguishable electrophysiological properties with WT GLIC (Prevost *et al.*, 2012). Bromoform is an analogue of chloroform containing three bromine atoms that produce a specific anomalous signal that can be observed by crystallography using X-rays with tunable wavelengths. The bromoform-bound structure was determined at a 2.95 Å resolution (figure 4.1).

It was completely superimposable with the apo-form of GLIC K33C L246C with a root mean square deviation of 0.77 Å over the 1555 C_α atoms. In the pore, a bromoform-binding site is indicated by a Fo-Fc electron density peak (7.0 σ) that overlaps with a bromine-specific anomalous peak (10.0 σ). Bromoform binds in the middle part of the pore between the I240 (I9') and S236 (S6') rings of residues (figure 4.1B), two critical rings of residues that are repectively involved in gating and ion permeation (Sauguet *et al.*, 2013b). This bromoform-binding site in GLIC is novel and is specifically occupied when the channel pore is closed. In contrast, when the channel pore is open, this location is filled of ordered water molecules that were found to be critical for ion permeation. Interestingly, bromoform was found to occupy a similar location in ELIC 's closed pore (Spurny *et al.*, 2013). A previous study revealed that bromoform occupies alternatively three poses in the intrasubunit cavity of the GLIC open-channel structure (named W1 to W3). In contrast, this intrasubunit cavity is remodelled in the GLIC locally-closed structure thus affecting the previously described bromoform binding sites (figure 4.1C-D). Indeed, despite the presence of an intrasubunit bromine anomalous signal in four out of five subunits, the absence of interpretable Fo-Fc difference electron density supports the possibility that bromoform may also bind at positions W1 and W2, but with too low occupancy or too high mobility to allow for confident model building. This is caused by the side chain of residue Y197 that alternates between two conformations. The second one induces a steric clash that prevents binding of bromoform at W1 and W2 sites (figure 4.2). In addition, the revolving motion of the M2-M3 loop partly occludes the intrasubunit cavity and prevents bromoform-binding at position W3. In summary, bromoform binding-sites are different in the locally-closed versus the open GLIC structures: while a novel site is observed in the pore, binding to the intrasubunit cavity is discouraged in the locally-closed form.

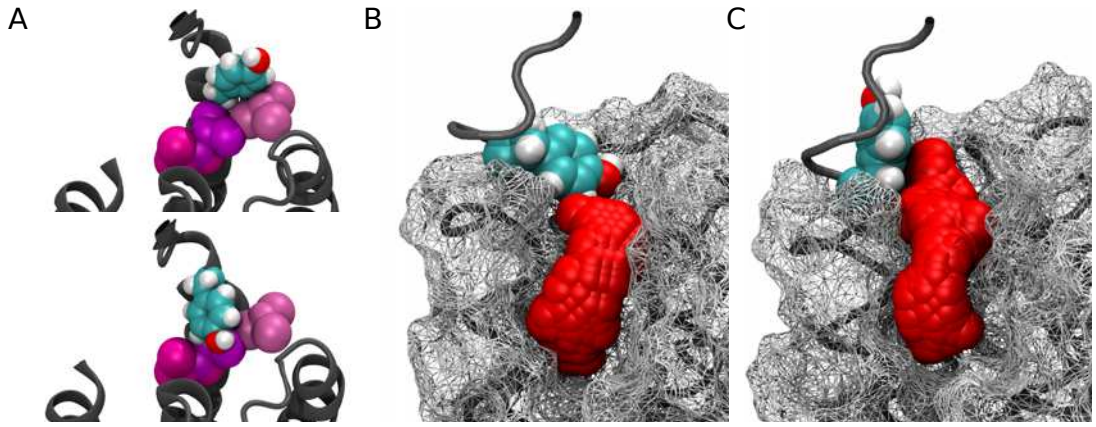

Figure 4.2 – Two distinct conformations of residue Y197. A) In open conformation (top), Y197 side chain (space filling representation colored by atom type) does not overlap any of the 3 intrasubunit bromoform binding sites (pink, purple and magenta spheres), while in down conformation (bottom) it partially overlaps site W2 and W3. B-C) Influence of Y197 side chain orientation on the intrasubunit pocket accessible volume (represented with red spheres and calculated with the Epock software, see section 3.3). The protein surface is represented as a white wireframe, its backbone as a grey tube.

4.1.2 Molecular Dynamics simulations to explore and quantify anesthetics binding

I combine three complementary simulation strategies to explore bromoform binding to GLIC. Firstly, I used data from several one microsecond long MD simulations of membrane inserted GLIC in an over-saturated bromoform solution to assess the spontaneous exploration of the system by the anesthetic and identify preferential bromoform binding sites. We subsequently refer to this type of simulations as *flooding* experiments. Three forms of GLIC were used, WT GLIC in open and LC conformation and the GLIC mutant F238A in the open conformation. Secondly, we ran 25 unconstrained 8-ns long MD simulations of GLIC F238A in open conformation starting from anesthetic locations in sites W1-2 and site B1. Considering that, at this timescale, the five subunits are independent, we accumulated a total of $25 \times 5 \times 8$ ns sampling per system. This one microsecond dataset for each location yielded an extensive exploration of both the intra- and the intersubunit binding pockets and allowed us to observe transitions between sites. Ten 8-ns simulations were run for GLIC WT in the LC form starting with bromoform in the pore site P1. Thirdly, we determined the bromoform affinity for each of the 6 binding sites in WT and F238A mutant for both open and LC variants of GLIC using alchemical free energy calculations. Sampling times are given in table 4.1, full technical details of all the simulation approaches are provided in section 2.6.

4.1.3 Crystallographic sites are spontaneously reachable

Flooding experiments reveal that all binding sites (i.e. W1, W2 W3, B1, B2, and P1), are spontaneously reachable in at least one of the three simulations. By design, the ion channel in the short MD simulations remains very close to the crystal structure, which makes for straightforward comparison with experiment. The observed occupancies for sites W1 and W2 are equivalent as was observed in the crystal (respectively 0.41 and 0.38, table 4.2). In contrast, the equilibrium is shifted in favor of the membrane exposed W1 site in flooding simulations. In both short and long MD simulations, the occupancy of the W1-2 sites is

Sequence	Form	Bromoform pore	Sampling time
Long MD simulations			
F238A	Open	Flooding	1 µs
WT	LC	Flooding	1 µs
WT	Open	Flooding	1 µs
Short MD simulations			
F238A	Open	Site W1-2	25 × 8 = 200 ns
F238A	Open	Site B1	25 × 8 = 200 ns
WT	LC	Site P1	10 × 8 = 80 ns
FEB calculations			
F238A	Open	W1-2, W3, B1, B2, P1	32 windows sampled for 3 to 10 ns each
F238A	LC		
WT	Open		
WT	LC		

Table 4.1 – Sampling time and studied systems for bromoform characterization.

markedly higher than that of site W3, which is consistent with crystallographic data. During flooding simulation of WT GLIC in LC conformation, site W3 occupancy was lower by one order-of-magnitude compared to the simulation starting from the open form and displayed a particularly low residence time (3.4 ns in average), consistently with crystallographic data. As the intersubunit cavity B1 does not exist in WT GLIC because of the presence of the bulky F238 residue sidechain, this site has only been reached in the simulation of the F238A mutant. Spontaneously, the pore site P1 has been reached in the flooding simulation of WT open GLIC only.

		Site W1	Site W2	Site W3	Site B1	Site B2	Site P1
F238A – O	Site W1	0.41	0.38	0.03	0.01	0.00	0.01
F238A – O	Site B1	0.00	0.00	0.00	0.94	0.00	0.00
WT – LC	Site P1	0.00	0.00	0.00	0.00	0.00	1.00
WT – O	Flooding	0.56	0.43	0.35	0.00	0.12	0.49
WT – LC	Flooding	0.45	0.24	0.03	0.00	0.00	0.00
F238A – O	Flooding	0.46	0.23	0.25	0.02	0.19	0.00

Table 4.2 – Bromoform binding site occupancy along MD simulations. The 3 first rows correspond to short MD simulations with bromoform being placed in site W1, B1 or P1, respectively. The 3 last rows refer to flooding simulations

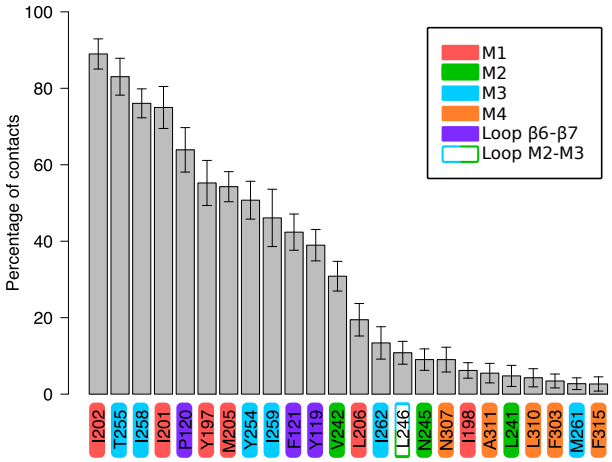

Figure 4.3 – Key residues of the intrasubunit pocket. Percentage frames in which a bromoform molecule contacts a residue of the intrasubunit pocket along short MD simulations.

4.1.4 All sites are interconnected, with gates between them

Considering the close vicinity of the experimentally observed sites W1,2,3 and B1, we characterized the dynamics of bromoform bound to each site. The simulations furthermore generate data to examine possible paths between these sites.

Site W1-2 may act as an entrance to other sites

In the flooding simulations, site W1 was the most occupied site mainly because it is exposed to the membrane. During the second half of the production run around 95 % of bromoform molecules were located in the membrane, as a consequence the W1 sites were the easiest to access and the first to be bound. In short simulations bromoform did not often penetrate much deeper into the intrasubunit pocket as depicted by both the low site W3 occupancy and the relatively low number of contacts with M2 residues (figure 4.3). Concerning site W3, its occupation in the flooding runs was significantly higher compared to the short simulations but only in the open form of GLIC. In the simulation of LC GLIC, the occupation of sites W1 and W2 were close to GLIC in open form, however site W3 was weakly occupied. In the LC form, the conformation of the M2-M3 loop and of the top part of the M2 helix prevent occupation of site W3, in particular residue T245 is in close contact with this site, preventing any binding (figure 4.4). A key observation is that, once site W3 has been reached, bromoform was able to enter the upper intersubunit B2 pocket, as previously mentioned. The same behavior was observed in flooding simulations of open GLIC (WT and F238A), which is reflected by the high occupancy of site B2.

Y197: a gate to the inner channel

Interestingly, Y197 which is not in the immediate environment of bromoform in the crystal structure, appears to dramatically modulate the volume of the intrasubunit pocket (figure 4.2), as proposed by Mowrey *et al.* (2013b). Therefore Y197 might control access to sites W3 and B2. In the crystal structure, the Y197 side chain is oriented toward the extracellular domain, the dihedral angle θ between Y197 C-C_α-C_β-C_γ atoms being equal to $(167.6 \pm 0.5)°$. Notably, in available crystal structures of open GLIC, all

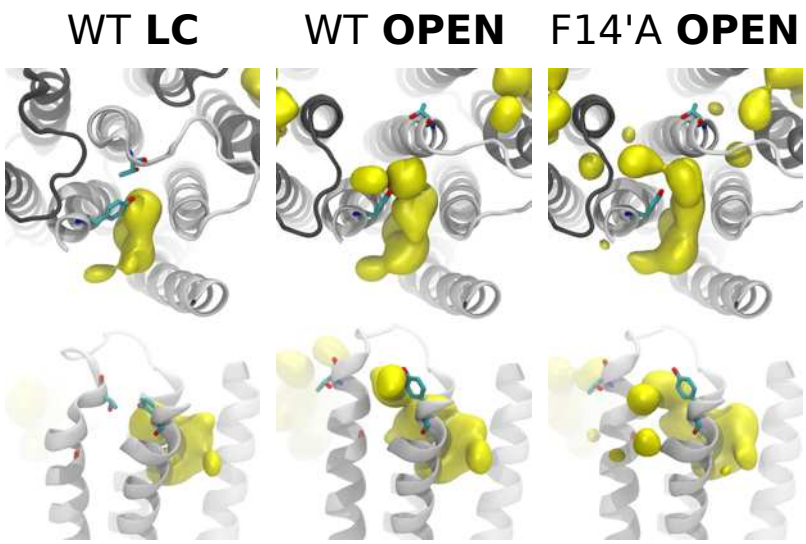

Figure 4.4 – Bromoform exploration in flooding simulations. Top (top pannel) and side (bottom pannel) views of a GLIC subunit in which bromoform occupancies are represented as yellow surfaces. Residues Y197 and T244 are represented with sticks colored by atom type.

Y197 residues display this *up* conformation (θ = 165.6 ± 3.8°), while closed structures of subtypes I and III respectively display four and two Y197 in *down* conformation, in which the Y197 side chain lies inside the intrasubunit pocket (θ =60.4 ± 8.5°). During the MD simulations, θ oscillated around these values with average orientations at 169.9° and 72.9° (figure 4.5). Notably, in down conformation, the Y197 side chain plunges into the intrasubunit pocket and occupies a large volume there, overlapping sites W2 and 3, hindering bromoform entering the deeper intrasubunit pocket (figure 4.2), and therefore ultimately prevents it from entering the intersubunit pocket. It is to be noted that the transition of the Y197 rotamer is a rare event with an average of 10 ± 6 transitions per microsecond calculated on the flooding simulations dataset.

Bromoform is confined within the intersubunit site

In the intersubunit cavity, bromoform mostly stays within 6.5 Å of the crystal structure location. Throughout the one-microsecond dataset provided by short MDs, a single transition from site B1 to site B2 was observed. Two residues, $L241_{i}$ and $E243_{i-1}$, restrain the available space in this region and therefore hinder crossing from site B1 to B2 (either way). During the F238A flooding simulation, the same inverse transition (B2 to B1) was observed once (figure 4.6). In that case, the molecule has been traveling from the membrane to the W1-2 site to W3, to B2, to finally reach site B1. Another intersubunit site was occupied by one molecule for more than 800 ns. This site is located 5 Å above B1 and slightly closer to the membrane. The bromoform molecule reaches that site from site W2 and stays there for the rest of the simulation.

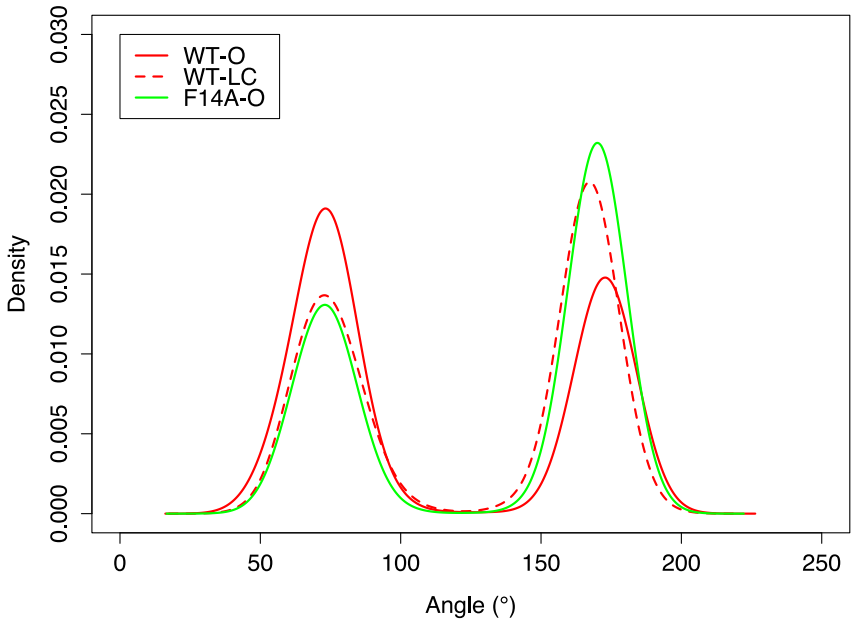

Figure 4.5 – Y197 side chain orientation along flooding simulations. Distribution of the Y197 side chain orientation along flooding MD simulations for each system namely wild-type (WT) open (O) and locally closed (LC) and the F14'A mutant in open conformation (F14A-O). Densities have been calculated over the a microsecond period with a time step of 0.5 ns, leading to a total of 10,000 points per density (2000 points per Y197 × 5 subunits).

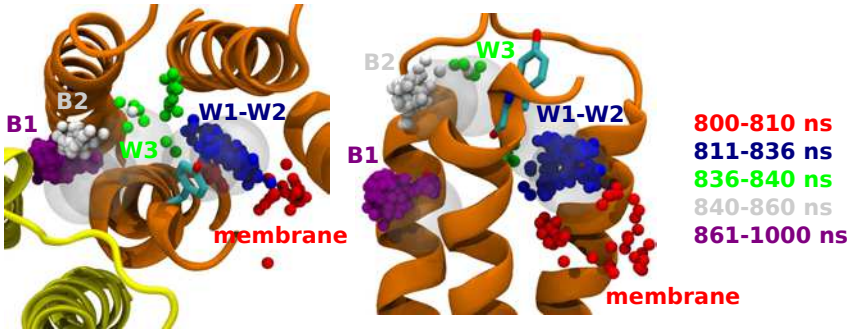

Figure 4.6 – Transition of a bromoform molecule from the membrane to the B1 site. Top (left) and side (right) views of a GLIC subunit. Bromoform center of mass is represented as spheres colored according to the time of the simulation. In 200 ns, this bromoform molecule could pass from the membrane to site W1-2, W3, B2 and ultimately B1.

Bromoform is stable in site P1

Short simulations of the locally closed form of GLIC with bromoform bound to the pore site P1 show that the anesthetic molecule remains close to this site, never leaving the region delimited by residues T226 (T2') and I233 (I9'), even in the absence of additional stabilizing bromoform molecules that were present in the flooding simulations but not in the short runs. Flooding simulations of WT open GLIC showed anesthetic molecules switching from the ECD vestibule to the pore site P1. As a result, the occupancy of site P1 in that simulation reaches 0.49 (table 4.2), with a total of three bromoform molecules present in the pore. One bound at the beginning of the simulation and stayed continuously in site P1 for more than 400 ns before binding an upper pore site (between I9' and A13') and coming back to site P1 twice for a few nanoseconds. The two other molecules were observed binding the upper site in the pore, the first one came early (after ~ 100 ns) from the vestibule and stayed in the upper cavity for the rest of the simulation with the exception of one 3 ns binding event in site P1 halfway through the simulation (~ 500 ns). The other molecule came from the intra subunit cavity (W1-2 for 205 ns then site W3 for 365 ns) at the very end of the simulation (~ 950 ns). In the LC form simulation, no anesthetics were shown to bind to the pore site. During the F238A mutant simulation, no anesthetics were shown to bind in site P1, however one molecule was binding the upper pore site in the second half of the simulation (~ 550 ns). The molecule came from the intra subunit cavity (W1-2 for 67 ns) and later an intermediate position between B2 and W3 for 63 ns. To be noted, none of the bromoform molecules, which bound in the pore, left the pore.

Bromoform binding affinities are favorable for all binding sites

Bromoform affinity for crystallographic binding sites ranges from −7.1 to −4.8 kcal/mol (figure 4.7). The open form appears significantly more favorable for sites W3, B1 and B2 with $|\Delta\Delta G|$ ranging from 1.3 to 2.8 kcal/mol, while P1 displays a higher affinity in the LC form than in the open form with $|\Delta\Delta G|$ of 1.7 and 2.4 kcal/mol respectively for the WT and the F238A channel. It should be noted that the pore site P1 of both WT and F238A channels displays a favorable FEB, comparable to site W1-2 in open form and even more favorable than W1-2 in the channel LC form.

Bromoform free energy of binding is sensitive to H235 protonation state.

The H235 residue is located on the pore-lining M2 helix (H11' in prime notation), close to the B1 site entrance. The protonation state of its side chain has been shown to be particularly difficult to determine with confidence (Laurent *et al.*, 2013). To assess whether the protonation state of H235 modulates the ligand binding affinity, we compared bromoform free energy of binding for neutral and protonated H235 (table 4.3). When neutral, we observe a difference of −1.9 kcal/mol between the WT intrasubunit site W1 (−6.3 ± 0.1 kcal/mol) and the F238A intersubunit site B1 (−8.2 ± 0.1 kcal/mol). When H235 is protonated, the FEB to B1 reduces to (−6.1 ± 0.1) kcal/mol. No effect was observed on the more remote site W1.

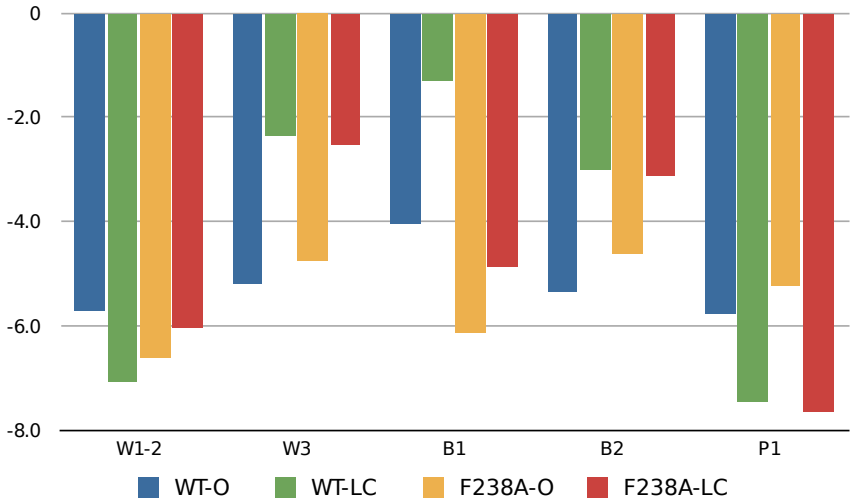

Figure 4.7 – Free energies of binding of bromoform to the five binding sites. Energies are given in kcal/mol. WT = "Wild-Type"; O = "Open"; LC = "Locally-Closed". Error estimates are all below or equal to 0.2 kcal/mol.

	H235 +	H235 n
Site W1 (intra)	−6.6	−6.3
Site B1 (inter)	−6.1	−8.2
Δ = W1 − B1	−0.5	+1.9

Table 4.3 – Free energy of binding of bromoform as a function of H235 protonation state Energies are given in kcal/mol; "+" stands for double charged, "n" stands for neutral.

4.2 Discussion

Globally, a picture of dynamically accessible and interconnected anesthetic binding sites emerges from this computational study, in excellent agreement with the available crystallographic data. The calculations reveal phenomena enriching the picture obtained from the experimental data such as the transitions between sites or the possible modulation of anesthetic binding affinities by pH.

4.2.1 Multi-site allosteric modulation, a delicate balance toward potentiation or inhibition

Evidence that anesthetics bind the intrasubunit site in the W1 region is strong. Crystal structures showed that bromoform (Sauguet *et al.*, 2013a), propofol and desflurane bind to this pocket (Nury *et al.*, 2011; Chiara *et al.*, 2014). The data collected in this work are in very good agreement, clearly showing that intrasubunit sites are spontaneously accessible from the membrane and display favorable FEB, especially for W1. Our data suggest that sites W1 and W2 should be considered as two marginally different poses of the same site. The ligand may switch from one to another with equal probabilities as is the case in short MD simulations. On a longer timescale, this equilibrium shifts in favor of site W1, probably because of a bias induced by its direct exposure to the membrane where most bromoform molecules accumulate.

In WT GLIC, site B1 does not exist since the presence of the bulky F238 side chain does not leave enough room for an anesthetic molecule, as it leaves barely enough space for a single water molecule. Interestingly the F238 residue is conserved in the human nicotinic acetylcholine receptor subunits α3,4,5 and β1,2 and 5-HTR subunits 3A,B. In the glycine and the $GABA_A$ receptors this residue is substituted by less bulky residues, respectively Q and L/I. Howard and coworkers showed that this substitution creates an intersubunit pocket (corresponding to the B1 binding site) in which ethanol can bind, explaining the potentiating effect on the channel. We observe in this study that the artificial intersubunit pocket created in GLIC by the mutation F238A can be reached from intrasubunit site W1. Calculated free energies of binding reflect the high affinity of bromoform for B1 in both open and locally closed forms of GLIC F238A. This observation suggests that the anesthetics' initial binding site could be site W1 for both inhibitory and excitatory channels. The anesthetic would then migrate to B1 in inhibitory channels, in this way stabilizing the open form, while remaining in the intrasubunit pocket in excitatory channels stabilizing the closed form. Such a scenario would support hypotheses proposing that the difference of action of anesthetics (and alcohols) on inhibitory and activating Cys-loop receptors might be found in the accessibility of the lower intersubunit pocket (Murail *et al.*, 2012).

Besides transitions from site W1-2 to B1, our data show a high mobility of bromoform inside the binding regions (see figure 4.8). Notably, we observe an important exchange rate between site W1-2 and site W3 over the microsecond period. The average occupancy of site W3 appears significantly lower than for site W1-2 in both short and long MD simulations, which is in very good agreement with crystallographic data. Importantly, we observe several transitions from site W3 to site B2, an intersubunit site described in GLIC on the basis of MD simulations (Nury *et al.*, 2011) and in $GABA_A$ by photolabeling (Yip *et al.*, 2013) with respectively desflurane and ortho-propofol diazirine. As we observe bromoform occupying this site on the microsecond timescale, we argue that site B2 is definitely occupied by antagonists at a timescale

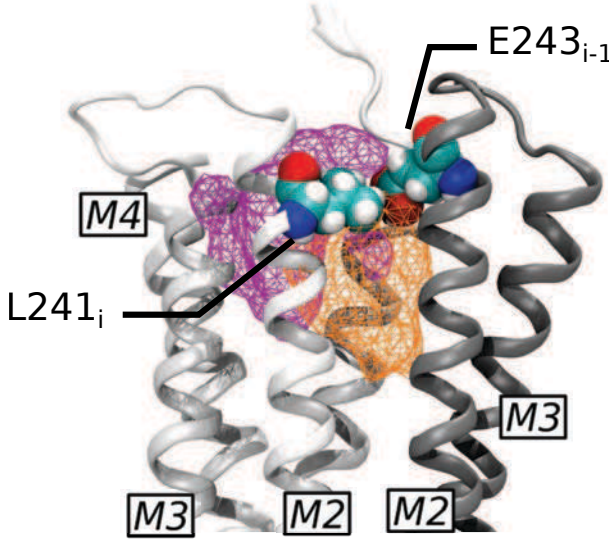

Figure 4.8 – Bromoform exploration of the intra- and intersubunit binding pockets in short MD simulations. Side view from the pore. The protein TMD backbone atoms are represented as cartoon colored by subunits. The area explored by the bromoform along short MD simulations are represented with meshes colored according to the bromoform starting location (purple and orange for intra- and intersubunit pockets, respectively). Two residues that seem to occlude the route between B1 and B2 intersubunit sites are represented with space filling spheres colored by atom type.

relevant for anesthesia. Still, the anesthetic has to be able to access site B2 through a route passing by sites W1-2 and W3, a route that can be occluded by Y197.

4.2.2 A residue gating the access to anesthetic allosteric binding sites

Our data strongly suggest that the orientation of the Y197 side chain is critical for anesthetic binding and insertion depth; therefore their transition to intersubunit cavities may be controlled in this way. The up conformation of Y197 as described above appears highly conserved in all GLIC open structures, while the down conformation is found in the majority of the locally closed structures. In addition, the M2 helices bending in the locally closed conformation move residues from the top of M2 and from the M2-M3 loop inside the intrasubunit pocket and in particular residue T245. Therefore, site W3 is not accessible anymore from site W1 in the locally closed channel conformation, as shown in the crystal structure of bromoform bound to LC GLIC presented here. Free energy calculations corroborate these data showing that site W1 is clearly more favorable to bromoform than site W3 and B2 in the LC structures, while in GLIC 's open form, this difference is less clear (figure 4.7). Importantly, structural alignments reveal that Y197 is highly conserved in Cys-loop receptors, including, nicotinic acetylcholine, 5HT3 and glycine receptors (Sauguet *et al.*, 2013a). Interestingly, in $GABA_AR$, the tyrosine is substituted by a phenylalanine, two residues with very similar side chains, especially considering their volume. We argue this residue might play a critical role in Cys-loop receptors' sensitivity to anesthetics. Our simulations show that Y197 orientation affects the F195 rotamer distribution. In the down conformation the Y197 residue prevents F195 from being in the same conformation, and vice versa. The F195 residue, given its direct interaction with the M2-M3 helix, may be a key residue to modulate GLIC gating. This residue is a glycine in all $GABA_A$

Y197 side chain orientation	GLIC subtype	Site W1	Site W2	Site W3
Up	WT – Open	0.16	0.28	0.31
	WT – LC	0.29	0.20	0.03
	F238A – Open	0.36	0.22	0.25
Down	WT – Open	0.40	0.15	0.00
	WT – LC	0.16	0.04	0.00
	F238A – Open	0.10	0.01	0.00

Table 4.4 – Bromoform occupancy of intrasubunit sites along flooding MD simulations according to the Y197 residue side chain orientation.

receptors except for subunit ρ and mutation of this residue G222F was affecting potentiation by propofol and inducing faster desensitization (Engblom *et al.*, 2002; Chang *et al.*, 2003). Given the timescale involved for the Y197 side chain transition, our simulations cannot accurately sample the two state populations as a function of GLIC's structural state (LC vs. open) or bromoform binding. However, even in open GLIC, site W2 and W3 occupancies are correlated with Y197 side chain orientation (table 4.4). Further studies need to be conducted to understand which factor induces the transition of Y197 toward one state or the other. Interestingly, we find that binding to site B1 can be affected by single residue properties, namely the protonation state of H235.

4.2.3 H11' protonation state impacts the potentiating site

GLIC 's activity has been shown to be modulated by residue protonation state (Cheng *et al.*, 2010). H235 is located on the M2 helix and is close to the intersubunit site B1 entrance. Its protonation state is still highly debated, some studies reporting neutralizing its side chain (Brömstrup *et al.*, 2013; Prevost *et al.*, 2012; Zhu and Hummer, 2012b), while others report a doubly protonated H235 (Mowrey *et al.*, 2013b; LeBard *et al.*, 2012; Bocquet *et al.*, 2009). Free energy of binding calculations indicate that, when the H235' side chain is neutral, the most favorable location for bromoform is B1, favored over W1 by 1.9 kcal/mol (see table 4.3), observation supported by Brömstrup *et al.*, 2013). However, this difference vanishes when H235 is doubly protonated ($|\Delta\Delta G|$ = 0.5 kcal/mol). The intrasubunit site W1 does not seem to be affected by H235 protonation state. Importantly, several groups reported this residue as being critical for the channel activity (Prevost *et al.*, 2012; Wang *et al.*, 2012). We demonstrate here that the protonation state of H235 may play a key role for ligand binding to the intersubunit site B1, reinforcing the hypothesis that this residue might play a pivotal role for channel function and modulation. We hypothesize that the F238A mutation perturbs the helix bundle. The double protonation of H235 might be an additional disruptive factor that would explain the significant difference in bromoform free energy of binding compared to the single protonated variant.

4.2.4 The "pore binding site" hypothesis is supported by crystallographic and simulation data.

Flooding simulations were carried out to assess the accessibility of the anesthetic binding sites. We demonstrated that bromoform could spontaneously reach the pore site (P1) from the ECD vestibule or from intra subunit site W1-2 and site W3/B2. Short MD simulations show that bromoform is stable in P1, consistently with free energy of binding calculations indicating a relatively strong binding to this site, especially in GLIC 's LC form. Moreover at least one more binding site was observed in the pore. While pore binding sites in pLGICs have been proposed both based on experimental (Hilf *et al.*, 2010; Spurny *et al.*, 2013) and theoretical data (Brannigan *et al.*, 2010; Chen *et al.*, 2010; Murail *et al.*, 2012), the mechanism of action of pLGIC inhibitors is not yet proven. Considering the narrowness of the pore, it is intuitive to think about inhibitors as "steric channel blockers", molecules that bind and obstruct the pore due to their size. This mechanism is particularly relevant for large inhibitors such as lidocaine that has been described as an open channel blocker consequently to its blockage of ion passage in the channel open state. However, an allosteric inhibition mechanism involving a pore site cannot be rigorously excluded. Bromoform, by increasing the pore hydrophobicity, may dry it, which would cause M2 helix bending and therefore channel closure. This hypothesis is supported by our MD data showing a spontaneous binding to the open channel and is consistent with several works on GLIC 's gating describing the channel drying prior/simultaneous to channel M2 helix bending and channel closure (Nury *et al.*, 2010; Zhu and Hummer, 2012b; Mowrey *et al.*, 2013b). This hypothesis is consistent with the crystallographic structure presented here, clearly displaying a bromoform molecule bound to GLIC 's locally closed state. Hence, bromoform may bind the channel resting state and stabilize it in the same fashion.

4.3 Conclusion

Supported by a high-resolution crystal structure of bromoform bound to the pore of the locally closed GLIC channel, I show using extensive calculations that this site is not only spontaneously accessible from the solvent, but is very favorable to bromoform binding. This is documented by free energy of binding calculations and binding site exploration along unconstrained molecular dynamics simulations. These data suggest that pore-induced channel closure can be at least partly responsible for channel inactivation by anesthetics. Data are however to be compared with calculations made in the same conditions on allosteric binding sites. Results show comparable accessibilities and affinities with the pore site. Since mutagenesis data unambiguously showed an effect of mutations in allosteric binding site regions on anesthetics action, the underlying allosteric mechanism might still be the key for a full understanding of anesthesia mechanisms at an atomistic level. Molecular dynamics simulations reveal that parts of the intrasubunit pocket are not accessible to bromoform in the channel locally closed conformation due to the structure of the M2-M3 loop. This work highlights the crucial role of residue Y197 which is likely to act as a gate for anesthetic binding to the intersubunit pocket by hindering its access *via* the intrasubunit pocket. Finally, routes are revealed between intra- and intersubunit binding sites suggesting that the intrasubunit pocket may act as an entrance to the channel.

The next chapter will investigate the binding of two other anesthetics to GLIC, propofol and desflurane.

These additional simulations will provide further support to generalize the results on the possible molecular mechanism of general anesthesia obtained for bromoform.

Propofol & desflurane simulations provide new insights into anesthetic action at the atomic scale

5

Propofol and desflurane are two widely used general anesthetics that have been shown to inhibit GLIC at clinically relevant concentrations (Weng *et al.*, 2010) and to bind the channel in the intrasubunit site W1 (Nury *et al.*, 2011). Propofol and desflurane inhibit WT GLIC with very similar efficiency but their action on the T255A GLIC mutant is opposite: propofol inhibits the GLIC T255A mutant more than WT GLIC while desflurane inhibits the T255A mutant dramatically less than WT GLIC (table 5.1).

I started my PhD project with the aim of understanding the role of the T255A mutation on GLIC's inhibition by anesthetics. For this purpose, I used the short MD simulation approach as developed in section 2.6.1. To acquire an extended picture, I studied propofol and desflurane bound to WT and T255A GLIC variants in both open and locally closed conformations, leading to a total of 8 systems or 200 independent MD simulations. Furthermore, I have been interested in studying the influence of the ligand binding symmetry on both its dynamics and the protein geometry. I therefore simulated systems in which only one anesthetic molecule was bound to GLIC, as opposed to all other short MD simulations I ran in which five anesthetic molecules were bound to the channel to maximize sampling.

As in January 2013 I switched to the characterization of bromoform's interactions with GLIC because of the novel crystallographic data that became available to me, I did not quite finalize this aspect of my project, mainly because of the large amount of data I generated (see section 3.1.3) and the inherent complexity of the statistical analysis of these data. This part is dedicated to highlighting some of the most important results I obtained characterizing propofol and desflurane interactions with WT and T255A GLIC variants,

	Desflurane	Propofol
WT	27 ± 13	24 ± 6
T255A	1400 ± 1100	2 ± 1

Table 5.1 – Inhibition of GLIC by two general anesthetics. Half maximal inhibitory concentration of desflurane and propofol for modulation of WT and T255A GLIC given in μM. Values taken from Nury *et al.* (2011).

	W1-2	W3	B1	B2	P1	Other
WT	0.48	0.05	0.00	0.01	0.00	0.45
WT-LC	0.56	0.02	0.00	0.00	0.00	0.42
T255A	0.58	0.04	0.00	0.01	0.00	0.37
T255A-LC	0.84	0.00	0.00	0.00	0.00	0.15

Table 5.2 – Propofol occupancy of crystallographic binding sites in short MDs.

in open and LC conformation.

5.1 Results

5.1.1 An extensive sampling close to the crystal structure

To ensure that GLIC's conformation stays close to the crystal structure without restraining its movements, we ran 25 slightly different short simulations of each system (see table 2.3). As each simulation is 8 ns long, we finally get a sampling of 200 ns per system. Considering that each anesthetic molecule acts independently from the ones in the neighboring subunits, a hypothesis that will be verified later on in this chapter, each simulation actually accounts for 5 times the sampling *i.e.* a total of 1 µs per system.

As shown on figure 5.1, GLIC's structure only drifts very little at the timescale of the simulation. The initial configurations of the propofol and desflurane systems have been extracted from previous simulations (section 2.6.1). As a result, the average structure RMSD value is in the order of 2 and 3 Å for the LC and open structure, respectively. This difference of stability between the two conformations can be easily explained by the fact that in the locally closed conformation the top of the M2 helices is bent toward the center of the pore, increasing the stability of the bundle compared to the open conformation in which M2 helices are more mobile. In contrast, in the simulations of bromoform bound to GLIC, the systems have been equilibrated with GLIC's conformation constrained to the crystal structure. The RMSD tends to increase along the simulation, which reflects the adjustment of the channel to its environment which is quite different from a crystalline packing. Importantly, the most flexible regions are loops in the upper part of the extracellular domain. The TMD remains very stable, especially the M2 helices bundle.

5.1.2 Anesthetics are mobile within the W1 binding site

Despite the large occupancy of site W1-2 (table 5.2), general anesthetics propofol and desflurane show high mobility within the intracellular binding pocket (figure 5.2). Notably, both anesthetics reached the intersubunit site B2 in open GLIC. In locally closed conformation, the accessibility of site B2 through site W3 appears jeopardized, as depicted with bromoform simulations.

However, marked anesthetic mobility differences have been revealed. Propofol exploration of the intrasubunit pocket is comparable to bromoform, in the sense that it explores mostly W1-2 and W3 regions and rarely exits the intrasubunit pocket to bind the protein surface (table 5.4 and appendix D.2). Note however that bromoform has markedly higher occupancy for site W1-2 compared to propofol.

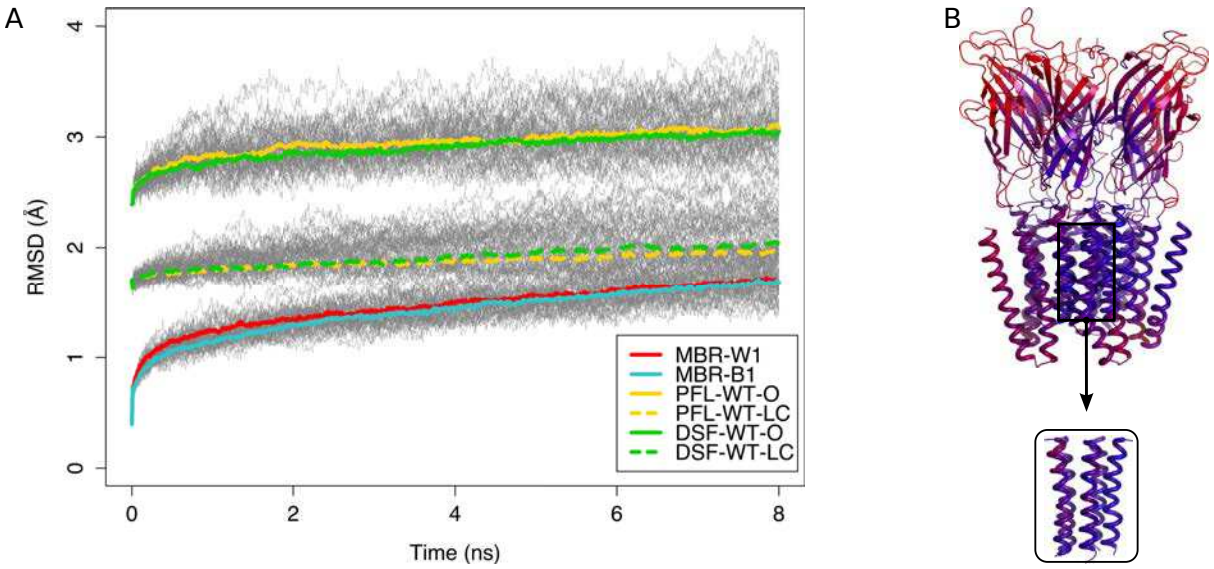

Figure 5.1 – A sampling close to the crystal structure. RMSD and RMSD per residue calculated against GLIC's high resolution structure (PDB-id 4HFI). A) RMSD Timeline average for different systems. The RMSD has been calculated on carbon α atoms of GLIC's structured regions. Individual RMSD timelines are shown in black transparent background, for visual perception of the raw data complexity. B) RMSD per residue the reference structure and the average structure of simulations of propofol and desflurane bound to open GLIC. Color scale ranges from 0.0 (blue) to 5.0 (red).

	W1-2	W3	B1	B2	P1	Other
WT	0.24	0.07	0.00	0.01	0.00	0.69
T255A	0.30	0.05	0.00	0.01	0.00	0.63
WT-LC	0.23	0.00	0.00	0.01	0.00	0.75
T255A-LC	0.54	0.00	0.00	0.00	0.00	0.46

Table 5.3 – Desflurane occupancy of crystallographic binding sites in short MDs.

Desflurane displays a dramatically increased mobility compared to the latters, as reflected by its lower occupancy of site W1, where it was initially posed (table 5.3). Desflurane appears more subject to leave the intrasubunit pocket and bind the surface of the protein in the TMD. More importantly, desflurane repeatedly entered an intersubunit region close to site B2, not passing through site W3 but between the M2 and M3 helices, while propofol did not.

As expected because of the presence of a phenylalanine residue at position 238 (F14') in both WT and T255A variants of GLIC, no binding to the intersubunit site B1 has been observed during these MDs, for none of both anesthetics.

5.1.3 Different mobilities impact the number of contacts with the receptor

To quantify the interactions between the anesthetics and GLIC, I calculated the number of contacts that are made between the two parties during the simulations (figure 5.3).

The majority of contacts are made by the residues that are the closest in the crystal structure, with

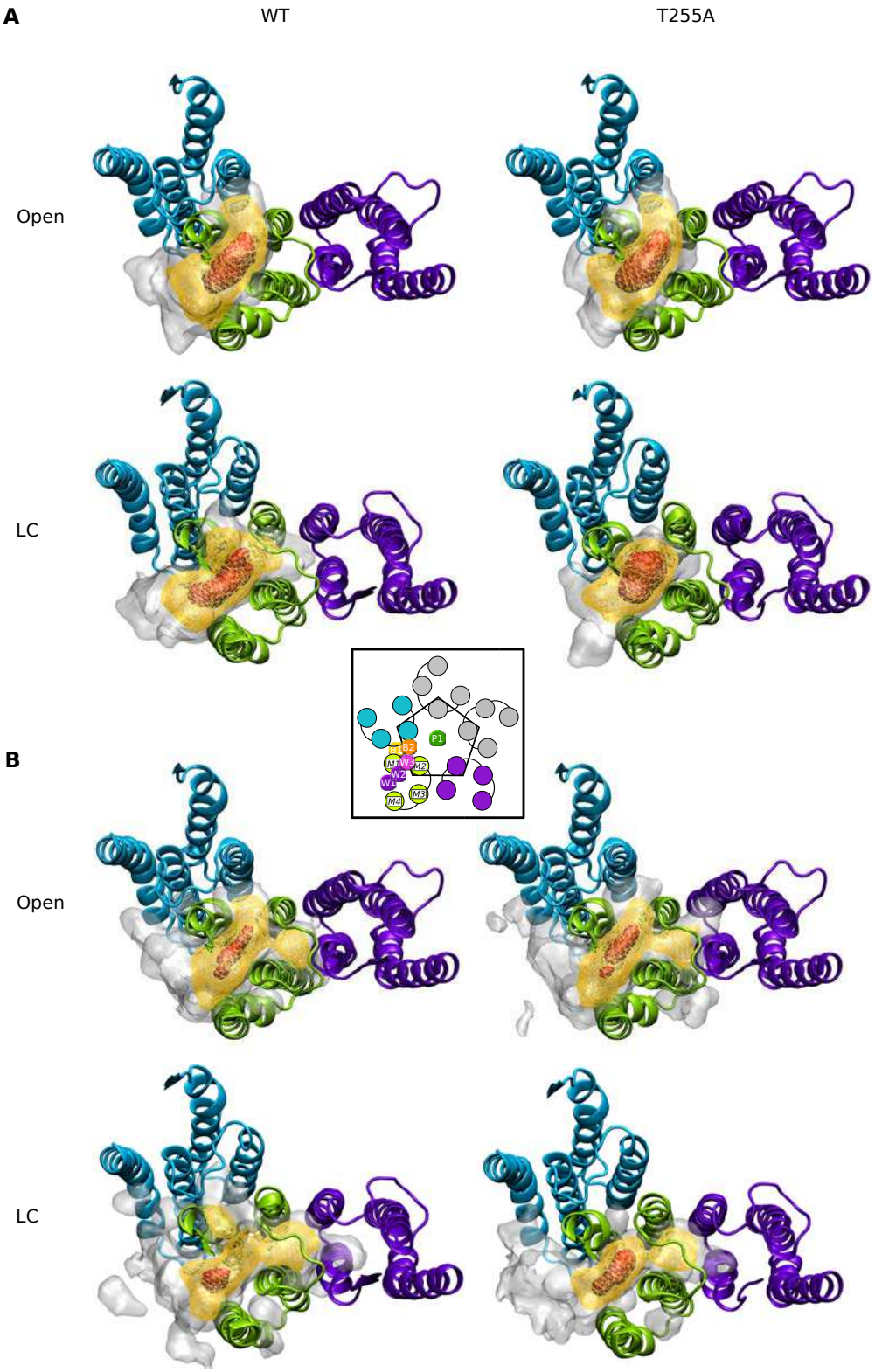

Figure 5.2 – Exploration of anesthetic molecules bound to W1. Exploration of propofol (A) and desflurane (B) bound to open, locally closed, WT and T255A variant of GLIC. The TMD of 3 subunits (blue and green and purple) are represented here for clarity. 3 occupancies corresponding to as many occupancy cutoffs are represented here: 0.1, 0.01 and 0.001 in red opaque surface, yellow mesh and white transparent surface, respectively.

	W1-2	W3
Bromoform	0.80	0.03
Propofol	0.48	0.05
Desflurane	0.24	0.07

Table 5.4 – Comparison of anesthetic occupancies of the intrasubunit pocket binding sites. Comparison of W1-2 and W3 occupancies in WT open GLIC (desflurane & propofol) and F238 open GLIC (bromoform).

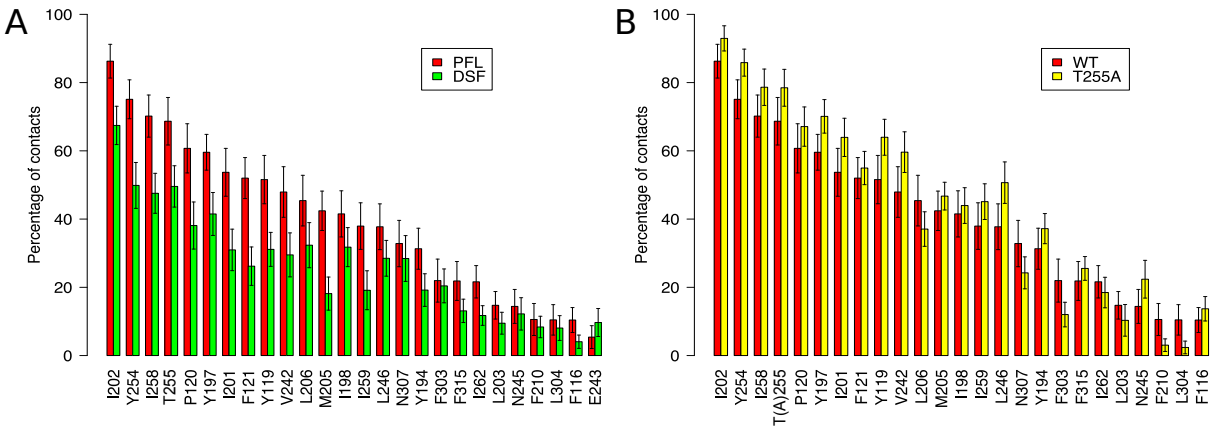

Figure 5.3 – Anesthetics contacts with open GLIC. Most frequent contacts between anesthetics and GLIC residues. A) Comparison between two anesthetics, propofol (PFL) and desflurane (DSF) in the WT open channel. B) Comparison of the contacts made by propofol in the WT and the T255A mutant channels. Error bars represent standard errors calculated from standard deviations with significance level $\alpha = 0.05$.

contributions from M1 (I201, I202, M205, and L206), M2 (V242), M3 (Y254, T255, I258, and I259), and from the β6-β7 loop (Y119, P120, and F121). Residues from the M4 helix (N307 and F303) appear less contacted than what has been presumed from the crystal sructures. Notably, residue Y197, whose side chain is oriented toward the ECD in the crystal structure and therefore does not seem to interact in any way with the anesthetic, appears here to be one of the most contacted residues.

The difference of mobility between propofol and desflurane is not without consequences on the contacts anesthetics make with the protein. Figure 5.3A shows numerous quantitative differences in the contacts desflurane makes with GLIC compared to propofol. Propofol hits the 10 most contacted residues significantly more than desflurane does. The same observation is made with T255A open GLIC, which shows no clear difference with the WT open form besides (figure 5.3B).

5.1.4 Ligand binding stretches the intrasubunit pocket

To investigate if anesthetic binding has an effect on the pocket geometry, I used the Epock software (see section 3.3) to calculate the volume of binding pockets along the trajectories.

The volume of the intrasubunit pocket W1 appears highly correlated to the anesthetic's volume: the larger the anesthetic, the wider the pocket (table 5.5 and figure 5.4). The same tendency is observed in the crystal structures. Note that, at this timescale, the pocket volume calculated for GLIC crystal structures

	Anesthetic volume (Å^3)	Pocket crystal volume (Å^3)	Pocket MD volume (Å^3)
Propofol	193	349 ± 22	408 ± 2
Desflurane	88	325 ± 26	309 ± 3
Bromoform	83	309 ± 29	284 ± 1

Table 5.5 – Volume of the intrasubunit pocket W1 in open GLIC in presence of anesthetic molecules. Average volume calculated on the last 3 ns of open GLIC short simulations. The anesthetic volume has been calculated with Epock. The pocket crystal volume confidence intervals are calculated from the five pockets that the pentamer displays.

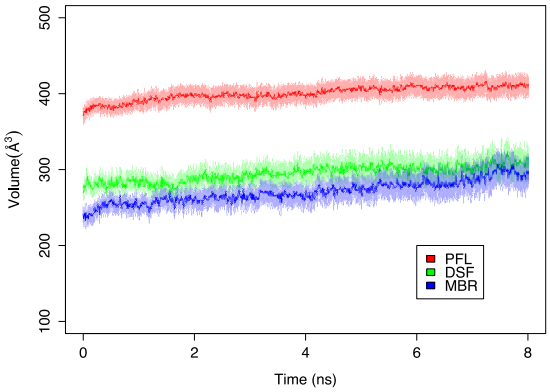

Figure 5.4 – Volume of the intrasubunit pocket W1 occupied by 3 different anesthetics. Volume of the pocket W1 occupied by propofol (PFL, red), desflurane (DSF, green) or bromoform (MBR, blue). Error bars represent standard errors calculated from standard deviations with significance level $\alpha = 0.05$.

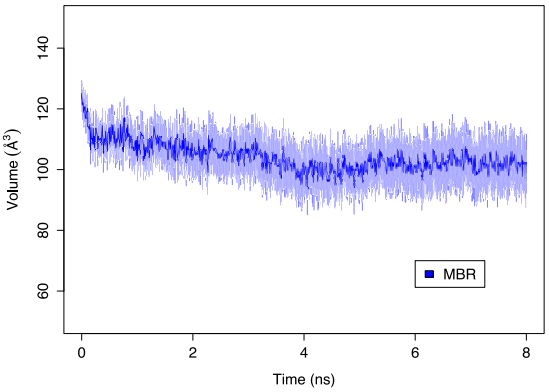

Figure 5.5 – Volume of the intersubunit pocket B1 occupied by bromoform. Volume of the intersubunit pocket created by the mutant F238A in GLIC. Error bars represent standard errors calculated from standard deviations with significance level $\alpha = 0.05$.

appears slightly lower than in the MD simulations, except for propofol for which the pocket volume is higher in the simulations than in the crystal (see appendix D.3).

The volume of the intersubunit pocket B1 in the GLIC F238A mutant in presence of a bromoform molecule is (101 ± 1) Å^3 compared to (102 ± 1) Å^3 in absence of bromoform (figure 5.5). Notably, this value is significantly lower than the volume observed in the crystal structure, in which the calculated volume of this pocket is (194 ± 12) Å^3.

5.1.5 Ligand binding does not impact neighboring cavities

I used simulations in which only one anesthetic molecule was bound in GLIC's intrasubunit pocket to assert if an effect of the anesthetic can be detected on the pockets located in the other subunits. Pockets that neighbor a cavity filled with an anesthetic will be referred to as *neighboring cavities*, while the two cavities in between two neighboring cavities will be referred to as *empty*.

Figure 5.6 shows the volume of intrasubunit pockets filled with an anesthetic, neighboring or empty. Neighboring cavities display an average volume of (148 ± 1) Å^3 and (125 ± 1) Å^3 for propofol and desflurane,

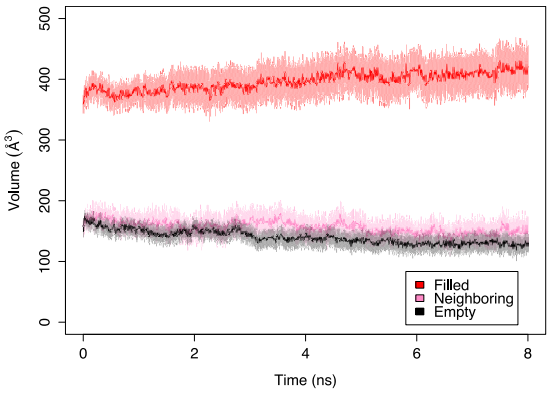

Figure 5.6 – Influence of propofol binding on neighboring pockets. Volume of the intrasubunit pocket W1 filled with propofol (red), neighboring a pocket filled with propofol (pink) or empty (black). Error bars represent standard errors calculated from standard deviations with significance level $\alpha = 0.05$.

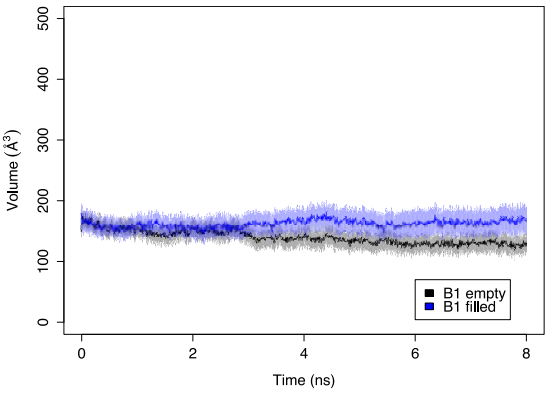

Figure 5.7 – Influence of bromoform binding to B1 on W1. Volume of the intrasubunit pocket if B1 is filled with a bromoform molecule (blue) or empty (black). Error bars represent standard errors calculated from standard deviations with significance level $\alpha = 0.05$.

respectively. Empty pockets in simulations display a shift of approximately the same range with (139 ± 1) Å^3 and (120 ± 2) Å^3 for propofol and desflurane, respectively.

The data collected also permitted to search for a potential effect of ligand binding in the intersubunit pocket on the intrasubunit pocket volume. Figure 5.7 compares the volume of the empty intrasubunit pocket when the intersubunit pocket B1 is filled with bromoform and when it is empty. The volume of the empty intrasubunit pocket is (129 ± 1) Å^3 (calculated by averaging of propofol and desflurane data) compared to (163 ± 1) Å^3 when a bromoform is bound to the intersubunit pocket.

5.1.6 Tyrosine 197 conformations are stabilized by hydrogen bonds

As observed with bromoform simulations, the Y197 residue side chain switched between up and down conformations in simulations of propofol and desflurane bound to GLIC. Each conformation turned out to be stabilized through hydrogen bonds.

The up conformation is defined by a Y197 side chain orientation toward the ECD. It is stabilized mainly by hydrogen bonding with D32 (figure 5.8), which is located on the β1-β2 loop. Since pK_a calculations suggest that D32 is most probably deprotonated, it would act as an acceptor of hydrogen bonds. While D32 and Y197 are at a distance of more than 5 Å in the open GLIC crystal structure, the flexibility of the β1-β2 loop makes hydrogen bonding with Y197 not only possible but quite frequent.

The down conformation is defined by a Y197 side chain orientation perpendicular to the M1 helix axis. Notably, in down conformation, Y197's side chain can face either the M3 helix and make hydrogen bonds with residue T255 or face the M2 helix and make hydrogen bonds with residue N245. In GLIC's open form, Y197 and N245 are more than 5 Å apart from each other and N245 faces the neighboring intersubunit. Hydrogen bonding between the two residues is greatly facilitated in GLIC's LC conformation since the bending of the top of the M2 helices brings a few residues, including N245, toward the intrasubunit pocket.

Finally, a fourth residue is found to make hydrogen bonds with Y197, namely Y119, which is located on

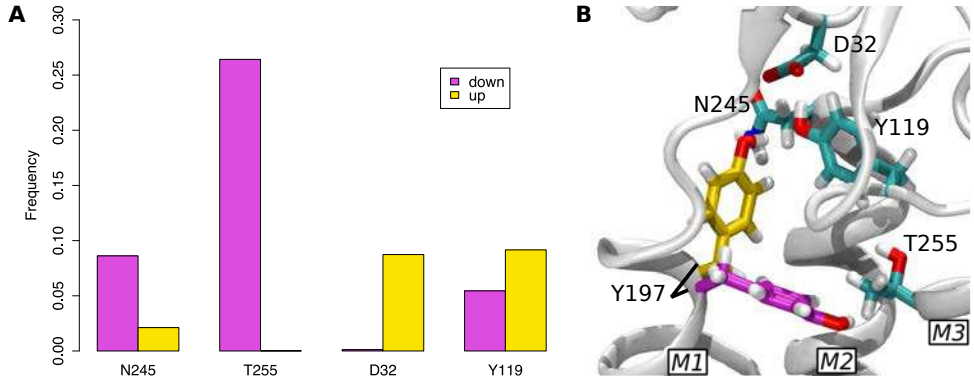

Figure 5.8 – Y197 hydrogen bonds to surrounding residues. A) Frequency of hydrogen bonds made with tyrosine 197 in down and up conformation in simulations of bromoform, propofol and desflurane bound to WT GLIC. Error estimates calculated by boostrapping are lower than 10^{-3}.
B) Residues forming hydrogen bonding with Y197. Y197 up and down conformations are colored in yellow and magenta, respectively. The protein backbone atoms are represented as a white cartoon.

the β6-β7 loop. Interestingly, this residue is close enough to Y197 to make hydrogen bonds when Y197 is in up as well as in down conformation. Y119 is itself stabilized by hydrogen bonds with D32. Y197's up conformation therefore depends directly and indirectly on the position of D32, and more generally of the β1-β2 loop.

5.1.7 Y197's stability is modulated by the T255A mutation

Simulations of open GLIC display a clear picture in which the two modes of the distributions of the Y197 side chain orientation are virtually perfectly centered around the same values: $(73.8 \pm 1.2)°$ for the down conformation and $(173.1 \pm 0.8)°$ for the up conformation (figure 5.9). For both desflurane and propofol and both WT and T255A GLIC, the up conformation is favored over the down conformation with a frequency of over 80 % (except for the desflurane bound to the T255A GLIC mutant simulations in which this probability reaches 77 %).

Transitions from up to down conformation have however been occasionally observed at this timescale and the down conformation appears quite stable as depicted by the lifetime of this conformation which is regularly greater than 5 ns.

Simulations of LC GLIC were carried out with Y197 side chains initially placed essentially in down conformation. Simulations of WT-LC GLIC display similar behaviors for both anesthetics with 21 % and 23 % in up conformation for desflurane and propofol, respectively. In contrast, simulations of T255A GLIC show a displaced equilibrium in favor of the up conformation, especially for propofol simulations that display 65 % in up conformation *vs* 36 % in up conformation for desflurane.

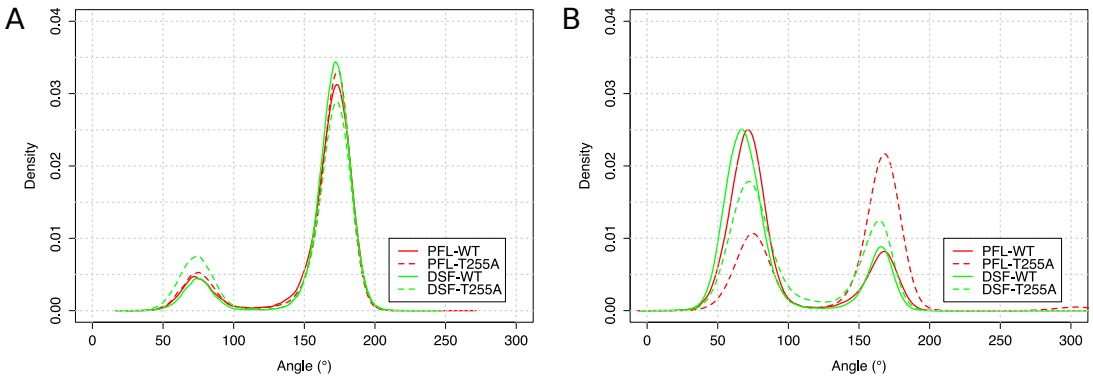

Figure 5.9 – Tyrosine 197 side chain orientation in open and LC GLIC. Distribution of the dihedral angle between Y197 $C - C_\alpha - C_\beta - C_\gamma$ atoms for propofol, desflurane, wild-type and T255A GLIC variants in open (A) and LC (B) conformations. Each distribution is made of 100,125 values.

5.2 Discussion

5.2.1 Binding to the intersubunit site B2 confirmed

A connection between the intrasubunit site W1 and intersubunit site B2 had been detected for a desflurane (DSF) molecule in Nury *et al.* (2011). However, this observation was based on a single 30 ns long MD simulation in which the movements of M2 helices might have been important. A major finding concerning the characterization of bromoform binding to GLIC is that this transition from site W1 to site B2 *via* site W3 has also been observed for bromoform in several 8 ns long MD simulations, meaning in a GLIC conformation close to the crystal structure. Still, bromoform as well as desflurane are relatively small molecules and a very important question to ask is: "can a bigger anesthetic, such as propofol, travel the same way and bind the intersubunit site B2?". The results presented here not only confirm that desflurane can bind to site B2 but that propofol, a much larger molecule (193 Å^3 *vs* 88 Å^3 *vs* 84 Å^3 for propofol, desflurane and bromoform respectively), can bind this site despite the steric hindrance. This result, supported by a recent photolabeling of propofol bound to the site B2 of a $GABA_AR$ (Yip *et al.*, 2013), suggests that general anesthetics could bind and play a role in pLGICs modulation through this site, despite the diversity of their physico-chemical properties. Although the exact consequence of GA binding to this site is to be explored, the work I realized on bromoform suggests that site B2 could be a transient location prior to pore entry (see chapter 4).

Another major point of this work is the high mobility of desflurane bound to GLIC. While all simulations started with anesthetics bound to the intrasubunit pocket, desflurane showed an increased tendency to leave the pocket and bind the surface of the protein, at the interface with the lipid phase. Importantly, desflurane could bind the intersubunit region of site B2 without passing through site W3 but passing between the M2 and M3 helices, a behavior that contrasts with both propofol and bromoform which almost never took this route, despite the sampling that could be considered higher for the two latter anesthetics, since they rarely exit the intrasubunit pocket. A possible explanation for this contrast between propofol and desflurane are the stronger steric constraints that apply to propofol compared to desflurane. The difference of behavior between desflurane and bromoform is to be investigated, as both anesthetics have

similar volume, desflurane is even slightly more bulky than bromoform.

5.2.2 A more detailed contact map: the role of Y197 confirmed

All simulations reflect the important number of contacts between anesthetics and Y197 compared to other residues. As it is a reasonable assumption that a residue which is often in contact with a ligand plays a role in its binding, it is quite important to investigate the role of Y197 because of the critical impact the orientation of its side chain has on the accessibility of both the intrasubunit pocket and therefore the intersubunit site B2.

Simulations of open GLIC show two well distinct Y197 side chain orientations, with modes that do not depend neither on the anesthetic nor GLIC type (WT or T255A mutant). Interestingly, while the side chain of Y197 is in up conformation more than 80 % of the time, the probability of contact with the anesthetic is high, over 40 % for desflurane and close to 60 % for propofol. There are two main reasons for this observation: i) the anesthetic can contact Y197's backbone when its side chain is oriented toward the ECD and ii) even when in up conformation, only small variations in orientation can result in contacts between the Y197 side chain and molecules inside the intrasubunit pocket, especially if molecules are bulky such as propofol. Finally, because of its location at the top of the M1 helix, Y197 has privileged relations with the ligand when bound, as just stated with respect to site W1-2, but also when bound to site W3 which is located between the M1 and M2 helices.

In contrast, simulations of LC GLIC display marked differences between the wild-type and the T255A forms of GLIC. As the Y197 side chain was initially imposed in down conformation, it is not surprising that simulations of WT GLIC show a higher propensity to adopt this conformation (or more probably to stay in this conformation as the simulations last 8 ns each). In the same fashion as in open GLIC the Y197 side chain switched from up to down conformation, in simulations of LC GLIC Y197 occasionally switched from down to up conformation. Notably, in the GLIC T255A mutant simulations, the equilibrium is displaced in favor of the up conformation. In the case of propofol, the probability of being in up conformation is even higher than the one of being in down conformation. Since the Y197 down conformation is mainly stabilized by hydrogen bonding to T255, it is expected that the lifetime of the Y197 down conformation would be reduced in a mutant in which T255 is replaced by a residue that cannot form an hydrogen bond with Y197. The reason why this behavior is enhanced in GLIC's locally closed conformation could be related to the number of contacts with the residues of the intrasubunit pocket: desflurane has 2 times more contacts with these residues in the T255A mutant compared to WT for LC GLIC; for propofol, this ratio increases to 7. More specifically, the number of contacts between the anesthetic and residue 255 is significantly higher in the T255A mutant than in the WT form. These observations support the hypothesis of a competition between Y197-T255 hydrogen bonding and anesthetic binding. More data is however needed to investigate this particular point, especially long simulations of anesthetics bound to the intrasubunit pockets in WT and T255A GLIC variants would generate valuable data.

5.2.3 Influence of the ligand binding symmetry

The data collected allowed to clearly demonstrate an effect of the anesthetic on the binding pocket geometry, especially its volume. This effect seems highly correlated with the anesthetic's volume, bigger anesthetics

causing more stretching of the binding pocket than smaller ones.

To assess whether this effect impacts neighboring cavities, I compared the volume of empty neighboring binding pockets to the volume of empty non-neighboring pockets. Interestingly, different average volumes were calculated in simulations of propofol and desflurane bound to GLIC, with (148 ± 1) Å^3 and (125 ± 1) Å^3, respectively. Theses values, however, are not consistent with the hypothesis of an impact of anesthetics on neighboring binding pockets since the expected effect would be a decrease of the volume due to anesthetic stretching of its own binding pocket, resulting in the compression of neighboring cavities. This hypothesis implies that neighboring pocket volume would be maximal when no anesthetic is bound to a neighboring pocket, lower when desflurane is bound to a neighboring pocket, and minimal when propofol is bound to a neighboring pocket, which is not the case here. Moreover, volume differences observed here are quite low, with a maximum difference of 23 Å^3 between propofol and desflurane. It is important to consider that this value is small at the atomic scale, as it is lower than the volume of a single water molecule which is on the order of 30 Å^3. The differences of volume observed here are therefore to be considered within the range of the method uncertainty and probably do not reflect a real effect of the anesthetic on neighboring cavities.

The data collected on bromoform binding to intersubunit B1 suggest a possible effect on the neighboring intrasubunit volume, as the volume of the intrasubunit pocket calculated in these simulations is 34 Å^3 lower when bromoform is bound to B1 and when it is not. However, the remarks on the method uncertainty apply here too and more data is needed to investigate a possible effect of ligand binding to B1 on the intrasubunit pocket volume. Simulations of larger anesthetic molecules (*e.g.* propofol) bound to B1 would be of great help for this purpose.

The data collected did not allow to unambiguously detect any significant difference in anesthetic behavior with either one or five anesthetic molecules bound to GLIC. Interestingly, an earlier study suggests that an asymmetric ligand binding would facilitate conformational transitions in pLGICs (Mowrey *et al.*, 2013b). In this work, the authors compared the speed of open GLIC pore dehydration in presence of a varying number of propofol molecules bound to the intrasubunit pocket to assess whether the symmetry of ligand binding impacts the channel closure. Based on 100 ns long MD simulations, the authors suggest that the channel closure speed is maximal when two to three propofol molecules are bound to GLIC. The results I present in this manuscript are not in contradiction with Mowrey *et al.*'s results, since my data is based on 8 ns long MD simulations. At this timescale, it is not surprising that, to date, I have detected no significant effect either on the anesthetic itself or the protein dynamics. However, these data require in-depth investigation to conclude with confidence.

5.2.4 Understanding anesthetic's action

Desflurane *vs* propofol *vs* bromoform

Desflurane showed markedly higher mobility than propofol. Desflurane was able to enter the intersubunit region of B2 by passing between the M2 and M3 helices, under the M2-M3 loop. Notably, bromoform, which is smaller than desflurane, did not show similar behavior. Several reasons may explain that. First, desflurane sampling may be considered higher than bromoform sampling, considering that the initial poses of the anesthetic were extracted from a 30 ns long MD simulation for desflurane whereas they were

randomly generated within a cutoff of the crystal binding sites for bromoform. Hence, desflurane poses are more varied than bromoform ones and desflurane may explore additional regions. Note that this argument is two-edged since, as bromoform residence time in the binding pocket is higher, the sampling could actually be considered higher for bromoform than for desflurane. Second, the protein initial conformation has been equilibrated without restraints for desflurane, which was not the case for bromoform. As a result, the protein dynamics in bromoform and in desflurane MD simulations are not perfectly comparable and may have influenced the results. Finally, desflurane's mobility may simply reflect a lower affinity for anesthetic binding sites in GLIC which would increase its relative mobility compared to bromoform. To summarize, the incapacity of bromoform to enter the B2 regions by passing between M2 and M3 helices as seen for desflurane is unlikely to be due to a steric effect. A sampling bias or significantly higher affinity for the intrasubunit pocket are more probable in this case. Our current data can however not answer this question. Propofol sampling, in contrast, is probably adequate. Although propofol entered the intersubunit site B2, it certainly did not pass between the M2 and M3 helices, but rather between the M1 and M2 helices, through site W3. As a consequence, propofol bound site B2 only in GLIC's open form since the position of both Y197 and the M2-M3 loop in the LC form prevent entering site B2 from site W3. Again, if steric contraints are more likely to apply to propofol, more sampling (*e.g.* longer simulations) might be needed to investigate propofol's capacity to enter the B2 region in the same fashion as desflurane.

This work clearly shows that anesthetics of diverse physico-chemical nature can enter the intersubunit site B2 in GLIC's open form. The role of this site in GLIC's inhibition process in still unclear. As i) desflurane enters B2 more often than propofol does and ii) desflurane binds B2 in both GLIC open and LC states, desflurane would probably be expected to be a stronger inhibitor of GLIC's activity than propofol. Yet, both anesthetics have very similar half maximal inhibitory concentrations (see table 5.1). Notably, propofol and desflurane are different molecules that may have different local effects on the protein, as I demonstrated on the volume of the intrasubunit binding pocket for example. The only capacity of a molecule to access a potential inhibitory site with ease is therefore probably not the only factor that affects GLIC's inhibition.

Wild-type *vs* T255A variants

Very few significant differences have been detected between WT and T255A GLIC variants. Among them, the number of contacts with the pocket residues in GLIC locally closed form is often higher for the T225A mutant than for the WT form (see appendix D.1). More importantly, a marked difference of the Y197 side chain orientation has been detected: in WT GLIC, the Y197 down conformation is stabilized by hydrogen bonding with T255 which, by definition, is not possible in the T255A mutant. The possible effect of Y197 would be to prevent anesthetics from binding the deeper intrasubunit cavity when in down conformation. As a result, the down conformation is expected to be disadvantaged in the T255A mutant, therefore the accessibility to sites W3 and B2 is favored. The inhibition of GLIC is then expected to be more important in the T255A mutant than in the WT form. Notably, propofol is one order of magnitude more efficient in the T255A mutant than in the WT form of GLIC, consistently with the mechanism suggested above. The same tendency is observed with desflurane which would imply the same effect on GLIC's inhibition. However, GLIC's inhibition by desflurane is at least one order of magnitude lower in the T255A mutant

than in WT GLIC.

This apparent paradox can be explained by assuming a central role of the Y197 side chain. For propofol, because of its bulky size, the passage is generally hindered by the down conformation of Y197 and the T255A mutant induces a higher frequency of possible passages, increasing the inhibiting effect of the anesthetic. Desflurane on the other hand, a small and very mobile anesthetic as we showed previously, has no trouble to overcome the Y197-controlled passage even in wild type. The Y197 down conformation may help to mechanically keep it within the intersubunit binding pocket once it has reached it, whereas the T255A mutant may increase the probability of escape for this very mobile ligand, hence the decrease in inhibition compared to WT. To test this hypothesis, directed mutagenesis experiments might be of great help. Unfortunately, this mutant is unlikely to transmit electrical current in electrophysiology experiments realized by Pierre-Jean Corringer's team at Institut Pasteur. Yet, other mutations may help to understand the role of Y197. Unlike the T255A mutant which probably favors the Y197 side chain up conformation, mutations D32A and Y119A (and possibly their combined action) are expected to favor Y197 side chain down conformation. According to the hypothesis formulated above, these mutants are expected to produce an opposite action at those of the T255A mutant as summarized in table 5.6.

	Desflurane	Propofol
WT	++	++
T255A	+	+++
D32A/Y119A	+++	+

Table 5.6 – Predictive effect of mutant D32A and Y119A on GLIC's inhibition by propofol and desflurane. Inhibition potency is illustrated with + symbols: the more symbols, the more inhibition potency. According to the hypothesis developed in section 5.2.4, these mutations would have an opposite action to those recognized of the T255A mutation.

5.3 Conclusion

This study aimed to

- verify whether the mechanisms highlighted in the context of the characterization of bromoform binding to GLIC were applicable to other anesthetics,
- understand differences of action of propofol and desflurane in WT and T255A GLIC variants.

Several observations made in the context of the bromoform study turn out to be applicable to both propofol and desflurane. The three anesthetics have been found to bind the site B2 located at the interface between subunits and between the channel ECD and TMD. This observation is a major result because propofol is significantly more bulky than bromoform and desflurane and would not have been expected to bind this site. In GLIC locally closed form, access to site B2 from site W3 (located between M1 and M2 helices) is prevented by both Y197 side chain orientation and residues from the M2-M3 loop that face the intrasubunit pocket specifically in the channel LC form. Desflurane is the only anesthetic we scanned that

was able to find a way to the intersubunit site B2 not passing between M1 and M2 helices but between M2 and M3 helices, below the M2-M3 loop.

The key residues of the intrasubunit cavity contacted by the three anesthetics are the same, a significant number of contacts being made with Y197. Consistently with crystallographic data, this residue is confirmed to adopt two distinct conformations, up and down, the down conformation preventing deep access to the intrasubunit pocket (*i.e.* site W3), therefore site B2. The up conformation is mainly stabilized by hydrogen bonding to residues D32 and Y119, while in the down conformation hydrogen bonds are made with N245 and T255. The T255A mutation is consequently found to modulate the up-down equilibrium in favor of the up conformation, especially in GLIC locally-closed form, as longer simulations of GLIC open form would provide adequate data to compare the up-down equilibrium in WT and T255A channels.

To date, more insights into pLGICs modulation by anesthetics and alcohols are needed to fully understand their action mechanism. More specifically, if this study highlights the difference of mobility between propofol and desflurane, to my knowledge no currently proposed mechanism of action of anesthesia can explain the dual modulation of GLIC T255A mutant by desflurane and propofol. However, I present in this study an innovative hypothesis in which residue Y197 would act as a gate, preventing propofol from entering the intrasubunit pocket on one hand, and preventing desflurane from exiting the intersubunit region on the other hand. This hypothesis requires directed mutagenesis coupled with electrophysiology measurements to be verified.

Concluding Remarks, Perspectives & Thoughts

6

6.1 Conclusions

In quite few years, it is amazing how our knowledge of pLGICs developed. In 2004, no complete structure of any member of this family was solved. Today, tens of structures have been released of GLIC, ELIC and GluCl including wild type, mutants, apo-form or co-crystallized structures with ligands. To date, the channel for which the most data has been acquired is probably GLIC, which is an amazing coincidence when thinking about it. *Gloeobacter violaceus* (Gv) is the only cyanobacteria that does not have thylakoids, an intracellular membrane-bound compartment where light-dependent reactions of photosynthesis take place in other cyanobacteria (and in chloroplasts). It is this unique feature that pushed scientists to sequence Gv's genome in the first place (Nakamura *et al.*, 2003), at a time where no intensive work was being done on the bacteria. The results of this work were published only a few months prior to Tasneem *et al.*'s search for bacterial homologs to the human nicotinic receptors which led to the discovery of GLIC. Few years later, when I joined Marc Baaden's group for my first master internship in 2009, the first two structures of GLIC had just been released (Bocquet *et al.*, 2009; Hilf and Dutzler, 2009) and when I started to work on this project, in October 2010, the crystal structures of desflurane and propofol bound to GLIC were not even released.

More than three years have passed now and the picture of pLGICs action mechanism is step by step made clearer. In particular, this work provides significant insights on interactions between anesthetics and GLIC, which I believe is a good model for the whole pLGIC superfamily. I think that it is clear now that anesthetics can bind, and probably actually do bind, multiple parts of those channels. Extensive free energy of binding calculations have proven the relatively high affinity of bromoform for GLIC's pore, intra- and intersubunit pockets, with roughly equal energies of binding. Unconstrained flooding simulations have proven the spontaneous accessibility of all pockets on the hundred nanoseconds timescale. Since anesthetics can bind the channel pore, it is very intuitive to think of anesthetics as channel blockers, that would either sterically block ion flow, either provoke the pore closing (or stabilize the closed state) by binding to the pore. However, the opposite influence of desflurane and propofol on the GLIC T255A mutant cannot be explained by anesthetics binding to the pore alone. A more distant allosteric mechanism initiated from within the intrasubunit pocket probably occurs. Moreover, the extensive sampling provided by both flooding and short molecular dynamics simulations allowed me to detect and quantify transitions

between binding sites. Thus, numerous transitions from the intrasubunit pocket to the intersubunit site B2 have been detected for all three anesthetic molecules I tested, providing reasonable confidence on the existence of this site in a biological context. Taken together, the study of bromoform, desflurane and propofol suggests that B2 cannot be entered by the bulkiest anesthetics in the closed state of GLIC, as residues from the M2-M3 loop occlude the passage between M1 and M2 helices. Less voluminous anesthetics may still access B2's region by passing between M2 and M3 helices and under the M2-M3 loop, as depicted for desflurane. Larger timescale simulations suggested that B2 could be a transient site since the intersubunit site B1, which has been described as potentiating site in the glycine and $GABA_A$ receptors and in GLIC's F238A mutant, can been reached from site B2. The same set of simulations showed that anesthetic molecules can as well enter the channel pore vestibule from site B2. Yet, the precise role of this intersubunit region remains to be further investigated.

The data collected also highlighted the critical role of the Y197 residue in the deep insertion of anesthetics in the intrasubunit pocket. This very conserved residue among cationic pLGICs[1] lies at the intrasubunit pocket entrance. I clearly show in this study that the Y197 side chain can adopt two distinct conformations referred to in this manuscript as *up* and *down*. In the down conformation, the Y197 side chain lies down into the intrasubunit pocket, partially overlapping sites W2 and W3 and therefore dramatically reducing the pocket volume. This conformation is stabilized by hydrogen bonding mainly with T255. The possible role of Y197 in anesthetic binding to GLIC is still actively investigated thanks to simulations of anesthetics bound to the GLIC T255A mutant. More mutagenesis data would be critical for a better understanding of Y197's role.

6.2 Perspectives

In my opinion, the understanding of the action mechanism of anesthetics would be made a lot easier by the detailed comprehension of the pLGIC gating process. The recent structure of GLIC's resting state will probably lead to major breakthroughs in a near future. Molecular dynamics simulations have proven their ability to improve our knowledge of atomic scale mechanisms and is, in my opinion, a tool of choice to study pLGICs at atomistic scale. Simulations of GLIC full opening and closure will allow to understand the details of these processes. Extensive simulations of GLIC's full gating with anesthetics bound to the intrasubunit, intersubunit and pore sites will probably allow to understand anesthetics actions.

However, running such simulations is challenging. A single full gating process would possibly take place between two and five microseconds in an MD simulation. As an example, let us use the performance values described in section 3.1: the simulation software was run on 544 cores, calculates 21 ns/day and uses 630,000 h/µs computed. A single gating simulation would then consume between 1.3 and 3.2 millions hours in CPU time and would take between 3 and 8 months to run. Of course, confidence in subsequent discoveries would only come with the simulation of several anesthetics bound to WT GLIC and possibly mutants, and with anesthetics bound to different binding sites, as stated above. Furthermore, the complexity of analyzing such an ensemble of simulations should not be underestimated.

To date, the data collected on anesthetic dynamics while bound to a member of the pLGIC family are

[1] In anionic pLGICs, this tyrosine is substituted by a phenylalanine, a residue that has similar properties, especially considering it's volume.

still insufficient to guide research of more efficient and less risked anesthetic molecule. Many years of effort will be necessary to understand the full complexity of anesthetic action mechanism, and it is not even clear that significant drug design could be made from subsequent discoveries, since anesthetics are likely to bind numerous targets on numbers of neuroreceptors. Still, major improvements in clinical anesthesia will most probably require this crucial step and the hypotheses and ideas developed within this thesis may be useful in this context.

Appendices

Appendices to chapter 1

Below are supplementary materials that complement chapter “Biological Background”.

A.1 Comparing nAChR and GLIC packing

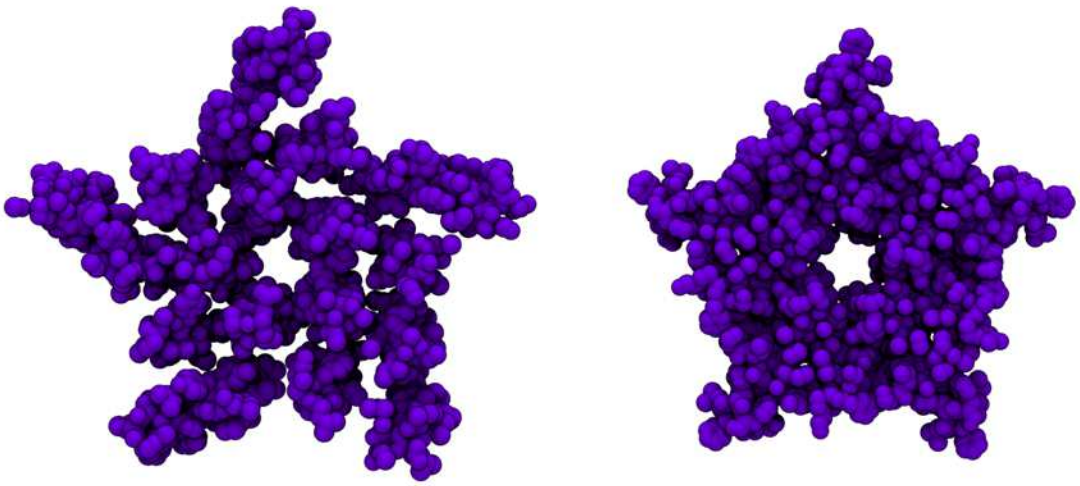

Figure A.1 – Compared packing of the nAChR and GLIC structures. View of the extracellular half of the transmembrane domains of the nAChR (left) and GLIC (right) receptors. The nAChR structure displays a significantly lower packing than GLIC inducing gaps in the structure that may be filled with cholesterol (see section 1.1.5).

Appendices to chapter 2

B

Below are supplementary materials that complement chapter “Molecular Modeling: Theory And Practice”.

B.1 Bromoform parameters in GROMACS format

```
;
; Gromacs file for Bromoform parameters calculated with TIP3P water.
;


[ atomtypes ]
; name at.num  mass   charge  ptype    sigma           epsilon
CX      6   12.0110  0.0  A         0.339967    0.457730
HCMM    1    1.0079  0.0  A         0.247135    0.0656888
BR     35   79.9040  0.0  A         0.347000    1.966480


[ moleculetype ]
; Name nrexcl
MBR 3

[ atoms ]
; nr type resnr resid atom cgnr charge mass

   1 CX   1  MBR C1      1  0.5841  12.0110
   2 BR   1  MBR Br1     2  0.0713  78.9183
   3 BR   1  MBR Br2     3  0.0713  78.9183
   4 BR   1  MBR Br3     4  0.0713  78.9183
   5 HCMM 1  MBR H1      5  0.3702   1.0079

[ bonds ]
; ai aj fu b0 kb, b0 kb
  1   2 1 0.1924  152300.1  0.1924   152300.1
  1   3 1 0.1924  152300.1  0.1924   152300.1
  1   4 1 0.1924  152300.1  0.1924   152300.1
  1   5 1 0.1110  287014.9  0.1110   287014.9

[ angles ]
; ai aj ak fu th0 kth ub0 kub th0 kth ub0 kub
  2   1   3 1  111.7000  658.21    111.7000  658.21
  2   1   4 1  111.7000  658.21    111.7000  658.21
  3   1   4 1  111.7000  658.21    111.7000  658.21
  2   1   5 1  107.1380  369.15    107.1380  369.15
  3   1   5 1  107.1380  369.15    107.1380  369.15
  4   1   5 1  107.1380  369.15    107.1380  369.15


#ifdef POSRES_LIGAND
[ position_restraints ]
; atom  type      fx      fy      fz
   1 1 1000 1000 1000
   2 1 1000 1000 1000
   3 1 1000 1000 1000
   4 1 1000 1000 1000
#endif
```

Figure B.1 – Bromoform parameters for GROMACS.

B.2 Pure and Applied Chemistry review

This paper is a review I wrote in 2012 and that has been published in Laurent, B., Murail, S., Da Silva, F., Corringer, P.-J., and Baaden, M. (2013). Modeling complex biological systems: From solution chemistry to membranes and channels. *Pure and Applied Chemistry*, 85(1):1–13. It aims at describing common difficulties in biological systems modeling and forms the basis for part of chapter 2.

Pure Appl. Chem., Vol. 85, No. 1, pp. 1–13, 2013.
http://dx.doi.org/10.1351/PAC-CON-12-04-10

Modeling complex biological systems: From solution chemistry to membranes and channels*

Benoist Laurent[1], Samuel Murail[1], Franck Da Silva[1], Pierre-Jean Corringer[2], and Marc Baaden[1,‡]

[1]Laboratoire de Biochimie Théorique, CNRS, UPR9080, Univ. Paris Diderot, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005 Paris, France; [2]Institut Pasteur, Channel-Receptors G5 Group, CNRS URA 2182, Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France

Abstract: Complex biological systems are intimately linked to their environment, a very crowded and equally complex solution compartmentalized by fluid membranes. Modeling such systems remains challenging and requires a suitable representation of these solutions and their interfaces. Here, we focus on particle-based modeling at an atomistic level using molecular dynamics (MD) simulations. As an example, we discuss important steps in modeling the solution chemistry of an ion channel of the ligand-gated ion channel receptor family, a major target of many drugs including anesthetics and addiction treatments. The bacterial pentameric ligand-gated ion channel (pLGIC) called GLIC provides clues about the functional importance of solvation, in particular for mechanisms such as permeation and gating. We present some current challenges along with promising novel modeling approaches.

Keywords: cys-loop receptors; biomolecular chemistry; membranes; molecular dynamics; solution chemistry.

INTRODUCTION

Biological solutions are intrinsically complex and crowded mixtures. Some groups aim at modeling such systems in their full complexity, as illustrated by a snapshot of a recent simulation by McGuffee and Elcock endeavoring to understand diffusion properties of the bacterial cytoplasm (see Fig. 1A; [1]). However, such studies are currently limited to simplified approaches, such as Brownian dynamics of rigid molecules. Here, we investigate the more common case of studying a single protein in a realistic environment, using a fully flexible atomistic force field representation. For this purpose, we describe the example of a pentameric ligand-gated ion channel (pLGIC) called GLIC, illustrating how modeling may nowadays attempt to capture the complex "solutions" involved in biological processes.

pLGICs are membrane receptors widespread in the animal kingdom. The members of this family, the γ-aminobutyric acid receptor of type A ($GABA_AR$), glycine (GlyR), serotonin ($5\text{-}HT_3R$), and nicotinic acetylcholine receptors (nAChRs), play a leading role in the nervous signal transduction. Upon specific ligand binding (drawing the receptor name), an allosteric transition occurs leading to an open-

Pure Appl. Chem.* **85, 1–305 (2013). A collection of invited papers based on presentations at the 32nd International Conference on Solution Chemistry (ICSC-32), La Grande Motte, France, 28 August–2 September 2011.
‡Corresponding author

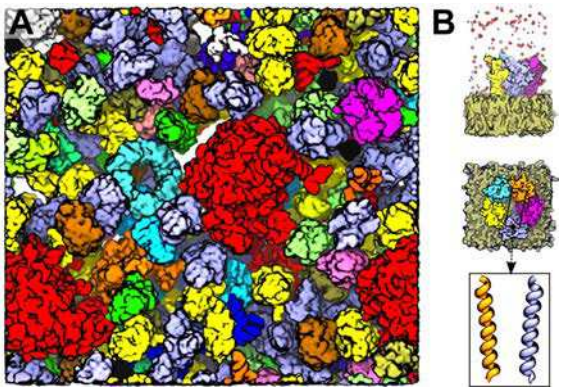

Fig. 1 (A) Snapshot of a bacterial cytoplasm model whose dynamics has been studied by Brownian dynamics (courtesy of Dr. Adrian Elcock [1]). (B) The GLIC system we study using MD simulations. The side view (top) displays the GLIC protein embedded in a phosphatidylcholine membrane in the presence of 150 mM NaCl. In the online version, the membrane is shown as an ochre surface, the protein is shown as yellow, cyan, orange, magenta, and blue surfaces, each color representing one of the five subunits. Sodium and chloride ions are displayed as blue and red spheres, while water is not shown for clarity. The top view (middle) depicts the five subunits, symmetrically arranged around the pore formed by the M2 helices (bottom panel showing two of the five helices lining the pore).

ing of the pore at the membrane level. This transition allows ions to enter the cell, producing a change in the electric potential across the membrane. The effect depends on the channel selectivity. In the case of cationic channels (nAChRs and 5-HT_3R), the entry of ions into the cell will lead to the depolarization of the cell and, if the membrane potential crosses a certain threshold (related to the number of open channels), the depolarization will propagate to initiate the action potential. On the other hand, the opening of anionic channels ($GABA_A$R and GlyR) will induce a hyperpolarization of the membrane potential and thus inhibits the action of cationic channels. Numerous compounds are known to modulate pLGIC function and channel activity, including some alcohols, steroids, cannabinoids, barbiturates, and general anesthetics. Membrane composition has also been shown to interfere with the activity of a few pLGICs. The example of nAChR is particularly striking, as cholesterol and anionic lipids are required for its activity [2].

Rational design of drugs acting on pLGICs recently came within reach owing to the crystallization of bacterial homologues, including the first open-channel structure of GLIC in 2009 [3,4]. While the role of the channel remains unclear for the bacteria, in vitro studies showed that GLIC is cation-selective and that its "gating", the process of opening and closing the channel, is regulated by pH variations [5]. The bacterial channel displayed several common characteristics with eukaryotic pLGICs, for instance, a very close 3-dimensional structure compared to a nematode glutamate-gated channel [6]. It also displayed sensitivity to alcohols and anesthetics [7,8].

pLGICs are transmembrane proteins constituted by five symmetrically organized subunits. The symmetry axis coincides with the central pore of the channel, where ion permeation takes place. The extracellular domain (ECD) of the protein contains the activating ligand-binding site and harbors a large vestibule for ions. The transmembrane domain (TMD) is structured in 4 helices per subunit called M1 to M4. The pore of the channel is lined by the M2 helices (Fig. 1B). Theoretical works suggest that the effective closure is due to a global twisting motion of the channel combined with local motions of the

transmembrane helices. In particular, the local motions comprise tilting associated with the bending of M2 [9–11].

Focusing on the case of GLIC, we discuss several challenges for the study of biological membrane protein solutions. Simplifications are necessary in constructing the model, in defining the composition of the system, and in choosing the concentration of each species. Uncertainties remain since neither experiments nor calculations can resolve issues such as reliably choosing the protonation states of each of GLIC's titratable groups. We discuss intrinsic properties of the models that remain unclear at this time such as the behavior of water in hydrophobic nanoconfinement.

THE COMPOSITION OF BIOLOGICAL "SOLUTIONS" IS COMPLEX

Biological solutions of interest contain several ingredients including proteins, nucleic acids, sugars, ions, water and other small molecules (alcohols, anesthetics, etc.). Depending on the goal a study aims to achieve, it might be necessary to include all these compounds in a model used for molecular dynamics (MD) simulations.

Cells communicate with each other thanks to membrane-embedded receptors whose hydrophobic TMDs are stabilized by the membrane environment. Hence, lipid bilayers or bilayer mimetics have to be included in a realistic model. Typical biological membranes are themselves complex ensembles composed of lipids, carbohydrates, proteins, and cholesterol in the case of eukaryotic cells [12]. The proportion of each molecular species depends on the cell type. Furthermore, different types of phospholipids are mixed in the membranes, with proportions again depending on the cell type and possible leaflet asymmetry. Matrix protein networks may further structure the membrane. To make the membrane picture more intricate, microdomains enriched in cholesterol, also called "lipid rafts", could play a role in the cellular function by compartmentalizing specific lipids and membrane proteins. However, the existence and role of such microdomains are very controversial.

Although some MD simulation studies now attempt to accurately model the complexity of, e.g., the *Escherichia coli* bacterial membranes [13], generally speaking such a complexity is very difficult to reproduce and the composition of the membrane is often simplified by using a single type of phospholipid. As an example, the GLIC simulation system is composed of the protein embedded in a fully hydrated phosphatidylcholine membrane, whereas a typical synaptic plasma membrane contains cholesterol (29 %), phosphatidylcholine (28 %), phosphatidylethanolamine (26 %), phosphatidylserine (10 %), phosphatidylinositol (2 %), sphingomyelin (2 %), and glycolipids (3 %) [14]. Unfortunately, our knowledge about lipid compositions of various organisms and cell types is still very limited, yet it is established that lipid "building blocks" have the potential to generate up to tens of thousands of different molecular species [15]. This situation may improve with recent efforts in the field of lipidomics [16].

When appropriate, ligands such as alcohols or general anesthetics are included in the model in order to study binding to the channel and modulation of its functional properties related to permeation and gating. The ligands may have indirect effects on the channel by interacting with the membrane and altering its behavior.

On the order of 150 mM NaCl or KCl ions are typically added to mimic physiologic electrolyte concentration. Sometimes higher concentrations up to 1 M are used to enhance the probability of processes such as ion permeation. These processes may be driven by cross-membrane potentials, which are typically in the range of –90 to –40 mV. To reproduce such a potential in the simulation may require special methods such as applying an electric field using additional forces on all charged particles [17]. Maintaining a charge imbalance between two solution compartments has initially been proposed by Sachs et al. [18]. This approach has subsequently been used by several groups to study the permeation of ion channels [19,20]. Herrera and Pantano proposed a variation of the method where ionic motion is restricted anisotropically to one side of the system [21]. Finally, Bostick et al. used vacuum slabs to separate the compartments of a single unit cell [22]. A set-up to study ion permeation through GLIC using

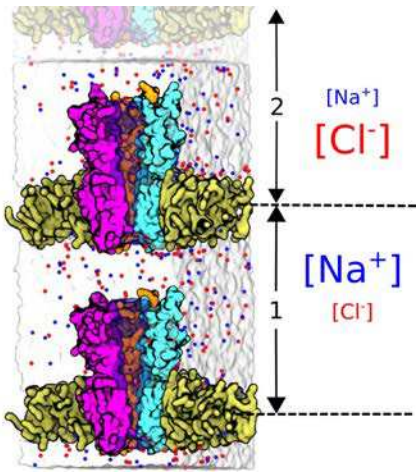

Fig. 2 System set-up used to study ion permeation through GLIC according to the double bilayer method [19]. The two compartments display different ion concentrations that create an electric potential between them, driving the ions through the channel. Only three of the five protein subunits and half of the membrane are represented in order to expose the pore interior, displayed as a blue transparent surface (in the online version).

the double bilayer approach is shown in Fig. 2. The intrinsic electrostatic field of GLIC guiding the ions through its center is illustrated in the following movie.

An animation of the electrostatic field around GLIC was generated using BioBlender (http://bioblender.eu/) followed by post-processing in our in-house UnityMol tool. A movie illustrates the flow of electric field lines through the ion channel. It is available at http://www.baaden.ibpc.fr/pub/glic/pac.mov.

CONCENTRATION

Defining molecular concentrations at the microscopic level is not straightforward. Most simulation studies aim to match experimental conditions. Yet, due to limited computational resources, simulation "boxes" are typically designed as small as possible, focusing on the immediate membrane environment of the channel of interest and minimizing the bulk solvent part. This introduces a bias in the calculation of concentrations.

In many experiments involving membrane proteins, a salt concentration of about 150 mM NaCl is required to warrant an appropriate osmotic pressure. The number of Na^+ and Cl^- ions to add to a simulation system may be calculated using eq. 1 with C_{water} = 55 M. The calculated number of ions is adjusted to neutralize the global charge of the system (i.e., protein charges), hence in this case $N_{Na} \neq N_{Cl}$. For the highly charged GLIC channel, there is an important imbalance between positive and negative ions, in our simulations typically about 90 Cl^- and 50 Na^+ are present. These numbers correspond to a concentration of about 150 mM when calculated on the basis of the Cl^- ions or 80 mM on the basis of the Na^+ ions. If the system was much larger (as in Fig. 3B), with an extensive bulk solvent part, this difference would progressively become negligible, tending toward $N_{Na} \sim N_{Cl}$.

$$N_{Na} = N_{Cl} = N_{water} \times C_{NaCl}/C_{water} \quad (1)$$

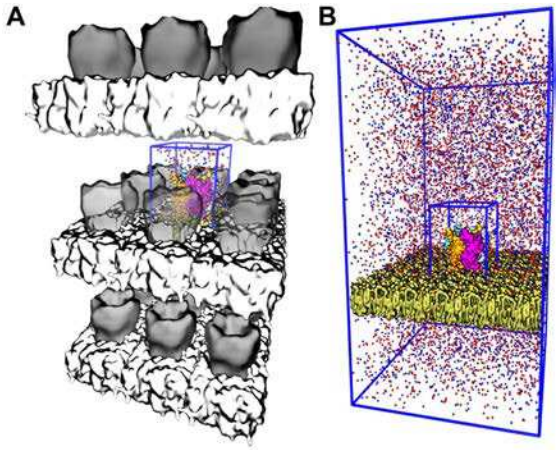

Fig. 3 (A) The 150 000 atoms atoms GLIC simulation system used for MD (colorized in the online version) surrounded by its periodic images. Since the simulation box is relatively tight, the replicates are very close to the original system, minimizing the bulk water and restricting membrane fluctuations. (B) An imaginary system in which the bulk water is extended and the membrane is large enough to allow for medium-range fluctuations, leading to a significant increase in the number of atoms beyond 4 000 000. For comparison, the periodic box of the 150 000 atoms system is also represented.

Ion concentration could in principle constitute a way to study the influence of pH by explicitly adding protons to the simulation system [23]. However, this approach is impractical in the present case because a huge simulation box would be required: at pH 4.6, a box containing 2 189 589 water molecules is necessary to observe a single H^+ ion, i.e., 50 times the actual simulation box size we use for GLIC. This echoes the fact that the number of water molecules in simulations is generally too low to mimic dilute solutions. A higher ratio of water to (protein + membrane) would be required to properly account for bulk concentrations.

Furthermore, concentration is a dynamic property. Brannigan et al. reported simulations in which they flooded both GLIC and nAChR receptors with isoflurane, a general anesthetic. During the simulation an important number of the hydrophobic isoflurane molecules partition into the lipid bilayer. Hence, at the end of the simulation, the concentration of isoflurane in the solvent is significantly lower (<10 mM) than when it started (>100 mM). This process is depicted on Fig. 4 along with data for alcohol partitioning [24].

Brannigan's study implies that a long equilibration of the system may be necessary before concentrations can be measured reliably in order to allow solute molecules to partition between aqueous and membrane phases. It may be debated whether concentrations should be calculated with respect to the water phase only, with respect to water and membrane or with respect to the entire simulation box.

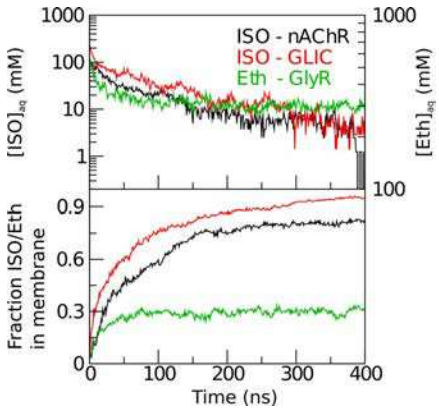

Fig. 4 Isoflurane (black and red in the online version) and ethanol (green in the online version) partition into the membrane during the equilibration of a GLIC, a nAChR and a GlyR system, respectively. The aqueous concentration of isoflurane (top) decreases for the benefit of the fraction in the membrane (bottom). The same behavior is observed for the partitioning of ethanol during the equilibration of a GlyR system (green in the online version) but to a lesser extent. Due to its more hydrophilic properties, ethanol concentration decreased to half the initial one (~300 mM) at the end of the simulation. This is in contrast to isoflurane: its concentration drops to less than 10 % of the starting one (~10 mM).

PROTONATION STATE

Knowing the protonation state of ionizable residues is a key issue to reliably model a protein. The protonation state depends on a residue's local environment. Standard pK_a values measured in bulk cannot be applied to buried protein residues, in particular for membrane proteins with an environment that largely differs from aqueous solution. GLIC is constituted of five symmetric protomers, and the location of its 81×5 ionizable residues is shown in Fig. 5. We may consider that equivalent residues in each subunit bear identical protonation states. This assumption leads to approximately $2^{81} = 10^{19}$ possible combinations of protonation states. Tang and co-workers suggest that the protonation state of some titratable groups may be different from one protomer to another leading to up to 10^{98} different combinations [25,26], a figure exceeding the number of particles in the universe!

The development of methods for calculating pK_a values of titratable groups in proteins was pioneered by Tanford and Kirkwood who proposed to represent the protein as an impenetrable sphere, which allows one to analytically solve the Poisson–Boltzmann equation (PBE) [27]. The increase in computing performances has facilitated the development of many PBE solvers, including the widely used APBS software [28,29]. Nielsen and co-workers showed that a finite difference Poisson–Boltzmann method yields better results when adding an explicit step to optimize the hydrogen-bond network [30]. Ideally, protein conformational flexibility should be taken into consideration for calculating pK_a values. Specific terms have been included in some algorithms [31] and, more recently, methods based on the λ-dynamics approach using constant pH MDs and replica exchange MD emerged [32–34]. These latter methods are currently still under development and have so far only been tested on small nonmembrane peptides or proteins. PROPKA [35,36] may be one of the most commonly used empirical approaches because it is very fast.

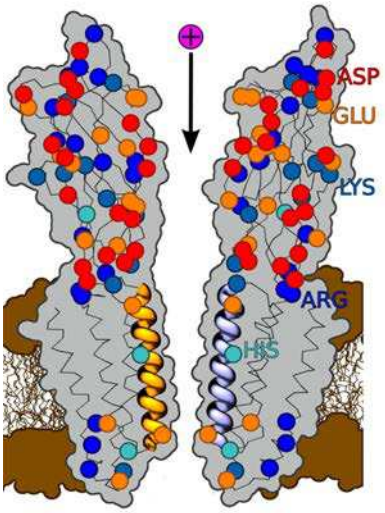

Fig. 5 Localization of ionizable residues shown in a cross-section of the GLIC ion channel (gray). M2 helices, in cartoon representation, line the pore through which cations (pink in the online version) cross the membrane (ochre in the online version).

In order to set up simulations of the GLIC system, we assessed the results of several widely used programs and web services. These pK_a predictions yielded widely varying pK_a shifts as illustrated in Fig. 6. We settled on the use of the Yasara software [37] mixing Ewald summation and hydrogen-bonding network optimization to determine if a titratable group should be protonated or not [38]. The Yasara results remain in a reasonable pK_a shift range, whereas some of the other methods suggest huge shifts (Fig. 6). We applied a consensus approach, only protonating residues that were simultaneously found to change ionization state in all five subunits. Many efforts in improving the crystallization protocol for GLIC recently lead to a higher-resolution structure in which ion binding can be predicted between residue D86 and D88. This is a strong indication that these residues should not be protonated (unpublished data). These findings allowed us to iteratively improve our protonation state estimate for GLIC.

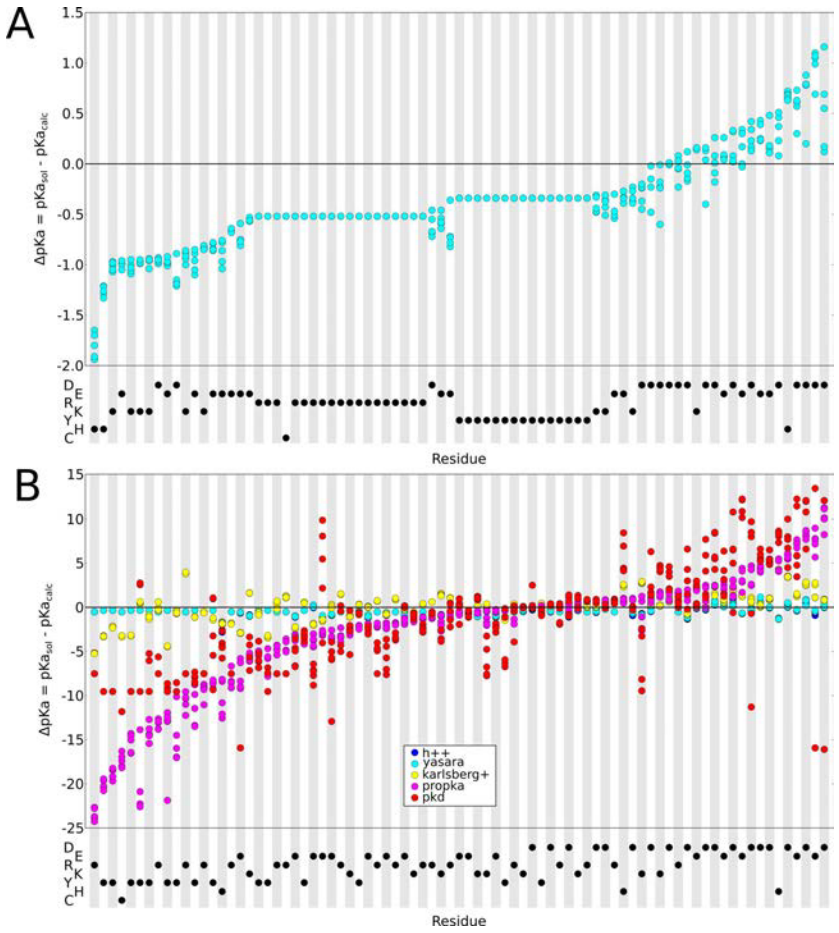

Fig. 6 pK_a shift predictions with respect to standard values for all ionizable residues in GLIC obtained using different software packages. Residues are ordered according to $\Delta pKa = (pKa_{solution} - pKa_{calc})$ with respect to the Yasara software (A) or the PROPKA software (B), respectively. The bottom part of each panel indicates the amino acid type of each residue.

SOLVATION IN SPECIAL/UNUSUAL ENVIRONMENTS

The complex shapes of proteins may feature channels and cavities providing special, potentially solvated nano-environments. Water in such a hydrophobic nanoconfinement may be particularly unstable, a phenomenon known as capillary evaporation. Several groups have observed and characterized dewetting transitions in MD simulations, for example, in the context of nanopores [39,40] or in the bacterial

mechanosensitive channels MscL and MscS [41,42]. Roth and co-workers suggest that capillary evaporation could constitute an intrinsic property of some channels [43] and may be a widespread biological mechanism.

In the case of GLIC, extensive sampling led us to observe an unexpected pore dewetting behavior (see Fig. 7A), as did several other groups [26,44]. Yet we cannot currently conclude whether GLIC belongs to a family of "bubble gated ion channels", since ongoing studies in our lab suggest that subtle changes in the simulation parameters may prevent dewetting from occurring (see Fig. 7B). Another very recent study is more affirmative [45]. It should be noted that forcefield parameters generally have not been tuned to reproduce the behavior of water in such special environments, which is in part due to the lack of experimental data.

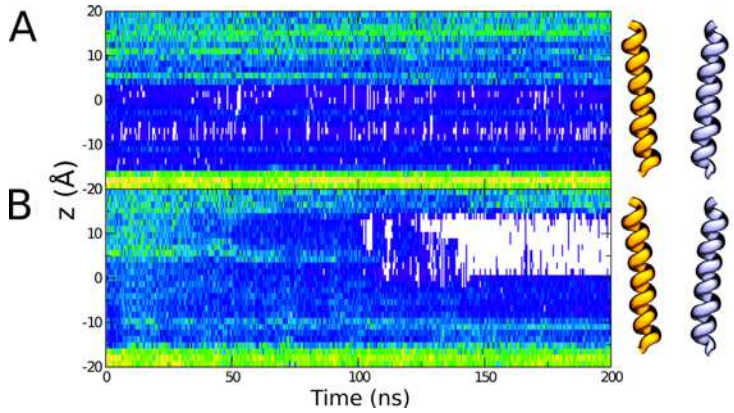

Fig. 7 Hydration traces of GLIC's pore during two representative simulations. Minor changes in the simulation parameters can make a noticeable difference between a fully hydrated channel (A) and a channel that dehydrates spontaneously in the upper part of the M2 helix-lined pore (B). For both simulations, the protonation states were identical [3], the only differences were the forcefield and the MD software used (amber99 and Gromacs for simulation A, vs. Charmm22 and NAMD2 for simulation B).

SAMPLING, STATISTICS, TIMESCALE

A fundamental question before starting any computational study is how to best spend the limited amount of available computing time. Strategies may vary in between two extremes: (1) running many short simulations from several starting points or (2) running an extended one-shot simulation. Shaw et al. recently showed that the result of the second approach matches experimental data very well, when the MD simulations are long enough [46].

In 2010, we studied GLIC gating in a 1 µs MD simulation suggesting a "domino" gating mechanism in which subunits sequentially switch from an open to a closed conformation [9]. Despite the large amount of computational resources (approx. 10 months of calculations on a supercomputer in 2009, i.e., tens of years on a recent desktop machine), only two protomers had fully undergone this transition to a closed state at the end of the simulation, suggesting that a much longer simulation was required to achieve a complete gating transition in all five protomers. Longer simulations are also needed to characterize processes such as ion permeation. Since GLIC has a low conductivity of 8 pS, one should observe an estimated passage of only 3 ions per microsecond at –65 mV. A computationally cheaper

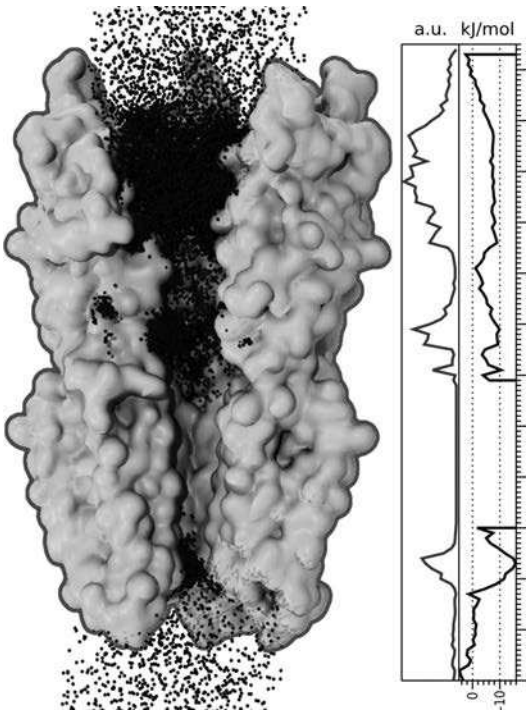

Fig. 8 Sodium ion occupancy and related relative Boltzmann energy accumulated during a 1-µs MD simulation.

alternative is to map the affinity of ions for a certain position along the channel pore. We have determined such a graph from the 1 µs gating simulation for a nonconductive state with a central barrier (Fig. 8). It is usually admitted that a simulation should be run at least 10 times longer than the slowest timescale of interest [47]. This is often impossible since many relevant biomolecular timescales exceed 1 µs. Typically, the neuromuscular acetylcholine receptor's gating is expected to be in the range of 1–10 µs [48] which implies MD simulations from 10 to 100 µs.

Running several short comparative simulations may be more appropriate for ligand binding studies, for example, involving drugs. An advantage of such short simulations is to remain close to a well-defined state, e.g., a crystal structure, rather than moving away from the experimentally backed conformation to some transient intermediate state. Furthermore, one may reduce the number of unproductive runs where the drug may diffuse out of the binding pocket into the solvent. Many short simulations with slightly different ligand starting conformations improve statistics and sampling. We employ such an approach to study two general anesthetics, propofol and desflurane, that have recently been co-crystallized with GLIC [49]. This study revealed a binding site in the upper part of the TMD of the protein. Other binding sites for general anesthetics and alcohols, including transmembrane, extracellular, and pore sites, have been suggested [8,25,50,51]. Channel blocking by charged quaternary ammonium compounds, divalent ions, and lidocaine has been shown using electrophysiology and X-ray

crystallography [52], also suggesting binding sites located in the pore of the channel. These observations pose the problem of sampling from a combinatorial point of view: multiplying the number of sites by the number of ligands, then by the number of mutants one wishes to test quickly leads to an intractable required total simulation time.

Lipids are crucial for the structure and function of membrane proteins. Bilayers with complex compositions pose a particular sampling challenge [53]. A misplaced lipid in a simulation set-up might have consequences on the whole trajectory, in particular if it were to play a specific biological role. With a diffusion coefficient of the order of 10^{-8} cm^2/s, a lipid embedded in a membrane is expected to have a mean-square displacement of 4 nm^2 for a 1-µs-long simulation. In our GLIC simulations, convergence for this value sets on beyond 100 ns and fully stabilizes at about 500 ns. Hence, the timescale of most current studies does not allow for an extensive reorganization of lipids around membrane proteins. Parton et al. recently addressed this problem while simulating a whole vesicle, demonstrating the importance of lipid diffusion for protein aggregation [54]. The authors, however, highlight that the coarse-grained models are highly simplified and inevitably approximate the nature of the protein–lipid and protein–protein interactions. De Meyer et al. previously suggested the role of cholesterol in protein clustering using dissipative particle dynamics Monte Carlo and a more simplified model [55].

At last, the problem of simulation convergence should be raised briefly. Methods for the quantification of sampling have been proposed for several decades, yet none has been widely adopted. In 2000, Berk Hess proposed a method based on principal component analysis [56] that has been used by other groups to evaluate the convergence of a set of MD simulations [57,58]. Faraldo-Gómez and coworkers focus on convergence of membrane protein simulations, and although the timescale is relatively short by today's standards, their main findings are likely still valid. Their work concludes that structured TMDs converge relatively fast, even on a 10 ns timescale, but more mobile parts are undersampled. Grossfield et al. calculated 26 independent 100 ns MD runs of rhodopsin and found similar results [58]. To date, despite new method proposals [59–61], sampling quality is often tentatively assessed based on several simple criteria. A single descriptor may be monitored along a simulation until it reaches a stable value. The root mean square deviation (RMSD), which is a descriptor for molecular deformation, is a common but controversial criterion. A variation consists in stopping a simulation after a descriptor reaches an experimental reference value and remains close to it for a certain time. Another approach is to use several independent MD simulations with different starting points. When these simulations converge to a similar state, sampling is considered sufficient.

CONCLUSION

Accurately simulating complex biological systems such as the GLIC ion channel in a realistic environment remains a puzzle with some missing pieces. Choosing the composition of the model requires a compromise between biological accuracy and technical constraints imposed by the limited size of simulation boxes. Even simple quantities such as concentrations are not easy to map from the macroscopic world to the microscopic representation. For some systems, protonation state assessment is a particularly tricky and largely unsolved problem. The behavior of water in special environments such as nanocavities and pores is another open issue. Underlying these specific points is the general question of sampling and timescale. Furthermore, the chosen GLIC example is a rather simple membrane system. Other more challenging ones may include double bilayers [62], entire vesicles [63,64], fully decorated virion particles [65–68], or huge bilayer patches, each adding a set of specific questions on its own.

ACKNOWLEDGMENTS

The authors would like to thank Jérôme Hénin and Grace Brannigan for providing the original data of Fig. 4. We also thank Marie Prévost for her expertness on experimental measurements, Tyler Reddy and Daniel Parton for their insight into recent lipidomics studies. We thank the French Agency for Research

for funding (Grant ANR-2010-BLAN-1534). Molecular simulations were performed using HPC computing resources from GENCI-IDRIS and GENCI-CINES (Grant 2011072292 to Marc Delarue). B.L. would like to thank his funding organisms: the Servier Labs and the "Ecole Doctorale Interdisciplinaire Européenne Frontières du vivant ED474" doctoral program Liliane Bettencourt.

REFERENCES

1. S. R. McGuffee, A. H. Elcock. *PLoS Comput. Biol.* **6**, e1000694 (2010).
2. A. W. Dalziel, E. S. Rollins, M. G. McNamee. *FEBS Lett.* **122**, 193 (1980).
3. N. Bocquet, H. Nury, M. Baaden, C. Le Poupon, J. P. Changeux, M. Delarue, P. J. Corringer. *Nature* **457**, 111 (2009).
4. R. J. Hilf, R. Dutzler. *Nature* **457**, 115 (2009).
5. N. Bocquet, L. Prado de Carvalho, J. Cartaud, J. Neyton, C. Le Poupon, A. Taly, T. Grutter, J. P. Changeux, P. J. Corringer. *Nature* **445**, 116 (2007).
6. R. E. Hibbs, E. Gouaux. *Nature* **474**, 54 (2011).
7. Y. Weng, L. Yang, P. J. Corringer, J. M. Sonner. *Anesth. Analg.* **110**, 59 (2010).
8. R. J. Howard, S. Murail, K. E. Ondricek, P. J. Corringer, E. Lindahl, J. R. Trudell, R. A. Harris. *Proc. Natl. Acad. Sci. USA* **108**, 12149 (2011).
9. H. Nury, F. Poitevin, C. Van Renterghem, J. P. Changeux, P. J. Corringer, M. Delarue, M. Baaden. *Proc. Natl. Acad. Sci. USA* **107**, 6275 (2010).
10. F. Zhu, G. Hummer. *Biophys. J.* **97**, 2456 (2009).
11. F. Zhu, G. Hummer. *Proc. Natl. Acad. Sci. USA* **107**, 19814 (2010).
12. G. van Meer, D. R. Voelker, G. W. Feigenson. *Nat. Rev. Mol. Cell Biol.* **9**, 112 (2008).
13. T. J. Piggot, D. A. Holdbrook, S. Khalid. *J. Phys. Chem. B* **115**, 13381 (2011).
14. W. G. Wood, M. Cornwell, L. S. Williamson. *J. Lipid Res.* **30**, 775 (1989).
15. A. Shevchenko, K. Simons. *Nat. Rev. Mol. Cell Biol.* **11**, 593 (2010).
16. H. Alex Brown. *Curr. Opin. Chem. Biol.* **16**, 221 (2012).
17. J. Gumbart, F. Khalili-Araghi, M. Sotomayor, B. Roux. *Biochim. Biophys. Acta* **1818**, 294 (2012).
18. J. N. Sachs, P. S. Crozier, T. B. Woolf. *J. Chem. Phys.* **121**, 10847 (2004).
19. C. Kutzner, H. Grubmuller, B. L. de Groot, U. Zachariae. *Biophys. J.* **101**, 809 (2011).
20. S. J. Lee, Y. Song, N. A. Baker. *Biophys. J.* **94**, 3565 (2008).
21. F. E. Herrera, S. Pantano. *J. Chem. Phys.* **130**, 195105 (2009).
22. D. Bostick, M. L. Berkowitz. *Biophys. J.* **85**, 97 (2003).
23. M. Baaden, M. Burgard, G. Wipff. *J. Phys. Chem. B* **105**, 11131 (2001).
24. S. Murail, B. Wallner, J. R. Trudell, E. Bertaccini, E. Lindahl. *Biophys. J.* **100**, 1642 (2011).
25. M. H. Cheng, R. D. Coalson, P. Tang. *J. Am. Chem. Soc.* **132**, 16442 (2010).
26. D. Willenbring, L. T. Liu, D. Mowrey, Y. Xu, P. Tang. *Biophys. J.* **101**, 1905 (2011).
27. C. Tanford, J. G. Kirkwood. *J. Am. Chem. Soc.* **79**, 5333 (1957).
28. F. Fogolari, A. Brigo, H. Molinari. *J. Mol. Recognit.* **15**, 377 (2002).
29. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon. *Proc. Natl. Acad. Sci. USA* **98**, 10037 (2001).
30. J. E. Nielsen, K. V. Andersen, B. Honig, R. W. Hooft, G. Klebe, G. Vriend, R. C. Wade. *Protein Eng.* **12**, 657 (1999).
31. E. G. Alexov, M. R. Gunner. *Biophys. J.* **72**, 2075 (1997).
32. S. Donnini, F. Tegeler, G. Groenhof, H. Grubmuller. *J. Chem. Theory Comput.* **7**, 1962 (2011).
33. S. L. Williams, P. G. Blachly, J. A. McCammon. *Proteins* **79**, 3381 (2011).
34. Y. Meng, A. E. Roitberg. *J. Chem. Theory Comput.* **6**, 1401 (2010).
35. H. Li, A. D. Robertson, J. H. Jensen. *Proteins* **61**, 704 (2005).
36. D. C. Bas, D. M. Rogers, J. H. Jensen. *Proteins* **73**, 765 (2008).
37. E. Krieger, G. Koraimann, G. Vriend. *Proteins* **47**, 393 (2002).

38. E. Krieger, J. E. Nielsen, C. A. Spronk, G. Vriend. *J. Mol. Graphics Model.* **25**, 481 (2006).
39. O. Beckstein, P. C. Biggin, M. S. P. Sansom. *J. Phys. Chem. B* **105**, 12902 (2001).
40. O. Beckstein, M. S. Sansom. *Proc. Natl. Acad. Sci. USA* **100**, 7063 (2003).
41. A. Anishkin, S. Sukharev. *Biophys. J.* **86**, 2883 (2004).
42. A. Anishkin, B. Akitake, K. Kamaraju, C. S. Chiang, S. Sukharev. *J. Phys.: Condens. Matter* **22**, 454120 (2010).
43. R. Roth, D. Gillespie, W. Nonner, R. E. Eisenberg. *Biophys. J.* **94**, 4282 (2008).
44. D. N. Lebard, J. Henin, R. G. Eckenhoff, M. L. Klein, G. Brannigan. *PLoS Comput. Biol.* **8**, e1002532 (2012).
45. F. Zhu, G. Hummer. *Biophys. J.* **103**, 219 (2012).
46. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers. *Science* **330**, 341 (2010).
47. D. M. Zuckerman. *Annu. Rev. Biophys.* **40**, 41 (2011).
48. S. Chakrapani, A. Auerbach. *Proc. Natl. Acad. Sci. USA* **102**, 87 (2005).
49. H. Nury, C. Van Renterghem, Y. Weng, A. Tran, M. Baaden, V. Dufresne, J. P. Changeux, J. M. Sonner, M. Delarue, P. J. Corringer. *Nature* **469**, 428 (2011).
50. Q. Chen, M. H. Cheng, Y. Xu, P. Tang. *Biophys. J.* **99**, 1801 (2010).
51. G. Brannigan, D. N. LeBard, J. Henin, R. G. Eckenhoff, M. L. Klein. *Proc. Natl. Acad. Sci. USA* **107**, 14122 (2010).
52. R. J. Hilf, C. Bertozzi, I. Zimmermann, A. Reiter, D. Trauner, R. Dutzler. *Nat. Struct. Mol. Biol.* **17**, 1330 (2010).
53. T. A. Soares, T. P. Straatsma. *Mol. Simul.* **34**, 295 (2008).
54. D. L. Parton, J. W. Klingelhoefer, M. S. Sansom. *Biophys. J.* **101**, 691 (2011).
55. F. J. de Meyer, J. M. Rodgers, T. F. Willems, B. Smit. *Biophys. J.* **99**, 3629 (2010).
56. B. Hess. *Phys. Rev. E* **62**, 8438 (2000).
57. J. D. Faraldo-Gomez, L. R. Forrest, M. Baaden, P. J. Bond, C. Domene, G. Patargias, J. Cuthbertson, M. S. Sansom. *Proteins* **57**, 783 (2004).
58. A. Grossfield, S. E. Feller, M. C. Pitman. *Proteins* **67**, 31 (2007).
59. A. Grossfield, D. M. Zuckerman. *Annu. Rep. Comput. Chem.* **5**, 23 (2009).
60. X. Zhang, D. Bhatt, D. M. Zuckerman. *J. Chem. Theory Comput.* **6**, 3048 (2010).
61. F. Zhu, G. Hummer. *J. Comput. Chem.* **33**, 453 (2012).
62. L. Darré, A. Tek, M. Baaden, S. Pantano. *J. Chem. Theory Comput.* **8**, 3880 (2012).
63. P. M. Kasson, E. Lindahl, V. S. Pande. *PLoS Comput. Biol.* **6**, e1000829 (2010).
64. H. J. Risselada, H. Grubmuller. *Curr. Opin. Struct. Biol.* **22**, 187 (2012).
65. A. Arkhipov, W. H. Roos, G. J. Wuite, K. Schulten. *Biophys. J.* **97**, 2061 (2009).
66. P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, K. Schulten. *Structure* **14**, 437 (2006).
67. E. R. May, C. L. Brooks III. *J. Phys. Chem. B* **116**, 8604 (2012).
68. W. H. Roos, I. Gertsman, E. R. May, C. L. Brooks III, J. E. Johnson, G. J. Wuite. *Proc. Natl. Acad. Sci. USA* **109**, 2342 (2012).

Appendices to chapter 3

Below are supplementary materials that complement chapter "High-Performance Computing And Large Scale Data Analysis".

C.1 A script to generate 25 slightly different simulations

This script aims to prepare several simulations for submission to run on a supercomputer. Prior to running the script, the start configuration should have been calculated and placed in a directory which path is defined in the variable `ROOTPATH` at the top of the script. The script handles the simulation preparation by making a directory for each simulation with the appropriate start conformation for each trajectory. It also takes care of the simulation input files, making all the names matching the actual filenames, which are different in each simulation. Additionally, the submission scripts are generated according to the variables defined at the top of the script that depend on the supercomputer I wanted to use.

Although these tasks might seem basic, it saved me an enormous amount of time in the preparation and submission of these jobs and similar scripts were used at different steps of the project.

newjobs.py **Page 1**

```
import os
import random
import shutil
import sys

ROOTPATH = "../01_prep/05_make_poses"
PARAMDIR = "../00_input"

USE_JADE = True
USE_TURING = not USE_JADE

# Option for runs on turing@idris.fr
NCPUS           = 512
BGSIZE          = 64
RANKS_PER_NODE  = 8

# Options for runs on jade@cines.fr
NNODES  = 23        # number of nodes to use
NCPUS   =  8        # number of CPUs to used per core
NMPI    =  8        # number of MPI processes per core
# So the total number of cores that will be used is NNODES * NCPUS


if USE_TURING:
    assert NCPUS == (BGSIZE * RANKS_PER_NODE)


def m4(s, **kwargs):
    for k, v in kwargs.iteritems():
        s = s.replace(k, v)
    return s


def make_job(prefix):
    inputdir = os.path.join(ROOTPATH, prefix)
    workdir = prefix

    # basenames
    coorbase    = prefix + "_mini_out.restart.coor"
    xscbase     = prefix + "_mini_out.restart.xsc"
    pdbbase     = prefix + ".pdb"
    psfbase     = "nachr_mbr46i.psf"
    ffparbase   = "par_all27_prot_lipid_na.inp"
    ligparbase  = "MBR.par"

    # full paths
    coor    = os.path.join(inputdir, coorbase)
    xsc     = os.path.join(inputdir, xscbase)
    pdb     = os.path.join(ROOTPATH, prefix, pdbbase)
    psf     = os.path.join(ROOTPATH, prefix, psfbase)
    ffpar   = os.path.join(PARAMDIR, ffparbase)
    ligpar  = os.path.join(PARAMDIR, ligparbase)

    # simulation output file prefix
    output_dir    = "output"
    output_prefix = os.path.join(output_dir, "mbr46.prod")

    # create the work directory
    os.mkdir(workdir)

    # create the output directory
    os.mkdir("output")

    # copy the run script in the work directory
    shutil.copy("run.bash", workdir)

    # copy the xsc and coor files from the minimization
    shutil.copy(xsc, workdir)
    shutil.copy(coor, workdir)

    # copy the pdb and psf files to the output directory
    shutil.copy(pdb, "output")
```

newjobs.py **Page 2**

```
    shutil.copy(psf, "output")

    # copy the parameter files in the work directory
    shutil.copy(ligpar, workdir)
    shutil.copy(ffpar, workdir)

    if USE_JADE:
        # manage the namd run files
        with open("glic.jade.namd.input", "rt") as f:
            content = f.read()
        content = m4(content,
                     __COOR__=coorbase,
                     __XSC__=xscbase,
                     __PDB__=os.path.join(output_dir, pdbbase),
                     __PSF__=os.path.join(output_dir, psfbase),
                     __FF_PAR__=ffparbase,
                     __LIG_PAR__=ligparbase,
                     __OUTPUT_PREFIX__=output_prefix,
                     __SEED__=str(random.randint(1, 5000)),
                     )
        with open(os.path.join(workdir, "glic0.namd"), "wt") as f:
            f.write(content)

        # manage the job submission script
        with open("namdjob.pbs.input", "rt") as f:
            content = f.read()
        content = m4(content,
                     __WALLTIME__="24:00:00",
                     __NNODES__=str(NNODES),
                     __NCPUS__=str(NCPUS),
                     __NMPI__=str(NMPI),
                     __JOBNAME__="prod_" + prefix,
                     __NAMD_SCRIPT__="glic0.namd",
                     )
        with open(os.path.join(workdir, "namdjob.pbs"), "wt") as f:
            f.write(content)

    else:
        # manage the namd run files
        with open("glic0.namd.input", "rt") as f:
            content = f.read()
        content = m4(content,
                     __COOR__=coorbase,
                     __XSC__=xscbase,
                     __PDB__=os.path.join(output_dir, pdbbase),
                     __PSF__=os.path.join(output_dir, psfbase),
                     __FF_PAR__=ffparbase,
                     __LIG_PAR__=ligparbase,
                     __OUTPUT_PREFIX__=output_prefix,
                     __SEED__=str(random.randint(1, 5000)),
                     )
        with open(os.path.join(workdir, "glic0.namd"), "wt") as f:
            f.write(content)

        with open("glic1.namd.input", "rt") as f:
            content = f.read()
        content = m4(content,
                     __PDB__=os.path.join(output_dir, pdbbase),
                     __PSF__=os.path.join(output_dir, psfbase),
                     __FF_PAR__=ffparbase,
                     __LIG_PAR__=ligparbase,
                     __OUTPUT_PREFIX__=output_prefix,
                     )
        with open(os.path.join(workdir, "glic1.namd"), "wt") as f:
            f.write(content)

        with open("glic2.namd.input", "rt") as f:
            content = f.read()
        content = m4(content,
                     __PDB__=os.path.join(output_dir, pdbbase),
                     __PSF__=os.path.join(output_dir, psfbase),
                     __FF_PAR__=ffparbase,
                     __LIG_PAR__=ligparbase,
```

newjobs.py **Page 3**

```
                    __OUTPUT_PREFIX__=output_prefix,
                    )
        with open(os.path.join(workdir, "glic2.namd"), "wt") as f:
            f.write(content)

        # manage the job submission script
        with open("namdjob.ll.input", "rt") as f:
            content = f.read()
        content = m4(content,
                    __WALLTIME__="20:00:00",
                    __BGSIZE__=str(BGSIZE),
                    __NCPUS__=str(NCPUS),
                    __RANKS_PER_NODE__=str(RANKS_PER_NODE),
                    __JOBNAME__="prod_" + prefix,
                    __NAMD_SCRIPT0__="glic0.namd",
                    __NAMD_SCRIPT1__="glic1.namd",
                    __NAMD_SCRIPT2__="glic2.namd",
                    )
        with open(os.path.join(workdir, "namdjob.ll"), "wt") as f:
            f.write(content)

    # move the output directory to the work directory
    os.rename("output", os.path.join(workdir, "output"))


if __name__ == '__main__':

    com = " ".join(arg for arg in sys.argv)

    if not com == "{0} go".format(sys.argv[0]):
        usage = "usage: python {0} go".format(sys.argv[0])
        usage += "\nModify the source code to control the script's behavior."
        print usage
        exit()

    first, last = 12, 24

    prefix_list = ["model_{0:03d}".format(i) for i in xrange(first, last + 1)]
    for prefix in prefix_list:
        print "preparing {0}...".format(prefix),
        sys.stdout.flush()
        make_job(prefix)
        print "done"
        sys.stdout.flush()
```

Appendices to chapter 5

D

Below are supplementary materials that complement chapter "Propofol & desflurane simulations provide new insights into anesthetic action at the atomic scale".

D.1 Contact maps

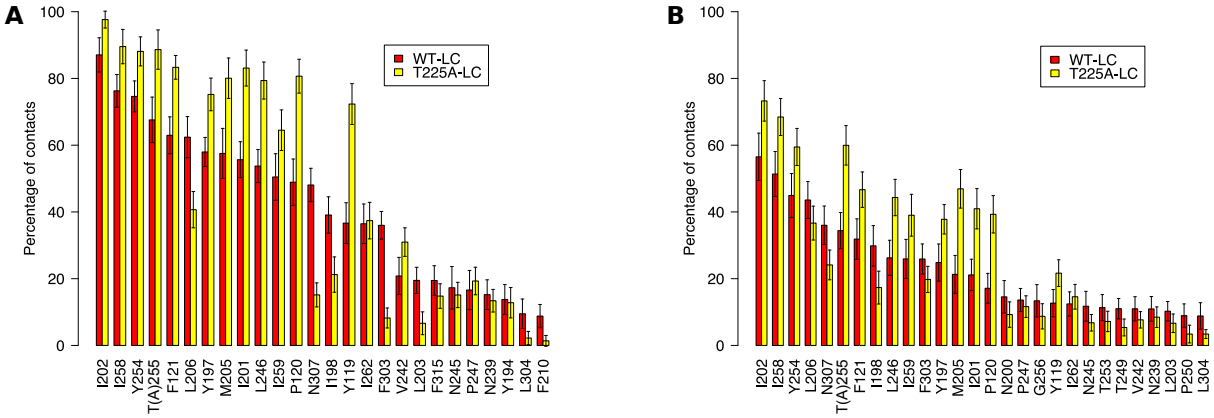

Figure D.1 – Contact maps of anesthetics bound to locally closed GLIC variants. Most contacted residues by anesthetics propofol (A) and desflurane (B) along short MD simulations.

D.2 Pocket exploration by bromoform

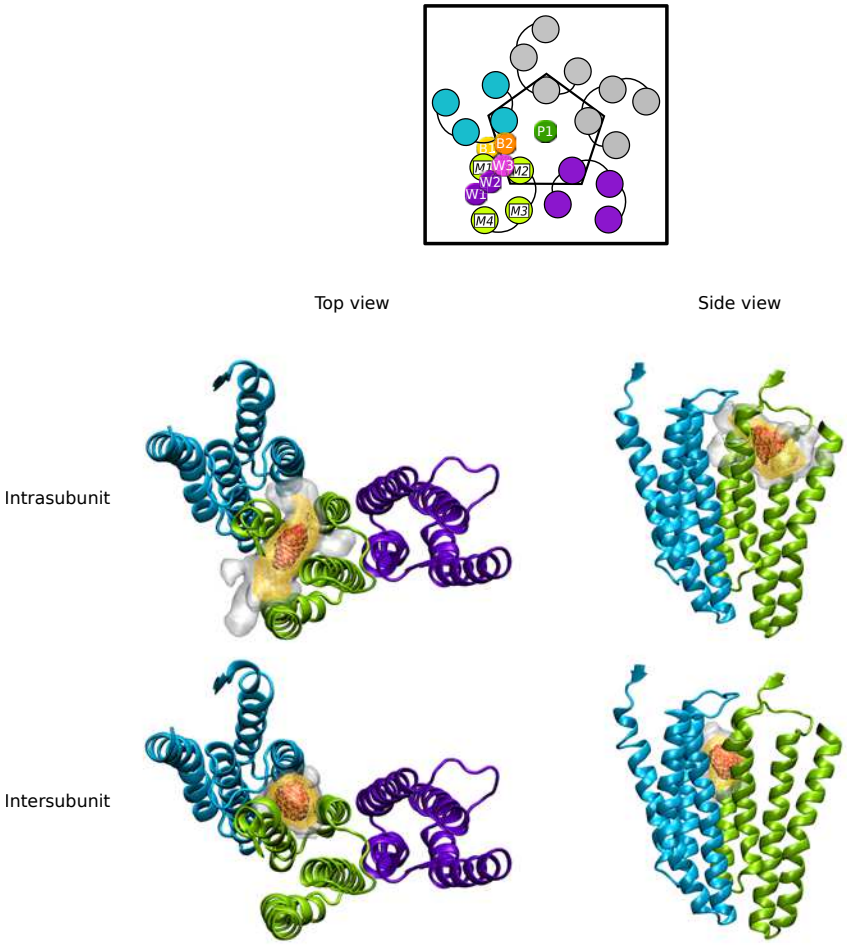

Figure D.2 – Pocket exploration by bromoform in short MD simulations. Top (left) and side (right) views of bromoform exploration of binding intra- (top) and intersubunit (bottom) binding pockets in open F238A GLIC. The TMD of 3 subunits (blue and green and purple) are represented here for clarity. 3 occupancies corresponding to as many occupancy cutoffs are represented here: 0.1, 0.01 and 0.001 in red opaque surface, yellow mesh and white transparent surface, respectively.

D.3 Pocket volumes in crystal structures of GLIC

Sequence	Form	Ligand	PDB id	Resolution (Å)	Volume (Å^3)
WT	Open	None	4HFI	2.40	273.95 ± 12.82
WT	Open	None	3EAM	2.90	268.73 ± 29.94
WT	Open	Desflurane	3P4W	3.20	324.73 ± 25.56
WT	Open	Propofol	3P50	3.30	348.55 ± 22.32
H235F	LC I	None	3TLT	3.30	328.70 ± 89.24
E243P	LC II	None	3TLS	3.20	556.60 ± 39.94
K33C-L246C	LC III	None	3TLV	2.90	586.68 ± 66.32
WT	Open	Bromoform	4HFH	2.75	308.92 ± 28.55
F238A	Open	None	4HFB	2.65	294.65 ± 16.22
F238A	Open	Bromoform	4HFD	3.10	343.23 ± 37.53
F238A	Open	Ethanol	4HFE	2.80	273.25 ± 15.75

Table D.1 – Volume of the intrasubunit pocket W1 in GLIC crystal structures. Average volume calculated on the five subunit with the Epock software. LC I, II, & III are locally closed subtype forms as depicted in Prevost *et al.* (2012).

Sequence	Form	Ligand	PDB id	Resolution (Å)	Volume (Å^3)
WT	Open	None	4HFI	2.40	42.22 ± 5.31
WT	Open	None	3EAM	2.90	59.32 ± 8.19
WT	Open	Desflurane	3P4W	3.20	53.57 ± 10.00
WT	Open	Propofol	3P50	3.30	63.90 ± 5.22
H235F	LC I	None	3TLT	3.30	12.15 ± 17.86
E243P	LC II	None	3TLS	3.20	11.43 ± 10.38
K33C-L246C	LC III	None	3TLV	2.90	5.35 ± 6.74
WT	Open	Bromoform	4HFH	2.75	52.10 ± 4.97
F238A	Open	None	4HFB	2.65	190.17 ± 12.97
F238A	Open	Bromoform	4HFD	3.10	193.80 ± 12.32
F238A	Open	Ethanol	4HFE	2.80	187.90 ± 11.58

Table D.2 – Volume of the intersubunit pocket W1 in GLIC crystal structures. See also table D.1.

Bibliography

Alex Brown, H. (2012). Lipidomics: when apocrypha becomes canonical. *Current opinion in chemical biology*, 16(1-2):221–226.

Alexov, E. G. and Gunner, M. R. (1997). Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophysical journal*, 72(5):2075–2093.

Allen, N. J. and Barres, B. A. (2009). Neuroscience: Glia - more than just brain glue. *Nature*, 457(7230):675–677.

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72(4):2384.

Anishkin, A., Akitake, B., Kamaraju, K., Chiang, C.-S., and Sukharev, S. (2010). Hydration properties of mechanosensitive channel pores define the energetics of gating. *Journal of physics. Condensed matter : an Institute of Physics journal*, 22(45):454120.

Anishkin, A. and Sukharev, S. (2004). Water dynamics and dewetting transitions in the small mechanosensitive channel MscS. *Biophysical journal*, 86(5):2883–2895.

Baaden, M., Burgard, M., and Wipff, G. (2001). TBP at the Water-Oil Interface: The Effect of TBP Concentration and Water Acidity Investigated by Molecular Dynamics Simulations. *The journal of physical chemistry. B*, 105(45):11131–11141.

Baenziger, J. E., Ryan, S. E., Goodreid, M. M., Vuong, N. Q., Sturgeon, R. M., and Dacosta, C. J. B. (2008). Lipid composition alters drug action at the nicotinic acetylcholine receptor. *Molecular pharmacology*, 73(3):880–890.

Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding & design*, 2(3):173–181.

Bahri, M. A., Seret, A., Hans, P., Piette, J., Deby-Dupont, G., and Hoebeke, M. (2007). Does propofol alter membrane fluidity at clinically relevant concentrations? An ESR spin label study. *Biophysical chemistry*, 129(1):82–91.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18):10037–10041.

Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. *Proteins*, 73(3):765–783.

Beckstein, O., Biggin, P. C., and Sansom, M. S. P. (2001). A Hydrophobic Gating Mechanism for Nanopores. *The journal of physical chemistry. B*, 105(51):12902–12905.

Beckstein, O. and Sansom, M. S. P. (2003). Liquid-vapor oscillations of water in hydrophobic nanopores. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):7063–7068.

Bennett, C. H. (1976). Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, 22(2):245–268.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

Bocquet, N., Nury, H., Baaden, M., Le Poupon, C., Changeux, J.-P., Delarue, M., and Corringer, P.-J. (2009). X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature*, 457(7225):111–114.

Bocquet, N., Prado de Carvalho, L., Cartaud, J., Neyton, J., Le Poupon, C., Taly, A., Grutter, T., Changeux, J.-P., and Corringer, P.-J. (2007). A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature*, 445(7123):116–119.

Bostick, D. and Berkowitz, M. L. (2003). The implementation of slab geometry for membrane-channel molecular dynamics simulations. *Biophysical journal*, 85(1):97–107.

Bradley, R. J., Sterz, R., and Peper, K. (1984). The effects of alcohols and diols at the nicotinic acetylcholine receptor of the neuromuscular junction. *Brain Research*, 295(1):101–112.

Brannigan, G., Hénin, J., Law, R., Eckenhoff, R., and Klein, M. L. (2008). Embedded cholesterol in the nicotinic acetylcholine receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38):14418–14423.

Brannigan, G., LeBard, D. N., Hénin, J., Eckenhoff, R. G., and Klein, M. L. (2010). Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(32):14122–14127.

Brömstrup, T., Howard, R. J., Trudell, J. R., Harris, R. A., and Lindahl, E. (2013). Inhibition versus Potentiation of Ligand-Gated Ion Channels Can Be Altered by a Single Mutation that Moves Ligands between Intra- and Intersubunit Sites. *Structure (London, England : 1993)*, 21:1307–1316.

Brünger, A. T., Adams, P. D., and Rice, L. M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. *Current opinion in structural biology*, 8(5):606–611.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101.

Calimet, N., Simoes, M., Changeux, J.-P., Karplus, M., Taly, A., and Cecchini, M. (2013). A gating mechanism of pentameric ligand-gated ion channels. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):E3987–96.

Cao, Z., Liu, L., and Wang, J. (2011). Why the OPLS-AA force field cannot produce the β-hairpin structure of H1 peptide in solution when comparing with the GROMOS 43A1 force field? *Journal of biomolecular structure & dynamics*, 29(3):527–539.

Chakrapani, S. and Auerbach, A. (2005). A speed limit for conformational change of an allosteric membrane protein. *Proceedings of the National Academy of Sciences of the United States of America*, 102(1):87–92.

Chang, C.-s. S., Olcese, R., and Olsen, R. W. (2003). A single M1 residue in the beta2 subunit alters channel gating of GABAA receptor in anesthetic modulation and direct activation. *The Journal of biological chemistry*, 278(44):42821–42828.

Chebaro, Y., Pasquali, S., and Derreumaux, P. (2012). The coarse-grained OPEP force field for non-amyloid and amyloid proteins. *The journal of physical chemistry. B*, 116(30):8741–8752.

Chen, Q., Cheng, M. H., Xu, Y., and Tang, P. (2010). Anesthetic binding in a pentameric ligand-gated ion channel: GLIC. *Biophysical journal*, 99(6):1801–1809.

Cheng, M. H., Coalson, R. D., and Tang, P. (2010). Molecular dynamics and brownian dynamics investigation of ion permeation and anesthetic halothane effects on a proton-gated ion channel. *Journal of the American Chemical Society*, 132(46):16442–16449.

Cheng, X., Ivanov, I., Wang, H., Sine, S. M., and McCammon, J. A. (2009). Molecular-dynamics simulations of ELIC-a prokaryotic homologue of the nicotinic acetylcholine receptor. *Biophysical journal*, 96(11):4502–4513.

Chiara, D. C., Gill, J. F., Chen, Q., Tillman, T., Dailey, W. P., Eckenhoff, R. G., Xu, Y., Tang, P., and Cohen, J. B. (2014). Photoaffinity Labeling the Propofol Binding Site in GLIC. *Biochemistry*, 53(1):135–142.

Christ, C. D., Mark, A. E., and van Gunsteren, W. F. (2010). Basic ingredients of free energy calculations: a review. *Journal of computational chemistry*, 31(8):1569–1582.

Cino, E. A., Choy, W.-Y., and Karttunen, M. (2012). Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *Journal of chemical theory and computation*, 8(8):2725–2740.

Corringer, P.-J., Poitevin, F., Prevost, M. S., Sauguet, L., Delarue, M., and Changeux, J.-P. (2012). Structure and pharmacology of pentameric receptor channels: from bacteria to brain. *Structure (London, England : 1993)*, 20(6):941–956.

Dalziel, A. W., Rollins, E. S., and McNamee, M. G. (1980). The effect of cholesterol on agonist-induced flux in reconstituted acetylcholine receptor vesicles. *FEBS letters*, 122(2):193–196.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092.

de Meyer, F. J. M., Rodgers, J. M., Willems, T. F., and Smit, B. (2010). Molecular Simulation of the Effect of Cholesterol on Lipid-Mediated Protein-Protein Interactions. *Biophysical journal*, 99(11):3629–3638.

Donnini, S., Tegeler, F., Groenhof, G., and Grubmüller, H. (2011). Constant pH Molecular Dynamics in Explicit Solvent with lambda-Dynamics. *Journal of chemical theory and computation*, 7(6):1962–1978.

Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W., and Cieplak, P. (2010). The r.e.d. tools: advances in resp and esp charge derivation and force field library building. *Physical Chemistry Chemical Physics*, 12(28):7821.

Durrant, J. D., de Oliveira, C. A. F., and McCammon, J. A. (2011). POVME: an algorithm for measuring binding-pocket volumes. *Journal of molecular graphics & modelling*, 29(5):773–776.

Engblom, A. C., Carlson, B. X., Olsen, R. W., Schousboe, A., and Kristiansen, U. (2002). Point mutation in the first transmembrane region of the beta 2 subunit of the gamma–aminobutyric acid type A receptor alters desensitization kinetics of gamma–aminobutyric acid- and anesthetic-induced channel gating. *The Journal of biological chemistry*, 277(20):17438–17447.

Faraldo-Gómez, J. D., Forrest, L. R., Baaden, M., Bond, P. J., Domene, C., Patargias, G., Cuthbertson, J., and Sansom, M. S. P. (2004). Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations. *Proteins*, 57(4):783–791.

Feller, S. E., Zhang, Y., Pastor, R. W., and Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621.

Fogolari, F., Brigo, A., and Molinari, H. (2002). The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *Journal of Molecular Recognition*, 15(6):377–392.

Franks, N. P. and Lieb, W. R. (1984). Do general anaesthetics act by competitive binding to specific receptors? *Nature*, 310(5978):599–601.

Gonzalez-Gutierrez, G., Cuello, L. G., Nair, S. K., and Grosman, C. (2013). Gating of the proton-gated ion channel from Gloeobacter violaceus at pH 4 as revealed by X-ray crystallography. *Proceedings of the National Academy of Sciences of the United States of America*.

Gonzalez-Gutierrez, G. and Grosman, C. (2010). Bridging the gap between structural models of nicotinic receptor superfamily ion channels and their corresponding functional states. *Journal of molecular biology*, 403(5):693–705.

Grossfield, A., Feller, S. E., and Pitman, M. C. (2007). Convergence of molecular dynamics simulations of membrane proteins. *Proteins*, 67(1):31–40.

Grossfield, A. and Zuckerman, D. M. (2009). Quantifying uncertainty and sampling quality in biomolecular simulations. *Annual reports in computational chemistry*, 5:23–48.

Gumbart, J., Khalili-Araghi, F., Sotomayor, M., and Roux, B. (2012). Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochimica et biophysica acta*, 1818(2):294–302.

Hamouda, A. K., Jayakar, S. S., Chiara, D. C., and Cohen, J. B. (2013). Photoaffinity Labeling of Nicotinic Receptors: Diversity of Drug Binding Sites! *Journal of molecular neuroscience : MN*.

Hansen, A. H., Sørensen, K. T., Mathieu, R., Serer, A., Duelund, L., Khandelia, H., Hansen, P. L., and Simonsen, A. C. (2013). Propofol modulates the lipid phase transition and localizes near the headgroup of membranes. *Chemistry and physics of lipids*, 175-176C:84–91.

Herrera, F. E. and Pantano, S. (2009). Salt induced asymmetry in membrane simulations by partial restriction of ionic motion. *The Journal of Chemical Physics*, 130(19):195105.

Hess, B. (2000). Similarities between principal components of protein dynamics and random diffusion. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 62(6 Pt B):8438–8448.

Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G. (1997). LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472.

Hibbs, R. E. and Gouaux, E. (2011). Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature*, 474(7349):54–60.

Hilf, R. J. C., Bertozzi, C., Zimmermann, I., Reiter, A., Trauner, D., and Dutzler, R. (2010). Structural basis of open channel block in a prokaryotic pentameric ligand-gated ion channel. *Nature structural & molecular biology*, 17(11):1330–1336.

Hilf, R. J. C. and Dutzler, R. (2008). X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature*, 452(7185):375–379.

Hilf, R. J. C. and Dutzler, R. (2009). Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature*, 457(7225):115–118.

Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins*, 33(3):417–429.

Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544.

Hoover, W. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical review. A*, 31(3):1695–1697.

Hormuzdi, S. G., Filippov, M. A., Mitropoulou, G., Monyer, H., and Bruzzone, R. (2004). Electrical synapses: a dynamic signaling system that shapes the activity of neuronal networks. *Biochimica et biophysica acta*, 1662(1-2):113–137.

Howard, R. J., Murail, S., Ondricek, K. E., Corringer, P.-J., Lindahl, E., Trudell, J. R., and Harris, R. A. (2011a). Structural basis for alcohol modulation of a pentameric ligand-gated ion channel. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29):12149–12154.

Howard, R. J., Slesinger, P. A., Davies, D. L., Das, J., Trudell, J. R., and Harris, R. A. (2011b). Alcohol-binding sites in distinct brain proteins: the quest for atomic level resolution. *Alcoholism, clinical and experimental research*, 35(9):1561–1573.

Huang, J. and MacKerell, A. D. (2013). Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of Computational Chemistry*, 34(25):2135–2145.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–8, 27–8.

Ito, M., Xu, H., Guffanti, A. A., Wei, Y., Zvi, L., Clapham, D. E., and Krulwich, T. A. (2004). The voltage-gated Na+ channel NaVBP has a role in motility, chemotaxis, and pH homeostasis of an alkaliphilic Bacillus. *Proceedings of the National Academy of Sciences of the United States of America*, 101(29):10566–10571.

Izuo, M. (2004). Medical history: Seishu hanaoka and his success in breast cancer surgery under general anesthesia two hundred years ago. *Breast Cancer*, 11(4):319–324.

Kandel, E., Schwartz, J., and Jessell, T. (2000). *Principles of Neural Science*. McGraw-Hill Medical.

Krieger, E., Dunbrack, R. L., Hooft, R. W. W., and Krieger, B. (2012). Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods in molecular biology (Clifton, N.J.)*, 819:405–421.

Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. *Proteins*, 47(3):393–402.

Krieger, E., Nielsen, J. E., Spronk, C. A. E. M., and Vriend, G. (2006). Fast empirical pKa prediction by Ewald summation. *Journal of molecular graphics & modelling*, 25(4):481–486.

Krikorian, A. D. (1975). Were the opium poppy and opium known in the ancient near east? *Journal of the History of Biology*, 8(1):95–114.

Kubo, R., Toda, M., and Hashitsume, N. (1992). *Statistical Physics II - Nonequilibrium Statistical Mechanics*. Springer New York.

Kutzner, C., Grubmüller, H., de Groot, B. L., and Zachariae, U. (2011). Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophysical journal*, 101(4):809–817.

Laube, B., Kuhse, J., Rundström, N., Kirsch, J., Schmieden, V., and Betz, H. (1995). Modulation by zinc ions of native rat and recombinant human inhibitory glycine receptors. *The Journal of physiology*, 483 (Pt 3):613–619.

Laurent, B., Murail, S., Da Silva, F., Corringer, P.-J., and Baaden, M. (2013). Modeling complex biological systems: From solution chemistry to membranes and channels. *Pure and Applied Chemistry*, 85(1):1–13.

Leach, A. R. (2001). *Molecular Modelling*. Principles and Applications. Pearson Education.

LeBard, D. N., Hénin, J., Eckenhoff, R. G., Klein, M. L., and Brannigan, G. (2012). General anesthetics predicted to block the GLIC pore with micromolar affinity. *PLoS computational biology*, 8(5):e1002532.

Lee, S.-J., Song, Y., and Baker, N. A. (2008). Molecular Dynamics Simulations of Asymmetric NaCl and KCl Solutions Separated by Phosphatidylcholine Bilayers: Potential Drops and Structural Changes Induced by Strong Na+-Lipid Interactions and Finite Size Effects. *Biophysical journal*, 94(9):3565–3576.

Li, H., Robertson, A. D., and Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4):704–721.

Lide, D. R. (2003). *CRC Handbook of Chemistry and Physics, 84th Edition*. CRC Press.

Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PloS one*, 7(2):e32131.

Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011). How fast-folding proteins fold. *Science (New York, N.Y.)*, 334(6055):517–520.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., and Ha, S. a. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B*, 102(18):3586–3616.

Martyna, G. J., Tuckerman, M. E., Tobias, D. J., and Klein, M. L. (1996). Explicit reversible integrators for extended systems dynamics. *Molecular Physics*, 87(5):1117–1157.

McGuffee, S. R. and Elcock, A. H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS computational biology*, 6(3):e1000694.

Meng, Y. and Roitberg, A. E. (2010). Constant pH replica exchange molecular dynamics in biomolecules using a discrete protonation model. *Journal of chemical theory and computation*, 6(4):1401–1412.

Meyer, K. H. (1937). Contributions to the theory of narcosis. *Trans. Faraday Soc.*, 33:1062.

Miyamoto, S. and Kollman, P. A. (1992). Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962.

Miyazawa, A., Fujiyoshi, Y., and Unwin, N. (2003). Structure and gating mechanism of the acetylcholine receptor pore. *Nature*, 423(6943):949–955.

Mnatsakanyan, N. and Jansen, M. (2013). Experimental determination of the vertical alignment between the second and third transmembrane segments of muscle nicotinic acetylcholine receptors. *Journal of neurochemistry*.

Molza, A.-E., Ferey, N., Czjzek, M., Le Rumeur, E., Hubert, J.-F., Tek, A., Laurent, B., Baaden, M., and Delalande, O. (2014). FD169: Innovative interactive flexible docking method for multi-scale reconstruction elucidates dystrophin molecular assembly. *Faraday Discussions*, pages –.

Mowrey, D., Chen, Q., Liang, Y., Liang, J., Xu, Y., and Tang, P. (2013a). Signal transduction pathways in the pentameric ligand-gated ion channels. *PloS one*, 8(5):e64326.

Mowrey, D., Cheng, M. H., Liu, L. T., Willenbring, D., Lu, X., Wymore, T., Xu, Y., and Tang, P. (2013b). Asymmetric ligand binding facilitates conformational transitions in pentameric ligand-gated ion channels. *Journal of the American Chemical Society*, 135(6):2172–2180.

Mulle, C., Léna, C., and Changeux, J. P. (1992). Potentiation of nicotinic receptor response by external calcium in rat central neurons. *Neuron*, 8(5):937–945.

Murail, S., Howard, R. J., Broemstrup, T., Bertaccini, E. J., Harris, R. A., Trudell, J. R., and Lindahl, E. (2012). Molecular mechanism for the dual alcohol modulation of Cys-loop receptors. *PLoS computational biology*, 8(10):e1002710.

Murail, S., Wallner, B., Trudell, J. R., Bertaccini, E., and Lindahl, E. (2011). Microsecond simulations indicate that ethanol binds between subunits and could stabilize an open-state model of a glycine receptor. *Biophysical journal*, 100(7):1642–1650.

Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Takeuchi, C., Yamada, M., and Tabata, S. (2003). Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 10(4):137–145.

Nielsen, J. E., Andersen, K. V., Honig, B., Hooft, R. W., Klebe, G., Vriend, G., and Wade, R. C. (1999). Improving macromolecular electrostatics calculations. *Protein engineering*, 12(8):657–662.

Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268.

Nosé, S. and Klein, M. L. (1983). Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076.

Nury, H., Poitevin, F., Van Renterghem, C., Changeux, J.-P., Corringer, P.-J., Delarue, M., and Baaden, M. (2010). One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. *Proceedings of the National Academy of Sciences of the United States of America*, 107(14):6275–6280.

Nury, H., Van Renterghem, C., Weng, Y., Tran, A., Baaden, M., Dufresne, V., Changeux, J.-P., Sonner, J. M., Delarue, M., and Corringer, P.-J. (2011). X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel. *Nature*, 469(7330):428–431.

Olsen, R. W., Li, G.-D., Wallner, M., Trudell, J. R., Bertaccini, E. J., Lindahl, E., Miller, K. W., Alkana, R. L., and Davies, D. L. (2013). Structural Models of Ligand-Gated Ion Channels: Sites of Action for Anesthetics and Ethanol. *Alcoholism, clinical and experimental research*.

Palma, E., Maggi, L., Miledi, R., and Eusebi, F. (1998). Effects of Zn2+ on wild and mutant neuronal 7 nicotinic receptors. *Proceedings of the National Academy of Sciences*, 95(17):10246–10250.

Pan, J., Chen, Q., Willenbring, D., Mowrey, D., Kong, X.-P., Cohen, A., Divito, C. B., Xu, Y., and Tang, P. (2012). Structure of the Pentameric Ligand-Gated Ion Channel GLIC Bound with Anesthetic Ketamine. *Structure (London, England : 1993)*, 20(9):1463–1469.

Paramo, T., East, A., Garzón, D., Ulmschneider, M. B., and Bond, P. J. (2014). Efficient characterisation of protein cavities within molecular simulation trajectories: trj_cavity. *Journal of chemical theory and computation*, page 140328121610008.

Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190.

Parton, D. L., Klingelhoefer, J. W., and Sansom, M. S. P. (2011). Aggregation of Model Membrane Proteins, Modulated by Hydrophobic Mismatch, Membrane Curvature, and Protein Class. *Biophysical journal*, 101(3):691–699.

Peters, J. A., Hales, T. G., and Lambert, J. J. (1988). Divalent cations modulate 5-HT3 receptor-induced currents in N1E-115 neuroblastoma cells. *European journal of pharmacology*, 151(3):491–495.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16):1781–1802.

Piggot, T. J., Holdbrook, D. A., and Khalid, S. (2011). Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *The journal of physical chemistry. B*, 115(45):13381–13388.

Prevost, M. S., Sauguet, L., Nury, H., Van Renterghem, C., Huon, C., Poitevin, F., Baaden, M., Delarue, M., and Corringer, P.-J. (2012). A locally closed conformation of a bacterial pentameric proton-gated ion channel. *Nature structural & molecular biology*.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics (Oxford, England)*, 29(7):845–854.

Ramerstorfer, J., Furtmüller, R., Sarto-Jackson, I., Varagic, Z., Sieghart, W., and Ernst, M. (2011). The GABAA receptor alpha+beta- interface: a novel target for subtype selective drugs. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(3):870–877.

Roth, R., Gillespie, D., Nonner, W., and Eisenberg, R. E. (2008). Bubbles, gating, and anesthetics in ion channels. *Biophysical journal*, 94(11):4282–4298.

Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341.

Sachs, J. N., Crozier, P. S., and Woolf, T. B. (2004). Atomistic simulations of biologically realistic transmembrane potential gradients. *The Journal of Chemical Physics*, 121(22):10847–10851.

Samson, R., Legendre, J. B., Christen, R., Fischer-Le Saux, M., Achouak, W., and Gardan, L. (2005). Transfer of Pectobacterium chrysanthemi (Burkholder et al. 1953) Brenner et al. 1973 and Brenneria paradisiaca to the genus Dickeya gen. nov. as Dickeya chrysanthemi comb. nov. and Dickeya paradisiaca comb. nov. and delineation of four novel species, Dickeya dadantii sp. nov., Dickeya dianthicola sp. nov., Dickeya dieffenbachiae sp. nov. and Dickeya zeae sp. nov. *International journal of systematic and evolutionary microbiology*, 55(Pt 4):1415–1427.

Sauguet, L., Howard, R. J., Malherbe, L., Lee, U. S., Corringer, P.-J., Adron Harris, R., and Delarue, M. (2013a). Structural basis for potentiation by alcohols and anaesthetics in a ligand-gated ion channel. *Nature communications*, 4:1697.

Sauguet, L., Poitevin, F., Murail, S., Van Renterghem, C., Moraga-Cid, G., Malherbe, L., Thompson, A. W., Koehl, P., Corringer, P.-J., Baaden, M., and Delarue, M. (2013b). Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *The EMBO journal*.

Sauguet, L., Shahsavar, A., Poitevin, F., Huon, C., Menny, A., Nemecz, À., Haouz, A., Changeux, J.-P., Corringer, P.-J., and Delarue, M. (2014). Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3):966–971.

Schlick, T. (2010). *Molecular Modeling and Simulation: An Interdisciplinary Guide*, volume 21. Springer New York.

Schmidtke, P., Bidon-Chanal, A., Luque, F. J., and Barril, X. (2011). MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics (Oxford, England)*, 27(23):3276–3285.

Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., and Wriggers, W. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science (New York, NY)*, 330(6002):341–346.

Shevchenko, A. and Simons, K. (2010). Lipidomics: coming to grips with lipid diversity. *Nature reviews. Molecular cell biology*, 11(8):593–598.

Smart, O. S., Goodfellow, J. M., and Wallace, B. A. (1993). The pore dimensions of gramicidin A. *Biophysical journal*, 65(6):2455–2460.

Soares, T. A. and Straatsma, T. P. (2008). Assessment of the convergence of molecular dynamics simulations of lipopolysaccharide membranes. *Molecular Simulation*, 34(3):295–307.

Spurny, R., Billen, B., Howard, R. J., Brams, M., Debaveye, S., Price, K. L., Weston, D. A., Strelkov, S. V., Tytgat, J., Bertrand, S., Bertrand, D., Lummis, S. C. R., and Ulens, C. (2013). Multisite binding of a general anesthetic to the prokaryotic pentameric Erwinia chrysanthemi ligand-gated ion channel (ELIC). *The Journal of biological chemistry*, 288(12):8355–8364.

Spurny, R., Ramerstorfer, J., Price, K., Brams, M., Ernst, M., Nury, H., Verheij, M., Legrand, P., Bertrand, D., Bertrand, S., Dougherty, D. A., de Esch, I. J. P., Corringer, P.-J., Sieghart, W., Lummis, S. C. R., and Ulens, C. (2012). Pentameric ligand-gated ion channel ELIC is activated by GABA and modulated by benzodiazepines. *Proceedings of the National Academy of Sciences*, 109(44):E3028–34.

Taly, A., Delarue, M., Grutter, T., Nilges, M., Le Novère, N., Corringer, P.-J., and Changeux, J.-P. (2005). Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophysical journal*, 88(6):3954–3965.

Tanford, C. and Kirkwood, J. G. (1957). Theory of Protein Titration Curves.I.General Equations for Impenetrable Spheres. *Journal of the American Chemical Society*, 79(20):5333–5339.

Tasneem, A., Iyer, L. M., Jakobsson, E., and Aravind, L. (2004). Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. *Genome biology*, 6(1):R4.

Tirion, M. M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters*, 77(9):1905–1908.

Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure (London, England : 1993)*, 16(5):673–683.

Unwin, N. (2005). Refined Structure of the Nicotinic Acetylcholine Receptor at 4Å Resolution. *Journal of molecular biology*, 346(4):967–989.

Vamparys, L. (2013). *Exploration de la reconnaissance de la courbure membranaire par le motif ALPS*. PhD thesis, Université Paris Diderot.

van Meer, G., Voelker, D. R., and Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nature reviews. Molecular cell biology*, 9(2):112–124.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and et al. (2009). Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *Journal of Computational Chemistry*, 31:671–690.

Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103.

Vernino, S., Amador, M., Luetje, C. W., Patrick, J., and Dani, J. A. (1992). Calcium modulation and high calcium permeability of neuronal nicotinic acetylcholine receptors. *Neuron*, 8(1):127–134.

Wang, H.-L., Cheng, X., and Sine, S. M. (2012). Intramembrane proton binding site linked to activation of bacterial pentameric ion channel. *The Journal of biological chemistry*, 287(9):6482–6489.

Weir, C. J. (2006). The molecular mechanisms of general anaesthesia: dissecting the GABAA receptor. *Continuing Education in Anaesthesia, Critical Care & Pain*, 6(2):49–53.

Weng, Y., Yang, L., Corringer, P.-J., and Sonner, J. M. (2010). Anesthetic sensitivity of the Gloeobacter violaceus proton-gated ion channel. *Anesthesia and analgesia*, 110(1):59–63.

Willenbring, D., Liu, L. T., Mowrey, D., Xu, Y., and Tang, P. (2011). Isoflurane alters the structure and dynamics of GLIC. *Biophysical journal*, 101(8):1905–1912.

Williams, S. L., Blachly, P. G., and McCammon, J. A. (2011). Measuring the successes and deficiencies of constant pH molecular dynamics: a blind prediction study. *Proteins*, 79(12):3381–3388.

Wood, W. G., Cornwell, M., and Williamson, L. S. (1989). High performance thin-layer chromatography and densitometry of synaptic plasma membrane lipids. *Journal of lipid research*, 30(5):775–779.

Yip, G. M. S., Chen, Z.-W., Edge, C. J., Smith, E. H., Dickinson, R., Hohenester, E., Townsend, R. R., Fuchs, K., Sieghart, W., Evers, A. S., and Franks, N. P. (2013). A propofol binding site on mammalian GABAA receptors identified by photolabeling. *Nature chemical biology*, 9(11):715–720.

Yoluk, Ö., Brömstrup, T., Bertaccini, E. J., Trudell, J. R., and Lindahl, E. (2013). Stabilization of the GluCl Ligand-Gated Ion Channel in the Presence and Absence of Ivermectin. *Biophysical journal*, 105(3):640–647.

Zhang, X., Bhatt, D., and Zuckerman, D. M. (2010). Automated Sampling Assessment for Molecular Simulations Using the Effective Sample Size. *Journal of chemical theory and computation*, 6(10):3048–3057.

Zhu, F. and Hummer, G. (2009). Gating transition of pentameric ligand-gated ion channels. *Biophysical journal*, 97(9):2456–2463.

Zhu, F. and Hummer, G. (2010). Pore opening and closing of a pentameric ligand-gated ion channel. *Proceedings of the National Academy of Sciences of the United States of America*, 107(46):19814–19819.

Zhu, F. and Hummer, G. (2012a). Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *Journal of computational chemistry*, 33(4):453–465.

Zhu, F. and Hummer, G. (2012b). Drying Transition in the Hydrophobic Gate of the GLIC Channel Blocks Ion Conduction. *Biophysical journal*, 103(2):219–227.

Zimmermann, I. and Dutzler, R. (2011). Ligand Activation of the Prokaryotic Pentameric Ligand-Gated Ion Channel ELIC. *PLoS biology*, 9(6):e1001101.

Zimmermann, I., Marabelli, A., Bertozzi, C., Sivilotti, L. G., and Dutzler, R. (2012). Inhibition of the Prokaryotic Pentameric Ligand-Gated Ion Channel ELIC by Divalent Cations. *PLoS biology*, 10(11):e1001429.

Zuckerman, D. M. (2011). Equilibrium Sampling in Biomolecular Simulations. *Annual Review of Biophysics*, 40(1):41–62.

Résumé

Pendant des milliers d'années, l'Homme a utilisé des décoctions de plantes et l'alcool pour leur vertu sédative. La découverte des anesthésiques dans les derniers siècles représente un progrès majeur, rendu possible principalement par l'observation empirique de leurs effets sur les humains et les animaux. Des expériences *in vitro* ont révélées les neurorécepteurs comme cibles possibles des anesthésiques. Ces récepteurs sont des canaux membranaires localisés sur les cellules cibles aux terminaisons nerveuses. Au cours des dernières années, des homologues bactériens de neurorécepteurs ont été identifiés. Le récepteur GLIC, un homopentamère homologue au récepteur nicotinique humain, a été co-cristallisé avec des anesthésiques généraux liés à lui, dont le bromoforme, le desflurane et le propofol.

Dans cette thèse, j'utilise les simulations de dynamique moléculaire et la programmation de logiciels pour caractériser les interactions entre les anesthésiques généraux et la forme sauvage de GLIC, ainsi qu'avec différents mutants. En 2011, le propofol et le desflurane ont été co-cristallisés dans un site de liaison intra-sous-unité localisé dans le domaine transmembranaire de GLIC. Plus récemment, il a été montré que le bromoforme se lie à ce site ainsi qu'à un site inter-sous-unités. Dans ce travail, je décris des simulations d'une nouvelle structure cristalline montrant un site de liaison supplémentaire situé dans le pore du canal. Des simulations dans lesquelles GLIC est noyé de bromoforme ont démontré l'accessibilité spontanée des sites cristallographiques dans un environnement non cristallin. Des calculs d'énergie libre exhaustifs corroborent ces données mettant en évidence des différences d'énergie de liaison entre les sites et entre des mutants de GLIC. Un échantillonnage complet des poches de liaison m'a permis de détecter un deuxième site de liaison inter-sous-unité duquel l'accessibilité est probablement modulée par un résidu en particulier. Ensemble, les données accumulées au cours de ce projet fournissent une image grandissante de l'action des anesthésiques à l'échelle atomique.

Abstract

For thousands of years, Humans have been using plant decoctions and alcohol for their sedative effect. The discovery of anesthetic molecules in recent centuries represents a notable advance, mostly enabled by empirically observing their effect on humans and animals. *In vitro* experiments uncovered neuroreceptors as possible target for anesthetic molecules. Those are membrane-bound ion channels located on the target cells at nervous endings. In the last few years, bacterial neuroreceptor homologs were identified. The GLIC receptor, a homopentamer homologue to the human nicotinic receptor, was co-crystallized with bound general anesthetics, including bromoform, desflurane and propofol.

In this thesis, I use molecular dynamics simulations and software programming to characterize interactions of general anesthetics with the wild type form of GLIC as well as with several mutants. In 2011, propofol and desflurane were co-crystallized in an intrasubunit binding site located in GLIC's transmembrane domain. More recently, bromoform was shown to bind this site as well as an intersubunit site. In this work I describe simulations of a new crystal structure displaying an additional binding site located in the channel's pore. Simulations in which GLIC is flooded by bromoform demonstrate the spontaneous accessibility of crystallographic binding sites in a non-crystalline environment. Exhaustive free energy calculations corroborate this data highlighting differences of binding energy between sites and between GLIC variants. Extensive sampling of binding pockets allowed me to detect a second intersubunit binding site, the accessibility of which is possibly modulated by a specific residue. Alltogether, data accumulated in this project provide a growing picture of anesthetic action at the atomic scale.