Thèse d'exercice

UNIVERSITÉ DE LIMOGES

Faculté de Pharmacie

2016 THÈSE N°

The use of computer simulations to serve a new approach of drug design

Application to the pentameric ligand gated ion channels family

THÈSE POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN PHARMACIE

présentée et soutenue publiquement

le 4 Novembre 2016

par

Nicolas MARTIN

Née le 21 Juillet 1990, à St Germain-en-Laye (78)

EXAMINATEURS DE LA THÈSE

M. le Professeur Nicolas Picard ..... Président

M. le Professeur Patrick Trouillas ..... Directeur

M. le Docteur Gabin Fabre ..... Juge

M. le Docteur Florent Di Meo ..... Juge

Thèse d'exercice

UNIVERSITÉ DE LIMOGES

Faculté de Pharmacie

2016 THÈSE N°

The use of computer simulations to serve a new approach of drug design

Application to the pentameric ligand gated ion channels family

THÈSE POUR LE DIPLÔME D'ÉTAT DE DOCTEUR EN PHARMACIE

présentée et soutenue publiquement

le 4 Novembre 2016

par

Nicolas MARTIN

Née le 21 Juillet 1990, à St Germain-en-Laye (78)

EXAMINATEURS DE LA THÈSE

M. le Professeur Nicolas Picard ..... Président

M. le Professeur Patrick Trouillas ..... Directeur

M. le Docteur Gabin Fabre ..... Juge

M. le Docteur Florent Di Meo ..... Juge

2 rue du Docteur Marcland
87025 Limoges cedex
T. 05 55 43 58 00
F. 05 55 43 58 01
S. http://www.unilim.fr

Faculté de Pharmacie

01.09.2016

DOYEN DE LA FACULTE : Monsieur le Professeur Jean-Luc **DUROUX**

VICE-DOYEN : Madame le Professeur Catherine **FAGNERE**

ASSESSEURS : Madame le Professeur Sylvie **ROGEZ**
Monsieur le Professeur Serge **BATTU**

PROFESSEURS :

BATTU Serge	CHIMIE ANALYTIQUE
CARDOT Philippe	CHIMIE ANALYTIQUE ET BROMATOLOGIE
DESMOULIERE Alexis	PHYSIOLOGIE
DUROUX Jean-Luc	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE
FAGNERE Catherine	CHIMIE THERAPEUTIQUE – CHIMIE ORGANIQUE
LIAGRE Bertrand	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MAMBU Lengo	PHARMACOGNOSIE
ROUSSEAU Annick	BIOSTATISTIQUE
TROUILLAS Patrick	CHIMIE PHYSIQUE – PHYSIQUE
VIANA Marylène	PHARMACOTECHNIE

PROFESSEURS DES UNIVERSITES – PRATICIENS HOSPITALIERS DES DISCIPLINES PHARMACEUTIQUES :

MOESCH Christian	HYGIENE HYDROLOGIE ENVIRONNEMENT
PICARD Nicolas	PHARMACOLOGIE
ROGEZ Sylvie	BACTERIOLOGIE ET VIROLOGIE
SAINT-MARCOUX Franck	TOXICOLOGIE

ASSISTANT HOSPITALIER UNIVERSITAIRE DES DISCIPLINES PHARMACEUTIQUES :

CHAUZEIX Jasmine	HEMATOLOGIE

MAITRES DE CONFERENCES :

BASLY Jean-Philippe	CHIMIE ANALYTIQUE ET BROMATOLOGIE
BEAUBRUN-GIRY Karine	PHARMACOTECHNIE
BILLET Fabrice	PHYSIOLOGIE
CALLISTE Claude	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE
CLEDAT Dominique	CHIMIE ANALYTIQUE ET BROMATOLOGIE
COMBY Francis	CHIMIE ORGANIQUE ET THERAPEUTIQUE
COURTIOUX Bertrand	PHARMACOLOGIE, PARASITOLOGIE
DELEBASSEE Sylvie	MICROBIOLOGIE-PARASITOLOGIE-IMMUNOLOGIE
DEMIOT Claire-Elise	PHARMACOLOGIE
FROISSARD Didier	BOTANIQUE ET CRYPTOGAMIE
GRIMAUD Gaëlle	CHIMIE ANALYTIQUE ET CONTROLE DU MEDICAMENT
JAMBUT Anne-Catherine	CHIMIE ORGANIQUE ET THERAPEUTIQUE
LABROUSSE Pascal	BOTANIQUE ET CRYPTOGAMIE

LEGER David	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MARION-THORE Sandrine	CHIMIE ORGANIQUE ET THERAPEUTIQUE
MARRE-FOURNIER Françoise	BIOCHIMIE ET BIOLOGIE MOLECULAIRE
MERCIER Aurélien	PARASITOLOGIE
MILLOT Marion	PHARMACOGNOSIE
MOREAU Jeanne	MICROBIOLOGIE-PARASITOLOGIE-IMMUNOLOGIE
MUSUAMBA TSHINANU Flora	PHARMACOLOGIE
PASCAUD Patricia	PHARMACIE GALENIQUE – BIOMATERIAUX CERAMIQUES
POUGET Christelle	CHIMIE ORGANIQUE ET THERAPEUTIQUE
VIGNOLES Philippe	BIOPHYSIQUE, BIOMATHEMATIQUES ET INFORMATIQUE

ATTACHE TEMPORAIRE D'ENSEIGNEMENT ET DE RECHERCHE :

FABRE Gabin (01.09.2016 au 31.08.2017)	CHIMIE PHYSIQUE – PHYSIQUE
LAVERDET Betty (1.09.2016 au 31.08.2017)	PHARMACIE GALENIQUE
PHAM Thanh Nhat (1.09.2016 au 31.08.2017)	CHIMIE ORGANIQUE – BIOCHIMIE

PROFESSEURS EMERITES :

BUXERAUD Jacques

DREYFUSS Gilles

OUDART Nicole

Remerciements

I would like to thank first the professor Nicolas Picard for accepting to be the president of my jury. Your comments helped me a lot in writing a better manuscript.

I also want to thank the professor Patrick Trouillas for endless reasons that I could not exhaustively enumerate here. You introduced me to computational sciences with all the care and patience my pharmaceutical background required. You gave me the chance to start doing research in your lab and to get excited about science. You always were of good advice and I needed a lot of advice! For all of the above and much more I'm grateful and wish that I would have the chance to collaborate again with you in my future research career.

Doctor Gabin Fabre, thank you for everything you did for me. You listened to my doubts, my complaints and my endless list of questions with all the world's patience and empathy. You were the best office mate and I must say my big brother when it comes to science. Thank you for all the great moments we spent together in the lab or outside, I wish everyone could have a Gabin on his team.

Doctor Florent Di Meo, thank you for pushing me to go deeper in my understanding of science but also of life in general. I loved the lively discussions we had about pretty much everything during the time we spent together in the lab. Thank you also for introducing me to quantum chemistry. I hope to get better at it someday.

I would also like to thank all the other members of the lab or partners of lunch and diners. Pierre-André, Thibault, Thomas, Gaëlle, Claire, thank to all of you.

I thank the members of my lab here in Strasbourg who introduced me to plenty of new methods, subjects, and to the exciting world of programming. Dr. Simone Conti, Dr. Nicolas Calimet, Dr. Jeremy Esque, Joel Montalvo, Florian Blanc and Adrien Cerdan, you all either taught me a lot or helped me clear my mind from work during long tea breaks. I also thank the Dr. Marco Cecchini, the current supervisor of my PhD, for his comments and help on this project.

I would like to thank all my classmates as well, who endured me for so many hours in the amphitheatres. The ones of the first year, Adrien, Etienne, Marion, Louise, Vivien, Anne-Caroline, Corentin, and the ones I only had the chance to meet later, Jean-Charles, Nicolas, Rémi, Julien, Sophie and Elodie (yes you belong together), Xavier, Cindy Maxime and Laurianne. And all the other I might forget. We had a great time all together and I wish we had a chance to rewind the tape and enjoy it once more!

I would also like to thank Marion. During the past 2 years, you have made my life easier, better and happier. You taught me a lot about myself and helped me to overcome the

tough moments of life. I cherish your love and support and hope that we will enjoy many “afternoon-tea-when-it’s-cold-outside” together in the years to come.

Finally I would like to thank my parents and family. Mom you always pushed me to do my best and to always keep in mind that, in the end, what matters is to be happy. You taught me to be curious and to dare living new experiences. I could never thank you enough for all your support and love. You are the main reason for which I could finally, after two long years of discontinuous writing, finish this manuscript and defend it. I hope someday to be as brave and strong as you are. Da’, I know you will always be here for me. You encouraged me and taught me to be ambitious and responsible, you prepared me for the world. You are a model for me and I wish someday to be, just like you, able to work as much as with my head as with my hands. You both gave me all I could need to be happy and so far it works!

Droits d'auteurs

Table des matières

Introduction

The communication between neurons allows the transmission of information from the brain to peripheral sites. The neurotransmission process involves the liberation and capture of neurotransmitters in the synaptic cleft. A wide range of pathologies can alter this mechanism and the function of the type of receptor involved will determine the corresponding syndrome.

Pentameric ligand gated ion channels (pLGICs) are a superfamily of receptors involved in fast synaptic transmission in the central nervous system. In mammals, they comprise $5HT_3$, $GABA_A$, and glycine receptors. Thus they contribute to major functions such as central autonomous control, sleep, anxiety, memory and attention, emotion or cognition and many others. Several mutations of the genes' sequence coding for the pLGICs has been identified and are responsible for severe neurological disorders including the congenital myasthenic syndrome, several epileptic syndromes, hyperekplexia and autism. In addition to the former, malfunction of nACh and $GABA_A$ receptors appears linked to neurodegenerative pathologies such as Alzheimer disease (AD) for which only few treatments are nowadays available.

Structurally they are complex transmembrane protein made of 5 homologous subunits arranged in a five-fold symmetry around an ion pore. For each subunit, one can describe a extracellular (EC) domain that capture the neurotransmitter diffusing in the synaptic cleft, a very organized transmembrane domain where other ligands can bind and an intracellular domain of which the function is still not fully understood.

Because of their crucial functions the pLGICs are the targets of many drugs. They can be activated by endogenous neurotransmitters but also by therapeutic molecules that regulate their action and bind at the level of an active site, i.e., orthosteric site, lodged between the EC domains of two adjacent subunits. Moreover this superfamily of receptors is capable of binding small molecules at a topographically different site, i.e., the allosteric site. The later is located in the transmembrane domain within a subunit, unlike the orthosteric site. Ivermectin (IVM) is known to bind in the allosteric site of most pLGICs and to positively modulate their action. General anaesthetics are also binding at the same site on $GABA_AR$ and act as positive allosteric modulators. Nevertheless the detailed mechanism of action of these allosteric binders, i.e., how they block or enhance the activity of a given receptor, is unknown.

In this manuscript we will first give a detailed review of the architecture and members of the pLGICs family. Then we will present their pharmacology and discuss the consequences of their dysfunction on human health. Finally we will present computer simulation of

biomolecules and a method and illustrate this approach with the development of a method based on free energy calculations that we developed to study the effect of IVM on a bacterial homologue of the pLGICs family.

1. Pentameric ligand gated ion channels: history and atomic structure.

1.1. Discovery

First named nicotinic receptors due to their properties discovered by John Newport Langley in 1905 [1], these receptors were then called cys-loop, due to the presence of a very highly conserved loop formed by a cysteine bridge in the extracellular domain. They are now called pentameric ligand gated ion channels referring to both their function and structure.

In 1970, the first pLGICs, the nicotinic acetylcholine receptor, was isolated from fish electric organ [2], however its detailed tri-dimensional structure is still unknown. The first atomistic structures of pLGICs were elucidated half a century later from prokaryotic homologues; GLIC from *gloebacter violaceus* [3] and ELIC from *Erwinia chrysantemi* [4]. It is only very recently that high resolution 3D structures of eukaryotic homologues [5,6] were published and led to a breakthrough in the field.

1.2. Phylogenetic of the pLGICS family

Until the discovery of Tasneem et al. in 2005 of a prokaryotic homologue of the pLGICs family it was thought that these membrane proteins were present only in complex eukaryotic organisms. In fact, they are present in most of the living organisms, i.e., from bacteria to human brain [7] . They are involved in major functions in the central nervous system (CNS) as in the muscles.

The genes coding for these receptors can be sequenced thanks to modern techniques and the results can be analysed to draw various conclusions. The amino acid sequences can be aligned and compared with other receptors within a given group to isolate subgroups among the family of pLGICs, as shown on Figure 1, namely: i) AChBP which are very similar to the rest of the members in the extracellular (EC) domain but the transmembrane (TM) domain is completely absent; ii) nAChR and 5-HT_3; iii) $GABA_{A,C}R$ and GlyR and iv) bacterial homologues.

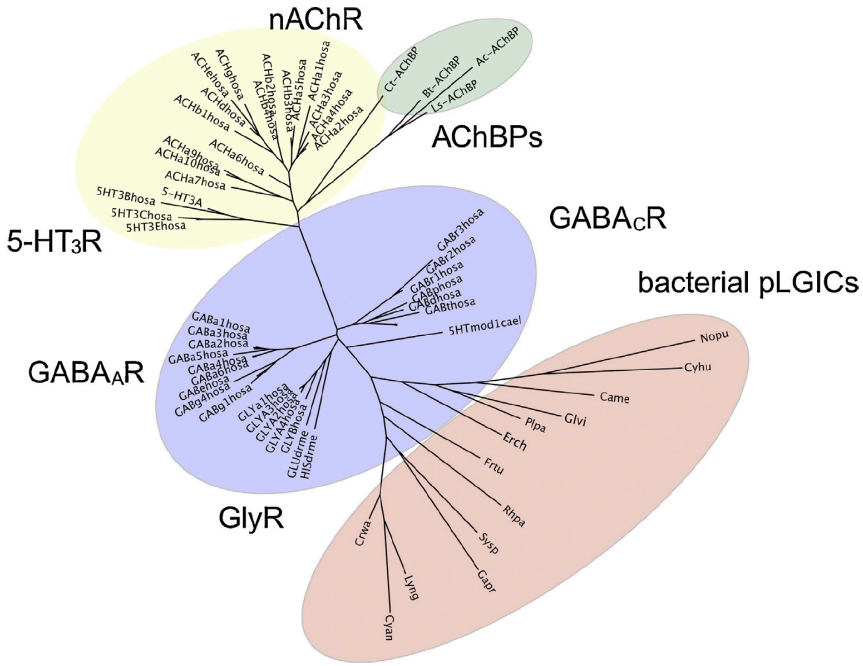


Figure 1: phylogenetic tree of the pLGICs family, including prokaryotic and eukaryotic receptors. Are shown in yellow the cation-selective eukaryotic; in green the acetylcholine binding proteins; in blue the anion selective eukaryote; and in pink the prokaryotic homologues. [8]

1.3. Human members of the pLGICs family

They consist in several human receptors, which are all major pharmacological targets:

- GABA$_A$R are anionic receptors. A recent structure of a human receptor was elucidated by X-ray crystallography in complex with benzamidine (a new antagonist). [9]
- 5-HT$_3$R are cationic receptors. A recent structure of a mouse receptor was elucidated in X-ray crystallography. It is complexed with nanobodies that bind at the interface between the subunits in the EC domain preventing the binding of the endogenous neurotransmitter, thus acting as antagonists. Nanobodies are also used in crystallography to reduce the entropy of the EC domain, which facilitates the formation of crystals.
- GlyR are anionic receptors. A recent set of structures obtained by X-ray crystallography and cryo-electron-microscopy (cryoEM) was published. See Table I for a complete presentation of these structures.
- nAChR are cationic receptors. So far no complete atomic-resolved structure of this receptor was elucidated. Only theoretical models based on prokaryotic homologues of this receptor are available. The acetylcholine-binding-protein (AChBP), a homologue of the EC domain of the nAChR, is extensively used to

explore the binding of small molecules to the nAChR by experimental techniques. Although, it cannot describe neither the effect of binding on the activity of the receptor nor describe the activation or deactivation mechanisms.

Receptors of the pLGICs family are composed of several α and β subunits (heteromeric receptors) which are slightly different in their sequence and thus in their 3D structure. This influences their pharmacological et kinetic properties [10]. Interestingly, bacterial pLGICs are often homomeric, meaning that they are composed of five subunits of the same type when human homologues are mostly heteromeric.

1.4. Non-human members of the pLGICs family

Because they are often homomeric membrane proteins and are expressed in prokaryotic or very simple eukaryotic organisms, and because of their very close homology to their human homologues, these receptors are used as models for computational as well as for electrophysiology studies.

ELIC: (*Erwinia* Ligand Gated Ion Channel) isolated from the bacteria *Dickeya dadantii (*previously known as *Erwinia chrysanthemi*) is responsible for the soft rot disease in foliage plants. It was the first structure ever elucidated of a pLGIC at high resolution (3.3 Å)[11].

GluCl: (Glutamate Gated Chloride Channel) the two structures published of this protein were expressed in *Caenorhabditis elegans* and were the first eukaryotic structure of a pLGIC. Although the open state is believed to be an active state due to its ligation with IVM and glutamate (positive allosteric modulator (PAM) and endogenous neurotransmitter respectively) it was recently stated that it could be a desensitized state [12]. The resting structure was crystallized in the absence of ligand and it is thus believed to be representative of a closed state of the channel.

GLIC: (*Gloeobacter* Ligand gated Ion Channel) isolated from the bacteria *Gloeobacter violaceus*. It is activated by protons, thus it is sensitive to changes in pH.

Table I: Sum up of the principal structures available of pLGICs and their characteristics.

Receptor	Organism	In complex with	Channel state	Resolution (Å)	PDBID
Glycine	Human	Strychnine	Close	3.04	5CFB
Glycine	Zebra fish	Strychnine	Close	3.9	3JAD
Glycine	Zebra fish	Glycine	Open	3.9	3JAE
Glycine	Zebra fish	Glycine, ivermectin	Open	3.8	3JAF
GluCl	Bacteria	-	Close	3.6	4TNV
GluCl	Bacteria	Glutamate, ivermectin	Open	3.35	3RIF
ELIC	Bacteria	-	Closed	3.3	2VL0
GLIC	Bacteria	Propofol/desflurane	Open	3.3	3P50
GLIC	Bacteria	-	Open	2.9	3EAM
GLIC	Bacteria	-	Closed	4.35	4NPQ
5-HT$_3$	Mouse	Nanobodies (VHH15)	Closed	3.5	4PIR
GABA$_A$	Human	Benzamidine	Closed	2.97	4COF

1.5. General organisation

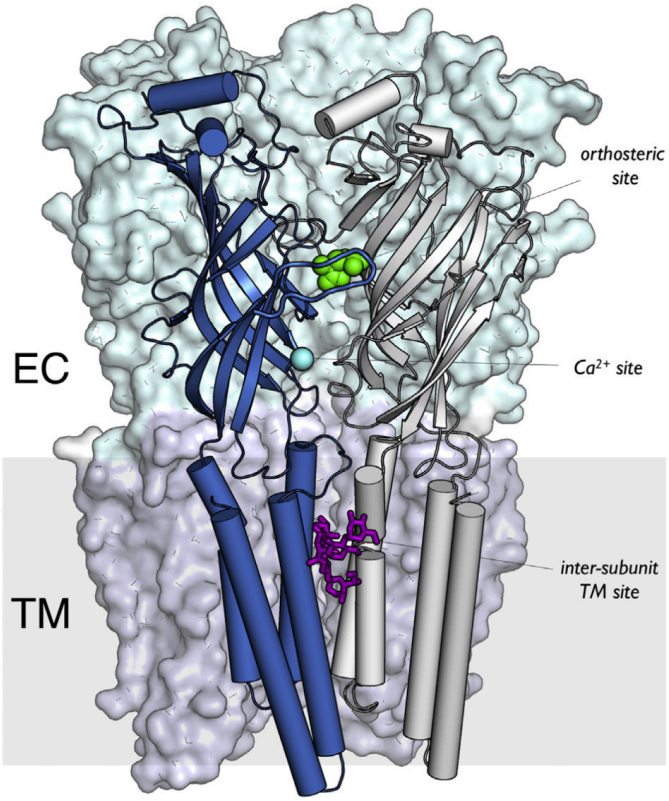


Figure 2: general architecture of the pLGICs, example of an eukaryotic homologue from *C. elegans*: GluCl.

pLGICs are trans-membrane proteins with a molecular weight of about 290 kDa, comprising five identical or homologous subunits symmetrically arranged around a central ionic channel with a five-fold axis perpendicular to the membrane plane.

They are found in bacteria as well as in human brain. In mammals, nAChRs can be composed of different types of subunits, namely nine β-subunits and three α-subunits. They assemble into various compositions, for instance $(\alpha4)_2(\beta2)_3$ or $(\alpha7)_5$, which impact the pharmacological and physiological properties [10]. Indeed, upon their location in the human body or even in the different parts of the brain (see Figure 3), nAChRs have different subunit compositions [13].

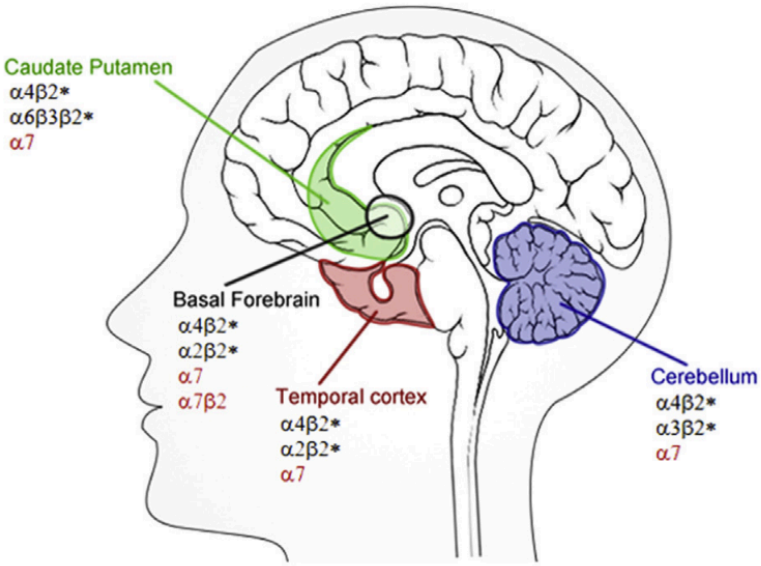


Figure 3: Repartition of nAChRs subtypes in the human brain. [13]

The overall structure of the pLGICs is well known due to crystallographic structures, each subunit consists in 3 well-defined domains: a large hydrophilic extracellular domain (EC), a transmembrane domain (TM) mainly lipophilic and an intracellular domain. At the interface between the TM and the EC domains one can describe a conserved cysteine bridge forming a loop referred to as the F-loop. The latter carries the canonical ΦPΦD motif where Φ are aromatic residues, P a proline and D an aspartic acid.

1.6. Extracellular domain

1.6.1. Orthosteric binding site

No X-ray structure of the main pharmaceutical target, i.e. nAChR, was elucidated so far. Nevertheless, one may extrapolate the structure of the binding site of human pLGICs from crystal structures of eukaryotic homologues such as GLIC, ELIC or GluCl or from electronic microscopy (EM) or even using the AChBP.

It is lodged in the upper part of the extracellular domain, at the interface between two subunits. The subunit holding the A, B, and C-loops is referred to as (+) or principal subunit, whereas the E, F, and G-loops are carried by the (-) or complementary subunit (see Figure 4). The binding pocket is barely accessible to solvent and formed by the loop of the (+) subunit on one side and by the β strands of the (-) subunit on the other.

In the binding site of GluCl L-glutamate forms a bridge between the principal and complementary subunits. It is tightly sandwiched between 2 tyrosines of the (+) subunit and two arginines on the (-) subunit.

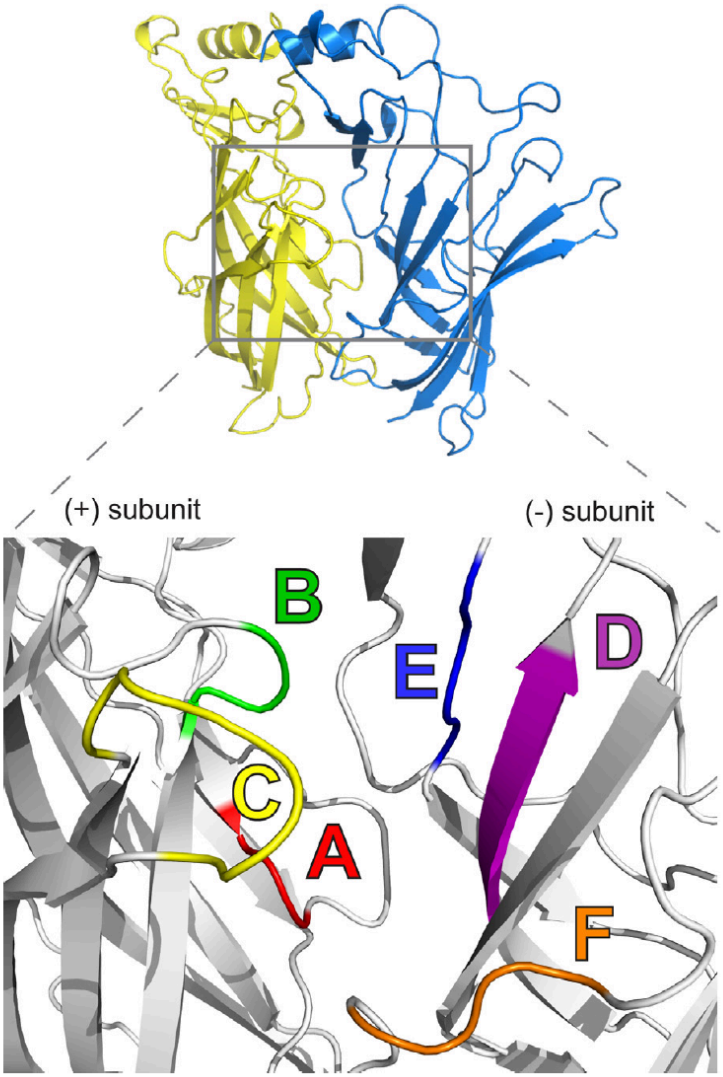


Figure 4: representation of the active site of the Ac-AchBP. The binding site is located at the intersubunit surface. (+) Subunit is shown in yellow and (-) in blue in the upper part of the figure. The six loops forming the active site are highlighted in colors in the lower part of the figure. [8]

Despite the diverse sources of structures, the orthosteric neurotransmitter-binding site of pLGICs seems to be remarkably conserved from bacteria to humans.

1.6.2. Loop C

The binding pocket is capped by a flexible segment of amino acids called C-loop, which is stabilized in a closed configuration by interacting with the ligand placed in the active site. When no ligand is bound, the C-loop is much more flexible and can sample a wide range of conformations going from completely open to completely close. While in an open configuration, the active site is exposed to the solvent and can thus facilitate the spontaneous binding of a ligand.

The flip of the C-loop from an open to a closed position was first thought to have a critical role in the allosteric activation of the receptor [8]. Although it is very likely that the closing of the loop is a consequence of the interactions between the ligand and the C-loop and is not what triggers the opening or closing of the channel 50 Å away from the orthosteric binding site.

1.7. Transmembrane domain

1.7.1. Cys-loop

The loop connecting helices M2 and M3 and located at the interface between the TM and EC domains is often referred to as cys-loop as it contains two cysteins forming a disulfide bridge. Moreover, it carries a very conserved proline believed to be a key residue in the opening or closing mechanism of the pLGICs.

It was recently stressed [14] that the proline located in the cys-loop is subjected to strong evolutionary pressure leading to a high conservation of this residue among the member of the family, both in bacteria and in human homologues. Interestingly, the cysteins residues on the cys-loop after which this family of receptors was named are less conserved. This would suggest renaming the cys-loop in Pro-loop receptors for a more consistent nomenclature.

1.7.2. Allosteric binding site

Several allosteric sites have been identified from pLGICs' structures. The first allosteric potentiator discovered for pLGICs was Ca^{2+} for $\alpha7$ and $\alpha4\beta2$ nAChRs [15,16]. Although, a wide variety of molecules such as ethanol and other alcohols, general anaesthetics, lipids, cholesterol, neurosteroids, ivermectin and other synthetic compounds [7,17–22] can bind to this site. These binders typically have low intrinsic activity but can drastically increase or decrease the physiological answer expected from a given receptor.

The allosteric binding sites are located about 40 Å away from the orthosteric site in the upper part of the TM domain. Two major allosteric binding sites were identified, although they partially overlap. The first one is a cleft and is formed by the helix M1 of the (-) subunit and M3 of the (+) subunit, and may host IVM and lipids. The second allosteric binding site is referred to as general anaesthetic binding site because it was elucidated on crystal structures of GLIC in presence of propofol and desflurane [23]. Unlike IVM binding site, it is lodged inside a subunit and is much smaller.

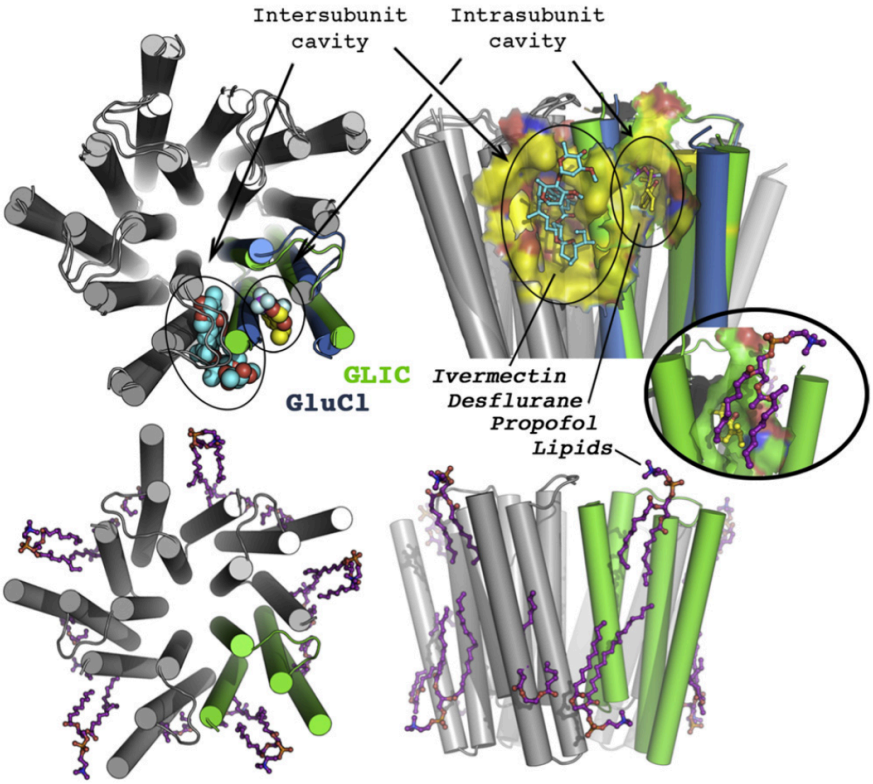


Figure 5: top and side views of the TM domain of GLIC (green) and GluCl (grey) showing the allosteric binding sites of IVM.[7]

1.7.3. The ion channel

The physiological answer to ligand binding of receptors of the family of pLGICs consists in the opening of the channel to let ions get through the cell membrane. This may lead to different endings such as excitation, inhibition, or metabolic signalling.

The ion channel is formed by five alpha helices, referred to as M2, oriented perpendicularly to the axis of the membrane and arranged symmetrically around the axis of the pore. To compare different structures, the residues lining the pore were numbered using a prime notation from -2' (inner part of the ion pore, facing the cytoplasm of the cell) to 20' (outer part of the ion pore, facing the outside of the cell). This ion channel consists in a hydrophobic domain, well conserved among the pLGICs, in the middle of the ion channel and two charged residues located at its edges.

The ion channel narrowest portion is called constriction point. In most pLGICs (GluCl, ELIC, GlyR, GABA and $5HT_3$,see ref [24–27] for detailed information) a ring of five leucine residues pointing toward the lumen of the pore form the gate at position 9'. The inward or outward displacements of these residues are responsible for the closing or the opening of the channel to the ion flux; respectively. Indeed, previous experiments [28–30] showed that if these

key leucine residues were mutated to hydrophilic residues, the opening rate of the channel is increased. This is explained by the fact that, when replaced by hydrophilic residues, the hydrophobic gate residues cannot play their role of electrostatic repulsion forming a barrier to the ions.

Even though the pLGICs family is homogeneous in terms of global architecture and in the overall composition of the ion channel, the selectivity for ions diverges. $GABA_AR$, GlyR and GluCl are anionic channels, when 5-HT_3R, nAChR and GLIC are cationic channels. The mechanism of the ion selectivity is not fully elucidated. It is however clear that unlike highly selective channels such as KcsA, the selectivity of this type of receptors lies in charge interactions and not in the ion radius. The residues thought to be responsible for the ion selectivity are called selectivity filter residues.

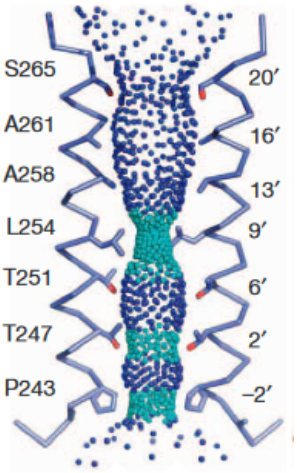


Figure 6: side view of the ion channel describing the residues lining the pore and their specific numbering for GluCl apo (PDBID 4NTV). M2 helices are represented in ribbons and the volume of the channel accessible to water is visualized by blue or cyan dots. [6]

1.7.4. The ion channel binding site

Different strategies can be used to control the activity of pLGICs. Among them, channel blockers, which bind directly inside the lumen of the pore. Unlike ortho and allosteric ligands, this type of molecules acts directly as sterical blockers of the channel. Thus their binding site and mechanism of action are peculiar. The binding sites of channel blockers are widely spread over the full transmembrane part of the channel. A recent structure of GluCl bound to picrotoxin [5] has shown that this channel blocker binds between position -2' and 2' (see Figure 6). It was also shown that tetraethylammonium and tetrabutylantimony bind in the middle of the channel at position 6' and that lidocaine binds near positions 9' and 6' [31].

1.8. Intracellular domain

The inner domain of these transmembrane proteins is absent from the prokaryotic members of the pLGICs family and thus is much less known than the extra and transmembrane domains. It is believed to act as a selectivity filter allowing only for the permeation of specific ions [32]. Due to the flexibility of the intracellular (IC) domain, very few data exist on its structure. In fact, in most of the few vertebrate structures present in the protein data bank (PDB) the IC domain has been cut out before the crystallization experiments. In 2014, Hassaine and co-workers crystallized a 5-HT_3R with a partial IC domain, which could show that the IC domain helices form a narrower constriction point that the one present at position 2'.

Despite the interest of the IC domain and the recent data published on its structure, its function is still barely known and its absence from bacterial homologues makes the elucidation of it's precise role much more complex.

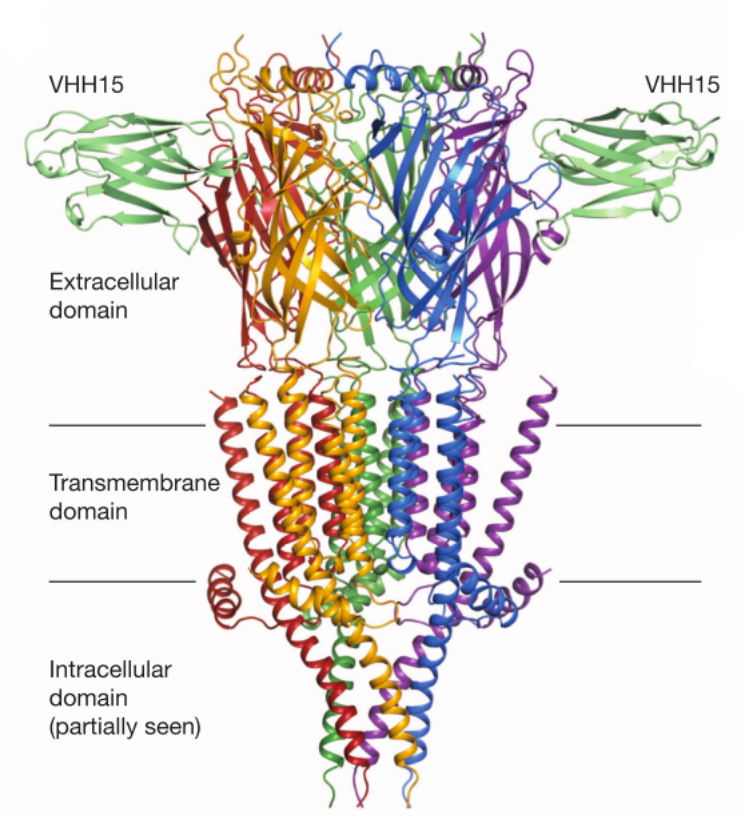


Figure 7: X-ray crystallographic structure of the 5-HT_3 receptor showing a partial intracellular domain and nanobodies (VHH15) binding at the level of the EC domain.[33]

2. Pharmacology of the pLGICs

2.1. Agonist and antagonist

Based on the work of Langley, Hill, and Clark, was defined the receptor occupancy model which described the notion of agonist and antagonist.

- An agonist is defined as a molecule that binds to a receptor and triggers a biological response. Such as the cleavage of an intracellular messenger, or an ion flux through the cell's membrane. Several kind of agonists can be cited:
    - Partial agonist: even when occupying 100% of the binding site the biological response triggered by the partial agonist is in intensity lower than the one obtained by the binding of the natural agonist.
    - Inverse agonist: when binding to the target receptor it triggers the opposite biological response to the agonist.
- An antagonist is on the other hand defined as the molecule capable of blocking the action of the agonist. There are several types of antagonists:
    - Reversible agonist: ligand of that subtype can bind to their target (blood protein, nuclear receptor, nuclear receptor etc…) and unbind freely. The typical interactions involved in such cases are H-bonds and Van der Waals interactions.
    - Irreversible antagonist: here the ligand binds covalently to its target. The target activity will be restored as soon as the cell produces new ones. One may site as an example the action of aspirin on COX-1 and COX-2.
    - Allosteric antagonist: not all targets possess the ability of being allosterically modulated. An allosteric antagonist (or agonist) can bind to a target on a topographically different site than the one of the endogenous neurotransmitter.
    - Orthosteric antagonist: unlike the allosteric antagonist, it binds to the target at the endogenous neurotransmitter site and compete with it for its binding.

To quantify and sort all the different agonists, antagonists, and subtypes, one can use the allosteric efficacy defined as follow:

$$\propto = \frac{[AR^*][R]}{[AR][R^*]}$$

Equation 1: definition of the allosteric efficacy

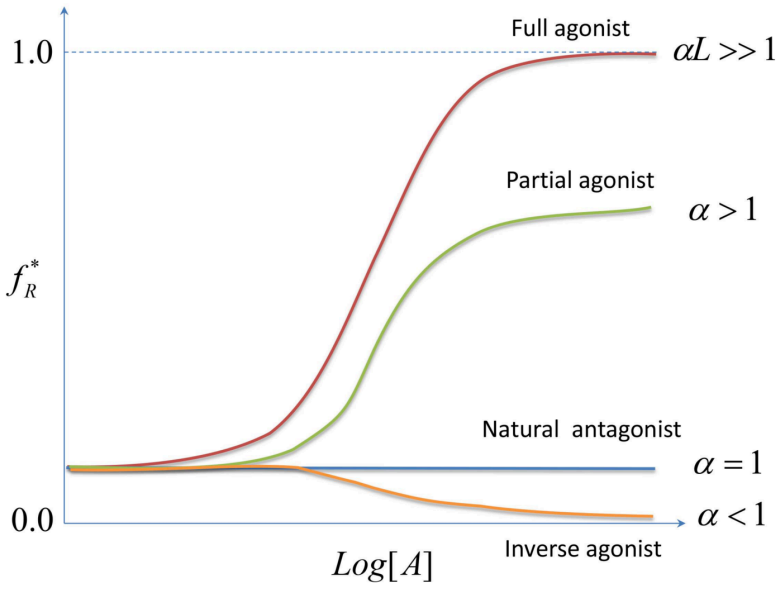


Figure 8: different types of ligands and their corresponding α values [34]. Where f_R^* is the fraction of receptors in the activated state and [A] is the concentration of the ligand.

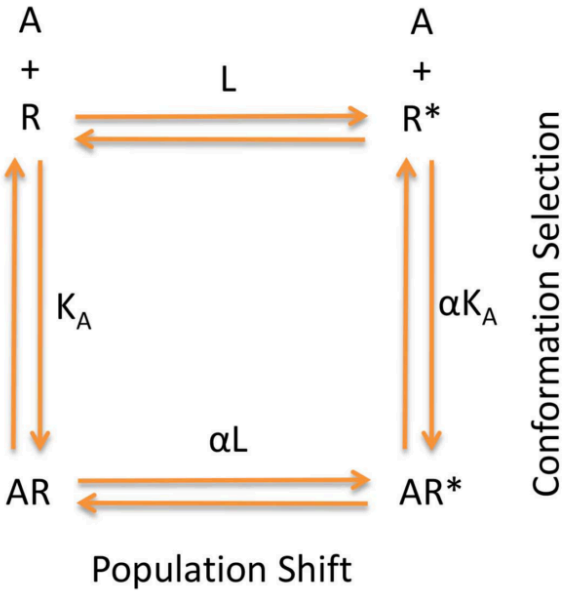
2.2. Allostery

Fifty years ago from now, with the mention of the adjective “allosteric” by Jacques Monod and Francois Jacob [35], a new fields of interest was named. Allosteric comes from Greek *allos* meaning “other, different”, and *stereos* standing for “solid”. This concept was first brought to explain the experiments of Jean-Pierre Changeux, on end product of the enzyme L-threonine deaminase [36]. In 1965, the publication of a landmark paper by Monod, Wyman and Changeux [37] revolutionized the field of allostery by proposing the WMC model. A year later, the premises of the a theory coined by Pauling in 1935 [38] were taken over by Koshland et al. to propose the KNF (Koshland Nemethy Filmer) model[39].

- *WMC model* (also referred to as conformational selection): in this description of receptor activation it is stated that both the resting and the active states pre-exist. This implies that the probability of populating a resting state in the absence of a ligand is not inexistent, even though it is very small. The binding of a ligand is in this model not responsible for the conformational changes but is the key that will lock or stabilize the pre-existing active conformation. It is assumed that the receptor is in dynamics equilibrium between the two conformations and that the ligand only selects one or the other leading to opposite physiological outcomes. In this model the capability of being active or resting lies in the protein sequence and

3D conformation and the ligand will only shift the equilibrium between the two states. The discovery of rare but existing ionic currents even in the absence of ligand in GluCl strengthened the claims of this model.

- *KNF model* (also referred to as induced fit): here is postulated that the activation process is not a property of the protein but is due to the binding of the ligand. In fact, the latter triggers a conformational wave that goes from the orthosteric or allosteric binding site to the action site and changes the conformation of the protein. In this model, a resting state of the receptor cannot exist in the absence of a ligand.



$$L = \frac{[R^*]}{[R]} \qquad K_A = \frac{[AR]}{[A][R]} \qquad \alpha K_A = \frac{[AR^*]}{[A][R^*]} \qquad \alpha L = \frac{[AR^*]}{[AR]}$$

Equation 2: Illustration of the allosteric two states model and its measurable quantities. [34]

2.3. Mechanism of action

Pentameric ligand gated ion channels can be activated by two different mechanisms. The first one is the binding of the endogenous neurotransmitter to its site. It can trigger alone the activation or deactivation of the receptor by respectively binding or unbinding to the orthosteric site. Each receptor can bind its endogenous neurotransmitter plus some other small molecules at this site, e.g., L-glutamate for GluCl, acetylcholine and nicotine for nAChR, Glycine for GlyR, serotonin for the $5HT_3R$ and γ-amino-butyric-acid for the $GABA_AR$.

The second mechanism is referred to as allosteric potentiation and can be triggered by positive or negative allosteric potentiators (PAMs/NAMs). They act by increasing or decreasing the activation of the receptor. This category of ligand binds to a different site, localized in the TM domain of the receptor.

There exist two ways of increasing the current of ions going through the channel of this family of receptor; i) the allosteric modulator can stabilize the active state and elongates the time in which it stay open, delaying the desensitization of the receptor, thus the closing of ion pore ii) enlarging the diameter of the transmembrane pore, thus increasing the quantity of ion going through the cell's membrane per unit of time.

According to the WMC model described before and which seems to be the most suitable to describe the mechanism of action of the pLGICs, one can describe the following; the cycle of activation starts with a receptor in a resting configuration with its active site not in a tridemensionnal conformation compatible with the binding of the ligand. The receptor then spontaneously go through a transition populating the active state f the receptor even though the ligand is not bound, and allows for the binding of the latter in its site. The ligand thus blocks the receptor in the active configuration leading to the creation of an ion flux and the physiological answer related. Depending on if bound to a PAM or not, the receptor will go shortly or after a longer time into desensitization, i.e. a state where the ligand is bound but no ion flux can be measured. The final step of the cycle consists in the unbinding of the ligand so the receptor can go back to the initial resting configuration.

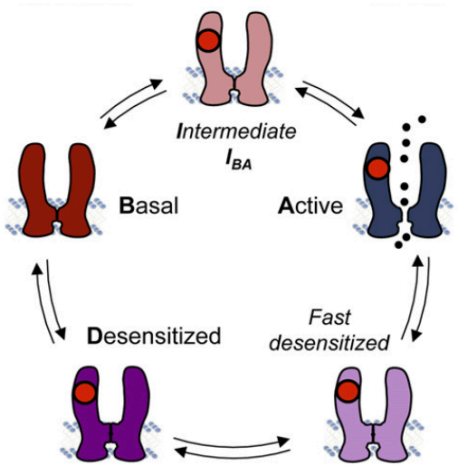


Figure 9: action cycle of pLGICs showing a KNF model where the biding of the ligand is responsible for the open of the channel. [7]

2.4. Gating mechanism

The atomic motions leading to quaternary changes in the structure of the receptor itself that trigger the closing or the opening of the channel is referred to as gating mechanism. It is of major interest to understand and describe in great details this key mechanism, which is responsible for the biological function of the pLGICs. Elucidating this mechanism, one could in theory enhance or decrease at will the activity of such receptors.

As first the understanding of the gating mechanism came through experimental studies such as electrophysiology and X-ray crystallography. Indeed the structure of prokaryotic pLGICs such as GLIC pH4 [23] (open channel), GLIC pH7 [40] (closed channel), GluCl open [41] and closed [6] or ELIC [11] provided the first insights on the gating mechanism. With the help of these structures, computational studies [42] brought a better understanding of the gating mechanism making possible to witness the process at the atomic scale. The latter techniques combined with experimental studies [3] allowed to put in light the first component of the gating, i.e. the twisting of the EC domain over the TM domain (see Figure 10 for a visual representation and Model 1 in Table II).

In a recent review by Cecchini and Changeux [43], the previous gating mechanism was completed by adding a second transformation of the EC domain: the blooming isomerization (see Figure 10 and Model 2 in Table II). It corresponds to a radial expansion of the 5 subunits in the outward direction. This motion is not mandatorily coordinated which means that all subunits can bloom with different kinetics. The presented model propose that the radial contraction of the EC domain promotes the opening of the ion pore at the transmembrane level and is followed by the un-twisting of the EC domain over the TM domain locking the ion pore in an open configuration.

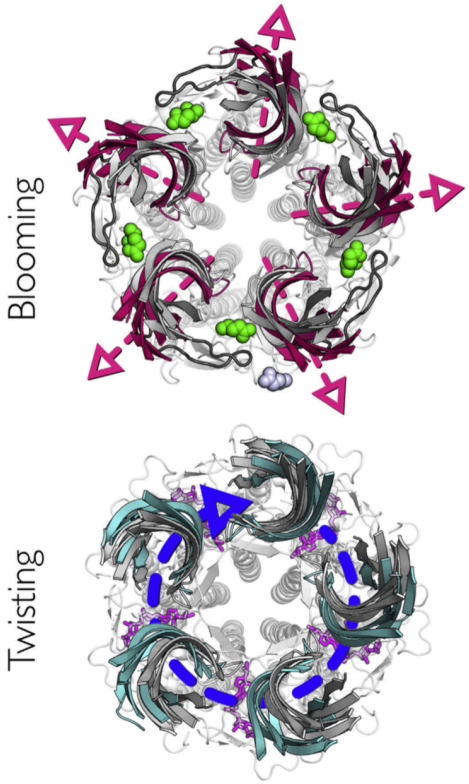


Figure 10: illustration of the two components of the gating mechanism, i.e. blooming and twisting. [43]

Table II: description of the two observables twisting and blooming in function for GluCl apo and active.

State	Channel	Model 1		Model 2	
		Twisting	Blooming	Twisting	Blooming
Apo	Closed	untwisted (24°)	bloomed (12°)	untwisted (24°)	-
Active	Open	twisted (12°)	unbloomed (8°)	twisted (12°)	-

2.5. Endogenous neurotransmitters

It is common practice to name a receptor after the molecule that allowed its discovery. As a first example, the nicotinic receptor, which was given its name due to its capability of binding nicotine. In fact the latter is not its natural ligand or endogenous neurotransmitter which is acetylcholine (Ach) but allowed its discovery.

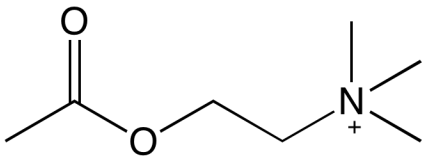


Figure 11: Acetylcholine 2D formula.

Table III: receptors of the pLGICs family and their corresponding endogenous neurotransmitter

Receptor	Endogenous nT
nAChR	Acetylcholine
$GABA_AR$	γ amino butyric acid
$5\text{-}HT_3R$	Serotonin
GlyR	Glycine
GLIC	H^+
ELIC	Various AA
GluCl	Glutamate

In all the receptors of the pLGICs family, the endogenous neurotransmitter binds to orthosteric binding site, localized in the ECD at the interface between subunits.

2.6. Associated pathologies

pLGICs are found in the human brain as they are found in muscles and in peripheral nervous system of smaller organisms and participate to the regulation of specific and crucial functions. For this reason, a wide range of associated pathologies can be described going from psychological and mental to physical disorders. These pathologies can either be due to mutations of the amino acids sequence of these receptors changing the response to normal stimulation or the lack or excess of these proteins at the surface of the cells' membrane.

2.6.1. Alzheimer's disease

AD accounts for about 70% of the senile dementia in the world, it is thought to be characterized by accumulation of misfolded amyloid-β peptide in the shape of extracellular plaques, intracellular polymerisation of tau protein, a loss of cholinergic tone and level of acetylcholine in the brain. The latter was hypothesized to be responsible for the cognitive decline observed in AD [44] [45].

It was first postulated that the memory loss shown in patients with AD was due to a deficit in muscarinic receptors mostly [46] [47]. Although, autoradiographic and histochemical studies of autopsy brain tissue [48] [49] [50] and brain imaging studies of AD patients lead toward a more specific loss of nicotinic receptors rather than muscarinic acetylcholine receptors. These data also show that muscarinic receptors including M2 receptors, are much less, if at all, reduced in AD.

2.6.2. Parkinson's disease

PD is a degenerative disorder of the central nervous system mainly affecting motor neurons in the brain. In patients with PD the level of dopamine in the brain is decreased compared to others. This is due to the abnormal death of the neurons generating dopamine in the part of the brain called *substantia nigra*. Moreover, already existing treatments for PD act on the liberation of dopamine by dopaminergic neurons, which can be stimulated by cholinergic receptors.

2.6.3. Schizophrenia

Schizophrenia is a disorder principally characterized by cognitive deficits and disorder thoughts. It is manifested by hallucinations and paranoia. It can have either genetic or environmental etiologies. Schizophrenic brain tends to show a robust decrease of the number of α7 nAChRs in the hippocampal region. Thus enhancing the response of nAChR to the binding of nACh using allosteric positive modulators could compensate the lack of α7 nAChRs.

2.6.4. Addiction

Tobacco use is one of the most important health issues in the world nowadays. Its use is increasing in the less developed countries and thus is responsible for an increase of the mortality. It was estimated in 1997 by the World Healthcare Organization (WHO) [51] that half of the adolescents who continued smoking throughout their whole life will die from smoking related diseases. Tobacco, due to its content in nicotine, is very addictive. Indeed, 80% of the attempts to quit smoking fail within a year.

Smoking cessation drugs acting on pLGICs have been shown efficient. Designing novel allosteric modulators could lead to the commercialization of other treatments possibly more receptor-specific and thus with less side effects.

2.7. Drugs acting on the pLGICs family

Various families of compounds act on pLGICs with a wide range of actions. One can cite as an example the benzodiazepines acting on $GABA_AR$. They have 5 well known activities; (i) sedative, (ii) hypnotic, (iii) anxiolytic, (iv) anticonvulsant and (v) myorelaxant. Many drugs, even though they are targeting one specific member of the pLGICs binds to the other receptor of the family, leading to different side effects. Although, even if one can describe the pharmaceutical target of one drug as done in Table III, it is very likely that this molecule will bind to other receptors. The orientation of the indication of a given drug lies on its specificity to a receptor.

Table IV: examples of drugs acting on the receptors of the pLGICs family and their indications.

Target	ICD	Indication
nAChR	Fluoxetine, Sertraline, Paroxetine, Citalopram	Depression
	Targacept	Alzheimer's disease
	Varenicline	Smoking cessation
	Carbachol	Glaucoma
$GABA_AR$	Phenobarbital	Anticonvulsant
	Diazepam	Anxiety
	Zolpidem, Zopiclone	Insomnia
	Etomidate, Propofol	General anaesthesia
$5\text{-}HT_3$	Tropisetron, Ondansetron, Granisetron, Dolasetron	Antiemetic
	Mirtazapine	Depression
	Alosetron*	Irritable Bowel Syndrome (IBS) in women only

* Commercialized in USA and withdrawn from the market in 2000 due to life-threatening adverse effects. Was reintroduced in 2002 with a restricted use.

3. Molecular modelling to understand the gating mechanism of pLGICs

3.1. Introduction to computer simulations

The use of computer to figure out molecular structures with an atomic resolution requires solvation of mathematical equations. The increasing power of computers through the past decades [52] has allowed scientists to tackle increasingly complex molecular systems. The solvation of physical-chemical equations and related approximations systematically requires numerical treatments, which is either infeasible or would require ages for human being whereas it can be done in short times by computers available nowadays. Thanks to this constant increasing of the power of CPU (central processing unit) and more recently GPU (graphical processing unit), one can simulate macromolecular systems as complex as lipid bilayers [53], proteins embedded in membranes [54], or even virus capsids [55] under physiological conditions, within a reasonable amount of time. One can even extrapolate from the Moore's law [52] that in 2050, we should be capable of simulating systems of billions of atoms e.g., a entire living cell.

3.1.1. Quantum mechanics

Light has been described throughout history as both a particle and a wave. Among others, De Broglie has rationalized this dual description and has extended this concept to any particle. This has been ideally formulated with the quantum mechanics formalisms, as particularly well exemplified by the Schrödinger equation:

$$\mathcal{H}\psi = E\psi$$

Equation 3: Schrödinger's equation

Where $\mathcal{H}$, the Hamiltonian operator, contains all information related to the energetics of the particle(s) and that is applied on ψ, the wavefunction associated to the particle(s) and E the energy of the studied particle(s).

In quantum mechanics, the movement of electrons is explicitly considered, which allows to study chemical reactions (e.g., formation or breaking of covalent bonds). The Schrödinger's equation can be solved analytically only for molecular systems containing one electron. To tackle problems of major biological relevance such as conformations of protein or ligand binding energies, several approximations are required.

3.1.2. The Born-Oppenheimer approximation

Considering that the mass of electrons is negligible with respect to that of nuclei, the Born-Oppenheimer approximation [56] suggests to decouple the motion of electrons and nuclei. Among other consequences of this approximation, the energy can be split into two independent contributions:

$$E_{tot}(nuclei, electrons) = E(electrons) + E(nuclei)$$

Equation 4: decomposition of the electronic and nuclear energy [57]

3.1.3. Molecular mechanics

Due to the current computer power limitations, the very accurate methods based on the quantum theory cannot be used for large molecular systems such as proteins. Molecular mechanics (MM), by definition neglects electronic motion, considers atoms as constant volume spheres, bonds as springs and obeys to Newtonian physics. These assumptions make MM much faster than solving Schrödinger's equation.

Within the MM formalism, the potential energy of a system is written as the sum of several contributions: (i) bond stretching; (ii) angle opening and closing; (iii) dihedral angle rotation; (iv) van der Waals interactions; (v) electrostatic interactions.

The MM formalism can be used in molecular dynamics and to perform energy minimization of a given system.

3.1.4. Force fields

In MM, the energy is obtained from the force field, i.e., the equation gathering all energetic contributions and all related parameters for all atom types. A force field is built as a table including all these parameters, which are derived from either experimental data or from highly accurate quantum calculations.

There exist many different force fields, all having particular specificities, e.g., Slipids [60] focuses on lipids, the different versions of AMBER[61] are well feature to DNA and proteins, so does CHARMM [62], and CGenFF [63] can be used for small molecules.

A general form of the potential energy is defined by the following equation [57], however it may vary with possible implementations :

$$E(r^N) = \sum_{bonds} \frac{k_i}{2}(l_i - l_0) + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_0) + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma))$$
$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left\{ 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right\}$$

Equation 5: decomposition of the potential energy used in most common force fields.

This general equation is divided into several terms :

- Bond stretching: for a given distance between two atoms, and based on a reference value, this function provides the potential energy between two atoms with respect to the inter-nuclear distance. Among the many functions available to calculate such quantity, the Morse law is particularly used (Figure 12). It shows that when the distance between two particles tends to zero, the interaction energy tends to infinity and that on the other side, if the distance between two particles tends to infinity, the interaction energy tends to zero. However, a simpler potential based on the harmonic oscillator approximation is used in most FF and provides a proper description of the interaction energy around the equilibrium distance.

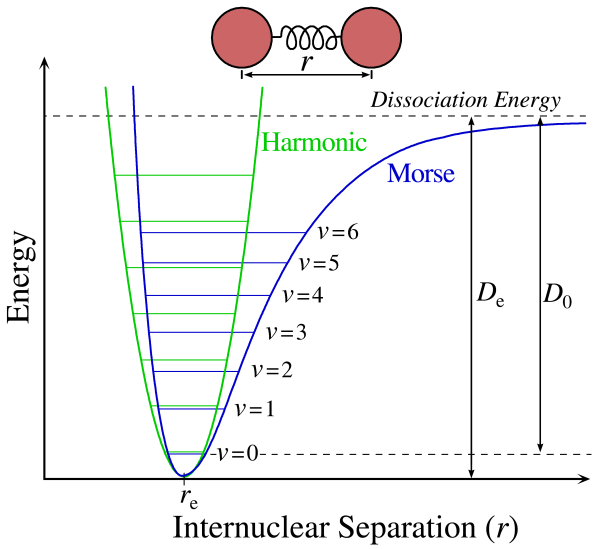


Figure 12: interaction energy in function of the inter-nuclear distance of a di-atomic system according to both harmonic oscillator approximation and Morse law.

- Angle bending: here, the evolution of the energy versus the angle is described by a harmonic potential. The potential energy increases quadratically with respect to the variation from the equilibrium angle.

- Dihedral angle torsions: Variation of dihedral angle may imply large energetic costs and may significantly drive the overall organisation of the molecule. They are usually described by cosinus functions, i.e., exhibiting periodicity.

  Non-bonded interactions: these interactions are mainly the van der Waals and electrostatic interactions, which are usually described by the Lennard Jones and Coulomb potentials, respectively.

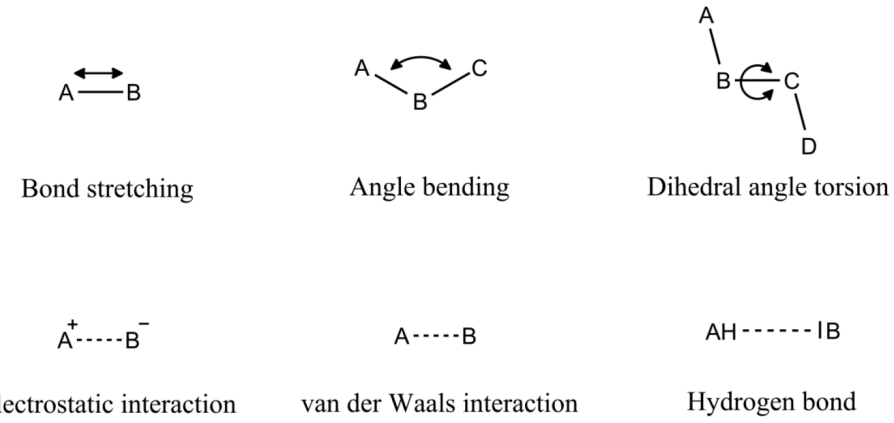


Figure 13: scheme of the different interactions described in a classical force field.

The major drawbacks of MM methods are that they are highly parameter-dependent. Indeed, when doing the above-mentioned approximation, the description of an atom type depends on chemical environment. For instance the ideal length of a sp3 carbon is not absolute and depends on the neighbouring atoms. Therefore, if not using the most adequate parameters, i.e., force field, for a given set of atoms, the obtained results are much questionable.

3.1.5. Molecular dynamics

When MM methods only allow for the calculation of a static property, molecular dynamics (MD) allows to study the motion of particles by solving Newton's second law of motion.

$$-\frac{dU}{dx_i} = m_i \frac{d^2x_i}{dt^2}$$

Equation 6: Newton's second law of motion

Where U is the potential energy calculated using the force field parameters, $-\frac{dU}{dx_i}$ the force acting on a particle i of mass m and of coordinates x and dt is an interval of time. To obtain a

trajectory, i.e., variation of x along time, the first step is to compute the forces applied on the initial coordinates. Then the forces applied can be computed as long as the corresponding velocities. Finally one calculates the new coordinates at time $t + dt$ from the velocities. The cycle can then be repeated with the new set of coordinates indefinitely.

In order to obtain an accurate and meaningful description of the motion of the atoms of a molecule, and because the applied force on a given atom are considered constant between two iterations of the integration cycle, one must use a dt (time step) as small as possible. In currently available molecular dynamics softwares the time step used by default is 1fs, which allows to describe even the fastest motions within a molecule, i.e., the stretching of bonds involving hydrogens.

Biological processes occurs on timescales of microsecond to minutes which implies that to reach that amount of sampling time, one should perform 10^{11} integration of the equation of motion for all atoms of the system. In order to decrease the huge number of cycle required, one can double the time step. However, this implied that the fastest degrees of vibration should be fixed. To this end, several algorithms such as SHAKE [58] or LINCS [59] have been developed to allow to restrain the stretching of covalent bounds involving hydrogens allowing for the use of a 2fs time step.

3.1.6. Energy landscape

The energy landscape is defined as the energy of a given conformation for all possible conformations of a system. For the sake of clarity, it is often represented as 2D curve, while it is a hyper surface of the energy vs. the 3N-6 degrees of freedom of a system of N atoms. For instance, in the case of a protein, when the quaternary structure changes due to activation or deactivation, the potential energy of the system increases or decreases forming two different basins in the energy landscape.

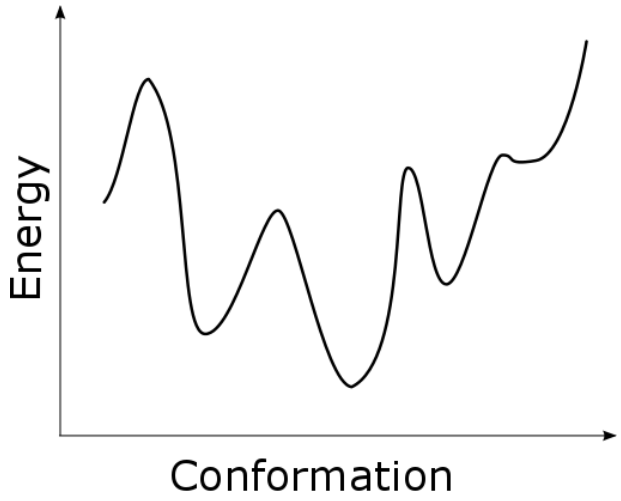


Figure 14: scheme of the mono dimensional energy landscape of a protein showing 4 basins

3.1.7. Umbrella sampling

Most of the biological processes occur within time scales ranging from microseconds to minutes. With the computer resources available nowadays it is not possible to reach such sampling time for system of thousands of atoms. Umbrella Sampling (US) is based on the concept that if one cannot observe a phenomenon in a computationally accessible amount of time, one can force the system, by applying a harmonic restrain, to undergo the expected transition and observe the evolution of its free energy. The applied restraining potential is of the following form:

$$W(r) = \frac{k}{2}(r - r_0)^2$$

Equation 7: harmonic biasing potential used in Umbrella sampling.

Where $W(r)$ is the energy applied by the bias, r is the current position, r_0 the equilibrium position and k the force constant of the harmonic oscillator.

In US the harmonic biasing potential is added when the potential energy of the system is computed (see Equation 5). This potential is used to flatten the energy space and thus encourage barrier crossing so the system can efficiently sample the full conformational space. Unlike with free simulations where the probability to obtain a given confirmation of the studied system depends on the systems, here it depends on the bias applied and for this reason in US this probability is a biased. Thus, one needs, in order to analyse the free energy derived from the probability calculated, to un-bias it, which can be done using the Weighted Histogram Analysis Method (WHAM); see below for further descriptions on this algorithm.

Before running US simulation one must choose the quantity ξ from which the biasing potential will be calculated. This quantity is often referred to as reaction coordinate. For instance if one wants to study the free energy of rotation of a dihedral, ξ will straightforwardly be the value of this dihedral angle.

Since the biasing potential used in US restrains the space explored by the system during MD around a centre or target value, one must run several simulations (called windows) with different centres in order to explore the full space of interest. The more the number of windows the more accurate will be the description of the reaction coordinate but the more will be the computational time required.

There exist no easy method for setting automatically neither the number or the centre of the windows nor the value of the force constant. These parameters must be defined based on *a priori* knowledge of the reaction coordinate and of the system of interest and out of trials and errors.

3.1.8. Potential of the mean force

3.1.8.1. Principle

Free energy is considered as a thermodynamic quantity of great importance in protein structures as it can explain and quantify complex mechanisms such as partitioning coefficients, ligand binding to a receptor or conformational changes of a protein. It is described by the following equation where F is the free energy, T the temperature, S the entropy and U the internal energy of the system:

$$F = U - T\Delta S$$

Equation 8: description of the free energy

Calculating the changes in free energy upon the variations along given reaction coordinate (RC) is known as potential of the mean force (PMF). It can be obtained through a variety of methods including US described before, metadynamics [64], accelerated molecular dynamics [65] or even free simulations at equilibrium. All the orthogonal reaction coordinates, i.e., which are not correlated with the RC studied, are thus integrated out. The free energy potential obtained through this method can be considered as the true energy potential only if one assumes that the simulations have reached convergence, i.e., are at equilibrium. Within the limit of this approximation, one can consider this energy potential to be the lower energy path along the chosen reaction coordinate. This particular point makes the choice of an appropriate and relevant reaction coordinate a matter of crucial importance.

3.1.8.2. Weighted Histogram Analysis Method (WHAM)

Running biased simulations using either US or other biasing methods such as metadynamics [64] , adaptive biasing force (ABF) [66] , or Accelerated Molecular Dynamics allows to compute biased probabilities of a given quantity. In order to unbiased these probabilities and recover the unbiased free energy along a given RC (PMF), one may use an automated and reproducible procedure based on the algorithm WHAM developed by Kumar et al. [67].

Using umbrella sampling to explore the energetics of a system along a RC, one runs several restrained simulations called windows (i) and each of them experiences a different biasing potential $U'(x)$. The unbiased free energy $A(x)$ of a given window can be expressed as follow:

$$A(x) = -k_B\,T \ln P'(x) - U'(x) + F$$

Equation 9: expression of the unbiased free energy

where k_B is the Boltzmann constant, $P'(x)$ is the biased probability to observe the value x and F is an unknown constant also referred to as offset.

$P'(x)$ and $U'(x)$ can easily be measured for all windows, but F, which depends on $U'(x)$ is unknown. $A(x)$ and the unbiased probability $P(x)$ are related by the following equation:

$$P(x) \propto e^{\frac{-A(x)}{k_B T}}$$

Equation 10 : relationship between the unbiased probability and the free energy

The WHAM algorithm allows to find the best F and $A(x)$ values by iteratively solving the self-consistent following equations:

$$\begin{cases} P(x) = \dfrac{\sum_{i=1}^{N_{wind}} n_i(x)}{\sum_{i=1}^{N_{wind}} N_i \exp\left(\dfrac{\left(F_i - U_i'(x)\right)}{k_B T}\right)} \\ F_i = -k_B\,T \ln\left(\sum_{x_{bin}=1}^{x_{bin}} P(x) \exp\left(\dfrac{U_i'(x)}{k_B T}\right)\right) \end{cases}$$

Equation 11 : WHAM equations

where N_{wind} is the total number of windows, x_{bin} is number of the bin, $n_i(x)$ is the number of time the value x is visited in window i and N_i is the total number of samples for window i.

The RC chosen is divided in n bins within which the PMF is assumed to be constant. This implies that the more the bins the better the resolution of the final PMF.

This method also implies that the probability distribution of the RC corresponding to a given window should overlap with its neighbours so WHAM is able to accurately reconstruct the full PMF.

3.2. Understanding the energetics of the twist of pLGICs using GluCl as a model

It was shown by Calimet et al. in 2013 [54] that upon removal of IVM (PAM), GluCl goes through a transition that shuts the ion pore 60 Å away from the binding site of the endogenous neurotransmitter, i.e. glutamate. As shown on **Erreur ! Source du renvoi introuvable.**, the twist angle (τ) move from 14 to 22 degrees when IVM is removed and remains around the value of the crystal structure, 14 degrees, when IVM is bound. This transition is associated with the closing of the ion pore.

Based on the knowledge of this transition, one may wonder if it is possible to block the twisting of the receptor so the ion pore remains open. In fact, this question is answered by the study of GluCl in the presence of IVM. In the latter case, the twist is locked at a value of 14 degrees forcing the ion pore to remain open thus increasing the overwall flux of chloride ion entering the cell.

Understanding the energetics behind this twist transition could not only allow to isolate intermediate states, which could be potential pharmacology targets, but also to design novel ligands that block or promote the twisting transition, thus modulating the activity of pLGICs.

To tackle this question we proposed an approach based on free energy calculations. Indeed, brut force MD is very informative but also very expensive in computer time. Free energy calculation based on enhanced sampling techniques such as US allow one to significantly decrease the cost of studying such complex biological mechanisms We performed PMF calculations along the twist for two different systems; i) GluCl bound to IVM ii) GluCl with IVM removed.

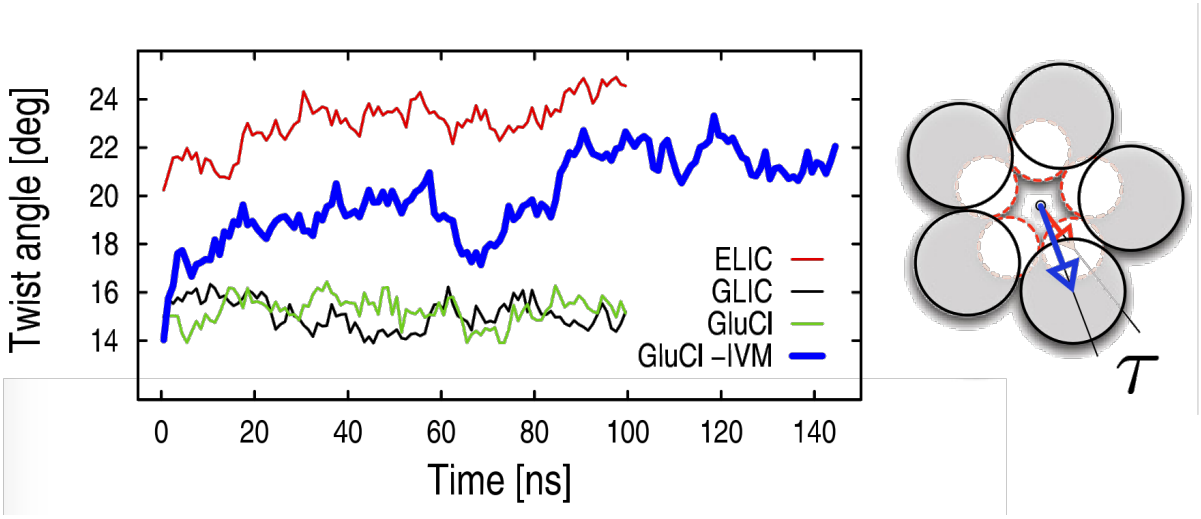


Figure 15: evolution of the twist in function of the time and its schematic representation for a simulation of GluCl with IVM removed (in blue), GluCl with IVM (in green) and for other pLGICs. (See Calimet et al. [54] for a complete description)

3.2.1. Presentation of the simulated system

3.2.1.1. GluCl

There exist no high-resolution structures of human pLGICs in the most commonly used proteins databases. Therefore, one has to tackle to rely on non-human structures or build-up a model predicting the structure [68] of human receptor. Considering the fact that these receptors are highly conserved among eukaryotes, we can study structures from other organisms with good confidence that the conclusions made will be applicable to the human homologues. There exist several structures of pLGICs from prokaryotes and only a few from eukaryotes. Among the latter, one can cite the structure of GluCl in complex with the allosteric agonist IVM and the endogenous neurotransmitter L-glutamate, which was published in 2011 [5] (PDBID 3RIF). A recently published structure of the *apo* state of GluCl [6] came to enrich the knowledge of the cys-loops receptors.

Unlike most human pLGICs, GluCl is homopentameric. Five identical subunits are organized around a pore, selective to chloride ions. In its wild structure, it can be either activated by the presence of L-glutamate alone or the binding of IVM can potentiate this activation.

X-ray crystallography produces structures of protein with no information about the hydrogens, neither about the membrane, thus, the modeller has to predict the position of the hydrogens and of the bilayer membrane. The protein studied was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane and explicit water molecules and ions were added. The system consisted in a total of about 200 000 atoms.

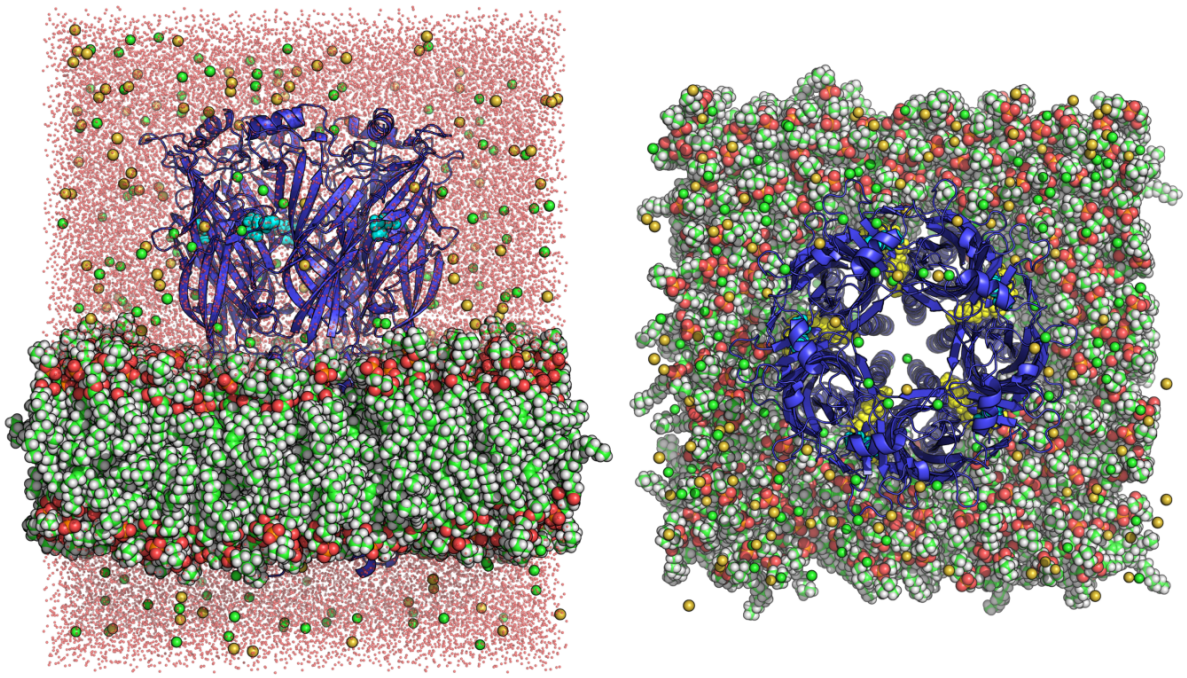


Figure 16: GluCl embedded in a POPC bilayer. (Side view on the right hand side and top view on the left hand side)

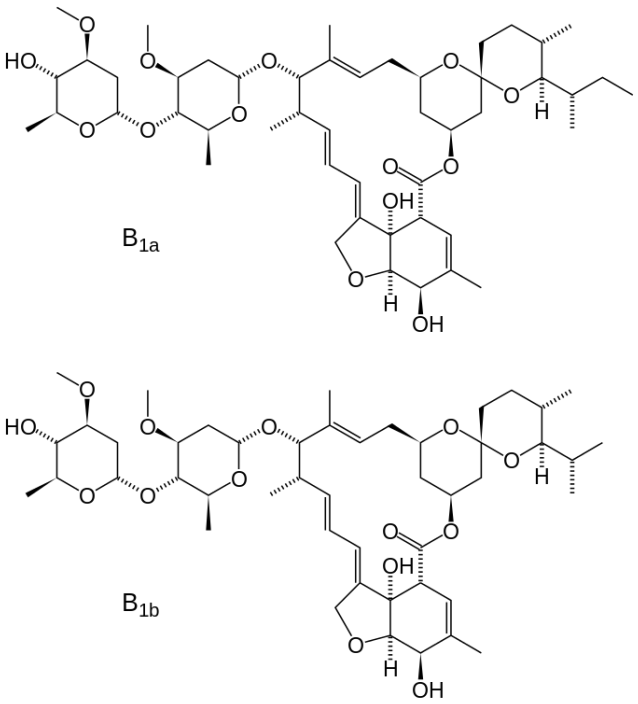
3.2.1.2. Ivermectin

As mentioned in previous sections, GluCl could be crystallized in the presence and in the absence of IVM.

Ivermectin belong to the avermectin family. It is known to be a broad spectrum anti parasitic drug commercialized under the name of STROMECTOL®. IVM is used for treating worm infestations or skin diseases such as scabies in both humans and animals. Its production is based on the culture of bacteria called *streptomyces avermectilis*.

In parasites, its activity is due to the binding of IVM to receptors located in the central nervous system and on the muscles. It is lethal by enhancing the inhibitory neurotransmission of the living organisms.

In humans, at therapeutic doses, only a small fraction of the drug crosses the blood brain barrier and is responsible for the majority of the side effects by interacting with receptors of the pLGICs family ($GABA_AR$, GlyR, nAchR etc…) and leads to CNS disorders.



Equation 12: 2D formula of the two conformers of ivermectin.

3.2.2. Preparing the system for running molecular dynamic simulations

3.2.2.1. pKa calculations

The available structures of proteins in the Protein Data Bank (PDB) are not suitable for running molecular dynamics since they do not include hydrogens. Indeed, the resolution of current X-ray experiments is not sufficient to allow their detection for this type of proteins. For this reason, one ignores the protonation states of the peculiar residues of the protein. Their exist amino acids called “titrable residues” that are sensitive to changes in the pH of the media. This group contains all amino acids that have a side chain charged or carrying an acidic/basic moiety, and the N and C termini.

One can find in the literature the pKa of all amino acids in water. Nevertheless, in the context of a protein, part of the residues can be buried inside the protein and not exposed to the bulk. In this case, the aforementioned amino acids can have an exotic protonation state. Moreover the proximity of charged AA could influence the protonation state of the surrounding AA. The challenge of pKa calculations is to predict the right protonation state for all the titrable residues.

To tackle this problem one can use the multiple titrable sites approach [69] or simpler methods available through online webservers such as ProPKA. (also implemented in the visualization software Pymol)

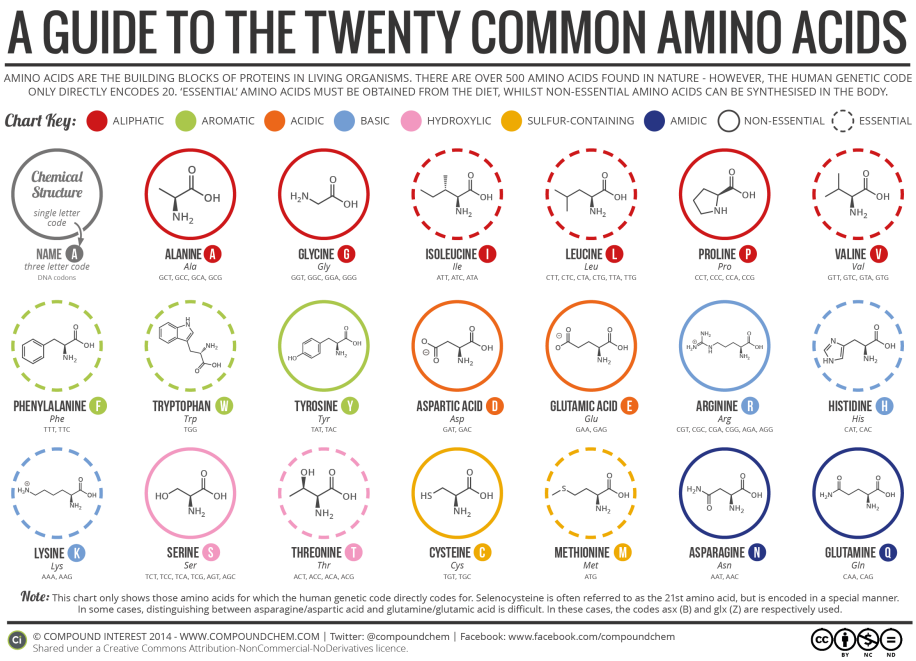


Figure 17: list of the common amino acids found in proteins stressing their chemical properties.

3.2.2.2. Build of the model

The available structures of proteins in the Protein Data Bank (PDB) are imperfect. They do neither include hydrogens since the resolution of current X-ray experiments is not sufficient to detect them, neither to see the solvent nor the membrane in which in embedded the protein. All these components have to be carefully reconstructed by the modeller.

In computer simulations of biomolecules, solvent can be treated implicitly or explicitly. The latter implies that water molecules or more exotic solvents are replaced by a continuous medium. Even though the calculations are then less accurate they allow for a drastic decrease of the computational cost and thus to reach much longer timescales. In the second model, actual water molecules are present during the simulation. To avoid the considerable increase in the complexity of the system due the great number of water molecules present, a simplified model of water molecules (TIP3 [70]), in which some of the internal degrees of freedom of the water molecules are fixed, was used.

As described before, the most probable protonation states of the titrable residues were predicted and introduced of the model. The structure then produced was embedded in a lipid bilayer membrane composed of POPC molecules. The initial configuration of the lipid bilayer was taken from an atomistic model pre-equilibrated in the NPT ensemble (constant pressure, number of mole and temperature) at 303K, i.e. physiological temperature of 25 degrees and using the CHARMM36 [71] force field for the lipids. This model, too small to our

purpose, was replicated before the insertion of our protein model. The force field CHARMM27 [72] was used for the protein and the parameters from CGenFF [63,73] for IVM when it was not removed.

To allow the use of periodic boundary conditions (PBC) an orthorombique box of 120x120x140Å centred on the protein was created.

Sodium and chloride counter ions were added to the system to compensate for the charge of the protein then an excess of ion was also added to reach the physiological reference concentration of 150mM NaCl.

Finally, the constructed system of GluCl with IVM removed consisted of 199,925 atoms (with 41,957 water molecules, 347 lipids, and 118 Na^+ and 138 Cl^- ions).

3.2.2.3. Minimization

As previously mentioned, tridemensionnal structures from obtained with X-ray crystallography are imperfect, thus one then needs to perform energy minimization, heating and equilibration to prepare the system for molecular dynamics simulations.

The energy minimization is the very first step after the building of the model. During the previous stages of the process, atoms were placed regardless to their neighbours and clashes, i.e. overlapping of two atomic positions in space, might have been be created. It is also called geometry optimization and will cause the system to fall in a local minimum of energy after several steps, relaxing the possible restrains induced by the previous steps. Energy minimization can be performed using different methods such as steepest decent (SD) or conjugate gradient (CG).

3.2.2.4. Heating

The second step is the heating or thermalization. In order to simulate meaningful motion of atoms, one needs to set an adequate temperature for the system. The energy minimized structures, in which no notion of temperature or motion is present, needs to be gradually heated to a target temperature. The latter is regulated all along the simulation to remain constant by coupling the system to an external heat bath [74] fixed at the desired temperature, which allows exchanges of heat.

This step is performed under position restrains on the backbone of the protein in order to allow only side chain movements and prevent from creating artefacts in the structure during the process.

3.2.2.5. Equilibration

The final step is the equilibration. So far the system was restrained to a starting position, thus to run a free molecular dynamic simulation one needs to gradually remove these restrains. In the protocol used, we gradually remove the backbone restrains to allow a smooth relaxation of the full system and then run several cycles without restrains to let the system equilibrate at the right temperature and with no restrains, making for a 2 ns long equilibration.

This step is crucial because it fills the gaps artificially created by the building process. Water molecules diffuse and start interacting with specific residues on the protein forming hydrogens bonds, lipids inside the membrane reorganise around the protein and side chains and loops of the protein are allow to rotate to find their minimal energy position.

3.2.3. Reaction coordinate

The activation or deactivation of a receptor implies structural changes in the 3D structure of the protein. These changes occur at different levels. The modification of an external stimulus such as the modification of the pH, the stop of the exposure to light or the unbinding of a neurotransmitter to his receptor, leads to atomics scale changes. They are then propagated to the whole protein and change the global 3D structure. It is often possible to decompose theses changes in several simple transformations such as tilting or twisting (see Figure 10).

In the case of GluCl it was shown [54] that the deactivation of the receptor would mainly show a twist of the extracellular domain over the transmembrane domain. Later on, based on the observation of the crystallographic structure of GLIC [75] and unpublished work, a two-step mechanism was proposed [43] involving the blooming and twisting motions.

The reaction coordinate chosen to be explored in this study was the twisting. We reproduced the values of twist observed by Calimet et al. in 2013, i.e. a variation of 10 degrees of twist angle, from 12 to 22 degrees.

3.2.3.1. Geometrical definition of the twist angle

The twist angle as defined previously, can be represented using two different approaches (see Figure 18 to confront the two definitions of the twist):

- through one dihedral angle per subunit (as shown on the left hand side of Figure 18). Indeed, a dihedral angle is composed of 4 points, a, b, c and d. In the case of our system, a is the centre of mass of the core residues of the EC

domain of a given subunit, b its projection on the axis defining the ion pore, d the centre of mass of the core residues of the TM domain of the same subunit and c its projection on the pore axis. The twist of the protein is expressed as the average over the five subunits.

- Through a global angle of rotation of the EC domain to a reference conformation, when the two structures are superimposed on their TM domains. Unlike the first description here we define a relative measure.

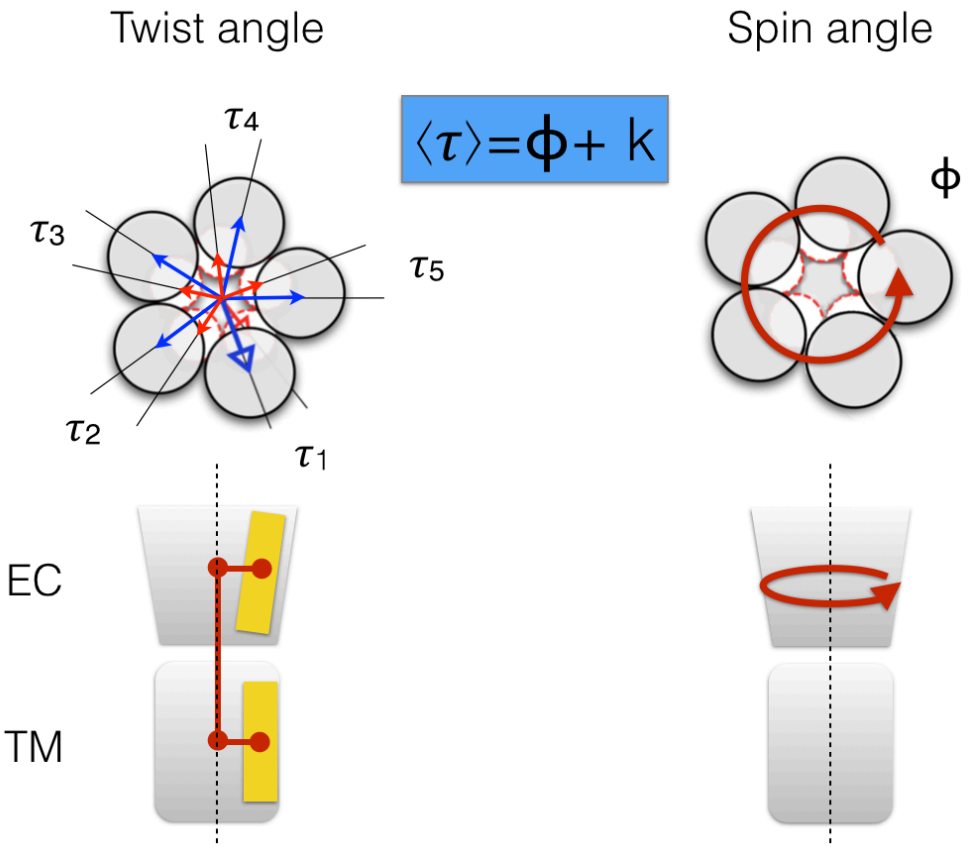


Figure 18: Scheme showing the differences between the twist defined as a dihedral (left hand side) and as a global rotation angle, i.e. spinangle (right hand side) using a pentameric receptor as an example.

3.2.3.2. Restraining the twist angle

Once the system was built and ready to be used for running restrained simulations and once we made the decision to study the twisting component of the gating, we had to design an experiment to do so. All simulations were run using NAMD2.9 and the collective variable tools [76]. In the case of GluCl it is not possible to restrain and monitor the twist motion along the simulations with a simple dihedral due to the way this restrain was implemented in the

molecular dynamics software used. We had to explore more complicated approach among which the spinangle collective variable, which corresponds to the second definition given in section 3.2.3.1 and which seemed to be adequate to describe the twisting of the receptor.

In order to use a spinangle instead of a twist angle we had to show that both quantity were strongly correlated or identical and that the energetics would not be biased by the change of the angle used. To allow the comparison between twist and spin angles, we had to find a system where both quantities could be monitored and restrained.

3.2.4. Method validation on a toy system: the butane molecule

3.2.4.1. Presentation of the protocol

When studying a simpler system such as the butane molecule, the twist angle of interest is the dihedral angle χ defined by the atoms 1,2,3,4 as shown on the following figure.

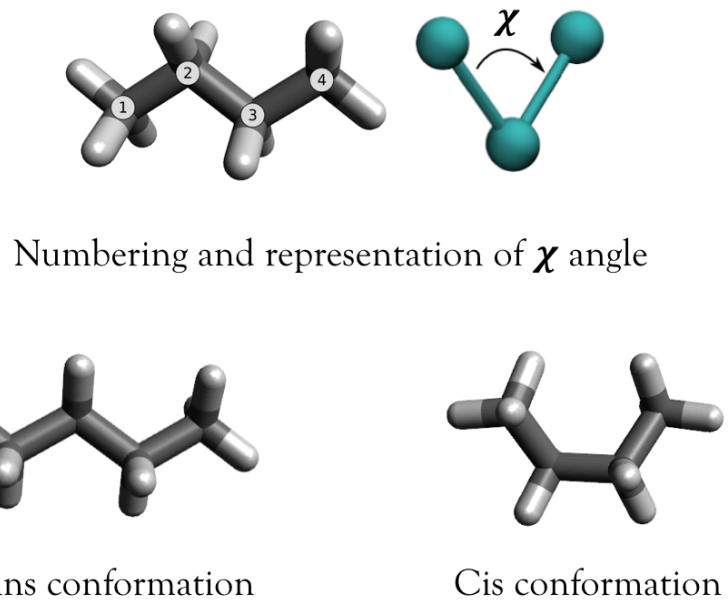


Figure 19: The spinangle is the rotation of atom 4 around the axis defined by atom 2 and 3. Twist and spin angle here corresponds to the exact same motion.

To validate the use of the spinangle tool we proposed to study the energetics of the rotation of the angle χ from -180 to 180 degrees and compute various quantities. First we wanted to show that while doing this transformation we obtained similar PMF profiles using both dihedral and spinangle restrains. Second that the two methods led to the same $\Delta F_{cis/trans}$, which is defined as follow:

$$\Delta F_{cis/trans} = -k_B\, T \ln\left(\frac{\sum_{i\in cis} p_i}{\sum_{j\in trans} p_j}\right)$$

Equation 13: Definition of the free energy difference between the two possible conformations of the butane molecule

where k_b is the Boltzmann constant, T the temperature, p_i is the unbiased probability of the ith bin belonging to the *cis* conformer and p_j is the unbiased probability of the jth bin belonging to the *trans* conformer.

We defined a *cis* conformation as any conformation with $\chi \in [-60:60]$ and *trans* as $\chi \in [-180:-60[\ or\ \chi \in\ ]60:180]$.

3.2.4.2. Results

Umbrella Sampling simulations were run in order to increase the sampling for high-energy conformations all along the reaction coordinate, i.e. χ angle, using both the dihedral and the spinangle approaches. We could then compute the PMF along the rotation for both methods, i.e., dihedral and spinangle restrains. As shown on Figure 20 the two profiles are extremely similar showing a global energy minimum for an angle of +/- 180 corresponding to a *trans* conformations and a local minimum around +/- 60 degrees corresponding to *cis* conformations.

Moreover while computing $\Delta F_{cis/trans}$ a very small difference of 0.0008 kcal/mol was observed showing that both methods can retrain and describe the twist angle in similar if not identical way.

Finally re-computing the twist and spin angles on the simulation trajectories of all windows allowed us to stress the obvious correlation between the two aforementioned quantities as shown on Figure 21 leading to the conclusion that one can use the spinangle to elucidate the energetic of the twist in the gating mechanism of pLGICs.

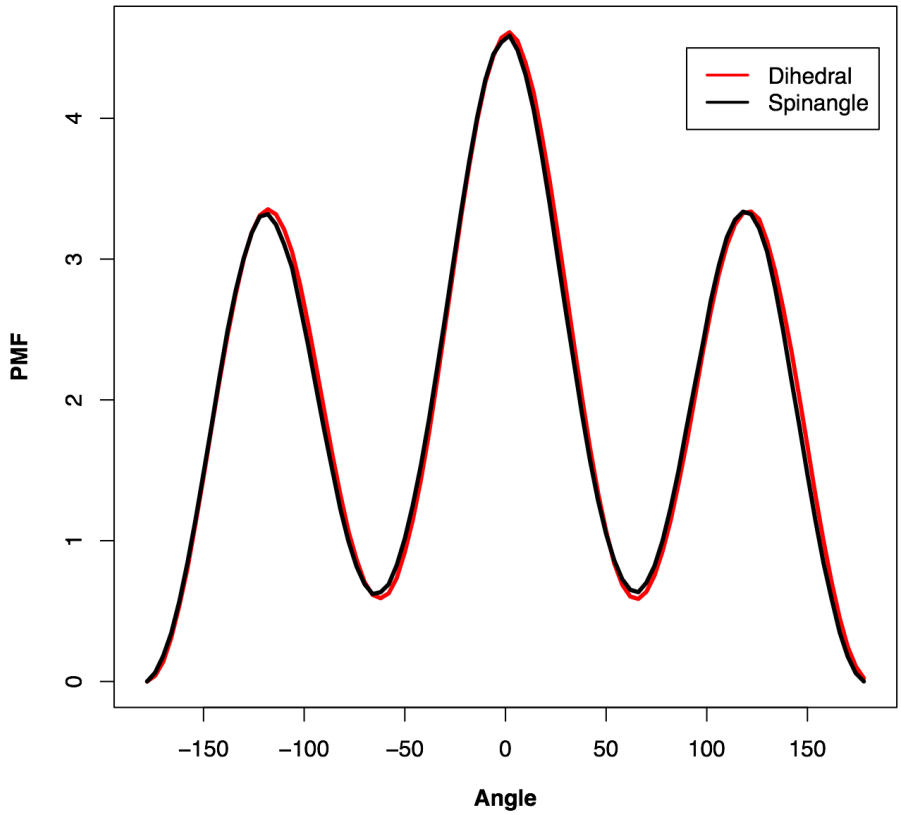


Figure 20: Potential of mean force for the full rotation of the dihedral angle of the butane computed restraining the dihedral angle and the spinangle.

Table V: results of the free energy difference between *cis* and *trans* conformations using two different methods.

Method	$\Delta F_{cis/trans}$ (Kcal/mol)
Spinangle	0.1794
Dihedral	0.1802

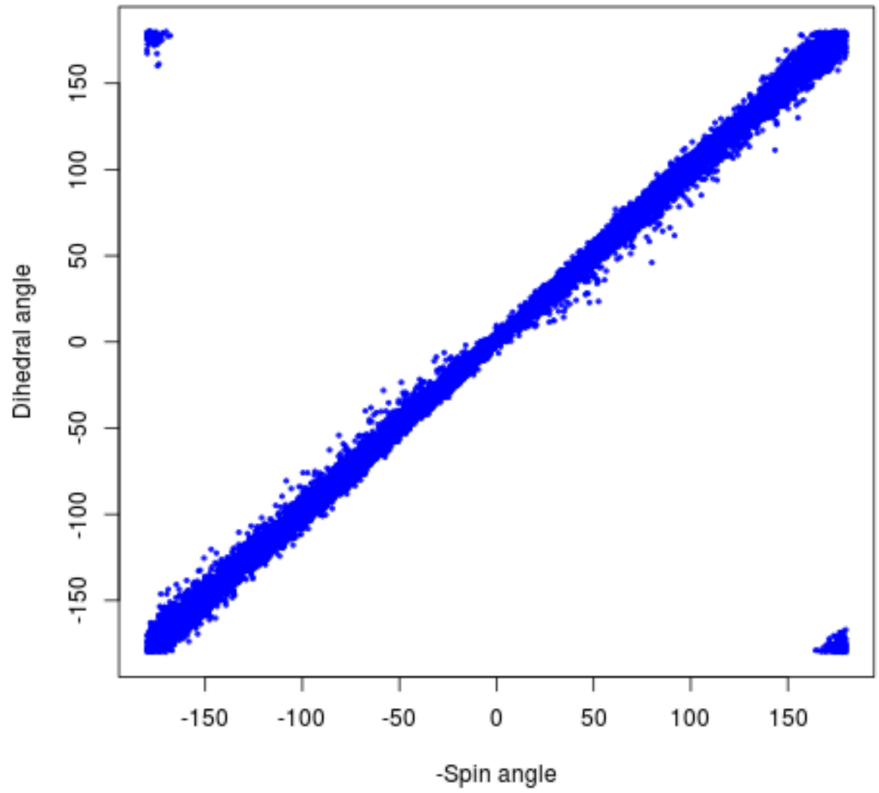


Figure 21: correlation between dihedral or twist angle and the spinangle for a set of restrained simulations for the butane molecule.

3.2.5. Correlation between spinangle and twist angle in the case of GluCl

As previously done for the butane molecule, twist and spin angles for a free simulation of GluCl with IVM removed were computed and their correlation plotted. One can clearly see from Figure 22 the correlation between the spin and the twist angles as it was shown before for the butane molecule (see Figure 21). The former being a relative measure of the rotation and the latter an absolute measure, one has to find the mathematical relation between the two. This can be done either by looking at the twist of the crystal structure of GluCl (used as a reference for computing the value of the spinangle) or extrapolated from the correlation plot. The relation between the two quantities is given by the formula:

$$\tau = \varphi + 12.1$$

Equation 14: correlation between twist and spinangle.

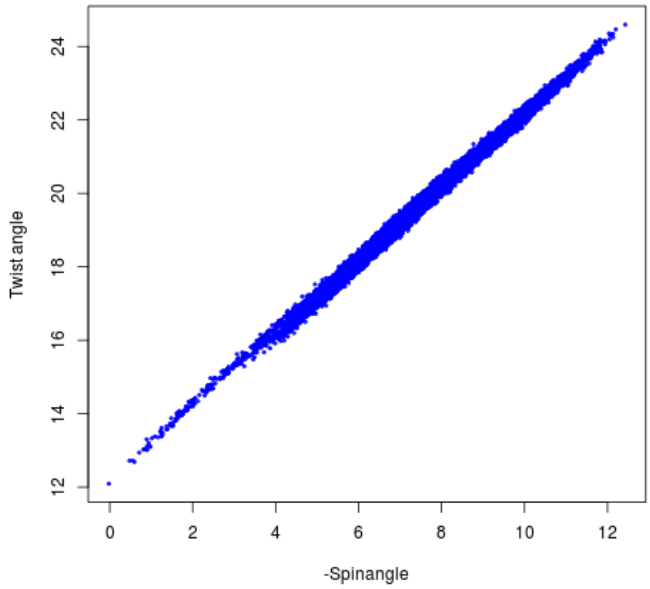


Figure 22: correlation between twist and spinangle for GluCl

3.2.6. PMF in the presence and the absence of IVM

3.2.6.1. Protocol

The delta of twist angle studied was 10 degrees as suggested by the transition from open to close channel observed by Calimet et al [54] in free simulations. The RC was divided in 21 windows spaced of 0.5 degrees. Two systems were studied: i) GluCl bound to L-glutamate and IVM and ii) GluCl with IVM removed.

The starting structures for each window were extracted from free MD and were chosen so the twist angle of the extracted structure would be as close as possible to the centre of the considered window. Thus all windows were started with a different initial structure. On the other hand, one could start with the crystal structure for all windows but the force applied to pull the protein to a specific value of twist angle would be very high in the first steps of the simulation and could create strain energy leading to non-physical configurations of the receptor.

Each windows was run for 1ns first then all simulations were extended to 10ns to check for convergence of the PMF computed. No heating or equilibration were run since the starting structures were extracted from a previous MD run.

The errors of the PMF were estimated using a Monte Carlo bootstrapping [77] method implemented in the WHAM software. Bootstrapping consists in resampling a given distribution several times to infer on the errors.

3.2.6.2. Umbrella sampling validation

Umbrella sampling protocol for exploring a RC implies that one has to divide that RC in several portions or windows. Moreover, in order to compute an accurate PMF, the sampling within each window should be long enough to allow for a sufficient overlap of the probability distribution of the quantity restrained between adjacent windows.

To assess for the quality of the sampling of all windows it is useful to compute the distribution of the restrained quantity per window, as represented on Figure 23. Ideally, and due to the harmonic bias used in US, one should observe Gaussian distributions around the centre of the windows and overall a homogeneous sampling of the full reaction coordinate as illustrated on Figure 23. In case some portion of the RC would not be explored it might be necessary to add windows *a posteriori* to cover the full RC space.

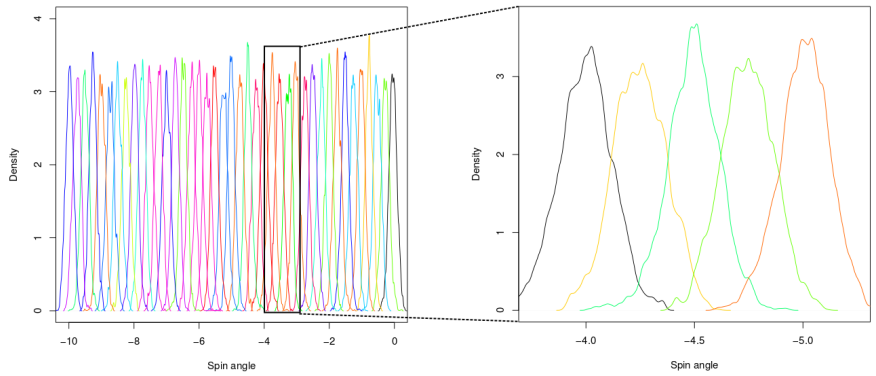


Figure 23: density distribution of the spinangle along the different windows for the butane molecule

3.2.6.3. Results

Electrophysiological studies have shown that when bound to IVM the overall current through the cell's membrane was increased [78] compared to when only the orthosteric ligand was bound. In fact, $GluCl_{\alpha}$ (crystallographic construct) seems to exhibit a peculiar behaviour compared to the other members of the pLGICs and to the physiological construct. Indeed, it is not activated by the binding of only L-glutamate but requires the prior addition of IVM to be activated unlike other pLGICs that are activated by their orthosteric ligands and potentiated by allosteric ligands such as IVM.

Free MD as long as X-ray studies have shown that the gating mechanism of the pLGICs involved a twisting of the EC domain over the TM domain and that IVM could be capable of blocking the twisting of the receptor. Thus one may want to study the free energy of the two systems i.e., GluCl bound to IVM and L-glutamate and GluCl without IVM to understand the role of IVM on the twisting.

When IVM is removed from its binding site, we can see that the global energy minimum is located around 16 to 18 degrees of twist (Figure 15, blue curve). On the other hand, when IVM remain bound (Figure 15, red curve), a global minimum can be observed around 14 degrees of twist. Thus the binding of IVM tends to stabilize structures with a lower value of the twist angle. Interestingly, if one look more careful at the PMF in the absence of IVM, one can identify a local minimum around 14 degrees that seems to correspond to the on observed in the PMF of GluCl bound to IVM. This leads to the conclusion that IVM could select this conformations of the receptor and stabilizes it as well as it could destabilizes structures with higher values of the twist as shown by the very fast increase of the free energy as twist increases on the red curve. Moreover it is interesting to see that when IVM is bound, the fluctuations of the twist are much smaller than when IVM is not bound showing that IVM enthalpically stabilizes the receptor.

It is also interesting to stress that the results of free MD (see Figure 15) and of free energy calculations both predict the same twist angle for the equilibrium structure of GluCl bound to IVM, i.e., 14 degrees. Surprisingly, free MD and free energy calculations are in contradiction when looking at GluCl without IVM. The former suggested a value of twist at equilibrium of 23 degrees when the latter suggests 16 to 18 degrees. Unlike in the case of GluCl with IVM bound our method base on free energy calculation fails at justifying energetically the equilibrium values observed in free MD.

The shift observed between the values of twist angle in free MD and the one predicted by free energy calculations could be due to several issues. Umbrella sampling allows to sample accurately along a given RC but orthogonal degrees of freedom are integrated out. Indeed, if the sampling time is not long enough to account for these degrees of freedom then, one might explore an irrelevant path. Recent studies have shown the crucial interest of the blooming RC to describe the gating mechanism. In fact, the RC was never taken into account during the discussed free energy calculation and could very well explain the shift observed in the PMF.

Overall, even though the location of the minima when IVM is removed is not coherent with free simulations, this experiments allowed discriminating between an IVM-bound and a non-bound state opening the door for further application such as designing new ligands to inhibit or enhance the activity of these family of receptors.

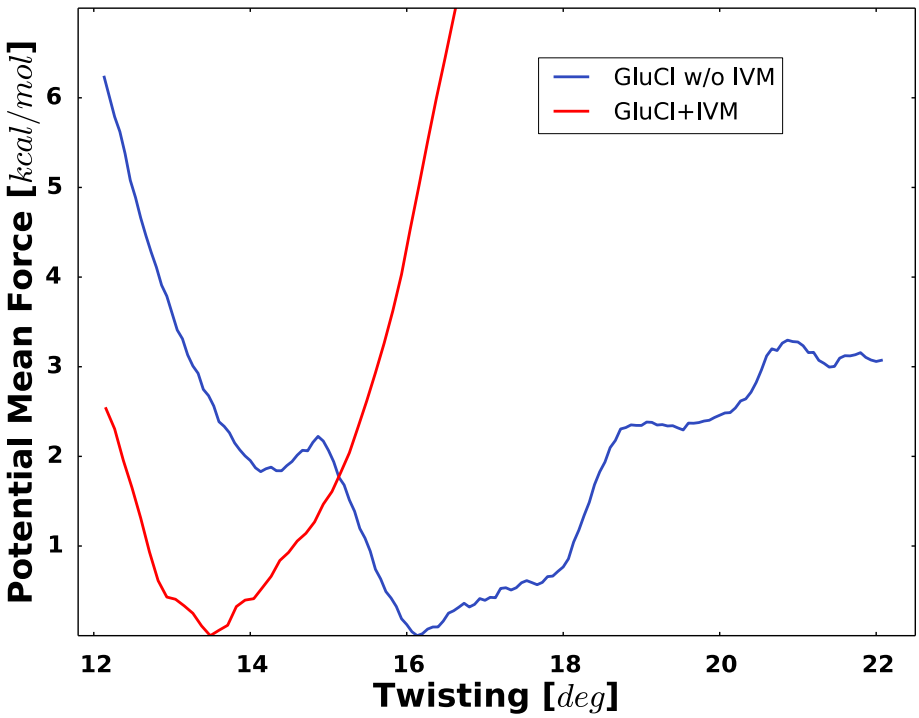


Figure 24: PMF of the twisting for two systems. In blue, GluCl with IVM removed and in red GluCl bound to IVM

Conclusion

Pentameric ligand gated ion channels are found as well as in bacteria as in the human brain and they are the targets of numerous treatments, with applications ranging from anti-parasites to Alzheimer disease. Although more and more crystallographic and cryo-electron microscopy data are available, their gating mechanism remains partially unknown. Studies have shown that they can be activated by specific endogenous neurotransmitters at the level of their orthosteric site (EC domain); however they also possess the very interesting capability of being modulated by the binding allosteric ligands a distinct binding site. For instance, ivermectin, a well known anti-helminthic drug, acts as a positive allosteric modulator of most of the pLGICs and its precise effect on the dynamics of the receptor is still not fully understood.

Advanced computer simulations techniques have been developed during the past decades in response to the increase of computers' power, allowing for the study of always bigger and more realistic systems. During the past fifty years, the field of computational chemistry has moved from studying di-atomic simple systems in vacuum to transmembrane protein embedded in lipid bilayer membrane and fully solvated by explicit water molecules. Among other computational science, structural biology has greatly gained benefits from the increase of computer power and it contributed to unravel complex biological processes.

In this work we have shown how a computational approach can lead to a better understanding of a biological system. We have developed a methodology in order to perform umbrella-sampling simulations on a full protein. Using the aforementioned method, we have performed free energy calculation along a reaction coordinate i.e., the twist, which was shown to be biologically relevant since it is correlated with the closing of the ion pore. This allowed us to study the effect of IVM on an eukaryotic channel, GluCl and to draw several conclusions. We have shown that IVM blocks the twist and thus the correlated closing of the pore by stabilizing an untwisted configuration (14 degrees of twist angle), when in the absence of IVM the behaviour was completely different. Thus, we have developed and applied a method capable of discrimination between an IVM-bound and a non-bound state, conclusion that could be extrapolated to discriminate any other type of ligand acting on the twist, even at a topographically distinct active site. Indeed, this method is capable of concluding on a ligand's action and not only on its capability of binding, unlike, for instance, docking. We believe that this method can help in designing *de novo* new allosteric modulators, which could potentially be synthetized and tested experimentally. Moreover, only very few or no treatments are nowadays available for helping patient with Parkinson or

Alzheimer disease and increasing the knowledge of the mechanisms of action and modulation of receptors such as $GABA_A$, nAChR or $5HT_3$ is a first and promising step toward the discovery of novel treatments.

Références bibliographiques

1. Langley, J. N. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *J. Physiol.* **33,** 374–413 (1905).
2. Changeux, J. P., Kasai, M. & Lee, C. Y. Use of a snake venom toxin to characterize the cholinergic receptor protein. *Proc. Natl. Acad. Sci. U. S. A.* **67,** 1241–1247 (1970).
3. Bocquet, N. *et al.* X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* **457,** 111–114 (2008).
4. Hilf, R. J. C. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457,** 115–8 (2009).
5. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective {Cys}-loop receptor. *Nature* **474,** 54–60 (2011).
6. Althoff, T., Hibbs, R. E., Banerjee, S. & Gouaux, E. X-ray structures of GluCl in apo states reveal a gating mechanism of Cys-loop receptors. *Nature* **512,** 333–337 (2014).
7. Corringer, P. J. *et al.* Structure and pharmacology of pentameric receptor channels: From bacteria to brain. *Structure* **20,** 941–956 (2012).
8. Nys, M., Kesters, D. & Ulens, C. Structural insights into Cys-loop receptor function and ligand recognition. *Biochem. Pharmacol.* **86,** 1042–1053 (2013).
9. Miller, P. S. & Aricescu, A. R. Crystal structure of a human GABAA receptor. *Nature* **512,** 270–5 (2014).
10. Albuquerque, E. X., Pereira, E. F. R., Alkondon, M. & Rogers, S. W. Mammalian Nicotinic Acetylcholine Receptors: From Structure to Function. **89,** 73–120 (2009).
11. Hilf, R. J. C. & Dutzler, R. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. **452,** 375–380 (2008).
12. Du, J., Lü, W., Wu, S., Cheng, Y. & Gouaux, E. Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **526,** 224–9 (2015).
13. Zoli, M., Pistillo, F. & Gotti, C. Diversity of native nicotinic receptor subtypes in mammalian brain. *Neuropharmacology* **96,** 302–311 (2014).
14. Jaiteh, M., Taly, A. & Hénin, J. Evolution of Pentameric Ligand-Gated Ion Channels: Pro-Loop Receptors. *PLoS One* **11,** e0151934 (2016).
15. Mulle, C., Lena, C. & Changeux, J. P. Potentiation of nicotinic receptor response by external calcium in rat central neurons. *Neuron* **8,** 937–945 (1992).
16. Vernino, S., Amador, M., Luetje, C. W., Patrick, J. & Dani, J. A. Calcium modulation and high calcium permeability of neuronal nicotinic acetylcholine receptors. *Neuron* **8,** 127–134 (1992).
17. Taly, A., Corringer, P.-J., Guedin, D., Lestage, P. & Changeux, J.-P. Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system. *Nat. Rev. Drug Discov.* **8,** 733–750 (2009).
18. Forman, S. a., Chiara, D. C. & Miller, K. W. Anesthetics target interfacial transmembrane sites in nicotinic acetylcholine receptors. *Neuropharmacology* **96,** 169–177 (2014).
19. Lobo, I. A. & Harris, R. A. Sites of Alcohol and Volatile Anesthetic Action on Glycine Receptors. *International Review of Neurobiology* **65,** 53–87 (2005).
20. Mihic, S. J. *et al.* Sites of alcohol and volatile anaesthetic action on GABA(A) and glycine receptors. *Nature* **389,** 385–389 (1997).
21. daCosta, C. J. B., Free, C. R., Corradi, J., Bouzat, C. & Sine, S. M. Single-Channel

and Structural Foundations of Neuronal 7 Acetylcholine Receptor Potentiation. *Journal of Neuroscience* **31,** 13870–13879 (2011).

22. Bertrand, D. & Gopalakrishnan, M. Allosteric modulation of nicotinic acetylcholine receptors. *Biochem. Pharmacol.* **74,** 1155–1163 (2007).
23. Nury, H. *et al.* X-ray structures of general anaesthetics bound to a pentameric ligand-gated ion channel. *Nature* **469,** 428–431 (2011).
24. Miyazawa, A., Fujiyoshi, Y. & Unwin, N. Structure and gating mechanism of the acetylcholine receptor pore. *Nature* **423,** 949–955 (2003).
25. White, B. H. & Cohen, J. B. Agonist-induced changes in the structure of the acetylcholine receptor M2 regions revealed by photoincorporation of an uncharged nicotinic noncompetitive antagonist. *J. Biol. Chem.* **267,** 15770–83 (1992).
26. Arevalo, E., Chiara, D. C., Forman, S. a, Cohen, J. B. & Miller, K. W. Gating-enhanced accessibility of hydrophobic sites within the transmembrane region of the nicotinic acetylcholine receptor's {delta}-subunit. A time-resolved photolabeling study. *J. Biol. Chem.* **280,** 13631–40 (2005).
27. Banks, M. I., White, J. a & Pearce, R. a. Interactions between distinct GABA(A) circuits in hippocampus. *Neuron* **25,** 449–457 (2000).
28. Chang, Y., Wang, R., Barot, S. & Weiss, D. S. Stoichiometry of a recombinant GABAA receptor. *J. Neurosci.* **16,** 5415–5424 (1996).
29. Labarca, C. *et al.* Channel gating governed symmetrically by conserved leucine residues in the M2 domain of nicotinic receptors. *Nature* **376,** 514–6 (1995).
30. Filatov, G. N. & White, M. M. The role of conserved leucines in the M2 domain of the acetylcholine receptor in channel gating. *Mol. Pharmacol.* **48,** 379–84 (1995).
31. Hilf, R. J. C. *et al.* Structural basis of open channel block in a prokaryotic pentameric ligand-gated ion channel. *Nat. Struct. Mol. Biol.* **17,** 1330–U184 (2010).
32. Hassaine, G. *et al.* X-ray structure of the mouse serotonin 5-HT3 receptor. *Nature* **512,** 276–281 (2014).
33. Hassaine, G. *et al.* X-ray structure of the mouse serotonin 5-HT3 receptor. *Nature* **512,** 276–281 (2014).
34. Tsai, C. J. & Nussinov, R. A Unified View of 'How Allostery Works'. *PLoS Comput. Biol.* **10,** (2014).
35. JACOB, F. & MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3,** 318–356 (1961).
36. Changeux, J.-P. The Feedback Control Mechanism of Biosynthetic L-Threonine Deaminase by L-Isoleucine. *Cold Spring Harb. Symp. Quant. Biol.* **26,** 313–318 (1961).
37. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12,** 88–118 (1965).
38. Pauling, L. The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation. *Proc. Natl. Acad. Sci. U. S. A.* **21,** 186–91 (1935).
39. Koshland, D. E., Némethy, G. & Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* **5,** 365–385 (1966).
40. Sauguet, L. *et al.* Crystal structures of a pentameric ligand-gated ion channel provide a mechanism for activation. *Proc. Natl. Acad. Sci.* **111,** 966–971 (2014).
41. Hibbs, R. E. & Gouaux, E. Principles of activation and permeation in an anion-selective {Cys}-loop receptor. *Nature* **474,** 54–60 (2011).

42. Taly, A. *et al.* Normal mode analysis suggests a quaternary twist model for the nicotinic receptor gating mechanism. *Biophys. J.* **88,** 3954–65 (2005).

43. Cecchini, M. & Changeux, J.-P. The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation. *Neuropharmacology* **96,** 137–149 (2014).

44. Lombardo, S. & Maskos, U. Role of the nicotinic acetylcholine receptor in Alzheimer ' s disease pathology and treatment. *Neuropharmacology* **96,** 255–262 (2015).

45. Giniatullin, R., Nistri, A. & Yakel, J. L. Desensitization of nicotinic ACh receptors : shaping cholinergic signaling. **28,** (2005).

46. Davis, K. L. & Yamamura, H. I. Cholinergic underactivity in human memory disorders. *Life Sci.* **23,** 1729–1733 (1978).

47. Levey, A. I. Muscarinic acetylcholine receptor expression in memory circuits: Implications for treatment of Alzheimer disease. *Proc. Natl. Acad. Sci.* **93,** 13541–13546 (1996).

48. Nordberg, A. & Winblad, B. Reduced number of [3H]nicotine and [3H]acetylcholine binding sites in the frontal cortex of Alzheimer brains. *Neurosci. Lett.* **72,** 115–120 (1986).

49. Smith, M. A., Sayre, L. M., Monnier, V. M. & Perry, G. Radical AGEing in Alzheimer's disease. *Trends Neurosci.* **18,** 172–176 (1995).

50. Whitehouse, P. J. *et al.* Nicotinic acetylcholine binding sites in Alzheimer's disease. *Brain Res.* **371,** 146–151 (1986).

51. Organization, W. H. & others. Tobacco or Health: a global status report. Geneva; 1997. (2009).

52. Moore, G. E. Cramming more components onto integrated circuits. *Proc. IEEE* **86,** 82–85 (1998).

53. Paloncýová, M. *et al.* Benchmarking of Force Fields for Molecule − Membrane Interactions. *J. Chem. Theory Comput.* Articles ASAP (2014). doi:10.1021/ct500419b |

54. Calimet, N. *et al.* A gating mechanism of pentameric ligand-gated ion channels. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E3987–96 (2013).

55. Miao, Y., Johnson, J. E. & Ortoleva, P. J. All-atom multiscale simulation of cowpea chlorotic mottle virus capsid swelling. *J. Phys. Chem. B* **114,** 11181–95 (2010).

56. Born, M. & Heisenberg, W. Zur Quantentheorie der Molekeln. *Ann. Phys.* **379,** 1–31 (1924).

57. Andrew R. Leach. Molecular Modelling: Principles and Applications (2nd Edition).

58. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

59. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18,** 1463–1472 (1997).

60. Jämbeck, J. P. M. & Lyubartsev, A. P. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B* **116,** 3164–3179 (2012).

61. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. a & Case, D. a. Development and testing of a general amber force field. *J. Comput. Chem.* **25,** 1157–1174 (2004).

62. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J.*

Comput. Chem. **25,** 1400–15 (2004).

63. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31,** 671–90 (2010).
64. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12562–12566 (2002).
65. Hamelberg, D., Mongan, J. & McCammon, J. A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **120,** 11919–11929 (2004).
66. Darve, E., Rodríguez-Gómez, D. & Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *J. Chem. Phys.* **128,** 144120 (2008).
67. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13,** 1011–1021 (1992).
68. Baker, D. & Šali, A. Protein structure prediction and structural genomics. *Science (80-.).* **294,** 93–96 (2001).
69. Bashford, D. & Karplus, M. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. *J. Phys. Chem.* **95,** 9556–9561 (1991).
70. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926 (1983).
71. MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102,** 3586–616 (1998).
72. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114,** 7830–7843 (2010).
73. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **33,** 2451–2468 (2012).
74. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, a & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).
75. Sauguet, L. *et al.* Structural basis for ion permeation mechanism in pentameric ligand-gated ion channels. *EMBO J.* **32,** 728–41 (2013).
76. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111,** 3345–3362 (2013).
77. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **7,** 1–26 (1979).
78. Cully, D. F. *et al.* Cloning of an avermectin-sensitive glutamate-gated chloride channel from Caenorhabditis elegans. *Nature* **371,** 707–711 (1994).

Table des figures

Table des tableaux

SERMENT DE GALIEN

Je jure en présence de mes Maîtres de la Faculté et de mes condisciples :

- d'honorer ceux qui m'ont instruit dans les préceptes de mon art et de leur témoigner ma reconnaissance en restant fidèle à leur enseignement ;
- d'exercer, dans l'intérêt de la santé publique, ma profession avec conscience et de respecter non seulement la législation en vigueur, mais aussi les règles de l'honneur, de la probité et du désintéressement ;
- de ne jamais oublier ma responsabilité, mes devoirs envers le malade et sa dignité humaine, de respecter le secret professionnel.

En aucun cas, je ne consentirai à utiliser mes connaissances et mon état pour corrompre les mœurs et favoriser les actes criminels.

Que les hommes m'accordent leur estime si je suis fidèle à mes promesses.

Que je sois couvert d'opprobre et méprisé de mes confrères, si j'y manque.

Nicolas Martin

The use of computer simulations to serve a new approach of drug design

Application to the pentameric ligand gated ion channels family

Résumé :

Dans ce travail sont présentés les protéines de la famille des récepteurs pentamériques canaux parmi lesquels on peut citer, nAChR, GlyR, 5-HT_3,$GABA_A$R ou GluCl. Leurs propriétés pharmacologiques sont présentés et le concept d'allosterie est détaillé. Dans la seconde partie nous nous intéressons à la modulation allosterique d'un homologue eucaryote des pLGICs, GluCl. Enfin nous présentons une méthode basée sur des calculs d'énergie libre qui permet la discrimination entre ligands allosteriques actif ou non actif sur GluCl.

Mots-clés : canaux ioniques pentamériques, pLGICs, allostrerie, dynamique moléculaire.

Abstract :

In this work are presented protein from the pentameric ligand gated ion channels family, among which one may cite, nAChR, GlyR, 5-HT_3,$GABA_A$R ou GluCl. Their pharmacological properties are then discussed and the concept of allostery is introduced. In the second part, we are interested in the allosteric modulation of an eukaryotic homologue of the pLGICs, GluCl. Finally we describe a method based on free enrgy calculation that allows discriminating between active and non-active allosteric binders of GluCl.

Keywords: Pentameric ligand gated ion channels, pLGICs, allostrery, molecular dynamics