Investigation of the domain assembly in ligand-gated and voltage-gated ion channels using computational methods

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
Dr. rer. nat.

der Fakultät für
Biologie
an der

Universität Duisburg-Essen

vorgelegt von
Guido Humpert

aus Düren

Oktober 2016

Die der vorliegenden Arbeit zugrunde liegenden Experimente wurden am Zentrum für Medizinische Biotechnologie (ZMB) in der Abteilung für Bioinformatik und Computational Biophysics der Universität Duisburg-Essen durchgeführt.

1. Gutachter: Prof. Dr. Daniel Hoffmann
2. Gutachter: PD Dr. Jürgen Schlitter

Vorsitzender des Prüfungsausschusses: Prof. Dr. Peter Bayer

Tag der mündlichen Prüfung: 22. Dezember 2016

Zusammenfassung

Ionenkanäle sind essentiell für die Erzeugung und Weiterleitung von elektrischen Signalen im menschlichen Organismus. Sie befinden sich hauptsächling in der Zellmembran von Nerven und Muskelzellen, in denen sie spezifische Aufgaben übernehmen und daher spezielle Eigenschaften entwickelt haben, die diesen Ansprüchen entsprechen. Die Konsequenz daraus ist eine Diversifizierung der Ionenkanal Strukturen, die Aufgrund ihrer Eigenschaften wie z.B. die Ionenselektivität oder die Art der Aktivierung, in Klassen und Familien eingeteilt werden können. Die zentrale Rolle der Ionenkanäle birgt wiederum das Risiko ernsthafter Störungen im Organismus, sobald Fehlfunktionen der Kanäle auftreten. Um dem entgegenwirken zu können, wird intensiv an der strukturellen Aufklärung der Ionenkanäle geforscht. Dies ist mit sehr großen Herausforderungen verbunden und bisher sind nur wenige Ionenkanal Strukturen bekannt.

In dieser Promotionsarbeit wurde die korrekte Anordnung von Ionenkanal Domänen zu einem vollständigen Kanal untersucht, und darüber hinaus die Interaktion zwischen transmembranen Helices innerhalb einer Domäne an einem entsprechenden Beispiel untersucht. Es werden Verfahren beschrieben, die zum einen computerbasierte Modelle von Ionenkanälen auf Basis verschiedenster Informationen erzeugen und zum anderen Rückschlüsse auf etwaige Eigenschaften und Funktionen aus den Modellen ableiten können.

Am Beispiel des 5-HT$_{3A}$ Rezeptors wurde die Assoziation der 4 transmembranen Helices zu einer Domäne analysiert. Als Basis dienten Ergebnisse aus Mutagenese Experimenten, womit einzelne, für die Funktion und Stabilität des Kanals wichtige, Residuen identifiziert wurden und als Kriterien für die Erstellung eines Modells genutzt werden konnten. Mit dem Modell wurden in der Folge Molekulardynamik Simulationen und Berechnungen der Bindungsenergien einzelner Residuen durchgeführt, um den Energiebeitrag der einzelnen Aminosäuren zur Assoziation der Helices zu bestimmen und mit den experimentellen Beobachtungen in Relation zu setzen. Die in der Zwischenzeit publizierte 5-HT$_{3A}$ Struktur diente abschließend zur Verifizierung der Ergebnisse.

Die Fragestellung beim heterotetrameren Na$_v$1.8 Kanal war, welche Anordnung die 4 untereinander verbundenen Domänen einnehmen, um einen funktionierenden Kanal zu bilden. Neben den in der Literatur diskutierten Varianten, gegen und im Uhrzeigersinn, wurden auch Modelle mit 4 weiteren theoretisch möglichen

Varianten mit kreuzweiser Domänenanordnung erzeugt und untersucht. Neben Molekulardynamik Simulationen und Energieberechnungen der Domänen Interfaces, wurde auch die sequenzbasierte Direct Coupling Analyse (DCA) verwendet, wodurch evolutionär miteinander gekoppelte Residuen identifiziert werden können. Mit der Kombination der verschiedenen Ansätze konnte eine klare Tendenz zu einer Anordnung der Domänen im Uhrzeigersinn erkannt werden, die im Einklang mit einer kürzlich veröffentlichten Kalziumkanal Struktur steht.

Contents

List of Abbreviations

5-HT$_{3A}$	5-Hydroxytryptamine receptor 3A
Å	Angstrom
AA	Amino Acids
AP	Action Potential
ATP	Adenosine Triphosphate
CCW	Counter Clock-Wise
cryo-EM	cryo-Electron Microscopy
CW	Clock-Wise
FDA	Food and Drug Administration
FF	Force Field
GPCR	G-Protein coupled receptor
HMM	Hidden Markov Model
IC$_{50}$	Half Maximal Inhibitory Concentration
IVS	Inverse Virtual Screening
LA	Local Anaesthetic
LGIC	Ligand Gated Ion Channels
MD	Molecular Dynamics
MR	Mineralocorticoid receptor
MSA	Multiple Sequence Alignment

PDB	RCSB Protein Data Bank
PME	Particle Mesh Ewald
POPC	1-Palmitoyl-2-Oleoylphosphatidylcholine
PSF	Protein Structure File
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SEA	Similarity Ensemble Approach
TM	Transmembrane Helix
VGCC	Voltage Gated Calcium Channel
VGIC	Voltage Gated Ion Channels
VGSC	Voltage Gated Sodium Channels
VSD	Voltage Sensing Domain

List of Figures

List of Tables

1 Introduction

This thesis describes the work I have done during the years of 2010 through 2016 as a guest (Ph.D.) student at the department of Bioinformatics at the University Duisburg-Essen and as a Discovery Informatics Expert at the Department of Drug Discovery at Gruenenthal GmbH. The first chapter introduces the biological background of excitable cells such as neurons or muscle fibres and describes the concept of signal transduction as well as the structure and mechanisms of the involved proteins. The second chapter explains computational concepts and methods, that were used throughout this study, including some examples of verification experiments. Chapter three and four contain the main experiments and results of this thesis. Finally, chapter five discusses possible follow-up experiments and speculates about chances or ways one could head in the future.

1.1 Specifics of Transmembrane Proteins

1.1.1 The Cell Membrane

The cell membrane, also called plasma membrane, is the barrier that separates the living cell from its surrounding. It controls the transfer of material between the cytoplasm and the extracellular medium, and its origin is probably one of the most important evolutionary steps for the development towards remarkably complex organisms. The first membrane model was proposed by Davson and Danielli in 1935 [44] representing a sandwich-like protein coated phospholipid bilayer. It was refined by Singer and Nicolson introducing their fluid-mosaic model in 1972 [180].

Cell membranes in general contain phospholipids, cholesterol and glycolipids, which present carbohydrates to the exterior for different purposes, such as cell recognition. The polar heads of the phospholipids are facing outwards to the plasma and to the exterior of the cell, and the hydrophobic lipid tails are buried inside the membrane to minimize water exposure. The proportions of these membrane elements vary, depending on the type of cell. Cells existing in a cold environment may require a more fluid bilayer, consisting of a high amount of unsaturated lipids and less cholesterol, compared to cells under warm conditions.

Proteins with various functions are embedded into the phospholipid bilayer or interact with the membrane's surface and trigger signalling cascades or perform transport processes. During evolution these proteins diverged continuously to adapt to their environment, resulting in specialized compartments, like organelles, taking over different functions in the cell.

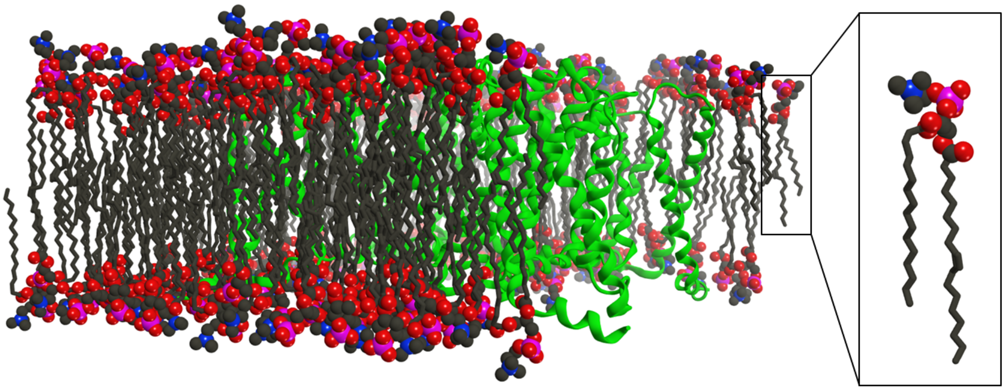


Figure 1.1: Cell membrane with ion channel. This model shows a cell membrane patch with a phospholipid bilayer (POPC) surrounding an Na$_v$1.8 ion channel (green). The polar lipid heads (red, blue) are facing outwards to the plasma and cell exterior, and the hydrophobic lipid tails are buried inside the membrane, forming the bilayer core (gray).

1.1.2 Channels and Transporters

Various types of proteins are embedded in the membrane with plenty of different functions. There are membrane receptors like G-protein-coupled receptors (GPCRs), which act like a molecular switch, passing signals between the internal and external environment of a cell when activated through binding to a G-protein [196]. Other transporters and pores balance the amount of sugars and amino acids or metabolic products between the cell and their compartments. For example, the aquaporins regulate the sometimes rapid movement of water and small uncharged solutes through the cells without affecting the electrochemical potential of the membrane.

One of the most important transporters in the membrane of excitable cells, like muscle fibres or nerve cells, is the sodium-potassium pump [182]. It maintains the membrane potential by shuttling potassium cations into and sodium cations out of the cell in a 2:3 ratio. The sodium-potassium pump is an ATPase enzyme that consumes energy in form of Adenosine triphosphate (ATP), for the transport of sodium and potassium ions in opposite directions, against their concentration gradients. In addition, the plasma membrane is equipped with various ion selective channels, which accelerate or restrain the diffusion of specific ions across the membrane, according to their activation states, like open, closed or inactivated. These ion channels are essential for spontaneous responses to electrical, chemical or physical stimuli that might cause fast reactions of the organism, like defensive reflexes activating spontaneous muscle contractions.

1.1.3 The Membrane Potential

The membrane potential is maintained by passive or facilitated diffusion through channels and by active energy consuming transport of ions as described in the previous section 1.1.2. The typical membrane potential lies within the range of -70 mV and -100 mV, meaning that the cytoplasm is negatively charged compared to the extracellular medium. The selective permeability of the lipid bilayer is a prerequisite for generating a membrane potential. This is essential for the operation of numerous membrane proteins like cotransporters, which use the electric potential as driving force to transport molecules against their concentration gradient across the membrane. For neuronal cells and muscle cells the membrane potential is required to propagate signals.

Accumulating electrical currents passing the membrane and thus causing a local depolarization, can trigger the event of an action potential (AP), if the threshold potential of -45 mV is reached. The consequence is an "all-or-nothing" response, which excitable cells use for signal transduction [12]. The initiation and propagation of an action potential up to the final signal transformation into neurotransmitter release is illustrated by a peripheral nerve fibre in Figure 1.2. This afferent neuronal cell transmits a stimulus from the location of its detection to the spinal cord for further processing.

Once such an AP occurs, Voltage-Gated Ion Channels (VGIC) open due to the change of membrane potential, and the flux of ions across the membrane increases dramatically. The quickly responding voltage-gated sodium channels (VGSC) open first to foster the action potential by Na^+ influx. This leads to further voltage alterations, sensed by neighbouring VGICs that are recruited to accelerate the depolarization. As rapidly as the depolarizing process started, it is also stopped by the fast inactivation of the VGSCs that limits the influx of Na^+ ions. In addition, the opening

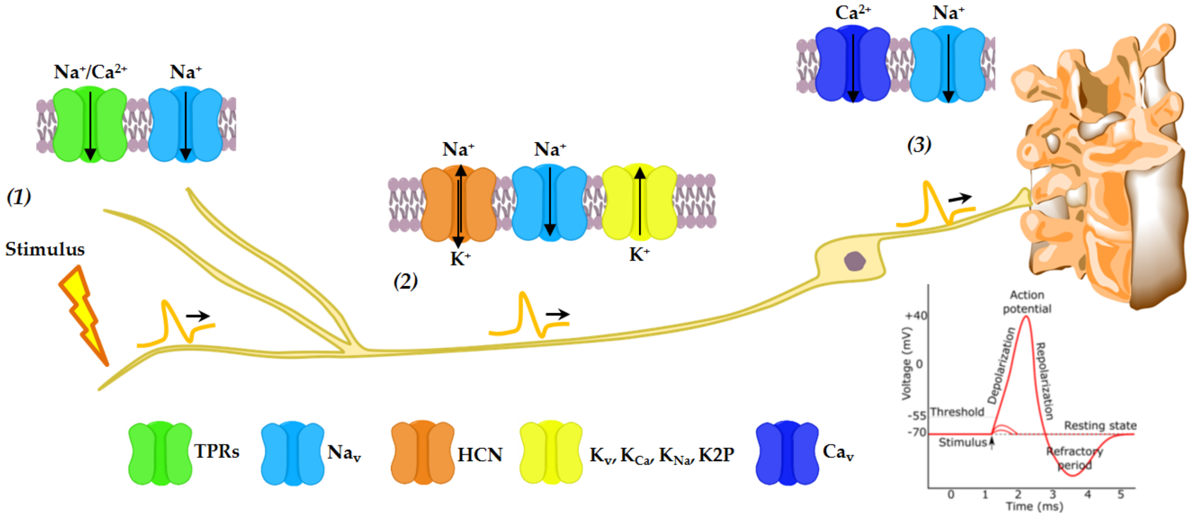


Figure 1.2: Ion channels in afferent neuronal cells. The different ion channel types involved in the signal transduction process are colour coded. (1) The activation of non-selective cation channels and voltage-gated sodium channels in free endings of neurons in response to certain stimuli depolarize the membrane through cation influx. An action potential (AP) is initiated when the depolarization reaches the threshold and additional sodium channels are opened. (2) The AP is propagated by local depolarization and recruitment of neighbouring ion selective channels, which thus are opened, while already activated sodium channels become inactive. Then membrane repolarization is triggered by potassium channels. (3) Finally the translation of AP into neurotransmitter release at the synapse leads to signal transduction to another neuron.

of slow activating voltage-gated potassium channels (VGPC) enables K^+ ions to leave the cell again. These effects help re-polarizing the voltage at the membrane, until the resting potential of -70 mV is reached after a short overrun.

Caused by this chain reaction of opening VGICs, the signal spreads along the neuron and upon reaching the synapse at the axon terminal, induces the release of neurotransmitters. Finally, the adjacent synaptic cleft, a gap between the synapse and the connected cell membrane, is flooded by these messenger molecules, such as serotonin and acetylcholine. They subsequently connect to Ligand-Gated Ion Channels (LGIC) on the membrane surface of the postsynaptic dendrite (Figure 1.3). These channel forming receptors are activated by their specific ligands resulting in a continuous signal transduction from one neuronal cell to another.

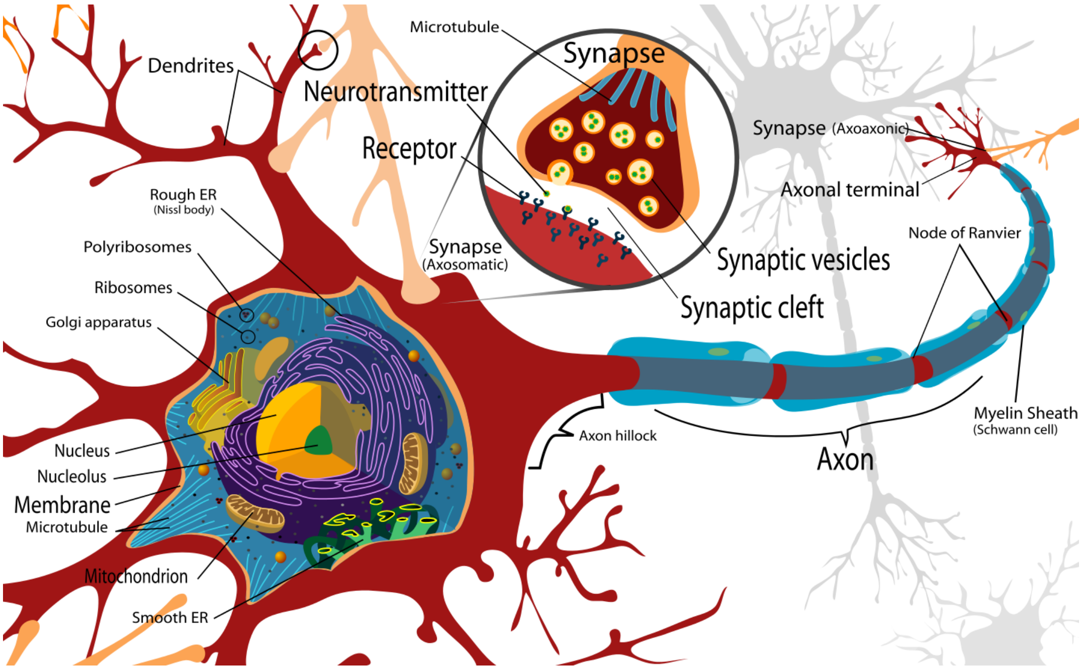


Figure 1.3: Illustration of neuronal cells. Signal transduction by electrical impulses and the translation into chemical transmission at the synaptic cleft. Neurotransmitters are released from the synapse of the presynaptic neuron and activate receptors of postsynaptic neurons, which continue the signal transmission [114].

1.2 Ion Channels

Ion channels are responsible for the signal transduction in excitable cells, like neurons or muscle cells. To fulfil the various functions involved in the signal transduction process, channel types, like the voltage-gated and ligand-gated ion channels, with specialized properties evolved.

Although voltage-gated and ligand-gated ion channels are classified into the large α-type channel class, as they predominantly consist of α-helices, they are separated into families, due to their different types of gating mechanisms. Especially the classification of the VGIC superfamily is further refined by distinguishing the specific ion permeability of the family members. A comprehensive source, organizing these protein types, is the Transporter Classification Database (TCDB) [`http://www.tcdb.org/`].

When traced back through evolution, the classified ion channel families are supposed to originate from a few common ancestors, some of which are still present in certain prokaryotes [48]. Studies led to the general assumption that the heterotetrameric Na^+ and Ca^{2+} channels, built out of one protein with four connected but different domains, evolved from homotetrameric K^+ channels, built out of four separate protein domain replicas [30, 219].

It is suggested that the genes encoding the heterotetrameric channel proteins were merged together through two consecutive gene duplications of the single protein domain of the homotetrameric K^+ channels. [6]. The similar structure of the pore-forming principal subunits supports this assumption. The phylogenetic tree in Figure 1.5 is based on a multiple sequence alignment (MSA) of VGIC pore-forming subunit sequences [218] and visualizes the evolutionary relation between VGICs.

A common evolutionary ancestor is also proposed for the ligand-gated ion channel family. These oligomers share a highly similar amino acid sequence, as well as a very homologous structure [149].

The descriptions of VGIC and LGIC do not cover the auxiliary β-subunits [92], which are associated to the pore building α-subunits (Figure 1.4). The β-subunits are considered to assist in functions like channel gating, but are not integrated in the membrane.

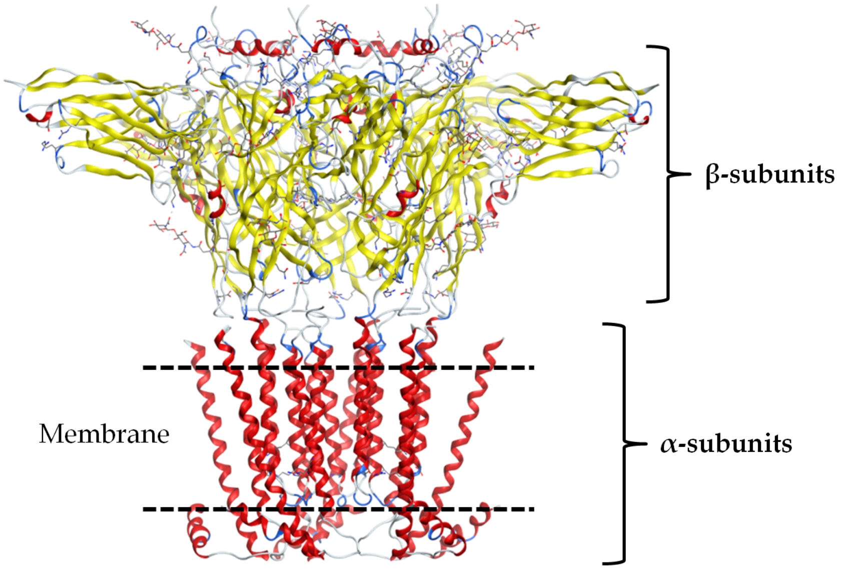


Figure 1.4: Illustration of α- and β-subunits. The α-subunit forms the pore of the ion channel and is embedded into the membrane, whereas the β-subunits are separate proteins, which are attached to the channel.

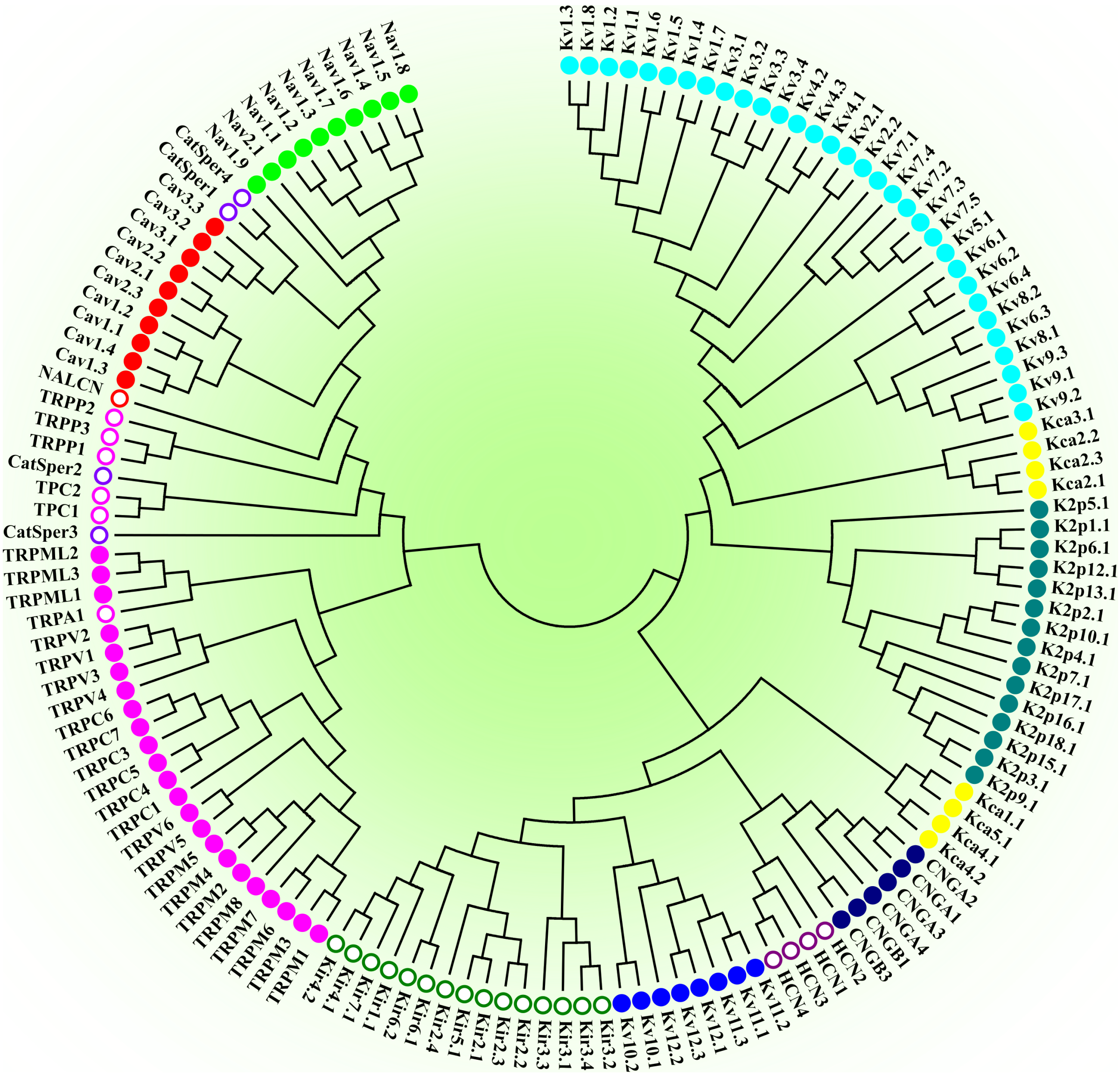


Figure 1.5: Phylogenetic tree of VGIC families. The multiple sequence alignment (MSA) is based on VGIC pore-forming subunit sequences. The sequences in the alignment included the transmembrane domain segments S5 and S6 and the P-loop, connecting S5 and S6.

1.2.1 Structure and Topology

1.2.1.1 Voltage-Gated Ion Channels

Voltage-gated ion channels have a complex structure, which is assembled from up to four α-subunits and multiple auxiliary subunits. The basic elements of the α-subunits are shared by all discovered VGIC family members. Different α-subunits can be combined to build heteromeric channels, but also homomeric channels exist, that contain identical α-subunits. Heterotetrameric ion channels consist of four different α-subunits. An α-subunit segment, containing two transmembrane helices (2-TM), split by a pore-forming P-loop, represents the minimum ion channel building block, the pore domain (PD), as shown in Figure 1.6 (A). This motif is sufficient to create a functional homotetrameric channel, like the inward-rectifier potassium channel K_{ir} [191]. Based on the 2-TM element the protein was subsequently extended by four additional transmembrane helices, resulting in the well known 6-TM domain, shown in Figure 1.6 (B), of which mammalian voltage-gated ion channels are assembled [6]. These six transmembrane helices are called S1 to S6.

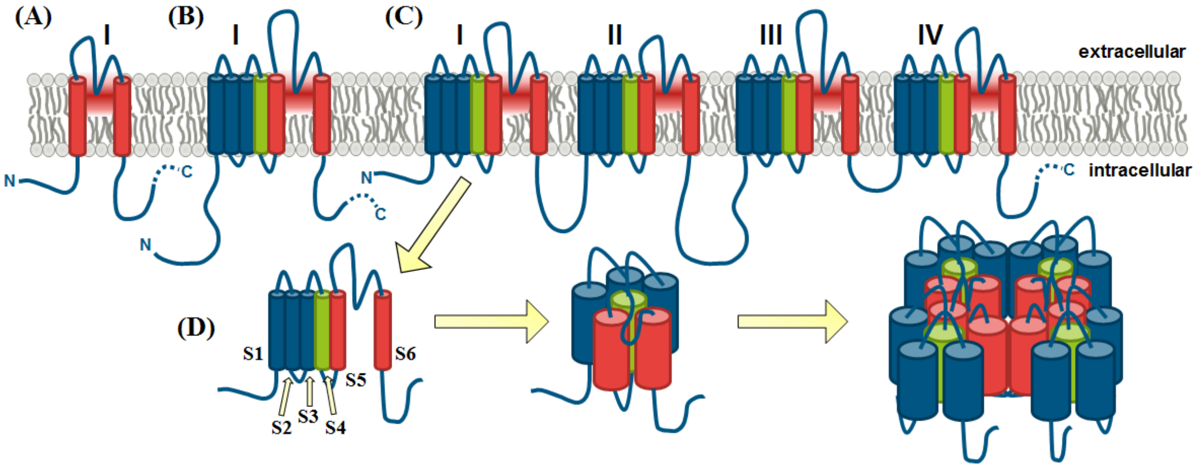


Figure 1.6: Composition and topology of transmembrane subunits. (A) 2-TM subunit, also described as pore domain (PD), represents the minimum ion channel building element including segments S5 and S6 and the pore-forming re-entrant P-loop. (B) 6-TM subunit with PD and attached voltage-sensing domain (VSD), that includes segments S1-S4. (C) 24-TM subunit containing four 6-TM subunits, described as domain I-IV, connected by intracellular loops. (D) Labelled 6-TM domain subunit and the schematic depiction of the stepwise assembly from single subunits to a complete channel.

The new transmembrane helices S1-S4 added the ability of voltage detection to the ion channels. Especially the voltage sensing S4 segment contains positively charged residues, which react to changes of the membrane potential. The so called voltage-sensing domain (VSD) spans from S1 through S4 and is connected to the PD by the S4-S5 linker helix. As shown in Figure 1.6 (C), mammalian heterotetrameric sodium- and calcium-selective voltage-gated ion channels are composed from a single continuous sequence of 1800 to 2500 amino acids, folded into a protein with four domains (DI-DIV), that are connected by cytoplasmic loops [218].

1.2.1.2 Ligand-Gated Ion Channels

Ligand-gated ion channels [37], in this context focusing on the Cys-loop receptor family, differ from VGIC in their assembly, due to the fact that they are pentameric proteins, composed by five homologous subunits, which are displayed in Figure 1.7 from a top view perspective. Each subunit contains four transmembrane helices M1-M4, which form the central pore, and β-sheets, composing the large extracellular vestibule, that contains multiple ligand binding sites. In eukaryotes, two opposing cysteine residues in the extracellular loops of each subunit form a disulphide bridge. Thus they are referred to as Cys-loop receptors [81].

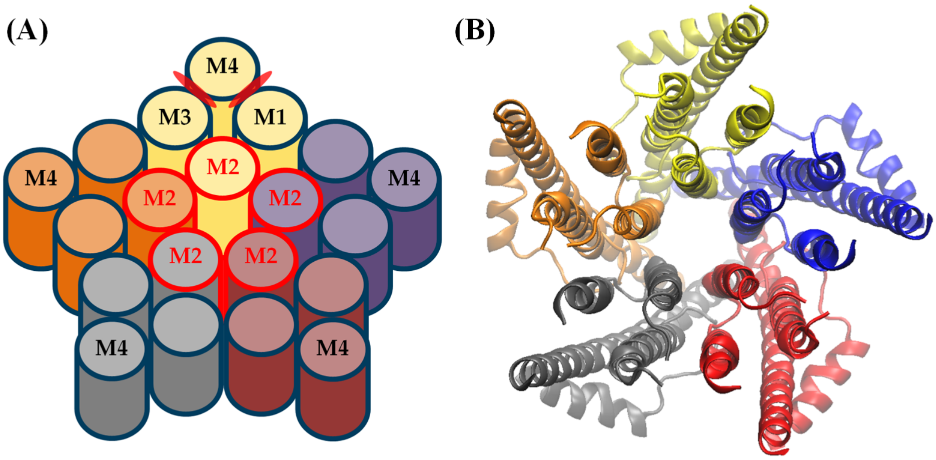


Figure 1.7: Assembly of the LGIC 5-HT$_3$. (A) Schematic illustration of the pentameric channel, composed of five domains shown in different colors. Each domain is assembled out of the four transmembrane helices M1 to M4. The pore building M2 helices are outlined in red. Examples of helix interfaces are indicated by red ovals. (B) Transmembrane pore domain of the 5-HT$_{3A}$ crystal structure (PDB code: 4PIR) in top view, coloured by subunit.

1.2.2 Structure Determination

The determination of structural information about transmembrane proteins remains a difficult task. Out of >120,000 structures organized in the RCSB Protein Data Bank (PDB) [`www.rcsb.org`][16], only ~5600 structures are resolved for proteins that are integral components of the membrane. Out of these, only 3321 are flagged as membrane proteins according to the manually annotated transmembrane dataset from mpstruc [`http://blanco.biomol.uci.edu/mpstruc/`]. The great majority of structures organized in the PDB originate from electron density maps, generated by x-ray crystallography. Out of currently 3321 PDB entries for membrane proteins, 3024 derive from x-ray experiments, which today is the standard technique for the experimental determination of protein structures. The first ion channel crystal structure was the KcsA potassium channel, that was resolved by Roderick MacKinnon, who won the Nobel Prize in Chemistry in 2003 for his work on ion channels. This demonstrates the value of information, x-ray structures provide to support the understanding of protein function. On the other hand, the method is limited to individual snapshots of a certain state a protein adopts in the crystal. These states might also be influenced by detergents or stabilizing mutations, required for crystallization, but potentially causing artificial conformations [128, 211].

An alternative method is the cryo-electron microscopy (cryo-EM) [27], which is featured in many publications and has gained enormous interest, since the first 3.4 Å structure of the TRPV1 ion channel was solved and published [28], soon followed by the human TRPA1 ion channel [153]. The number of membrane protein structures in the PDB, derived with cryo-EM, has doubled in 2015 and increases further in 2016. The total numbers of publicly available structures from x-ray and cryo-EM experiments are shown in Figure 1.8. The plots indicate a linear yearly increase of x-ray structures, whereas the number of cryo-EM structures per year shows an exponential growth. This suggests that the technology develops and it becomes accessible to a fast growing number of labs. Furthermore, the recently published first-in-class voltage-gated calcium channel cryo-EM structure of $Ca_v1.1$ with a 4.2 Å resolution [212] indicates that the cryo-EM method avoids some major challenges in solving ion channels, compared to the x-ray crystallography. The structure provides insight into another, yet structurally unexplored ion channel family.

The proceeding development of cryo-EM techniques enables scientists to also analyse protein dynamics by flash-freezing proteins in different functional states, as demonstrated in an impressive movie of the V-ATPase, which was created using 100,000 cryo-EM images [221].

Despite recognizable progress in solving challenging proteins and in developing technical innovations like the cryo-electron microscopy, the number of available ion channel structures remains scarce. Out of 39 resolved ion selective channels, only 16 are mammalian and 6 of them human, as listed in Table 1.1. All other structures are derived from bacterial sources.

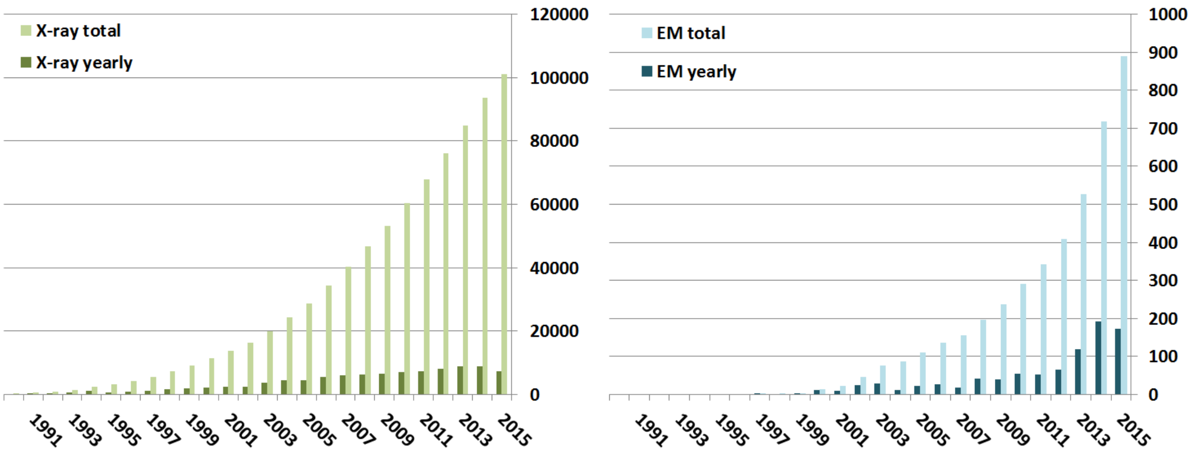


Figure 1.8: RCSB Protein Data Bank statistics. Growth of released structures in the Protein Data Bank (PDB) separated by experimental methods. Left: The increase of x-ray structures slows down whereas (Right:) the structures resolved by electron microscopy increase exponentially.

Channel Name	PDB Code	Reference
Two-Pore Domain Potassium Channel K2P1.1 (TWIK-1): *Homo sapiens* (expressed in Pichia pastoris), 3.40 Å	3UKM	Miller & Long (2012)
Two-Pore Domain Potassium Channel K2P4.1 (TRAAK): *Homo sapiens* (expressed in Pichia pastoris), 3.80 Å	3UM7	Brohawn et al. (2012)
Two-Pore Domain Potassium Channel K2P10.1 (TREK-2), up state: *Homo sapiens* (expressed in S. frugiperda), 3.20 Å	4BW5	Dong et al. (2015)
Kv1.2 Voltage-gated potassium Channel: *Rattus norvegicus* (expressed in Pichia pastoris), 2.9 Å	2A79	Long et al. (2005)
Kv1.2/Kv2.1 Voltage-gated potassium channel chimera: *Rattus norvegicus* (expressed in Pichia pastoris), 2.4 Å	2R9R	Long et al. (2007)
Human BK (SLO1) Channel Ca2+-activation apparatus: *Homo sapiens* (expressed in S. frugiperda), 3.0 Å	3MT5	Yuan et al. (2010)
SLO3 K+ Channel pH-sensitive Gating Ring: *Homo sapiens* (expressed in S. frugiperda), 3.40 Å	4HPF	Leonetti et al. (2012)
GIRK1 (Kir3.1) cytoplasmic domain: *Mus musculus* (expressed in E. coli), 1.8 Å	1N9P	Nishida & MacKinnon (2002)
Kir3.1-Prokaryotic Kir Chimera: *Mus musculus & Burkholderia xenovornas* (expressed in Escherichia coli), 2.2 Å	2QKS	Nishida et al (2007)
Kir3.1 cytoplasmic domain: *Mus musculus* (expressed in E. coli), 2.0 Å	3K6N	Xu et al (2009)
GIRK2 (Kir3.2) G-protein-gated K+ channel: *Mus musculus* (expressed in Pichia pastoris), 3.60 Å	3SYO	Whorton & Mackinnon (2011)
Hv1 chimeric (VSOP/Hv1) voltage-gated proton channel: *Mus musculus* (expressed in S. frugiperda), 3.45 Å	3WKV	Takeshita et al. (2014)
NaV1.7 VSD4 voltage dependent sodium channel: *Homo sapiens* (expressed in Trichoplusia ni), 3.5 Å	5EK0	Ahuja et al. (2015)
RyR1 ryanodine receptor, closed state in complex with FKBP12. Cryo-EM structure: *Oryctolagus cuniculus*, 3.8 Å	3J8H	Yan et al. (2015)
InsP3R1 Inositol-1,4,5-trisphosphate receptor: *Rattus norvegicus*, 4.7 Å	3JAV	Fan et al. (2015)
CaV1.1 voltage-gated calcium channel: *Oryctolagus cuniculus*, 4.2 Å	3JBR	Wu et al. (2015)

Table 1.1: List of mammalian ion channel structures. PDB codes and descriptions adapted from the "Membrane Proteins of Known 3D Structures" website [`http://blanco.biomol.uci.edu/mpstruc/`]. The first published structure for each channel type is listed.

1.2.3 Activation and Gating

The fundamental structural differences between VGIC and LGIC have been discussed in the previous chapter, which is followed by the detailed description of their activation and gating mechanisms in this section. It was pointed out earlier, that most ion channels are specialized in conducting only specific ions and their permeability changes very quickly in response to a distinct signal. This rapid adaptation to environmental conditions is enabled by conformational changes the channels perform to establish defined states. For VGICs these resting, partially activated, open, and inactivated states are visualized in Figure 1.9, as a cycle of transitions between the different activation states. For LGICs similar conformational states are described, even though the mechanisms to adopt to those states differ.

1.2.3.1 Voltage-Gated Ion Channels

The activation of VGICs is mediated by the four-helix bundle of the VSD and primarily by the S4 helix, which contains an evolutionary conserved, positively charged arginine residue in almost every third position [31]. The adverse position of these charges in the middle of the non-polar lipid membrane requires shielding by the surrounding S1-S3 helices [94]. However, the position of S4, in the resting state of the channel, is far from static and displaceable by a small impulse. During a depolarizing voltage change across the membrane, the S4 helix is attracted by the negative turning exterior and leaves its resting position. Through its connection to the S4-S5 linker, a conformational change in the pore domain of the ion channel is induced [122]. This partially activated state is still not conducting ions, until the cascade of conformational changes results in the opening of the channel pore. This open state enables ions to pass through the channel, according to the voltage gradient, until the channel undergoes an inactivation and the flow of ions is stopped.

The inactivation of ion channels is necessary to end the flow of ions and maintain the voltage gradient of the cell membrane [9]. Two different mechanisms of ion channel inactivation are known. The fast inactivation is observed to be important for membrane potential repolarization [35], and the slow inactivation regulates the excitability of cells, such as muscle fibres [167]. In VGSCs the intracellular loop between DIII and DIV acts as a fast responding inactivation gate, containing hydrophobic residues, called IFM (Ile-Phe-Met) motif [134]. These residues bind to the channel pore, as kind of a plug, also referred to as ’hinged-lid’ or ’ball-and-chain’ mechanism, and inactivate the channel within milliseconds.

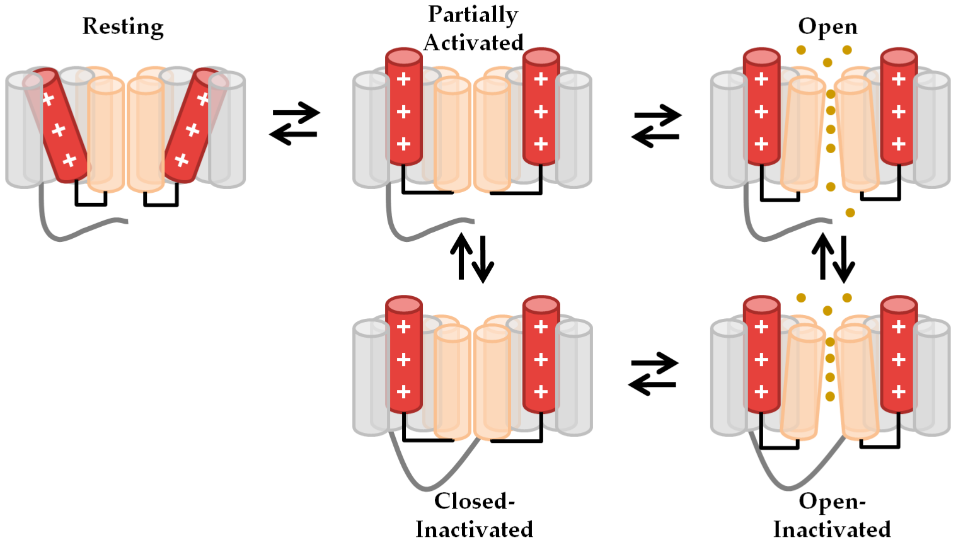


Figure 1.9: Cycle of Activation States. Two schematic opposing subunits of a fourfold symmetric channel are shown. The voltage sensing domain is coloured orange and red, with indicated positive charges, and the pore-forming domain is grey. In the resting state the pore is closed and VSDs are located on membrane level. Upon membrane depolarization the VSDs move out of the membrane plane and the pore opens until ions (orange dots) can pass the pore. The inactivation starts when the 'hinged-lid' (dark grey loop) binds to the channel pore and stops the ion conduction. Once the membrane potential normalizes, the VSDs move back to their resting position and the 'hinged-lid' relieves from the pore.

Slow inactivation is achieved by a conformational change in the channel pore, independent from the intracellular 'hinged-lid' loop of the fast inactivation [104]. As the name implies, slow inactivation runs on a longer timescale and regulates the activation state of the ion channels and thus the membrane excitability and frequency of ion conduction. However, the mechanism is not yet fully understood and under discussion. Beyond these two basic inactivation mechanisms, also modulations by subunits and terminal regions are described, which alter the kinetics of the inactivation process to specific needs, like for cardiac channels [74].

1.2.3.2 Ligand-Gated Ion Channels

Ligand-gated ion channels are opened or closed through binding of a ligand, fitting into a specific binding site [198, 1]. The agonist, binding to the large N-terminal extracellular domain (ECD), rapidly modifies the configuration of the resting channel and the transmembrane pore opens. This process is called gating isomerization [7]. Upon binding at the receptor site of the ECD, far away from the transmembrane domain

(TMD), conformational changes are initiated that need to be efficiently transduced to the TMD. This mechanism is described as a concerted opposite-twist rotation of the ECD relative to the TMD around the five-fold symmetry axis [176]. This results in the opening of the central pore by an outward movement of the loop connecting TM2 and TM3 in the TMD. When binding an antagonist, the channel adopts a closed conformation, blocking or terminating the ion flux through the pore.

In cases where the ligand is not released from the binding site, the LGIC changes its functional state from the active agonist-bound open conformation to a closed agonist-bound conformation, the desensitized state. This process can be triggered through binding of an effector molecule to an allosteric binding site, leading to interactions between TM2 and TM3 in the TMD, which limits the ion conduction and controls the postsynaptic current, initially induced by neurotransmitter release [72].

1.2.4 Ion Selectivity

Ion selectivity is characteristic for some ion channels, like sodium or potassium ion channels, that are selectively permeable for particular ions. Five types of selectivity have been observed in voltage-gated ion channels: proton (H^+), sodium(Na^+), potassium (K^+), calcium (Ca^{2+}) and chloride (Cl^-) selectivity is related to ion channels, but often channels are not restricted to one single ion type.

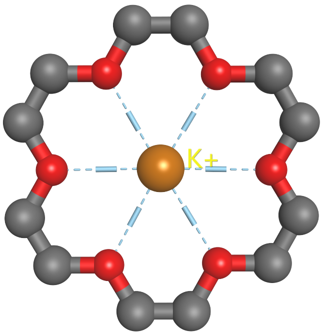


Figure 1.10: K^+ ion in complex with 18-Crown-6 ether

Hydrated ions cannot pass the hydrophobic membrane without assisting mechanisms or guided paths, such as those provided by ion channels or membrane pores. The water filled pores of ion channels connect intra- and extracellular medium and ions can diffuse into the pore according to their concentration gradient.

The selectivity filter of an ion channel is formed by the P-loops, which link the pore-forming helices TM5 and TM6. For potassium channels, a conserved set of residues (TVGYG) line the selectivity filter, that constricts the pore at the extracellular end [52]. K^+ ions need to strip their hydration shell of water molecules when entering the selectivity filter. Inside the selectivity filter, the K^+ ions are coordinated at defined temporal binding sites by oxygen atoms of polar residues, until the next entering ion pushes it further to the next binding site. Figure 1.11 (A) shows an ion occupied selectivity filter of a potassium channel. The proposed basic principle of a selectivity filter in VGICs can be compared to the function of a crown ether, a cyclic chemical structure forming complexes with cations [57]. Certain dedicated crown ethers show high affinities for different ions, as shown in Figure 1.10 for a 18-crown-6 ether in complex with a K^+ ion, which is equivalent to the ion coordination in potassium channels.

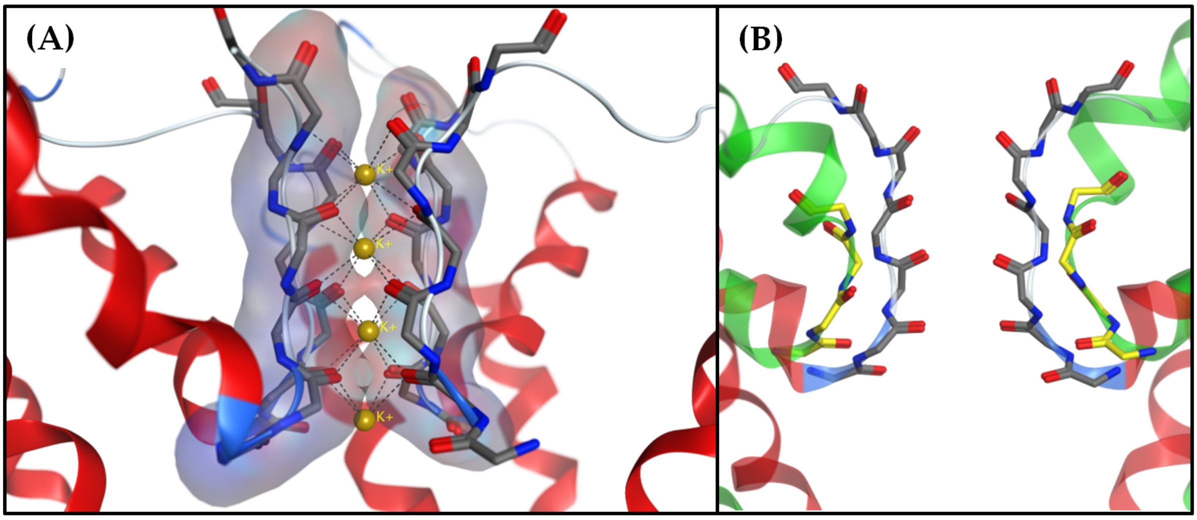


Figure 1.11: Selectivity filters of K^+ and Na^+ channels. (A) Potassium channel selectivity filter occupied with dehydrated K^+ ions. The ions are strictly coordinated between the oxygens of polar residues. (B) Overlay of a potassium (grey, red) and sodium (yellow, green) channel selectivity filter. The diameter of the channel openings clearly differ, according to the sizes of the passing ions.

As this similar mechanism explained the observations made for potassium channels, an analogous mechanism was assumed for sodium ion channels, until the first x-ray structures for VGSCs were solved. The structures indicated a broader and shorter selectivity filter than in potassium channels. In procaryotic homotetrameric sodium channels the selectivity filter forms a ring of four glutamate side chains (EEEE) reaching into the channel entry, whereas the selectivity filter of eucaryotic heterotetrameric sodium channels contains conserved residues called DEKA motive. Figure 1.11 (B) shows a backbone overlay of the potassium channel (grey, red) and sodium channel (yellow, green) selectivity filter, demonstrating the different dimensions. As opposed to the K^+ ions, Na^+ ions permeate the ring of the sodium selectivity filter and the channel pore in its hydration shell. This process is described as knock-on mechanism [32, 40]. The proportions of K^+ and Na^+ ions are illustrated in Figure 1.12, which shows the size of the hydrated ions and the interaction distances in a potassium selectivity filter. The hydrated Na^+ is smaller than the hydrated K^+ ion. Thus, the water molecules are more effectively stripped off the K^+ ion when reaching the potassium selectivity filter as it mimics the K^+ ion hydration shell, which simultaneously hinders the passage of the hydrated Na^+ ion.

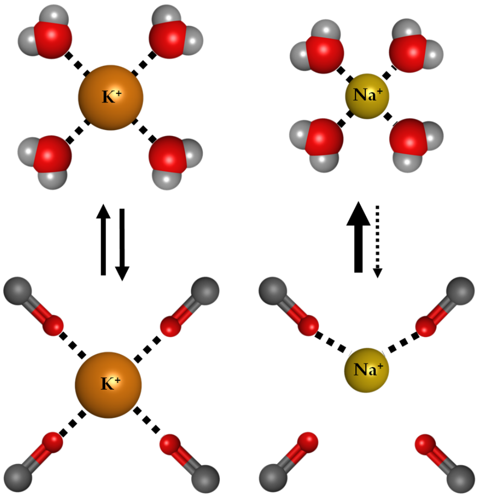


Figure 1.12: Proportions of hydrated K^+ and Na^+ ions. Hydrated potassium and sodium ions compared to the distances of oxygen atoms of residues forming the selectivity filter of a potassium ion channel.

LGIC are generally not ion type selective. A principal selectivity filter, formed by specific charged residues on the M2 transmembrane helices, only discriminates between anions and cations, as shown in Figure 1.13 using $GABA_AR$ and 5-$HT_{3A}R$ channel subunits as examples. The anion/cation selectivity is also regulated by the pore diameter, which depends on the presence of a specific proline residue. A larger pore diameter results in weak selectivity whereas a narrow pore increases selectivity [105].

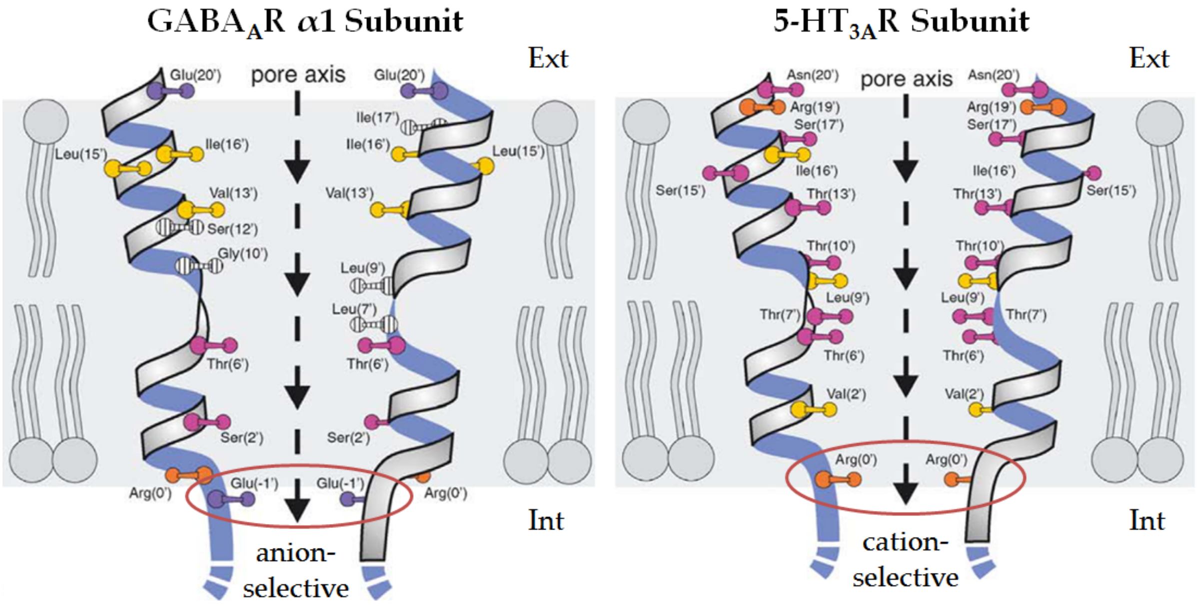


Figure 1.13: LGIC pore and selectivity filter. Schematic representation of two M2 helices of GABA$_A$R and 5-HT$_{3A}$R channel subunits. The side chains are colour coded as follows: basic residues (red), acidic residues (blue), polar residues (purple), non-polar residues (yellow). The charged key residues, accountable for the anion/cation selectivity, are marked with orange circles [105].

1.2.5 Pharmacology - Toxins and small molecules

Ion channels account for key functions in many physiological processes by transporting ions across cell membranes, as described in the previous sections. Due to their central role in physiological mechanisms, studies have indicated that ion channel proteins are the primary factor in a number of channelopathis (diseases caused by channel dysfunction), originating from mutations in the ion channel or caused by interacting with abnormal proteins [101]. Figure 1.14 gives an overview on ion channel drug targets and related therapeutic indications.

Selective modulation of ion channel function by pharmacological drugs is a promising approach in the treatment of channelopathis. Nevertheless, compared to other pharmacologically relevant protein families, like kinases and GPCRs, the total number of released ion channel drugs is small. In total 185 compounds were launched for ion channels over the past decades, according to Thompson Reuters Integrity Database, whereas for the favoured GPCR proteins, the database counts 536 launched compounds. Figure 1.15 shows the number of compounds in different phases of phar-

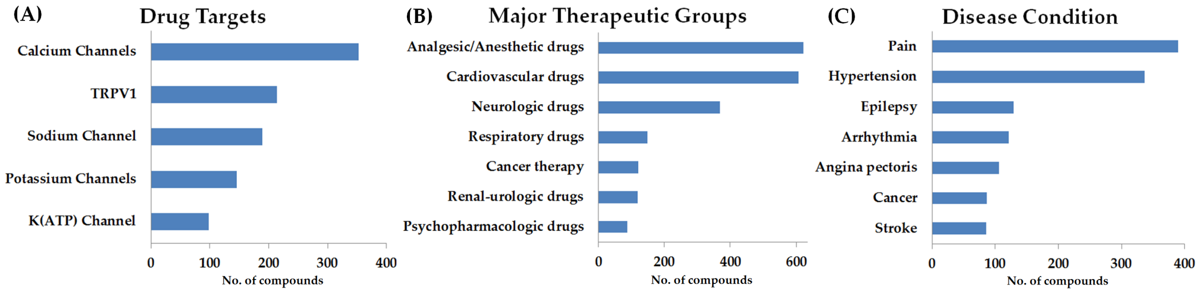


Figure 1.14: Ion channel related drug targets and pharmaceutical application. These figures are based on compounds that reached at least preclinical development status. (A) shows individual ion channels and ion channel classes ranked by the number of related compounds. (B) shows the distribution of ion channel compounds on major therapeutic groups, and (C) classifies the compounds according to disease conditions in which they are or might be effective. *Data source: Thomson Reuters Integrity*

macological development. The numbers indicate that research in the field of ion channels is more complex and less successful in terms of finding promising compounds. Another reason might be that the field of ion channels, due to scarce detailed information, was underestimated or neglected by research groups and the pharmaceutical industry in the past.

An obvious start for the search of new ion channel compounds are natural toxins, which are available in the venoms of various animals, and can be used as scaffolds in compound development, trying to mimic parts of the toxins while increasing the selectivity for specific ion channel targets. These natural neurotoxins are very potent ion channel blockers, showing, in some cases, reasonable selectivity for certain ion channel species. The well characterized tetrodotoxin (TTX) [138], present mainly in sea animals such as puffer fish, blocks VGSCs, except for $Na_v1.5$, $Na_v1.8$ and $Na_v1.9$, with IC_{50} values in a nanomolar range. The TTX resistant $Na_v1.x$ channels are lacking one aromatic side chain, close to the selectivity filter, which is crucial for TTX binding. A nice overview on the therapeutic potential and pharmacological properties of VGSCs is provided by Ruiz et al. [47] and Nardi et al. [143].

Beside the known indications in which ion channels are generally involved, $Na_v1.7$ obtained increasing interest in pharmaceutical research, when Cox et al. [42] described a congenital inability to pain sensation (CIP) phenotype in human. This inability to pain sensation results from loss-of-function mutations in different parts of the $Na_v1.7$ α-subunits. Before, it has already been shown that $Na_v1.7$ and $Na_v1.8$ are involved in certain inflammatory and neuropathic pain pathways [144, 100]. The in-

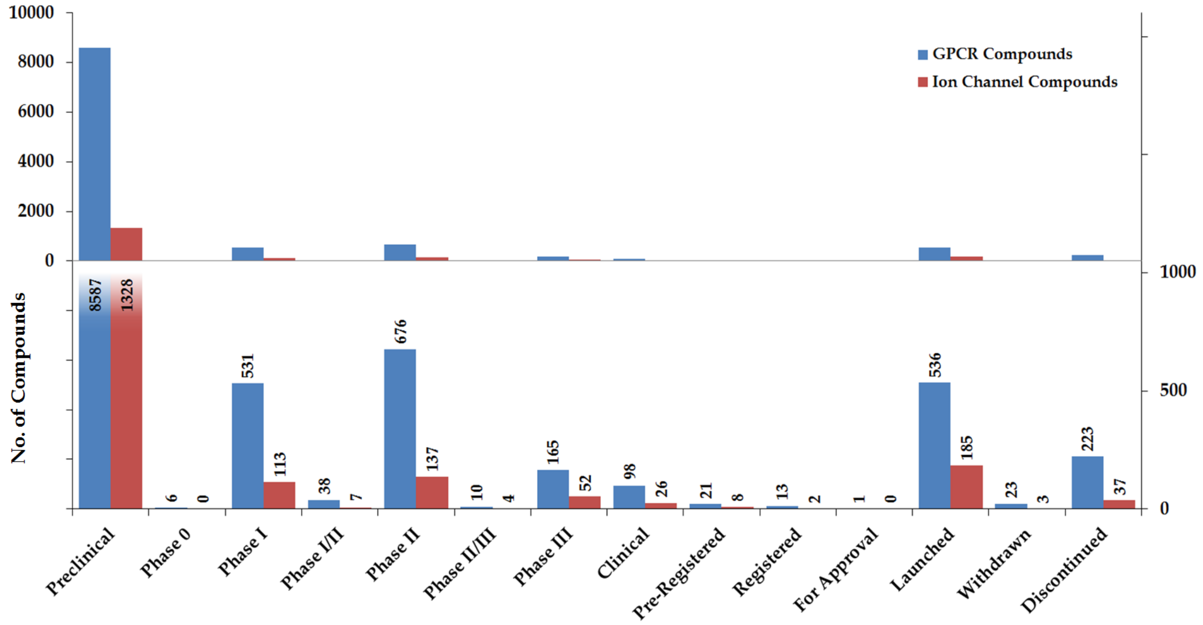


Figure 1.15: Compounds in different phases of development. This panel bar chart shows the number of compounds in different phases of development, for ion channels and GPCRs in comparison. The top panel shows the absolute numbers and the bottom panel is scaled up to visualize the smaller numbers. Through all phases of development, more compounds can be found for GPCRs than for ion channels. This raises questions about the complexity of ion channel drug development and how it could be promoted by new insights from ion channel research. *Data source: Thomson Reuters Integrity*

terest for this target can also be noticed by the high number of 788 drugs and biologics currently under development, according to Thomson Reuters Integrity.

1.3 Research Motivation

Ion channels are involved in several processes like nerve and muscle excitation, hormone secretion, cell proliferation and sensory transduction. Due to their central positions in many physiological pathways, dysfunctions of ion channels lead thus to severe channelopathies like particular pain syndromes [101, 60, 42], epilepsy [102] or cystic fibrosis [162], and are therefore of special interest as drug targets [53, 60]. As described before, ion channels share a common overall structure. For example, the same number of helical transmembrane domains, and the protein sequences in single sub-families are quite homologous. This leads to major challenges in the discovery and development of selective compounds, which are further exacerbated by the lack of detailed structural knowledge necessary for rational drug design.

The motivation for this work was to expand the knowledge about the structure of certain ion channels. For this purpose, different computational methods were used to create and analyse ion channel models, which will be described in Chapter 2. Chapter 3 deals with the analysis of the assembly of transmembrane helices in a 5-HT$_{3A}$ pore-forming domain. For this, homology models were built, using constraints from site directed mutagenesis experiments. During the homology modeling efforts, the question about the domain association of heterotetrameric ion channels emerged. Since the structure of these ion channels is build by one continuous amino acid sequence containing four unequal domains, as illustrated in Figure 1.6, different ways of domain association leading to a complete channel structure are possible. To identify the most probable association type of the four channel domains, an analysis using the Na$_v$1.8 voltage-gated sodium channel as an example was performed. The results and the applied workflow, which integrates different computational methods, are described and discussed in Chapter 4. Finally, Chapter 5 covers some preliminary activities on the detection and translation of ligand binding sites from ion channel unrelated x-ray structures onto ion channel models, to discover new starting points for ion channel drug discovery. Further, some speculations and possible follow-up experiments are discussed.

2 Molecular Modeling Methods

Molecular modeling methods and techniques enable scientists to model and simulate states and dynamics of molecules or molecular systems. Today, a broad spectrum of methods can be applied on topics like sequence analysis, homology modeling, molecular dynamics simulations and ligand docking. The high performance of modern computer systems and increasing amounts of data from various screening and sequencing initiatives, facilitate the application of molecular modeling methods on larger and more complex biological systems. The next sections describe some fundamental methods of molecular modeling, and presents examples and results of their application that should help to estimate the feasibility and value of the methods when applied on ion channels.

2.1 Sequence Alignment and Secondary Structure Prediction

Sequence alignment methods are used to compare proteins on sequence level and evaluate their similarity that might indicate structural, functional or evolutionary correlations. Proteins with similar sequences tend to show similar secondary and tertiary structures, often leading to similar functions. Based on the observed similarity and identified conserved sequence regions or patterns, proteins can be categorized into protein families and classes. A sequence alignment is commonly illustrated as a matrix. Each row in the matrix represents a sequence, and each column contains residues that are identical or similar in the aligned sequences. Gaps are placed into the sequences to achieve the best feasible match between residues.

Alignments can be distinguished into local (Smith-Waterman algorithm) [183] and global (Needleman-Wunsch algorithm) [146] alignment algorithms. Global alignments are appropriate for sequences of similar length that share a certain similarity to minimize the number of inserted gaps, as the algorithm tries to match residues along the whole sequences. A local alignment is typically applied for sequences with unequal length or containing isolated regions of similarity, for example, multiple domains or repeats. Both algorithms are implemented using dynamic programming techniques, which apply substitution matrices containing defined sores for matching or non-matching residue pairs. On the other hand, penalty scores are assigned for opening and extending gaps in the alignment to avoid heavily scattered sequence domains or patterns. A BLAST [4] search, as an example for a sequence similarity searching method, applies alignments based on heuristic methods to find sequences in a database, which are similar to a query sequence of interest and therefore might share a common structure or function. This approach needs an efficient method, like the k-tuple method, to search large sequence databases. The k-tuple method defines pieces of the sequence as query words, and if words are found in both sequences, detailed alignment comparison is carried out.

The basic type of alignment is the pairwise sequence alignment. Two sequences are compared to each other to identify common regions or patterns indicating a relationship. When aligning more than two sequences, the multiple sequence alignment (MSA) method is used. From its result, information about sequence homology and evolutionary links, like conserved regions, can be derived. The phylogenetic tree in Figure (1.5) is based on an MSA of the VGIC family. The Uniprot [38] database was used to collect the sequences for the ion channel MSA by searching for the keyword "Ion channel" and the organism "Homo sapiens", resulting in 322 sequences marked as reviewed entries. Due to extreme differences in the length of the channel sequences, they were reduced to the sequence parts of pore-forming helices and loops, as those contain specific sequence patterns that allow discrimination between the ion channels. The MSA created with MUSCLE [59] was then translated into a phylogenetic tree using MEGA6 [189].

One way of initializing an alignment of sequences is to focus on structurally important regions and perform a local multiple sequence alignment (MSA). In case of ion channels, the alignment can be initially focused on the channel's pore region (S5 Helix - P-loop - S6 Helix), as described before, and the voltage sensing domain, which contains lots of charged residues. These conserved and well described parts are fairly easy to find in the sequence and usually existent in different types of ion channels. Less conserved dissimilar sequence parts can then be aligned in a second step within the borders of these anchor points, resulting in fewer and better manageable alignment variations.

The observation that sequence similarity links to structural and functional similarity of proteins is the basis for homology modeling, an advanced method that is introduced in section 2.2.

Secondary structure prediction methods can provide structural information that can help to align structurally similar parts of sequences such as α-helices or β-sheets, which are not obviously related according to sequence similarity. The secondary structure is predicted from the primary structure. The primary structure is a sequence of residues with a high number of degrees of freedom to adopt a certain conformation. Based on data about features and characteristics of secondary structures derived from x-ray crystal structures, the method tries to identify sequence regions that are likely α-helices, β-strands or loops. Popular tools like PSIPRED [96] and Jpred [55] improved their prediction accuracy to over 80% using trained neural networks and scoring matrices. Also specialized tools like TMHMM (Transmembrane Hidden Markov Model) [111] have been developed to predict transmembrane helices, which is of great value when working with transmembrane proteins like ion channels, transporters or G-protein-coupled receptors. Such predictions provide valuable constraints for homology modeling [43], ab initio [26] and tertiary structure prediction methods [129].

2.2 Homology Modeling

Homology modeling, also called comparative modeling, is based on the assumption that proteins with a similar amino acid sequence share a similar 3D structure [214, 73]. A sequence alignment of the target sequence (unknown structure) and template sequence (known structure) is required and important for the quality of the resulting homology model, because the alignment defines the translation of a target residue to a structural coordinate of the template structure. The principle of homology modeling is illustrated in Figure 2.1. In case of low sequence similarity or large alignment gaps between template and target structures, the approach becomes more challenging. As described earlier, secondary structure prediction can provide additional information for the alignment. By predicting secondary structures, a structural alignment can be used to focus on structural similarity instead of sequence information only, which allows explicit matches between parts of the protein sequences. When an alignment is defined the geometry of each template residue is applied to the corresponding target residues. For target residues located in regions of inserts or loops, and aligned with gaps of the template structure, coordinates can be generated by dedicated loop modeling methods. These modeling methods use collections of known loops or fragments as templates to find a suitable conformation or apply *ab initio* methods based on statistical models.

In case of very closely related proteins it can be sufficient to only mutate certain residues in the template structure that differ from the target sequence and relax the side-chains to create a model of the target protein. This is also an option for crystal structures lacking certain side-chains, if they have not been identified in the electron density from an x-ray experiment. By self-homology modeling, these missing residue parts can be added.

An often discussed question is how the quality of a homology model depends on the sequence similarity between target protein and template [91]. Alignments with more than 50% sequence similarity are assumed to generate reliable models, whereas the "twilight zone" below 30% of similarity is considered to be increasingly error-prone [137]. Nevertheless, examples like hemoglobin and myoglobin combine a low sequence similarity of 24% with high structural similarity measured by a root mean squared deviation (RMSD) of ∼1.35 Å based on Cα atoms. This indicates that the three-dimensional structure of proteins is evolutionary more conserved than the sequence similarity, which has a higher variability [185]. Among the homology models for five GPCR target proteins, generated using 10 diverse GPCR template structures, the best models were not derived from the closest related templates according to sequence similarity [161]. Many encouraging efforts based on methods with continually improved quality of prediction show that it is feasible to consider models based on distantly related structures [181, 71].

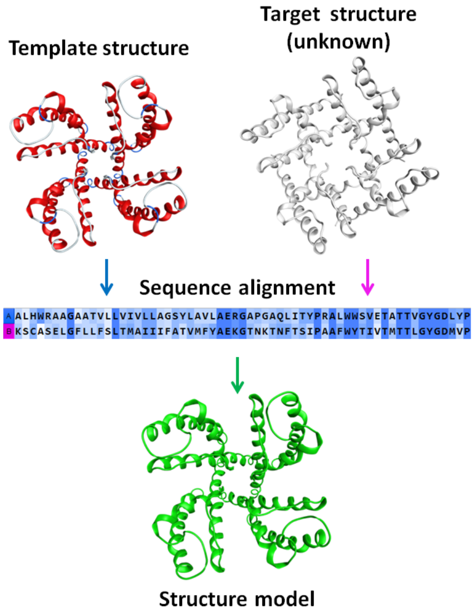


Figure 2.1: Principle of homology modeling. A sequence alignment of the template and target structure sequences is used to match the target residues to coordinates of the template, creating a structure model containing the target residues.

2.2.1 Homology Modeling Test Case

To evaluate the accuracy of homology modeling methods for transmembrane proteins, a cross modeling experiment with the known voltage-gated potassium channel structures of KcsA (PDB: 1K4C) with a resolution of 2 Å and KirBac1.1 (PDB: 1P7B) with a resolution of 3.65 Å was preformed, as shown in Figure 2.2. The KirBac1.1 structure shows an open conformation, whereas the KcsA channel structure is closed. The channel structures' RMSD is 15.7 Å. The MOE software suite [33] was used to model each of the two protein sequences on the structure template of the other protein. In a first step the templates were superposed and the resulting alignment of the sequences was used for the modeling steps. The sequence alignment revealed a similarity of 39% for a global and 50% for a local alignment.

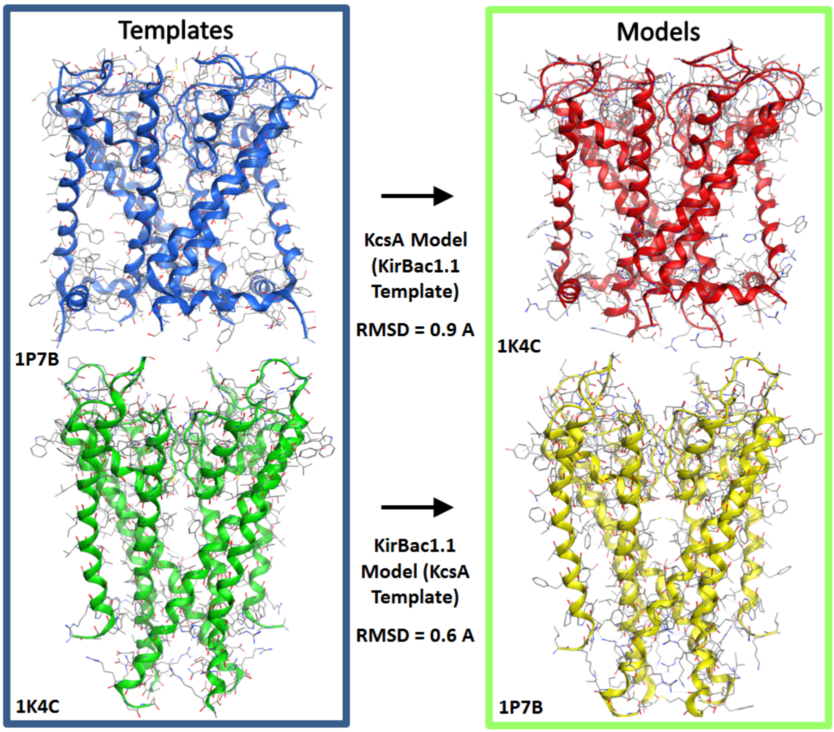


Figure 2.2: Assessment of homology modeling. Cross homology modeling of KcsA and KirBac1.1 voltage-gated potassium channels. Both structures were used as template for the sequence of the other channel respectively. The RMSD values between templates and models were < 1 Å.

On each of the two templates a model was built, using the sequence of the other protein respectively, representing the target structure. The quality of the generated homology models needed to be assessed to estimate their use for subsequent analysis.

Methods to check homology models are, for example, the calculation of RMSD values as comparison measure to a related experimental structure, if one is known. The potential energy of a model or the number of outliers in a Ramachandran plot [158] are other criteria to rate a homology model.

The model of KcsA built on PDB structure 1P7B as template has an RMSD value of 0.95 Å to the 1P7B template and 15.8 Å RMSD to the original 1K4C structure of KcsA. The model of KirBac1.1 built on PDB structure 1K4C as template has an RMSD value of 1.75 Å to the 1K4C template and 15.7 Å RMSD to the original 1P7B structure of KirBac1.1. Compared to each other the models show an RMSD value of 15.8 Å. The potential energy of both models are on the same level as the respective templates and also the Ramachandran plot shows almost all residues satisfying the protein geometry constraints except for a few outliers in loop or terminal regions. This basic experiment demonstrates that homology models stick close to the template conformation, which is considered to be a similar structure, but might not be the natural conformation of the model protein. Such models are valid and helpful starting points, but they may require further refinement to converge to native like conformations. One approach to achieve this is molecular dynamics simulation, which is covered in the next section.

2.3 Molecular Dynamics Simulation and Energy Calculation

Molecular dynamics (MD) simulates the movement and interaction of atoms in a defined system. As a result, MD simulations generate a series of snapshots of the simulated system over a defined period of time, the so-called trajectory, which allows the analysis of a molecule's behaviour under predefined conditions.

In general, the properties of atoms and bonds in a molecular dynamics system are defined by a force field (FF) that incorporates information like atomic mass, van der Waals parameters, partial charges, bond lengths, bond angles and force constants. These information vary according to the parameterization of a force field, which can be based on electronic structure calculations like the Merck Molecular Force Field (MMFF) [79], developed for small molecules. More often experimental data are used for the parameterization, like for the Charmm force field [126] that is applied in all simulations of this thesis. Other commonly used force fields are Amber [39], Engh-Huber [62], Gromos [148] and OPLS [99]. The force field parameters are used to compute a potential energy function that describes the molecular system. This potential energy function $U(\vec{r})$ is the summation of bonded and non-bonded energy terms.

$$
\begin{aligned}
U(\vec{r}) = \quad & \sum_{bond} k(r_{ij} - r_0)^2 && (2.1)\\
+ \quad & \sum_{angle} k_\theta(\theta - \theta_0)^2 + k_{UB}(r_{ac} - r_{UB})^2 && (2.2)\\
+ \quad & \sum_{tors} k(1 + cos(n\psi + \phi)) && (2.3)\\
+ \quad & \sum_{impr} k(\psi - \phi)^2 && (2.4)\\
+ \quad & \sum_{vdW} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] && (2.5)\\
+ \quad & \sum_{elec} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_R r_{ij}} && (2.6)
\end{aligned}
$$

The potential energy of covalent bonds and angle bending, represented by equations 2.1 and 2.2, are modeled by a harmonic oscillator and often exemplified as two particles connected by a spring, where k is the spring constant, r_{ij} is the bond length between atoms i and j, and r_0 is the reference bond length. Accordingly, θ is the bond angle of three connected atoms a, b, and c and the ideal angle θ_0. The Urey-Bradley term for angle bending accounts for the distance r_{ac} between the angle atoms a and c and is used in the Charmm FF [127]. Equation 2.3 describes the torsional or dihedral angle ψ around the central bond of four connected atoms, with n minima at positions of the angle ϕ. The planarity of structures build by four connected atoms is maintained by equation 2.4. The fourth and fifth equation represent potential energy terms for non-bonded interactions. Equation 2.5 introduces the van der Waals interactions as 6-12-Lennard-Jones potential, where ϵ_{ij} is the minimum potential, σ_{ij} is the distance between atoms for which the energy is zero, and r_{ij} is the atoms' distance. The electrostatic potential for atomic charges is described in 2.6, with charges q of atom i and j, the vacuum permittivity ϵ_0 and the dielectric constant ϵ_R.

The dynamics of the system is then determined by solving Newton's second law of motion (2.7) to calculate the interactions between atoms, where $\vec{F_i}$ is the force acting on atom i with mass m_i, acceleration $\vec{a_i}$ and position $\vec{r_i}$.

$$\vec{F_i} = m_i \cdot \vec{a_i} = m_i \cdot \frac{d^2\vec{r_i}}{dt^2} \tag{2.7}$$

The force $\vec{F}_i$ on each atom i in the system is calculated by differentiating the potential energy function $U(\vec{r})$ (2.8).

$$\vec{F}_i = -\nabla_i U(\vec{r}_i, ..., \vec{r}_N) \tag{2.8}$$

In practice this is implemented as an integration algorithm using short time intervals of 1-2 femtoseconds, which leads to the before mentioned MD trajectory.

During an MD simulation, the arrangement of atoms in the simulated system is continuously changed, and the potential energy changes accordingly. Usually this leads to a decreasing energy slowly striving towards a local or global minimum. The process of finding an energy minimum for a simulated system can be accelerated by energy minimization methods, which focus on reducing atomic forces regardless of a physically meaningful trajectory.

Among several software packages that are available for MD simulations, like Gromacs [15], Amber [29], CHARMM [24] or YASARA [110], the NAMD [155] software was chosen for all MD simulations in this work. Together with the closely interacting analysis and visualization tool VMD [90], the CHARMM36 [126] adaptive forcefield including parameters for lipids and the CHARMM general forcefield extension for drug-like molecules, NAMD provides a suitable platform featuring the capability for GPU computing.

2.3.1 Molecular Dynamics Test Case

In order to test the capabilities and features of the NAMD suite, data from the Na$_v$M ion channel crystal structures, co-crystallized with a set of small molecule ligands, were used as an example. The related publication from Bagneris et al. [8] discussed the properties of the crystal structures in detail and thereby provided suitable information for MD simulations, even though it was not described and maybe not considered by the authors. Consequently, Na$_v$M MD simulations were set up and executed in this study.

Na$_v$M is a prokaryotic homotetrameric sodium channel containing the pore-forming helices S5 and S6, but missing the voltage sensing unit. Through experiments with brominated ligands, Bagneris et al. found four symmetrically located potential binding sites in the channel pore. The bromide atom of the ligands produces strong diffraction signals well recognizable in electron density maps.

The provided x-ray structures were used to evaluate the setup of a molecular dynamics system, including the channel protein, a lipid bilayer patch, water and ions using VMD as prerequisite for a MD simulation in NAMD. First, the crystal structure 4CBC (2.67 Å, apo) was inserted into a 70 Å^2 POPC lipid membrane, created with VMD's integrated membrane builder plugin applying CHARMM36 topology parameters. Next, the protein and the surrounding membrane was placed into a TIP3P [98] water box of 70 x 70 x 90 Å, and the system was adjusted to a neutralizing ion concentration by replacing random waters with Na$^+$ and Cl$^-$ ions, resulting in a total number of around 46,000 atoms.

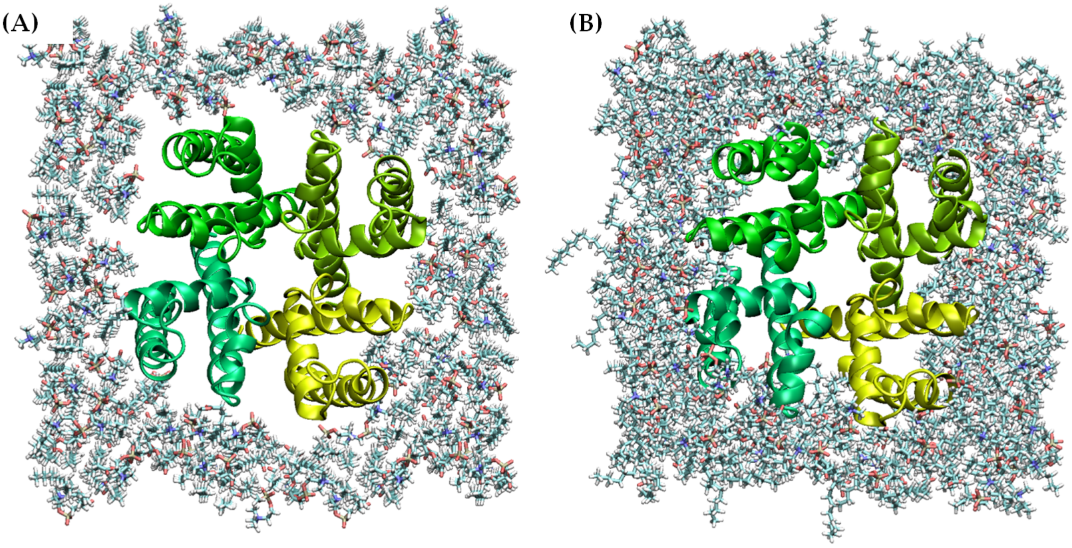


Figure 2.3: Na$_v$M in POPC membrane patch. (A) Na$_v$M structure (green colours) embedded in a membrane patch. Visible gaps are caused by removed protein overlapping lipids. (B) Energy minimized system, showing a tight packing between Na$_v$M and membrane lipids.

All MD simulations were performed using NAMD version 2.9 [155] applying the CHARMM36 forcefield parameters including the additive parameters for lipids and ions, such as sodium cations and chloride anions. As shown in Figure 2.3 (A), the initial MD system contained gaps between the protein and the unrelaxed membrane patch created by VMD's membrane builder. A close packing between Na$_v$M and membrane lipids needed to be achieved by relaxation and energy minimization. To remove restraints and frictions from the MD system, a stepwise protocol was used, oriented at NAMD's membrane proteins tutorial [3].

First the membrane lipid tails were relaxed with 1000 minimization steps, using the Newton-Raphson conjugate gradient algorithm [217] and a 0.5 ns dynamics run, while keeping all other atoms fixed. The same approach was then applied to the lipid tails after removing their constraints, to reduce gaps between channel TMs and lipid molecules. Constraints on all other atoms of the system were subsequently reduced to minimize and equilibrate the protein's environment further. At first by releasing water molecules to freely distribute in the simulation box, except for entering spaces of the protein lipid interface, and finally by dropping all constraints to relax the whole system as preparation for the productive simulation. The subsequent productive simulations were performed for 20 ns without any constraints, using a 2 fs time step and periodic boundary conditions (PBC). The Particle Mesh Ewald (PME) [45] method and the Shake [168] algorithm were used for all atoms. The cutoff for van der Waals and long-range electrostatic interactions was set to 12 Å. A constant temperature of 300 K was kept using a Langevin thermostat [213], and the constant pressure at 1 atm was controlled by Langevin piston [67] Nosé-Hoover [132] method.

The final coordinates of the Na$_v$M MD, showing a compact and relaxed conformation (Figure 2.3 (B)), were used as starting-point for subsequent simulations, which included the ligand lamotrigine as described among other ligands in the Na$_v$M publication. Lamotrigine was parameterized for the CHARMM FF using the MATCH [216] web server and was manually placed to one of the four proposed binding sites in the Na$_v$M channel pore. Residues T207 and F214 that were identified by mutagenesis to effect the channel block were in contact to the ligand when the simulation started, as shown in Figure 2.5 (A).

Two independent MD runs of 20 ns with random initial seeds were executed for the protein ligand complex. The RMSD traces for both complexes determined for Cα atoms and referring to the coordinates of the first frames show little variability in the global conformations and follow a similar profile in the course of the trajectory, as shown in Figure 2.4 (A). The traces in Figure 2.4 (B) show the RMSD of the lamotrigine ligands in both MD runs.

During the first phase of both simulations, lamotrigine stays at its initial position with main interactions to Met204 and Thr207 as exemplified by one of the starting poses shown in Figure 2.5 (A), until an abrupt conformational change occurs, indicated in the graph at around 8 ns. These steps in the RMSD traces correspond to lamotrigine slipping out of the binding site into the channel pore, which can be observed for both simulations.

This effect is also visualized by MOE's Protein-Ligand Interaction Fingerprints (PLIFs) in Figure 2.5 (C) for one of the MD simulations. Each row in the graph represents a MD frame in chronological order from top to bottom, whose

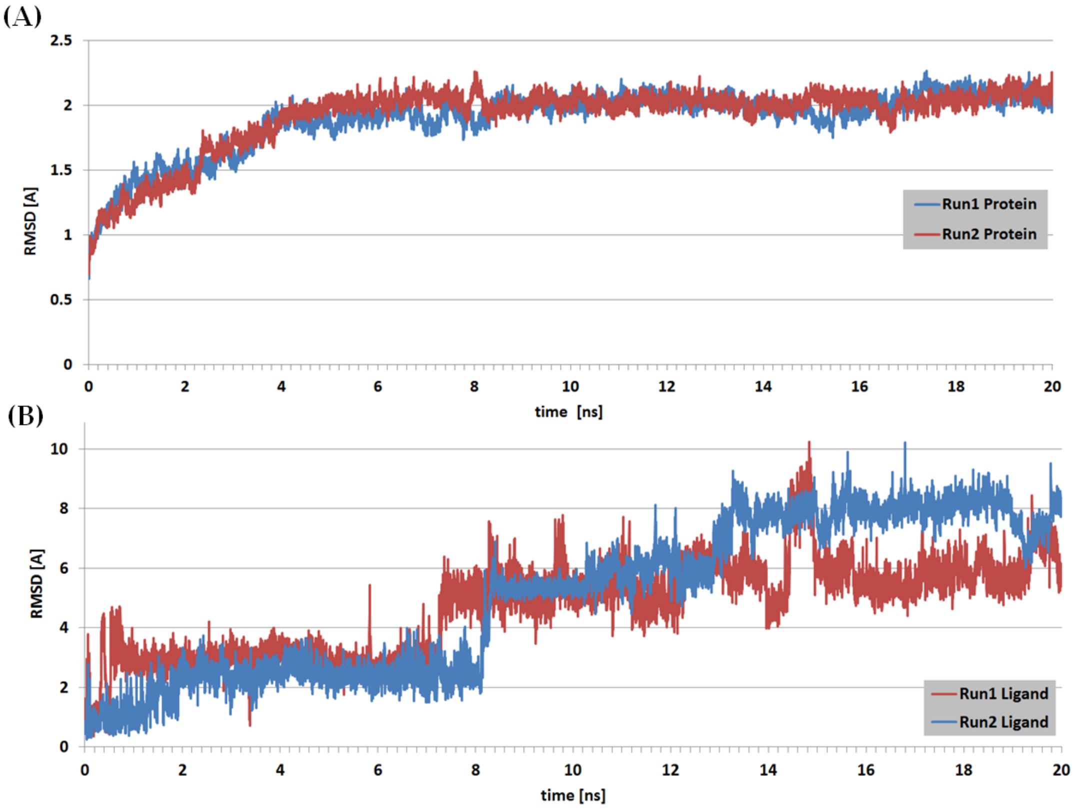


Figure 2.4: Trajectories of Na$_v$M MD simulation. (A) MD trajectories of the Na$_v$M complex. (B) RMSD traces of lamotrigine in two MD simulations.

interactions between the ligand and the receptor's residues are displayed as black lines in the respective row of the fingerprint. Every residue is assigned to one or more columns representing certain interaction types. The initial position of lamotrigine is illustrated in the plot by similar contact patterns that were found for the first frames. Then the protein ligand contacts suddenly change to Leu211 and Met175, which correlates to a positional change of the ligand in the channel while establishing new interactions, located towards the channel pore on another channel domain. The interaction to Met204 dissolved and lamotrigine tends to move towards a neighbouring binding site between the next domain pair. Another conformational change is indicated in the graph at around 14 ns, when the ligands in both simulations drift further into diverging positions. A final position of lamotrigine, resulting from one of the simulations, is shown in Figure 2.5 (B). The superposed temporary poses lamotrigine established during the MD simulation are shown in Figure 2.5 (D) demonstrating the drifting position of the ligand. By the end of the trajectory, lamotrigine and residues of the next binding site are in close distance,

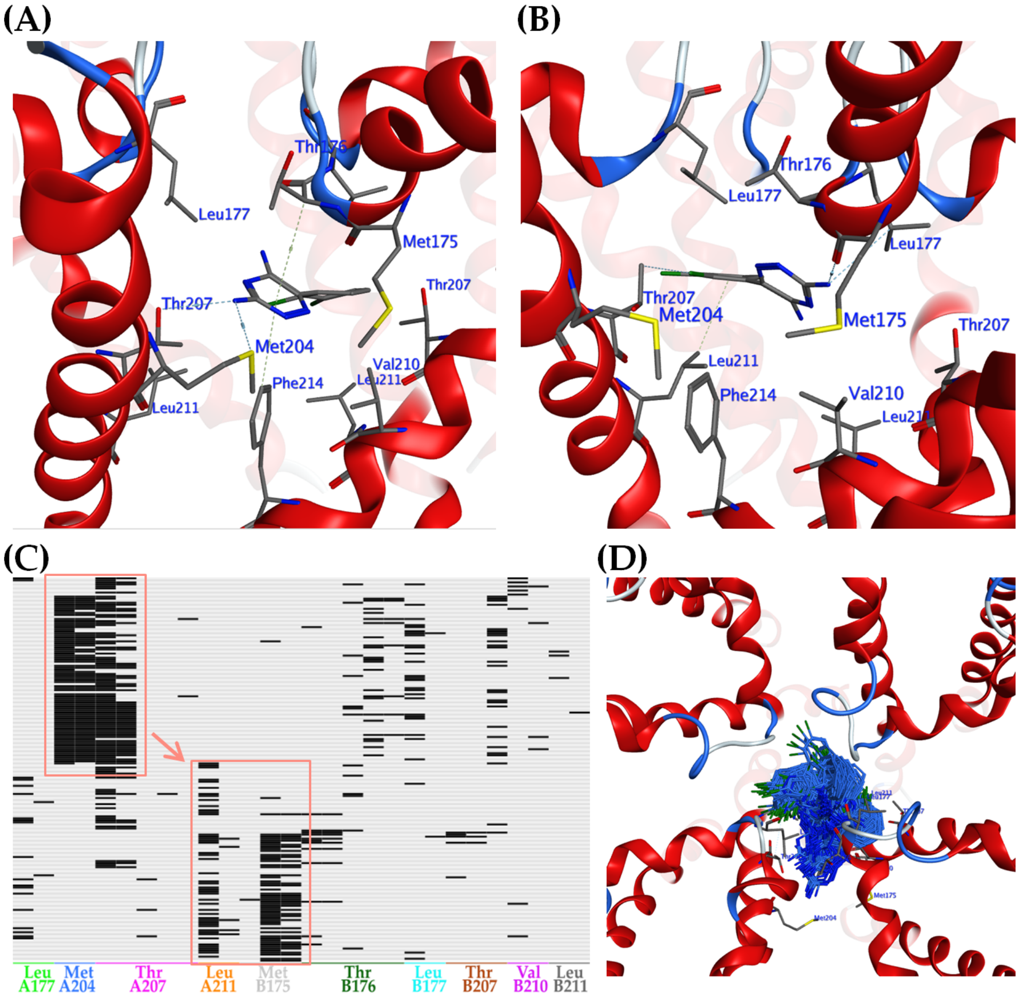


Figure 2.5: Lamotrigine in the Na$_v$M binding sites. (A) Lamotrigine was placed to the proposed binding site in Na$_v$M as starting position for the MD simulations. (B) Position of lamotrigine at the end of one of the MD simulations. (C) Protein-Ligand Interaction Fingerprints (PLIFs) illustrating the interactions between channel residues and the ligand in the course of the simulation. Each row represents a protein ligand pair and an interaction is indicated as a black line. Every residue is assigned to one or more columns representing certain interaction types. (D) Superposed lamotrigine binding poses of 197 MD trajectory frames, shown in top view.

creating evidence for a full transition of the ligand into the next cavity. This observation is in line with Bagneris et al. who proposed four identical binding sites between the channel's homologous domains with an occupancy rate of 0.25 for a single drug molecule.

To further explore the proposed position of the binding sites by another method, a docking experiment was performed using snapshots from the Na$_v$M MD simulation and the set of investigated ligands. Some general information about docking methods and the results of the Na_vM docking runs are described in the next section.

2.4 Molecular Docking

Molecular docking is an established computational approach to predict the conformation and position of a ligand in a binding site of the target receptor. Furthermore, it is used to estimate the interaction strength between a ligand and the residues in the receptor pocket. A scoring function evaluates the probability of a certain ligand pose of being realistic. The macromolecular structure can be derived from crystal structures, homology models or MD-simulation trajectories. Modern docking tools provide advanced features like flexible receptor binding sites, different scoring functions and the inclusion of solvation effects. Some algorithms are optimized for high throughput docking and are suitable for structure-based virtual screening [41], in which libraries of small molecules are docked into a target receptor to find potential new high affinity ligands for the protein of interest. A variety of docking suites and software packages are available for free or under commercial licenses, for example, AutoDock [142], AutoDock Vina [195], GOLD [97], Glide [70] and FlexX [85], to name some popular tools. A variation of the docking approach is used for protein-protein docking, to predict the formation of protein complexes by proposing the proteins' interface conformations and interactions. The proposed docking solutions are also ranked by a scoring function to estimate the most likely one. Known tools for protein-protein docking are, for example, ZDOCK [156] and HADDOCK [50].

Throughout this study AutoDock Vina was used for molecular docking of ion channel structures and their respective ligands, and to estimate the quality of structure models by matching docking scores and experimental data. AutoDock Vina uses a grid map based algorithm. The target receptor is assigned to a grid, and a probe atom is used to calculate the energy between the target and the probe by placing the probe atom consecutively onto every grid point. This generates a table of energies, which is applied to the receptor during the docking cycles. A Lamarckian genetic algorithm is used for conformational searching [141], followed by the prediction of binding free energies between the ligand and the macromolecular target in an empirical free energy force field.

Although great advances were made regarding speed and precision of the docking algorithms, their strength still lies in identifying reasonable docking poses rather than the evaluation of ligand-receptor affinities. As discussed in [160], high correlations of predicted and measured affinity data, that are based on common molecular docking approaches, are questionable and might be by chance. Especially binding properties of similar ligands like enantiomers, analysed in the cited study, are very challenging to discriminate. Furthermore the prediction of free energies of binding is estimated to have an error of ±2 kcal/mol [89], when using methods like Vina.

2.4.1 Molecular Docking Test Case

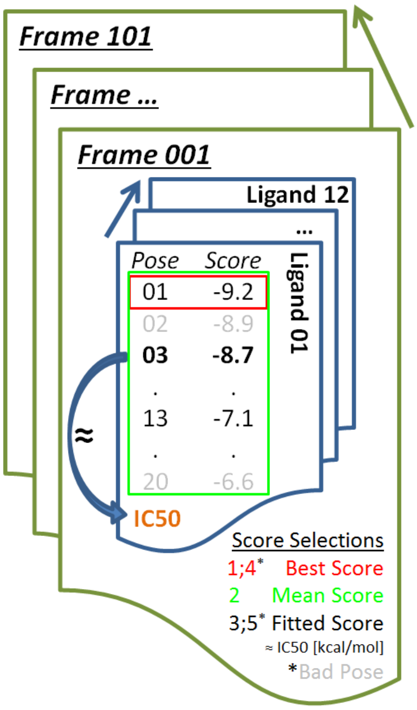


Figure 2.6: Data structure of docking results. Different result selections are color coded. The fitted scores are most similar to the IC_{50}. Selection 3 and 5 neglect bad poses not in contact with binding site residues.

To assess the usage of molecular docking with AutoDock Vina, a set of compounds listed in the afore mentioned Na_vM publication [8] was docked into various conformations of the Na_vM model structures, derived from the molecular dynamics simulation described in section 2.3. Due to the fact that the electrophysiologically measured IC_{50} data were published for the ligand set, an evaluation of Vina's accuracy for the specific protein class of ion channels was possible. The comparison between measured and predicted affinities should reveal an estimate of correlation that can be expected in later docking approaches on ion channels.

The Vina docking was performed for 101 Na_vM receptor structures extracted from the MD trajectory at evenly distributed time points. As the MD frames are different according to their backbone and side chain positions, with a mean RMSD of 1.432 Å across all frames and a maximum RMSD of 1.9 Å between the most diverse frames, the docking run enclosed structural variability of the receptor. This variability could mimic potential states of the residue side chains that might favour or reduce the chance for a ligand to interact with the binding site. This approach is similar to an inverse virtual screening (IVC), in which Vina is used to screen a large collection of target receptors against a library of compounds [116]. The grid box defining the area where the docking algorithm places the ligands and estimates the interactions to the receptor was limited to the pore region of the channel models, the proposed location of the ligand binding sites [8].

Each of the 1212 docking runs resulting from combining 101 frames and 12 ligands to unique ligand-receptor pairs delivered up to 20 scored docking poses with varying ligand orientations within the channel. The data structure of the results per ligand-receptor pair is illustrated in Figure 2.6. The results were analysed based on five different score selections, created by applying different selection criteria, displayed by color code in Figure 2.6. Those score sets were then correlated with the ligands' measured IC_{50} data [36, 76]. Under the assumtion that relative activities should correspond to relative free energies of binding for closely related ligands binding to the same site, the measured IC_{50} data were converted into approximate energy values ΔG [kcal/mol], applying the formula [49, 165, 201]

$$\Delta G \approx RT\ln(IC_{50})$$

where R is the gas constant of 1.986 x $10^{-3} kcal K^{-1} mol^{-1}$, and T is the room temperature of 298.15 K.

The formula was initially developed to translate K_i results to ΔG, and its applicability to IC_{50} values of VGICs was tested with a set of IC_{50} and K_i data from the ChEMBL Database [`https://www.ebi.ac.uk/chembl`][14], related to the VGSC family. The ChEMBL search returned compounds having both types of results, which allowed the correlation of the values, as shown in Figure 2.7, that were converted with the respective formula. The degree of correlation between IC_{50} and K_i data was excepted due to the absence of published K_i values for most of the VGIC. This procedure was in agreement with the good correlation of IC_{50} and K_i values Benjamin et al. [13] showed for $Na_v1.2$. Even though this application might only be an approximation, it was most feasible to allow a comparison between experimental and in silico data, as it was shown before in the referenced studies.

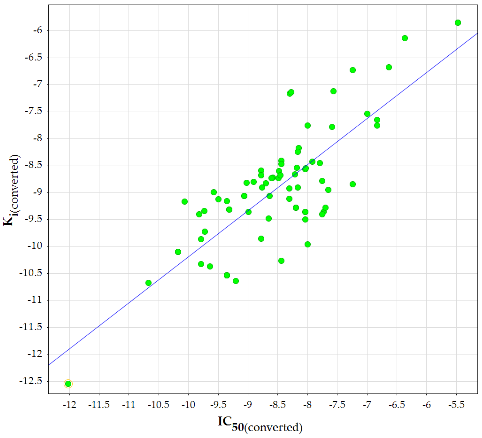


Figure 2.7: Correlation of converted Na_v IC_{50} and K_i values. The converted values showed a correlation of R^2 = 0.67.

The first analysed Na_vM docking result selection contained the best binding affinity score for each ligand-receptor pair. The second analysis was based on the mean values of the ligand-receptor related docking scores. The scores most similar to the measured IC_{50} data in kcal/mol were selected for the third analysis. The fourth and fifth analysed scoring value selection considered only scores of ligand poses that formed an interaction to at least one out of a set of specified residues. Despite this extra filter, the score selections were equivalent to the first and third selection, respectively.

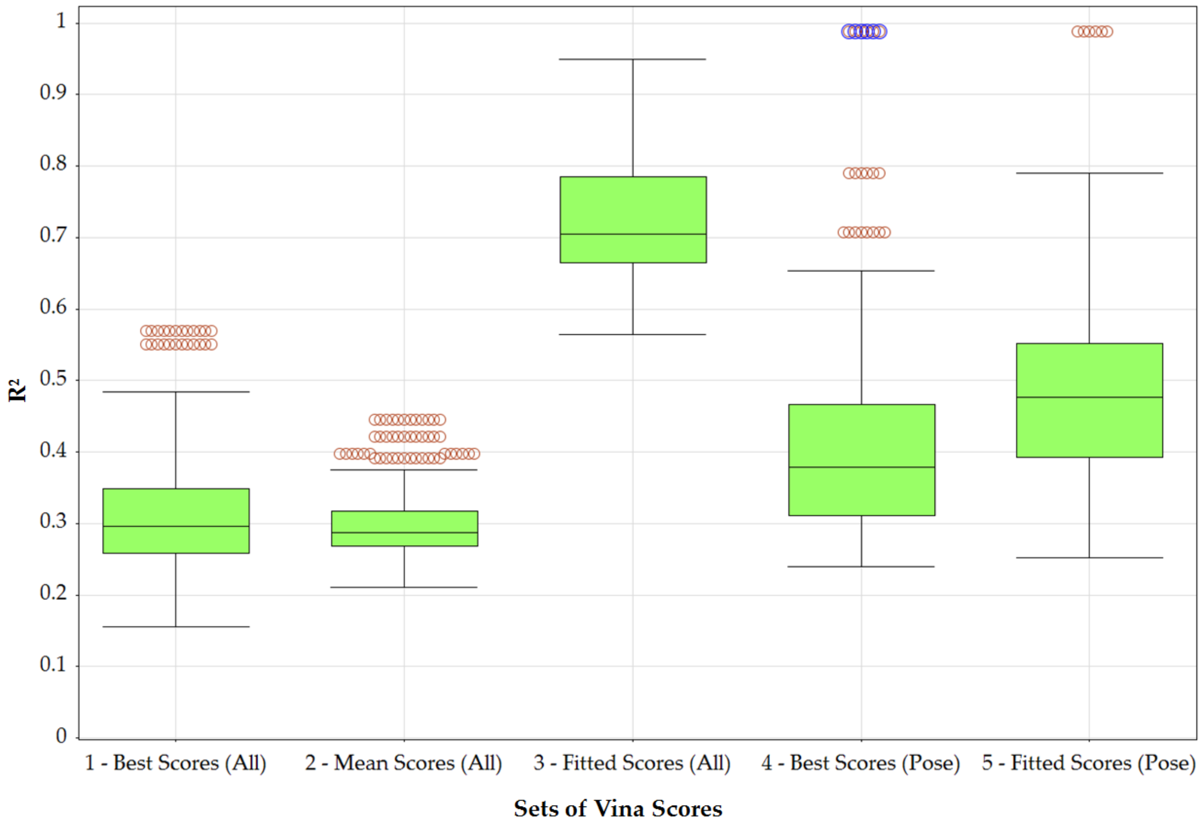


Figure 2.8: Distributions of Na_vM correlation coefficients. R^2s of measured IC_{50} data and five different sets of predicted binding affinity scores composed by applying different selection criteria. 1) Best docking score of each ligand-receptor pair, 2) mean of all docking scores of each ligand-receptor pair, 3) best fitting score, compared to measured data, of each ligand-receptor pair, 4) best docking score of specific poses for each ligand-receptor pair, 5) best fitting score of specific poses, compared to measured data, for each ligand-receptor pair. The outliers in 4 and 5 refer to correlations with reduced numbers of individuals caused by the pose filter.

The correlations between each selection of binding affinity scores and the IC_{50} data were described with Pearson's coefficient of determination R^2. The distributions of the R^2 values for each selection of docking scores are shown as box plots in Figure 2.8.

It was observed that the way of selecting docking scores from the total data set has an impact on the correlation with experimental data. As expected, the best correlations were achieved for docking score selection 3 of Figure 2.8, built by minimizing the difference between the calculated docking scores and the IC_{50} values, which is obviously a hypothetical approach to figure out the best possible correlations for the docking results. Only low correlations were seen when taking the best and mean docking scores, respectively, which misses a check for reasonable poses, such as neglecting high scored unrealistic ligand-receptor conformations. Whereas the correlations improved when focusing on ligand poses that involved certain binding site residues. As this latter analysis approach was most conclusive, it was adopted as standard analysis for later docking experiments, although the resulting mean R^2 remains below 0.5. A final comparative application of Spearman's rank correlation generated even worse results and therefore showed no improvement for the analysis. A reason for the low correlations in this particular test case might be the affinity range of the measured ligands of only 5.11 kcal/mol (Figure 2.9). In addition the individual ligands' affinity measures were not evenly distributed throughout the affinity range, and small affinity differences could not be resolved considering the proposed accuracy of ±2 kcal/mol of Vina's scoring algorithm.

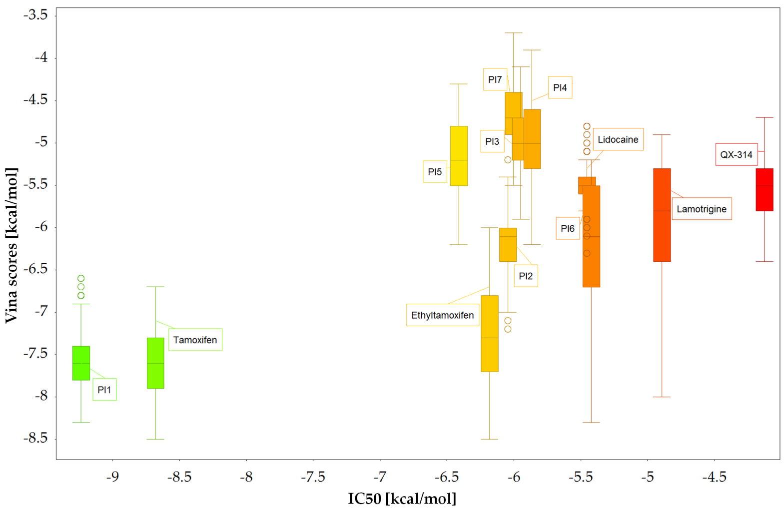


Figure 2.9: Na_vM IC_{50} data in comparison to Vina docking scores. Distributions of Vina scores are shown as box plots for each ligand. The range of measured affinities is not evenly occupied and some ligands show similar IC_{50} values. The color gradient indicates the IC_{50} range from low (green) to high (red).

Another reason might be the lack of strong local interactions in the binding site, also indicated by lamotrigine when leaving the binding site to enter another site during the MD simulation. With respect to the moderate potency of the docked ligands with IC_{50} measures between 178 nM and 455 μM, strong key interactions in a mostly hydrophobic environment were furthermore hard to detect for methods like Vina, which perform best in identifying binding poses as mentioned earlier [160]. Nevertheless, these docking experiments might generate supportive information about preferred ligand target pairs or the ligand binding response of different receptor conformations. But the ligand set needs to be diverse in structure and potency to increase the probability of being predictive.

2.5 Direct Coupling Analysis

The Direct Coupling Analysis (DCA) [139, 106, 147] is based on the concept of Direct Information (DI) that was first described by Weigt et al. [206]. Evolutionary Coupling (EC) [124] methods represent an alternative to energy estimates and conformational measures, often used to analyse protein structures. The DCA method makes only use of the evolutionary information contained in the protein sequence. It was observed that functionally coupled residues, which build strong conserved interactions, undergo coordinated evolutionary changes. These changes can be found in a Multiple Sequence Alignment (MSA) of a protein family of interest. Often the co-evolution of a residue pair maintains or even refines certain interactions necessary for the function of the protein. For example, mutations that occur at one residue position may result in compensatory mutations at other, structurally interacting residue positions during evolution. In a similar way, cross mutations in site directed mutagenesis experiments are often used to show direct interaction between two residues.

The method can be used to build a three dimensional model and is most applicable to proteins without known structure. Based on key residue pairs in the protein sequence showing a strong evolutionary coupling, it can be predicted that these residues are most likely in close three-dimensional distance to each other. These constraints supply valuable information to increase accuracy in model design.

A couple of EC methods are provided via web services, which use different DCA algorithms or combinations of those. While early local statistical models were not able to distinguish between direct and indirect residue interactions, which lead to misinterpretation of structural models, more accurate global statistical models were developed while the number of available sequences increased. These sophisticated methods, like the pseudo likelihood maximization Direct Coupling Analysis (plmDCA) and multivariate Gaussian modeling [10] can distinguish direct from indirect residue couplings. The RaptorX web server [205] combines the group graphical lasso method for contact prediction with evolutionary coupling analysis and Random Forest machine learning. The supervised learning method allowed accurate suggestions even for proteins with limited available sequence homologs. Other hybrid methods include DCA results as constraints to generate structural ensembles from MD simulations [187], indicating its contribution to *de novo* structure prediction.

3 Modeling of 5-HT$_{3A}$ Domain Association

The work presented in this chapter was performed in collaboration with the lab of Prof. Schmalzing, heading the department of molecular pharmacology at the RWTH Aachen. The Schmalzing group designed and executed all assay based experiments, and provided data based on the results of alanine scan mutation and expression experiments. Accounting to these data, computational methods were used to predict and design structural models that fulfil the observed constraints of certain interacting residues.

3.1 Introduction

The assembly of membrane proteins in the lipid bilayer was investigated in various studies [20, 208]. The commonly proposed model of a two-stage mechanism of protein folding is the first step in this assembly process, followed by events like oligomerization, loop folding or binding of prosthetic groups [61]. The process starts with the formation of α-helices in the hydrophobic environment of the lipid membrane. This is followed by the second stage, in which the individually stable helices developed protein-protein interactions to build higher order structures, such as homo- and hetero-oligomeric assemblies or subunits of a single receptor. This principal of protein folding is proposed for oligomeric membrane proteins, such as pentameric LGICs [1.2.3.2].

As described earlier [1.2.3.2] LGICs are neurotransmitter receptors that convert chemical signals into membrane potential changes, and their malfunction leads to various severe effects like psychiatric or neurodegenerative disorders or muscle dystrophy [188]. Their general structure (Figure 1.7) is composed of five identical or homologous subunits, each containing four membrane spanning helices M1-M4. The M2 helices are the actual channel forming segments lining the pore.

Members of the LGIC super family are, for example, the nicotinic acetylcholine receptor (nAChRs) [199], prokaryotic pentameric LGIC (ELIC) [84], human glycine receptor (hGlyR) [88] and the serotonin-gated 5-hydroxytryptamine receptor 3A (5-HT_{3A}) [81].

The 5-HT_{3A} receptor was used as an example throughout the study to analyse the helices' domain assembly. In particular, the involvement of leucin-zipper [77] and aromatic network interactions between transmembrane helices was investigated. The formation of an interhelical network of aromatic residues was already proposed for nAChR [78]. Similar to these results, the interaction between the transmembrane helices were also highlighted as a requirement for function and surface expression of α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors [171].

The key residues for the domain assembly of the 5-HT_{3A} receptor were identified with alanine scanning experiments and the analysis of the protein mutants' expression. In addition, the serotonin response was determined in terms of EC_{50} data as well as electrophysiological measures of the maximal currents I_{max}. In contrast to the expression values, the variations in EC_{50} and I_{max} measures were not related to a modified binding to serotonin, as its binding site is located on the β-subunit of the channel, not directly affected by the alanine scanning mutations.

At the time of the study the x-ray structure of m5-HT_{3A} was not published, and the computational analyses and predictions were based on homology models built on related protein structures. Now, the solved crystal structure provides a good opportunity for result validation, as discussed at the end of this chapter.

3.2 Experimental Basis

3.2.1 Mutation, Expression and Purification Methods

The assay experiments were conducted with *Xenopus laevis* oocytes, used to express different m5-HT_{3A} receptor subunit mutants, which contained single alanine substitutions (alanine scanning)[207] in sequence areas predicted to build transmembrane helices. For later purification of the expressed protein mutants via affinity chromatography, a C-terminal hexa histidine tag was fused to the sequences [21]. The expressed proteins were also metabolically labelled with ^{35}S isotope, that is incorporated into the protein during overnight incubation in L-[^{35}S]-methionine reagent [64]. The cells were further treated with the membrane-impermeant fluorescent cyanine (Cy5) dye

[19], additionally labelling the cell surface expressed proteins. Then, the membrane proteins were extracted by cell lysis, using the mild non-ionic lysis detergent digitonin [220]. The lysate was purified on a metal affinity resin, and the proteins were finally released in non-denaturing buffer. The oligomeric state of the purified protein mutants were analysed by blue native PAGE (BN-PAGE) along with the analysis of surface and total expression, respectively, by SDS-PAGE.

The same 5-HT$_{3A}$ receptor mutants were expressed for two-electrode voltage clamp (TEVC) experiments, recording the mutants' current traces caused by the channel response at saturating serotonin concentration (1 μM) in comparison to the wildtype.

3.2.2 Alanine Scanning Results

Analysing the 26 residue substitutions to alanine (Y456-W481) in the M4 transmembrane helix of 5-HT$_{3A}$, a high rate of successfully assembled mutant subunits was identified in BN-PAGE (Figure 3.1 A). Further, the functional protein assemblies were also recorded at the cell membrane by SDS-urea-PAGE (Figure 3.1 B). Only four mutants (V436A, D438A, L446A and L456A) were not able to assemble a recognizable number of stable and functional conformations, observed by missing expression signals in PAGE results shown in Figure 3.1 (A-B). These four mutations also reduced the total protein expression level of these mutants as seen in Figure 3.1 (C). However, the TEVC recordings detected currents (Figure 3.1 D) for two of these mutants, which means that at least some functionally assembled channels were built despite the point mutation. The measured data for mutant L456A were equal to those from wild-type channel, but for V436A the measures were significantly reduced (Table 3.1). For the latter it can not be distinguished, if the reduced number of channels or a mutation related reduction in functionality caused the low recorded current. No response in TEVC was recorded for D438A and L446A that were hence rated as non-functional in Table 3.1.

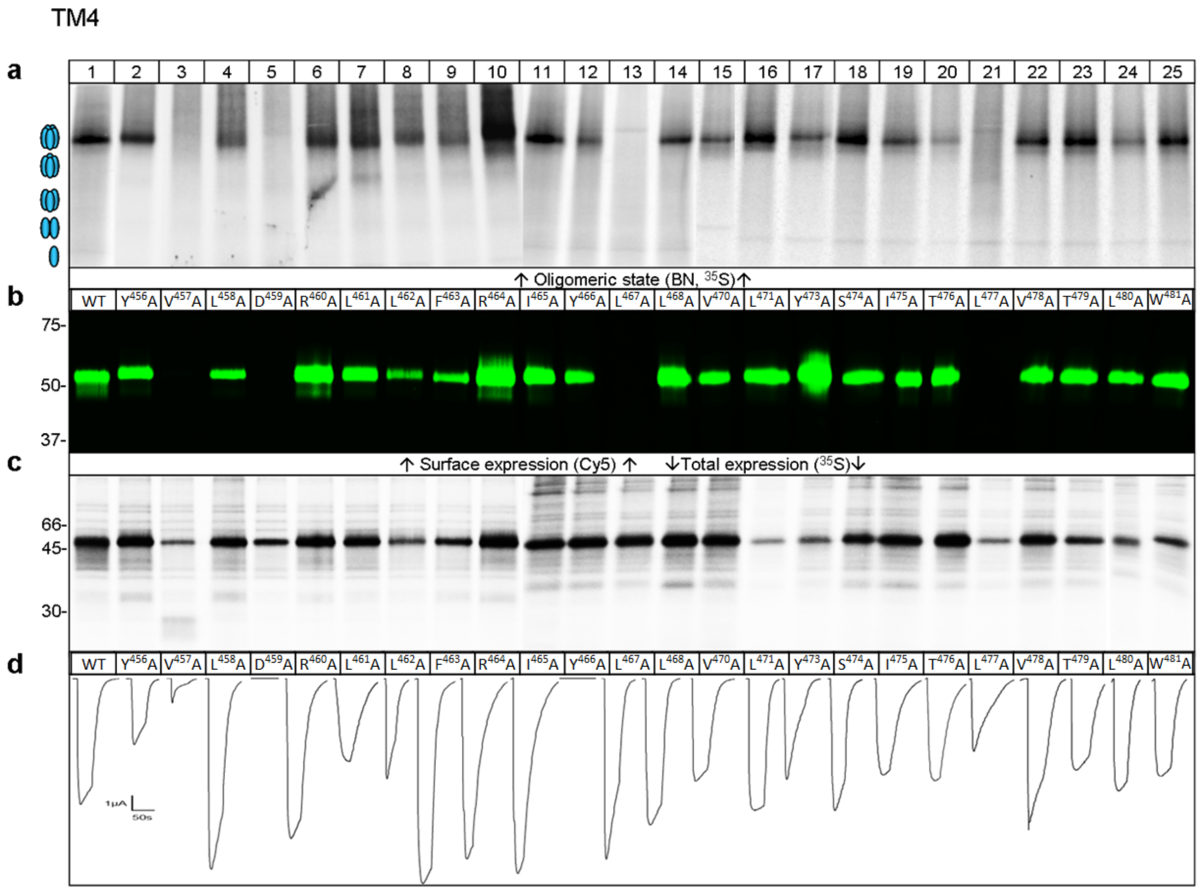
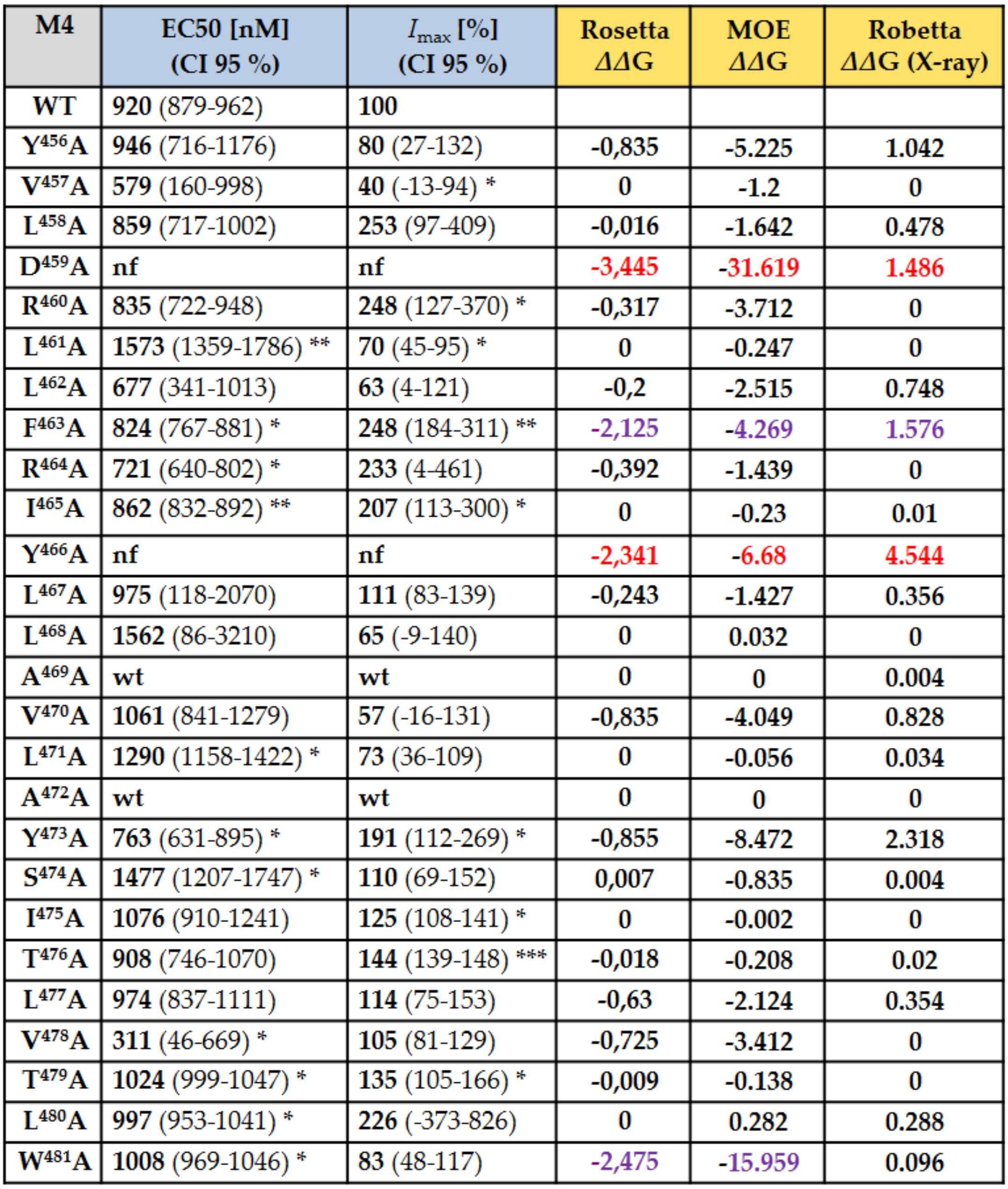


Figure 3.1: Effects of alanine substitutions in M4 of 5-HT$_{3A}$. Receptor mutants labeled with L-[35S]-methionine were resolved by BN-PAGE to display their oligomeric state (a). A reducing SDS-urea-PAGE displays Cy5-labeled surface expressed 5-HT$_{3A}$ mutants (b) and total amount of L-[35S]-methionine-labeled 5-HT$_{3A}$ mutant subunits expressed in the cells (c). Blue ovals near (a) schematically illustrate the oligomeric states, built by the five subunits. TEVC results on whole-oocytes in saturating concentration (1 μM) of serotonin elicited typical current traces (d). Each mutant was analysed at least twice with identical results. *With permission of Prof. Schmalzing*



M4	EC50 [nM] (CI 95 %)	I_{max} [%] (CI 95 %)	Rosetta ΔΔG	MOE ΔΔG	Robetta ΔΔG (X-ray)
WT	920 (879-962)	100			
Y^{456}A	946 (716-1176)	80 (27-132)	-0,835	-5.225	1.042
V^{457}A	579 (160-998)	40 (-13-94) *	0	-1.2	0
L^{458}A	859 (717-1002)	253 (97-409)	-0,016	-1.642	0.478
D^{459}A	nf	nf	-3,445	-31.619	1.486
R^{460}A	835 (722-948)	248 (127-370) *	-0,317	-3.712	0
L^{461}A	1573 (1359-1786) **	70 (45-95) *	0	-0.247	0
L^{462}A	677 (341-1013)	63 (4-121)	-0,2	-2.515	0.748
F^{463}A	824 (767-881) *	248 (184-311) **	-2,125	-4.269	1.576
R^{464}A	721 (640-802) *	233 (4-461)	-0,392	-1.439	0
I^{465}A	862 (832-892) **	207 (113-300) *	0	-0.23	0.01
Y^{466}A	nf	nf	-2,341	-6.68	4.544
L^{467}A	975 (118-2070)	111 (83-139)	-0,243	-1.427	0.356
L^{468}A	1562 (86-3210)	65 (-9-140)	0	0.032	0
A^{469}A	wt	wt	0	0	0.004
V^{470}A	1061 (841-1279)	57 (-16-131)	-0,835	-4.049	0.828
L^{471}A	1290 (1158-1422) *	73 (36-109)	0	-0.056	0.034
A^{472}A	wt	wt	0	0	0
Y^{473}A	763 (631-895) *	191 (112-269) *	-0,855	-8.472	2.318
S^{474}A	1477 (1207-1747) *	110 (69-152)	0,007	-0.835	0.004
I^{475}A	1076 (910-1241)	125 (108-141) *	0	-0.002	0
T^{476}A	908 (746-1070)	144 (139-148) ***	-0,018	-0.208	0.02
L^{477}A	974 (837-1111)	114 (75-153)	-0,63	-2.124	0.354
V^{478}A	311 (46-669) *	105 (81-129)	-0,725	-3.412	0
T^{479}A	1024 (999-1047) *	135 (105-166) *	-0,009	-0.138	0
L^{480}A	997 (953-1041) *	226 (-373-826)	0	0.282	0.288
W^{481}A	1008 (969-1046) *	83 (48-117)	-2,475	-15.959	0.096

Table 3.1: Experimental measures and in-silico results of 5-HT$_{3A}$ M4 mutants. The fist column indicates the generated protein mutants, followed by the experimental EC$_{50}$ and I_{max} data in the second and third column. The data represent means with a 95% confidence interval and significance level of $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). In-silico data are marked by yellow column headers. The results in terms of binding energy differences (ΔΔG) estimated with Rosetta and MOE were based on the m5-HT$_{3A}$ homology model. More negative values indicate a larger decrease in binding energy caused by the Ala mutation. Numbers indicating large energy differences are colored in red, when correlating to severe effects in an experiment, or in purple in case of moderate effects or potential false positives. The Robetta ΔΔG results are based on the m5-HT$_{3A}$ x-ray structure, and decreasing binding energies due to Ala mutations are indicated by numbers, increasing to more positive values.

3.3 Computational Modeling

The experimental results in hand, indicating some crucial mutations for the association and function of the m5-HT$_{3A}$ receptor, the next step was to investigate how the results translate to the structure of m5-HT$_{3A}$. As no structure of m5-HT$_{3A}$ was solved so far, structural insight had to be generated by building a homology model. This model was considered to represent the wildtype and was then mutated using computational alanine scanning according to the experimentally expressed mutants. The mutations' effect was finally analysed in respect to the experimental results.

The initial approach for the determination of energy changes caused by the alanine mutations was based on targeted MD simulation [174]. The transition between the complete m5-HT$_{3A}$ subunit and the separated helices (Figure 3.2) was guided by external forces during the simulation, pulling the M4 helix away from the M1-M3 helix bundle. In a subsequent step it was planned to calculate the free energy of binding from the potential of mean force (PMF), as described by Chen and Kuyucak [34] for a KcsA-Charybdotoxin complex and MacCullum et al. [125] for the association of α-helices. However, the simulation of a full thermodynamic cycle for each of the mutants turned out to be infeasible, taking into account the available computational capacity in the department.

Since it was shown that the conformational entropy has a small impact on side-chain conformation [87] and is primarily important for absolute free binding energy prediction rather than to estimate relative binding energies suitable for the correlation to binding affinities [86], the entropy was largely neglected for energy calculations in favour of a computationally less expensive approach. Therefore, the Rosetta software suite [117] was used, which provides the InterfaceAnalyzer application [118, 23] for the estimation of binding energy, expressed in Rosetta energy units (REU). The energy function is only partly physics-based (mostly for enthalpic energy terms), and was built on available crystal structures with the goal to create results that match experimental data [140]. In addition to Rosetta, a second energy estimation based on the potential energy differences of the bound and separated mutant subunit helices was calculated in MOE [33] using the Amber FF.

The resulting ΔG values of these two methods were compared to the experimental data with the attempt to match the observed results and guide the model creation process.

3.3.1 Modeling, Simulation and Energy Calculation Methods

3.3.1.1 Homology Modeling

At the time of this work no structural information such as x-ray or cryo-EM structures were available for the m5-HT$_{3A}$ receptor, and a homology model was generated as initial starting point. The target sequence of the murine 5-HT$_{3A}$ (UniProtKB: P23979) was aligned with the sequences of four possible templates, namely GLIC (PDB-Codes: 3P4W and 3EHZ), GluCl (PDB-Code: 3RHW) and nAChR (PDB-Code: 2BG9). The sequence alignment revealed a sequence identity of 10%, which is not sufficient to apply a straightforward modeling approach and energetic evaluation. Therefore, an iterative modeling approach was applied to include information from different related structures.

The GLIC structure 3EHZ was chosen for the first modeling iteration, showing the best x-ray resolution of 3.1 Å among the then available structures and a sequence identity of ∼12.5% towards the target sequence. The sequence alignment between the template and the model sequence was created using a modified version of the Needleman and Wunsch [146] alignment algorithm implemented in MOE (Ver. 2011.10) [33]. This algorithm generates an alignment by optimizing a function based on residue similarity scores. The function uses the amino acid substitution matrix BLOSUM62 [83] and applies penalty values of 7 for the opening and 1 for the extension of a gap in the alignment. To compensate the low sequence identity of ∼10% between the transmembrane domains of the target and template sequences, Uniprot's information derived from secondary structure predictions were used to refine the alignment. Therefore, the predicted transmembrane helical parts of the target sequence were manually aligned to corresponding regions of the template sequence. An initial set of homology models of the murine 5-HT$_{3A}$ receptor was created based on this alignment to the GLIC template 3EHZ. The standard modeling process [145], as implemented in MOE, included the transfer of initial coordinates from the template chains to the aligned target chains, followed by the modeling of target sequence parts lacking corresponding template coordinates.

The homology models from this first modeling iteration contained the transmembrane helices of interest (M1-M4), but omitted the large extra- and intracellular loops, as they do not functionally contribute to the assembly of the receptor [163].

According to the relatively low sequence identity, a consensus modeling approach was applied in a second iteration, which also included the three aforementioned templates of GLIC, GluCl and nAChR and allowed the partial refinement of the initial models. However, the m5-HT$_{3A}$ target sequence of the M4 helix did not fit well on the consensus template structures, and no structure coordinates were available to match the endings of the M4 model helix. Nevertheless, this helix part contains interesting residues analysed in the mutagenesis experiment, and was therefore modelled into a general α-helix conformation as proposed by secondary structure predictions from Uniprot. Finally, the created homology models contained all relevant residues for which experimental results are available.

The modeling approach generated 250 intermediate homology models, protonated with the Protonate3D algorithm in MOE [113]. Further refinement of the models followed, using smooth graded energy minimization in an Amber99 force field, which was terminated when a root-mean-square gradient of < 0.01 Å was reached [186]. The electrostatic solvation energy was used to score the 250 models and was calculated using a generalized born-volume integral method [112].

The best m5-HT$_{3A}$ homology model was derived from these 250 intermediate models, according to the best energy score values and the overall structural conformation, which was analysed with Ramachandran plots.

3.3.1.2 MD Simulations

To test the m5-HT$_{3A}$ protein model's stability, a molecular dynamics simulation was performed using the NAMD molecular dynamics suite (Version 2.6) [155]. Therefore, an explicit 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid membrane patch of 46 Å x 46 Å was created using VMD's membrane builder plugin. In addition, a protein structure file (psf), as input for NAMD, needed to be generated for the m5-HT$_{3A}$ model using the *AutoPSF* plugin in VMD that applied CHARMM27 parameters for atoms, bonds and angels. Next, the protein model structure was combined with the membrane patch, both aligned to their principal axis, by removing all POPC phospholipid molecules overlapping with a protein atom. Finally, a surrounding TIP3P waterbox of 46 Å x 46 Å x 77 Å with a neutralizing concentration of Na^+ and Cl^- ions was built around the protein membrane system. After all, the whole system contained $\sim$13600 atoms entering the MD simulation.

To obtain a close packing between the embedded protein model and the phospholipid membrane, 1000 minimization steps and 0.5 ns of simulation were performed, during which all system atoms, except the lipid tails, were restrained [178]. Then, the lipid heads were released from restraints and relaxed using the same minimization and simulation process as in the step before, followed by a relaxation step with freely moving water molecules that were only kept from entering gaps in the protein or membrane. During the final relaxation run, all constraints of the system were dropped to prepare the system for production simulation run.

The production simulation run of ~25 ns was simulated without any constraints using a 1 fs time step and periodic boundary conditions (PBC). The Particle Mesh Ewald (PME) [45] method and the Shake [168] algorithm were used for all atoms. The cutoff for van der Waals and long-range electrostatic interactions was set to 12 Å. A constant temperature of 300 K was kept using a Langevin thermostat [213], and the constant pressure at 1 atm was controlled by Langevin piston [67] Nose-Hoover [132] method.

3.3.1.3 Interaction Energy Evaluations

The interaction energy was evaluated for computationally generated mutant models based on the m5-HT$_{3A}$ homology model. The mutant models correspond to the experimental alanine scanning protein mutants described previously. Computational alanine mutations of the model's M4 helix, resulted in 26 individual mutant models plus a wildtype version. The models were mutated with the MOE Protein Builder tool, which allows the substitution of residue side chains including options like minimization and repacking to optimize the side chain placement. To assess the energetic contribution of the individual M4 residues in the assembly of 5-HT$_{3A}$ channel subunits, the binding energy of the interhelical interface and the change in binding energy caused by alanine point mutations in the complex were calculated [207, 159]. The cycle in Figure 3.2 illustrates the different binding states of the domain helices. The

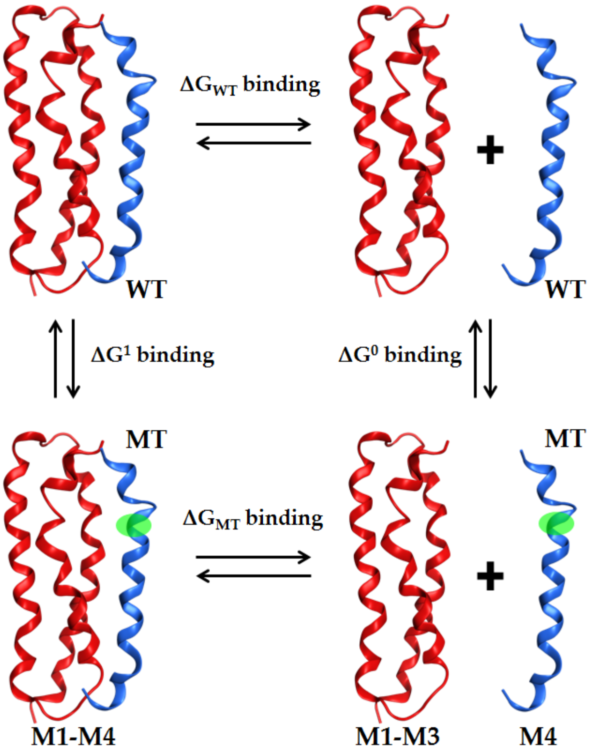


Figure 3.2: Estimation of binding energy ΔΔG. Wildtype (WT) and mutant (MT) in bound and unbound states.

binding energy changes ($\Delta\Delta$G) for the mutant models in relation to the wildtype could not be determined in a single step. First, the binding energy ΔG for every mutant model is derived by calculating the energy for the assembled M1-M4 subunit and for the separated subunits, when M4 is not bound to M1-M3, respectively (Figure 3.2). The energy difference of these two states equals the binding energy ΔG. The differences of binding energies ΔG between the mutant models and the wildtype correspond to the single alanine mutations shown as $\Delta\Delta$G values in Table 3.1. Residues on M4 with high $\Delta\Delta$G values are supposed to contribute strongly to the interhelical binding in the protein models, as the alanine substitute is a quite neutral residue and is usually maintaining only parts or none of the original interactions.

The $\Delta\Delta$G values in Table 3.1 originate from two approaches that were applied to 5-HT$_{3A}$ mutant models: 1) Rosetta InterfaceAnalyzer to examine the interface and its binding energy for the combined and separated subunit helices [118, 22], and 2) Potential energy evaluation of the combined and separated subunit helices in an Amber99 force field [202] as implemented in MOE.

The Rosetta InterfaceAnalyzer determines ΔG directly, when the interface is specified accordingly. The energies are derived with mostly knowledge-based energy functions, complemented with physics based enthalpy terms, as implemented in Rosetta. The model's interface is examined in consideration of the burial solvent-accessible surface area (SASA), and the number and location of unsatisfied hydrogen bonds in the interface.

The ΔG calculation in MOE needed individual steps. First, the energy for the complete model (M1-M4) was calculated. Then, the energies for the separated parts M1-M3 and M4, respectively, were determined. According to Figure 3.2 the ΔG was finally calculated out of the three energy values.

3.3.2 Molecular Modeling Results

During the MD simulation the 5-HT$_{3A}$ homology model showed a low conformational variability, indicated by a mean RMSD of 1.757 Å for nine snapshots picked from the trajectory at evenly distributed time points as shown in Figure 3.4. Moreover, no disturbing events were observed in the profiles of the different energy contributions of the simulated protein in the course of the trajectory shown in Figure 3.3. As the MD simulation was used to check the homology model for major steric or energetic issues, these results indicated the model to be suitable for further analysis. Consequently, the model was used to perform the recently described alanine scanning experiments, which produced the ΔΔG values shown in Table 3.1.

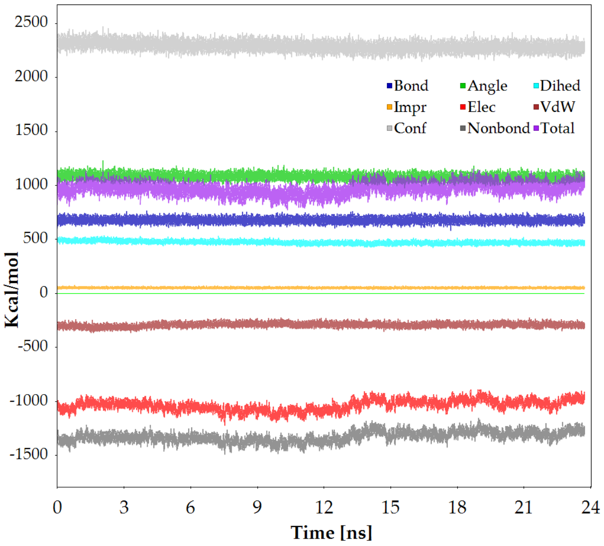


Figure 3.3: Energy traces of 5-HT$_{3A}$ MD simulation. The different energy contributions of the 5-HT$_{3A}$ protein model are shown in the course of the simulation.

Further, the concordance between the ΔΔG values and the measured effects of the experimentally mutated residues were investigated. The in-silico alanine scanning with Rosetta's InterfaceAnalyzer derived the two largest changes in binding energy (ΔΔG) for mutation D459A and Y466A. This observation matches the results from TEVC recordings, which could not measure a conduction in these channel mutants and rated them as non-functional (Table 3.1). In case of D459A, the high ΔΔG can be explained by a strong ionic interaction to R277, located on the opposite pore lining helix M2. This so-called ionic lock [11] was also observed between helices in G-protein coupled receptors (GPCR), stabilizing a certain conformation state. In one of these GPCRs, the adrenergic receptor (AR), the ionic lock was absent in the crystal structures, but the salt bridge (ionic lock) between an Arg and Glu formed during MD simulation [200], comparable to the energy minimizing steps in the here presented homology modeling approach. Such buried ionic interactions have a short distance range and contribute up to 5 kcal/mol [5] to the overall free energy.

For mutant Y466A, the 5-HT$_{3A}$ model illustrates a hydrophobic environment between M4 and M3, interacting with the aromatic side chain of Y466, which is slightly rotated away from the intrahelical opening between M1 and M3. Comparable interactions between aromatic and alkyl side chains (Leu, Val, Ile) have been shown to stabilize the intramolecular contacts with about 2 kcal/mol for each interaction [197].

The impact of these two mutants (D459A, Y466A) on the conformation and function of the receptor could imply that the loss of strong contacts between certain key residues leads to reduced protein assembly and the loss of channel function.

However, the F463A mutant was also rated with a high ΔΔG value, but the channel assembles into a functional conformation. In contrast to D459A and Y466A, the I_{max} current of F463A increased about 150% percent, but a nearly unchanged EC$_{50}$ value

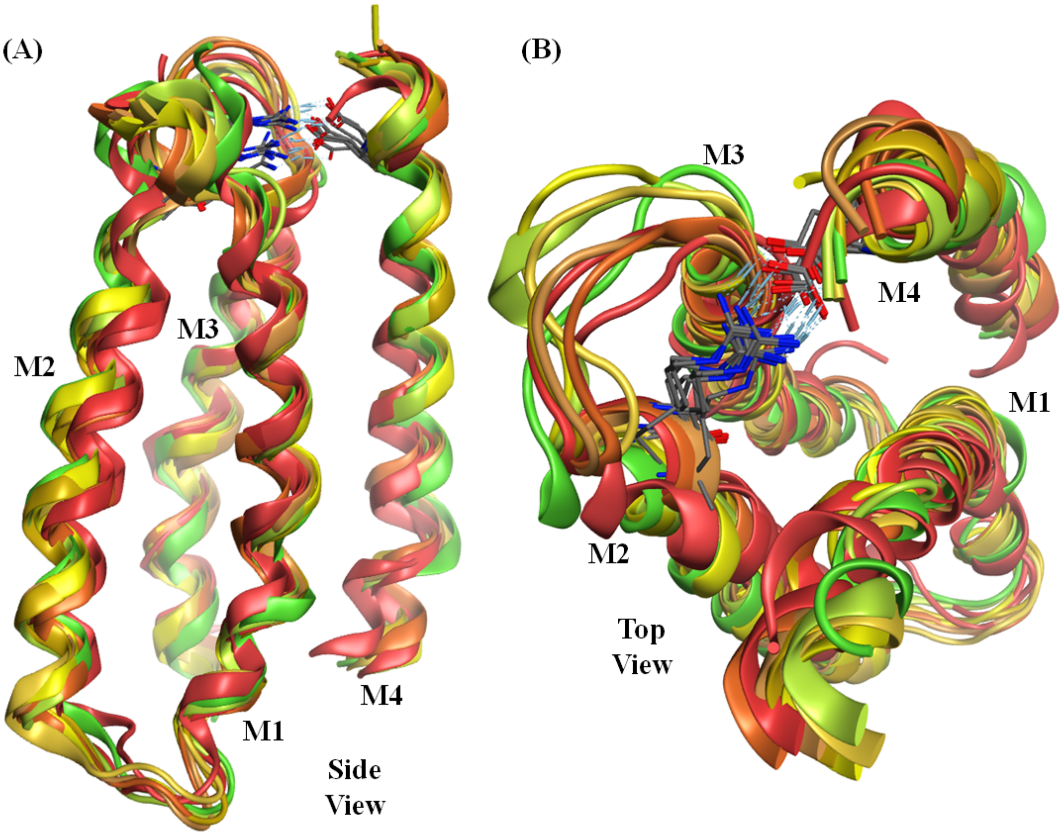


Figure 3.4: 5-HT$_{3A}$ MD snapshots overlay. Nine MD snapshots picked from equally distributed time points were superposed, showing a mean RMSD of 1.757 Å. The domain models are shown in side view (A) and from the top view perspective (B). The ionic lock interaction between D459 and R277 is present in all snapshots and indicated as blue lines.

was observed, compared to wildtype channel (Table 3.1). The aromatic side chain is again part of hydrophobic interaction, as described for Y466, but now between M4 and M1. The measures indicate that F463 as well as D459 and Y466 are involved in key interactions responsible for the correct function of the channel, but their mutation leads to oppositional channel behaviour.

While the ΔΔG and I_{max} values for the three aforementioned mutants could be discussed conclusively, only D459A of the four mutants (V457A, D459A, L467A, L477A), causing reduced or no subunit expression levels (Figure 3.1), correlates with in-silico alanine scanning. The measured EC$_{50}$ and I_{max} for L467A and L477A, respectively, show only minor differences compared to the wildtype, thus seem not to be involved in the channel's functionality. Their positions in the m5-HT$_{3A}$ structure model are not beneficial for strong interhelical interactions, so the model provides no evidence to explain the reduced expression observed in PAGE experiments. Also V457 was not identified by in-silico alanine scanning, causing no change in binding energy, and its side chain does not point to the helix interface, as captured in the structure model. However, the PAGE results showed a clear reduction of the assembly and surface expression rates for the mutant. The I_{max} is reduced by 60%, which indicates at least an involvement in channel function. But again, this could not be explained by the homology model.

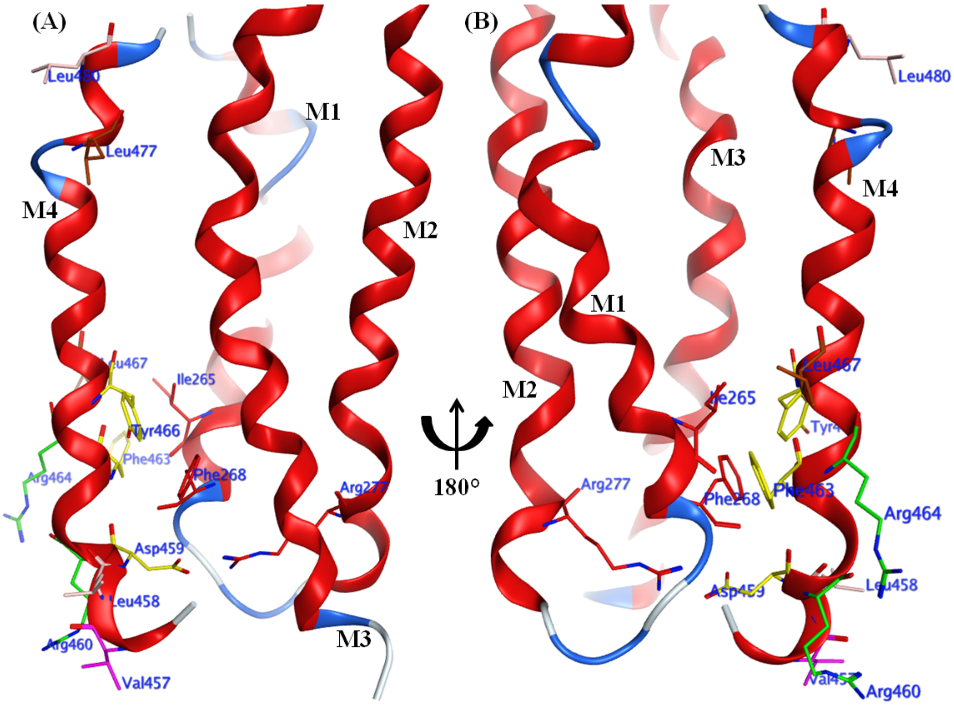


Figure 3.5: 5-HT$_{3A}$ model structure. The structure is displayed from two opposite sides, turned by 180°. All discussed residues are highlighted. The yellow residues showed the largest effects in mutagenesis experiments and energy estimations. The green arginines interact with lipid heads, and the pink valine connects to the MX helix, observed later in the 5-HT$_{3A}$ x-ray structure.

3.3.2.1 Comparison to m5-HT$_{3A}$ x-ray structure

In 2014 the first x-ray structure of the mouse 5-HT$_{3A}$ receptor (PDB-Code: 4PIR) was published [81] with a resolution of 3.5 Å, providing the opportunity to verify the experimental results and computational predictions. Thus, the x-ray structure was compared to the here reported m5-HT$_{3A}$ homology model, and the energy contribution of the M4 helix residues in the x-ray structure was analysed by alanine scanning using the Robetta web server [`http://robetta.bakerlab.org`][109]. The Robetta web server uses Rosetta software, and its results for the x-ray structure are comparable to the results of the m5-HT$_{3A}$ homology model, generated with the Rosetta InterfaceAnalyzer. The crystal structure was prepared for alanine scanning by adding missing residues to the short MX helix, followed by a short relaxation of the side chains with fixed backbone structure. All five subunits resolved in the crystal structure were used for alanine scanning, and Table 3.1 shows the mean values of their Robetta $\Delta\Delta G$ results along with the results for the m5-HT$_{3A}$ homology model.

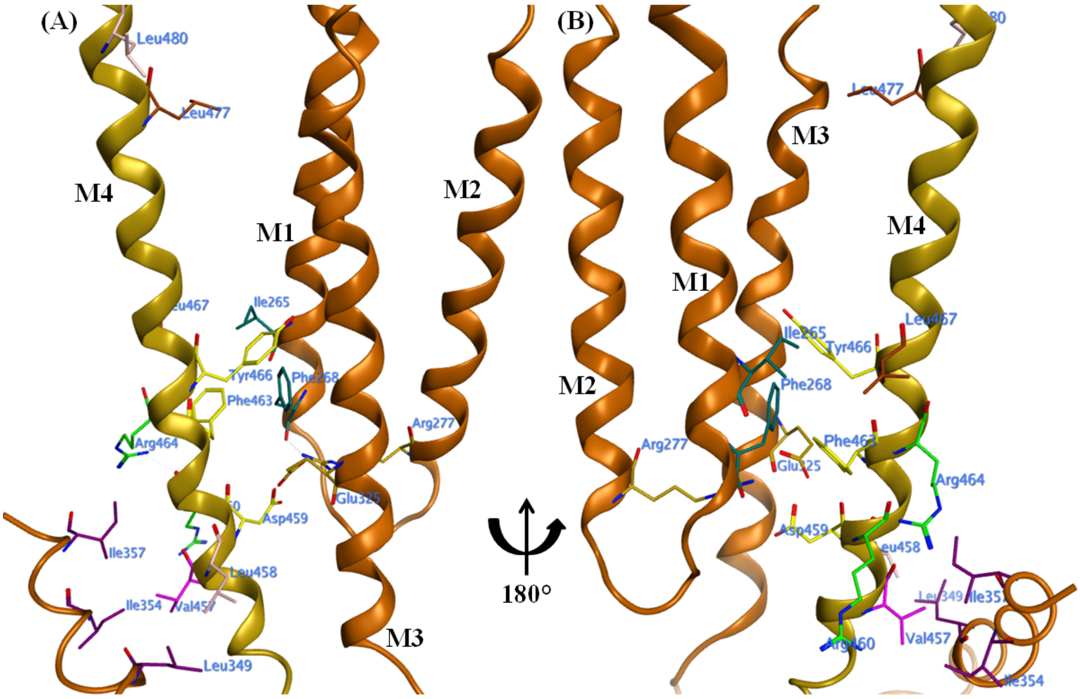


Figure 3.6: 5-HT$_{3A}$ x-ray structure. The structure is displayed from two opposite sides, turned by 180°. All discussed residues are highlighted. The yellow residues showed the largest effects in mutagenesis experiments and energy estimations. The green arginines interact with lipid heads, and the pink valine connects to residues of the MX helix. The other residues did not show noticeable effects.

The superposition of the M4 helix from the m5-HT$_{3A}$ x-ray structure and the homology model reveals a RMSD of 1.524 Å when ignoring the terminal residues. The terminal residues of the homology model did not adopt a native conformation during the minimization, because no stabilizing loops or helices are present. Therefore the residues' positions in the homology model diverge from their positions in the x-ray structure. In a structure superposition of the complete structures, a shift of approximately 1/2 helix turn in relation to the helix length is visible for M2 and M4. The structural comparison showed an overall RMSD of 2.994 Å.

The positions of the previously discussed residues in the homology model are in agreement with the m5-HT$_{3A}$ x-ray structure, supporting certain interactions as proposed by interface analysis. For instance, a variant of the ionic lock between D459 and R277, involving E325 from M3, can be observed in the x-ray structure. As described in [200], the full ionic lock might form by small backbone adaptations, which can be simulated with MD. The $\Delta\Delta$G of D459A in the x-ray structure is noticeable, but not as high as for the homology model. This could be caused by the incomplete ionic lock between D459 and R277, the interaction distance of which is not optimal in the x-ray structure.

The backbone position of Y466, whose mutation caused non functional channels as well, is nearly the same in the model and the x-ray structure, but in the latter the side chain points right into the cleft between M1 and M3. This position provides a favourable buried cavity for hydrophobic interactions, which is reflected by a higher $\Delta\Delta$G for Y466A, compared to the homology model.

The predicted location of the I_{max} boosting residue F463 in the homology model also matches the location in the x-ray structure, as well as its hydrophobic interaction to M1 residues like F268 or I265. The $\Delta\Delta$G calculated for the F463A mutation in the x-ray structure is one of the highest, although it is lower than for the homology model. The interacting F268 is described in [81] as conserved residue across the different available crystallized LGIC receptors. The location of F268 is in close distance to F463, while the F463A mutation was leading to a dramatic increase of I_{max}. This might indicate a critical contact, for example, when the channel opens, and the interaction is involved in limiting the I_{max} current or in reshaping the closed channel conformation.

Residues L467 and L477, inspected in the crystal structure, do not interact with residues from neighbouring helices. Only intrahelical contacts to side chains from M4 can be observed, which stabilize the helix but correspond to low $\Delta\Delta$Gs, indicating little effect of the mutations in the model. Although L467 and L477 might not contribute to interactions between helices, they are in a position that allows hydrophobic contacts to membrane lipids, effecting the channel stability as indicated by PAGE experiments.

New insights are provided by the m5-HT$_{3A}$ crystal structure regarding V457, which caused reduced expression levels when mutated to alanine. This was not reproducible by alanine scanning considering only the homology model. Now, an interaction of V457 with the MX helix could be proposed when inspecting the x-ray structure. The MX helix is connected to the post M3-loop and clamps the M4-MA helix to the helical domain bundle. Due to insufficient electron density of MX helix side-chains, only the backbone chain of the MX helix was placed in the structure [81]. After modeling the missing MX residue side-chains, a cavity around V457 occurred (Figure 3.7), created mainly by hydrophobic residues like L349, I354 and I357. These hydrophobic interactions might play an essential role during domain assembly by stabilizing the location of the MX helix, which acts like a clamp, tightening the contact between M1-M3 and M4, including M4's extending MA helix. The MX helix also seems to have a kind of shielding function, maintaining an hydrophobic environment for M4 residues located at the outer membrane region where the lipid heads might offer disturbing interactions. The expression data for this V457A mutant show a heavily reduced channel assembly, which adds to the assumption of an important interaction of V457. Nevertheless, the measures from electrophysiology show a small but existing current indicating some functional channels. The reduced current might be caused by the decreased number of assembled channels. In addition, the V457A mutation in the assembled channels might lead to malfunction in pore gating, also detectable in a reduced current.

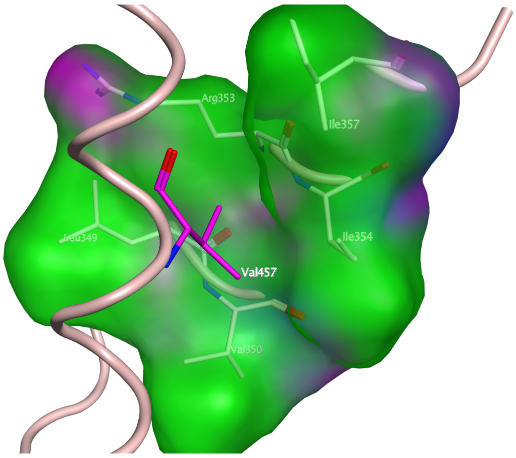


Figure 3.7: Val457 interacting to a hydrophobic pocket of the MX helix.

In general high I_{max} values can be observed mainly for m5-HT$_{3A}$ mutations of residues located in the outer parts of the M4 helix, such as L458, R460 and R464 in intracellular direction and L480 towards the extracellular surface. Their side chains point towards the membrane, possibly interacting with phospholipids. These interactions are not captured by the in-silico alanine scanning that included only the protein helices. However, the interactions could be observed during MD simulation of the homology model as illustrated in Figure 3.8 and in the lipid bilayer embedded m5-HT$_{3A}$ crystal structure, created and provided by the MemProtMD database [`http://sbcb.bioch.ox.ac.uk/memprotmd/beta/`].

The side chains of R460 and R464 build ionic interactions to the polar lipid heads of the phospholipids and hydrogen bonds with water molecules. In total, these bonds sum up to a high interaction energy, able to influence structural rearrangements. The interactions of residues like arginine or lysine with lipid phosphate groups contribute to the function and stabilization of channel domains within the membrane, as studied for potassium channels in molecular simulations [172, 175]. Further it was shown that certain lipids like PIP2 (Phosphatidylinositol 4,5-bisphosphate) even act like an agonist through binding to certain ion channels and modulating their gating mechanism [175]. Removing the interactions between the lipid heads and R460 and R464 by mutations to alanine caused increasing I_{max} currents, although the I_{max} measure for R464A in Table 3.1 shows no statistical significance. This large variance in the I_{max} measures is also true for L458A and L480A mutants. Despite this, their proposed position in the protein model and crystal structure indicate hydrophobic interactions of the leucines with the membrane. However, the function of L458 and L480 could not be clarified in this analysis.

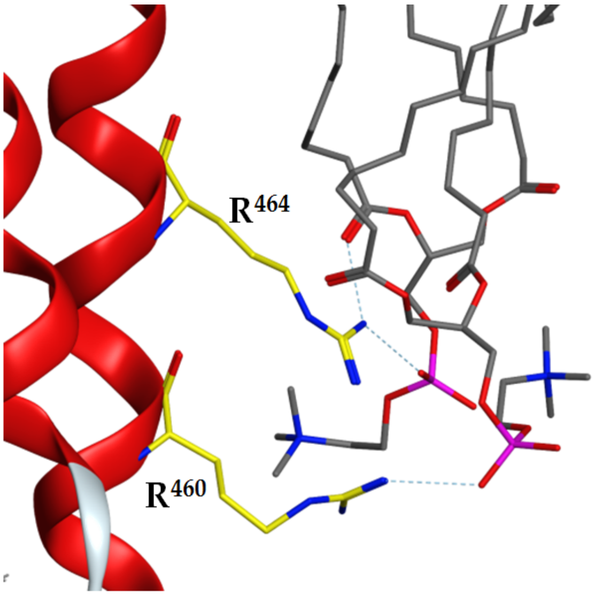


Figure 3.8: Arginin interaction to membrane phospholipid heads.

3.3.3 Conclusion

The results of the mutagenesis experiments were partially reproducible by estimates of the binding energies, based on a consensus homology model. The model was suitable to conclude certain key interactions, like the ionic lock between Asp459 and Arg277, not described previously for the 5-HT$_{3A}$ channel. Interactions of, for example, Arg460 and Arg464 to the phospholipid heads of the membrane were observed due to the short MD simulation used to check the model's stability. As the x-ray structure became available during this study, missing links between the mutagenesis data and the structure model were closed, like the interactions of Val457 with the MX helix, not present in the structure model. After structural comparison of the model and the x-ray structure, it can be concluded, that although the similarity of target and template sequences were low, a consensus homology modeling approach and the inclusion of mutagenesis data allowed the design of a reasonable homology model. However, the observed I_{max} increasing mutations could only be explained rudimentary, based on the structure model or the short MD simulation. With respect to some mutations associated with significant changes of I_{max}, one could speculate that these residues are involved in mechanisms like channel opening and regulation of the ion current or the desensitization process of terminating the ion conduction. On the other hand, parts of the I_{max} results show a large variance, limiting the basis for such assumptions.

In general the analysis of the m5-HT$_{3A}$ structure gave no evidence for the hypothesis of a leucine zipper motif, as described for other proteins [115, 77], which showed alternating leucine residue interactions between transmembrane helices. Most leucine residues of the M4 helix in the m5-HT$_{3A}$ model and x-ray are facing the membrane and do not interact with the M1-M3 protein helices. More likely is the formation of an aromatic network, as proposed by Haeger et al. [78] for Cys-loop receptors using GlyR α1 as an example. The x-ray of m5-HT$_{3A}$ shows some aromatic residues pointing towards neighbouring helices and the F463A and Y466A mutants had an impact on experimental measures and computational energy calculations. These mutagenesis experiments need to be extended to other helices to increase evidence for this domain stabilizing aromatic network. In addition, a mutation that substitutes Asp459 and Arg277 against each other could be useful to validate the observed ionic lock interaction.

4 Heteromeric Assembly of the Voltage-gated Sodium Channel $Na_v1.8$

4.1 Introduction

Working with voltage-gated and ligand-gated ion channels and learning about their fascinating mode of operation as well as engineered looking structure, continuously raised questions, which turned out to be still under discussion or need further investigations. One fundamental question that was barely addressed in literature relates to the domain associations of heterotetrameric voltage-gated ion channels. The topology of a voltage-gated sodium channel described earlier in Figure 1.6, which is built by a sequence, containing four interconnected domains that form the circular channel and pore, creates the question about the correct direction of domain association. The channel pore lining of heterotetrameric ion channels presents certain residues that build particular binding sites, which may vary extensively according to the direction of the channel assembly. This seems to be crucial for voltage-gated sodium channels, in contrast to homotetrameric channels with strictly symmetrical composition.

In the literature, only one group [58, 121] was found that investigated this question by performing mutant cycle analysis experiments using the neurotoxic peptide μ-conotoxin GIIIA [119] and the rat adult skeletal muscle Na^+ channel. Na^+ channel mutants and μ-conotoxin mutants were created in the experiments by altering key residues of the toxin and at the pore entry of the channel, presumably involved in μ-conotoxin interaction. The differences in the measured binding affinities between the toxin and channel mutants led to the conclusion of a clockwise channel arrangement. However, the presented data leave room for interpretations, because in some of the mutant experiments no results could be created, leading to an incomplete dataset. This fact and the lack of complementary experiments from other research groups gave the impulse to further explore this topic, using the computational methods described in chapter 2.

As sample for this analysis, the human voltage-gated sodium channel Na$_v$1.8 was selected, which is of interest as target for the pharmaceutical industry, because it is responsible for various diseases related to indications like pain and inflammation [190, 80].

To better understand the molecular features and interactions that contribute to the channel domain association, several homology models of the human sodium channel Na$_v$1.8 were built to study their domain assembly. To sample all possible arrangements of the four individual domains, models in clockwise, counter-clockwise, and four different crosswise orientations were built as illustrated in Figure 4.1. These variants are theoretically possible due to large interdomain loops with a minimum length of 50 AA that allows in principle to form various structural motives like up to 3-4 α-helices of average size [203]. For example, Jpred [56] predicted two helices and a sheet in the shortest cytoplasmic sequence between domain III and IV, all connected by loops and able to span the farthermost parts of the channel protein.

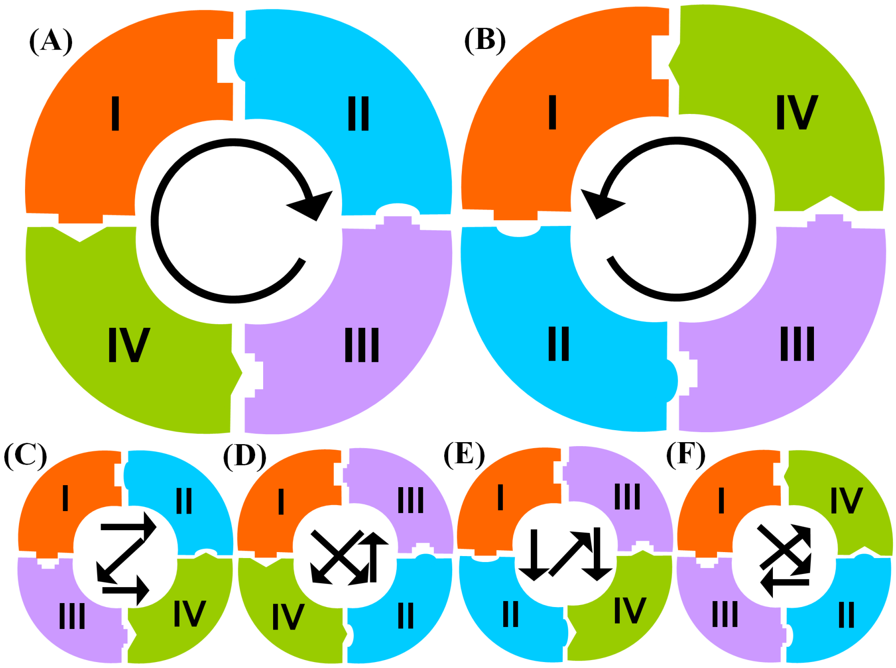


Figure 4.1: All possible Na$_v$1.8 domain associations. In the literature the ion channel domain associations of (A) clockwise (cw) and (B) counter-clockwise (ccw) are described. In addition the four crosswise associations (C) C1243, (D) C1324, (E) C1342 and (F) C1423 were considered in this study.

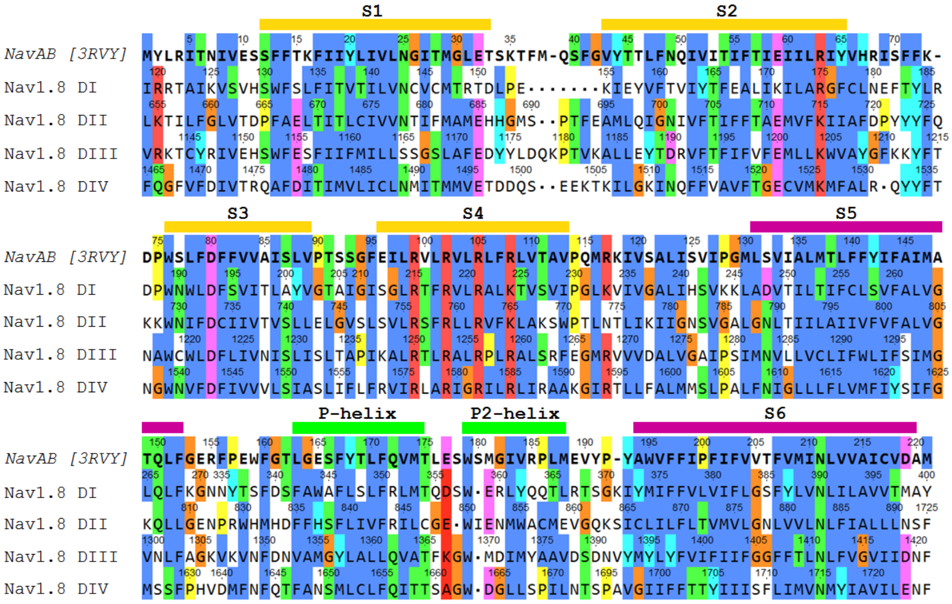
The μ-conotoxin GIIIA, used in the afore mentioned studies together with the rat adult skeletal muscle Na^+ channel, is not effective on $Na_v1.8$ [210] and therefore might not establish strong interactions to the channel. A direct in-silico reproduction of these experiments with $Na_v1.8$ was thus not reasonable. A workflow for building and analysing the six different domain association type models, composed out of different computational methods, is described in the next sections. The workflow starts with the sequence and template selection for the homology modeling of the different channel association types, followed by molecular dynamics simulations of the best ranked model from each of the six types. The simulations created trajectories, containing snapshots with sampled side chain conformations that were further clustered for domain interface analysis. In addition, the sequence based Direct Coupling Analysis (DCA) was applied to identify noticeable conserved residues that were further analysed regarding their contribution to interface interactions, which might be important for domain association. Finally, a set of known inhibitors were docked into the models to compare the binding properties of the channel models with literature data.

The last section describes the analysis of the $Na_v1.8$ models and the results compared to results generated for the cryo-EM structure of the VGSC related voltage-gated calcium channel (VGCC) $Ca_v1.1$, published in 2015 [212]. This comparison provided useful information about the model systems and allowed to estimate the limits of the applied computational methods.

4.2 Methods and Theoretical Basis

4.2.1 $Na_v1.8$ Sequence Alignment and Modeling

The construction of the human $Na_v1.8$ homology models started with the search for an appropriate protein structure template. As discussed earlier, the collection of structurally solved VGSCs in public domain is quite small and in fact, at the time of performing this analysis, the PDB contained only two suitable structures of VGSC related proteins, namely Na_vAb (PDB-Code: 3RVY) and Na_vM (PDB-Code: 4F4L). An overlay of both structures resulted in a good fit, yielding an RMSD of ∼1.45 Å. However the Na_vAb structure provided not only the two pore building helices (S5, S6) like Na_vM, but all six domain helices (S1 to S6). The Na_vAb structure resolution of 2.7 Å also surpassed that of the Na_vM structure with 3.49 Å, so Na_vAb was consequently picked as template for homology modeling. The human Na$_v$1.8 α-subunit sequence (Uniprot: Q9Y5Y9) was selected as target and manually aligned to the Na$_v$Ab sequence, according to known motives like the selectivity filter (DEKA) [120] and the positively charged arginines in the voltage sensing domain (S4), the P- and P2-helix, the linker between S4 and S5, as well as translated information from published alignments for modeling closely related sodium channel subtypes [215, 193, 194]. Refinements of the initial alignment, like handling gaps and inserts, were made with the MUSCLE sequence alignment software [59].



```
NavAB [3RVY]   MYLRITNIVESSFFTKFIIYLIVLNGITMGLETSKTFM-QSFGVYTTLFNQIVITIFTIEIILRIYVHRISFFK-
Nav1.8 DI      IRRTAIKVSVHSWFSLFITVTILVNCVCMTRTDLPE·······KIEYVFTVIYTFEALIKILARGFCLNEFTYLR
Nav1.8 DII     LKTILFGLVTDPFAELTITLCIVVNTIFMAMEHHGMS··PTFEAMLQIGNIVFTIFFTAEMVFKIIAFDPYYYFQ
Nav1.8 DIII    VRKTCYRIVEHSWFESFIIFMILLSSGSLAFEDYYLDQKPTVKALLEYTDRVFTFIFVFEMLLKWVAYGFKKYFT
Nav1.8 DIV     FQGFVFDIVTRQAFDITIMVLICLNMITMMVETDDQS··EEKTKILGKINQFFVAVFTGECVMKMFALR·QYYFT

NavAB [3RVY]   DPWSLFDFFVVAISLVPTSSGFEILRVLRVLRLFRLVTAVPQMRKIVSALISVIPGMLSVIALMTLFFYIFAIMA
Nav1.8 DI      DPWNWLDFSVITLAYVGTAIGISGLRTFRVLRALKTVSVIPGLKVIVGALIHSVKKLADVTILTIFCLSVFALVG
Nav1.8 DII     KKWNIFDCIIVTVSLLELGVSLSVLRSFRLLRVFKLAKSWPTLNTLIKIIGNSVGALGNLTIILAIIVFVFALVG
Nav1.8 DIII    NAWCWLDFLIVNISLISLTAPIKALRTLRALRPLRALSRFEGMRVVVDALVGAIPSIMNVLLVCLIFWLIFSIMG
Nav1.8 DIV     NGWNVFDFIVVVLSIASLIFLFRVIRLARIGRILRLIRAAKGIRTLLFALMMSLPALFNIGLLLFLVMFIYSIFG

NavAB [3RVY]   TQLFGERFPEWFGTLGESFYTLFQVMTLESWSMGIVRPLMEVYP-YAWVFFIPFIFVVTFVMINLVVAICVDAM
Nav1.8 DI      LQLFKGNNYTSFDSFAWAFLSLFRLMTQDSW·ERLYQQTLRTSGKIYMIFFVLVIFLGSFYLVNLILAVVTMAY
Nav1.8 DII     KQLLGENPRWHMHDFFHSFLIVFRILCGE·WIENMWACMEVGQKSICLILFLTVMVLGNLVVLNLFIALLLNSF
Nav1.8 DIII    VNLFAGKVKVNFDNVAMGYLALLQVATFKGW·MDIMYAAVDSDNVYMYLYFVIFIIFGGFFTLNLFVGVIIDNF
Nav1.8 DIV     MSSFPHVDMFNFQTFANSMLCLFQITTSAGW·DGLLSPILNTSPAVGIIFFTTYIIISFLIMVNMYIAVILENF
```

Figure 4.2: Alignment of Na$_v$1.8 sequence and Na$_v$Ab structure template. The sequence alignment of Na$_v$AB against human Na$_v$1.8 alpha-subunit sequence is shown in Clustal-X colour coding. It shows sodium channel specific conserved motives like the positively charged arginines of the voltage sensing domain (VSD) in S4, coloured in red, and the DEKA motive of the selectivity filter between P- and P2-helix. The large intra- and extracellular loops of the Na$_v$1.8 sequence were truncated, as they are not present in the crystal structure for modeling.

The large intra- and extracellular loops of the sequence could not be mapped to the template, as they are not present in the crystal structure. Consequently these loops were truncated while maintaining all domain interactions in the transmembrane area and the small molecule binding sites, respectively, that are mainly located near the helices [143]. Thus, the initial alignment similarity of ~30% between Na$_v$1.8 sequence and Na$_v$Ab structure template was substantially facilitated as shown in Figure 4.2.

All alignments according to the six different domain associations (Figure 4.1) have been generated for the subsequent homology modeling step. For each domain association type, 200 models were built using the *allhmodel()* class derived from the *automodel()* class of MODELLER 9v8 [69] that builds all-hydrogen models and was executed in parallel mode. MODELLER needs a variety of input files; the sequence alignment file needs to be in .pir format, indicating the template and target sequences and files in the first line of each alignment block. Although the Na$_v$1.8 channel is composed by one long sequence, the modeling was set up like for a multi chain model, matching the truncated and hence split sequence of Na$_v$1.8 DI-DIV separately to the Na$_v$AB template sequence.

All 200 models of each association type were initially ranked by MODELLER scores followed by manual inspection of Ramachandran plots with a special focus on the orientation of residue side chains, such as Tyr1717, whose interaction with small molecules have been observed by site directed mutagenesis experiments [25]. Table 4.1 shows a list of these residues whose side chains are assumed to be exposed to the channel pore, accessible for ligand binding.

Residue No.	Domain	Position	Reference
D 356	I	DEKA	[173]
I 381	I	LA site	[25]
N 390	I	LA site	[25, 8]
L 393	I	LA site	[8]
E 849	II	DEKA	[173]
K 1367	III	DEKA	[173]
L 1410	III	LA site	[25]
N 1411	III	LA site	[25, 8]
V 1414	III	LA site	[25, 8]
I 1433	III	Inactivation gate	[173]
F 1434	III	Inactivation gate	[173]
A 1661	IV	DEKA	[173]
I 1706	IV	LA site	[25, 8]
F 1710	IV	LA site	[173, 25, 8]
Y 1717	IV	LA site	[173, 25, 8]

Table 4.1: List of key Na$_v$1.8 residues. Residues involved in the local anaesthetics (LA) binding site or part of the DEKA motive of the channel's selectivity filter, identified by mutagenesis experiments.

4.2.2 Molecular Dynamics

Molecular Dynamics simulations were conducted for each top ranked model representing one of the six association types. The models were prepared using MOE [33] and VMD 1.9.2 [90] by executing the following steps.

First, all model pdb files coming from MODELLER had to be edited, which implied to provide proper chain names, residue names, and even the numbers and tags of hydrogens had to be adapted according to the CHARMM36 force field from the MacKerell Lab [126] used in NAMD [155]. To be applicable as input for NAMD, additional protein structure files (PSF) needed to be generated by the *AutoPSF* plug-in provided in VMD. This tool applies CHARMM36 parameters to pdb coordinate files and creates .psf files that contain sections of information for atoms, bonds, angles etc., describing properties like charge or mass of atoms and certain bonds for a list of atom pairs in the molecule. Next, an explicit POPC (1-Palmitoyl-2-oleoylphosphatidylcholine) phospholipid bilayer membrane patch with the dimensions of 120Å x 120Å was created with VMD's integrated membrane builder plug-in, applying CHARMM36 topology parameters. Each of the six channel models were aligned to a copy of the membrane patch by their principal axis, placing the model in the centre of the patch, leaving at least 15 Å distance to the membrane edges to avoid effects from periodic images. Models were embedded into the lipid bilayer by discarding all lipid molecules that clashed with parts of the ion channel model. After creating a TIP3P [98] water box with the dimensions of 120Å x 120Å x 90Å surrounding the system, a neutralizing ion concentration was applied by replacing random waters with Na^+ and Cl^- ions, resulting in an average number of $\sim$109000 atoms for each of the six different prepared model systems of domain association types.

MD simulations were performed using NAMD Version 2.9 [155] applying the CHARMM36 forcefield parameters including lipid and ion additive parameters like sodium cations and chloride anions. To remove restraints and frictions from the MD system, a stepwise protocol was used, oriented on NAMD's membrane proteins tutorial [3].

First the membrane lipid tails were relaxed in 1000 minimization steps, using Newton-Raphson [217] conjugate gradient algorithm and a 0.5 ns dynamics run, while keeping all other atoms fixed. The same approach was then applied to the lipid heads, after removing their constraints, to reduce gaps between channel TMs and lipid molecules. Constraints on all other atoms of the system were subsequently reduced to further minimize and equilibrate the protein's environment. At first by

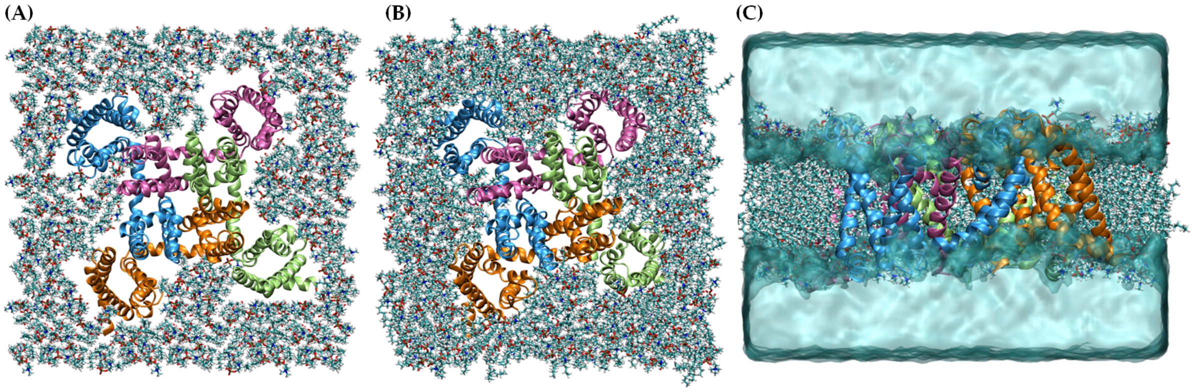


Figure 4.3: Na$_v$1.8 embedded in a POPC membrane. (A) Top view of the Na$_v$1.8 structure (coloured ribbons) embedded into a not equilibrated membrane patch. The gaps between protein and lipids were caused by removing symmetrically organized lipids that were overlapping with protein atoms, during protein insertion. To compensate any shrinking of the simulation box due to gap filling of the lipids [184], the periodic cell dimensions were initially chosen slightly smaller than the membrane size. (B) Energy minimized system, showing a tight packing between Na$_v$1.8 and the membrane lipids. (C) Side view of Na$_v$1.8 with lipid membrane and water box.

releasing water molecules to freely distribute in the simulation box, except for entering spaces of the protein lipid interface and finally by dropping all constraints to relax the whole system as preparation for the production run.

The subsequent productive simulations were performed for 20 ns without any constraints, using a 2 fs time step and periodic boundary conditions (PBC). The Particle Mesh Ewald (PME) [45] method and the Shake [168] algorithm were used for all atoms. The cutoff for van der Waals and long-range electrostatic interactions was set to 12 Å. A constant temperature of 300 K was kept using a Langevin thermostat [213] and the constant pressure at 1 atm was controlled by the Langevin piston [67] Nose-Hoover [132] method. Two independent MD runs with random initial seeds were executed for each of the six domain association types, writing out snapshots every 4 ps. Thus, the MD simulations created twelve MD trajectories, shown in Figure 4.4, to be taken into the next analysis step. However only frames from the second part of the MD runs went into the clustering for further analysis, as those showed RMSD profiles within a range of <1.5 Å, while containing some variability in side chain orientation and contacts. This variability was utilized to generate ensembles of protein conformations for all domain association type models by clustering snapshots from the MD trajectories. A similar approach was described by Tarcsay et al. [192] to generate discrete protein conformations for GPCRs.

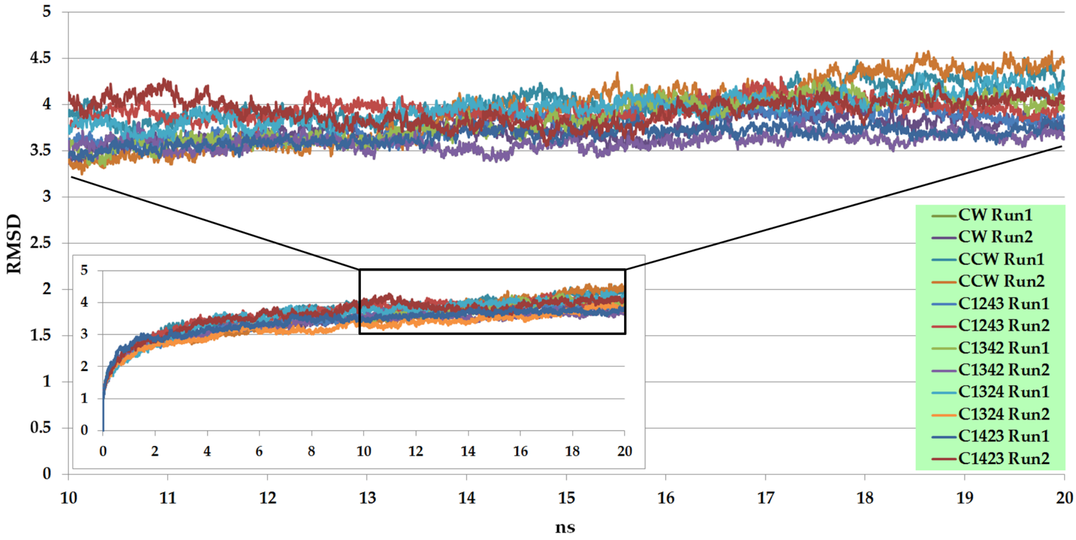


Figure 4.4: RMSDs of all Na$_v$1.8 MD trajectories. The starting conformation of the channel models have been subjected to multiple MD runs with different initial seeds (small graph). Only the second parts of the MD trajectories, showing RMSD profiles within a range of <1.5 Å have been used for further analysis.

4.2.3 Clustering from Molecular Dynamics Trajectories

Prior to the clustering of snapshots from the MD trajectories, the protein was isolated from the rest of the simulation system, since especially the phospholipid membrane was primarily set up to mimic a physiological surrounding for the ion channel and to facilitate realistic interface interactions between the domains DI to DIV in a hydrophobic environment. Each MD trajectory contained 2500 frames to be analysed. A common procedure to reduce the number of individual objects is the cluster analysis [151, 177], which groups similar objects into the same cluster. A prerequisite for this is the comparison of each object with all others, generating a similarity matrix, in this case measured by RMSD. The distance matrix shown in Figure 4.5 was calculated using the VMD extension iTrajComp.

Based on the distance matrix an agglomerative hierarchical clustering algorithm [17] was used to group the MD snapshots by similarity. This iterative algorithm combines single objects or clusters based on their average linkage, which determines the mean similarity between two objects, or all objects of two compared clusters. On the lowest cluster aggregation level, these objects represent single frames, while later objects represent clusters containing multiple similar frames. For each MD run the frames were aggregated until 50 clusters where created, which contained at least 20

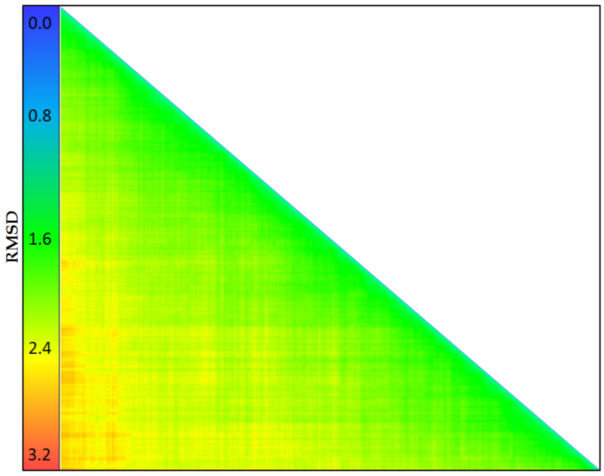


Figure 4.5: Matrix of RMSD measures. Calculation of a distance matrix based on RMSD values for the last 2500 frames of each MD run.

frames, as shown in Figure 4.6 (A). The related dendrogram is shown in in Figure 4.6 (B) and illustrates the cluster hierarchy. Then, the average frame of each cluster was selected, based on the RMSD of all frames in the respective cluster, ending up with 50 structure models from each MD simulation.

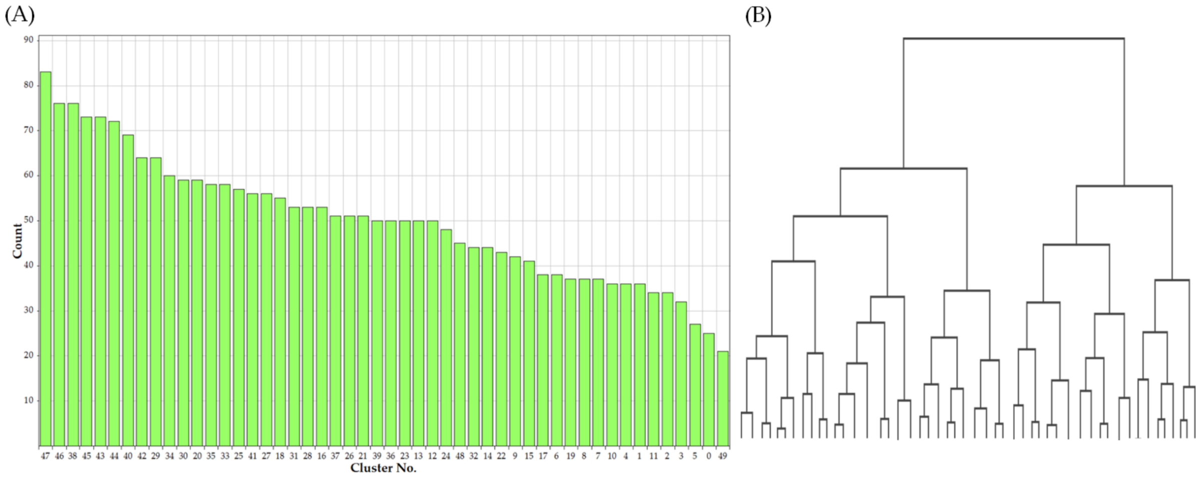


Figure 4.6: Cluster distribution and hierarchical clustering dendrogram. (A) Example distribution of MD frames in 50 clusters containing between 25 and 79 individuals. (B) The clustering dendrogram was truncated bottom-up until a set of 50 clusters were generated.

4.2.4 Rosetta InterfaceAnalyzer

The ion channel models delivered by previous cluster analysis were dissected using the InterfaceAnalyzer module of Rosetta v3.3 [117]. The InterfaceAnalyzer, already described for another application in Section 3.3.1.3, calculates the Rosetta energy of a protein model in different states of domain association. First the Rosetta energy of the complete protein model is determined. Then the protein domains are separated in two units upon prior definition, followed by a second energy calculation. The change in Rosetta energy from complexed to separated states indicates the ΔG value that is used to energetically characterize and compare the models of the six different domain association types. In Figure 4.7 an interface of one domain facing towards the other domains is illustrated as well as a schema of the complexed and separated channel protein, respectively. To permit suitable conditions for the interface analysis within Rosetta, all model structures were adjusted with a Rosetta relaxation protocol to take into account the peculiarities of the Rosetta forcefield used for energy calculation.

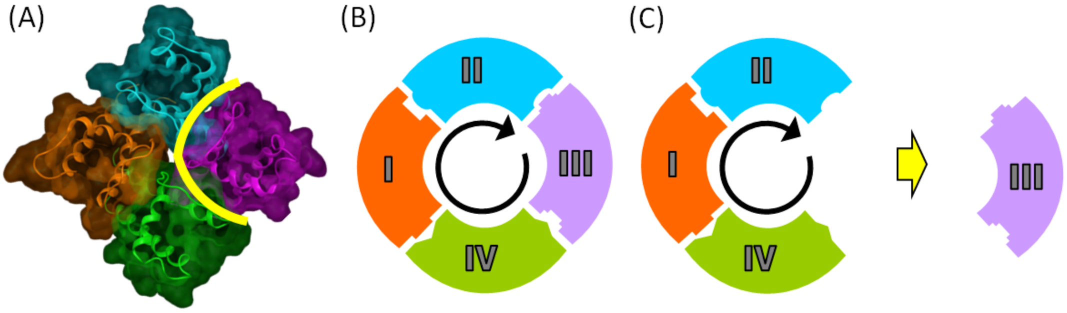


Figure 4.7: Interface analysis and ΔG estimation. (A) Image of a channel pore built by four domains displayed in different colours. The yellow line indicates the interface of one domain to the other domains. (B) Schematic figure of an ion channel in CW domain association. This configuration is used to calculate the energy of the assembled channel. (C) Illustrates the channel configuration with one separated channel domain, used to calculate separated energy portion of the single domain and the remaining channel domains.

The resulting interface interaction energies (ΔG) were determined for all four domain interfaces of each model. This was done for all six ion channel association types. Finally, each individual ion channel model could be described as a vector of four ΔG values.

$$\vec{X}_{model} = [\ \Delta G_{D(I|II,III,IV)}\ \Delta G_{D(II|I,III,IV)}\ \Delta G_{D(III|I,II,IV)}\ \Delta G_{D(IV|I,II,III)}\ ]$$

These vectors were aggregated column wise using matrix algebra by first creating a matrix containing all vectors, and then calculating the arithmetic mean of elements for each of the four columns resulting in an average vector like:

$$m' = [\, \bar{X}_1 \; \bar{X}_2 \; \bar{X}_3 \; \bar{X}_4 \,]$$

In Figure 4.13 the average vectors of all six association types are plotted as well as the top level aggregation with averaged ΔG values from each domain interface of a dedicated association type.

4.2.5 Pareto Ranking

Pareto ranking is a sophisticated analysis method dealing with multi-objective data sets. It allows the comparison of individuals described by multiple attributes to find one or more "best" solutions. In this study Pareto ranking was used to compare all individual model describing vectors of ΔG values to each other, in order to find the best domain association solutions in terms of Rosetta energy measures. Vectors were compared with each other to identify those that were *Pareto optimal*, which means that they are not dominated by any other vector. For instance, the vector $\vec{a} = (\,2\;4\;3\,)$ dominates $\vec{b} = (\,1\;3\;2\,)$, because all attributes in $\vec{a}$ are better (higher) than in $\vec{b}$. Whereas $\vec{a}$ and vector $\vec{c} = (\,2\;3\;4\,)$ are not dominated by one another, because $a_2 > c_2$ and $a_3 < c_3$ while $a_1 = c_1$, and so both can be stated as *Pareto optimal*. *Pareto optimal* vectors can be placed on the *Pareto frontier*, illustrated as red line in a two dimensional example in Figure 4.8, where a and b are *Pareto optimal*, and c is dominated by b.

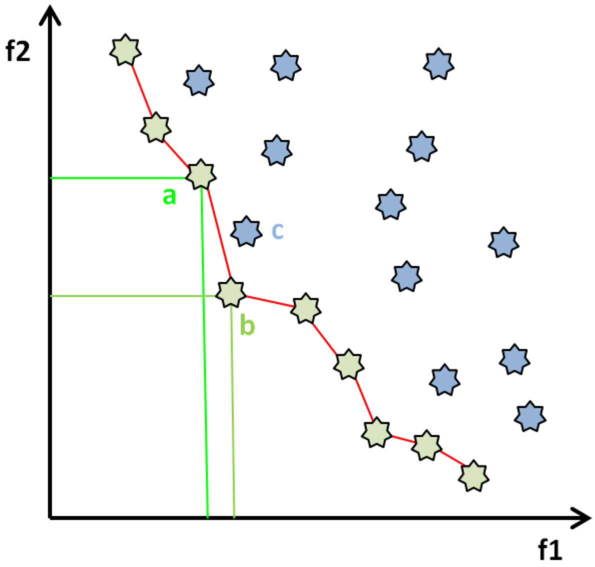


Figure 4.8: Pareto Ranking. Solution b is *Pareto optimal*, because it is not dominated by any other solution, whereas solution c is dominated by solution b, whose functions are lower (better) than the ones of c.

Applied to the domain association models' vectors, the Pareto ranking method revealed a set of best domain association solutions that represents the first *Pareto frontier*. Upon removing this set of non-dominated vectors from the collection of all vectors, the next *Pareto* *frontiers* were created subsequently. The first ten *Pareto ranks* are plotted as histogram on Figure 4.14, showing the distribution of domain association models by *Pareto ranks*.

4.2.6 Direct Coupling Analysis

To obtain further evidence for the most likely association model the direct coupling analysis method [46] was applied. This method makes use of the evolutionary information contained in the $\text{Na}_v1.8$ protein sequence only, neglecting energetic measures or structural conformations. It calculates evolutionary coupling (EC) scores for every pair of residues in the protein sequence. High EC scores indicate a high probability for the residue pairs to be coupled and in close contact. When mapping the EC scores of all residue pairs to their related position in the different association models (e.g. CW and CCW), it is assumed that the correct model shows more high scored residue pairs located close together. For this evolutionary sequence analysis of $\text{Na}_v1.8$ three individual web services were used, providing different approaches and methods of direct coupling analysis (DCA). Details about the DCA or evolutionary coupling analysis methods are described in Section 2.5.

All of the three web servers limit the input sequence of the target protein to a length of 600 AA, whereas $\text{Na}_v1.8$ has 1956 AA and thus had to be split to match the web servers requirements. The split created four sequences, composed of the domain pairs A-B, B-C, C-D and D-A. Additionally, all sequence sections that were not present in the structure models and therefore not part of any domain interface, were also truncated to satisfy the servers' sequence length limitations.

The EVfold web server [`http://evfold.org`][130] provides the EVcouplings service, which makes use of a multiple sequence alignment (MSA) to calculate evolutionary coupled residues with the plmDCA method. By default the EVfold web server expects the Uniprot sequence ID of the target protein or its sequence in fasta format as input, which is used to find a suitable MSA in the Pfam domain database [68], or to generate a MSA using Hidden Markov Model (HMM) based homology sequence searching algorithms, such as HHblits [164] or jackhammer [95]. In case of the $\text{Na}_v1.8$ sequences, the EVfold web server was not able to find or generate a MSA via the default option. For this reason, a MSA was generated with ClustalOmega [179] based on InterPro's [`https://www.ebi.ac.uk/interpro`] protein family IPR001696 of VGSCs, counting 1744 sequences. After splitting the aligned sequences to match the 600 AA limit described earlier, the MSAs were separately uploaded into the advanced settings area of the web server as fixed alignment input with the target sequence in the first row of the alignment and additionally provided in the GUI of the web service.

A similar EC analysis algorithm as EVfold is used by the GREMLIN web server [`http://gremlin.bakerlab.org`][150], which also needs a MSA of the protein target family. Hence, the previously prepared MSA was also used with GREMLIN. This time the MSA serves as initial information on which the algorithm tries to add missing sequences, and to optimize the MSA composition by removing sequences identified as very similar (>90% identical) or duplicates. This optimization was done with HHblits [164], that was run with an E-value of 1E-06 and 8 iterations of sequence searching. The number of homologous sequences was finally around 5 times the sequence length and so barely at the minimum count for accurate predictions.

As third web service the RaptorX [205] Contact Prediction web server [`http://raptorx.uchicago.edu/ContactMap`] was used, which applies a joint EC analysis on the target sequences. It combines the protein target family sequences with additional related protein families, which may have less similar sequences but comparable structural folds in order to enrich the sequence collection with otherwise unconsidered information. In contrast to EVfold and GREMLIN, the RaptorX web server only needs the target protein sequence as input and collects all related sequences automatically. If a sufficient number of sequences is available, the DCA is started.

For all three web servers EC-scores were generated based on a MSA of $\mathrm{Na}_v1.8$. These EC-scores were further matched to the corresponding residue pairs of CW and CCW structure models. The only restriction was that EC-scores were only assigned to residue pairs containing residues, which were located on different channel domains and were potentially involved in interface interactions. Then all distances d_{ij} (in Å) between the Cα atoms of all possible residue pairs ij, located on different channel domains, were measured in the CW and CCW model. As it was supposed that residue pairs showing a long distance in one model type and a short distance in the other type, are more predictive in discriminating between the CW and CCW domain association model, a weighting factor incorporating this distance difference was considered.

Thus, the absolute differences between these pairwise distances from CW and CCW $|d_{ij}|_{cw,ccw}$ was calculated and used as weighting factor for the EC-scores to calculate a weighted EC-score E_g according to formula 4.1, where E_{ij} is the EC-score of a residue pair ij:

$$E_g = |d_{ij}^{cw} - d_{ij}^{ccw}| * E_{ij} \tag{4.1}$$

The residue pairs were further clustered by the measured pairwise distance between their Cα atoms, resulting in ten clusters for each domain association type.

Residue pairs showing short distances and high EC-scores most likely relate to a correct position, indicating a favourable domain association type model. Assigning a high weighting factor to residue pairs that are located close together on one domain association type model and far apart in the other model, helps to distinguish the CW and CCW model, as the high weighted EC-score is sorted into the first, short distances cluster for one of the models and into a high distance cluster for the other model. Residues located, for example, in the pore region show relatively short distances between each other in both association type models. Their weighting factor is quite low, according to their distance differences in the CW and CCW models. An example is visualized in Figure 4.9 showing two residue pairs and their relative position in the CW and CCW model.

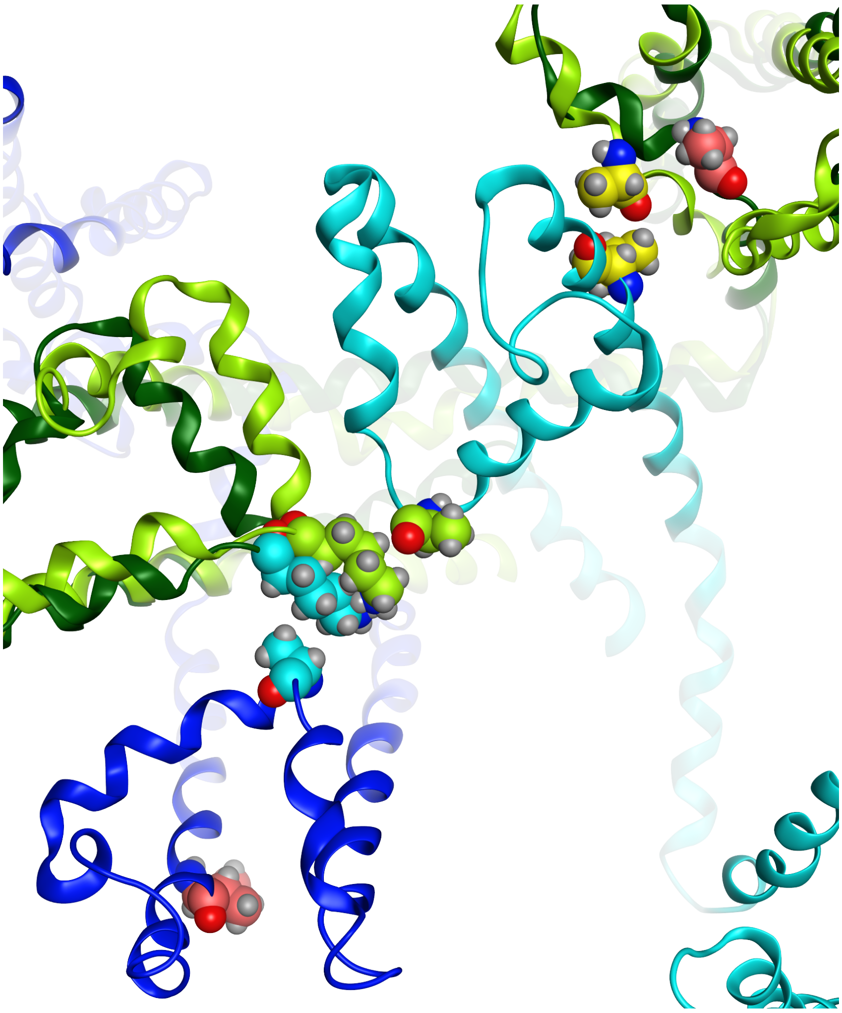


Figure 4.9: Position of residue pairs in CW and CCW models. Green ribbons: Chain A of CW is light green and chain A of CCW is dark green. Chains are superposed. Blue ribbons: Chain B of CW is light blue and chain B of CCW is dark blue. Residue pair Lys-Ala is coloured green in CW and blue in CCW. Their position in the CW and CCW model is very similar and not helpful to discriminate the association types. On the other hand, residue pair Ala-Ile, coloured yellow in CW and red in CCW, is far apart in the CCW model, but the residues are in close contact in the CW model, indicating the correct association type.

4.2.7 EC-score to ΔG Mapping

The relation between the EC-scores and the binding energy contribution of the single residues to the domain interface was evaluated next. Rosetta's InterfaceAnalyzer does not allow to break down the ΔG value of an analysed interface into the energy contribution of single residues, without extra effort and the involvement of additional software. Single residue energy contributions for the domain interfaces of the CW and CCW domain association types were therefore estimated with MOE. First, the CW and CCW models were adapted to MOE's AmberETH forcefield by energy minimization of the side chains, until the structure change from one minimization step to the next resulted in an RMSD of ≤ 0.1. Then, all residues located at the domain interfaces and involved in interchain connections were identified. An in-silico alanine scan was performed for these residues to estimate their energy contribution to the interface interactions. The resulting energy differences (ΔG) were linked to the single residue EC-scores of RaptorX, as the RaptorX dataset enclosed all channel residue pairs. The analysis focused on residues with EC-scores ≥ 0.3 and ΔG values ≥ 1. This mapping protocol was also tested using the Ca$_v$1.1 structure for validation (Figure 4.29).

4.2.8 Docking

In order to analyse and compare the ion channel domain association types further, a docking approach with Na$_v$1.8 ligands was performed. The generated docking scores and binding modes of the ligands were assumed to reflect binding site differences, caused by the domain composition in the different domain association type models. For this evaluation, the coordinates of multiple different model conformations (frames) derived from the trajectories of each CW and CCW MD run were assembled as described in 4.2.3, and used to estimate and compare their binding affinity to selected ligands. Therefor, the ligands were docked into the pore region of the channel where all four different domains are close together, and binding sites spread across two domains were most likely. All key residues listed in Table 4.1 are located in this region. Other known binding sites investigated and described by binding toxins like Tetrodotoxin (TTX) [138] have been neglected for this analysis, as they are typically occupied by lager peptides, not covered by the selected sets of compounds. Two sets of ligands, one containing primarily high effective but sparsely characterized compounds, and the other comprised of less effective ligands with mainly known binding locations were prepared for docking.

The first ligand set was collected using the ChEMBL Database [`https://www.ebi.ac.uk/chembl`][14] as well as Thomson Reuters' Integrity [`http://thomsonreutersintegrity.com/`]. These sources cover published literature data and structure scaffolds from patents, and include compounds from all phases of development. The number of ligands was reduced with various filters to compile a set of ligands with suitable properties for docking as described next.

Filtering for high Na$_v$1.8 subtype selectivity was not successful, as most of the retrieved ligands were also claimed to be active on other channels. High subtype selective compounds would have been beneficial for predicting the correct domain association, as they tend to bind only to specific binding sites, which provide a fitting contour with an unique composition of residues [204, 136]. This requires the correct order of residues and side chains of the binding site to enable best interactions with the ligand and might only be present in the right domain association model. The next applied filter was a 10 μM cutoff for the ligands' IC_{50} values. This is thought to be beneficial for the use in a docking approach, as most scoring functions applied in docking algorithms are trained on existing strong binding ligands crystallized in a complex with a receptor [135]. Since the binding affinity in hydrophobic sites already increases solely with the ligand's size [63], likewise observed in ion channel pore cavities, affinity variations caused by specific interactions might be masked in case of large ligands. Consequently, the ligands' molecular weight was kept below 700 g/mol. The conformational space was limited by tolerating a maximum of ten rotatable bonds, and ChEMBL structures needed an assigned confidence score of greater than 9.

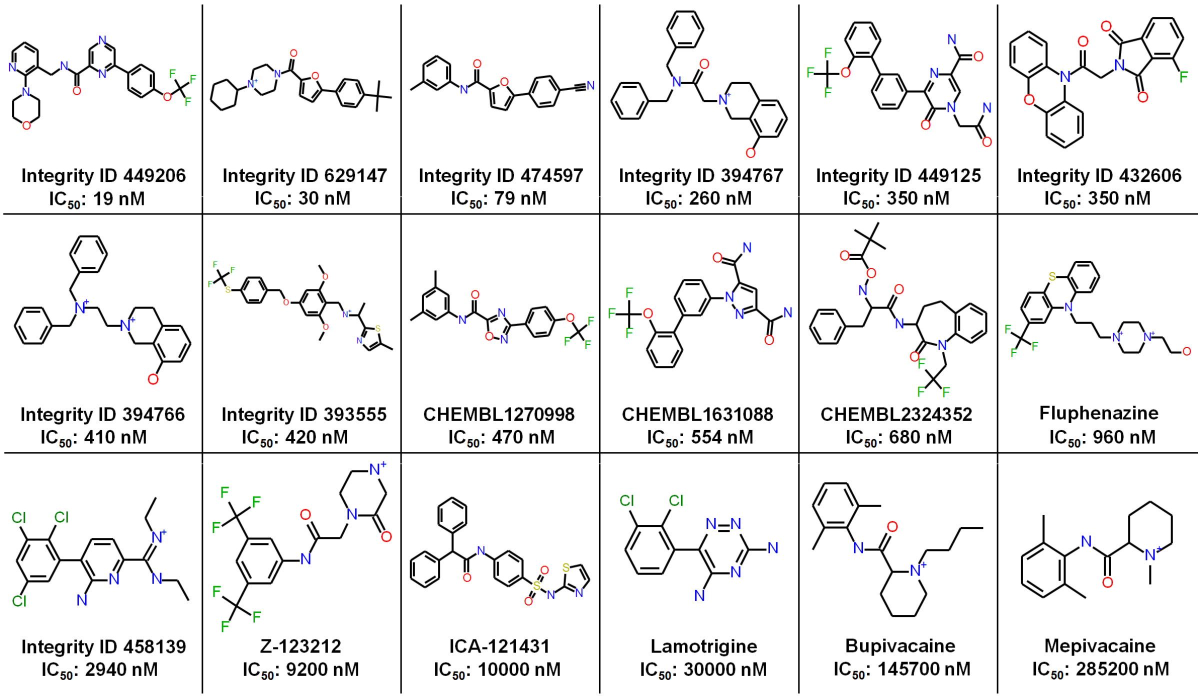


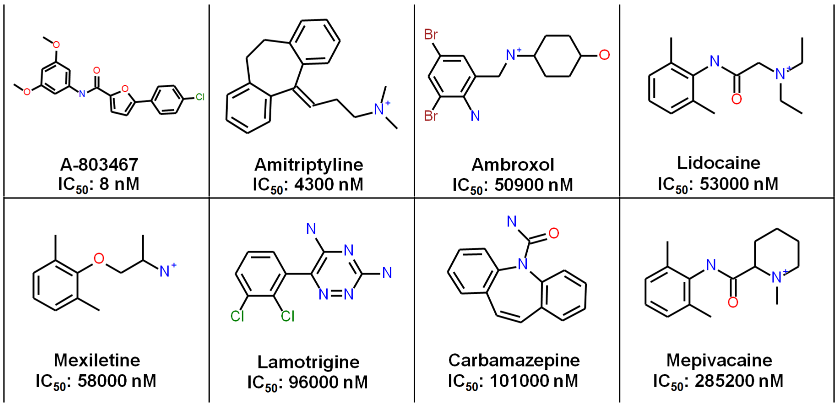
Figure 4.10: Diverse set of Na$_v$1.8 ligands. Ligand set collected from ChEMBL and Thomson Reuters' Integrity by applying different selection criteria.

Finally, the ligand collection was enriched with few well characterized reference compounds that were missing in the data sources or filtered out in the selection process, although they might provide useful docking poses.

To further reduce the number of ligands without losing chemical diversity, MACCS fingerprints were generated for each ligand followed by a clustering. The fingerprint representation of the chemical structures allowed a fast comparison and clustering of the ligands [209], resulting in a structurally diverse collection of 18 compounds depicted in Figure 4.10, that have been used for the first docking approach. Although the ligands were associated to Na$_v$1.8 in the data sources, no information about their actual binding sites were available.

The second set of ligands (Figure 4.11) was gathered by literature searches for local anaesthetic compounds, which presumably bind in the ion channel pore where the local anaesthetic binding sites are located (Table 4.1). This was beneficial for docking compared to the first ligand set and could compensate the reduced binding affinity of the local anaesthetics.

All ligands' 3D structures have been generated using Corina [169], and the frequently contained basic nitrogens were protonated. To prepare the ligands and receptors for Autodock Vina, the python scripts *prepare_ligand4.py* and *prepare_receptor4.py* are provided with AutoDockTools as part of the MGLTools package (1.5.6) [142]. They were used to generate PDBQT files, which are required by Autodock Vina version 1.1.2 and contain additional charge (q) and atom type (t) columns compared to regular PDB files.



A-803467
IC$_{50}$: 8 nM

Amitriptyline
IC$_{50}$: 4300 nM

Ambroxol
IC$_{50}$: 50900 nM

Lidocaine
IC$_{50}$: 53000 nM

Mexiletine
IC$_{50}$: 58000 nM

Lamotrigine
IC$_{50}$: 96000 nM

Carbamazepine
IC$_{50}$: 101000 nM

Mepivacaine
IC$_{50}$: 285200 nM

Figure 4.11: Collection of local anesthetic ligands. Ligand set merged by literature searches for local anesthetic compounds.

The docking procedure additionally requires the definition of a grid box that de-limits the search space onto the crucial protein areas of interest in order to reduce computational time. A grid box of 36 x 36 x 50 Å^3 has been designed with MGLTools, fitting the central pore region of the ion channel models as well as the S5 and S6 helices and adjacent parts. The configuration of the actual docking runs allowed a maximum number of 20 binding modes and respective scores as results for each ligand.

Those score sets were then correlated with the ligands' measured IC_{50} data [36, 76]. In order to compare the docking scores from AutoDock Vina to published experimental IC_{50} data in terms of binding energies in *kcal/mol*, the latter had to be transformed using the formula previously described and discussed in the docking test case 2.4.1.

According to the analysis approach also described in the previous test case 2.4.1, the resulting docking poses and scores of both ligand sets were analysed, only considering the scores that are related to docking poses close to the expected binding site or showing interactions between the docked ligand and binding site residues. The Pearson correlation between the docking scores and the experimentally measured affinity values from the literature was calculated for each structure model (frame). The distribution of these correlations is visualized as box plot for the CW and CCW domain association types, respectively.

4.3 Results and Discussion

4.3.1 Models

Structure model sets for all six types of domain assemblies were derived from MD simulations with relatively short runtimes. In contrast to achievable microsecond scale simulation times, described in MD focused publications [82], the intention for including an MD simulation step in this analysis was basically the relaxation and, to a certain extent, a conformational sampling of residue side-chains, as they are predominantly involved in domain contacts and ligand interactions.

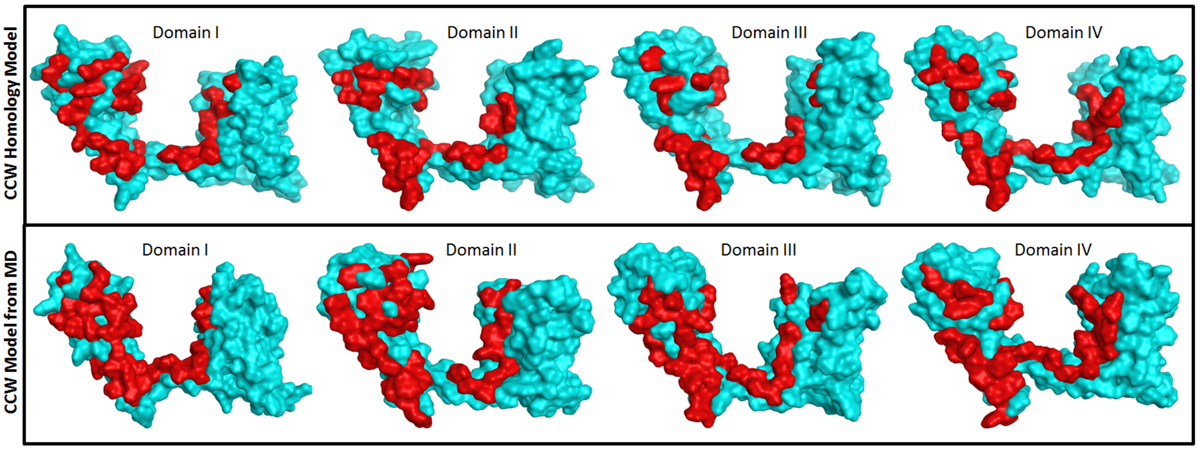


Figure 4.12: Comparison of domain interface contacts. Interface surfaces coloured by residues interacting with neighbouring domains. Red patches indicate residues involved in bonds with other chains, whereas blue areas do not participate in domain interactions. The models coming from MD simulation show larger areas of domain interactions.

Model	Hydrophobic Bond	Hydrogen Bond	Ionic Bond
CCW Homology Model	111	10	5
CCW MD Model	138	24	5

Table 4.2: Counts of different bond types. The counts quantify the increase of contacts during MD simulation from the CCW homology model (first line) to the final MD model (second line). Similar increases of the bond counts were observed for the other domain association types.

Especially the phospholipid membrane was primarily set up to mimic physiological surroundings for the ion channel and facilitate realistic interface interactions between the protein domains (DI-DIV) in an hydrophobic environment. In order to maintain the conformational structure of the template and maintain the general fold of the transmembrane segments, large backbone movements of the model were not intended during the simulation. The models derived from the MD trajectory showed RMSDs of CA backbone atoms (Figure 4.4) in the range of <1.5Å, as well as noticeable side chain sampling observed by superposing various models from different time points and visual inspection. Domain interactions improved in the course of MD simulation, measured by the number of domain interface contacts as shown in Figure 4.12. Since the template structure of Na_vAb [154] is in a quasi-closed state, no open state conformation models of Na_v1.8 have been generated because this would have required an additional modeling abstraction level. In fact, the closed state conformation represents a good basis for this investigation, because the resting position of a VGSC can be noted as energetically favourable and most stable, as it needs external forces like the membrane potential to change its conformational state [152].

4.3.2 Interfaces

The interface interaction energies have been evaluated with Rosetta InterfaceAnalyzer on the basis of 100 models from two independent MD runs for each of the six domain association types, resulting in 600 models to be analysed. For every interface between the four channel domains free enthalpy values were computationally determined using the InterfaceAnalyzer.

The interaction energies for all domain association types are shown in Figure 4.13. The results of the domain-wise aggregated ΔG values in Figure 4.13 (A) are quite diverse and the standard deviations range from 6 to 13 kcal/mol. Figure 4.13 (B) shows an aggregation of the domain related values to estimate the total mean ΔG for every domain association type.

The aggregated mean ΔG value of the CCW domain association type shows the highest absolute value among all other types, whereas the C1342 conformation was calculated with the lowest value. Compared to the other association types, the standard deviation for the CCW association type is narrow, indicating a balanced strength of interface contacts across the four protein domains. The aggregated mean ΔG values of the other four association types are on an equal level, only showing diverse energies for the individual domains.

The unequal interface interaction energy distribution among the domains of such a channel protein might be responsible for inducing changes of the protein structure conformation by external forces, destabilizing the weakest link of the domain assembly. Accordingly, the CCW type might need to pass a higher energy threshold to

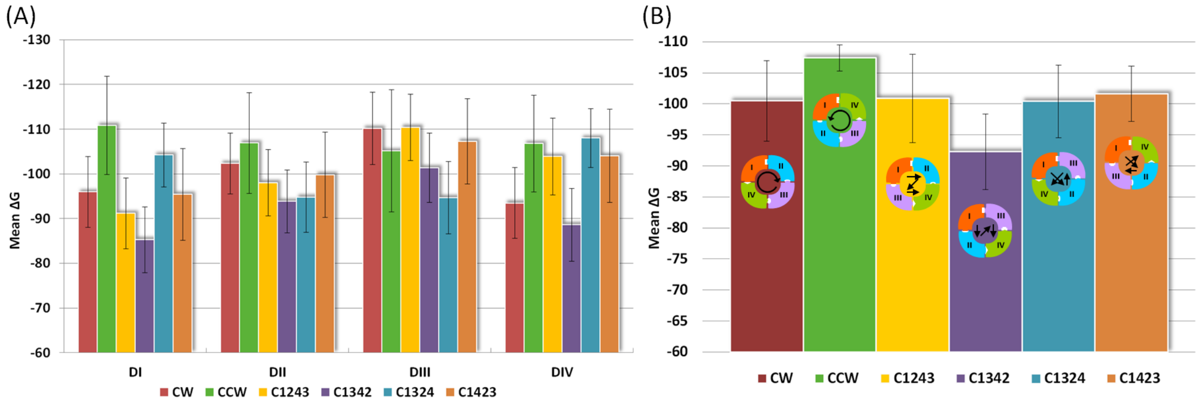


Figure 4.13: Aggregated mean ΔG results. (A) Domain-wise aggregated ΔG values, based on 100 model for each color coded domain association type. (B) Aggregated mean ΔG values for each association type. The error bars in both graphs indicate the standard error of the mean.

change its structure conformation, which would not be beneficial for a fast responding ion channel. On the other hand C1342, being probably the most unstable association type, might therefore be unable to maintain certain functional states.

The pareto analysis of the domain interaction energies provided results that were consistent to the previous observations. As shown in Figure 4.14 most of the models, that were sorted into the first pareto rank belong to the CCW association type. Only one model from C1324 and C1423, respectively, is present in the first rank as well. The CCW association type was also dominant in ranks two, three and four, where a considerably higher number of models of this type could be found. The first single CW model was found in rank two, and the first C1342 models occur not before rank six.

A comparison by structure superpositioning of the C1342 and the CCW type model revealed a more open state of C1342 than the CCW type model. The aromatic residues at the end of the four S6 helices of the C1342 model are separated from each other by diagonal diameters of ∼15 Å and ∼18 Å measured from the backbone CA atoms. This loosely packed open like conformation could be related to the weak domain interaction energies of C1342. Whereas the channel gate of the CCW type model shows diagonal diameters of ∼10 Å and ∼15 Å between CA atoms, which is likely to be more related to a closed state conformation, in particular when compared to the open channel structure of Na_vM measuring ∼19 Å and ∼23 Å in diameter at the pore exit.

The data presented in the Figures 4.13 and 4.14 suggest that the CCW domain association type could be the most favourable conformation of the Na_v1.8 channel, al-

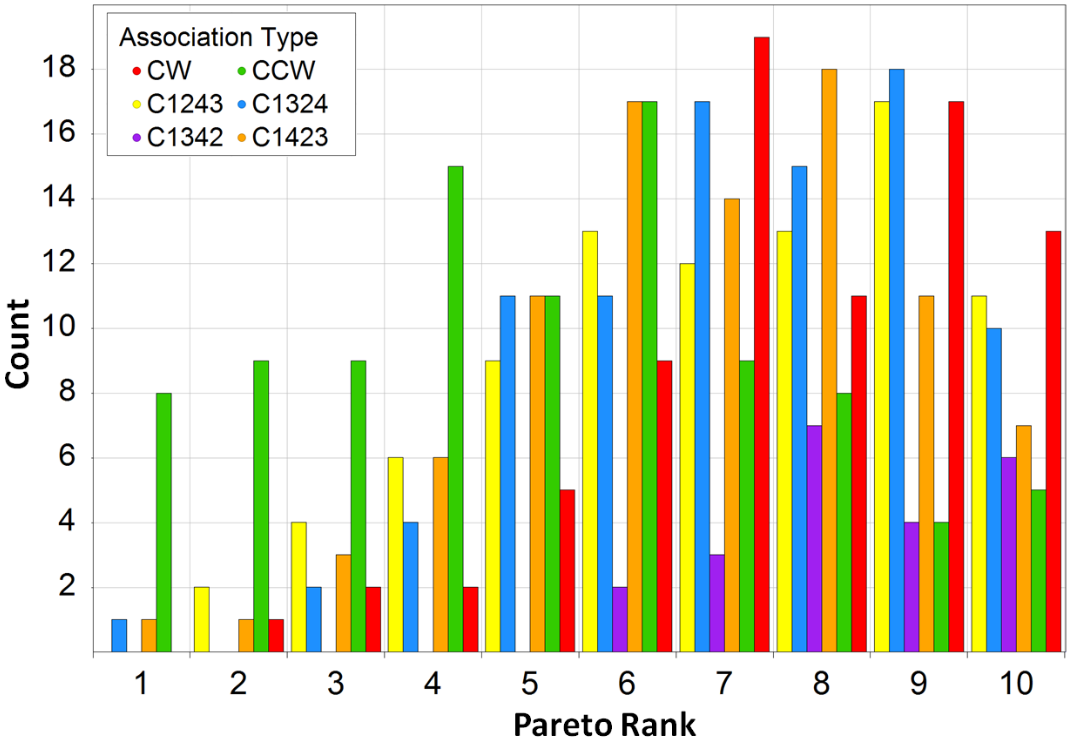


Figure 4.14: Pareto ranking of domain association types. The domain association type models are sorted into Pareto ranks by the Pareto multi-objective ranking method, according to their four different domain interaction energies.

though the CW type was proposed in previous studies [58, 121]. The hypothetical cross-wise domain association types, built as a proof of concept, showed similar energetic properties as those of the CW and CCW type. However, the cross-wise models were dropped from further in depth analysis, because the CW and CCW models seemed to be more likely and were the main content of debate in the literature. Subsequently the CW and CCW association types were analysed in more detail by inspecting the individual frames of each pareto rank. Figure 4.15 shows the frames over time, along with the pareto rank they were assigned to.

In Figure 4.15 (A) it is shown that the number of CW models assigned to better pareto ranks increase over time. This slight trend is also observed for the CCW models, but in the opposite direction, meaning more models are assigned to adverse pareto ranks in the course of simulation time. The histograms in Figure 4.15 (B) display the sum of models per pareto rank, separated into the first (upper histogram) and second (lower histogram) half of the simulation. The graphs show that the first set of models, the CCW type dominates the first pareto ranks, whereas in the second set the

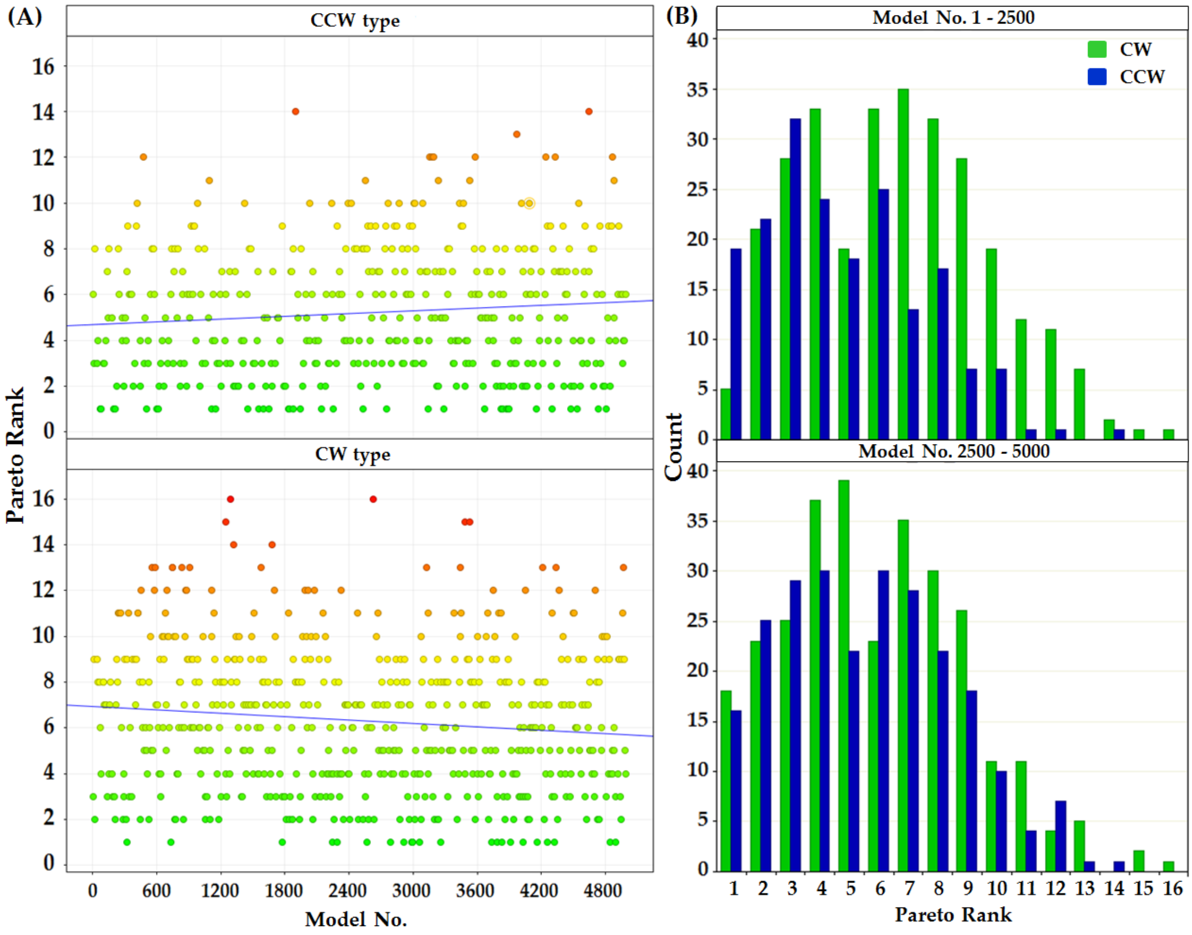


Figure 4.15: Time-resolved pareto ranked Na$_v$1.8 models. (A) The domain association type models of CW and CCW are plotted by simulation time. The CW type shows a decreasing trend towards better pareto ranks, whereas the CCW type models, in the same time, are assigned to adverse pareto ranks. (B) The counts of CW and CCW type models are shown for each pareto rank in the first (upper histogram) and second (lower histogram) half of the analysed time period. Among the first set of models, the CCW type dominates the first pareto ranks, whereas in the second set the CW domain association type models are already dominant in the first rank. In addition a general shift of CW towards better ranks can be observed.

CW domain association type models are already dominant in the first rank. In addition a general shift of CW towards better ranks can be observed. This indicates that longer simulation times could strengthen the identified trends, leading to a better discrimination between CW and CCW. In case of the CW domain association type, the rearrangement of residue side chains to form additional inter-domain contacts, that additionally contribute to the domain interfaces, could be further optimized, whereas interactions in the CCW models might be released due to conformational changes or simply become less high rated in comparison to the CW models. Because such long

MD simulation periods, that would have been required to analyse theses effects, were beyond the scope of this work, aside from insufficient available computational capabilities, another approach was applied, that utilizes evolutionary information covered in protein sequences.

4.3.3 Direct Coupling

4.3.3.1 Results from DCA methods

The DCA results, generated with the three web servers Evolfold, GREMLIN and RaptorX, were different in terms of data structure, number of results and result values, describing the EC-scores. The results are presented as follows: i) The Evofold web server generated an EC-score for every residue pair in the sequence. The EC-score range covers positive and negative values, meaning more positive results indicate a higher probability of strong evolutionary coupling. Figure 4.16 shows the distribution of mean EC-scores clustered by the distance between the corresponding residues.

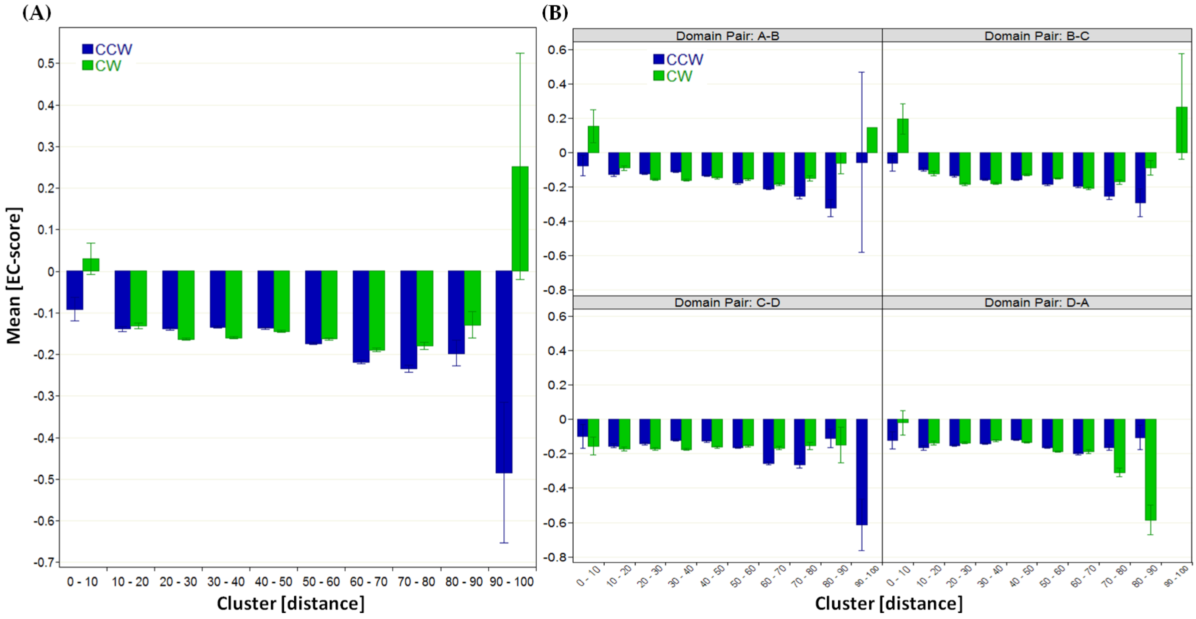


Figure 4.16: Evofold Na$_v$1.8 mean EC-scores. The EC-scores are clustered according to the distance of the related residue pairs. The higher the EC-score value the stronger a residue pair is coupled. (A) shows aggregated data for the complete channel, whereas (B) displays the EC-scores split by domain pair. The error bars in all graphs indicate the standard error of the mean.

ii) GREMLIN provided only a reduced set of results due to internal filters truncating the results according to predefined thresholds. For example, the GREMLIN result list is trimmed to 1.5 times the sequence length, resulting in only 662 EC-scores for a sequence of 441 AA. Consequently, the plotted data represent few individuals per cluster, which leads to large error bars in the plot shown in Figure 4.17.

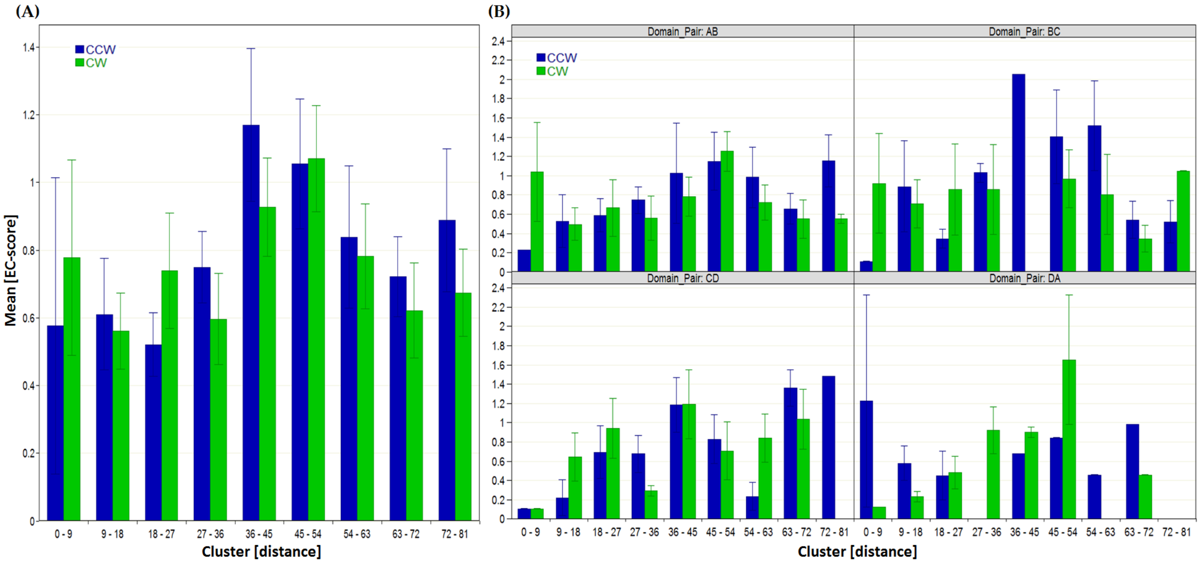


Figure 4.17: GREMLIN Na$_v$1.8 mean EC-scores. The EC-scores are clustered according to the distance of the related residue pairs. The higher the EC-score value the stronger a residue pair is coupled. (A) shows aggregated data for the complete channel, whereas (B) displays the EC-scores split by domain pair. The error bars in all graphs indicate the standard error of the mean.

iii) The RaptorX output was the most comprehensive one, as it extended the set of sequences to be analysed through integration of related protein families' sequences, returning an EC-score for every residue pair of the input sequence as shown in Figure 4.18.

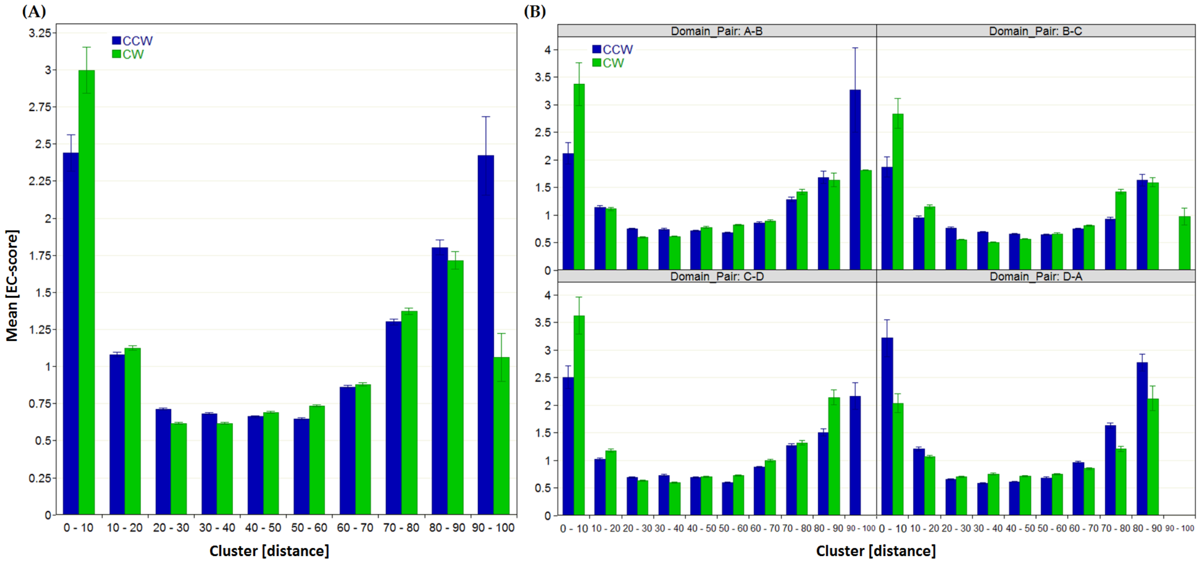


Figure 4.18: RaptorX Na$_v$1.8 mean EC-scores. The EC-scores are clustered according to the distance of the related residue pairs. The higher the EC-score value the stronger a residue pair is coupled. (A) shows aggregated data for the complete channel, whereas (B) displays the EC-scores split by domain pair. The error bars in all graphs indicate the standard error of the mean.

All plots solely contain data from residue pairs, whose members are located on different channel domains. All intra-domain residue contacts were neglected in this respect, as they are predominantly involved in processes such as secondary structure or domain structure folding. Figures 4.16, 4.17 and 4.18 present the same layout, showing one histogram (A) containing aggregated data of the whole channel, split by the CW and CCW domain association types. A trellis by domain pairs shows four histograms (B), further breaking down the mean weighted EC-scores.

In general, the results from all three DCA web services indicate, that the CW domain association type is the favourable model. The mean EC-scores aggregated over all domain pairs, as shown in the histograms (A), are higher for the CW type compared to CCW. A drill down towards the individual data of the four domain pairs shows that the high EC-scores for CW are not dominating among all domain pairs. For domain pairs A-B, B-C and partially C-D, the CW association type shows higher EC-scores for residue pairs in the first distance cluster of closely located residues. The

EC-score distribution for domain pair D-A in Figure 4.17 (B) and 4.18 (B) shows a higher value for CCW in the first cluster than for CW.

These findings can be explained assuming that folding of the channel structure starts with domains A, B and C at the beginning of the sequence. Strong interactions between key residues could guide the process of folding and domain assembly into the clockwise orientation. Once the direction is set there is limited freedom of progressing into another orientation. Accordingly, the channel closure between domains D and A does not require assistance.

Nevertheless the high EC-scores between domain D and A in the CCW association type raises the questions: What are the corresponding residues in CW, how far are they apart, and what might be their function?

The high values in the last cluster, leading to long bars in the histograms, can be explained by the instance, that the weighting factor can become higher, the larger the basic distance of a residue pair is in one of the models. This might result in huge distance differences, assigned as weights to the respective EC-score. Additionally, few individuals are assigned to these last clusters, as shown in Figure 4.19, plotting the count of residue pairs of each cluster.

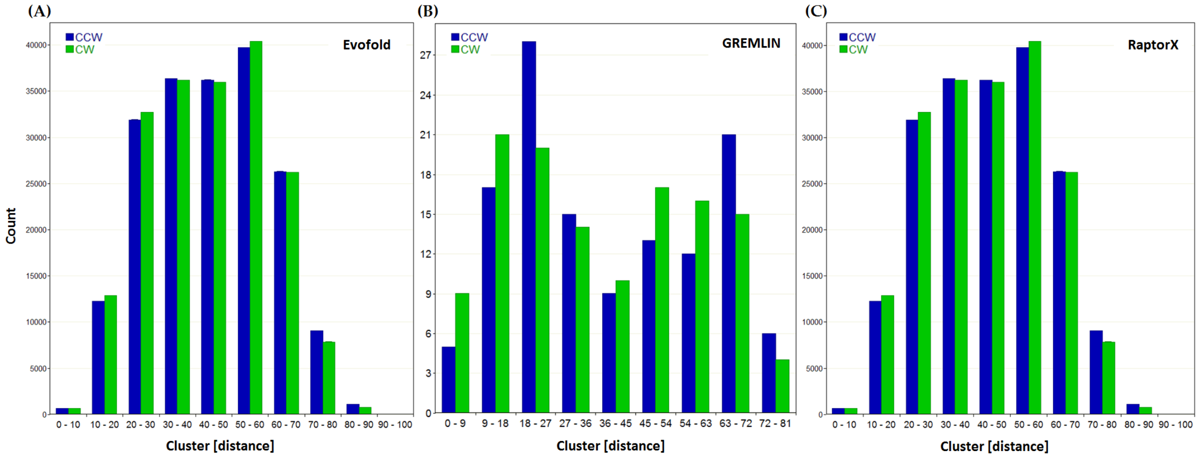


Figure 4.19: Count of residue pairs with an EC-score. Count of EC-scores for each distance cluster, shown for all three DCA methods. Few individuals were assigned to the first clusters, but even less to the last clusters, collecting residue pairs with very long distances. (A) and (C) are very similar, as both web services delivered nearly the same number of results. (B) displays far less EC-scores in the results set, which was pre-filtered by the web service.

4.3.3.2 Single residue EC-scores vs. ΔG results

A link between the single residue EC-scores and the ΔG results from Rosetta's InterfaceAnalyzer could not be drawn directly, as one ΔG describes the complete interface of a channel domain. Therefore, single energy contributions of the interface lining residues were estimated with MOE and the results were correlated with RaptorX generated EC-scores. The results are plotted in Figure 4.20 (A). All residues with ΔG results greater than 5 kcal/mol were identified as tryptophanes. This fact and the clear offset towards the lower values in the plots indicate a residue specific effect during the alanine scanning and energetic evaluation of these big aromatic side chains in MOE. Residues with high EC-scores and ΔG results $\geq$ 1 kcal/mol were only present in the CW model, marked by a blue frame in the scatter plot. The positions of the respective residues were highlighted in the CW structure model as bold atoms and coloured by domain, as illustrated in Figure 4.20 (B). The locations of these residues are distributed across all domains, which allows the assumption that these residues might have a key function for the assembly or stability of the channel. In the CCW model, these residues are not in contact to each other and were not detected by the alanine scanning, which supports the conclusion of CW being the favourable domain association type.

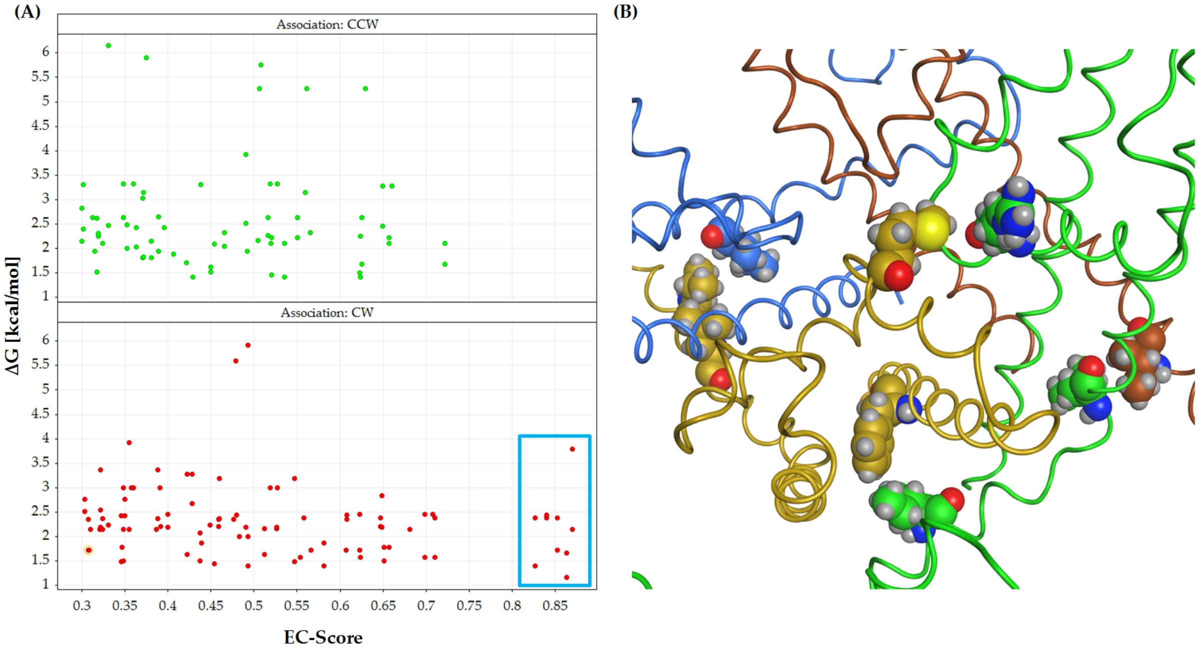


Figure 4.20: Na$_v$1.8 key residues by EC-score and ΔG results. (A) RaptorX EC-scores plotted against ΔG results from MOE alanine scanning. The high ΔG results of exclusively Trp residues seem to be related to the energy evaluation method or the forcefield, MOE applied. The residues in the blue frame, showing high EC-scores and ΔG values over threshold, are highlighted in the CW structure model, shown in (B). The key residues are shown as bold atoms and coloured by domain.

Finally a single measure for each domain association type was determined using the following formula:

$$AT_{measure} = \sum \Delta G_i * E_i \tag{4.2}$$

The $AT_{measure}$ sums up all products of the energy difference ΔG for each residue i and the corresponding EC-Score E. This measure combines structural and energetic results with evolutionary information and should allow a simple illustration of the most likely association type. Based on this $AT_{measure}$, once again CW turns out to be the correct model as shown in Figure 4.21.

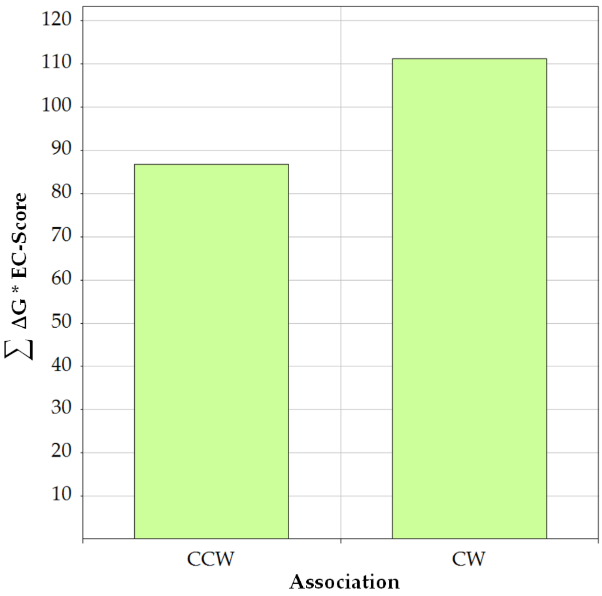


Figure 4.21: Association type measure based on EC-score and ΔG results. Each bar represents the summation of the product between ΔG and EC-score of each analysed residue.

4.3.4 Docking

These experiments were performed to investigate the differences of an incorrect model like it is currently assumed for the CCW association type and the correct CW model, when used for ligand docking.

Two sets of ligands were used for docking, containing compounds selected based on different criteria, as described in 4.2.8. The first set of compounds was selected to include a good diversity of structural scaffolds and moieties, different shapes of the ligands, and substituents that allow interactions with all residue types.

The results of the docking runs, using this diverse ligand set, were correlated with the literature data using Pearson's correlation. The distribution of the R^2 results are shown in Figure 4.22 (A) for both domain association types. No distinction between CW and CCW can be observed from the box plots, with a resulting mean R^2 of below 0.6 for the correlations of both association types.

Compared to the docking test case introduced in 2.4.1, similar correlations were expected. However, a better discrimination of the association types was intended, but the differences in the correlations are limited to little more outliers with low R^2 values in the CCW plot. The results indicate that the environment in the channel pore for the docked ligands is similar in CW and CCW, which let to similar docking scores. Furthermore, the assumption that the ligands' binding sites are located in the pore

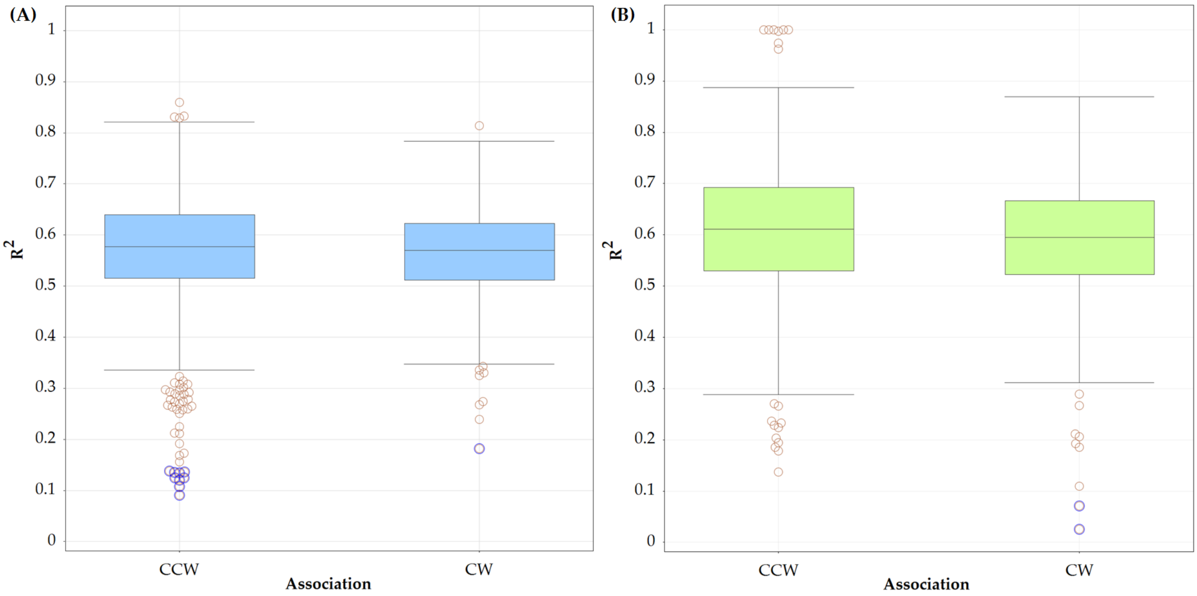


Figure 4.22: Correlation distributions of docking results based on the diverse (A) and local anaesthetic (B) ligand sets. Averaged distribution of the R^2 values for both association types. Red circles show suspected outliers within 1.5 IQR of the whiskers, whereas blue circle outliers are beyond this range.

region most likely affected the results. For instance, ligand ICA-121431 was recently discovered to bind between the helices of the voltage-sensor domain far away from the pore region [2].

As the determination of a favourable association type was not successful using the first set of ligands, the docking was repeated with the set of local anaesthetic compounds. The correlation distributions of the second run are shown in Figure 4.22 (B).

Although, the mean R^{2}s of the correlation distributions were slightly better than of the previous results, no notable difference between the box plots of CW and CCW can be identified.

This result supports the assumption, that the surroundings for the docked ligands in the CW and CCW models are similar to that extend to which the binding affinities calculated by AutoDock Vina are nearly the same. As already described in 2.4.1, the prediction of free energies of binding is estimated to have an error of ±2 kcal/mol [89], when using methods like Vina. This seems not to be accurate enough to discover bind site differences in the Na$_v$1.8 structures.

Although the correlation data are not conclusive, the individual ligands might show differentiating docking scores on the CW and CCW association type. Therefore a simple comparison of the docking scores was performed by calculating the difference between the experimental affinity and the docking scores for each association type model and ligand. The box plots in Figure 4.23 illustrate to what extend the measured affinities were matched by the docking scores.

Interestingly, the most potent ligand A-803467 showed the largest differences between docking scores and measured affinity, while the results for other ligands like Lidocaine are in good agreement. A trend towards larger differences between docking sore and experimental measures can be observed with increasing IC_{50} values. Although the local anaesthetics binding site is most likely located in the channel pore, which was shown by various studies (Table 4.1) investigating the binding modes of pore blocking compounds, no experimental structure of a complex containing a ligand bound to the pore of a heterotetrameric VGSC is available. This information would be useful to tailor further docking approaches towards relevant protein re-

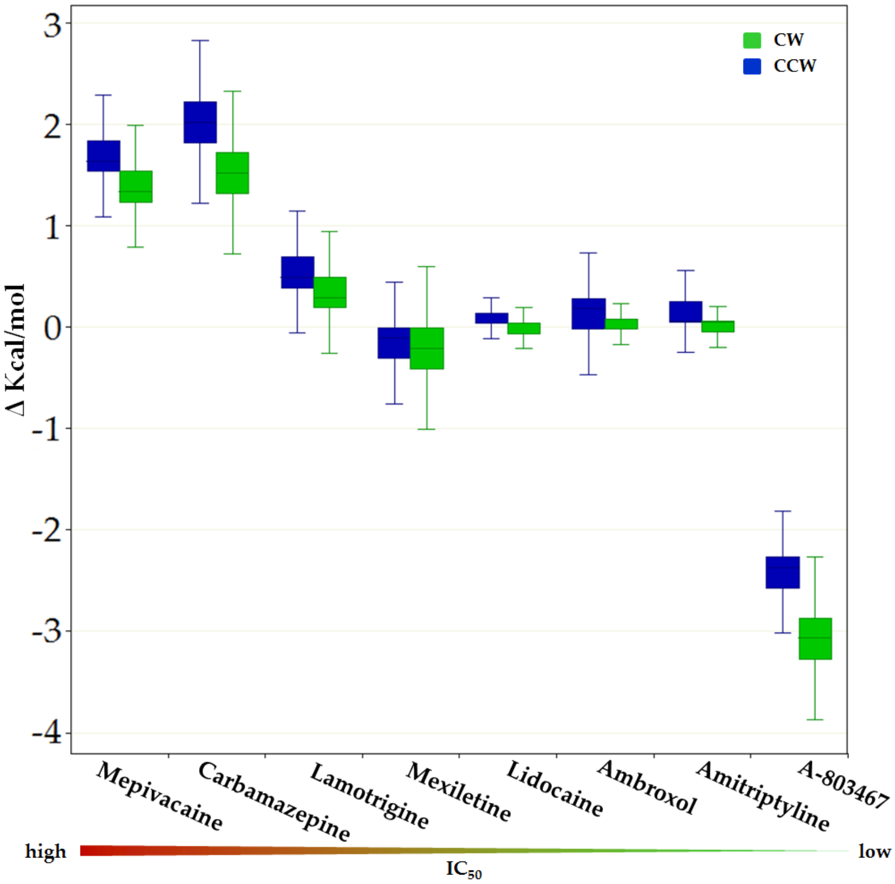


Figure 4.23: Ligand-wise comparison between docking scores and affinity data. The box plots represent the differences between each docking score calculated for a ligand and the ligand's measured value. The ligands are ordered according to their IC_{50} values. The CCW type is coloured blue and CW is coloured green.

gions. The first supportive information provided the study of Bagneris et. al [8] that showed brominated co-crystallized compounds in fenestrations of the pore cavity in crystal structures of the Na_vM channel. These results were in concordance to the previously described binding sites in Table 4.1, which is illustrated by a superpositioning of the Na_vM crystal structure and a $Na_v1.8$ model shown in Figure 4.24. The observed binding modes in the Na_vM structure could be translated to the pore region of the $Na_v1.8$ channel.

Another influencing factor on the docking approach is the general conformation of the channel structure, as the models cover only a certain range of conformational states the channel can adopt, such as open or closed states. Studies have shown that for instance hERG open state models deliver better results in docking experiments [65, 133], because ligand binding residues become more exposed to the pore, and even side chain orientations change to favourable positions for ligand interactions. Furthermore it is known that ligands can act state dependent and only bind to specific receptor conformations [157]. Finally, it can be concluded that no discrimination of the CW and CCW domain association models was achieved with this docking approach.

4.3.5 Comparison to $Ca_v1.1$ Structure

In order to use the recent $Ca_v1.1$ cryo-EM structure (PDB: 3JBR) in comparison to the $Na_v1.8$ models, the structure had to be prepared by removing channel subunits not considered in this study, and by adding residue side chains, that were not resolved in the electron density of only 4.2 Å resolution. The CW and CCW association type models of $Ca_v1.1$ were created using self homology modeling and the rabbit $Ca_v1.1$ α-subunit sequence (Uniprot: P07293). Thus, the alignment of $Ca_v1.1$ sequence and structure was straight forward and gaps in the structure of chain A and chain D were bridged by connecting the TM3 and TM4 during the modeling process without introducing residues that could not be matched to structure coordinates. This avoided free modeling of additional loops, which do not participate in domain interaction and would have caused a divergence of the model structures from the template structure. The alignment was manually created in MOE [33] and the output was adapted to serve as input for MODELLER [69]. For each association type, 200 models were created using the *allhmodels()* option of MODELLER. The deviation parameter, which controls the randomization of coordinates during model building, was in this case increased to 8.0 (default: 4.0) to generate a diverse set of homology models, as especially the self-modeled CW solutions stick very close to the template structure. The mean

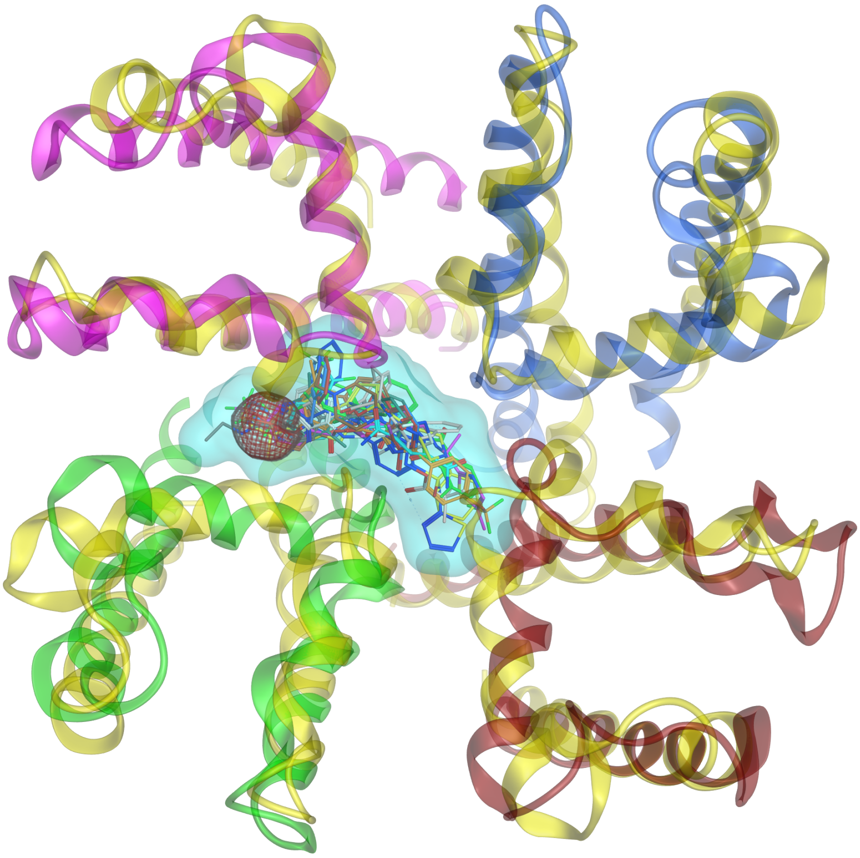


Figure 4.24: Superpositioning of Na$_v$M structure and Na$_v$1.8 model. Structural alignment of Na$_v$M crystal structure (yellow) and Na$_v$1.8 model (colored domains). Docked ligands are displayed with blue surface to show the area of binding in the pore of the Na$_v$1.8 model. The red lattice ball indicates the bromide density included in the Na$_v$M x-ray structure.

molpdf score, based on MODELLER's objective function, is used to rank the models according to their native-like conformation. The distribution of all MODELLER scores (Figure 4.25) shows on average better scores for CW models than for CCW models.

Further, the best ranked model for the CW and CCW domain association type was selected and processed according to the workflow described for the Na$_v$1.8 models. In short, both Ca$_v$1.1 models were embedded into a lipid membrane surrounded by a water box for a 20 ns MD simulation. The trajectories were first analysed in VMD, using the iTraj plug-in to identify chain contacts at the four domain interfaces. Figure 4.26 shows the number of contacts for CW model interfaces, that increase over time, starting at a lower count than the CCW model.

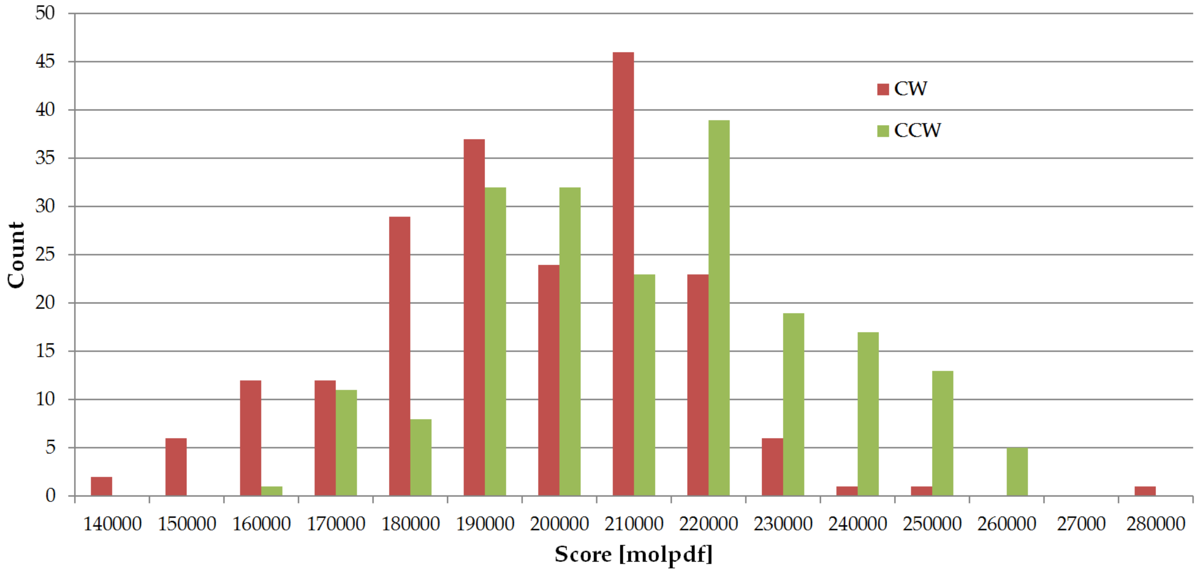


Figure 4.25: MODELLER Scores for Ca$_v$1.1 CW and CCW homology models. Molpdf scores are clustered into 15 categories and plotted as histogram. A low score number indicates a high ranked homology model. The distribution of all MODELLER scores shows in average lower scores for CW models than for CCW models, which indicates more native-like models in CW configuration.

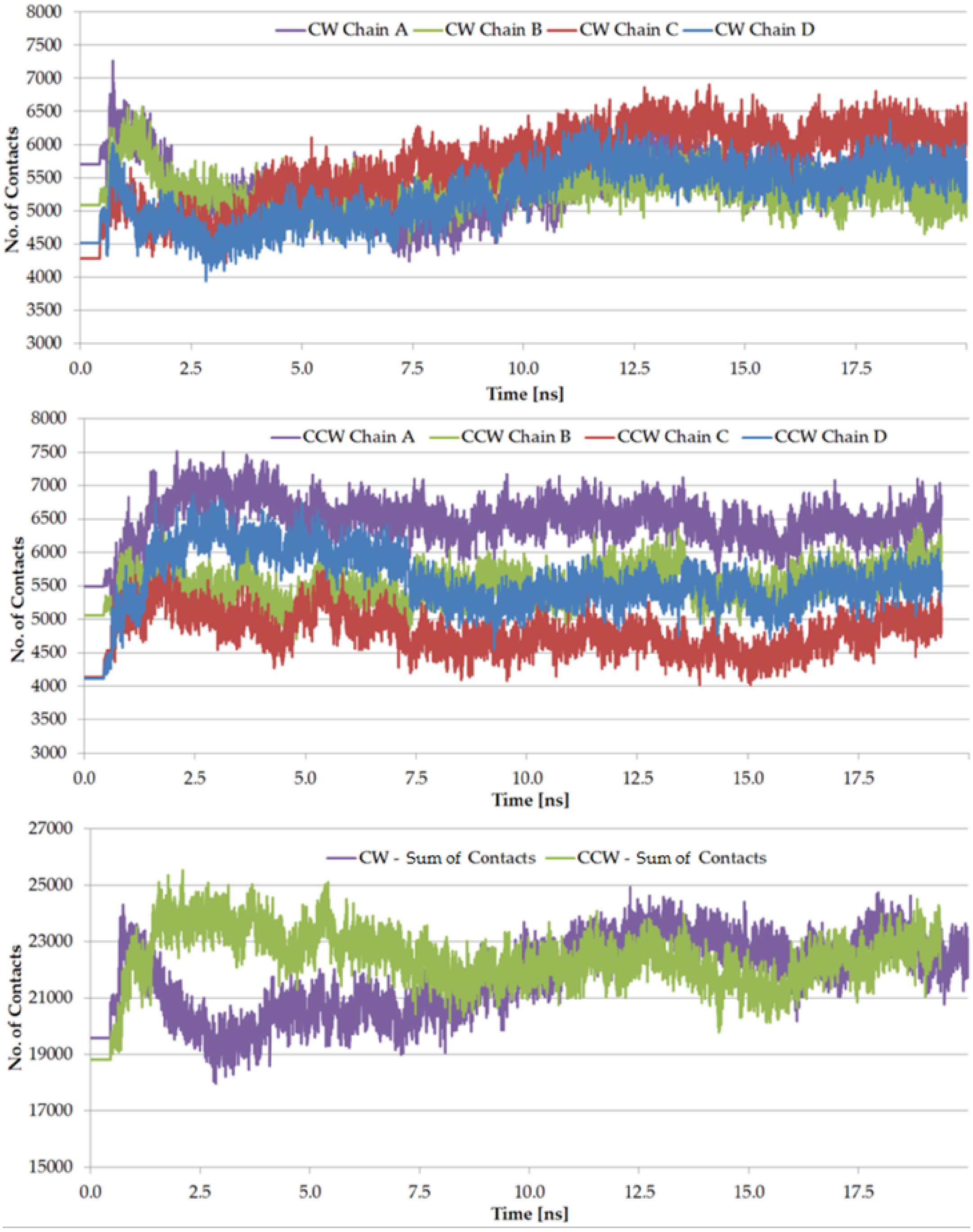


Figure 4.26: Domain contact count for Ca$_v$1.1 CW and CCW homology models. Interdomain contacts for each chain interface. The counts derive from the MD trajectories of the CW and CCW models. Larger variations of the interface contacts can be observed in the CCW model, assuming non homogeneous interfaces between the chain domains. The interface contacts in the CW model are more evenly distributed. The sum of the interface contacts over all chain domains are similar from the middle of the simulation onwards.

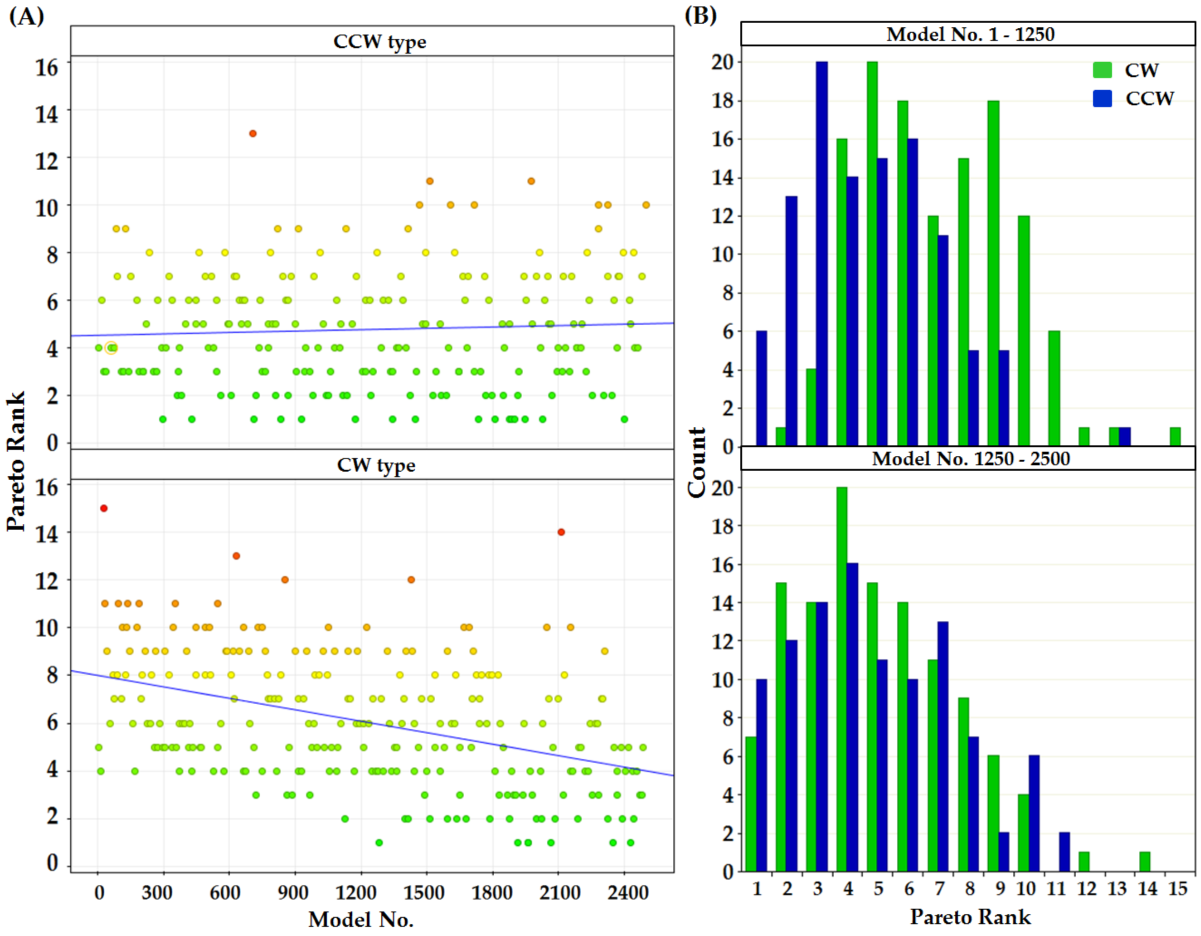


Figure 4.27: Time-resolved pareto ranked Ca$_v$1.1 models. (A) The domain association type models of CW and CCW are plotted by simulation time. The CW type shows a decreasing trend towards better pareto ranks, whereas the CCW type models, in the same time, are assigned to adverse pareto ranks. (B) The counts of CW and CCW type models are shown for each pareto rank in the first (upper histogram) and second (lower histogram) half of the analysed time period. Among the first set of models, the CCW type dominates the first pareto ranks, whereas in the second set the CW domain association type models are already equally present in the first two pareto ranks.

Pareto ranked ΔG values were binned and plotted as histogram shown in Figure 4.27 (B). The upper histogram, showing models for each pareto rank in the first half of the analysed time period, counts more CCW models than CW models in the first three pareto ranks, indicating the CCW model to be energetically favourable over the CW model. However, in the lower histogram, showing the second set of ranked models, the CW domain association type models are already equally present in the first two pareto ranks. This could be presumed, when examining the ΔG values plotted over simulation time in Figure 4.27 (A). It can be observed that the "mean" rank of CW models improve in the course of the simulation, resulting in CW models which are assigned to higher pareto ranks than the CCW models. A possible explanation for this could be that the starting model of CW has an unfavourable conformation compared to the CCW starting model, which was intensively modelled and relaxed while creating the CCW association, whereas the CW model is optimized during simulation and implicit structural relaxation. With extended simulation time it can be assumed, that the CW models begin to dominate the CCW models, showing best fitting interface contacts between the channel domains.

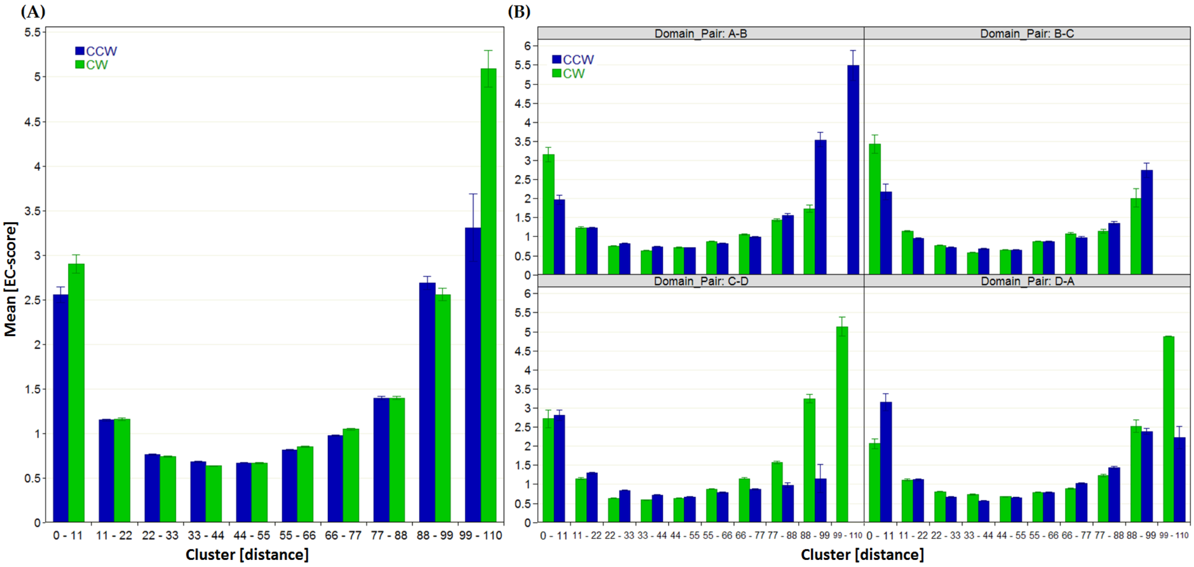


Figure 4.28: RaptorX Ca$_v$1.1 mean EC-scores. The EC-scores are clustered according to the distance of the related residue pairs. The higher the EC-score value the stronger a residue pair is coupled. (A) shows aggregated data for the complete channel, whereas (B) displays the EC-scores split by domain pair.

Evolutionary coupling scores were also calculated for the Ca$_v$1.1 sequence to evaluate the predictive capability of the method when applied on the ion channel protein class, just reaching the minimum numbers of required sequences but knowing the clockwise orientation of the Ca$_v$1.1 structure. The RaptorX web server, among the three web servers employed before, was used for this analysis, because it provided the most convenient user interface, requiring only the Ca$_v$1.1 protein sequence containing 1646 AA, in this case originated from rabbit (Uniprot P07293). The Ca$_v$1.1 sequence was truncated to a length of 591 AA, and 2423 sequences were identified and merged to a MSA by RaptorX for the direct coupling analysis.

The results were analysed according to the procedure described for Na$_v$1.8 (4.2.6) and visualized in Figure 4.28. As expected due to the experimental structure, the results show higher EC-scores for the CW model in the first cluster, even though the difference between CW and CCW is moderate. The residue-wise mapping of the EC-scores to ΔG results from MOE alanine scanning also documents the $Ca_v1.1$ CW type to be correct. Finally it can be stated that the DCA method delivers useful results, even for proteins with a long sequence but marginal number of sequences.

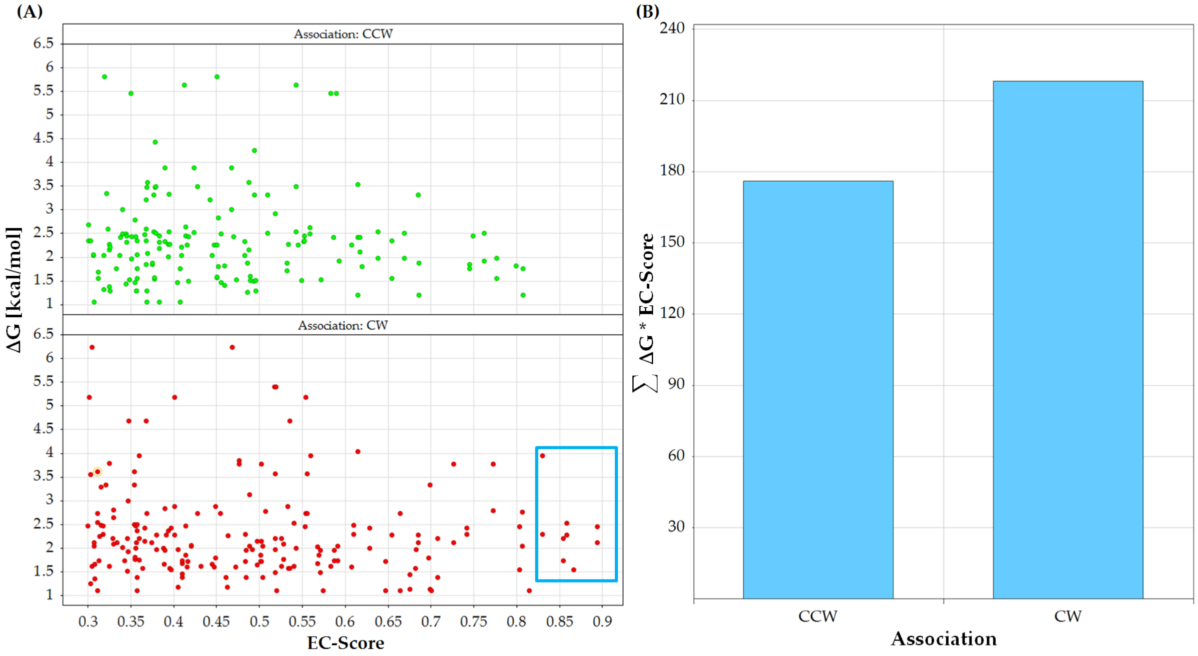


Figure 4.29: Association type measure based on EC-score and ΔG results for $Ca_v1.1$. (A) RaptorX EC-scores plotted against ΔG results from MOE alanine scanning. The high ΔG results are exclusively related to Trp residues, and this effect seems to be related to the energy evaluation method or the forcefield MOE applied. Comparable to $Na_v1.8$, residues in the blue frame show high EC-scores and ΔG values for the CW model, but were not observed for CCW. (B) Each bar represents the summation of the product between ΔG and EC-score of each analysed residue.

4.4 Conclusion

Voltage-gated sodium channels play a central role in the generation of action potentials and signal transduction for important physiological processes. The VGSC $Na_v1.8$ was chosen for this analysis, because significant reduction in mechanical allodynia and hyperalgesia [51] was shown in rat inflammatory and neuropathic pain models for this channel subtype. These findings make the target class of general interest for the identification of new medicines. In the past, most approaches to identify new compound classes have been ligand based, for instance with pharmacophore models [54, 170]. Especially in cycles of lead structure optimization, precise knowledge of protein-ligand interactions can guide the identification of suitable substituents. Within the scope of mammalian heterotetrameric ion channels and the current lack of crystal structure templates to build models of such an ion channel, it was investigated how to build models of the VGSC $Na_v1.8$ on the basis of homomeric x-ray structures using a comprehensive in silico approach. At the same time the understanding of the domain association in voltage-gated ion channels is essential to build accurate homology models for the drug discovery process. So far, reported experimental results on the basis of mutagenesis data were contradictory, and the provided results were not sufficient to exclude all other possible associations like the counter-clockwise orientation entirely. Therefore, all possible domain associations for the sodium voltage-gated ion channel $Na_v1.8$ have been investigated by computational methods. Homology models in clockwise and counter-clockwise domain association as well as all possible cross-wise variations have been applied to time-resolved protein-protein interaction analysis and various energetic measures to characterize the differences between these associations. Experimental measures were incorporated in the evaluation of the computational models and for comparison of all results generated based on these models. The cryo-EM structure of $Ca_v1.1$, published right after this study, provided additional information for an evaluation of the applied methods and results, although it is only a related ion channel type and not necessarily identical to $Na_v1.x$ channels regarding the domain association. In summary, taking all generated and analysed results into account, the findings suggest a clockwise association of the four domains in human heterotetrameric voltage-gated sodium channels, which is in agreement with published mutagenesis results that conclude a clockwise orientation.

In order to support this conclusion, the $Na_v1.8$ MD simulations could be extended by using additional computational capacity or applying methods like coarse-grained MD simulation, which reduces the number of elements in the simulated system and enables longer simulation times, although the reconstruction of the system to an atomic level needs careful practice as described in this extensive review [108]. As indicated in Figures 4.15 and 4.27, longer simulations and even models of different channel states might lead to a better discrimination between the association types, illustrated by an increasing number of CW models in the first pareto rank.

5 Summary and Outlook

5.1 Summary

This work dealt with questions about the assembly of ion channel domains and their association to form a functional ion channel pore. The questions were addressed with different computational methods comprising sequence and structure based methods such as homology modeling and direct coupling analysis (DCA), as well as methods to estimate energetic properties like molecular dynamics simulation, alanine scanning and molecular docking.

The analysis of the LGIC 5-HT$_{3A}$ revealed key residue interactions for the assembly of a domain subunit. Results from an alanine scanning mutagenesis experiment were used to guide a model creation process, as no 5-HT$_{3A}$ structure was available at that time. The 5-HT$_{3A}$ model was used to map the mutated residues that showed a severe effect on the assembly and function of the ion channel. These critical interactions of residues on the TM4 helix were energetically estimated, and the results were in agreement with the experimental data. Nevertheless, not all experimental results could be explained with the help of the 5-HT$_{3A}$ model, which did not represent all parts of the domain. This was finally possible when the 5-HT$_{3A}$ x-ray structure was published, and all results generated by then were verified and appeared to be in good agreement to the 5-HT$_{3A}$ structure, which indicated that the presented modeling approach was suitable to conclude certain key interactions, like the ionic lock between Asp459 and Arg277, not described for 5-HT$_{3A}$ before.

In case of the Na$_v$1.8 VGSC, measures to assess the correct domain association of the heterotetrameric protein were investigated. Based on models of all possible domain associations types, the domain interface interactions were analysed to identify the best fitting domain association. Measuring the binding energies between the interfaces on a feasible complexity level due to computational limitations, was not conclusive by itself. Thus, the sequence based and therefore complementary DCA method was used to identify key residues that could be matched to structure models and correlated to single residue binding energy estimates. The combination of the DCA results and models' structural information allowed a discrimination between the clockwise and counter-clockwise models, which was increased by incorporating estimated single binding energies. A verification of the results became possible to a certain extent, by a published heterotetrameric calcium channel, indicating a clockwise conformation and thus was in agreement to the prediction for Na$_v$1.8.

5.2 Outlook - Chemogenomics

The analysed question of the correct domain association appeared while preparing a chemogenomics approach [166] on the VGIC family, for which homology models were built to identify and compare potential binding sites. Chemogenomics integrates structural information of target receptors with available activity data of potential drug molecules, and gains increasing attention due to the growing amount of data generated. The principle of "similar ligands bind to similar binding sites" [107] is the basis for certain drug design approaches [131]. In the pharmaceutical industry special interest lies in off-target activity a compound might show, causing side effects when applied to a patient. It is intended to design target specific molecules with no or strongly reduced side effects. Nevertheless, in some occasions side effects of a molecules offer the chance to optimize the activity towards the off-target and develop a drug for an initially unintended indication. A well know example for observing an unintended indication is Sildenafil, which is now marketed as Viagra for the treatment of penile erectile dysfunction, but was initially developed as treatment for cardiovascular disorders, until significant side effects, discovered during a trial, guided the scientists to another indication [123].

Two basic approaches representing different perspectives are described in the context of chemogenomics.

Ligand Based Approach An example for a ligand based chemogenomics approach is described by Fauzi et al. [66], who characterize the mode of action for compounds from traditional Chinese medicines. The authors created a target prediction tool, which used fingerprints based on bioactivity data from ChEMBL to classify new compounds according to their similarity to known compounds and related bioactivity profiles. In principle this tool contains a list of compounds with associated targets, and new compounds are compared to all know molecules in the list while an algorithm predicts potential targets for this compound. A comparable method called Similarity Ensemble Approach (SEA) is used by Keiser et al. [103] to quantitatively cluster proteins based on the chemical similarity of 65000 annotated ligands. Some unexpected relationships were tested experimentally and the results were consistent with the predictions.

Target Based Approach A target based approach was used by Martin et al. [131] to discover a novel somatostatin receptor subunit 5 (SST5R) antagonist. Due to missing reference compounds for the target receptor, the approach was based on a sequence analysis of the binding sites of GPCRs closely related to SST5R. The active reference compounds of GPCRs with highly homologous binding site sequences were screened against SST5R, and the identified hits provided starting point for a drug design process.

Preliminary work The binding sites in VGIC were only sparsely described in the literature, which was a major limitation for docking approaches using the afore mentioned homology models, as the search space could not be specified precisely for efficient pose sampling. Therefore, ligand based chemogenomics was applied in a first step by searching for ligands that are effective on VGICs as well as on another protein classes. In addition, these ligands were checked against the PDB in order to find a protein ligand complex structure, which contains information about the binding mode of the ligand and thus a defined binding site. This binding site was then used as input for a search against the collection of generated VGIC homology models, changing the perspective from the ligand to the target space. Furthermore, the complete PDB was scanned to identify additional protein structures, which are related to VGICs through the searched binding site. This binding site search was performed with the Epitope-Match software [93] [`http://www.epitopematch.org/`], which was developed in Prof. Hoffmann's research group and implements algorithms for binding site comparisons of a query pocket against the PDB or custom structure sets.

The preliminary work on this study resulted in a first hit, the Spironolactone, which is crystallized in the Mineralocorticoid receptor (MR) 2AB2 [18] and is related to the potassium channels $K_v1.5$, $K_v4.3$ and $K_v7.1$ as active molecule [75]. The binding site of MR was then used to search a similar topology of residues in the target space of the PDB, and identified the K_v structure of 1ORQ as a probable binding site.

The fit between spironolactone and the K_v binding site looks promising, even though both structures remained in their original conformations. A flexible docking

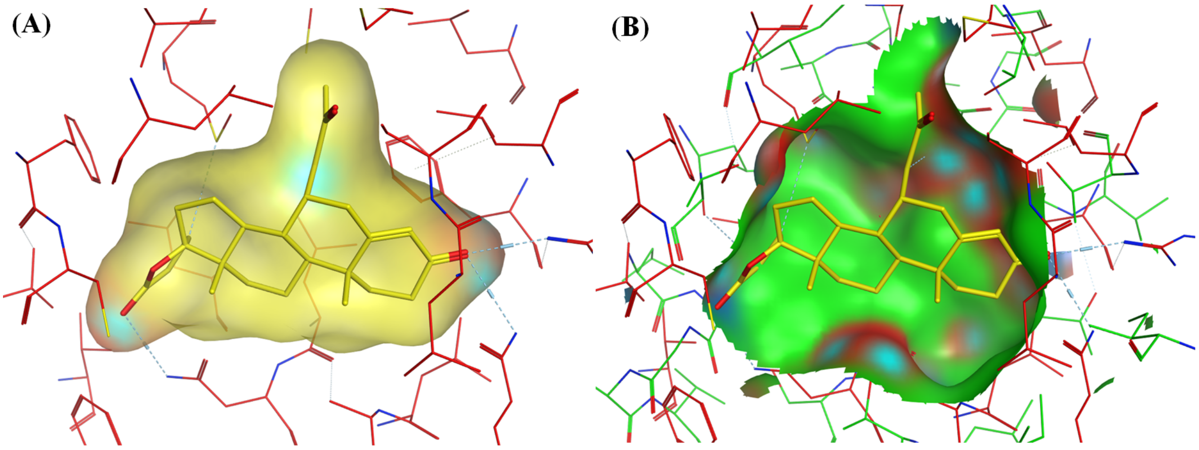


Figure 5.1: Spironolactone in MR and potential K$_v$ binding site. (A) Spironolactone shown with a molecular surface in the MR binding site of structure 2AB2. (B) Spironolactone placed in a potential K$_v$ binding site. MR (red sticks) and K$_v$ (green sticks) structures are superposed.

of spironolactone or an MD simulation of the complex might lead to an optimized fit and generates additional information about the binding site. On the other hand, the binding site could be used for ligand screening that could lead to valuable molecular scaffolds for ion channel drug design.

Bibliography

[1] N. L. Absalom, T. M. Lewis, and P. R. Schofield. Mechanisms of channel gating of the ligand-gated ion channel superfamily inferred from protein structure. *Experimental Physiology*, 89(2):145–153, feb 2004. 14

[2] S. Ahuja, S. Mukund, L. Deng, K. Khakh, E. Chang, H. Ho, S. Shriver, C. Young, S. Lin, J. P. Johnson, P. Wu, J. Li, M. Coons, C. Tam, B. Brillantes, H. Sampang, K. Mortara, K. K. Bowman, K. R. Clark, A. Estevez, Z. Xie, H. Verschoof, M. Grimwood, C. Dehnhardt, J.-C. Andrez, T. Focken, D. P. Sutherlin, B. S. Safina, M. A. Starovasnik, D. F. Ortwine, Y. Franke, C. J. Cohen, D. H. Hackos, C. M. Koth, and J. Payandeh. Structural basis of Nav1.7 inhibition by an isoform-selective small-molecule antagonist. *Science*, 350(6267):aac5464–aac5464, dec 2015. 89

[3] A. Aksimentiev, M. Sotomayor, D. Wells, and Z. Huang. *Membrane Proteins Tutorial*. University of Illinois, Beckman Institute for Advanced Science and Technology, Urbana-Champaign, feb 2012. 30, 63

[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, oct 1990. 23

[5] D. E. Anderson, W. J. Becktel, and F. W. Dahlquist. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry*, 29(9):2403–2408, mar 1990. 50

[6] P. A. Anderson and R. M. Greenberg. Phylogeny of ion channels: clues to structure and function. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 129(1):17–28, may 2001. 6, 8

[7] A. Auerbach. The energy and work of a ligand-gated ion channel. *Journal of Molecular Biology*, 425(9):1461–1475, may 2013. 14

[8] C. Bagneris, P. G. DeCaen, C. E. Naylor, D. C. Pryde, I. Nobeli, D. E. Clapham, and B. A. Wallace. Prokaryotic NavMs channel as a structural and functional model for eukaryotic sodium channel antagonism. *Proceedings of the National Academy of Sciences*, 111(23):8428–8433, may 2014. 29, 35, 36, 62, 91

[9] R. Bähring and M. Covarrubias. Mechanisms of closed-state inactivation in voltage-gated ion channels. *The Journal of Physiology*, 589(3):461–479, jan 2011. 13

[10] C. Baldassi, M. Zamparo, C. Feinauer, A. Procaccini, R. Zecchina, M. Weigt, and A. Pagnani. Fast and accurate multivariate gaussian modeling of protein families: Predicting residue contacts and protein-interaction partners. *PLoS ONE*, 9(3):e92721, mar 2014. 39

[11] J. A. Ballesteros, A. D. Jensen, G. Liapakis, S. G. F. Rasmussen, L. Shi, U. Gether, and J. A. Javitch. Activation of the 2-Adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *Journal of Biological Chemistry*, 276(31):29171–29177, may 2001. 50

[12] M. W. Barnett and P. M. Larkman. The action potential. *Practical Neurology*, 7(3):192–197, 2007. 3

[13] E. R. Benjamin, F. Pruthi, S. Olanrewaju, V. I. Ilyin, G. Crumley, E. Kutlina, K. J. Valenzano, and R. M. Woodward. State-dependent compound inhibition of Nav1.2 Sodium channels using the FLIPR Vm dye: On-target and off-target effects of diverse pharmacological agents. *Journal of Biomolecular Screening*, 11(1):29–39, oct 2005. 36

[14] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, M. Nowotka, G. Papadatos, R. Santos, and J. P. Overington. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42(D1):D1083–D1090, nov 2013. 36, 73

[15] H. Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, sep 1995. 29

[16] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, jan 2000. 9

[17] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, and B. Wiswedel. KNIME: The Konstanz Information Miner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*. Springer, 2007. 65

[18] R. K. Bledsoe, K. P. Madauss, J. A. Holt, C. J. Apolito, M. H. Lambert, K. H. Pearce, T. B. Stanley, E. L. Stewart, R. P. Trump, T. M. Willson, and S. P. Williams. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *Journal of Biological Chemistry*, 280(35):31283–31293, jun 2005. 101

[19] M. D. Blower, E. Feric, K. Weis, and R. Heald. Genome-wide analysis demonstrates conserved localization of messenger RNAs to mitotic microtubules. *J Cell Biol*, 179(7):1365–1373, dec 2007. 42

[20] A. C. Borel and S. M. Simon. Biogenesis of polytopic membrane proteins: Membrane segments assemble within translocation channels prior to membrane integration. *Cell*, 85(3):379 – 389, 1996. 40

[21] J. A. Bornhorst and J. J. Falke. [16] purification of proteins using polyhistidine affinity tags. In *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification*, volume 326 of *Methods in Enzymology*, pages 245 – 254. Academic Press, 2000. 41

[22] T. Braun, A. Orlova, K. Valegård, A.-C. Lindås, G. F. Schröder, and E. H. Egelman. Archaeal actin from a hyperthermophile forms a single-stranded filament. *Proceedings of the National Academy of Sciences*, 112(30):9340–9345, jun 2015. 49

[23] J. R. Brender and Y. Zhang. Predicting the effect of mutations on protein-protein binding interactions through structure-based interface profiles. *PLOS Computational Biology*, 11(10):e1004494, oct 2015. 45

[24] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983. 29

[25] L. E. Browne, F. E. Blaney, S. P. Yusaf, J. J. Clare, and D. Wray. Structural determinants of drugs acting on the Nav1.8 channel. *Journal of Biological Chemistry*, 284(16):10523–10536, feb 2009. 62

[26] C. Bystroff and Y. Shao. Fully automated ab initio protein structure prediction using i-SITES, HMMSTR and ROSETTA. *Bioinformatics*, 18(Suppl 1):S54–S61, jul 2002. 24

[27] E. Callaway. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature*, 525(7568):172–174, sep 2015. 10

[28] E. Cao, M. Liao, Y. Cheng, and D. Julius. TRPV1 structures in distinct conformations reveal mechanisms of activation. *Nature*, 504(7478):113–118, dec 2013. 10

[29] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. Amber 12, 2012. 29

[30] W. A. Catterall. Structure and function of voltage-gated ion channels. *Trends in Neurosciences*, 16(12):500–506, dec 1993. 6

[31] W. A. Catterall. Ion channel voltage sensors: Structure, function, and pathophysiology. *Neuron*, 67(6):915–928, sep 2010. 13

[32] N. Chakrabarti, C. Ing, J. Payandeh, N. Zheng, W. A. Catterall, and R. Pomes. Catalysis of Na+ permeation in the bacterial sodium channel NavAb. *Proceedings of the National Academy of Sciences*, 110(28):11331–11336, jun 2013. 16

[33] Chemical Computing Group Inc., 1010 Sherbrooke St. West, Suite ♯910, Montreal, QC, Canada, H3A 2R7. *Molecular Operating Environment (MOE)*, 2015.10 edition, 2016. 26, 45, 46, 63, 91

[34] P.-C. Chen and S. Kuyucak. Accurate determination of the binding free energy for KcsA-charybdotoxin complex from the potential of mean force calculations with restraints. *Biophysical Journal*, 100(10):2466–2474, may 2011. 45

[35] S. Y. Chiu, J. M. Ritchie, R. B. Rogart, and D. Stagg. A quantitative description of membrane currents in rabbit myelinated nerve. *The Journal of Physiology*, 292(1):149–166, jul 1979. 13

[36] L. D. Colibus, X. Wang, J. A. B. Spyrou, J. Kelly, J. Ren, J. Grimes, G. Puerstinger, N. Stonehouse, T. S. Walter, Z. Hu, J. Wang, X. Li, W. Peng, D. J. Rowlands, E. E. Fry, Z. Rao, and D. I. Stuart. More-powerful virus inhibitors from structure-based analysis of HEV71 capsid-binding molecules. *Nature Structural & Molecular Biology*, 21(3):282–288, feb 2014. 36, 76

[37] G. L. Collingridge, R. W. Olsen, J. Peters, and M. Spedding. A nomenclature for ligand-gated ion channels. *Neuropharmacology*, 56(1):2–5, jan 2009. Ligand-Gated Ion Channels. 9

[38] T. U. Consortium. UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, oct 2014. 23

[39] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, may 1995. 27

[40] B. Corry and M. Thomas. Mechanism of ion permeation and selectivity in a voltage gated sodium channel. *J. Am. Chem. Soc.*, 134(3):1840–1846, jan 2012. PMID: 22191670. 16

[41] S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell, and A. J. Olson. Virtual screening with AutoDock: theory and practice. *Expert Opinion on Drug Discovery*, 5(6):597–607, apr 2010. 34

[42] J. J. Cox, F. Reimann, A. K. Nicholas, G. Thornton, E. Roberts, K. Springell, G. Karbani, H. Jafri, J. Mannan, Y. Raashid, L. Al-Gazali, H. Hamamy, E. M. Valente, S. Gorman, R. Williams, D. P. McHale, J. N. Wood, F. M. Gribble, and C. G. Woods. An SCN9A channelopathy causes congenital inability to experience pain. *Nature*, 444(7121):894–898, dec 2006. 19, 20

[43] J. A. R. Dalton and R. M. Jackson. An evaluation of automated homology modelling methods at low target-template sequence similarity. *Bioinformatics*, 23(15):1901–1908, may 2007. 24

[44] J. F. Danielli and H. Davson. A contribution to the theory of permeability of thin films. *J. Cell. Comp. Physiol.*, 5(4):495–508, feb 1935. 1

[45] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993. 31, 48, 64

[46] D. de Juan, F. Pazos, and A. Valencia. Emerging methods in protein co-evolution. *Nature Reviews Genetics*, 14(4):249–261, mar 2013. 69

[47] M. de Lera Ruiz and R. L. Kraus. Voltage-gated sodium channels: Structure, function, pharmacology, and clinical indications. *J. Med. Chem.*, 58(18):7093–7118, sep 2015. PMID: 25927480. 19

[48] C. Derst and A. Karschin. Evolutionary link between prokaryotic and eukaryotic K+ channels. *Journal of Experimental Biology*, 201(20):2791–2799, oct 1998. 5

[49] A. Dhanik, J. S. McMurray, and L. Kavraki. On modeling peptidomimetics in complex with the SH2 domain of Stat3. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Institute of Electrical and Electronics Engineers (IEEE), aug 2011. 36

[50] C. Dominguez, R. Boelens, and A. M. J. J. Bonvin. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, 125(7):1731–1737, feb 2003. 34

[51] X.-W. Dong, S. Goregoaker, H. Engler, X. Zhou, L. Mark, J. Crona, R. Terry, J. Hunter, and T. Priestley. Small interfering RNA-mediated selective knockdown of NaV1.8 tetrodotoxin-resistant sodium channel reverses mechanical allodynia in neuropathic rats. *Neuroscience*, 146(2):812–821, may 2007. 97

[52] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon. The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. *Science*, 280(5360):69–77, apr 1998. 15

[53] J. P. Drenth, R. H. te Morsche, G. Guillet, A. Taieb, R. L. Kirby, and J. B. Jansen. SCN9A mutations define primary erythermalgia as a neuropathic disorder of voltage gated sodium channels. *Journal of Investigative Dermatology*, 124(6):1333–1338, jun 2005. 20

[54] O. Dror, D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson. Novel approach for efficient pharmacophore-based virtual screening: Method

and applications. *Journal of Chemical Information and Modeling*, 49(10):2333–2343, oct 2009. 98

[55] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*, 43(W1):W389–W394, apr 2015. 24

[56] A. Drozdetskiy, C. Cole, J. Procter, and G. J. Barton. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res*, 43(W1):W389–W394, apr 2015. 59

[57] T. Dudev and C. Lim. Determinants of K+ vs Na+ selectivity in potassium channels. *J. Am. Chem. Soc.*, 131(23):8092–8101, jun 2009. PMID: 19456150. 16

[58] S. C. Dudley, N. Chang, J. Hall, G. Lipkind, H. A. Fozzard, and R. J. French. μ-conotoxin Giiia interactions with the voltage-gated na+ channel predict a clockwise arrangement of the domains. *The Journal of General Physiology*, 116(5):679–690, oct 2000. 58, 80

[59] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, mar 2004. 23, 61

[60] E. C. Emery, A. P. Luiz, and J. N. Wood. Nav1.7 and other voltage-gated sodium channels as drug targets for pain relief. *Expert Opinion on Therapeutic Targets*, 20(8):975–983, apr 2016. PMID: 26941184. 20

[61] D. M. Engelman, Y. Chen, C.-N. Chin, A. Curran, A. M. Dixon, A. D. Dupuy, A. S. Lee, U. Lehnert, E. E. Matthews, Y. K. Reshetnyak, A. Senes, and J.-L. Popot. Membrane protein folding: beyond the two stage model. *FEBS Letters*, 555(1):122–125, oct 2003. 40

[62] R. A. Engh and R. Huber. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Cryst Sect A*, 47(4):392–400, jul 1991. 27

[63] D. A. Erlanson, W. Jahnke, R. Mannhold, H. Kubinyi, and G. Folkers. *Fragment-based Drug Discovery: Lessons and Outlook*. Wiley-VCH, 1 edition, feb 2016. 74

[64] A. M. Esposito and T. G. Kinzy. In vivo [35 S]-methionine incorporation. In *Methods in Enzymology*, volume 536 of *Methods in Enzymology*, pages 55–64. Elsevier BV, 2014. 41

[65] R. Farid, T. Day, R. A. Friesner, and R. A. Pearlstein. New insights about hERG blockade obtained from protein modeling, potential energy mapping,

and docking studies. *Bioorganic & Medicinal Chemistry*, 14(9):3160–3173, may 2006. 91

[66] F. M. Fauzi, A. Koutsoukas, R. Lowe, K. Joshi, T.-P. Fan, R. C. Glen, and A. Bender. Chemogenomics approaches to rationalizing the mode-of-action of traditional chinese and ayurvedic medicines. *Journal of Chemical Information and Modeling*, 53(3):661–673, mar 2013. 100

[67] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks. Constant pressure molecular dynamics simulation: The langevin piston method. *The Journal of Chemical Physics*, 103(11):4613–4621, 1995. 31, 48, 64

[68] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta. Pfam: the protein families database. *Nucleic Acids Research*, 42(D1):D222–D230, nov 2013. 69

[69] A. Fiser, R. K. G. Do, and A. Švali. Modeling of loops in protein structures. *Protein Sci.*, 9(9):1753–1773, 2000. 62, 91

[70] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 47(7):1739–1749, mar 2004. 34

[71] S. Genheden. Are homology models sufficiently good for free-energy simulations? *Journal of Chemical Information and Modeling*, 52(11):3013–3021, nov 2012. 25

[72] M. Gielen, P. Thomas, and T. G. Smart. The desensitization gate of inhibitory Cys-loop receptors. *Nature Communications*, 6:6829, apr 2015. 15

[73] K. Ginalski. Comparative modeling for protein structure prediction. *Current Opinion in Structural Biology*, 16(2):172–177, apr 2006. 24

[74] A. L. Goldin. Mechanisms of sodium channel inactivation. *Current Opinion in Neurobiology*, 13(3):284–290, 2003. 14

[75] R. Gómez, L. Núñez, R. Caballero, M. Vaquero, J. Tamargo, and E. Delpón. Spironolactone and its main metabolite canrenoic acid block hKv1.5, Kv4.3 and Kv7.1 +minK channels. *British Journal of Pharmacology*, 146(1):146–161, sep 2005. 101

[76] A. Gupta, N. Chaudhary, K. R. Kakularam, R. Pallu, and A. Polamarasetty. The augmenting effects of desolvation and conformational energy terms on the predictions of docking programs against mPGES-1. *PLOS ONE*, 10(8):e0134472, aug 2015. 36, 76

[77] R. Gurezka, R. Laage, B. Brosig, and D. Langosch. A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *Journal of Biological Chemistry*, 274(14):9265–9270, apr 1999. 41, 57

[78] S. Haeger, D. Kuzmin, S. Detro-Dassen, N. Lang, M. Kilb, V. Tsetlin, H. Betz, B. Laube, and G. Schmalzing. An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nat Struct Mol Biol*, 17(1):90–98, dec 2009. 41, 57

[79] T. A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.*, 17(5-6):490–519, apr 1996. 27

[80] C. Han, J. Huang, and S. G. Waxman. Sodium channel Nav1.8: Emerging links to human disease. *Neurology*, 86(5):473–483, jan 2016. 59

[81] G. Hassaine, C. Deluz, L. Grasso, R. Wyss, M. B. Tol, R. Hovius, A. Graff, H. Stahlberg, T. Tomizaki, A. Desmyter, C. Moreau, X.-D. Li, F. Poitevin, H. Vogel, and H. Nury. X-ray structure of the mouse serotonin 5-HT3 receptor. *Nature*, 512(7514):276–281, aug 2014. 9, 41, 52, 54, 55

[82] M. Havrila, M. Zgarbová, P. Jurečka, P. Banáš, M. Krepl, M. Otyepka, and J. Šponer. Microsecond-scale MD simulations of HIV-1 DIS kissing-loop complexes predict bulged-in conformation of the bulged bases and reveal interesting differences between available variants of the AMBER RNA force fields. *The Journal of Physical Chemistry B*, 119(49):15176–15190, dec 2015. PMID: 26548477. 77

[83] S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89(22):10915–10919, nov 1992. 46

[84] R. J. C. Hilf and R. Dutzler. X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature*, 452(7185):375–379, mar 2008. 41

[85] D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey, and T. Lengauer. Two-stage method for protein-ligand docking. *J. Med. Chem.*, 42(21):4422–4433, oct 1999. PMID: 10543886. 34

[86] T. Hou, J. Wang, Y. Li, and W. Wang. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. the accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 51(1):69–82, jan 2011. 45

[87] X. Hu and B. Kuhlman. Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences. *Proteins: Structure, Function, and Bioinformatics*, 62(3):739–748, nov 2005. 45

[88] X. Huang, H. Chen, K. Michelsen, S. Schneider, and P. L. Shaffer. Crystal structure of human glycine receptor-$\alpha3$ bound to antagonist strychnine. *Nature*, 526(7572):277–280, sep 2015. 41

[89] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, 28(6):1145–1152, 2007. 35, 89

[90] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, feb 1996. 29, 63

[91] K. Illergård, D. H. Ardell, and A. Elofsson. Structure is three to ten times more conserved than sequence-a study of structural response in protein cores. *Proteins*, 77(3):499–508, nov 2009. 25

[92] L. L. Isom, K. S. D. Jongh, and W. A. Catterall. Auxiliary subunits of voltage-gated ion channels. *Neuron*, 12(6):1183–1194, jun 1994. 6

[93] S. Jakuschev and D. Hoffmann. A novel algorithm for macromolecular epitope matching. *Algorithms*, 2(1):498–517, mar 2009. 101

[94] V. Jogini and B. Roux. Dynamics of the Kv1.2 Voltage-gated k+ channel in a membrane environment. *Biophysical Journal*, 93(9):3070–3082, nov 2007. 13

[95] L. S. Johnson, S. R. Eddy, and E. Portugaly. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, 11(1):431, 2010. 69

[96] D. T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*, 292(2):195–202, sep 1999. 24

[97] G. Jones, P. Willett, and R. C. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of Molecular Biology*, 245(1):43–53, jan 1995. 34

[98] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926, 1983. 30, 63

[99] W. L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, mar 1988. 27

[100] S. K. Joshi, J. P. Mikusa, G. Hernandez, S. Baker, C.-C. Shieh, T. Neelands, X.-F. Zhang, W. Niforatos, K. Kage, P. Han, D. Krafte, C. Faltynek, J. P. Sullivan, M. F. Jarvis, and P. Honore. Involvement of the TTX-resistant sodium channel Nav1.8 in inflammatory and neuropathic, but not post-operative, pain states. *Pain*, 123(1):75–82, jul 2006. 19

[101] R. S. Kass. The channelopathies: novel insights into molecular and genetic mechanisms of human disease. *Journal of Clinical Investigation*, 115(8):1986–1989, aug 2005. 18, 20

[102] J. Kearney, N. Plummer, M. Smith, J. Kapur, T. Cummins, S. Waxman, A. Goldin, and M. Meisler. A gain-of-function mutation in the sodium channel gene Scn2a results in seizures and behavioral abnormalities. *Neuroscience*, 102(2):307–317, 2001. 20

[103] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, and B. K. Shoichet. Relating protein pharmacology by ligand chemistry. *Nature Biotechnology*, 25(2):197–206, feb 2007. 101

[104] S. Kellenberger, J. W. West, W. A. Catterall, and T. Scheuer. Molecular analysis of potential hinge residues in the inactivation gate of brain type IIA Na+ channels. *The Journal of General Physiology*, 109(5):607–617, may 1997. 14

[105] A. Keramidas, A. J. Moorhouse, P. R. Schofield, and P. H. Barry. Ligand-gated ion channels: mechanisms underlying ion selectivity. *Progress in Biophysics and Molecular Biology*, 86(2):161 – 204, 2004. 17, 18

[106] D. E. Kim, F. DiMaio, R. Y.-R. Wang, Y. Song, and D. Baker. One contact for every twelve residues allows robust and accurate topology-level protein structure modeling. *Proteins*, 82:208–218, sep 2013. 39

[107] T. Klabunde. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *British Journal of Pharmacology*, 152(1):5–7, sep 2007. 100

[108] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid, and A. Kolinski. Coarse-grained protein models and their applications. *Chemical Reviews*, 116(14):7898–7936, jul 2016. 98

[109] T. Kortemme, D. E. Kim, and D. Baker. Computational alanine scanning of protein-protein interfaces. *Science Signaling*, 2004(219):pl2–pl2, feb 2004. 53

[110] E. Krieger, G. Koraimann, and G. Vriend. Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. *Proteins*, 47(3):393–402, may 2002. 29

[111] A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, jan 2001. 24

[112] P. Labute. The generalized born/volume integral implicit solvent model: Estimation of the free energy of hydration using london dispersion instead of atomic surface area. *J. Comput. Chem.*, 29(10):1693–1698, feb 2008. 47

[113] P. Labute. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins*, 75(1):187–205, sep 2008. 47

[114] LadyofHats. Diagram of a typical myelinated vertebrate motor neuron. en.wikipedia.org, jul 2007. Public Domain. 5

[115] W. Landschulz, P. Johnson, and S. McKnight. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, 240(4860):1759–1764, jun 1988. 57

[116] G. Lauro, A. Romano, R. Riccio, and G. Bifulco. Inverse virtual screening of antitumor targets: Pilot study on a small database of natural bioactive compounds. *J. Nat. Prod.*, 74(6):1401–1407, jun 2011. 35

[117] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, and P. Bradley. Rosetta3. In *Computer Methods, Part C*, volume 487 of *Methods in Enzymology*, pages 545–574. Elsevier BV, 2011. 45, 67

[118] S. M. Lewis and B. A. Kuhlman. Anchored design of protein-protein interfaces. *PLoS ONE*, 6(6):1–14, jun 2011. 45, 49

[119] R. A. Li and G. F. Tomaselli. Using the deadly μ-conotoxins as probes of voltage-gated sodium channels. *Toxicon*, 44(2):117–122, aug 2004. 58

[120] G. M. Lipkind and H. A. Fozzard. Voltage-gated Na channel selectivity: The role of the conserved domain III lysine residue. *The Journal of General Physiology*, 131(6):523–529, may 2008. 61

[121] Y. Liu, M. E. Jurman, and G. Yellen. Dynamic rearrangement of the outer mouth of a K+ channel during gating. *Neuron*, 16(4):859–867, 1996. 58, 80

[122] S. B. Long, E. B. Campbell, and R. MacKinnon. Voltage sensor of Kv1.2: Structural basis of electromechanical coupling. *Science*, 309(5736):903–908, aug 2005. 13

[123] B. M, A. MJ, B. SA, G.-A. S, M. GJ, N. AM, O. IH, and G. C. Sildenafil: an orally active type 5 cyclic GMP-specific phosphodiesterase inhibitor for the treatment of penile erectile dysfunction. *J Impot Res.*, 8(2):47–52, jun 1996. 100

[124] J. Ma, S. Wang, Z. Wang, and J. Xu. Protein contact prediction by integrating joint evolutionary coupling analysis and supervised learning. *Bioinformatics*, 31(21):3506–3513, aug 2015. 39

[125] J. L. MacCallum, M. S. Moghaddam, H. S. Chan, and D. P. Tieleman. Hydrophobic association of α-helices, steric dewetting, and enthalpic barriers to protein folding. *Proceedings of the National Academy of Sciences*, 104(15):6206–6210, apr 2007. 45

[126] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, apr 1998. 27, 29, 63

[127] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, apr 1998. 28

[128] M. Manna, W. Kulig, M. Javanainen, J. Tynkkynen, U. Hensen, D. J. MÃijller, T. Rog, and I. Vattulainen. How to minimize artifacts in atomistic simulations of membrane proteins, whose crystal structure is heavily engineered: β 2 -adrenergic receptor in the spotlight. *J. Chem. Theory Comput.*, 11(7):3432–3445, jul 2015. PMID: 26575777. 10

[129] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander. Protein 3D structure computed from evolutionary sequence variation. *PLoS ONE*, 6(12):e28766, dec 2011. 24

[130] D. S. Marks, T. A. Hopf, and C. Sander. Protein structure prediction from sequence variation. *Nature Biotechnology*, 30(11):1072–1080, nov 2012. 69

[131] R. E. Martin, L. G. Green, W. Guba, N. Kratochwil, and A. Christ. Discovery of the first nonpeptidic, small-molecule, highly selective somatostatin receptor subtype 5 Antagonists: A chemogenomics approach. *J. Med. Chem.*, 50(25):6291–6294, dec 2007. 100, 101

[132] G. J. Martyna, D. J. Tobias, and M. L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994. 31, 48, 64

[133] M. Masetti, A. Cavalli, and M. Recanatini. Modeling the hERG potassium channel in a phospholipid bilayer: Molecular dynamics and drug docking studies. *J. Comput. Chem.*, 29(5):795–808, 2008. 91

[134] J. C. McPhee, D. S. Ragsdale, T. Scheuer, and W. A. Catterall. A critical role for transmembrane segment IVS6 of the sodium channel α subunit in fast inactivation. *Journal of Biological Chemistry*, 270(20):12025–12034, may 1995. 13

[135] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui. Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer Aided-Drug Design*, 7(2):146–157, jun 2011. 74

[136] M. Michino, T. Beuming, P. Donthamsetti, A. H. Newman, J. A. Javitch, and L. Shi. What can crystal structures of aminergic receptors tell us about designing subtype-selective ligands? *Pharmacological Reviews*, 67(1):198–213, dec 2014. 74

[137] M. J. Mizianty and L. Kurgan. Modular prediction of protein structural classes from sequences of twilight-zone identity with predicting sequences. *BMC Bioinformatics*, 10(1):414, 2009. 25

[138] E. G. Moczydlowski. The molecular mystique of tetrodotoxin. *Toxicon*, 63:165–183, mar 2013. 19, 73

[139] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49):E1293–E1301, nov 2011. 39

[140] R. Moretti. Comment-8155. Rosetta Commons - The hub for Rosetta modeling software, dec 2014. www.rosettacommons.org/comment/8155#comment-8155. 45

[141] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662, nov 1998. 34

[142] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, and A. J. Olson. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16):2785–2791, dec 2009. 34, 75

[143] A. Nardi, N. Damann, T. Hertrampf, and A. Kless. Advances in targeting voltage-gated sodium channels with small molecules. *ChemMedChem*, 7(10):1712–1740, sep 2012. 19, 61

[144] M. A. Nassar, L. C. Stirling, G. Forlani, M. D. Baker, E. A. Matthews, A. H. Dickenson, and J. N. Wood. Nociceptor-specific gene deletion reveals a major role for Nav1.7 (PN1) in acute and inflammatory pain. *Proceedings of the National Academy of Sciences*, 101(34):12706–12711, aug 2004. 19

[145] A. Nayeem. A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science*, 15(4):808–824, mar 2006. 46

[146] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, mar 1970. 22, 46

[147] A. F. Neuwald. Gleaning structural and functional information from correlations in protein multiple sequence alignments. *Current Opinion in Structural Biology*, 38:1–8, jun 2016. 39

[148] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. V. Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25(13):1656–1676, 2004. 27

[149] M. O. Ortells and G. G. Lunt. Evolutionary history of the ligand-gated ion-channel superfamily of receptors. *Trends in Neurosciences*, 18(3):121–127, mar 1995. 6

[150] S. Ovchinnikov, H. Kamisetty, and D. Baker. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife*, 3:e02030, may 2014. 70

[151] R. D. Paris, C. V. Quevedo, D. D. Ruiz, O. N. de Souza, and R. C. Barros. Clustering molecular dynamics trajectories for optimizing docking experiments. *Computational Intelligence and Neuroscience*, 2015:1–9, 2015. 65

[152] M. M. Pathak, V. Yarov-Yarovoy, G. Agarwal, B. Roux, P. Barth, S. Kohout, F. Tombola, and E. Y. Isacoff. Closing in on the resting state of the Shaker K+ channel. *Neuron*, 56(1):124–140, oct 2007. 78

[153] C. E. Paulsen, J.-P. Armache, Y. Gao, Y. Cheng, and D. Julius. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature*, 525(7570):552–552, jul 2015. 10

[154] J. Payandeh, T. Scheuer, N. Zheng, and W. A. Catterall. The crystal structure of a voltage-gated sodium channel. *Nature*, 475(7356):353–358, jul 2011. 78

[155] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005. 29, 30, 47, 63

[156] B. G. Pierce, K. Wiehe, H. Hwang, B.-H. Kim, T. Vreven, and Z. Weng. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773, feb 2014. 34

[157] D. Ragsdale, J. McPhee, T. Scheuer, and W. Catterall. Molecular determinants of state-dependent block of Na+ channels by local anesthetics. *Science*, 265(5179):1724–1728, sep 1994. 91

[158] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, jul 1963. 27

[159] V. Ramadoss, F. Dehez, and C. Chipot. AlaScan: A graphical user interface for alanine scanning free-energy calculations. *Journal of Chemical Information and Modeling*, 56(6):1122–1126, jun 2016. 48

[160] D. Ramírez and J. Caballero. Is it reliable to use common molecular docking methods for comparing the binding affinities of enantiomer pairs for their protein target? *IJMS*, 17(4):525, apr 2016. 35, 38

[161] K. Rataj, J. Witek, S. Mordalski, T. Kosciolek, and A. J. Bojarski. Impact of template choice on homology model efficiency in virtual screening. *Journal of Chemical Information and Modeling*, 54(6):1661–1668, jun 2014. 25

[162] F. Ratjen and G. Döring. Cystic fibrosis. *The Lancet*, 361(9358):681–689, 2003. 20

[163] G. L. Reddy, T. Iwamoto, J. M. Tomich, and M. Montal. Synthetic peptides and four-helix bundle proteins as model systems for the pore-forming structure of channel proteins. II. Transmembrane segment M2 of the brain glycine receptor is a plausible candidate for the pore-lining structure. *Journal of Biological Chemistry*, 268(20):14608–14615, jul 1993. 46

[164] M. Remmert, A. Biegert, A. Hauser, and J. Söding. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods*, 9(2):173–175, dec 2011. 69, 70

[165] R. C. Rizzo, J. Tirado-Rives, and W. L. Jorgensen. Estimation of binding affinities for HEPT and nevirapine analogues with HIV-1 reverse transcriptase via Monte Carlo simulations. *J. Med. Chem.*, 44(2):145–154, jan 2001. 36

[166] D. Rognan. Chemogenomic approaches to rational drug design. *British Journal of Pharmacology*, 152(1):38–52, sep 2007. 100

[167] R. L. Ruff, L. Simoncini, and W. StÃijuhmer. Slow sodium channel inactivation in mammalian muscle: A possible role in regulating excitability. *Muscle & Nerve*, 11(5):502–510, may 1988. 13

[168] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, mar 1977. 31, 48, 64

[169] J. Sadowski, J. Gasteiger, and G. Klebe. Comparison of automatic three-dimensional model builders using 639 x-ray structures. *Journal of Chemical Information and Modeling*, 34(4):1000–1008, jul 1994. 75

[170] S. Sakkiah, S. Thangapandian, and K. W. Lee. Pharmacophore modeling, molecular docking, and molecular dynamics simulation approaches for identifying new lead compounds for inhibiting aldose reductase 2. *Journal of Molecular Modeling*, 18(7):3267–3282, jan 2012. 98

[171] C. L. Salussolia, A. Corrales, I. Talukder, R. Kazi, G. Akgul, M. Bowen, and L. P. Wollmuth. Interaction of the M4 segment with other transmembrane segments is required for surface expression of mammalian α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors. *Journal of Biological Chemistry*, 286(46):40205–40218, sep 2011. 41

[172] M. Sansom, P. Bond, S. Deol, A. Grottesi, S. Haider, and Z. Sands. Molecular simulations and lipid-protein interactions: potassium channels and other membrane proteins. *Biochm. Soc. Trans.*, 33(5):916–920, oct 2005. 56

[173] H. Scheib, I. McLay, N. Guex, J. J. Clare, F. E. Blaney, T. J. Dale, S. N. Tate, and G. M. Robertson. Modeling the pore structure of voltage-gated sodium channels in closed, open, and fast-inactivated conformation reveals details of site 1 toxin and local anesthetic binding. *Journal of Molecular Modeling*, 12(6):813–822, mar 2006. 62

[174] J. Schlitter, M. Engels, and P. Krüger. Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*, 12(2):84–89, jun 1994. 45

[175] D. Schmidt, Q.-X. Jiang, and R. MacKinnon. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature*, 444(7120):775–779, nov 2006. 56

[176] Z. shan Wu, H. Cheng, Y. Jiang, K. Melcher, and H. E. Xu. Ion channels gated by acetylcholine and serotonin: structures, biology, and drug discovery. *Acta Pharmacologica Sinica*, 36(8):895–907, aug 2015. 15

[177] J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.*, 3(6):2312–2334, nov 2007. 65

[178] I. H. Shrivastava and M. S. Sansom. Simulations of ion permeation through a potassium channel: Molecular dynamics of KcsA in a phospholipid bilayer. *Biophysical Journal*, 78(2):557–570, feb 2000. 47

[179] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, 7(1):539–539, apr 2014. 69

[180] S. J. Singer and G. L. Nicolson. The fluid mosaic model of the structure of cell membranes. *Science*, 175(4023):720–731, feb 1972. 1

[181] J. Skolnick, H. Zhou, and M. Gao. Are predicted protein structures of any value for binding site prediction and virtual ligand screening? *Current Opinion in Structural Biology*, 23(2):191–197, apr 2013. 25

[182] J. C. Skou. The influence of some cations on an adenosine triphosphatase from peripheral nerves. *Biochimica et Biophysica Acta*, 23:394–401, jan 1957. 2

[183] T. Smith and M. Waterman. Identification of common molecular subsequences. *Journal of Molecular Biology*, 147(1):195–197, mar 1981. 22

[184] B. Sommer. MEMBRANE PACKING PROBLEMS: A SHORT REVIEW ON COMPUTATIONAL MEMBRANE MODELING METHODS AND TOOLS. *Computational and Structural Biotechnology Journal*, 5(6):1–13, feb 2013. 64

[185] K. Sousounis, C. E. Haney, J. Cao, B. Sunchu, and P. A. Tsonis. Conservation of the three-dimensional structure in non-homologous or unrelated proteins. *Human Genomics*, 6(1):10, 2012. 25

[186] C. M. Summa and M. Levitt. Near-native structure refinement using in vacuo energy minimization. *Proceedings of the National Academy of Sciences*, 104(9):3177–3182, feb 2007. 47

[187] L. Sutto, S. Marsili, A. Valencia, and F. L. Gervasio. From residue coevolution to protein conformational ensembles and functional dynamics. *Proceedings of the National Academy of Sciences*, 112(44):13567–13572, oct 2015. 39

[188] A. Taly, P.-J. Corringer, D. Guedin, P. Lestage, and J.-P. Changeux. Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system. *Nature Reviews Drug Discovery*, 8(9):733–750, sep 2009. 40

[189] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12):2725–2729, oct 2013. 23

[190] M. Tanaka, T. R. Cummins, K. Ishikawa, S. D. Dib-Hajj, J. A. Black, and S. G. Waxman. SNS Na+ channel expression increases in dorsal root ganglion neurons in the carrageenan inflammatory pain model. *NeuroReport*, 9(6):967–972, apr 1998. 59

[191] X. Tao, J. L. Avalos, J. Chen, and R. MacKinnon. Crystal structure of the eukaryotic strong inward-rectifier K+ channel Kir2.2 at 3.1 Å resolution. *Science*, 326(5960):1668–1674, dec 2009. 8

[192] Á. Tarcsay, G. Paragi, M. Vass, B. Jójárt, F. Bogár, and G. M. Keserű. The impact of molecular dynamics sampling on the performance of virtual screening against GPCRs. *Journal of Chemical Information and Modeling*, 53(11):2990–2999, nov 2013. PMID: 24116387. 64

[193] D. B. Tikhonov and B. S. Zhorov. Architecture and pore block of eukaryotic voltage-gated sodium channels in view of NavAb bacterial sodium channel structure. *Molecular Pharmacology*, 82(1):97–104, apr 2012. 61

[194] T. Tomašić, B. Hartzoulakis, N. Zidar, F. Chan, R. W. Kirby, D. J. Madge, S. Peigneur, J. Tytgat, and D. Kikelj. Ligand- and structure-based virtual screening for clathrodin-derived human voltage-gated sodium channel modulators. *Journal of Chemical Information and Modeling*, 53(12):3223–3232, dec 2013. PMID: 24215100. 61

[195] O. Trott and A. J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31(2):455–461, 2009. 34

[196] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski, and S. Filipek. Action of molecular switches in GPCRs - theoretical and experimental studies. *CMC*, 19(8):1090–1109, mar 2012. 2

[197] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, and A. Fujii. Magnitude and directionality of the interaction energy of the aliphatic CH/π interaction: significant difference from hydrogen bond. *J. Phys. Chem. A*, 110(33):10163–10168, aug 2006. 50

[198] N. Unwin. Neurotransmitter action: Opening of ligand-gated ion channels. *Cell*, 72:31–41, jan 1993. 14

[199] N. Unwin. Refined structure of the nicotinic acetylcholine receptor at 4Å resolution. *Journal of Molecular Biology*, 346(4):967 – 989, mar 2005. 41

[200] S. Vanni, M. Neri, I. Tavernelli, and U. Rothlisberger. Observation of "ionic lock" formation in molecular dynamics simulations of wild-type β 1 and β

2 Adrenergic receptors. *Biochemistry*, 48(22):4789–4797, jun 2009. PMID: 19378975. 50, 54

[201] V. Vyklicky, T. Smejkalova, B. Krausova, A. Balik, M. Korinek, J. Borovska, M. Horak, M. Chvojkova, L. Kleteckova, K. Vales, J. Cerny, M. Nekardova, H. Chodounska, E. Kudova, and L. Vyklicky. Preferential inhibition of tonically over phasically activated NMDA receptors by pregnane derivatives. *Journal of Neuroscience*, 36(7):2161–2175, feb 2016. 36

[202] J. Wang, P. Cieplak, and P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000. 49

[203] J. Wang and J.-A. Feng. Exploring the sequence patterns in the α-helices of proteins. *Protein Engineering Design and Selection*, 16(11):799–807, nov 2003. 59

[204] Q. Wang, R. H. Mach, R. R. Luedtke, and D. E. Reichert. Subtype selectivity of dopamine receptor ligands: Insights from structure and ligand-based methods. *Journal of Chemical Information and Modeling*, 50(11):1970–1985, nov 2010. 74

[205] Z. Wang and J. Xu. Predicting protein contact map using evolutionary and physical constraints by integer programming. *Bioinformatics*, 29(13):i266–i273, jun 2013. 39, 70

[206] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa. Identification of direct residue contacts in protein-protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72, dec 2008. 39

[207] G. A. Weiss, C. K. Watanabe, A. Zhong, A. Goddard, and S. S. Sidhu. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proceedings of the National Academy of Sciences*, 97(16):8950–8954, jul 2000. 41, 48

[208] S. H. White and G. von Heijne. The machinery of membrane protein assembly. *Current Opinion in Structural Biology*, 14(4):397–404, aug 2004. 40

[209] P. Willett. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11(23-24):1046–1053, dec 2006. 75

[210] M. J. Wilson, D. Yoshikami, L. Azam, J. Gajewiak, B. M. Olivera, G. Bulaj, and M.-M. Zhang. μ-conotoxins that differentially block sodium channels Nav1.1 through 1.8 identify those responsible for action potentials in sciatic nerve. *Proceedings of the National Academy of Sciences*, 108(25):10302–10307, jun 2011. 60

[211] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS Journal*, 280(22):5705–5736, sep 2013. 10

[212] J. Wu, Z. Yan, Z. Li, C. Yan, S. Lu, M. Dong, and N. Yan. Structure of the voltage-gated calcium channel Cav1.1 complex. *Science*, 350(6267):aad2395–aad2395, dec 2015. 10, 60

[213] X. Wu and B. R. Brooks. Self-guided langevin dynamics simulation method. *Chemical Physics Letters*, 381(3-4):512–518, nov 2003. 31, 48, 64

[214] Z. Xiang. Advances in homology protein structure modeling. *Current Protein & Peptide Science*, 7(3):217–227, jun 2006. 24

[215] V. Yarov-Yarovoy, P. G. DeCaen, R. E. Westenbroek, C.-Y. Pan, T. Scheuer, D. Baker, and W. A. Catterall. Structural basis for gating charge movement in the voltage sensor of a sodium channel. *Proceedings of the National Academy of Sciences*, 109(2):E93–E102, dec 2011. 61

[216] J. D. Yesselman, D. J. Price, J. L. Knight, and C. L. Brooks. MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.*, 33(2):189–202, nov 2011. 31

[217] T. J. Ypma. Historical development of the Newton–Raphson method. *SIAM Review*, 37(4):531–551, dec 1995. 30, 63

[218] F. H. Yu and W. A. Catterall. The VGL-chanome: A protein superfamily specialized for electrical signaling and ionic homeostasis. *Science Signaling*, 2004(253):re15–re15, sep 2004. 6, 8

[219] F. H. Yu, V. Yarov-Yarovoy, G. A. Gutman, and W. A. Catterall. Overview of molecular relationships in the voltage-gated ion channel superfamily. *Pharmacological Reviews*, 57(4):387–395, dec 2005. 6

[220] Y. Zhang and X.-Q. Xie. Biosynthesis, purification, and characterization of a cannabinoid receptor 2 fragment (CB2$_{271-326}$). *Protein Expression and Purification*, 59(2):249–257, jun 2008. 42

[221] A. Zhou, A. Rohou, D. G. Schep, J. V. Bason, M. G. Montgomery, J. E. Walker, N. Grigorieff, and J. L. Rubinstein. Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *eLife*, 4:e10180, aug 2015. 10

List of Publications

Journal Articles

Humpert G, Hoffmann D, Kless A. Protein interfaces and assembly of the heteromeric $\mathrm{Na}_v1.8$ ion channel. *in preparation*.

Zimmer E, Dopychai A, Kless A, Humpert G, Schmalzing G. Leucine Zipper and Aromatic Network are Key Determinants in the Assembly of Cys Loop Receptor Types 5-HT3AR and GABAAR. *in preparation*.

Presentation

Humpert G. Protein interfaces and assembly of the heteromeric $\mathrm{Na}_v1.8$ ion channel. *Molecular Modeling Workshop 2015*, 3rd in Lecture Award, Erlangen, Germany, 2015.

Poster

Humpert G, Hoffmann D, Kless A. Heteromeric assembly of voltage-gated ion channels. *ICCS 2014*, Noordwikerhoud, Netherlands, 2014.

Acknowledgements

> «Ich sage nur ein Wort: Vielen Dank!»
> «I'll just say one word: Many Thanks!»
>
> – Andreas Brehme, Football world champion 1990

My colleagues at Grunenthal Michael Engels, Henning Steinhagen and all colleagues from HR who offered me the chance to educate my scientific skills further and enabled the studies for this thesis alongside my regular work.
My managers Stefan Oberbörsch and Sven Kühnert who appreciated my extra work and gently pushed me towards the finish line.
Tsega Melake and Elvira Dahlke for being great colleagues, who absorbed the workload peaks in my spare time as if it was natural without ever complaining.

My supervisors and mentors Daniel Hoffmann who did not hesitate to support my scientific adventure, even though a return for his steady and patient work was not guaranteed.
Achim Kless who mimicked my academic working group, provided help and advice, and was a good example in being a scientist.

My family and friends Who always supported me by respecting my absence at multiple occasions, never stopped to encourage me in so many ways and always believed in me. Nikolaj Dybowski and Holger Franken for being interested in my doctoral project and finally proofreading this thesis.

My wife and daughter For tolerating the countless hours I spend in my basement office. Thank you Claudia for taking so much work, thoughts and trouble off my shoulders, rarely complaining about too little family time and your firm confidence in me being successful. Anna for your basement visits, all the joy you spread, and for just being there to remember me of what is important and will from now on get all the time it deserves!

Curriculum Vitae

For reasons of confidentiality, the curriculum vitæ is not included in the online version of this work.

Declarations

Erklärung:
Hiermit erkläre ich, gem. § 6 Abs. (2) f) der Promotionsordnung der Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung der Dr. rer. nat., dass ich das Arbeitsgebiet, dem das Thema "Investigation of the domain assembly in ligand-gated and voltage-gated ion channels using computational methods" zuzuordnen ist, in Forschung und Lehre vertrete und den Antrag von Guido Humpert befürworte und die Betreuung auch im Falle eines Weggangs, wenn nicht wichtige Gründe dem entgegenstehen, weiterführen werde.

Essen, den ____________ ______________________________________

Unterschrift eines Mitgliedes der Universität Duisburg-Essen

Erklärung:
Hiermit erkläre ich, gem. § 7 Abs. (2) c) + e) der Promotionsordnung Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung des Dr. rer. nat., dass ich die vorliegende Dissertation selbständig verfasst und mich keiner anderen als der angegebenen Hilfsmittel bedient habe.

Essen, den ____________ ______________________________________

Unterschrift des Doktoranden

Erklärung:
Hiermit erkläre ich, gem. § 7 Abs. (2) d) + f) der Promotionsordnung der Fakultäten für Biologie und Geografie, Chemie und Mathematik zur Erlangung des Dr. rer. nat., dass ich keine anderen Promotionen bzw. Promotionsversuche in der Vergangenheit durchgeführt habe und dass diese Arbeit von keiner anderen Fakultät/Fachbereich abgelehnt worden ist.

Essen, den ____________ ______________________________________

Unterschrift des Doktoranden