



MASTERARBEIT

Titel der Masterarbeit

„Influence of aromatic side chain assignment on the stability
of ion channel homology models“

Verfasserin ODER Verfasser

Eva-Maria Zangerl

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2012

Studienkennzahl lt. Studienblatt:

A 066 862

Studienrichtung lt. Studienblatt:

Chemie

Betreuerin / Betreuer:

ao. Univ.-Prof. i.R. Dr. Karl Peter Wolschann

Contents

I. Acknowledgment.....	1
II. Abstract.....	2
III. Zusammenfassung.....	4
1 Introduction.....	6
1.1 Ion channels.....	6
1.2 Ion channels: crystal structures.....	7
1.3 Structure of Ion channels.....	9
1.4 Mechanism of Ion channel channel gating.....	11
1.5 Mechanism of Ion selectivity.....	11
1.5.1 Ion conductance.....	11
1.5.2 Gating.....	12
1.6 Channelopathies.....	14
2 Computational structure prediction.....	15
2.1 Side chain predictions.....	15
2.2 Homology Modeling.....	16
2.2.1 Mutation.....	17
2.2.2 Alignment and Model building.....	17
2.2.3 Minimization, Equilibration and Simulation.....	18
2.2.4 Model Evaluation.....	20
3 Molecular dynamics simulations.....	21
3.1 Force fields.....	21
3.2 Simulation details	22
3.2.1 Periodic boundaries.....	22
3.2.2 Thermostat.....	22
3.2.3 Barostat.....	22
3.3 Description of the used programs.....	22
3.3.1 Swiss Pdb Viewer.....	22

3.3.2 Modeller.....	23
3.3.3 VMD.....	23
3.3.4 PYMOL.....	23
3.3.5 WHATCHECK.....	24
3.3.6 PROCHECK.....	24
3.3.7 Gromacs.....	24
3.3.8 Gromacs file formats.....	24
3.3.9 Used Gromacs tools 78.....	25
3.3.10 Xmgrace/Grace.....	26
4 Methods.....	27
4.1.1 Homology modeling.....	27
4.1.2 Alignment of the sequences.....	27
4.1.3 Insertion of the model into the membrane.....	28
4.1.4 Inclusion of Ions.....	28
4.1.5 Minimization and Equilibration.....	29
4.1.6 Molecular dynamics simulations.....	29
4.2 Evaluations.....	30
4.2.1 Inspection of MD simulations.....	30
4.2.2 RMSD.....	30
4.2.3 RMSF.....	30
4.2.4 Packing quality.....	31
5 Results.....	32
5.1 Homology modeling.....	32
5.2 Alignment of the sequences.....	33
5.3 Visual MD analysis.....	38
5.3.1 Conclusion.....	40
5.4 RMSD.....	41
5.4.1 Conclusion.....	41
5.5 RMSF.....	44
5.5.1 KcsA in the closed state.....	48

5.5.2 KcsA in the open state.....	49
5.5.3 MthK.....	49
5.5.4 Kir2.2.....	50
5.5.5 Nav.....	50
5.5.6 NaK.....	51
5.5.7 Mlotik1.....	51
5.5.8 Conclusion.....	52
5.6 Packing quality.....	53
5.6.1 Conclusion.....	53
6 Conclusion.....	55
7 List of figures.....	57
8 List of tables.....	58
9 Lebenslauf.....	59
10 Bibliography.....	60

I. Acknowledgment

Special thanks go to my supervisor Univ. Prof. Dr. Karl Peter Wolschann and Mag. Dr. Anna Weinzinger for the idea to this thesis, her support, incitations and patience to share her enormous knowledge about ion channels and their functions with me.

I want to thank my colleges Mag. Tobias Linder and Song Ke, MSc for their help with diverse problems and for a great working atmosphere.

My biggest thanks goes to my family for their support and encouragement, especially to my brother Peter for his creative help with diverse problems. I want to thank my partner Christof for his patience and supportive nature. Special thanks also go to my friends for their support all over my study.

II. Abstract

Ion channels are proteins that exist in every biological organism. They serve as a “communication tool” between the intra- and the extracellular side of cells. Ion channels coordinate electrical signals in most tissues and are thus involved in every heartbeat, every movement, and every thought and perception. All these functions require rapid and accurate transmission of information among the cells and a fast coordination of different and remote functions.¹ Dysfunctions of ion channels can lead to severe diseases – the so called channelopathies.²

Up to now, only a limited number of K^+ and Na^+ channels have been crystallized.³ Therefore, homology modeling techniques are widely applied to aid in studying the structure and function of ion channels.

One particular challenge in modeling ion channels is the correct assignment of side chain conformations. This might be especially tricky, when the content of aromatic side chains is high. This is the case in most ion channels. For example the bacterial potassium channel KcsA contains 11% aromatic side chains (tryptophan, tyrosine, phenylalanine) and the bacterial sodium channel NaV contains 19% aromatic amino-acids. This unusually high content of aromatic residues complicates modeling, since incorrect side chain assignment might influence the stability of the generated homology models.

To test this thesis we performed modeling on seven ion channels using the Modeller software⁴, a widely recommended program used for homology modeling of three-dimensional structures. The task was to see how useful Modeller generated side chain conformations are.

Therefore, we let Modeller generate 20 homology models per ion channel based on a target – template alignment of an alanine-mutant form (all aromatic residues were mutated to alanines, performed with Pdb-Swiss Viewer⁵) of the crystal structures and the crystal structure itself. The three to four homology models with the lowest root-mean-square-deviation (RMSD, this was calculated between the crystal structure and each generated homology model), the crystal structure and the alanine-mutant were used to perform molecular dynamics (MD) simulations. For each of them we performed five 20ns MD simulations. As we used 7 ion channels we performed 180 MD simulations.

We evaluated the data by assessing the stability of the generated models, checking the root-mean-

II. Abstract

square-deviation, root-mean-square-fluctuation (RMSF) and the packing quality. Therefore we always compared the crystal structure, alanine-mutant and the Modeller generated homology model molecular dynamics simulations.

The alanine-mutant was used to gain information about the importance of the aromatic amino acids to the stability of the protein. We could demonstrate that the aromatic amino acids play an important role in filter stability, but their effect on the stability of other helical segments is negligible.

To our surprise Modeller did not only assign side chain conformations to mutated residues, but it also changed the conformation of all the other side chains even when they are identical to the crystal structure template. This is problematic when only view amino acids in the target structure are changed since the crystal structure side chain conformations usually show a good packing quality, low RMSD and RMSF and a good behavior in molecular dynamics simulations.

The best results were achieved by using the crystal structures. Modeller generated homology models show the highest RMSD and RMSF and the lowest packing quality. We showed that Modeller seems to be non-recommendable when it is used for molecular dynamics simulations without previous side chain refinement.

III. Zusammenfassung

Ionenkanäle sind Membranproteine die in jedem biologischen Organismus vorhanden sind. Sie dienen als Kommunikationsmittel zwischen dem extra- und dem intrazellulären Raum. Ionenkanäle koordinieren elektrische Signale und sind damit beteiligt an jedem Herzschlag, jeder Bewegung und auch an Gedanken und Wahrnehmung. Diese Funktionen benötigen eine schnelle und genaue Weiterleitung von Informationen zwischen Zellen, sowie eine schnelle Koordination von unterschiedlichen und unter Umständen auch entfernten Funktionen.¹ Krankheiten, die auf Fehlfunktionen von Ionenkanälen zurückzuführen sind, werden als Channelopathien bezeichnet.² Bis heute gibt es nur eine begrenzte Anzahl an kristallisierten Ionenkanälen.³ Deshalb werden Methoden zum computerunterstützten Generieren von Modellen (sogenannte Homology Models) herangezogen, um mit deren Hilfe die Funktionen und Struktur von Ionenkanälen zu untersuchen. Eine der Schwierigkeiten bei der Modellierung von Ionenkanälen besteht darin, dass Ionenkanäle einen hohen Anteil an aromatischen Aminosäuren (Phenylalanin, Tyrosin und Tryptophan) besitzen. So beträgt dieser Anteil zum Beispiel für den Kalium Bakterienkanal KcsA 11% und für den Natrium Bakterienkanal NaV 19%. Dieser hohe Prozentsatz erschwert das Modellieren, da speziell bei der Generierung von Homology Models das Erstellen von Seitenkettenkonformationen wichtig ist. Grund dafür ist deren Einfluss auf die Stabilität des Proteins.

Um Homology Models zu generieren, verwendeten wir Modeller⁴, ein oft zitiertes Programm zum Modellieren von drei-dimensionalen Strukturen. Die Aufgabenstellung bestand darin zu überprüfen, wie gut die vom Modeller generierten Seitenkettenkonformationen sind.

Um dieses zu testen, haben wir mit Modeller Homology Models basierend auf einem Target – Template Abgleich generiert. Dafür mussten wir eine Mutation erstellen. Diese wurde mit Hilfe des Pdb-Swiss Viewers⁵ generiert – alle aromatischen Aminosäuren wurden durch Alanin ersetzt (dies bildet das sogenannte Template). Dieses Template wurde dann mit der Kristallstruktur (dem sogenannten Target) abgeglichen. Wir führten sowohl mit den Kristallstrukturen und den Alanin-mutierten Modellen als auch mit drei bis vier aus zu Anfang 20 generierten Homology Models je fünf Molekulardynamicsimulationen (MD Simulationen) von je 20ns durch. Da wir insgesamt 7 Ionenkanäle behandelten, belief sich die Anzahl der MD Simulationen auf 180.

Als Auswertungskriterien analysierten wir die Stabilität der Proteine, berechneten die root-

III. Zusammenfassung

mean-square-deviation (RMSD), die root-mean-square-fluctuation (RMSF) und die Qualität der Packung.

Die Alanin-Mutation wurde in erster Linie dazu verwendet, Informationen über die Aufgabe der aromatischen Aminosäuren und deren Einfluss auf die Stabilität des Proteins zu gewinnen. Dabei konnten wir zeigen, dass die aromatischen Aminosäuren eine wichtige Rolle für die Stabilität der Filter-Region spielen, ihr Einfluss auf die Stabilität der helikalen Segmente jedoch nur sehr gering ist.

Anders als von uns erwartet, generiert Modeller nicht nur die Seitenkettenkonformationen an den mutierten Aminosäuren, sondern ändert auch die Seitenkettenkonformationen aller anderen Aminosäuren, selbst wenn diese identisch mit dem Kristallstruktur Template sind. Das ist speziell in Fällen in denen nur ein paar Aminosäuren der Target - Struktur geändert werden sollten problematisch, da die Kristallstrukturkonformationen meist eine gute Qualität der Packung, niedrigen RMSD und RMSF sowie eine hohe Stabilität in MD Simulationen aufweisen.

Allgemein konnten die besten Ergebnisse mit der Kristallstruktur erzielt werden. Die von Modeller generierten Homology Models hingegen weisen den höchsten RMSD und RMSF sowie die schlechteste Qualität der Packung auf.

In dieser Arbeit konnte gezeigt werden, dass es nicht empfehlenswert ist, ein von Modeller generiertes Homology Model für molekular dynamische Simulationen zu verwenden ohne eine vorhergehende Remodellierung der Seitenketten vorzunehmen.

1 Introduction

1.1 Ion channels

Ion channels are membrane proteins that allow ions to cross the lipid bilayer membrane and move into the cells. They are ubiquitous in every biological cell. Their functions are the transmission of information through every biological organism, regulation of the concentration of electrolytes and they are also responsible for multiple biological functions including the propagation of nerve impulses and the contraction of muscles.^{6,7} Investigations in understanding the mechanisms of ion channels are made especially because of their pharmaceutical importance, as a huge number of diseases – so called channelopathies - are caused by mutations of genes expressing ion channels and result in different permeability functions of the channels.¹

Our basic understanding of membrane proteins is the result of the framework of Hodgkin and Huxley in 1952.⁸ They first developed voltage clamps to measure ion movements in a cell as an electrical current. Hodgkin and Huxley inferred that this membrane current could be ascribed to Na⁺ and K⁺ permeability mechanisms of the cells.⁷ With this technique they could first develop the three main key features that help characterizing the sodium channels:

1. voltage dependent activation,
2. rapid inactivation,
3. selective ion conductance.⁹

They also stated some important features for potassium movement into the cells:

1. the activating molecules have an affinity to potassium and not sodium,
2. they move more slowly,
3. they are not blocked or inactivated.⁸

Hodgkin and Huxley demonstrated that the ability of the membranes to selectively regulate the passage of ions is provided by the electrical signaling of nerve cells.¹⁰

Nowadays much more information about ion channels is available. It is clear, that ion channels have two main functions: selective ion conductance and the ability to gate ions in response to an appropriate stimulus.⁹

1 Introduction

Ion channels differ in the way of their gating mechanism. They are pH gated, voltage-gated, mechanosensitive, ligand-gated or activated by an intracellular ligand. There are 8 different types of voltage-gated ion channels that are assumed to have a similar pore structure: voltage-gate sodium, potassium and calcium channels, calcium activated potassium channels, cyclic nucleotide-modulated ion channels, transient receptor potential channels, inwardly rectifying potassium channels and two-pore potassium channels.¹¹

One of the most important features of ion channels is that they show a high selectivity in permeability. As an example, all K^+ channels show a selectivity sequence of $K^+ \approx Rb^+ > Cs^+$. The permeability for Na^+ and Li^+ is immeasurably low in these channels. The movement of Na^+ ions through K^+ channels is at least 10,000 times less than the movement of K^+ , which is an essential feature for the function of potassium channels. This seems to imply strong energetic interactions between K^+ ions and the pore of the channels.¹²

1.2 Ion channels: crystal structures

Crystallizing ion channels is a very popular topic. The first potassium ion channel crystallized in 1998 by Doyle et al.¹² was the bacterial potassium channel KscA from *Streptomyces lividans*. Until now KcsA is the best investigated ion channel. Nowadays 13 potassium and sodium channel unique (= include channels of the same type from different species) structures of Na^+ , K^+ and Ca^{2+} channels exist.³

In this thesis we used the ion channels (with different resolutions) shown in Table 1.

As a matter of fact most channels have been crystallized in one state (open or closed) only. The only channel where we investigated two (open and closed) conformations was KscA. For all the other channels analyzed, only one gating conformation was available.

1 Introduction

Name	Conformational state	Pdb-code	Resolution [Å]	activation	Permeable ions	cells
KcsA ¹³	closed	1K4C	2.00	pH-gradient	K ⁺	<i>Streptomyces lividans</i>
KcsA	open	3F7V	3.20	pH-gradient	K ⁺	<i>Streptomyces lividans</i>
MthK ¹⁴	open	3LDC	1.45	Calcium-activated	K ⁺	<i>Methanobacterium thermoautotrophicum</i>
Kir2.2 ¹⁵	closed	3JYC	3.11	Inward rectified	K ⁺	<i>Gallus gallus</i>
NaK ¹⁶	closed	2AHY	2.40	cyclic-nucleotide	Na ⁺ and K ⁺ (non-selective)	<i>Bacillus cereus</i>
Nav ¹⁷	closed	3RVY	2.70	voltage-gate	Na ⁺	<i>Arcobacter butzleri</i>
Mlotik1 ¹⁸	closed	3BEH	3.10	non voltage-gate	K ⁺	<i>Mesorhizobium loti</i>

Table 1: Information of the used crystal structures

The sequence identities of some ion channels are listed in Table 2.¹⁹

Table 2 shows that sequence identities of this ion channels are quite low, but as we see in Figure 1 their structures are quite similar. This will be discussed later on (see 2 Computational structure prediction).

	%ID	RMSD
KcsA-MlotiK	39	2,5
KcsA-NaK	25	2,5
KirBac-MlotiK	24	2,0
NaK-MlotiK	23	3,3
KcsA-KirBac	15	3,3
KirBac-NaK	15	2,8

Table 2: Sequence identity (ID in percent) and RMSD of chosen ion channels¹⁹

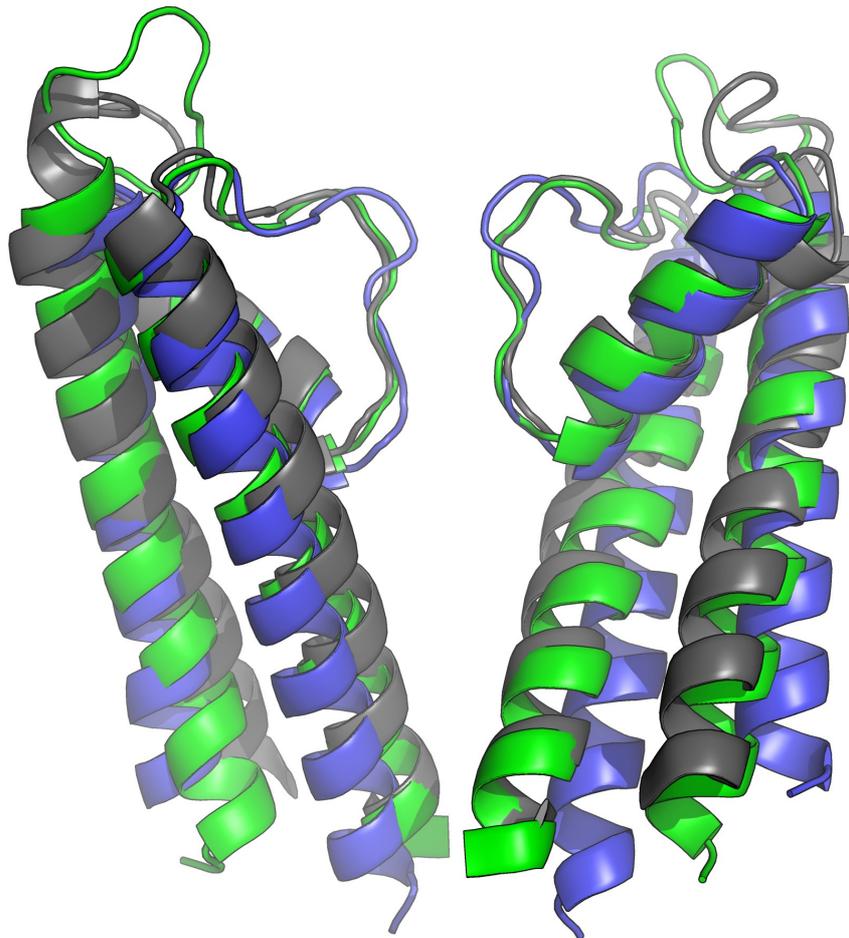


Figure 1: Structure superimposition of three ion channels listed in Table 2 (KcsA (green), MlotiK (gray), and NaK (blue))

1.3 Structure of Ion channels

In order to provide different ion selectivity and conduction properties ion channels have slightly different structures (i.e. in the filter-region). However, the global architecture (as will be described below) is similar.

Doyle et al. 1998 wrote about an “inverted teepee” architecture of the four inner helices and that the pore helices are slotted in between the poles of this teepee, when they established the first X-ray structure of the potassium channel KcsA with a resolution of 3.2 Ångstroms.¹²

The architectures of the ion channel families consist of four variations built on a common pore-forming structural theme. The voltage-gated sodium channel's principal α -subunits are composed

1 Introduction

of four homologous domains (I to IV) that form a common motif. Each domain consists of six regions that form α -helices (termed segments S1 to S6) and a membrane loop between the S5 and S6 segments.¹¹ The helices embedded in the membrane (S5, S6 and the P-helix) build the pore section (see Figure 2).¹⁷

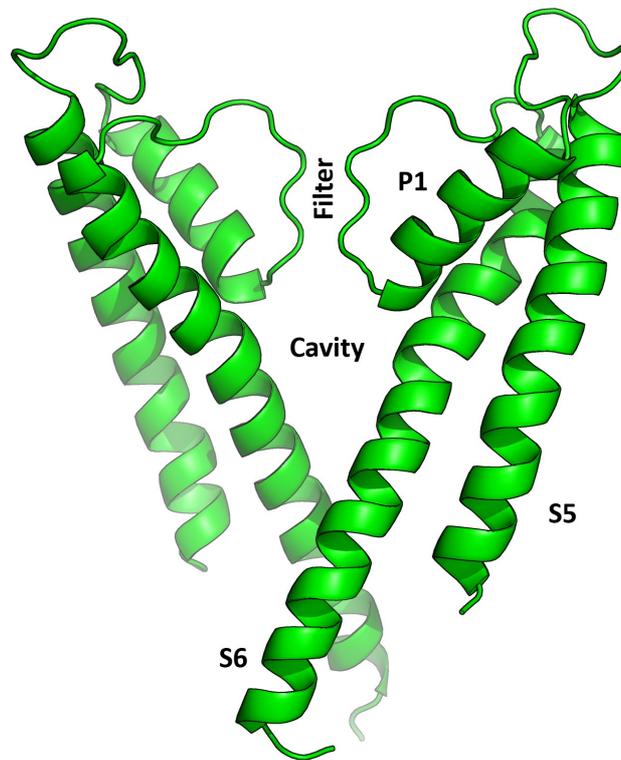


Figure 2: This figure shows two opposing pore-forming segments of the closed KcsA. The helices and the filter are labeled.

The voltage-sensing segments S1 to S4 form a truss that interacts with the pore-forming S5 - P - S6 segments from the neighboring subunit. These two separate structural elements within each domain are connected by the long S4-S5 linker. The linker is placed along the intracellular surface of the channel.²⁰

At the extracellular end of the channel there is the selectivity filter. Its structure is highly conserved among channels conducting similar ions. Potassium channels have the highly conserved TVGY(F)G sequence motif in the filter domain. Between the gate and the filter is the cavity that is filled with water molecules. The number of solvent molecules might vary as ions pass during the channel opening.²¹

1 Introduction

1.4 Mechanism of Ion channel channel gating

The mechanisms of the gating can differ, as the ion channels have different structures. There are three main categories of gating mechanisms: voltage-gating, pH-dependent gating and ligand-binding.²¹ Ion channels can work very fast. As an example, K⁺ channels have a very high throughput of 10⁸ ions per second, which nearly approaches the diffusion limit.¹²

The driving force for the conformational change of the conduction pore between open and closed states depends on the type of the ion channel:

- pH-gradient: e.g. KcsA
- voltage: it converts stored energy from the membrane electric field into mechanical work that opens the pore; e.g. Kv1.2, NavAB;
- ligand-binding: chemomechanical or electromechanical coupling between the pore unit and the gating unit leads to the opening;⁹ e.g. Kir2.2;

To get a clear idea of how the opening mechanism works, high resolution crystal structures of ion channels in different channel-states are needed. After the high resolution X-ray structure of the open state of MthK was established, groups tried to construct an open KcsA state based on the MthK structure to get a better idea of the gating mechanism itself.²²

1.5 Mechanism of Ion selectivity

1.5.1 Ion conductance

Ion conductance stands for the rate of ions passing through a channel. Ions that move fast through the channel have a high conductance, slow passing means low conductance. Not every ion can pass every filter, as the filter-regions of the channels are very selective. To enter the filter the ion must partially dehydrate. The carbonyl oxygen atoms of the filter partially replace the water oxygen atoms. The transmembrane helices form the cavity and the pore. In the cavity the ions are rehydrated with water to overcome the electrostatic destabilization of the low dielectric bilayer.¹² All this is known for potassium selective channels so far. The conductance mechanism of other channels still remains unknown.

1 Introduction

1.5.2 Gating

The helices can move away from the center of the channel to enable channel opening or gating.²³ The most flexible part are the S6 helices – this are the helices that move most during the pore opening. To let ions pass through an opening of 2-3 Å is required (see Figure 3).²¹ In voltage gated ion channels, pore opening is initiated by transmembrane potentials charges. This leads to outward movement of positive charges on the S4 transmembrane segments. This is catalyzed by formation of ion pair building of the S1 to S4 helices with the S5 helix. The exact mechanism of the opening is not known until today, especially because of the lack of a crystal structure of the voltage-sensor segments in the open state. The movement bends the S6 segment and opens the pore.^{24,25}

This structural motif is the building block of the voltage-gated calcium and potassium channels as well as a large family of related ion channels, including the cyclic nucleotide-gated channels and calcium-activated potassium channels.²⁶

KcsA channels which lack the voltage sensing domains (S1-S4) gate with a drop in pH.²¹ High concentration of the intracellular protons allow the channel to open at pH < 5.5. The pH sensor is suggested to consist of a network of ionizable amino-acids at the cytoplasmic end of transmembrane helices TM1 (S5) and TM2 (S6) which form complex inter- and intrasubunit interactions.²⁷

Besides opening and closing of ion channels inactivation mechanisms prevent ion flow:

Inactivation can occur either:

- from the open state at strong depolarized membrane potentials
- or from pre-open closed states at hyperpolarized or depolarized membrane potentials.

Voltage-gated Na⁺, K⁺, Ca²⁺ channels utilize both inactivation mechanisms. In an inactivated state ion-flow is prevented by certain channel conformations, involving either the filter region or might include plugging of the intracellular gate by a linker residue. Recovery of the inactivated state only occurs with returning of the membrane potential to its initial state.²⁸ The exact structural mechanisms underlying inactivation gating are still unknown.

All parts of the channel need to work precisely together in order to obtain a functional channel.

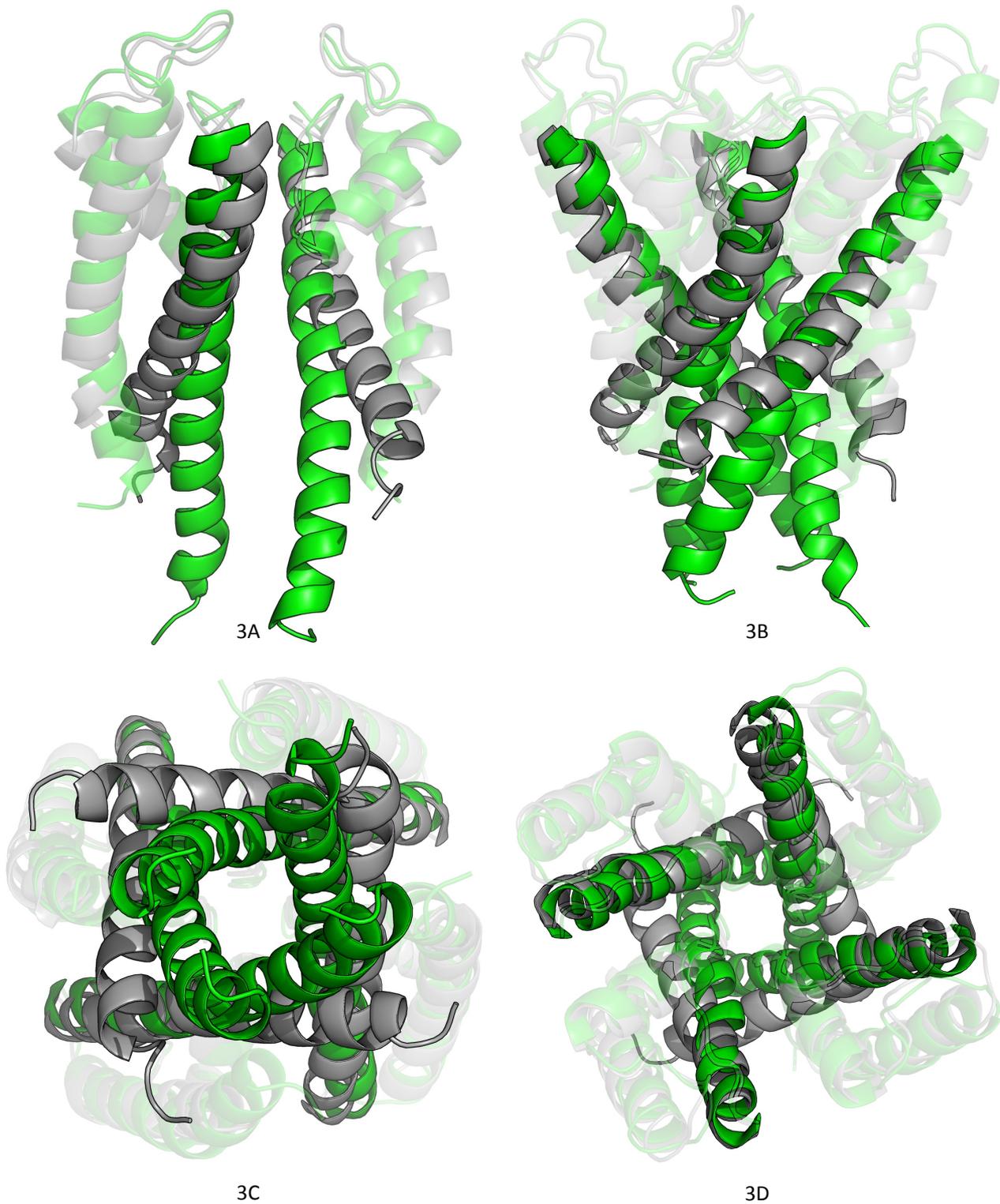


Figure 3: Alignments of KcsA in the closed and the open conformation. Green colored is the KcsA in the closed conformation, gray colored is the channel in the open conformation. 3A shows 2 opposing domains in the side view and 3B show the whole channels in the side view. 3C shows the bottom view and 3D the top-view of the aligned channels.

1.6 Channelopathies

Channelopathies are diseases caused by mutations of genes expressing K^+ , Na^+ , Ca^{2+} and Cl^- ion channel subunits. Studies of such mutations have given important insights into molecular mechanisms and help to understand the mechanisms of channelopathies and to find new pharmaceutical targets.²

The first discovered sodium channelopathies were the paramyotonia congenita and the familial periodic paralysis syndromes hyperkalemic periodic paralysis. These diseases are caused by dominant skeletal muscle sodium channel mutations.²⁶

As an example, a mutation in SCN9A, the gene encoding the $Na_v1.7$ sodium channel, causes nonactivating sodium currents in $Na_v1.7$ channels that lead to paroxysmal extreme pain disorder. It induces repetitive firing in peripheral nerves that contain pain information.²⁹ Mutations of this SCN9A region can also lead to a congenital inability to experience pain – a very rare phenotype. The resulting mutations cause loss of function of $Na_v1.7$. Affected individuals often suffer permanent injuries during their childhood because they do not develop a pain-avoiding behavior.³⁰ This role of $Na_v1.7$ in the experience of pain might exhibit these channels as a drug target for treatment of chronic pain.^{29,30}

Another disease is caused by a loss-of-function mutation of the $Na_v1.1$ and causes myoclonic epilepsy in babies.² There are many other examples of ion channel mutations causing diseases, like coexisting epilepsy and paroxysmal dyskinesia induced by a mutation of the conductance calcium-activated potassium channel³¹, hearing loss caused by K_v7 mutations³², non-insulin dependent diabetes mellitus³³, night blindness caused by a mutation of a Ca_v ³⁴ channel or light-threatening arrhythmia (like ventricular fibrillation) that may occur in response to cardiac and non-cardiac drugs³⁵. This shows that there is great interest in modifying the gating behavior of ion channels with pharmaceuticals to treat diseases. Since crystallizing these membrane bound proteins is still rather challenging homology modeling is an important tool to obtain information about the effect of different mutations on the function of ion channels.

2 Computational structure prediction

Homology modeling is important because - even though some crystal structures are available - many structures still remain unknown. As previously shown, ion channels on one hand differ a lot in their sequence identity (see Table 2), but on the other hand their structure is quite similar (see Figure 1). This gives hope that even without crystallographic structures reasonably accurate models can be built by homology modeling. The global fold of ion channels is similar, allowing for the generation of reasonable models. Nevertheless, since the side chains differ a lot and packing might not be conserved when the sequence identity drops to 30% or less, accurate side chain predictions may influence the quality of the generated models.³⁶

2.1 Side chain predictions

Protein folding is an important and still unsolved problem. It regards the sequence of the structure and the molecular function.^{37,38} Basically, given the structure of a wild type protein and a point mutation, an accurate method should be able to restore the structural changes.³⁹ We decided to use Modeller, as it is an often used program for homology modeling.

Side chain conformation studies were used ever since the first protein crystal structures were available. As this number increased, the most common side chain conformations could be identified by statistical analysis. This data improved the protein structure prediction.⁴⁰

The idea to explore every single side chain conformation for each side chain of a protein would simply be too CPU-intensive.⁴¹ Thus collections of the most common conformational data were established, so called rotamer libraries. They contain information about the conformation and the frequency of a certain conformation. One rotamer (stands for “rotational isomer”) contains information of a single side chain angle, its dihedral angle degree of freedom. It is most commonly a conformation in a local minimum.⁴⁰ The precision of side chain conformation predictions depends on the quality of the used rotamer library and the used force-field, as this includes the proper parametrization of interactions between the atoms.^{42,43} Which rotamers will be used for a protein sequence with given backbone coordinates depends on the defined energy function.⁴⁴ Different groups developed methods to predict side chain conformations.^{37,41,42,44-50}

2 Computational structure prediction

In this thesis we used the Modeller software to create homology models. No side chain refinement was performed, since the aim of this thesis is to evaluate how accurate the side chain predictions of the Modeller software are. This question is rather important, since many non-specialist groups use homology modeling programs such as Modeller as “black box” tools without performing complex and time-consuming evaluations and refinements on the obtained homology models. We evaluated the usefulness of such unrefined/unchecked structure models to address questions where molecular dynamics simulations are performed.

2.2 Homology Modeling

Evolutionary related proteins share a similar structure and the structural information is better conserved than the amino acid sequence. Homology modeling methods make use of this fact.⁵¹

To make homology models of a 3-dimensional structure of a given protein (target) one usually bases them on an alignment to one or more proteins of known structure (template). Those templates should be evolutionary related to the target and therefore share a similar structure.

The prediction process consists of the following steps which are repeated until a satisfactory model is obtained:^{52,53}

1. Identification of template(s)
2. Target- template alignment
3. Model building
4. Model evaluation

In this work it was not necessary to perform steps one and two since we built models of ion channels for which the corresponding crystal structures are available. In order to test the ability of Modeller to accurately predict side chains, we mutated the aromatic side chains in selected crystal structures to alanine and used these structures to generate wild type channel homology models.

Thus our protocol consisted of the following steps (see also Figure 4):

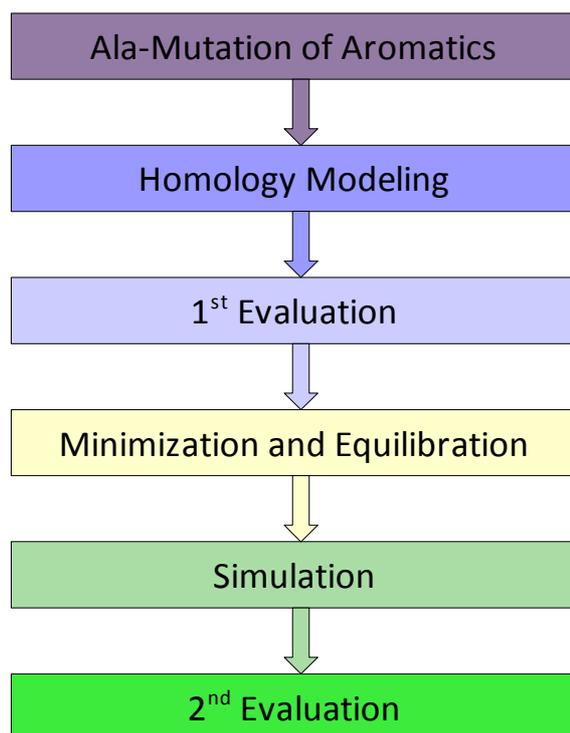


Figure 4: Overview of the performed steps

2.2.1 Mutation

All aromatic side chains (Tryptophan, Thyrosine and Phenylalanine) were mutated using the Swiss-PdbViewer Program. In our example we mutated all aromatic side chains to Alanines. This process is straight forward, since alanine has only one rotameric state.

2.2.2 Alignment and Model building

The target – template alignment was done with Modeller. To do so Modeller needs a sequence of the target (the alanine mutated form of the protein) and the sequence of the template (the corresponding crystal structure). Modeller then calculates the new models based on the alanine-mutant structures (we chose 20 as an output for each ion channel).

2 Computational structure prediction

Before performing molecular dynamics simulations with the created homology models, we evaluated them with the program WHATCHECK (for explanation of this program see chapter 3.3.5 WHATCHECK). In each case we selected three to four Modeller generated homology models of the respective ion channels. Those choice was based on their root-mean-square-deviation (RMSD) to the original crystal structures and the quality of the Ramachandran-Plots. The best models were then further investigated using molecular dynamics simulations (MD simulations). MD simulations were also performed on the original crystal structures and the constructed alanine mutants.

2.2.3 Minimization, Equilibration and Simulation

All structures (wild type, alanine-mutants and Modeller generated wild type homology models (from now on called: homology models)) were placed in a DOPC (dioleoylphosphatidylcholine) lipid bilayer. The double bilayer consists of an already equilibrated and fully hydrated system (from previous studies) for which proper force-field parameters are available. It was made sure, that the protein was placed in the center of the system (in its x- and y-position) and at the right z-position of the lipid bilayer, so that the extracellular and intracellular parts of the ion channels were extending into the solvent surrounding the lipids. After that, a Gromacs script to embed the protein into the membrane (`g_membed`; stands for embedding a protein within a membrane) was used. As a first step, this tool shrinks the protein and all the molecules that overlap are removed. After this, a molecular mechanics run of 1,000 steps is performed to grow the protein slowly to its original size. While doing so all the overlapping molecules are again removed (see Figure 5).⁵⁴

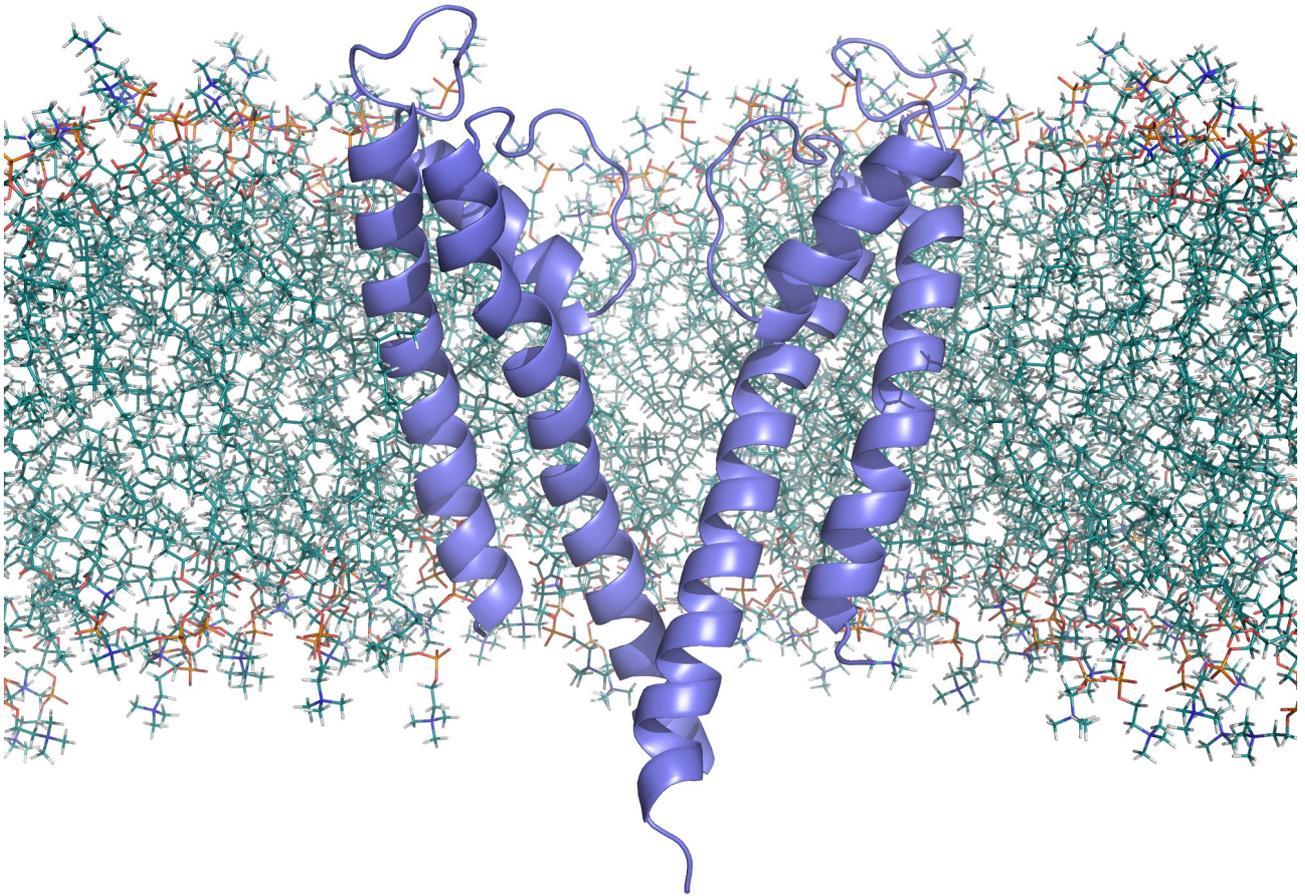


Figure 5: The figure shows two opposing domains of the KcsA channel in the closed state embedded in a lipid bilayer (DOPC).

To equilibrate the system we first set backbone restrains to hold the protein in its position. Afterwards, ions (K^+ for KscA, MthK, Kir2.2 and Mlotik or Na^+ for Nak and NavAb; and Cl^- ions as countercharge in all systems) were added to the system to maintain electroneutrality and obtain a concentration of ions of 100 mMol in the system. During the process a number of water molecules is exchanged by the same number of ions; thus, the number of water molecules can differ from simulation to simulation.

To prepare the simulations the systems had to be minimized twice and equilibrated (first in temperature (see 3.2.2), then in pressure (see 3.2.3)). Finally the simulations were started with a time-limit of 20 ns.

2 Computational structure prediction

2.2.4 Model Evaluation

Assessing the quality of models is difficult. In this thesis, after the molecular dynamics simulations, most of the evaluations we applied were performed with Gromacs tools (e. g. calculation of the root mean square deviation (RMSD), root mean square fluctuation (RMSF), angles and others). But there are also other programs available, which are used to check geometric properties of models (e. g. WHATCHECK, PROCHECK see chapter 3.3).

An often used validation in molecular dynamics simulations is the RMSD, which shows if the conformation of the protein is stable during the simulation process.

(All the used programs are described in detail in section 3.3 Description of the used programs)

3 Molecular dynamics simulations

Molecular dynamics (MD) simulations rely on Newton's law of motions. It describes the time evolution of a molecular system. It was first introduced in 1957.⁵⁵ As of today molecular dynamics simulations remain to be the only method able to precisely account for the high complexity of proteins. To calculate the dynamic behavior of proteins it is important to know the forces acting on each atom. Therefore, special force fields are implied that describe bonded (e.g. covalent bond-stretching, angle-bending, improper dihedrals and proper dihedrals) and non-bonded interactions (e.g. Lennard-Jones and Coulomb; this interactions are calculated by a neighbor list (a list of non-bonded atoms within a certain distance)) and restraints (e.g. position restraints, distance restraints, angle restraints and dihedral restraints) between the atoms.⁵⁶

A limitation of molecular dynamics simulations is that the typical 10-100 ns timescale is not sufficient to explore all the conformational spaces and processes of cell biological importance.^{19,57}

3.1 Force fields

As of today, there are four main force fields that are commonly used in simulations of biological macromolecules: AMBER⁵⁸, OPLS⁵⁹, CHARMM⁶⁰ and GROMOS⁶¹. Which force field should be chosen depends on the level of details that should be included. They provide different kind of force field for every type of atom in a system, such as all-atom (= treats every atom explicitly and also includes hydrogen atoms. Each atom has many degrees of freedom, as a consequence, the calculations quickly become very time consuming.³⁷), united-atom (= treats all the atoms of methyl groups as a single interaction center) or coarse-grained (= larger molecular units are described. This can increase the speed of computation by several orders of magnitude.⁶²) force fields. Throughout the simulations we used the AMBER99SB all atom force field (supported by Gromacs⁶³) was used.

3 Molecular dynamics simulations

3.2 Simulation details

3.2.1 Periodic boundaries

As we use a small box for the simulation we need to introduce periodic boundary conditions. This means, that any atom/molecule leaving the box on one side will come into the box on the opposite side. Thereby the conditions in the box stay the same (concentration, charge,...). To assure that a molecule never sees its periodic image (otherwise the forces calculated will be useless), we need to set a solute-box distance of 1.0 nm (this is called the minimum image convention). This results in at least 2.0 nm between any two periodic images which is sufficient for any cutoff scheme that is commonly used in simulations.⁶⁴

3.2.2 Thermostat

During an MD simulations systems exchange energy with its surroundings. This exchange of energy must be implemented in the system and therefore a canonical (NVT) ensemble is introduced. This means, that the temperature is a given variable and therefore not fluctuating at all. The fluctuation parameter is the pressure. Usually a Berendsen thermostat is used for such a problem.^{65,66}

3.2.3 Barostat

By slightly modifying the box size throughout the simulation with the Berendsen barostat⁶⁵ constant pressure is guaranteed. The pressure coupling rescales the coordinates with every step. In such a isothermal-isobaric (NPT) ensemble the pressure is given and the volume fluctuates.⁶⁶

3.3 Description of the used programs

3.3.1 Swiss Pdb Viewer

As a molecular graphic program the Swiss Pdb Viewer provides a user friendly interface. One of its most powerful features is that several proteins can be analyzed at the same time. The

3 Molecular dynamics simulations

superimposition of multiple proteins enables to easily deduce structural alignments and compare their active sites. Additional analytical tools allow to measure and modify torsion angles, mutate amino acids and compute molecular surfaces just to mention a few examples.^{5,67} Another advantage of this program is that it is free of charges for academics.

3.3.2 Modeller

To construct three-dimensional protein structure-models (homology modeling) the Modeller software serves as a powerful program. The user input usually consists of an alignment of a sequence with a known structure (target) with one or more templates of known structure. Additionally, the program can search for protein sequences, construct multiple sequence alignments of protein sequences and/or structures and perform fully automatic modeling. Furthermore, optimization of various parts of protein structures including loops can be performed with Modeller.^{4,68} This software is also available free of charge for academics.

3.3.3 VMD

VMD is an acronym for Visual Molecular Dynamics. It is a molecular graphics software designed for modeling, visualization and analyzing of large biomolecular systems such as proteins. VMD acts as a graphic front end for an external MD program (NAMD) by animating and displaying a molecule undergoing simulation on a remote computer.^{69,70} This program is free of charge for academics.

3.3.4 PYMOL

PyMOL is a Python-enhanced molecular graphics open source tool. Like VMD it can be used for visualization and analyzing large systems. In this theses we used PyMOL especially to create figures.^{71,72}

3 Molecular dynamics simulations

3.3.5 WHATCHECK

What_Check is part of the WHAT IF software and comprises several tools for protein structure verification. Quality checks like side chain conformations, bond angles, torsion angles, bond length and others that concern the geometry and stereochemistry of the protein are reported in detail. For this thesis our special interest was in the Ramachandran plot (= visualization of the backbone dihedral angles of a protein) of the proteins and the packing quality. The program is available via a web server or as a standalone version.⁷³

3.3.6 PROCHECK

The stereochemical quality of the crystal structure and modeled structures was analyzed using Procheck. This program produces plots analyzing the overall and residue-by-residue geometry including main-chain bond length distribution, Ramachandran plots, side chain rotamer plots etc.^{74,75}

3.3.7 Gromacs

Gromacs stands for GRONingen Machine for Chemical Simulation and was developed at the University of Groningen, Netherlands. It is a very powerful tool for molecular dynamics simulations and it is very fast. Force fields like AMBER, OPLS, GROMOS, CHARMM and ENCAD can be used. Gromacs comes with a large selection of different utilities and analysis programs. It can be run in parallel on big computer systems to save time. Gromacs is primarily designed for big biochemical (macro)molecules like proteins or lipids.^{76,77}

3.3.8 Gromacs file formats

- .pdb = Protein-structure file containing atom number, atom type, amino acid, chain number, number of the amino acid and coordinates

3 Molecular dynamics simulations

- .gro = Protein-structure file containing the same information as a .pdb file in different order
- .itp = files included into the topology file with details about the set position restraints
- .top = topology file containing information about the used .itp files, the number of DOPC, water molecules and ions
- .mdp = configuration file for the parameters for the running MD simulations (including simulation time, electrostatics, temperature, pressure and more)
- .ndx = index files containing information about different groups (like backbone, side chains, lipids, ions,...)

3.3.9 Used Gromacs tools ⁷⁸

- grompp: creates a .tpr file, requires: .mdp, .top, .pdb / .gro
- mdrun: starts a MD simulations
- editconf: converts and manipulates structure files
- pdb2gmx: converts pdb files to topology and coordinate files
- g_membed: embeds proteins into a membrane
- genrest: generates restraints for index groups
- genion: generates ions
- make_ndx: creates an index-file
- g_confrms: computes the RMSD of two structures after fitting one structure to another
- g_rms: compares two structures by calculating the RMSD
- g_rmsf: calculates the RMSF of atomic positions
- g_chi: computes phi, chi, omega and psi dihedrals of the backbone and side chains
- g_dist: calculates the center of mass distance of two atom groups as a function of time
- g_rama: calculates the phi/psi dihedral combinations as a function of time

3 Molecular dynamics simulations

3.3.10 Xmgrace/Grace

This is a free of charge program to create two-dimensional plots of numerical data. Xm stands for the X window system (X11 – Unix graphical environment) and Motif (the widget toolkit on Unix systems), Grace stands for GRaphing, Advanced Computation and Exploration of data.^{79,80} We used this program to create RMSD-, RMSF- and Ramachandran-Plots.

4 Methods

4.1.1 Homology modeling

We only investigated the pore domain comprising S5 – P1 – S6 section of every ion channel. Each aromatic residue (phenylalanine, tyrosine and tryptophan) was then mutated to Alanine (see 2.2.1). We call this the “alanine-mutant“. Table 3 shows the number of aromatic amino acids that were mutated and their total percentage in the different ion channels. In Figure 7 and Figure 8 we show the positions of the aromatic amino acids in the channels.

	number of AA	Phenylalanin	Tyrosine	Tryptophan	% aromatic AA
KcsA_closed	411	8	16	20	11
KscA_open	360	8	16	16	11
MthK	328	24	20	8	16
Kir2.2	412	44	8	8	15
NaV	432	52	16	12	19
NaK	334	40	12	0	16
Mlotik1	448	28	12	16	13

Table 3: Number of all the amino acids (AA) of each ion channel, the number of the aromatic amino-acids phenylalanine, tyrosine and tryptophan and their percentage in the different ion channels.

Subsequently the mutated ion channels were modeled as described in Chapter 2.2.2. Hence they are referred to as homology models.

We wanted to compare the structures with as little difference in the membrane positioning as possible. To ensure this we made an alignment of the channels via PyMOL. We aligned each mutated and homology model with the wild type.

4.1.2 Alignment of the sequences

For the creation of the alignments we used ClustalX⁸¹. To illustrate the results in color we used ESPript 2.2^{82,83} (see Figure 6).

4 Methods

4.1.3 Insertion of the model into the membrane

As a lipid bilayer we used DOPC (dioleoylphosphatidylcholine) in an already equilibrated state. The box size was $x = 9.5192$ nm, $y = 9.5717$ nm, $z = 7.5803$ nm.

To embed the protein in the membrane the Gromacs tool `g_membed` was used (see Chapter 2.2.3). Therefore the protein-structure was shrunk to 10% of its original size and re-extended to its normal size using 1000 steps to perform this. As the execution of `g_membed` is time consuming, we chose to use it only for the wild type model. To speed things up we used the Gromacs tool `g_confrms` and replaced the wild type model with each mutated and homology model form. To exchange the proteins we used the C-alpha backbone as a reference, so we ensured to preserve the conformation of the side chains. Both methods will yield the same results as the minimization and equilibration starts after this step.

To set position restraints we used the `g_genrestr` tool of Gromacs. The force constants were set to $5000 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ in every direction for the backbone.

4.1.4 Inclusion of Ions

Ions were inserted by using the `g_genion` tool of Gromacs. We wanted to obtain a concentration of 0.1 Mol and maintain electroneutrality. The number of required ions differed between each simulation. Ions were added by us if they were not given in the filter region of the wild type crystal structure. We ensured that at least one water molecule was between adjacent ions in the filter. Throughout the simulations, ions were not restrained. In the potassium channels the filter ions were placed in S0, S2 and S4 position as recommended by Åqvist and Luzhkov⁸⁴. In sodium channels no such positions are recommended until now. For NaK sodium and calcium ions are already included in the crystal structure. The sodium ions are beaded starting in the filter and ending at the upper end of the cavity. We only used the first and third sodium ion and the calcium ion for this channel. The other two ions were manually replaced by water molecules, to gain conditions similar to potassium ion channels. For NaV we used a model already established by our group. As there are no ions in the crystal structure, we used the free molecular dynamics data (the corresponding simulations were performed by Ke Song, MSc) where ions moved into the filter during the simulation.

4 Methods

4.1.5 Minimization and Equilibration

After this, 2 steepest descent minimizations were performed with a maximum number of 50,000 steps. The cut-off of the neighboring list was set to 1.2 nm.

Before equilibration the backbone restraints were removed again. The first equilibration step was executed for 100 ps using a modified Berendsen thermostat (see 3.2.2) with a reference temperature of 323 K. The second equilibration was performed using a pressure coupling NPT ensemble (see 3.2.3). The calculation time was set to 1,000 ps. The reference pressure was adjusted to 1.0 bar.

4.1.6 Molecular dynamics simulations

For the molecular dynamics simulations the process time was set to 20 ns. We performed five 20 ns MD simulations for each wild type, alanine mutant and homology model model.

Used parameters:

- cut offs: 1.0 nm
- reference temperature: 300 K
- reference pressure: 1.0 bar
- Particle Mesh Ewald for long distance electrostatics

Between each step we used VMD as a visualization tool to take a look at the resulting models and ensure that the protein maintained its conformation. The protein as well as the ions in the filter were confirmed to stay in their positions.

We performed 180 MD simulations. All together the simulation time of our calculations amounts to 3,600 ns.

We performed these simulations either on VSC1 (Vienna Scientific Cluster) or on VSC2 to speed it up. One calculation like this takes 52h on 48 CPU's. To perform all the simulations we needed more than 9,400 CPU-hours.

4 Methods

4.2 Evaluations

We analyzed the structure and models using the following approaches:

- inspection of MD simulations
 - stability of the proteins (e.g. unfolding of helices etc.)
 - movement of ions in the filter
- RMSD (root-mean-square-deviation)
- RMSF (root-mean-square-fluctuation)
- packing quality (WHATCHECK)

4.2.1 Inspection of MD simulations

To visualize the trajectories we used VMD (see 3.3.3). Each simulation was classified according to the stability of the ions inside the filter, the stability of the filter and the stability of the protein. As this analysis gives only a first impression of the quality of the protein, we also used other analyzing parameters (see below).

4.2.2 RMSD

A commonly used method to compare structures is the root-mean-square-deviation. When computing the RMSD one calculates the average distance between the atoms of superimposed proteins. Gromacs supports this with its `g_rms` tool.

4.2.3 RMSF

The root-mean-square-fluctuation gives information about the average fluctuation of some defined average position over a defined period of time. We compared the root-mean-square-fluctuation of each domain by using the Gromacs `g_rmsf` tool. To divide each domain an index-file with the corresponding atom numbers was created.

We applied the RMSF to compare the average of the crystallographic structure to the alanine-

4 Methods

mutant and the homology models average.

Another comparison was the averages of the backbone RMSF of the wild type to the homology model. When large varieties appeared in the backbone RMSF, the RMSF of the side chains were compared with each other (see Figure 14).

4.2.4 Packing quality

We analyzed the packing quality using the WHAT IF web interface⁸⁵. The used method is called Directional Atomic Contact Analysis. The idea of this method is to divide the amino acids into fragments (those fragments are as large as possible but do not contain rotational degrees of freedom around dihedral angles; for example: the 2 aromatic rings of Tryptophan are one fragment). The amino acids are also segmented into special atom types. Then the contact probability densities for each atom class around each fragment type are calculated on grid points arranged in a cubic box around each fragment. The probabilities of occurrence of certain atom types at the grid points around a fragment are estimated. This gives the packing quality. The higher the packing value the better the packing quality. G. Vriend and C. Sander came to the conclusion, that structures with a quality index below -2.5 are very probably wrong and structures with a quality index below -1.2 should be treated with great caution.⁸⁶ This is valid for globular protein structures with a crystal structure resolution of less than 1 Å. For transmembrane proteins like ion channels which have resolutions of 2 to 4 Å, packing quality values lower than -4 are most probably wrong.⁸⁵

5 Results

5.1 Homology modeling

We performed homology modeling on seven ion channels as described in the Methods section. We used Modeller to generate 20 homology models for each ion channel.

		RMSD [nm]						
		KcsA_closed	KcsA_open	MthK	Kir2.2	NaV	NaK	Mlotik1
wild-type	hom_mod_1	0.364572	0.164226	0.304302	0.165312	0.198178	0.165312	0.173228
wild-type	hom_mod_2	0.367243	0.168688	0.305502	0.163252	0.208578	0.163252	0.162582
wild-type	hom_mod_3	0.361110	0.170938	0.307865	0.155597	0.210047	0.155597	0.164024
wild-type	hom_mod_4	0.365567	0.177365	0.305384	0.169849	0.212608	0.169849	0.175763
wild-type	hom_mod_5	0.362548	0.179862	0.303387	0.168560	0.207448	0.168560	0.168611
wild-type	hom_mod_6	0.365727	0.185838	0.305110	0.168388	0.212468	0.168388	0.172336
wild-type	hom_mod_7	0.366046	0.172486	0.305170	0.153938	0.206737	0.153938	0.171972
wild-type	hom_mod_8	0.365815	0.173111	0.305031	0.166036	0.206482	0.166036	0.169382
wild-type	hom_mod_9	0.366044	0.163625	0.305706	0.172759	0.200820	0.172759	0.172159
wild-type	hom_mod_10	0.369006	0.182753	0.303688	0.162325	0.212665	0.162325	0.165170
wild-type	hom_mod_11	0.362374	0.169401	0.308347	0.165420	0.202801	0.165420	0.178010
wild-type	hom_mod_12	0.362579	0.162507	0.305122	0.166896	0.210973	0.166896	0.184568
wild-type	hom_mod_13	0.369686	0.179347	0.303189	0.167305	0.213656	0.167305	0.165104
wild-type	hom_mod_14	0.365528	0.162209	0.303194	0.162957	0.206327	0.162957	0.170573
wild-type	hom_mod_15	0.360565	0.180512	0.308222	0.173004	0.198329	0.173004	0.165869
wild-type	hom_mod_16	0.365602	0.168326	0.305791	0.161122	0.197326	0.161122	0.171938
wild-type	hom_mod_17	0.364824	0.173701	0.306606	0.156464	0.203821	0.156464	0.181946
wild-type	hom_mod_18	0.366523	0.174192	0.304585	0.159923	0.200312	0.159923	0.174285
wild-type	hom_mod_19	0.360146	0.161490	0.308855	0.162460	0.197536	0.162460	0.176981
wild-type	hom_mod_20	0.315893	0.162371	0.305269	0.146285	0.209762	0.146285	0.173303

Table 4: RMSD of the homology models of each ion channel. “Hom_mod” stands for the homology model.

The comparison was done between the wild type and each homology model. The lower the RMSD, the closer are the models to the crystallographic structure. The colored ones were submitted to further calculations.

As shown in Table 4, the RMSD of the homology models was already high even before the simulations, especially for KcsA in the closed state and MthK. During the simulations the RMSD values of KcsA in the closed state, MthK, NaV and NaK decreased, while they increased for KcsA in the open state, Kir2.2 and Mlotik1. This might be due to the resolution of the crystal structures, as KcsA in the closed state, MthK, NaV and NaK have a higher resolution than the other three structures.

5 Results

5.2 Alignment of the sequences

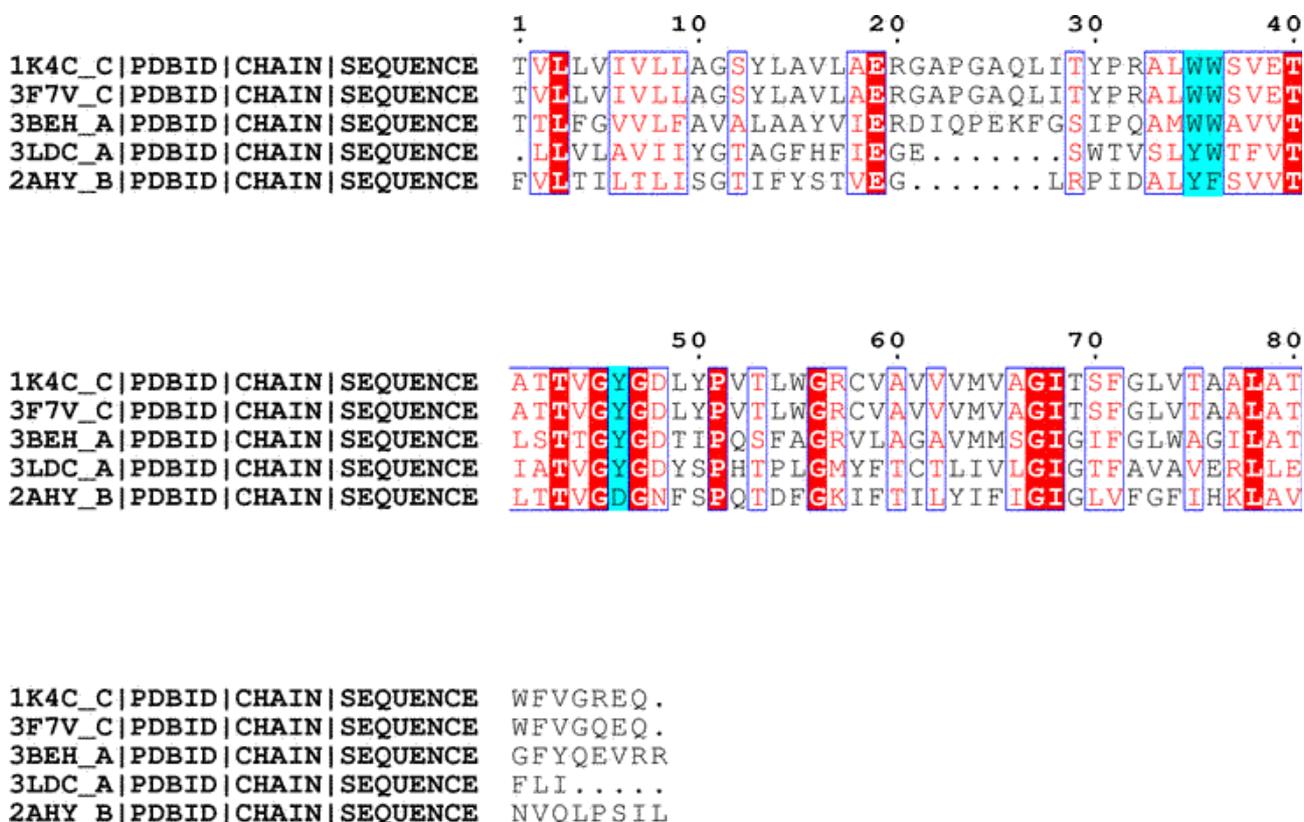
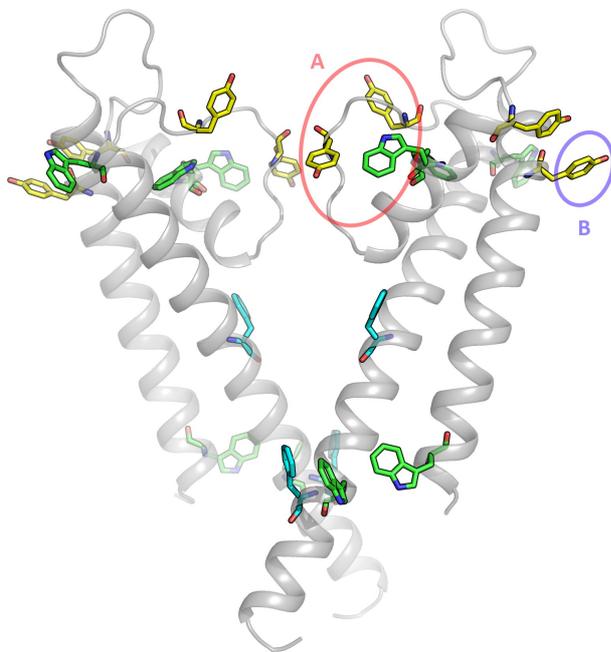


Figure 6: Multiple sequence alignment of the used crystal structures. The high conserved regions are colored in red, the low conserved regions are written in red. For the alignment we did not use NaV, as this is the most distant channel (it is an sodium channel). The cyan highlights indicate the strongly conserved aromatic amino acids in and close to the filter.

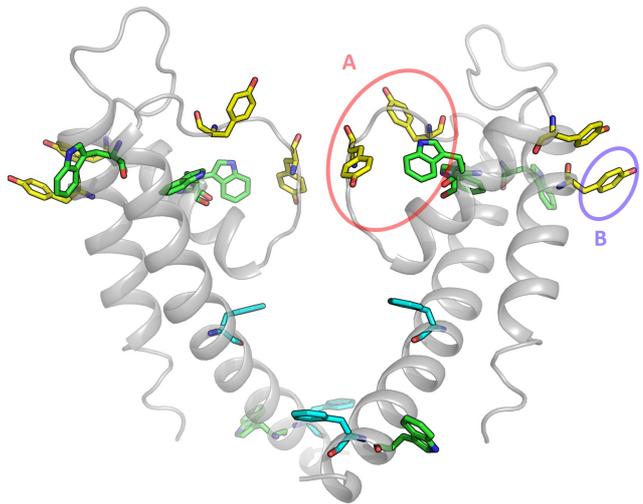
There are a lot of similarities in this channels (NaV is no shown in the alignment, as it is a sodium channel). The high conserved aromatic amino acids highlighted in Figure 6 are essential for the stability of the filter (as can be seen in Figure 16 to 18 and Figure 21 - the alanine-mutant has a higher RMSF of the filter region then the wild type).

5 Results

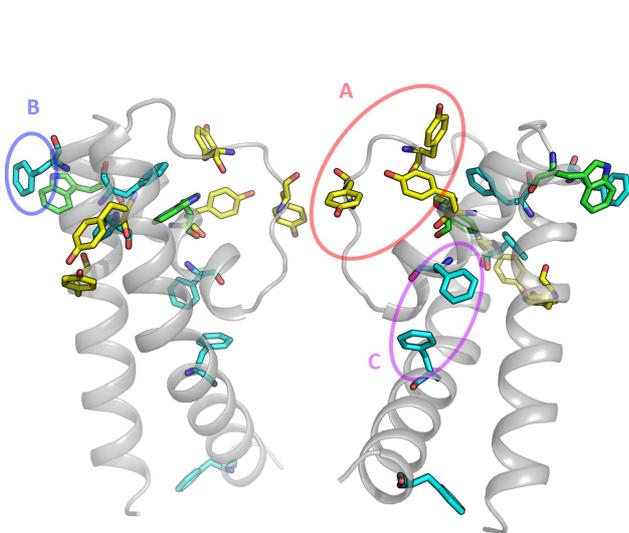
In our further analysis we often refer to the aromatic amino acids of the ion channels. Therefore they are illustrated in the following two figures (Figure 7 and Figure 8).



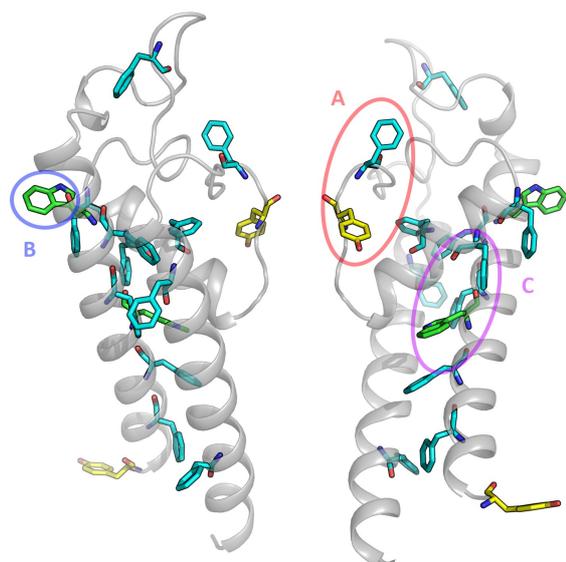
7A



7B



7C



7D

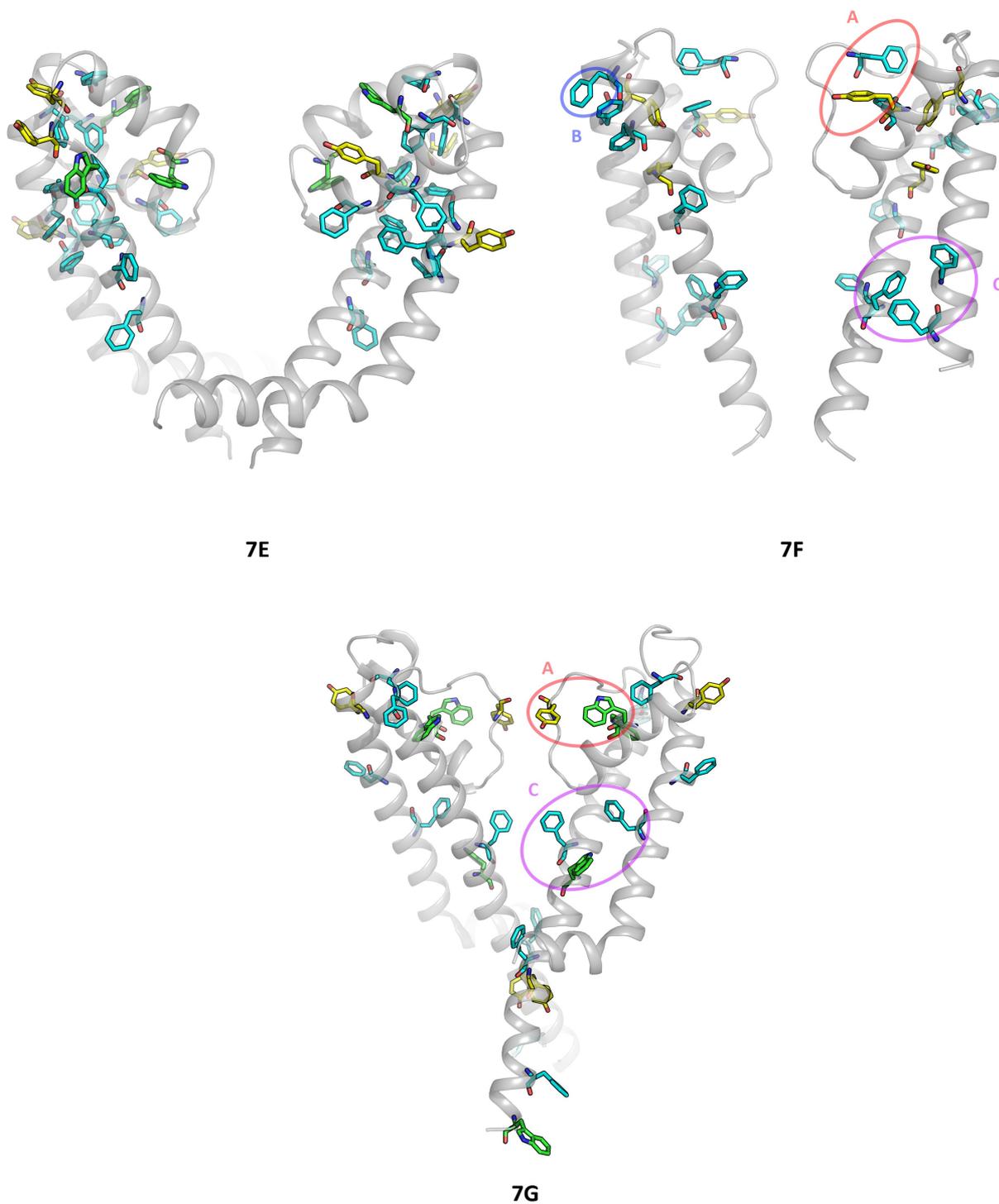
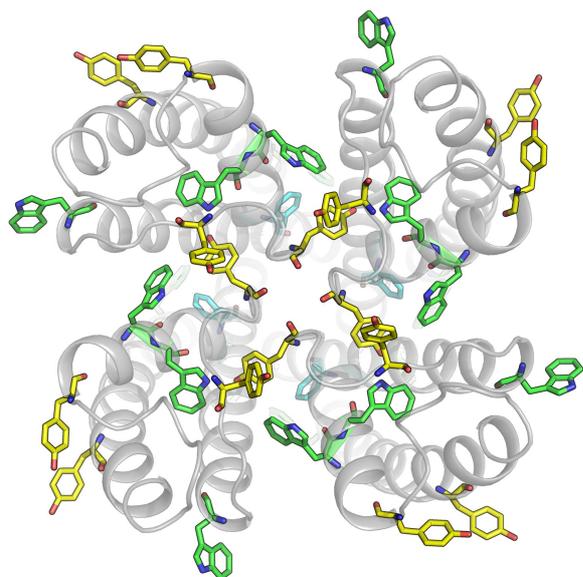


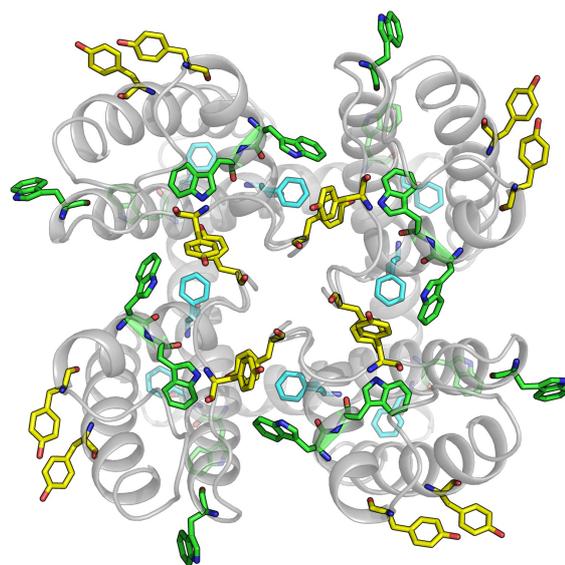
Figure 7: Sideview on two domains of the ion channels and their aromatic amino acids illustrated as sticks (phenylalanine = cyan, tyrosine = yellow and tryptophan = green). 7A = KcsA in the closed conformation, 7B = KcsA in the open conformation, 7C = MthK, 7D = Kir2.2, 7E = NaV, 7F = NaK and 7G = Mlotik1.

5 Results

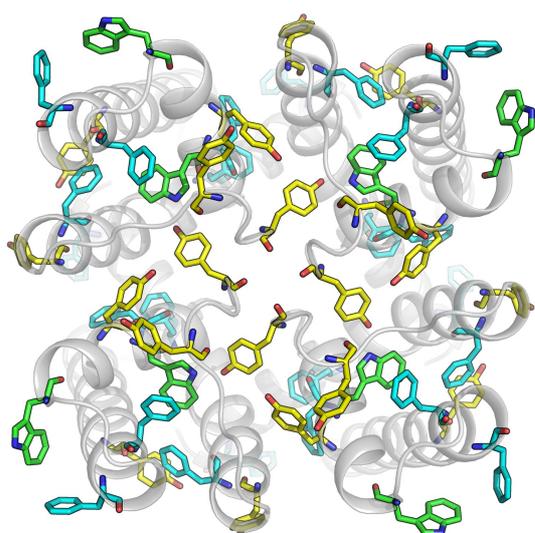
In Figure 7 we highlighted some special regions: region A indicates two to three aromatic residues that are important for the stability of the filter (red ellipse), region B indicates the membrane anchoring points (blue ellipse) and region C shows possible aromatic residues for interaction between the different helices (violet ellipse).



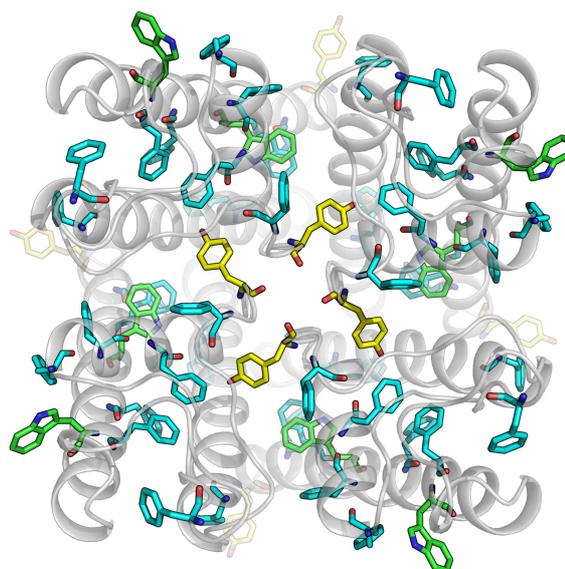
8A



8B



8C



8D

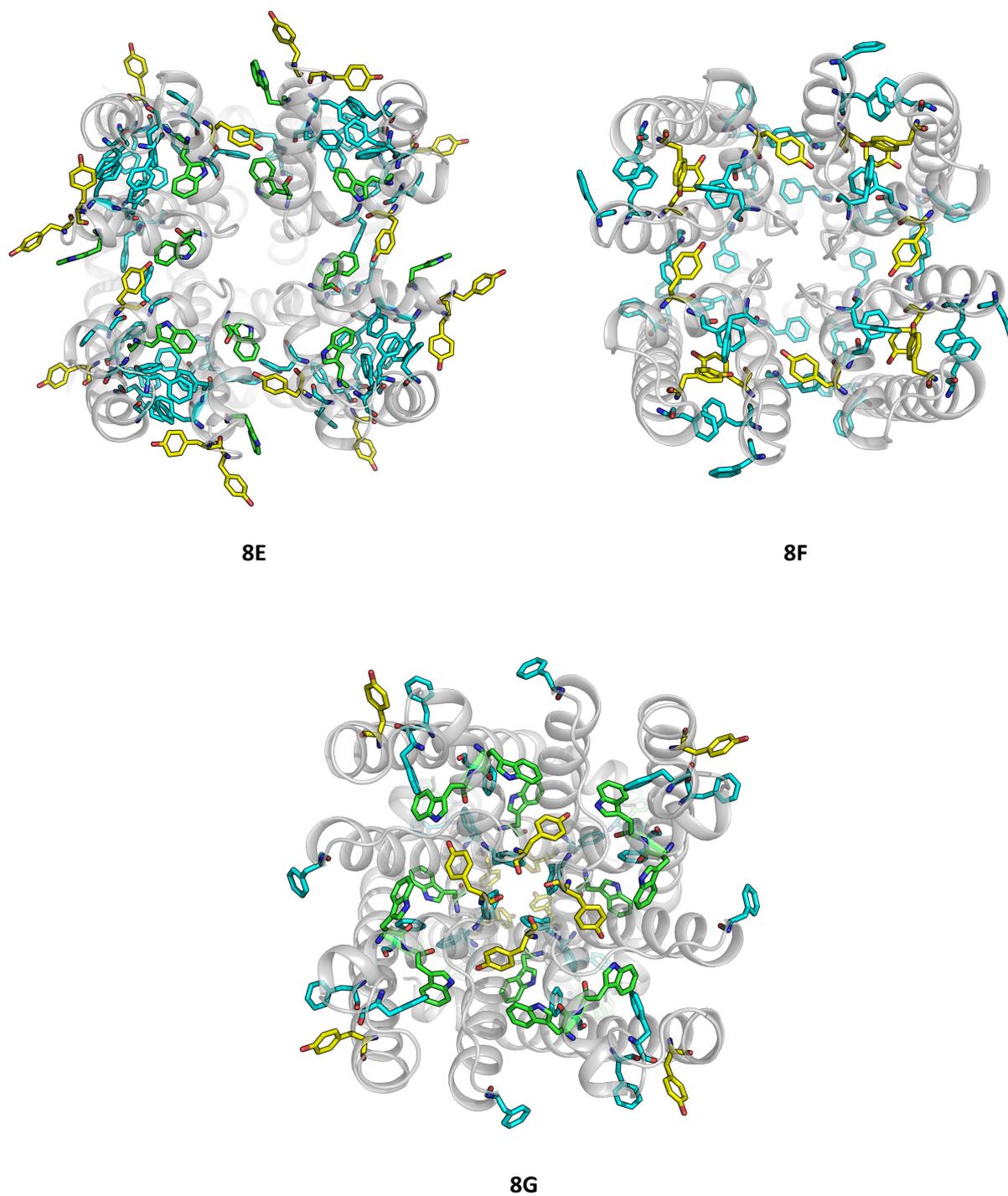


Figure 8: Topview on the ion channels and their aromatic amino acids illustrated as sticks (phenylalanine = cyan, tyrosine = yellow and tryptophan = green). 8A = KcsA in the closed conformation, 8B = KcsA in the open conformation, 8C = MthK, 8D = Kir2.2, 8E = NaV, 8F = NaK and 8G = Mlotik1.

5 Results

5.3 Visual MD analysis

In this analysis we focused on the number of ions in the filter, the stability of the filter and the stability of the rest of the protein as well as the movement of the protein in the embedding membrane during the simulations. The number of ions and other parameters (like partially unfolding of some helices or the movement of the whole protein in the membrane) were averaged over all the simulations.

We divided the behavior of the protein during the simulation into 7 different stability groups:

- A) The protein is stable (no unfolding (see Figure 9) of the helices); the filter is stable;
- B) The protein is less stable (1 or more helices are unfolding); the filter is stable;
- C) The protein is less stable (1 or more helices are unfolding); the filter is unstable;
- D) The protein is stable (no unfolding of the helices); the filter is unstable;

For the results see Table 5.

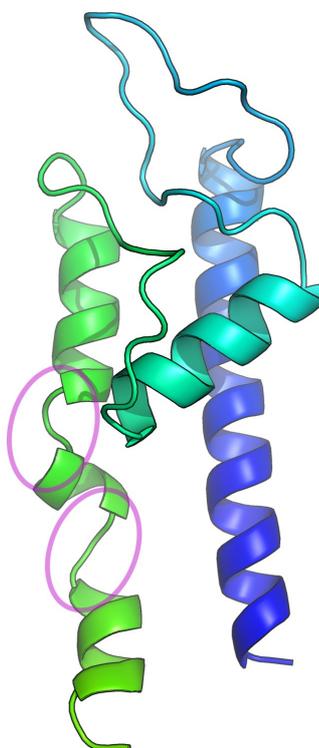


Figure 9: Example for partial unfolding of a S6 helix of Kir2.2 in the wild type. The highlights show the unwinded parts.

KcsA_closed_state	stability group	# Ions
wild-type	A	3
ala-mut	D	2–3
hom-mod	B	2–3

KcsA_open_state	stability group	# Ions
wild-type	A	3
ala-mut	3 × A, 2 × D	2
hom-mod	B	2

MthK_open_state	stability group	# Ions
wild-type	A	2–3
ala-mut	D	2–3
hom-mod	D	1

Kir2.2_closed_state	stability group	# Ions
wild-type	1 × A, 4 × C	2
ala-mut	B	2
hom-mod	B	2–3

NaV_closed_state	stability group	# Ions
wild-type	A	1
ala-mut	B	1
hom-mod	B	1

NaK_closed_state	stability group	# Ions
wild-type	D	3
ala-mut	D	3
hom-mod	D	3

Mlotik1_closed_state	stability group	# Ions
wild-type	B	2
ala-mut	B	2
hom-mod	B	2

Table 5: The table shows the stability groups of each channel (wild type, alanine-mutant (ala-mut) and the homology models (hom-mod)), as well as the number of ions in the filter.

5 Results

5.3.1 Conclusion

Kir2.2 and Mlotik1 show partial unfolding in the wild type – this is probably due to their low crystal structure resolution.

In contrast to previous studies⁸⁴, our research show that even when the filter appears to be unstable, the ions remain in the filter.

The wild type, alanine-mutant and homology model simulations show the same amount of ions in the filter, so their influence on the simulations can be ignored in the following calculations.

The visual MD analysis was done for preselection, as we based our further checks on this results.

5 Results

5.4 RMSD

We calculated the RMSD of the P1, S5, S6 and backbone of each model (see Table 6 and for illustration Figure 10). We therefore compared the P1, S5 and S6 with the backbone. The first 10 ns of the simulations were excluded as equilibration. The first and last 2 amino acids of each segment were omitted because they are the most flexible part of the protein and would increase the RMSD. To select the regions we had to create an index file, containing the information of the atom numbers for each group.

5.4.1 Conclusion

We compared the backbone RMSD of the wild type with the crystallographic resolutions. Our results suggest, that there is a correlation between resolution and the backbone RMSD of the simulations. MthK has the best resolution (1.45 Å) and the lowest backbone RMSD, while Mlotik1 (3.10 Å) has a high RMSD. The only notable exception to this finding is the KcsA in the open state which has the worst resolution (3.20 Å) but the third-best backbone RMSD.

As we can see in Table 6, the RMSD of the S6 has mostly the highest value. By comparing the RMSD of the crystal structures (= wild type) of the closed and open conformation of KcsA, we might be able to explain this by the conformational space needed for the opening and closure of the channels. During this process the S6 moves most, while the S5 and P1 hardly move at all.

In 5 out of 7 ion channels, the RMSD average of the alanine-mutant has the highest value. This indicates, that the alanine-mutant is not as stable as the wild type and the homology-models.

Kir2.2 and NaV show a low RMSD (Kir2.2: bad S6 and backbone RMSD, NaV: bad S5 and S6 RMSD) and a low packing quality (see Table 7). This indicates a trend, that structures with unfavorable side chain interactions have a high RMSD. As we only used the pore segment of each channel, the RMSD of Kir2.2 might be influenced by the loss of its large cytoplasmic domain, which was not included in our simulations.

The crystal structure has the most stable RMSD in every simulation, except for 3 homology model RMSDs (NaK backbone RMSD, Mlotik1 S6 and backbone RMSD).

5 Results

KcsA_closed	P1	S5	S6	BB
wild_type	0,132558	0,141320	0,139809	0,130920
mutant	0,149237	0,148240	0,177504	0,165022
average hom_mod	0,202649	0,192831	0,243015	0,216036

KcsA_open	P1	S5	S6	BB
wild_type	0,123357	0,148046	0,191383	0,128989
mutant	0,150411	0,194530	0,213868	0,169503
average hom_mod	0,167477	0,189714	0,255659	0,190693

MthK	P1	S5	S6	BB
wild_type	0,075982	0,139565	0,157481	0,098990
mutant	0,131229	0,164359	0,194538	0,163269
average hom_mod	0,150059	0,169812	0,188326	0,137294

Kir2.2	P1	S5	S6	BB
wild_type	0,110484	0,177228	0,308584	0,185012
mutant	0,138359	0,214404	0,435924	0,266191
average hom_mod	0,170605	0,240011	0,332167	0,225249

NaV	P1	S5	S6	BB
wild_type	0,118183	0,213578	0,236441	0,151342
mutant	0,198352	0,292836	0,306999	0,230156
average hom_mod	0,163257	0,266726	0,256610	0,182970

NaK	P1	S5	S6	BB
wild_type	0,110146	0,150016	0,194017	0,109416
mutant	0,143372	0,193756	0,209691	0,144207
average hom_mod	0,111863	0,176120	0,208268	0,107506

Mlotik1	P1	S5	S6	BB
wild_type	0,123398	0,203911	0,454793	0,233515
mutant	0,154103	0,252075	0,483025	0,286987
average hom_mod	0,171037	0,204222	0,440069	0,230591

Table 6: RMSD of P1, S5, S6 and backbone (BB) of each ion channel of the wild type, alanine-mutant and the averaged homology model (average hom_mod).

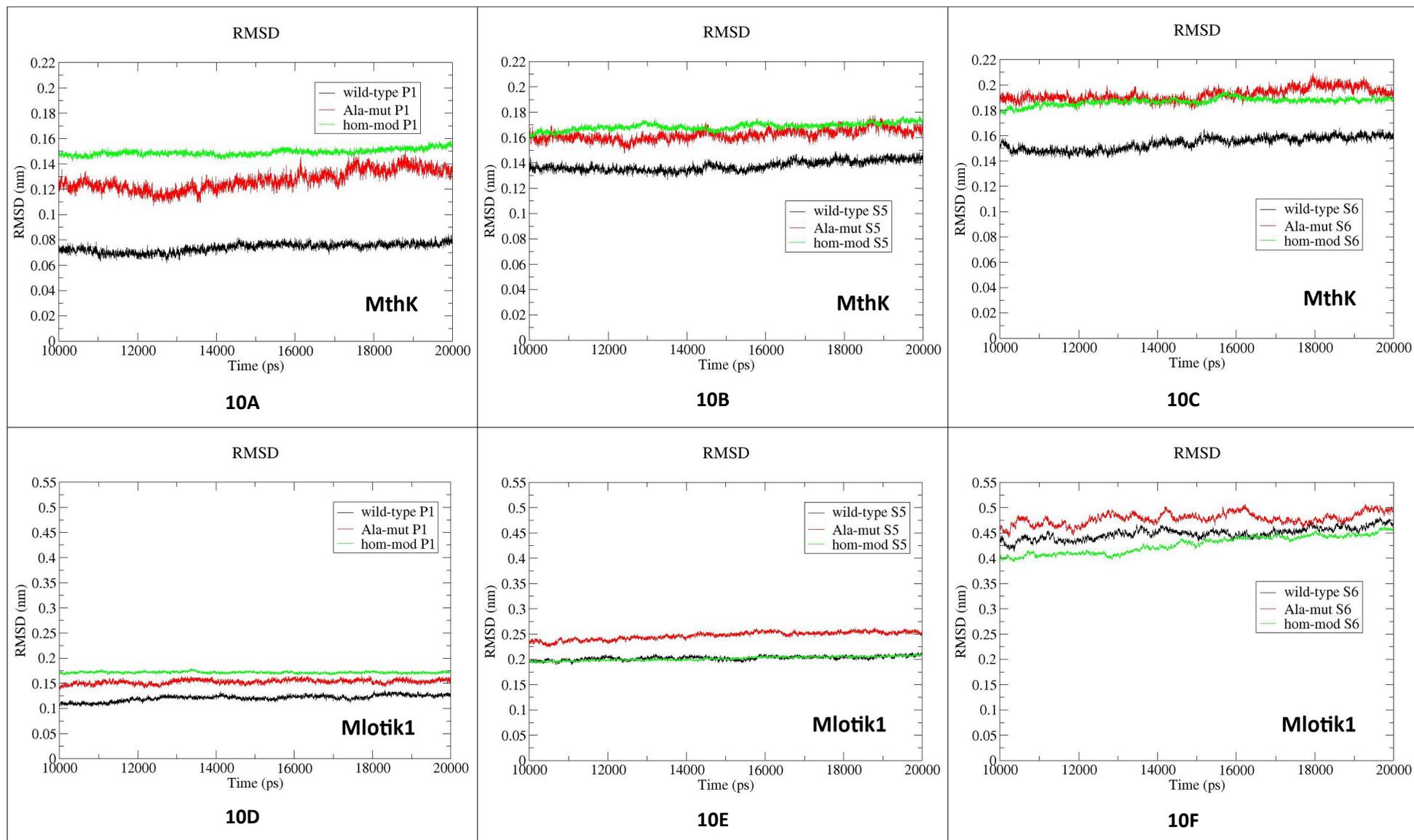


Figure 10: Root-mean-square deviations of MthK (10A RMSD of P1, 10B RMSD of S5 and 10C RMSD of S6) and Mlotik (10D RMSD of P1, 10E RMSD of S5, 10F RMSD of S6).

5.5 RMSF

The plots give detailed information about the fluctuation of each amino acid of the channel. The higher the value, the larger the fluctuation of the amino acid. An example is shown in Figure 12. It is easy to see that some regions are fluctuating more than others.

For the evaluation of the RMSF we implemented separations between: S5 – loop between S5 and P1 – P1 – filter – loop between filter and S6 – S6 (see Figure 11 and Figure 13).

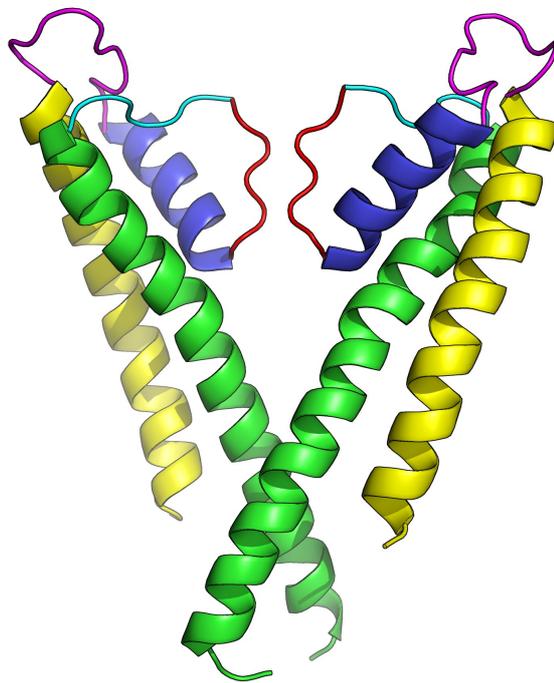


Figure 11: Separations made for RMSF evaluation in colors: S5 (yellow), loop between S5 and P1 (magenta), P1 (blue), filter-region (red), loop between filter-region and S6 (cyan) and the S6 (green)

RMS fluctuation

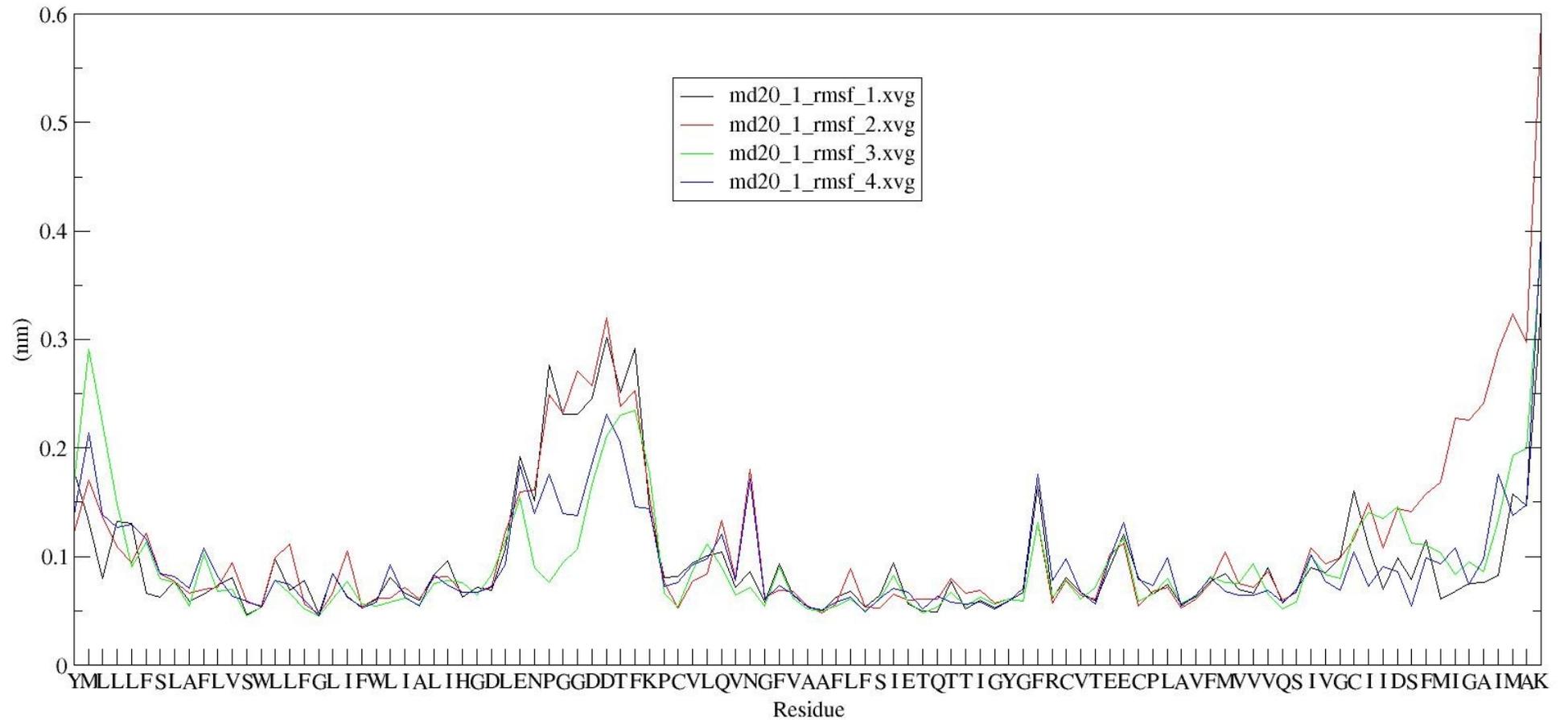


Figure 12: This figure shows the RMSF of Kir2.2 in the wild type model (first simulation). The four different graphs show the four domains. The letters on the x-axis are the short letter codes for the amino acid. On the y-axis the fluctuation is indicated in nm.

RMS fluctuation

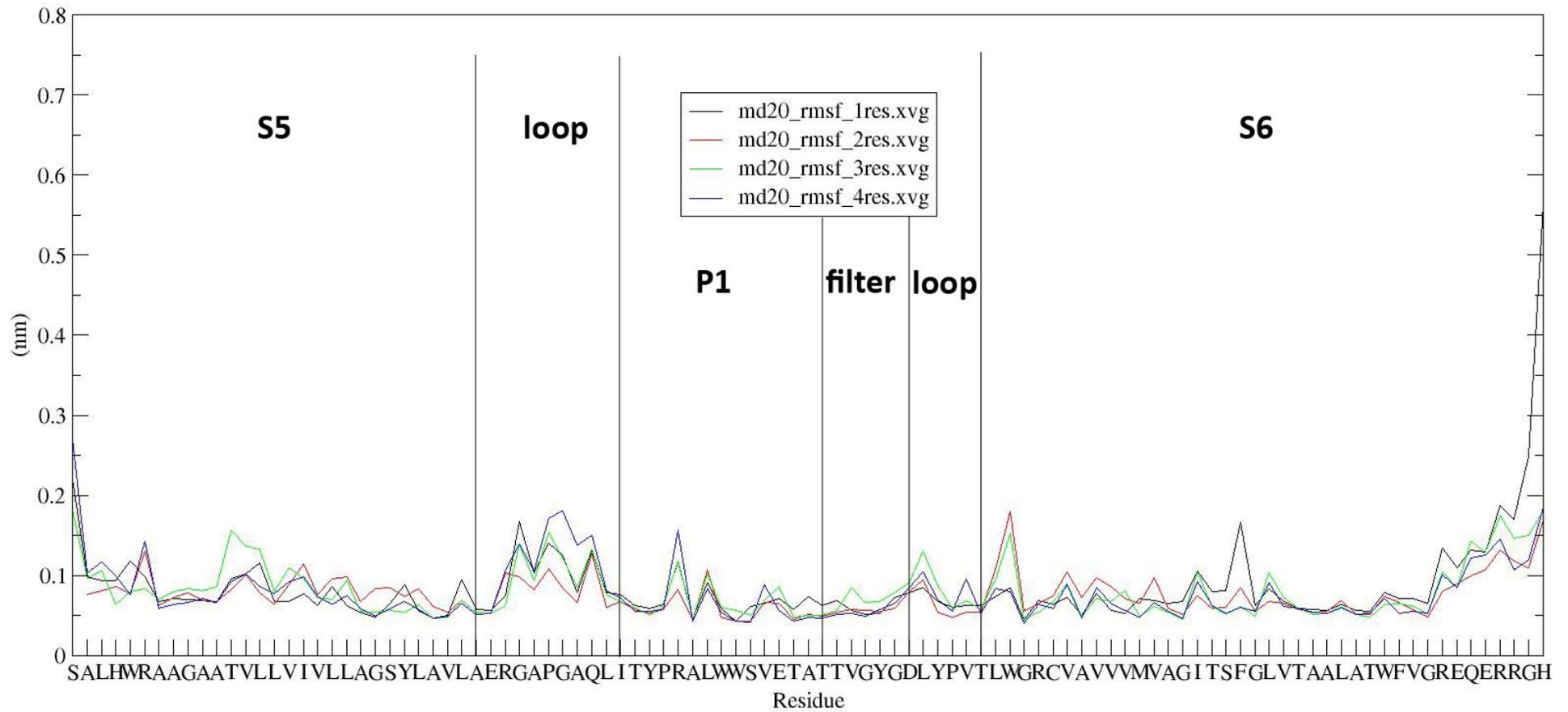


Figure 13: RMSF of the closed conformation of the KcsA wild type (first simulation). The four different graphs show the four domains. The graph is sectioned in different regions of each domain.

RMS fluctuation

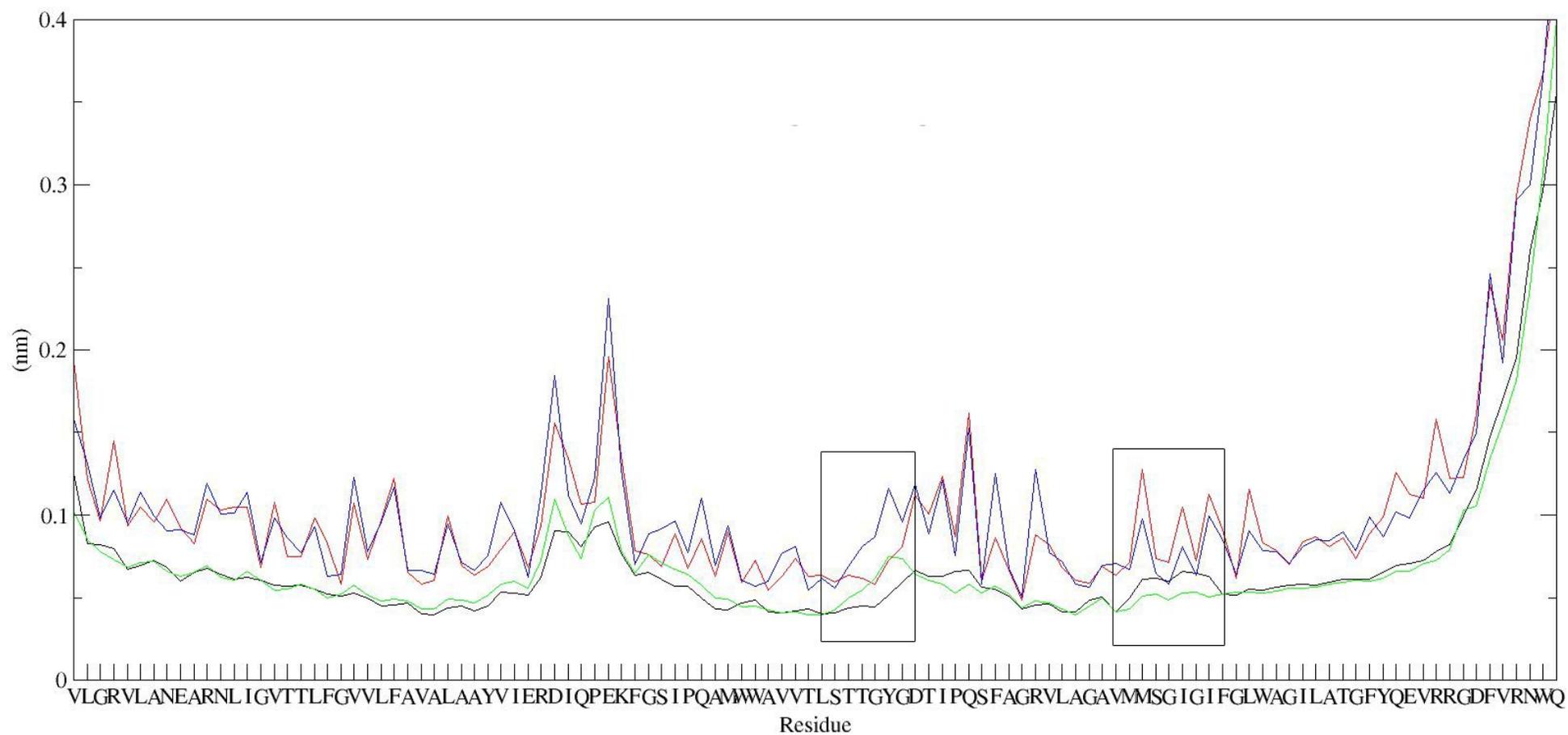


Figure 14: MlotiK1. RMSF of the average of the backbones and side chains of the wild type and the homology model 2. Colors: black: backbone average of the wild type; red: side chain average of the wild type; green: backbone average of the homology model 2; blue: side chain average of the homology model 2; The inserted boxes indicate the highest differences of the RMSF of the backbone averages and the differences of the side chain RMSFs.

5 Results

In the further evaluation we distinguished between the S5, P1, S6, filter and the loop regions (as described in 5.5 RMSF and Figure 11).

In Figures 15 to 21 we show the average of the RMSF of the wild type, the alanine-mutant and the homology models and we highlighted the highest varieties of these simulations. The caption of the x-axes in the plots valid for the wild type and the homology model, not for the alanine-mutant, as all the aromatic residues (phenylalanine (F), tyrosine (Y) and tryptophan (W)) are replaced by alanines (A).

5.5.1 KcsA in the closed state

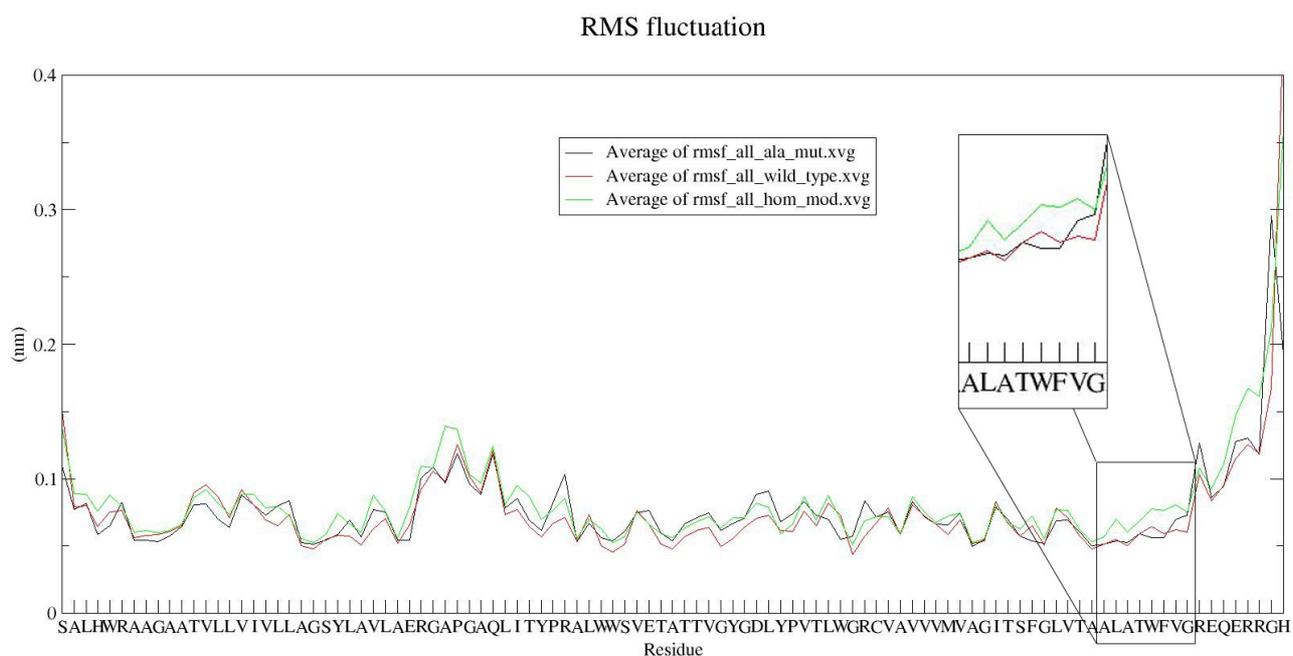


Figure 15: Averages of the RMSF of KcsA in the closed conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5 Results

5.5.2 KcsA in the open state

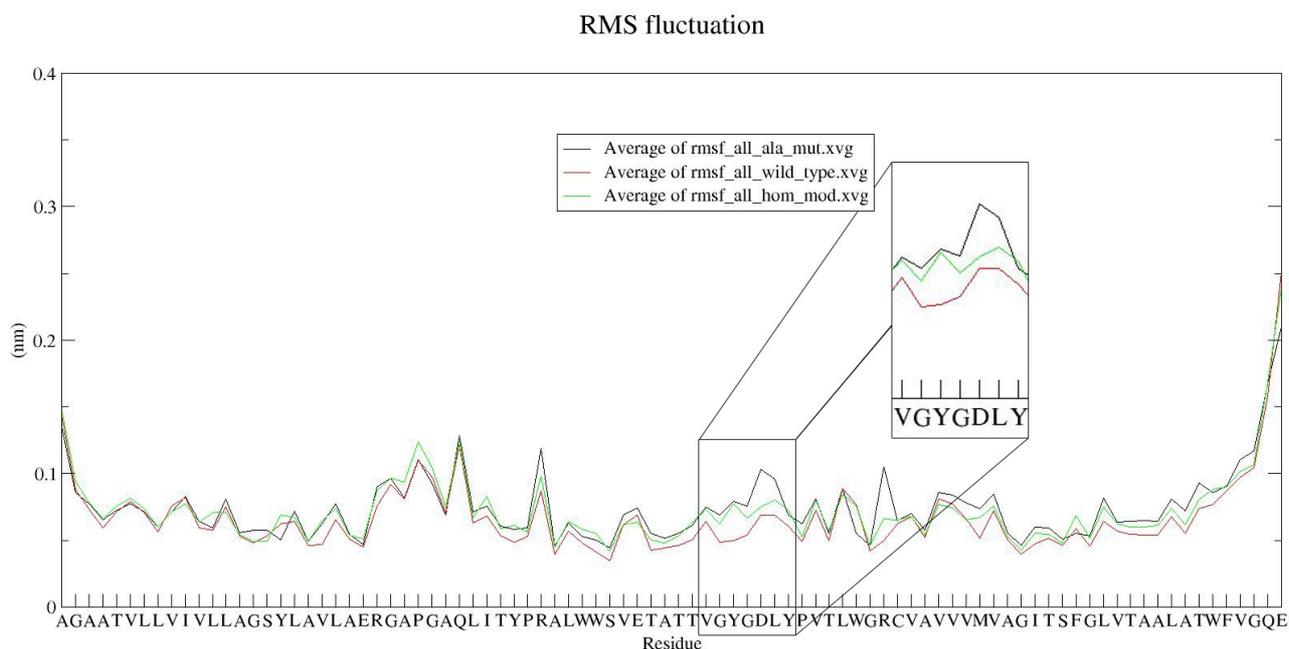


Figure 16: Averages of the RMSF of KcsA in the open conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5.5.3 MthK

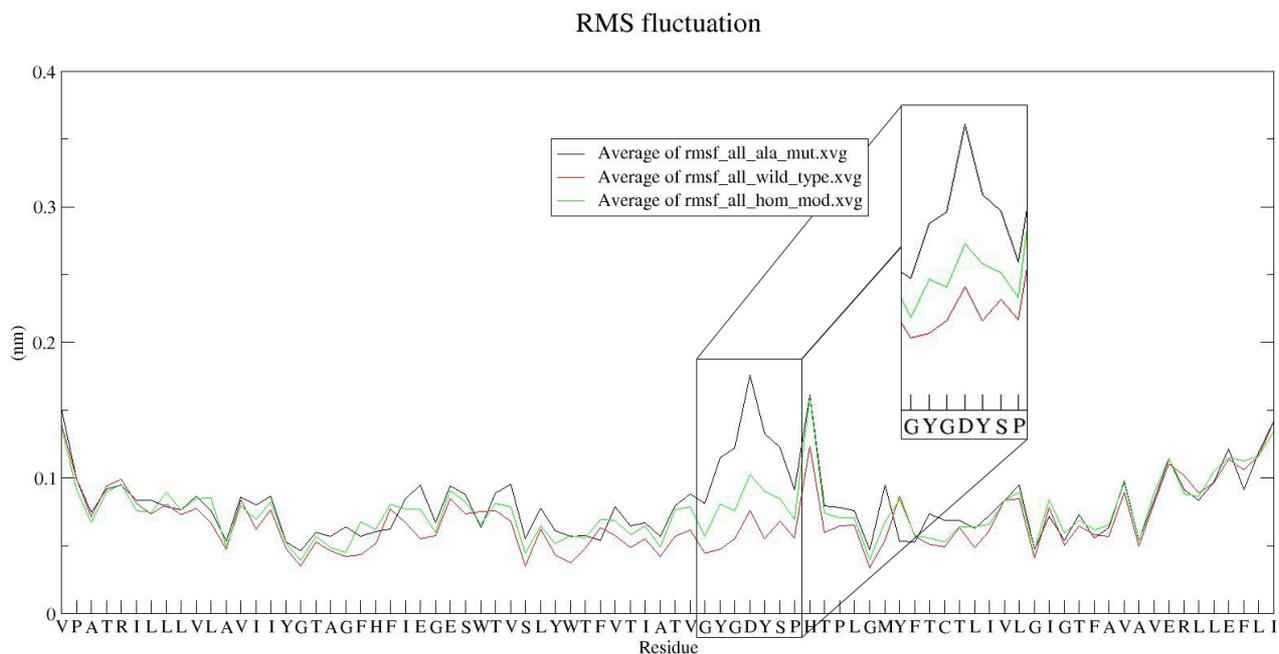


Figure 17: Averages of the RMSF of MthK in the open conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5 Results

5.5.4 Kir2.2

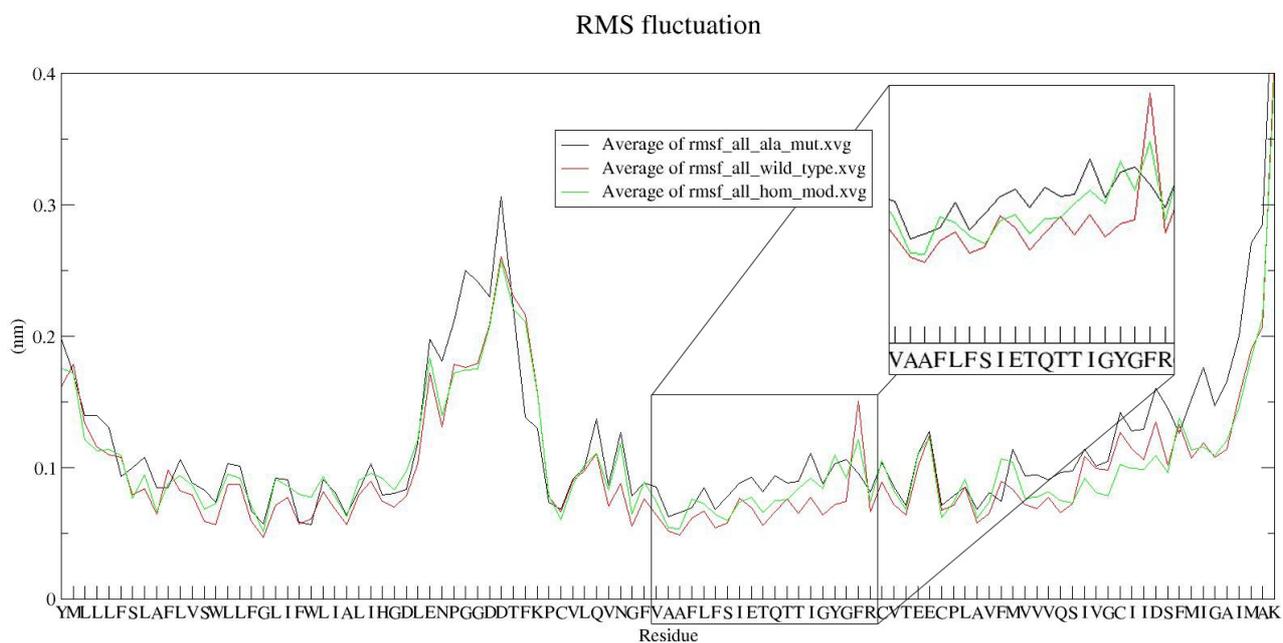


Figure 18: Averages of the RMSF of Kir2.2 in the closed conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5.5.5 Nav

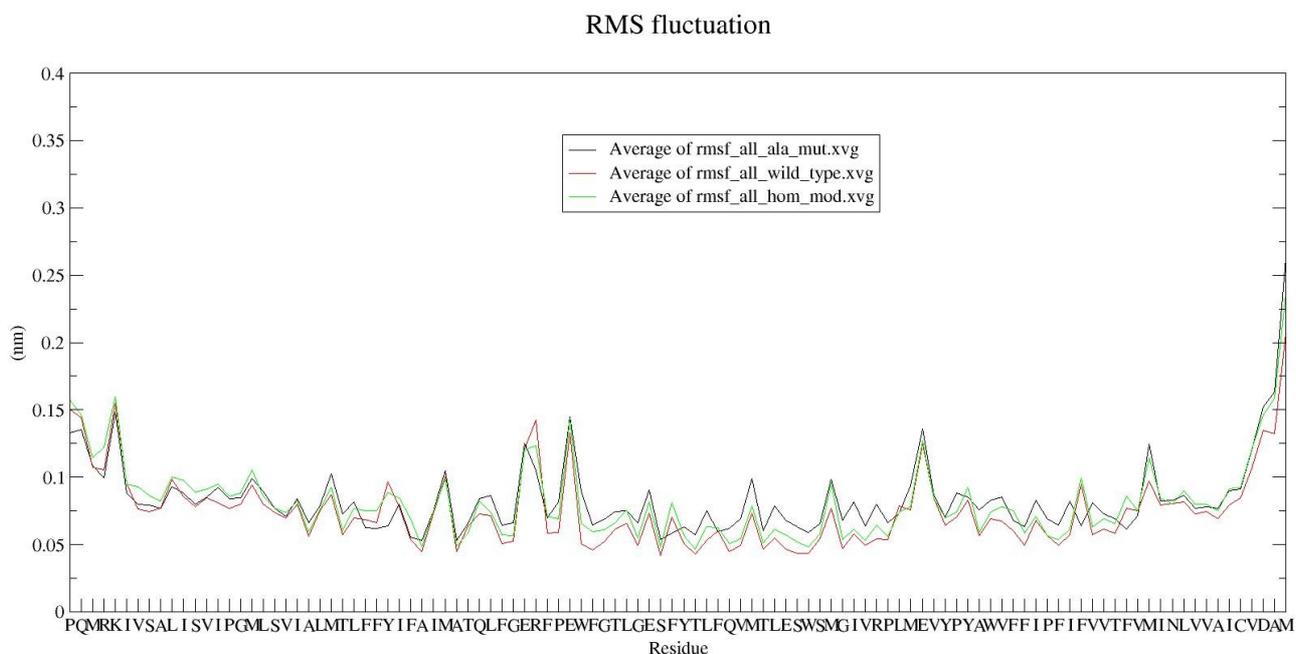


Figure 19: Averages of the RMSF of Nav in the closed conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5 Results

5.5.6 NaK

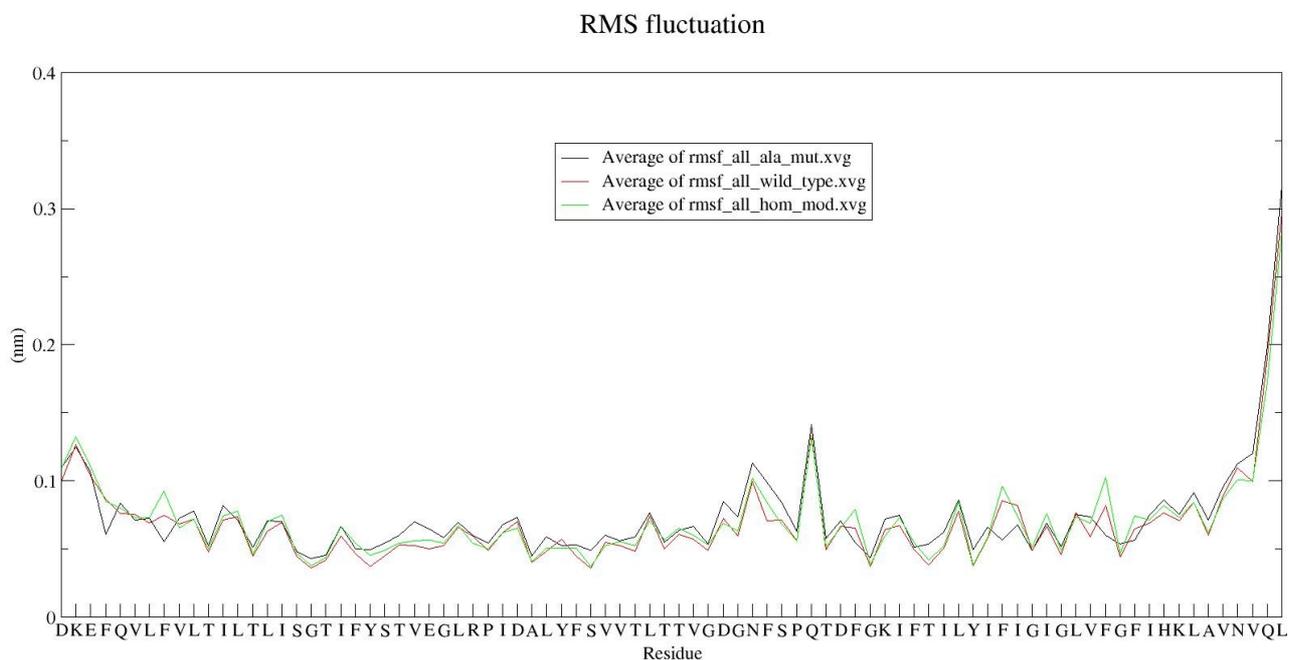


Figure 20: Averages of the RMSF of NaK in the closed conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5.5.7 Mlotik1

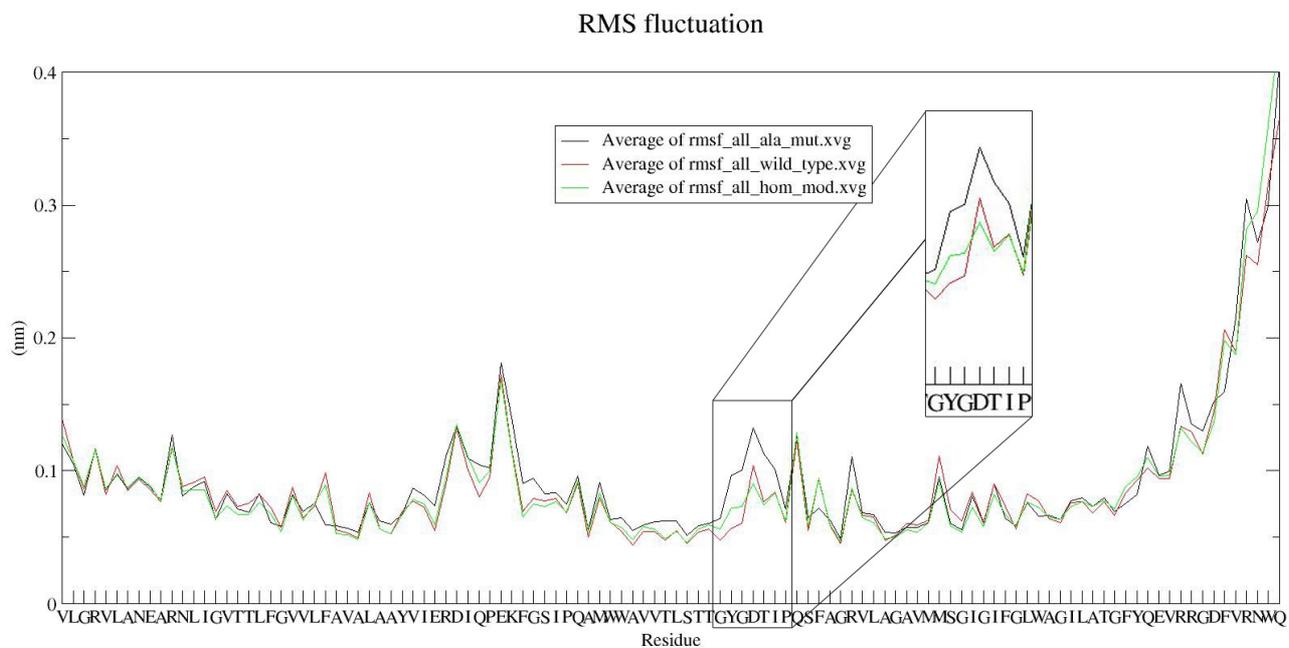


Figure 21: Averages of the RMSF of Mlotik1 in the closed conformation (wild type (green), alanine-mutant (black) and homology model (red)).

5 Results

5.5.8 Conclusion

As expected, the loop regions of the channels are the most flexible parts of the protein, this is the reason why those regions usually have a high RMSF. Kir2.2 has a long turret loop and therefore it shows a very high RMSF.

The wild type RMSF of the helical segments are approximately 0.5 to 1.5 Å.

Usually the filter is stabilized by nearby aromatic residues. In NaV (Figure 19) and NaK (Figure 20) the filter is not stabilized by aromatic residues and in this region hardly a difference in the RMSF of the wild type and the alanine-mutant can be reported. Interestingly the RMSF of the filter region of KcsA in the closed state shows the same behavior. The exact reasons remain unclear.

The alanine-mutant simulations contain regions with higher RMSF values than the wild type. This might be due to the effect of the loss of hydrogen bonds and interactions built by the aromatic residues of the wild type with other surrounding amino acids.

Kir2.2 seems to be less tolerant to alanine-mutations in the P and S6 segments, as the alanine-mutant RMSF value is noticeably higher in this region than the one of the wild type. As highlighted in Figure 7D (region C, violet ellipse) the P-helix builds aromatic interactions with aromatic residues of the S6. This might lead to a stabilization of the protein.

In Figure 7D (MthK), Figure 7F (NaK) and Figure 7G (Mlotik1) we indicated the possible aromatic interactions (region C, violet ellipse) between the P, S5 and S6 helices – we already described the effect of these interactions on the Kir2.2 channel. By comparing the RMSF results of those channels with the other channels one can see that these interactions are not decisive for the stability, as the other channels (KcsA in the closed state and NaV) show the same RMSF stability without a possibility for such aromatic interactions. It seems that generally the flexibility of the S6 is very high and therefore conformational changes are highly tolerated.

By comparing the averages of the backbone and the side chain RMSF of the wild type and the homology model (see Figure 14) one can see that even small changes in the backbone RMSF increase the RMSF of the side chains.

5 Results

5.6 Packing quality

kcsa_closed_state	Average	kcsa_open_state	Average
wild-type	-1.728	wild-type	-1.872
ala_mut	-1.437	ala_mut	-1.753
hom_mod_1	-2.310	hom_mod_14	-2.084
hom_mod_3	-2.515	hom_mod_19	-2.200
hom_mod_19	-2.469	hom_mod_20	-2.256
hom_mod_20	-2.402		

mthk_open_state	Average	Kir2.2_closed_state	Average
wild-type	-1.777	wild-type	-3.892
ala_mut	-2.015	ala_mut	-3.891
hom_mod_1	-2.134	hom_mod_4	-4.073
hom_mod_13	-2.327	hom_mod_5	-4.188
hom_mod_14	-2.362	hom_mod_7	-4.208

nav_closed_state	Average	nak_closed_state	Average
wild-type	-4.358	wild-type	-3.262
ala_mut	-5.031	ala_mut	-3.506
hom_mod_1	-4.283	hom_mod_3	-3.639
hom_mod_16	-4.411	hom_mod_7	-3.217
hom_mod_19	-4.538	hom_mod_20	-3.006

mloitk_closed_state	Average
wild-type	-2.597
ala_mut	-2.678
hom_mod_2	-2.812
hom_mod_3	-2.742
hom_mod_13	-2.707

Table 7: Packing quality of all used ion channels. Here the averages of the 5 simulations per wild type, alanine mutant (ala_mut) and homology models (hom_mod) are shown.

5.6.1 Conclusion

As we stated in Evaluations (chapter 4.2), every packing quality index lower than -4 should be treated with great caution. In Table 7 we listed our packing values. Some of them are lower than -4, especially the packing values of NaV and Kir2.2.

In 5 out of 7 channels the alanine-mutant shows a better packing quality value than the homology models. This might be explained by the small amino acid used for the mutation. The aromatic amino acids need more space than alanine.

5 Results

The aromatic amino acids are big, so they have the highest influence on the side chain packing. As we did not refine any side chain conformations of the homology models, we can conclude that the conformations generated by Modeller are of low quality.

All together this evaluation turned out to be less informative than we expected. The packing quality of the alanine-mutant is a bit lower than the one of the wild type. The homology models packing quality was mostly the worst.

6 Conclusion

One of the biggest disadvantages we see in Modeller is rather poor refinement of side chains. Our expectation was that Modeller would only generate the new conformations of the aromatic residues and take the other side chain conformations from the original structure file. But to the contrary, it also remodels every other side chain conformation. This leads to homology models with bad side chain conformations.

As this reduces the validity of the comparison of the alanine-mutant and the homology models, the alanine-mutant is only important to answer the question about the function of the aromatic residues in the channels and their contribution to its stability.

By making an alignment we could see that we are working with six highly related structures – NaV is an exception, as it is a sodium channel.

We performed a visual inspection of the MD simulations as a preselection. An interesting finding in this analysis is that even when the filter appears to be unstable, the ions stay in the filter. The wild type, alanine-mutant and the homology model of a channel have the same amount of ions in the filter.

We could find a trend saying that structures with unfavorable side chain interactions have a high RMSD. Therefore, the crystal structures has the lowest RMSD in every simulation. The RMSD of the S6 mostly obtains the highest value. This might be explained by the conformational space needed for the opening and closure of the channels.

By comparing the RMSF plots we could see, that the loops are the most flexible part of the proteins. The RMSF value of the wild type is something between 0.5 and 1.5 Å.

As the RMSF value of the wild type and the alanine-mutant filter region is quite different in some cases (exceptions are the NaV and NaK – both have no aromatic amino acid in their filter region), we can conclude that the stability of the filter depends on the aromatic residues. The aromatic clusters between the S5, P and S6 helices do not significantly change the stability of the protein. An increase in the backbone RMSF also increases the side chain RMSF.

The packing quality of the homology models was poor (in some cases even bad).

6 Conclusion

As a conclusion, we strongly recommend to perform a side chain refinement by hand or use programs like SCWRL⁸⁷ before using a Modeller generated conformation for molecular dynamics simulations.

7 List of figures

Figure 1: Structure of three superimposition ion channels.	9
Figure 2: Two opposing pore-forming segments of the closed KcsA.	10
Figure 3: Alignments of KcsA in the closed and the open conformation.	13
Figure 4: Overview of the performed steps.	17
Figure 5: Two domains of the KcsA channel in the closed state embedded in a lipid bilayer.	19
Figure 6: Multiple sequence alignment of the used crystal structures.	33
Figure 7: Sideview on the different ion channels and their aromatic amino acids.	34-35
Figure 8: Topview on the different ion channels and their aromatic amino acids.	36-37
Figure 9: Partial unfolding of a S6 helix of Kir2.2 in the wild type.	38
Figure 10: Root-mean-square deviations of MthK.	43
Figure 11: Separations made for RMSF evaluation in colors.	44
Figure 12: RMSF of Kir2.2 in the wild type model.	45
Figure 13: RMSF of the closed conformation of the KcsA wild type.	46
Figure 14: RMSF of the average of the backbones and side chains of the wild type and the homology model of Mlotik1.	47
Figure 15: Averages of the RMSF of KcsA in the closed conformation.	48
Figure 16: Averages of the RMSF of KcsA in the open conformation.	49
Figure 17: Averages of the RMSF of MthK in the open conformation.	49
Figure 18: Averages of the RMSF of Kir2.2 in the closed conformation.	50
Figure 19: Averages of the RMSF of NaV in the closed conformation.	50
Figure 20: Averages of the RMSF of NaK in the closed conformation.	51
Figure 21: Averages of the RMSF of Mlotik1 in the closed conformation.	51

8 List of tables

Table 1: Information of the used crystal structures.	8
Table 2: Sequence identity (ID in percent) and RMSD of chosen ion channels.	8
Table 3: Number of all the amino acids and aromatic amino acids of each ion channel.	27
Table 4: RMSD of the homology models of each ion channel.	32
Table 5: Stability groups of each channel.	39
Table 6: RMSD of P1, S5, S6 and backbone of each ion channel.	42
Table 7: Packing quality of the used ion channels.	39

9 Lebenslauf

Persönliche Daten:

Eva-Maria Zangerl

Geboren am 25.01.1987 in Zams in Tirol

Ledig

Ausbildung:

1994-1998	Volksschule Strengen
1998-2005	Bundesrealgymnasium Landeck/Perjen
2005	Abschluss mit Matura
2005	Beginn des Studiums der Chemie an der Universität Wien
2010	Abschluss des Bachelor-Studiums der Chemie
2010	Beginn des Master-Studiums der Chemie

10 Bibliography

1. Kass, R.S. Review series introduction The channelopathies : novel insights into molecular and genetic mechanisms of human disease. **115**, (2005).
2. Catterall, W. A, Dib-Hajj, S., Meisler, M.H. & Pietrobon, D. Inherited neuronal ion channelopathies: new windows on complex neurological diseases. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **28**, 11768-77 (2008).
3. Membrane Proteins of Known Structure. at <<http://blanco.biomol.uci.edu/mpstruc/listAll/list>>
4. About MODELLER. at <<http://salilab.org/modeller/>>
5. swiss pdb viewer. at <<http://spdbv.vital-it.ch/>>
6. Catterall, W. A *et al.* Voltage-gated ion channels and gating modifier toxins. *Toxicon : official journal of the International Society on Toxinology* **49**, 124-41 (2007).
7. Armstrong, C.M. & Hille, B. Voltage-gated ion channels and electrical excitability. *Neuron* **20**, 371-80 (1998).
8. Description, Q. & Hodgkin, B.Y.A.L. Ik), (i,). *Current* 500-544 (1952).
9. Jiang, Y. *et al.* Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **417**, 515-22 (2002).
10. Armstrong, C. NEUROSCIENCE: The Vision of the Pore. *Science* **280**, 56-57 (1998).
11. Yu, F.H. & Catterall, W. A The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Science's STKE : signal transduction knowledge environment* **2004**, re15 (2004).
12. Doyle, D. a. The Structure of the Potassium Channel: Molecular Basis of K⁺ Conduction and Selectivity. *Science* **280**, 69-77 (1998).
13. Morais-Cabral, Ä.H., Kaufman, A. & MacKinnon, R. hydration revealed by a K channel ± Fab Ê resolution complex at 2 . 0 Å. *Nature* **414**, 43-48 (2001).
14. Ye, S., Li, Y. & Jiang, Y. Novel insights into K⁺ selectivity from high-resolution structures of an open K⁺ channel pore. *Nature structural & molecular biology* **17**, 1019-23 (2010).

10 Bibliography

15. Tao, X., Avalos, J.L., Chen, J. & MacKinnon, R. Crystal structure of the eukaryotic strong inward-rectifier K⁺ channel Kir2.2 at 3.1 Å resolution. *Science (New York, N.Y.)* **326**, 1668-74 (2009).
16. Shi, N., Ye, S., Alam, A., Chen, L. & Jiang, Y. Atomic structure of a Na⁺- and K⁺-conducting channel. *Nature* **440**, 570-4 (2006).
17. Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. a The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353-8 (2011).
18. Clayton, G.M., Altieri, S., Heginbotham, L., Unger, V.M. & Morais-Cabral, J.H. Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1511-5 (2008).
19. Anishkin, A., Milac, A.L. & Guy, H.R. Symmetry-restrained molecular dynamics simulations improve homology models of potassium channels. *Proteins* **78**, 932-49 (2010).
20. Yu, F.H., Yarov-Yarovoy, V., Gutman, G.A. & Catterall, W.A. Overview of Molecular Relationships in the Voltage-Gated Ion Channel Superfamily. *Pharmacological Reviews* **57**, 387-395 (2005).
21. Kariev, A.M. & Green, M.E. Voltage gated ion channel function: gating, conduction, and the role of water and protons. *International journal of molecular sciences* **13**, 1680-709 (2012).
22. Jogini, V. & Roux, B. Electrostatics of the intracellular vestibule of K⁺ channels. *Journal of molecular biology* **354**, 272-88 (2005).
23. Boiteux, C., Kraszewski, S., Ramseyer, C. & Girardet, C. Ion conductance vs. pore gating and selectivity in KcsA channel: modeling achievements and perspectives. *Journal of molecular modeling* **13**, 699-713 (2007).
24. DeCaen, P.G., Yarov-Yarovoy, V., Scheuer, T. & Catterall, W. A Gating charge interactions with the S1 segment during activation of a Na⁺ channel voltage sensor. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 18825-30 (2011).
25. Catterall, W. A Ion channel voltage sensors: structure, function, and pathophysiology. *Neuron* **67**, 915-28 (2010).
26. Catterall, W. A From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* **26**, 13-25 (2000).

10 Bibliography

27. Thompson, A.N., Posson, D.J., Parsa, P.V. & Nimigean, C.M. Molecular mechanism of pH sensing in KcsA potassium channels. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6900-5 (2008).
28. Bähring, R. & Covarrubias, M. Mechanisms of closed-state inactivation in voltage-gated ion channels. *The Journal of physiology* **589**, 461-79 (2011).
29. Catterall, W. A & Yu, F.H. Painful channels. *Neuron* **52**, 743-4 (2006).
30. Cox, J.J. *et al.* An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **444**, 894-8 (2006).
31. Du, W. *et al.* Calcium-sensitive potassium channelopathy in human epilepsy and paroxysmal movement disorder. *Nature genetics* **37**, 733-8 (2005).
32. Lv, P., Wei, D. & Yamoah, E.N. Kv7-type channel currents in spiral ganglion neurons: involvement in sensorineural hearing loss. *The Journal of biological chemistry* **285**, 34699-707 (2010).
33. Rajan, a S. *et al.* Ion channels and insulin secretion. *Diabetes care* **13**, 340-63 (1990).
34. Hering, S. *et al.* UKPMC Funders Group Pore stability and gating in voltage-activated calcium channels. *Channels* **2**, 61-69 (2011).
35. Napolitano, C. *et al.* Evidence for a cardiac ion channel mutation underlying drug-induced QT prolongation and life-threatening arrhythmias. *Journal of cardiovascular electrophysiology* **11**, 691-6 (2000).
36. Sánchez, R. *et al.* Protein structure modeling for structural genomics. *Nature structural biology* **7 Suppl**, 986-90 (2000).
37. Grahnen, J. A, Kubelka, J. & Liberles, D. a Fast side chain replacement in proteins using a coarse-grained approach for evaluating the effects of mutation during evolution. *Journal of molecular evolution* **73**, 23-33 (2011).
38. Voelz, V. A, Bowman, G.R., Beauchamp, K. & Pande, V.S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *Journal of the American Chemical Society* **132**, 1526-8 (2010).
39. Kellogg, E.H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing

10 Bibliography

- mutation-induced changes in protein structure and stability. *Proteins* **79**, 830-8 (2011).
40. Chase, F. & Avenue, B. Rotamer libraries in the 21 st century Roland L Dunbrack Jr. *Proteins* 431-440 (2002).
41. Xiang, Z. & Honig, B. Extending the accuracy limits of prediction for side-chain conformations. *Journal of molecular biology* **311**, 421-30 (2001).
42. Matthes, D. & de Groot, B.L. Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes. *Biophysical journal* **97**, 599-608 (2009).
43. Samudrala, R. & Moult, J. Determinants of side chain conformational preferences in protein structures. *Protein engineering* **11**, 991-7 (1998).
44. Canutescu, A.A., Shelenkov, A.A. & Jr, R.L.D. A graph-theory algorithm for rapid protein side-chain prediction. *Society* 2001-2014 (2003).doi:10.1110/ps.03154503.Sander
45. Zhang, Y. Protein structure prediction: when is it useful? *Current opinion in structural biology* **19**, 145-55 (2009).
46. Dunbrack_1993.pdf.
47. Chase, F. & Avenue, B. Rotamer libraries in the 21 st century Roland L Dunbrack Jr. *Proteins* 431-440 (2002).
48. Betancourt, M.R. Comparison between molecular dynamic based and knowledge based potentials for protein side chains. *Journal of computational biology: a journal of computational molecular cell biology* **17**, 943-52 (2010).
49. Liang, S. Side-chain modeling with an optimized scoring function. 322-331 (2002).doi:10.1110/ps.24902.tasks
50. Mendes, J., Baptista, A M., Carrondo, M. A & Soares, C.M. Improved modeling of side-chains in proteins with rotamer-based methods: a flexible rotamer model. *Proteins* **37**, 530-43 (1999).
51. Lesk, A. M. & Chothia, C.H. The Response of Protein Structures to Amino-Acid Sequence Changes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **317**, 345-356 (1986).

10 Bibliography

52. Sánchez, R. & Sali, A. Advances in comparative protein-structure modelling. *Current opinion in structural biology* **7**, 206-14 (1997).
53. Mart, M.A., Stuart, A.C., Roberto, S., Melo, F. & Sali, A. C p s m g g. *Biochemistry* 291-325 (2000).
54. Wolf, M.G., Hoefling, M., Aponte-Santamaría, C., Grubmüller, H. & Groenhof, G. g _ membed : Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation. *Journal of Computational Chemistry* (2010).doi:10.1002/jcc
55. Alder, B.J. & Wainwright, T.E. Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics* **27**, 1208 (1957).
56. Manual, U. User manual. *Bioinformatics*
57. Bond, P.J., Holyoake, J., Ivetac, A., Khalid, S. & Sansom, M.S.P. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. *Journal of structural biology* **157**, 593-605 (2007).
58. Ponder, B.J.A.Y.W. & Å, D.A.C. FORCE FIELDS FOR PROTEIN SIMULATIONS I . Introduction Molecular dynamics and Monte Carlo simulations of proteins , which directions the field is taking in developing new models . The most commonly used protein force fields incorporate a relatively simple. *Advances* **66**, 27-85 (2003).
59. Kaminski, G.A., Friesner, R.A., Tirado-rives, J. & Jorgensen, W.L. Comparison with Accurate Quantum Chemical Calculations on Peptides †. *Quantum* **2**, 6474-6487 (2001).
60. Manuscript, A. NIH Public Access. **30**, 1545-1614 (2010).
61. Oostenbrink, C., Villa, A., Mark, A.E. & van Gunsteren, W.F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of computational chemistry* **25**, 1656-76 (2004).
62. Tozzini, V. Coarse-grained models for proteins. *Current opinion in structural biology* **15**, 144-50 (2005).
63. Hornak, V. *et al.* Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Bioinformatics* **725**, 712-725 (2006).

10 Bibliography

64. gromacs tutorial. at <http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/03_solvate.html>
65. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, a. & Haak, J.R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684 (1984).
66. Binder, K., Horbach, J., Kob, W., Paul, W. & Varnik, F. Molecular dynamics simulations. *Journal of Physics: Condensed Matter* **16**, S429-S453 (2004).
67. Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23 (1997).
68. Fiser, A. & Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods in enzymology* **374**, 461-91 (2003).
69. Humphrey, W., Dalke, a & Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **14**, 33-8, 27-8 (1996).
70. VMD - Visual Molecular Dynamics. at <<http://www.ks.uiuc.edu/Research/vmd/>>
71. Grell, L., Parkin, C., Slate, L. & Craig, P. a EZ-Viz, a tool for simplifying molecular viewing in PyMOL. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* **34**, 402-7 (2006).
72. pymol. at <<http://www.delanoscientific.com/>>
73. Joosten, R.P. *et al.* A series of PDB related databases for everyday needs. *Nucleic acids research* **39**, D411-9 (2011).
74. Laskowski, R. A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).
75. PROCHECK home page. at <<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>>
76. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *Journal of computational chemistry* **26**, 1701-18 (2005).
77. Gromacs - Gromacs. at <<http://www.gromacs.org/>>

10 Bibliography

78. gromacs 4.0 online reference.
at <<http://psc.edu/general/software/packages/gromacs/online.html>>
79. Grace User's Guide (for Grace-5.1.22).
at <<http://plasma-gate.weizmann.ac.il/Grace/doc/UsersGuide.html#ss1.1>>
80. grace home. at <<http://plasma-gate.weizmann.ac.il/Grace/>>
81. Larkin, M. a *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)* **23**, 2947-8 (2007).
82. Gouet, P. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**, 3320-3323 (2003).
83. ESPript. at <<http://esript.ibcp.fr/ESPript/ESPript/>>
84. Aqvist, J. & Luzhkov, V. Ion permeation mechanism of the potassium channel. *Nature* **404**, 881-4 (2000).
85. What if interface.
<http://swift.cmbi.ru.nl/servers/html/index.html>
<<http://swift.cmbi.ru.nl/servers/html/index.html>>
86. Protein, N. Quality Control of Protein Models : Directional Atomic Contact Analysis The method Fragment types. 47-60 (1993).
87. SCWRL4 - Protein Side Chain Prediction :: Dunbrack Lab. at <<http://dunbrack.fccc.edu/scwrl4/>>