University of South Bohemia, České Budějovice

Faculty of Sciences

Structure and function of cation translocation systems: modeling and simulations

Ph.D. Thesis

Vasilina Zayats, MSc.

Supervisor: Prof. RNDr. Rüdiger Ettrich PhD

University of South Bohemia, České Budějovice

and Institute of Nanobiology and Structural Biology,

Global Change Research Center,

Academy of Sciences of the Czech Republic

Co-Supervisor: Dr. Mag.Thomas Stockner, PhD.

Institute of Pharmacology,

Medical University of Vienna, Austria

Nové Hrady 2014

Dedicated to my beloved parents

This thesis should be cited as:

Zayats V., 2014: Structure and function of cation translocation systems: Modeling and simulations. Ph.D. Thesis. University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic, 126 pp.

Annotation

Ion translocation systems are the basis of any cell life from unicellular to multicellular organisms. They provide the communication of cells with the environment, maintain cell integrity, control the cell signaling, maintain the membrane potential and underlie the basis of all physiological processes. Using a combined bioinformatics and molecular dynamics approach this thesis contributes to the understanding of different aspects of structure and function of cation translocation systems, using a set of different representative systems specialized on the translocation of mono- and divalent cations. These systems are either interconnected in function or structurally related. The first studied system is a non-selective cation channel, the transient receptor potential A1 (TRPA1), a membrane protein involved in pain sensation and regulated by PIP_2. TRPA1 regulates association of the dynamic calcium signal transducers STIM (stromal interaction molecules) and the calcium release-activated calcium channel Orai (a pore domain of store-operated calcium entry) and itself associates with store-operated calcium entry proteins. Thus the second system is the human Orai1 channel. The third system is the potassium-selective translocation system Trk1 of *Saccharomyces cerevisiae*. This ion translocation system traces back in evolution to a common ancestor protein with the TRP channels that is assumed to be close to the potassium-selective channel KcSa, and its study contributes to a clearer understanding of the evolutionary relationship between the different groups of cation translocation systems. The results deepen the theoretical understanding of mechanisms behind cation translocation provided by these systems.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 21.09.2014 .....

Vasilina Zayats

This thesis originated from a partnership of the Faculty of Science, University of South Bohemia, and the Institute of Nanobiology and Structural Biology, Global Change Research Center, Academy of Sciences of the Czech Republic, supporting doctoral studies in the Biophysics study programme.

Financial support

This work was funded by the Ministry of Education, Youth and Sports of the Czech Republic (projects No. ME09062 and MSM6007665808), the Academy of Sciences of the Czech Republic (AVOZ60870520), and the Czech Science Foundation, grants P207/10/1934 and 13-21053S. Access to the computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Infrastructure for Research, Development, and Innovations” (LM2010005), is acknowledged.

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. Iryna Kishko, Balasubramanian Harish, **Vasilina Zayats**, David Reha, Brian Tenner, Dhananjay Beri, Tobias Gustavsson, Rüdiger Ettrich, Jannette Carey (2012) Biphasic kinetic behavior of FMN-dependent NAD(P)H:quinone oxidoreductase WrbA from *E. coli* PLOS One 7 (8): e43902. (IF=3.534)

VZ *performed ligand docking, participated in analysis of the docking results, figure preparation and writing of the manuscript.*

II. Vasilina Zayats, Abdul Samad, Babak Minofar, Katherine Roelofs, Thomas Stockner, Rudiger Ettrich (2013) Regulation of the transient receptor potential channel TRPA1 by its N-terminal ankyrin repeat domain Journal of Molecular Modeling 19: 4689–4700. (IF=1.867)

VZ *conceived the idea (jointly with RE* and *TS), participated in developing the computational strategies, performed sequence analysis, homology modeling and molecular dynamics simulations, participated in data analysis and in writing the manuscript.*

III. Saurabh K Pandey, David Reha, **Vasilina Zayats**, Milan Melichercik, Jannette Carey, Rudiger Ettrich (2014) Binding-competent states for L-arginine in E. coli arginine repressor apoprotein. Journal of Molecular Modeling 20 (7): 2330. (IF=1.867)

VZ *performed and analysed a subset of the molecular dynamics simulations. Participated in writing the manuscript.*

IV. Irene Frischauf*****, **Vasilina Zayats***, Michael Deix, Anna Hochreiter, Barbara Lackner, Barbora Svobodova, Christoph Romanin, Rüdiger Ettrich, Rainer Schindl (2014) Extracellular calcium sink in the Orai1 channel tunes calcium permeation and initiation of NFAT transcription. *prepared for submission. (* contributed equally)*

VZ *conceived the modeling idea (jointly with RE* and *RS), performed all homology modeling, molecular dynamics simulations, and data analysis. Participated in figure preparation and writing the manuscript.*

V. Vasilina Zayats, Thomas Stockner, Saurabh K Pandey, Rüdiger Ettrich, Jost Ludwig (2014) A refined atomic scale model of the Saccharomyces cerevisiae K+-translocation protein Trk1p combined with experimental evidence confirms the role of selectivity filter glycines and other key residues. Submitted 9/2014 to BBA Biomembranes (IF=3.431).

VZ *conceived the modeling idea (jointly with JL* and *RE), developing the computational strategies (jointly with TS and RE), performed sequence analysis, homology modeling and molecular dynamics simulations, data analysis and participated in writing the manuscript.*

Prof. RNDr. Rüdiger H. Ettrich, Ph.D.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Rudiger Ettrich for introducing me into the world of bioinformatics and molecular modeling, for his encouraging ideas and great solutions of any scientific or life's problem, for help, support and patience throughout all these years!

I would like to thank Dr. Thomas Stockner for his invaluable help with understanding membrane proteins, for showing me the way through the colorful letters of sequence alignment and for his enthusiasm that was always inspiring for me!

I would like to thank Dr. Jost Ludwig, who came with the idea that we have to build Trk1 that is either symporter or transporter (we need to find it out!) for all interesting and always funny discussions, help with understanding experimental results. Finally, I would like to thank for all discussions and enormous help during writing my PhD thesis!

I would like to thank Dr. Babak Minofar, who got me started with simulations of membrane proteins and for the help during writing my PhD thesis!

I would like to thank my colleagues Dr. David Reha and my fellow PhD student Saurabh Pandey for the friendly atmosphere during the time we shared the office and for their help with MD, docking and QM/MM, as a lot of work in my thesis was done in close collaboration with them and for the help during writing my PhD thesis!

I would like to thank all my colleagues with whom I started together: Morteza Khabiri, Natalia Kulik, Dhiraj Sinha and Zofie Sovova for all the help with modeling and simulations! Especially I would like to thank Milan Melicherchik for a lot of technical support!

I would like to thank our collaborators from the experimental laboratory from Linz for their enthusiasm and inspiring ideas, I absolutely enjoyed your extremely helpful and collaborative attitude!

I would like to thank Anamika Mishra and Kumud Mishra for the helpful and useful discussions during preparations for the biophysical and state exam and support during writing my PhD thesis!

Although not directly connected to my research work, I would like to thank my schoolteacher Borisova Natalia Andreevna who first introduced science for me and gave me a chance to present my research results!

I thank my dear friends for all happy and unforgettable moments we had together, for their support and attention, who were always with me even if the long distance was between us!

I would like to express my feeling of gratitude to all my family: my brother and sister in law, my uncles and their families, my dear grandma and parents for their love and care that I always feel and bear with me!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADP	Adenosine Diphosphate
ANK	Ankyrin
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
BLOSUM	BLOcks SUbstitution Matrix
BQ	BenzoQuinone
CRAC	Calcium-Release Activated Currents
DAG	DiAcylGlycerol
DNA	Deoxyribonucleic Acid
EM	Energy minimization
FMN	Flavin MonoNucleotide
IP3	Inositol trisphosphate
IUPHAR	International Union of Basic and Clinical Pharmacology
MD	Molecular Dynamics
MM	Molecular Mechanics
NAD	Nicotinamide Adenine Dinucleotide
NMR	Nuclear Magnetic Resonance
PAM	Point Accepted Mutations
PDB	Protein Data Bank
PH	Pleckstrin homology
PIP2	Phosphatidylinositol 4,5-bisphosphate
PLC	Phospholipase C
PME	Particle mesh Ewald
POPC	1-Palmitoyl-2-oleoylphosphatidylcholine
QM	Quantum Mechanics
SOCE	Store-Operated Calcium Entry
TRP	Transient Receptor Potential

Preface

Bioinformatic tools are nowadays widely applied to study, compare and align protein sequences, to identify conserved motifs or domains, predict secondary structure or the protein fold, and to built phylogenetic trees. Molecular dynamics allows the simulation of physical movements of atoms and molecules in a given environment, and thus brings in a dynamic component, as proteins are not static, but "kicking and screaming" (Weber 1975) which means they are sampling a large set of conformations in time. Cation translocation through the cell membrane is one of the most essential features in living cells, as appropriate intracellular concentrations of cations are of key importance for living cells, since they determine the potential across the plasma membrane, the cell volume, the intracellular pH, and many other important cellular parameters necessary for the function of all living cells. Using a combined bioinformatics and molecular dynamics approach we were eager to contribute to the understanding of different aspects of structure and function of cation translocation systems, using a set of different representative systems, that are nevertheless interconnected in function, such as membrane cation translocation systems and cytoplasmic regulatory proteins. Cation translocation systems, such as ion channels and transporters are located in the plasma membrane in all living organisms and their strict coordinated interplay is of vital importance for the cell integrity and a basis of all physiological processes. A theoretical understanding of mechanisms behind cation translocation provided by these systems is not only interesting from an academic point of view but also of crucial importance in biomedicine as most of these channels and transporters are responsible for various kinds of health disorders, consequently these proteins are main targets in pharmacology as well. A detailed understanding of protein structure and function would allow shedding light on many physiological and pathophysiological processes. Although, membrane proteins are of very high importance, they comprise only a small fraction of all solved structures in protein data bank (PDB). Therefore, the structural information about various channel and transporters is still missing, and modeling of a target protein based on the coordinates of a structural homolog is often the only chance to gain a structural insight. Given a reasonable high sequence identity, homology modeling provides a powerful tool to study proteins without yet experimentally resolved structure.

I) In the first studied system we focused on a non-selective cation channel, the transient receptor potential A1 (TRPA1), a membrane protein involved in pain sensation and regulated by PIP_2. TRPA1 consists of three main domains:

1) N-terminal cytoplasmic domain which contains 17 predicted ankyrin repeats and one putative EF-hand-like calcium binding motif;
2) a tetrameric transmembrane domain; and
3) a cytoplasmic C-terminal domain.

None of the structures of these domains has been experimentally solved yet, and we aimed for providing a structural insight into the organization of the TRPA1 N-terminal and transmembrane domains and their functions:

- to model N-terminal predicted ankyrin repeat domain (chapter 1.5.3 and 3.1)
- to model a predicted EF-hand motif (chapter 1.5.4 and 3.1)
- to investigate the effect of calcium binding to the putative EF-hand (chapter 3.1)
- to model a hypothetical half of intermolecular PIP_2-binding domain that would complementary interact with PLC gamma1 (chapter 1.5.5 and 3.2)
- to model the transmembrane domain (chapter 3.1)
- to simulate the dynamic behavior of the N-terminal domain and investigate its implications on the channel function (chapter 3.1)

Earlier predictions assigned 14 to 18 ankyrin repeats to the N-terminal domain of TRPA and proposed an EF-hand loop within this ankyrin repeat domain. The transmembrane domain of TRPA1 shows a similar fold and quaternary structure arrangement as potassium voltage-gated channels. As only in case of high sequence identity (more than 50% identity), homology modeling is straightforward, and in case of TRPA1 the sequence identity between the N-terminal domain with the template structures is at the lower border (25%) where homology modeling is possible at all, careful manual interference and curation is needed, including additional sequence analysis to identify correctly secondary structure elements. The structure of the TRPA1 TM domain was predicted for a long time based on the available crystal structure of Kv channels. We utilized homology modeling to build a 3D structure based on Kv, Nav and TRPV1 templates. The general characteristics of the TRPA1 channel are described in chapter 1.5.

II) Interestingly, several TRP channels including TRPA1 associate with store-operated calcium entry (SOCE) proteins. TRP channels can be activated by store depletion, however only under

particular conditions. TRPA1 regulates association of the dynamic calcium signal transducers STIM (stromal interaction molecules) and the calcium release-activated calcium channel (a pore domain of store-operated calcium entry) and itself associates with SOCE proteins. A general characterisation of the SOCE can be found in chapter 1.6. So, logically in the second system we focused on the human Orai1 channel. Additionally, this also gave the unique opportunity to utilize the recent crystal structure of the Drosophila Orai channel that opened the structural view on a completely structurally unknown family of channels so far. This unique possibility motivated our collaboration with the Christopher Romanin group in Linz (Austria) to combine experimental and theoretical approaches in order to investigate protein structure-function relationship. The high sequence identity (64%) between the template structure (*Drosophila* Orai) and the target protein human Orai1 allowed to obtain a corresponding hOrai1 model of a good atomistic quality, what enabled us further to focus on protein function. Using molecular dynamics simulations we aimed to investigate:

- the role of negatively charged residues at the pore entrance (chapter 3.4)
- electrostatic interactions between extracellular loops
- the selectivity of the pore for calcium ions vs. sodium ions
- identify a cholesterol binding site within transmembrane domain (chapter 3.3)

All studies were done in close collaboration with the experimental group from Linz, where in parallel a real hOrai1 channel was studied using disulphide crosslinking and electrophysiological recordings.

III) TRP channels, which are found in almost all kingdoms, as well fungal Trk1 protein trace back in evolution to a common ancestor protein that might have been close to the potassium-selective channel KcsA. The crystal structure of KcsA published in 1998 (Doyle et al. 1998) for a long time was a template structure for these very distant proteins. The early models of TRPA1 and Trk1 were built based on the KcSa channel. The Trk1 transporter inherited a potassium-selective filter with conserved glycine residues in it. The TRPA1 channel inherited a tetrameric architecture and pore forming domains but not selectivity properties. So, in our last system we focused on the potassium-selective translocation system Trk1 of *Saccharomyces cerevisiae*. The project is running in collaboration with the experimental group of Jost Ludwig in our institute. The structure of Trk1 is not known; however the fungal Trk1 protein is distantly related to prokaryotic Trk and Ktr and plant HKT proteins. Our studies on Trk1 were additionally motivated by the publication of a structure of the first protein related to Trk1 protein (TrkH). Using combined theoretical and experimental approaches we aimed to:

- obtain a three-dimensional model of Trk1 transporter satisfying experimental results
- to investigate a distinct role of conserved glycine residues of the selectively filter
- examine a structural/functional role of conserved salt bridge between first pore-domain and the last transmembrane helix

The initial model of Trk1 built on distantly related TrkH protein enabled us to design experimental studies in order to verify a predicted positions of key residues in the model. Information about Trk1 can be found in chapter 1.7 and 3.4.

The thesis organized in a following way. The Introduction part (chapter 1) describes different biological aspects of cation translocation systems from bacterial to animal cell and their role in living organisms. The general characteristics of the proteins we studied is written in chapter 1.5, 1.6 and 1.7. The evolutionary of cation channels is described in chapter 1.4. The methodology is described in chapter 2. The chapter 3 "results and discussion" summarizes the story line and discusses the main findings from the papers and additionally contains and discusses unpublished results. The thesis finishes with a short conclusion. The Appendix contains the published results including the papers submitted and ready for submission.

1. INTRODUCTION

1.1 The role of the plasma membrane in energy metabolism and ion homeostasis

Unicellular organisms and the cells of multicellular organisms are separated from the environment by special borders that help to maintain particular internal conditions that are not easily influenced by external factors. Such border in living organisms serves a plasma membrane. The plasma membrane consists of two layers of lipids, that are organized in such a way that polar headgroups of lipids are oriented toward outside and inside the cell, while hydrophobic parts (hydrocarbon chains) of both layers are oriented toward each other. Due to hydrophobic nature of lipids, this membrane is freely permeable only to lipophilic compounds. These compounds, such as steroid hormones or small uncharged molecules like CO_2, N_2 could cross the membrane according to the concentration gradient by means of simple diffusion (Berg et al. 2002). However, in order to realize complex requirements of the cell, which needs as well polar molecules to come in or out, membrane gained special proteins that fulfill such transport functions. (Fig.1) There are two main types of such proteins: channels and transporters. Transporters can mediate the translocation of molecules uphill their electrochemical gradient. Since such process is energetically unfavorable, transporters use either directly available energy sources (ATP, light, etc.) in a process that is called primary active transport, or couple the "uphill" transport of the one substance with the downhill transport of another substance (symport or antiport). Channels allow molecules to move down their electrochemical gradient, however, only when cell needs it (Berg et al. 2002; Augustine et al 2004).

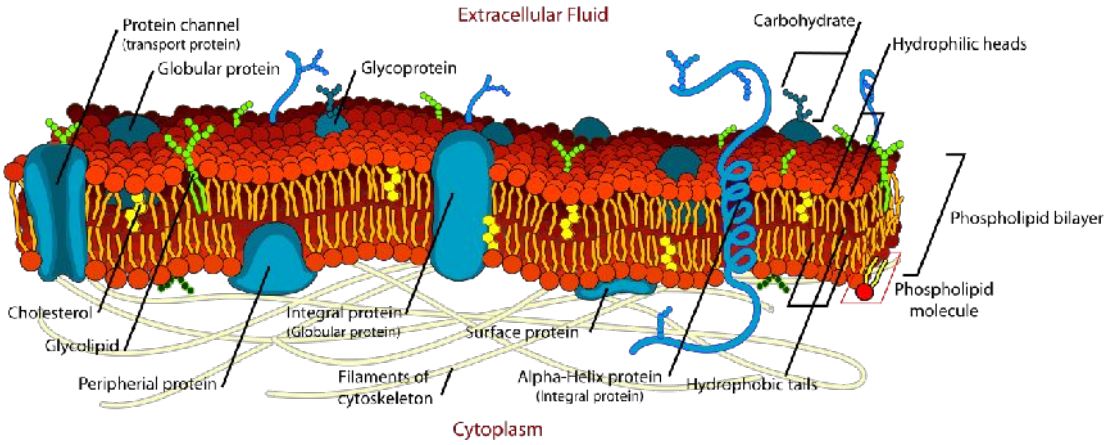


Fig.1: Biological membranes are composed out of a lipid-bilayer (consisting largely of phospholipids). Associated and integrated are proteins. The figure illustrates this "fluid mosaic model" that was originally proposed by Singer & Nicolson (1972).

Organisms from non-animal kingdoms as Bacteria, Fungi and Plants have got an additional, a mechanical protection from the outside - a cell wall. All organisms belonging to the Animal kingdom possess only a plasma membrane covered with a glycocalix (facing outside carbohydrate tails of glycolipids and glycoproteins), which offer some additional protection as well. Since, plants, bacteria and animals are organized fundamentally differently and their cells face different environments membranes had to adapt accordingly.

1.1.2 Bacterial cell

The metabolism of many bacterial cells involves the citric acid cycle (Krebs-Martin cycle) which produces reduction equivalents (NADH, FADH, etc.). The electrons from these reduction equivalents are then transferred via the respiratory chain to molecules with lower redox potentials. During this process protons are transferred across the plasma membrane and thus an (outside to inside) electrochemical proton gradient is generated. These protons can re-enter the cell via H+-ATPases and drive the synthesis of ATP from ADP to Pi (Inorganic phosphate). The driving force in that case is not only the chemical proton gradient but also the membrane potential that is generated by the transport of H+ from the inside to the outside of the cell. This membrane potential is estimated to be as high as -150 mV (inside to outside) and thus significantly contributes to the driving force for protons (the so called proton motive force (PMF) (Armstrong 2002).

In order to keep enzymes active, the cell also has to keep the ion concentration in its cytosol withi a reasonable range. The main monovalent cation that has to be present in a suitable concentration in the cytosol is the potassium ion (K^+). The plasma membrane has got special proteins, which can let selectively through only potassium ions. In order to adapt to different external conditions, where the potassium concentration may vary from µM to high mM, bacteria regulates the internal K+ by moving it in or out using potassium transporters. When potassium concentration is optimal, potassium ions just move down their gradient via selective potassium channels (Armstrong 2002).

1.1.3 Animal cell

Animal cells inherited the similar ion concentration inside the cell as bacteria with low Ca2+, Mg2+, Cl- and high K+ and preserved a negative membrane potential inside, although this is usually, much lower than that of the bacterial cell (~ -50 – 60 mV). Bacteria use cell wall for protection and plasma membrane to produce energy. The animal cell is not surrounded by a solid cell wall but only by plasma membrane, for the energy production specialized intracellular organelles – mitochondria – are present (Armstrong 2002). The inside and outside of a cell faces

aqueous solutions with different concentration of solutes. In order to keep cell integrity water should not leave or enter the cell in a big amount. The water always tends to move in a direction of less water concentration, to dissolve a solute for example. The condition when no net movement of water through the membrane takes place is called osmotic equilibrium. In order to stabilize waters, the solute concentrations inside and outside have to be stabilized as well. Such an osmotic equilibrium can be achieved when the charge on both side of membrane is equal. This is regulated via complex molecular mechanisms that involve membrane channels and transporters (Armstrong 2002).

1.1.4 Membrane potential

As mentioned already, the interior of the cell contains a high concentration of potassium ions. Other ions, such as sodium and calcium are present outside in a higher concentration. The difference in ion composition and concentrations inside and outside is called concentrations gradient. On the other hand an (inside and outside) negative membrane potential exists. Both, concentration gradient and membrane potential contribute to the direction of the driving force for a given ion (the electrochemical gradient for that ion) according to the Nernst equation. According to the Second law of thermodynamics, the ions have to move down their electrochemical gradient to reach an electrochemical equilibrium on both sides of the membrane. This equilibrium however is usually not reached because of the work of ion translocation proteins that use energy to establish different ion potentials across the membrane. The most important ion translocation system in animal cells is the Na/K-pump, which transfers (by utilizing ATP as energy source) three sodium ions outside the cell and bring two potassium ions inside in turn. Due to this process, also a difference in electrical potential across the membrane is generated leading to a membrane potential which is in the range of -40 to -90 mv (inside to outside). The charge difference exists because the negatively charged molecules (such as chloride ions, phosphate groups or some amino acids) are still present inside the cell constantly. Such a difference in charge on the sides of membrane is called **resting membrane potential**.

This membrane potential is relatively constant in somatic cells. However, in excitable cells (like in the brain, nerves and muscles) electrical signals have to be generated and transmitted. Therefore the resting state of the cell is interrupted in case of excitation. This involves a depolarization caused by the influx of sodium ions. Depolarization of membrane is a basis of **action potential** and nerve signal transfer. The local depolarization of membrane causes the depolarization at the neighboring region and so on. In order to speed up the transfer of signal the long extensions of the nerve cells,

the axons, often possess myelinated membranes. In such cells, the signal “jumps” over myelinated (thus electrically almost completely insulated) regions between the nodes of Ranvier, therefore signal conduction is much faster than in unmyelinated nerve fibers (Augustine et al. 2004).

The electrochemical gradient of K^+ is usually almost in equilibrium in animal cells. For Na^+ however, a gradient from outside to inside exists. The Na^+/K^+ pump keeps the concentration of these ions constant in resting conditions. However, when excitable cells are excited by some factor sodium channels are activated and allow the influx of Na^+ what leads to a depolarization. When this depolarization reaches a certain the threshold value, voltage-gated potassium-selective channels open, sodium channels are inactivated and the initial condition are slowly restored, bringing the cell back to the resting potential.

1.2 The role of transporters and channels

The whole process from resting to action membrane potential is possible due to complex combination of membrane permeability to different ion species. As it was mentioned already the membrane has got two types of proteins that are responsible for membrane potential. The first group is ion transporters; they establish concentration gradient of ions across the membrane. The second group comprises ion channels which control the ion permeability.

1.2.1 Transporters

In order to keep particular ion concentrations inside the cell, transporters might have to use energy to move ions in and out irrespective to the electrochemical gradient of ions. Transporters are divided into two classes on a basis of energy source they use to conduct ions. Some transporters rely on the energy source stored in ATP molecules, such as Na^+/K^+ ATPase, which transfers three sodium ions outside and two potassium ions inside by expenses of one ATP molecule (which is under standard conditions 30 kJ/ mol of energy). Two different types of binding sites are available in this transporter, which distinguish ions by their size. When ATP molecule is bound, three Na^+ ions are carried outside, making the channel to open potassium binding site on another side of membrane. When two K^+ are bound, the phosphate leaves and the protein converts to back to it's “initial” conformation, again, exposing the Na^+ binding sites intracellularly. The other type of transporters that use energy to translocate substances against their electrochemical gradient are called secondary active transporters. These proteins transport two substances simultaneously. While one substance is transported “down” its electrochemical gradient, the other substance is transported against its gradient. Thus, the electrochemical gradient of one substance is used to trasnport another

substance. This process can happen in two forms. When both substances are moved in the same direction it is called a symport, the protein is symporter. When the substances go to opposite directions, it is called antiport, the proteins are antiporters.

1.2.2 Channels

The squid axon extensively studied with voltage-clamp techniques revealed different properties of resting membrane potential and action potential. Voltage-clamping allows measuring the current across the membrane of the whole neuron. Later the patch-clamp technique was developed which enabled analysis of single channel current. Both techniques gave complementary results showing that combined single channel behavior during action potential corresponds exactly to the overall respond of the whole cell. Therefore indicating that interplay of sodium and potassium channels is responsible for the action potential formation (Augustine et al. 2004).

Other types of channels exist, which are responsible for conductance of other ions, such as calcium, magnesium, chloride and other ions. Channels can be activated by variety of different factors. Based on the nature of channel activation, they are combined in several groups: a) voltage-gated b) ligand-gated c) mechano-gated channels. Additionally, so called “leak channels” exists that are always open.

a) Voltage-gated ion channels are mostly involved in generation of action potential, therefore in order to give an accurate response they are usually selective for particular ion type. Many other types of ion channels can be selective to more than one ion species or conduct e.g. all types of cations unselectively. Such ion channels bear different function that action potential generation. They can be activated by various factors, mostly by the binding of ligands (ligand-gated) or mechanical stimuli (mechano-gated channels).

b) Ligand-gated ion channels possess ligand binding site either on the intracellular or the extracellular side. They open when a particular compound is bound. This class includes channels binding neurotransmitters, which are important for transfer information between neurons. Other ligand-gated channels are involve in cell-cell signaling (Augustine et al. 2004).

c) The class of mechano-gated channels are essential in responses to cold or hot temperatures, to skin stretches, or light. These channels are activated by variety of mechanical stimuli and convert these into electrical signal of ions which can be recognized by cells (Augustine et al. 2004).

1.3 Structural characteristics of cation channels

1.3.1 The structure of a K+ channel

The first information about the structure of ion channels came in 1998 when the crystal structure of a K^+ channel (with 3.2 Å resolution) from the bacterium *Streptomyces lividans* was published (Doyle et al., 1998). The K^+ channel gene encodes one subunit which consists of two transmembrane helices and one short pore helix (Fig.2). The whole channel is a homo-tetramer of four subunits. The overall structure from outside resembles a cone with upper part being more wide than lower. The inside channel forms a pore, which possess several "rooms". The narrowest "room" of the channel is selectivity filter of 12 Å length located at the extracellular side of the channel, which allows passing only K^+ ions. This selectivity filter is formed by short loops of five residues (T-V-G-Y-G) which restricts the diameter of the pore to approximately 3 Å. This five-residue sequence highly conserved throughout the potassium channels and transporters. Especially the glycine residues are more or less invariant in K+-transporters and channels from bacteria to human. K+ ions are coordinated by backbones atoms of selectivity filter. The widest part of the channel is called central cavity. It is composed of hydrophobic residues (Doyle et al. 1998).

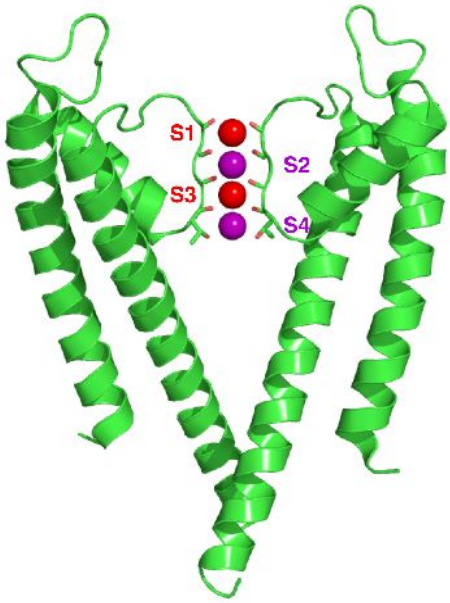


Fig.2: The structure of a potassium channel. Shown is the structure of the bacterial KcsA potassium channel (PDB 1K4C; Zhou et al. 2001). The selectivity filter (consisting of the co-ordination sites S1-S4) residues is drawn as stick model. Only 2 of the 4 subunits are shown. Within the selectivity filter 2 K^+ ions (purple) and 2 oxygen atoms from water (red) are shown.

1.3.2 The structure of voltage-gated K+ and Na+ channels

The crystal structure of mammalian voltage-gated potassium channel in open conformation of 2.9 Å resolution was published in 2005 (Long et al. 2005). The overall structure of the ion conducting pore is similar to bacterial potassium channel described above (Fig.2) with the identical composition of K+ selective filter. In addition to the pore, voltage-gated channel gains an additional membrane domain, the voltage-sensor. Each monomer of such channel is composed of: *a)* pore-forming elements: two transmembrane alpha helices (S5 and S6, short pore helix and selectivity filter loop) and *b)* four transmembrane alpha helices (S1 to S4) forming the voltage-sensing domain (Long et al. 2005).

The structure of voltage-gated sodium channel from *Arcobacter butzleri* (NavAb) was solved in closed conformation with 2.7 Å resolution (Payandeh et al. 2011). As it was predicted earlier based on sequence similarity, the structure of voltage-gated channel preserved general architecture of potassium voltage-gated channel (Fig.3). The main difference between channels lies within the selectivity filter. The sodium channel selectivity filter is 4.6 Å long, which is shorter than in potassium channels. The diameter is in contrast wider, which allows passing partially hydrated sodium ions. The narrowest site of the filter is formed by side chains of glutamate residues, which directly interact with sodium ions (Payandeh et al. 2011).

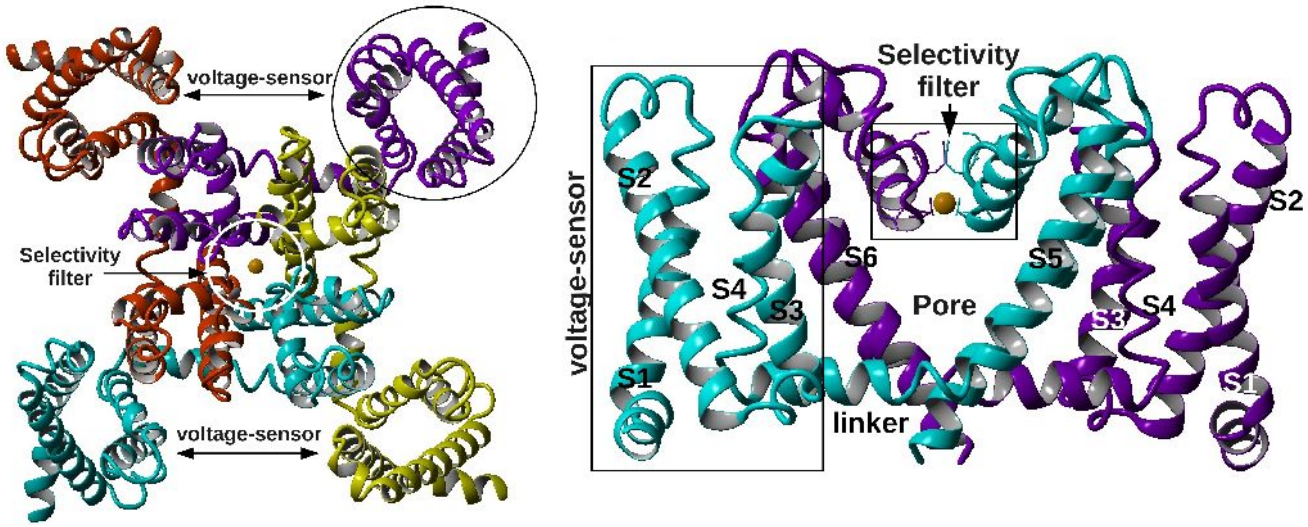


Fig.3 The structure of voltage-gated sodium channel from *Arcobacter butzleri* (NavAb). Top and side view. Figure prepared in YASARA.

1.3.3 The mechanism of cation selectivity

The crystal structure of KcsA revealed two K^+ bound in selectivity filter with the distance of 7.5 Å between them (Doyle et al. 1998; see also Fig.2). This allowed hypothesizing about the mechanism

of ion conduction through the channel. The backbone oxygens from the selectivity filter can bind a dehydrated K^+ ion. However, when the single ion is present in the filter, the binding energy is high enough to keep ion within the selectivity filter. If a second K^+ reaches the selectivity filter, the first ion moves due to repulsion forces between the first and the second K^+ (Doyle et al. 1998; see also Berg et al. 2002).

Potassium channels selectively let through K+ ions with the permeability being approximately 100-fold higher to K^+ than to Na^+ (Berg et al. 2002). Since both ions bear single positive charge, the selectivity bases on the difference in ionic radii and dehydration energy. The free energy required to dehydrate a potassium ion equals 203 kJ x mol^{-1}, 301kJ x mol^{-1} is needed to dehydrate sodium. The role of water oxygens of the hydration shell of cations play the oxygen atoms of the backbone or sidechains of residues within the selectivity filter. Dehydrated K^+ fits tightly to the filter as the filter radius coincides with the sum of ionic radii (1.4 (Oxygen) + 1.33 (Potassium); Shannon 1976). The sodium ion has a smaller radius (ionic radius for Na^+ is 0.95), therefore dehydrated Na^+ is too small and cannot be stabilized by backbone oxygens of the selectivity filter (Berg et al. 2002 and Fig.4).

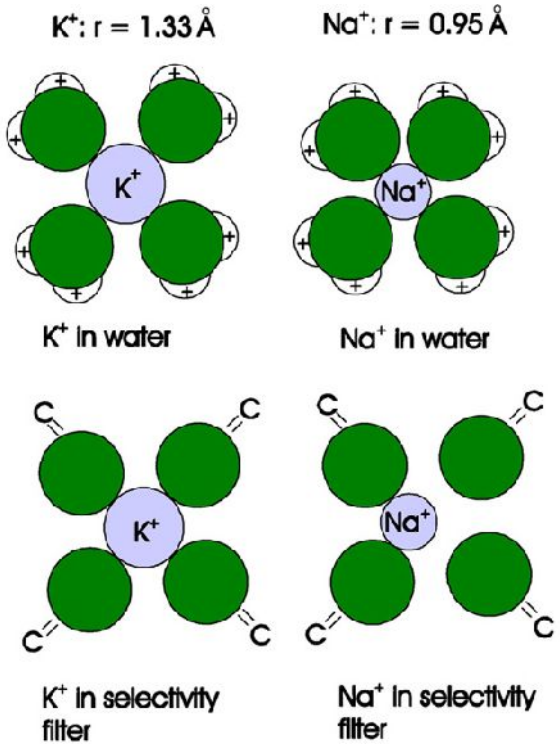


Fig.4. Mechanism of K+/Na+ discrimination in the selectivity filter of K channels. Shown are the hydrations of K^+ (left) and Na^+ (right) in water (top) and the coordination of these ions the selectivity filter of a K channel (bottom). Whereas K^+ is perfectly coordinated by oxygens, Na^+ does not "fit" due to its smaller ion radius. Reprinted with permission from AAS from Armstrong (2003).

1.4 Evolution of cation channels

Multiple phylogenic analyses based on known sequences and structures predict the first potassium channel to be an ancestor of known cation channels (Fabio Franciolini et al. 1989; Yu & Caterrall, 2005; Fig.5). Evidences decoding the secrets of cation channels evolution are increasing. Nowadays X-ray structures of potassium KcsA and voltage-gated channels are solved, and already along with wild types, mutant structures are available in protein data banks (PDB). In parallel with experimental approaches theoretical studies are conducted in order to understand structure-functional relationships.

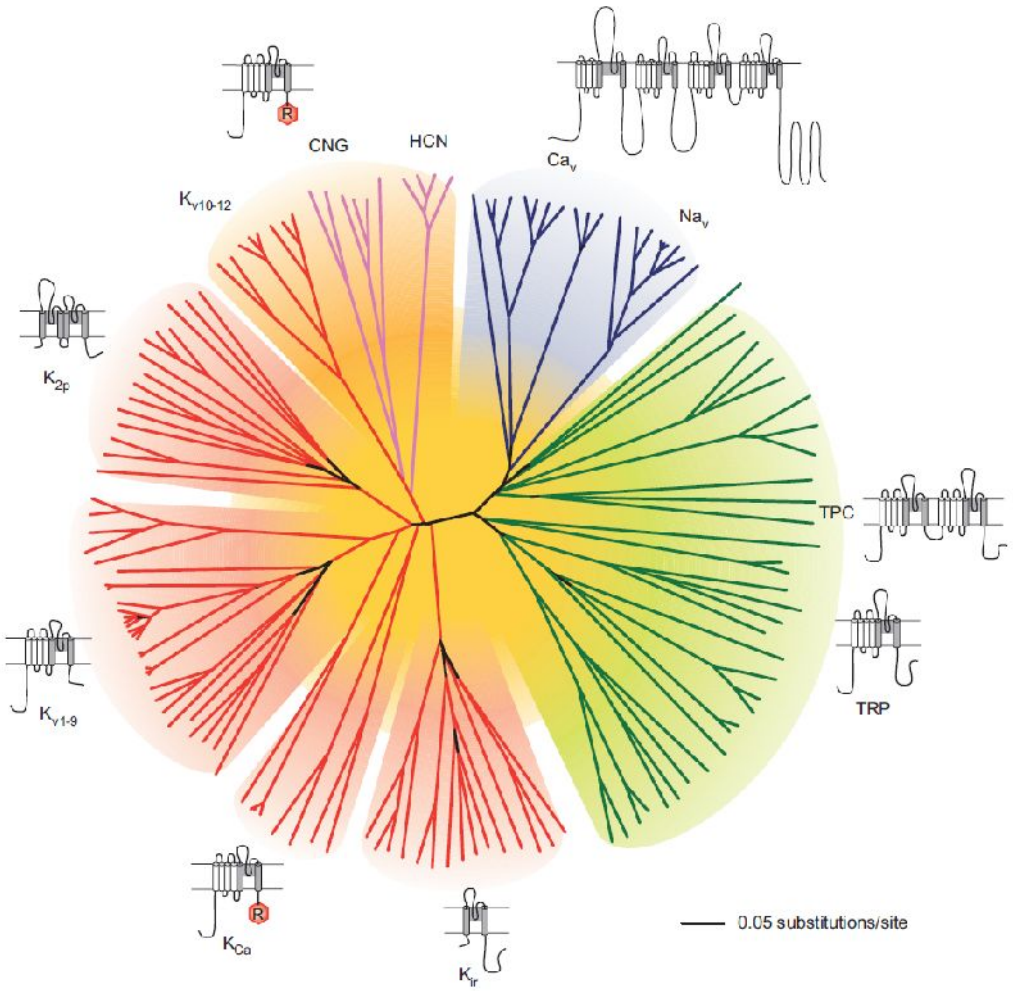


Fig.5: Evolutionary relations between different types/classes of cation channels. Reprinted with permission from AAS from Yu & Catterall (2004; Fig.1)

In 1999, Durell and colleagues published the hypothesis of an evolutionary relationship between transporter families (prokaryotic Trk and KtrAB, fungal Trk1,2, wheat HKT1) and the bacterial potassium channel (Durell et al. 1999). Based on sequence analysis and structural information of a single KcsA structure solved by that time, they concluded that transporters consist of a four-fold

repeat of the single transmembrane helix - pore – single transmembrane helix (MPM) motif of the bacterial potassium channel (Fig.6). The alignment also pointed on a single Gly residue within the pore preserved among potassium channels and transporters. This led them to the conclusion that on the early steps of evolution only the protein consisting of a single MPM motif existed (modern KcsA). Later, two gene duplications caused the emergence of a new proteins with repeated four MPM motives within a single domain (gene, monomer), as it is found to be in ion transporters (Durell et al. 1999).

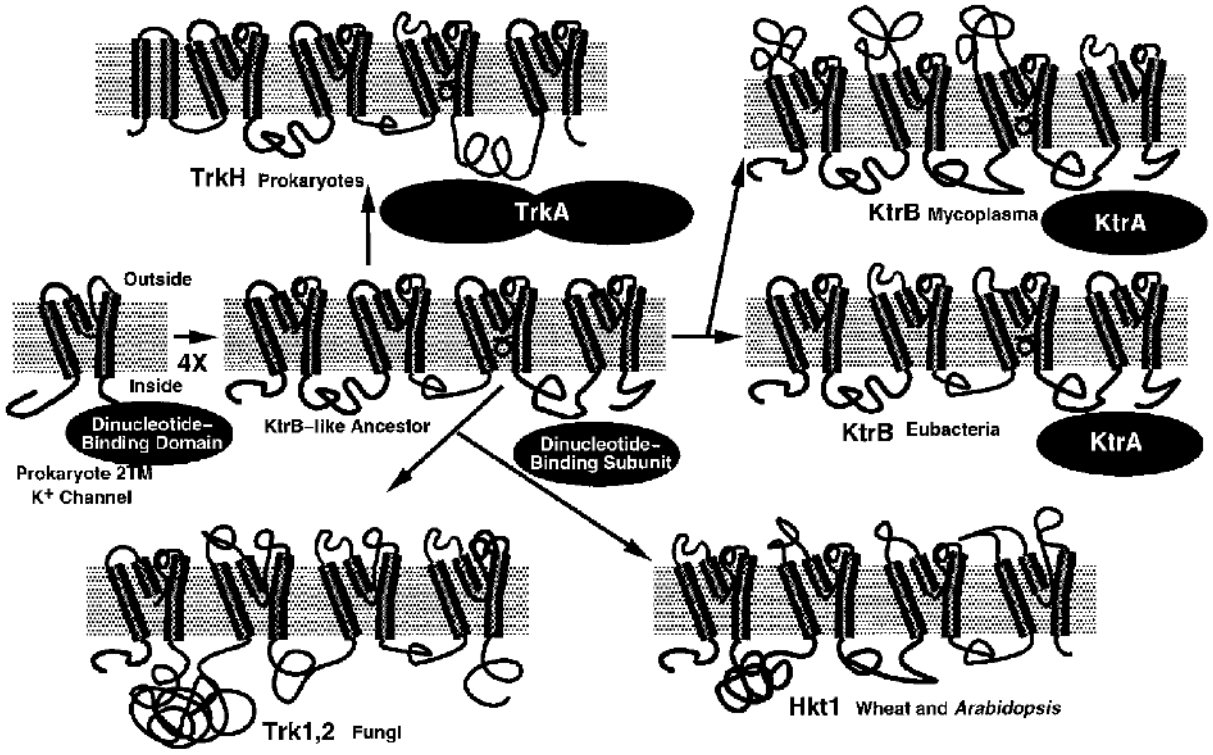


Fig.6: Possible evolution leading from an ancestral K-channel to the K translocating systems TrkH, KtrB, Hkt1 and Trk. Figure reprinted with permission from Elsevier from Durell et al., (1999, Fig.3).

Transporters are not the only group of membrane proteins predicted to be evolved from potassium KcsA channels. During the subsequent processes of genetic transformation voltage-gated cation channels must have appeared. These encompass diverse proteins, which can be grouped based on their preference for different cations. The most abundant are K^+, Na^+, $Ca2^+$ and non-selective voltage-gated channels. Their basic structure consists of six transmembrane helices, where first, second, third and especially fourth form a voltage-sensing domain and fifth and sixth helices form the ion conducting pore. Initially, with limited sequence information available, it was predicted that voltage-gated channels evolved from KcsA based on similarity of pore forming elements (domains five and six) (see. e.g., Jan & Jan 1992; Strong et al. 1993, Derst & Karschin 1998; Plummer &

Meisler 1999 and Fig.5). Although, in general the "phenotypes" of Kv and Nav/Cav are similar, the genetic information significantly differs. The voltage-gated potassium channels are tetramers, each monomer consists of transmembrane domain (TM domain) of six helices. In contrast, Cav and Nav channels are monomers already containing four times repeated TM domain in their primary amino acid sequence. Interestingly, every single repetition of the 6-TM domains from Nav/Cav channels is similar to a Kv monomer (Anderson et al. 2001). Also, it was observed from sequence alignments, that the first and third domain of Nav/Cav channels are more similar to each other compared to domains two and four. This is also true for the second and fourth domains (Strong et al. 1993). Similarly, as in the case of transporters, here also a gene duplication processes took place. The first event created a two-domain gene (first and second domain), while in the next event the existing two-domain gene was duplicated again. This sufficiently explains why first and third domains are more similar to each other then to the second and fourth domains, and vice versa (Anderson et al. 2001).

Although, the pore forming elements of voltage-gated channels and KcsA channels are similar, the former have acquired a voltage-sensing domain during evolution, which contains additional four transmembrane helices. Therefore the next question was how and when this element appeared in the long way of life development. The answer to this question arose when structures of voltage sensing phosphatases (Ci-VSP, Murata et al. 2005) and voltage-gated proton channels and Hv1 (Ramsey et al. 2006) were solved (Matsuda et al. 2011; Fujiwara et al. 2013). These proteins itself contain domains that are similar to the S1-S4 voltage-sensors of voltage-gated channels but do not contain the MPM motif from ion channels and transporters. Therefore the latter could have appeared in early steps of evolution due to gene fusion of ancestors of voltage-sensitive protein with four transmembrane helices and potassium channels with two helices forming the pore.

Supporting the theory that cation channels evolved through gene duplication and fusion is the existence of "intermediate links": Proteins, consisting of different combination of pore forming domains and voltage-sensors. For example, KCNK potassium channels possess subunits with two repeats of the pore forming MPM motif (tandem pore domain potassium channels). Another example is the voltage-gated potassium channel Tok1 from *Saccharomyces cerevisiae*, which contains one "full" transmembrane domain consisting of six transmembrane helices (like voltage-gated K-channels from higher eukaryotes) and additionally a domain two transmembrane helix MPM motif. A possible intermediate between the voltage-gated Kv channels and Nav/Cav channels, TPC proteins (Ishibashi et al. 2000; Calcraft et al. 2009; Morgan et al. 2011) are endosomal/lysosomal voltage-gated Nav/Cav-channels (Wang et al. 2012) with two repeated "copies" of a "Kv" monomer, while (as mentioned before) "standard" Nav/Cav have got four

repetitions (cf. Fig.5).

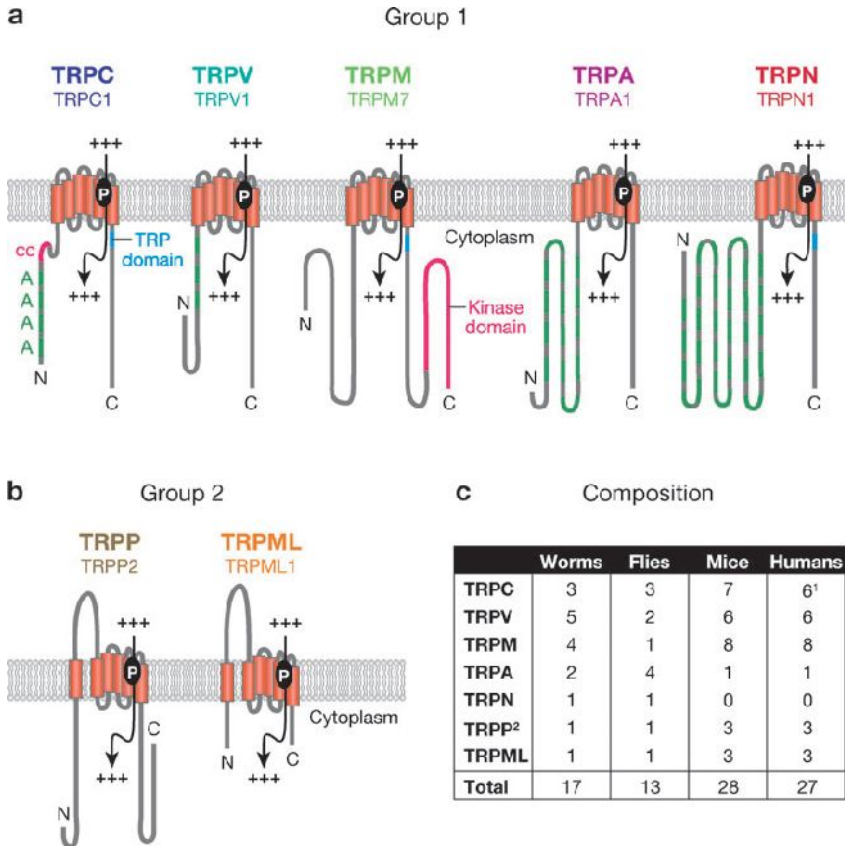
Another important question is when the change in ion selectivity occurred (Anderson & Greenberg, 2001). If Nav/Cav channels appeared due to gene duplication, then they must have been evolved from proteins consisting of single MPM monomers, but with altered ion selectivity. Taking into account the diversity of single monomer channels, particular Kv, it could be assumed, that the Kv ancestor underwent multiple mutations, which produced a huge variety of Kv channels and on the other hand, mutations in the pore which still produced functional channel but with alternated selectivity resulted in various groups of non-selective, monovalent and divalent cation-selective channels. The cyclic nucleotide-gated channels (CNG channels) were discussed as a possible candidates for being Nav/Cav channels ancestor based on information that they bear a voltage sensor and can conduct divalent and monovalent cations (Anderson et al. 2001; Gauss et al. 1998; Krieger et al. 1999)

1.5 The superfamily of transient receptor potential channels (TRP)

TRP channels share the 6 transmembrane domain motif and the composition out of four subunits with voltage gated K channels and cyclic nucleotide-gated channels. They are found in many organisms including fungi, worms, flies and mammals (Montell, 2005; Venkatachalam & Montell 2007) The name "transient receptor potential originates" from a *Drosophila* mutant in which photoreceptors after stimulation reacted only with a "transient", i.e. inactivating membrane current ("receptor potential") instead of the non-inactivating current in wild type photoreceptors (Cosens & Manning, 1969). The TRP gene was cloned in 1989 and found to encode a membrane protein of 1275 residues, which is present in photoreceptor cells. (Montell & Rubin 1989) Later, it was shown that *Drosophila* TRP is indeed an ion channel which is required for phototransduction function in flies by regulation of $Ca2^{+}$ entry. (Hardie & Minke, 1992). Subsequently, *Drosophila* TRP initiated a large number of studies in order to characterize the protein responsible for store-operated calcium entry (SOCE) (for review see e.g., Montell 2005; Venkatachalam & Montell 2007)

The superfamily of TRP channels consists of two large, distantly related groups (Fig.7). The first group comprises proteins with high sequence homology to the first identified TRP – *Drosophila* TRP and is subdivided into the families TRPC, TRPV, TRPM, TRPA, TRPN. Drosophila TRP is the founding member of the TRPC ("canonical") family. The Second group includes the TRP proteins with low homology to classical TRP, TRPP and TRPML (Montell 2005; Venkatachalam & Montell 2007). Out of these seven families only a member of TRPN was not found in humans. As mentioned above, **TRPC** refers to the first (Canonical or Classical) TRP member. This family

consists of seven members (TRPC1 – TRPC7). The **TRPV** family got its name from a **v**anilloid compound, capsaicin, the main pungent substance from chili pepper. There are six vanilloid TRPs identified so far: TRPV1 – TRPV6. Another group of TRPs, known as "melastatin" (**TRPM**) was originally related to melanoma aggressiveness. (Hunter et al. 1998; Duncan et al. 1998). This family consists of eight members: TRPM1 – TRPM8. The **TRPA** family has only a single member, the TRPA1 channel. "A" stands for ankyrin, due to unusual high number of ankyrin repeats in the cytoplasmic N terminus of the channel. **TRPN** is the only member of the family, so far found in worms, frogs and zebrafish. It is required for mechanotransduction. (Duggan et al. 2000). Mutations in **TRPML1** (originally also dubbed mucolipidin by Bassi et al 2000), the first member of the TRPML family have been found to be associated with mucolipidosis (Bassi et al 2000; Sun et al. 2000; Bargal et al 2000).. This family comprises of three members (TRPML1, TRPML2, and TRPML3, Clapham et al., 2003, Clapham, 2003). There are also three members of the **TRPP** family, mutation of which causes polycystic kidney disease. Therefore these proteins are also known as polycystins (Clapham et al. 2003; Sharman et al. 2013). The **TRPY** proteins from yeast and other fungi (Palmer et al. 2001, Zhou et al. 2005) mentioned earlier is a distant relative of other TRPs and therefore in not included in either group.



	Worms	Flies	Mice	Humans
TRPC	3	3	7	6[1]
TRPV	5	2	6	6
TRPM	4	1	8	8
TRPA	2	4	1	1
TRPN	1	1	0	0
TRPP[2]	1	1	3	3
TRPML	1	1	3	3
Total	17	13	28	27

Fig.7. The TRP family of ion channels. Figure taken from Venkatachalam & Montell (2007, Fig.1). Reprinted with permission from Annual Reviews.

TRP channels are diverse properties that are often involved in sensory physiology. They are activated by various physical and chemical stimuli. TRPs gained variant cation selectivity, ranging from highly Ca2+ -selective to non-selective channels. They are involved in multiple physiological functions such as hearing, vision, smell, touch. (reviewed e.g. in Montell 2005, Venkatachalam & Montell 2007). TRP channels attracted lots of attention due to their involvement in pain physiology and diseases. Pain sensation is a protective mechanism that informs the organism about harmful situations. Pain sensation has been developed into specialized mechanism which starts by nociceptors – unspecialized ("free") nerve cell endings (Augustine et al. 2004). Nociceptors bring information from the periphery to the particular department of the spinal cord via dorsal root ganglions (Augustine et al. 2004). There are two types of nociceptors: Ad-fibers and C-fibers. Ad-fibers are myelinated, therefore they can pass information relatively fast in contrast to C-fibers which are unmyelinated. The latter are responsible for long-lasting pain. These two types of pain are known as "first" and "second" pain accordingly. There are special receptors associated with Ad and C fibers, which are activated by noxious stimuli. Many members of the TRP superfamily were found in nociceptors. The first identified was TRPV1, responsible for sensation of temperatures about 45 C and capsaicin, a vanilloid compound from chilli pepper, which produces burning feeling. TRPV1 exists in both types of nociceptors – Ad- and C- fibers. Another member, TRPV2 found only in "fast" fibers is activated by higher temperature, 52°C. (Augustine et al 2004) TRPV3 and TRPV4 responding to unnoxious body temperatures (30° to 39° C and >25°C, accordingly) are expressed in keratinocytes and dorsal root ganglia, allowing the skin to "understand" non-dangerous temperatures (Wang & Woolf 2005). Other TRPs, TRPM8 and TRPA1 are also temperature-sensitive. TRPM8 is responsible for sensing cool temperatures in a range of 23 to 28 °C, while TRPA1 is a so called "cold" receptor; it is activated at temperatures below 18 °C (Wang & Woolf 2005) TRP channels can be activated either directly by noxious stimuli, like temperature, or they can be activated by chemicals released from damaged cells. Interestingly, some TRPs are known to produce painful feeling in response to food compounds, for example, TRPV1 to capsaicin from chilli pepper, TRPA1 to allicin from mustard. It seems likely that these compounds share structural features with substances released from injured cell.

1.5.1 General characteristics of TRPA1

TRPA1 is a non-selective cation channel encoded by *TRPA1* gene with a molecular mass of 127.4 kDa in human. The exact number of residues varies between species: 1119 in human, 1125 in rat and mouse, 1197 in fly (Garcia-Anoveros & Nagata 2007). It is expressed in excitable cells:

nociceptive neurons of peripheral ganglia, such as dorsal root, trigeminal and nodose ganglia, in sensory epithelia of inner ear; as well as in non-excitable cells like keratinocytes. TRPA1 is activated by multiple stimuli including cold temperature, mechanical stimuli, exogenous and endogenous chemicals (Garcia-Anoveros & Nagata 2007). TRPA1 consists of a cytoplasmic N-terminal ankyrin repeat domain, transmembrane domain and a cytoplasmic C-terminal domain.

1.5.2 The role of TRPA1 in pain response and diseases

Among the TRP channels TRPA1 has attracted special attention due to its crucial role in pain caused by various diseases of almost all organs and direct involvement of the channel in inflammation process. (Fanger et al. 2010; Nilius et al. 2012).

Familial Episodic Pain Syndrome, shown as episodic debilitating upper body pain, was related to a single mutation within TRPA1 channel. Gene sequencing revealed that the exchange of asparagine 855 to serine is responsible for the disorder. The mutation led to five fold increased inward currents upon activation at resting membrane potential (Kremeyer at al. 2010; Nilius et al. 2012)

TRPA1 is highly expressed in dorsal root sensory neurons. Under various pathophysiological conditions patients suffer from hypersensitivity caused by mechanical stimulations. Blockage of TRPA1 here helps to decrease the pain (Nilius et al. 2012). Spinal TRPA1 is also involved in formation of neurogenic inflammation. It was demonstrated that blockage of the receptor reduces such type of inflammation (Nilius et al. 2012; Pertovaara & Koivisto2011). TRPA1 expressed in trigeminal nerve which innervates face, was related to the tooth sensation (Nilius et al. 2012). A role of TRPA1 in cardiovascular and respiratory diseases was also reported. TRPA1 is activated by various endogenous compounds which provoke the inflammation process during asthma. Inhibition of the channel in asthma patients plays a key role in treatment of this disorder (Nilius et al. 2012).

TRPA1 also plays an important role in the generation of painful inflammation response. It was proposed that TRPA1 does not only directly signals about damage via sensory neurons, but also indirectly signals damage to immune cells. Immune cells are activated in response to vasodilation and other inflammatory factors produced upon released of neuropeptides and neurotransmitters due to local neuronal activity. On the other hand, immune cells will subsequently induce TRPA1 activation by production of multiple TRPA1 agonists, which in turn will help to keep the inflammation process active (Fanger et al. 2010; Nilius et al. 2012).

Moreover, TRPA1 is also responsible for irritating sensation – itching. Particularly, histamine-independent itching can be treated by inhibition of TRPA1 receptor. (Nilius et al. 2012).

Among the thermosensistive TRPs TRPA1 is known as a “cold” receptor. It is activated by

temperatures lower than 15-18 °C and produces painful feeling (Nilius et al. 2012).
These all demonstrates the critical importance of TRPA1 in production of pain of different nature. Therefore nowadays TRPA1 is a main pharmacological target for analgesic and anti-inflammatory drugs.

1.5.3 Cytoplasmic N-terminal ankyrin repeat domain

Many TRP channels possess several ankyrin repeats (up to eight) in their N-terminal domain. TRPA1 and TRPN1 channels however have even more ankyrin repeats (29 in TRPN1). TRPN1 is involved in mechanosensation in flies, fishes and nematodes. Due to its also large number of ankyrin repeats it was suggested that TRPA1 might be also involved in mechanosensation in mammals (Montell 2005).
The ankyrin repeat was first recognized in the yeast Swi6/Cdc10 protein and in Notch, a signaling protein of *Drosophila* (Mosavi et al. 2004) It was named ankyrin after it was characterized in ankyrins, proteins functioning as adapters in a cell (Sedgewick & Smerdon 1999). Each ankyrin repeat is a short motif of 31-33 residues. The three-dimensional structure of a single repeat consists of two anti-parallel helices connected by beta-hairpin or a long loop. Ankyrin repeats are well recognized by their consensus sequence: [G-(X)-TPLH-(X)-A-(X3)-G-(X7)-LL-(X2)-GA-(X5)] (Michaely et al. 2002). It was shown that residues of consensus sequence are mostly located at the inner part of repeats and therefore responsible for a correct folding. In contrast, residues on the surface which are involved in protein interaction are highly solvent accessible and not conserved. The studies of idealized short sets ankyrin repeats (from one to four) demonstrated that proteins of three and four repeats are well-folded and highly thermostable, while a single repeat does not fold at all and the structure of two repeats showed only limited folding (Mosavi et al. 2002; Mosavi et al. 2004). The chain of repeats is stabilized by inter-ankyrin hydrophobic interactions and additionally by hydrogen bonds between beta-hairpin/long loops of nearby repeats. The two helices within a single ankyrin repeat are of different length. Therefore, when several repeats are joined together the final construction is not straight but adopts a curved shape. The shape of the curve varies and depends strongly on the type of side chain of residue 10 in the first helix (Mosavi et al. 2002; Mosavi et al. 2004).
The ability of TRPA1 and TRPN1 to sense mechanical stimuli was ascribed to a long chain of ankyrin repeats. In a study of M. Sotomayor and colleagues (2005) two possible candidates for mechanotransduction were studied: cadherin and ankyrin repeats. Both structures consist of multiple repetitions of the same motif. Cadherins (that are important in cell adhesion, for review see

e.g. Oda & Takeichi, 2011) consist of antiparallel beta-strands (Shapiro et al. 1995) while ankyrins have the helical structure described above. It was demonstrated that ankyrins are able to respond to a small applied force by elongation without losing secondary structure and revert to the initial structure after the external force is removed. The observed behavior corresponds well to the properties of mechanical springs. The overall structure of an ankyrin repeat domain has a supercurved structure (see above), the elongation involved the change in the curvature without changing of secondary structure of a single motif. In contrast, the beta-stranded cadherins were not able to respond to the same small force. Therefore due to their elastic properties namely ankyrin repeats are considered to work as a molecular spring (Sotomayor et al. 2005).

1.5.4 N-terminal predicted EF-hand-like motif

TRPA1, as other members of TRP superfamily, is potentiated and inactivated by intracellular calcium ions (Wang et al. 2008). Patch-clamp experiments conducted in both Ca^{2+}-free conditions and in various $[Ca^{2+}]_{in}$ concentrations, revealed a direct activation of TRPA1 by Ca^{2+} (Zurborg et al. 2007). In order to identify the mechanism of calcium-mediated activation of TRPA1, it was studied, whether the Ca^{2+}-activation of TRPA1 was Calmodulin-dependent. However, despite the important role of Calmodulin for many TRP channels, the TRPA1 calcium-dependent activation retained unaffected by a mutated version of Calmodulin which is unable to bind Ca^{2+} (Zurborg et al. 2007). Several studies demonstrated that calcium activates TRPA1 by docking to EF-hand-like motif, a single mutation within a loop of EF-hand abolished channel activation (Doerner et al. 2007; Zurborg et al. 2007).

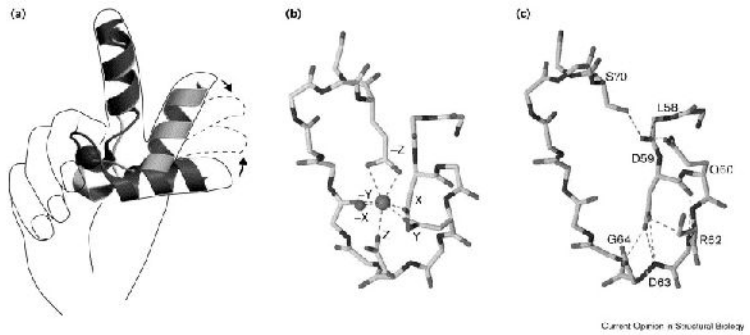


Fig.8. The Ef-hand, a common Ca^{2+}-binding motif in proteins. Reprinted with permission from Elsevier from Lewit-Bentley & Réty (2000)

The EF-hand (Fig.8) is a very common calcium-binding motif found in various proteins, such as calmodulin, tropominC, S100, parvalbumin, spectrin, calretinin (Nelson & Chazin 2009) The classical motif consists of two helices, named “E” and “F”, and the calcium-binding loop between them (Creighton 2011). The structure itself is schematically represented as a hand where helices E

and F are fingers (forefinger and thumb respectively) which hold a 12 amino acid string for coordination of calcium ion. Due to the particular arrangements of the "fingers" the curvature of the "string" provides seven oxygen atoms (both backbone and sidechain) for chelating Ca^{2+} with the resulting octahedral arrangement of the binding pocket (Creighton 2011).

The canonical EF-hand can be identified by its consensus sequence, where the residue at each position of the motif has a particular function. The most conserved region of the motif is the calcium-binding loop. The positions 1, 3, 5, 7, 9 and 12 are needed to bind ion, they are named as X, Y, Z, -Y, -X, -Z in octohedral space. The two helices preserve only non-polar residues which have to be on an interface between E and F helices. The classical EF-hand usually exists as a pair of helix-loop-helix motifs (Creighton 2011; Finn et al. 2014). However, except of the classical version, also **pseudo EF-hands** and **EF-hand-like** motifs have been described (PFAM database). The pseudo EF-hands contain a 14-residue loop where only backbone oxygens are involved in ion binding. The EF-hand-like structures have a different arrangement of E and F "fingers" in comparison to classical one (Finn et al. 2014).

1.5.5 Regulation of TRPA1 by PIP2

There are three classes of lipids which comprise biological membranes: phospholipids, glycolipids, sterols. The composition of membranes varies depending on cell type. However, phospholipids usually supply the largest fraction (Lodish et al. 2004). Among the various types of membrane phosphalipids (phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, phosphatidic acid) usually more than 50% are phosphatidylcholines. Phosphatidylcholine and sphingolipids play structural roles in the membrane; others are also involved in alternative cell functions. Particularly, phosphatidylinositol (4,5-bisphosphate, PIP2), which contributes only 5% of membrane lipids, takes part in signaling processes (Qin 2007; van Meer et al. 2008). Phosphatidylionositol is a target (substrate) for phospholipase C (PLC) protein. When activated PLC hydrolyzes it into two products: diacylglycerol (DAG) and inositol-1,4,5-triphosphate (IP3). Both compounds are second messengers which further initiate different cell processes (cf. Fig.9). The lipophilic part of PIP2 – DAG – is retained in the membrane where it activates protein kinase C. Inositol-1,4,5-triphosphate being soluble separates from the membrane and reaches its target receptors on the endoplasmic reticulum (ER). This in turn initiates the release of calcium ions from the stores through TRP channels (Qin 2007) PIP2 has multiple functions. In addition to its role in cell signaling, a molecule of PIP2 itself is involved in anchoring of proteins to

plasma membrane, coordination of endocytosis and exocytosis processes and interaction with ion channels (Qin 2007).

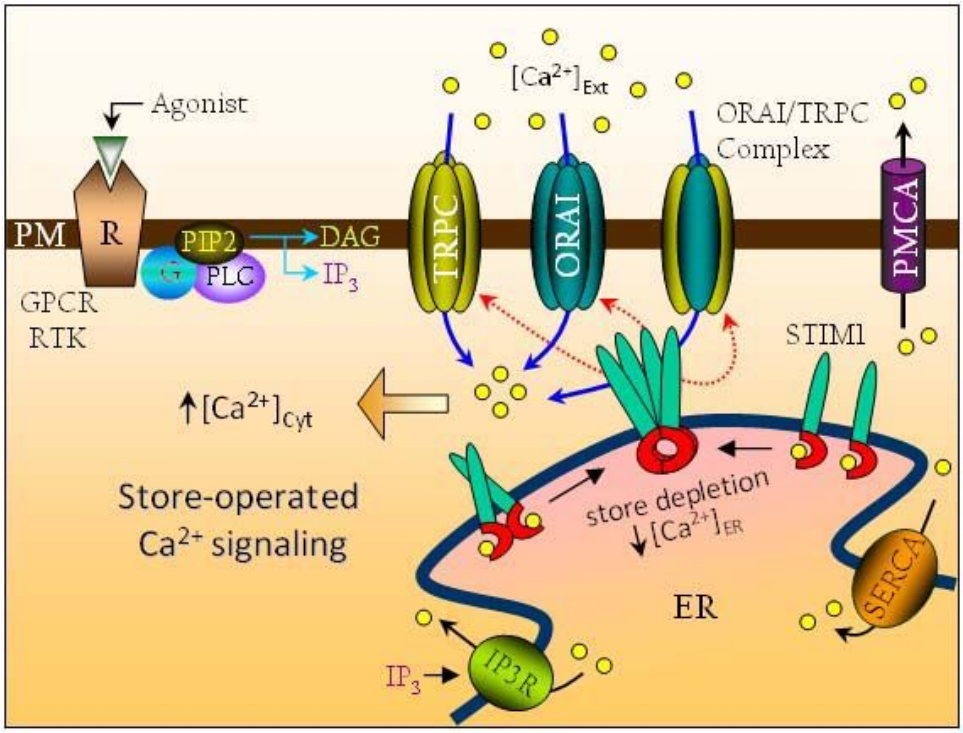


Fig.9 Storage operated calcium entry (SOCE) is mediated via the STIM, TRPC and Orai proteins upon aczivation of the Phospholipase C (PLC) pathway. Reprinted with permission from Frontiers in Biosciences from Pani et al. (2012, Fig. 1)

The first hint for TRP-PIP interaction came out on a basis of evidences that *Drosophila* TRP is involved in phototransduction. This process in *Drosophila melanogaster* requires PLC to break the PIP2 to DAG and IP3, which leads to calcium release via TRP and TRPL proteins. (Hardie, 2001). A pain receptor, TRPV1, passes a noxious stimulus about tissue damage after the intracellular signaling pathway was initiated by inflammatory substances, such as bradykinin and NGF. Activation of these substances leads to stimulation of PLC via G-coupled receptors. TRPV1 is activated in response to the decreased levels of PIP2 due to hydrolysis by PLC. The thermal hypersensitivity caused by inflammation did not arise in mice with TRPV1 deletion (Chuang et al. 2001, Qin 2007). Several members of the TRPM family were also found to be regulated by PIP2. TRPM5 is stimulated by increased intracellular Ca^{2+} concentrations followed by release from internal stores upon PIP2 hydrolysis (Lui & Liman 2003). PIP2 was shown to directly activate a cold receptor (TRPM8) and regulate the rundown of the channel. Interestingly, that while heat-sensor TRPV1 is inhibited by PIP2, cold-sensitive TRPM8 is directly activated (Lui & Qin 2005). The role of PIP2 in regulation of TRPA1 channel was initially predicted due its activation by

inflammatory peptide bradykinin and later with a potential role of TRPA1 in mechanotransduction (Qin 2007).

Several mechanisms of TRP modulation by PIP2 were described (Nilius et al. 2008). These interactions occur with TRPC, TRPV and TRPM family members. It has been hypothesized that mostly the calmodulin/IP3 receptor binding (CIRB), the highly conserved TRP domain in the proximal C-terminus of TRP channels and partial pleckstrin homology (PH) - (like) domains are involved in TRP/PIP interaction. Although, experiments done on TRPA1 shows that PIP also plays role in the channel's regulation, still there is no enough information about possible interaction sites.

Work of our experimental collaborators (Viki Vlachova, Prague) indicated that TRPA1 does not have a TRP domain which could be responsible for PIP-binding. They performed point-mutation experiment on all positively charged amino acids of the C-terminus and concluded that TRPA1 does not interact with PIP2 through the C-terminal TRP domain or other amino acids of the C-terminus. However their studies did not exclude the possibility that TRPA1 could interact with PIP through an intermolecular PH domain as was hypothesized for TRPC3 channel *(*Viki Vlachova, personal communication*)*. The involvement of a split pleckstrin homology domain in PIP2 binding was hypothesized first for TRPC3 (van Rossum et al. 2005). According to this hypothesis, PIP2 binding could be possible due to the formation of an intermolecular PH- domain consisting of the N-terminal part provided by the partial TRPC3 PH-like domain and a C-terminal part provided by the split PH-domain from PLC γ1. Recently, this hypothesis was examined by (Wen et al. 2006). They showed that mixing of the N-terminal half of the predicted TRPC3 PH-like domain with the C-terminal half of the split PH domain of PLC γ1 did not produce the canonical fold. In contrast, the mixing of two isolated halves of the split PH domain of PLC γ1 led to the original folding of two parts, pointing out that probably only two cognate halves could fold properly (Wen et al. 2006). The Nilius lab analyzed the role of PIP2 for TRPA1 and identified four putative split PH-domains located at the N- and C-termini of TRPA1 (Karashima et al. 2008). This result motivated us to model a predicted half of PH-domain and further investigate a possibility of intermolecular interaction between a TRPA1 PH-half and another complementary half of a PH-domain.

1.6 Store-operated Ca^{2+} entry (SOCE)

Calcium ions are involved in many functions on different levels of organization from cell to whole body processes: cell signaling, secretion, cell division, synaptic transmission, bone formation, muscle contraction (Clapham 2007). A particular calcium concentration is required for correct realization those multiple functions. On the cell level the movement of Ca^{2+} in or out the cell is regulated by calcium-selective or non-selective ion channels and transporters. Inside the cell, binding of calcium ions to Ca-regulated/modulated proteins promotes conformational changes enabling the protein to fulfill its function. The communication between cells is also often achieved involving Ca2+. Here, C2 ions are transported via ion channels which open in response to transmitter release or voltage change and thus send signals from one cell to another (Clapham 2007). The activation of phospholipase (PLC) initiates the phosphoinositide pathway (see also above and Fig.9) by converting PIP2 into DAG and IP3 which is crucial in regulation of calcium movement in the cell. Subsequently, DAG promotes calcium entry into the cell and IP3 triggers calcium release from internal stores - (the endoplasmic reticulum, ER) - via activation of IP3 receptors in the ER membrane (Smyth et al. 2006). In general, calcium entry into the cell is mediated by two independent mechanisms. One involves activation of plasma membrane channels by second messengers (such as DAG), so called receptor-operated channels (ROCs). The other mechanism is strictly coupled to the release of Ca^{2+} stored in the ER (Smyth et al. 2006). Under some condition, such as IP3 activation, calcium is released very fast. leading to almost complete Ca2+ depletion pf the ER. After some time (a few seconds) however, another process replenishing ER calcium levels starts. This process which senses exclusively the drop in $[Ca^{2+}]$ in the ER independently of the cytosolic Ca^{2+} level, is known as store-operated Ca entry (SOCE) and the cannels involved channels are the store-operated channels (Putney 2005; Clapham 2006; Smyth et al. 2006) mediating the Calcium-release activated currents (CRAC). These currents are characterized by being highly calcium-selective having low single channel conductance (~1000-fold lower than that of other ion channels) (Clapham 2006; Prakriya & Lewis, 2006; Soboloff et al. 2012).

In a search for protein responsible for SOCE, TRP channels were proposed as candidates and the first known TRP channel – *Drosophila* TRP was identified. The involvement of TRPs in SOCE was hypothesized based on the activation of TRP channel by a product of PLC activation – DAG (The PLC pathway is involved in light sensing in *Drosophila* (Salido e al. 2011). Despite some circumstantial evidence, for the involvement of TRPC in SOCE (see alos below), TRPC mediated currents do not fully comply with the biophysical characteristics of CRAC currents (Smyth et al.

2006; Salido et al. 2011). Thus, the molecular identity of the proteins underlying the SOCE and CRAC currents has been discussed controversial for a rather long time.

Meanwhile it is general consensus that the proteins primarily responsible for SOCE have been identified correctly: A protein of the endoplasmic reticulum, STIM was characterized in 2005 (Roos et al. 2005). Later, in 2006 its partner - the Orai channel - located in the plasma membrane was also discovered. STIM and Orai co-expressed together are able to reconstitute the properties of CRAC currents (Feske et al. 2006; Soboloff et al. 2012; Fig.9).

1.6.1 The STIM-Orai complex (Fig.9)

STIMs are highly conserved single-span membrane proteins, expressed in all cell types and mostly located in the ER. The N-terminus of STIM is located at the lumenal side. It contains two EF-hands, and protein interacting SAM-domain ("sterile α-motif") (Soboloff et al. 2012). The lumenal N-terminal domain detects the changes in calcium concentration in the ER and facilitates intermolecular interactions. The single transmembrane helix crosses the membrane and the C-terminal part is located in the cytoplasm. This part of STIM comprises a coiled-coil domain, a poly-basic region and a serine-proline rich region. When the lumenal calcium concentration drops, calcium ion dissociates from EF-hand domains of STIM. This in turn initiates the change in the EF-hand/SAM structure that enables oligomerization of multiple STIM proteins (Smyth et al. 2010). At rest (normal [Ca^{2+}] in the ER), STIM is located in tubular structures of ER membrane. Within a few seconds after activation, oligomerized STIM translocates to punctate structures where ER is located close to plasma membrane, the ER-plasma membrane junctions. The cytoplasmic C-terminal coiled-coil region comprise an approximately 100 residue segment which couple with and activate the plasma membrane Orai channel.. The whole process starting from STIM oligomerization induced by dissociation of Ca from N-terminal EF-hand domain and following Orai channel activation is reversible. When the calcium concentration inside ER is restored, Orai channels are deactivated and STIM leaves the ER-plasma-membrane junctions (Smyth et al. 2010; Soboloff et al. 2012).

Orai, a protein which reconstituted CRAC currents was first identified in Severe Combined Immunodeficiency (SCID) patients. In parallel, using genome-wide RNAi screens in *Drosophila* three proteins, three human homologues were found. These proteins were named Orai – the keepers of the gates of heaven in Greek mythology (Eunomia, Dike and Eirene; Feske et al. 2006).

Orai channels are highly selective for calcium ions over sodium or potassium under physiological conditions, but (as usual for Ca-channels) let pass monovalent cations in the absence of calcium.

The Orai protein consists of four transmembrane domains with cytosolic N- and C-terminal parts. The crystal structure of *Drosophila* Orai channel, the closest homolog of human Orai1, showed a new protein fold: Six four-transmembrane subunits are assembled together into hexameric structure around a central pore. The transmembrane helices 1 form the central ~ 55 A long pore with a diameter that is different along the pore axis. At the extracellular side the pore starts with highly conserved acidic residues, followed hydrophobic region. In the internal part of the pore is characterized by multiple positively charged residues (Hou et al. 2012).

1.6.2 Participation of TRP channels in SOCE and ROCE

However, although, the whole or focused genome screens did not identify any of the TRP channels as likely being responsible for SOCE and ectopically expressed TRPs are not able to reproduce CRAC currents (Smyth et al. 2006), an involvement of TRPs in SOCE as well as in rceptor operated Calcium entry (ROCE) seems to be still likely. Particularly, the members of the TRPC family were studied extensively in that respect. A participation of all members of TRPC family in SOCE has been confirmed in multiple experimental studies such as gene inactivation or gene expression silencing, application of neutralizing antibodies (Salido et al. 2011). Yet the involvement of TRPCs in SOCE is not as obvious as it is for Orai/STIM. It has been shown that depending on the expression level TRPCs could respond differently even in the same cell type. TRPC could either act as store-operated channel at low expression level or - at high expression level - be activated by PLC and behave as a receptor-operated channel (Salido et al. 2011). The multiple roles of TRPC channels might be explained by their interaction with other proteins involved in SOCE or ROCE. A number of studies starting from 2006 reported direct interactions of STIM1 with TRPC1. This interaction has been shown to enable the store-operated function of TRPC1. On the other hand, when STIM1 was not present, TRPC1 associated with other members of the TRPC family and these heteromers acted as receptor-operated channels (Salido et al. 2011). Moreover, an interaction of TRPCs and Orai has also been reported, which suggests the formation of heteromeric complexes functioning together as SOC channels (Salido et al. 2011).

The unambiguous and final answer to the question which proteins are to which degree responsible for SOCE is furthermore complicated by the fact that SOCE can be modulated and regulated by various factors. These include temperature, STIM1 phosphorylation, redox status, as well as the action of proteins involved in regulation of intracellular calcium homeostasis (Albarran et al. 2013). These proteins (such as TPC2, SERCA, TRPC3) can influence the STIM1-Orai1 association and therefore regulate SOCE. Recently it has been reported that not only TRPC channels but also

TRPA1 channels can associate/interact with STIM1 and Orai1. This interaction diminishes SOCE in human megakaryoblastic cell lines (Albarran et al. 2013).

1.7 The fungal TRK K^+-uptake systems

Establishment of evolutional relationship is critical in order to understand protein structures. As the number of solved protein structures is limited, one structure could be a template for reliably modeling many closely and even distantly related proteins. However, modeling of distantly related proteins is complicated and uncertain for the structure in total. But since the functionally important regions of proteins are usually conserved during evolution, the known structure of a remote relative/ancestor is often sufficient to predict the "core" structure of a related protein with unknown structure correctly (i.e. arrange the most conserved elements in a similar way).

On the basis of the crystal structure of the prokaryotic potassium channel KcsA (Doyle et al. 1998, pdb: 1BL8, for description see chapter 1.3.1 ("structure of a K channel") ; Fig.2) and with multiple sequence alignments, Durell and colleagues in 1999 proposed a relationship between potassium channels and (presumed) K^+ transporter families (TrkH, KtrB, HKT and Trk1,2). They noticed the similarity of putative selectivity filter and pore helices and proposed that these K+-translocation systems could have appeared due to gene duplication and fusion originating from a gene encoding a potassium channel transmembrane subunit (Durell et al., 1999; cf. Fig.6). Sequence analysis led them to the conclusion that among transporter family members studied in their work KtrB is the most closely related to the ancestor prokaryotic 2TM K^+ channel. Prokaryotic TrkH (with 2 additional TM helices at the N terminal part of the TM domain) diverged more. Eukaryotic Trk1,2/HKT gained cytoplasmic loops, one long loop is in between MA and MB subunits and the shorter one between MC and MD subunits. Such loops are neither present in TRKH, nor in the KscA channel. In parallel work three-dimensional structures of K^+ transporters were developed by Durell and Guy (Durell & Guy 1999). The potassium channel structure 1BL8, the only available structure of potassium channel at that time, was used as a structural template. Helical wheel representations of three-dimensional models derived by Durell and Guy demonstrate the distribution of conserved residues (Fig.10). The most conserved part locates at the pore region and the selectivity filter possess invariant glycine residue (in red color in Fig.10). Results of modeling supported the hypothesis of evolutional relationship of K^+ "transporters" and channels. The modeling of the selectivity filter (SF) based on KscA indicated that it was very likely that their SF does not have the outer binding site which exists in potassium channels (Durell & Guy 1999).

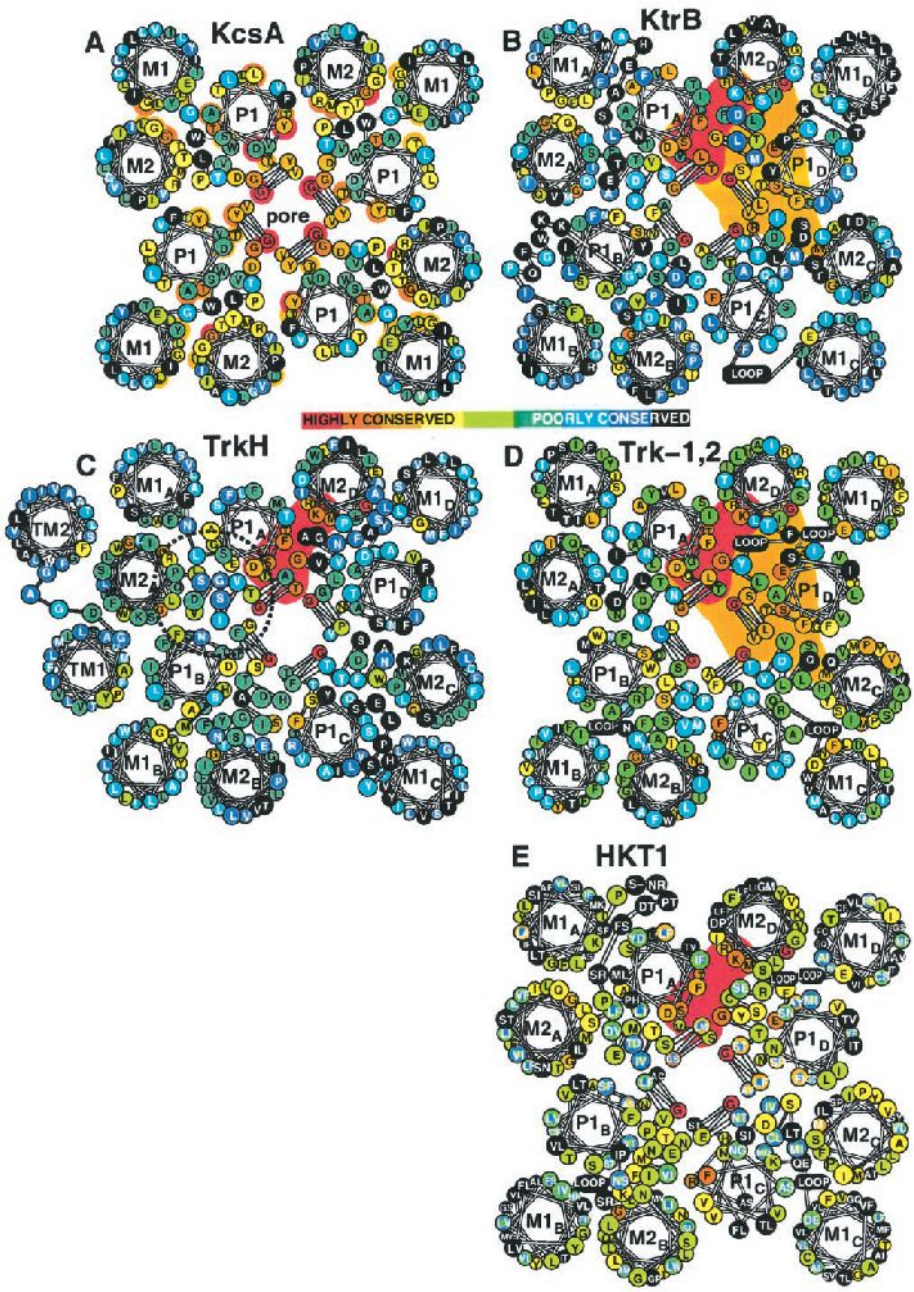


Fig.10 Helical wheel represenations of structural models of K-translocation proteins developed by Durell and Guy, 1999. Figure reprinted with permission from Elsevier from Durell & Guy (1999, Fig. 1).

Meanwhile the X-ray structures of two K^+ transporter representatives, TrkH/TrkA and KtrAB are available (Cao et al. 2011, Cao et al. 2013, Vieira-Pires eat al. 2013) and confirm the hypotheses made by Durell and Guy in 1999. A three-dimensional overlay of KcsA, KtrB and TrkH allows to fit them on the basis of the SF, where the bottom half is strongly conserved (until 3rd residue). This 3rd residue in SF is a conserved Gly. After this Gly the KtrB and TrkH selectivity filters start to deviate more from their putative "ancestor". This relation is also reflected in a multiple sequence

alignment. Electron density map allowed to describe explicitly only one binding site. However, the authors of the studies mentioned supposed that architecture of SF allow containing three binding sites, which correspond to those dubbed S4, S3, S2 in KcsA (cf. Fig.2). The site S1 from KcsA is not present in TrkH and KtrB and therefore most likely also missing in other related transporter families. This conclusion is also supported by multiple sequence alignments, which show that the SF within "transporter" families is more similar then to the potassium channel.

The X-ray structure revealed several differences between the K^+ "transporters" and the K^+ channel. Namely, these are 1) an internal loop filling the inside cavity that is present in KcsA and serves to fill and stabilize an internal cavity and 2) an arginine residue which is only present in K^+-transporters that islocated on TM helix $M2_D$ oriented towards the center of the pore (the ion path). The functional importance of this arginine was confirmed experimentally. Rb^+ flux (Rb^+ was used as tracable K^+ analogue) was significantly increased in a mutated TrkH protein in which this key Arg was substituted by Ala (Cao et al. 2011) The opposite effect, however, was observed earlier for the analogous mutation in KtrB: Here a decreased K^+ flux was observed (Kato et al. 2007).

Despite the difference in effects of Arg mutations, the involvement of this residue in ion passage was confirmed. Also, the important role of the internal loop in K^+ transport was supported experimentally for KtrB. (Cao et al. 2011; Hänelt et al. 2010). While the mentioned arginine likely is a common feature of K^+ transporters according to mutagenesis studies and multiple sequence analysis, the internal loop might be a specific feature of TrkH/KtrB transporters because such a conserved region is not observed through all transporters sequences. No similarly high conserved region is present in other parts of fungal Trk and plant or fungal HKT transporter sequences. Here arises a question: If Trk and HKT do not possess the internal loop that is present in TrkH and KtrB, how is the cavity inside filled? Did fungal Trk and HKT diverge so much from KtrH/TrkB and their common ancestor that allowed them to develop another mechanism of K+ transporting? If we assume that the Trk/HKT proteins are organized in a similar way as TrkH/KtrB, the internal cavity stays “empty”, which is not possible. Therefore, in order to fill internal cavity structural elements should be organized in a different manner. Within this thesis, it was thus attempted to develop a refined and more realistic atomic scale model of fungal Trks (specifically of yeast Trk1).

2. MATERIALS AND METHODS

2.1 What does a protein sequence tell us?

Proteins are composed of amino acids which are connected via peptide bonds to form a long chain counting hundreds up to thousands of residues. The sequence of residues in a protein is called primary structure. Each amino acid possesses different chemical and physical features. Combined together in a sequence amino acids define local and the overall different properties of proteins. The primary sequence of residues organizes itself into a limited number of regularly ordered 3-D structure elements – the secondary structure. Secondary structure of proteins is represented by helices, beta-strands, loops and turns. Depending on their specific side chains amino acids mostly prefer to adopt one secondary structure type over another. Finally, helices, strands, loops and turns fold into unique shape that gives the protein its full three-dimensional structure and is called tertiary protein structure (Xiong 2006).

During the long period of evolution proteins changed and diverged due to multiple spontaneous mutations. If a change occurred to a functionally important residue which properties, such as positive or negative charge, are particularly required for the protein to fold or perform its function, the change might lead to death of the whole cell, organism or severe disease and therefore these mutations were not accepted as the prevailing form of the protein sequence. If a mutation could change a residue to another one with similar properties and therefore didn't produce any affect on the overall protein structure there is no evolutionary pressure on this mutation to be changed to another amino acid and it can become even the prevailing form of the protein sequence. This type of mutation is called neutral, and most of accepted point mutations in proteins are based on this type of mutation, which led to the neutral theory of evolution in contrast to Darwin's selection (Motoo 1983).

Some mutations have changed the protein properties in a way that allowed to perform its function more efficiently or to better adopt to changes of environment. Such mutations, called positive, had given some advantage to its owners and therefore were reproduced successfully and become the prevaling from. It is possible to observe the traces of evolution when sequences of proteins are compared. This process is known as protein sequences alignment.

The number of sequences in the UniProtKB/Swiss-Prot database currently contains 540958 sequence entries and this number is continuously growing [http://web.expasy.org] (Artimo et al. 2012). At the same time the number of protein structures is also increasing. Nowadays 87138 protein structures are available in Protein Data Bank (PDB) [http://www.rcsb.org] (Berman et al. 2000). Although, each protein possess a unique sequence and properties it does not necessarily have an exceptional fold, but shares the fold with other family or even superfamily (fold-family)

members. According to PDB statistics 1393 unique folds exist in 87138 solved and deposited structures.

Protein sequence alignment is a crucial step in order to characterize a new protein, as the above mentioned statistics demonstrate that there is a limited number of three-dimensional folds, while the primary sequence has nearly endless variations. This means that proteins that differ significantly in their primary sequence nevertheless can share a common three-dimensional structure. Given a relatively high sequence identity between two proteins usually – usually at least 30 % - one can predict that both proteins belong to the same family and thus share a similar three-dimensional structure. This "law" is a basic principle of protein modeling, which gave an opportunity to design and investigate yet unsolved protein structures if the three-dimensional structure of at least one family member is known.

2.1.1 Sequence alignment

Sequence alignment serves to find an evolutionary relationship between proteins. All known protein sequences are stored in databases, where they can be retrieved, compared, analyzed. BLAST (Basic Local Alignment Search Tool) is a powerful engine designed to compare protein or DNA of interest to the datasets of structures and sequences (Altschul et al. 1990).

Conclusions about relationship between proteins can be drawn on the basis of sequence identity and similarity, which is calculated directly between two sequences. Identity defines the percentage of identical residues shared by two sequences. Percentage of similarity includes additionally the number of residues with common physico-chemical features that share the same position in the compared sequences.

Two main approaches of alignment exist: global and local. The use of one or another depends entirely on the question asked. Global alignment is aimed to analyze the sequences along whole length and find the best match from the beginning to the end. This method is useful to compare related sequences of approximately the same length or for determining the evolutionary distance between two or more sequences. When a protein is unknown or is composed of various domains it is good to check not only the entire sequence but also any region that could be similar to already characterized proteins. Local alignment methods allow the comparison of sequences of different length and degree of similarity are used to find distantly related sequences, conserved motives and domains.

Both local and global alignment are aimed to get the best possible match between two sequences, the underlying algorithms in both cases are similar and use dynamic programming.

2.1.2 Dynamic Programming

In dynamic programming method a two-dimensional matrix of the compared sequences is created in the first step. Starting from the first row, each residue of sequence 1 is compared to all residues in sequence 2. When the first row is compared the scoring matrices and gap penalties are applied to justify the likelihood of residue matches. Thus each combination of residues gets a score and also the neighboring amino acids are traced for each position. Once the matrix is completed, backtracing is used to find the optimal alignment.

The dynamic programming method for global alignments is the Needleman-Wunsch algorithm (Needleman & Wunsch 1970). This method guarantees to find the optimal global alignment, however there might be several optimal solutions between which the algorithm cannot distinguish. Its useful for aligning closely related sequences of approximately the same length, but is unable to identify short motifs that are not in the same position in sequence, or domain shifts. This problem is solved by the Smith-Waterman algorithm, that differs from the Needleman Wunsch only in a small detail, which nevertheless has a huge effect on the outcome (Smith & Waterman 1981): It does not allow the score to become negative. In case it would get negative it becomes just zero. Thus it allows the backtracking to start and end at any sequence position, and not necessarily only at the C-terminal, and is therefore able to identify motifs or domains in otherwise unrelated proteins.

2.1.3 Substitution matrices

The most simplest scoring scheme for aligning two sequences would simply differenciate between match and mismatch. Nevertheless, not every mismatch necessarily changes the three-dimensional structure and amino acids with similar properties (size, hydrophobicity, charge) can be substituted with a 95% confidence that this would not change the three-dimensional structure. In order to differentiate the potential impact of a specific substitution on the protein structure and function, substitution matrices have been developed that score the possible pairs of amino acids. Therefore naturally occurring substitutions were studied in sequences of the same protein family and statistical data derived from it evolved into 20X20 substitution matrices. There are two main types: PAM (Point Accepted Mutation), which were developed already in the 1970s by analysing closely related proteins to estimates what rate of substitution would be expected if 1% of the amino acids had changed, and then extrapolating the PAM1 matrices to the various degrees of similarity and BLOSUM (BLOcks SUbstitution Matrix), which were developed using a specific set of related proteins and global alignments for each degree of similarity. Substitution matrices thus justify the probability of one residue to be exchanged with another one when two sequences are compared

(Dayhoff & Eck 1968; Altschul 1991; Henikoff & Henikoff 1992).

In order to estimate the statistical significance of the alignment one has to estimate the probability that the alignment is not just random, but that the shared residues are indeed from a common ancestor. For that purpose the score of the final alignment is compared to the scores of multiple random alignments of the same amino acid composition. Residues of both sequences are randomly shuffled, and new alignments are generated using fake sequences. The distribution of scores of random and true alignments is constructed. If the score of target alignment is on the extreme border of the distribution in contrast to scores of random sequences, the alignment can be considered as true.

Often, the alignment of more than two sequences is required, for example, the alignment of all sequences within one family. This enables for example to identify highly conserved regions within whole family, which could be a functional domain, motif, binding site, that could not be found on a basis of only two sequences. Multiple alignment use a guide tree to find the best matching within all aligned sequences. Given a multiple number of targets, first pairwise alignments of all possible combinations of sequences are constructed and all matches are scored according to a particular scoring matrix. Then the easy alignments are done first and then step by step the more complicated (means more diverged sequences) are included following the guide tree. The resulting multiple alignment is always a global alignment.

2.2 Homology modeling

Homology modeling, or comparative modeling, also called knowledge-based modeling became an important approach in protein science, as practically all fold families are known nowadays and there is at least one family member with known 3D structure in most of the protein families. This is reflected by PDB statistics, where not only the number of folds is much lower in comparison to the overall number of structures but also practically no new folds were observed over the last years despite the ever increasing number of structures. Homology modeling allows to analyze a protein on atomic level, when its structure is not resolved yet. As the name suggests, modeling involves the generation of the three-dimensional structure of a new protein using a homologous already resolved structure, or in other words it requires a three-dimensional template (homolog) and a sequence alignment.

Creating a model involves several steps. The crucial step in homology modeling is the selection of an appropriate template, which will be used as the frame for a new model. At this step the target

sequence of an unknown protein is compared against all known sequences of solved protein structures. Generally, this is done with BLAST (Basic Local Alignment Search Tool). When several structures of the same family exist, preference for one template over another is considered on the basis of sequence identity between compared proteins as well as on several parameters of the structure, such as: structure resolution, R-value. Additionally, the conformation of the template might play a role, too, as proteins might have various conformations or might have been resolved with various co-factors, ligands or substrates. The degree of sequence identity determines the quality of the three-dimensional prediction. A common threshold for proteins considered to be homologous lays at 30 % of sequence identity. Additionally, the range between 20 to 30 % is the so-called twilight zone, where proteins still could be assumed as related in some cases. However, its not always straightforward to conclude whether proteins are homologous or not (Xiong 2006; Sternberg 1996).

Once the template is identified, the most critical point in modeling is the proper alignment of the target sequence with the template(s). For a given high degree of identity, this step can be done confidently well even using automatic procedures. When the percentage of sequences identity falls to 30 % or below, the alignment becomes a challenge and is the most crucial step in the whole modeling process. Generally spoken, a bunch of approaches is applied to improve the alignment. The initial alignment could be created using different algorithms, that are more sensitive to particular features, or include structural information into alignment. At another level, secondary structure prediction using several different algorithms could be done to provide more insight into the structural features of the unknown protein. When dealing with membrane proteins, the protein topology should be predicted to enable the correct identification of membrane segments borders. If more templates are available or several sequences of the target protein family, the construction of multiple alignment helps to better justify overall alignment and especially regions of low similarity.

In principal, the modeling procedure itself narrows down to backbone, loop and side chain modeling. The approaches for creating a complete full-atomic model differs mostly at a level of backbone formation. The use of one or another depends on a particular protein of interest and information available. There are two main methods of comparative modeling: fragment-based and restraint-based.

Fragment-based method identify sequence conserved regions (SCRs) and sequence variable regions (SVRs) in the target protein. For the SCRs the backbones of all templates, are superimposed and converted into “framework”, a weighted mean structure. In the next step loops are added to connect the individual backbone pieces. Finally, side chains are added and optimized. Loop and side-chain modeling might produce structural clashes or energetically unfavourable angles and bond lengths,

and is therefore followed by a consequent structural optimization of the whole target protein that includes energy minimization, hydrogen-bond network optimization, and molecular dynamics. The impact of every template in framework depends on square sequence identity with the target sequence. (Sternberg 1996; Srinivasan & Blundell 1993, Sutcliffe et al. 1987).

Loops are the most complicated part of protein to model, as they are usually poorly conserved and contain most of deletions and insertions. Main criteria should be satisfied in order to select a loop template: sequence similarity, loop length and also end-to-end distance because a new loop should fit well into anchoring structure. Finally, a loop should not overlap with the rest of the structure. For short missing segments only fulfilling geometrical parameters is enough to produce reliable conformation. (Sternberg 1996).

Modeling of side chains of identical residues is straightforward, as coordinates could be copied to the new structure. Analysis of database of protein structures revealed that side chains prefer to accommodate a particular torsion angle. The rotamer libraries were derived using quantitative values of possible phi and psi angles. Theoretically, side chains of unknown residues could be built by trying all possible conformations. However, its computationally very expensive and not always needed as many side chain orientations depends on a backbone conformation. Therefore several methods were developed in order to optimize searching for a appropriate torsions. COMPOSER, fragment-based method uses probability and preferred side chain conformation corresponding to secondary structure type. An example of a fragment-based method which was used in this thesis is the Yasara homology model module (Krieger et al. 2002).

Thw restraint-based method produces a model by satisfaction of a set of different spacial restraints: interatomic distances, angles, chirality and other structural features. (Sternberg 1996; Havel & Snow 1991; Sali & Blundell 1993). The program used here, Modeller, applies spacial restraints for Ca-Ca and mainchain N-O distances, and mainchain and sidechain dihedral angles. Probability density functions are calculated for each type of restraints based on 17 families of closely related proteins. The model is calculated on a basis of a given alignment between template structure and target sequence. Probability density functions (Pdf) are applied to each restraint feature (Ca-Ca and mainchain N-O distances, and mainchain and sidechain dihedral angles). Pdfs are expressed as function of similarity of target and template residues. In this approach the target protein is complete from the very beginning, which means that loops and sidechains are included while fulfilling the restraints. In the process of modeling the Pdf is optimized employing methods of conjugate gradients and molecular dynamics with simulated annealing.

At the last step of homology modeling procedure the final model has to be evaluated to reveal all the problematic regions and errors. There are different approaches existing to analyze the theoretical

correctness of the model. Procheck generates Ramachandran plot and describes all stereochemical characteristics of main chains, side chains: bond length and angles, phi and psi angles. It gives a summarized value for each type of protein geometry and deviations from standard values (Laskowski et al. 1993). ProSa analyses overall quality of a new model and how well it fits into normal values obtained for experimentally solved protein structures. It gives also knowledge-based energy plot as a function of residue number, so that it is possible to trace what are the problematic regions of a model (Wiederstein & Sippl 2007). Another useful approach is WHATIF server. It characterizes bond lengths and angles, creates Ramachandran plot (Vriend 1990).

2.3 From static model to molecular dynamics

Molecular dynamics and quantum mechanics are two approaches to study biological systems *in silico*. Molecular dynamics is used to study systems at the atomic level, where the smallest entity of the system is an atom (all atom dynamics) or a group of atom (coarse grained). Quantum mechanics is used to describe a system on the level of electrons and protons. In each case various qualities of the systems are analyzed. For example, molecular dynamics allow to study systems consisting of thousands atoms, such as a big cytoplasmic protein or ion channel embedded in a lipid bilayer. In such systems the interaction between molecules, rotation of domains or transport of water or ions via ion channels can be investigated. Coarse grained simulations allows to study such properties as a diffusion of lipids. In order to describe chemical reactions (breaking and formation of covalent bond) between molecules quantum mechanics is required. However, quantum mechanics is computationally very expensive, therefore it is limited to small molecules. When information on electron level is needed for a big system such as proteins a combined approach of QM/MM is used. Within the QM/MM approach a system is divided into two parts, the QM part (usually ligand and binding site) is described by using QM methods and the MM part where the rest of the system (the rest of the protein and solvent) is calculated using classical molecular dynamics.

2.3.1 Molecular dynamics

In atomistic molecular dynamics atoms are the smallest elements that can be represented as hard spheres bound to each other by means of covalent bonds commonly modeled as springs. The energy of the system is possible to derive from this simple model (atoms as charged balls connected by covalent bonds as harmonic springs). Molecular dynamics basis on Newton's equation of motion:

$F = m * a$,

where F – force, m – mass, α – acceleration

Movements of atoms are enabled by applying integration of Newton's equation, where the force on one particle depends on position of that particle and relative position of other particles in the system. [Andrew R. Leach]

There are two types of interaction between atoms in a system: *a)* atoms joined to each other by bonds and *b)* atoms which are not bound directly. These two "groups" of interactions define potential energy (*U*) of the system:

$$U = \sum U_{bonded} + \sum U_{nonbonded}$$

Specific parameters for each type of atom and bond, such as bond length or angle, are described by force fields.

2.3.1.1 Force fields

Force fields describe the energy of the system as a function of the atomic coordinates. In case of the classical molecular mechanical force fields the evaluated energy depends on the atomistic empirical parameters. The main dynamic features of a molecule described by force fields include bond stretching, angle bending and changing of torsion angle. The force fields used in this thesis OPLS (Optimized Potential for Liquid Simulations) was developed in order to better reproduce simulations in a solvent. Equilibrium parameters for bonds and angles were obtained from solved crystal structures. Bonds and angles are harmonically restrained with the force constant corresponding to experimental data of vibrational frequency (Jorgensen et al. 1996). Dihedral parameters are derived using optimization of dihedral parameters and calculation of potential energy for each conformation using *ab initio* and molecular dynamics. The charges are calculated on an equilibrium conformation of molecule using RESP (Restrained Electrostatic Potential) method based on *ab initio* calculations. Within this method atomic charges are fitted on electrostatic potential obtained from *ab initio* QM calculations (Jorgensen et al. 1996). In order to reproduce better simulation in solvents, Jorgensen repulsion term in the Lennard-Jones potential is modified in order to avoid a strong repulsion (Jorgensen et al. 1996).

2.3.1.2 Bonded interactions

Bonded interactions include: bond stretching, angle bending, torsional angle and improper angle.

Bond stretching

The bond lengths are identified experimentally from solved X-ray structures. The values for small

molecules can be obtained from quantum mechanical calculations. The bonds are represented as springs, where changes in a bond length (vibrations) are defined by Hook's law:

$$E_{bond} = \sum_{bonds} \kappa_r (r - r_{eq})^2 ,$$

where E - potential, *k* - spring constant, r_{eq} equilibrium value for a bond length.

Hook's law allows only a small deviation from the reference bond length which is within 0.1A (Schlick 2002; Leach 2001; Kaminski et al. 2001).

Angle bending

Parameters for a particular angle in a molecule depend on the hybridization state of the central atom. Therefore the values for a angle will depend on the particular atom type. The fluctuations of an angle about its reference value can be also described by a harmonic potential.

$$E_{angle} = \sum_{angles} \kappa_\theta (\theta - \theta_{eq})^2 ,$$

where θ_{eq} equilibrium value for an angle.

However, the angle bending requires less energy to deviate from its equilibrium value in comparison to bond stretching terms (Schlick 2010; Leach 2001; Jorgensen et al. 1996).

Torsion angle

Bond stretching and angle bending are very firm terms. The biggest movements between atoms in molecules are achieved due to torsions and non-bonded interactions. However, definition of bond stretching and angle bending are not enough to describe correctly the geometry between C1-C4 atom in C1-C2-C3-C4 chain. Therefore torsion should be included into potential energy function as well (Schlick 2002; Leach 2001; Jorgensen 1996).

$$E_{torsion} = \sum_i \left(\frac{V_1^i}{2} [1 + cos(\varphi_i + f_i 1)] + \frac{V_2^i}{2} [1 - cos(2\varphi_i + f_i 2)] + \frac{V_3^i}{2} [1 + cos(3\varphi_i + f_i 3)] \right),$$

where φ_i – dihedral angle, *f1*, *f2*, *f3* are phase angles, V_1, V_2, V_3 – coefficients in the Fourier series.

Improper angles

Improper angles are designed for special cases when all described above properties cannot define correctly the geometry of the whole molecule due to lack of electronic information. Mostly improper torsion angles are defined between a three-atom plane (C1-C2-C3) connected via a central atom with “out-of-plane” bond with a central atom (C2-C4). Improper angles improve energy profiles and geometries of molecules (Schlick 2002). Improper dihedrals of harmonic type in GROMACS (GROningen MAchine for Chemical Simulations) are described by Hook's law $E_{(\omega)} = k(1 - cos 2\omega)$, and the periodic type of dihedrals is defined as a proper dihedral angle (Berendsen et al. 1995).

2.3.1.3 Non-bonded interactions

The non-bonded terms describe the interaction between different molecules and atoms within the same molecule that are not bonded directly. The total energy for non-bonded interactions in the OPLS force fields is defined by the equation (Jorgensen et al. 1996):

$$E_{nonbond} = \sum_{i}^{N} \sum_{j<i} [\frac{q_i q_j}{r_{ij}} + 4\varepsilon_{ij}(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}})] f_{ij}$$, where f_{ij} *is scaling factor* f_{ij} *= 0.5 if i, j are for 1,4 atoms pair (atoms over three covalent bonds far; 1,2 and 1,3 atom pairs (atoms connected by covalent bond or over two bonds far) are not included at all); otherwise* f_{ij} *=1.0; N – total number of atoms;* q_i *– atomic charge;* ε_{ij}, σ_{ij} *– Lannard-Jones parameters for atomic pair;* r_{ij} *– distance between two atoms;* $E_{nonbond}$ *– total nonbonded energy*

The first term in this equation describes electrostatic interactions expressed by **Coulomb potentials** and characterizes an interplay between atomic charges (Leach 2001). Energy decreases when two oppositely charged atoms coming closer and increases for same charge atoms.

The second term of non-bonded interactions between atoms is described by the Lennard-Jones potential (Van der Waals interactions) which joins together the attraction and repulsion terms of interaction. The attractive term includes London dispersion. The repulsive force describes Pauli repulsion due to overlapping occupied electron orbitals. They are defined by van der Waals radii that are calculated as a half distance between two nuclei of atoms of the same nature observed in a crystal. The resulting Lennard-Jones potential for two atoms indicates the minimum of energy that corresponds to equilibrium distance between both atoms. If the distance between atoms gets smaller, the repulsion term starts to work which is characterized by a steep jump in energy. The equilibrium distance between particles depends on the atom type (Israelachvili 1991).

2.3.1.4 Periodic boundary conditions

In order to simulate natural infinite conditions where particles do not meet walls, or boundaries, a special method was developed. The simulation box is replicated in all directions, so that the central original box is surrounded by its copies. This corresponds to the concept of an ensemble in statistical thermodynamics, having an identical composition of atoms and other properties (volume, temperature, pressure). [Alan Hinchliffe, 2003; Andrew R. Leach, 2001] During the simulation when one atom or molecule leaves the box (as there is no boundaries), the copy of this particle appears on another side of the box, so keeping the constant number of atoms in the box. The force on an atom is computed taking into account all atoms in the original box as well as atoms in surrounded boxes (Hinchliffe 2003).

2.3.1.5 Cut-off

The next position of an atom is computed with respect to positions of other particles in a system that exert a force on that atom. The atoms in the neighborhood have a higher influence than atoms at the opposite side of the box. When we deal with periodic system, the atoms in each copy cell should be taken into account. Therefore the number of non-bonded interactions that should be considered for calculation of potential energy is very high and very computationally expensive. In order to deal with such situation the cut-off are introduced that breaks the interaction which is beyond a particular distance (Leach 2001) It is generally assumed that a cutoff distance should be smaller than the shortest box length. For a short-range interaction minimal image convention is implemented in GROMACS, that allow to "see" only the nearest copy of atom, with the cut-off value of 10 Å. While short-range Lennard-Jones potential *feels only closest atoms* and cut-off can be used without correction, the electrostatic interactions decay very slowly and "feels" atoms at a longer distance and use of simple cut-off where the potential is cut to zero at some distance give artifacts to energy calculations. Therefore different approaches were designed to deal with such artifatcs and to decrease the cost of calculation of long range electrostatic interactions. In GROMACS the **Particle Mesh Ewald (PME)** method is used to deal with electrostatic potentials. The Particle Mesh Ewald (PME) method provides a slow decay of the potential, that is more natural than the cut-off method. PME is based on the Ewald summation, where the electrostatic interaction is divided into real (short-range interactions) and reciprocal space (long-range interaction) (Darden & Pedersen 1993).

2.3.1.6 Neighbor search

It is computationally very demanding to calculate all non-bonded interactions every step, since the distance between every atom pair should be determined. Interacting particles, however, do not change every step during simulation. On basis of this idea the *neighbor search* method is designed. All particles within a particular distance to particle ***i*** are detected together with particles that are slightly behind the cut-off. The forces are calculated only for atoms with distance shorter than cut-off radius, this information is stored as pair list. For neighbor search in a system with periodic boundary conditions a grid search method is implemented in Gromacs (Verlet 1967; Bekker et al. 1993; Pronk et al. 2013).

2.3.1.7 Integrating Newton's equation of motion

As a first step of molecular dynamics simulation, potential energy of the system is calculated using

initial atomic coordinates, which encompasses bonded and non-bonded terms. The kinetic energy is included in form of atomic velocities, that are generated from random numbers using Maxwell-Boltzmann distribution for desired temperature. Then the force on each atom is computed as a gradient of potential energy against coordinates. The next position of an atom is calculated by integrating Newton's equation of motion taking into account current position, velocity and force. When a molecule is moved, new forces are calculated taking into account new positions and velocities of all atoms. These steps are repeated as long as simulation time set, positions and velocities of atoms in time are recorded into trajectories. (McQuarrie 1976; Leach 2001)

The most popular integrator methods used in molecular dynamics are based on Verlet algorithms (Verlet 1967). In Verlet algorithm position, velocities and accelerations are written as Taylor series expansions. The new position of an atom is calculated from *a)* position and acceleration at time point ***t*** and *b)* position at previous time point in a following way:

$$r(t+\delta t)=2r(t)-r(t-\delta t)+\delta t^2 a(t),$$

where t – time, r – position, $r\ (t+\delta t)$ – position at the next step, $r\ (t-\delta t)$ - position at the previous step, α -acceleration.

The disadvantage of Verlet algorithm is that velocities are not calculated simultaneously with position, but only after the next coordinates are calculated. Also, Verlet requires a previous step in order to calculate all quantities, that's why for initiation of procedure some approach should be used to calculate position at previous step. Several other methods were developed based on Verlet algorithm to avoid these inconveniences. For example, *leap-frog* and *velocity Verlet methods* implemented in GROMACS integrate Newton's equation in a modified way (van Gunsteren & Berendsen 1988; Swope et al. 1982).

In the *leap-frog* method velocities at the next step are computed from previous half-step velocities and acceleration at time point ***t.*** The next coordinates of atoms are calculated then using new velocities and atom position at time ***t***.

$$r(t+\delta t)=r(t)+\delta t\, v(t+\frac{1}{2}\delta t)$$

$$v(t+\frac{1}{2}\delta t)=v(t-\frac{1}{2}\delta t)+\delta t\, a(t)$$

Therefore, velocities and position in the leapfrog algorithm are calculated not simultaneously, but at different time.

$$v(t)=\frac{1}{2}[v(t+\frac{1}{2}\delta t)+v(t-\frac{1}{2}\delta t)]$$

2.3.1.8 Temperature coupling and pressure coupling

Usually, biological systems have to be studied under constant temperature (T) and constant pressure (P), what correspond to natural conditions. However, in order to simulate biological molecules and analyze thermodynamic properties, it is required to set up a system that would correspond to isothermal-isobaric thermodynamic ensemble. Several methods exist which allow to mimic constant temperature and pressure conditions (Hinchliffe 2003; Schlick 2002; Leach 2001).

Temperature coupling

The temperature and kinetic energy are connected via equation

$$\langle K \rangle_{av\,NVT} = \frac{3}{2} N k_B T ,$$

where $\langle K \rangle_{av\,NVT}$ - average kinetic energy at constant NVT (number of particles, volume and

temperature), N – number particles, moles, k_B – Boltzmann constant.

According to the above equation the most straightforward way to control the temperature is to modify the velocities in the system. The magnitude of velocity modification is defined by a scaling factor. (Woodcock, 1971; Leach 2001)

Berendsen temperature coupling scheme applies an external heat bath to the system, which "senses" the temperature inside and correct it in order to keep it constant. The correction of temperature is done by adjusting atomic velocities. Whether velocities will slow down or speed up is given by a parameter lambda λ, which contains the information about the differences in current and desired temperatures (Leach 2001; Berendsen et al. 1984).

Nose-Hoover thermostat is intended to reproduce a correct canonical ensemble. Kinetic energy of reservoir is connected to Q parameter, which determines how much thermal energy should be given to the system (Leach 2001). GROMACS implies a different parameter to keep the system at desired temperature, which is related to Q, but the strength of coupling is directly related to fluctuations of kinetic energy (Nose 1984; Hoover 1985).

Pressure coupling

Barostates are designed to maintain the pressure of the system. The concept of a barostat in molecular dynamics relies on the virial theorem, which connects the average kinetic and potential energy of the system which can be expressed in terms of positions and potential energy (Leach 2001). According to this virial theorem the position will have a greater effect on a system than

internal energy, that it directly influences volume and therefore pressure. Consequently, constant pressure conditions can be achieved by controlling the volume (Leach 2001).

In **Berendsen's** version of pressure coupling the volume is changed via adjustment of atomic positions by scaling factor. Scaling factor is represented by a matrix, which can be calculated either isotropically (for one direction), anisotropically (along X, Y and Z axes independently), or semi-isotropically (along chosen axes). Scaling factor includes isothermal compressibility value, that reflect the degree of volume fluctuation under particular pressure (Leach 2001; Berendsen et al. 1984).

Parrinello-Rahman

However, the Berendsen barostat does not generate a correct isothermal isobaric (NPT) ensemble, in order to reach a true NPT ensemble Parrinello-Rahman pressure coupling can be used. With the Parrinello-Rahman method box vectors are used to construct a 3x3 matrix. Thus the volume will be treated as a fully dynamic variable. In a similar way to the *Nose-Hoover* thermostat an additional virtual object with its own degree of freedom is incorporated into system that serve as a pressure piston (Parrinello & Rahman 1981; Schlick 2002)

2.4 Energy minimization

The potential energy of the system can be represented as potential energy surface, which has a multidimensional shape. The high energy of the system corresponds to instability in the system due to some unfavorable conformation. In order to obtain lower energy the geometry of the molecules of the system has to be changed. The process of searching for lower potential energy surface for the system is known as energy minimization (Leach 2001). System minimization is usually achieved by calculation of derivation of the energy function with respect to atomic coordinates. The points where this first derivative is zero are known as stationary points and they correspond to the minima (or transition states of different order) of the energy function. Using the first derivative method the direction of the nearest minimum of energy function and as well as magnitude of the slope can be identified. The second derivative method provides more detail information about nature of the potential energy surface, so that direction of the function can be predicted (Leach 2001). An example of first derivative method implemented in GROMACS is **Steepest Descent**, which

minimizes the system in a following way. On a basis of a starting conformation potential energy of the system and forces acting on each atom are calculated. The initial maximum displacement should be given as an input parameter. Using initial coordinates, information about forces and initial displacement, the new coordinates are calculated. Then forces and potential energy are calculated for the new putative conformation. New conformations are accepted if the potential energy is smaller in comparison to the previous step or rejected if the energy gets higher or equals. The maximum number of minimization steps is given by the user in the input parameters file (Pronk et al. 2013).

2.5 Ligand docking

Docking algorithms predict the most probable binding conformation of two molecules. Most often, a small molecule (ligand) is bound to the large target molecule (protein). There are two main steps in docking of the ligand to the protein. At first, a most probable conformation of the ligand is found by exploring all allowed degrees of freedom of the ligand using a search algorithm. At this step the ligand can be treated either like a rigid or flexible molecule while keeping the protein fixed. Additionally, if needed, the degree of freedom can be also assigned to the side chains of the protein in the active site. When conformations with a suitable geometry of protein-ligand complex are found, the binding energies for each conformation inside the binding pocket are computed and then sorted according to their energy values using scoring function. The best conformation corresponds to the lowest free energy of binding. Docking methods differ in their search algorithms for possible conformation and in calculations of energies. The method used in AutoDock program applies Genetic search algorithm and semiemprical force fields for calculations of the free energy of binding (Morris et al. 2009; Huey et al. 2007).

2.6 Quantum Mechanics

While molecular mechanics treats the atoms as charged spheres and bond between atoms as springs with constant parameters, quantum mechanics can distinguish smaller components of atoms: electrons and nuclei. The system is described by the time independent **Schrödinger equation**: $\hat{H}\psi_n = E_n\psi_n$. The left-hand part of Schrodinger equation is known as Hamiltonian operator, which is the sum of kinetic and potential energy of the system. However, exact solution of Schrödinger equation is possible in limited cases only, like hydrogen atom. For molecular systems **Born-Oppenheimer** approximation of the Schrödinger equation is used. According to Born-Oppenheimer approximation, motions of electrons and nuclei are treated separately. In order to calculate electron wave function, positions of nuclei are considered unchanged. Since the mass of electron is much smaller, position of nuclei will not adjust every electron movement. The moment of nuclei are treated classically. The total energy (hamiltonian) is defined as the sum of nuclear and electronic energy. The electronic energy is represented by kinetic of electrons, potential energy of electrons (attractive) in the electrostatic filed of nuclei in particular configuration, fixed position and repulsion energy between electrons. Nuclear energy is defined as electrostatic repulsion between nuclei. The electronic wavefunction is then calculated for the given geometry of nuclei (Leach 2001; Schlick 2002). The many electron wavefunctions is further simplified by the product of single electron wafefunctions (molecular orbitals - MO) in the form of the Slater determinant. This approximation is utilized in the so called Hartee-Fock (HF) method, where the Schrödinger equation for electrons can be rewritten as $\hat{F}\phi_n = \epsilon_n\phi_n$, where φ stands for single electron wavefunction (MO), ε for orbital energies and F for the Fock operator. Finally the single electron wavefunction can be expressed as a linear combination of fixed functions (usually atomic orbitals) called basis set. Thus the function is expressed as a vector and operator as a matrix, transforming the Schrödinger equation to the algebraic form. Apart of HF method, the other approximate method for solving the Schrodinger equation, called the Density Function Theory (DFT) is widely used. Within this method the total electron density rather then many electron wavefunction is used, however the formalism of its practical mathematical solution is very similar to the HF.

3. RESULTS AND DISCUSSION

3.1 *h*TRPA1 transmembrane domain and N-terminal modeling

The initial model of the TRPA1 channel was prepared in our group based on the first structure of a voltage-gated potassium channel (PDBID: 2A79), which was the only available template at that time. However, this crystal structure had many unresolved residues denoted as 'X' for which only backbone information was available. Therefore, when a newer structure that included most of the side chains was released we refined the earlier low resolution model and finally generated a new model of TRPA1 based on the template structure (PDBID: 3LUT). This model then was used to connected with the model of the N-terminal 17 ankyrin repeat structure and the calcium-binding EF-hand. This work was motivated by the potential mechanical properties of ankyrin repeat domains that allow it to function as a soft spring that can be extended over a large range while maintaining structural integrity(Lee et al. 2006). As the calcium binding EF-hand is part of the domain, we were especially interested how calcium binding might alter the mechanical properties. In our simulations we found a strong effect of calcium binding on the dynamics: the calcium-bound state is by far more rigid as compared to the calcium-free state, in which the end-to-end distance can change by almost 50% in the apo form. Our work shows that this increase in stiffness that constraints the end- to-end distance in the holo form is predicted to affect the force acting on the gate of the TRPA1 channel, thereby changing its open probability. Additionally, the simulations show that residue N855 in the transmembrane domain of TRPA1, which has been associated with familial episodic pain syndrome (Kremeyer et al. 2010), forms a strong link between the S4-S5 connecting helix and S1, thereby creating a direct force link between the N-terminus and the gate. The N855S mutation weakens this interaction, thereby reducing the communication between the N-terminus and the transmembrane part of TRPA1. All details of this work are reported in attached **paper II**, Zayats, JMM, 2012.

Later, a first structure of voltage-gated sodium channel from *Arcobacter butzleri* (PDBID: 3RVY) became available. Several phylogeny studies suggested that Na^{+}/Ca^{2+} voltage-gated channels and TRP channels had evolved from the same ancestor protein and are closer to each other than to potassium channels (Egla et al. 2009). Therefore, in addition to the classical potassium channel Kv1.2, a new template NavAb, which gave additional information and refinement, was considered for modeling. In order to improve the transmembrane topology assignment and the prediction of residues corresponding to the selectivity filter, a multiple alignment between the given templates and target channels was constructed. The multiple alignment consisted of several members of every TRP family (TRPA1; TRPV1-6; TRPC1,4-7; TRPM1,3-6,8), Voltage-gated sodium channels

(NavAb, Nav1.7, Nav1.8, Nav1.9) and voltage-gated calcium channels (Cav2.1, Cav3.2). The initial alignment was built using the T-Coffee server Expresso structure-based mode. The transmembrane topology for every TRP channel was assigned to correctly identify TM helices. All six transmembrane regions were aligned accordingly and adjusted manually in every sequence. It was not possible to identify a common pattern of every transmembrane helix or pore-helix in all aligned sequences due to the low sequence identity between the proteins. However, as a guidance, a conservation of the hydrophobic pattern could be traced. There are several rules that apply to transmembrane proteins that should be considered during homology modeling. These rules were developed by Guy and Durrell (Guy & Durrel, 1994, 1996) on the basis of membrane protein analysis. It is particularly challenging to align selectivity filter (SF) sequences, due to the difference in amino acid composition and selectivity properties of TRP channels. Several theoretical and experimental works were reported that describe and identify key residues of SF (Banke et al. 2009; Wang et al. 2008). Concerning TRPA1, mutagenesis studies on rat TRPA1 identified Asp 918 (Asp 915 in human) being responsible for permeation of Ca2+. Mutation of Asp 918 to Ala produced a channel almost $Ca2^+$ impermeable, while Asp 918 to Glu mutation increased the $Ca2^+$ conductance in comparison to the wild type (Wang et al. 2008). A multiple sequence alignment of SF and S6 domains of ion channels published by Jensen et al. 2009 predicts the Asp 915 in line with analogous Asp/Glu of other TRP channels in the center of the selectivity filter and aligned to the tyrosine of conserved GYGD motif of the K^+ SF.

3.1.1. Selectivity filter (SF) structure prediction

The consensus sequence GYGD of the potassium selectivity filter (SF) is insistently preserved in potassium channels from bacteria to mammals. On the one hand, this indicates the exclusive necessity of potassium ion being strictly regulated, on the other hand, the invariance of selectivity filter residues shows that this combination of amino acids provides geometry and physicochemical properties required for exclusive potassium selection.

Changes in the amino acid composition of the SF which were accepted in evolution altered or changed the channel function. For example, the altered sequence (TVGDG) of the selectivity filter of the NaK channel changed the geometry of the selectivity filter and therefore selectivity itself changed. The first X-ray structure of a NaK channel provided detailed information about a structure of the SF which does not distinguish between NA^+ and K^+ ions. The very regular and straight SF of the potassium channel gained a different shape in NaK channels due to only one mutated amino

acid. The NaK channel preserved sites S3 and S4 of SF similar to KcsA channel, while sites S1 and S2 are no longer present due to different packing of main-chains and side-chains of the GDG segment. Instead of sites S1 and S2, GDG now forms one site were carbonyl oxygens are not oriented toward the inside of the filter (Shi et al. 2006) An additional ion-binding site is formed above, by the carbonyl group of Gly 67. The structure of NaK was solved in the presence of K+ as well as in presence of Na^+. In both cases the shape of SF is the same and ions occupy the same position, what additionally indicates that channel has no preference neither for K+ nor for Na+. In contrast, the SF of potassium channel at low K^+ concentration undergoes structural changes (Shi et al. 2006; Zhou & MacKinnon 2003).

Another example is the SF of NavAb channel. In comparison to earlier described selectivity filters of NaK and Kv channels, NavAb channel had lost all intristic potassium binding sites in SF (S1 to S4) and the length of SF became shorter. Instead of three twists of the SF loop in potassium channel Nav has got only 1 twist. Also, the diameter of the selectivity filter increased. In Kv the diameter was just wide enough to accommodate dehydrated K+ ion while in Nav there is space for conducting hydrated Na+ (Yarov-Yarovoy et al. 2011) Similar to NaK channels, a negatively charged residue is present in the middle of the SF loop. Mutation of this residue alters channel selectivity. Sequence alignment of NavAb Nav and Cav channels suggest that perhaps a similar position Glu or Asp is present in Cav channels as well. (Clayton et al. 2008; Yarov-Yarovoy et al. 2011). Another distinguishing feature of Nav channel is the presence of a second pore helix which connects the SF loop and S6 at the extracellular side instead of the loop as in Kv and NaK channels. (Yarov-Yarovoy et al. 2011)

3.1.1.1 TRPA1 SF based on various templates

Prior to the transmembrane domain modeling in order to better understand what conformation selectivity loop of TRPA1 could adopt, we utilized homology modeling to analyze the geometry of SF residues under different structural conditions. For this purpose we selected several available templates of TRP-related cation channels, such as Kv1.2, NaK, CNG and NavAb channels. All reasonable variants of alignments were prepared and modeled (**Fig.11**) All SF-models were inspected visually to reveal obvious clashes between residues. Geometry of each structure was analyzed with Procheck.

Kv-like selectivity filter

Two SF models were created based on the potassium channel. In the first variant the ASP 915 residue was aligned with TYR residue of 3LUT channel, while in the second variant ASP 915 was

aligned with the corresponding ASP of Kv channel. The overall G-factor value for the first variant model is -0.39, in comparison, the template has got -0.05. Also, the ILE residue C-N-CA angle highly deviates from the ideal value according to Ramachandran plot. In the second alignment the ASP residue was aligned with the aspartic acid residue of Kv channel. The G-value (+0.09) is largely improved for this model, phi and psi angles statistics does not recognize any residues which highly deviate from the standard values.

NaK-like selectivity filter

The amino acid composition of selectivity filters in the NaK channel and the TRPA1 channel is very similar. Both contain an ASN residue with an aromatic residue in the next position (TYR in TRPA1 and PHE in NaK) and a GD segment in the middle of the filter. However, modeling of the selectivity filter based on the NaK channel reveals difficulties in the positioning of ILE (in segment DIN of TRPA1). In all created models (10) the side chain of ILE is oriented towards the inside the pore and creates steric clashes. The least side chain overlap inside the pore is shown in **Fig.11**. The ring of four glycines residues at this position creates the narrowest part of the selectivity filter, where a residue with a large side chain as ILE cannot fit. The overall G factor of SF geometry in the template is +0.18, while in the model it is -0.38. Particularly, the mentioned ILE residue shows a high deviation of bond angles C-N-CA and N-CA-CB. Therefore, despite of the similar amino acid composition of selectivity filters, TRPA1 has a different structure of the selectivity filter.

CNG-like selectivity filter

TRPA1 and CNG channel have similar conducting properties, therefore it was predicted that TRPA1 has a similar selectivity filter features as well. The GD residues of TRPA1 were aligned with GD residue s of CNG channel (Wang et al. 2008). The resulting model showed good stereo-chemical values. None of the residues were found in disallowed regions and no deviation either for bond length or bond angles were detected. The overall G-factor for the representative model structure is -0.03, for template +0.18.

Nav-like selectivity filter

The selectivity filter of the sodium channel significantly differs from potassium channel (3LUT) and NaK/CNG channels (2AHY (NaK), 3K04 NaK/CNG mutant). According to the experimental studies the ASP 915 residue is predicted to be a part of TRPA1 selectivity filter. However, it is not clear whether ASP915 is located at the mouth of the SF, similarity as in Kv channels, or inside the filter like in other channels (GDGN (NaK), GETP (CNG), TLES (bacterial Nav)). Also, the Cav

channel is also predicted to have acidic residue inside the pore as in the NavAb channel.

Two models were built on basis of a voltage-gated sodium channel. In the first variant negatively charged residues were aligned (ASP in TRPA1 and GLU in NavAb). In the second variant the TRPA1 sequence was shifted by one residue forward to align ASP with SER. According to geometry analysis the first variant lead to a model with an overall G-factor -0.31 and two residues (ASN and TYR) with unnatural phi and psi angles. In contrast, the second variant gave a satisfactory resulting model with overall G-factor -0.11 and none of residues in disallowed regions of Ramachandran plot.

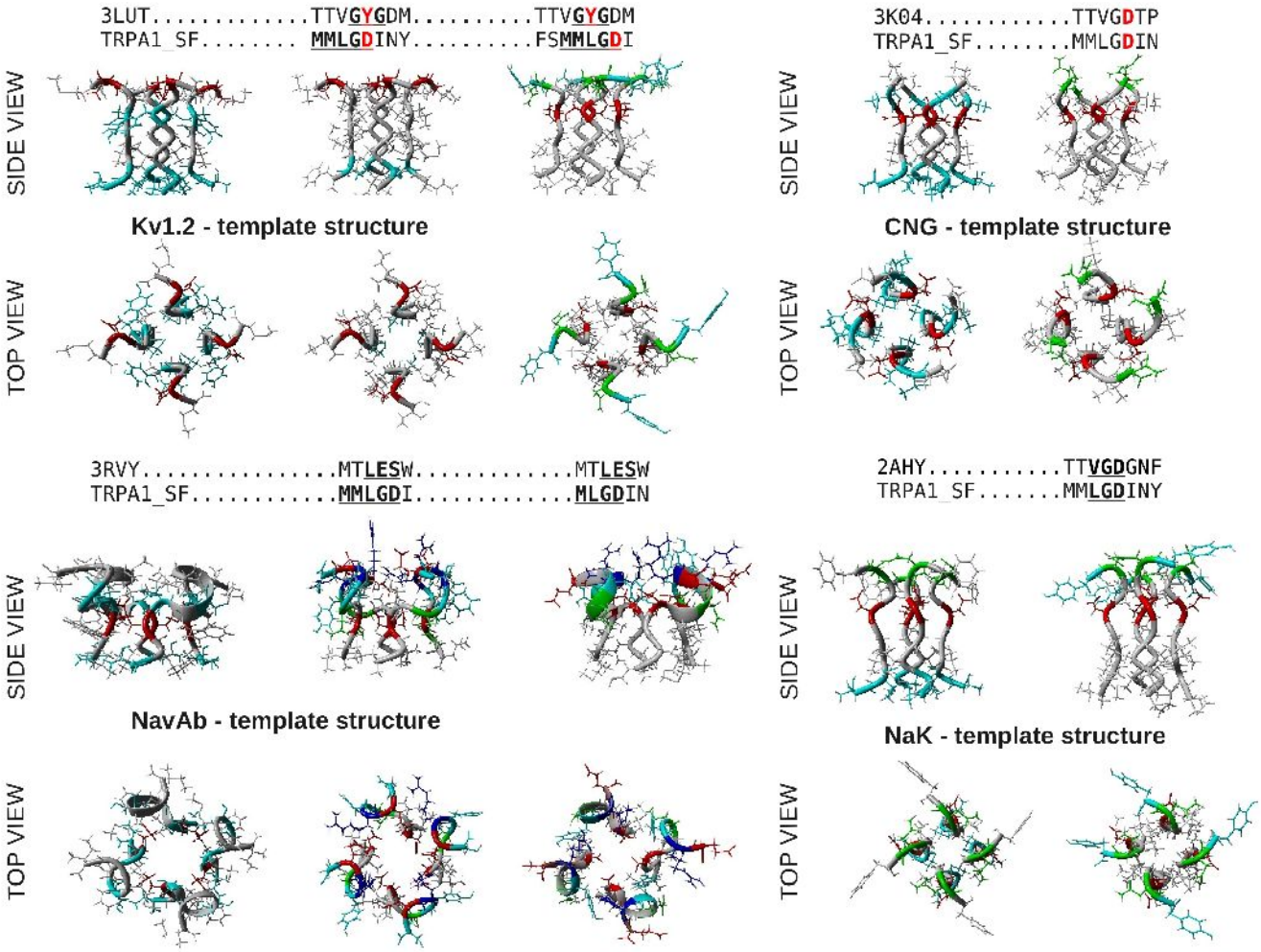


Fig.11 Homology modeling of the selectivity filter of TRPA1 based on various template structures.

Conclusions

Based on our analysis of the selectivity filter alone we could conclude concluded that the location of the negatively charged key residue D915 of the selectivity filter is rather located at the mouth of the selectivity filter than in the center of the filter, where it is more sterically restricted. All models with aspartates built inside of the filter resulted in unusual and unfavorable geometries.

3.1.2. Modeling of the TRPA1 transmembrane domain

Further we carried on modeling of the full transmembrane domain. Multiple sequence alignments of four TRP families were used to identify structural elements of the transmembrane domain. However, given the low sequence identity between target channels and template structure a single solution satisfying all condition and requirements can not be found easily. The possibility of any conservation pattern was investigated across all sequences in every TM helix. Several reasonable hints were found, where we could fix and improve the alignments.

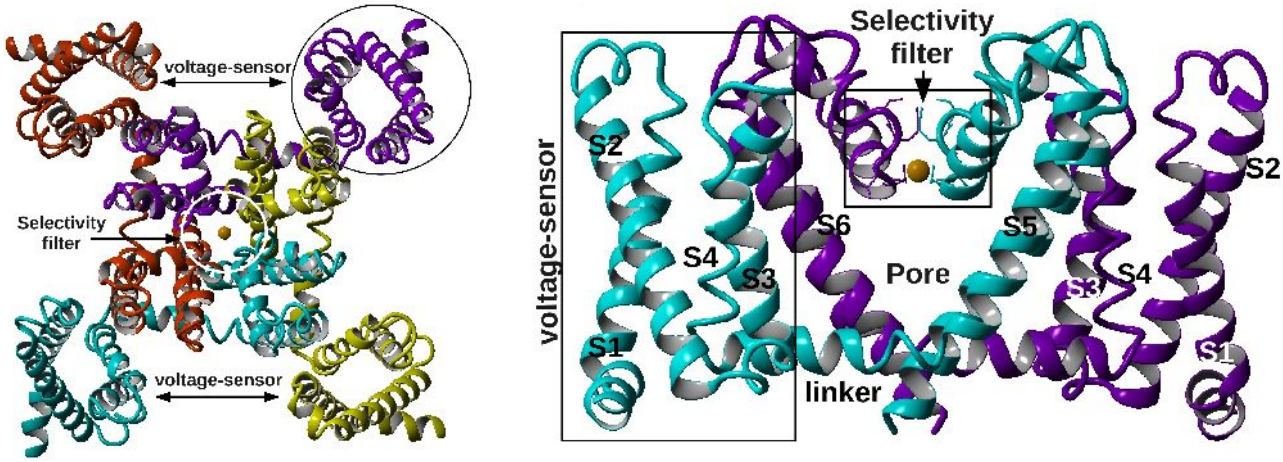


Fig.12. Structure of voltage-gated sodium channel (PDBID:3RVY). Panel A: top view, B: side view. Figure prepared with Yasara.

TM helix S1 was aligned in a way to match the regions of hydrophobic residues, while keeping charged residues at the edges. At the N-terminus of Nav/Cav, TRPM and TRPC families Pro residue presumably could serve as a start of the TM helix.

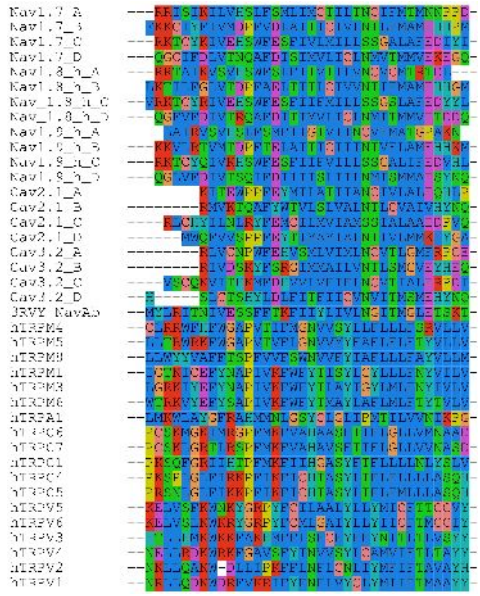


Fig.13 Multiple sequence alignment for TM S1 helix.

TM helix S2 is the second helix of the voltage-sensor ans has a clear pattern of E-XXX-K. TRP

channels have got a region of charged residues which reminds E-XXX-K of Nav/Cav channels, therefore sequence were aligned to match this pattern. Additionally, the N-terminus of S2 reveals a conservation of negatively charged GLU prior to the hydrophobic part. (**Fig.14**)

In **TM helix S3** a highly conserved N-XX-D residues were found in TRPM, TRPMC and TRPA1 and partially appeared in TRPV families. Therefore sequences were aligned in order to match those residues.(**Fig.14**)

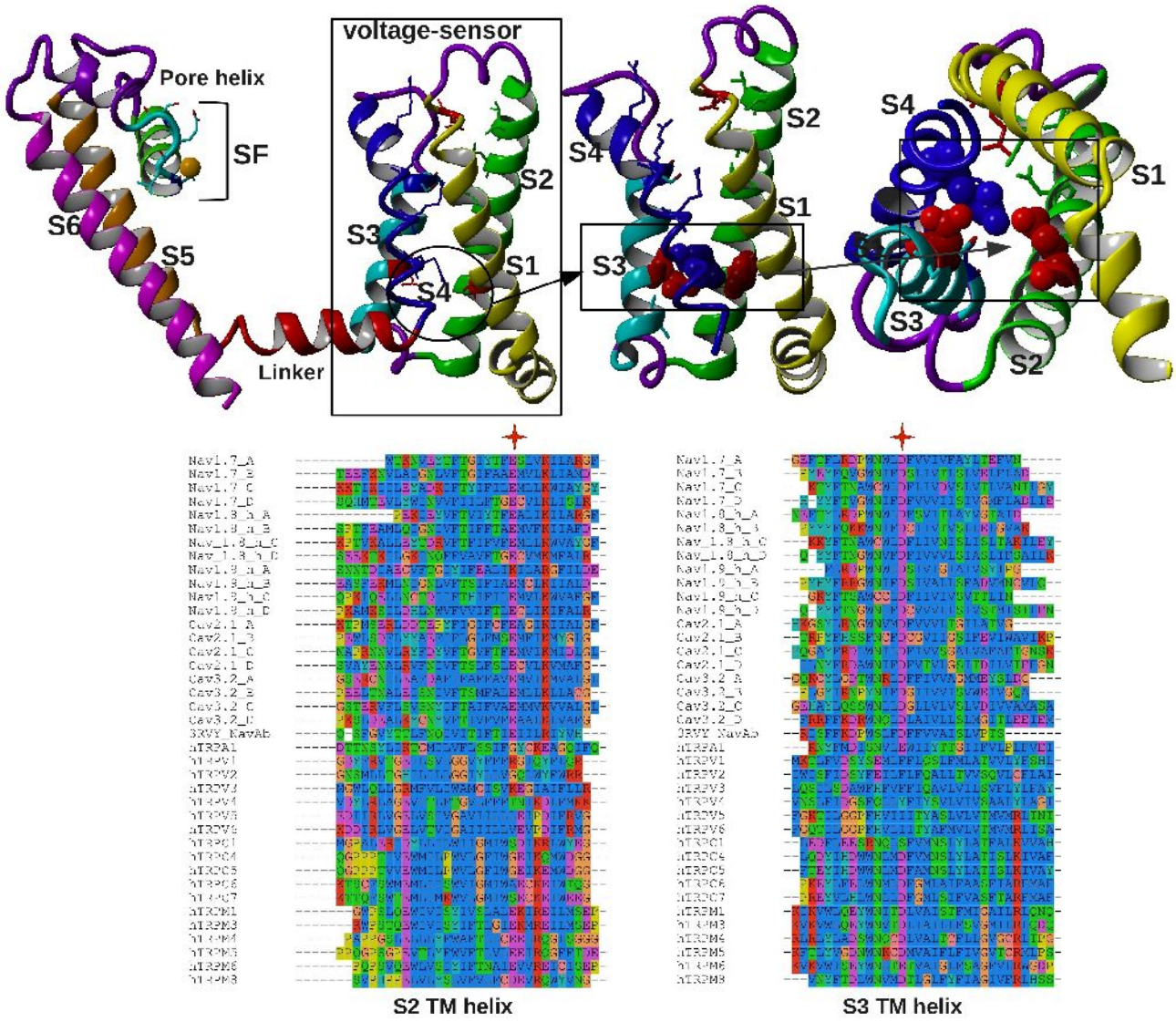


Fig.14. Structure of a single monomer of NavAb (conserved GLU from S2 and ASP from S3 are shown in ball representation). Bellow corresponding multiple sequence alignment for transmembrane helices S2 and S3. Conserved GLU from S2 and ASP from S3 are marked with a star sign.

Alignment of the S4-S5-linker and following S5 TM helices as well as the alignment of the selectivity filter was the most challenging part of the alignment, as in these cases no single solution could be found which would be satisfactory for all TRP channels.

TM helix S6 contains a highly conserved ASN residue across all Nav/Cavs and the TRP family with the only exception of the TRPC1 channel. (**Fig.15**)

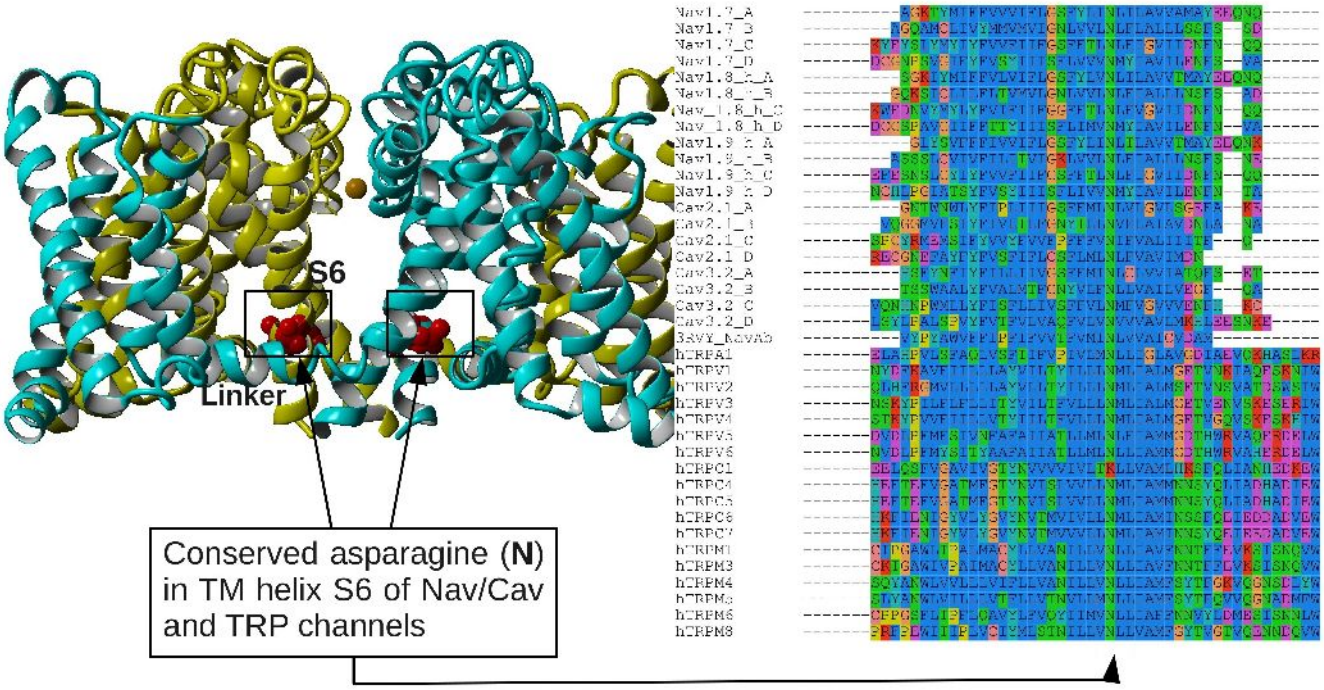


Fig.15 (right) Structure of NavAb (3RVY). Conserved ASN is shown in red. (left) Multiple sequence alignment for the S6 helix throughout the Nav/Cav and TRP families.

3.1.2.1 Modeling of S4-linker-S5 helices

Taking everything into account, we kept the alignment of S1, S2, S3 and S6 helices unvaried and focused further on S4-linker-S5 region. With the help of homology modeling we aimed to answer questions regarding structure and function of TRPA1 channel.

Do TRPs have a voltage-sensor?

Voltage dependence of TRP channels is not yet determined as their response to voltage is not obvious as in Nav/Cav/Kv channels (Sharman et al 2013). Evidently, TRP channels do not possess a classical voltage-sensor, since the consensus four R/K residues are not present in any of the TRP. However, there are arginine, lysine and other polar residues in the corresponding part of the TRP sequence. Therefore, we tried to identify if some of these charged residues could occupy the same

positions as classical voltage-sensing R/K. Multiple sequence alignments revealed conserved negative charges in S2 and S3 helices, which could couple with positive charges from S4. Therefore a model should allow R/K residues from helix-sensor to interact with the corresponding residues of S2 and S3. In order to perform its function, a typical S4-sensor (as in voltage-gated channels) partially adopts a 3_{10} helical conformation. Since TRPs do not have 4xR/K pattern in S4, does it exist in 3_{10} or alpha-helical conformation?

Where is ASN 855 located, a residue responsible for familial episodic pain syndrome?

In our previous TRPA1 model based on Kv1.2 (PDBID: 3LUT) we predicted the position of ASN855 residue in the loop connecting S4 helix and linker, where it could possibly make a link via an interaction with residue ASN 722 of S1. Mutation of this residue to SER produces a channel with an increased current flow in comparison to the activated wild type channel, what causes the familial episodic pain syndrome (Kremeyer et al. 2010). The structure of the S4-linker-S5 helices in NavAb differs from Kv1.2. The linker itself consist of four turns, instead of three in Kv1.2 structure, therefore junctions between S4 and the linker also differ. That is why the previous conformation might not be acceptable in a new model.

Why menthol activated human TRPA1, but blocks mice TRPA1?

Transmembrane helix S5 is involved in menthol sensitivity of TRPA1. TRPA1 reacts on menthol in a species-dependent manner, human TRPA1 is activated in response to it, mouse TRPA1 is blocked at high menthol concentrations, while *Drosophila* TRPA1 in not sensitive at all. In a search for the "menthol" site in TRPA1, replacement of the mouse or human S5 domain with *Drosophila* S5 produced a chimeric channel that did not respond to menthol (Xiao et al. 2008). Multiple sequence alignment showed 17 amino acids to be different between mammalian and *Drosophila* proteins. Nine residues within the S5 helix were shown to be in charge of menthol sensitivity in mice and human as distinct from *Drosophila*. However, only two residues were shown to be responsible for the different answer to menthol between human and mouse channels. Particularly, VAL875 is responsible for species-specific response to menthol, while residues SER873 and THR874 determine sensitivity to menthol. This could mean that while residues 873 and 874 are accessible for menthol molecule VAL875 is potentially involved in interactions within the protein.

Structural aspects. G&P-helix-breaking residues

Additional information we need to take into account for modeling is the presence of helix-breaking residues such as proline and glycine. All TRP sequences contain glycine residue and most of them

(but not all!) in addition proline residues in the next position (GP). Combinations of these two could be most probably found in the turn or loop region rather than in helix, unless this motif has some particular function.

Structural aspects. Charged residues.

The location of positively charged residues is of particular importance in membrane proteins. A charged residue within the membrane should be involved in interaction with another charged residue or the side chain of charged residue should face the solvent and not the membrane. Charges within hydrophobic regions would destabilize the protein. All TRP channels contains multiple charged residues within the S4-linker-S5 region, that are not all conserved. The model aims to find a solution where all charges in each TRP channel are “neutralized”.

In order to find a reasonable structural solution for all TRP channels, three variants of alignment were selected according to several criteria. Every variant has its advantages and disadvantages, that are described in detail in the following paragraphs, and investigated in their structural consequences using homology modeling.

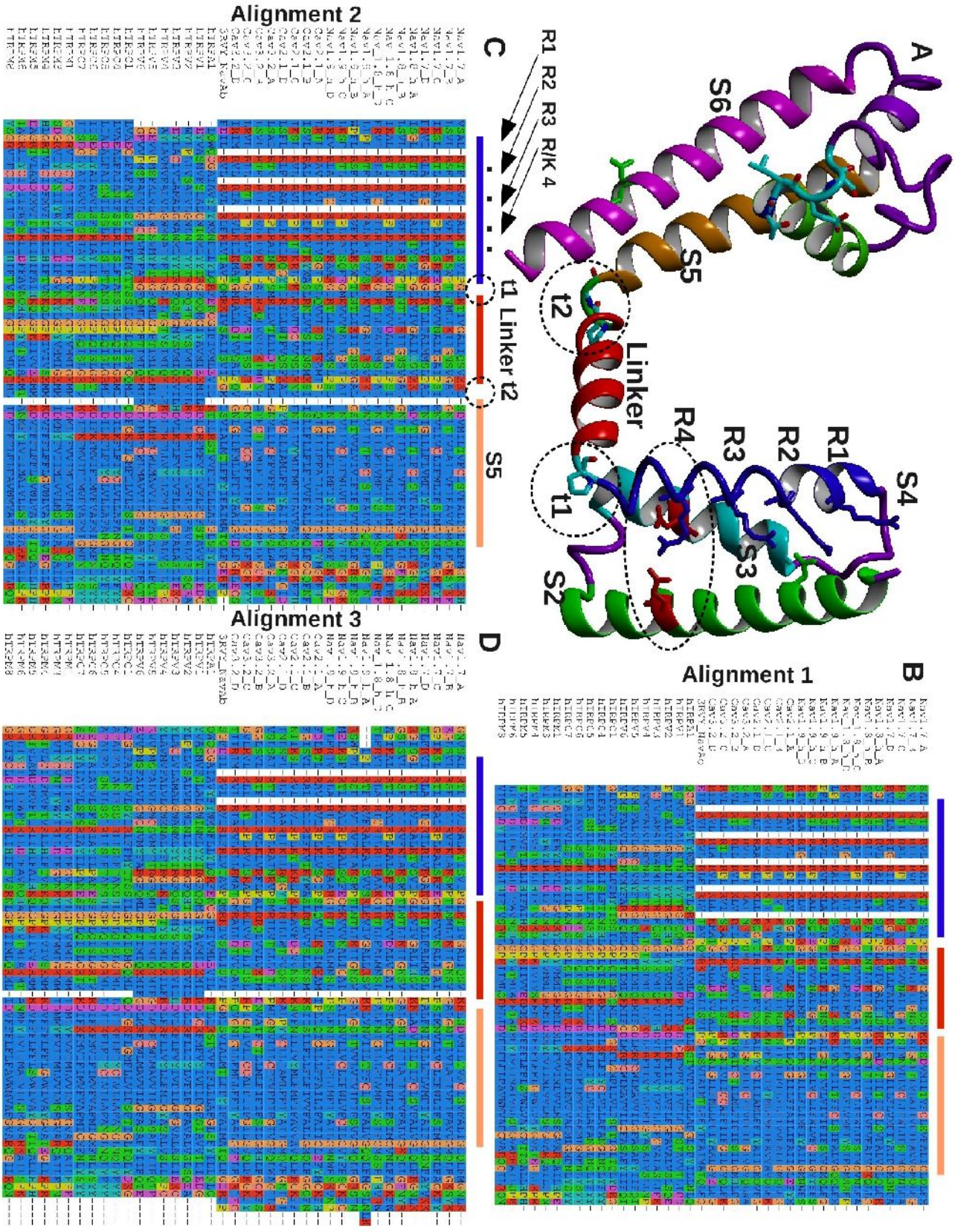


Fig.16. A. Single monomer of NavAb (3RVY) structure, helices S2 to S6 are shown in different colors. **B.** Multiple sequence alignment variant I. **C** Multiple sequence alignment variant II. **D** Multiple sequence alignment variant III.

Three variants of alignment for TRP channels modeling (Fig. 16).

Alignment I (Fig.16 B)

The first alignment was based on following reasoning:

1. In this alignment we predict that the S4 helix is an alpha-helix due to the lack of conserved 4xR/K-pattern of the voltage-sensor.
2. ASN 855 of TRPA1 is located in the turn between TM S4 and linker as it was predicted in the previous TRPA1 model, where coupling of ASN 855 to ASN 724 of the S1 helix could create a strong link between N-terminal domain and S4-S5 linker helix.
3. Charged residues from the linker are oriented towards the protein so that they could interact with conserved ASN from helix S6.
5. Conserved charged residues RQ at the C-terminus of S5 helix in TRPM family could possible interact with negatively charged residues from S4 helix.
6. Conservation of the third ARG residue in the voltage-sensor which could interact with negatively charged residues of helices S2 and S3.
7. Helix-breaking residues GP are placed within the turn S4-linker.

Alignment II (Fig.16 C)

The second alignment was based on following reasoning:

1. S4 helix is partially 3_{10}/alpha helix.
2. Conservation of one salt bridge in the voltage sensor which is aligned well with the template NavAb, were fourth ARG interacts with conserved S2 and S3 helices.
3. According to this alignment the charges in the linker coincide well with the charged residues in template.
4. The template structure 3RVY contains PQ residues in the turn between S4 and linker, in this version of alignment some glycine and proline residues in TRPs appears as well.
5. Conserved RQ motif of TRPM families moves up along the helix and faces lipids head group, where it could have a favorable environment.
6. There is a conserved charge in Nav/Cav and Kv channels in the turn between linker and S5, which can be aligned well with TRP proteins.

Alignment III (Fig.16 D)

The third alignment was based on following reasoning:

1. S4 helix is partially 3_{10}/alpha helix, which could be required for voltage response.
2. Conservation of the one salt bridge in the voltage sensor between charged residue at position of R3 and conserved negative charges from S2 and S3 helices.
3. The conserved helix-breaking GP motif is placed in the beginning of the linker helix, in a similar way as PG motif in NavAb at the C-terminus of the linker.
4. In this alignment the conserved RQ residues of TRPM channels moves up along the S5 helix, so they are no longer present within hydrophobic region.
5. Charged residues K/R of TRPA1/TRPCs/TRPVs are aligned with the conserved charge of the turn between linker and S5 helix in Nav/Cav.

Model preparation

At the next step of homology modeling of TRPA1 we utilized Modeller to analyze the structure prediction based on our multiple sequence alignment. For every variant of the alignment a model of every TRP channel present in the alignment was generated in order to find a solution that would be satisfactory for all members of TRP superfamily (18 models). All models were visually inspected, superimposed and compared using Yasara and VMD. The analysis of stereo-chemical parameters was done in Procheck.

3.1.2.2 Analysis of TRP models based on NavAb channel

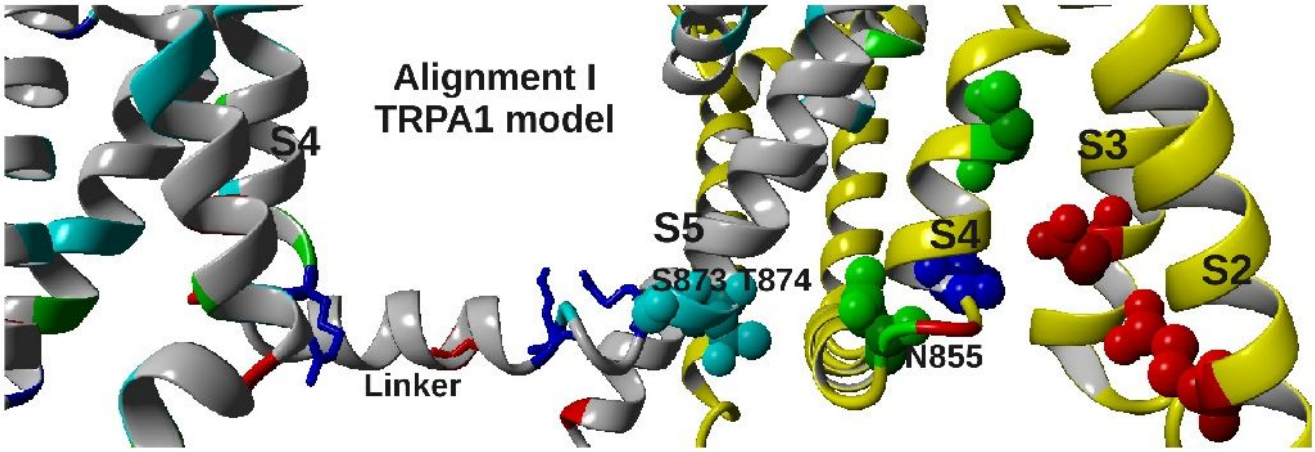


Fig.17 Magnified view on the model of TRPA1. Asparagine residues are in green color, ball representation, positively charged residues are in blue, negatively charged residues are in red color.

Model based on alignment I (Fig.17)

1. Putative interaction within the voltage-sensor between ARG852 and conserved negative charge from TM S2 and TM S3 (GLU788 GLU 808) is not formed.

2. There are multiple charged residues within the linker helix in most of TRPs. In this case, they are oriented inside the protein towards helix S6 (not shown on figure X), which contains only one charged residue at this position, a conserved ASN 954, which can not simultaneously create contacts with three and more residue.

3. ASN 855 within the turn between S4 and linker, with the distance approximately 9 Å (Ca-Ca atoms) from the ASN 722 helix S1 (6 Å between side chains) and approximately 11 Å (Ca-Ca atoms) from THR 874 of the linker-S5 turn.

4. Menthol-sensitive residues SER873 and THR874 are placed within the turn between linker and S5 helix. In this position SER-THR residues could be accessible for menthol molecule from the cytosol.

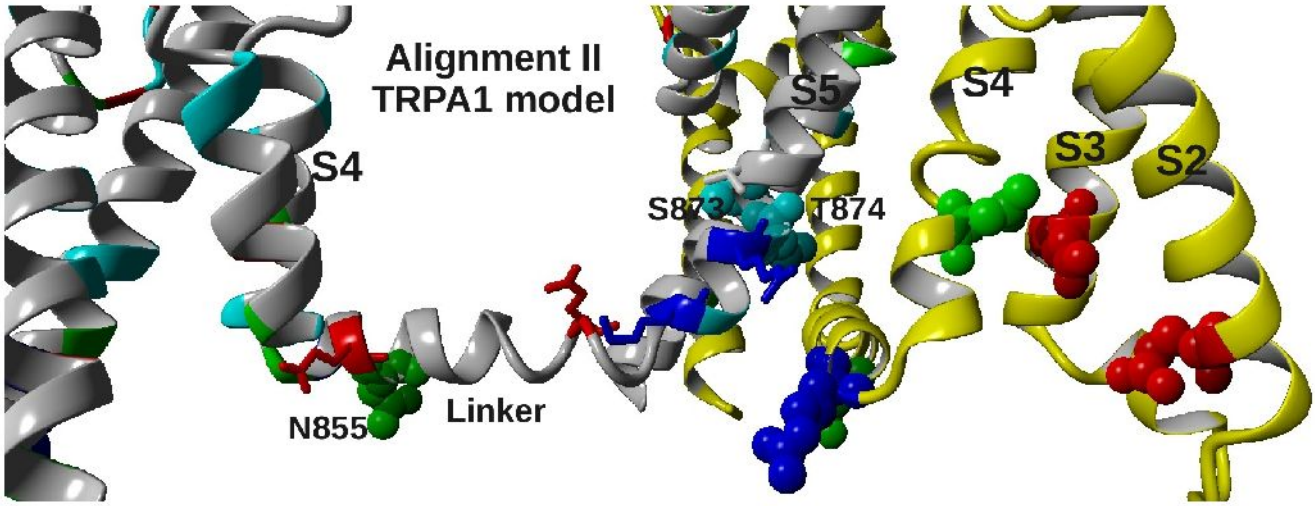


Fig.18 Magnified view on the model of TRPA1. Asparagine residues are in green color, ball represantation, positively charged residues are in blue, negatively charged residues are in red color.

Model based on alignment II (Fig.18)

1. TRPA1 model does not have any ARG852 residue within S4-helix, instead position R4 occupied by asparagine.

1. Multiple charged residues in most of TRP families are located within first and second turns of S5 helix.

2. ASN855 is oriented to cytoplasm, where it does not interact with the protein.

3. Menthol-sensitive residues SER873 and THR874 are located in the third turn of S5 helix, where they are perhaps not very accessible for interaction with some molecules from the cytoplasm.

4. Gly-Pro helix-breaking residues are in the second turn of the linker.

5. TRPC have an ASN residue in the S4 helix oriented outside of the voltage-sensor oriented towards the hydrophobic environment.

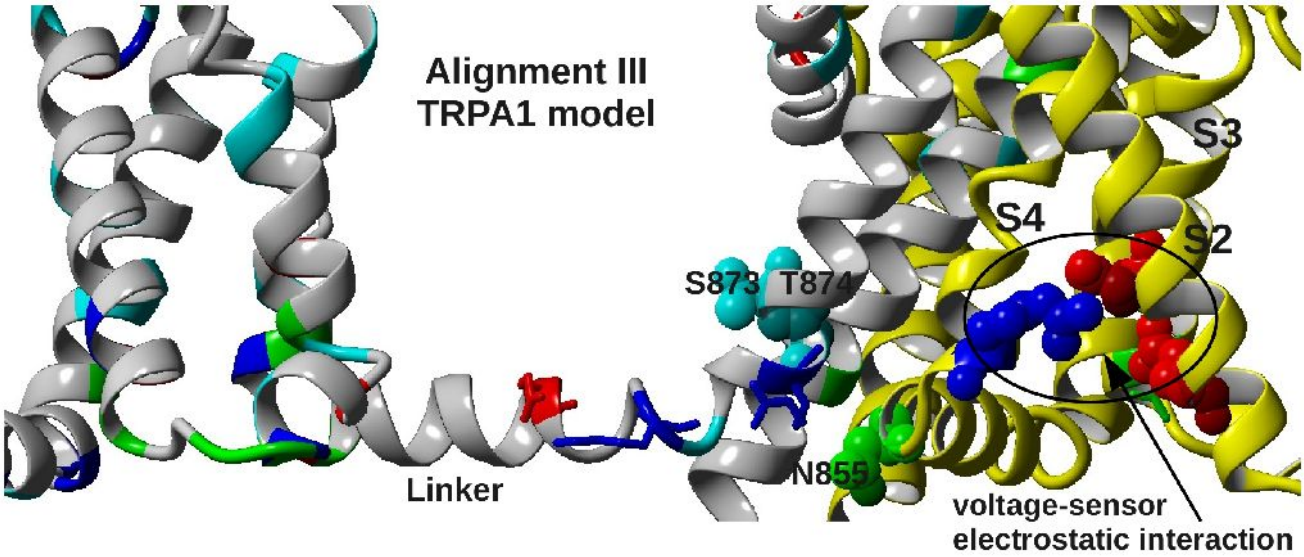


Fig.19 Magnified view on the model of TRPA1. Asparagine residues are in green color, ball representation, positively charged residues are in blue, negatively charged residues are in red color.

Model based on alignment III (Fig.19)

1. Menthol-sensitive residues SER873 and THR874 are located in the first turn of S5 helix, where they are accessible from the cytoplasm and the internal pore.
2. ASN 855 is the first residue in the linker-helix. At this position the ASN855 Ca carbon is located 7Å from residue ARG872 of the S5 linker in TRPA1.
3. A salt bridge within the voltage-sensor is formed between S4 ARG and S2-S3 helices.
4. Multiple charged residues of the linker helix do not interfere with the hydrophobic environment.

Conclusions

The main disadvantage of the model based on alignment I is in the position of the positively charged residues within the linker helix, where they do not have interacting partners. In this case, when the charges are not complemented, these residues could destabilize the whole S6 helix. The major problem of alignment II solely based on sequence information is the location of the helix-breaking residues GP within the alpha-helix (the second turn of the linker). Another disadvantage of this alignment lies in the fact that the multiple charged residues from the linker are now shifted to the beginning of the S5 helix, where they again create a charge within a hydrophobic region, that could destabilize the protein structure. Modeling of TRP channels based on alignment III lead to the best solution for multiple charged residues in every built TRP channel. The final TRPA1 model shows that residue ASN855 could be located near the ARG872 from the firth turn of the S5 helix. C alpha carbons are located a distance of 7Å from each other and the side chains of residues are long

enough for interaction. Additionally, the third alignment predicts a partially 3_{10} S4-helix and preserves the only electrostatic interaction within the voltage-sensor between ARG 852 and negatively charged GLU788 and GLU 808 from S2 and S3 helices.

Afterword

Shortly after the TRPA1 model based NavAb structure was modeled, the TRPV1 channel structure as the first of TRP superfamily was determined by electron cryo-microscopy (Liao et al. 2013). Although, TRP channels are very diverse in sequence, the structure of TRPV1 provides a great opportunity to model more precisely the structures of other TRP channels. We analyzed the results of TRP channels modeling based on NavAb channel and compared with the structure of TRPV1.

3.1.2.3 Superimposition of TRPV1 model and TRPV1 structure

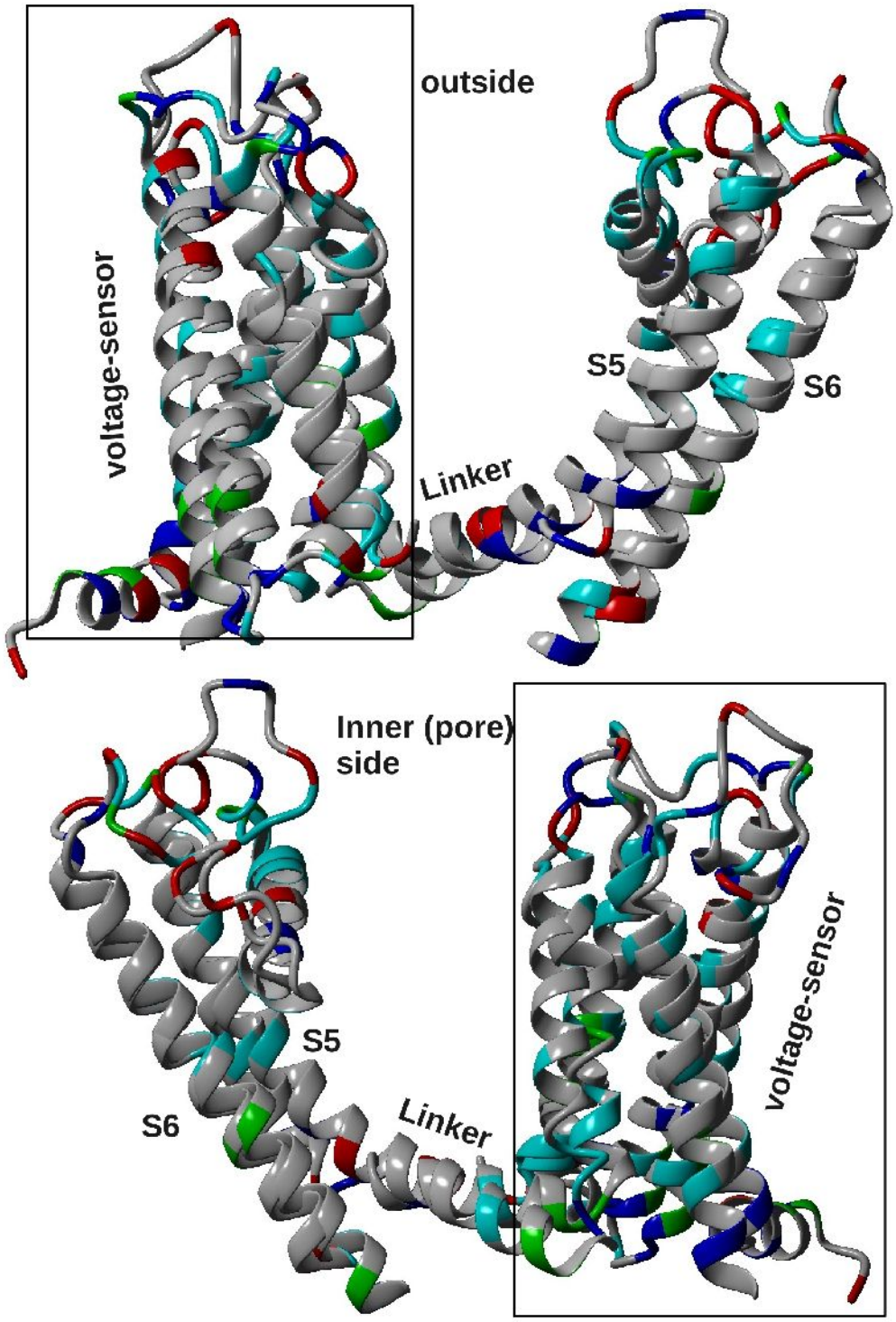


Fig.20. Superimposition of TRPV1 model based on NavAb and structure of TRPV1.

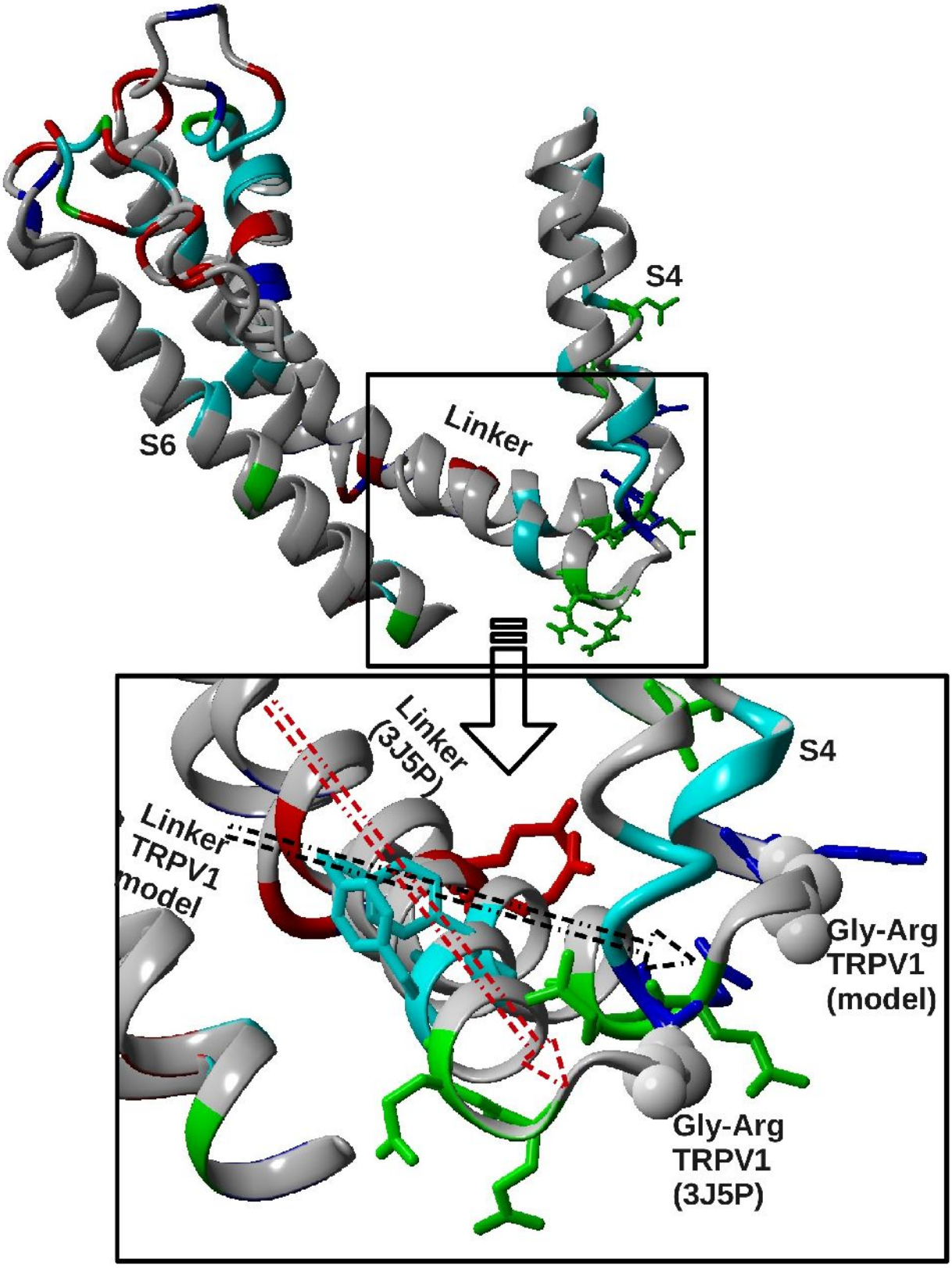


Fig.21. The difference in the TRPV1 structure and TRPV1 model.

3.1.2.4 Homology model of TRPA1 based on TRPV1 structure

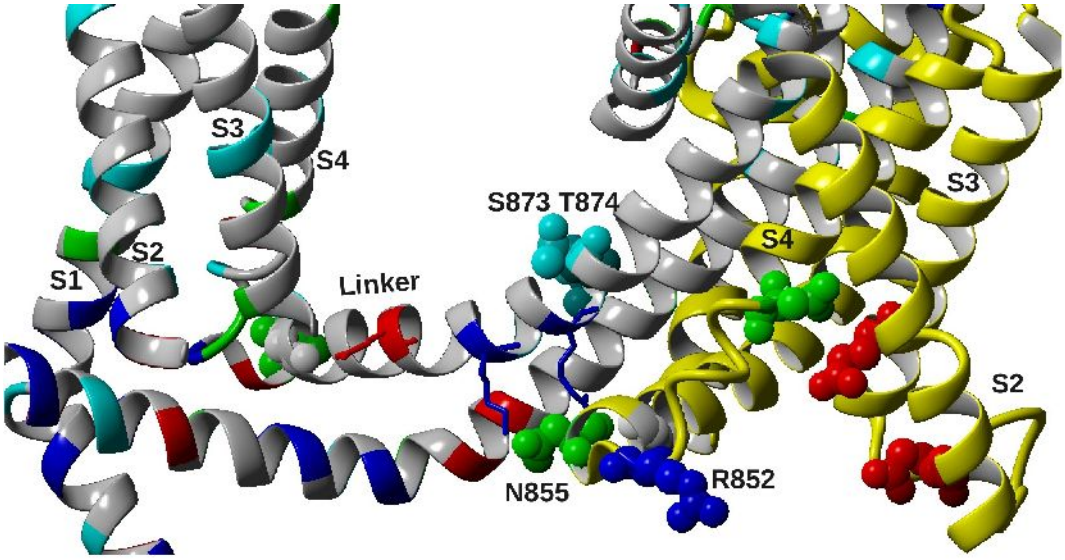


Fig.22. Homology model of TRPA1 based on TRPV1 structure. Magnified view. Asparagine residues are in green color, ball representation, positively charged residues are in blue, negatively charged residues are in red color.

In general, the structure of the TRPV1 channel is organized in a similar way as the Kv and Nav/Cav channels as was predicted earlier. Four monomers each consisting of six TM helices S1 to S6 are organized symmetrically around the central axis. The architecture of the selectivity filter reminds the wide filter of NaAb channel. S1-S4 helices form a voltage-sensing domain. S5-S6 helices form the cation conducting pore. However, the S6 helix does not go straightly to the cytosol, but make a turn and "enters" the voltage-sensor domain from the bottom, so that the helical C-terminus goes up. The linker helix and the C-terminus of S6 helix are very rich in charged residues.

The structure of TRPV1 shed light on many questions regarding the structure and functions of the whole TRP superfamily and offered the solution for the questions we have asked prior while doing the TRPA1 modeling using the NavAb structure. While the S4 helix of NavAb consists of two alpha-helical and three 3_{10} helical turns, S4 in TRPV1 makes four alpha-helical turns and two 3_{10} turns at the C-terminus. The junction between the S4 helix and linker in NavAb and TRPV1 differs as well. There is a three-residue turn between S4 helix and the linker in the TRPV1 structure, while only one residue separates S4 and linker in sodium channel. Due to a longer turn between S4 and the linker ARG 852 does not make a contact with S2 and S3 helices, but rather faces the lipids head groups. In a TRPA1 model based on TRPV1 (**Fig.22**) ASN 855 would be located in the first turn of the linker helix in a close proximity of ARG 872 (Ca-Ca distance is 8 Å), similar as it was predicted

in our model based on alignment III. All charged residues of the linker helix face the cytoplasm or other charged residues of the prolonged C-terminal helix of S6. Menthol-sensitive residues S873 and T874, however, are not easily accessible from the cytosol according to the actual conformation of the channel. Nevertheless, our model based on NavAb structure was surprisingly close to the new crystal structure and most conclusions we draw can be still considered as correct.

3.2 Prediction of a potential pleckstrin homology (PH) domain

A full pleckstrin homology domain (Haslam et al. 1993) consists of approximately 120 amino acids. Although the amino acid sequence of the domain is not conserved, all PH domains fold into a canonical structure consisting of two perpendicular beta sheets followed by an alpha helix. PH domain preserves a canonical fold despite the sequence variability and despite of possible insertions of other domains. Moreover, an insertion of another domain can interrupt the PH domain in different places. Although, in the primary sequence such a domain is splitted into two halfs with the insertion in between, it does not affect the canonical fold of PH domain (Wen et al. 2006; Koshiba et al. 1997).

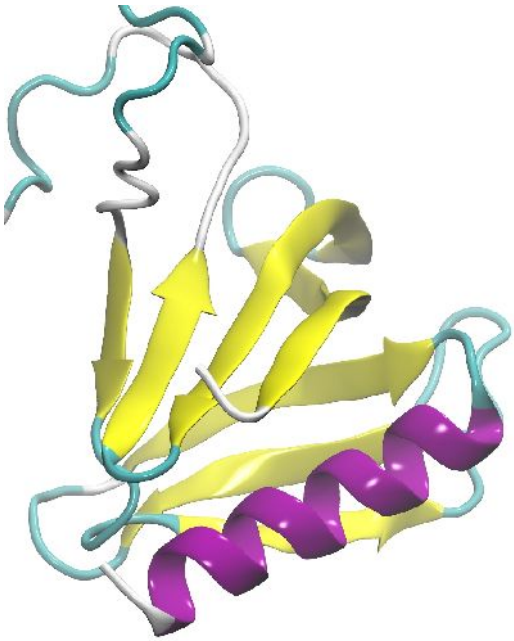


Fig.23 Canonical fold of pleckstrin homology domain (figure prepared in VMD).

3.2.1 Sequence analysis and conserved domain prediction

A putative mechanism of TRPA1 channel regulation by PIP_2 via a hidden N-terminal part of an PH domain was proposed for TRPA1, where the N-terminal PH-half couples with the C-terminal domain of PLC gamma 1 in order to create an inter-molecular domain which then would bind PIP_2. We analyzed all available information in order to reveal a hidden pleckstrin domain in the TRPA1 N-terminal domain. Our final aim was to complex an N-terminal PH-domain model of TRPA1 with the C-terminal PH-domain structure of PLC gamma 1 in order to test the stability and robustness of the interaction and the possible binding of PIP_2.

The majority of PH domains capable of lipid binding possess a conserved lipid-binding motif

consisting of several positively charged residues (K-Xn-(K/R)-X-R). Four split PH-domains were predicted based on the conserved PIP_2 binding motif for the TRPA1 channel (Karashima et al. 2008). We analyzed all four predicted PH domains using multiple sequence analysis of only mammalian members of TRPA1 (rat, mouse and human) in a first step. As a result we found that only two of the predicted PH-domains (II and III) are conserved within mammalian TRPA1s, while the other were not even conserved within this group making them highly improbable. However, both of the conserved predictions coincide with the ankyrin repeat domain, which is again highly improbable, as the ankyrin repeats cannot resemble the canonical PH domain fold and thus there would be either an ankyrin repeat or a PH-domain.

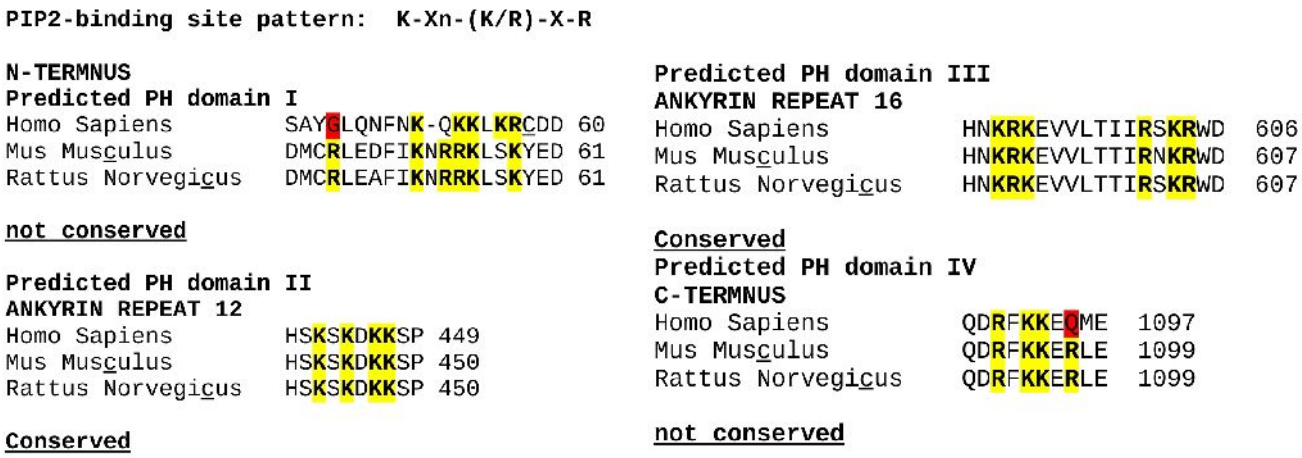


Fig.24 Predicted PH domains in TRPA1 sequence based on canonical lipid binding site.

At the next step, we repeated the procedure described by B. van Rossum (van Rossum et al. 2005) on TRPA1 protein to search for a potential hidden PH- domain in our target sequence. The complementary N-terminal PH-domain half was predicted for residues 591-634 (**Fig.25**). This prediction also coincides with the third PH domain (III) (**Fig.24**) identified based on lipid-binding site consensus sequence. To be able to get a feeling for the quality of our prediction, we then checked the robustness and sensitivity of the NCBI domain prediction algorithm (Marchler-Bauer et al. 2007) that we used by *"in silico"* mutating a single amino acid in the predicted domain at a time. Making a single change, the algorithm no longer gave a prediction for a complete PH domain. Our analysis showed that the prediction does not happen by chance nor is very tolerant to variations, but is very specific for the sequence.

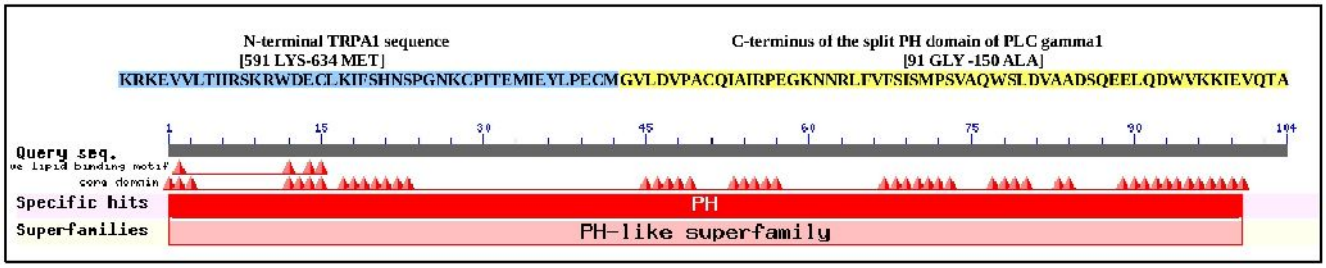


Fig.25 Conserved Domain Database prediction of complete PH domain

In order to identify if the N-terminal domain could contain regions with beta-stranded secondary structure (as ankyrin repeats are helical in contrast to the PH-domain) we made the secondary structure prediction. Few beta-strands were predicted for two regions within N-terminus: [LYS7 to SER43] and [ALA572 to THR624]. The second prediction partially overlapped with the NCBI CDD prediction of a full PH domain.

Taking all together these results suggest that a putative hidden N-terminal part of a split PH-domain could be located within residues 591-634 as it was predicted using Conserved Domain Database (CDD) search and the conserved lipid-binding site. Additionally, the secondary structure prediction also shows a reasonable probability for beta-stranded structure in the same region.

3.2.2 PH domain modelling and simulation

The template for homology modeling (PDBID: 1PMS) was taken as it was predicted by CDD results. We prepared a model of the N-terminal ankyrin repeat domain together with the predicted PH-domain in order to analyze whether the N-terminal half could preserve secondary structure within another domain. A set of models was built in Modeller and analyzed with Procheck. The model with appropriate secondary structure and reasonable quality of the overall structure was selected for MD simulations. Molecular dynamics was performed in a standard way including solvation, energy minimization, equilibration of the system following by production run. Molecular dynamics results showed the increase in root mean square fluctuation (RMSF) starting with amino acids 450, the region which contains the putative PH domain (**Fig.26**). The model of the N-terminal domain with an PH domain does - in contrast to the ankyrin repeat domain simulated with or without calcium ion in its EF-hand- not equilibrate during 20 ns of simulation (**Fig.27**).

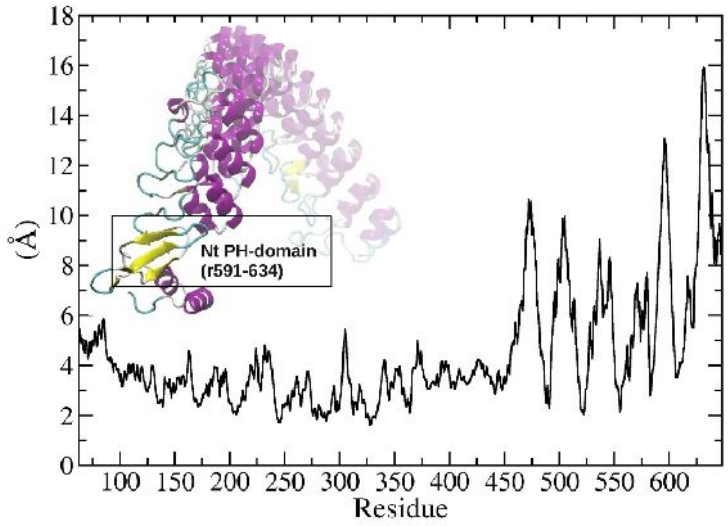


Fig.26 Root mean square fluctuation of N terminal ankyrin repeat domain with inserted PH domain.

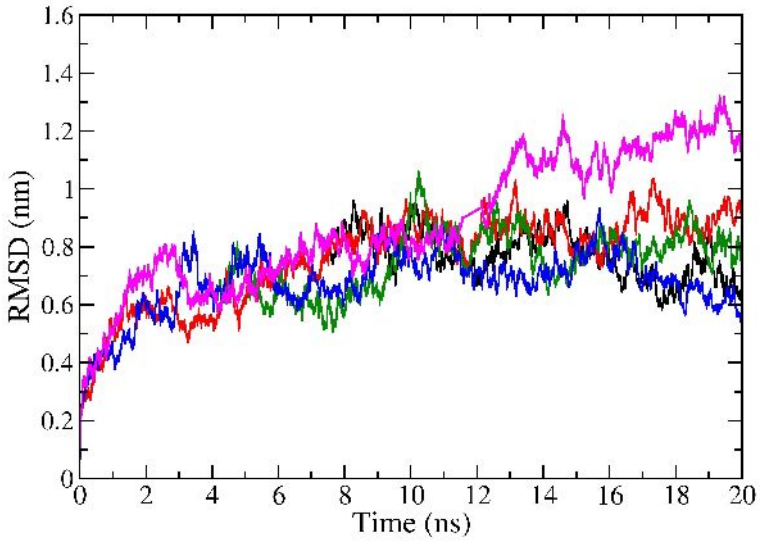


Fig.27 Root mean square deviation (RMSD) of the N-terminal domain built together with predicted PH domain (magenta) in comparison to RMSD of N-terminal domain with EF-hand domain in presence (black and red) of calcium and without (blue and green).

CONCLUSIONS

The existence of a hidden half of a PH-domain is highly improbable for the N-terminal of TRPA1 and even the site with the highest probability has been demonstrated to be less stable in molecular dynamics than an alternative ankyrin repeat in the same place.

3.3 Docking of cholesterol to hOrai1 model

Cholesterol plays an important role in ion channel regulation, trafficking proteins into the membrane and their distribution in the membrane (Levitan et al. 2010; Baier et al. 2011). An increased amount of cholesterol in the membrane has been shown to suppress activity of many voltage-gated channels, while depletion can inhibit functions of TRP and some types of sodium and potassium channels (Levitan et al. 2010). The association of STIM and Orai channels involved in store-operated calcium entry takes place within lipid rafts rich in cholesterol. Also, it has been shown that association of STIM and TRPC1 proteins within lipids rafts is regulated by the ER calcium stores (Derler et al. 2012; Dionisio et al. 2010).

Cholesterol interacts with membrane proteins via specific cholesterol binding motifs, such as CRAC, CARC, CCM or a tilted peptides (Fantini et al. 2013; Baier et al. 2011)

The CRAC motif is very common among membrane proteins. It possesses a consensus sequence (**L/V-$X_{(1-5)}$-Y/F-$X_{(1-5)}$-R/K**) and locates at the protein-lipid interface of a transmembrane domain. Also, it was shown that an inverted CRAC motif, named as CARC also exists. The CARC motif, in contrast to CRAC, is oriented upside down with the positively charges residue (R or K) being at the starting position of the motif (**R/K-$X_{(1-5)}$-Y/F-$X_{(1-5)}$-L/V**) (Baier et al. 2011). Cholesterol as part of the membrane should be located either in outer or inner leaflet of the membrane with its head groups (-OH group in cholesterol) being in a row with the head groups of the other lipids in the upper or lower leaflet.

The role of cholesterol in regulation of Orai1 channel was studied in our collaborator's group (Christoph Romanin, Linz) and motivated the current study to predict a potential interaction site of cholesterol with Orai1.

In order to identify potential binding sites for cholesterol in our model of human Orai1 we performed global docking. Global docking is aimed to predict an unknown binding site, therefore the whole protein was analysed, to allow search inside of the protein, on the surface and in inter-facial regions. According to our docking results, the most populated pocket for cholesterol binding was found on protein-lipid interface. **Fig.28.A**

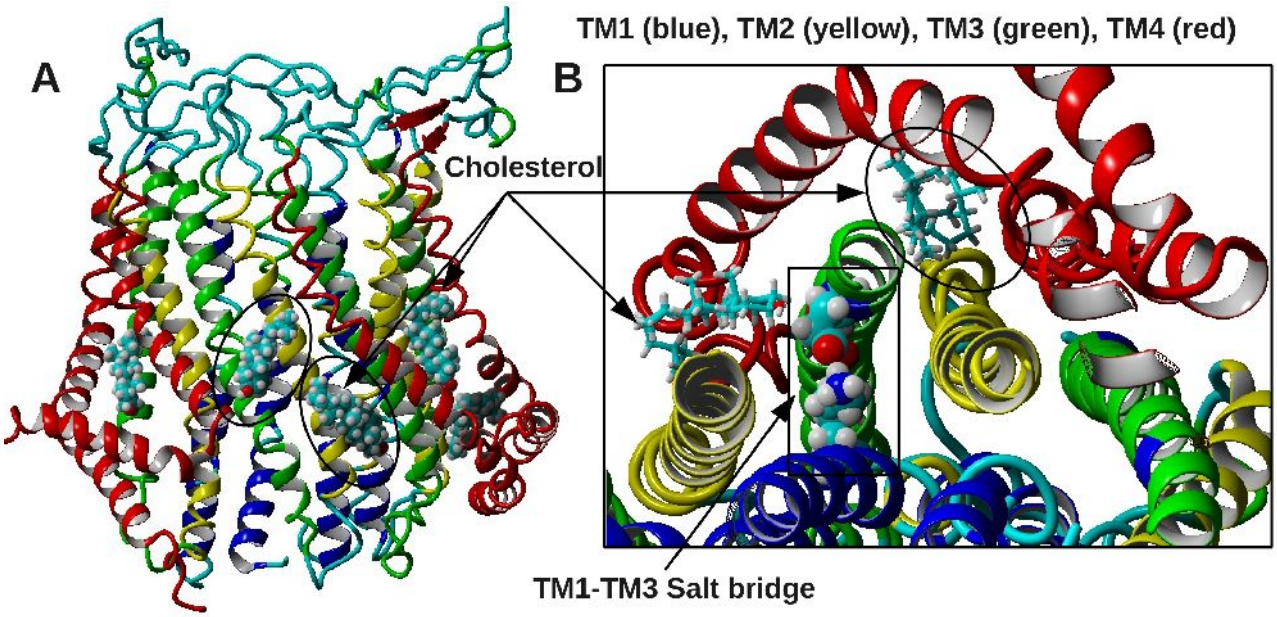


Fig.28.A. Global docking of cholesterol. B. Magnified view on docked molecule of cholesterol from the bottom.

Cholesterol molecules are bound to helices TM2 and TM3. Due to a bent of the helix TM4, Orai1 structure possesses two structurally different surfaces: a) a dimer interface with crossed TM4 helices and b) a dimer interface without crossed TM4, which creates a kind of groove on the surface. The surface with crossed helices was more populated with docked cholesterol molecules in comparison to the other one. However, in general, a similar cholesterol orientation was found at both surfaces. We have then studied the predicted binding sites for a cholesterol conformation that would fulfill cholesterol binding requirements. As an example, from the ensemble we got from docking we would like to describe here in more detail one possible conformation of cholesterol binding, that resembles similar properties as reported three-dimensional cholesterol binding sites **(Fig.29.A) (**Hanson et al. 2008; Fantini & Barrantes, 2013) In this pocket at the crossed surface cholesterol is mostly bound to TM helix 3, with the -OH group next to positively charged His171 (TM3) and Arg268 (TM4), sterane group with its alpha or beta surface is next to aromatic residues Phe 178 (TM3) and the tail is bound by Ile 182 (TM3). In addition Val133 and Phe136 from TM2 participate in binding. Therefore, these residue create so called 3D consensus motif, where residues from different helices could bind together one cholesterol molecule.

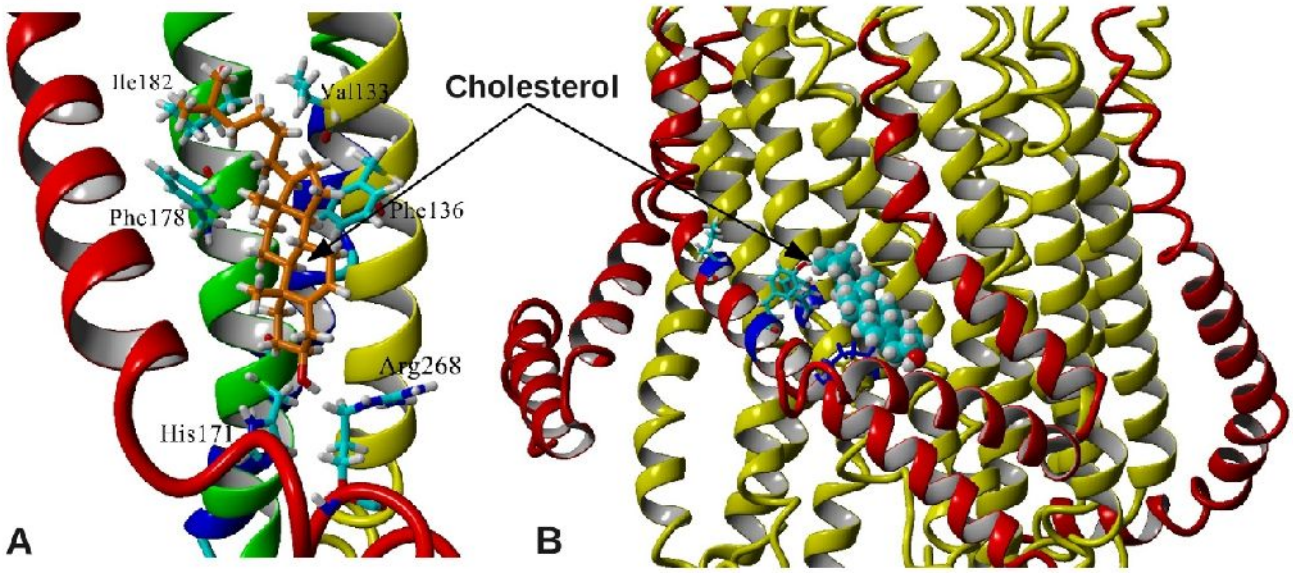


Fig.29.A Docking of cholesterol to 3D binding pocket. B Cholesterol bound to CRAC motif

We also studied whether any docked cholesterol comes close to the sequence containing the alternative CRAC/CARC motif. Interestingly, helix TM4 contains the CRAC motif **V195-X_2-Y198-X_6-K**, which is located on the protein-lipid interface. These residues are located on the same side of the helix and are easily accessible from membrane. The binding sites found by docking are often located either on left or right side of TM4. **Fig.29.B**

CONCLUSIONS

According to our global docking results, cholesterol is mostly bound to helices TM2 and TM3. Helices TM2 and TM3 form a ring around helix TM1 which forms the pore. Although, TM1 is not easily accessible from outside, at least in the present (presumably the closed) conformation, and is tightly bound to TM2 and TM3 by hydrophobic interactions. Beside of hydrophobic interaction, TM1 is connected to TM2 via a salt bridge which is formed some times by residues GLU149 and LYS78. TM1 and TM3 helices are coupled via a stable salt bridge, residues LYS85 and GLU173 **(Fig.28.B)**, additionally a salt bridge between LYS78 and GLU166 can be formed (however, it is present only occasionally in our simulations) The molecule of cholesterol is docked very close to this salt bridge. **Fig.28.B** Therefore, cholesterol binding could be tightly connected to protein function.

3.4 Extracellular calcium sink in the Orai1 channel

In our simulations of human Orai1, the hexameric channel maintained its secondary and tertiary structure and also kept structural symmetry of all six subunits, and thus it became a good system to study the distribution of Ca^{2+}ions close to the channel with ion concentrations used by the Romanin group in their experiments. We were using an concentration 10mM Ca^{2+}, and were able to observe persistent Ca^{2+} binding in a region right above the selectivity filter. Binding involved the aspartates D110 and D112 in the loop1 adjacent to the selectivity filter. Additionally to Ca^{2+} we observed also Na^{+} ion binding at the pore entrance, however with a reduced occupancy of less than 2 %. Ion density, calculated with horizontal stacks across the simulation box through the channel, revealed two major ion density peaks, one around 9 nm corresponding to the loop1 segment and the second one in the membrane around 7.5 nm matching the position of the selectivity filter. We estimate the local Ca^{2+} concentrations to be around 2.5 M within a $2nm^3$ volume close to the pore entrance, compared to the 10mM Ca^{2+} concentration in the bulk. Our understanding of this phenomenon is that this represents a local Ca^{2+} sink placed just right above the selectivity filter (ring of glutamate 106) with a local energy barrier and a distance of ~ 1.2 nm, and that its role is the maintenance of a high local Ca^{2+} concentration at the pore entrance. The experimental results by our co-workers based on site directed mutagenesis and our computational results showed that the efficiency of this Ca^{2+} sink is fine-tuned by transient electrostatic interactions with the outer loop3 of Orai1. The experimental results further demonstrated and confirmed experimentally that this Ca^{2+} sink can act as a regulatory Ca^{2+} binding domain that controls Ca^{2+} permeation and subsequently downstream Ca^{2+} signals, including NFAT dependent transcription. The details of this work are described in detail in **Paper IV.**

3.5 Structure prediction of yeast TRK1

Although the Trk proteins are related to prokaryotic Trk and Ktr and plant HKT K^+ transport systems, the overall sequence similarity is too low for simple homology modeling and a combination of homology modeling, secondary structure and fold prediction and integration of experimental restrains needs to be applied if one wants to generate a useful structural model. A key feature of such a model must be its stability in molecular dynamics simulation, when it is embedded in a lipid bilayer and simulated with a realistic solvent above and below the membrane. In **Paper V** we present a refined structural model of the ***Saccharomyces cerevisiae*** Trk1 protein based on homology modeling, molecular dynamics simulation and experimental evidence through functional analysis of mutants. The experimental results were all gained in Jost Ludwigs group and we applied a forth and back work-flow between the experimental group and the modelling.

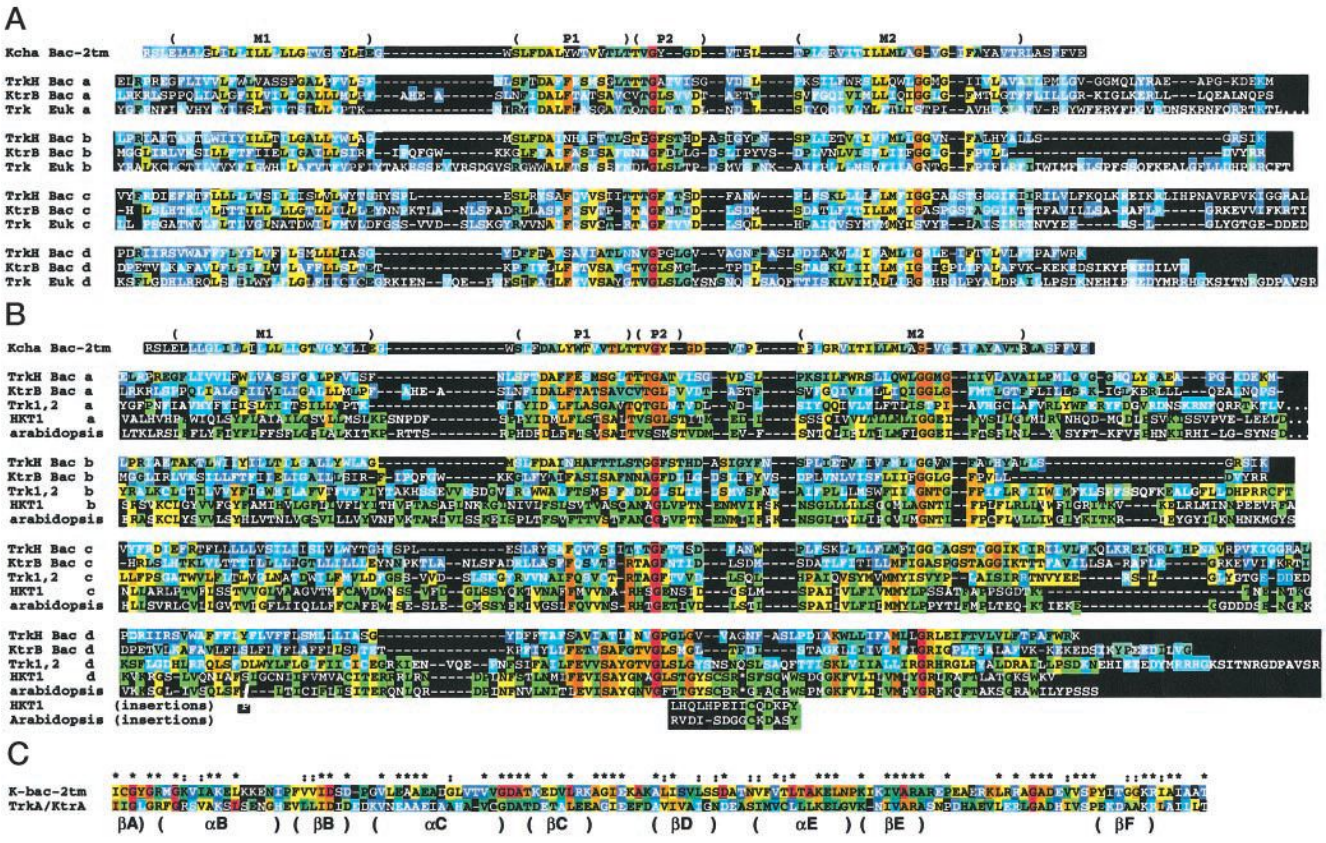


Fig.29 Consensus sequence alignments of putative 2TM bacterial K1 channels with the three K1 transporter families. (A) The residues are colored coded according to the number of identical residues in the 13 consensus sequences, i.e., red 5 13; reddish orange 5 10–12; yellowish orange 5 8–9; yellow 5 6–7; yellowish green 5 5; green 5 4; cyan 5 3; blue 5 2; black 5 1. (B) The residues are colored according to the degree of conservation within each family or MPM motif, and between families of symporters. Colors were determined by counting the number of residue types that occur at each position in the alignment of the sequences for a given family. Figure reprinted with permission from Elsevier from Durell et al., (1999, Fig.3).

Taken together the model and the experimental results showed that among the K-channel/transporter family conserved glycine residues (**Fig.29**) within the selectivity filter are not only important for protein function but also govern correct folding/membrane targeting, which was an unexpected result, as site-directed mutagenesis is used very often to test the functionality of channels, and rarely folding problems of the mutants are discussed.

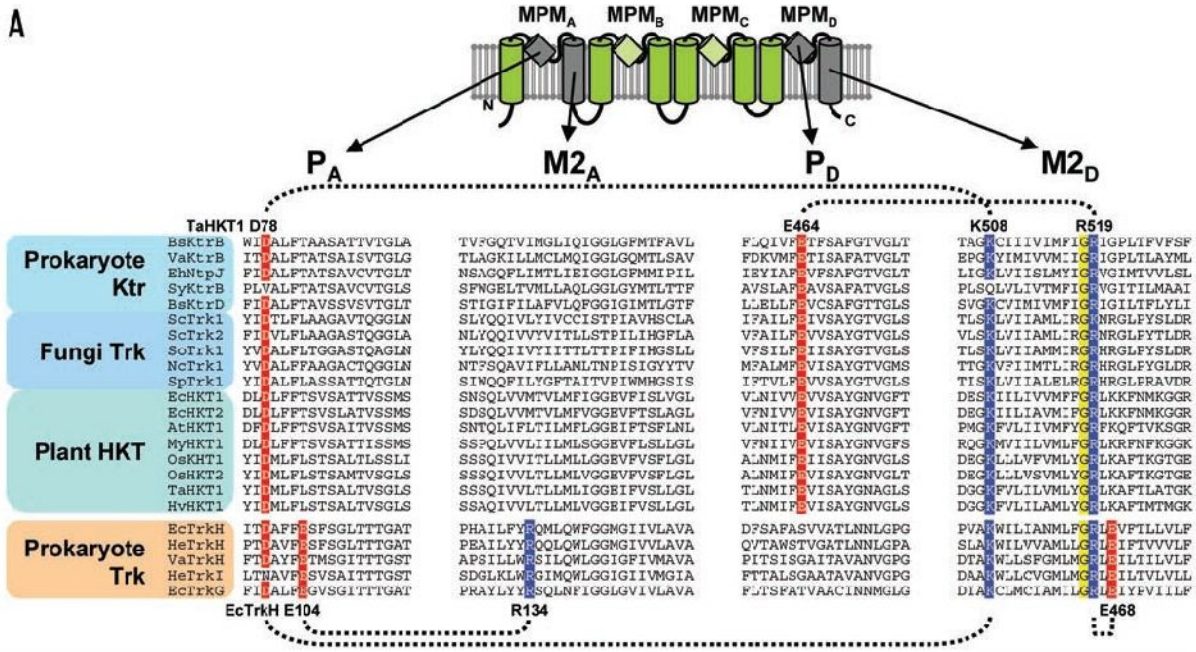


Fig.30 Alignment of conserved regions of presumed K^+ transporters of different families. Figure reprinted from Kato et al. (2007, Fig. 8). No permission needed.

However, in our case protein does not have the typical tetrameric arrangement, but we have one long covalent chain that has to fold into four domains, which might be more proune to alterations by site-directed mutagenesis. Additionally to the selectivity filter a couple of other residues in key positions were tested (**Fig.30**): a proposed interaction between a conserved aspartate (D79 in Trk1) in the A domain pore region (PA) and a lysine (K1147 in Trk1) in the D-domain helix 2 (M2D) was observed in the simulations and tested experimentally. However, an additional interaction of Asp79 with another residue close by seems to be present in the functional Trk1 protein. Taken all results together, we can conclude that our model correctly represents correctly key features of TrK1 and is consistent with the experimental data so far. As protein folding seems to be a critical issue for TrK1 and mutations easily lead to folding defects, the model might serve in the future to predict modest mutations that nevertheless would lead to measurable alterations of the functionality. All details of this work are given in **paper V**.

3.6 Other applications of computational modeling (WrbA, ArgR)

The gained experience in computational modeling, docking as well as advanced analysis was applied to two additional non-membrane systems, the ***E. coli*** proteins WrbA and ArgR. WrbA is an FMN-dependent NAD(P)H:quinone oxidoreductase and according to the present knowledge might play a role in oxidative defense. It is a tetrameric protein, in which always three subunits participate in forming a binding pocket that contains the co-factor FMN and additional space to accommodate the substrate NAD(P)H or various quinones. When we started our work, it was still unclear if the enzyme uses a ping-pong mechanism, which means that in the binding site would be in the same spot bound either the substrate NAD(P)H or the quinone substrate, as the full reaction cycle consist of a reductive half (quinone to hydroquinone) and an oxidate half (NAD(P)H to NAD). The second possibility would be the existence of two binding sites close to the co-factor FMN in which both substrates could be present simultaneously. Our docking of NADH and NAD did not only show the difference in substrate preference in dependence of the redox state of FMN, and demonstrated a preference of FMN for NADH and FMNH2 for NAD, but also demonstrated that benzoquinone, which is present in the crystal structure, would overlap with the docked NADH if both were present. Thus, we can clearly can exclude this possibility on basis of our structural and computational findings. The active-site chamber, though very large, is clearly not large enough to accommodate both NADH and BQ at the same time. Additionally, we could demonstrate that the NADH position that is found in the crystal structure has a lower affinity than the actual relevant binding site in the active site that allows the reaction to take place. Therefore it is earlier in non-physiological position, or the crystal structure represents an intermediate when the substrate approaches or the product leaves the active site. Additionally, as the protein can occur also in its dimeric form, though with a lower enzymatic activity, we also described changes of the binding pocket with the docked substrate upon dimerization. My computational results in this paper assisted the experimentalists in the structural explanation of their kinetic data and understanding of the enzymatic mechanism. All details and the used methodology are explained in **Paper I.**

The second system, Arginine repressor of *E. coli* or ArgR is a multifunctional hexameric protein that provides feedback regulation of arginine metabolism upon activation by the negatively cooperative binding of L-arginine. This protein shows allostery for the binding of its activator L-arginine. The binding shows negative cooperativity. After one arginine has bound, binding of the subsequent five ligands is with a 100fold lower affinity as for the first. A trimer-trimer rotation has been shown earlier to play a major role in this effect, as the first binding event stops the rotation and

shifts the system back to a less rotated state. The aim of this work was to simulate the transition from the initial holo state with all six arginine bound to the fully rotated state, the identification of binding-competent conformations in the ensemble of conformations sampled in the simulation, the docking of the first L-Arg into the bidning competent pocket and to simulate the consequence of the first binding event. So we used 100 ns molecular dynamics simulations in presence and absence of the six L-arg ligands that bind at the trimer-trimer interface. The analysis of correlation maps, center of mass shifts, arg-asp saltbridges and visual inspection together supports the possibility that formation and release of the saltbridge between Arg110 and Asp128 is mainly responsible for trimer-trimer rotation. All other interactions found in the apoArgRC simulations were also present in holoArgRC conformers and thus did not influence the rotation. Thus the formation and release of this arg-asp salt bridge opens up the cavity and provides rotational dynamics to the system, which might help to bring the system to conformations in which favorable for L-arginine binding. Our simulations of ArgR with the docked L-Arg reveal for the first time that the binding of the first ligand results in a holoprotein-like (or unrotated) conformational distribution. This work is the main research focus of my colleague Saurabh Pandey and I was involved in setting up some of the simulations and brought in my expertise in docking and advanced data analysis. All details are described in **Paper III.**

4. CONCLUSIONS

Combined sequence analysis and molecular modeling approaches were applied to discover a structural characteristics of ion channels specialized in the translocation of mono- and divalent cations. Establishment of evolutionary relationships between different groups of cation channels enabled us to predict and model unknown three-dimensional structures of transmembrane domains of three various cation channels/transporters: human TRPA1, a non-selective cation channel involved in pain and cold sensation; yeast Trk1, a putative potassium transporter; human Orai1 channel, a pore domain of store-operated calcium entry. Every of these proteins is unique with many undiscovered and challenging aspects. We applied the combination of different approaches in order to uncover structural properties underlying important functions of these systems.

Conclusions 1. Structural modeling of TRPA1 channel.

1.1. N-terminal domain of TRPA1.

The cytoplasmic N-terminal domain was reported to contain from 14 to 18 ankyrin repeats. mutagenesis experiments also predicted a calcium binding motif EF-hand. We analyzed the sequence using various bioinformatics approaches and found 17 true ankyrin repeats within the N-terminus. Thsese predicted 17 ankyrin repeats and the calcium binding motif between 12^{th} and 13^{th} ankyrin repeats were modeled using homology modeling procedure. Furthermore, the ankyrin repeat domain kept structural integrity during molecular dynamics simulations and demonstrated elastic properties attributed to known ankyrin repeats. The insertion of the EF-hand motif did not affect the N-terminal integrity and dynamic properties, whereas, the binding of calcium ion to EF-hand produced a strong effect on the overall curvature of the N-terminal domain. While the end-to-end distance and overall curvature in the apo-structure fluctuates freely, the calcium binding restricted all the “macro” movements. We predict that such a sudden restriction of movement obtained by calcium binding propagates to the channel gate and increases the open probability of the channel.

1.2 Transmembrane domain of TRPA1.

Several models of the transmembrane domain were built in order to get a better representative model of the channel. Starting with the modeling based on the distantly related templates, we utilized all the additional resources of information in order to make a more accurate prediction.

Later, a TRPA1 model based on new structure of a related TRPV1 channel allowed us to verify earlier predictions and conclude that even from the distantly related template, the TRPA1 model was built with a high accuracy.

1.3 Pleckstrin homology (PH) domain prediction.

The prediction of the complementary half of the N-terminal PH-domain was an ambiguous task, since there were no experimental indications of the domain location and no significant similarity with known structures of PH domain. We combined a conserved domain prediction, secondary structure assignment, and sequence analysis in order to find a putative hidden domain. Further, predicted domains were modeled and verified in molecular dynamics simulations. Modeling of the PH domain gave a chance to evaluate the sensitivity of modeling tools. Already on the level of homology modeling many models were rejected due to failure in forming secondary structure elements (beta-strands). Only correctly folded models with correctly formed secondary structure were further evaluated in molecular dynamics. At this level the validity of the best predictions was tested. Our results argued against the existence of the putative complementary hidden N-terminal half of PH domain. Molecular dynamics showed that in contrast to a consistent "native" ankyrin repeat domain discussed above, the model with an embedded PH domain has a strong negative impact on the overall model structure stability.

Therefore, our results showed that while sequence analysis and prediction tools give important hints for the analysis of protein of unknown structure, molecular modeling is a powerful technique in order to actually test the possibility that prediction was done correctly by assessment of the structural integrity and stability.

Conclusions 2. Structure prediction of Trk1

The evolutionary relationship between yeast Trk1 and bacterial TrkH/KtrB transporters enabled us to build a three-dimensional model of yeast Trk1 channel. Although, the low sequence similarity did not allow finding a single optimal solution for modeling, it provideds a good basis for experimental studies that in turn helped to improve the model. A key residue that enabled to relate Trk1, TrkH and HKT potassium transporters to a common ancestor is the invariant glycine residue in the selectivity filter. The experimental evidences proving that this conserved glycine belongs to the selectivity filter was given for TrkH and HKT proteins, but not yet for Trk1. Through the

functional analysis of mutants the location of predicted conserved glycine residues within the selectivity filter was proved as well for Trk1 within this work. This in turn strengthens the hypothesis of evolutionary relationship between Trk1 and bacterial potassium transporters TrkrH and KtrB, and also indicates the validity of the model. Furthermore, our experimental results based on the model of Trk1 showed that Trk1 structure as well as TkrH and KtrB possess a conserved salt bridge between Asp 79 and Lys 1147. This result provided additional evidence supporting the predicted structure.

Therefore, we conclude that combined theoretical and experimental studies, mutually interconnected, help to design experiments and evaluate the predicted models.

Conclusions 3. Investigation of Orai1

3.1 Transmembrane domain of Orai1.

In contrast to the previous proteins, a high sequence identity between human Orai1 and template structure enabled us to obtain a high quality model and study the actual channel functions. Particularly, we provided structural explanations for the experimental results showing an exclusive role of extracellular loops in regulation of calcium ion permeation. Electrostatic interactions between three negatively charged aspartic acid residues of the loop1 and positively charged residues of the loop 3 were identified in molecular dynamics simulations and confirmed by cross-linking experiments. Statistical analysis of all interactions made by each negatively charged residue of the loop 1 either with the loop 3 or with the ions revealed an exclusive role of each aspartatic acid of loop1. The first one is almost exclusively involved in ion chelating and binding, the second aspartate interacts with both: ions and loop3, while the third aspartate is mostly involved in intramolecular loop-loop interactions, These finding provided the structural explanation for results of mutagenesis studies, which showed that mutation of the first aspartate have a strongest effect on calcium permeation.

3.2 Cholesterol binding site.

The docking results of the cholesterol binding to the Orai1 channel revealed two putative binding pockets on the protein-lipid interface, that fulfill the requirements of the cholesterol binding sites. Cholesterol is required for the successful association of STIM1 and Orai1 and other channels involved store-operated calcium entry.

Finally, we conclude, that our approach of bioinformatics and molecular dynamics is a powerful and useful tool in the study of ion translocations systems if combined with carefully designed experiments that constraints and refine the modeling. Our work demonstrates the importance of the dynamics in understanding the mechanism behind the channel function, where a static model does not give the comprehensive picture.

5. PUBLICATIONS

(APPENDIX)

5.1

Biphasic kinetic behavior of FMN-dependent NAD(P)H:quinone oxidoreductase WrbA from *E. coli*

[1]Iryna Kishko, [2]Balasubramanian Harish, [1]Vasilina Zayats, [1]David Řeha,
[2]Brian Tenner, [2]Dhananjay Beri, [3]Tobias Gustavsson,
[1]Rüdger Ettrich*, and [2]Jannette Carey*

5.2

Regulation of the transient receptor potential channel TRPA1 by its N-terminal ankyrin repeat domain.

Vasilina Zayats[1], Abdul Samad[1], Babak Minofar[1], Katherine E. Roelofs[2], Thomas Stockner[3]* and Rudiger Ettrich[1]

5.3

Binding-competent states for L-arginine in E. coli arginine repressor apoprotein

Saurabh Kumar Pandey[1,2], David Řeha[1,2], Vasilina Zayats[1,2], Milan Melichercik[1,3], Jannette Carey[1,4*], Rüdiger Ettrich[1,2*]

5.4

Extracellular calcium sink in the Orai1 channel tunes calcium permeation and initiation of NFAT transcription.

Irene Frischauf*, **Vasilina Zayats***, Michael Deix, Anna Hochreiter, Barbara Lackner, Barbora Svobodova, Christoph Romanin, Rüdiger Ettrich, Rainer Schindl (2014) . *prepared for submission. (* contributed equally)*

5.5

A refined atomic scale model of the *Saccharomyces cerevisiae* K^+-translocation protein Trk1p combined with experimental evidence confirms the role of selectivity filter glycines and other key residues.

Vasilina Zayats[1,2], Thomas Stockner[4], Saurabh K. Pandey[1], Rüdiger Ettrich[1,2*], Jost Ludwig[1,3*]

6. REFERENCES

References

Adamian, L., Naveed, H. & Liang, J. (2011) Evolutionary conservation of lipid-binding sites in membrane proteins. *Biochim Biophys Acta Biomembranes,* **1808**, 1092–1102.

Aggarwal, S. K. & MacKinnon, R. (1996) Contribution of the S4 segment to gating charge in the Shaker K+ channel. *Neuron,* **16**, 1169–1177.

Albarrán, L., Lopez, J. J., Dionisio, N., Smani, T., Salido, G. M. & Rosado, J. A. (2013) Transient receptor potential ankyrin-1 (TRPA1) modulates store-operated Ca^{2+} entry by regulation of STIM1-Orai1 association. *Biochim. Biophys. Acta*, **1833**, 3025-3034.

Altschul, S. F., Gish, W., Miller, W., Myers, E.W., Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol,* **215**, 403–410.

Altschul, S. F. (1991) Amino acid substitution matrices from an information theoretic perspective. *Journal of Molecular Biology*, **219**, 555-565.

Anderson, P. A. V. & Greenberg, R.M. (2001) Phylogeny of ion channels: clues to structure and function. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, **129**, 17-28.

Armstrong, C. M. (2002) Channels and pumps early in evolution. In: Pumps, transporters and ion channels. Studies on their structure, function and cell biology, Sepulveda, F.V., Bezanilla, F. eds., *Kluwer Academic / Plenum Publishers*, New York.

Armstrong, C.M. (2003) Voltage-Gated K Channels. *Sci. STKE*, **188**, re10.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. & Stockinger, H. (2012) ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*, **40**(W1):W597-W603.

Augustine, G. J., Chikaraishi, D. M., Ehlers, M. D., Einstein, G. M., Fitzpatrick, D. et al., (2004)

Neuroscience 3rd ed., Augustine, G.J., Fitzpatrick, D., Hall, W.C., La Mantia, A-S., Purves, D. eds., *Sinauer Associates*, Mas, USA.

Baier, C. J., Fantini, J. & Barrantes, F. J. (2011) Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. *Sci. Rep.,* 1: 69.

Banke, T. G., Chaplan, S. R. & Wickenden, A. D. (2010). Dynamic changes in the TRPA1 selectivity filter lead to progressive but reversible pore dilation. *Am. J. Physiol. Cell Physiol.* **298**, C1457–C1468.

Bargal, R., Avidan, N., Ben-Asher, E., Olender, Z., Zeigler, M., Frumkin, A., Raas-Rothschild, A., Glusman, G., Lancet, D. & Bach, G. (2000) Identification of the gene causing mucolipidosis type IV. *Nat. Genet*., **26**, 118-123.

Bassi, M. T., Manzoni, M., Monti, E., Pizzo, M. T., Ballabio, A. & Borsani, G. (2000) Cloning of the gene encoding a novel integral membrane protein, mucolipidin-and identification of the two major founder mutations causing mucolipidosis type IV. *Am. J. Hum. Genet.*, **67**, 1110-1120.

Bekker, H., Berendsen, H. J. C., Dijkstra, E. J., Achterop, S., van Drunen, R., van der Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., Renardus, M. K. R. Gromacs: A parallel computer for molecular dynamics simulations (1993). In *Physics Computing,* **92**, de Groot, R. A., Nadrchal, J., eds. World Scientific.

Berendsen, H. J. C., Postma, J. P. M., DiNola, A., Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.,* **81**, 3684-3690.

Berendsen, H. J. C., van der Spoel, D., van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.,* **91**, 43-56.

Berg, J. M., Tymoczko, J. L. & Stryer, L. (2002) Membrane Channels and Pumps in: Biochemistry, 5th edition, *W.H. Freeman and Company*, New York, USA.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. &

Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res,* **28**, 235-242.

Calcraft, P. J., Ruas, M., Pan, Z., Cheng, X., Arredouani, A., Hao, X., Tang, J., Rietdorf, K., Teboul, L., Chuang, K. T., et al. (2009) NAADP mobilizes calcium from acidic organelles through two-pore channels. *Nature*, **459**, 596–600.

Cao, E., Liao, M., Cheng, Y. & Julius, D. (2013) TRPV1 structures in distinct conformations reveal activation mechanisms. *Nature,* **504**, 113–118.

Cao, Y., Jin, X., Huang, H., Derebe, M. G., Levin, E. J., Kabaleeswaran, V., Pan, Y., Punta, M., Love, J., Weng, J., Quick, M., Ye, S., Kloss, B., Bruni, R., Martinez-Hackert, E., Hendrickson, W. A., Rost, B., Javitch, J. A., Rajashankar, K. R., Jiang, Y. & Zhou, M. (2011) Crystal structure of a potassium ion transporter, TrkH. *Nature*, **471**, 336-340.

Cao,Y., Pan, Y., Huang, H., Jin, X., Levin, E. J., Kloss, B. & Zhou, M. (2013) Gating of the TrkH ion channel by its associated RCK protein TrkA. *Nature*, **496**, 317-322.

Carson, M. (1997) Ribbons. *Methods Enzymol.* **277**, 493-502.

Catterall, W. A. (2010) Ion channel voltage sensor: structure, function, and pathophysiology. *Neuron,* **67**, 915–928.

Chuang, H. H., Prescott, E. D., Kong, H., Shields, S., Jordt, S. E., Basbaum, A. I., Chao, M. V. & Julius, D. (2001) Bradykinin and nerve growth factor release the capsaicin receptor from PtdIns(4,5)P2-mediated inhibition. *Nature*, **411**, 967-962.

Clapham, D. E. (2003) TRP channels as cellular sensors. *Nature*, **426**, 517–524.

Clapham, D. E. (2007) Calcium signaling. *Cell*, **14**, 1047-1058.

Clapham, D. E., Montell, C., Schultz, G. & Julius, D. (2003) International Union of Pharmacology XLIII Compendium of voltage-gated ion channels: transient receptor potential channels. *Pharmacol. Rev.*, **55**, 591–596.

Clayton, G. M., Altieri, S., Heginbotham, L., Unger, V. M., Morais-Cabral, J. H. (2008) Structure of the transmembrane regions of a bacterial cyclic nucleotide-regulated channel. **105***,* 1511–1515.

Cosens, D. J. & Manning, A. (1969) Abnormal electroretinogram from a *Drosophila* mutant. *Nature*, **224**, 285-287.

Creighton, T. E. (2011) The biophysical chemistry of proteins. *Helvetian Press*.

Darden, T., York, D. & Pedersen, L. (1993) Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J Chem Phys,* **98,** 10089–10092.

Dayhoff, M. O. & Eck, R. V. (1968). Atlas of protein sequence and structure. *National Biomedical Research Foundation*, Silver Spring, Maryland, **3***,* 33.

Derst, C. & Karschin, A. (1998) Evolutionary link between prokaryotic and eukaryotic K^+ channels. *Exp. Biol*., **201**, 2791-2799.

Derler, I., Schindl, R., Fritsch, R. & Romanin C. (2012) Gating and permeation of Orai channels. *Front. Biosci.,* **17**, 1304-22.

Dionisio, N., Galán, C., Jardín, I., Salido, G.M. & Rosado, J.A. (2011) Lipid rafts are essential for the regulation of SOCE by plasma membrane resident STIM1 in human platelets, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1813**, 431-437.

Doerner, J. F., Gisselmann, G., Hatt, H. & Wetzel, C. H. (2007) Transient receptor potential channel A1 is directly gated by calcium ions. *J. Biol. Chem*., **282**, 13180 -13189.

Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science,* **280**, 69-77.

Duggan, A., García-Añoveros, J. & Corey, D. P. (2000) Insect mechanoreception: what a long, strange TRP it's been. *Curr. Biol*., **10**, R384-387.

Duncan, L. M., Deeds, J., Hunter, J., Shao, J., Holmgren, L. M., Woolf, E. A., Tepper, R. I. & Shyjan, A. W. (1998) Down-regulation of the novel gene melastatin correlates with potential for melanoma metastasis. *Cancer Res.*, **58**, 1515-1520.

Durell, S. R., Hao, Y., Nakamura, T., Bakker, E. P. & Guy, H. R. (1999) Evolutionary Relationship between K^+ Channels and Symporters. *Biophysical journal*, **77**, 775–788.

Durell, S. R. & Guy, H. R. (1999) Structural models of the KtrB, TrkH, and Trk1,2 symporters based on the structure of the KcsA K^+ channel. *Biophys. J.*, **77**, 789–807.

Enyedi, P. & Czirják, G. (2010) Molecular background of leak K^+ currents: two-pore domain potassium channels. *Physiol. Rev.*, **90**, 559-605.

Fanger, C. M., del Camino, D. & Moran, M. M. (2010) TRPA1 as an Analgesic target. *Open Drug discovery J.,* **2**, 64-70.

Fantini, J. & Barrantes, F. J. (2013) How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Front. Physiol.* **4**, 31.

Feske, S., Gwack, Y., Prakriya, M., Srikanth, S., Puppel, S. H., Tanasa, B., Hogan, P. G., Lewis, R. S., Daly, M. & Rao, A. (2006) A mutation in Orai1 causes immune deficiency by abrogating CRAC channel function. *Nature*, **441**, 179-185.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al.(2014) Pfam: the protein families database. *Nucleic Acids Res*, **42**, D222–D230.

Franciolini, F. & Petris, A. (1989) Evolution of Ionic Channels of Biological Membranes. *Mol. Biol. Evol., **6,*** 503-513.

Fujiwara, Y., Takeshita, K., Nakagawa, A. & Okamura, Y. (2013) Structural characteristics of the redox-sensing coiled coil in the voltage-gated H^+ channel. *J. Biol. Chem.*, **288**, 17968-17975.

Garcia-Anoveros, J. & Nagata, K. (2007) TRPA1. *Handb. Exp. Pharmacol.,* **179**, 347-362.

Gauss, R., Seifert, R. & Kaupp, U. B. (1998) Molecular identification of a hyperpolarization-activated channel in sea urchin sperm. *Nature,* **393**, 583-7.

Hänelt, I., Löchte, S., Sundermann, L., Elbers, K., Vor der Brüggen, M. & Bakker, E. P. (2010) Gain of function mutations in membrane region M2C2 of KtrB open a gate controlling K^+ transport by the KtrAB system from *Vibrio alginolyticus*. *J. Biol. Chem.*, **285**, 10318-10327.

Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y., Velasquez, J., Kuhn, P. & Stevens, R.C. (2008) A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure,* **13**, 897–905.

Hardie, M. C. (2001) Phototransduction in *Drosophila melanogaster*. *J. Exp. Biol.*, **204**, 3403-3409.

Hardie, R. C. & Minke, B. (1992) The trp gene is essential for a light-activated Ca^{2+} channel in *Drosophila* photoreceptors. *Neuron*, **8**, 643-651.

Havel, T. F. & Snow, M. E. (1991) A new method for building protein conformations from sequence alignments with homologues of known structure. *Journal of Molecular Biology*, **217**, 1-7.

Henikoff, S. & Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci.,* **89**, 10915–10919.

Hinchliffe, A. (2003) Molecular Modelling for Beginners. *Wiley*, Chichester, West Sussex, England.

Hoover, W. G. (1985) Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, **31**, 1695-1697.

Hou, X., Pedi, L., Diver, M. M. & Long, S. B. (2012) Crystal structure of the calcium release-activated calcium channel Orai. *Science*, **338**, 1308-1313.

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

Hunter, J. J., Shao, J., Smutko, J. S., Dussault, B. J., Nagle, D. L., Woolf, E. A, Holmgren L. M., Moore, K. J. & Shyjan AW (1998) Chromasomal localization and genomic characterization of mause melastatin gene (Mlsn1). *Genomics*, **54**, 116-123.

Ishibashi, K., Suzuki, M. & Imai, M. (2000) Molecular cloning of a novel form (two-repeat) protein related to voltage-gated sodium and calcium channels. *Biochem. Biophys. Res. Commun.*, **270**, 370-376.

Israelachvili, J. N. (1991) *Intermolecular and Surface Forces.* Academic, New York, 2nd edn.

Jan, L. Y. & Jan, Y. N. (1992) Tracing the roots of ion channels. *Cell*, **69**, 715-718.

Jegla, T. J., Zmasek, C. M., Batalov, S. & Nayak, S. K. (2009) Evolution of the human ion channel set. Comb. Chem. *High Throughput Screen,* **12**, 2–23.

Jensen, M. Ø., Borhani, D.W., Lindorff-Larsen, K., Maragakis, P., Jogini, V., Eastwood, M.P., Dror, M. P. & Shaw, D.E. (2010) Principles of conduction and hydrophobic gating in K^+ channels. *Proc. Natl. Acad. Sci. USA*. **107***,* 5833–5838.

Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. **1996**, **118**, 11225-11236.

Jorgensen, W. L. & Tirado-Rives, J. (1988) The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.*, **110**, 1657-1666.

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, **105**, 6474-6487.

Karashima, Y., Prenen, J., Meseguer, V., Owsianik, G., Voets, T. & Nilius, B. (2008) Modulation of the transient receptor potential channel TRPA1 by phosphatidylinositol 4,5-biphosphate manipulators. *Pflugers Arch*. **457**, 77-89.

Kato, N., Akai, M., Zulkifli, L., Matsuda, N., Kato, Y., Goshima, S., Hazama, A., Yamagami, M., Guy, H. R. & Uozumi, N. (2007) Role of positively charged amino acids in the M2D transmembrane helix of Ktr/Trk/HKT type cation transporters. *Channels (Austin)*, **1**, 161-171.

Kim, D., Cavanaugh, E. & Simkin, D. (2008) Inhibition of transient receptor potential A1 by phosphatidylinositol-4,5-bisphosphate. *Am. J. Physiol. Cell Physiol.*, **295**, C92-C99.

Koshiba, S., Kigawa, T., Kim, J. H., Shirouzu, M., Bowtell, D. & Yokoyama, S. (1997) The solution structure of the pleckstrin homology domain of mouse Son-of-sevenless 1 (mSos1). *J. Mol. Biol.,* **269**, 579-91.

Kremeyer, B., Lopera, F., Cox, J. J., Momin, A., Rugiero, F., Marsh, S., Woods, C. G., Jones, N. G., Paterson, K. J., Fricker, F. R., Villegas, A., Acosta, N., Pineda-Trujillo, N. G., Ramírez, J. D., Zea, J., Burley, M. W., Bedoya, G., Bennett, D. L., Wood, J. N. & Ruiz-Linares, A. (2010) A gain-of-function mutation in TRPA1 causes familial episodic pain syndrome. *Neuron*, **10**, 671-680.

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

Krieger, J., Strobel, J., Vogl, A., Hanke, W. & Breer H. (1999) Identification of a cyclic nucleotide- and voltage-activated ion channel from insect antennae. *Insect Biochem. Mol. Biol.,* **29**, 255-67.

Kuhr, F.K., Smith, K.A., Song, M.Y., Levitan, I. & Yuan, J.X. (2012) New mechanisms of pulmonary arterial hypertension: role of Ca^{2+} signaling. *Am. J. Physiol. Heart. Circ. Physiol.*, **302**, H1546-H1562.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*, **26**, 283–291.

Leach, A. R. (2001) Molecular Modelling: Principles and Applications. *Prentice Hall*, 2nd edition.

Lee, G., Abdi, K., Jiang, Y., Michaely, P., Bennett, V. & Marszalek, P. E. (2006) Nanospring behaviour of ankyrin repeats. *Nature*, **440**, 246–249.

Levitan, I., Fang, Y., Rosenhouse-Dantsker, A. & Romanenko, V. (2010) Cholesterol and ion channels. *Subcell. Biochem.,* **51**, 509–549.

Lewit-Bentley, A., Réty, S. (2000) EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol.*, **10**, 637-643.

Liu, B. & Qin, F. (2005) Functional control of cold- and menthol-sensitive TRPM8 ion channels by phosphatidylinositol 4,5-bisphosphate. *J. Neurosci*., **25**, 1674-1681.

Lodish, H., Berk, A., Zipursky, L. S., *et al.* (2004) *Molecular Cell Biology* (4th ed.). New York: Scientific American Books.

Long, S. B., Campbell, E. B. & Mackinnon, R. (2005) Crystal structure of a mammalian voltage-dependent Shaker family K^+ channel. *Science*, **309**, 897-902.

Lui, D. & Liman, E. R. (2003) Intracellular Ca^{2+} and PIP2 regulate the taste transduction in TRPM5. *Proc Natl Acad Sci U S A*., **100**, 15160-15165.

L.V. Woodcock (1971) Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters*, **10**, 257-261.

Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., DeWeese-Scott, C., Gonzales, N. R., Gwadz, M., Hao, L., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Krylov, D., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Lu, S., Marchler, G. H., Mullokandov, M., Song, J. S., Thanki, N., Yamashita, R. A, Yin, J. J., Zhang, D. & Bryant, S.H. (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.,* **35** (Database issue):D237-40.

Matsuda, M., Takeshita, K., Kurokawa, T., Sakata, S., Suzuki, M., Yamashita, E., Okamura, Y. & Nakagawa, A. (2011) Crystal structure of the cytoplasmic phosphatase and tensin homolog (PTEN)-like region of *Ciona intestinalis* voltage-sensing phosphatase provides insight into substrate

specificity and redox regulation of the phosphoinositide phosphatase activity. *J. Biol. Chem.*, **286**, 23368-23377.

McQuarrie, D. A. (1976) Statistical Mechanics. *Harper & Row*, New York.

Michaely, P., Tomchick, D. R., Machius, M. & Anderson, R. G. (2002) Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J.*, **21**, 6387-6396.

Minke, B. (2010) The history of the Drosophila TRP channel: the birth of a new channel superfamily. *J. Neurogenet*., **24**, 216-233.

Montell, C. (2005) The TRP superfamily of Cation Channels. *SciSTKE*, **272**, re3.

Montell, C. & Rubin, M. (1989) Molecular characterization of the Drosophila trp locus: A putative integral membrane protein required for phototransduction. *Neuron,* **2**, 1313-1323.

Morgan, A. J., Platt, F. M., Lloyd-Evans, E. & Galione, A. (2011) Molecular mechanisms of endolysosomal Ca^{2+} signalling in health and disease. *Biochem. J.*, **439**, 349–374.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R. K., et al. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30,** 2785-2791.

Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. Y. (2004) The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.*, **13**, 1435-1448.

Mosavi, L. K., Minor, D. L. Jr. & Peng, Z. Y. (2002) Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. U S A.*, **99**, 16029-16034.

Motoo, K. (1983) The neutral theory of molecular evolution. *Cambridge University Press*. Cambridge.

Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. (2005) Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature*, **435**, 1239-1243.

Nelson, M. & Chazin, W. (2008) EF-Hand Calcium-Binding Proteins Data Library. Vanderbilt University. Retrieved 2009-08-29.

Needleman, S.B. & Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443–453.

Ngata, K., Duggan, A., Kumar, G. & Garcia-Anoveros, J. (2005) Nociceptor and hair cell transducer protperties of TRPA1, a channel for pain and hearing. *J. Neurosci*, **25**, 4052-4061.

Nilius, B., Appendino, G. & Owsianik, G. (2012) The transient receptor potential channel TRPA1: from gene to pathophysiology. *Pflugers Arch.*, **464**, 425-458.

Nilius, B., Owsianik, G. & Voets, T. (2008) Transient receptor potential channels meet phosphoinositides. *EMBO J.*, **27**, 2809-2816.

Nilius, B., Appendino, G. & Owsianik, G. (2012) The transient receptor potential channel TRPA1: from gene to pathophysiology. *Pflügers Arch.*, **464**, 425-458.

Nose, S. (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.,* **52**, 255-268.

Oda, H. & Takeichi, M. (2011) Structural and functional diversity of cadherin at the adherens junction. *J. Cell Biol*., **193**, 1137-1146.

Palmer, C. P., Zhou, X. L., Lin, J., Loukin, S. H., Kung, C. & Saimi, Y. (2001) A TRP homolog in *Saccharomyces cerevisiae* forms an intracellular Ca^{2+}-permeable channel in the yeast vacuolar membrane. *Proc. Natl. Acad. Sci. U S A.,* **98**, 7801–7805.

Pani, B., Bollimuntha, S. & Singh, B.B. (2012) The TR(i)P to Ca^{2+} signaling just got STIMy: an update on STIM1 activated TRPC channels. *Front Biosci (Landmark Ed.)*, **17**, 805-823.

Parrinello, M. & Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular

dynamics method. *J. Appl. Phys.*, 52, 7182-7190.

Payandeh, J., Scheuer, T., Zheng, N., Catterall, W. A. (2011) The crystal structure of a voltage-gated sodium channel. *Nature*, **475**, 353-359.

Perozo, E., Marien, D., Cortes & Cuello, L. G. (1999) Structural rearrangements underlying K+-Channel Activation Gating. *Science*, **285,** 73-78.

Pertovaara, A. & Koivisto, A. (2011) TRPA1 ion channel in the spinal dorsal horn as a therapeutic target in central pain hypersensitivity and cutaneous neurogenic inflammation. *Eur J Pharmacol.,* **666**, 1-4.

Plummer, N. W. & Meisler, M. H. (1999) Evolution and diversity of mammalian sodium channel genes. *Genomics*, **57**, 323-331.

Prakriya, M. & Lewis, R. S. (2006) Regulation of CRAC channel activity by recruitment of silent channels to a high open-probability gating mode. *J. Gen. Physiol.*, **128**, 373-386.

Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R. Smith, J.C., Kasson, P. M., van der Spoel, D., Hess, B., Lindahl, E. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics,* **29**, 845-854.

Putney, J. W. Jr. (2005) Capacitative calcium entry: sensing the calcium stores. *J. Cell Biol.*, **169**, 381-382.

Qin, F. (2007) Regulation of TRP ion channels by phosphatidylinositol-4,5-bisphosphate. *Handb. Exp. Pharmacol.,* **179**, 509-525.

Ramsey, I. S., Moran, M. M., Chong, J. A. & Clapham, D. E. (2006) A voltage-gated proton-selective channel lacking the pore domain. *Nature*, **440**, 1213-1216.

Roos, J., DiGregorio, P. J., Yeromin, A. V., Ohlsen, K., Lioudyno, M., Zhang, S., Safrina, O., Kozak, J. A., Wagner, S. L., Cahalan, M. D., Veliçelebi, G. & Stauderman, K. A. (2005) STIM1, an

essential and conserved component of store-operated Ca^{2+} channel function. *J. Cell Biol.*, **169**, 435-445.

Sali, A. & Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.,* **234**, 779-815.

Salido, G. M., Jardín, I. & Rosado, J. A. (2011) The TRPC ion channels: association with Orai1 and STIM1 proteins and participation in capacitative and non-capacitative calcium entry. *Adv Exp Med Biol.,* **704**, 413-433.

Samad, A., Sura, L., Benedikt, J., Ettrich, R., Minofar, B., Teisinger, J. & Vlachova, V. (2010) The C-terminal basic residues contribute to the chemical- and voltage-dependent activation of TRPA1. *Biochem J*., **433**, 197–204.

Schlick, T. (2002) Molecular Modeling and Simulation: An Interdisciplinary Guide, Interdisciplinary applied mathematics, **21**, *Springer-Verlag*, NY.

Sedgewick, S. G. & Smerdon, S. J. (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.*, **24**, 311-316.

Shannon, R. D. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica,* **32**, 751-767.

Sharman, J. L., Benson, H. E., Pawson, A. J., Lukito, V., Mpamhanga, C. P., Bombail, V., Davenport, A. P., Peters, J.A., Spedding, M., Harmar, A.J., & NC-IUPHAR. (2013) IUPHAR-DB: updated database content and new features. *Nucl. Acids Res.* **41** (Database Issue): D1083-8.

Shi, N., Ye, S., Alam, A., Chen, L., Jiang, Y. (2006) Atomic structure of a Na^{+}- and K^{+}-conducting channel. *Nature,* **440**, 570–574.

Shin, Y. C., Shin, S.Y., So, I., Kwon, D. & Jeon, J. H. (2011) TRIP Database: a manually curated database of protein–protein interactions for mammalian TRP channels. Nucleic acids research, 39, D356-D361.

Singer, S.J. & Nicolson, G.L. (1972) The fluid mosaic model of the structure of cell membranes. *Science*, **175**, 720-731.

Smith, T. F., Waterman, M. S. (1981) Identification of common molecular subsequences. *J Mol Biol,* **147**, 195–197.

Smyth, J. T., Dehaven, W. I., Jones, B. F., Mercer, J. C., Trebak, M., Vazquez, G. & Putney, J. W. Jr. (2006) Emerging perspectives in store-operated Ca^{2+} entry: roles of Orai, Stim and TRP. *Biochim. Biophys. Acta.*, **1763**, 1147-1160.

Smyth, J. T., Hwang, S. Y., Tomita, T., DeHaven, W. I., Mercer, J. C. & Putney, J. W. (2010) Activation and regulation of store-operated calcium entry. *J. Cell. Mol. Med.*, **14**, 2337-2349.

Soboloff, J., Rothberg, B. S., Madesh, M. & Gill, D. L. (2012) STIM proteins: dynamic calcium signal transducers. *Nat Rev Mol Cell Biol.*, **13**, 549-565.

Sotomayor, M., Corey, D. P. & Schulten, K. (2005) In search of the hair-cell gating spring: elastic properties of ankyrin and cadherin repeats. *Structure*, **13**, 669-682.

Spoel, D., Sijbers, A., Keegstra, H., Reitsma, B., Renardus, M. K. R. (1993) Gromacs: A parallel computer for molecular dynamics simulations. In Physics Computing, 92, de Groot, R. A., Nadrchal, J., eds. World Scientific.

Srinivasan, N., Blundell, T. L. (1993) An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng.,* **6**(5), 501–512.

Sternberg, M. J. E. (1996) Protein Structure Prediction - A Practical Approach. *Oxford University Press, Oxford*.

Strong, M., Chandy, K. G. & Gutman, G. A. (1993) Molecular evolution of voltage-sensitive ion channel genes: on the origins of electrical excitability. *Mol. Biol. Evol.*, **10**, 221-242.

Sun, M., Goldin, E., Stahl, S., Falardeau, J. L., Kennedy, J. C., Acierno, J. S. Jr., Bove, C., Kaneski, C. R., Nagle, J., Bromley, M. C., Colman, M., Schiffmann, R. & Slaugenhaupt, S. A. (2000) Mucolipidosis type IV is caused by mutations in a gene encoding a novel transient receptor potential channel. *Hum. Mol. Genet.*, **9**, 2471-2478.

Sutcliffe MJ, Hayes FR, Blundell TL (1987) Knowledge based modeling of homologous proteins, part II: Rules for the conformations of substituted side chains. *Prot Eng*, **1**, 385–92.

Swartz, K. J. (2004) Towards a structural view of gating in potassium channels. *Nat. Rev. Neurosci.,* **5**, 905-16.

Swofford, D. L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4, Sinauer Associates, Sunderland, MA.

Swope, W. C., Andersen, H. C., Berens, P. H., Wilson, K. R. (1982) A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, **76**, 637–649.

Tholema, N., Bakker, E. P., Suzuki, A., Nakamur, T. (1999) Change to alanine of one out of four selectivity filter glycines in KtrB causes a two orders of magnitude decrease in the affinities for both K^+ and Na^+ of the Na^+ dependent K^+ uptake system KtrAB from *Vibrio alginolyticus*. *FEBS letters*, **450**, 217–220.

Tholema, N., Vor der Brüggen, M., Mäser, P., Nakamura, T., Schroeder, J. I., Kobayashi, H., Uozumi, N. & Bakker, E. P. (2005) All four putative selectivity filter glycine residues in KtrB are essential for high affinity and selective K^+ uptake by the KtrAB system from *Vibrio alginolyticus*. *Journal of biological chemistry*, **280**, 41146–41154.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.,* **25**, 4876-4882 .

van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman,

D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen R. & Berendsen, H. J. C. (2010) Gromacs User Manual version 4.5.6, www.gromacs.org.

van Gunsteren, W. F., Berendsen, H. J. C. (1988) A leap-frog algorithm for stochastic dynamics. *Mol. Sim.,* **1**,173–185.

van Meer, G., Voelker, D. R. & Feigenson, G. W. (2008) Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.*, **9**, 112-124.

van Rossum, D. B., Patterson, R. L., Sharma, S., Barrow, R. K., Kornberg, M., Gill, D. L. & Snyder, S. H. (2005) Phospholipase Cγ1 controls surface expression of TRPC3 through an intermolecular PH domain. *Nature*, **434**, 99–104.

Venkatachalam, K. & Montell, C. (2007) TRP Channels. *Annu. Rev. Biochem*., **76**, 387–417.

Verlet, L. (1967) Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **159**, 98-103.

Vieira-Pires, R. S., Szollosi, A. & Morais-Cabral, J. H. (2013) The structure of the KtrAB potassium transporter. *Nature*, **496**, 323-328.

Voets, T., Owsianik, G., Janssens, A., Talavera, K. & Nilius, B. (2007) TRPM8 voltage sensor mutants reveal a mechanism for integrating thermal and chemical stimuli. *Nat. Chem. Biol.*, **3**, 174-182.

Vieira-Pires, R. S. & Morais-Cabral, J. H. (2010) 3_{10} helices in channels and other membrane proteins. *J. Gen. Physiol.,* **136**, 585–592.

Vriend G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.,* **8**, 52-56.

Wang, H. & Woolf, C. J. (2005) Pain TRPs. *Neuron*, **46**, 9-12.

Wang, X., Zhang, X., Dong, X. P., Samie, M., Li, X., Cheng, X., Goschka, A., Shen, D., Zhou, Y., Harlow, J., Zhu, M. X., Clapham, D. E., Ren, D. & Xu, H. (2012) TPC proteins are

phosphoinositide- activated sodium-selective ion channels in endosomes and lysosomes. *Cell*, **151**, 372-383.

Wang, Y. Y., Chang, R. B., Waters, H. N., McKemy, D. D. & Liman, E. R. (2008) The nociceptor ion channel TRPA1 is potentiated and inactivated by permeating Ca ions. *J. Biol. Chem.*, **283**, 32691-32703.

Weber, G. (1975) Energetics of ligand binding to proteins. *Adv. Protein Chem.,* **29**, 1-83.

Wen, W., Yan, J. & Zhang, M. (2006) Structural Characterization of the Split Pleckstrin Homology domain in Phospholipase C-γ1 and Its interaction with TRPC3. *J. Biol. Chem.,* **281**, 12060–12068.

Wiederstein, M. & Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.,* **35,** 407-410.

Xiao, B., Dubin, A. E., Bursulaya, B., Viswanath, V., Jegla, T. J., et al. (2008) Identification of transmembrane domain 5 as a critical molecular determinant of menthol sensitivity in mammalian TRPA1 channels. *J. Neurosci.,* **28**, 9640–9651.

Xiong, J. (2006) Essential Bioinformatics. Cambridge University Press, ISBN: 0-521-60082-0; 339.

Yarov-Yarovoy, V., DeCaen, P. G., Westenbroek, R. E., Pan, C. Y., Scheuer, T., Baker, D. & Catterall, W. A. (2012) Structural basis for gating charge movement in the voltage sensor of a sodium channel. *Proceedings of the National Academy of Sciences of the United States of America* **109**,E93–E102.

Yu, F.H. & Catterall, W.A. (2004) The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci. STKE*, **253**, re15

Zhou, X. L., Loukin, S. H., Coria, R., Kung, C. & Saimi, Y. (2005) Heterologously expressed fungal transient receptor potential channels retain mechanosensitivity in vitro and osmotic response in vivo. *Eur. Biophys. J.*, **34**, 413–422.

Zhou, Y. & MacKinnon, R. (2003) The occupancy of ions in the K+ selectivity filter: charge

balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J Mol Biol,* **333**, 965–975.

Zhou, Y., Morais-Cabral, J.H., Kaufman, A. & MacKinnon, R. (2001). Chemistry of ion coordination and hydration revealed by a K^+ channel-Fab complex at 2.0 A resolution. *Nature*, **414**, 43–48.

Zurborg, S., Yurgionas, B., Jira, J. A., Caspani, O. & Heppenstall, P. A. (2007) Direct activation of the ion channel TRPA1 by Ca^{2+}. *Nat. Neurosci.,* **10**, 277-279.