TECHNISCHE
UNIVERSITÄT
DARMSTADT

# *IN SILICO* STUDIES ON PROTEINS FOR SYNTHETIC BIOLOGY

vom Fachbereich Biologie
der Technischen Universität Darmstadt

zur Erlangung des Grades
Doctor rerum naturalium
(Dr. rer. nat.)

Dissertation
von

Christine Groß, M.Sc.

Erstgutachter: Prof. Dr. Kay Hamacher
Zweitgutachter: Prof. Dr. Gerhard Thiel

Darmstadt 2018

Groß, Christine: *In Silico* Studies on Proteins for Synthetic Biology

Darmstadt, Technische Universität Darmstadt

Jahr der Veröffentlichung der Dissertation auf TUprints: 2019

URN: urn:nbn:de:tuda-tuprints-83488

Tag der mündlichen Prüfung: 13.12.2018

## ABSTRACT

Synthetic biology develops artificial biomolecules or biological systems with novel functionalities for diverse applications in research, medicine or industry. This thesis focuses on *in silico* studies of three proteins that are promising candidates for enzymatic plastic waste treatment and highly sensitive biosensors, respectively.

The first candidate is the enzyme *Fusarium solani* Cutinase [122], which is able to degrade synthetic polymers, like PET. It allows for the development of an environmental friendly and sustainable solution for plastic waste treatment on an industrial scale. As the wildype enzyme loses its activity during the process of PET degradation, a rational design approach was followed, to improve the activity of this enzyme for PET as substrate. Via MD simulations and linear response theory (LRT) [84] based on coarse-grained elastic network models, the reason for the loss of activity could be identified. Based on the knownledge gained, mutants with improved activity for PET were proposed. In the context of this study, an extension for the LRT method similar to that of a previous study [104] was developed.

The second protein system, the hyperpolarization-activated cyclic nucleotide-gated cation (HCN) channel [169], regulates the flux of ions across biological membranes by changes in the membrane voltage and binding of the ligand cAMP. Hence, it is an ideal model for studying the interplay of different domains during the gating process. Together with plenty of other ion channels, it can also serve as building blocks for the assembly of different domains to design synthetic ion channels with novel functionalities. To understand the complex mechanism of HCN gating, the extension of the LRT method was adjusted to work for a tetramer and was used to determine the conformational changes that occur upon binding of the ligand cAMP. In this context, movements in the transmembrane domains that are involved in the gating process were discovered for the first time. They provide important information on the complex gating mechanism and enable a directed planning of further experimental and theoretical investigations.

Small viral pore forming proteins also enable the flux of ions across biological membranes and therefore can be seen as viral companions of ion channels. The third protein is such a pore forming protein from HIV and simian relatives SIV, called Vpu [34]. As this small protein is less complex than ion channels but also exhibits ion channel function, it is another candidate to serve as building block for the design of artificial ion channels. To consider the Vpu protein as possible building block, the formation of an ion conducting pore has to be a reliable property. In this thesis, the evolutionary conservation of ion channel formation was proved by computing the Shannon entropy [192] for involved residues based on a multiple sequence alignment. Although the study could not clarify the role of the ion channel function for virus release or replication, the detected evolutionary conservation serves as proof for the functional significance. Hence, this protein reliably forms an ion conducting pore and can be further considered as possible building block for the assembly of synthetic ion channels.

# ZUSAMMENFASSUNG

Die synthetische Biologie entwickelt artifizielle Biomoleküle oder ganze biologische Systeme mit neuartigen Funktionen für Anwendungen in der Forschung, Medizin oder Industrie. Der Fokus dieser Arbeit liegt auf *in silico* Studien dreier Proteine, die vielverprechende Kandidaten für die enzymatische Verwertung von Plastikmüll bzw. hochsensitive Biosensoren darstellen.

Der erste Kandidat ist das Enzym *Fusarium solani* Cutinase [122], welches in der Lage ist, synthetische Polymere, wie PET, abzubauen. Es ermöglicht daher die Etablierung einer umweltfreundlichen und nachhaltigen Lösung zum Abbau von Plastikmüll in einem industriellen Maßstab. Da das Wildtyp Enzym während des Abbaus von PET an Aktivität verliert, wurde ein rationaler Design Ansatz verfolgt, um die Aktivität des Enzyms gegenüber PET als Substrat zu verbessern. Mittels MD Simulationen und der *Linear Response Theory* (LRT) [84] basierend auf reduzierten elastischen Netzwerkmodellen konnte die Ursache für den Aktivitätsverlust identifiziert werden. Basierend auf den gewonnenen Erkenntnissen wurden Mutanten mit einer erhöhten Aktivität gegenüber PET vorgeschlagen. Im Rahmen dieser Studie wurde für die LRT Methode eine Erweiterung ähnlich der aus einer früheren Studie [104] entwickelt.

Das zweite Proteinsystem, der *hyperpolarization-activated cyclic nucleotide-gated* (HCN) Kanal [169], reguliert den Ionenfluss durch biologische Membranen basierend auf Änderungen der Membranspannung sowie durch Binden des Liganden cAMP. Dieser Ionenkanal ist ein ideales Modell, um das Zusammenspiel verschiedener Domänen während des Schaltvorgangs zu untersuchen. Desweiteren bildet er mit einer Vielzahl anderer Ionenkanäle eine Art Baukasten, sodass nach dem Baukastenprinzip verschiedene Domänen zu einem synthetischen Ionenkanal mit neuartigen Eigenschaften zusammengesetzt werden können. Um den komplexen Schaltmechanismus des HCN Kanals zu verstehen, wurde die für ein Monomer implementierte Erweiterung der LRT Methode an Tetramere angepasst und verwendet, um die resultierenden Konformationsänderungen nach Binden des Liganden cAMP zu bestimmen. In diesem Zusammenhang wurden erstmals Bewegungen in den transmembranen Bereichen beobachtet, die am Kanalschalten beteiligt sind. Diese liefern wichtige Hinweise zum mechanistischen Ablauf des komplexen Kanalschaltens und ermöglichen die gezielte Planung weiterer experimenteller und theoretischer Untersuchungen.

Kleine virale porenbildende Proteine ermöglichen ebenfalls den Fluss von Ionen durch biologische Membranen und können daher als virale Pendants zu Ionenkanälen betrachtet werden. Das dritte Protein ist solch ein porenbildendes Protein von HIV und seinem affenartigen Verwandten SIV und wird als Vpu Protein [34] bezeichnet. Da dieses kleine Protein weniger komplex ist als Ionenkanäle aber ebenfalls eine Ionenkanalaktivität aufweist, ist es ein weiterer Kandidat zur Erweiterung des Baukastens für die Entwicklung artifizieller Ionenkanäle. Um das Vpu Protein als möglichen Baustein in Betracht zu ziehen, muss das Bilden der ionenleitenden Pore eine zuverlässige Eigenschaft sein. In dieser Arbeit wurde die evo-

lutionäre Konserviertheit der Ionenkanalbildung durch Berechnung der *Shannon entropy* [192] für betroffene Residuen basierend auf einem multiplen Sequenzalignment nachgewiesen. Obwohl die Studie nicht aufklären konnte, welche Rolle die Funktion als Ionenkanal für die Virusfreisetzung und Replikation spielt, kann die evolutionäre Konserviertheit als Beweis für eine funktionale Signifikanz gedeutet werden. Somit bildet das Vpu Protein zuverlässig ionenleitende Poren und kann weiterhin als möglicher Baustein für den Zusammenbau synthetischer Ionenkanäle berücksichtigt werden.

CONTENTS

# 1

# GENERAL INTRODUCTION

In the era of *omics* researchers come up with plenty of new techniques that enable even more detailed and faster research to understand the function of biomolecules and how they contribute to functional organisms. *Genomics*, for example, benefits from the use of more efficient sequencing methods, like next generation sequencing (NGS) [27, 105, 143, 164], and also from a recent breakthrough in enzymatic DNA synthesis [148]. For *proteomics* and *structural biology* on the other hand, the growing number of high resolution crystall/NMR structures found in the protein data bank (`www.rcsb.org`) [21] and recent advances in cryo-electron microscopy (cryoEM) [12, 167] build an excellent basis for structural and functional studies on proteins. In this context, computer simulations become more and more relevant and play an essential role, e.g., for structure analysis and prediction [55, 187], for the prediction of conformational changes and protein dynamics [39, 40, 187], for creating design hypotheses [66] or for drug discovery and design [29, 179]. Since the development of (super)computers in the 1980s started, the field of computational biology has gained momentum and is still developing with enormous speed [175]. This results in a vast spectrum of simulation and computational approaches (*in silico* methods) ranging from molecular dynamics (MD) simulations [94] over normal mode analysis (NMA) [75, 197], elastic network models (ENMs) [8], to Monte-Carlo (MC) simulations [6, 93]. Computational biology not only benefits from the development of faster hard- and software, but also from improvements or new developments in experimental approaches, which lead to more efficient data collection. On the other hand, lots of methods for data collection, e.g., high-throughput screenings (HTS), are dependent from computational support and bioinformatics [68]. These data can further be used for training of machine learning approaches [198]. This shows how wet-lab methods and computational biology go hand in hand and support each other.

Whenever biology not only aims at understanding biological systems and processes, but also wants to influence them and develop novel functionalities, we talk about *synthetic biology* [7, 17]. Together with engineering, this new interdisciplinary branch has been emerging and aims at building artificial biological systems or modules for diverse applications in research, medicine or industry. Therefore, it combines various disciplines, like molecular engineering, biophysics, or systems engineering [56, 98, 160]. Synthetic biology tools that have been invented during the recent years are ranging from synthetic promotors for transcriptional control [26, 145] or artificial ribozymes or riboswitches for translational control [20, 41, 65, 222] over complex synthetic circuits, like e.g., for cancer cell recognition and destruction [225]. Further applications are still driven by the development of novel techniques, e.g. like the CRISPR-Cas gene editing tool [45, 176] or an artificial cell-cell communication tool [77]. In contrast to that, when synthetic biology directly focusses on the protein level, protein engineering plays a significant role to increase the activity of enzymes under production conditions, alter the substrate

and product specifity or alter regulatory elements of proteins [56]. Thus, biological compounds may be single molecules, proteins, DNA/RNA or even whole cells that have a special function or contribute to such a function so that they are interesting for the use in a technical application [7, 160]. The biological compounds can also serve as models for synthetic replicas or they can be modified, e.g., in terms of mutagenesis, prior to the application in a technical device [189].

An interesting biological compound for the use in synthetic biology is the small enzyme, called *Fusarium solani* Cutinase (*Fs*C). It is a very promising candidate for the application of a sustainable solution of plastic waste treatment on an industrial scale due to its ability to hydrolyze synthetic polymers [31, 78, 122, 166]. This application would contribute to a reduction of the worldwide pollution by plastic waste and microplastics, but prior to this, current limitations in the enzymatic activity for synthetic substrates have to be circumvented.

Ion channels are also interesting biological compounds for the use in synthetic biology, as they serve as biological models for highly selective and controllable synthetic nanopores. The development of such synthetic nanopores with novel properties is pursued by the LOEWE Research Cluster "iNAPO – ion conducting nanopores" for the application in a biosensor, e.g., for environmental analytics, process sensors or medical diagnostics [48]. Ion channels are present in biological membranes and can be selectively switched on and off by various factors to allow for a highly selective flux of ions through the membrane [80]. There are lots of different ion channels that share the same general architecture of the central ion conducting pore, but exhibit different additional domains, that are responsible for their typical properties [126]. Thus, ion channels can not only serve as models for synthetic nanopores; they can also serve as building blocks for the reassembly of different domains to synthetic nanopores. On the other hand, there exist viral channel forming proteins (VCPs), that also exhibit ion channel function. Though their simplicity, they also assemble to a quarternary structure, like ion channels, and make the membrane permeable for ions [54, 115]. For this reason, VCPs are interesting models to study the formation of ion channels and in the context of synthetic biology they are also interesting as an alternative building block for the central ion conducting pore for synthetic ion channels. However, the gating processes and underlying mechanisms of the different ion channels or VCPs have to be discovered first in detail.

In the scope of this thesis, *in silico* methods are used to better understand the above introduced proteins (*Fs*C, ion channels and VCPs) as promising candidates for applications in the context of synthetic biology. The following sections give an overview on these candidates and discuss their potentials and current limitations for the applicability in industrial or technical devices. Subsequently, the research challenges and corresponding goals of this thesis are presented. Afterwards, section 1.3 gives a short introduction into conceivable methods of computational biology and discusses how the goals of this thesis are pursued.

## 1.1 FUSARIUM SOLANI CUTINASE FOR ENZYMATIC PET DEGRADATION

### 1.1.1 *Overview*

The worldwide environmental pollution by plastic waste steadily increases and causes fatal damages in biosystems [137]. Most hazardous are the plastic particles, with diameters on a micro- or nanometer scale, which we cannot even see. Such microplastics and nanoplastics are spread all over the world and already pollute our air, sediments, oceans etc. [3, 134]. Nanoparticles of plastic material that are detected in the world's oceans in great quantities can accumulate toxic substances on their surface [63, 106]. Recent studies show that such particles have already arrived in our food chain [9, 13, 106]. Therefore, the development of efficient sustainable solutions for plastic waste treatment on an industrial scale is advisable. What makes plastics so convenient for everyday life, e.g., the non-biodegradability, high durability or resistance to moistures, makes them even harder to be decomposed when they ought to be recycled [106]. While existing chemical and mechanical methods suffer from the use of environmentally harmful substances or high energy costs [58, 210], the enzymatic degradation would allow for an energy-efficient and biocompatible degradation process. The mentioned cutinase (*Fs*C) is a secreted enzyme from the plant pathogenic fungi *Fusarium solani pisi*, that easily enters the plant by hydrolyzing cutin, the main component of the plant cuticle [31, 122]. *Fs*C belongs to the group of serine esterases and has a classical Ser-His-Asp (S-H-D) catalytic triad in their active site, which is exposed to the solvent (see Figure 1.1).

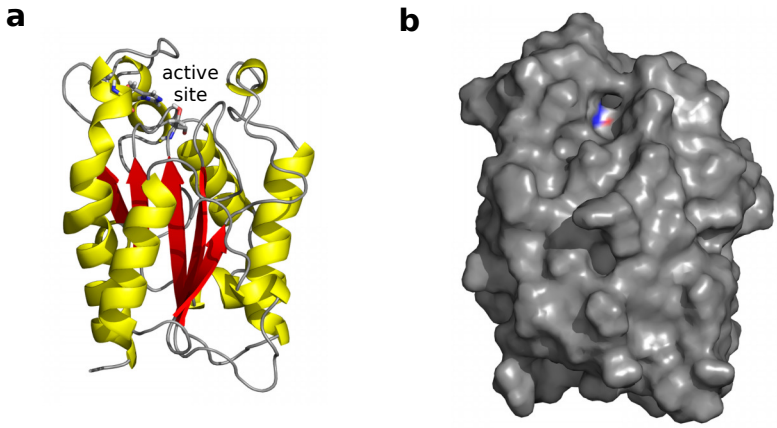


Figure 1.1: (**a**) *Fs*C shown in cartoon representation (PDB code 1CEX) [123] with helices highlighted in yellow and $\beta$-sheets highlighted in red. The catalytic triad (S120, D175, H188) in the active site is shown as stick representation. (**b**) *Fs*C shown in surface representation with the exposed catalytic triad residues visible as blue/red shading.

As *Fs*C is able to degrade a natural high-molecular weight polyester, like cutin, it became interesting for the enzymatic degradation of synthetic high-molecular weight polyesters. In fact, it was shown that this enzyme is able to degrade polyethylene terephthalate (PET), the synthetic polymer most commonly used worldwide [31, 122]. Unfortunately, the activity of *Fs*C toward the degradation of PET and other synthetic polymers is not high enough to be applicable for industrial

use. Instead, the already low activity further decreases over time during the PET degradation process. Thus, researchers seek at optimizing the wildtype enzyme by mutagenesis to improve the activity toward PET degradation [78, 166]. To avoid a trial and error procedure or random mutagenesis studies [31, 78, 122, 166] which have been numerously undertaken but without significant success, in this thesis, this challenge is addressed by using a rational design approach.

### 1.1.2 *Research Challenges and Goals*

For the optimization of *Fs*C toward an improved activity of PET degradation by rational design, it is necessary to understand what happens during the process of PET hydrolysis and why a loss of activity occurs. The fact that *Fs*C is generally able to degrade PET shows that the size of the active site cannot be the reason. Instead, the loss of activity must be caused by an instance that occurs during the ongoing degradation process and does not exist at the beginning of this process. This leads to the assumption that the increasing amounts of cleavage products play a key role in the reduction of activity. The cleavage products do not exist at the beginning of the process and do also not arise during the degradation of the natural substrate of *Fs*C, because cutin has a totally different composition. In contrast to cutin, that consists of fatty acids of different lengths [31], PET is cleaved into oligomers and monomers of ethylene glycol (EG) and terephthalic acid (TPA) [210]. To confirm this assumption, the influences of the cleavage products on the activity of *Fs*C have to be investigated. For the simulation of enzymatic PET degradation into its cleavage products, combined quantum and molecular mechanical (QM/MM) methods [153] were necessary, but they are prohibitively expensive for a system of that size, and classic MD simulations generally are not able to simulate enzymatic reactions [185, 231]. Thus, possible effects of cleavage products on the activity of *Fs*C have to be studied indirectly. This could be achieved by studying the effects of the cleavage products on the solvent, because altered solvent properties consequently might influence the enzyme properties and functionality. While the cleavage product TPA is insoluble in water and probably directly sediments in the reaction solution, the small water analogue EG is more likely to influence the properties of the solvent and accordingly of *Fs*C. It is also known as antifreezing agent and in this context typically increases the viscosity and density of solvents [178, 195, 205]. Thus, the effects of increasing amounts of EG in the solvent on the structure and dynamics of *Fs*C should be investigated. Consequently, it should be elucidated, whether these possible changes in structure and dynamics of *Fs*C are in connection with its activity. As this activity cannot be measured nor simulated, other ways to make a link between possibly affected regions of *Fs*C by EG and the loss of activity must be found.

Therefore, the goal of the thesis is to provide a simulation approach that is able to (i) investigate the effects of increasing amounts of EG on the structure and dynamics of *Fs*C, (ii) make a link between structure and dynamics of *Fs*C and its activity and (iii) investigate whether and how the activity is affected. Based on the findings, the thesis aims at providing an estimate or multiple estimates for mutants that do not exhibit the loss of activity during the process of PET degradation, and thus, may be applicable for PET degradation in a large scale industrial process.

## 1.2 BIOLOGICAL MODELS FOR SYNTHETIC NANOPORES

### 1.2.1 *Overview*

There exist two promising classes of candidates that serve as biological models and building blocks for synthetic nanopores in this thesis: (i) ion channels and (ii) small viral channel forming proteins (VCPs).

ION CHANNELS are protein complexes of two or four subunits that are located in biological membranes and enable the flux of ions from one side of the membrane to the other side through their central ion conducting pore [124, 131]. Most ion channels are highly selective for different types of ions, like sodium, potassium or chloride ions and allow for transport rates of the corresponding ions of up to $10^8$ ions/s [124, 227]. This makes them very important for various functions in living organisms, e.g., they are responsible for the signal transduction between neurons in our nervous system, control our heart beat, or regulate the level of blood sugar [89]. The opening and closing of ion channels, which is called *gating*, underlies different mechanisms. Some ion channels have a voltage-dependent gating [119], while others show mechano-sensitivity [144], pH-dependence [38], or gate upon ligand-binding [83]. Due to their properties as highly selective and controllable biological nanopores, ion channels are ideal models for synthetic biologists, that aim at building synthetic nanopores for the application in biosensors or technical devices. While ion channels share the same general architecture in the central ion conducting pore, they differ in additional domains, for e.g. ligand-binding or voltage-detection. Because of their modularity, ion channels can serve as building blocks for the reassembly of synthetic nanopores or the integration of artificial properties into existing ion channels.

In this thesis, the focus lies on voltage-dependent ion channels, as their gating process has not yet been sufficiently understood. In particular, the hyperpolarization-activated cyclic nucleotide-gated (HCN) channels, that come in four isoforms (HCN1-4), are focused [154]. They are voltage-dependent potassium/sodium channels that are also influenced by binding of a small ligand, named cyclic adenosine monophosphate (cAMP) [22]. HCN channels contribute to the pacemaker activity in the heart as well as in neural networks in the brain [22, 154, 165]. They are very interesting ion channels, as they open or close when changes in the membrane voltage occur but in the opposite manner as the majority of voltage-dependent ion channels do. The typical voltage-dependent potassium channels (Kv channels) open upon depolarization of the membran, i. e. when the potential is less negative then the resting potential [119]. HCN channels, in contrast, open upon hyperpolarization, which means that the membrane potential must be more negative then the resting potential. Although both types of potassium channels have a similar architecture of channel pore and voltage-sensing domain (VSD) their gating behavior is totally different, and this invites researchers to gain insights into the underlying mechanisms. For Kv channels, a conclusive mechanism of voltage-dependent gating has been proposed based on experimental data that is strengthened by a set of MD simulations [90, 120, 121]. Due to the poor availability of molecular structure information of HCN channels, even plenty of

experimental data were not yet sufficient to fully clarify the complex HCN gating mechanism [16, 36, 156, 170, 209]. Since the first molecular structure of a full HCN channel (HCN1) was published by Lee and MacKinnon in 2017 (Figure 1.2), it was possible to compare the structure of Kv channels and HCN channels in detail and first differences in their architecture have been described [112]. Now, it has to be investigated, whether these differences are the cause for the diverging gating behaviors.

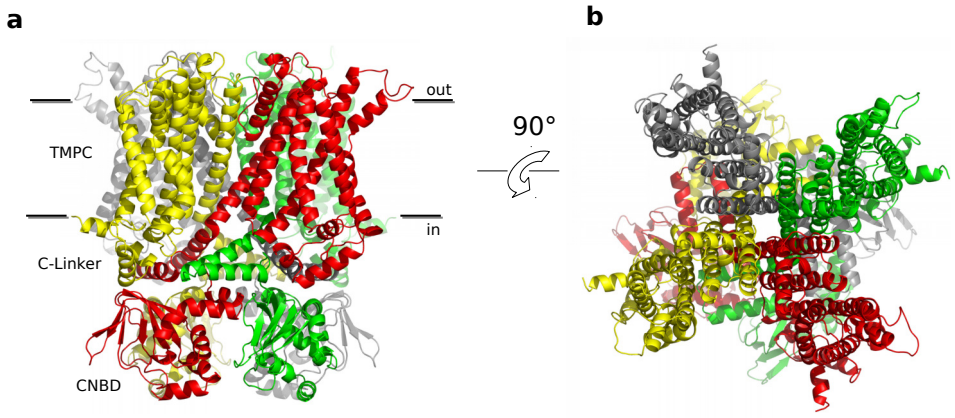


Figure 1.2: HCN1 channel in cartoon representation with differently colored subunits (PDB code 5U6O [112] with remodeled loop regions) (**a**) Side view to demonstrate how the channel is embedded in the membrane. The localization of the different domains with the terms used in this thesis are additionally shown (TMPC, C-Linker, CNBD). (**b**) Top view to demonstrate how the assembly of four subunits forms a central ion conducting pore.

VIRAL CHANNEL FORMING PROTEINS (VCPS) can assemble and integrate into membranes and subsequently form pores that enable the flux of ions across the membranes [54, 115]. Due to this property, such VCPs might serve as alternative models or building blocks for the central ion conducting pore in synthetic ion channels. In this thesis, the 81 amino acid type-I integral membrane protein Vpu encoded by HIV-1 and SIV [34, 128, 194] is investigated. It consists of three helices, an N-terminal hydrophobic membrane-spanning helix and two cytoplasmic helices that are located in plane with the membrane (Figure 1.3). This Vpu protein has multiple functions for viral infectivity and is responsible for augmented virus release, e.g. by counteracting the restriction factor tetherin or degrading $CD_4$ receptors [146, 206, 230]. Besides these functions, the transmembrane helix assembles with transmembrane helices of other Vpu proteins so that they form an ion conducting pore. So far, it is not clear whether and how this ion channel activity contributes to the above mentioned funtions and whether it is evolutionary conserved or just an epiphenomenon of the Vpu protein from one HIV-1 type. For considering the Vpu protein as alternative pore for synthetic ion channels, its functionality and the underlying molecular mechanism have to be better understood.

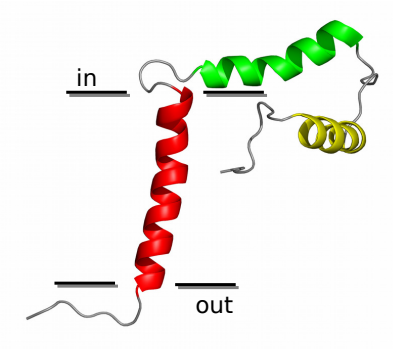


Figure 1.3: Vpu protein (PDB code 2N28) [230] in cartoon representation with the membrane-spanning helix shwon in red and the two cytoplasmic helices that are in plane with the membrane shown in green and yellow.

### 1.2.2 *Research Challenges and Goals*

In the context of this thesis, computational approaches are pursued to make a link between the experimental data obtained for HCN channels and Vpu proteins and the still open questions of HCN channel gating and Vpu ion channel activity, respectively.

The open questions of HCN channel gating especially concern the allosteric nature of the gating mechanism by voltage and binding of cAMP. While the cyclic-nucleotide binding domain (CNBD) and the binding of cAMP has been well studied [44, 118, 170, 229], it has not yet been understood how both regulatory signals are processed in the context of the whole protein - especially because the bigger part of the ion channel, including the VSD, where changes in membrane voltage are detected, is membrane-embedded. For this transmembrane portion of the channel (TMPC), no molecular structure has been available until the full structure of HCN1 was published recently [112]. Experimental data as well as molecular structure information of the CNBD with and without cAMP bound to the binding pocket already revealed conformational changes in the CNBD and connected parts [163, 229]. Now, having the full structure (see Figure 1.2), again with and without cAMP bound to the CNBD [112], the open questions can be addressed. They all mainly deal with the central questions of how the conformational changes caused by cAMP binding are transmitted from the CNBD to the TMPC and how they are related to gating movements in the VSD and central ion conducting pore. Unfortunately, both full structures of HCN1 only show the closed configuration of the channel, due to the fact that the process of structure determination takes place without impressed voltage. Thus, gating movements cannot be deduced by comparing the cAMP-free and cAMP-bound HCN1 structures.

Therefore, the goal of this thesis is to circumvent the problem of limited structural information by simulating the channel movements in the TMPC based on the known movements in the CNBD. With these simulations, it should be answered (i) how the conformational information is transmitted between the CNBD and the TMPC, (ii) how these conformational changes in the CNBD are related to the gating movements in the VSD and central pore, (iii) what are the movements in the VSD (especially S4 domain) and central pore (especially S6 domain) and (iv) how

may cAMP binding favor the opening of the channel gate. The here gained results hopefully contribute to a possible mechanistical model for the gating process of HCN channels.

The open questions on Vpu protein function especially concern the role of ion channel activity in the context of viral replication and release. It is not clear whether ion channel activity has any functional significance in the context of virus release or whether the formation of an ion conducting pore is just an epiphenomenon of one HIV-1 type. For considering the Vpu protein as alternative building block for synthetic nanopores, the formation of such a pore must be a reliable property. An important hint for the functional significance is the evolutionary pressure that leads to conservation of functionally relevant structures. Therefore, as a first step to better understand the ion channel function of the Vpu protein and to be able to decide whether it can further be considered as alternative building block for synthetic nanopores, in this thesis, the degree of evolutionary conservation of the involved residues is studied.

## 1.3 METHODOLOGY

To address the above research challenges and successfully reach the research goals, simulation approaches that are suitable from a conceptual perspective under the given availability of data have to be applied. This means that the chosen methods must be able to simulate and analyze the biochemical and physical aspects of the concrete research question based on the given data, like molecular structures, activity measures, etc. In this context also the degree of abstraction plays an essential role. While some research questions address phenomena that take place on the atomic level, like e.g. side chain interactions with the solvent, others only address movements of larger domains, so that simulations with a reduced resolution may be sufficient. Furthermore, it is important to be aware of the timescale over which a phenomenon occurs in nature and to chose a method that is able to simulate phenomena of this timescale. In the following section computational methods with the potential to be used for the above research goals are introduced to get an overview about their fields of application and which timescales they are able to simulate.

### 1.3.1 *Overview of Potential Computational Methods*

A main differentiator of computational methods is whether the molecular structure of the biomolecule, e.g. from X-ray crystallography, is available or not. As this thesis only focuses on proteins, in the following only proteins are mentioned. Of course, for some methods, the same is also true for other types of biomolecules.

The availability of molecular structures of proteins allows for studying their dynamics and corresponding movements to better understand their functionalities. (*Ab-initio*) quantum mechanics (QM) as well as the combined method of quantum mechanics with *semi-empirical* molecular mechanics (QM/MM) are able to simulate chemical reactions but they are prohibitively expensive for large systems as they are based on the solution of the Schrödinger equation [149, 185]. *All-atom MD simulations*, that integrate Newton's equations of motion over time steps of

few femtoseconds, rely on *semi-empirical* force fields and provide a very realistic representation of biomolecular dynamics. Unfortunately, the high resolution also requires lots of computational ressources [40, 94]. While they are able to study short-term molecular motions, like side-chain rotations or loop motions (Figure 1.1), for large proteins, MD simulations hardly reach timescales of micro- to milliseconds [152], over which larger domain movements, like folding/unfolding, allosteric transitions or binding occur (Figure 1.1) [52]. Only with a supercomputer, like Anton [186], trajectories can reach the millisecond scale, but of course the access to Anton is limited.

Hence, for studying large-scale movements, reduction of the atomistic resolution is necessary. This can be achieved by *coarse-graining*, where groups of atoms or molecules are reduced to abstract shapes. For MD simulations on such a coarse-grained level, special force fields, e.g. the `MARTINI` force field for proteins [139], have been developed. They allow for running simulations that are 2-3 orders of mangnitudes longer compared to atomistic simulations, but with a reduction of accuracy.

There exist further coarse-grained methods, that do not rely on MD simulations. In an *elastic network model* (ENM), the protein is reduced to an elastic network of beads and springs, where the residues are represented by beads and the molecular interactions (bonded or nonbonded) between them are represented as springs with corresponding spring constants [8, 43, 142]. Whether residues are connected or not, is defined by a given threshold for spatial distances between them (e.g. based on the positions of the C$\alpha$ atoms). Depending on how the fluctuations of the residues are treated, we distinguish between *Gaussian network models* (GNMs) and *anisotropic network models* (ANMs) [8, 202]. Furthermore, the *normal mode analysis* (NMA) is a successful method to study collective motions of biomolecules on the atomic as well as on a coarse-grained level [8, 40, 197]. For this, the Hessian matrix (second derivates of the potential function) of the molecular structure is diagonalized and the deformation of the system along the low frequency normal modes is studied. In conjunction with an ANM, we have the advantage, that the structure coordinates directly provide an analytical expression for the Hessian matrix [8, 11, 40]. Figure 1.4 gives an overview of the timescales over which some molecular motions in proteins occur and which timescales are reachable by MD simulations and NMA.

If the molecular structure of a protein is not known, it is also possible to study its structure and function based on the underlying amino acid sequence. In this context, *multiple sequence alignments* (MSAs) can give important insights into the structure or function based on evolutionary information, like e.g., the correlation of residues. Nevertheless, these methods only work, when enough, i.e. several hundreds or more, sequences of homologous proteins from different organisms are aligned to each other. Therefore, the choice of the search tool (e.g., `BLAST` [4], `FASTA` [150]) for homologous protein sequences, as well as the choice of the MSA algorithm (e.g., `MUSCLE` [47], `MAFFT` [95]) and the underlying substitution matrix (e.g., `BLOSUM` [76], `PFASUM` [97]) influence the resulting MSA. In such an MSA, several measures can be computed to gain information on the protein structure and function. The degree of evoluationary conservation of residues can be determined by computing the Shannon entropy [192]. This measure often correlates with the

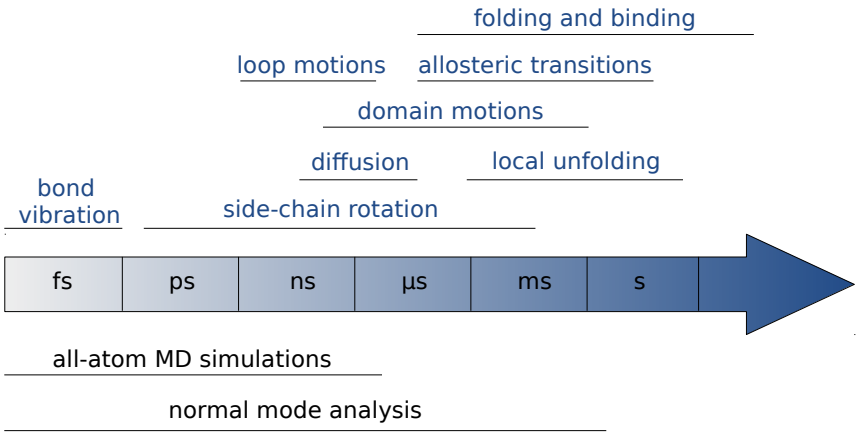


Figure 1.4: Timescales of computational methods for studying the dynamics of proteins (below) in comparison with the timescales over which some typical protein movements occur (above). The illustration has been adapted from [52].

relevance of the amino acid for the protein function. Furthermore, compensatory mutations and coevolving residues can provide informations on which residues must be in local proximity because they interact with each other although in the amino acid sequence they are far away. In this context, the computation of Mutual Information (MI) [60, 69] or direct coupling analysis (DCA) [141] are common methods. As these methods are not able to discrimiate between direct and indirect contacts, and therefore, do not work very well in predicting the structure of proteins, advanced methods, like higher-order MI [218] or three-body DCA [177] have been developed. Of course, such methods based on the protein sequence provide less information on the mechanisms of protein function, but for a protein with unknown structure they provide important insights on how the structure might look like and which residues probably are essential for the protein function. Furthermore, they can provide important information for e.g. homology modeling.

### 1.3.2 *Mix of MD Simulations and Linear Response Theory for Rational Design of FsC*

For the rational design approach that aims at optimizing the activity of *Fs*C toward PET degradation, it is not possible to simulate the reaction of PET degradation within the active site of *Fs*C to learn what happens during that process. This is because chemical reactions (bond breaking/ bond formation) in general cannot be simulated via MD simulation and for QM/MM simulations the size of the system is far too big [149, 185]. Therefore, as described above, this process has to be investigated indirectly by studying the effects of increasing amounts of the degradation product EG on the structure and dynamics of *Fs*C and on the solvent properties. This can be done via short MD simulations with different concentrations of EG, performed by Sven Jager (Computational Biology & Simulation Group, TU Darmstadt), which substitute the repetitive cycles of hydrolysis reactions during which increasing amounts of EG arise. Based on the MD trajectories, the effects on the structure and dynamics of *Fs*C are further investigated by computing common measures, like root mean square deviation (RMSD) and

root mean square fluctuation (RMSF). To check for possible accumulations on the protein surface, surface density calculations (SDC) are computed. The effects on the solvent properties, e.g. the viscosity, are studied based on the MD trajectories by computing further parameters.

To make a link between the influences of EG on the structure and dynamics of *Fs*C and its activity, a mechanical model on a coarse-grained level, called linear response theory (LRT) [84], is applied. With this model, which is elucidated in Chapter 2, it is possible to determine conformational changes within the whole protein that occur after binding of the substrate to the active site. This method is based on an ANM, and thus, can be used to simulate such movements that naturally occur upon timescales that cannot be reached via MD simulations. With LRT, it can be investigated whether substrate binding to the active site triggers characteristic conformational changes, like e.g. an induced fit. The observations can afterwards by interpreted in combination with the results from the MD simulations, to find out, how the activity of *Fs*C may be influenced by EG in the solvent. To validate the significance of the obtained results, an extension of the LRT model with random perturbations similar as in a previous study [104], is applied. With this reference model (LRT null model), which is precisely introduced in Chapter 2, it can be statistically assessed which movements are actually caused by substrate binding and do not belong to the normal modes of the protein. Based on this mix of multi-scale simulations possible mutations in *Fs*C that lead to an improved activity will be pointed out.

### 1.3.3 *Adaption of Linear Response Theory and the Null Model for HCN Channels*

For understanding the gating process of HCN channels, it is necessary to investigate the movements in the transmembrane parts of the channels that occur during channel opening. Due to the limited structural information of these parts, these movements have to be simulated based on the closed conformation of the HCN channel structure. While it would generally be possible to use MD simulations, we don't have the ressources to reach the neccessary timescales for investigating all relevant movements in the channel protein that are involved in the gating process. First, the simulation system with a tetrameric ion channel embedded into a lipid membrane and salt solution on both sides of the membrane is very large. Second, simulating voltage-dependent gating of the HCN channel requires the simulation of applied voltages on the membrane. Fortunately, this is implemented in the MD simulation software GROMACS [1, 15, 19, 79, 116, 147, 207], but multiple simulations with different voltages would have to run. Even reducing the accuarcy of the simulation system in terms of coarse-graining would not help to reduce the required compuational time effectually. Instead, it implies further problems, because the tiny filter of the ion channel, that is build by individual atoms of the filter residues, cannot be represented by a coarse-grained approach, where groups of atoms are reduced to abstract shapes. Backmapping from coarse-grained systems to all-atom systems does also not work very well for ion channels, as the mentioned atoms of the tiny filter region are not correctly oriented during the backmapping process [14]. Furthermore, coarse-grained force fields are not parameterized for simulating membrane voltages. Even if these problems could be

circumvented or a functional filter region could be forgone, the optimization of MD simulations for the present research challenge is a huge research project itself and is definitely time-consuming.

Hence, within the scope of this thesis, the conformational changes during HCN channel gating are studied by applying the same coarse-grained LRT model as for the above research project. Based on the known movements in the CNBD of HCN1 after cAMP binding, the LRT model is used to simulate the mechanically induced movements in the rest of the protein. In contrast to a previous study [219], where LRT was applied to study the release of cAMP from the CNBD of an *in silico* model of HCN4, the present study is built on the actual structure of HCN1 [112], that significantly differs in the tertiary arrangement of the subunits. Furthermore, cAMP binding is not simulated directly in the CNBD, where six external forces had to be applied to mimic the interactions between cAMP and the binding site. Instead, the reference model of LRT with random perturbations is used to identify which external force applied on a single central position in HCN1 best simulates cAMP binding. Therefore, the simulation results have to be compared with the known conformational changes in the CNBD. Furthermore, the simulation results must be able to explain or verify experimental data on diverse positions in the HCN channel. After having identified and verified the necessary external force acting on a central position, the induced conformational changes in the TMPC, especially in the S4 domain of the VSD and the S6 domain in the central pore, are analyzed. For this purpose, the extended LRT model, that has been developed for a monomeric protein, consequently, has to be adopted for the use in rotationally symmetrical homo tetrameric proteins, like HCN channels. Although the pursued simulations do not simulate changes in the membrane voltage, that are responsible of the gating movements, they identify the mechanical connections within the protein. This gives important insights into which movements are realistic and possible when changes in the membrane voltage occur.

### 1.3.4 *Mix of Patch-Clamp Measurements of Vpu Proteins and Computation of Shannon's Entropy in an MSA of Vpu Proteins*

For investigating the evolutionary conservation of the ion channel activity of the Vpu protein, a mix of electrophysiological measurements and computation of Shannon's entropy [192] in an MSA of homologous Vpu proteins is used. The electrophysiological patch-clamp measurements in HEK293T cells, performed by Timo Greiner (Membrane Biophysics Group, TU Darmstadt), are necessary to prove the ion channel activity of Vpu proteins of different HIV and SIV strains. The computation of Shannon's entropy points out which residues are highly conserved throughout evolution.

## 1.4 THESIS OUTLINE

The thesis is structured as follows:

CHAPTER 2 presents the rational design approach for the optimization of *Fs*C toward the activity of PET with a mix of multiscale simulations as published in [66]. In this chapter, the applied reference model of LRT (LRT null model) which is similar to that of a previous study [104] and plays an essential role in the context of this thesis, is introduced. Furthermore, it is shown how the LRT null model is able to prove the significance of obtained results by LRT simulations and how the robustness of the LRT null model can be proved. As a base for further LRT simulations in this thesis, the influences of different cutoff radii for connected residues in the coarse-grained model are analyzed and compared.

CHAPTER 3 presents the study on HCN channel gating as published in [67]. The LRT null model, that was used in Chapter 2, is adopted for a rotationally symmetrical homo tetrameric protein, so that it can be applied on the HCN1 channel. In this context, the LRT null model is used to identify the perturbation direction, that leads to realistic movements via LRT, when applied to a central position in the protein. This is proved by known conformational changes and experimental data. Afterwards, conformational changes in other domains are predicted, that contribute to a mechanistical model.

CHAPTER 4 presents the study on the Vpu protein, as published in [64]. In contrast to the methodology of the studies in the two previous chapters, here, the procedure relies on the sequence information of homologous Vpu proteins in an MSA to study the evoluationary conservation of the ion channel function of this Vpu protein.

CHAPTER 5 summarizes the results of this thesis and recapitulates the research goals that have been addressed by the work at hand. Subsequently, the main contributions of this thesis are pointed out followed by a discussion and an outlook into possible future work.

# 2

# RATIONAL DESIGN OF FSC

**This chapter has been published in:**

Christine Groß, Kay Hamacher, Katja Schmitz and Sven Jager (2017). Cleavage Product Accumulation Decreases the Activity of Cutinase during PET Hydrolysis, J. Chem. Inf. Model, 57(2):243-255. doi: 10.1021/acs.jcim.6b00556



**Contributions:**
The initial concept of studying the influences of degradation products on the activity of cutinase was given by Sven Jager, the senior author of this paper. Together, we further specified the concept of this study and defined reasonable evaluation methods to conceive this paper. My focus thereby lay on studying the mechanical connections within the protein via coarse-grained elastic network models. Therefore, I implemented the method of Linear Response Theory upon ligand binding in `R`. This method, introduced by Ikeguchi et al. [84], computes conformational changes in proteins after applying an external force. For our study, I enhanced the LRT method by further statistics in a reference model (null model) with which we can simulate randomized external stimuli at the same point of perturbation. I proved the reference model for robustness by a 1000-fold repetition of the clustering approach and evaluated the optimal spatial threshold for connected residues in the elastic network model to sufficiently account for the short as well as longrange interactions within the protein.
In this context, I was responsible for the content and preparation of Figures 7-10 in the paper and Figures S1-S2 (here:A.1-A.2) and S9-S11 (here: A.9-A.11) in the Supplementary Material. Figures 1 and 4 were prepared together with Sven Jager. All results were analyzed and discussed together and the paper was written together with support from all co-authors. To make the applied methods accessible to the community, we included the scripts of the implemented LRT method and its expansion into an `R` Bibliography, named `LRTNullModel`, and published it together with our study.

To retain consistency throughout the whole thesis, changes of order and renaming of section titles compared to the published article may occur.

## 2.1 ABSTRACT

The *Fusarium solani* cutinase (*Fs*C) is a promising candidate for the enzymatic degradation of the synthetic polyester polyethylene terephthalate (PET), but still suffers from a lack of activity. Using atomic MD simulations with different concentrations of cleavage product ethylene glycol (EG), we show influences of EG on the dynamic of *Fs*C. We observed accumulation of EG in the active site region reducing the local flexibility of *Fs*C. Furthermore, we used a coarse-grained mechanical model to investigate whether substrate binding in the active site causes an induced fit. We observed this supposed induced fit or "breath-like" movement during substrate binding indicating that the active site has to be flexible for substrate conversion. This guides rational design: mutants with an increased flexibility near the active site should be considered to compensate the solvent-mediated reduction in activity.

## 2.2 INTRODUCTION

To reduce the worldwide increasing environmental pollution by plastic waste, new methods to convert polymers back into monomers are needed. For the degradation of synthetic polymers, like polyethylene terephthalate (PET) or polyamide (PA), enzymatic degradation is to be favored over chemical or mechanical methods that suffer from the use of environmentally harmful chemicals or high energy costs [58]. In this context, the use of hydrolytic enzymes called cutinases, which are secreted by plant pathogenic fungi or bacteria, is a quite promising approach [31, 78, 122, 166]. Due to their ability to degrade the natural high-molecular weight polyester cutin, the main component of the plant cuticle, some cutinases are also able to degrade synthetic polyesters [31, 122]. Cutinases have been subject to numerous activity and mutation studies regarding the degradation of several synthetic polymers [31, 78, 122, 166].

PET, the synthetic polymer most commonly used worldwide, is a main target of enzymatic polymer degradation studies. PET waste causes environmental damage worldwide, while the overall PET production steadily increases [137]. The great interest to find a sustainable solution for PET waste treatment is underlined by the growing number of publications regarding this topic during the last decade. In Google Scholar we obtain 965 hits for the combined keywords "polyethylene terephthalate" and "cutinase" for the last decade, 683 of them for the last five years. A number of comprehensive studies have been undertaken for the *Fusarium solani* Cutinase (*Fs*C; EC: 3.1.1.74) by combining experimental studies on enzyme kinetics with experimental and computational studies on structure and dynamics [31] – especially since the molecular structure of the *Fs*C has been solved at a resolution of 1 Å [123]. The good quality of the X-ray structure allowed for the detailed analysis of the time dependent behavior of *Fs*C via molecular dynamics (MD) simulations, at timescales up to 15 ns [133]. As opposed to other cutinases requiring extreme conditions, the catalytic optimum of *Fs*C lies at 40 °C, which makes it the ideal candidate for an environmentally sustainable process.

PET hydrolysis leads to oligomeric fragments as well as to monomeric terephthalic acid (TPA) and ethylene glycol (EG). *Fs*C is able to catalyze this process, but its hydrolysis rate is quite low for the wildtype and converges to zero after a period of 24 to 96 h so that it only achieves a total weight loss of PET film of 5% [166]. Previous mutation studies only considered structure or shape guided design, to enlarge the active site. In these mutants large residues were replaced by smaller ones as reviewed by Chen et al. [31] and references cited therein. This resulted in activity enhancement for high-molecular weight polyesters but not for low-molecular monoesters. While structure guided design is based on the structure of the protein, a rational design approach uses additional simulations, modelling, or statistics to predict a promising mutant. Furthermore, the issue of activity loss over time has not been addressed so far. It is important to fully understand the limitations of wildtype *Fs*C and to carve out a clear hypothesis about the requirements to the mutants.

Our study focuses the low and decreasing activity of wildtype *Fs*C during PET hydrolysis. We assume that the increasing amount of the cleavage products plays a key role in this context. The small polar water analogue EG appears to be a reasonable candidate as it increases the viscosity and density of the solvent and may alter the hydration of the protein [178, 195, 205]. TPA is unlikely because it is not soluble in water so that its concentration in the solution is negligible. The effect of EG monomers on the structure and dynamics of proteins has not yet been sufficiently studied. Thus, in our study we investigated the effect of increasing concentrations of EG on the structure and the dynamics of *Fs*C at a molecular level by combination of multiscale simulations.

In the first part of our study, we used all-atom MD simulations to analyze the influence of different concentrations of EG on the enzyme dynamics. We chose MD to study the allosteric effects of solvent molecules as a state-of-the art method to determine protein dynamics and its solvent interactions [135]. In the second part, we used a coarse-grained model to investigate possible conformational changes of *Fs*C upon binding of a high-molecular weight polyester within the active site. Coarse-graining matches experimental data of small proteins or RNA structures (e.g. thermodynamics of the bovine trypsine inhibitor) up to huge biological complexes (e.g. assembly of the ribosomal subunits) [70, 72, 86]. Furthermore, it overcomes the limitations of MD simulations regarding the required timescale, which means that less computational effort is needed to simulate larger timescales [135]. In particular, we used the linear response theory [84] (LRT) to simulate the substrate binding in the active site by an external force vector to investigate possible structural changes. Based on our findings, we point out changes in the protein structure that may lead to improved enzyme activity.

## 2.3 MATERIALS AND METHODS

### 2.3.1 *Molecular Dynamics (MD) Simulations*

MD simulations were performed using the native *Fs*C structure (PDB-Code: 1CEX) with a resolution of 1 Å [123]. The simulation box with dimensions of $x = 59.90$ Å, $y = 57.16$ Å, and $z = 66.83$ Å was filled with TIP3P water molecules and varying amounts of EG up to final concentrations of 0% (0 molecules), 2% (41 molecules), 3% (61 molecules), 5% (103 molecules), 10% (210 molecules), and 20% (420 molecules). To neutralize the simulation box, 9 $Na^+$ and 12 $Cl^-$ ions were added to a final concentration of 0.9%. The simulations were performed with the Yasara software suite [107] and the AMBER03 force field [46] at constant temperature of 313 K, constant pressure of 1 bar, and constant pH of 7.4.

We used a van der Waals cutoff of 10 Å. Long range Coulomb interactions were calculated using the particle mesh Ewald algorithm. Grid points for the PME evaluation were evenly spaced in each dimension (27 grid points). For temperature control we used a velocity rescale thermostat which keeps the time average macroscopic temperature at the requested value by rescaling the atom velocities using a Berendsen thermostat [18]. For pressure control we chose the Manometer barostat in Yasara [107].

Prior to the simulations, the simulation box including the *Fs*C structure was filled with the defined number of EG molecules, then filled with water, and at the end with counter ions. Possible clashes were removed via energy minimization using the steepest descent algorithm with subsequent simulated annealing until convergence, i.e. energy improvement of less than 0.01 kJ/mol per atom over 200 steps. After 500 ps of solvent equilibration the simulation with 2 fs time-steps was run for an overall simulation time of 100 ns.

During the simulation we did not use rototranslational constraints, but prior to trajectory analysis we preprocessed the trajectory files using `trjconv` function in gromacs [159] in order to correct drift and rotation of the proteins from their initial positions in the simulation boxes.

#### 2.3.1.1 *Parametrization for EG*

For the parametrization of EG we used GAFF (General AMBER Force Field) [214] atom types and force field parameters followed by a calculation of semi-empirical AM1 Mulliken point charges [190] and a geometry optimization with the COSMO solvation model [103]. Furthermore we improved the AM1 charges for EG with the "AM1 Bond Charge Correction" [87]. This parametrization procedure is carried out by Yasara [107].

### 2.3.2 *Tetrahedral Order Parameter*

To account for the ability of water to form hydrogen bonds with adjacent water molecules and, thus, to establish a tetrahedral network, the tetrahedral order parameter is defined as follows [49, 109, 125]:

$$Q_i = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left[ \cos(\psi_{jik}) + \frac{1}{3} \right]^2 . \tag{2.1}$$

The index $i$ denotes the considered oxygen atom and $j, k$ the nearest oxygen neighbours (not necessarily hydrogen bonded to the local atom). The time average $\left\langle \frac{1}{N} \sum_i Q_i \right\rangle$ of a system with $N$ water molecules is 0 for random configurations and 1 for perfect tetrahedral orientation of all molecules. It can range form -3 to 1.

### 2.3.3 *Mean Square Displacement*

The mean square displacement (MSD) is a measure for the quantification of the dynamics of molecules [102].

$$r^2(\tau) = \frac{1}{N} \sum_{i=1}^{N} [\vec{r}_i(\tau) - \vec{r}_i(0)]^2 \tag{2.2}$$

$\vec{r}_i(\tau)$ is the current position of particle $i$ at timestep $\tau$. $r^2(\tau)$ is the MSD at timestep $\tau$ for a system of $N$ particles compared to the inital positions $\vec{r}_i(0)$. We used this measure to indirectly determine the viscosity of the solvent, as high viscosity correlates with reduced dynamics and vice versa. This antiproportional relation is given by the diffusion coefficient in the approximation of a sphere $D = \frac{k_B \cdot T}{6 \cdot \pi \cdot \eta \cdot R_0}$ according to the Einstein-Stokes equation

$$r^2(\tau) = 2n \cdot D \cdot \tau = \frac{n \cdot k_B \cdot T \cdot \tau}{3 \cdot \pi \cdot \eta \cdot R_0} \tag{2.3}$$

with the Boltzmann constant $k_B$, temperature $T$, timestep $\tau$, viscosity $\eta$ and particle radius $R_0$ in a $n$-dimensional system. For our MD simulations we computed the MSD of the waters' oxygen atoms in order to indirectly quantify the solvent viscosity.

### 2.3.4 *Root Mean Square Deviation*

In order to quantify structural differences and conformational changes of the overall protein structure we used the root mean square deviation (RMSD). The RMSD can be computed as

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [\vec{r}_i(a) - \vec{r}_i(b)]^2} \tag{2.4}$$

where $N$ is the number of atoms and $\vec{r}_i(a)$ is defined as the coordinates of atom $i$ in conformation $a$. The coordinates of all atoms in one conformation are compared

to the coordinates of all atoms in another conformation. We used all C$\alpha$ atoms and `trjconv fit rot+trans` as well as pbc corrections using `trjconv` function in gromacs [159] to process the trajectory. For the RMSD computation we compared the initial structure of the production run to all remaining frames. For curve smoothing we used the locally weighted scatterplot smoothing (LOESS) algorithm [33].

### 2.3.5 *Root Mean Square Fluctuation*

The residual flexibility was analyzed by the root mean square fluctuation (RMSF). It represents the average movement per residue during the simulation of $T$ frames by measuring the deviation of the coordinates $\vec{r}_i(t)$ of a specific residue $i$ from the average coordinates $\widetilde{\vec{r}}_i$.

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}[\vec{r}_i(t) - \widetilde{\vec{r}}_i]^2} \tag{2.5}$$

### 2.3.6 *RDF: Radial Distribution Function*

The radial distribution function [226] $g(\vec{r})$ gives the probability of finding a particle $\vec{R}_j$ within a spherical shell of radius $\vec{r}$ from another particle $\vec{R}_i$ within an infinitesimal thickness with $N$ particles in a system of volume $V$.

$$g(\vec{r}) = \frac{V}{N^2} \cdot \sum_{i \neq j} \delta\big(\vec{r} - (\vec{R}_i - \vec{R}_j)\big) \tag{2.6}$$

For our analysis, we used the protein C$\alpha$ atoms and computed the radial distribution of EG O atoms within radii from 2 to 30 Å and a thickness of 2 Å.

### 2.3.7 *Surface Density Calculations*

In our study we used $\mathcal{T} = 40,000$ frames for each simulation. We denoted the number of amino acids as $\mathcal{Q}$ and of EG mass centers as $\mathcal{E}$. We defined $V_{t,i}$ as the volume between amino acid $i$ (represented by its C$\alpha$ atom) and the maximum interaction distance $d$ to the solvent normalized by its Solvent Accessible Surface Area (SASA) at frame $t$

$$V_{t,i} = SASA_{t,i} \cdot d \tag{2.7}$$

The SASA computations were carried out using the parameter optimized surface calculator (POPS) [30]. For each simulation frame, a SASA, at amino acid level, as well as for the whole protein, was calculated. This resulted in a time series of SASAs for each amino acid. Local densities of solvent molecules could be quantified as the number of particles $\rho_i$ inside a volume fraction $V_{i,t}$.

To define the number of molecules occupying $V_{i,t}$ in a trajectory, we calculated the number of particles within the volume of every backbone amino acid $i$ for every frame $t$. This yielded to the accumulation tensor $M$ (Eq. 2.8) with the axis defined by simulation frame $t$, the backbone amino acid $a_i$, and the EG mass center $e_j$. We denoted $\|\vec{a}_{it} - \vec{e}_{jt}\|$ as the Euclidean norm of the C$\alpha$ coordinates ($\vec{a}_i$) and the mass center coordinates of an EG molecule ($\vec{e}_j$) at frame $t$. The entries of an accumulation tensor $M$ are defined as

$$M_{ijt} = \begin{cases} 0 & \text{if } \|\vec{a}_{it} - \vec{e}_{jt}\| > d \\ 1 & \text{if } \|\vec{a}_{it} - \vec{e}_{jt}\| \leq d, \end{cases} \tag{2.8}$$

with $d$ set to 7 Å, which corresponds to the first coordination shell of an amino acid. Now we could approximate the average local EG density $\rho_i$ at amino acid $i$ over a whole simulation by

$$\rho_i = \frac{1}{\mathcal{T}} \sum_{t}^{\mathcal{T}} \sum_{j}^{\mathcal{E}} M_{t,i,j} \cdot V_{t,i}^{-1}. \tag{2.9}$$

Furthermore we defined the time dependent density ($\rho_t$) for the complete protein by

$$\rho_t = \frac{1}{\mathcal{Q} \cdot \mathcal{E}} \sum_{i}^{\mathcal{Q}} \sum_{j}^{\mathcal{E}} M_{t,i,j} \cdot V_{t,i}^{-1}. \tag{2.10}$$

For density comparisons the mean density over a defined period of time was used

$$\tilde{\rho} = \frac{1}{\mathcal{T}} \sum_{t}^{\mathcal{T}} \rho_t. \tag{2.11}$$

### 2.3.8 *Linear Response Theory (LRT)*

The linear response theory (LRT) introduced by Ikeguchi et al. [84] is a model to predict the structural changes of a protein upon ligand binding. It is based on the normal mode analysis (NMA), which is a well suited method to study the collective motions in proteins [197]. To reduce computational effort, this method can also be applied on coarse-grained structures where proteins are reduced to a network of beads and springs. The beads represent the amino acids of the protein and the springs represent bonded or nonbonded interactions between several amino acids when their spatial distances fulfill a given cutoff criteria. Such a network is called an elastic network model (ENM) [8].

Instead of treating the fluctuations as isoptropic, like a Gaussian network model (GNM) does, the LRT is based on an anisotropic network model (ANM) that considers anisotropic fluctuations of amino acids [202]. It could be shown that combinations of low frequency modes correspond to the protein structural changes upon ligand binding. Hence, using the LRT the direction of a structural change of a protein can be predicted via the formula [84]

$$\Delta\vec{R}_i \simeq \beta \cdot \sum_j \langle \Delta\vec{R}_i \cdot \Delta\vec{R}_j \rangle_0 \cdot \vec{f}_j \tag{2.12}$$

where $\Delta\vec{R}_i$ represents the predicted translocation of atom $i$ after the perturbation and $\langle \Delta\vec{R}_i \cdot \Delta\vec{R}_j \rangle$ denotes the covariance matrix of atomic fluctuations in the ligand free state. $\vec{f}_j$ represents the external force vector mimicking ligand binding and $\beta$ is $1/k_B T$ with the Boltzmann factor $k_B$. For the computation of the coordinate changes, the covariance matrix derived from an ANM or extracted from a MD simulation can be used. The covariance matrix can be computed as the Moore-Penrose pseudoinverse [140, 151] of the $3N \times 3N$ Hessian Matrix $\mathscr{H}$ that describes the second derivatives of the harmonic potential of the ANM with N residues [8]

$$\mathscr{H} = \begin{pmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} & \dots & \mathcal{H}_{1N} \\ \mathcal{H}_{21} & \mathcal{H}_{22} & \dots & \mathcal{H}_{2N} \\ \vdots & & & \\ \cdot & & & \\ \mathcal{H}_{N1} & \mathcal{H}_{N2} & \dots & \mathcal{H}_{NN} \end{pmatrix} \tag{2.13}$$

with a $3 \times 3$ super element

$$\mathcal{H}_{ij} = \begin{pmatrix} \frac{\partial^2 V}{\partial x_i \partial x_j} & \frac{\partial^2 V}{\partial x_i \partial y_j} & \frac{\partial^2 V}{\partial x_i \partial z_j} \\ \frac{\partial^2 V}{\partial y_i \partial x_j} & \frac{\partial^2 V}{\partial y_i \partial y_j} & \frac{\partial^2 V}{\partial y_i \partial z_j} \\ \frac{\partial^2 V}{\partial z_i \partial x_j} & \frac{\partial^2 V}{\partial z_i \partial y_j} & \frac{\partial^2 V}{\partial z_i \partial z_j} \end{pmatrix}. \tag{2.14}$$

Regarding *Fs*C, we used this model to predict the structural changes (induced fit) during substrate binding to the active site serine (S120). We obtained the covariance matrix from a heterogeneously parameterized ANM using the energy minimized structure of *Fs*C with spatial cutoffs of 7 and 13Å, respectively for connected residues (see Figure 2.1). We used the matrix for intra-chain interactions between amino acids by Miyazawa and Jernigan [138] as well as the matrix for inter-chain interactions of amino acids by Keskin et al. [96] provided by the `R` [161] package `BioPhysConnectoR` [81].

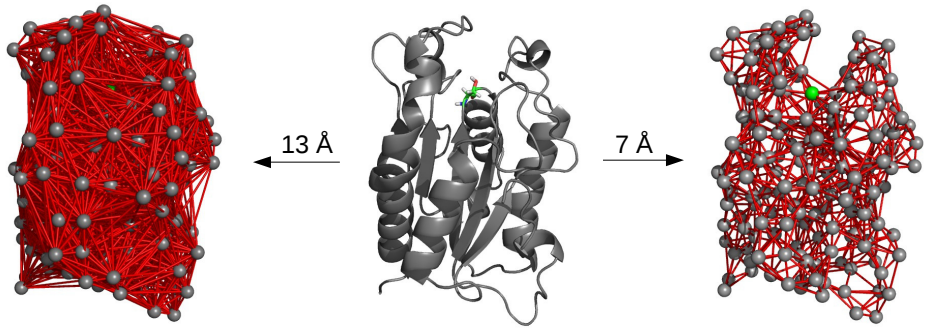


Figure 2.1: Illustration of the elastic network models with distance cutoffs of 7 and 13Å for connected residues in comparison with the all-atom model of *Fs*C in cartoon representation. The C$\alpha$ atoms are reduced to gray spheres and the connections between atoms are shown as red lines. The catalytic S120 is highlighted in green.

To mimic the substrate binding, the external force vector was directed towards S120 from a possible position for the substrate's carbonyl carbon upon formation of the tetrahedral intermediate. This position was randomly chosen from a cluster of accessible positions in the binding pocket. Note that no substrate or solvent is present in the ANM setup.

### 2.3.9 *LRT Null Model*

The implementation of a null model has proven beneficial to study the statistical significance of a computational approach [217]. To investigate the influences of the direction of the force vector as well as the significance of the above chosen force direction, we used a reference model of isotropic perturbation with 1000 force vectors randomly originating from different points on a sphere around S120, similar to a previous study [104]. We chose spherical coordinates $\phi$ and $\theta$ uniformly distributed with $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi]$ to generate 1000 different force vectors. The force vectors $\vec{f_j} = (x_j, y_j, z_j)$ were parameterized as follows:

$$f_{xj} = f_0 \cdot \sin(\theta) \cdot \cos(\phi) \tag{2.15}$$
$$f_{yj} = f_0 \cdot \sin(\theta) \cdot \sin(\phi) \tag{2.16}$$
$$f_{zj} = f_0 \cdot \cos(\theta) \tag{2.17}$$

with $f_0$ being an arbitrary scaling factor that eventually has no impact on our subsequent results. The induced fit of the enzyme substrate complex was demonstrated by perturbing S120 with 1000 different external force vectors $\vec{f_j}$ with repulsive and attractive forces in comparison.

To check for clustering, we clustered the displacement vectors of selected residues after perturbing S120 from each random direction. The selected residues for displacement calculation were residues 80-90, 179-187, and 42-45, as they are reported to participate in the functional behavior of *Fs*C [37]. Force directions were clustered by applying the `kmeans` algorithm from Hartigan and Wong [74] on the $1000 \times (3 \cdot 24)$ matrix of the $x$-, $y$-, and $z$-displacements of the C$\alpha$ atoms of the selected residues. The optimal number of clusters was investigated by comparing the log values of maximal within-cluster sum of squares (maximum withinss) from `kmeans` clustering as a function of number of clusters.

### 2.3.10 *Software Contribution*

`R` [161] is an environment for statistical analysis of data that offers many additional packages, especially for computational biology. We implemented the method of linear response upon substrate binding [84] in `R` using the `BioPhysConnectoR` [81] and `bio3d` [62] packages and enhanced the model by further statistics in our null model. We included both in the `LRTNullModel` package in `R` to make the applied methods accessible to the community.
Software link: `http://www.cbs.tu-darmstadt.de/LRTNullModel.tar.gz`

## 2.4 RESULTS

In order to investigate the influence of increasing concentrations of the cleavage product EG in the reaction solution on the activity of *Fs*C, we performed MD simulations with different EG concentrations (0%, 2%, 3%, 5%, 10%, and 20%) in the solvent. The trajectories were analyzed regarding two different aspects. First, we were interested in the influences of increasing amounts of EG on the overall dynamics of *Fs*C. Second, we analyzed the results with respect to local accumulations of EG on the surface of *Fs*C.

### 2.4.1 *Increasing EG Concentrations Reduce the Overall Dynamics of FsC*

As a measure for the overall dynamics of *Fs*C during the simulation, we compared the RMSD values of the different runs (Figure 2.2 A). It is noticeable that increasing concentrations of EG in the solvent reduce the overall dynamics of *Fs*C. Just the small change from 2% to 3% EG in the solvent causes a remarkable drop of the RMSD values, restricting movements to at least half of the range found in pure water. Furthermore, we analyzed the effect on the residual fluctuations of *Fs*C in terms of RMSF (Figure 2.2 B). With increasing concentrations of EG the RMSF also declines for all residues.

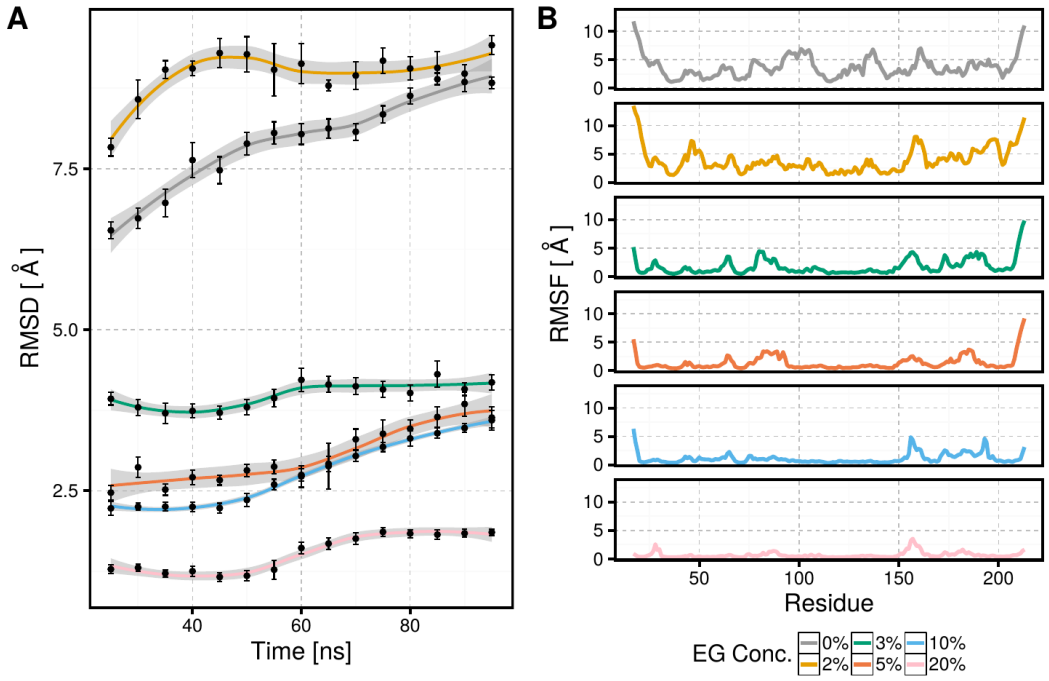


Figure 2.2: RMSD (in steps of 5 ns, that is, 2000 frames) and RMSF of the MD simulations with different concentrations of EG in comparison. (**A**) The RMSD curves were smoothed using the LOESS [33] algorithm. The gray area is the uncertainty of the fit, while the error bars represent the standard deviation of 2000 frames each. The higher the EG concentration, the lower the average RMSD (except for the simulation with 0%) with a remarkable drop in overall dynamics of *Fs*C from 2% to 3% of EG; (**B**) The residual flexibility also strongly reduces with increasing concentrations of EG.

### 2.4.2 *Accumulation of EG near the Active Site*

To find out if and where EG accumulates on the surface of *Fs*C, we performed surface density calculations based on the MD trajectories. Figure 2.3 A shows the mean occurrence of EG particles on the surface of *Fs*C for the trajectories with different EG concentrations. Furthermore, we distinguished between surface residues in the active site region and remaining surface residues. Residues with a SASA < 1 Å$^2$ were not considered as surface residues. A 12.4-fold higher slope of particles per volume with rising EG concentration was observed for the surface residues near the active site than for the remaining surface residues (Figure 2.3 B). This indicates, that EG accumulates near the active site with increasing concentrations of EG.

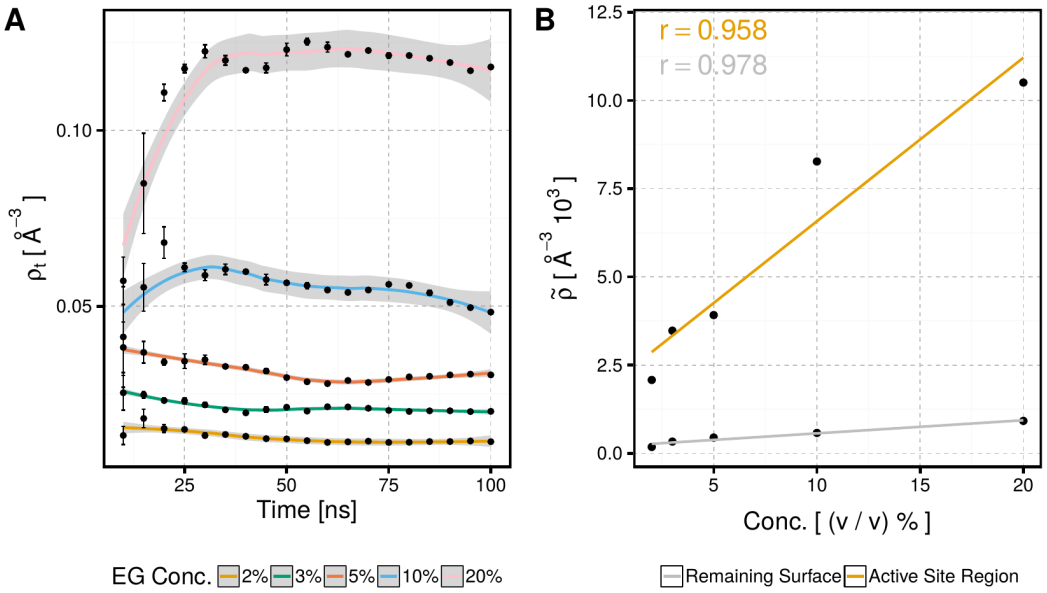


Figure 2.3: (**A**) Time-dependent densities of EG particles $\rho_t$ for the simulations with different concentrations of EG in the solvent. The curves were smoothed using the LOESS [33] algorithm. The gray area is the uncertainty of the fit, while the error bars represent the standard deviation of 2000 frames each; (**B**) Linear regression of mean densities $\tilde{\rho}$ of EG particles (60-100 ns) as a function of the concentration of EG in the solvent distinguished for surface residues near the active site (slope= 0.460, $p$-value= 0.0038) and the remaining surface residues (slope= 0.037, $p$-value= 0.04).

For visual analysis, the average EG densities on the *Fs*C surface are depicted in Figure 2.4 via color-coded surface representations. With increasing EG concentrations, we observed higher densities of EG on the surface of *Fs*C. Furthermore it is noticeable that for increasing concentrations EG accumulates near the active site, which agrees with the results taken of Figure 2.3 B. These observations give a first indication why the activity of *Fs*C decreases during the hydrolysis of PET as increasing amounts of the cleavage product EG are generated.

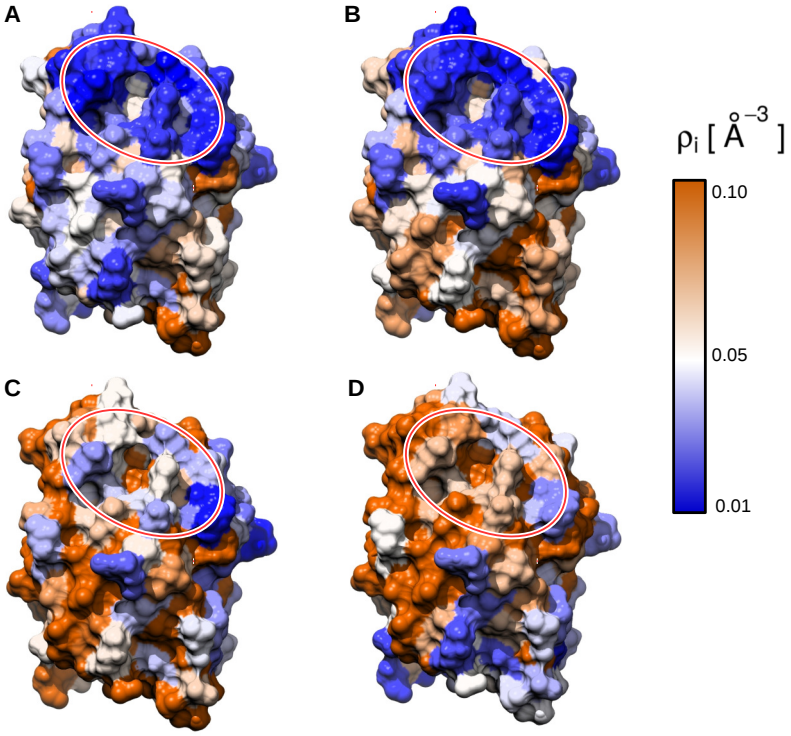


Figure 2.4: Accumulation of EG on the surface of *Fs*C. Blue denotes low densities of EG whereas red denotes high densities (in Particles/Å$^3$). From (**A**) to (**D**) increasing concentrations (3%, 5%, 10%, and 20%) of EG were used for the density calculations. With increasing concentrations we observe a movement of the regions of EG accumulation towards the active site (red ellipsoid).

### 2.4.3 *EG Accumulations Reduce the Local RMSF of Catalytic H188*

As EG accumulates near the active site, we focus on the catalytic triad and the oxyanion hole of *Fs*C (S120, D175, H188, Q121, S42, N84). We observed increasing EG densities on each catalytic residue when the EG concentrations of the solvents increased (Figure 2.5 A). The effects on the residual flexibility are pointed out by the local RMSF values (Figure 2.5 B). The accumulation of EG most effects the flexibility of H188 (drop from 5.5 to 0.5 Å), while for low EG concentrations H188 hardly encounters any EG at all.

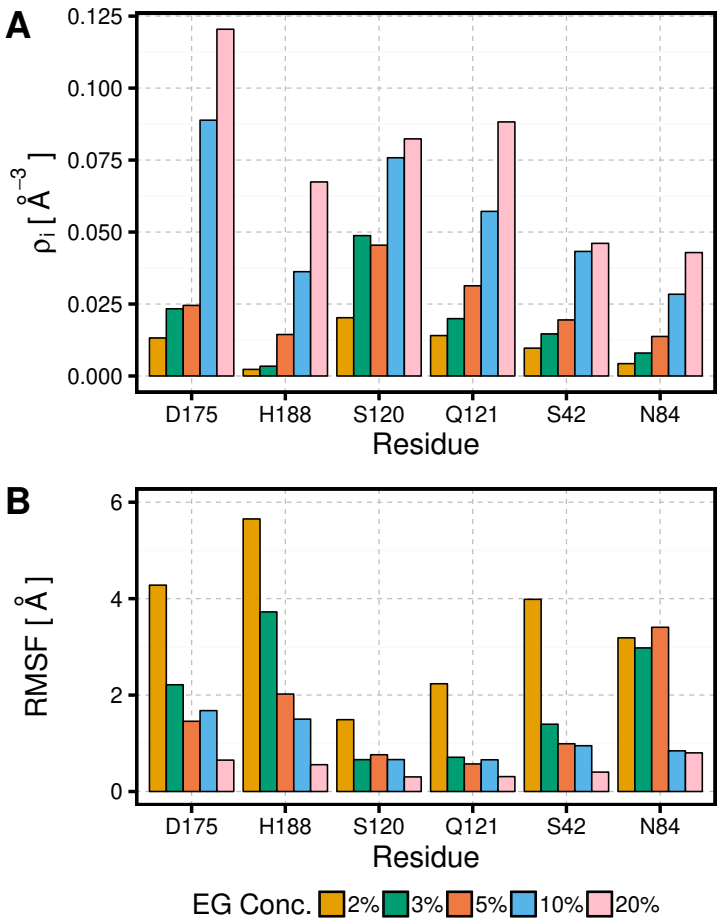


Figure 2.5: Densities of EG (**A**) and RMSF (**B**) for the residues in the active site and oxyanion hole based on the MD simulations of *Fs*C with different concentrations of EG in water.

### 2.4.4 *Characterization of the FsC Environment*

To elucidate the influence of EG to the solvent properties, we investigated the mean square displacement (MSD) of the water molecules as an indirect measure of the solvent viscosity as well as the tetrahedral order parameter of water molecules to study the effects on the hydrogen bonding network. The MSD with EG does not significantly differ from the MSD without EG which indicates that the viscosity of the solvent is marginally influenced by EG (Figure 2.6 A). Interestingly, the distribution of the tetrahedral oder parameter $Q_i$ is right shifted in the simulations with EG compared to the simulation without EG (Figure 2.6 B). The hydrogen bonding network is more structured, when EG is added to the solvent, no matter at which concentration.

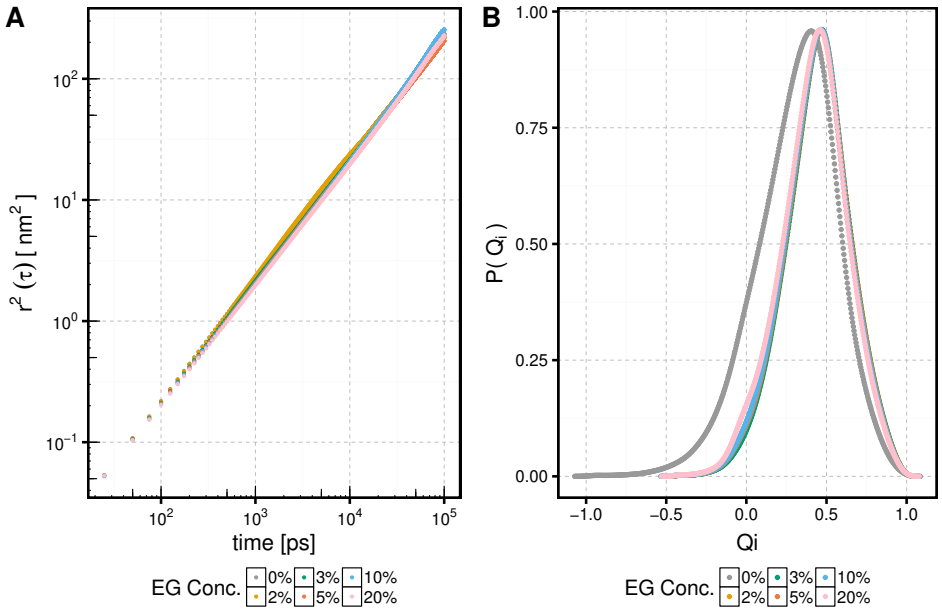


Figure 2.6: Characterization of the solvent for different concentrations of EG. (**A**) Double logarithmic representation of MSD of the water molecules as indirect measure of the viscosity. The constant increase of the MSD during the simulation describes normal diffusion. The MSD for the different EG concentrations does not significantly differ, which means that the viscosity of the solvent is not influenced by EG; (**B**) Tetrahedral order parameter $Q_i$ to quantify the ability of water molecules to form hydrogen bonds to adjacent water molecules. In all simulations with EG, the distribution of $Q_i$ values is right shifted compared to the simulation without EG. This indicates that with EG the hydrogen bonding network is in a more orderly state.

### 2.4.5 *LRT Reveals the Need for more Flexibility in the Active Site Region*

The results above show that with increasing amounts of EG in the solvent EG accumulates near the active site and decreases the flexibility of the catalytic residue H188 and causes a reduction in overall protein structural dynamics. Whether these observations are sufficient as explanations for the low activity of *Fs*C with progressing cleavage product release, had to be further evaluated. As QM/MM simulations [153] are prohibitively expensive for the simulation of enzymatic reactions in the bulk, a coarse-grained approach based on an ANM [8] was chosen. Using linear response theory (LRT), it is possible to investigate the structural response of proteins to a mechanical stimulus, e.g. due to ligand binding. We used this method to predict the structural change of *Fs*C upon substrate binding in the active site (formation of the tetrahedral intermediate) in order to investigate to what extent flexibility in the active site is actually required.

We computed the LRT model for distance cutoffs of 7 and 13 Å, respectively. The model with 7 Å only includes the connections with residues in the first coordination shell, whereas larger cutoffs (e.g. 12-15 Å) better reproduce experimental observations [8]. For the ENM with a cutoff distance of 13 Å for connected residues, Figures 2.7 A and C show the structural response of *Fs*C upon mechanical perturbation at S120 (gray sphere) with increasing forces. Attractive force vectors with increasing forces were used to simulate the release of the cleavage product (Figure 2.7 B and D).

It is remarkable that for both cases mainly the loop regions near the active site are affected, while the $\alpha/\beta$-core remains stable, although S120 is located at the interface of a $\beta$-sheet to an $\alpha$-helix within the $\alpha/\beta$-core. In case of the repulsive forces (substrate binding) the loops around the active site move closer to each other, leading to a closing of the binding pocket. In contrast to that, the attractive forces cause movements of the loops away from each other corresponding to an opening of the binding pocket. These observations confirm the hypothesis of a "breath-like" movement of *Fs*C during the hydrolysis reaction proposed by Longhi et al. [123].

### 2.4.6 *Analysis of LRT Results Based on a Null Model*

To prove the significance of the above results with a randomly chosen force direction, we used a reference model of isotropic perturbation. We clustered the 1000 different force directions regarding their resulting displacements of selected residues after perturbation of S120 with the corresponding force vector (for more information see "Materials and Methods" section). The development of the log values of maximal cluster sum of squares (maximum withinss) after clustering as a function of the number of clusters using the `kmeans` algorithm (Figure A.1) convergences within ten iterations and has the most obvious drop from three to four clusters. Therefore, the use of four clusters is plausible.

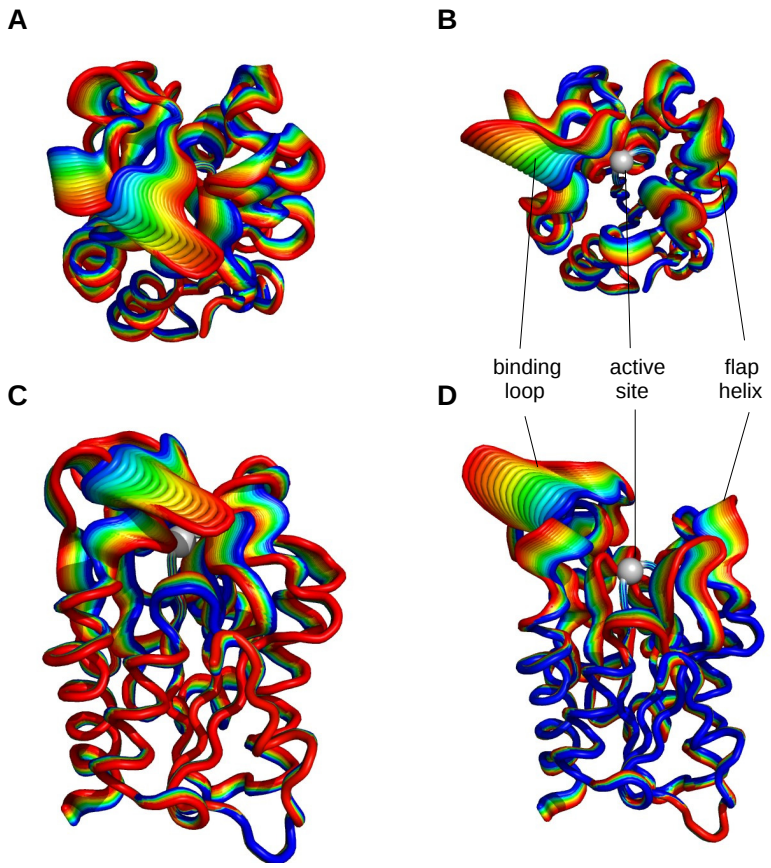


Figure 2.7: LRT model with repulsive (**A**) + (**C**) and attractive (**B**) + (**D**) force vectors with different forces between 0 (blue) and 3000 (red) in arbitrary units from top and side view (right angle to each other). The repulsive forces represent substrate binding and formation of the tetrahedral intermediate whereas the attractive forces represent the release of the cleaved substrate. The observed motions strengthen the idea of a "breath-like" movement as reported by Longhi et al. [123].

The result of the clustering is shown in Figure 2.8. The 1000 different force directions represented as 1000 different end points on a sphere around S120 (C$\alpha$ atom) are colored according to their assignment to a cluster. All force directions of a cluster result in the same conformational change of the functionally relevant regions of *Fs*C.

Interestingly, the force directions belonging to one cluster are located in four different spatial areas. This means that functional clustering is linked with the spatial distribution of the force directions. It is remarkable, that the possible force directions for the substrate to bind to S120 from a sterical side of view, exclusively belong to the green cluster. This proves the insensitivity of the obtained structural changes upon substrate binding in our LRT model towards small inaccuracies in the structural mechanical model.

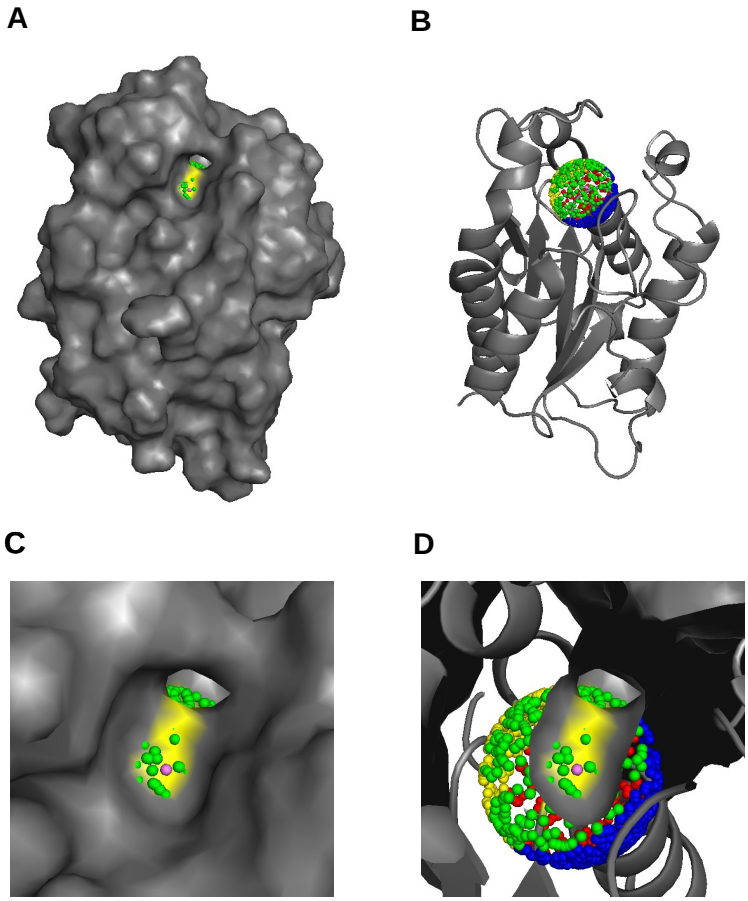


Figure 2.8: LRT null model with 1000 force directions for the external force vector represented as small spheres colored according to the cluster. In this illustration, the spheres have a distance of 3.9 Å to the C$\alpha$ atom of S120 in order to visualize possible positions for the substrate on the surface of the binding site. The realistic force directions from which the substrate can perturb the S120 during the catalytic intermediate, all belong to one cluster (green). The randomly selected force direction within the group of realistic ones, which was used for the LRT model in the previous section, is highlighted in violet. (**A**) and (**C**) Surface representation; (**B**) Cartoon representation with a sphere of 1000 force directions; (**D**) Zoom into the surface to see the non-realistic force directions of the other clusters.

To assess the robustness of the clusterings by our LRT reference model, we compared the probability to be in cluster $i$ by the Kullback-Leibler divergence [108]: $D_{KL}(P||Q) = \sum p_i \cdot ln(\frac{p_i}{q_i})$, where $p_i$ denotes for the probability to be in $i$ in one run of the LRT protocol and $q_i$ is the probability to be assigned to $i$ in an independent repetition. Our 999 000 comparisons showed a mean (maximum) $D_{KL}$ of 0.03 nat (0.14 nat). These rather small values suggest high robustness of the implied clusterings.

Figure 2.9 shows the representative movements of the two main loops near the active site after perturbing S120 in a schematic manner. Here, the results for repulsive forces are shown which mimic the binding of the substrate and formation of the tetrahedral intermediate. For attractive forces, the opposite displacements occur.

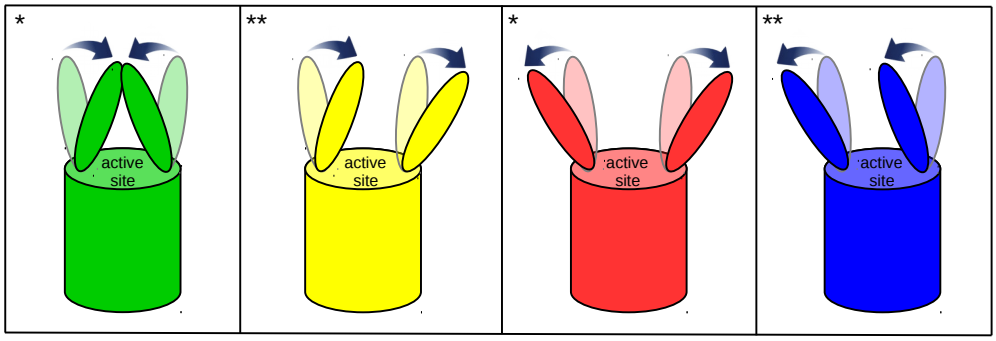


Figure 2.9: Schematic representation of the loop displacements near the active site after perturbing S120 with repulsive forces from other clusters. The small representations are colored regarding the respective clusters in Figure 2.8. Note: The parallel loop motions caused by perturbation from directions in the yellow and blue clusters correspond to normal modes of *Fs*C, while the breath-like motions caused by perturbation from possible directions in the active site pocket do not correspond to normal modes. They only occur by external perturbation due to substrate binding or cleavage.

To ensure that the observed breath-like motions are actually caused by substrate binding/cleavage, we compared them to the low frequency normal modes of *Fs*C. Low frequency normal modes (eigenvectors of the Hessian matrix) underlie equilibrium dynamics of a protein and represent intrinsically accessible motions (conformational changes) without external forces acting on the protein [10]. In fact, the normal modes only comprise motions of the loop regions near the active site that either move parallel to the left or parallel to the right (Figure A.2 and Movies M1 and M2), similar to those observed by perturbations from the yellow or blue cluster. Normal modes do not comprise breath-like motions that are similar to those caused by the perturbation of S120 from possible directions in the active site pocket.

## 2.5 SUMMARY AND DISCUSSION

The aim of our study was to investigate the reasons for the low and decreasing activity of wildtype *Fs*C during hydrolysis of PET in order to find means of improving enzyme activity. For this purpose, we focused on the influence of increasing concentrations of the cleavage product EG on (1) the overall dynamics of *Fs*C and (2) the accumulation of EG on the surface of *Fs*C via explicit all-atom MD simulations. We found, that increasing concentrations of EG result in reduced flexibility of *Fs*C caused by EG accumulation near the active site. The local flexibility of the catalytic H188 is most affected by this accumulation. With our simulations we can

confirm important residual fluctuations measured in NMR studies of *Fs*C [155, 157, 158] (Figure A.3). These NMR studies already pointed out the catalytic H188 to be highly flexible, which also goes along with our RMSF and surface density calculations, irrespective of the different timescales. As shown by Prompers et al. [157] H188 actually requires this flexibility to enable the enzymatic reactions.

The high RMSD values of the MD simulations with 0% and 2% EG in the solvent could lead to the assumption that unfolding events could have occured. However, structural analysis of our protein (Figure A.4) proves that our MD simulations are stable over time.

On the basis of the characterization of the protein environment during the MD simulations, we sum up that the hydrogen bonding network of the water molecules in the solvent is more orderly with EG than without. Interestingly, the dynamics of water molecules, which we analyzed in terms of MSD, were only marginally reduced. At first glance these observations do not match, but having a look at the radial distribution function of protein C$\alpha$ to EG O, we see that EG mainly accumulates near *Fs*C (Figure A.5 A), which fits well with our surface density computations. This means that there are not many EG molecules left in the bulk to reduce the viscosity significantly. To make sure that the reduced flexibility of *Fs*C is actually caused by the accumulation of EG and not only by solvent mediated effects caused by the hydroxy functionality of an alcohol in general, we additionally made MD simulations with the same simulation setup and force field but with 5% methanol (MeOH) or 5% ethanol (EtOH) respectively, instead of EG (Figures A.5-A.7). MeOH strongly increases water dynamics and, thus, strongly reduces the viscosity. Furthermore, it increases the protein RMSD as well as RMSF. Interestingly, it shows the same effect on the hydrogen bonding network of the water molecules as with 5% EG, but MeOH is mostly located in the bulk and does not accumulate near *Fs*C. For EtOH we observe hardly any effect compared to the simulation without any alcohol added to the solvent. It seems that, at least at this concentration, EtOH behaves like water in the solvent. We only see a slight accumulation of EtOH near *Fs*C. This suggests that few EtOH molecules alter the dynamics of the protein, which is evident from Figure A.5 B.

To account for the preferred type of interaction between EG and the active site residues, we determined the distribution of densities of EG near hydrophobic and hydrophilic surface residues in the active site and for all remaining surface residues in comparison. We found that EG densities are significantly higher for hydrophilic than for hydrophobic surface residues in the active site (Figure A.8). For the remaining surface, EG densities are quite similar for both types with a slight tendency towards hydrophobic residues. This shows that the accumulation of EG monomers within or near the active site is mainly based on hydrophilic interactions, that is, hydrogen-bonding interactions.

The importance of the flexibility of the region near the active site actually was also shown via the LRT model simulating the induced fit upon substrate binding. While all-atom simulations of enzymatic reactions are not possible with the available computational resources, our coarse-grained model is a valid method to investigate the required flexibility during the reaction and to demonstrate the induced fit. Contrary to previous X-ray studies [132] predicting a preformed oxyanion hole, Prompers et al. [157] observed an induced fit during NMR experiments, which was

our motivation for using the LRT method to investigate the mechanical principles behind the induced fit. We were able to demonstrate significant motions of the loop regions near the active site that do not correspond to the intrinsically accessible normal modes of *Fs*C and, thus, confirm the breath-like movement proposed by Longhi et al. [123].

The significance of our LRT results was further validated via a LRT null model. We found four main conformational changes of the loop regions near the active site. Interestingly, the random directions belonging to the same clusters are located in discrete spatial areas. By using this null model we were able to demonstrate that (1) our model is realistic and (2) that the presented results of the LRT model are reliable, as all physically possible force directions on the active site surface belong to the same cluster so that small perturbations would yield the same result in our modelling approach.

The observed movements of the loop regions near the active site appear straightforward for the LRT model based on the cutoff distances of 13 Å, as there are direct connections (springs) between S120 and the loop regions (Figure 2.1, left). However, our model with the smaller distance cutoff of 7 Å shows similar structural conformations of the loop regions near the active site (Figure A.9), although there is no direct connection (Figure 2.1, right). This demonstrates that the mechanism of mechanical force transfer is highly complex and forces can be transferred over a wide range of edges within the whole network. The corresponding null model is shown in Figure A.10. To account for long-range interactions Figure A.11 shows the displacements for 15, 17, and 21 Å cutoff, which do not differ in their direction, only in their magnitude. This demonstrates that the 13 Å model does not disregard possible deviating long-range interactions.

We suppose that bigger substrates correspond to higher forces acting on S120 during the formation of the tetrahedral intermediate. This means, that in the LRT model, the red conformations (high forces) correspond to the induced fit caused by binding of high-molecular weight polyesters, whereas the conformations near the blue one (low forces) correspond to the induced fit caused by low-molecular weight substrate binding. Therefore, our results underline the necessity for flexibility of the active site regions for larger substrates.

## 2.6 CONCLUSION AND OUTLOOK

In our study we found that (1) with increasing concentrations EG accumulates near the active site and reduces the overall flexibility of *Fs*C and (2) that the loop regions near the active site perform a breath-like motion during substrate binding and cleavage. From this, we conclude that increasing accumulation of EG negatively affects the activity of *Fs*C which is based on the following mechanism: The arrangement of residues of the active site and the residues stabilizing the tetrahedral intermediate via H-bonds within the oxyanion hole is crucial for the success of the catalytic mechanism (Figure 2.10). In order to enable the nucleophilic attack of S120 at the carbonyl carbon of the substrate's ester bond, the nucleophilicity of the hydroxy group of S120 has to be increased. This is achieved by the hydrogen bond network within the catalytic triad (D175, H188, S120) [31]. Due to the nucleophilic attack, a covalent bond between S120 and the substrate is formed and the proton

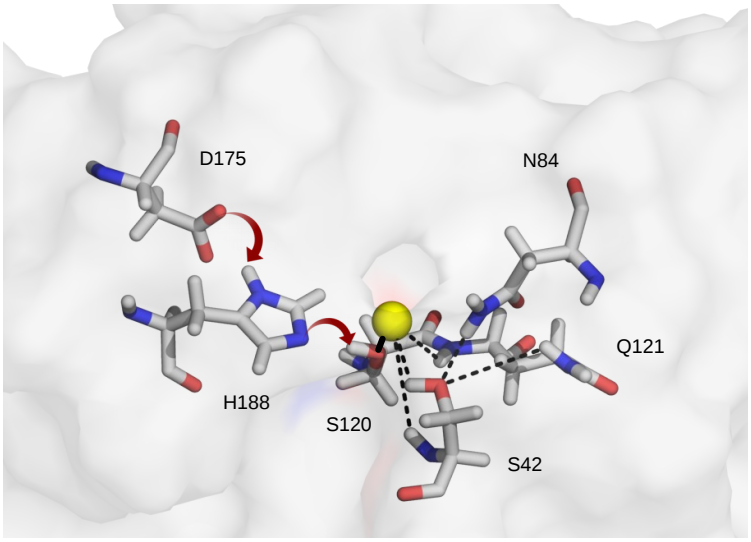


Figure 2.10: Active site (D175, H188, S120) and oxyanion hole (S42, Q121, N84) of *Fs*C as stick representation. Note that the residues are in close spatial proximity even though they are located at remote positions in the protein sequence. In the catalytic triad, red arrows demonstrate the hydrogen bond network, so that the more nucleophilic S120 can attack the substrate (carbonyl group of the ester bond simplified as yellow sphere). The covalent bond of the tetrahedral intermediate is shown as a black solid line. The stabilizing hydrogen bonds in the oxyanion hole are shown as black dashed lines. To illustrate the orientation within the protein its transparent surface is shown.

is transferred to the adjacent H188. Residues S42 and Q121 stabilize the negatively charged carbonyl oxygen of the tetrahedral intermediate via H-bonds, so that further steps of the catalyic mechanism are facilitated [31]. N84 and Q121 further stabilize the position of S42 [123]. During the loop movements, especially upon opening of the active site to accomodate large substrates, the respective residues are pulled appart from the perfect arrangement. A loss of flexibility of these residues impairs their instantaneous self-rearrangement to the positions required for the catalyic triad and oxyanion hole. This distortion is likely to reduce binding and substrate conversion rates. With further accumulation of EG and increased rigidity, even the loop motions may abate so that the binding of high-molecular substrates, like PET, is completely prevented.

One intuitive solution of the problem of EG accumulation would be the improvement of the industrial degradation process towards removing the cleavage products during the process. A further solution is to improve the enzyme properties by means of rational design in order to counteract this cleavage product mediated effect. For this purpose, mutants with increased flexibility near the active site should be favored. Furthermore, mutants with an increased hydrophobicity in the active site might be promising candidates as we found that the accumulation of EG in the active site is mainly based on hydrogen-bonding interactions. The initiation of EG accumulation could be prevented by reducing the number of possible hydrogen-bonding partners. As our LRT model demonstrates that the mechanical

correlations are rather complex, more sophisticated methods than structure guided design should be considered.

Our conceptual results have to be complemented by experiments investigating the enzymatic activity in terms of $K_m$ and $k_{cat}$, for example, colorimetric assays [28]. In these experiments standard 4-nitrophenyl esters (like pNPA, pNPB, pNPP, etc.) are commonly used to determine the activity of *Fs*C towards the hydrolysis of low-molecular esters but not necessarily towards the hydrolysis of PET. We propose to search for more complex model substrates for 4-nitrophenyl assays, that mimic PET more closely.

During the preparation of the current manuscript the discovery of bacteria that are supposed to be able to digest PET has been published [228]. The responsible enzymes may be quite interesting for enzymatic PET degradation, but a lot of basic research has to be done in order to reach the current state of knowledge that *Fs*C has in its community.

### ACKNOWLEDGEMENTS

C.G. and K.H. thank the LOEWE project iNAPO funded by the Ministry of Higher Education, Research and the Arts (HMWK) of the Hessen state for support. S.J. thanks by iGEM@TU- Darmstadt 2012-2015 for support.

### SUPPORTING INFORMATION

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.6b00556.

Figures S1-S11 depicting `kmeans` clustering, lowest-frequency normal mode, comparison with NMR data, secondary structure analysis, radial distribution functions, RMSD and RMSF using other alcohols, solvent characterization using other alcohols, distributions of EG densities, LRT model with 7 Å, LRT null model with 7 Å, LRT with long-range interactions (PDF)
Movies M1 and M2 depicting simulatenous motions of all ten lowest frequency normal modes using two different force fields (ZIP)

In this thesis, the Supplementary Figures can be found in Appendix A1 as Figures A.1-A.11.

# 3

# INVESTIGATION OF HCN CHANNEL GATING

**The following chapter has been published in:**

Christine Groß, Andrea Saponaro, Bina Santoro, Anna Moroni, Gerhard Thiel and Kay Hamacher (2018). Mechanical transduction of cytoplasmic-to-transmembrane-domain movements in a hyperpolarization-activated cyclic nucleotide–gated cation channel, J. Biol. Chem. 293(33):12908-12918.
doi: 10.1074/jbc.RA118.002139



**Contributions:**
To understand the gating mechanisms in ion channels, the initial concept of studying the underlying conformational changes via the previously implemented LRT null model, was given by me. Together with the co-authors, the study was further specified in the context of the allosteric gating of HCN channels. For this study, I adjusted the previously published `R` Biobliography `LRTNullModel` to work for homotetramers, like HCN channels. I performed all simulations and analyzed the data. I was further responsible for the preparation of all figures of the paper and the Supplementary Material. The co-authors especially contributed to interpreting the structural data in the context of HCN function. Based on an initial draft that was written by me, all co-authors helped to write the final manuscript.

To retain consistency throughout the whole thesis, changes of order and renaming of section titles compared to the published article may occur.

## 3.1 ABSTRACT

Hyperpolarization-activated cyclic nucleotide–gated cation (HCN) channels play a critical role in the control of pacemaking in the heart and repetitive firing in neurons. In HCN channels, the intracellular cyclic nucleotide–binding domain (CNBD) is connected to the transmembrane portion of the channel (TMPC) through a helical domain, the C- linker. Although this domain is critical for mechanical signal transduction, the conformational dynamics in the C-linker that transmit the nucleotide-binding signal to the HCN channel pore are unknown. Here, we use linear response theory to analyze conformational changes in the C-linker of the human HCN1 protein, which couple cAMP binding in the CNBD with gating in the TMPC. By applying a force to the tip of the so-called "elbow" of the C-linker, the coarse-grained calculations recapitulate the same conformational changes triggered by cAMP binding in experimental studies. Furthermore, in our simulations, a displacement of the C-linker parallel to the membrane plane (i.e. horizontally) induced a rotational movement resulting in a distinct tilting of the transmembrane helices. This movement, in turn, increased the distance between the voltage-sensing S4 domain and the surrounding transmembrane domains and led to a widening of the intracellular channel gate. In conclusion, our computational approach, combined with experimental data, thus provides a more detailed understanding of how cAMP binding is mechanically coupled over long distances to promote voltage-dependent opening of HCN channels.

## 3.2 INTRODUCTION

Out of an estimated 200 genes encoding ion channels in mammals, hyperpolarization-activated cyclic nucleotide-gated cation (HCN) channels are the only channels that open on membrane hyperpolarization but conduct a depolarizing inward current [22, 165]. HCN channels are also the only voltage-gated channels regulated by the direct binding of cyclic nucleotides. By virtue of these properties, HCN channels play unique and essential roles in a variety of physiological processes, the most important being the generation of spontaneous electrical activity in the heart and the regulation of synaptic transmission in the brain [22, 154, 165]. These channels, which are present in humans in the four isoforms HCN1-4, share the general architecture of voltage-gated $K^+$ channels [154]. Their monomers are built of six transmembrane domains (TMDs) of which the $4^{th}$ TMD (S4) comprises the voltage sensor, and the last two TMDs (S5-S6) contribute to the ion-conducting pore [169]. A specific feature of HCN channels is the presence of a cyclic nucleotide-binding domain (CNBD) at their cytosolic C-termini. HCN channels are activated by membrane hyperpolarization, and this voltage-dependent activation is further modulated in an allosteric manner by binding of cyclic nucleotides to the CNBD [44]. As a result of cAMP binding, the voltage dependency of channel opening is shifted to lower (less hyperpolarized) potentials. Thus, an intracellular increase in the cytosolic concentration of cAMP causes an earlier membrane depolarization and hence an acceleration of pacemaking [44].

A full understanding of the allosteric nature of HCN channel regulation by voltage and ligands requires insights into the mechanism responsible for processing both regulatory signals in the context of the whole protein. One component in this scenario, the CNBD, has been well studied. Its structure was solved at atomic resolution for most HCN isoforms in the presence of cAMP and for HCN2 also in the absence of cAMP [118, 170, 229]. These data show that this domain is built from an eight-stranded $\beta$-roll, which is connected to one $\alpha$-helix on the N-terminal side (A-helix) and two additional $\alpha$-helices on the C-terminal side (B- and C-helices). The cAMP-binding site is composed by two elements within the CNBD: the distal C-helix and the phosphate binding cassette; the latter is embedded within the $\beta$-roll and contains the short P-helix. The CNBD is connected to the channel pore via a helical domain, the C-linker, which in turn is composed of six $\alpha$-helices (A'-F'). The C-linker is critical for the transmission of conformational information between the CNBD and the transmembrane portion of the channel (TMPC) and is thus responsible for the communication between the two domains in channel regulation.

Recently, Lee and MacKinnon [112] obtained the first high-resolution structure of the full-length HCN1 channel in the cAMP-free and cAMP-bound state using cryo-electron microscopy (cryo-EM). The voltage-sensitive domain (VSD) of HCN1 is positioned next to the pore domain of the same subunit (non-swapped), an arrangement similar to that of the closely related Eag1 (Kv10.1) and CNG channels [113, 220]; this is different from the $K^+$ channels (Kv channels) where the VSD (S1–S4) is positioned near the pore domain of the neighboring subunit (domain-swapped). This latter arrangement of the VSD correlates with a much shorter S4-S5 linker, which in HCN1 is significantly shorter than the stretch of ~15 amino acids (AA) typical of Kv channels. Despite providing significant advances in our understanding of HCN channel structure-function relations, the seminal study by Lee and MacKinnon [112] left the question of how conformational information is transmitted between the CNBD and the transmembrane portion of the channel largely unanswered. Purification of the HCN1 protein in 0 mV conditions resulted in a channel locked in a closed conformation, independent on the cAMP-free or cAMP-bound configuration of the CNBD, preventing the required analysis of protein movements within the TMPC.

To address open questions on the conformational dynamics in HCN channels, here and elsewhere, we use linear response theory (LRT). This mechanical model, which was introduced by Ikeguchi et al. [84], can help to calculate the direction of conformational changes in a protein upon external perturbation, e.g. by ligand binding. LRT is a coarse-grained modeling technique [8] that requires much shorter computational time than molecular dynamics (MD) simulations to obtain insight into protein dynamics around the native state. Despite the simplicity, this approach still generates results that match very well with experimental data [66, 70, 72, 86] or results from all-atom MD simulations. This also holds true for the analysis of membrane proteins in models where the membrane is neglected, in order to further simplify the mechanical model [40].

To understand the mechanical connections between the HCN channel CNBD and the TMPC (i.e. the six transmembrane domains and the connecting linkers), in a previous study we employed LRT to model a synthetic channel in which the HCN channel C-linker/CNBD was connected to the available structure of the TMPC of the Kv2.1 channel [219]. This is because, at the time of the study, there was no structural information available yet on the HCN channel TMPC. The simulation predicted that release of cAMP from its binding site triggers a quaternary twist in the cytosolic portions of the four subunits in the channel tetramer [219]. This prediction turned out to be in good agreement with the cryo-EM structures of HCN1, which similarly suggest the predicted quaternary twist in the cytosolic parts in response to cAMP binding/release [112].

With the new structural information available on the entire HCN1 channel, we now revisit the open questions on the gating mechanism of HCN channels. In particular, we want to understand how the information from cAMP binding is transmitted to the TMPC and how conformational changes in the cytosolic domain are related to gating movements. For example, what is the movement of the $S_4$ domain in response to cAMP binding? What is the movement of the S6 domain in response to cAMP binding? How may cAMP binding favor opening of the channel gate?

Because the mechanical connections and the directions of conformational changes are not known for most proteins, including HCN channels, we recently developed a reference model for LRT (LRT null model) for a monomeric protein [66]. This method provides a way to uncover the mechanical response in proteins in an unbiased and efficient manner. Thus, instead of using information derived from a "typical" interaction, a functionally relevant residue in a protein is subjected to a set of random perturbations from any possible direction. The resulting responses, in the form of residue displacements, can then be clustered according to the general directions of displacement they impose on the protein structure. In the following step, perturbations from different clusters can be evaluated with respect to the plausibility of the direction of perturbation, which can be judged by comparing the computational data with experimental results obtained with the same protein.

In this study, the LRT null model was adjusted to work for the homotetrameric HCN1 channel. To identify the most plausible perturbation direction, we relied on previous experimental studies in which conformational changes in the cytosolic domains of HCN channels were monitored after binding of cAMP to the CNBD [5, 129, 170]. Based on these studies, we decided to apply an external force at a single position located in the bend between the A'- and B'-helices of the C-linker. As stated above, the C-linker is in a strategic position (Figure 3.1) as it connects the CNBD to the S6-helix of the pore, and thus, it is thought to transmit the conformational changes in the CNBD after cAMP binding/release to the pore [35, 117, 229].

We first show that all our modeling results are in good agreement with experimental data and thus confirm the assumption that the tip of the C-linker "elbow" [229] is an important position for the mechanical transduction of information from the CNBD to the HCN channel pore and vice versa. Next, we employ LRT analysis to predict the conformational changes in the orientation of the transmembrane helices, as well as the movement of the inner gate of HCN1, in

response to cAMP binding. More specifically, we show how tilting movements in the transmembrane helices, predicted by the LRT model, generate two important outcomes implied by prior experimental studies: 1) an increase in the distance between the voltage sensor (S4) and the surrounding transmembrane domains; and 2) a widening of the intracellular channel gate. Altogether, the combination of a high-resolution structure and LRT analysis provides a valuable tool for uncovering the short- and long-range mechanical connections in HCN channels, which are relevant for their gating by ligands.

## 3.3 MATERIALS AND METHODS

### 3.3.1 *Linear Response Theory*

This mechanical model [84] predicts conformational changes in a protein upon ligand binding. It is based on either the normal mode analysis (NMA) or elastic network models (ENMs), both of which extract collective motions in proteins [85, 197]. NMA, which is based on traditional MD force fields, can be applied to all-atom structures as well as to reduced, coarse-grained networks of projected beads and springs. The alternative approach, the so called ENMs [8], directly starts from a given (PDB) structure of a biomolecule. In the latter, the protein residues are represented by beads, and their covalent and noncovalent interactions are modeled as springs with corresponding spring constants. Interacting residues are defined by a cutoff threshold for spatial distances. Fluctuations of amino acids in ENMs can either be treated as isotropic, like in Gaussian network models, or as anisotropic, as in anisotropic network models (ANMs) [8, 191, 202]. Because LRT is based on ANMs, it is able to predict the direction of a structural change of a protein upon ligand binding; ligand binding is in this case mimicked by an external force vector [84].

The translocation $\Delta\vec{R}_i$ of atom $i$ can be predicted by Equation 3.1,

$$\Delta\vec{R}_i \simeq \beta \cdot \sum_j \langle \Delta\vec{R}_i \cdot \Delta\vec{R}_j \rangle_0 \cdot \vec{f}_j \tag{3.1}$$

where $\langle \Delta\vec{R}_i \cdot \Delta\vec{R}_j \rangle$ represents the covariance matrix of the fluctuations of the protein in the ligand-free state, and $\vec{f}_j$ denotes the external force vector that mimics ligand binding. $\beta$ is $1/k_BT$ with the Boltzmann constant $k_B$. The covariance matrix can be computed as the Moore-Penrose pseudoinverse [140, 151] of the Hessian matrix of an ANM [8].

### 3.3.2 *Structure Preparation*

All computations used the recently published cAMP-free structure of the HCN1 channel (PDB code 5U6O)[112] to simulate the effects of cAMP binding. The missing residues in the loop regions 201-202 and 243-251 of this cryo-EM structure were generated using `MODELLER` (version 9.15) employing the loop model function with fast loop refinement and restraint protein [55, 168]. Afterward, the remodeled structure was energy-minimized using `GROMACS` (version 2016.1/2016.2) [1] during 25,000 steps with steepest descent *in vacuo* to solve possible clashes in or with

the remodeled domains. As there were some tiny conformational changes in the C-terminal region after energy minimization in vacuum (C$\alpha$ root mean square deviation of 0.168 Å), the remodeled structure of residues 94-405 was combined with the original structure of residues 406-586, where no residues were missing. This was done by identifying overlapping residues in the C-linker region (AA 405 and 406). Subsequently, the coordinates of residues 94-405 of one structure was merged with residues 406-586 of the second structure into one PDB file. This combined structure of HCN1 was used for further computations.

The same procedure was carried out with the cAMP-bound form of HCN1 (PDB code 5U6P) to study the actual motions, which are occurring by cAMP binding. After remodeling of the missing loop regions, no conformational changes occurred in the C-terminal region. The resulting energy-minimized structure was used for further comparative computations.

### 3.3.3 *Perturbation of HCN1 Structure by External Forces*

The curated cAMP-free HCN1 structure was reduced to a heterogeneously parameterized anisotropic network model (ANM) [8] with a cutoff for connected residues of 13 Å, which was shown in a previous study to be a good choice [66]. For intra-chain interactions between amino acids, we used the matrix of spring constants from Miyazawa and Jernigan [138], and for inter-chain interactions, we used the matrix for spring constants from Keskin et al. [96] as derived in Ref. [71]. The ANM, which is illustrated in Figure A.14, was afterward perturbed at Ala-425 at the tip of the elbow of the C-linker from 1000 different directions using 1000 external force vectors with a force strength in arbitrary units. Therefore, the LRT null model introduced in a previous study for a small monomeric protein [66] was adjusted to a homotetrameric protein like HCN1. To account for the rotational symmetry of the tetramer, Ala-425 of each subunit was simultaneously perturbed by the same 1000 (also rotated) force directions. Afterward, clustering of the random force directions into four clusters was done based on the induced displacements of residues 446-465 (C′ and D′ helices), which represent the "shoulder" of the C-linker. The displacement vectors of the selected residues after perturbation were clustered by applying the `kmeans` algorithm from Hartigan and Wong [74]. The corresponding force directions were accordingly clustered and visualized by different colors in Figure 3.1. Robustness of clustering was proved by a 1000-fold repetition of the clustering approach [66].

With this strategy we investigated possible influences of the movement of the elbow (A′- and B′-helices) of the C-linker toward the shoulder of the C-linker. The decision for four clusters was based on the comparison of log values of maximal within-cluster sum of squares (maximum withinss) from `kmeans` clustering as a function of number of clusters, which was explained in detail in Ref. [66] (Figure A.15).

For each cluster, one representtative force direction was chosen to determine the resulting displacements. Therefore, perturbations for each selected force direction were computed for repulsive forces with varying strengths to determine a trend for increasing force strengths.

### 3.3.4 *Computation of Inner Gate Radii*

Computation of the inner gate radii was done by using the program HOLE [188]. The computations were restricted to the region from Val-390 to Gln-398. The values of minimal radii within this region were extracted from the HOLE output.

## 3.4 RESULTS

### 3.4.1 *C-Linker Movement after Perturbation at the Elbow*

The C-linker is an important structure in HCN channels for the coupling of conformational changes in the CNBD, which are generated by cAMP binding, with the channel pore [211, 213]. In all HCN structures available, the C-linkers are found tightly packed in the tetramer in the so-called "elbow on the shoulder" conformation in which the A'- and B'-helices (elbow) of one subunit interact with the C'- and D'-helices (shoulder) of the adjacent subunit [112, 229]. Based on experimental data, which related channel activation of cyclic nucleotide-gated channels with structural features in the CNBD [35, 36, 91], a model was proposed in which, upon cAMP binding, the elbow moves with respect to the shoulder of the adjacent subunit in an overall centrifugal motion, away from the central axis of the channel, which causes a widening of the inner pore [88].

To simulate the proposed C-linker movement, we perturbed the anisotropic network model (ANM) of cAMP-free HCN1 structure at the tip of the elbow (Ala-425 in the bend between the A'- and B'-helices) with forces from different directions. The elbow was chosen to perturb the protein because it seems to undergo a large conformational change during cAMP binding [88]. To investigate which of these perturbation directions causes conformational changes that best match the experimentally observed C-linker movement, the force directions were first clustered into groups of similar displacement effects. Clustering, which is described in detail in "Materials and Methods", was based on the effect of the elbow movement on the underlying shoulder, represented by residues 446-465 on the C'- and D'-helices. The shoulder domain was chosen for clustering because it is directly coupled to the elbow and hence responds in a direct manner to perturbations at the elbow [88]. Clustering based on movements of all residues would be less meaningful as unstructured loop regions can move very randomly. The clustering of perturbation directions at the tip of the elbow, which induce different movements of the shoulder, is illustrated in Figure 3.1. The displacements observed in response to one representative force direction for each of the four different clusters are shown in Figure 3.2, where the coloring of the displacement vectors corresponds to the coloring of clusters in Figure 3.1 (and in all following figures), and the length of the arrows is proportional to the size of the LRT displacements (rendered in arbitrary units).

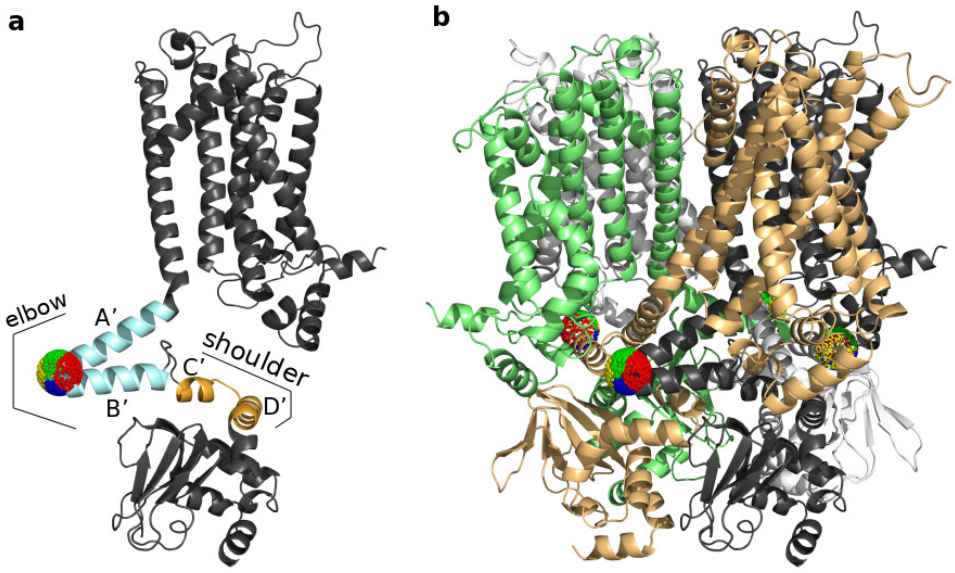


Figure 3.1: LRT null model of cAMP-free HCN1 channel with clustering of different perturbation directions. (**a**) Clustering of the different perturbation directions imposed on Ala-425 at the tip of the elbow (light blue) based on the displacement of the shoulder (orange). Both elements are part of the C-linker, which connects the CNBD to the S6 domain of the channel. For clarity, a single subunit of HCN1 is shown. Perturbation directions on a sphere around Ala-425 that belong to the same cluster are represented in the same color (red, blue, yellow, or green); (**b**) Clustering of the perturbation directions shown for all four subunits of HCN1. To illustrate contacts between individual subunits, the four monomers are shown in different colors. The dark gray color of the subunit in **b** corresponds to the same color in **a**.

The data show in a side view perspective that horizontal (in the plane of the membrane) and vertical (perpendicular to the plane of the membrane) perturbations cause distinct displacements in the C-linker/CNBD domain as well as in the TMPC. Notably, the strongest displacements (longer arrows) throughout the protein are elicited by perturbations applied in a horizontal direction parallel to the plane of the membrane (Figure 3.2 a and 3.2 b). These perturbations (Figure 3.2 a and 3.2 b, red and yellow arrows) show the best match with the previously proposed model for C-linker movements in response to cAMP binding [35, 36, 91]. The predicted iris-like movement of the C-linker of HCN1 can indeed be seen in the simulations. A top view of the C-linker shows an iris-like motion as a result of the horizontal displacements of the elbow (Figure 3.3 a and b). Vertical perturbations, as represented by the force direction vectors in the other two clusters, do not induce such movements (Figure 3.3 c and d).

To further validate the conformational changes, which are predicted in our computational results, we used the high-resolution structures of HCN1 in the cAMP-free and cAMP-bound conformation for comparison (superposition, Figure 3.3 f) [112]. Figure 3.3 e shows the position of the elbow and shoulder in the cAMP-free form. The arrows in Figure 3.3 e represent the movements in the protein, which are required for the transition from the experimentally determined cAMP-free into the cAMP-bound structure [112]. A comparison between the simulated and

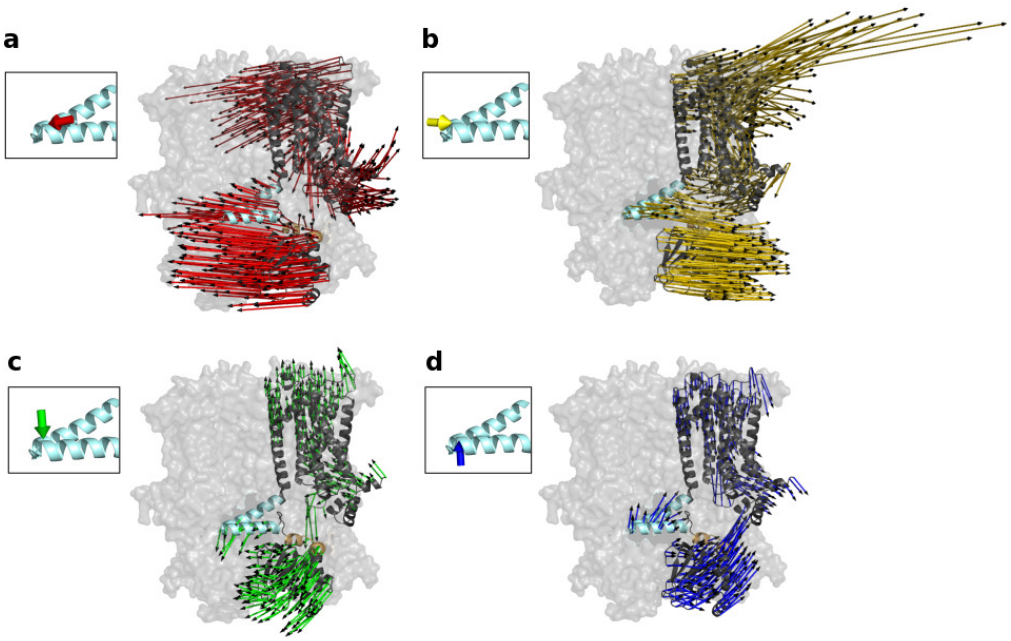


Figure 3.2: Perturbation of the ANM from HCN1 in the cAMP-free form. Perturbations at the tip of the elbow from four different clusters lead to distinct horizontal and vertical displacements of the C-linker. Displacements of HCN1 after perturbation at Ala-425 from one representative direction (**a**-**d**) for each cluster correspond to the coloring in Figure 3.1. For clarity, the displacements are only shown for one subunit, which is highlighted in cartoon representation. The direction of force application is illustrated in the insets. The other subunits are illustrated in transparent surface representation. The thin arrows demonstrate the displacement of each residue. As in Figure 3.1, the elbow is highlighted in light blue, and the shoulder is highlighted in orange.

experimental data shows that the displacement in Figure 3.3 b (yellow arrows) in which a force on the tip of the elbow is applied pushing toward the central axis of the protein (inward) reveals a very good match with the direction of the conformational changes of the C-linker observed after cAMP binding. As the tip of the elbow is pushed toward the central axis of the protein, the helices of the elbow and the underlying shoulder respond by all moving in the overall same direction away from the center in a centrifugal motion, but at different angles. This indeed generates an iris-like rotational movement, seen in the top view perspective of Figure 3.3 b and e, as a counter-clockwise rotation.

Whereas the horizontal movement of the elbow in Figures 3.2 b and 3.3 b recapitulates the experimentally determined iris-like transition of the C-linker in response to cAMP binding, it is important to note that in all HCN structures resolved thus far the C-linker is found in the resting (nonactive) position. Thus, in the full-length HCN1 structure (obtained at 0 mV), the depolarized position of VSD always locks the channel in a closed state, whether cAMP is present or not. This is likely to impose a severe limitation on the range of movement the C-linker is able to undergo (see Figure 3.3, legend). Similarly, it has been suggested that in all available crystal structures for the cAMP-bound C-linker/CNBD fragment, the C-linker is found in a resting state, as inferred by the presence of a set of saline

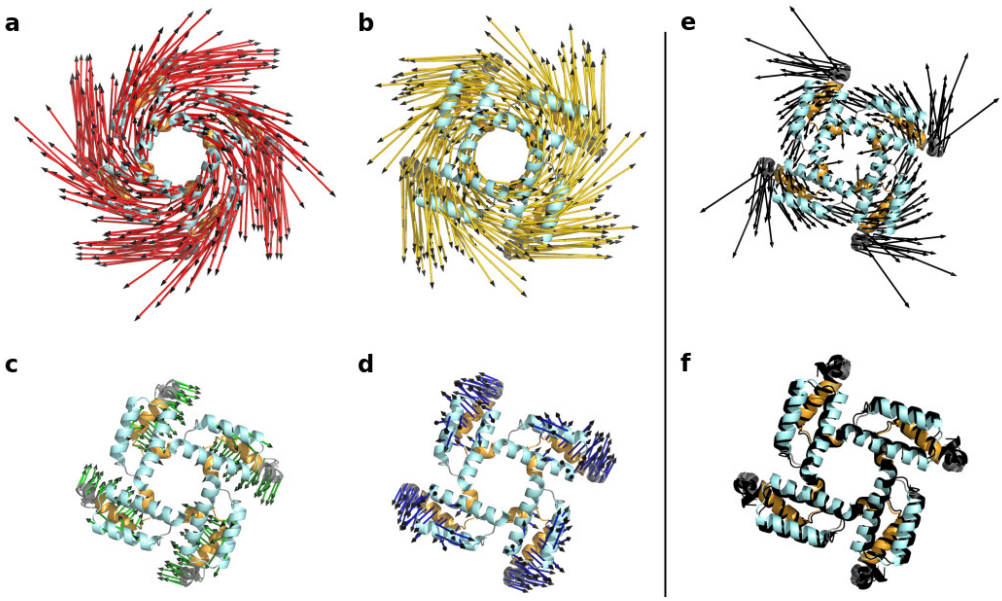


Figure 3.3: Comparison of predicted and measured displacements of the C-linker in cAMP-free and cAMP-bound HCN1 channel. Displacements of the C-linker of HCN1 after perturbation at Ala-425 from one representative direction (**a**-**d**) for each cluster. Coloring corresponds to clusters in Figures 3.1 and 3.2. (**e**) Experimentally observed displacement of the C-linker after superposition of the full-length cAMP-free and cAMP-bound HCN1 structures. Because the displacements are very small in the experimental structures, the arrows shown represent an arbitrary multiple of the actual displacement so as to allow to see the directions properly and compare them with the LRT displacements; (**f**) Superposition of the cAMP-free (colored) and cAMP-bound (black) HCN1 structures showing the actual displacements. In all visualizations the proteins are shown in top view from an extracellular perspective, so that the movements of all four subunits can be seen. The elbow again is highlighted in blue, and the shoulder is highlighted in orange. The remaining residues are colored in gray.

bridges in the structure, which are postulated to break in the open state of the HCN channel [35, 36]. Our modeling data are able to simulate a much wider range of C-linker movements independent of the presence or absence of these critical salt bridges and to document these effects of movements on the rest of the protein. Although arbitrary, such a wider range of movements is likely to provide a useful representation of the scope of motions found in the actual protein.

Thus, the simulations also show how a horizontal force applied in the outward direction (red arrows) causes a displacement in the opposite direction (Figures 3.2 a and 3.3 a). This displacement likely reflects the direction of conformational changes the protein may undergo upon cAMP release, and indeed, very similar residue displacements are observed when the same force is applied on the ANM of the cAMP-bound HCN1 structure (Figure A.12). Based on these results, we proceeded to further analyze all conformational changes under the assumption that displacements elicited by the inwardly directed horizontal force applied to the elbow of the cAMP-free HCN1 structure (as represented by the yellow arrows in Figures 3.2 and 3.3) reflect the structural changes that are induced by cAMP binding in the actual HCN channel protein.

### 3.4.2 *Effects of C-Linker Movement on the CNBD*

The modeling data suggest that a horizontal movement of the C-linker transmits the conformational changes in the CNBD, which originate from cAMP binding, toward the TMPC. Because LRT and elastic network modeling, approximate the mechanics of a protein around a stable configuration, we must assume that any movement, which is induced by a local perturbation, is bidirectional. This means that any force, which leads to a distinct structural response, must be reproduced by the same structural perturbation. Using this logic, we can assume that the same perturbation at the elbow, which mimics cAMP binding, should also cause a realistic conformational change in the opposite direction, namely toward the cAMP-binding domain. To test this prediction, we analyzed the effect of the C-linker movement after the perturbation mimicking cAMP binding (Figure 3.2 b, yellow arrows) on the CNBD. The computational prediction can be compared to the conformational changes observed in the experimental HCN structures in the presence and absence of cAMP [170, 229]. Figure 3.4 a illustrates cAMP within the binding pocket of the CNBD. The yellow arrows in Figure 3.4 b illustrate the predicted movements in the CNBD in response to a force on the elbow, which mimics cAMP binding. The corresponding movements from experimental data are represented by black arrows in Figure 3.4 c. The relative movement of the domains can be appreciated from a superposition of cAMP-free and cAMP-bound structures (Figure 3.4 d). Comparison of both data sets (Figure 3.4 b and c) shows a similarity but also differences. The similarity is that cAMP binding induces *in silico* as well as in the experimental structures an overall outward movement of the CNBD, driven by the outward movement of the shoulder described above; representative arrows which point in the same direction in both structures are highlighted in red. Also in this case, the computational model recapitulates the experimental data. Differences between the two predictions occur mostly within the C-helix, where the directions of the theoretical and experimentally predicted displacements point in opposite directions (highlighted by the blue arrows; Figure 3.4 b and c). The deviations between the two approaches for the C-helix come as no surprise; the experimental data were obtained in the presence and absence of cAMP, respectively, so that the conformational changes reflect the sum of ligand binding/release plus the subsequently triggered conformational changes in the CNBD which include folding of the C-terminal portion of the C-helix and formation of the P-helix [170]. The computational data, however, only capture the conformational changes in the protein occurring as a result of cAMP binding. Therefore it should be expected that the computational and the experimental structures differ in some of the elements that make direct contact with cAMP, because the latter may undergo conformational changes during ligand binding [170].

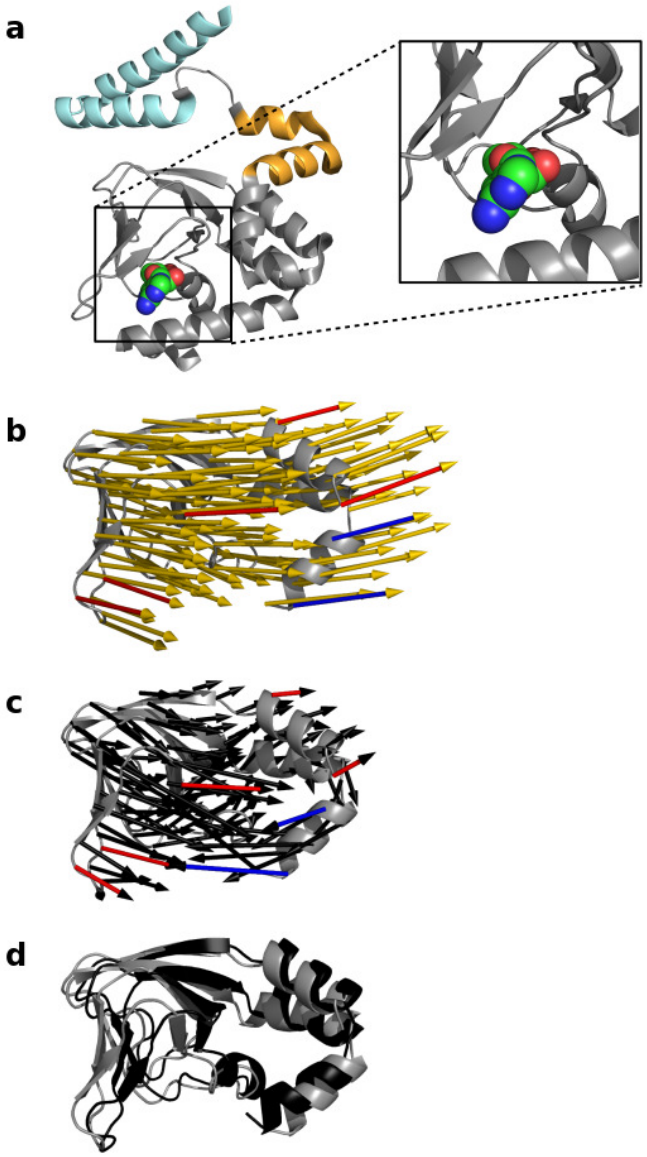


Figure 3.4: Comparison between predicted and measured displacements of CNBD in cAMP-free and cAMP-bound HCN1 channel. (**a**) cAMP-bound HCN1 structure (residue 401-608, cytosolic domain) showing the orientation of cAMP within the binding pocket. The HCN1 structure is shown in cartoon representation with elbow and shoulder of the C-linker colored according to Figure 3.1. The cAMP molecule is highlighted by sphere representation. **b**-**d** Comparison of the displacements in the CNBD (residues 480-586) after perturbing the elbow from the most realistic direction (**b**, yellow arrows) to the experimentally resolved displacement between cAMP-free (gray) and cAMP-bound (black) HCN1 structure (**c**, black arrows; **d**, superposition). The length of the arrows was chosen such that the length in **b** and **c** are similar; the absolute length has no quantitative meaning. The red and blue arrows highlight exemplary displacements, which are similar or different between the computational model and the experimental data, respectively. The experimental data ±cAMP are from Ref. [112]. For clarity, only one subunit of HCN1 is shown.

### 3.4.3 *Effects of cAMP Binding on Conformations in the Transmembrane Portion of the Channel*

COUPLING OF C-LINKER TO S4-S5 LINKER Several studies have highlighted a role for the HCN channel VSD in the allosteric regulation of cAMP affinity for the CNBD [110, 201, 223, 224]. These effects could be potentially explained by the suggested interaction between the S4-S5 linker and the C-linker [42]. Several studies have proposed that the relative orientation of these two domains changes during channel gating [111, 156]. In a recent study, the distance between an AA of the S4-S5 linker (Phe-359) and the A'-helix in the C-linker of a HCN channel from sea urchin sperm, called spHCN, was monitored using transition metal ion FRET [5]. The measurements showed that cAMP binding reduced the distance between the S4-S5 linker and the C-linker. To compare these experimental results with the predictions from our computational study, we assessed the distance between AAs Tyr-289 and Lys-412 in the HCN1 ANM structure, the former position corresponding to Phe-359 in spHCN and the latter to the center of the A'-helix in the C-linker. The data in Figure 3.5 show that the LRT calculation exhibits the same conformational change predicted from the experimental study; indeed, application of an appropriately directed force to the tip of the elbow, i.e. a force that triggers the conformational changes of cAMP binding (yellow arrows in Figures 3.2 and 3.3), results in a reduced distance between the C-linker and the S4-S5 linker. This distance decreases in an exponential fashion with the strength of the force. It is important to note that the LRT only provides qualitative results in terms of the directions of displacements in a protein. For this reason, we can only show whether the distance decreases or increases; the actual magnitudes of displacements cannot be computed, and thus, the distances are only given in arbitrary units.

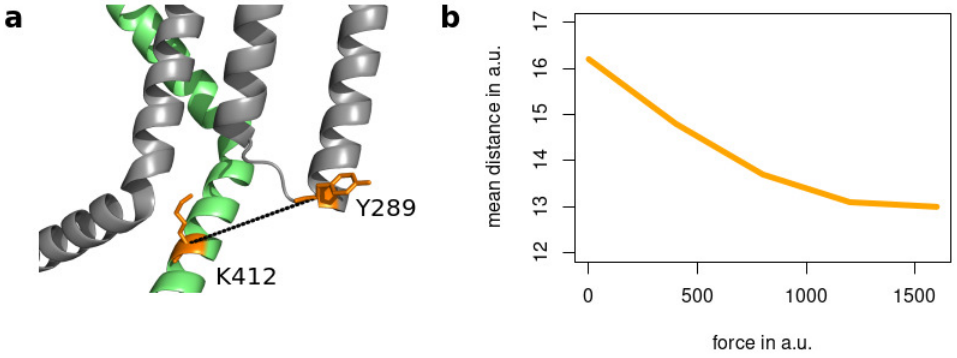


Figure 3.5: Simulation predicts that the C-linker and the S4-S5 linker move closer together after cAMP binding. (**a**) Visualization of the distance between Tyr-289 of one subunit (gray) and Lys-412 of the neighboring subunit (lime green); (**b**) Computed distance between the CA positions of Tyr-289 and Lys-412 of two neighboring subunits after perturbation at Ala-425 with forces of increasing strengths (yellow cluster in Figures 3.2 and 3.3). Both the forces as well as the distances are given in arbitrary units (a.u.) and thus can only show a trend.

The results of these experiments suggest that a cAMP-induced conformational change in the CNBD is mechanically transmitted via the C-linker to the S4-S5 linker. A potential physical interaction between the two domains provides a plausible mechanism for a reciprocal communication between the cytosolic domain and the TMPC.

COUPLING OF C-LINKER TO VSD AND PORE MODULE The experimental structures of full-length HCN1 reveal hardly any differences in the TMPC (VSD and pore module) between the cAMP-free and cAMP-bound form. This would seem to imply that the movement of the C-linker only generates, at least in the absence of a membrane voltage, subtle changes in the TMPC. To obtain more information on how cAMP can then so effectively modulate the voltage-dependent gating of the channel, we visualized the displacements in HCN1 after perturbation at Ala-425 from the direction that simulates cAMP binding (yellow arrows in Figures 3.2 and 3.3). The resulting displacements of the six helices S1-S6 of the TMPC are illustrated in Figure 3.6 in a front and side view and in Figure A.13 from a top and bottom perspective. This analysis shows that the C-linker movement has indeed a distinct effect on the TMPC. All six helices undergo a tilting in the sense that the upper parts move into the opposite direction from the lower parts (Figures 3.6 and A.13 a-f). It is worth noting that the angles and directions of movement are somewhat different for each TMD. The results of this analysis imply that cAMP binding may facilitate voltage-dependent opening of the channel by inducing a differential tilting like movement of the TMDs. The modeling data underscore a remarkable degree of movement at the extracellular end of the TMDs and interconnecting loops. This observation may provide a potential explanation for prior experimental studies, which have demonstrated a critical role for these elements in the modulation of HCN channel gating (see "Discussion").

Our computational data, however, show no indication for a vertical displacement of the transmembrane helices, including S4, in response to protein perturbations mimicking cAMP binding. This, of course, does not rule out that vertical displacements of S4, similar to what is seen in Kv channels, may occur in HCN channels in response to voltage. Indeed, a vertical translocation of S4 in response to voltage was suggested by an early study using cysteine accessibility methods in the sea urchin HCN channel [129]. Later studies employing similar methods in the mammalian HCN1 channel, however, have cast some doubt on this simple model [16, 209]. These studies demonstrated that the intracellular region of S4 that displays state-dependent modification is much larger than the state-dependent extracellular S4 region and that the observed modification rate shifts are much more pronounced for residues at the intracellular compared with the extracellular end of S4 [16, 209]. These incongruences have been variously explained either by postulating an "unwinding" of the S4-helix upon hyperpolarization, such that the lower end is able to undergo a vertical translation while the upper end remains relatively stable [209]. As an alternative scenario, it has been postulated that a rearrangement in the transmembrane segments surrounding S4 may cause formation or collapse of a water-filled internal "gating" canal in response to negative or positive voltages, respectively. As a consequence, the shape of the membrane field surrounding the S4 segment would correspondingly be altered [16]. In light

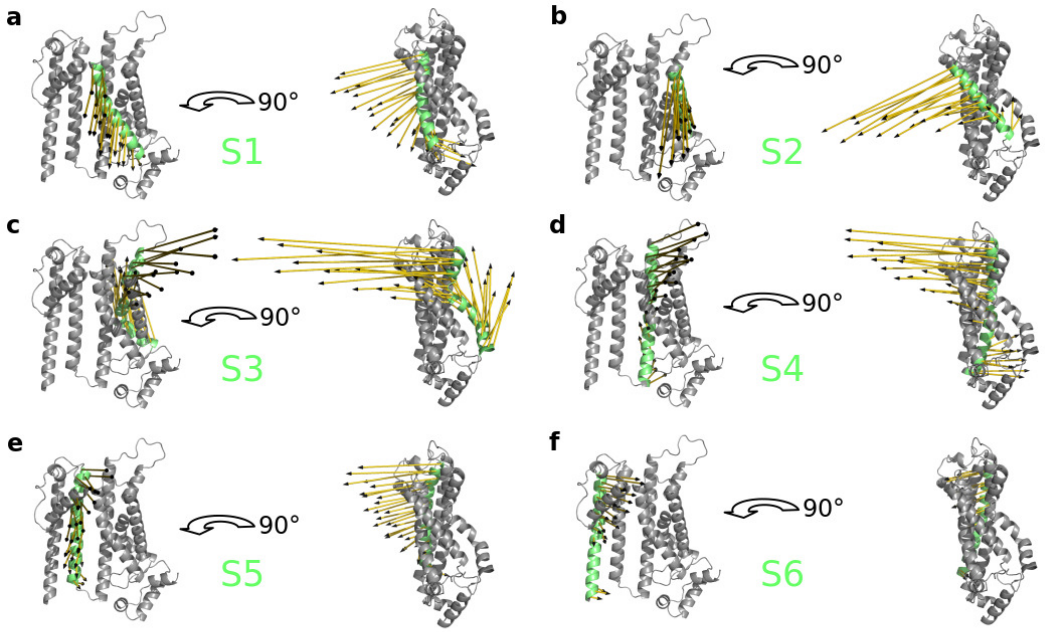


Figure 3.6: Simulation predicts tilting movements of transmembrane helices S1–S6 after cAMP binding. Displacements of S1 to S6 helices of the TMPC (**a**-**f**) after perturbing the elbow of the C-linker from the most realistic perturbation direction. The corresponding helix is highlighted and labeled in lime green, and the displacement is visualized as yellow arrows. For clarity, only one subunit (residue 94-402) is shown from the front and side views.

of the latter hypothesis, we tested the possibility that cAMP binding might alter gating by affecting the width of this internal aqueous gating canal. We therefore measured the distance of reference residues within S4 with partner residues on S5 and S3 (Figure 3.7 a). In both previous studies, the residues in the mouse HCN1 channel corresponding to Leu-265 and Ser-272 in human HCN1 (Figure 3.7) became accessible to sulfhydryl-modifying methanethiosulfonate reagents upon hyperpolarization [16, 209]. Our present data show that application of a force on the elbow, in a direction which simulates cAMP binding, affects the distance between S4 and the surrounding transmembrane domains (Figure 3.7 b). Again, distances can only be given in arbitrary units, and thus, we only focus on whether the distances between residues of interest decrease or increase. Increasing force augments a separation between S4 and S5 and between S4 and the lower segment of S3 (S3a). Together with the experimental data, this suggests that negative voltage and cAMP binding may act in the same manner in that they open up an internal gating canal. This could provide an explanation for the allosteric function of ligand binding and voltage toward the opening of HCN channels.

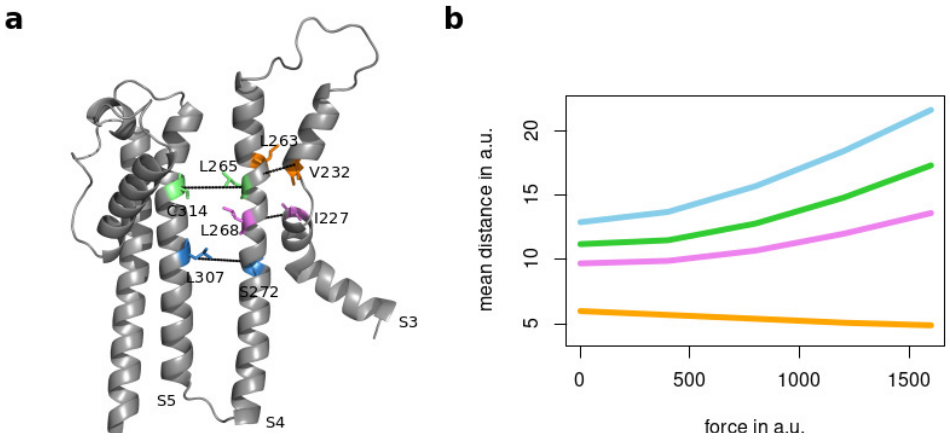


Figure 3.7: cAMP binding predicts a local increase in the distances between S4 and surrounding transmembrane domains. (**a**) HCN1 monomer with reference amino acids in S3, S4, and S5; (**b**) Computed distances between these amino acids in the same subunit after perturbation of Ala-425 with forces of increasing strengths (yellow cluster in Figures 3.2 and 3.3). The colors in **b** cross-reference with colors in **a**. Forces as well as the distances are again given in arbitrary units (a.u.).

COUPLING OF C-LINKER TO THE INNER GATE The current view on HCN channel activation is that they undergo a voltage-dependent transition from a resting state to a nonconductive active state before opening in a final voltage-independent transition. The latter step is allosterically modulated by cAMP binding to the CNBD, which in turn favors the opening transition in the pore via C-linker movements [32, 212]. It has been proposed that cAMP binding stabilizes the open state by a rotational movement of the C-linkers and a consequent widening of the inner channel gate [35, 36]. Having determined that a lateral pushing force on the C-linker elbow results in transmission of conformational forces to the TMPC in our model, we further proceeded to examine the associated changes in the radius of the inner gate. The respective region, comprising residues 390-398 of the S6 helix at the constriction of the inner gate, is shown in Figure 3.8 a. Figure 3.8 b shows the development of the minimal inner gate radius with forces of increasing strengths acting on Ala-425. The analysis shows that only perturbations from directions of the yellow cluster, which simulates cAMP binding, lead to a relative pore widening. Perturbations from all other directions cause in contrast a narrowing of the inner gate.

The results of this analysis further supports the assumption that the movements in the protein, which are caused by a force direction from the yellow cluster, mimic the conformational changes that take place after cAMP binding. Thus, the cAMP-dependent rearrangements, which originate in the CNBD, are ultimately transmitted to the TMPC in such a manner as to favor the opening of the inner gate. In contrast, gate opening is opposed and fully reversed by the conformational changes, which are induced by cAMP release from the CNBD (Figures 3.2 a and 3.3 a, red arrows).

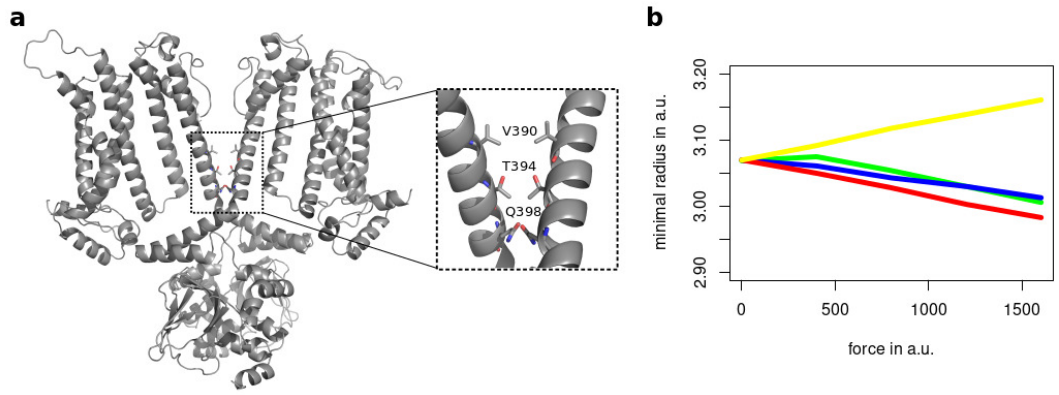


Figure 3.8: Only a horizontal displacement of the elbow causes widening of inner channel gate. (**a**) Location of inner gate (inset) in the global structure of HCN1; (**b**) Development of minimal inner gate radii in the region from Val-390 to Gln-398 with forces of increasing strengths acting on Ala-425. Force direction is represented according to the color coding used in Figures 3.2 and 3.3. Both the forces applied and the minimal radii are given in arbitrary units (a.u.) as the LRT model only provides a trend for protein movements, in a qualitative manner but not quantitatively. Upon perturbations from directions of the yellow cluster, a widening of the inner gate can be observed, whereas for perturbations from other directions, the inner gate becomes narrower.

## 3.5 DISCUSSION

The current challenge in understanding HCN channel function is to integrate information from static, high-resolution structures with a wealth of functional data on the dynamics of channel gating. Here, we show that LRT provides an alternative route to understand how ligand binding in the CNBD is related to channel gating in the transmembrane part of the protein. This coarse-grained computational approach offers less molecular details than conventional MD simulations and also ignores the impact of the membrane on the protein of interest. But in spite of these simplifications the LRT approach still provides valuable insights into the mechanic connections within the protein. Based on the available high-resolution HCN1 structure [112], the LRT computations reproduce without any bias [66] many of the conformational changes, which were observed in structural experiments in response to cAMP binding. Our approach is based on the assumption that movements in the elbow of the C-linker are mechanically connecting, in a bidirectional manner, conformational changes in the CNBD and in the TMPC. This model, which was originally proposed on the basis of experimental data [35, 223, 224], is perfectly reproduced by the computational data. The simulations show that only a horizontal inwardly directed force applied to the tip of the elbow causes the same rotation and iris-like opening of the C-linker domain, which had been first proposed based on functional experiments [35, 36] and then observed in structural experiments on HCN channels in response to cAMP binding [112]. In the LRT computations, this rotational movement propagates in the direction of the TMPC, where it causes a tilting of the TMDs and a small but distinct widening of the inner gate at the cytoplasmic end of the pore. This movement is overall consistent

with the conformational changes between the cAMP-free and cAMP-bound HCN1 structure, which also suggest a concerted rotation of the C-linker and displacement of the S6-helix in favor of channel opening [112].

Our computational results furthermore show that the same horizontal force application on the elbow also propagates in the opposite direction toward the CNBD. There it generates distinct conformational rearrangements in the CNBD. The conformational changes, which occur in response to an imposed inward directed force at the tip of the elbow, are overall very similar to those observed in response to cAMP binding to the CNBD in experimental studies [170].

Collectively, the results of these analyses underscore a very good agreement between predictions from the computational simulation and experimental data. With this support for the predictive power of the computational method, we can now address the question of how cAMP binding may facilitate opening of HCN channels. The simulation data show that the rotation and iris-like opening of the cytosolic domain, which is triggered by cAMP binding, propagates into the TMPC, where it causes a distinct tilting of the TMDs. A close scrutiny of the S4 domain shows that a cAMP induced movement causes only a lateral but no vertical displacement of the voltage sensor. Such finding is consistent with previous studies, which have suggested that a vertical movement of the S4 domain may not be of central importance in the gating of HCN channels [16] but rather that the voltage dependency of HCN channels may be modulated by a rearrangement of the TMDs surrounding the S4 domain. A central argument in this scenario is that the formation, or collapse, of a water-filled crevice (internal gating canal) could alter the shape of the electrical field around the S4 segment [16, 209]. In light of this model, it was very intriguing to find that the computational data highlight a general tilting type movement of all TMDs in response to cAMP binding. The analysis of distances between critical residues in S4 with respect to reference residues in S3 and S5 shows that cAMP binding indeed augments a space between the voltage sensor domain and its surrounding helices. This effect of cAMP binding might lower the energy barrier for the opening of an internal gating canal and in this way facilitate HCN activation by negative voltages [16]. Whereas our computational data do not rule out the possibility that S4 may undergo a vertical translational movement in response to voltage, as some other authors have suggested [129], the results certainly lend support to the notion that important lateral displacements may occur in S4 and its surrounding TMDs during HCN channel gating. Our LRT calculations also show that, in response to cAMP binding, the A'-helix of the C-linker moves closer to the S4-S5 linker. Several previous experimental studies have speculated that such a movement could similarly be important for the facilitated activation of HCN channels in the presence of cAMP [42, 111, 156].

A second central result of the simulations is that the rearrangement of the TMDs, which is triggered by movements in the CNBD and C-linker, causes a progressive widening of the inner gate at the intracellular end of the channel pore. Very intriguingly, we find that such widening of the inner gate is only achieved upon a horizontal inward movement of the elbow, which follows cAMP binding; the entire process is reversed by a movement in the opposite direction, which is presumably triggered by cAMP release from its binding site. It is also important to note, in this context, that a movement of the elbow in the vertical direction has only a minor

impact on the width of the inner gate and that there is little difference between an upward or downward movement. The results of these data therefore suggest that an upward or downward movement of the C-linker elbow is not immediately apparent in cAMP-mediated channel gating.

The conformational change at the cytoplasmic end of the HCN1 channel pore is a potential mechanism by which cAMP binding may modulate the voltage-dependent opening of HCN channels, as was also suggested on the basis of previous experimental data [35, 112]. Because LRT only provides a qualitative and not quantitative indication about the trend of protein movement, we cannot gauge whether the induced widening observed in the simulation would be sufficient to open the intracellular channel gate. Other mechanisms may very well be at play, including additional movements induced by changes in voltage across the membrane, which are not captured in our simulation. Nonetheless, the bidirectional effect on C-linker dynamics in response to cAMP binding/release represents a perfect reversible mechanism for the long-distance modulation of HCN gating.

Finally, on a more speculative note, the LRT simulation revealed a remarkably wide scope of motion at the extracellular end of the HCN1 protein TMDs. Although very little is known about the role of movement at the top of the TMDs and interconnecting loops (S1-S2, S3-S4, and S5-S6, including the selectivity filter), several studies have implicated these elements in the modulation of HCN channel gating. Residues in the S1-S2 loop influence the gating kinetics of HCN channels, and thus contribute to determine the different properties of the HCN1, HCN2, and HCN4 channel isoforms [85, 191]. Variations in the sequence and length of the S3-S4 loop prominently modulate the voltage dependence of HCN1 channels [203, 204]. More recently, a mutation located at the extracellular end of S6, identified in a Brugada syndrome patient, was found to shift the voltage dependence of the HCN4 channel by nearly 10 mV [23]. These results collectively suggest that critical interactions are likely to occur around the outer mouth of the HCN channel pore, which may regulate movement during HCN channel gating. Future simulations, coupled with experimental structure-function analysis, may be able to provide insights into this question, and the many other open questions still remaining on the dynamics in HCN channels.

## ACKNOWLEDGEMENTS

This work was supported by the Ministry of Higher Education, Research and the Arts (HMWK) of the Hessen State in LOEWE Project iNAPO, by the European Reseach Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program Grant 695078 noMAGIC ERC 2015AdG, and by National Institutes of Health Grant R01 NS036658 (to B. S.). We further thank Daniel Bauer for modeling the missing loop regions in the HCN1 structures.

## AUTHOR CONTRIBUTIONS

C. G. data curation; C. G. formal analysis; C. G. visualization; C. G. and K. H. methodology; C. G., A. S., B. S., A. M., G. T., and K. H. writing-original draft; A. S., B. S., A. M., G. T., and K. H. conceptualization; A. M., G. T., and K. H. funding acquisition; K. H. resources; K. H. software; K. H. supervision

## SUPPORTING INFORMATION

The Supporting Information is available free of charge on the JBC Publications website at DOI: 10.1074/jbc.RA118.002139
This article contains Figures S1-S4.

In this thesis, the Supplementary Figures can be found in Appendix A2 as Figures A.12-A.15.

# 4

# INVESTIGATION OF ION CHANNEL ACTIVITY OF VPU PROTEINS

**The following chapter has been published in:**

Timo Greiner, Sebastian Bolduan, Brigitte Hertel, Christine Groß, Kay Hamacher, Ulrich Schubert, Anna Moroni and Gerhard Thiel (2016). Ion Channel Activity of Vpu Proteins Is Conserved throughout Evolution of HIV-1 and SIV, Viruses, 8(12):325. doi: 10.3390/v8120325



**Contributions:**

In this study I contributed as co-author. In this context, I was responsible for performing the information theoretic calculations, i.e. computation of Shannon entropy, for gaining insights into the conservation state of several residues that are involved in the ion channel function. Therefore, I extracted a multiple sequence alignment of the Vpu protein family from the PFAM database and further processed the full alignment for computing the Shannon entropy. I was responsible for the content and preparation of Figure 1 and Table 1 in the paper and for Figure S2 (here: A.17) in the Supplementary Material. I analyzed and discussed the data in the context of the whole study and helped to write the paper. My focus thereby lay on the methods and results parts of the information theoretic calculations and associated results.

To retain consistency throughout the whole thesis, changes of order and renaming of section titles compared to the published article may occur.

## 4.1 ABSTRACT

The human immunodeficiency virus type 1 (HIV-1) protein Vpu is encoded exclusively by HIV-1 and related simian immunodeficiency viruses (SIVs). The transmembrane domain of the protein has dual functions: it counteracts the human restriction factor tetherin and forms a cation channel. Since these two functions are causally unrelated it remains unclear whether the channel activity has any relevance for viral release and replication. Here we examine structure and function correlates of different Vpu homologs from HIV-1 and SIV to understand if ion channel activity is an evolutionary conserved property of Vpu proteins. An electrophysiological testing of Vpus from different HIV-1 groups (N and P) and SIVs from chimpanzees ($SIV_{cpz}$), and greater spot-nosed monkeys ($SIV_{gsn}$) showed that they all generate channel activity in HEK293T cells. This implies a robust and evolutionary conserved channel activity and suggests that cation conductance may also have a conserved functional significance.

## 4.2 INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) accessory protein Vpu is a 15–20 kDa oligomeric type 1 integral membrane phosphoprotein [34, 128, 194], which is encoded exclusively by HIV-1 and related simian immunodeficiency viruses (SIVs), but not by the majority of SIVs and HIV-2. It has been shown that Vpu augments virus release by counteracting the human host cell restriction factor tetherin [146, 206]. Moreover, Vpu has been shown to induce degradation of the CD4 receptor by the endoplasmic reticulum (ER)-associated protein degradation (ERAD) pathway [24, 127, 162, 180]. The cytoplasmic domain of Vpu contains a pair of serine residues (at positions 52 and 56), which are constitutively phosphorylated by the casein kinase 2 (CK-2) [182]. CK-2 mediated phosphorylation of the two serine residues is critical to induce CD4 degradation [57, 183, 184] and for the assembly of transmembrane domains as homo-oligomers [114]. Computational studies advocate a model according to which a putative ion conducting pore is formed by the transmembrane domains of monomers, which assemble in a dynamic manner in a pseudo symmetry axis [114].

Experimental studies have confirmed that the Vpu protein can indeed generate a weakly cation selective ion channel activity [51, 181] with so far unknown biological function [51, 181]. It was previously shown that ion channel activity is not required for down-regulation of tetherin from the cell surface [25, 181]. This once again raised the question of whether ion channel function of Vpu has any functional significance [25, 193]. However, it was recently shown that the novel antiviral drug BIT225 (*N*-carbamimidoyl-5-(1-methyl-1*H*-pyrazol-4-yl)-2-naphthamide) inhibits Vpu generated ion channel activity. This block of Vpu channel activity can be correlated with a BIT225 mediated inhibition of HIV-1 replication in myeloid dendritic cells suggesting that a block of channel activity could be used in therapy for limiting viral spread [50, 99]. A recent double-blind, placebo-controlled, randomized clinical phase 1b/2a study in 21 HIV-1-infected antiretroviral therapy (ART)-naive subjects, has shown that BIT225 treatment can indeed significantly reduce the viral burden in myeloid lineage cells and the level of monocyte activation

[100, 221]. These results gave the first indications that ion channel activity might play an important role in myeloid cells.

The recent analysis of many Vpu sequences has highlighted a considerable variability in this protein, placing it among the most highly variable proteins in the HIV-1 proteome [61]. Further studies have shown that this structural variability is causally related to the ability of different Vpus in down regulating CD4 and bone marrow stromal cell antigen 2 (BST2) proteins. Here we want to extend this study on examining the relationship between Vpu polymorphism and ion channel activity. The key question is whether channel function has been evolutionary conserved or whether it is just an epiphenomenon of the Vpu protein from M type HIV-1. To address this question, we examine the electrical properties of Vpu proteins from related immunodeficiency virus isolates. Vpu genes were originally only found in HIV-1 and SIVs from chimpanzees ($SIV_{cpz}$) but not in HIV-2. It is now well established that Vpus are also present in SIVs originating from other primates. The evolution of the Vpu proteins presumably reflects the evolution of SIV and HIV viruses. A detailed phylogenetic analysis of SIV/HIV [101] suggests that $SIV_{cpz}$ rose to the pandemic (M, main) and non-pandemic (O, outlier and N, non-M, non O) groups of HIV-1; also the related SIV's from gorillas ($SIV_{gor}$) and the closely related and only recently detected HIV-1 group P, which can be traced back to the same origin. The $SIV_{cpz}$ itself is presumably a product of a series of cross-species transmission and recombination events, which involved precursors of today's SIV from various monkeys, namely the greater spot-nosed ($SIV_{gsn}$), mona ($SIV_{mon}$), Dent's mona ($SIV_{den}$), and Mustached ($SIV_{mus}$) monkey. Hence, it is reasonable to assume that all vpu genes originate from a common ancestor of the $SIV_{gsn/mus/mon/den}$ linage of primate lentiviruses [101].

After analyzing, in a previous study, the conductive properties of a Vpu protein from an M type HIV-1 [25] we analyze here the ion channel activities of different Vpu proteins of HIV-1 and SIV. The Vpu proteins represent different HIV-1 groups (N, P) and $SIV_{cpz}$ and $SIV_{gsn}$ [171, 172]. Heterologous expression and electrophysiological characterization of the Vpu homologs clearly showed that all Vpus investigated generate comparable ion channel function in HEK293T cells. The results of these experiments imply a robust and evolutionary conserved ion channel activity, which suggests that a cation conductance may also have a conserved functional significance.

## 4.3 MATERIALS AND METHODS

### 4.3.1 *Bioinformatics*

For information theoretic calculations, we used a multiple sequence alignment of the Vpu protein family from the PFAM database (PFAM id PF00558) [53]. The full alignment contained 9232 sequences with an overall length of 168 positions including gaps. 9084 of these sequences are assigned to HIV whereas 23 sequences are assigned to the simian variant SIV. To investigate the conservation state of several residues in HIV Vpu, the HIV sequences of the alignment were further processed and used for the calculation of Shannon entropy. First, sequences with letters that do not encode natural amino acids (X, Y, Z, B, O, U) as well as se-

quences with less than 50 residues were deleted. Furthermore, positions with a gap content bigger than 60% were also deleted. This resulted in an alignment of 6947 sequences with 80 positions. Shannon entropy was calculated using the `R` package `BioPhysConnectoR` (version 1.6-10) [81]. The sequences of SIV Vpu were extracted, realigned with `Clustal W` (version 2.1), and processed the same way.

Selected Vpu sequences from HIVs and SIVs, which were previously examined for their ability to antagonize tetherin function [172], were aligned with `Clustal W` [200] using default parameters and afterwards manually optimized in `Jalview` (version 2.0) [215].

### 4.3.2 *Heterologous Expression of Vpus*

Human embryonic kidney 293 T cells (HEK293T) (DSMZ, Braunschweig, Germany) were grown at 37 °C in a humidified 95% air/5% $CO_2$ incubator in Dulbecco's Modified Eagle Medium (DMEM; SIGMA-Aldrich, Taufkirchen, Germany) supplemented with 10% v/v heat-inactivated fetal bovine serum, 100 U/mL penicillin G and 100 g/mL streptomycin sulfate (all from Invitrogen, Carlsbad, CA, USA). Cells were passaged after reaching 70%–80% confluence every 2–3 days.

Vpu variants from HIVs and SIVs, which were extensively described before [171, 172], were a generous gift from Dr. Sauter (University of Ulm, Ulm, Germany). The Vpu genes were inserted as reported previously [171, 172] in the pCG-IRES-GFP vector, which is expressing AU1-tagged Vpu together with green fluorescent protein (GFP) from bicistronic mRNA.

For expression, HEK293T cells were transiently transfected with aforementioned bicistronic vector using the liposomal transfection reagent TurboFect™ (Fermentas, Waltham, MA, USA). For all constructs we generally found a transfection efficiency between 10% and 20% judging from the green fluorescence (Figure A.16). Twenty four hours post transfection cells were washed with phosphate-buffered saline (PBS), dispersed with trypsin (SIGMA-Aldrich) and seeded into new culture dishes with lower density. For patch clamp recordings, only isolated and adherent cells were considered. This ensures that recordings are from intact cells and that the currents reflect the conductance of a single cell of interest.

The expression of AU1 tagged Vpu proteins in HEK293T cells was tested by Western blotting. HEK293T cells were 48 h after transfection pelleted and subsequently lyzed. Vpu was detected by a monoclonal AU1 antibody (Covance, Munich, Germany); an anti-mouse immunoglobulin (Ig)-alkaline phosphatase antibody served as secondary antibody. Immune complexes were washed with wash buffer and separated in a 15% sodium dodecyl sulfate (SDS) gel.

### 4.3.3 *Electrophysiological Characterization*

For the electrophysiological measurements, cells were transferred into a measuring chamber. The culture medium was subsequently removed and replaced by bath solution. The bath solution contained: 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, 5 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 50 mM KCl. The pH was adjusted to 7.4 with KOH. The osmolarity was adjusted to 330 mOsmol with mannitol. The pipette solution contained: 130 mM D-potassium-gluconic acid,

10 mM NaCl, 5 mM HEPES, 0.1 mM guanosine triphosphate (Na salt), 4 mM $CaCl_2$, 2 mM $MgCl_2$, 5 mM phosphocreatine, and 2 mM adenosine triphosphate (ATP, Na salt); the pH was adjusted to 7.4 with KOH, the osmolarity was adjusted to 330 mOsmol with mannitol. The measuring chamber was placed on an inverted epifluorescent microscope (Axiovert 100, Zeiss, Oberkochen, Germany) for patch clamp measurements. Cells were inspected under normal light for selecting isolated cells. Also, the fluorescence of cells was monitored by exiting GFP with blue light (390 ±10 nm) from a monochromator (Till Photonics, Munich, Germany) and observing fluorescence > 525 nm after passing the emitted light through a band pass filter (MF525-39, ThorLabs, Munich, Germany). Only green fluorescent cells were used for patch clamp measurements.

Whole-cell patch-clamp recordings were performed at room temperature using standard methods [73] with an EPC-9 patch-clamp amplifier (HEKA, Lamprecht, Germany). Cells were clamped from a holding voltage (0 mV) to test voltages between +80 and -160 mV and back to a post potential (-80 mV). The data were acquired and analyzed with the software (version 2*73, HEKA, Lamprecht, Germany). The number of cells, which were measured for each construct, is reported in Figure 4.3. Each construct was measured from $\geq$ 2 independent transfections.

## 4.4 RESULTS

### 4.4.1 *Sequence Variability of Vpu Proteins*

At first, the sequence variability of Vpu proteins of HIV-1 and SIV was analyzed. For this purpose, we used the sequence of the HIV-1 NL4-3 Vpu [2], to browse for an alignment of the Vpu protein family within the PFAM database release 28.0 [53] and extracted 6947 sequences for an optimized alignment. Figure 4.1 A shows the 25 most similar sequences. The gray bars on the top quantify the Shannon entropy for each position (column) of the full alignment. Since this entropy value is the average of variability contained in each column, low values indicate strong conservation of amino acids and high values indicate a greater variability - and thus less selective pressure on the respective position - among all known sequences.

The channel-forming region of the Vpu protein is in the N-terminal domain [130, 136] between amino acid residues 4–26. In this region, we find in the alignment of 6947 sequences six positions with entropy values < 0.5 bit; three of these have entropy values < 0.25 bit. Interesting to note is that the Ser23, an amino acid, which was previously identified as crucial for channel function HIV-1 NL4-3 Vpu [25, 136], is not among the highly conserved amino acids. Notably, some of the highly conserved positions are less conserved in the Vpu sequences from non-pandemic HIV strains (Figure 4.1 B; Figure A.17). It is possible that these deviations could be related to the functional differences of these Vpus. It was found that Vpus from the nonpandemic HIV-1 O strains exhibit only poor counteracting activity against tetherin while those from the rare group N viruses are not able to degrade CD4. From these data, it was deduced that only HIV-1 M evolved a fully functional Vpu during three independent cross-species transmissions [171, 172].

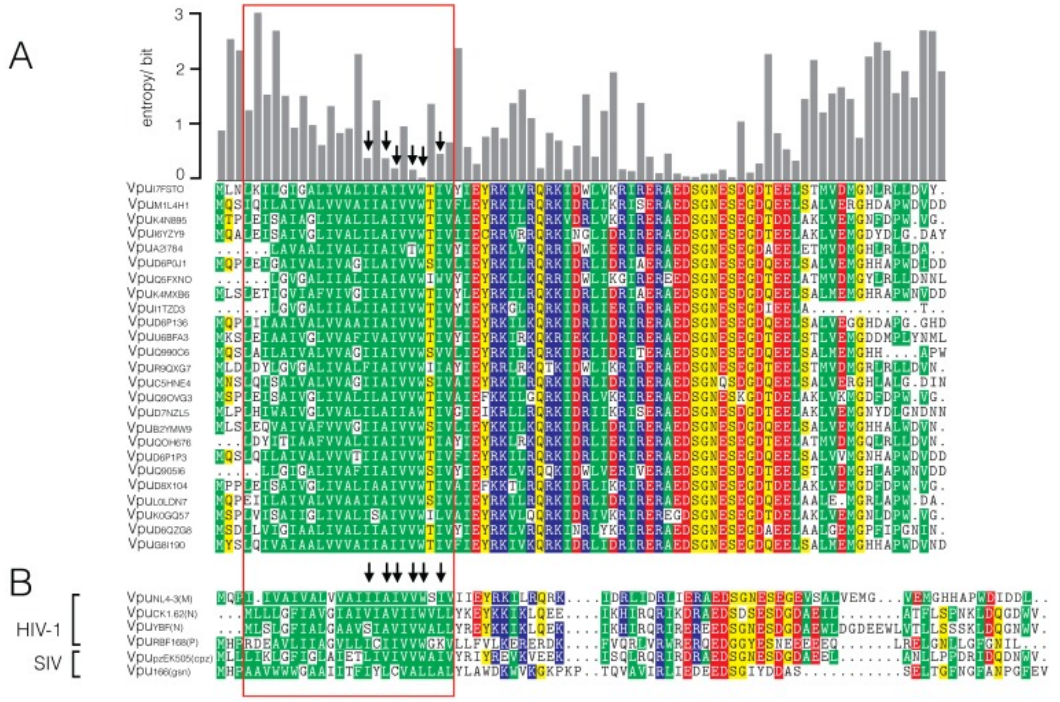


Figure 4.1: Sequence variability of Vpu proteins. (**A**) Alignment of the 25 most similar sequences out of 6947 sequences analyzed. Red frame indicates position of the transmembrane domain. Gray bars on the top give Shannon entropy for each column position from alignment of 6947 sequences. Arrows indicate amino acid positions with low entropy values (< 0.5 bit) indicate strong conservation and hence high selective pressure among all known sequences; (**B**) Alignment of Vpu proteins of human immunodeficiency virus type 1 (HIV-1) and simian immunodeficiency virus (SIV), which were tested for channel function. They include Vpu proteins from different HIV-1 groups namely M group (VpuNL4-3 [AAB60577.1]), N group (VpuYBF30 [O91085.1], VpuCK1.62 [ACX46167.1]), and P group (VpuRBF168 [ACT66828.1]) as well as SIVs from chimpanzees ($SIV_{cpz}$ [ABD19498.1]) and greater spot-nosed monkeys ($SIV_{gsn}$ [AAM90236.1]); the accession numbers for the Vpus are given in square brackets. Colors in alignment indicate acidic (red), basic (blue), polar uncharged (yellow), and hydrophobic nonpolar (green) amino acids.

Each of the selected sequences in Figure 4.1 differs at only few positions within the transmembrane domain (TMD) from the query sequence. Among all of the 6947 HIV sequences, the same amino acid substitutions can also be found in other Vpu homologs but with very low probability. In the Vpu homologs from SIVs however, the respective amino acids are more frequently found in the corresponding positions (Figure A.16). The results of this analysis suggest that the high degree of conservation in parts of the TMD is not essential for the ion channel activity but may be required for other functional properties of Vpu. Table 4.1 shows the distribution (relative frequencies) of the most frequent amino acids of the highly-conserved positions of the HIV alignment — for SIV and HIV in comparison.

Table 4.1: Most frequent amino acids at highly conserved positions in the transmembrane domain of Vpu from human immunodeficiency virus type 1 (HIV) and simian immunodeficiency virus (SIV). Data are given as relative frequencies in % of the respective alignment (HIV vs. SIV) *.

| Position | 17 | | | 19 | | | 20 | | 22 | | | 23 | 25 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HIV | I | | | A | | | I | | V | | | W | I | |
| | 94.8 | | | 94.4 | | | 97.3 | | 97.9 | | | 99.6 | 93.5 | |
| SIV | I | L | A | V | A,N | T | I | V | V | I | A | W | K | I |
| | 36.4 | 22.7 | 22.7 | 27.3 | 22.7 | 18.2 | 63.6 | 27.3 | 40.9 | 27.3 | 18.2 | 95.5 | 40.9 | 27.3 |

* Note, that in all positions an almost perfect conservation (>90%) in HIV is reduced to still noticeable abundance (~20–60%).

### 4.4.2 *Various Vpu Proteins from Human Immunodeficiency Virus (HIV) and Simian Immunodeficiency Virus (SIV) Generate Channel Function*

Against the background of apparent sequence variability among Vpus (Figure 4.1) and an apparent sensitivity of channel function to the fold of the transmembrane domain [25, 136] we examined the channel activities of Vpus from HIV-1 and SIV. Candidate Vpu proteins were expressed in HEK293T cells and their electrical activity measured in positively transfected - i.e., green fluorescent - cells by patch clamp. The data in Figure 4.2 A confirm that all Vpu constructs were indeed expressed in HEK293T cells. The level of expression among the different constructs was very variable between different experiments. Since we also found different efficiencies in transfection between different experiments and between different constructs, it remains unclear whether the constructs of interest have different levels of expression in HEK293T cells.

As a reference, we first monitored the currents of control cells, which were mock-transfected with GFP only. Figure 4.2 B shows the current responses of a typical control cell to voltage steps from a holding voltage at 0 mV to test voltages between +80 mV and -160 mV in a bath solution with 50 mM KCl. The recorded currents and the corresponding current/voltage (I/V) relation shown in Figure 4.2 B are characteristic for HEK293T cells [199]. They show very small negative currents and only slightly larger positive currents.

As a result of the low conductance at negative voltages, the I/V curve is hardly distinguishable from the voltage axis (Figure 4.2 B). Because of an endogenous outward rectifying channel in these cells, the I/V relation exhibits a small slope at positive voltages. For further analysis, the I/V relations of control cells can be best quantified by the ratio of currents at +80 mV versus that at -140 mV ($I_{+80}/I_{-140}$). In $n = 9$ mock-transfected cells we measured under the same experimental conditions a mean ratio $I_{+80}/I_{-140}$ of 7.0 ±1.4 (Figure 4.3 A).

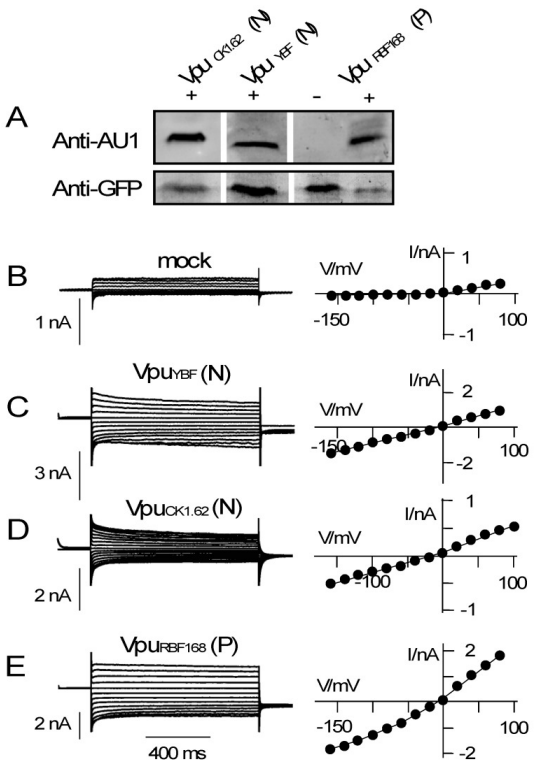


Figure 4.2: Conductive properties of Vpus from different HIVs. (**A**) Expression of different Vpu proteins in HEK293T cells was analyzed after cell lyses by Western blot, using anti-AU1 antibody. Successful transfection of cells was verified by anti-GFP antibody. The three Vpu proteins of interest comprised an AU1 tag (+). VpuRBF168 without AU1 tag was used as negative control (-). Example current responses (**left**) and corresponding I/V relations (**right**) of HEK293T cells transfected with either GFP alone (mock **B**) or with bicistronic vector for GFP plus VpuYBF (**C**), VpuCK1.62 (**D**), or VpuRBF168 (**E**). Letter in brackets indicates group of HIV-1 virus from which Vpu originates.

Currents of cells, which express a Vpu protein from non-pandemic HIV-1, namely of the N and P group, are different from those of control cells. Figure 4.2 C–E shows representative current recordings and I/V curves of HEK293T cells, which were transfected with the respective HIV-1 Vpus. The transfected cells are typically exhibited in response to the same voltage clamp steps, which were also used for the controls, much larger positive and negative currents than the controls. While the absolute current density was variable between different cells and different constructs (e.g., Figure 4.2 B, C) the additional currents rendered the I/V relations approximately linear over the window of test voltages (Figure 4.2 C–E). The same overall increase in membrane currents with a near linear I/V relation was confirmed in other HEK293T cells, which express one of the aforementioned Vpus. The linearity of the I/V curve is reflected in the $I_{+80}/I_{-140}$ ratio, which is in average for all Vpus $\leq 1.2$ (Figure 4.3 A). A comparison of the present results with similar measurements in which a Vpu from an M group HIV-1 ($Vpu_{NL4-3}$) was expressed in HEK293T cells exhibits no appreciable differences. Also, $Vpu_{NL4-3}$ generates an elevated inward current with a near linear I/V relation after heterologous expression in HEK293T cells (Figure 4.3 A [25]).

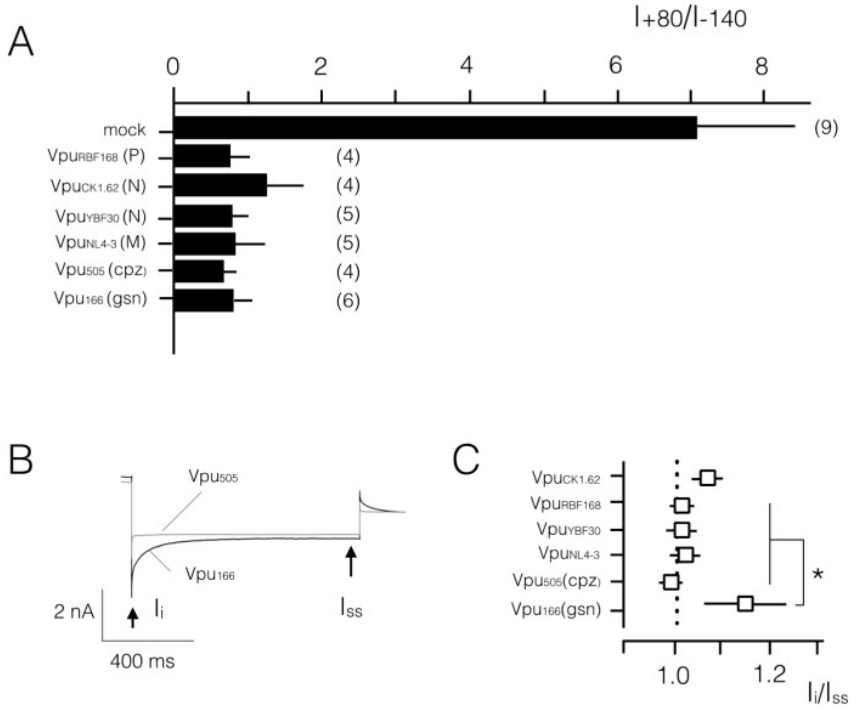


Figure 4.3: Analysis of current/voltage properties of Vpus from different HIVs and SIVs. (**A**) Mean ratios ($I_{+80}/I_{-140}$) with standard deviations (SD) of currents at +80 mV versus currents at -140 mV of mock transfected HEK293T cells and cells transfected with Vpu homologs. The high ratio indicates that mock-transfected cells exhibit an outward rectifying I/V relation while a value around 1 shows that the expression of all Vpu homologs generates an approximately linear I/V relation; (**B**) Representative current responses of HEK293T cells transfected with either $Vpu_{cpz}$ or $Vpu_{gsn}$ to voltage step from 0 mV to -160 mV. While the former exhibits no apparent time dependency, the latter decreases with time; (**C**) Kinetics of current responses to negative voltage steps in cells expressing different Vpus is quantified by ratio ($I_i/I_{ss}$) of current at start of voltage pulse ($I_i$) divided by steady state current ($I_{ss}$) at end of pulse. Numbers in brackets in **A** report the number of measured cells; the same data were used for the analysis in **B** and **C**. A Student's *t*-test shows that the data are different between mock transfected cells and Vpu expressing cells in **A** ($p < 0.0005$) and between indicated Vpu constructs and $Vpu_{166}$(gsn) in **C** ($p < 0.05$, *). Data on the $Vpu_{NL4-3}$ generated increase in membrane conductance in HEK293T cells are re-plotted from [25] for comparison with other Vpus from HIV1 or SIVs.

Our data show that all Vpu proteins of HIV-1 generated irrespective of the sequence variations in the transmembrane domain a current. These currents must reflect a channel activity of the tested Vpus in the plasma membrane of the HEK293T cells. This interpretation is supported by the fact that a Vpu mutant, which contains a fully randomized TM sequence or a critical single point mutation, do not generate any currents in HEK293T cells [25]. Furthermore, there are to the best of our knowledge no endogenous currents in HEK293T cells, which resemble those recorded in cells expressing different Vpus [208]. In particular, the TASK channel, which was suggested as an interaction partner of Vpu [82] is an outward rectifier [174] and not voltage independent like those recorded here. Finally, the

Vpu generated conductance in HEK293T cells, which is shown in Figure 4.3, is also in good agreement with measurements in *Xenopus* oocytes. In this alternative expression system Vpu generates a similar conductance with a quasi-linear I/V relation [173]. The diversity of endogenous channels in HEK293T cells and *Xenopus* oocytes [208, 216] makes it unlikely that the conductances in Vpu expressing cells are caused by an upregulation of endogenous channels.

The apparent channel activity of different Vpus from HIV-1 and the phylogenetic relationship between HIV and SIV further suggest that also the Vpus from SIVs should generate a channel conductance. To test this hypothesis, we expressed Vpus from related SIVs namely, $SIV_{cpz}$ and $SIV_{gsn}$, in HEK 293T cells (Figure 4.4). In terms of evolution, the latter can be considered the oldest Vpu allele among those considered here [59].

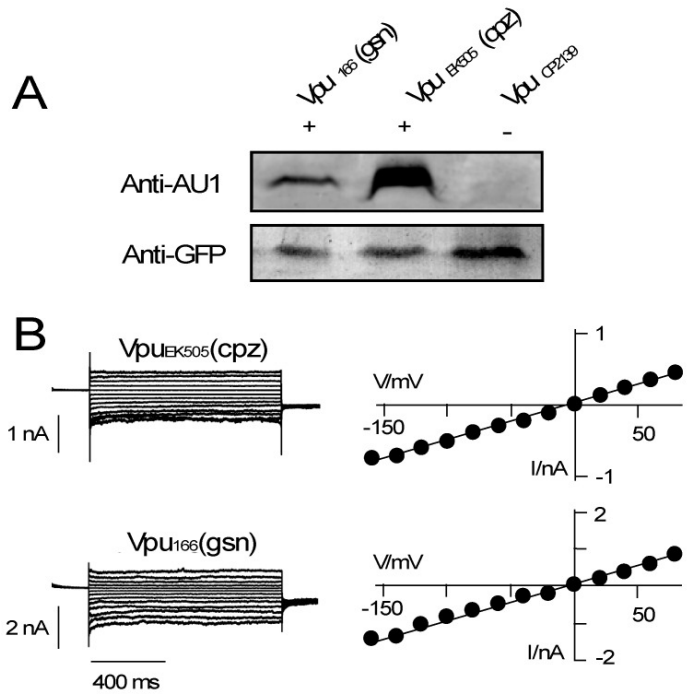


Figure 4.4: Conductive properties of Vpus from two SIVs. (**A**) Expression of different Vpu proteins in HEK293T cells was analyzed by anti-AU1 and anti-GFP antibodies as in Figure 4.1. The Vpu proteins of interest ($Vpu_{EK505}$, $Vpu_{166}$) comprised an AU1 tag (+). $Vpu_{CP2139}$ without AU1 tag was used as negative control (-); (**B**) Example current responses (left panel) and corresponding I/V relations (right panel) of HEK293T cells transfected with either, $Vpu_{EK505}$ from chimpanzee (cpz), or $Vpu_{CK1.62}$ from greater spot-nosed monkey (gsn).

The Western blots in Figure 4.4 A demonstrate that the Vpus from the two SIVs were indeed expressed in HEK293T cells. In the electrophysiological assay both Vpu proteins again generated a significant inward current (Figure 4.4 B), which was larger than that measured in control HEK293T cells (Figure 4.2 A). A plot of the mean $I_{+80}/I_{-40}$ ratios shows that also the Vpus from the primate SIVs generated a quasi-linear I/V relation in HEK293T cells (Figure 4.3 A). Both Vpus from $SIV_{cpz}$ and $SIV_{gsn}$ generate in HEK293T cells roughly the same current density (Figure 4.4 C).

Currents from all Vpus showed an overall similar phenotype but with some subtle variations. One subtle difference between the currents is visible in their kinetics. A scrutiny of the current responses shows in some cases a slight current inactivation at negative voltages (e.g., Figure 4.2 C, D and Figure 4.4 B). This is not consistent among all cells transfected with the same Vpu variant though (Figure 4.3 B). Only in the case of the $Vpu_{gsn}$ protein we find a robust difference between the initial ($I_i$) and steady state current ($I_{ss}$) (Figure 4.3 B). These data imply that the respective $Vpu_{gsn}$ undergoes an inherent inactivation at negative voltages.

## 4.5 DISCUSSION

The key question of the present study was to examine whether the ion channel function, which has been found for the Vpu from M type HIV-1 [25, 181], has been conserved through the evolution of this protein. The data show that Vpu proteins from different origins contain some highly conserved amino acid positions in the transmembrane domain, e.g., the domain, which is responsible for channel formation. These amino acids are not fully conserved in Vpu homologs from non-pandemic HIVs or Vpus from SIVs. The most highly conserved Trp23 for example is substituted by a Leu in $Vpu_{gsn}$. A functional testing of the homologs however shows that all the tested Vpu homologs generated a current in HEK293T cells; channel activity is maintained irrespective of sequence variations in the TMD. Already $Vpu_{gsn}$, the evolutionary oldest of the proteins tested here [59], generates channel function. Collectively, this implies that channel formation is an inherent and evolutionary old property of the Vpu protein. Negative effect of mutations on channel function like an exchange of Ser in position 23 in $Vpu_{NL4\text{-}3}$ [25, 136] must have been compensated by other mutations in the protein in order to maintain channel activity.

The present data, which were collected by overexpressing Vpus in a model mammalian cell, do not provide any answer on the functional significance of the channel activity of the Vpu proteins in virus replication. However, the fact that this function has been maintained throughout evolution of different HIV and SIV viruses implies an evolutionary pressure and hence a function of channel activity at one stage of replication.

While the data show that different Vpus function as channels, the sequence variations between the proteins are not entirely insignificant for channel function. They seem to affect in some cases the kinetics of the channel in a heterologous expression system. Apart from these small differences in the kinetics, the data show no striking difference between the Vpus from various HIVs and SIVs. In this sense, the present results also underscore previous data, which have shown that the ability of Vpu to degrade CD4 and antagonize tetherin is not connected to its ion channel function [25]. Here we show that ion channel activity is highly conserved among the Vpu proteins of HIV-1 and SIV and this function is independent of the previously tested ability of Vpu to counteract tetherin [172].

## ACKNOWLEDGEMENTS

We thank Sylvia Haase and Mirja Manthey for excellent technical assistance. We gratefully acknowledge financial support by German Israeli Foundation for Scientific Research and Development (grant #1086-13.11/2010 to GT), the Deutsche Forschungsgemeinschaft (HA 5261/3-1), the LOEWE project iNAPO of the Hessen State Ministry of Higher Education, Research and the Arts and grants SFB643-A1, IZKF grant A70, and the NIH grant RO1 DK81553 to US.

## AUTHOR CONTRIBUTIONS

T.G., S.B., and B.H. performed experiments and analyzed data; C.G. and K.H. contributed bioinformatics analysis; U.S., A.M., and G.T. conceived and designed the experiments and wrote the paper.

## SUPPORTING INFORMATION

The following are available online at www.mdpi.com/1999-4915/8/12/325/s1, Figure S1: Transfection of HEK293T cells Vpu from HIV-1, Figure S2: Alignment of Vpu orthologs from SIV with consensus sequence of Vpu from HIV-1.

In this thesis, the Supplementary Figures can be found in Appendix A3 as Figures A.16-A.17.

# 5 GENERAL DISCUSSION

## 5.1 THESIS SUMMARY

Synthetic biology aims at developing artificial biomolecules or entire biological systems with novel functionalities. It therefore combines biology and engineering [56, 98, 160]. In this context, computational biology also plays an essential role for supporting experimental approaches or providing insights, where experiments come to their limits. In the scope of this thesis, the functionalities or current limitations of three proteins that are promising candidates in the context of synthetic biology were investigated via computational approaches. To this end, the mechanical connections of *Fs*C and the HCN1 channel were elucidated via coarse-grained elastic network models and the evolutionary conservation of ion channel formation of the Vpu protein was analyzed based on an MSA. Consequently, this thesis provided further insights into the functionalities and underlying mechanisms of the three proteins as prerequisite for being optimized or to serve as biological model.

Chapter 1 gave a short introduction into computational as well as synthetic biology. In addition, the three proteins that are focused on in this thesis, were introduced with respect to their intended applications in the context of synthetic biology and their current limitations. Subsequently, the research goals of this thesis were defined and it was explained how these goals should be addressed via computational methods. In this context, a brief overview of conceivable *in silico* methods was given. This overview faced which timescales the different methods are able to simulate and over which timescales typical biological movements of proteins, like loop motions or side-chain rotation, occur.

Chapter 2 presented the rational design approach of *Fs*C, that is a promising candidate for a sustainable solution for enzymatic plastic waste treatment on an industrial scale. This chapter comprised a mix of multiscale simulations in order to find out, why *Fs*C decreases its activity with ongoing hydrolysis of PET and how the enzyme activity could be improved toward degradation of synthetic polymers. In this context, the LRT null model, similar to that of a previous study [104], as extension for the LRT approach [84] was introduced. Subsequently, the benefits of the LRT null model were shown. Based on the presented findings, this chapter pointed out which changes in the protein structure may lead to an improved enzyme activity for PET as substrate.

Chapter 3 presented the investigations on the gating mechanism of the HCN1 channel, that serves as biological model and building blocks for the development of synthetic nanopores. Based on two recently published molecular structures of HCN1, this chapter analyzed how binding of a ligand (cAMP) is transmitted into further domains via mechanical connections and how this may facilitate voltage-dependent opening and closing of the ion channel. Therefore, the previously applied LRT null model was adjusted to work for homotetrameric ion channels. In

this context, a further application for the LRT null model compared to the previous study on *Fs*C was demonstrated. With the here presented findings insights into the complex allosteric gating mechanism of HCN1 were given, based on which further computational and experimental approaches can be planned and performed.

Chapter 4 presented the investigations on the viral channel forming protein Vpu, that may also serve as building block for the development of synthetic nanopores. As the formation of an ion conducting pore is not the main functionaltiy of Vpu, this chapter aimed at understanding the role of the ion channel activity in the context of virus replication and release. Therefore, the degree of evolutionary conservation of the ion channel formation was investigated via computing the Shannon entropy of the concerned residues based on an MSA. This was an important first step, to better understand this protein to being further considered as alternative building block for the central ion conducting pore in synthetic ion channels.

## RECAPITULATION OF THE RESEARCH GOALS

In Chapter 1, the research goals for each protein, that is addressed by this thesis, were defined. In the following these research goals will be recapitulated and it will be reviewed how effectively they have been fulfilled.

*To understand why* FsC *loses its activity during the process of PET degradation and to be able to overcome these current limitations for an industrial application, this thesis aimed at providing a simulation approach, that is able to (i) investigate the effects of increasing amounts of EG on the structure and dynamics of* FsC*, (ii) make a link between structure and dynamics of FsC and its activity and (iii) investigate whether and how the activity is affected. Based on the findings, this thesis aimed at providing an estimate or multiple estimates for mutants that do not exhibit the loss of activity during the process of PET degradation.*

To address these goals, a mix of atomistic MD simulations and coarse-grained LRT computations was used. The 100 ns long MD simulations, performed by Sven Jager (Computational Biology & Simulation Group, TU Darmstadt), were able to uncover accumulations of the cleavage product EG on the surface of *Fs*C which negatively influence the overall dynamics of *Fs*C and the flexibility of affected surface residues, especially near the active site. With the LRT computations a link between structure and dynamics of *Fs*C and its activity could be made. The LRT simulations showed which conformational changes during substrate binding are neccessary for a functional enzyme and in this context could prove the hypothesis of an induced-fit [123]. Interestingly, regions that require high flexibility during substrate conversion are predominantly reduced in their flexibility due to the EG accumulations. Therefore, it can be suggested to find mutants that allow for an increased flexibility near the active site. To find out which mutants this may be, a look at the preferred type of interaction between EG and the active site residues was helpful. In fact, especially in the active site mainly hydrophilic interactions could be detected. Thus, it can be concluded that the reduction of hydrogen-bonding partners of EG near the active site may help to improve the activity of

*Fs*C for synthetic substrates. Possible mutants should therefore exhibit residues with increased hydrophobicity near the active site.

*To understand the complex allosteric gating mechanism of HCN1, this thesis aimed at circumventing the problem of limited structure information by simulating the channel movements in the TMPC based on known movements in the CNBD. With these simulations it should be answered (i) how the conformational information is transmitted between the CNBD and the TMPC, (ii) how these conformational changes in the CNBD are related to the gating movements in the VSD and central ion pore, (iii) what are the movements in VSD (especially S4 domain) and central pore (especially S6 domain) and (iv) how cAMP binding may favor the opening of the channel gate. The findings should contribute to a possible mechanistical model for the gating process of HCN channels.*

To address these goals, the previously introduced LRT null model was applied to indirectly simulate conformational changes upon cAMP binding. With this method it was possible to determine an external force acting on the central C-linker between CNBD and TMPC that perfectly reproduced known conformational changes from other studies that occur in the CNBD and C-linker upon binding of cAMP. Therefore, it can be assumed that this method also reliably simulates conformational changes in other parts of HCN1, especially in the TMPC, that are caused by cAMP binding. Although changes in the membrane voltage, that are necessary for channel opening, could not be simulated, the presented results provide some insights into possible movements in the TMPC during the allosteric gating process. The first insight was that the transmembrane domains S1-S4 of the voltage sensor as well as the transmembrane domains S5-S6 of the central conducting pore show a tilting. Especially the observed tilting and not vertical displacement of the S4 domain strengthened the hypothesis of a possible water filled gating canal that changes the electrical field around the S4 domain upon changes in the membrane voltage [16]. The analyzed distances between critical residues in S4 with respect to reference residues in S3 and S5 also showed that cAMP binding indeed augments a space between the S4 and its surrounding helices. The next insight was that the A' helix of the C-linker moves closer to the S4-S5 linker upon binding of cAMP. This strengthened the hypothesis of some experimental studies, which also proposed such a movement to facilitate the activation of HCN channels [5, 42, 111, 156]. The most interesting insight was that the simulated rearrangements caused by cAMP binding lead to a progressive widening of the inner gate at the extracellular ends of the S6 domains. With these results, it was not yet possible to provide an entire mechanistical model for the allosteric gating of HCN channels, but the gained insights can contribute to a mechanistical model and provide a good base for further experimental and computational studies to complement the current knowledge about HCN channel gating.

*To understand the role of ion channel activity of the small Vpu protein in the context of virus replication and to further consider it as building block for synthetic ion channels, it should be investigated, whether the ion channel function is an evolutionary conserved property of Vpu proteins.*

To address this research goal, a mix of electrophysiological patch-clamp measurements, performed by Timo Greiner (Membrane Biophysics Group, TU Darmstadt), and computation of the Shannon entropy based on an MSA of homologous Vpu protein sequences from different origins was used. With the patch-clamp measurements the ion channel activity of several Vpu proteins from different HIV and SIV strains could be proved. With the computation of the Shannon entropy some highly conserved amino acid positions in the transmembrane domain of Vpu, which is responsible for formation of the ion channel, could be detected. Unfortunately, this study could still not provide any answer on the function of the channel activity of the Vpu proteins in the context of virus replication, but the high degree of evolutionary conservation of some amino acid positions implies an evolutionary pressure. Hence, the ion channel function must play a significant role at one stage of virus replication, even if not uncovered so far, and, thus, the Vpu protein can be further considered as building block for synthetic ion channels.

## 5.2 CONTRIBUTIONS AND DISCUSSION

After having recapitulated the above defined research goals and how they have been fulfilled, now consequently the main contributions of this thesis are pointed out and discussed.

*I: Proposal of mutants for an improved activity of FsC toward PET degradation*

As a mandatory requirement for the applicability in an industrial application for enzymatic plastic waste treatment, the reason for the reduced activity of *Fs*C toward PET degradation was found. In this context, the type of interaction between amino acids near the active site and the degradation product EG, which mainly restrains the neccessary flexibility for the hydrolysis of huge substrates, was identified. Based on these findings, the mutation of residues near the active site by amino acids with increased hydrophobicity was proposed to allow for the required flexibility in this region. Accordingly, mutants that fulfill this requirement should be tested in an experimental assay for activity measurements on relevant substrates (PET or derivatives) as possible candidates for improved activity toward PET degradation.

*II: Insights into the allosteric gating mechanism of HCN1 gating*

Studying the mechanical connections in HCN1 provided important insights into the allosteric gating mechanism of HCN1 channel, and thus, led to a better understanding of the interplay of different domains in a complex ion channel that serves as biological model and building blocks for synthetic nanopores.

While the mechanical model neglects the impact of the membrane on the ion channel protein and therefore does not simulate changes of the membrane voltage, it demonstrates which conformational changes are generally possible based on the mechanical connections within the protein. Despite these simplifications of the computational model, the obtained results collectively underscore a good agreement to existing experimental data. Furthermore, they provide insights into con-

formational changes for the rest of the ion channel protein, where no experimental data exist, to revisit the still open questions on HCN channel gating: Binding of cAMP in the CNBD is mechnically transmitted via the C-linker to the TMPC and enables a tilting like movement of the transmembrane domains S1-S6. These observations strengthen the theory of the formation of a water-filled gating-canal around S4, that changes the electrical field upon changes in membrane voltage [16]. This explains the opposite gating behavior of Kv channels and HCN channels. Furthermore, the results show that the conformational changes caused by cAMP binding favor the opening of the inner gate at the cytoplasmic ends of the S6 domain, which then actually takes place upon hyperpolarization of the membrane.

The significance of the presented results, and thus, the predictive power of the applied method is substantiated by an experimental study that was published contemporaneously to our study. Via enhanced patch-clamp techniques Sunkara et. al [196] found a turning momentum of the CNBDs in the HCN2 channel upon binding of cAMP. By using concatemers with one to four functional subunits, they found a concerted rotation of the tetrameric CNBD/C-Linker region at cAMP binding. With their experimental results, they substantiate the here presented results of an iris-like movement of the C-linker and CNBD region and further underline that the simulations show conformational changes which are triggered by binding of cAMP and which are mandatory, so that hyperpolarizing membrane voltages are able to open HCN channels. On the other hand, Sunkara et al. found that binding of all four ligands cannot replace hyperpolarizing voltages to open HCN channels. This further strengthens the fact, that the LRT simulations cannot show an opening of the HCN channel because of the missing membrane voltage. The applied LRT simulations can only provide insights into conformational changes upon binding of cAMP, and thus, they only show a tendency toward opening of the inner gate.

*III: Development of LRT null model*

For studying the mechanical connections in *Fs*C as well as in the HCN1 channel, an advanced model of the LRT method, the LRT null model, was developed based on the idea of an LRT null model in a previous study [104]. While LRT only simulates the perturbation of the ANM with an external force from one single direction [84], the LRT null model simultaneously simulates the perturbation of the ANM with external forces from random directions. In contrast to the previously developed LRT null model by Sabine Knorr in 2014 [104], in the here applied LRT null model the perturbation directions are afterwards clustered regarding the displacements they induce to a set of selected residues. This allows for reducing the perturbation directions to only few clusters with distinct influences on the protein movements, and hence, finding the perturbation direction (cluster), which is necessary to simulate known conformational changes in a protein, e.g. due to substrate or ligand binding. Subsequently, the conformational changes in other domains of the same protein, which are still unknown, can be studied by this model. For the application on *Fs*C as well as on HCN1, the LRT null model was implemented in `R` for monomers as well as for rotationally symmetric tetramers. The `R` package `LRTNullModel4` comprises the previously published `R` package `LRTNullModel`

for a monomer [66] with the adjusted functions for a tetramer. Software link: `http://www.cbs.tu-darmstadt.de/LRTNullModel4.zip`.

As mentioned above, the predictive power of this method has been proved by another study on HCN channels, which arrived at the same conclusions via experimental procedures [196]. This shows that this method is also suitable for studying conformational changes in membrane proteins, although neglecting the impacts of the membrane as also done in other reduced models [40]. To sum up, the `LRTNullModel4` can be used to study the conformational changes of any monomeric or rotationally symmetric homotetrameric protein that are induced by an external perturbation. At least the C$\alpha$-coordinates of the protein residues must be available as *.pdb* file. The only requirement is that either the direction of the external perturbation at one position in the protein is known, or that conformational changes in at least one domain of the protein are known, to be able to extract the realistic perturbation direction from a set of random perturbation directions.

*IV: Prove of evoluationary conservation of ion channel function in Vpu proteins*

Computing the Shannon entropy of amino acid positions of the Vpu protein based on an MSA of homologous Vpu sequences proved the evolutionary conservation of important residues in the pore forming region. Together with the results of patch-clamp experiments, where ion channel activity of all analyzed Vpu proteins could be detected, this proves evolutionary conservation of the ion channel function. Because of this evolutionary pressure on the ion channel function, it can be deduced that it must play a significant role at one step of virus replication independent of the ability of the Vpu protein to degrade $CD_4$ and counteract tetherin.

## 5.3 CONCLUSION AND FUTURE WORK

In this thesis, a mix of different *in silico* methods has been used to study the functions of the here focused proteins. The decision for atomistic simulations or the use of reduced models on a coarse-grained level is a trade-off between accuracy and computational costs or simulated timescales. While atomistic simulations, e.g. QM or MD simulations, provide a high degree of accuracy, they raise high computational costs or can only simulate very short timescales. On the other hand, reduced models on the coarse-grained level have a significant reduction in accuracy but also a significant increase in the timescales they are able to simulate.

For studying the interactions of *Fs*C with the PET cleavage product EG, atomistic MD simulations of 100 ns were used. The high accuracy and short-timescales were appropriate to find out which amino acids interact with EG and cause accumulations of the latter on the protein surface, especially near the active site. For studying the conformational changes in *Fs*C and HCN1 upon substrate or ligand binding, respectively, the decision fell on a coarse-grained model based on an ANM. The presented results show that the choice of LRT and the LRT null model was appropriate for answering nearly all of the stated research questions. In future work, mutants of *Fs*C that fulfil the here determined requirements should be

screened for enhanced activity for PET. The most promising mutants could then be tested for up-scaling for an industrial application of enzymatic PET degradation.

With respect to the HCN channels, the interplay of different domains during the complex gating process could be elucidated. Furthermore, we got an idea of which other domains than the here investigated, play an important role in the gating mechanism. In future research, further attention will be payed to the top of the TMDs and the interconnecting loops as these parts show a remarkably wide scope of motion in the LRT simulations. Furthermore, they have been found in recent studies to contribute to different aspects in the complex mechanism of HCN channel gating [23, 85, 191, 203, 204]. Further attention will also be payed to the N-terminal HCN domain, which has not been focused in the present study but also undergoes strong conformational changes in the LRT simulations. Last but not least, effects of cAMP binding on the selectivity filter will be investigated to complete the investigations on HCN channel gating.

Although changes in the membrane voltage could not be simulated in this study, it is conceivable that such voltages changes could be simulated indirectly. This might be done by mimicking the forces that act on the S4 domain when the supposed water-filled gating canal changes the electrical field around S4. For such a procedure, the effect of voltage changes on this gating canal has to be deeper understood, but then, conformational changes caused by hyperpolarizing voltages might additionally be simulated by LRT.

After having elucidated the interplay of all relevant channel domains, selected domains, e.g. the CNBD, could be replaced by, e.g. other ligand binding domains, that induce similar conformational changes in the rest of the protein as binding of cAMP did. Then, with this modified ion channel, gating could be regulated by other ligands than cAMP. As a results, a biosensor based on such a modified ion channel could be used to detect these other ligands in a test solution.

In this thesis, the investigations on the Vpu protein only rely on sequence information. The presented results underline the importance of the ion channel function for the virus and proves that the formation of an ion conducting pore is not just an epiphenomenon. The certainty of evoluationary conservation is fundamental for further considering the Vpu protein as constant structural and functional building block for synthetic nanopores. Nevertheless, we still have to better understand this small protein and its function. First, the assembly of the Vpu monomers to pore forming quaterny structures or patches has to be understood in detail. Investigations on this questions, which already have been done by Lin et al. [114], could not yet detect a single conclusive pathway of pore formation. Second, it has to be understood how gating in Vpu proteins occurs and whether and how the selectivity for different ions is allowed. Last but not least, it has to be investigated how parts of the Vpu protein could be coupled with other domains of ion channels with respect to building an artificial nanopore with novel gating properties. While for real ion channels, which have been studied more elaborately, the level of knowledge is clearly higher and the use in synthetic ion channels is more likely, the Vpu protein and other VCPs are still worth to be considered and studied in the context of synthetic biology. Therefore, LRT and the here introduced LRT null model could be used to study the mechanical connections within the Vpu protein or maybe later within an *in silico* assembly of the Vpu protein with other ion channel domains.

# BIBLIOGRAPHY


[1] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers." In: *SoftwareX* 1 (2015), pp. 19–25.

[2] A. Adachi, H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin. "Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone." In: *Journal of Virology* 59.2 (1986), pp. 284–291.

[3] O. S. Alimi, J. Farner Budarz, L. M. Hernandez, and N. Tufenkji. "Microplastics and nanoplastics in aquatic environments: aggregation, deposition, and enhanced contaminant transport." In: *Environmental Science & Technology* 52.4 (2018), pp. 1704–1724.

[4] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. "Basic local alignment search tool." In: *Journal of Molecular Biology* 215.3 (1990), pp. 403–410.

[5] T. K. Aman and W. N. Zagotta. "HCN Channel Gating Studied with tm-FRET and a Fluorescent Noncanonical Amino Acid." In: *Biophysical Journal* 112.3 (2017), 249a.

[6] J. G. Amar. "The Monte Carlo method in science and engineering." In: *Computing in Science & Engineering* 8.2 (2006), pp. 9–19.

[7] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. "Synthetic biology: new engineering rules for an emerging discipline." In: *Molecular Systems Biology* 2.1 (2006).

[8] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." In: *Biophysical Journal* 80.1 (2001), pp. 505–515.

[9] C. G. Avio, S. Gorbi, and F. Regoli. "Plastics and microplastics in the oceans: From emerging pollutants to emerged threat." In: *Marine Environmental Research* 128 (2017), pp. 2–11.

[10] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava. "Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins." In: *Chemical Reviews* 110.3 (2009), pp. 1463–1497.

[11] I. Bahar and A. J. Rader. "Coarse-grained normal mode analysis in structural biology." In: *Current Opinion in Structural Biology* 15.5 (2005), pp. 586–592.

[12] X.-Chen. Bai, G. McMullan, and S. H. W. Scheres. "How cryo-EM is revolutionizing structural biology." In: *Trends in Biochemical Sciences* 40.1 (2015), pp. 49–57.

[13] L. G. A. Barboza and B. C. G. Gimenez. "Microplastics in the marine environment: current trends and future perspectives." In: *Marine Pollution Bulletin* 97.1-2 (2015), pp. 5–12.

[14] D. Bauer. "MD simulations of voltage-gated potassium channels." Master's thesis. Teschnische Universität Darmstadt, Darmstadt, Germany, 2016.

[15] H. Bekker, H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. von Drumen, D. van der Spoel, A. Sijbers, H. Keegstra, and M. K. R Renardus. "GROMACS - A Parallel Computer For Molecular-Dynamics Simulations." In: *Physics Computing '92*. World Scientific Publishing, 1993, pp. 252–256.

[16] D. C. Bell, H. Yao, R. C. Saenger, J. H. Riley, and S. A. Siegelbaum. "Changes in local S4 environment provide a voltage-sensing mechanism for mammalian hyperpolarization–activated HCN channels." In: *The Journal of General Physiology* 123.1 (2004), pp. 5–20.

[17] S. A. Benner and A. M. Sismour. "Synthetic biology." In: *Nature Reviews Genetics* 6.7 (2005), pp. 533–543.

[18] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular Dynamics with Coupling to an External Bath." In: *Journal of Chemical Physics* 81.8 (1984), pp. 3684–3690.

[19] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation." In: *Computer Physics Communications* 91.1 (1995), pp. 43–56.

[20] C. Berens and B. Suess. "Riboswitch engineering—making the all-important second and third steps." In: *Current Opinion in Biotechnology* 31 (2015), pp. 10–15.

[21] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. "The protein data bank." In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242.

[22] M. Biel, C. Wahl-Schott, S. Michalakis, and X. Zong. "Hyperpolarization-activated cation channels: from genes to function." In: *Physiological Reviews* 89.3 (2009), pp. 847–885.

[23] S. Biel, M. Aquila, B. Hertel, A. Berthold, T. Neumann, D. DiFrancesco, A. Moroni, G. Thiel, and S. Kauferstein. "Mutation in S6 domain of HCN4 channel in patient with suspected Brugada syndrome modifies channel function." In: *Pflügers Archiv-European Journal of Physiology* 468.10 (2016), pp. 1663–1671.

[24] J. Binette, M. Dubé, J. Mercier, D. Halawani, M. Latterich, and E. A. Cohen. "Requirements for the selective degradation of CD4 receptor molecules by the human immunodeficiency virus type 1 Vpu protein in the endoplasmic reticulum." In: *Retrovirology* 4.1 (2007), p. 75.

[25] S. Bolduan, J. Votteler, V. Lodermeyer, T. Greiner, H. Koppensteiner, M. Schindler, G. Thiel, and U. Schubert. "Ion channel activity of HIV-1 Vpu is dispensible for counteraction of CD317." In: *Virology* 416.1-2 (2011), pp. 75–85.

[26] J. E. Bronson, W. W. Mazur, and V. W. Cornish. "Transcription factor logic using chemical complementation." In: *Molecular BioSystems* 4.1 (2007), pp. 56–58.

[27] H. P. J. Buermans and J. T. Den Dunnen. "Next generation sequencing technology: advances and applications." In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1842.10 (2014), pp. 1932–1941.

[28] O. Buß, S. Jager, S.-M. Dold, S. Zimmermann, K. Hamacher, K. Schmitz, and J. Rudat. "Statistical Evaluation of HTS Assays for Enzymatic Hydrolysis of $\beta$-Keto Esters." In: *PloS One* 11.1 (2016), e0146104.

[29] I. Callebaut, B. Hoffmann, P. Lehn, and J.-P. Mornon. "Molecular modelling and molecular dynamics of CFTR." In: *Cellular and Molecular Life Sciences* 74.1 (2017), pp. 3–22.

[30] L. Cavallo, J. Kleinjung, and F. Fraternali. "POPS: A Fast Algorithm for Solvent Accessible Surface Areas at Atomic and Residue Level." In: *Nucleic Acids Research* 31.13 (2003), pp. 3364–3366.

[31] S. Chen, L. Su, J. Chen, and J. Wu. "Cutinase: Characteristics, Preparation, and Application." In: *Biotechnology Advances* 31.8 (2013), pp. 1754–1767.

[32] S. Chen, J. Wang, L. Zhou, M. S. George, and S. A. Siegelbaum. "Voltage Sensor Movement and cAMP Binding Allosterically Regulate an Inherently Voltage-independent Closed- Open Transition in HCN Channels." In: *The Journal of General Physiology* 129.2 (2007), pp. 175–188.

[33] W. S. Cleveland. "Robust Locally Weighted Regression and Smoothing Scatterplots." In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836.

[34] E. A. Cohen, E. F. Terwilliger, J. G. Sodroski, and W. A. Haseltine. "Identification of a protein encoded by the vpu gene of HIV-1." In: *Nature* 106 (1988), pp. 532–534.

[35] K. B. Craven, N. B. Olivier, and W. N. Zagotta. "C-terminal movement during gating in cyclic nucleotide-modulated channels." In: *Journal of Biological Chemistry* 283.21 (2008), pp. 14728–14738.

[36] K. B. Craven and W. N. Zagotta. "Salt bridges and gating in the COOH-terminal region of HCN2 and CNGA1 channels." In: *The Journal of General Physiology* 124.6 (2004), pp. 663–677.

[37] L. D. Creveld, A. Amadei, R. C. van Schaik, H. A. M. Pepermans, J. de Vlieg, and H. J. C. Berendsen. "Identification of Functional and Unfolding Motions of Cutinase as Obtained from Molecular Dynamics Computer Simulations." In: *Proteins: Structure, Function and Genetics* 33.2 (1998), pp. 253–264.

[38] L. G. Cuello, J. G. Romero, D. M. Cortes, and E. Perozo. "pH-dependent gating in the Streptomyces lividans K+ channel." In: *Biochemistry* 37.10 (1998), pp. 3229–3236.

[39] K. C. Cunha, V. H. Rusu, I. F. T. Viana, E. T. A. Marques, R. Dhalia, and R. D. Lins. "Assessing protein conformational sampling and structural stability via de novo design and molecular dynamics simulations." In: *Biopolymers* 103.6 (2015), pp. 351–361.

[40] A. Das, M. Gur, M. H. Cheng, S. Jo, I. Bahar, and B. Roux. "Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model." In: *PLoS Computational Biology* 10.4 (2014), pp. 1–17.

[41] T. L. Deans, C. R. Cantor, and J. J. Collins. "A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells." In: *Cell* 130.2 (2007), pp. 363–372.

[42] N. Decher, J. Chen, and M. C. Sanguinetti. "Voltage-dependent gating of hyperpolarization-activated, cyclic nucleotide-gated pacemaker channels: molecular coupling between the S4–S5 and C-linkers." In: *Journal of Biological Chemistry* 279.14 (2004), pp. 13859–13865.

[43] Y. Dehouck and A. S. Mikhailov. "Effective harmonic potentials: insights into the internal cooperativity and sequence-specificity of protein dynamics." In: *PLoS Computational Biology* 9.8 (2013), pp. 1–11.

[44] D. DiFrancesco and P. Tortora. "Direct activation of cardiac pacemaker channels by intracellular cyclic AMP." In: *Nature* 351.6322 (1991), pp. 145–147.

[45] J. A. Doudna and E. Charpentier. "The new frontier of genome engineering with CRISPR-Cas9." In: *Science* 346.6213 (2014), p. 1258096.

[46] Y. Duan et al. "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations." In: *Journal of Computational Chemistry* 24.16 (2003), pp. 1999–2012.

[47] R. C. Edgar. "MUSCLE: multiple sequence alignment with high accuracy and high throughput." In: *Nucleic Acids Research* 32.5 (2004), pp. 1792–1797.

[48] W. Ensinger. "Ion Conducting Nanopores for (Bio) molecular Sensing: the iNAPO Project." In: *Proceedings of the 2nd World Congress on New Technologies, Budapest, Hungary, August 2016*.

[49] J. R. Errington and P. G. Debenedetti. "Relationship between Structural Order and the Anomalies of Liquid Water." In: *Nature* 409 (2001), pp. 318–321.

[50] G. D. Ewart, N. Nasr, H. Naif, G. B. Cox, A. L. Cunningham, and P. W. Gage. "Potential new anti-human immunodeficiency virus type 1 compounds depress virus replication in cultured human macrophages." In: *Antimicrobial Agents and Chemotherapy* 48.6 (2004), pp. 2325–2330.

[51] G. D. Ewart, T. Sutherland, P. W. Gage, and G. B. Cox. "The Vpu protein of human immunodeficiency virus type 1 forms cation-selective ion channels." In: *Journal of Virology* 70.10 (1996), pp. 7108–7115.

[52] R. B. Fenwick, S. Esteban-Martın, and X. Salvatella. "Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles." In: *European Biophysics Journal* 40.12 (2011), pp. 1339–1355.

[53] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, et al. "Pfam: the protein families database." In: *Nucleic Acids Research* 42.D1 (2013), pp. D222–D230.

[54] W. B. Fischer, Y.-T. Wang, C. Schindler, and C.-P. Chen. "Mechanism of function of viral channel proteins and implications for drug development." In: *International Review of Cell and Molecular Biology*. Vol. 294. Elsevier, 2012, pp. 259–321.

[55] A. Fiser, R. K. G. Do, and A. Sali. "Modeling of loops in protein structures." In: *Protein Science* 9.9 (2000), pp. 1753–1773.

[56] J. L. Foo, C. B. Ching, M. W. Chang, and S. S. J. Leong. "The imminent role of protein engineering in synthetic biology." In: *Biotechnology Advances* 30.3 (2012), pp. 541–549.

[57] J. Friborg, A. Ladha, H. Göttlinger, W. A. Haseltine, and Cohen E. A. "Functional analysis of the phosphorylation sites on the human immunodeficiency virus type 1 Vpu protein." In: *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* 8.1 (1995), pp. 10–22.

[58] B. Geyer, G. Lorenz, and A. Kandelbauer. "Recycling of Poly(ethylene terephthalate)–A Review Focusing on Chemical Methods." In: *Express Polymer Letters* 10.7 (2016), pp. 559–586.

[59] R. J. Gifford, A. Katzourakis, M. Tristem, O. G. Pybus, M. Winters, and R. W. Shafer. "A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution." In: *Proceedings of the National Academy of Sciences* 105.51 (2008), pp. 20362–20367.

[60] G. B. Gloor, L. C. Martin, L. M. Wahl, and S. D. Dunn. "Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions." In: *Biochemistry* 44.19 (2005), pp. 7156–7165.

[61] M. E. González. "Vpu Protein: The Viroporin Encoded by HIV-1." In: *Viruses* 7.8 (2015), pp. 4352–4368.

[62] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves. "Bio3d: An R Package for the Comparative Analysis of Protein Structures." In: *Bioinformatics* 22.21 (2006), pp. 2695–2696.

[63] M. R. Gregory and A. L. Andrady. "Plastics in the marine environment." In: *Plastics and the Environment* 379 (2003), pp. 389–390.

[64] T. Greiner, S. Bolduan, B. Hertel, C. Groß, K. Hamacher, U. Schubert, A. Moroni, and G. Thiel. "Ion channel activity of Vpu proteins is conserved throughout evolution of HIV-1 and SIV." In: *Viruses* 8.12 (2016), p. 325.

[65] F. Groher and B. Suess. "Synthetic riboswitches—a tool comes of age." In: *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1839.10 (2014), pp. 964–973.

[66] C. Groß, K. Hamacher, K. Schmitz, and S. Jager. "Cleavage product accumulation decreases the activity of cutinase during PET hydrolysis." In: *Journal of Chemical Information and Modeling* 57.2 (2017), pp. 243–255.

[67] C. Groß, A. Saponaro, B. Santoro, A. Moroni, G. Thiel, and K. Hamacher. "Mechanical transduction of cytoplasmic-to-transmembrane-domain movements in a hyperpolarization-activated cyclic nucleotide–gated cation channel." In: *Journal of Biological Chemistry* 293.33 (2018), pp. 12908–12918.

[68] K. Hamacher. "Coarse-grained molecular models for high-throughput and multi-scale functional investigations." In: *Chemistry Central Journal* 2.1 (2008), S14.

[69] K. Hamacher. "Relating sequence evolution of HIV1-protease to its underlying molecular mechanics." In: *Gene* 422.1 (2008), pp. 30–36.

[70] K. Hamacher. "Free Energy of Contact Formation in Proteins: Efficient Computation in the Elastic Network Approximation." In: *Physical Review E, Statistical, Nonlinear and Soft Matter Physics* 84.1 (2011), pp. 1–6.

[71] K. Hamacher and J. A. McCammon. "Computing the amino acid specificity of fluctuations in biomolecular systems." In: *Journal of Chemical Theory and Computation* 2.3 (2006), pp. 873–878.

[72] K. Hamacher, J. Trylska, and J. A. McCammon. "Dependency Map of Proteins in the Small Ribosomal Subunit." In: *PLoS Computational Biology* 2.2 (2006), pp. 80–87.

[73] O. P. Hamill, A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. "Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches." In: *Pflügers Archiv* 391.2 (1981), pp. 85–100.

[74] J. A. Hartigan and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 28.1 (1979), pp. 100–108.

[75] S. Hayward and B. L. De Groot. "Normal modes and essential dynamics." In: *Molecular Modeling of Proteins* (2008), pp. 89–106.

[76] S. Henikoff and J. G. Henikoff. "Amino acid substitution matrices from protein blocks." In: *Proceedings of the National Academy of Sciences* 89.22 (1992), pp. 10915–10919.

[77] S. Hennig, G. Rödel, and K. Ostermann. "Artificial cell-cell communication as an emerging tool in synthetic biology applications." In: *Journal of Biological Engineering* 9.1 (2015), p. 13.

[78] E. Herrero Acero et al. "Enzymatic Surface Hydrolysis of PET: Effect of Structural Diversity on Kinetic Properties of Cutinases from *Thermobifida*." In: *Macromolecules* 44.12 (2011), pp. 4632–4640.

[79] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation." In: *Journal of Chemical Theory and Computation* 4.3 (2008), pp. 435–447.

[80] B. Hille et al. *Ion channels of excitable membranes*. Vol. 507. Sinauer Sunderland, MA, 2001.

[81] F. Hoffgaard, P. Weil, and K. Hamacher. "BioPhysConnectoR: Connecting Sequence Information and Biophysical Models." In: *BMC Bioinformatics* 11 (2010), p. 199.

[82] K. Hsu, J. Seharaseyon, P. Dong, S. Bour, and E. Marbán. "Mutual functional destruction of HIV-1 Vpu and host TASK-1 channel." In: *Molecular Cell* 14.2 (2004), pp. 259–267.

[83] F. Hucho and C. Weise. "Ligand-gated ion channels." In: *Angewandte Chemie International Edition* 40.17 (2001), pp. 3100–3116.

[84] M. Ikeguchi, J. Ueno, M. Sato, and A. Kidera. "Protein Structural Change Upon Ligand Binding: Linear Response Theory." In: *Physical Review Letters* 94.7 (2005), p. 078102.

[85] T. M. Ishii, M. Takano, and H. Ohmori. "Determinants of activation kinetics in mammalian hyperpolarization-activated cation channels." In: *The Journal of Physiology* 537.1 (2001), pp. 93–100.

[86] S. Jager, B. Schiller, T. Strufe, and K. Hamacher. "StreAM-T_g: Algorithms for Analyzing Coarse Grained RNA Dynamics Based on Markov Models of Connectivity-Graphs." In: *International Workshop on Algorithms in Bioinformatics*. Springer. 2016, pp. 197–209.

[87] A. Jakalian, D. B. Jack, and C. I. Bayly. "Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation." In: *Journal of Computational Chemistry* 23.16 (2002), pp. 1623–1641.

[88] Z. M. James and W. N. Zagotta. "Structural insights into the mechanisms of CNBD channel function." In: *The Journal of General Physiology* (2017), jgp–201711898.

[89] L. Y. Jan and Y. N. Jan. "Voltage-sensitive ion channels." In: *Cell* 56.1 (1989), pp. 13–25.

[90] M. Ø. Jensen, V. Jogini, D. W. Borhani, A. E. Leffler, R. O. Dror, and D. E. Shaw. "Mechanism of Voltage Gating in Potassium Channels." In: *Science* 336.6078 (2012), pp. 229–233.

[91] J. P. Johnson Jr and W. N. Zagotta. "Rotational movement during cyclic nucleotide-gated channel opening." In: *Nature* 412.6850 (2001), pp. 917–921.

[92] W. Kabsch and C. Sander. "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." In: *Biopolymers* 22.12 (1983), pp. 2577–2637.

[93] M. H. Kalos and P. A. Whitlock. *Monte carlo methods*. John Wiley & Sons, 2008.

[94] M. Karplus and G. A. Petsko. "Molecular dynamics simulations in biology." In: *Nature* 347.6294 (1990), pp. 631–639.

[95] K. Katoh, K. Misawa, K. Kuma, and T. Miyata. "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform." In: *Nucleic Acids Research* 30.14 (2002), pp. 3059–3066.

[96] O. Keskin, I. Bahar, A. Y. Badretdinov, O. B. Ptitsyn, and R. L. Jernigan. "Empirical Solvent-Mediated Potentials Hold for Both Intra-Molecular and Inter-Molecular Inter-Residue Interactions." In: *Protein Science* 7 (1998), pp. 2578–2586.

[97] F. Keul, M. Hess, M. Goesele, and K. Hamacher. "PFASUM: a substitution matrix from Pfam structural alignments." In: *BMC Bioinformatics* 18.1 (2017), p. 293.

[98] A. S. Khalil and J. J. Collins. "Synthetic biology: applications come of age." In: *Nature Reviews Genetics* 11.5 (2010), p. 367.

[99] G. Khoury, G. Ewart, C. Luscombe, M. Miller, and J. Wilkinson. "Antiviral Efficacy of the Novel Compound BIT225 against HIV-1 Release from Human Macrophages." In: *Antimicrobial Agents and Chemotherapy* 54.2 (2010), pp. 835–845.

[100] G. Khoury, G. Ewart, C. Luscombe, M. Miller, and J. Wilkinson. "The antiviral compound BIT225 inhibits HIV-1 replication in myeloid dendritic cells." In: *AIDS Research and Therapy* 13.7 (2016).

[101] F. Kirchhoff. "Is the high virulence of HIV-1 an unfortunate coincidence of primate lentiviral evolution?" In: *Nature Reviews Microbiology* 7 (2009), pp. 467–476.

[102] F. Klameth. "From Brownian Motion to Supercooled Water in Confinements - A Molecular Dynamics Simulation Study." PhD thesis. Darmstadt, Germany: Technische Universität Darmstadt, 2015.

[103] A. Klamt. "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena." In: *Journal of Physical Chemistry* 99.7 (1995), pp. 2224–2235.

[104] S. Knorr. "In Silico Strategies to Modulate DNA Damage Response." PhD thesis. Darmstadt, Germany: Technische Universität Darmstadt, 2015.

[105] D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson, and E. R. Mardis. "The next-generation sequencing revolution and its impact on genomics." In: *Cell* 155.1 (2013), pp. 27–38.

[106] R. Koshti, L. Mehta, and N. Samarth. "Biological Recycling of Polyethylene Terephthalate: A Mini-Review." In: *Journal of Polymers and the Environment* (2018), pp. 1–10.

[107] E. Krieger and G. Vriend. "Increasing the Precision of Comparative Models with YASARA NOVA — a Self-Parameterizing Force Field." In: *Proteins: Structure, Function and Genetics* 47 (2002), pp. 393–402.

[108] S. Kullback and R. A. Leibler. "On Information and Sufficiency." In: *Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86.

[109] P. Kumara, S. V. Buldyrevb, and H. E. Stanleyc. "A Tetrahedral Entropy for Water." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.52 (2009), pp. 22130–22134.

[110] J. Kusch, C. Biskup, S. Thon, E. Schulz, V. Nache, T. Zimmer, F. Schwede, and K. Benndorf. "Interdependence of receptor activation and ligand binding in HCN2 pacemaker channels." In: *Neuron* 67.1 (2010), pp. 75–85.

[111] D. C. H. Kwan, D. L. Prole, and G. Yellen. "Structural changes during HCN channel gating defined by high affinity metal bridges." In: *The Journal of General Physiology* 140.3 (2012), pp. 279–291.

[112] C.-H. Lee and R. MacKinnon. "Structures of the Human HCN1 Hyperpolarization-Activated Channel." In: *Cell* 168.1 (2017), pp. 111–120.

[113] M. Li, X. Zhou, S. Wang, I. Michailidis, Y. Gong, D. Su, H. Li, X. Li, and J. Yang. "Structure of a eukaryotic cyclic-nucleotide-gated channel." In: *Nature* 542.7639 (2017), pp. 60–65.

[114] M. H. Lin, C. P. Chen, and W. B. Fischer. "Patch formation of a viral channel forming protein within a lipid membrane–Vpu of HIV-1." In: *Molecular BioSystems* 12.4 (2016), pp. 1118–1127.

[115] M.-H. Lin, C.-P. Chen, and W. B. Fischer. "Patch formation of a viral channel forming protein within a lipid membrane–Vpu of HIV-1." In: *Molecular BioSystems* 12.4 (2016), pp. 1118–1127.

[116] E. Lindahl, B. Hess, and D. van der Spoel. "GROMACS 3.0: a package for molecular simulation and trajectory analysis." In: *Molecular Modeling Annual* 7.8 (2001), pp. 306–317.

[117] M. Lolicato, A. Bucchi, C. Arrigoni, S. Zucca, M. Nardini, I. Schroeder, K. Simmons, M. Aquila, D. DiFrancesco, M. Bolognesi, et al. "Cyclic dinucleotides bind the C-linker of HCN4 to control channel cAMP responsiveness." In: *Nature Chemical Biology* 10.6 (2014), pp. 457–462.

[118] M. Lolicato, M. Nardini, S. Gazzarrini, S. Möller, D. Bertinetti, F. W. Herberg, M. Bolognesi, H. Martin, M. Fasolini, J. A. Bertrand, et al. "Tetramerization dynamics of C-terminal domain underlies isoform-specific cAMP gating in hyperpolarization-activated cyclic nucleotide-gated channels." In: *Journal of Biological Chemistry* 286.52 (2011), pp. 44811–44820.

[119] S. B. Long, E. B. Campbell, and R. MacKinnon. "Crystal structure of a mammalian voltage-dependent Shaker family K+ channel." In: *Science* 309.5736 (2005), pp. 897–903.

[120] S. B. Long, E. B. Campbell, and R. MacKinnon. "Voltage sensor of Kv1.2: structural basis of electromechanical coupling." In: *Science* 309.5736 (2005), pp. 903–908.

[121] S. B. Long, X. Tao, E. B. Campbell, and R. MacKinnon. "Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment." In: *Nature* 450.7168 (2007), pp. 376–382.

[122] S. Longhi and C. Cambillau. "Structure-Activity of Cutinase, a Small Lipolytic Enzyme." In: *Biochimica et Biophysica Acta* 1441.2-3 (1999), pp. 185–196.

[123] S. Longhi, M. Czjzek, V. Lamzin, A. Nicolas, and C. Cambillau. "Atomic Resolution (1.0 Å) Crystal Structure of *Fusarium solani* Cutinase: Stereochemical Analysis." In: *Journal of Molecular Biology* 268 (1997), pp. 779–799.

[124] U. Lüttge, M. Kluge, and G. Thiel. *Botanik: Die umfassende Biologie der Pflanzen*. Wiley-VCH, 2010.

[125] R. M. Lynden-Bell and P. G. Debenedetti. "Computational Investigation of Order, Structure, and Dynamics in Modified Water Models." In: *Journal of Physical Chemistry B* 109 (2005), pp. 6527–6534.

[126] R. MacKinnon. "Potassium channels." In: *FEBS Letters* 555.1 (2003), pp. 62–65.

[127] J. G. Magadán, F. J. Pérez-Victoria, R. Sougrat, Y. Ye, K. Strebel, and J. S. Bonifacino. "Multilayered Mechanism of CD4 Downregulation by HIV-1 Vpu Involving Distinct ER Retention and ERAD Targeting Steps." In: *PLOS Pathogens* 6.4 (2010), pp. 1–18.

[128] F. Maldarelli, M. Y. Chen, R. L. Willey, and K. Strebel. "Human immunodeficiency virus type 1 Vpu protein is an oligomeric type I integral membrane protein." In: *Journal of Virology* 67.8 (1993), pp. 5056–5061.

[129] R. Männikkö, F. Elinder, and H. P. Larsson. "Voltage-sensing mechanism is conserved among ion channels gated by opposite voltages." In: *Nature* 419.6909 (2002), pp. 837–841.

[130] F. M. Marassi, C. Ma, H. Gratkowski, S. K. Straus, K. Strebel, M. Oblatt-Montal, M. Montal, and S. J. Opella. "Correlation of the structural and functional domains in the membrane protein Vpu from HIV-1." In: *Proceedings of the National Academy of Sciences* 96.25 (1999), pp. 14336–14341.

[131] G. F. Marban E.and Tomaselli. "Ion channels as enzymes: analogy or homology?" In: *Trends in Neurosciences* 20.4 (1997), pp. 144–147.

[132] C. Martinez, P. De Geus, M. Lauwereys, G. Matthyssens, and C. Cambillau. "*Fusarium solani* Cutinase is a Lipolytic Enzyme with a Catalytic Serine Accessible to Solvent." In: *Nature* 356.6370 (1992), pp. 615–618.

[133] M. Y. Matak and M. E. Moghaddam. "The Role of Short-Range Cys171-Cys178 Disulfide Bond in Maintaining Cutinase Active Site Integrity: A Molecular Dynamics Simulation." In: *Biochemical and Biophysical Research Communications* 390.2 (2009), pp. 201–204.

[134] K. Mattsson, S. Jocic, I. Doverbratt, and L.-A. Hansson. "Nanoplastics in the aquatic environment." In: *Microplastic Contamination in Aquatic Environments*. Elsevier, 2018, pp. 379–399.

[135] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu. "Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics." In: *PLoS Computatioanl Biology* 12.4 (2016), e1004619.

[136] T. Mehnert, Y. H. Lam, P. J. Judge, A. Routh, D. Fischer, A. Watts, and W. B. Fischer. "Towards a mechanism of function of the viral ion channel Vpu from HIV-1." In: *Journal of Biomolecular Structure and Dynamics* 24.6 (2007), pp. 589–596.

[137] Merchant Research & Consulting Ltd. *Polyethylene Terephthalate (PET): 2016 World Market Outlook and Forecast up to 2020*. 2016.

[138] S. Miyazawa and R. L. Jernigan. "Residue-Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading." In: *Journal of Molecular Biology* 256 (1996), pp. 623–644.

[139] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. "The MARTINI coarse-grained force field: extension to proteins." In: *Journal of Chemical Theory and Computation* 4.5 (2008), pp. 819–834.

[140] E. H. Moore. "On the Reciprocal of the General Algebraic Matrix." In: *Bulletin of the American Mathematical Society* 26.9 (1920), pp. 385–396.

[141] F. Morcos, T. Hwa, J. N. Onuchic, and M. Weigt. "Direct coupling analysis for protein contact prediction." In: *Protein Structure Prediction*. Springer, 2014, pp. 55–70.

[142] K. Moritsugu and J. C. Smith. "REACH coarse-grained biomolecular simulation: transferability between different protein structural classes." In: *Biophysical Journal* 95.4 (2008), pp. 1639–1648.

[143] O. Morozova and M. A. Marra. "Applications of next-generation sequencing technologies in functional genomics." In: *Genomics* 92.5 (2008), pp. 255–264.

[144] C. E. Morris. "Mechanosensitive ion channels." In: *The Journal of Membrane Biology* 113.2 (1990), pp. 93–107.

[145] K. F. Murphy, G. Balázsi, and J. J. Collins. "Combinatorial promoter design for engineering noisy gene expression." In: *Proceedings of the National Academy of Sciences* 104.31 (2007), pp. 12726–12731.

[146] S. J. Neil, T. Zang, and P. D. Bieniasz. "Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu." In: *Nature* 451.7177 (2008), pp. 425–430.

[147] S. Páll, C. Abraham M. J.and Kutzner, B. Hess, and E. Lindahl. "Tackling exascale software challenges in molecular dynamics simulations with GROMACS." In: *International Conference on Exascale Applications and Software*. Springer. 2014, pp. 3–27.

[148] S. Palluk, D. H. Arlow, T. de Rond, S. Barthel, J. S. Kang, R. Bector, H. M. Baghdassarian, A. N. Truong, P. W. Kim, A. K. Singh, et al. "De novo DNA synthesis using polymerase-nucleotide conjugates." In: *Nature Biotechnology* 36 (2018), pp. 645–650.

[149] L. Pauling and E. B. Wilson. *Introduction to quantum mechanics with applications to chemistry*. Courier Corporation, 2012.

[150] W. R. Pearson. "Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms." In: *Genomics* 11.3 (1991), pp. 635–650.

[151] R. Penrose. "A Generalized Inverse for Matrices." In: *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (1955), pp. 406–413.

[152] D. Perez, R. Huang, and A. F. Voter. "Long-time molecular dynamics simulations on massively parallel platforms: A comparison of parallel replica dynamics and parallel trajectory splicing." In: *Journal of Materials Research* 33.7 (2018), pp. 813–822.

[153] J. A. Pople. "Quantum Chemical Models (Nobel Lecture)." In: *Angewandte Chemie International Edition* 38.13-14 (1999), pp. 1894–1902.

[154] O. Postea and M. Biel. "Exploring HCN channels as novel drug targets." In: *Nature Reviews Drug Discovery* 10.12 (2011), pp. 903–914.

[155] K. R. Poulsen, T. K. Sørensen, L. Duroux, E. I. Petersen, S. B. Petersen, and R. Wimmer. "The Interaction of *Fusarium solani pisi* Cutinase with Long Chain Spin Label Esters." In: *Biochemistry* 45.30 (2006), pp. 9163–9171.

[156] D.L. Prole and G. Yellen. "Reversal of HCN channel voltage dependence via bridging of the S4–S5 linker and Post-S6." In: *The Journal of General Physiology* 128.3 (2006), pp. 273–282.

[157] J. J. Prompers, A. Groenewegen, C. W. Hilbers, and H. A. M. Pepermans. "Backbone Dynamics of *Fusarium solani pisi* Cutinase Probed by Nuclear Magnetic Resonance: The lack of Interfacial Activation Revisited." In: *Biochemistry* 38.17 (1999), pp. 5315–5327.

[158] J. J. Prompers, C. W. Hilbers, A. Groenewegen, R. C. van Schaik, and H. A. M. Pepermans. "1H, 13C, and 15N Resonance Assignments of *Fusarium solani pisi* Cutinase and Preliminary Features of the Structure in Solution." In: *Protein Science* 6.11 (1997), pp. 2375–2384.

[159] S. Pronk et al. "GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit." In: *Bioinformatics* 29.7 (2013), pp. 845–854.

[160] P. E. M. Purnick and R. Weiss. "The second wave of synthetic biology: from modules to systems." In: *Nature Reviews Molecular Cell Biology* 10.6 (2009), pp. 410–422.

[161] R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2008.

[162] F. Maldarelli R. L. Willey and, M. A. Martin, and K. Strebel. "Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4." In: *Journal of Virology* 66.12 (1992), pp. 7193–7200.

[163] H. Rehmann, A. Wittinghofer, and J. L. Bos. "Capturing cyclic nucleotides in action: snapshots from crystallographic studies." In: *Nature Reviews Molecular Cell Biology* 8.1 (2007), pp. 63–73.

[164] J. A. Reuter, D. V. Spacek, and M. P. Snyder. "High-throughput sequencing technologies." In: *Molecular Cell* 58.4 (2015), pp. 586–597.

[165] R. B. Robinson and S. A. Siegelbaum. "Hyperpolarization-activated cation currents: from molecules to physiological function." In: *Annual Review of Physiology* 65.1 (2003), pp. 453–480.

[166] A. Ronkvist, W. Xie, W. Lu, and R. A. Gross. "Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate)." In: *Macromolecules* 42.14 (2009), pp. 5128–5138.

[167] H. R. Saibil. "Macromolecular structure determination by cryo-electron microscopy." In: *Acta Crystallographica Section D: Biological Crystallography* 56.10 (2000), pp. 1215–1222.

[168] A. Šali and T. L. Blundell. "Comparative protein modelling by satisfaction of spatial restraints." In: *Journal of Molecular Biology* 234.3 (1993), pp. 779–815.

[169] B. Santoro and G. R. Tibbs. "The HCN gene family: molecular basis of the hyperpolarization-activated pacemaker channels." In: *Annals of the New York Academy of Sciences* 868.1 (1999), pp. 741–764.

[170] A. Saponaro et al. "Structural basis for the mutual antagonism of cAMP and TRIP8b in regulating HCN channel function." In: *Proceedings of the National Academy of Sciences* 111.40 (2014), pp. 14577–14582.

[171] D. Sauter, S. Hué, S. J. Petit, J.-C. Plantier, G. J. Towers, F. Kirchhoff, and R. K. Gupta. "HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef." In: *Retrovirology* 8.1 (2011), p. 103.

[172] D. Sauter, M. Schindler, A. Specht, W. N. Landford, J. Münch, K.-A. Kim, J. Votteler, U. Schubert, F. Bibollet-Ruche, B. F. Keele, et al. "Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains." In: *Cell Host & Microbe* 6.5 (2009), pp. 409–421.

[173] D. Sauter, S. Schwarz, K. Wang, R. Zhang, B. Sun, and W. Schwarz. "Genistein as antiviral drug against HIV ion channel." In: *Planta Medica* 80.08/09 (2014), pp. 682–687.

[174] M. Schewe, E. Nematian-Ardestani, H. Sun, M. Musinszki, S. Cordeiro, G. Bucci, B. L. de Groot, S. J. Tucker, M. Rapedius, and T. Baukrowitz. "A non-canonical voltage-sensing mechanism controls gating in K2P K+ channels." In: *Cell* 164.5 (2016), pp. 937–949.

[175] T. Schlick. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Vol. 21. Springer Science & Business Media, 2010.

[176] F. Schmidt and R. J. Platt. "Applications of CRISPR-Cas for synthetic biology and genetic recording." In: *Current Opinion in Systems Biology* 5 (2017), pp. 9–15.

[177] M. Schmidt and K. Hamacher. "Three-body interactions improve contact prediction within direct-coupling analysis." In: *Physical Review E* 96.5 (2017), p. 052405.

[178] R. Schmitz, N. Müller, S. Ullmann, and M. Vogel. "A Molecular Dynamics Simulations Study on Ethylene Glycol-Water Mixtures in Mesoporous Silica." In: *Journal of Chemical Physics* 145.10 (2016), p. 104703.

[179] C. Scholz, S. Knorr, K. Hamacher, and B. Schmidt. "DOCKTITE A Highly Versatile Step-by-Step Workflow for Covalent Docking and Virtual Screening in the Molecular Operating Environment." In: *Journal of Chemical Information and Modeling* 55.2 (2015), pp. 398–406.

[180] U. Schubert, L. C. Antón, I. Bacík, J. H. Cox, S. Bour, J. R. Bennink, M. Orlowski, K. Strebel, and J. W. Yewdell. "CD4 glycoprotein degradation induced by human immunodeficient virus type 1 Vpu protein requires the function of proteasomes and the ubiquitin pathway." In: *Journal of Virology* 72.3 (1998), pp. 1–18.

[181] U. Schubert, A. V. Ferrer-Montiel, M. Oblatt-Montal, P. Henklein, K. Strebel, and M. Montal. "Identification of an ion channel activity of the Vpu transmembrane domain and its involvement in the regulation of virus release from HIV-1-infected cells." In: *FEBS Letters* 398.1 (1996), pp. 12–18.

[182] U. Schubert, P. Henklein, B. Boldyreff, E. Wingender, K. Strebel, and T. Porstmann. "The Human Immunodeficiency Virus Type 1 Encoded Vpu Protein is Phosphorylated by Casein Kinase-2 (CK-2) at Positions Ser52 and Ser56 within a Predicted $\alpha$-Helix-Turn-$\alpha$-Helix-Motif." In: *Journal of Molecular Biology* 236.1 (1994), pp. 16–25.

[183] U. Schubert, T. Schneider, P. Henklein, K. Hoffmann, E. Berthold, H. Hauser, G. Pauli, and T. Porstmann. "Human-immundeficiency-virus-type-1-encoded Vpu protein is phosphorylated by casein kinase II." In: *European Journal of Biochemistry* 204.2 (1992), pp. 875–883.

[184] U. Schubert and K. Strebel. "Differential activities of the human immunodeficiency virus type 1-encoded Vpu protein are regulated by phosphorylation and occur in different cellular compartments." In: *Journal of Virology* 68.4 (1994), pp. 2260–2271.

[185] H. M. Senn and W. Thiel. "QM/MM methods for biomolecular systems." In: *Angewandte Chemie International Edition* 48.7 (2009), pp. 1198–1229.

[186] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, et al. "Anton, a special-purpose machine for molecular dynamics simulation." In: *Communications of the ACM* 51.7 (2008), pp. 91–97.

[187] D. E. Shaw et al. "Atomic-level characterization of the structural dynamics of proteins." In: *Science* 330.6002 (2010), pp. 341–346.

[188] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. P. Sansom. "HOLE: a program for the analysis of the pore dimensions of ion channel structural models." In: *Journal of Molecular Graphics* 14.6 (1996), pp. 354–360.

[189] M. B. Sprinzak D.and Elowitz. "Reconstruction of genetic circuits." In: *Nature* 438.7067 (2005), pp. 443–448.

[190] J. J. P. Stewart. "MOPAC: A Semiempirical Molecular Orbital Program." In: *Journal of Computer-Aided Molecular Design* 4.1 (1990), pp. 1–103.

[191] J. Stieber, A. Thomer, B. Much, A. Schneider, M. Biel, and F. Hofmann. "Molecular basis for the different activation kinetics of the pacemaker channels HCN2 and HCN4." In: *Journal of Biological Chemistry* (2003).

[192] B. J. Strait and T. G. Dewey. "The Shannon information entropy of protein sequences." In: *Biophysical Journal* 71.1 (1996), pp. 148–155.

[193] K. Strebel. "HIV-1 Vpu - an ion channel in search of a job." In: *Biochimica et Biophysica Acta* 1838.4 (2014), pp. 1074–1081.

[194] K. Strebel, T. Klimkait, and M. A. Martin. "A novel gene of HIV-1, vpu, and its 16-kilodalton product." In: *Science* 24.4870 (1988), pp. 1221–1223.

[195] T. Sun and A. S. Teja. "Density, Viscosity, and Thermal Conductivity of Aqueous Ethylene, Diethylene, and Triethylene Glycol Mixtures between 290 K and 450 K." In: *Journal of Chemical & Engineering Data* 48.1 (2003), pp. 198–202.

[196] M. R. Sunkara, T. Schwabe, G. Ehrlich, J. Kusch, and K. Benndorf. "All four subunits of HCN2 channels contribute to the activation gating in an additive but intricate manner." In: *The Journal of General Physiology* (2018), jgp–201711935.

[197] F. Tama and Y. H. Sanejouand. "Conformational Change of Proteins Arising from Normal Mode Calculations." In: *Protein Eng.* 14.1 (2001), pp. 1–6.

[198] A. L. Tarca, V. J. Carey, X. Chen, R. Romero, and S. Drăghici. "Machine learning and its applications to biology." In: *PLoS Computational Biology* 3.6 (2007), e116.

[199] P. Thomas and T. G. Smart. "HEK293 cell line: a vehicle for the expression of recombinant proteins." In: *Journal of Pharmacological and Toxicological Methods* 51.3 (2005), pp. 187–200.

[200] J. D. Thompson, D. G. Higgins, and T. J. Gibson. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." In: *Nucleic Acids Research* 22.22 (1994), pp. 4673–4680.

[201] S. Thon, E. Schulz, J. Kusch, and K. Benndorf. "Conformational flip of nonactivated HCN2 channel subunits evoked by cyclic nucleotides." In: *Biophysical Journal* 109.11 (2015), pp. 2268–2276.

[202] M. M. Tirion. "Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis." In: *Physical Review Letters* 77.9 (1996), pp. 1905–1908.

[203] S. Y. Tsang, H. Lesso, and R. A. Li. "Critical intra-linker interactions of HCN1-encoded pacemaker channels revealed by interchange of S3–S4 determinants." In: *Biochemical and Biophysical Research Communications* 322.2 (2004), pp. 652–658.

[204] S. Y. Tsang, H. Lesso, and R. A. Li. "Dissecting the structural and functional roles of the S3-S4 linker of pacemaker (hyperpolarization-activated cyclic nucleotide-modulated) channels by systematic length alterations." In: *Journal of Biological Chemistry* 279.42 (2004), pp. 43752–43759.

[205] N. G. Tsierkezos and I. E. Molinou. "Thermodynamic Properties of Water+ Ethylene Glycol at 283.15, 293.15, 303.15, and 313.15 K." In: *Journal of Chemical & Engineering Data* 43.6 (1998), pp. 989–993.

[206] N. Van Damme, D. Goff, C. Katsura, R. L. Jorgenson, R. Mitchell, M. C. Johnson, E. B. Stephens, and J. Guatelli. "The Interferon-Induced Protein BST-2 Restricts HIV-1 Release and Is Downregulated from the Cell Surface by the Viral Vpu Protein." In: *Cell Host Microbe* 3.4 (2008), pp. 245–252.

[207] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, Alan E. Mark, and H. J. C. Berendsen. "GROMACS: Fast, flexible, and free." In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1701–1718.

[208] A. Varghese, E. M. TenBroek, J. Coles Jr, and D. C. Sigg. "Endogenous channels in HEK cells and potential roles in HCN ionic current measurements." In: *Progress in Biophysics and Molecular Biology* 90.1-3 (2006), pp. 26–37.

[209] S. Vemana, S. Pandey, and H. P. Larsson. "S4 movement in a mammalian HCN channel." In: *The Journal of General Physiology* 123.1 (2004), pp. 21–32.

[210] S. Venkatachalam, S. G. Nayak, J. V. Labde, P. R. Gharal, K. Rao, and A. K. Kelkar. "Degradation and recyclability of poly (ethylene terephthalate)." In: *Polyester*. InTech, 2012.

[211] B. J. Wainger, M. DeGennaro, B. Santoro, S. A. Siegelbaum, and G. R. Tibbs. "Molecular mechanism of cAMP modulation of HCN pacemaker channels." In: *Nature* 411.6839 (2001), pp. 805–810.

[212] J. Wang, S. Chen, M. F. Nolan, and S. A. Siegelbaum. "Activity-dependent regulation of HCN pacemaker channels by cyclic AMP: signaling through dynamic allosteric coupling." In: *Neuron* 36.3 (2002), pp. 451–461.

[213] J. Wang, S. Chen, and S. A. Siegelbaum. "Regulation of hyperpolarization-activated HCN channel gating and cAMP modulation due to interactions of COOH terminus and core transmembrane regions." In: *The Journal of General Physiology* 118.3 (2001), pp. 237–250.

[214] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. "Development and Testing of a General Amber Force Field." In: *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174.

[215] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. "Jalview Version 2—a multiple sequence alignment editor and analysis workbench." In: *Bioinformatics* 25.9 (2009), pp. 1189–1191.

[216] W.-M. Weber. "Endogenous ion channels in oocytes of Xenopus laevis: recent developments." In: *The Journal of Membrane Biology* 170.1 (1999), pp. 1–12.

[217] P. Weil, F. Hoffgaard, and K. Hamacher. "Estimating Sufficient Statistics in Co-Evolutionary Analysis by Mutual Information." In: *Computational Biology and Chemistry* 33.6 (2009), pp. 440–444.

[218] S. Weißgraeber. "Hyperpolarization-Activated cyclic nucleotide-gated channels - structure and evolution." PhD thesis. Technische Universität Darmstadt, Darmstadt, Germany, 2015.

[219] S. Weißgraeber, A. Saponaro, G. Thiel, and K. Hamacher. "A reduced mechanical model for cAMP-modulated gating in HCN channels." In: *Scientific Reports* 7 (2017), p. 40168.

[220] J. R. Whicher and R. MacKinnon. "Structure of the voltage-gated K+ channel Eag1 reveals an alternative voltage sensing mechanism." In: *Science* 353.6300 (2016), pp. 664–669.

[221] J. Wilkinson, G. Ewart, C. Luscombe, K. McBride, W. Ratanasuwan, M. Miller, and R. L. Murphy. "A Phase 1b/2a study of the safety, pharmacokinetics and antiviral activity of BIT225 in patients with HIV-1 infection." In: *Journal of Antimicrobial Chemotherapy* 71.3 (2016), pp. 731–738.

[222] M. N. Win and C. D. Smolke. "RNA as a versatile and powerful platform for engineering genetic regulatory tools." In: *Biotechnology and Genetic Engineering Reviews* 24.1 (2007), pp. 311–346.

[223] S..n Wu, W. Gao, C. Xie, X. Xu, C. Vorvis, F. Marni, A. R. Hackett, Q. Liu, and L. Zhou. "Inner activation gate in S6 contributes to the state-dependent binding of cAMP in full-length HCN2 channel." In: *The Journal of General Physiology* 140.1 (2012), pp. 29–39.

[224] S. Wu, Z. V. Vysotskaya, X. Xu, C. Xie, Q. Liu, and L. Zhou. "State-dependent cAMP binding to functioning HCN channels studied by patch-clamp fluorometry." In: *Biophysical Journal* 100.5 (2011), pp. 1226–1232.

[225] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson. "Multi-input RNAi-based logic circuit for identification of specific cancer cells." In: *Science* 333.6047 (2011), pp. 1307–1311.

[226] J. L. Yarnell, M. J. Katz, R. G. Wenzel, and S. H. Koenig. "Structure Factor and Radial Distribution Function for Liquid Argon at 85 K." In: *Physical Review A* 7.6 (1973), p. 2130.

[227] G. Yellen. "The voltage-gated potassium channels and their relatives." In: *Nature* 419.6902 (2002), pp. 35–42.

[228] S. Yoshida, K. Hiraga, T. Takehana, I. Taniguchi, H. Yamaji, Y. Maeda, K. Toyohara, K. Miyamoto, Y. Kimura, and K. Oda. "A Bacterium that Degrades and Assimilates Poly(ethylene terephthalate)." In: *Science* 351.6278 (2016), pp. 1196–1199.

[229] W. N. Zagotta, N. B. Olivier, K. D. Black, E. C. Young, R. Olson, and E. Gouaux. "Structural basis for modulation and agonist specificity of HCN pacemaker channels." In: *Nature* 425.6954 (2003), pp. 200–205.

[230] H. Zhang, E. C. Lin, B. B. Das, Y. Tian, and S. J. Opella. "Structural determination of virus protein U from HIV-1 by NMR in membrane environments." In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1848.11 (2015), pp. 3007–3018.

[231] Y. Zhang. "Pseudobond ab initio QM/MM approach and its applications to enzyme reactions." In: *Theoretical Chemistry Accounts* 116.1-3 (2006), pp. 43–50.

# APPENDIX

This chapter provides Supporting Information for the publications in Chapters 2, 3 and 4. To retain consistency troughout the whole thesis, the numbering of Supplementary Figures has been changed compared to the published Supporting Information files.

The Appendix is structured as follows:

**A1**: Supporting Information to

> Christine Groß, Kay Hamacher, Katja Schmitz and Sven Jager (2017). Cleavage Product Accumulation Decreases the Activity of Cutinase during PET Hydrolysis, J. Chem. Inf. Model, 57(2):243-255. doi: 10.1021/acs.jcim.6b00556

**A2**: Supporting Information to

> Christine Groß, Andrea Saponaro, Bina Santoro, Anna Moroni, Gerhard Thiel and Kay Hamacher (2018). Mechanical transduction of cytoplasmic-to-transmembrane-domain movements in a hyperpolarization-activated cyclic nucleotide–gated cation channel, J. Biol. Chem., 293(33):12908-12918. doi: 10.1074/jbc.RA118.002139

**A3**: Supporting Information to

> Timo Greiner, Sebastian Bolduan, Brigitte Hertel, Christine Groß, Kay Hamacher, Ulrich Schubert, Anna Moroni and Gerhard Thiel (2016). Ion Channel Activity of Vpu Proteins Is Conserved throughout Evolution of HIV-1 and SIV, Viruses, 8(12):325. doi: 10.3390/v8120325

## A.1 SUPPORTING INFORMATION TO CHAPTER 2

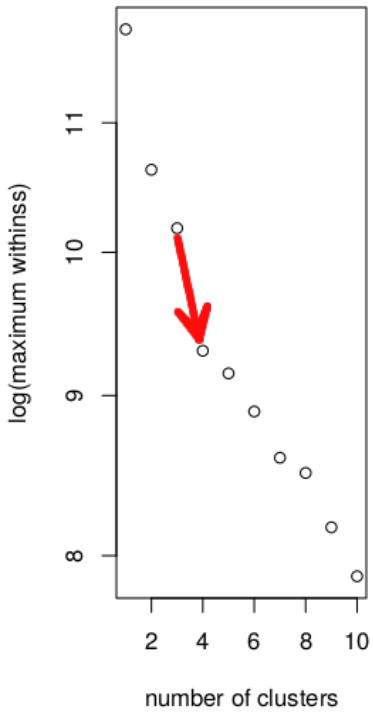


Figure A.1: `kmeans` clustering of the 13Å null model with maximal within-cluster sum of squares (withinss) as a function of number of clusters with the red arrow highlighting the drastic drop using four instead of three clusters.

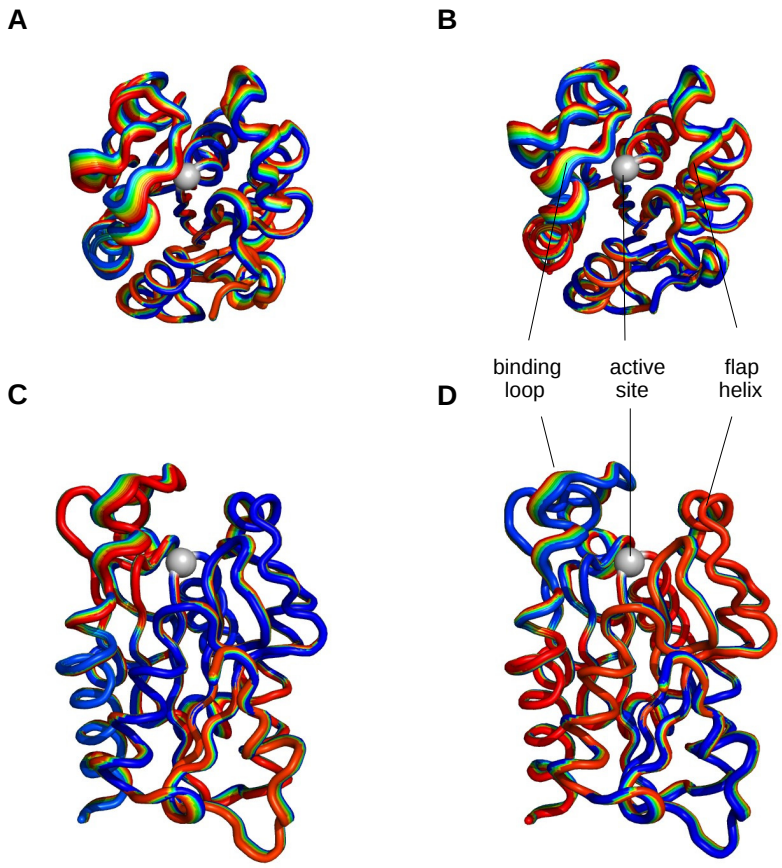


Figure A.2: Lowest frequency normal mode of *Fs*C in one direction (**A + C**) and in the opposite direction (**B + D**) in top and side view (right angle to each other). In contrast to the "breath-like" motions after perturbation of S120 by substrate binding/cleavage, the normal mode loop motions are in a parallel manner (both loops move to the left or both loops move to the right).

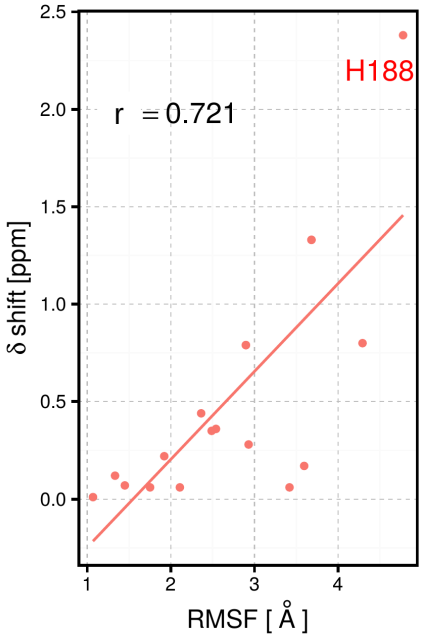


Figure A.3: Comparison of NMR data with RMSF values of the wildtype simulation with Pearson's $r = 0.721$, p-value= 0.00043. H188 is labeled in red. $\delta$ shift values are obtained from Prompers *et al*. [158]

For secondary structure quantification we used standardized secondary structure assignment, `Define Secondary Structure of Proteins` (short: `DSSP`). `DSSP` begins by identifying the intra-backbone hydrogen bonds of the protein using a purely electrostatic definition [92]. We computed `DSSP` for each frame and computed the mean occurrence of secondary stucture elements of the complete production run of the simulation using `R` [161] and `bio3d` [62].

We measured the following secondary structural states derived from `DSSP` (Figure A.4 D):
B = residue in isolated $\beta$-bridge,
E = extended strand, participates in $\beta$ ladder,
G = 3-helix ($3_{10}$ helix),
H = $\alpha$-helix,
I = 5 helix ($\pi$-helix),
S = bend,
T = hydrogen bonded turn,
U = loop region.

Furthermore, to see that the protein does not unfold during the simulation time, we plotted the percentages of $\alpha$-helices, residues in isolated $\beta$-bridges, and bends over time (Figure A.4 A-C).

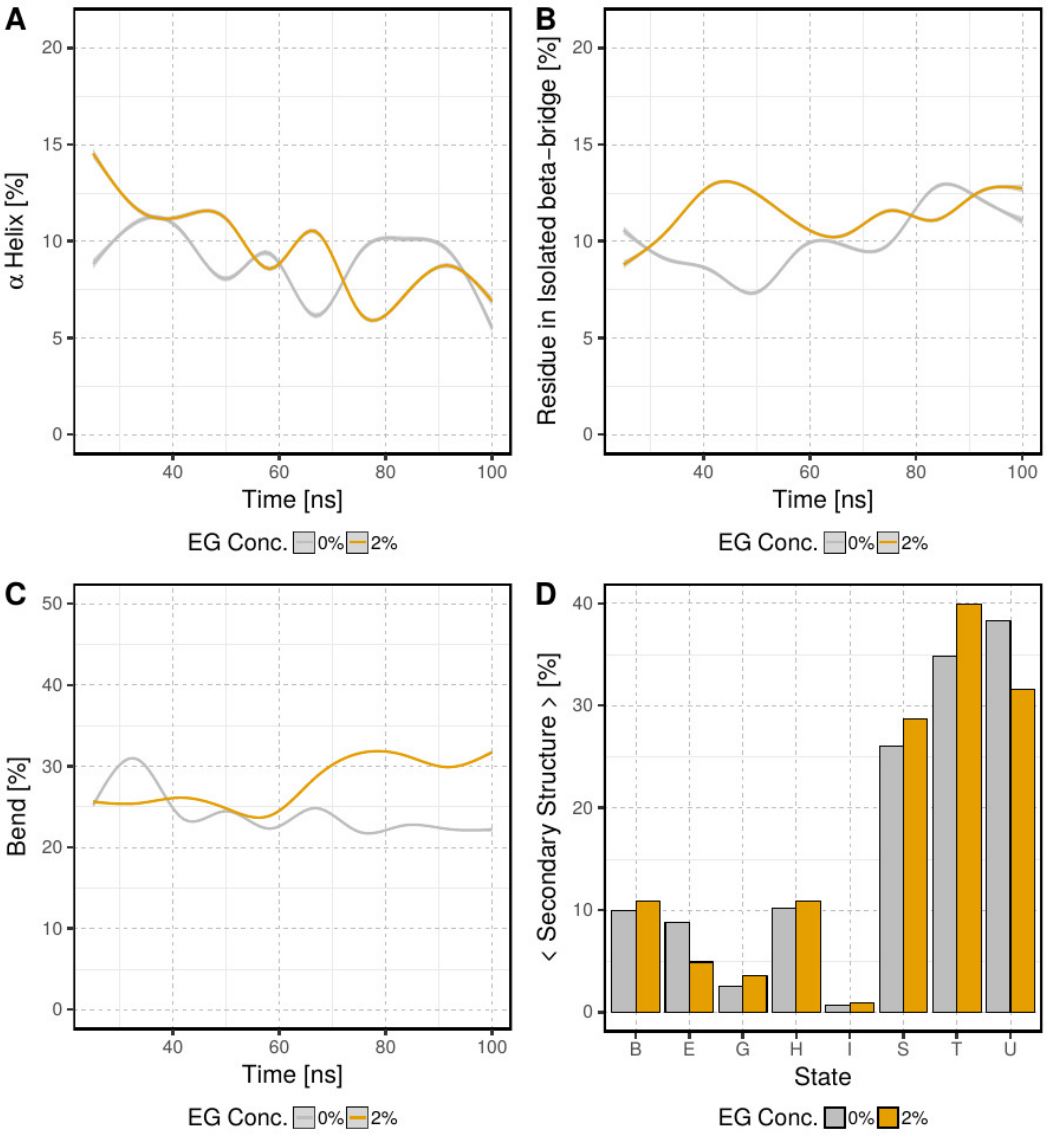


Figure A.4: Secondary structure analysis for the MD simulations with 0% and 2% EG in the solvent. (**A-C**) Main secondary structure elements over time. The curves were smoothed using the LOESS [33] algorithm. (**D**) Average percentages of all observed secondary structure elements.

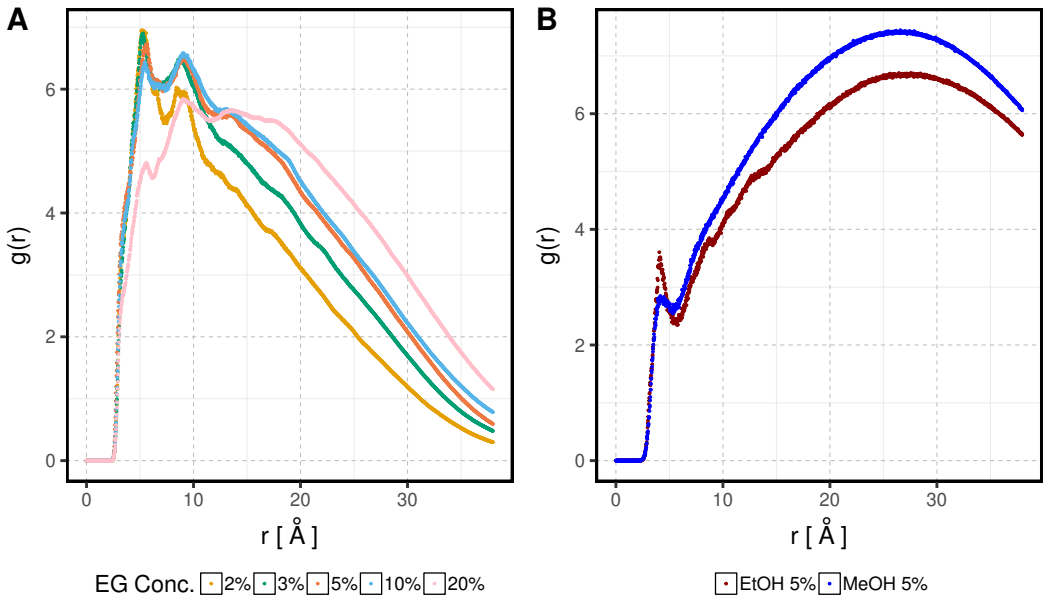


Figure A.5: (**A**) Radial distribution function (RDF) of protein C$\alpha$ to EG O for the simulations with different concentrations of EG in the solvent. For all EG concentrations we see a peak around 5-7 Å which is the first coordination shell of an amino acid followed by a second peak at around 10 Å. With further distances the radial distribution strongly decreases. These results indicate accumulation of EG near the *Fs*C surface while the amount of EG in the bulk is minor. (**B**) RDF af protein C$\alpha$ to EtOH O or MeOH O, respectively, for the simulations with other alcohols in the solvent. For both alcohols we see a tiny peak in the first coordination shell but in contrast to EG most of the EtOH and MeOH molecules are located in the bulk far away from the protein surface.

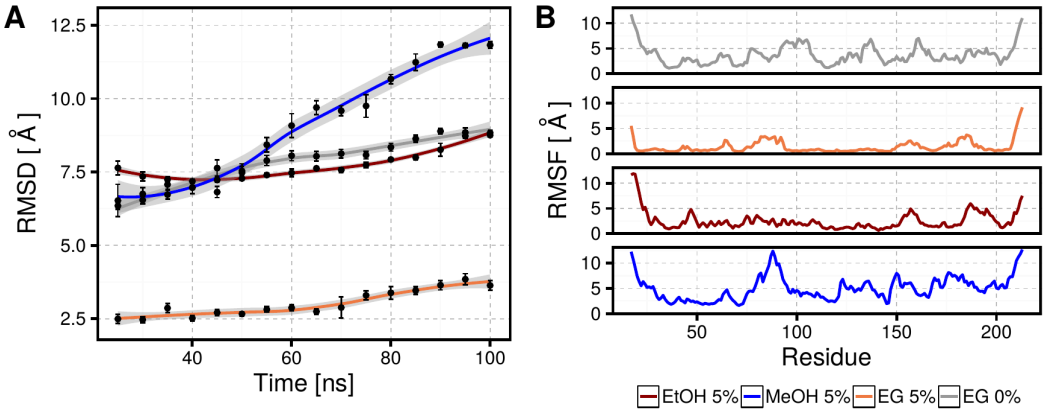


Figure A.6: Comparison of RMSD (**A**) and RMSF (**B**) for 5% EG, 5% EtOH, and 5% MeOH in the solvent. To see the effect of the different alcohols in the solvent, the corresponding RMSD and RMSF without any alcohol (0% EG) is also plotted. The curves were smoothed using the LOESS[33] algorithm.

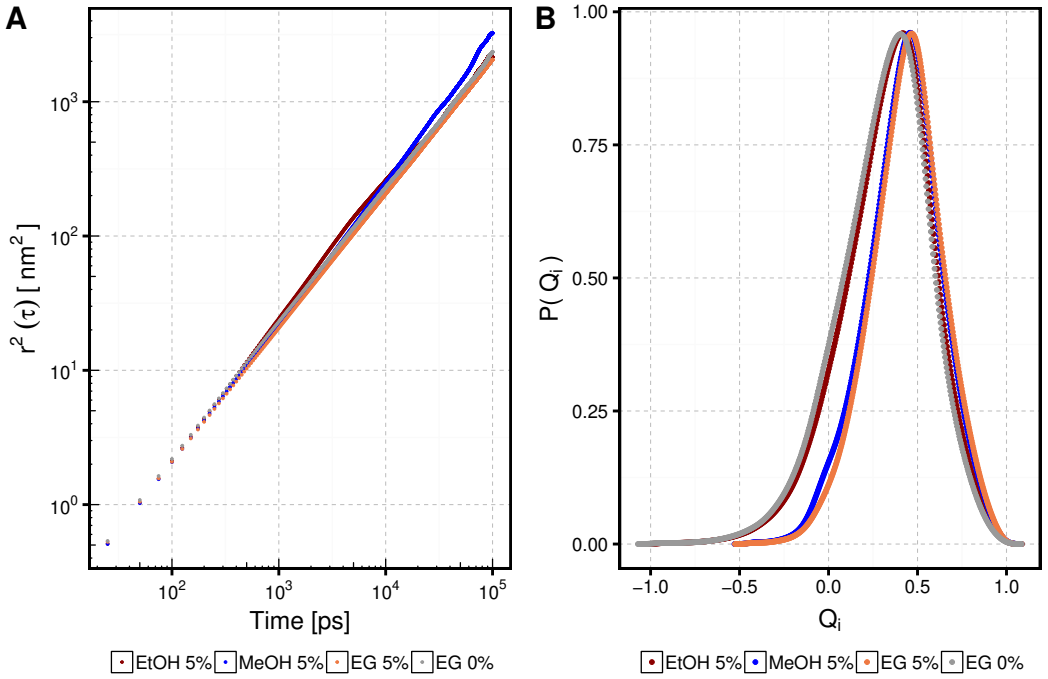


Figure A.7: Characterization of the solvent for the simulations with 5% EtOH, 5% MeOH, and 5% EG in comparison to the simulation without any alcohol added to the solvent. (**A**) Double logarithmic representation of MSD of the water molecules as indirect measure of the viscosity. The constant increase of the MSD during the simulation describes normal diffusion. While the MSD of the simulation with EG is slightly decreased compared to the simulation with water only, the MSD for the simulation with MeOH is increased and the MSD of the simulation with EtOH behaves as with water only. (**B**) Tetrahedral order parameter $Q_i$ to quantify the ability of water molecules to form hydrogen bonds to adjacent water molecules. The distribution of $Q_i$ is right shifted for EG and MeOH while EtOH does not significantly influence the hydrogen bonding network of the water molecules.

To determine whether the interactions between EG and the surface residues are of hydrophobic or hydrophilic nature, we made surface density calculations based on Equations 7-11 with $d$ set to 7 Å. The subsets of surface residues were grouped as follows:

*active site residues:*
41 42 81 84 119 120 121 150 175 177 183 184 185 186 188

*remaining surface residues:*
1 2 3 4 5 6 7 8 9 10 11 12 13 14 17 18 24 27 28 29 30 31 32 33 34 35 36 37 38 44 45 46 47 48 49 50 51 52 53 59 60 61 62 63 64 65 66 67 68 69 70 71 72 75 76 77 79 80 83 85 86 87 88 89 90 91 92 94 95 96 97 105 114 115 116 117 118 122 123 124 126 127 128 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 151 152 153 154 156 157 158 159 160 162 163 164 165 166 167 168 169 170 171 172 173 174 176 178 179 180 181 182 187 189 190 191 192 193 194 195 196 197

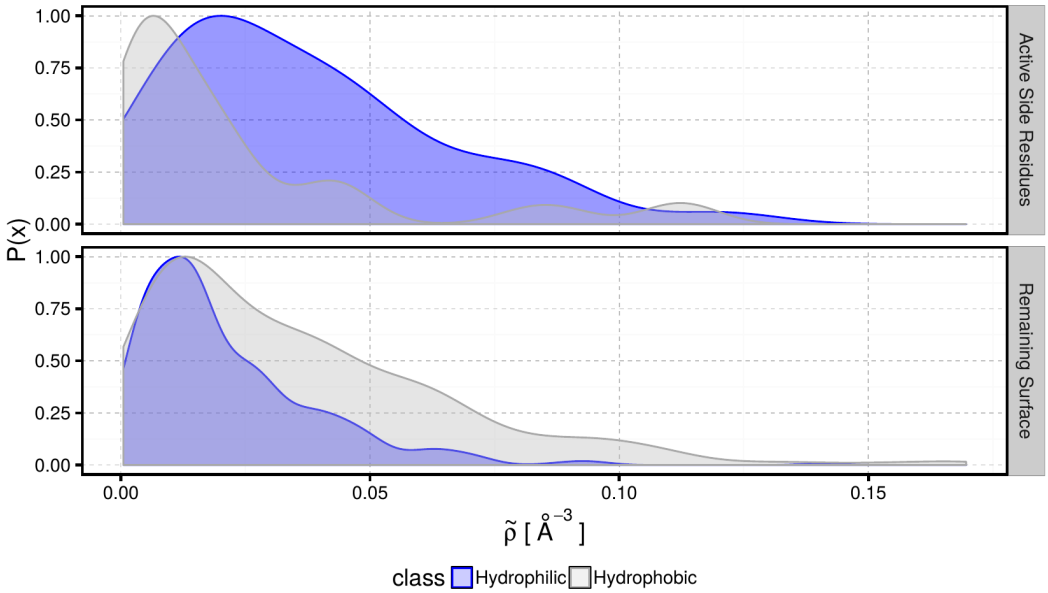


Figure A.8: Distributions of EG densities at surface residues in the ative site in comparison to the EG densities at remaining surface residues. For both surface subsets the densities are seperately plotted for hydrophilic (blue) and hydrophobic (transparent) residues. While the EG densities at the remaining surface are quite similar for hydrophobic and hydrophilic residues, the densities of EG in the active site are significantly higher at hydrophilic residues. This indicates that the EG accumulation in the active site is mainly based on hydrogen-bonding interactions.

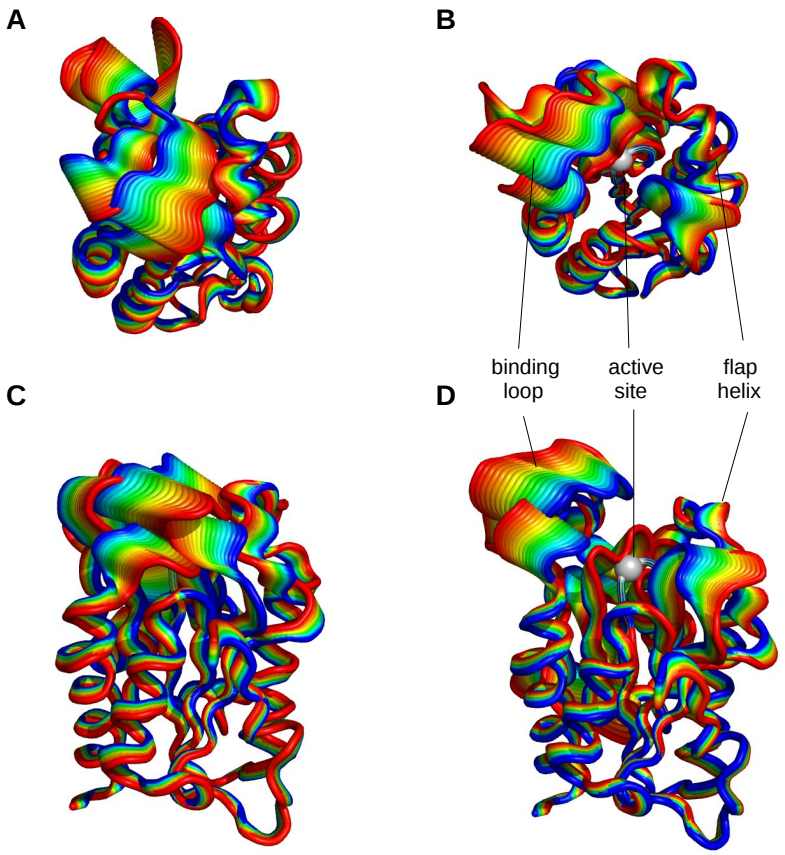


Figure A.9: LRT model (cutoff 7 Å) with repulsive (**A** + **C**) and attractive (**B** + **D**) force vectors with different forces analogous to the model with 13 Å in Figure 2.7. Here forces from 0 (blue) to 60 (red) in arbitrary units were applied.

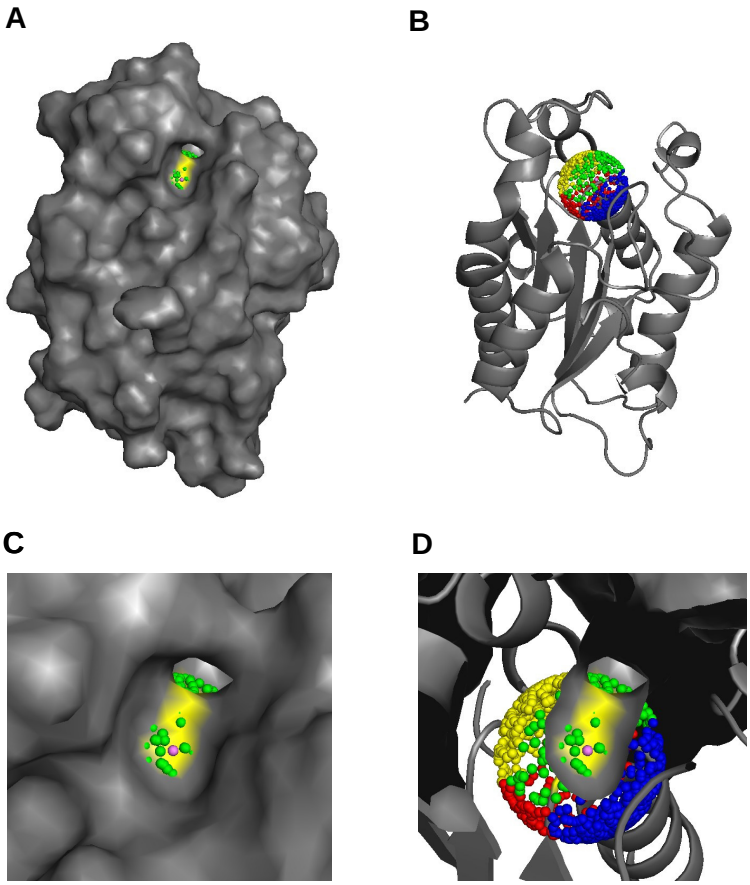


Figure A.10: LRT null model (cutoff 7 Å) with 1,000 force directions for the external force vector analogous to the model with 13 Å in Figure 2.8.

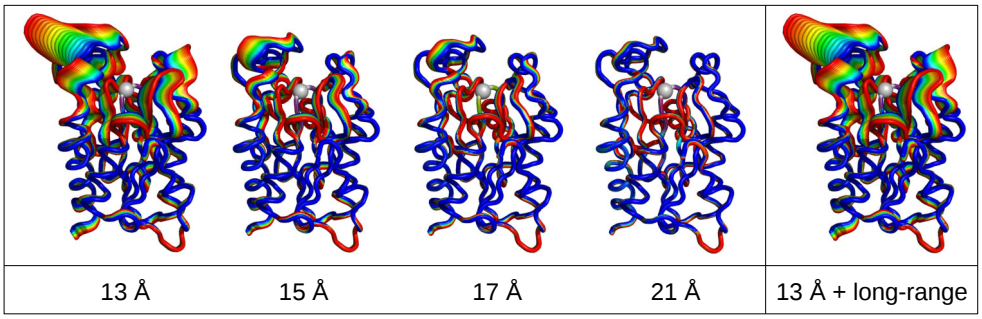


Figure A.11: Comparison of the displacements of anisotropic network models with different cutoffs for connected residues to account for long-range interactions. The displacements only differ in their magnitude, which shows that the 13 Å modell does not disregard possible deviating long-range interactions. The 13 Å + long-range model was computed with decreasing interactions between residues with larger distances than 13 Å. Nevertheless, the displacement is the same as for the original 13 Å model.

## A.2 SUPPORTING INFORMATION TO CHAPTER 3

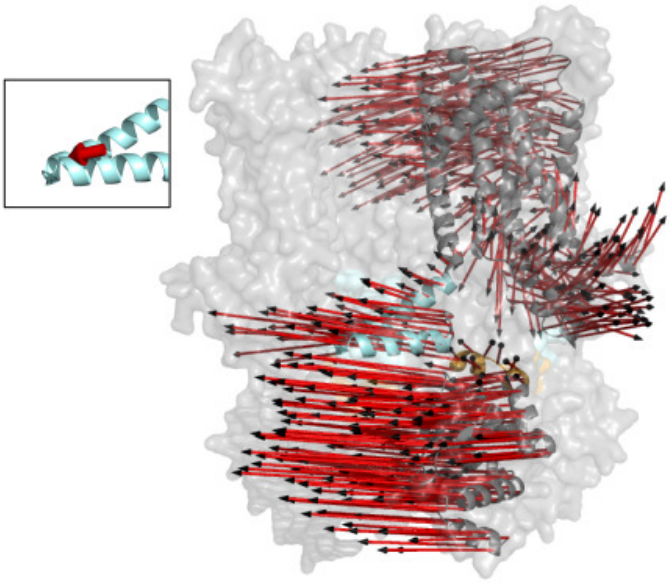

Figure A.12: Perturbation of the ANM from HCN1 in the cAMP-bound form. The force is applied at the tip of the elbow from the red cluster. This perturbation intends to represent displacements of the cAMP-bound structure in response to cAMP release from the binding site. Note that the red cluster points in the opposite direction of the perturbation from the yellow cluster, which simulates cAMP binding. For clarity, the displacements are only shown for one subunit as in Figure 3.2 of main text. Color coding of the "elbow" and the "shoulder" are the same as in Figure 3.1 of main text.

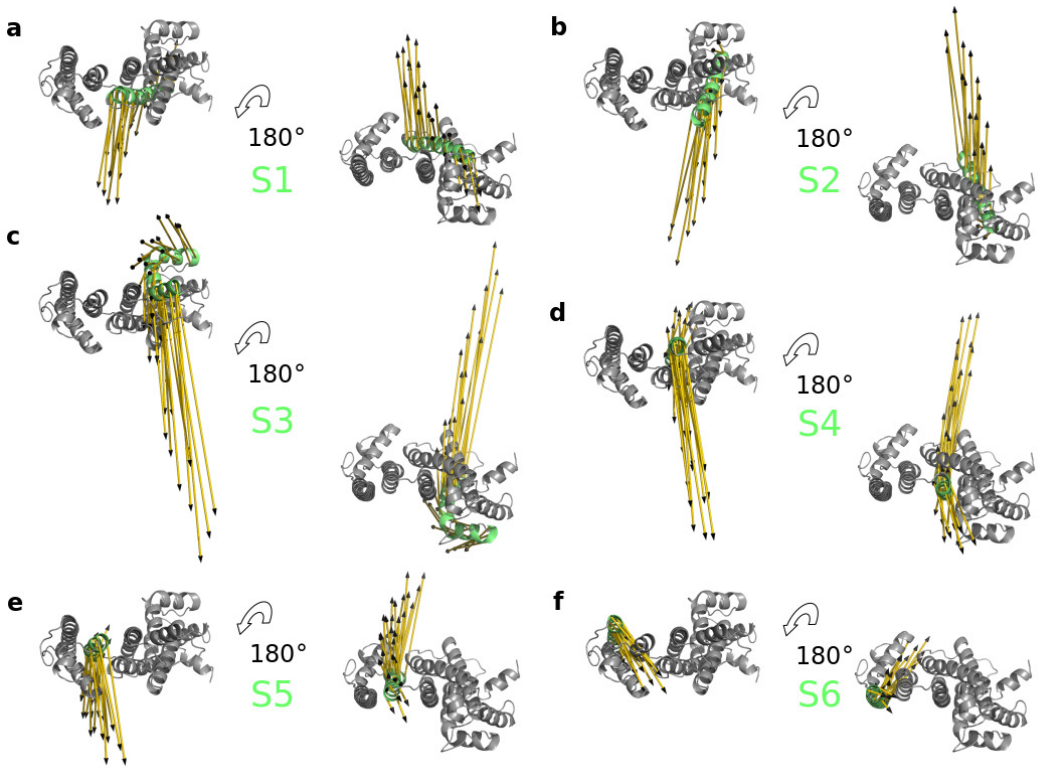


Figure A.13: Displacements of S1 to S6 helices of the TMPC. (**a**-**f**), seen in top and bottom view after perturbing the "elbow" of the C-Linker from the most realistic perturbation direction. The corresponding helix is highlighted and labeled in lime green and the displacement is visualized as yellow arrows. For clarity only one subunit (residue 94-402) is shown.

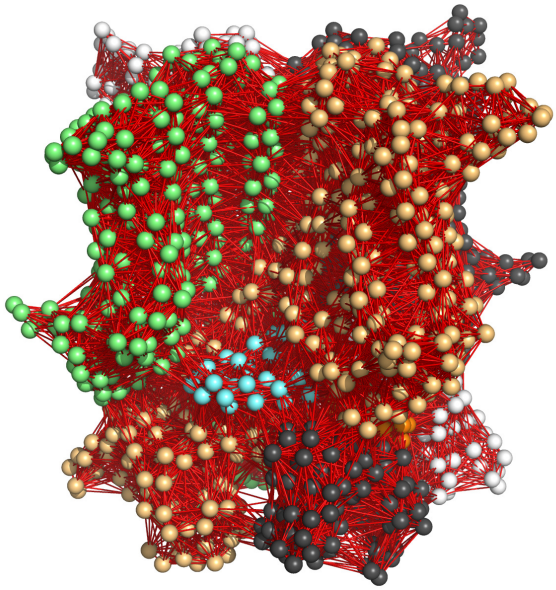

Figure A.14: Elastic network model of curated cAMP-free HCN1 structure with a distance cutoff of 13 Å for connected residues. The C$\alpha$ atoms of each residue are reduced to spheres (coloring of the subunits as in Figure 3.1 b) and the connections between atoms are shown as red lines. The elbow domain of the gray subunit is highlighted in light blue and the corresponding shoulder domain in orange as in Figure 3.1 a of main text.

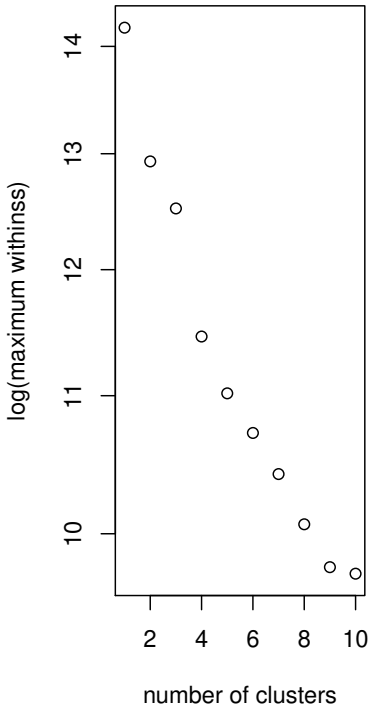


Figure A.15: `kmeans` clustering of the LRT null model with maximal within-cluster sum of squares (withinss) as a function of number of clusters. The most drastic drop is from three to four clusters.

## A.3 SUPPORTING INFORMATION TO CHAPTER 4

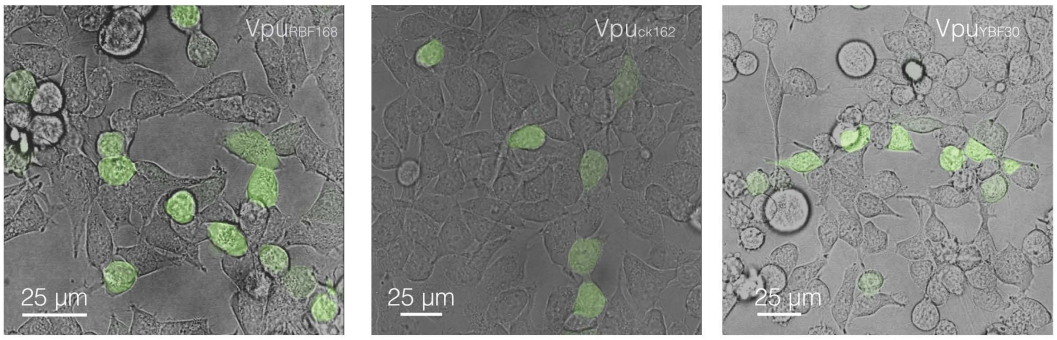


Figure A.16: Transfection of HEK293T cells with Vpus from HIV-1. Exemplary images of HEK293T cells transiently transfected with bi-cistronic vector containing different Vpus (from left to right: VpuRBF168, VpuCK1.62, VpuYBF30) together with GFP. Images are overlays of bright field images revealing cell contours and fluorescent images showing positive GFP signal in individual cells.

```
HIV_consensus   MQPLEI.LAIVALVVALIIAIVVWTIVYIEYRKILRQRKIDRLIDRIRERAEDSGNESEGDTEELSALVEMGHLAPWDVND
B0LB90_SIV      MLTW.EQIGLIALGIEGIIATVVWGIAFITWRRRKIEQKIIDLLERIRARQDDSGNESDGEDQRQMA.....YGFDNPMFD
Q9Q073_SIV      MLTW.EQIGLIGIGIEIIIAIVAWGIAFKEWRKGKEEQKIRTLIERIISRQEDSGNESDGEDQRQLDREIHVYGFDNPMFD
Q9ICT3_SIV      MLIW.EQIGLIALGIELIIVIVVWGIAYKKWREIKEEQRIYNLAERIRIRQEDSGNESNGEEEAQLDNLLHANGFDNPMFD
Q9YL64_SIV      MLNW.FEIGLIALGIEGILVVIIWGLVARLWRQIKIKQEIQNLLERIRIREEDSGNESDGEEEETLAKLLSSLELDNPRIE
Q5K6J1_SIV      MLSM.WVIGLIGIGTLLVINIVVWGIVYKRWKRHKEEQRIIDLIIKITERAEDSGNESDGEDKETLATLLHNNGFDNPMFE
Q1A253_SIV      MDL..IEIGLIGLVIELIIVIVVWLKAYQLYKENIRQKAINKLIERIRERGEDSGNESEGDMDE.LHAIIR....GDPEL.
B0LB99_SIV      MEIF.IILGLIGIVIELVIAIVVWLKAYECYKALKRQERRDQLIDRIRERAEDSGNESDGDTEE.LEAIL....IEDRHV.
S4SU71_SIV      MTGLEI.IGLIGIVIELSIAIGAWIVAYNNYQEYKRGERIDCLIKRIRERAEDSGNESDGDKEE.LCTLL....LSDPEF.
I2BF64_SIV      MLLL.IKIGFIGLAIETLIVIVVWAIVYRIYREVKVEEKISQLRQRIRDRAEDSGNESDGDAEE.LANLL.....PDRIE.
VPU_SIVEK       MLLL.IKIGFIGLAIETLIVIVVWAIVYRIYREVKVEEKISQLRQRIRDRAEDSGNESDGDAEE.LANLL.....PDRIE.
VPU_SIVMB       MDIV.QQVGLVVLIIELVIVIVIWVKVYKLCKEDRRQKIDRLIARIRERAEDSGNESDGDTEE.LQDLIT...E.DNLM.
E8ZER9_SIV      M...HL.VYLIISIILILLNIVLWSKVWIYYQHIKVKEKIQRLAIRIRERAEQDSGNESNGE...ILEHLVQKHGFANPLFD
E8ZER2_SIV      M...HL.VYLIISIILILLNIVVWSKVWIDYQHIKVKEKIQRLAIRIRERAEQDSGNESNGE...ILEHLVQKHGFANPLFD
VPU_SIVCZ       MTL..L.VGLVLILVGLIANICIWGYGYRRYYKRHRLETEIERLNLILRERAEDSGNESNGE...RLEQLIHNNHFANPMFD
Q5MYY3_SIV      MI...L.IALGCLAIALILNIFIWRNIWRLCKQNKLEVEIENLALRLTERAEDSGNESDGG...KLQRLLHDHGFANPLFD
Q1A235_SIV      M..QEI.VLIILFIALMLVAIFAWIAAYKEYKKLQQVRRIERLQDRIRSRAEDSGNESDGDEILLVEELMQVHQHVNPDWM
B9V2T0_SIV      MHPRDI.IVIIIGITLLAVTVIIWLKIFALYLRDRRERRFFDRLERLLSNKEDEGYESNEEEAAELMEMGNELGF..FNLH
B9V2S1_SIV      MHPRDI.IVIIIGITLLAATVIVWLKALALYLRDRRERRFFNRLERLLSTKEDEGYESNEEEAAELMERGNELGF..FNL.
B9V2W6_SIV      MHPRDI.LVIIIGIILLAVTVISWLKALALYLRDRRERRFFDRLERLISNKEDEGYESNEEEAAELMERGNELGF..FNL.
B9V2V5_SIV      MHSREI.AIIIIGSILLAVTVVLWVKIWLIYQANKREEQFVHRVERLLERKEDEGYESDGSEETDLMELGNRLGF..HML.
G8ZON4_SIV      MQLDAI..HFLALIVIIIELGVIIGACWWGYTQCKAYNQFDKLVQRIIERAEDSGNESEGDEEEA.LELLGARLGLPSPFP.
Q6VG35_SIV      MNYWSL.VAITYSLILIALPVAAWAW.WRYYKITKRFKRIDQEIQRLIQIHERRRHDSGVDTESESEQHEETHGFVNPVF.
```

Figure A.17: Alignment of Vpu orthologs from SIV with consensus sequence of Vpu from HIV-1. Consensus sequence (top lane) for Vpu protein was constructed from alignment of 6,497 HIV Vpu entries in PFAM database. The SIV sequences within the PFAM alignment were extracted realigned amongst each other, and processed the same way as the HIV sequences. This resulted in the above alignment of 22 SIV sequences.

# B

## ACRONYMS

| | |
|---|---|
| AA | amino acid |
| ANM | anisotropic network model |
| ART | antiretroviral therapy |
| BIT225 | *N*-carbamimidoyl-5-(1-methyl-1*H*-pyrazol-4-yl)-2-naphthamide |
| BST2 | bone marrow stromal cell antigen 2 |
| cAMP | cyclic adenosine monophosphate |
| $CD_4$ | cluster of differentiation 4 |
| CK-2 | casein kinase 2 |
| CNDB | cyclic-nucleotide binding domain |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| cryoEM | cryo-electron microscopy |
| DCA | direct coupling analysis |
| DMEM | Dulbecco's Modified Eagle Medium |
| DNA | deoxyribonucleic acid |
| EG | ethylene glycol |
| ENM | elastic network model |
| ER | endoplasmic reticulum |
| ERAD | endoplasmic reticulum-associated protein degradation |
| *Fs*C | *Fusarium solani* Cutinase |
| GAFF | general AMBER force field |
| GFP | green fluorescent protein |
| GNM | Gaussian network model |
| HCN channel | hyperpolarization-activated cyclic nucleotide-gated channel |
| HEK293T cells | human embryonic kidney 293 T cells |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| HIV | human immunodeficiency virus |

| | |
|---|---|
| HTS | high-throughput screening |
| Ig | immunoglobulin |
| I/V | current/voltage |
| Kv channel | voltage-dependent potassium channel |
| LRT | linear response theory |
| MC | Monte-Carlo |
| MD | molecular dynamics |
| MI | Mutual Information |
| MM | molecular mechanics |
| mRNA | messenger ribonucleic acid |
| MSA | multiple sequence alignment |
| MSD | mean square displacement |
| NGS | next generation sequencing |
| NMA | normal mode analysis |
| NMR | nuclear magnetic resonance |
| PA | polyamide |
| PBS | phosphate-buffered saline |
| PDB | protein data bank |
| PET | polyethylene terephthalate |
| PME | partical mesh Ewald |
| POPS | parameter optimized surface calculator |
| QM | quantum mechanics |
| RDF | radial distribution function |
| RMSD | root mean square deviation |
| RMSF | root mean square fluctuation |
| RNA | ribonucleic acid |
| SASA | solvent accessible surface area |
| SD | standard deviation |
| SDC | surface density calculations |
| SDS | sodium dodecyl sulfate |

| | |
|---|---|
| SIV | simian immunodeficiency virus |
| TMD | transmembrane domain |
| TMPC | transmembrane portion of the channel |
| TPA | terephthalic acid |
| VCP | viral channel forming protein |
| Vpu | viral protein u |
| VSD | voltage-sensing domain |

# C

## DANKSAGUNG

An dieser Stelle möchte ich mich noch bei allen bedanken, die mich auf dem Weg hierher und während meiner Zeit als Doktorandin begleitet und unterstützt haben.

Zuallererst danke ich meinem Doktorvater **Prof. Dr. Kay Hamacher** für die Möglichkeit zur Promotion in seiner Arbeitsgruppe, die Bereitstellung der sehr spannenden Forschungsthemen und das mir entgegengebrachte Vertrauen. Ein besonderer Dank geht auch an meinen Zweitgutachter **Prof. Dr. Gerhard Thiel** für die vielen konstruktiven Gespräche und das jederzeit offene Ohr bei Fragen rund um Ionenkanäle sowie die Unterstützung beim Erstellen und Einreichen von Publikationen.

Ein großer Dank geht auch an **Prof. Dr. Katja Schmitz** für die Möglichkeit in Ihrem Labor auch experimentelle Untersuchungen zur Cutinase durchzuführen sowie die Möglichkeit in diesem Zusammenhang auch ein Forschungspraktikum und eine Masterarbeit betreuen zu können. Vielen Dank auch an meine Studentin **Sabrina** und die mitbetreuende Doktorandin **Julia** für die tolle Zusammenarbeit auf diesem Thema.

Ein großes Dankeschön geht auch an alle aktuellen und ehemaligen Mitglieder der Arbeitsgruppe Computational Biology & Simulation für viele wertvolle Diskussionen und die super Unterstützung bei Fragen jeder Art. Danke an **Sven** für die tolle Zusammenarbeit am Cutinase Thema und die vielen Tipps und Tricks fürs Programmieren. Danke an **Ben** für ein jederzeit offenes Ohr und wertvolle Hilfe rund um alle Fragen zu Mathe und Physik. Danke an **Sebastian** für seine Hilfe und die vielen wertvollen Tipps rund um alle Fragen zu Mathe und IT-Zeugs. Danke an **Daniel** für die tolle Zusammenarbeit im Rahmen von iNAPO. Danke an alle AG Mitglieder, auch die, mit denen ich inhaltlich wenig zu tu hatte, für die tolle Arbeitsatmosphäre. Vielen Dank natürlich auch an **Caroline** für alle organisatorischen Dinge und ihr jederzeit offenes Ohr. Danke auch an **Laura** und **Robert** aus unserer Nachbar-AG für den ein oder anderen Kaffee und schöne Pausen im botanischen Garten.

Ein großes Dankeschön gebührt natürlich auch meinen **Freunden** und meiner **Familie**, die für die nötige Abwechselung im Privaten gesorgt haben. Ein besonderer Dank geht an **meine Eltern**, die die Grundlage für alles geschaffen und mich in allen Lebenslagen begleitet und unterstützt haben. Gleiches gilt für **meine Großeltern**, die immer für mich da waren und das Ende meiner Promotionszeit leider doch nicht mehr miterleben dürfen. Vielen Dank auch an **meine Schwester** und **ihren Freund**, sowie an **meine Schwiegereltern**, **meine Schwägerin**, **deren Mann** und **deren Kinder** für viele schöne Momente und ausreichend Ablenkung. Zu guter Letzt danke ich **meinem Mann** für seine großartige Unterstützung, Liebe, Verständnis und Geduld, die er mir in den letzten Jahren entgegengebracht hat, sowie **unserem Sohn** für das Versüßen meiner Promotionszeit und die vielseitige Ablenkung von der Arbeit.

# D

## EHRENWÖRTLICHE ERKLÄRUNG

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

*Darmstadt, den*

Christine Groß, M.Sc.