Molecular Dynamics Simulations of the Bacterial Outer Membrane Channels TolC and OprM & dxTuber, a Biomolecular Cavity Detection Tool based on Protein and Solvent Dynamics

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von **Martin Raunest** aus Essen
Bonn, 2013

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhems-Universität Bonn

1. Gutachter: Priv.-Doz. Dr. C. Kandt
2. Gutachter: Prof. Dr. M. Wiese
Tag der Promotion: 27. Juni 2013
Erscheinungsjahr: 2013

Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben.

Bonn, den 9.4.2013

Danksagungen

Ich danke Herrn Dr. Christian Kandt für die Möglichkeit, diese Arbeit unter seiner Leitung anzufertigen. Er gewährte mir Freiräume und nahm mit vielen Anregungen einen entschiedenen Einfluss auf diese Arbeit.

Bei Herrn Prof. Dr. Wiese möchte ich mich für die freundliche Übernahme des Koreferates bedanken.

Für das Mitwirken in meiner Prüfungskommision möchte ich mich bei Herrn Prof. Dr. Thorsten Lang und bei Herrn Prof. Dr. Arne Lützen bedanken.

Mein besonderer Dank gilt meinen Kollegen aus der Arbeitsgruppe: Nadine Fischer, Thomas Schmidt und Dennis Koch für die angenehme Zusammenarbeit und produktive Arbeitsumgebung.

Insbesondere möchte ich meiner Lebensgefährtin Laura Pascharat für die gemeinsame Zeit und den Rückhalt danken.

Bei meiner Familie, besonders meinen Eltern und Geschwistern, möchte ich mich für ihre Geduld, Motivation und ständige Unterstützung bedanken.

Abstract

The multidrug resistance of bacteria is a serious phenomenon in current medical treatment. Beginning with the introduction of antibiotics more and more bacterial strains achieved resistance against these chemical compounds and over the years a competition between antibiotic drug discovery and bacterial drug resistance arose. The well studied Gram-negative bacteria *Escherichia coli* and *Pseudomonas aeruginosa* serve in this work as a model organisms for bacterial resistance against antibiotics. Both bacteria evolved multidrug resistant strains through several strategies, including the expelling of harming compounds through efflux systems. The over expression of these efflux systems in the bacterial membranes are responsible for resistance against many antibiotic compounds.

The AcrA/B–TolC efflux system induces resistance of *E.coli* against a broad range of antibiotics. Ranging from the inner membrane towards the outer membrane, the efflux system spans the entire periplasmic space. The system consists of the inner membrane transporter AcrB, the membrane fusion protein AcrA and the outer membrane channel TolC. TolC itself cooperates with several inner membrane transporters and facilitates the export of harming compounds across the outer membrane. Due to this versatility TolC could become a target of drug treatment. A disabled or blocked TolC could prevent drug extrusion via systems that use TolC as an exit gate. At the time of writing the gating functionality of TolC is not known in detail. To gain insights into TolC functionality two series of unbiased molecular dynamics (MD) simulations were performed. Whereas the first series was carried out in absence of AcrB the second one was executed in presence of the AcrB docking domain (AcrB-DD).

For the first series unbiased MD simulations between 150-300 ns in a Palmitoyloleoylphosphatidylethanolamine (POPE) / NaCl / water environment were calculated. In most of these simulations TolC opens and closes freely on extracellular side hinting at the absence of a gating functionality on this side. On periplasmic side a double aspartate ring restricts substrate passage in all simulations and grasping-like motions were noticed for the tip loops of helix 7 & 8. A consecutive binding of two sodium ions inside the lower periplasmic part of TolC occured in one simulation, which induced a stabilized closed state on periplasmic side. TolC remained closed on periplasmic side unless all ions were removed from the simulation box indicating a sodium dependent lock on this side.

For the second series of MD simulations we added the AcrB-DD to the previously described system setup based on orientations of a previously published data driven modeled structure. Four unbiased 150 ns MD simulations were calculated and in one of these simulations the docking domain spontaneously docks onto TolC. The latter simulation was extended to a simulation time of 1.05 μs resulting in a tighter

binding between AcrB and TolC with regards to the modeled structure. A preferred open conformation on extracellular hints analogue to TolC only simulations at the absence of a lock on extracellular side. On the AcrB-facing side TolC's tip loops located at helix 7 & 8 opened up and were stabilized by the AcrB docking domain. However, the double aspartate ring remained closed until the end of the simulation, meaning that either the simulation time is too short to observe an opening of TolC or that another part of the AcrA/B–TolC efflux system is missing to open TolC.

In *Pseudomonas aeruginosa* OprM had been identified as a TolC homologue protein. OprM is part of the multidrug efflux system MexA/B–OprM and acts as an exit duct for several inner membrane transporters. Also for OprM the gating mechanisms are not known in detail at time of writing. To explore OprM's gating mechanisms it has been simulated in a POPE / NaCl / water environment. During all five 200 ns long MD simulations OprM opens and closes freely on extracellular side suggesting also for OprM the absence of a gating mechanism on extracellular side. The tip loops of helix 7 & 8 on periplasmic side open up in a way comparable to TolC simulations and in contrast to TolC no closing motions were noticed for these helices for OprM. In OprM a single aspartate ring limits substrate passage on the inner membrane facing side of OprM. In contrast to TolC simulations a slight opening of this aspartate ring was measured in all five simulations. The absence of heightened sodium densities near the periplasmic entrance regions could mean that either longer simulation time is needed to observe a sodium induced closure of OprM or that the periplasmic access is regulated only by the aspartate ring. Despite the absence of heightened sodium densities in the aspartate ring region, clear peaks of high sodium densities identified sodium pockets between the equatorial region and the aspartate ring region formed by Asp171 and Asp230.

Voids inside of proteins can indicate substrate binding sites, ion pockets, pathways through channel proteins, their open and closed states and active sites. Over the years numerous cavity detection tools have been introduced to identify and highlight these voids. All available cavity detection tools were based on static structures and present cavities for single protein conformations only.

With dxTuber we developed and introduced a novel cavity detection tool based on an ensemble of protein conformations. It uses averaged protein and solvent density maps, which are derived from MD trajectories, as input. With this technique protein dynamics are taken into account and cavities are detected through the separation of protein external solvent from protein internal solvent. Protein internal solvent can be grouped into cavities and stored in the commonly used PDB file format. Individual cavities can be separated via the *atom name field* of the PDB file format. dxTuber itself can calculate cavity volume and the cross-sectional area of a single cavity along a principle axis. For convenience a graphical user interface (GUI) and a command line interface (CLI) of dxTuber are released under the GPL v2.

Abbreviations

AcrA	Acriflavin resistance protein A
AcrB	Acriflavin resistance protein B
AcrB-DD	AcrB docking domain
AP	Adapter protein
BN	Bottleneck
CLI	Command line interface
ED	Efflux duct
ESV	External solvent voxel
GPL	General Public License
GPU	Graphics processing unit
GUI	Graphical user interface
IM	Inner membrane
IMA	Inner membrane drug antiporter
ISV	Internal solvent voxel
MD	Molecular Dynamic
MFP	Membrane fusion protein
OM	Outer membrane
OMP	Outer membrane protein
PBC	Periodic boundary conditions
PME	Particle mesh Ewald
POPE	Palmitoyloleoylphosphatidylethanolamine
RND	resistant nodulation devision
RMSD	Root mean square deviation
SMP	Symmetric multiprocessing
TCA	Triangular coss-sectional area
TolC	Tolerance to colicins
VMD	Visual molecular dynamics

Contents

Chapter 1

Introduction

1.1 Antibiotics

In 1889 Paul Vuillemin described the concept of antibiosis, which means a process by which life can be used to destroy life (Vuillemin, 1889). Later Ernest Duchesne a military physician discovered that moulds can kill bacteria (Duchesne, 1897). The discovery of the first and most prominent natural antibiotic, penicillin was reported by Fleming (1929), who received the Nobel price in 1945 (Fleming, 1945). Six years later (1935) *Protonsil*^{TM} a sulfonamide, the first commercially available antibacterial antibiotic, was released by Bayer AG and later in 1939 Gerhard Domagk, inventor of *Protonsil*^{TM}, received the Nobel price for his work on antibiotics (Domagk, 1939).

Shortly after the introduction of sulfonamides the mortality rate caused by crude infections sank dramatically (Shlaes, 2011). This initial success of antibiotics was slowed down by multidrug resistant strains induced by the broad administration of antibiotics, which results in a challenge of drug discovery against bacterial drug resistance (Alanis, 2005). On the one hand new antibiotics arouse evolutionary pressure on bacteria, which on the other hand evolved survival strategies to achieve resistance against antibiotics.

Main strategies against DNA attacking drugs are point mutations and the import of homologue genes from other species, causing a loss of affinity of the drug to the targeted DNA. This lowered affinity makes the drug inefficient. Drugs affecting proteins can be disabled by newly evolved enzymes or loose binding affinity due to alterations of the targeted protein. To prevent drug access, bacteria evolved several transporters and multidrug efflux systems, which are capable to expel a broad range of harming chemical compounds (figure 1.1) (Nikaido, 2009).

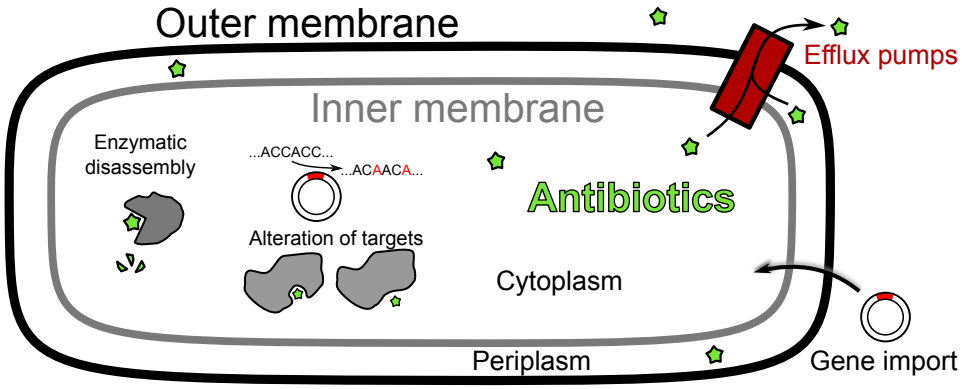

Figure 1.1: Overview of bacterial defense strategies against antibiotic compounds; enzymes render antibiotics harmless, DNA or proteins are modified via point mutations, post-translational modification or via gene import and antibiotic compounds can be expelled via efflux systems.

1.2 Bacterial Multidrug Resistance

Bacteria are classically divided into Archaebacteria and Eubacteria and the latter one in Gram-positive and Gram-negative bacteria. Both feature a cytosolic membrane followed by a peptidoglycan cell wall, which is thicker in Gram-positive bacteria (Lugtenberg and Alphen, 1983; Osborn, 1969). In contrast to Gram-positive bacteria, Gram-negative bacteria own a second bilayer - the outer membrane (OM) - consisting of lipopolisaccarids, phospholipids and proteins (Osborn et al., 1974). In this work the bacteria *Escherichia coli* and *Pseudomonas aeruginosa* serve as model organism for antibiotic resistance.

1.2.1 *Escherichia coli*

Escherichia coli (*E. coli.*) belongs to the Gram-negative bacteria, features two lipid membranes an 6.87 nm outer (OM), an 5.84 nm thick inner (IM) one and a 6.35 nm thick peptido-glycan layer in between (Typas et al., 2010). IM and OM comprise the 20.99 nm wide periplasmic space (Matias et al., 2003). *E. coli.* is facultative anaerobic, lives symbiotically in the lower intestine of mammals and produces vitamin K. *E. colis* massive population in the intestine hinders other bacteria to harm the hosting mammal (Actor, 2007; Römer et al., 2012). After the introduction of antibiotics in the 1930s beside pathogenetic bacteria also strains of *E. coli.* achieved multiple resistances to antibiotics including the oportunistic pathogens EPEC, STEC, EHEC, ETEC, EIEC, DEAC and EAEC, which are taking an advantage of a weakened immune system causing infections that could be suppressed by a healthy host (Blank et al., 2002; Day and Maurelli, 2002; Elsinghorst, 2002; Nataro and Steiner, 2002; Thorpe et al., 2002).

A major role in multidrug resistance of *E. coli.* plays the AcrA/B-TolC efflux

system, which is capable to export a broad range of toxic substrates (Nikaido, 2009).

The AcrA/B–TolC Efflux System

The over expression of the AcrA/B–TolC system prevents toxins and antibiotics to pass the inner membrane making the efflux system responsible for many multidrug resistant strains of *E.coli* (Murakami, 2008; Nikaido, 2009). The system consists of the membrane fusion protein (MFP) AcrA, the inner membrane proton drug antiporter (IMA) AcrB and the efflux duct (ED) TolC.

Whereas individual crystal structures of AcrA, AcrB and TolC have been published (Akama et al., 2004; Bavro et al., 2008; Das et al., 2007; Eicher et al., 2012; Higgins et al., 2004; Koronakis et al., 2000; Mikolosko et al., 2006; Murakami et al., 2006, 2002; Nakashima et al., 2011; Pei et al., 2011; Seeger et al., 2006; Sennhauser et al., 2007; Su et al., 2006; Törnroth-Horsefield et al., 2007; Yu et al., 2005, 2003a,b), the entire system was not solved at time of writing. Based on cross linking studies (Lobedanz et al., 2007; Tamura et al., 2005) and homology modeling on MexA Symmons et al. (2009) assembled a complete model for the AcrA/B–TolC efflux system (see figure 1.2).

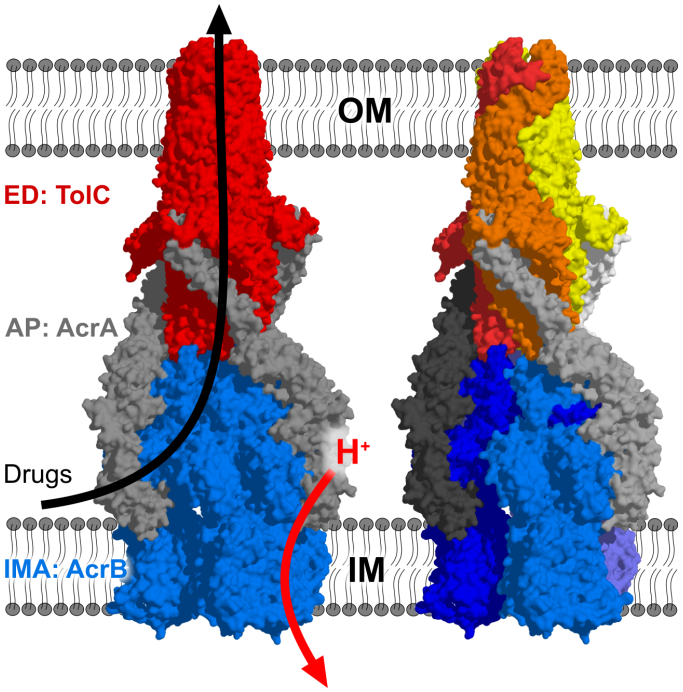

Figure 1.2: The AcrA/B–TolC multidrug efflux system is composed of the inner membrane drug antiporter (IMA) AcrB, the efflux duct (ED) TolC and the adapter protein (AP) AcrA. Although individual X-ray structures of the participating proteins are available, their structures in complex are not determined yet. Based on cross linking studies, Symmons et al. (2009) introduced the currently best available atomistic model for the efflux system.

AcrA The acriflavin resistance protein A (AcrA) is anchored in *E. coli.'s* inner membrane by an N-terminal lipid anchor, consists of 373 amino acids and is segmented in the α-helical hairpin, the lipoyl, the β-barrel and the membrane proximal domain (Ge et al., 2009; Symmons et al., 2009). The α-helical domain has two coiled-coil α-helices and is 58 Å long. The central part of AcrA, the lipoyl domain, consists of four β-strands, which interlink with one another to form a β-sandwich and connects the hairpin domain with the β-barrel domain. The latter one is formed by six antiparallel β-strands and a short α-helix (Mikolosko et al., 2006).

Whereas the hairpin domain interacts with TolC (Lobedanz et al., 2007; Touzé et al., 2004), all three remaining domains form contacts with AcrB (Krishnamoorthy et al., 2008; Tikhonova and Zgurskaya, 2004). The crystal structure 2FIM lacks the membrane proximal domain (Mikolosko et al., 2006), which later was spatially resolved via homology modeling based on X-ray crystallography data of the AcrA homologue MexA (2V4D) (Symmons et al., 2009).

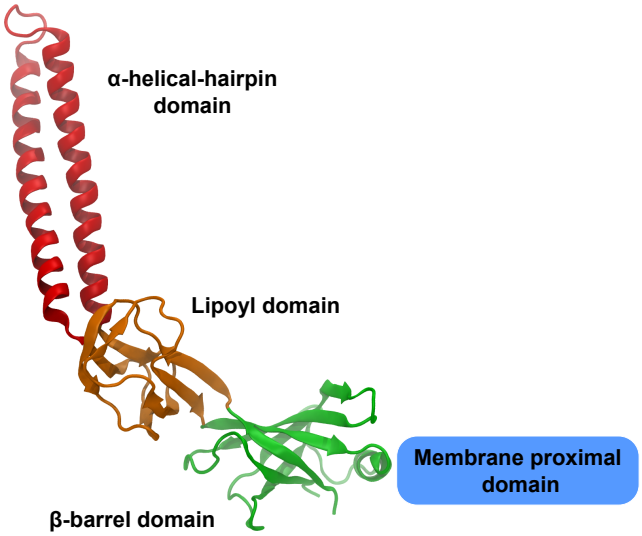

Figure 1.3: The adapter protein (AP) AcrA consists of four domains; the α-helical-hairpin domain (red), the lipoyl domain (orange), the beta barrel domain (green) and the membrane proximal domain, which is not solved in any available crystal structure.

Classified as a membrane fusion protein (MFP) AcrA acts as a stabilizing and stimulating adapter protein for AcrB and TolC (Tikhonova et al., 2011; Zgurskaya et al., 2009). The initial inner membrane AcrA/B complex assembles without TolC and as a second step TolC completes the efflux system (Zgurskaya, 2009).

The stoichiometry of the AcrA–AcrB/TolC complex is not clear. While crosslinking and complex modeling studies report three AcrAs per TolC (Bavro et al., 2008; Symmons et al., 2009), biochemical assays of AcrA-AcrB, AcrA-TolC and AcrA-AcrB complexes show preferred stoichiometry of six AcrAs per AcrB/TolC and fluorescence studies of the homologues MexA–OprM complex observe ranges from

two to six MexAs per OprM (Reffay et al., 2009).

AcrB The acriflavin resistance protein B (AcrB), member of the resistance nodulation division (RND) super family, an efflux pump located in *E. coli.'s* inner membrane, is capable of trans-locating a wide range of chemical compounds. AcrB, initially called AcrE (Ma et al., 1993), was discovered in 1993 and is divided into the TolC docking domain (DD), the pore domain (PD) and the transmembrane domain (TMD) (Nikaido, 2011). The latter is built by twelve transmembrane α-helices and empowers the pump mechanism by proton transfer across the inner membrane (Fischer and Kandt, 2011; Pos, 2009). The PD is divided into the subdomains PN1, PN2, PD1 and PD2 and facilitates the substrate pumping mechanism. The DD is built by the subdomains DN & DC and arranges the docking onto TolC (Murakami et al., 2002).

AcrB is able to pump wide range of substrates including cationic dyes such as acriflavine, crystal violet, ethidium bromide, and rhodamine 6G; antibiotics such as penicillins, cephalosporins, fluoroquinolones, macrolides, chloramphenicol, tetracyclines, novobiocin, fusidic acid, oxazolidinones, and rifampicin; detergents such as Triton X-100, sodium dodecylsulfate, bile acids and even to simple solvents, such as Methodspentane and cyclohexane (Nakajima et al., 1995; Nikaido, 2011; White et al., 1997). Whereas the first available atomistic structure containing conformational identical monomers was solved by (Murakami et al., 2002), in 2006 two asymmetric structures revealed three conformational stages of the pump mechanism: "loose / access", "tight / binding" and "open / extrusion" (Murakami et al., 2006; Seeger et al., 2006).

Driven by proton motive force each monomer undergoes conformational changes into the next state of the transport mechanism to export toxins and drugs from the periplasmic space towards TolC the exit gate leading outwards *E. coli.* (Murakami, 2008).

TolC The tolC gene of *E. coli.* was discovered through its deletion, which leads to a resistance of the bacteria against toxic colicins and was also the origin of its name; ***tol****erence to* ***c****olicins* (Nomura and Witten, 1967). While colicins use TolC as a back door to bypass the outer membrane (Cascales et al., 2007; Zakharov et al., 2004), TolC acts as an efflux duct (ED) for inner membrane transporters of the ATP binding cassette (ABC), resistance nodulation devision (RND) super family and the major facilitator (MF) type (Moussatova et al., 2008; Nikaido, 2011; Saier and Paulsen, 2001).

Assembled as a trimer of 471-residue protomers, TolC is divided in the β-barrel, the α-helical and the equatorial domain (Koronakis et al., 2000). The 40 Å long, twelve stranded, right twisted β-barrel embedded in *E. coli.'s* outer membrane has three lid like loops on extracellular side, which occlude the extracellular entrance of TolC (Vaccaro et al., 2008). The main body of TolC, a 100 Å cylinder, with a

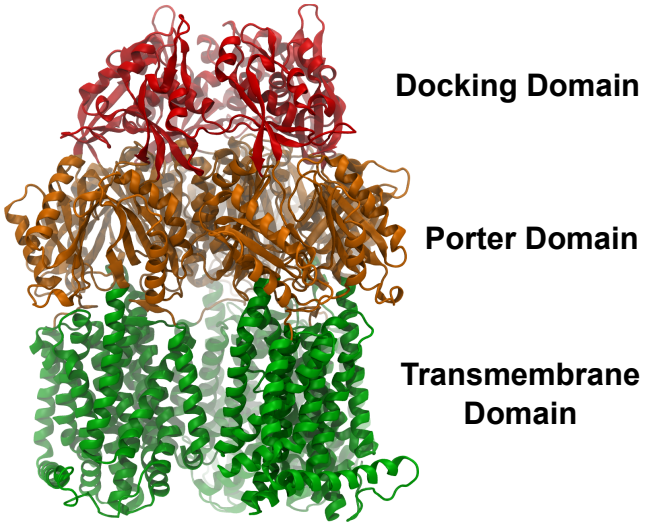

Figure 1.4: AcrB member of the resistant nodulation devision (RND) super family is the inner membrane drug antiporter (IMA) of the AcrA/B–TolC system. AcrB is a trimer and is composed of the transmembrane domain (green), the porter domain (orange) and the docking domain (red).

volume of 43,000 Å^3, ranging from the outer membrane into the periplasm, is built by the left twisted α-helical and the equatorial domain, a mixture of short α-helices and β-strands (Andersen et al., 2001). The lower end of this cylinder is closed by the twisted α-helices.

TolC occurs in at least two conformations, one permitting and one preventing the passage of substrates (Koronakis et al., 2004; Zgurskaya et al., 2011). Combined *in vitro* mutagenesis-conductance experiments revealed two bottlenecks, an inner one (BNI) built by an double aspartate ring at Asp371 and at Asp374 and an outer one (BNII) stabilized by inter- and intra-monomeric hydrogen bonds and salt bridges at Tyr362 and Arg367 (see figure 1.6) (Andersen et al., 2002a,b). The inner bottleneck BNI can be occluded temporarily by trivalent ions and blocked by administration of hexaminecobalt ($Co(NH_3)_6^{3+}$) to the extracellular side (Andersen et al., 2002a; Higgins et al., 2004). Mutations disturbing BNII lead to (a)symmetric "partially open" crystal structures of TolC induced by an outwards shift of helix 7 (Bavro et al., 2008; Pei et al., 2011). MD simulation based-studies on TolC mutants affecting BNII report a heightened flexibility compared to wild-type TolC at the periplasmic mouth region, indicated by the triangular cross sectional area (TCA) of the Gly365 Cα atoms (Vaccaro et al., 2008). Additionally Schulz and Kleinekathöfer (2009) identified a potassium pocket at Thr152, Asp 153 and Asp367Glu in a mutated TolC. A proposed allosteric opening mechanism on periplasmic side is based on an iris-like movement of the inner helices 7 & 8 to realign with the outer helices 3 & 4 (Zgurskaya et al., 2011).

Due to the fact that TolC cooperates with several inner membrane transporters

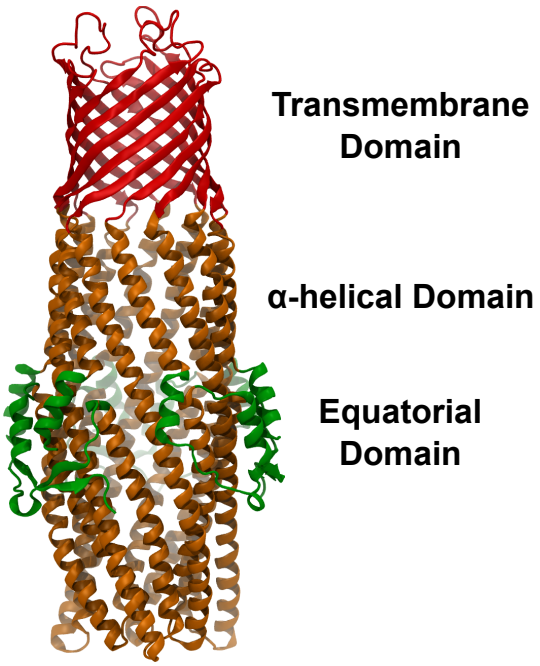

Figure 1.5: The efflux duct (ED) TolC is a homo trimer and consists of a 40 Å β-barrel domain followed by an 100 Å long α-helical domain pointing into the periplasmic space. The α-helical domain is surrounded by the equatorial domain, which is formed by a mixture of α-helices and β-sheets.

TolC becomes to a possible target of drug research. Drugs affecting the gating mechanisms of TolC could interfere its main function. A blocked, disabled or malfunctioning TolC could deny substrate ejection across the outer membrane and therefor disturb efflux systems, which are using TolC as an efflux duct. With the loss of drug efflux functionality bacteria could also loose resistance against several drugs. To target TolC with drugs the knowledge of TolC's gating functionality is a prerequisite and represents one aim of the current work.

1.2.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa belongs like *E. coli.* to the Gram-negative bacteria and is encapsulated by two lipid membranes, an 7.48 nm outer (OM) an 5.95 nm thick inner (IM) one and a 2.41 nm thick peptido-glycan layer in between. IM and OM border the 23.89 nm width periplasmic space (Matias et al., 2003). *Pseudomonas aeruginosa* is facultative anaerobic and can cause diseases in humans and animals. It thrives on most man made surfaces so that this bacterium is also found on medical equipment causing bacterial infections in hospitals (Balasubramanian et al., 2013; Lyczak et al., 2000). Over the past decades several strains of *Pseudomonas aeruginosa* gained resistance to many antibiotics due to intensive treatment with antibiotics (Livermore, 2002). A major role in multidrug resistance of *Pseudomonas*

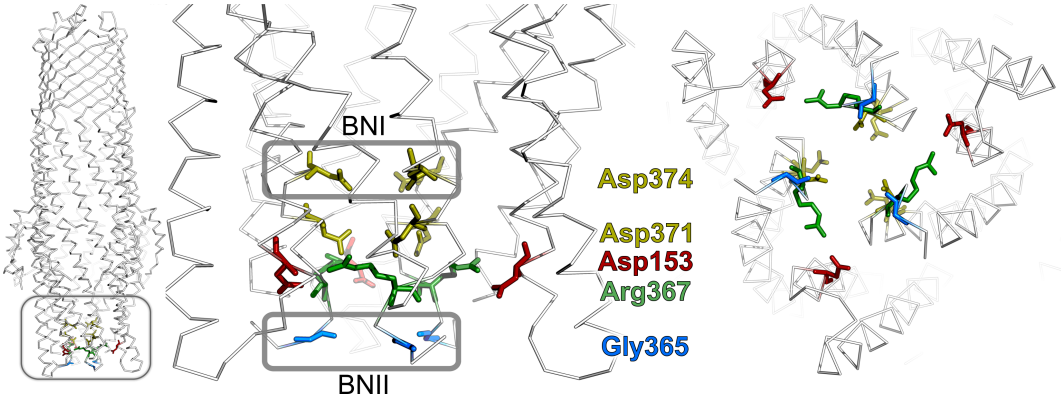

Figure 1.6: Detailed view of the periplasmic region of TolC. Asp371 and Asp374 form an aspartate double ring and build a constriction for ions as described in Andersen et al. (2002b) (BNI). Asp153 and Arg367 stabilize the closed conformation on periplasmic side by an intermonomer salt bridge (Andersen et al., 2002a; Bavro et al., 2008; Pei et al., 2011). Gly365 is located on the loop between helices 3 & 4 and was used as an indicator for the open and close states of the outer bottleneck (BNII) (Bavro et al., 2008; Raunest and Kandt, 2012b; Schulz and Kleinekathöfer, 2009). Opening states of the inner and outer periplasmic bottlenecks were monitored via distant measurements between Asp374 and Gly365 (Bavro et al., 2008; Raunest and Kandt, 2012b; Schulz and Kleinekathöfer, 2009). A more detailed overview of periplasmic key residues can also be found in table 3.2.

aeruginosa plays the MexA/B–OprM efflux system capable to export a broad range of toxic substrates (Lomovskaya et al., 2001).

The MexA/B–OprM Efflux System

The MexA/B–OprM system is a functional homologue to the AcrA/B–TolC system and act as an multidrug efflux protein complex in *Pseudomonas aeruginosa*. The underlying sequence identity between AcrA/B–TolC and MexA/B–OprM is 55%, 69% and 19%, respectively (Phan et al., 2010). While MexA is the adapter protein (AP), MexB acts an inner membrane proton drug antiporter (IMA) expelling chemical compounds towards OprM the efflux duct (ED) (figure 1.2).

OprM The outer membrane protein OprM operates, similar to TolC, as an efflux duct (ED. figure 1.2) for several inner membrane transporters and despite the low sequence identity (19%), the structural similarity with a Cα RMSD of 1.6 Å shows a close relationship between TolC and OprM (Gotoh et al., 1998; Koch et al., 2013; Masuda et al., 2000a; Zhao et al., 1998).

OprM is a 472-residue homo trimer and is composed of an β-barrel transmembrane domain followed by the α-helical domain pointing into the periplasmic space. The equatorial domain consists of a mixture of short α-helices and β-strands and surrounds the α-helical domain. Analogue to TolC, at least two conformations should

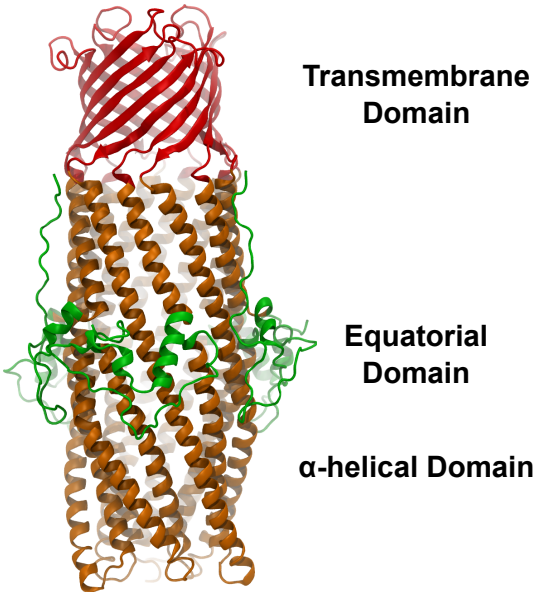

Figure 1.7: The 472-residue homo trimer OprM, a homologue protein of TolC in *Pseudomonas aeruginosa* is formed by the β-barrel transmembrane domain followed by the α-helical domain, which is surrounded by the equatorial domain. In contrast to the available TolC crystal structures, for OprM parts of the membrane anchor could be resolved.

exist for OprM, one permeable and one impermeable for substrates. OprM's access is restricted on extracellular side by three inwards pointing loops and by the coiled-coil α-helices 3,4,7 and 8 on periplasmic side in both available X-ray structures (1WP1, 3D5K) (Akama et al., 2004; Phan et al., 2010).

For the inner membrane-facing coiled-coil helix bundle an iris-like opening was proposed due to the similarity to TolC (Zgurskaya et al., 2011). Twisting and stretching conformational changes were proposed for OprM's gating mechanism based on elastic network normal-mode analysis (Phan et al., 2010). Both proposals do not consider the possible role of ions and their influences on OprM gating and the underlying gating mechanisms still remain unclear.

Similar to TolC, OprM interacts with several inner membrane transporters and becomes through this versatility to a possible target of drug discovery Masuda et al. (2000b)Masuda et al. (2000a). Chemical compounds affecting OprM's gating mechanisms might disturb OprM's role as an exit duct. A blocked or malfunctioning OprM could permit substrate transport across the outer membrane and therefor interfere efflux systems, which are using OprM as an exit gate. With a loss of drug efflux ability bacteria could also loose resistance against drugs. Whereas the understanding of the gating mechanisms of OprM is crucial to develop drugs against OprM, the exploration of these mechanisms are another aim of this work.

1.3 Cavity Detection

In the early 90s of the last century the Human Genome Project started to determine and to identify the human genes and published a complete human base pair sequence in 2001 (Lander et al., 2001). Based on that and partner projects containing the genomes of *E.coli*, the fruit fly and the laboratory mouse sequence analysis began to identify genes, their role and functionality in the organisms.

The products of this sequence oriented research; proteins and peptides can be spatially solved via X-ray crystallography, NMR and electron microscopy. In 1971 the protein data bank (PDB) was established at *Brookhaven National Laboratories* (BNL) as an archive for biological macromolecular crystal structure (Bernstein et al., 1977). In the first decades until the invention of the world wide web and the migration of the data to the *Research Collaboratory for Structural Bioinformatics* (RSCB) in 1998, the usage of the PDB was limited to a small group of researchers involved in structural research via data transfer on magnetic media (Berman et al., 2000).

Alongside the PDB, hosted in the USA, structural databases arose in Europe at the *European Bioinformatics Institute* (EBI) and in Japan at the *Institute for Protein Research*. In 2003 these databases were condensed into the *world wide protein data bank* (wwPDB) a single and uniform database for structural data (Berman et al., 2003). The open access of the wwPDB enables researches around the globe to create structure based studies exploring protein functionality. Over the years the number of solved structures raise up to 89,212 at the time of writing (2013).

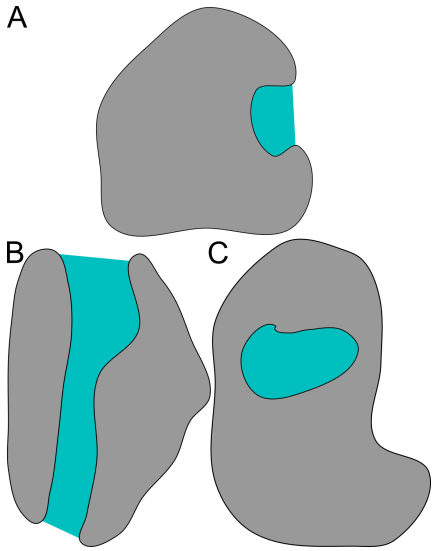

Figure 1.8: Cavity types: Clefts on the protein surface (A), Tunnel like cavities (B), Protein interior cavities (C)

Beside the atomistic details, solved structures (PDB entries) do not describe voids in a direct way. These voids or cavities are often hints for active sites and can

identify open and close states of proteins or explore binding pockets located in clefts of the protein surface (figure 1.8). Thus the knowledge of cavities onside and inside the protein can be very helpful in understanding protein functionality.

Table 1.1: Commonly used cavity detection tools and their methods

		Rolling sphere	Voronoi diagram	Pathway	Fill voids
3V	(Voss and Gerstein, 2010)	x			
Cast	(Liang et al., 1998)		x		
CAVER 1&2	(Petrek et al., 2006)		x	x	
Chunnel	(Coleman and Sharp, 2009)	x		x	
dxTuber	(Raunest and Kandt, 2011a)				x
Grasp	(Nicholls et al., 1991)	x			
HOLE	(Smart et al., 1996)			x	
Hollow	(Ho and Gruswitz, 2008)	x			x
LIGSITE	(Hendlich et al., 1997)				x
MOLE	(Petrek et al., 2007)		x	x	
MoleAxis	(Yaffe et al., 2008b)		x	x	
PROPORES	(Lee and Helms, 2011)				x
PoreWalker	(Pellegrini-Calace et al., 2009)		x	x	
PyMOL	(Delano, 2002)	x			
SURFNET	(Laskowski, 1995)				x
unnamed	(Exner et al., 1998)				x
unnamed	(Kim et al., 2008)	x	x		
Voidoo	(Kleywegt and Jones, 1994)				x

1.3.1 Geometric Cavity Detection Methods

Over the last decades a wide range of prediction tools for active sites and cavity detection tools have been introduced. While tools like SILCS (Guvench and MacKerell, 2009) and Q-SiteFinder (Laurie and Jackson, 2005) use chemical properties of the specimen to identify active sites and binding pockets, cavity detection tools use geometrical approaches to explore cavities in their entirety. Based on different geometrical approaches cavity detection tools can be divided into four major groups (see table 1.1). In the following a brief overview of these groups is given and prominent members are listed in table 1.1.

Rolling Sphere Approach

The rolling sphere approach introduced by Richards (1977) creates a molecular surface representation by rolling a probe sphere over the van der Waals (vdW) representation of atoms. In 1983 Connolly (1983b) developed an algorithm to determine the three dimensional surface representation of the solvent accessible surface (SAS) and of the solvent excluded surface (SES) for molecules (figure 1.9).

Cavity detection tools applying this approach typically use a mixture of varying probe radii in combination of SAS- and SES-analysis to identify cavities (Delano, 2002; Voss and Gerstein, 2010) or combine this approach with voronoi or pathway approaches as described in (Coleman and Sharp, 2009; Ho and Gruswitz, 2008; Kim et al., 2008).

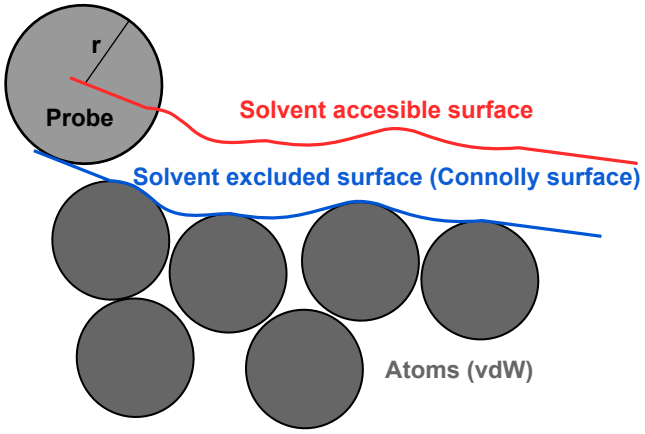

Figure 1.9: In order to create a surface representation a probe sphere is rolled over the van der Waals (vdW) representation of a given structure. While the surface described by the edge of the sphere generates the solvent excluded surface (SES) or Connolly surface, the surface based on the center of the probe is defined as solvent accessible surface (Richards, 1977).

Voronoi Diagram Approach

A Voronoi diagram consists of:

- A set of seeds $A = \{a_1, a_2, \ ... \ a_n\}$
- Voronoi regions $VR = \{VR_1, VR_2, \ ... \ VR_n\}$
- Voronoi edges $E = \{e_1, e_2, \ ... \ e_n\}$
- Voronoi vertices $V = \{v_1, v_2, \ ... \ v_n\}$

Based on the set of seeds A a Voronoi diagram divides a given space into Voronoi regions VR. Each region has a central seed $a_i \in A$ and consists of all points, which are closer to a_i than to other seeds. Two neighboring Voronoi regions $VR_i, VR_j \in VR$ are limited by a Voronoi edge $e_i \in E$, which is equidistant to the seeds $a_i, a_j \in A$ at all of its points. Voronoi edges E are limited by Voronoi vertices V, which are equidistant to three neighboring Voronoi seeds $a_i, a_j, a_k \in A$ (figure 1.10).

Voronoi diagrams are often used to identify possible paths through voids located inside proteins (Chovancova et al., 2012). Some algorithms combine pathway approaches with Voronoi diagrams. Beginning from an initial coordinate algorithms follow the Voronoi edges in order to find possible exit paths from the initial coordinate (Chovancova et al., 2012; Petrek et al., 2007; Yaffe et al., 2008b). Beside the exit path approaches Voronoi methods are combined with rolling sphere approaches to identify pockets on the protein surface (Kim et al., 2008).

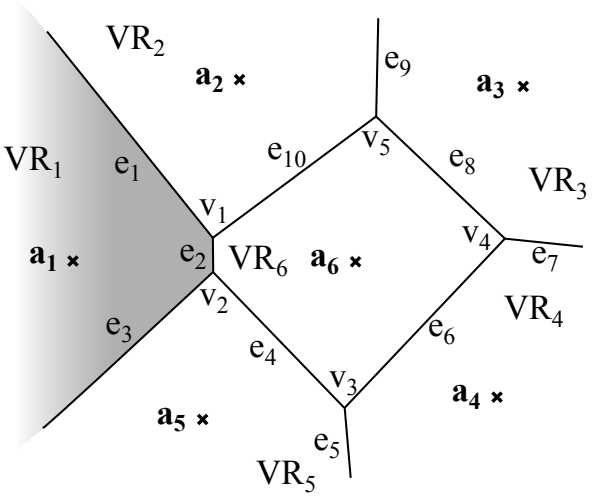

Figure 1.10: Simplified 2D representation of a Voronoi diagram for the six seeds a_1 to a_6 containing ten Voronoi edges e_1 to e_{10}, five Voronoi vertices v_1 to $_{10}$ and six Voronoi regions VR_1 to VR_6. The Voronoi region VR_1 is defined by its central seed a_1 and all points that are closer to a_1 than to other seeds. Voronoi regions VR_1 and VR_2 are limited by the Voronoi edge e_1, which is equidistant to the seeds a_1 and a_2. the Voronoi edge e_1 is limited by the Voronoi vertex v_1, which is equidistant to the seeds a_1, a_2 and a_6.

Pathway Approach

Pathway approaches explore cavities via paths and place spheres along the discovered path to fill and describe the underlying cavity. These paths are either detected via cost functions (Pellegrini-Calace et al., 2009; Petrek et al., 2006) or via Voronoi diagrams (Chovancova et al., 2012).

Some tools also need parameters like user specified initial coordinates of the cavity and number of exit paths and analyze only one cavity at once (Petrek et al., 2006; Yaffe et al., 2008a).

Fill Voids Approach

Tools, using the filling voids approach, usually fill cavities via grid based algorithms. Grid points are classified as protein internal (cavity) or protein external points. Depending on the method a range of altering sphere radii (Kleywegt and Jones, 1994; Laskowski, 1995) or voxels or spheres of the same size (Exner et al., 1998; Hendlich et al., 1997; Ho and Gruswitz, 2008; Lee and Helms, 2011; Raunest and Kandt, 2011a) are used to fill the initially detected cavities.

1.3.2 Dynamics and Cavity Detection Tools

The increasing use of MD simulations of molecular complexes discovered more and more the underlying dynamics of proteins and peptides and indirectly also the dynamics of their voids. MD simulations provide an ensemble of protein conformations, which could be used to describe cavities over time. For example active membrane transporters undergo a set of stages to trans-locate substances across a membrane. These conformational changes have an effect on its cavities, which were used to distinguish between open and close states of a transporter (Kandt and Tieleman, 2010). Since no cavity detection tool based on dynamics was available at that time, Kandt and Tieleman (2010) calculated mass weighted protein density (MPD) maps for time frames of different MD simulation setups and identified open and close states of the transporter through MPD slices and residues located at the entrance region. This cavity detection has the disadvantage that one MPD slice only represents only one layer of protein density and that the detection was done by hand. One aim of this work is to develop a new cavity detection tool based on solvent and protein dynamics since all prior introduced cavity detection tools do not consider protein and water dynamics during their cavity analysis and thus cavity detection tools are limited to static structures.

1.4 Objectives of this Work

1.4.1 TolC – Gating

The first aim of this work is to explore TolC's gating mechanisms on extracellular and periplasmic side. For this purpose multicopy (Caves et al., 1998; Das et al., 2000; Kandt and Tieleman, 2010; Kandt et al., 2006) unbiased molecular dynamics (MD) simulations were calculated for membrane embedded TolC in a NaCl / water environment (Raunest and Kandt, 2012b). On extracellular side TolC opens and closes freely suggesting the absence of a gating mechanism. On periplasmic side initially open motions were measured at the tip loops of TolC (BNII). The successive binding of two sodium ions in one simulation after 90 ns of simulation time triggered the closing of BNII and caused a more closer conformation of TolC compared to available X-ray structures. The inner bottleneck formed by an Asp double ring (Asp371 and Asp374) did not open up in the simulations unless all NaCl was removed. This removal of all ions induced also reopening of BNII and a slight opening of BNI pointing to a sodium dependent lock on the periplasmic side.

As a second step we introduced, to the prior TolC setup, the docking domain of AcrB (AcrB-DD) and placed it 1 nm apart from TolC based on orientations of a docked model published by Symmons et al. (2009). Multicopy MD simulations were calculated and whereas in three simulations AcrB-DD drifted apart in one the AcrB-DD docked onto TolC and results in a tighter binding after 1.05 μs of simulation time. On extracellular side again TolC opens and closes freely and moreover prefers an open conformation in presence of the AcrB-DD on extracellular side. On periplasmic side the AcrB-DD stabilized the open conformation of the tip loops between helix 7 & 8 but neither opened nor relax the inner bottleneck (BNI) formed by Asp371 and Asp374 meaning that either longer simulation time is needed or another part is needed to open TolC on periplasmic side.

1.4.2 OprM – Gating

Analogue to the first aim of this work unbiased multicopy MD simulations were calculated for OprM in a membrane / NaCl–water environment. Similar to TolC OprM opens and closes freely on extracellular side, which also negate a gating function on extracellular side.

However, on periplasmic side a different behavior in comparison to TolC was noticed for OprM. After the initial opening of the tip loops neither heighten sodium densities nor closing motions was noticed during the simulations. Moreover, the inner bottleneck located at the aspartate ring located at Asp416 opens slightly during all simulations. Additionally a new sodium hotspot located between the equatorial domain and the aspartate ring was identified. The absence of closing motions for the lower tip loops of OprM and the low sodium densities hints for a different gating functionality in comparison to TolC.

1.4.3 Cavity Detection

The third aim of this work is to develop a novel cavity detection tool, which takes protein and solvent dynamics into account. All cavities of a macromolecular complex should be detected during one analysis and should be saved distinguishable. Additional further, results should be stored in a well known file format in order to allow users to post process resulting cavity data. For this purpose it was assumed, that solvent molecules of MD trajectories can be used for cavity detection. To put it at its simplest: Solvent molecules highlight biomolecular cavities through their diffusion during a MD simulation and the separation of internal solvent molecules from exterior ones represent the distribution of cavities inside a biomolecular structure. For this purpose we developed dxTuber a novel cavity detection tool (Raunest and Kandt, 2011a).

Using the VolMap plugin of VMD (Humphrey et al., 1996) average mass weighted spatial density maps of solvent and protein are calculated from trajectories and stored in the OpenDX file format in a grid representation. dxTuber uses these density maps as input for cavity detection. In a first step interior solvent voxels (ISVs) are separated from exterior solvent voxels (ESVs). As a second step adjacent ISVs are assembled to cavities. These resulting cavities can be stored either individually in OpenDX or in PDB file format. In the latter case individual cavities are separated via the *atom name field* of the PDB file and solvent densities are stored as B-factors.

After cavity detection dxTuber can extract cavity volume and the cross-sectional area along a principle axis. For convenience dxTuber analysis can be executed via graphical user interface (GUI) or via command line interface (CLI). Both versions of dxTuber are licensed under the GPL v2[1] and can be downloaded at `http://www.csb.bit.uni-bonn.de/dxtuber.html` or at `http://code.google.com/p/dxtuber/`.

[1] `http://www.gnu.org/licenses/licenses.html#GPL`

Chapter 2

Methods

The current chapter describes the theoretical background of molecular dynamics (MD) simulations. Detailed descriptions of the applied simulations can be found in chapters 3, 4, 5 and 6.

2.1 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations describe the movement of atoms in a given biomolecular simulation system based on Newtons equation of motions $F = m * a$ with F as force acting on an atom m as mass and a as the acceleration. The forces acting on each atom can be divided in internal and external ones. While bonds, (dihedral-) angles, electrostatics and van der Waals (vdW) interactions belong to the internal forces, pressure and temperature are part of the external forces. The main intention of MD simulations is to gain and quantify macromolecular dynamics, which can neither be achieved nor observed in classical experimental procedures in such detail.

Based on collisions Alder and Wainwright (1957) introduced in the late 1950's the first molecular dynamics algorithm. From that on many MD packages have been developed (Brooks et al., 2009; Case et al., 2005; Hess et al., 2008; Johnston et al., 2005; Phillips et al., 2005); computers become faster and algorithms were improved fulfilling the requirements for establishing MD simulations of soluable proteins (Karplus and McCammon, 2002) and membrane proteins (Gumbart et al., 2005; Kandt et al., 2007). Furthermore groups reach simulation times of milliseconds for small molecules (Shaw et al., 2009) and an atomistic simulation of a complete virus of more than 50 ns (Freddolino et al., 2006).

2.1.1 Equations of Motion

Molecules can be described by the Schrödinger equation,

$$\mathbf{H}\Psi(\mathbf{R}, \mathbf{r}) = E\Psi(\mathbf{R}, \mathbf{r}) \tag{2.1}$$

with **H** as the Hamilton operator considering all atomic nuclei of the system **r** and their electrons **R**, the wave function Ψ regarding the dynamics, **R** and **r** representing the positions of the electrons and nuclei and E the energy of the system. Equation 2.1 can only be solved for the hydrogen atom and approximations are necessary to describe bigger systems. Since the nucleus of an atom represents more than 99% of the atom mass the electrons move much faster than the nucleus. Born and Oppenheimer divided equation 2.1 into *electronic* and *nuclear* components and built the basis for computational simulation through their approximation (Born and Oppenheimer, 1927).

The force F_i acting on each atom i can be calculated by taking the negative gradient of the potential energy function ν_i. The latter one consists of a set of equations that empirically describe (non-) bonded interaction, artificial restrains and external forces like barostats and thermostats. ν_i is called *force field* (equation 2.2) (Kandt et al., 2007).

$$\begin{aligned} F_i &= -\nabla\nu_i \\ \nu_i &= \sum \nu_{bonded\ interactions} + \sum \nu_{non-bonded\ interactions} \\ &\quad + \sum \nu_{external\ forces} + \sum \nu_{restraints} \end{aligned} \tag{2.2}$$

The application of these forces molecules allows to calculate the atom movement over time based on Newton's law of motion, with F_i as force, m_i as the mass of the atom i and a_i as acceleration.

$$F_i = m_i * a_i \tag{2.3}$$

Starting from the atomic positions $r_i(t)$, the forces $F_i(t)$ acting on each atom at the time t are calculated for all atoms i. By knowing the forces $F_i(t)$, the accelerations a_i can be calculated. Based on the assumption, that the forces $F_i(t)$ acting on the atoms i are constant during the time step Δt and initial velocities at $t_0 = 0$ are obtained by Gaussian or Boltzman distributions to achieve a simulation system at a desired temperature, new atom positions $r_i(t + \Delta t)$ and velocities $v_i(t + \Delta t)$ are calculated. Several on Taylor series based algorithms have been invented to integrate the equations of motions (Hockney, 1970; Schofield, 1973; Verlet, 1967). The most common one is the *verlet leap frog algorithm* from Hockney (1970). Based on the initial velocities at $t_0 = 0$, it calculates stepwise coordinates $r_i(t + \Delta t)$ and new velocities v_i for each atom i for each timestep $t + \frac{1}{2}$ based on the velocities at $t - \frac{1}{2}$ as described in equations 2.4. Thus the atomistic movements can be calculated over small increments of time.

$$\begin{aligned} v_i\left(t + \frac{1}{2}\Delta t\right) &= v_i\left(t - \frac{1}{2}\Delta t\right) + \Delta t\ a(t) \\ r_i\left(t + \Delta t\right) &= r_i(t) + \Delta t\ v_i\left(t + \frac{1}{2}\Delta t\right) \end{aligned} \tag{2.4}$$

The vibration of covalently bond hydrogens represents the fastest movements in biological MD simulations and limits the integration step Δt to 1 fs. To achieve

longer MD simulations methods like SHAKE (Ryckaert et al., 1977) and LINCS (Hess et al., 1997) eliminate this degree of freedom and allow an integration step size of 2 fs for all atom simulations.

2.1.2 Force Field and Potential Energy Function

Through force fields the forces acting on each atom i are calculated at each time step. A force field is built by a sum of bonded-, non-bonded- and optional restraining-potential energy functions ν describing the interactions between all atoms i based on experimental measured reference values and empirically determined force constants (equation 2.5).

$$\nu(i) = \underbrace{\sum_i \nu_{bonds} + \sum_i \nu_{angles} + \sum_i \nu_{dihedrals}}_{bonded\ interactions} + \underbrace{\sum_i \nu_{electrostatic} + \sum_i \nu_{van\ der\ Waals}}_{non-bonded\ interactions} + \sum_i \nu_{restraints} \tag{2.5}$$

Bonded Interactions

Covalent bonds fluctuate around an average length (Haynes, 2012) and are modeled as harmonic potentials, that describe the increase of energy as the bond length b_{ij} deviates from its reference value b_{ij_0}. The empirical determined force constant k^b_{ij} describes the rise of the function dependent on atom types of atoms i and j (equation 2.6 and figure 2.1a).

$$\sum \nu_{bonds}(ij_n) = \sum_{n=1}^{bonds} \frac{1}{4} k^b_{ij} \left(b^2_{ij_n} - b^2_{ij_0}\right)^2 \tag{2.6}$$

Analogue to bonds, angles θ_{ijk} also fluctuate around a reference θ^0_{ijk} and are also described by an empirical determined force constant κ^θ_{ijk} of involved atoms i, j and k (equation 2.7 and figure 2.1b).

$$\sum \nu_{angles}(\theta_{ijk_n}) = \sum_{n=1}^{angles} \frac{1}{2} \kappa^\theta_{ijk} \left(\cos\left(\theta_{ijk_n}\right) - \cos\left(\theta^0_{ijk}\right)\right)^2 \tag{2.7}$$

Torsions along a covalent bond are implemented via the angle ϕ_{ijkl} between two planes, which are spanned by the atoms i, j, k and l (see figure 2.1c), the empirical determined force constant k^ϕ_{ijkl}, m the number of minima between 0 to 360 degrees and the first minima given by ϕ^0_{ijkl} (equation 2.8 and figure 2.1c).

$$\sum \nu_{dihedrals}\left(\phi_{ijkl_n}\right) = \sum_{n=1}^{dihedals} k^\phi_{ijkl} \left(1 + \cos\left(m\phi_{ijkl_n} - \phi^0_{ijkl}\right)\right) \tag{2.8}$$

Improper dihedral angles ξ_{ijkl} are used to stabilize planar groups like aromatic rings and peptide bonds. ξ_{ijkl} is spanned by the atoms i, j, k and l. Atoms $i-l$ are either consecutively bond in case of the peptide bond or atoms $j-l$ are bound radial to atom i for ring structures. The improper angle ξ_{ijkl} is located between the planes of i, j, k and j, k, l and is implemented as a harmonic potential with ξ^0_{ijkl} as the reference angle and $\kappa_{\xi_{ijkl}}$ as the empirical force constant (equation 2.9 and figure 2.1c, d).

$$\sum \nu_{improper\ dihedrals}\left(\xi_{ijkl_n}\right) = \sum_{n=1}^{\substack{improper\\dihedrals}} \frac{1}{2}\kappa_{\xi_{ijkl}}\left(\xi_{ijkl_n} - \xi^0_{ijkl}\right)^2 \tag{2.9}$$

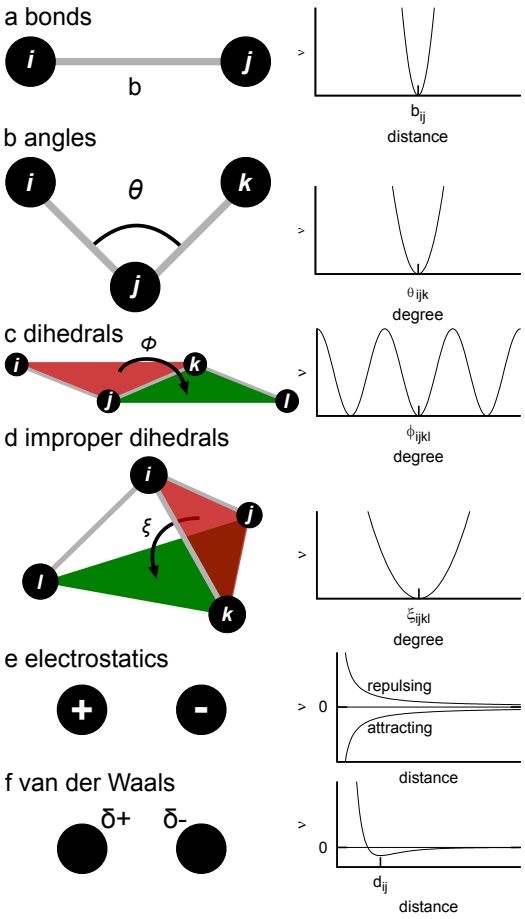

Figure 2.1: Bonded interactions (a-d) are implemented in GROMACS (Berendsen et al., 1995a) via harmonic (a,b,d) and periodic potentials (c). Coulomb- (e) and Lennard-Jones potentials (f) describe the non-bonded interactions (e,f).

Non-bonded Interactions

Non-bonded interactions are divided into electrostatics and van der Waals (vdW) interactions. Electrostatics describe attractive and repulsive interactions between charged atoms and are modeled using Coulomb potentials with the atomic charges q_i and q_j, their distance d_{ij} and ϵ_0 the dielectric constant (equation 2.10 and figure 2.1e).

$$\sum \nu_{Coulomb}\left(d_{ij_n}\right) = \sum_{n=1}^{pairs(ij)} \frac{q_i q_j}{4\pi\epsilon_0 d_{ij_n}} \tag{2.10}$$

In atoms the asymmetric distribution of electrons induce transient dipoles, which cause short range electrostatics. These electrostatics, also called van der Waals interactions and are divided in an attractive and repulsive term. While the approximation of two dipoles represents the attractive term, the repulsive one describes the repelling force when the distance of atoms is to small and electron radii overlap. vdW forces can be described via Lennard-Jones potentials (Lennard-Jones, 1931) with $\left(\sigma_{ij}/d_{ij}\right)^{12}$ as the repulsive and $\left(\sigma_{ij}/d_{ij}\right)^{6}$ as the attractive term, σ_{ij} the distance at which the potential is zero, ε the depth of the potential well and d_{ij} the distance between atoms i and j (equation 2.11 and figure 2.1f).

$$\sum \nu_{Lennard-Jones}\left(d_{ij_n}\right) = \sum_{n=1}^{pairs(ij)} 4\varepsilon \left(\left(\frac{\sigma_{ij}}{d_{ij_n}}\right)^{12} - \left(\frac{\sigma_{ij}}{d_{ij_n}}\right)^{6} \right) \tag{2.11}$$

2.1.3 Restraints

Potentials can additionally be added to the force field either to avoid unwanted movements or to steer the simulation system. For membrane proteins position restraints are a common method to equilibrate the membrane after the insertion of the protein into the membrane. In this case position restraints are applied on the protein for a short MD simulation in order to relax the membrane. Position restraints are described by the reference position R_i, the force constant k_i and the current position r_i of the atom i (equation 2.12).

$$\sum \nu(r_{i_n})_{postition\ restraints} = \sum_{n=1}^{restraints} \frac{1}{2} k_{i_n} \left| r_{i_n} - R_{i_n} \right|^2 \tag{2.12}$$

2.1.4 Periodic Boundary Conditions

In computational simulations the "world" or the experiment is limited by a simulation box and within this box edge effects will distort results more and more over time. To eliminate this error translated copies of the simulation box are placed surrounding the original box effecting that atoms passing an edge will enter on an opposite side into the original box (figure 2.2). The box size or the unit cell should be

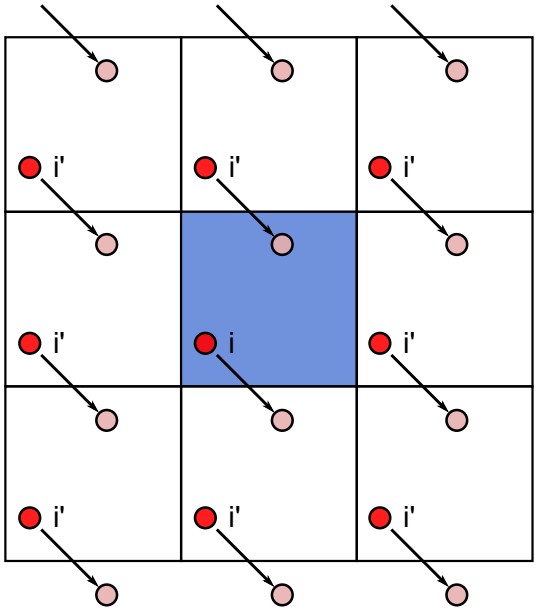

Figure 2.2: Two dimensional representation of periodic bondary conditions (PBC). Atoms that leave a simulation system a side enter the original system on the opposite side.

chosen wisely, because within a too small box protein atoms interact with translated copies from other boxes. These interactions can be avoided by a bigger simulation box, but the increase of the box size causes also an increase of computational time for a simulation due to the increased amount of solvent inside the system.

2.1.5 External Forces

The described force field implementation (section 2.1.2) assumes so far a constant number of particles (**N**), a constant volume (**V**) and energy (**E**), while pressure (**P**) and temperature (**T**) are variable. This **NVE** simulating engine is not compatible with experiments *in vivo*, where usually pressure and temperature are constant. To achieve a simulation engine with a constant number of particles (**N**), a constant pressure (**P**) and constant temperature (**T**) temperature and pressure need to be controled by the simulation engine. The kinetic energy E_{kin} of a system can be calculated at each timestep based on the product of the masses m_i and their velocities v_i. This energy is translated into temperature via the Boltzmann constant k and the degrees of freedom N_{df} (equation 2.13).

$$\begin{aligned} E_{kin} &= \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 = \frac{1}{2} N_{df} kT \\ \frac{dT}{dt} &= \frac{T_0 - T}{\tau} \end{aligned} \tag{2.13}$$

Constant temperature is achieved by means of coupling the system to an external temperature bath, as introduced by Berendsen et al. (1984). Here the temperature T is slowly corrected based on a reference external bath T_0 with the time constant τ. The modified temperatures result in a rescaling of the velocity vectors v_i (equation 2.13).

The pressure P of the entire system is calculated via the difference of the virial tensor (Clausius, 1870) and the kinetic energy $\mathbf{E}_{kin}$ and the volume V of the system. The virial itself is defined by the sum of the dot products of the atom positions r_i and the acting forces F_i (equation 2.14).

$$\begin{aligned} \mathbf{P} &= \frac{2}{V}\left(\mathbf{E}_{kin} - \Xi\right) \\ (excluding\ pbc:)\ \Xi &= -\frac{1}{2}\sum_{i=1}^{N} r_i \otimes F_i \\ P &= \frac{trace\left(\mathbf{P}\right)}{3} \end{aligned} \tag{2.14}$$

Analogue to temperature (equation 2.13) pressure is kept constant via an external *reference pressure bath* P_0 leading to rescaled atom positions r_i (equation 2.15).

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_p} \tag{2.15}$$

2.2 Energy Minimization

An initial system usually contains unfavored atom contacts which lead to a high kinetic energy of involved atoms. Through energy minimization algorithms these locally misplaced atoms can be rearranged.

The combination of atom positions of the initial system represents a point on a potential energy landscape. The *steepest descent method* (Petrova and Solov'ev, 1997) follows iteratively the steepest descent in a zig-zag manner to find a local minimum on this potential energy landscape. The method is a common energy minimization algorithm to minimize the potential energy of an initial system. For this purpose the *steepest descent method* is applied on the potential energy surface to find a local minimum.

2.3 Workflow of an MD Simulation

MD simulations are based on static structures gained either from experimental procedures such as X-Ray crystalography, nuclear magnetic resonance spectroscopy (NMR), electron microscopy or from theoretical models like molecular modeling, ab initio modeling or structures taken from previously calculated MD simulations.

Then the compounds of interest are inserted into a simulation box. In case of an membrane protein the protein needs to be embedded into a lipid bilayer. Then the system is salvated into a water and salt solution. The next step is to minimize the potential energy of the system to eliminate sterical clashes. For membrane proteins an additional membrane equilibration simulation, introducing position restraints for the protein, over several nanoseconds is recommended. Afterwards the main loop of an MD simulation can be executed:

1. Calculate the potential energy functions
2. Determine forces acting on the atoms
3. Calculate the velocity vectors
4. Scale velocity vectors by means of temperature coupling
5. Compute new atom coordinates
6. Scale atom coordinates by means of pressure coupling
7. Append coordinates, forces and velocities to output files (if required)
8. If end of simulation is not reached: go to 1
9. Store coordinates, forces and velocities in output files

2.4 GROMACS

The Groningen Machine for Chemical Simulations (GROMACS) is a molecular dynamics (MD) package capable to run MD simulations on symmetric multiprocessing (SMP) systems, clusters, supercomputers and on graphics processing units (GPU) via Open MPI and OpenMM (Eastman and Pande, 2010; Gabriel et al., 2004; Hess et al., 2008). Beside the simulation routines, a set of tools for the analysis of resulting simulation data are part of the GROMACS package.

While initially GROMACS was released by Herman Berendsen's group, department of Biophysical Chemistry of Groningen University (Berendsen et al., 1995a), today a global community under the leadership of *Stockholm Center for Biomembrane Research & Biomedical Centre* in Uppsala develops GROMACS. It is licensed as free software under the GNU General Public License[1] and can be compiled on Macintosh, Unix/Linux and Windows machines.

[1] http://www.gnu.org/licenses/licenses.html#GPL

2.5 Molecular Viewer

In the field of MD simulations molecular viewers are used to generate protein representations to illustrate results and to discuss biological implications. Beside the spatial representation of biomolecular structures, most viewers also contain a broad set of tools for further analysis like e.g. distance measurements, cavity detection or RMSD calculations.

2.5.1 PyMOL

PyMOL is an open source molecular viewer and was introduced in 2000 by Warren Lyford DeLano (Delano, 2002). In this study many images were rendered via PyMOL (e.g. all CAVER & SURFNET representation in chapter 6).

2.5.2 VMD

Visual molecular dynamics (VMD) is a molecular viewer and acts as a graphical interface for NAMD (Not (just) Another Molecular Dynamics program Phillips et al. (2005)). It was introduced by Humphrey et al. (1996) and is distributed free of charge for non-commercial use. In the current work most images were rendered via VMD and most angles and atom distances were monitored with VMD (see chapters 3, 4 and 5 for detailed descriptions). Density maps created by the VolMap plugin of VMD are mandatory input files of the cavity detection tool dxTuber, which is introduced in chapter 6.

Chapter 3

Locked on One Side Only: Ground State Dynamics of the Outer Membrane Efflux Duct TolC

The outer membrane channel TolC acts as an efflux duct for several inner membrane transporters. This versatility makes TolC to a possible target of drug research. A blocked, disabled or malfunctioning TolC could deny substrate export across the outer membrane and therefor disturb efflux systems which are using TolC as an efflux duct. The understanding of TolC's gating functionality is a prerequisite to target TolC with drugs. Hence we examined TolC's gating mechanism on extracellular and periplasmic side in the current chapter.

Adapted from:
Raunest, M. and Kandt, C. (2012b). Locked on one side only: ground state dynamics of the outer membrane efflux duct tolc. *Biochemistry*, 51(8):1719–1729

3.1 Abstract

Playing a major role in the expulsion of antibiotics and the secretion of cell toxins in conjunction with inner membrane transporters of three protein superfamilies, the outer membrane channel TolC occurs in at least two states blocking or permitting the passage of substrates. The details of the underlying gating mechanism are not fully understood. Addressing the questions of extracellular access control and periplasmic gating mechanism, we conducted a series of independent, unbiased 150 - 300 ns molecular dynamics simulations of wild-type TolC in a phospholipid membrane/150 mM NaCl water environment.

We find that TolC opens and closes freely on the extracellular side, suggesting the absence of a gating mechanism on this side in the isolated protein. On the periplasmic side, we observe the outer periplasmic bottleneck region adopting in all simulations a conformation more open than the TolC wild-type crystal structures until in one run the successive binding of two sodium ions induces the transition to a conformation more closed than any of the available TolC X-ray structures. Concurrent with a heightened sodium residence probability near Asp374, the inner periplasmic bottleneck region at Asp374 remains closed throughout the simulations unless all NaCl is removed from the system, inducing a reopening of the outer and inner bottleneck. Our findings suggest that TolC is locked only on the periplasmic side in a sodium-dependent manner.

3.2 Introduction

A member of the outer membrane efflux protein (OEP) family, TolC is a channel protein in the outer membrane of *Escherichia coli* (Koronakis et al., 2004; Zgurskaya et al., 2011) facilitating the transport of various toxic molecules (Benz et al., 1993; Fralick, 1996). Playing a major role in the expulsion of antibiotics (Fralick and Burns-Keliher, 1994; Sulavik et al., 2001) and the secretion of cell toxins (Delgado et al., 1999; Haynes, 2012; Wandersman and Delepelaire, 1990), TolC is highly versatile, functioning in conjunction with inner membrane transporters of the ATP binding cassette (ABC), resistance nodulation division (RND), or major facilitator (MF) type (Moussatova et al., 2008; Nikaido, 2011; Saier and Paulsen, 2001). Interaction of the inner membrane transporter and outer membrane efflux duct is mediated by specialized periplasmic adaptor or membrane fusion proteins anchored to the inner membrane (Dinh et al., 1994; Zgurskaya, 2009).

A prominent example of a TolC-dependent transport system is the AcrAB-TolC multidrug efflux pump employing the energy of a proton gradient over the inner membrane to power the expulsion of a broad range of substrates from the cell (Nikaido, 2011). Although predominantly involved in substrate export, TolC is also known to be used as a backdoor by cell toxins like colicins to penetrate the outer membrane (Cascales et al., 2007; Zakharov et al., 2004). Whereas channels like the cell toxin α-hemolysin, whose individual and unfolded subunits are also exported through TolC (Zgurskaya et al., 2011), are constantly open, TolC access is regulated with the protein occurring in at least two states permitting or blocking the passage of substrates (Koronakis et al., 2004; Zgurskaya et al., 2011).

The details of the underlying gating mechanism are not fully understood (Zgurskaya et al., 2011). The first crystal structure of the wild-type protein (Koronakis et al., 2000) revealed that TolC is a homotrimer, resembling the shape of a hollow cylinder organized in three domains. From a membrane-embedded β-barrel, cooperatively formed by all three monomers, an α-helical domain extends $\approx$ 100 Å into the periplasmic space surrounded in its middle section by the ringlike structure of an α/β equatorial domain (Koronakis et al., 2004; Zgurskaya et al., 2011).

TolC access was found to be restricted on both ends with three extracellular loops oriented toward the center of the β-barrel, limiting access to the channel to compounds with diameters of ¡ 7 - 8 Å in the wild type (Higgins et al., 2004; Koronakis et al., 2000) and 7.2 - 10 Å in the mutant X-ray structures (Bavro et al., 2008; Pei et al., 2011). On the periplasmic side, the dense packing of the tip regions of 12 α-helices constricts the TolC channel to a diameter of $\approx$ 4.4 Å in the wild type (Higgins et al., 2004; Koronakis et al., 2000) and 4 - 7.8 Å in the mutant crystal structures (Bavro et al., 2008; Pei et al., 2011).

Whereas the occurrence of gating on the extracellular side is unclear (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011), combined mutagenesis-conductance experiments (Andersen et al., 2002a,b) and subsequent

wild-type (Higgins et al., 2004) and mutant TolC crystal structures (Bavro et al., 2008; Pei et al., 2011) indicated the existence of two periplasmic bottlenecks: an inner bottleneck (BNI) at Asp371 and Asp374 where the binding of TolC-blocking hexaamminecobalt occurs (Andersen et al., 2002b; Higgins et al., 2004) and an outer bottleneck (BNII) at Tyr362 and Arg367 involved in inter- and intramonomeric hydrogen bonds and salt bridges stabilizing a closed conformation (Andersen et al., 2002a). Single and double mutations of these residues lead to increased ion conductance (Andersen et al., 2002a) as well as crystal structures showing partial symmetric (Pei et al., 2011) and asymmetric (Bavro et al., 2008) opening toward the periplasm (table 3.1).

Computational studies of TolC focused on molecular dynamics (MD) simulations comparing the wild type and BNII mutants (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008), as well as elastic network normal-mode analyses exploring possible opening mechanisms (Phan et al., 2010). In a 20 ns MD study of wild-type and Tyr362Phe/Arg367Ser TolC, the mutant was reported to exhibit heightened flexibility in the periplasmic mouth region, while for the extracellular loops, a gating function was proposed on the basis of the observed closing motions (Vaccaro et al., 2008). Another study compared wild-type, Tyr362Phe/Arg367Glu, and Tyr362Phe/Arg367Asp TolC in a series of 20 - 30 ns MD simulations reporting wild-type-like closed periplasmic mouth conformations stabilized by potassium ions coordinated by Thr152, Asp153, and Glu/Asp367 in the mutant structures. Only when the potassium binding sites were emptied using an outer electric field was a BNII opening trend observed (Schulz and Kleinekathöfer, 2009). Beyond TolC, other multidrug efflux pump components have also recently been investigated computationally (Fischer and Kandt, 2011; Phan et al., 2010; Schulz and Kleinekathöfer, 2009; Schulz et al., 2011; Vaccaro et al., 2006; Vargiu et al., 2011; Yao et al., 2010).

Addressing the questions of extracellular access control and periplasmic gating mechanism, we performed nine independent, unbiased MD simulations of membrane-embedded wild-type TolC in a 150 mM NaCl solution sampling TolC conformational dynamics on a 150 - 300 ns time scale. Opening and closing freely on the extracellular side, TolC opens in the BNII region on the periplasmic side until the successive binding of two sodium ions preferably interacting with Asp371, Thr366, Thr368, and Asp153 induces closure.

The resulting BNII conformation is more closed than any of the available crystal structures. Concurrent with a third site of heightened sodium residence probability at Asp374, TolC remains closed in the BNI region unless the removal of all NaCl from the system induces an opening response of BNI followed by a reopening of BNII. Displaying a so far unreported high degree of conformational dynamics in the channel mouth regions, our findings suggest that TolC is locked only on the periplasmic side in a sodium-dependent manner.

3.3 Experimental Procedures

Molecular Dynamics Simulations. MD simulations were performed employing GROMACS version 4.0.3 (Berendsen et al., 1995a; Hess et al., 2008) and the GROMOS96-53a6 force field (Oostenbrink et al., 2004), using the 1EK9 TolC crystal structure (Koronakis et al., 2000) as a starting structure. The protein was inserted in a pre-equilibrated 9.6 nm x 9.6 nm palmitoyloleoylphosphatidylethanolamine (POPE) bilayer patch (Tieleman and Berendsen, 1998) using INFLATEGRO (Kandt et al., 2007). The system was solvated with simple point charge water molecules (Berendsen et al., 1981) and 170 Na^+ and 152 Cl^- ions, yielding a 150 mM NaCl solution and a total system charge of zero (figure 3.1A).

Standard protonation states were assumed for titratable residues. After a 20 ns membrane equilibration with protein atoms position-restrained using a force constant of 1000 $\frac{kJ}{mol\ nm^2}$, five independent MD runs with different starting velocities were performed, each 150 ns in length. Of these initial simulations, one was extended in four independent copies to 300 ns: two copies without any modifications, one copy with four Na^+ ions removed, and one copy from which all NaCl ions had been deleted. In the simulations, all bond lengths were constrained by LINCS (Hess et al., 1997) so that an integration time step of 2 fs could be chosen.

Systems were simulated at 310 K, maintained separately for protein, lipids, and water by a Berendsen thermostat (Berendsen et al., 1984) with a time constant (τ_T) of 0.1 ps. Pressure coupling was done employing a Berendsen barostat (Berendsen et al., 1984) using a 1 bar reference pressure and a time constant of 4 ps. Semi-isotropic pressure coupling was employed to permit bilayer fluctuations in the membrane plane. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation (Darden et al., 1993; Essmann et al., 1995), and twin range cutoffs of 1.0 and 1.4 nm were applied for computing the van der Waals interactions.

Analysis

Similar to the analysis described in ref (Schulz and Kleinekathöfer, 2009), the TolC opening state on the periplasmic side was monitored by calculating the triangular cross-sectional area (TCA) defined by the α-carbons of Asp374 and Gly365. On the extracellular side, the opening state was monitored through the dihedral angle θ spanned by the Cα atoms of Asp56 located in the β-barrel and Ala270 at the tip of each extracellular loop (figure 3.2). As the outwardly closed 1EK9 crystal structure displays an average θ value of 89.8° (figure 3.2A), an extracellular loop was regarded as closed as or more closed than 1EK9 whenever $\theta \leq 90°$. Conversely, a loop for which $\theta > 90°$ was considered more open than the 1EK9 X-ray structure.

In each simulation, θ was monitored for each extracellular loop (figure 3.2B, first panel) and subsequently converted to a binary representation of "more open" than 1EK9 (a value of 0) or "more closed" (assigned a value of 1) (figure 3.2B, second

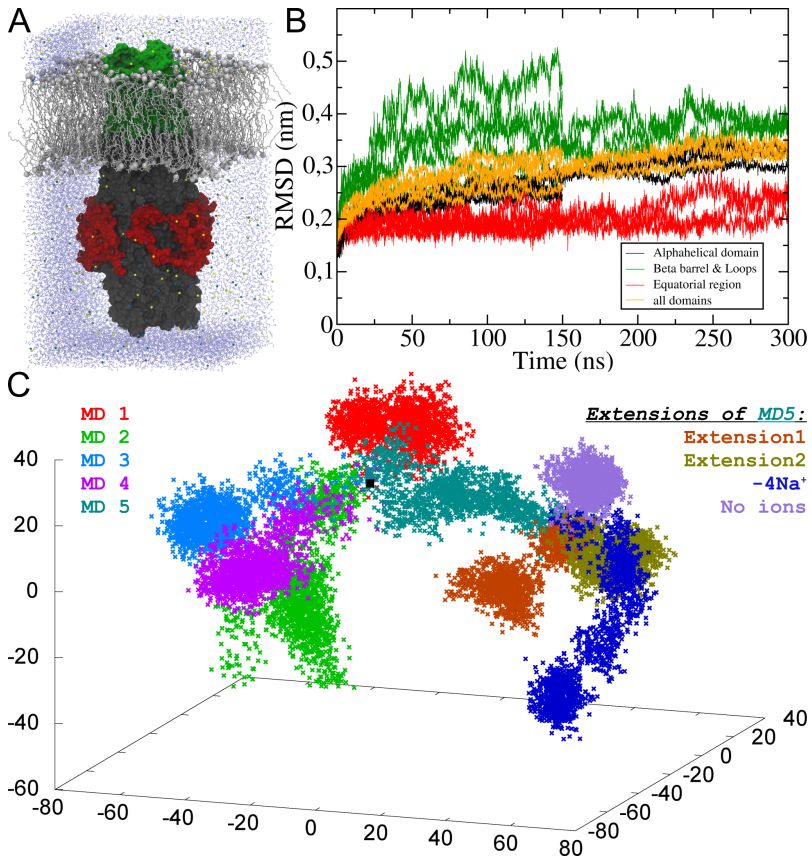

Figure 3.1: (A) Wild-type 1EK9 TolC was simulated in a POPE phospholipid/water environment at 150 mM NaCl. (B) As indicated by Cα root-mean-square deviations after respective least-squares alignment with the starting structure, the largest conformational changes occur in the region of the β-barrel and the extracellular loops. (C) Using the protein's α-carbons, TolC's path through conformational space in all simulations was mapped onto the first three eigenvectors in a principal component analysis.

panel). As a summary for each run, the number of loops in the closed conformation was determined (figure 3.2B, third panel) and converted to a frequency histogram (figure 3.2B, fourth panel) reflecting for each simulation the percentage occurrence of closed loop conformations (figure 3.2C). To obtain an overview of the extracellular loop conformations visited throughout our simulations, we calculated a histogram showing the θ distribution in all unmodified runs (figure 3.2E).

To analyze the distribution of sodium throughout the simulations, both one-dimensional (1D) Na^+ density profiles along the membrane normal and average spatial sodium distributions were computed. The former was done using the GROMACS tool g_density, and for the latter, we employed the VolMap function in VMD version 1.9 (Humphrey et al., 1996) using a spatialresolution of 1 $Å^3$ to analyze the distribution of sodium in the periplasmic bottleneck region at five density levels

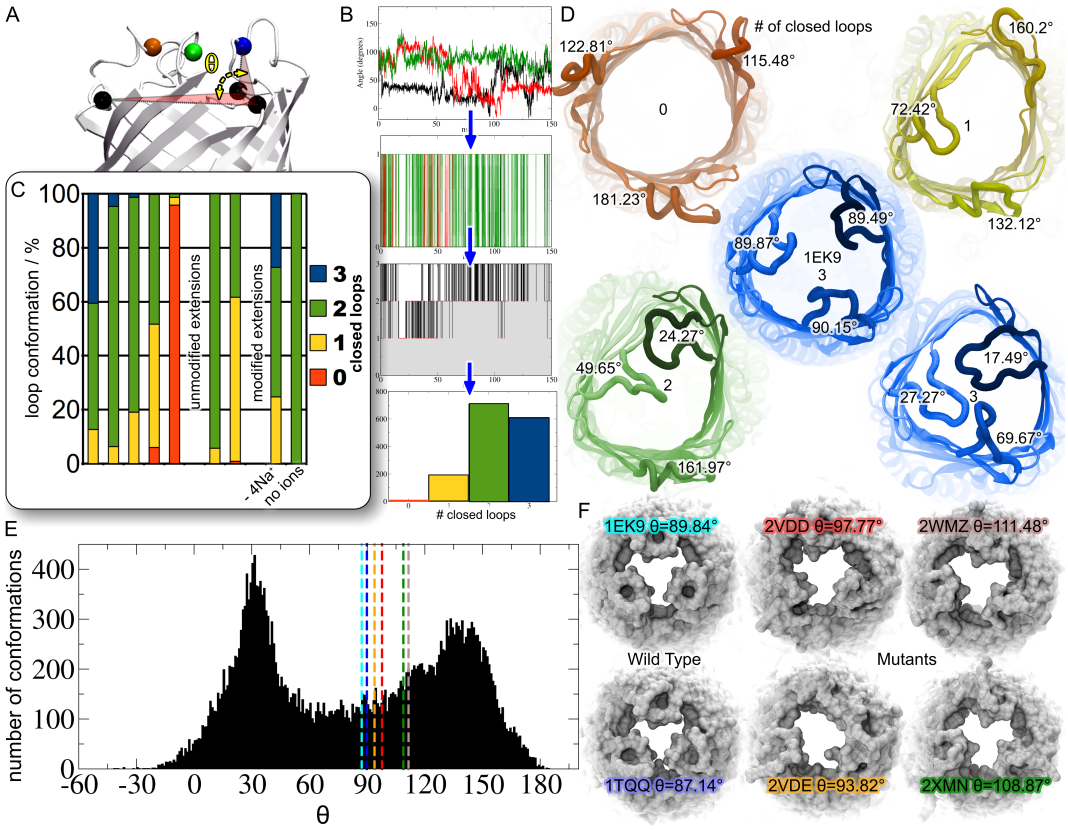

Figure 3.2: On the extracellular side, the TolC opening state was monitored using the dihedral angle θ defined by the α-carbons of Asp56 in the β-barrel and Ala270 in the tip of each extracellular loop. As the 1EK9 starting structure exhibits θ dihedrals of 89.8° (A), an extracellular loop was regarded closed or more closed than 1EK9 when θ was $\leq 90°$ (B). For each simulation, we calculated the number of closed extracellular loops and the percentage of simulation time spent in that conformation (C). Whereas TolC opens and closes throughout the simulations as illustrated by X-ray and simulation snapshots (D), partially open with one or two loops closed is the preferred conformation. As indicated by a θ histogram calculated over all unmodified simulations (E), the palette of extracellular loop conformations visited throughout our simulations exceeds the range of conformations observed in the known TolC X-ray structures (F).

ranging from 0.01 to 0.6 $Na^+/Å^3$.

Sodium binding and unbinding were monitored analyzing the Z-coordinate trajectory of each Na^+ ion and computing for each TolC monomer the percentage of simulation time a residue of the periplasmic mouth region (residues 134 - 158 and 352 - 374) comes into contact with at least one sodium ion in at least one monomer. For this analysis, we employed a residue-Na distance cutoff of ≤ 3 Å. Both the sodium distributions and the residue contact analyses were performed using time frames of (I) 1 - 150 ns for the four simulations during which the outer periplasmic

bottleneck (BNII) was open, (II) 125 - 300 ns for the two simulations during which BNII was closed, and (III) 150 - 300 ns for the run during which BNII remained closed after removal of four Na^+ ions from the bottleneck region.

As described in Fischer and Kandt (2011), simulation average structures were calculated using an iterative scheme of calculating the average conformation and realigning the trajectory to that average structure before computing a new average structure. This procedure was repeated until the average structure stopped changing.

Table 3.1: TolC Opening States As Observed in the Crystal Structures and Our Molecular Dynamics Simulations

	#closed	bottleneck Gly365 (BNII)		bottleneck Asp374 (BNI)	
TolC structure	loops‡	⌀ TCA† (Å^2)	⌀ Cα distance (Å)	⌀ TCA† (Å^2)	⌀ Cα distance (Å)
X-ray					
1EK9, wild type	3	64.4	12.2	58.6	11.6
2VDD, Y362F/R367E	0	203.8	21.8	76.7	13.8
2VDE, Y362F/R367E	2	202.1	21.6	71.2	13.8
2WMZ, R367S	0	154.8	18.9	65.7	12.3
2XMN, Y362F/R367S	0	195.6	21.3	99.5	15.2
1TQQ, wild-type-bound hexaamminecobalt	3	61.3	11.9	64.6	12.2
unmodified Simulations					
BNII open	0-3	195.1 ± 42.6	21.6 ± 3.4	64.4 ± 8.2	12.3 ± 1.1
		(max of 334.8)	(max of 31.3)	(max of 98.4)	(max of 17.3)
BNII closed	0-2	32.0 ± 17.7	11.7 ± 1.6	58.7 ± 7.4	11.7 ± 1.2
		(min of 0.9)	(min of 3.3)	(min of 40.2)	(min of 8.3)
modified Simulations					
-4 Na^+, BNII closed	1-3	19.3 ± 7.9	10.7 ± 1.2	64.7 ± 6.7	12.3 ± 0.6
		(min of 0.9)	(min of 3.3)	min of 46.0)	(min of 9.0)
no NaCl, BNII closed	2	66.8 ± 31.7	14.4 ± 2.5	88.7 ± 11.6	14.4 ± 0.9
BNI and BNII reopening		(max of 133.5)	(max of 25.6)	(max of 125.3)	(max of 19.3)

† TCA is the triangular cross-sectional area
‡ As closed as or closer than 1EK9 with $\theta \leq 90°$

3.4 Results

Protein Stability and Conformational Sampling. To monitor protein stability and conformational sampling throughout the simulations, we computed for each run Cα root-mean-square deviations (RMSD) of the entire protein, the β-barrel with extracellular loops, the α-helical domains, and the equatorial domain after least-squares fitting to the α-carbons of the crystal structure (figure 3.1B).

With RMSDs ranging from 3.5 to 5.1 Å, the largest conformational changes occur in the β-barrel-extracellular loop regions, whereas the smallest changes take place in the equatorial domains with RMSDs between 2 and 2.5 Å. Both α-helical domains and the entire protein display intermediate RMSDs of 3 - 3.5 Å. Within 300 ns, no stable RMSD plateau is reached. As is evident from the Cα-based principal component analysis (figure 3.1C) in each simulation, TolC samples different regions of conformational space.

3.4.1 Gating

Extracellular Side

To analyze TolC's opening state on the extracellular side, we monitored the dihedral angle θ formed by the Cα atoms of Asp56 in the β-barrel and Ala270 at the tip of each extracellular loop (figure 3.2). Using the 1EK9 starting structure's average θ dihedral of 89.8° as a reference, we determined the number of loop conformations as closed or more closed than the aforementioned crystal structure.

Throughout our simulations, the extracellular loops adopt a variety of conformations (figure 3.2C) ranging from zero to three loops in the closed conformation (figure 3.2D). As indicated by the percentage occurrence of the number of closed loop conformations in each run, "partially open" with one or two loops closed is the preferred conformation in our simulation (figure 3.2C). Ranging from -30° to 180° with two distinctive peaks at 30° and 135°, the histogram of distributions calculated for all unmodified simulations (figure 2E) indicates that the range of θ dihedrals in our simulations exceeds the range of 87.1 - 111.5° observed in all known wild-type and mutant TolC crystal structures (figure 3.2F).

Periplasmic Side

On the periplasmic side, we monitored the TolC opening state by calculating the triangular cross-sectional area (TCA) spanned by the Cα atoms of Asp374 and Gly365 (figure 3.3) representing the inner (BNI) and outer periplasmic bottleneck region (BNII) (figure 3.3A). In all runs, BNII opens within 10 ns with the TCA subsequently fluctuating around 195 ± 43 Å^2 (figure 3.3B and table 3.1). Concurrent with the successive binding of two sodium ions, in one simulation the TCA decreases in two steps below the 1EK9 crystal structure's TCA of 64.4 Å^2 (table 3.1), reaching a new average plateau of 32 ± 18 Å^2 in the two unmodified extensions (figure 3.3B, bold dark green lines, and table 3.1) and 19 ± 8 Å^2 in the extension where four sodium ions were removed from the bottleneck regions (figure 3.3B, dashed green line, and table 3.1).

When all NaCl is removed from the system, the TCA increases again, reaching 105 Å^2 by the end of the simulations (figure 3.3B, dotted green line). With average BNI TCAs of 64 ± 8 Å^2 (BNII open) and 59 ± 7 Å^2 (BNII closed) in the unmodified simulations (figure 3.3B, bold red line, and table 3.1) and 65 ± 7 Å^2 after the removal of four Na^+ atoms from the bottleneck region (figure 3.3B, dashed red line, and table 3.1), the inner bottleneck remains closed displaying a TCA similar to that of the 1EK9 X-ray structure except for the NaCl-free extension: here the TCA increases to 125 Å^2, settling at 80 Å^2 by the end of the simulations (figure 3.3B, dotted red line).

Whereas figure 3.3C shows the conformation of BNI and BNII in the crystal structure, panels D-F of figure 3.3 are simulation snapshots illustrating the maximal

(figure 3.3D) and minimal BNII opening (figure 3.3E) of 334 and 0.9 Å^2, respectively (table 3.1), as well as a maximal BNI opening of 125 Å^2 after the complete removal of NaCl (figure 3.3F and table 3.1).

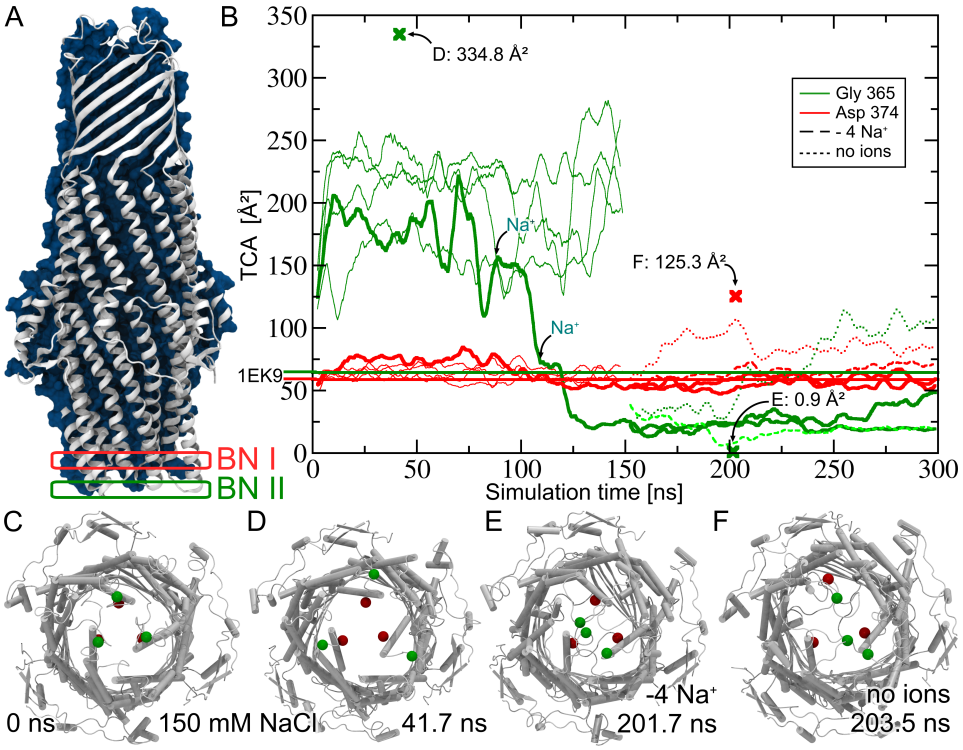

Figure 3.3: On the periplasmic side, the TolC opening state was monitored using the triangular cross-sectional area (TCA) spanned by the α-carbons of Asp374 and Gly365 representing the inner (BNI) and outer bottleneck region (BNII) (A). BNII visits open and closed conformations, with maximal closure occurring after a successive binding of two sodium ions (B). Simulation snapshots (C-F) illustrate periplasmic TolC conformations in the X-ray structure (C), at the maximal opening (D) and closure (E) of BNII at Gly396, and at the maximal opening of BNI, 53 ns after the removal of all NaCl from the system (F). For the sake of clarity, the TCAs have been smoothed using a running average filter of 5 ns. Bold crosses mark the unfiltered TCA maxima and minima corresponding to the simulation snapshots shown in panels D-F.

Sodium Distributions

To determine the distribution of sodium ions and check for potential binding sites, we calculated 1D Na^+ density profiles along the Z-axis (figure 3.4A) and computed average spatial Na^+ density distributions focusing on sodium in the periplasmic bottleneck regions (figure 3.4B-D). Computed separately for the open and closed BNII conformations, the results were averaged over the respective runs using time frames of 0 - 150 ns (BNII open), 125 - 300 ns (BNII closed), and 150 - 300 ns (BNII

closed after the removal of four Na^+ atoms from the bottleneck region).

As indicated by the peaks in the sodium density profile (figure 3.4A), two preferred Na^+ residence sites become apparent: the lipid headgroups and the region of the periplasmic bottlenecks where the density maximum is most pronounced in the unmodified simulations when BNII is closed. Zooming in on the bottleneck region, we find a heightened three-dimensional Na density near Asp374 and Asp371 exceeding 0.1 $Na^+/Å^3$ regardless of whether BNII is open (figure 3.2B) or closed (figure 3.2C,D).

When BNII is closed, three density maxima become apparent in the unmodified runs (figure 3.2C): two distinctive ones exceeding 0.6 $Na^+/Å^3$ at Asp371, Thr368, Thr366, and Asp153 and a smaller one between 0.2 and 0.3 $Na^+/Å^3$ near Asp374. When BNII is closed and four sodium ions are removed from the bottleneck region (figure 3.2D), the 0.1 $Na^+/Å^3$ density is more smeared out and the former maxima are less pronounced.

3.4.2 Sodium Binding

Monitoring the Z-trajectories of all sodium ions in the system, we find that next to the lipid headgroup region an additional sodium trace becomes apparent near 14 nm, forming after 100 ns and remaining stable until the end of the simulations (figure 3.5A). Zooming in on this region, we find that starting at 94 ns in the unmodified simulations the sodium trace is at first formed by single Na^+ ions (figure 3.5B,C), whereas after 120 ns, two Na^+ ions contribute simultaneously to the trace with individual sodium residence times ranging up to 180 ns. In the simulation where all (four) sodium ions have been removed from the periplasmic bottleneck region after 150 ns, the sodium trace reappears after 5 ns (figure 3.5D). During the remaining 145 ns of simulation time, single-ion occupancy is predominant and individual Na^+ residence times do not exceed 30 ns. Via comparison of the Na^+ trajectories with the BNII closure observed in the TCA analysis (figure 3.3), it becomes evident that the binding of the first stable sodium at 94 ns precedes the first decrease in TCA in BNII closure by 8 ns, whereas the binding of the second sodium at 118 ns is concurrent with the beginning of the second decrease in TCA completing BNII closure.

To determine which residues are involved in sodium binding, we calculated for all residues in the periplasmic tip region their frequency of sodium contact (figure 3.5E). Exceeding in at least one monomer 15% of observation time, Asp371, Thr366, Thr368, Asp153, and Asp374 were identified as main sodium interaction sites with ion residence times ranging from 20 to 82%. Whereas Asp374 showed in all simulations similar Na^+ contact times in all monomers, the other residues display an asymmetry in Na^+ interaction characterized by Na^+ residence times clearly higher in one monomer than in the other two.

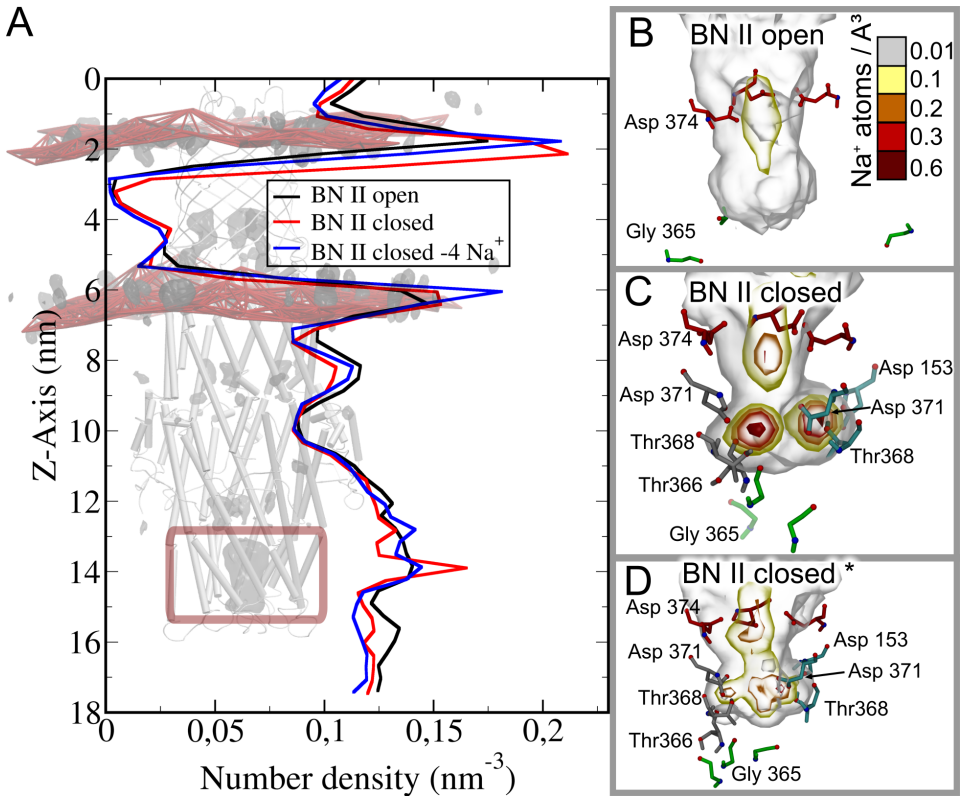

Figure 3.4: As indicated by partial sodium densities (A) calculated over all simulations where the outer bottleneck BNII is open (black) or closed during the unmodified runs (red) or after removal of the initially bound sodium ions (blue), there are two preferred sodium binding sites: at the lipid headgroups and between BNI and BNII. Spatial sodium density distributions in this area (B-D) indicate that Na densities exceeding 0.01 $Na^+/Å^3$ (transparent white isosurfaces) are present regardless of whether BNII is open or closed (B-D). When BNII is closed and the simulation system is unmodified, two distinctive maxima of Na^+ density are present at Asp371, Thr368, Thr366, and Asp153, suggesting distinctive interaction sites (C). A third maximum of smaller density occurs near Asp374. After the removal of four sodium ions from the bottleneck regions, the 0.1 $Na^+/Å^3$ density is more smeared out and the former maxima are less pronounced (D). However, the reoccurrence of maximal sodium densities in this region suggests an immediate reoccupation of the interaction sites.

3.4.3 Simulation versus X-ray

Throughout our simulations, wild-type TolC adopts conformations on the extracellular and periplasmic side that have not been reported in the published crystal structures (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011). To provide possible evidence explaining this discrepancy, we calculated the simulation average structure over all unmodified simulations and compared it to the 1EK9 crystal structure (figure 3.6). As indicated by superimposition of the simu-

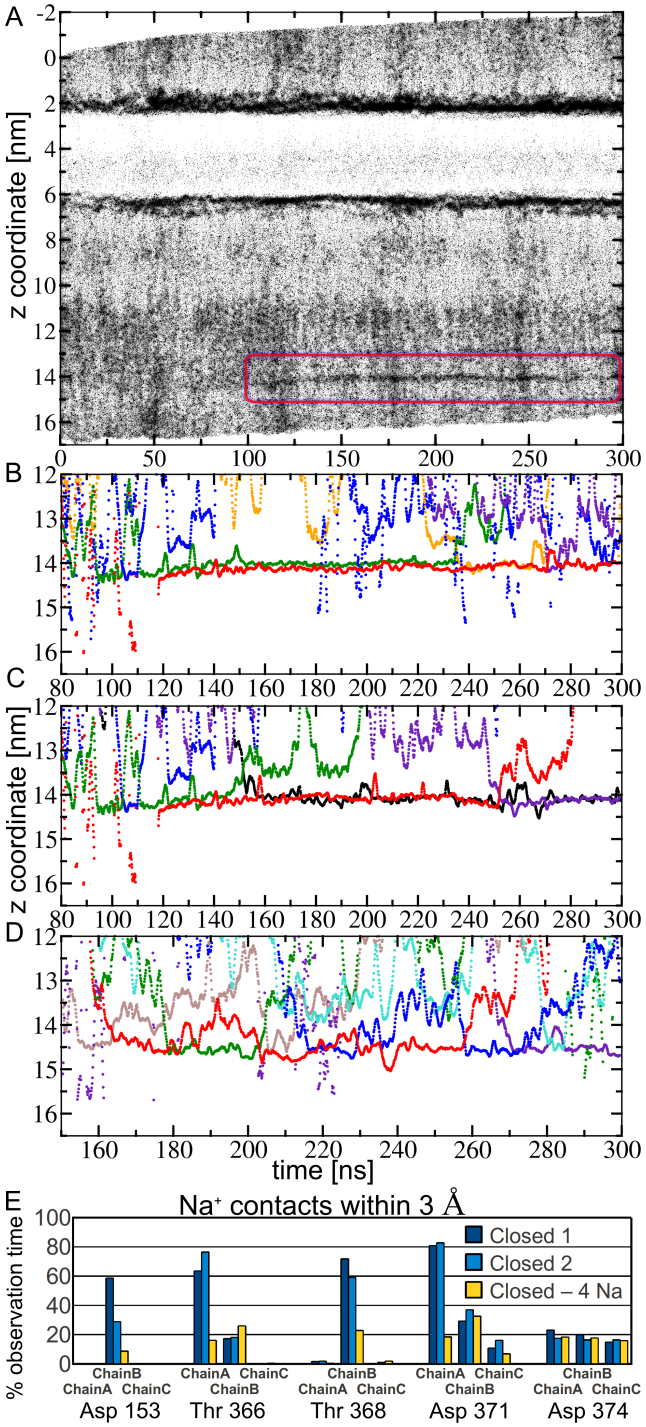

Figure 3.5: Sodium Z-trajectories (A-D) and frequency of Na^+ residue contact (E). Monitoring the Z-coordinate of each sodium ion in the system, we find next to the lipid headgroups another preferred sodium residence site becomes apparent as indicated by a stable sodium trace near 14 nm forming after 100 ns (A). The following three panels are close-ups of this region showing individual sodium ions contributing to the stable Na^+ trace in the unmodified simulations (B and C) and in the run following the removal of four Na^+ ions from the bottleneck region after 150 ns (D). After 90 ns, at least one Na^+ is present between 14 and 14.5 nm, whereas in the unmodified simulations after 120 ns, this region is preponderantly occupied by two Na^+ ions (B and C) whose individual residence times range up to 180 ns. When Na^+ is removed from the bottleneck region, the Na^+ interaction region is reoccupied within 5 ns. Now, however, single-ion occupancy is predominant, and individual occupancy times are much shorter (D). Sodium-interacting residues exhibiting Na^+ contacts of a minimum of 15% observation times in at least one monomer are summarized in panel E.

lation average and X-ray structure after Cα least-squares fitting (figure 3.6A) and calculation of Cα displacements for each monomer (figure 3.6B), with an overall RMSD of 1.5 Å, the conformational differences are small and the largest deviations occur in the extracellular loops and the periplasmic tip region.

Residues exceeding Cα displacements of 3 Å (highlighted in dark cyan in figure

3.6C) partially coincide with 4 Å crystal contacts (van der Waals representation in figure 3.6C) in the 1EK9 X-ray structure. Additionally, we have analyzed the B-factors in all wild-type and mutant TolC X-ray structures and plotted their Cα root-mean-square fluctuations along the membrane normal (figure 3.7).

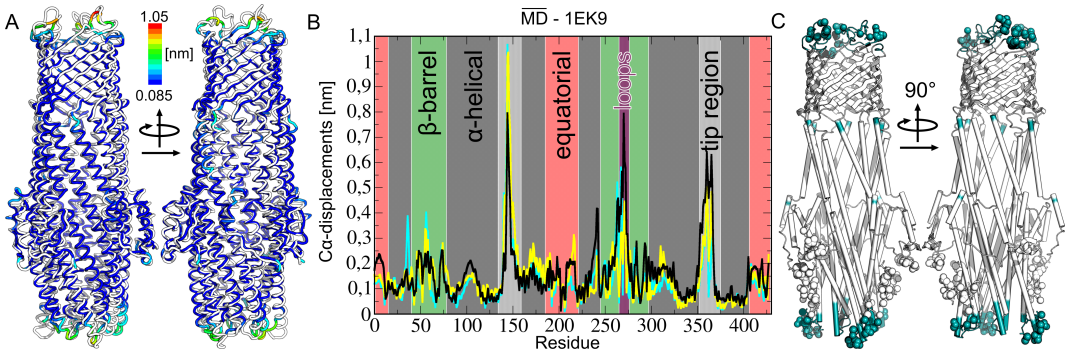

Figure 3.6: As indicated by the superposition of the 1EK9 X-ray structure (white) and TolC simulation average structure calculated over all unmodified simulations and colored by Cα displacement (A) and Cα displacements graphs for each monomer (B), the largest conformational differences occur in the extracellular loops and the periplasmic tip regions. The overall Cα RMSD between the simulation average and crystal structre is 1.5 Å. Residues exceeding Cα displacements of 3 Å (dark cyan) partially coincide with 4 Å crystal contacts (van der Waals representation) in the 1EK9 X-ray structure (C).

3.5 Discussion

In this study, we report unbiased 150 - 300 ns MD simulations of wild-type TolC in a phospholipid/water environment in which TolC visits conformations on the extracellular and periplasmic side that have not been observed in any of the available wild-type and mutant TolC crystal structures (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011). We find TolC freely opening and closing on the extracellular side as well as in the region of the outer periplasmic bottleneck (BNII), monitored at Gly365, until the subsequent binding of two sodium ions interacting with Asp371, Thr366, Thr368, and Asp153 induces a closure resulting in a BNII conformation more closed than any of the available crystal structures. In the inner bottleneck (BNI) region at Asp374, TolC remains closed unless all NaCl is removed from the system, inducing an opening response of BNI followed by a reopening of BNII. We begin this section discussing the limitations of our approach and then proceed to our findings and their biological implications.

Table 3.2: TolC Residues Affecting Conductance and Tertiary Structure and Involved in Ligand Binding

Residue	single-channel conductance	X-ray crystallography	MD simulation
Thr152			part of the potassium binding pocket (Schulz and Kleinekathöfer, 2009)
Asp153	D153A increases conductance (Andersen et al., 2002a)		part of the potassium binding pocket (Schulz and Kleinekathöfer, 2009)
			high sodium binding affinity leads to a closure of BNII (this work)
Tyr362	Y362F increases conductance (Andersen et al., 2002a)	Y362F/R367E double mutants in 2VDD and 2VDE show an asymmetrically open BNII (Bavro et al., 2008) Y362F/R367S double mutants in 2WMZ and 2XMN show a symmetrically open BNII (Pei et al., 2011)	Y362F/R367S double mutant increases flexibility in the tip region (Vaccaro et al., 2008)
Thr366			high sodium binding affinity leads to a closure of BNII (this work)
Arg367	R367S increases conductance (Andersen et al., 2002a)	Y362F/R367E double mutants in 2VDD and 2VDE show an asymmetrically open BNII (Bavro et al., 2008) Y362F/R367S double mutants in 2WMZ and 2XMN show a symmetrically open BNII (Pei et al., 2011)	part of the potassium binding pocket (Schulz and Kleinekathöfer, 2009)
Thr368			high sodium binding affinity leads to a closure of BNII (this work)
Asp371	D371A increases conductance(Andersen et al., 2002b)	binding site for hexaamminecobalt(Higgins et al., 2004)	high sodium binding affinity leads to a closure of BNII (this work)
Asp374	D374A increases conductance(Andersen et al., 2002b)	binding site for hexaamminecobalt(Higgins et al., 2004) Asp374 ring as an indicator for BNI closure (Pei et al., 2011)	high sodium binding affinity (this work)

† TCA is the triangular cross-sectional area
‡ As closed as or closer than 1EK9 with $\theta \leq 90^\circ$

3.5.1 Limitations of Our Approach

As an outer membrane protein, TolC's microenvironment is characterized by a heterogeneous and asymmetric lipid composition containing lipolysaccharides in the outer leaflet, while the protein interacts with the peptidoglycan layer on the periplasmic side (Lugtenberg and Alphen, 1983). In our simulations, we approximated the outer membrane by a homogeneous POPE bilayer omitting a representation of the peptidoglycan layer. While it would certainly be desirable to have both components fully included in simulation studies, the development of the appropriate lipopolysaccharides and peptidoglycan molecular topologies is still in its infancy (Kotra et al., 1999; Lins and Straatsma, 2001; Meroueh et al., 2006; Piggot et al., 2011; Straatsma and Soares, 2009; Vollmer and Bertsche, 2008; Vollmer et al., 2008a), and at the time of writing, such simulation parameters were not yet available for *E. coli*. It will be interesting to see the influence these components have on the dynamics of outer membrane proteins especially because all crystallographic (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011) and TolC conductance experiments (Andersen et al., 2002a,b) were conducted on the isolated protein.

On the other hand, our choice of a simple uniform membrane model exclud-

ing the peptidoglycan is also justified by other simulation studies of outer membrane proteins making biologically relevant predictions using similar simplifications (Cuesta-Seijo et al., 2010; Gumbart et al., 2009; Hajjar et al., 2010). Whereas one could argue that the TolC behavior we observe on the extracellular side could be an artifact of our simplified membrane model, it is encouraging that one of the first lipopolysaccharide simulations studying outer membrane protein OprF (Straatsma and Soares, 2009) reported that the lipopolysaccharids have a stabilizing effect on the open conformation of the OprF extracellular loops that exhibit an architecture similar to that in TolC.

With any molecular dynamics study, the question of whether the simulated time has been sufficiently long with respect to the problem under investigation arises. Given the findings reported by Grossfield and Zuckerman (Grossfield and Zuckerman, 2009) that 1.6 μs atomistic MD simulation of membrane-embedded rhodopsin was not enough for the protein structure to converge, our observations that TolC Cα RMSDs are still increasing after 300 ns (figure 3.1B), each run sampling different regions of conformational space (figure 3.1C), and that the protein structure has therefore not equilibrated yet are not surprising, underscoring the need for long-time simulations as well as a careful restraint in using the term "equilibrated" when addressing protein structures in MD simulations.

However, given that (a) our study's focus is on exploring wild-type TolC ground state dynamics near the crystal structure and (b) within the simulated time of 150 - 300 ns the protein already displayed unreported conformations (figures 3.2 and 3.3) and ion interaction (figures 3.4 and 3.5) both of which might provide potential new insights into the TolC functional mechanism (see below), we consider the amount of sampling achieved (figure 3.1C) adequate for the purpose of this investigation, providing insights into wild-type TolC dynamics on a time scale that is 7.5 - 10 times longer than those of any previous TolC simulations (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008).

3.5.2 Simulation and X-ray

As summarized in figure 3.8, throughout our simulations, wild-type TolC opens and closes on the extracellular (figure 3.2) and periplasmic (figure 3.3) side, adopting conformations that have not been reported in the wild-type (Higgins et al., 2004; Koronakis et al., 2000) and mutant TolC crystal structures (Bavro et al., 2008; Pei et al., 2011). A possible explanation for why these conformations have not been detected yet could be that the crystal environment hinders the conformational changes we observe. Next to crystal type-dependent conformational differences of the same TolC mutant (Bavro et al., 2008), this hypothesis is supported by the distribution of 4 Å crystal contacts in the 1EK9 X-ray structure (figure 3.6C), overlapping with residues where the conformational differences between the simulation average and crystal structure are maximal (figure3.6A,B).

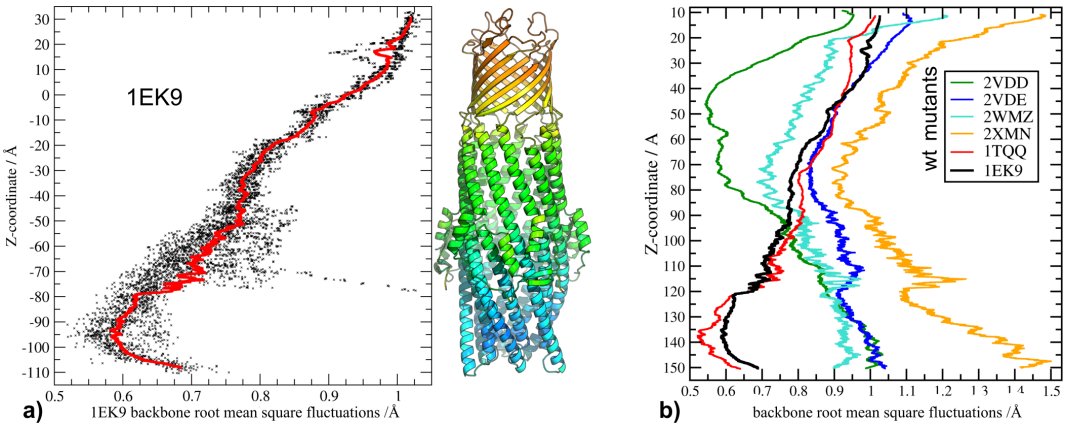

Figure 3.7: Crystallographic B-factor distribution in the TolC wild type crystal structure (Koronakis et al., 2000) used as starting structure for our molecular dynamics simulations (a) and all currently available wild type (Higgins et al., 2004; Koronakis et al., 2000) and mutant (Bavro et al., 2008; Pei et al., 2011) TolC X-ray structures (b).

Furthermore, with an overall Cα RMSD of 1.5 Å, the conformational difference between the X-ray and simulation average structure is very small, suggesting that similar to the results in Kandt et al. (2006) the protein crystal could already contain the open and closed TolC conformations we observe in our simulations. In that case, these conformations either constitute only a minority of the total conformational ensemble in the 1EK9 crystal structure or on average are canceling each other out. Either way, the B factor maximum in the 100K wild-type TolC crystal structures (Higgins et al., 2004; Koronakis et al., 2000) already implies a heightened degree of flexibility or crystal disorder in the extracellular loop region (figure 3.7).

3.5.3 Extracellular Access

Whereas the available structural (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011) and computational data (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008) suggest that TolC access is restricted on both sides, our findings imply that TolC gating occurs only on the periplasmic side (figures 3.3 and 3.8) as TolC freely opens and closes on the extracellular side, suggesting here the absence of a gating mechanism in the isolated wild-type protein (figures 3.2 and 3.8). The broad range of extracellular loop conformations visited throughout our simulations includes loop orientations both more closed and more open (figure 3.2D,E) than the available crystal structures (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011) (figure 3.2E,F).

As indicated by the distribution of θ dihedral angles that we used to monitor the extracellular loops' opening state (figure 3.2E), with θ values located between the two simulation maxima of 30° and 135°, the X-ray conformations represent only

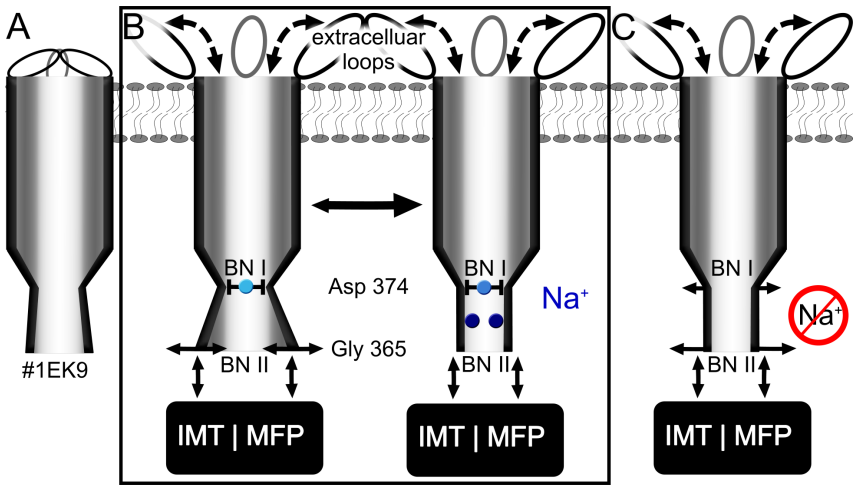

Figure 3.8: Whereas the available wild-type crystal structures restrict TolC access on both sides (A), throughout our simulations TolC opens and closes on the extracellular side. On the periplasmic side, the outer bottleneck BNII fluctuates around a conformation that is more open than the wild-type X-ray structures until the binding of two sodium ions induces a transition to a conformation that is more closed than in any of the known crystal structures. In both cases, the inner bottleneck BNI remains closed (B), unless all sodium is removed from the system, inducing a reopening of BNII and a beginning opening of BNI (C). If our simulations are correct, this observation could imply a locklike mechanism that is dependent on sodium or similar monovalent cations. Whereas the fluctuations of the outer bottleneck might play a role in the interaction with inner membrane transporters (IMT) or membrane fusion proteins (MFP), the principal accessibility from the extracellular side could hint at a novel mode of action for pharmaceutics specifically targeting the TolC interior.

a small section of possible loop conformations. Whereas the crystal structures limit extracellular access to compounds with diameters of up to 7 - 8 Å in the wild-type (Higgins et al., 2004; Koronakis et al., 2000) and 7.2 - 10 Å in the mutant crystal structures (Higgins et al., 2004; Koronakis et al., 2000) (which we estimated by computing Connolly surfaces employing different probe sphere radii), we find that TolC accessibility is sterically only limited by the inner diameter of the β-barrel itself when all three loops are open (figure 3.2D).

If our simulations are correct, these results could either indicate the presence of additional regulation mechanisms in vivo or hint at the possibility of designing a novel group of TolC-directed drugs specifically targeting the protein interior. Possible experimental scenarios for testing the hypothesis of extracellular TolC accessibility could involve double-spin-label electron spin resonance spectroscopy measuring the distance between the tip regions of the extracellular loops or fluorescence spectroscopy experiments in which first a fluorescence dye-specific cross-linker is introduced into the TolC interior and then fluorescence activity is checked after external

dye application, using dyes of different sizes, and protein extraction and purification. Beyond that, TolC-dependent colicin import (Hwang et al., 1997; Zakharov et al., 2004) already provides evidence of the principal occurrence of an outwardly open conformation, and our findings might indicate that the extracellular entrance does not constitute a rate-limiting barrier in the colicin uptake process as wild-type TolC is capable of opening by itself.

As we observe both opening and closing of the extracellular loops (figure 3.2), our findings are also compatible with previous simulations (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008) reporting extracellular closure. The fact that extracellular opening has not been reported in previous TolC simulations is likely due to the amount of conformational sampling achieved, which is larger in this study because of longer simulation times and the usage of multicopy MD (Caves et al., 1998; Das et al., 2000; Fischer and Kandt, 2011; Kandt and Tieleman, 2010; Kandt et al., 2006).

3.5.4 Periplasmic Access and a Sodium-Dependent Lock

Exhibiting an inner (BNI, monitored at Asp374) (Andersen et al., 2002b) and an outer periplasmic bottleneck (BNII, monitored at Gly365) (Andersen et al., 2002a), inwardly closed TolC has been shown by conductance experiments to be stabilized by inter- and intramonomeric hydrogen bonds and salt bridges (Andersen et al., 2002a). Mutating residues involved in this network (table 3.2) led to crystal structures more open than the wild type (Bavro et al., 2008; Pei et al., 2011) (table 3.1), whereas 20 - 30 ns MD simulations of the mutants reported increased flexibility of the periplasmic tip region (Vaccaro et al., 2008), potassium binding sites (Schulz and Kleinekathöfer, 2009), but no unforced opening of BNI or BNII (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008). Only when the effect of an outer electric field was simulated, accelerating cations toward the extracellular medium, could an opening of BNII be induced (Schulz and Kleinekathöfer, 2009).

In our simulations, Na^+ binding sites and spontaneous BNII opening and closure were observed in the wild-type protein without applying any biasing forces. However, as high-residence probability sodium binding sites were observed in only one simulation after 120 ns, the presence of potassium sites in wild-type TolC cannot be excluded on the basis of the available data. Furthermore, given that (a) the ionic and hydration radii of sodium lie between those of potassium and lithium (Cordero et al., 2008; Nightingale, 1959; Shannon, 1976), for which a conductance-lowering effect was reported compared to potassium (Andersen et al., 2002a), and (b) TolC-NaCl conductivity experiments have not been reported yet to the best of our knowledge, different K- and Na-TolC interactions cannot be excluded on the basis of the currently available data.

Although they display similar opening states in the wild-type crystal structures (figure 3.3 and table 3.1), BNI and BNII behave differently in our simulations.

Whereas BNI remains closed unless all NaCl is removed from the system (figure 3.3B), BNII occurs in two conformational states that have not been reported in previous wild-type X-ray structures or MD simulations (figures 3.3 and 3.8 and table 3.1).

In the open state, BNII is more open than the wild type and comparably open compared to the mutant crystal structures (table 3.1), whereas in the closed state, BNII is more closed than any of the available TolC crystal structures but still permits the passage of ions. Though open is the predominant state in our simulations, in one run we observe a spontaneous transition to closed state BNII completed after 120 ns that is induced and subsequently stabilized by two sodium ions (figure 3.3) interacting with residues located in a region where a network of hydrogen bonds and salt bridges has been identified as stabilizing inwardly closed TolC (Andersen et al., 2002a; Bavro et al., 2008; Pei et al., 2011) (figures 3.4 and 3.5 and table 3.2).

Closed BNII does not reopen during the following 180 ns unless all NaCl is removed from the system. In total, we find three preferred sodium residence sites (figure 3.4B-D): two sharply defined ones with Na^+ preferentially interacting with Asp371, Thr366, Thr368, and Asp153 of monomer A or B when BNII is closed and a less pronounced and centrally located site near Asp374 that does not show any monomer preference and is present regardless if BNII is open or closed (figures 3.4 and 3.5). If our simulations are correct, our findings of distinctive sodium interaction sites and the bottlenecks responding to the presence (BNII closure) or absence (opening of BNI and BNII) of sodium imply that periplasmic TolC access is restricted by a sodium-dependent lock.

Whereas so far ion binding sites have been reported for only hexaamminecobalt (Higgins et al., 2004) and potassium (Schulz and Kleinekathöfer, 2009), the presence of sodium binding sites in thebottleneck region is plausible given the high concentration of negative surface charges in this area (Koronakis et al., 1997), as well as the locationand preferred residue interaction of our proposed sodium sites overlapping with residues shown to be involved in stabilizing an inwardly closed TolC conformation (Andersen et al., 2002a; Bavro et al., 2008; Pei et al., 2011) (table 3.2). Furthermore, given that it is not uncommon that ion electron density is misinterpreted as water in the determination of X-ray crystallographic structures (Faust et al., 2008), the presence of several water molecules in the region of our proposed sodium interaction sites in the wild-type 1EK9 crystal structure (Koronakis et al., 2000) could be interpreted as further evidence supporting our hypothesis, assuming some of the crystal water positions actually correspond to sodium.

Depending on whether the observed BNII closure is reversible without requiring further interaction partners, the sodium-dependent lock mechanism we propose would involve a single (BNI) or double bolt (BNII). Although closed state BNII remains stable for 180 ns throughout our simulations under unaltered (figure 3.3B, bold lines) and modified (figure 3.3B, dashed line) conditions, our results cannot exclude the reopening of BNII after, for example, 1800 ns. On the other hand, the

increasing TCA observed during the last 50 ns in one of the unmodified extensions (figure 3.3B, bold line) could indicate a beginning reopening of BNII, and the presence of only a single sodium trace instead of two (figure 3.5B) supports a stabilizing sodium effect. However, as the question of whether BNII opens and closes freely cannot be decided on the basis of our current simulation data, further experiments to measure the periplasmic opening state in the Gly365 region using, for example, double-spin-label EPR spectroscopy are required.

Whether sodium-induced BNII closure would be observable in electrophysiological conductance measurements, which to the best our knowledge have so far not been reported yet investigating the effect of NaCl on TolC conductivity, depends on (a) the lifetime of the BNII closed state, (b) the difference in Na^+ flow (BNII closure in our simulations only slows ion flow but does not lead to a full blockage) and whether that difference is detectable in conductance experiments, and (c) the effect of the voltage used in the measurements and if the presence of an outer electric field hinders Na^+ binding. Whereas the absence of open and closed state BNII conformations in wild-type crystal structures (Higgins et al., 2004; Koronakis et al., 2000) and simulation studies (Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008) could be explained as an effect of the crystal environment and the amount of conformational sampling as discussed above, plausible functional interpretations are conceivable for both fluctuating open state BNII and closed state BNII with regard to the interaction of TolC with its target proteins.

The graspinglike motions of open state BNII (figure 3.8B, left) could be interpreted as a first stage of initiating contact with the inner membrane transporter or membrane fusion protein, whereas closed state BNII (figure 3.8B, right) could be relevant when TolC couples to an already assembled CusBA-like transporter-membrane fusion protein complex (Su et al., 2011) in which the outer membrane efflux duct has to slide into the hoselike arrangement of membrane fusion proteins to complete the assembly of the entire pump, regardless of whether direct interaction occurs between the outer membrane efflux duct and the inner membrane transporter. Displaying a continuously heightened sodium residence probability near Asp374 (figures 3.4B-D and 3.8B) and a consistently X-ray-like closed conformation in all runs unless complete NaCl removal induces a so far unreported opening of first BNI and then BNII (figures 3.3 and 3.8C and table 3.1), BNI could constitute a bolt in our proposed sodium-dependent lock mechanism regardless of whether BNII opens and closes freely.

Further underscoring the crucial role of the aspartate ring region in TolC gating (Andersen et al., 2002b; Higgins et al., 2004), our findings could be interpreted as showing that in vivo complex formation of TolC with its respective inner membrane transporter and membrane fusion protein induces conditions in the periplasmic TolC interior that hinder prolonged sodium interaction in the observed Asp374 region contributing to the unlocking of TolC. Without sodium, the three presumably deprotonated carboxyl groups of Asp374 would repel each other, inducing a sequential

opening of BNI and BNII as observed in the NaCl-free simulation. Although this result could also imply an involvement of chloride in stabilizing an inwardly closed TolC conformation, the absence of any preferred interaction sites speaks against a chloride involvement.

Whereas TolC conductance studies in the presence and absence of sodium could provide experimental evidence to test the hypothesis of a sodium-dependent lock, computer simulations of TolC in the presence of the inner membrane transporter or membrane fusion protein could yield insights into the involvement of the other components in the TolC opening mechanisms. MD simulations of TolC in presence of the isolated AcrB docking domain and the entire AcrB-AcrA complex are currently underway in our lab. Whereas we observe TolC opening motions on the periplasmic side, our data so far do not provide enough evidence to decide if periplasmic TolC opening occurs in a symmetrical (Pei et al., 2011) or asymmetrical manner (Bavro et al., 2008). However, the observation of two sodium interaction sites preferentially interacting only with two monomers of the TolC trimer (figure 3.5E) could be interpreted as favoring the asymmetric opening hypothesis. Either way, one should keep in mind the possibility that the different opening conformations observed crystallographically (Bavro et al., 2008; Pei et al., 2011) could simply represent two states of the same process.

3.6 Conclusions

Conducting a series of independent, unbiased 150 - 300 ns MD simulations of wild-type TolC in a phospholipid membrane / 150 mM NaCl water environment, we find that TolC opens and closes freely on the extracellular side, suggesting the absence of a gating mechanism on this side in the isolated protein.

On the periplasmic side, we observe the outer periplasmic bottleneck region adopting in all simulations a conformation more open then the TolC wild-type crystal structures until in one run the successive binding of two sodium ions induces the transition to a conformation more closed than any of the available TolC X-ray structures.

Concurrent with a heightened sodium residence probability near Asp374, the inner periplasmic bottleneck region at Asp374 remains closed throughout the simulations unless all NaCl is removed from the system, inducing a reopening of outer and inner bottlenecks. Our findings suggest that TolC is locked only on the periplasmic side in a sodium-dependent manner.

Chapter 4

Influences of the AcrB Docking Domain on TolC Gating

As discussed in chapter 3 and (Raunest and Kandt, 2012b) TolC seems to be locked only on periplasmic side in a sodium dependent manner and as suggested in section 3.5.4, we applied molecular dynamic simulations of TolC in presence of the docking domain of the inner membrane transporter AcrB in order to explore the influences of AcrB on TolC's gating behavior.

4.1 Abstract

The outer membrane channel TolC functions with a broad range of inner membrane transporters facilitating the export of toxins and antibiotics and occurs in at least two states, one blocking and one passable for substrates. TolC gating and docking mechanisms especially on the periplasmic side are not fully understood yet. To explore the influences of AcrB onto TolC gating we computed an unbiased 1.05 μs molecular dynamic simulation of TolC embedded in a phospholipid membrane and the docking domain of AcrB in a 150 mM NaCl water environment. In this simulation protein orientations were adapted from a previously published docked model, AcrB was reduced to its docking domain (AcrB-DD) and was moved 1 nm apart from TolC along the membrane normal. We observe initial contacts between both protein after 3 ns of simulation time and measure a tight binding at the end of the simulation. On extracellular side TolC's loops favor an open conformation in presence of the AcrB-DD indicating the absence of a gating functionality on this side. On periplasmic side the outer bottleneck (BNII) outlined by Gly365 turns into an open conformation before the docking of the AcrB-DD and stays open until the end of the simulation. The inner bottleneck (BNI) defined by a double aspartate ring (Asp371 and Asp374) remained closed throughout the entire simulation indicating that another component is missing to open TolC on periplasmic side or longer simulations are needed to observe an opening of TolC.

4.2 Introduction

As a member of the outer membrane efflux protein family the channel protein TolC acts as an efflux duct for a broad range of toxic molecules in *Escherichia coli* (Benz et al., 1993; Fralick, 1996; Koronakis et al., 2004; Zgurskaya et al., 2011) and plays a major role in drug expulsion (Fralick and Burns-Keliher, 1994; Sulavik et al., 2001) and the secretion of cell toxins (Delgado et al., 1999; Hwang et al., 1997; Wandersman and Delepelaire, 1990). TolC is compatible to inner membrane transporters of the ATP binding cassette (ABC), to the resistance nodulation division (RND) and to the major facilitator (MF) type (Moussatova et al., 2008; Nikaido, 2011; Saier and Paulsen, 2001). A prominent example of TolC mediated transport is the AcrAB-TolC multidrug efflux system using the proton gradient between periplasm and cytoplasm to power the extrusion of a broad range of harming substances out of the bacterial cell (Nikaido, 2011). The first TolC crystal introduced by (Koronakis et al., 2000) revealed that the protein is a homotrimer, resembling the shape of a hollow cylinder organized in 3 domains. From a membrane-embedded β-barrel, cooperatively formed by all 3 monomers, an α-helical domain extends ≈ 100 Å into the periplasmic space surrounded in its middle section by an α/β equatorial domain (Koronakis et al., 2004; Zgurskaya et al., 2011).

On extracellular side in this and all succeeding X-ray structures (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011) the access of TolC is limited by 3 inward oriented loops. Whereas for these loops closing motions are reported in a computational study of 20 ns molecular dynamic (MD) simulation time suggesting a gating functionality on the extracellular side (Vaccaro et al., 2008), a study of multi copy MD of 150-300 ns simulation time showed a broad range of loop conformations indicating a random like movement of the extracellular loops (Raunest and Kandt, 2012b).

On periplasmic side combined mutagenesisconductance experiments (Andersen et al., 2002a,b), wild-type (Koronakis et al., 2000) and mutant TolC crystal structures (Bavro et al., 2008; Pei et al., 2011) indicated the existence of two periplasmic bottlenecks: An inner bottleneck (BNI) at Asp371 and Asp374, where the binding of blocking hexamminecobalt occurs (Andersen et al., 2002b; Higgins et al., 2004) and an outer bottleneck (BNII) at Gly365, which is stabilized by Tyr362 and Arg367 involved in inter- and intramonomeric hydrogen bonds and salt bridges inducing a closed conformation (Andersen et al., 2002a). Single and double mutations of Tyr362 and Arg367 lead to increased ion conductance (Andersen et al., 2002a) as well as mutated crystal structures showing symmetric (Bavro et al., 2008) and asymmetric (Pei et al., 2011) "partially opening" toward the periplasm.

Computational studies on mutations disturbing the network of salt bridges and hydrogen bonds of BNII reported an increased flexibility (Vaccaro et al., 2008) of BNII and identified a potassium ion pocket (Schulz and Kleinekathöfer, 2009). Furthermore a sodium dependent lock of BNI and BNII was proposed by multi copy

150 - 300 ns molecular dynamic (MD) simulations for wild-type TolC (Raunest and Kandt, 2012b). Mutations affecting BNII as well as the absence of sodium ions in wild-type simulations did not result in a complete opening of TolC (Bavro et al., 2008; Pei et al., 2011; Raunest and Kandt, 2012b). The underlying gating mechanism of TolC especially of BNI is not fully understood yet and it seems that another component, maybe an inner membrane transporter like AcrB, is needed to open TolC for substrate passage.

To prove that AcrB and TolC interact Tamura and coworkers applied cysteine cross linking experiments for the ranges a) TolC's tip region of helix 3 & 4 (139, 142, 145, 147 and 150) and AcrB DD's tip loop (254 - 257) and b) TolC's tip region of helix 7 & 8 (360, 363 - 366) and AcrB DD's β-hairpin (794 - 796) (Tamura et al., 2005). Their cross linking results built the basis for the data driven docking of AcrB and TolC in the published model of the AcrAB-TolC efflux system (Symmons et al., 2009). In this docked model AcrB (Seeger et al., 2006) and TolC (Koronakis et al., 2000) crystal structures were conformationally adjusted. Helix 7 of TolC was shifted outwards into a "partially open" conformation adapted from mutated X-ray structures (Bavro et al., 2008) effecting an open state of BNII. Additionally, AcrB's β-hairpin (residues 794-796) was bend into an open conformation (Symmons et al., 2009).

We initially reduced AcrB to its AcrB-DD (residues 182 - 272, 724 - 812) taken from Seeger et al. (2006) to avoid the complexity of two phospholipid bilayers. Focusing on the influence of AcrB on TolC, we computed an 1.05 μs MD simulation of membrane-embedded wild-type TolC 1EK9 (Koronakis et al., 2000) and AcrB's docking domain (AcrB-DD) (Seeger et al., 2006), in a 150 mM NaCl solution. Both protein orientations were adapted from the docked model (Symmons et al., 2009), whereas the AcrB-DD was moved 1 nm apart from TolC along the membrane normal. Additionally we performed an unbiased 100 ns MD simulation of the isolated AcrB-DD in a 150 mM NaCl solution to validate AcrB-DD's stability in comparison to previously published AcrB simulations (Fischer and Kandt, 2011, 2013).

In the 1.05 μs simulation we observed, that the initially closed BNII of TolC opens within the first 3 ns causing a "partially open" (Bavro et al., 2008) conformation induced by a shift of helix 7 stabilized by the spontaneous docking of the AcrB-DD. This opening on the outer bottleneck II of TolC supports the conformational changes applied in the data driven model of Symmons (Symmons et al., 2009). Although we notice an opening of the outer bottleneck (BNII), AcrB-DD was not able to open the inner bottleneck (BNI) after 1.05 μs simulation time, suggesting that either another component is missing or the AcrB-DD on its own is not able to open TolC within the simulated time of 1.05 μs.

4.3 Methods

4.3.1 Molecular Dynamics Simulations

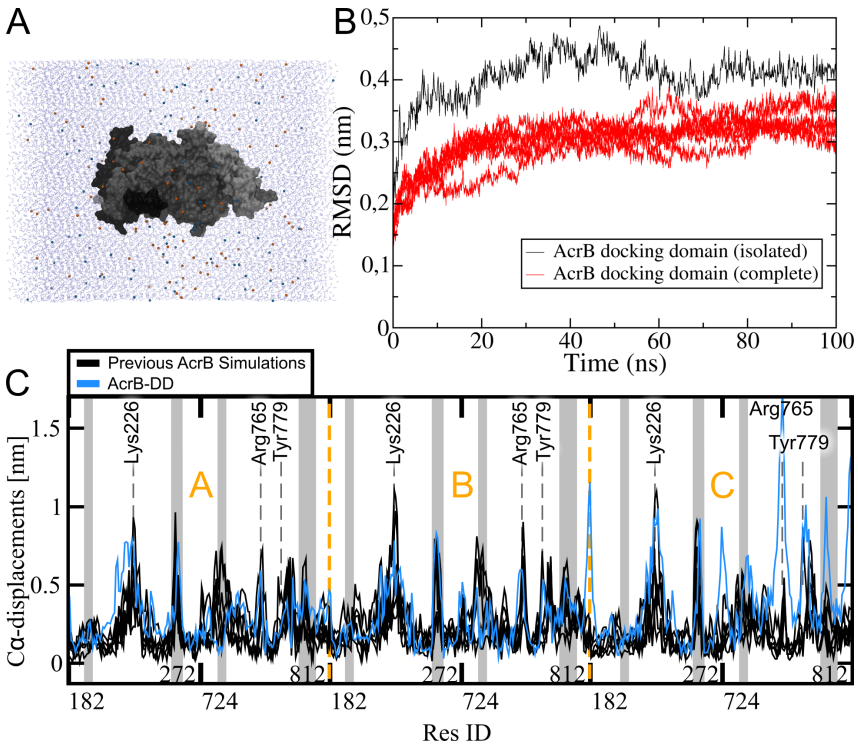

Figure 4.1: A 100 ns MD simulation in a 150 mM NaCl / water environment was applied for the AcrB docking domain (AcrB-DD) to verify its stability (A). Cα-RMSDs were calculated after least square fitting on the starting structure for the current and for previous AcrB simulations (Fischer and Kandt, 2013)(B). Cα-displacements between the final conformations and the crystal are shown in (C). We identified residues located in the interface of TolC and AcrB by a distance cutoff of 1 nm applied on the Symmons model (Symmons et al., 2009) and highlighted them as gray bars in (C).

MD simulations were performed employing GROMACS version 4.0.3 (Berendsen et al., 1995a; Hess et al., 2008) and the GROMOS96-53a6 force field (Oostenbrink et al., 2004). To test the stability of the AcrB docking domain (AcrB-DD) we isolated residues 182 - 272 and 724 - 812 from the crystal structure 2GIF (Seeger et al., 2006) and performed a 100 ns MD simulation in a 150 mM NaCl / water environment containing 108 Na^+ and 111 Cl^- ions in a 10.6 nm x 10.6 nm x 10.6 nm simulation box. For the TolC AcrB-DD simulations the starting structure of membrane embedded TolC of our prior study (Raunest and Kandt, 2012b) and the AcrB-DD were aligned on the Symmons AcrAB-TolC model (Symmons et al., 2009). Afterwards the AcrB-DD was moved 1 nm apart from TolC along the membrane normal, the system was solvated in a 150 mM Na^+ / Cl^- H_2O solution including

205 Na^+ and 190 Cl^- ions and 47423 H_2O molecules in a simulation box of 10.06 nm x 9.97 nm x 21.0 nm. This starting structure was used for four unbiased 150 ns MD simulations. In one of them we noticed a spontaneous docking of the AcrB-DD on TolC's periplasmic end and extended this simulation up to 1.05 μs.

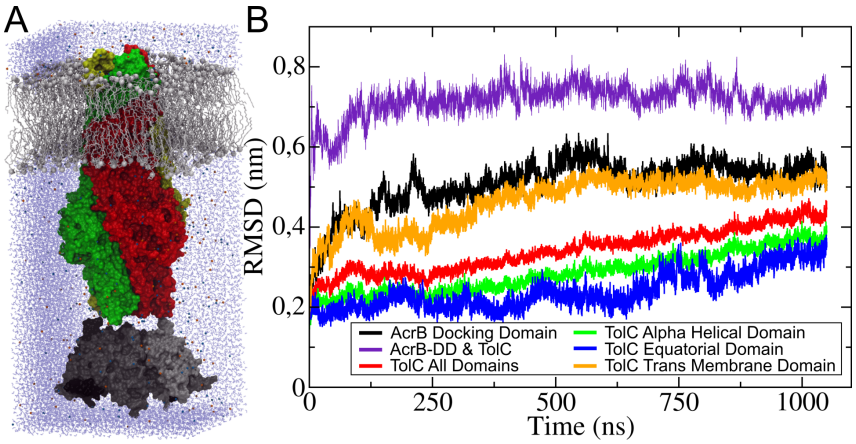

Figure 4.2: TolC and POPE phospholipid bilayer coordinates were adapted from our previous work (Raunest and Kandt, 2012b). Whereas orientations of TolC and AcrB-DD are based on a docked model from Symmons (Symmons et al., 2009), the AcrB-DD was moved 1 nm apart from TolC. Afterwards both proteins were solvated into a 150 mM NaCl / water environment (A). Cα-RMSD's were calculated after least square fitting on the individual domain / protein (B).

In all simulations bond lengths were constrained by LINCS (Hess et al., 1997) so that an integration time step of 2 fs could be chosen and standard protonation states were assumed for titratable residues. Systems were simulated at 310 K, maintained separately for protein, lipids, and water by a Berendsen thermostat (Berendsen et al., 1984) with a time constant (τ_T) of 0.1 ps. Pressure coupling was done employing a Berendsen barostat (Berendsen et al., 1984) using a 1 bar reference pressure and a time constant of 4 ps. While for the AcrB-DD solvent simulation isotropic pressure coupling were used, semiisotropic pressure coupling was employed to permit bilayer fluctuations in the membrane plane for the AcrB-DD TolC simulation. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation, (Darden et al., 1993; Essmann et al., 1995) and twin range cutoffs of 1.0 and 1.4 nm were applied for computing the van der Waals interactions.

4.3.2 Analysis

To monitor the stability of the TolC AcrB-DD complex, the domains of TolC and AcrB-DD we calculated Cα root-mean-square deviations (RMSD) after least square fitting on Cα carbons of the starting structure (figure 4.1 and 4.2). Additionally, we calculated Cα displacements for the last conformations of the 100 ns simulation containing the AcrB-DD in absence of TolC and of prior AcrB simulations Fischer and

Kandt (2013) after least square fitting on the AcrB-DD of 2GIF (Seeger et al., 2006) (figure 4.1C). We identified interfacing Cα carbons in the docked model (Symmons et al., 2009) by a 1 nm distance cutoff for TolC and AcrB-DD and marked them as gray bars in figure 4.1C.

Based on these reference atoms we calculated Cα-RMSDs for each time frame the 1.05 μs simulation (figure 4.3B) in order to compare the AcrB-DD TolC MD interfaces with the interface of the docked model (figure 4.3B). To quantify the docking of AcrB-DD on TolC we monitored the buried surface area between both proteins. Therefor the solvent accessible surface areas (SASA) of AcrB-DD, TolC and the entire complex were determined via NACCESS (Hubbard and Thornton, 1993). Afterwards the complex SASA was subtracted from the sum of AcrB-DD and TolC to achieve the buried surface area (figure 4.3C).

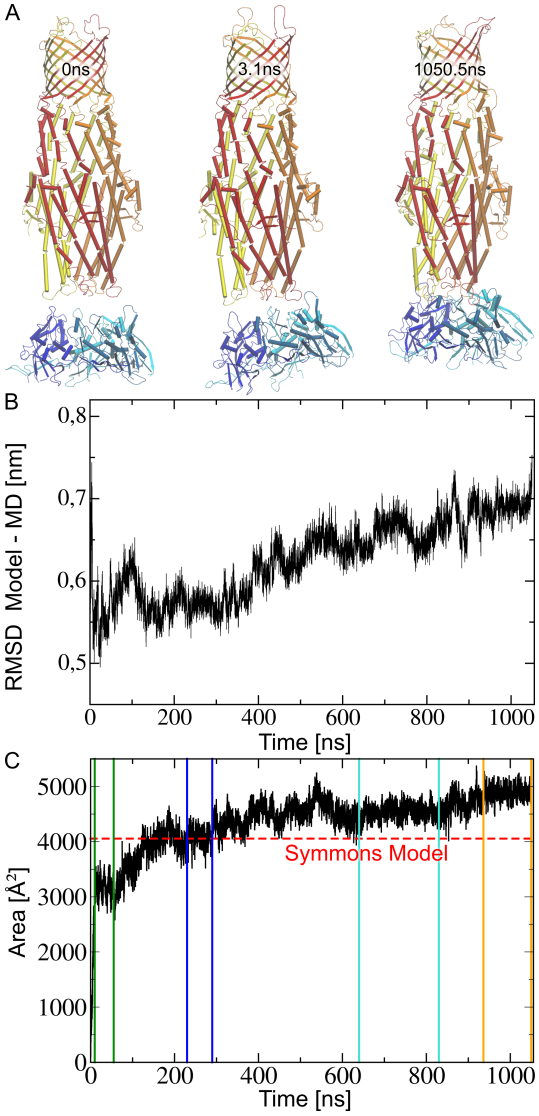

Figure 4.3: In the beginning of the MD simulation AcrB is located 1 nm apart von from TolC. After 3 ns first contacts between AcrB and TolC occur and result in a tight binding after 1.05 μs (A). Residues of the interface between AcrB-DD and TolC were identified by a 1 nm distance cutoff regarding the Cα coordinates of the docked model. Based on these Cα atoms we calculated Cα-RMSDs for our simulation after least square fitting (B). To quantify the docking of AcrB-DD onto TolC we calculated the buried surface area lying in between via NACCESS (Hubbard and Thornton, 1993). The interface area increase over time exceeds the interface area of the Symmons model (4054 $Å^2$) after 300 ns of simulation time and ends with a value of 4701 $Å^2$ (C). For further analysis we identified four plateau regions marked in green, blue, cyan and orange.

To identify frequently interacting residue pairs (FIRP), atom distances of TolC and AcrB were measured by the GROMACS tool g_dist with a 2 Å cutoff for the four plateau regions of figure 4.3B. The small threshold of 2 Å was chosen in order to name residue pairs which are close together and therefore highlight interacting residues. Multiple distances below 2 Å during a time step between the same residues of AcrB and TolC were counted as 1 in order to achieve a percentage representation (figure 4.4). We mapped FIRPs on average structures of the individual plateaus (figure 4.4A), which were calculated as described in Fischer and Kandt (2011) using an iterative scheme of calculating the average conformation and realigning the trajectory to that average structure before computing a new average structure. This procedure was repeated until the average structure stopped changing. In the grid representation of figure 4.4B only FIRPs above a frequency of 20% are listed and additionally the frequencies of monomer contacts are highlighted by gray backgrounds. Cα distances of prior tested (Tamura et al., 2005) interface residue pairs were calculated for each time step of plateau 4 by g_dist (figure 4.4B). Distances below 0.647 nm were assumed to be able to form a cystein cross link as reported in (Schmidt et al., 2006) and are listed in table 4.1 in percent of observed simulation time.

Table 4.1: Cystein crosslinks between AcrB & TolC

		TolC Gln139	Gln142	Arg143	Asn145	Gly147	Ala150
AcrB	Asn254	N 0%	N 96.2%	N 97.5%	N 100%	N ¡ 0.1%	N 98.8%
	Gln255	N ¡ 0.1%	Y 97.4%	Y 83.3%	N 53.4%	Y 0%	Y 92.5%
	Asp256	N 0.2%	Y 79.1%	Y 0%	N ¡ 0.1%	Y 0%	N 22.1%
	Gly257	N 0%	N 0.3%	N 0%	N 0%	Y 0%	N 0.1%
		TolC Ala360	Ser363	Val364	Gly365	Thr366	
AcrB	Ala794	N 0%	N 0%	N 6.4%	N 0%	N 5.2%	
	Asp795	N 0%	N 0%	N 0%	Y 0%	N 0%	
	Gly796	N 0%	N 0%	N 0%	Y 99.2%	N 99.7%	

Y stands for reported cys cross links, N stands for negative responses in cross linking experiments (Tamura et al., 2005). Observed Cα distances below 0.647 nm in the fourth plateau region in % simulation time. Green cells indicate an agreement between our simulation and Tamura et al. (2005). Red cells highlight residue pairs that do not behave similar in our simulation and in Tamura et al. (2005)

On the extracellular side the opening state of TolC's extracellular loops was monitored through the dihedral angle θ spanned by the Cα atoms of Asp56 located in the β-barrel and Ala270 at the tip of each extracellular loop (figure 4.5). As the 1EK9 crystal structure displays an average θ value of 89.8° (figure 4.5), an extracellular loop was regarded as closed as or more closed than 1EK9 whenever $\theta \leq 90°$. Conversely, a loop for which θ ¿ 90° was considered more open than the 1EK9 X-ray structure. θ was monitored for each extracellular loop (figure 4.5A) and subsequently converted into a binary representation of "more open" than 1EK9 (a value of 0) or "more closed" (assigned a value of 1). The number of loops in the

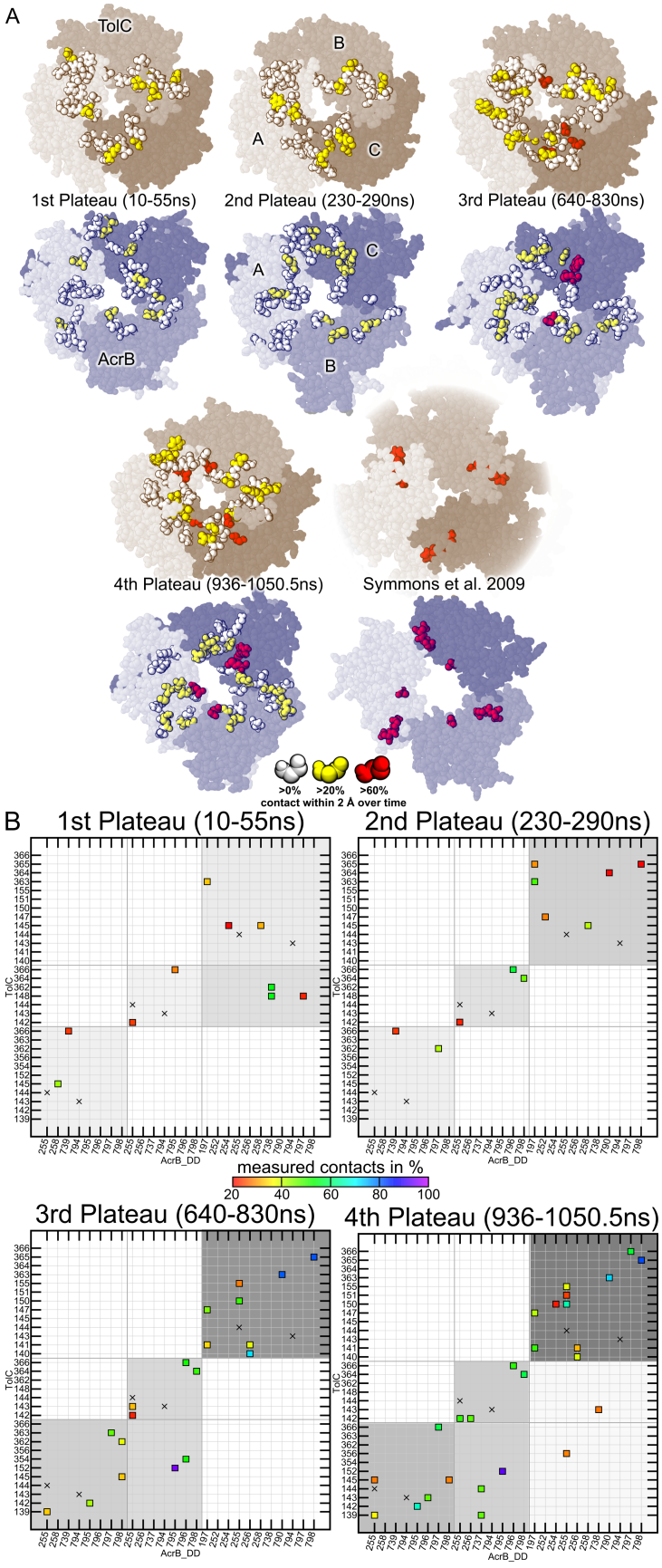

Figure 4.4: To monitor the docking of AcrB and TolC, we determined frequently interacting residue pairs (FIRPs) by a distance cutoff of 2 Å for the four plateau regions in figure 4.3C. In (A) 3D representations show the location and frequency of FIRPs. The grid representation in (B) focuses on FIRPs with frequencies above 20% and while the highest amount of FIRPs (monomers C-C in plateau 4) were colored in gray (50% black), the remaining monomers were colored by their amount of FIRPs in gray scale (B). Contacts of the docked model (Symmons et al., 2009) are shown in panel A (right) and are marked as 'X' in the grid representation (B). Both representations show an increase of quantity and frequency of FIRPs over time.

closed conformation was determined and in figure 4.5, the θ angle distribution of the current and our TolC only work (Raunest and Kandt, 2012b) are colored in black and gray, respectively. On periplasmic side TolC opening state was monitored similar to (Raunest and Kandt, 2012b; Schulz and Kleinekathöfer, 2009) by calculation of the triangular cross-sectional area (TCA) defined by the Cα atoms of Asp374 and Gly365 (figure 4.5B and C).

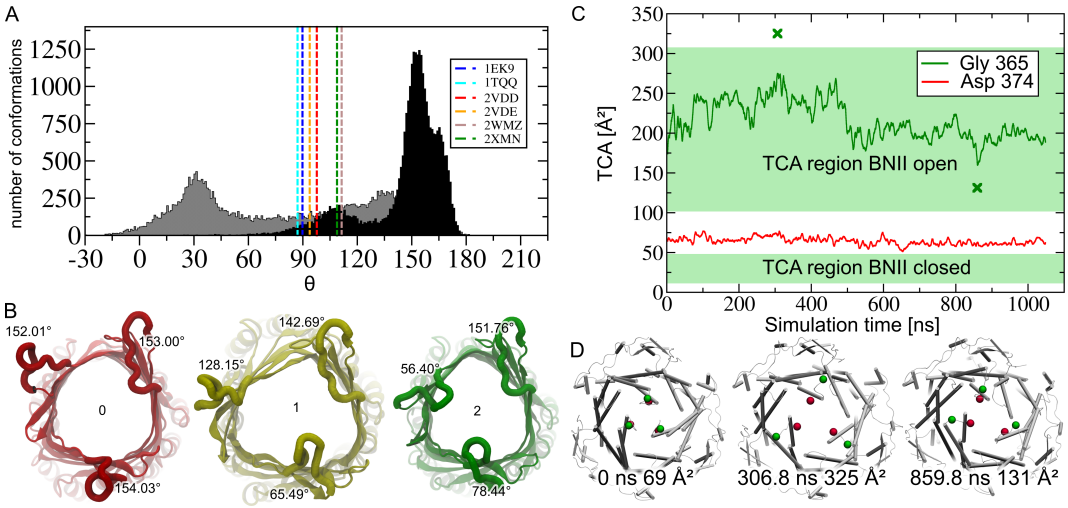

Figure 4.5: Similar to Raunest and Kandt (2012b) we monitored TolC open state using the dihedral angle θ build by the Cα atoms of Asp56 in the β-barrel and Ala270 in in the tip of each extracellular loop. For a better comparison we show previous TolC only results in gray (Raunest and Kandt, 2012b) and our current results in black (A). The extracellular loops prefer with a peak at 153° an open conformation. Simulation snapshots display 0, 1, 2 closed loop(s) and 1EK9 conformations (B). On periplasmic side, TolC open states were monitored via the triangular cross sectional areas (TCA) build by the Cα atoms of Asp374 and Gly365 highlighting the inner and outer bottleneck region, respectively (C). Simulation snapshots show initial, maximum and minimum conformations for BNII (D left, middle, right panel). For the minimum state of BNII we did not regard the first 10 ns, where the docking of the AcrB-DD happens. TCA's have been smoothed using a running average filter of 5 ns. Bold crosses mark the unfiltered maximum and minimum snapshots in panels D center and right.

To analyze the distribution of sodium throughout the simulation spatial sodium distributions were computed via the VolMap plugin in VMD version 1.9 (Humphrey et al., 1996) using a spatial resolution of 1 $Å^3$ to analyze the distribution of sodium in the periplasmic bottleneck region at the two density levels 0.01 and 0.1 Na/$Å^3$ (figure 4.6). To validate our simulation we compared final conformations of TolC and AcrB-DD to the underlying crystal structures. Therefor we superimposed the final simulation conformation on 1EK9 and 2GIF after Cα least square fitting on the crystal structures and calculated Cα displacements (figure 4.7B and D).

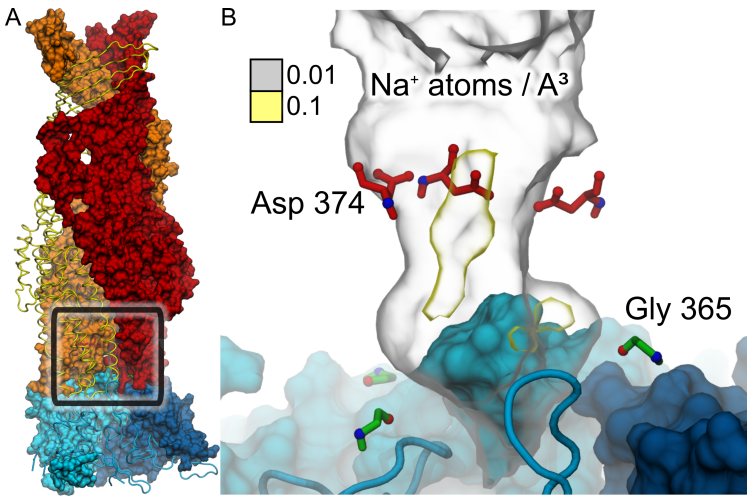

Figure 4.6: Focused on the BNI and BNII regions 3D spacial residence probabilities densities above 0.01 $Na^+/ Å^3$ occur for both bottlenecks and while densities above 0.1 $Na^+/Å^3$ were measured they did not exceed 0.2 $Na^+/Å^3$ (B). The measured sodium distribution is comparable to BNII open state simulations in Raunest and Kandt (2012b).

4.4 Results

4.4.1 AcrB Docking Domain Stability

To monitor the stability of the isolated AcrB docking domain (AcrB-DD) we performed a MD simulation of the isolated AcrB-DD in a 150 mM NaCl solution (figure 4.1A). Afterwards we computed Cα root-mean-square deviations (RMSDs) after least square fitting on the Cα atoms of the starting structure and compared it to previously published simulations of AcrB (Fischer and Kandt, 2013). While the isolated AcrB-DD reaches a plateau after 80 ns and ends with an RMSD of 0.41 nm, the Acrb's DDs reach plateaus in one case after 20 ns and for another two simulations after 90 ns. The end conformation RMSDs range from 0.29 nm to 0.36 nm (figure 4.1B). The Cα-displacements between end conformations of prior AcrB (Fischer and Kandt, 2013) and of current AcrB-DD were calculated after least square fitting on 2GIF (Seeger et al., 2006). TolC interfacing residues are marked as gray bars in figure 4.1C.

Beside the ends (residues 182, 272, 724 and 812), Arg765 and Lys226 part of the inter monomer connecting-loop (ICL (Pos, 2009)), the AcrB-DD shows comparable Cα-displacements to previous AcrB simulations (Fischer and Kandt, 2011). Whereas these divergent regions are located on the TolC averted side of the ArcB-DD, the interface residues show Cα-displacements close or below the displacements of previous simulations (Fischer and Kandt, 2013) (figure 4.1C).

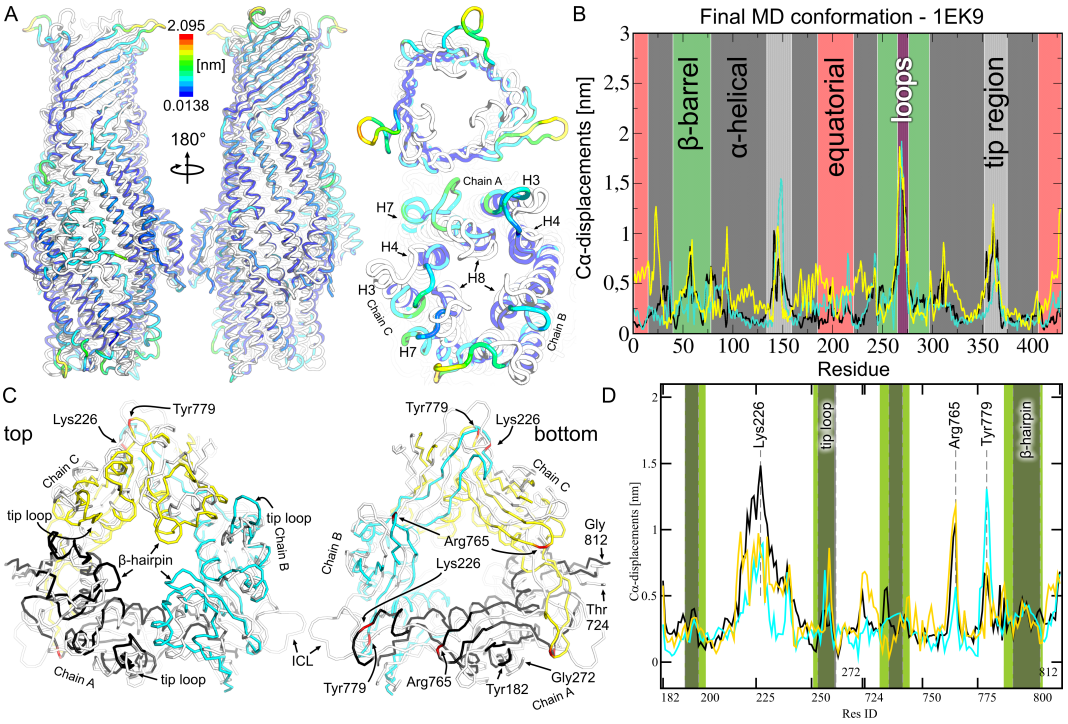

Figure 4.7: To compare our findings with experimental results we superimposed 1EK9 with the last simulation conformation after least square fitting (A), calculated their Cα-displacements and colored them monomer wise in black (monomer A), cyan (monomer B) and yellow (monomer C) (B). The largest conformational differences occur in the extracellular loop region followed by the periplasmic region leading to an open conformation on extracellular side and to an open state of BNII / "partial open state" (Bavro et al., 2008) as already shown in figure 4.5. The final conformation of AcrB-DD in complex with TolC was superimposed on 2GIF after least square fitting (C). Cα-displacements were recorded in (CD) and colored by monomer. Interface residues between AcrB and TolC of the Symmons model and of the last conformation of our simulation were identified by a distance cutoff of 1 nm and are highlighted by gray and green bars, respectively (D). The largest differences occur at Lys226 member of the intermonomer connecting loop (ICL (Pos, 2009)) (CD). Whereas the β-hairpins (794 - 796) move inwards, the tip loops (255 - 257) keep their orientation (C).

4.4.2 Complex Stability

To examine the complex stability we calculated RMSDs of the entire complex, the AcrB-DD, the complete TolC, the α-helical domain, the transmembrane domain and the equatorial domain (figure 4.1). With an RMSD of 0.83 nm the largest conformational changes occur at 369 ns of simulation time in the TolC-AcrB-DD complex, which reaches a plateau of 0.72 nm after 500 ns of simulation time. Followed by the AcrB-DD domain, which reaches a plateau after 800 ns of simulation time. The transmembrane domain achieves a plateau after 700 ns and ends with changes of

0.56 nm. The RMSDs of the complete TolC, the α-helical and the equatorial domain do not reach plateaus and end with RMSDs of 0.45 nm, 0.39 nm and 0.35 nm.

4.4.3 Interface between AcrB-DD and TolC

Within the first 3 ns we observe initial contacts between AcrB-DD and TolC follwed by an asymmetric docking, which results in a tight binding after 1.05 μs simulation time (figure 4.3A). To quantify the similarity between the interface of the docked model (Symmons et al., 2009) and the interfaces of our simulation we identified facing Cα atoms by a distance cutoff of 1 nm for the docked model. Based on this selection we calculated Cα-RMSDs for our simulation after least square fitting on the coordinates of the docked model. The initial RMSD between our simulation and the docked model is 0.65 nm. It decreases until 200 ns of simulation time, afterwards it rises until the the end of the simulation and reaches a value of 0.75 nm (figure 4.3B).

The buried surface area between AcrB-DD and TolC raises throughout the simulation and exceeds the interface area of the Symmons model (Symmons et al., 2009) after 300 ns (figure 4.3C). To gain insight into the docking interface we focus on the four plateau regions of the interface area highlighted in green, blue, cyan and orange in figure 4.3C. For these plateaus we calculated frequently interacting residues pairs (FIRPs) and analogue to the results of the interface area analysis we observe an increase of FIRPs. Beginning with the third plateau (640–830 ns) residue pairs above 60% frequency appear (figure 4.4).

While asymmetric monomer contacts are barely perceptible in figure 4.4A, the grid representation indicates asymmetric monomer contacts for the plateaus 1, 3 and 4 highlighted by gray backgrounds (figure 4.4). FIRPs of the Symmons model (Symmons et al., 2009) are shown in figure 4.4A, as 'X' in B. In our simulation none of the Symmons (Symmons et al., 2009) FIRPs were observed (figure 4.4B).

To identify facing residues between AcrB-DD and TolC, Tamura and coworkers applied cystein cross linking experiments pair wise for a) TolC's tip region of helix 3 & 4 (139, 142, 145, 147 and 150) and AcrB-DD's tip loop (254 - 257) and for b) TolC's tip region of helix 7 & 8 (360, 363 - 366) and AcrB-DD's β-hairpin (794 - 796) (Tamura et al., 2005). a) Our data are in agreement with experimental results for AcrB-DDs residues 254 - 267 except the contacts between 254AcrB-DD & 142,143,145,150TolC, 255AcrB-DD & 145,147TolC, 256AcrB-DD & 143,147,150TolC and 257AcrB-DD & 147TolC. We observe a tendency that AcrB-DD's Asn254 interacts with TolC tip region of helix 3 & 4 (142,143,145TolC). b) For the tip region of TolC's helix 7 & 8 our data are in agreement with experimental results except the contacts between 795AcrB-DD & 365TolC and 796AcrB-DD & 366TolC (table 4.1), respectively. To sum it up we measured $^{27}/_{39}$ (69.2 %) accordances in our simulation and Tamura et al. (2005) and $^{12}/_{39}$ (30.8 %) variations to residue pairs Tamura et al. (2005) tested in their cross linking study (table 4.1).

4.4.4 TolC Gating

Extracellular Side

To analyze TolC's opening state on the extracellular side, we monitored the dihedral angle θ formed by the Cα atoms of Asp56 in the β-barrel and Ala270 at the tip of each extracellular loop (figure 4.5A and B). Using the 1EK9 average θ dihedral of 89.8° as a reference, we determined the number of loop conformations as closed or more closed than the aforementioned crystal structure. The histogram ranges from 30° to 180° of θ distributions (figure 4.5A) and shows three peaks at 110°, 153° and 165° indicating that the open conformation is preferred in our simulation.

Periplasmic Side

On the periplasmic side we monitored the TolC opening state by calculating the triangular cross-sectional area (TCA) spanned by the Cα atoms of Asp374 and Gly365 representing the inner (BNI) and outer periplasmic bottleneck region (BNII) (figure 4.5C and D). Similar to Raunest and Kandt (2012b) BNII opens immediately within the first 3 ns. The average TCA of BNII in the 1.05 μs is with 216 ± 26 Å^2 slightly bigger and fluctuates lesser than in (BNII open)-TolC only simulations (195.1 $\pm$ 42.6 table 3.1, figures 3.3B, 4.5C).

For the inner bottleneck, we measure a TCA of 64 ± 6 Å^2. It remains closed and displays a TCA similar the 1EK9 X-ray structure (59 Å^2). Whereas the left panel in figure 4.5D shows the start conformation of BNI and BNII, panels D middle and right illustrate simulation snapshots of the maximal and minimal BNII opening of 352 and 131 Å^2, respectively.

4.4.5 Sodium Distribution

To determine the distribution of sodium ions and check for potential binding sites, we calculated the average spatial Na^+ density distribution. Zooming in on the bottleneck region, we find a heightened three-dimensional Na^+ density near Asp374 and Asp371 exceeding 0.1 Na^+/Å^3 (figure 4.7).

4.4.6 Simulation versus X-ray

To compare the docked structure to the crystal structures 1EK9 and 2GIF, we aligned the final conformation of the simulation to the crystal structures after least square fitting and calculated Cα displacements for the individual proteins (figure 4.7). In case of TolC the open conformation of the extracellular loops cause the highest displacements of 1.9 nm, followed by displacements of the tip region reaching 1.5 nm at helix 3 & 4 of monomer B. Slight outwards movements of helix 7 in monomer B and C caused displacements of 1 to 1.3 nm (figure 4.7). For AcrB-DD the largest conformational changes occur at the ICL of monomer A reaching a

displacement of 1.5 nm followed by displacements of 1.3 nm at Tyr 779 of monomer B (figure 4.7). While our simulations indicate an inward movement of the beta hairpins (residues 794 - 796), the tip loops of AcrB-DD (residues 255 - 257) keep their orientation except an outward bending of the tip loop in monomer B.

4.5 Discussion

In this work, we report an unbiased 1.05 μs MD simulation, where AcrB's docking domain (AcrB-DD) docks asymmetrically on TolC. We measure a tighter binding of AcrB-DD on TolC in comparison to a previously published docked model (Symmons et al., 2009). On extracellular side TolC prefers an open state for its extracellular loops in presence of the AcrB-DD. While on periplasmic side TolC's outer bottleneck (BNII) opens due to a shift of helix 7, the inner bottleneck (BNI) remains closed (figure 4.5C), supporting the idea of an "partially open state" of TolC during the assembly of the AcrAB-TolC efflux system (Bavro et al., 2008). We begin with the limitations of our approach and then precede for our finding and their biological implications.

4.5.1 Limitations of our Approach

As an outer membrane protein, TolC's micro environment is characterized by a heterogeneous lipid composition containing lipolysaccharides in the outer leaflet, while the protein interacts with the peptidoglycan layer on the periplasmic side (Lugtenberg and Alphen, 1983). In our simulation we approximated the outer membrane by a homogeneous POPE bilayer omitting a representation of the peptidoglycan layer.

While it would certainly be desirable to have both components fully included in simulation studies, the development of the appropriate lipopolysaccharides and peptidoglycan molecular topologies is still in its infancy (Kotra et al., 1999; Lins and Straatsma, 2001; Meroueh et al., 2006; Piggot et al., 2011, 2012; Straatsma and Soares, 2009; Vollmer and Bertsche, 2008; Vollmer et al., 2008b) and at the time of writing, such simulation parameters were not available for *E. coli*. Whereas one could argue, that the TolC behavior we observe on the extracellular side could be an artifact of our simplified membrane model, it is encouraging, that one of the first lipopolysaccharide simulations studying the outer membrane protein OprF (Straatsma and Soares, 2009), reported that the lipopolysaccharids have a stabilizing effect on the open conformation of the OprF extracellular loops, that exhibit an architecture similar to that of TolC.

We assume that the isolated AcrB-DD is a good model to explore the influence of AcrB onto the periplasmic gating mechanisms of TolC, because the Cα displacements and the RMSDs of AcrB's DD show that the isolated AcrB-DD behavior is comparable to previously 100 ns MD simulations of AcrB (Fischer and Kandt, 2013).

With any molecular dynamics study, the question of whether the simulated time

has been sufficiently long with respect to the problem under investigation arises. Given the findings reported by Grossfield and Zuckerman (Grossfield and Zuckerman, 2009), that 1.6 μs atomistic MD simulation of membrane-embedded rhodopsin was not enough for the protein structure to converge, our observations that TolC Cα RMSDs are still increasing after 1.05 μs (figure 4.2B), and that the protein structure has therefore not equilibrated yet, are not surprising. Underscoring the need for long-time simulations as well as a careful restraint in using the term "equilibrated" when addressing protein structures in MD simulations.

However, given that a) our study's focus is on the influences of the AcrB-DD on TolC's gating mechanisms, b) in one of four simulations AcrB-DD docs spontaneously onto TolC, c) AcrB-DD stabilizes the open conformation of TolC's BNII, we consider the amount of sampling achieved adequate for the purpose of this investigation providing insights into AcrB-DD docking and its influence on TolC on a time scale that is 3.5 to 35 times longer than those of any previous TolC simulations (Raunest and Kandt, 2012b; Schulz and Kleinekathöfer, 2009; Vaccaro et al., 2008).

4.5.2 Simulation and X-ray

To compare the last conformation of the 1.05 μs simulation to the X-ray structures 1EK9 and 2GIF we superimposed the final simulation conformations of TolC and AcrB-DD on the underlaying crystals structures after least square fitting and calculated Cα displacements (figure 4.7).

TolC

As shown in figure 4.5A, the open conformation for TolC's extracellular loops is favored in our simulation causing the highest displacements of 2.1 nm in presence of the AcrB-DD after 1.05 μs (figure 4.7A). On periplasmic side the tip region show the second most displacements reaching 1.5 nm induced by an outward shift of helix 7 in all monomers in the final conformation (figure 4.7A), which is also indicated by the TCA of BNII (figure 4.5C).

Both ends of TolC, the extracellular loops and the outer bottleneck (BNII), show wider open conformations in our simulation than in the crystal structure. A possible explanation of these discrepancies to the crystal structure is that the crystal environment may hinder TolC to open up on both sides. In a previous study we identified 4 Å crystal contacts on both ends, which might stabilize the closed state on both sides of TolC (Raunest and Kandt, 2012b) (see chapter 3 figure 3.6C). These crystal contacts could induce the predominant closed conformations in the wild-type crystal structures.

Moreover with a rising RMSD until the end of the simulation TolC did not reach an equilibrium state hinting that TolC undergoes conformational changes until the end of the simulation and maybe longer simulation time is needed to observe a complete opening of TolC on periplasmic side.

AcrB-DD

For AcrB-DD the highest displacements occur on the TolC averted side at an inner- and at the intermonomer connecting-loop (ICL). While we notice an increased Cα displacement of the ICL in prior AcrB (Fischer and Kandt, 2013), the inwards movement of the inner loop at Arg765 occurred only for the isolated AcrB-DDs simulations and could be caused by the isolation of the AcrB-DD (figure 4.7C and D). However the inward movement of the loop at Arg765 had little to no effect on the TolC facing side of AcrB-DD, since we measure comparable interface regions for our simulation and the Symmons model (figure 4.7D).

4.5.3 Interface between AcrB and TolC

In our simulations initial contacts between the AcrB-DD and TolC occur within the first 3 ns of simulation time resulting into a docked conformation where all monomers of AcrB-DD are in contact with TolC after 1.05 μs of simulation time (figure 4.3A). To measure the similarity between our simulation and the Symmons model (Symmons et al., 2009) we used facing Cα atoms within a 1 nm cutoff of AcrB and TolC in the docked model (Symmons et al., 2009) as basis for RMSD calculations. The initial decrease of the RMSDs until 200 ns indicates a conformational convergence is followed by a continuously increase of the RMSDs after 200 ns show conformational distinctions (figure 4.3B). The buried surface area rises until the end of the simulation and exceeds the interface area of the Symmons model (4054 $Å^2$) after 300 ns of simulation time and ends with a value of 4701 $Å^2$ (figure 4.3C). The increase of the frequently interacting residue pairs (FIRPs) in frequency and quantity indicate a tighter binding of the AcrB-DD over time (figure 4.4).

Beside this tighter binding we notice an asymmetric docking of the AcrB-DD onto TolC as highlighted by gray quadrants in figure 4.4B. The 2 Å contacts of the Symmons model (144TolC-255AcrB and 143TolC-794AcrB), marked as 'X' in figure 4.4B, could not be confirmed in our simulation. A possible explanation of this is that the asymmetric docking of the AcrB-DD on TolC disturbs the 2 Å contacts of the Symmons model and induced other short range interactions. Building only one conformation of the AcrAB-TolC complex and as indicated by a smaller buried surface area maybe the computational docking routine avoids atom distances below 2 Å. Previous point mutation / cystein cross linking experiments identified facing residues of a) TolC's tip region of loop 3 & 4 and AcrB-DD's tip loop and b) TolC's tip region of loop 7 & 8 AcrB-DD's β-hairpin (Tamura et al., 2005). In their work a closed conformation of helix 7 & 8 at the outer bottleneck (BNII) and a two stage docking mechanism for AcrB and TolC has been proposed. We measured Cα distances of the same residue ranges and assumed, that distances below 0.647 nm are able to form a cystein cross link as reported in Schmidt et al. (2006). Results of the cross linking study and the current study are compared in table 4.1 and display an 69.2 % accordance between Tamura et al. (2005) and our measurements.

The simultaneous occurrence of false-positives and false-negatives in table 4.1 could mean that the docking of AcrB-DD on TolC has not reached its final stage in our simulation. Since the interface area is rising until the end of the simulation longer simulation time is needed to achieve this final stage.

4.5.4 Extracellular Access

Available X-ray structures (Bavro et al., 2008; Higgins et al., 2004; Koronakis et al., 2000; Pei et al., 2011) and a computational study suggest, that the access could be restricted on both sides (Vaccaro et al., 2008). Whereas in our current simulation we find a predominant open state for the extracellular loops in presence of the AcrB-DD (figure 4.5A). Maybe the docking of the AcrB-DD onto TolC is responsible for the constant open state of the extracellular loops of TolC. Previous simulations have shown flexible extracellular loops (figure 4.5A gray), which means, that the sterical accessibility is only limited by the inner radius of the β-barrel.

4.5.5 Periplasmic Access

In 2002 Andersen and coworkers identified two bottlenecks on the periplasmic side via conductance experiments: An inner one outlined by a double aspartate ring build by the residues 371 and 374 of the 3 monomers (BNI), and an outer one (BNII), stabilized by network of an inter monomer salt bridge and intra monomer hydrogen bonds. Mutations, affecting BNI or BNII, lead to higher conductance (Andersen et al., 2002a,b). Based on these experiments mutations of BNII have been introduced into the crystal structures 2VMD, 2VDD, 2WMZ and 2XMN (Bavro et al., 2008; Pei et al., 2011) causing an opening of BNII, but not of BNI suggesting that BNII is able to open and close independently from BNI (Raunest and Kandt, 2012b).

Dividing TolC open states into an inner (BNI) (Andersen et al., 2002b) and outer periplasmic bottleneck (BNII) (Andersen et al., 2002a), the question whether TolC is open or closed on periplasmic side during the efflux system assembly needs to be stated more precisely in order to differentiate between BNI and BNII states (Tikhonova et al., 2011).

The inner Bottleneck (BNI)

Identified by Andersen et al. (2002b), the inner bottleneck (BNI) remains closed during our simulation and no opening motions were measured (figure 4.5C). Analogue to Raunest and Kandt (2012b) increased sodium residence probabilities of 0.1 Na^+ atoms / $Å^3$ occur near BNI, which might stabilize the closed conformation BNI (figure 4.6). Maybe a flush of solvent induced by the AcrB pump could push the ions outward of the Asp371 and Asp374 double ring and therefore relax and open the inner bottleneck. Schulz and Kleinekathöfer (2009) applied an electric field to pull potassium ions out of an ion pocket in order to open TolC's BNII. However,

our simulation shows, that the AcrB docking domain (AcrB-DD) is not able to open BNI during 1.05 μs, meaning that either the complete AcrB, eventually including also a bound drug, or both AcrA and AcrB or a longer simulation is necessary to open BNI.

The outer Bottleneck (BNII)

While the outer bottleneck (BNII) is stabilized by a network of salt bridges and hydrogen bonds (Andersen et al., 2002a), mutations disturbing this network lead to open conformations of BNII / "partial open conformation" in crystal structures 2VMD, 2VDD, 2WMZ and 2XMN (Bavro et al., 2008; Pei et al., 2011). Monitored BNII by the TCA of Gly365 Cα atoms, simulation studies report a lock of BNII in a potassium and sodium dependent manner for mutants (Schulz and Kleinekathöfer, 2009) and wild-type TolC (Raunest and Kandt, 2012b). In presence of the AcrB-DD the sodium distribution near the bottlenecks is similar to sodium distributions of the open state of BNII in (Raunest and Kandt, 2012b). Comparing the average TCA of BNII of the 1.05 μs simulation to the BNII open time frames of chapter 3, we measured a higher average TCA of BNII with lesser fluctuations for the 1.05 μs simulation meaning that the AcrB-DD stabilizes the open conformation of BNII. Analogue to Raunest and Kandt (2012b) we measure an initial opening of BNII within 3 ns and observe initial contacts between AcrB-DD and TolC after 3 ns of simulation time (figure 4.3A). This means that BNII is already open before the AcrB-DD docks completely on TolC. This hypothesis can be validated by locking the periplasmic side of TolC as described in Eswaran et al. (2003) and apply afterwards the very same mutational cystein cross linking experiments with the locked TolC and AcrB as described in Tamura et al. (2005). If BNII is closed before AcrB docks on TolC, the results should be comparable to results of Tamura et al. (2005).

4.6 Conclusions

Computing an unbiased 1.05 μs MD simulation containing AcrB's docking domain (AcrB-DD) and TolC in a phospholipid membrane / 150 mM NaCl water environment, we report a tighter but asymmetric binding of the AcrB docking domain (AcrB-DD) onto TolC in comparison to the best available docked model (Symmons et al., 2009). On TolC's extracellular side, analogue to (Raunest and Kandt, 2012b), we find freely open and closing motions for the extracellular loops and moreover we notice a favored open conformation of these loops in presence of AcrB-DD hinting for the absence of a gating functionality on extracellular side.

We measured on periplasmic side an open state of TolC's outer bottleneck (Andersen et al., 2002a) (BNII) supporting the "partially open" / BNII open conformation of the docked model published by Symmons et al. (2009). Furthermore, our data indicate, that the outer bottleneck (BNII) is open before or at least during

the assembly of TolC and AcrB. This open state of BNII is stabilzed by the docked AcrB-DD. The inner bottleneck (BNI) (Andersen et al., 2002b) remained closed throughout the simulation meaning that AcrB-DD could not open TolC during the simulation and that additional components are missing or a longer simulation time is needed to open TolC.

Chapter 5

Unilateral Access Regulation: Ground State Dynamics of the Pseudomonas aeruginosa Outer Membrane Efflux Duct OprM

OprM is a TolC homologue protein located in the outer membrane of Pseudomonas aeruginosa. Similar to TolC it interacts with several inner membrane transporters and facilitates the export of many harming substances. Through this circumstances OprM may become a target of drug research, because a disabled blocked or malfunctioning OprM could deny substrate expulsion across the outer membrane. With the loss of this export functionality strains could loose resistance against drugs. The understanding of the gating mechanisms of OprM are a prerequisite to target OprM with drugs. In the following chapter Dennis Koch performed MD simulations as part of his Master thesis under my co-supervision to explore the gating mechanisms of OprM.

Adapted from:
Koch, D. C., Raunest, M., Harder, T., and Kandt, C. (2013). Unilateral access regulation: Ground state dynamics of the pseudomonas aeruginosa outer membrane efflux duct oprm. *Biochemistry*, 52(1):178–187

5.1 Abstract

Acting as efflux duct in the MexAB-OprM multidrug efflux pump, OprM plays a major role in the antibiotic resistance capability of Pseudomonas aeruginosa, trafficking substrates through the outer cell membrane. Whereas the available crystal structures showed restricted OprM access on both ends, the underlying gating mechanism is not fully understood yet.

To gain insight into the functional mechanism of OprM access regulation we carried out a series of five independent, unbiased molecular dynamics simulations, computing 200 ns dynamics samples of the wild type protein in a phospholipid membrane / 150 mM NaCl water environment. On extracellular side OprM opens and closes freely under the simulated conditions suggesting the absence of a gating mechanism on this side of the isolated protein. On periplasmic side we observe an opening of the tip regions at Val408 and to a lesser degree at Asp416 located 1.5 nm further inward the channel, leading to OprM end conformations up to 3 respectively 1.4 more open than the asymmetric crystal structure.

If our simulations are correct our findings imply that periplasmic gating involves only the Asp416 region and that in vivo additional components, absent in our simulation, might be required for periplasmic gating if the observed opening trend near Asp416 is not negligible. In addition to that we identified in each monomer a previously unreported sodium binding site in the channel interior coordinated by Asp171 and Asp230 whose function role remains to be investigated.

5.2 Introduction

The aerobic, Gram-negative bacterium Pseudomonas aeruginosa is a common pathogen that can be found in almost all kind of environments (Ramos, 2010). Especially for humans with weak immunity or damaged tissue P. aeruginosa can cause severe medical conditions (Mesaros et al., 2007). One of the most troublesome characteristics of P. aeruginosa is the strong antibiotic resistance capability which often originates from the expression and activity of multidrug efflux pumps (Putman et al., 2000; Strateva and Yordanov, 2009). Until now several multidrug efflux pump systems have been characterized in P. aeruginosa, including the most prominent ones MexA-MexB-OprM (Poole et al., 1993), MexC-MexD-OprJ, MexE-MexF-OprN, and MexX-MexY-OprM (Köhler et al., 1997; Masuda et al., 2000a,b). Displaying different substrate specifities, these systems extrude a wide range of antimicrobial agents, including quinolones, tetracyclines and most penicillins (Masuda et al., 2000b).

In wild-type cells MexAB-OprM is the only efflux system that is expressed constantly (Nakajima et al., 2002). The corresponding operon encodes MexA, a periplasmic adaptor or membrane fusion protein (MFP), MexB, an inner membrane drug-proton antiporter of the resistance nodulation division (RND) protein super family, and the outer membrane factor (OMF) OprM (Hancock and Brinkman, 2002; Wong et al., 2001; Zgurskaya and Nikaido, 2000). Assembling transiently these components form a tripartite efflux system with MexB acting as engine and active transporter expelling various compounds out of the cell via the outer membrane efflux duct OprM.

MexA is assumed to couple the MexB and OprM in the assembled pump (Xu et al., 2012). Beyond interacting with MexA and MexB, OprM also functions as efflux duct in conjunction with other inner membrane transporters (Gotoh et al., 1998; Masuda et al., 2000a; Zhao et al., 1998). At the time of writing two OprM crystal structures have been solved, the first of a single monomer (Akama et al., 2004), the second of the entire trimer(Phan et al., 2010), showing the efflux duct at resolutions of 2.56 Å and 2.4 Å respectively.

OprM is a homo-trimer forming a hollow cylinder and displays a very high degree of conformational similarity to its *Escherichia coli* homologue TolC with an α-carbon root-mean-square deviation (RMSD) of 1.6 Å despite a low sequence identity of 19% (Phan et al., 2010). Like TolC OprM exhibits a β-barrel domain embedded in the outer membrane from which an α-helical domain extends 100 Å into the periplasmic space. A ring-like equatorial domain surrounds the α-helical domain's middle section while the protein's N-terminus is attached to the outer membrane via a palmitoyl lipid anchor (figure 5.1A). Acting as efflux duct in a transiently assembled transport system, OprM occurs in at least two different states: one permitting and one blocking the passage of substrates. While in the available X-ray structures OprM access is restricted on extracellular and periplasmic side, the underlying mechanisms of gating and access regulation are not fully understood yet.

Similar to TolC an iris-like opening mechanism has been proposed on periplasmic side (Zgurskaya et al., 2011), whereas elastic network normal mode analyses suggested additional conformational changes involving twisting and stretching motions of OprM (Phan et al., 2010). For TolC recent computer simulations suggested that channel access is regulated only on periplasmic side in a sodium-dependent manner while on extracellular side the isolated wild type protein opens and closes freely (Raunest and Kandt, 2012b).

Addressing the question of OprM access regulation and gating mechanism as well as exploring the potential influence of sodium ions, we conducted a series of five independent, unbiased 200ns molecular dynamics (MD) simulations of membrane-embedded wild-type OprM in a 150 mM NaCl solution. On extracellular side we find OprM opening and closing freely, suggesting the absence of a gating mechanism on this side in the isolated protein. Assuming a similar architecture as in its *E. coli* homologue TolC, comprising an inner and outer bottleneck region, we monitored OprM's periplasmic opening state using Asp416 and Val408 which we selected based on their proximity to their TolC counterparts after superimposing OprM and TolC X-ray structures. In all simulations an opening of both bottlenecks occurs.

However, the effect is stronger pronounced in the outer bottleneck region reaching end conformations up to 3 times more open than the starting crystal structure. At the same time the inner bottleneck displays post-simulation conformations only 1.1 to 1.4 times more open than in the starting structure. If our simulations are correct our findings imply that periplasmic gating occurs only in the inner bottleneck region at Asp416 and that in vivo additional components, absent in our simulation, might be required for periplasmic gating if the observed opening trend at Asp416 is not negligible. In addition to that we identified in each monomer a previously unreported sodium binding site in the channel interior coordinated by Asp171 and Asp230. Apparently uninvolved in gating or structure stabilization the functional role of the Na site remains to be investigated.

5.3 Experimental Procedures

5.3.1 Molecular Dynamics Simulations

MD simulations were performed employing GROMACS version 4.0.3 (Berendsen et al., 1995a; Hess et al., 2008) and the GROMOS96-53a6 force field (Oostenbrink et al., 2004) using the 3D5K OprM crystal structure (Phan et al., 2010) as a starting structure. The protein was inserted in a pre-equilibrated 9.6 nm x 9.6 nm palmitoyl-oleoylphosphatidylethanolamine (POPE) bilayer patch (Tieleman and Berendsen, 1998) using INFLATEGRO (Kandt et al., 2007). The system was solvated with simple point charge water molecules (Berendsen et al., 1981) and 181 Na^+ and 166 Cl^- ions, yielding a 150 mM NaCl solution and a total system charge of zero (figure 5.1A). Standard protonation states were assumed for titratable residues.

After a 20 ns membrane equilibration with protein atoms position-restrained using a force constant of 1000 kJ mol-1 nm-2, five independent MD runs with different starting velocities were performed, each 200 ns in length. In the simulations, all bond lengths were constrained by LINCS (Hess et al., 1997) so that an integration time step of 2 fs could be chosen. Systems were simulated at 310 K, maintained separately for protein, lipids, and water by a Berendsen thermostat (Berendsen et al., 1984) with a time constant (τ_T) of 0.1 ps. Pressure coupling was done employing a Berendsen barostat (Berendsen et al., 1984) using a 1 bar reference pressure and a time constant of 4 ps. Semiisotropic pressure coupling was employed to permit bilayer fluctuations in the membrane plane. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation, (Darden et al., 1993; Essmann et al., 1995) and twin range cutoffs of 1.0 and 1.4 nm were applied for computing the van der Waals interactions.

5.3.2 Analysis

To monitor protein stability and conformational drift throughout the simulations we computed Cα RMSDs after least square fitting to the starting structure (figure 5.1). This was done for the entire protein as well as the trans-membrane β-barrel including the extracellular loops, the α-helical domain, the equatorial domain and the membrane-anchoring N-terminus. To assess the overall amount of conformational sampling throughout the simulation a principal component analysis was performed on the α-carbon trajectories of all simulations.

Similar to Raunest and Kandt (2012b) the OprM opening state on extracellular side was monitored through the dihedral angle θ spanned by the Cα atoms of Ser113 located in the β-barrel and Thr106 at the tip of each extracellular loop (figure 5.2). As the outwardly closed 3D5K crystal structure displays an average θ value of 99.6° (figure 5.2C), an extracellular loop was regarded as closed as or more closed than 3D5K whenever $\theta \leq 99.6°$. Conversely, a loop for which θ ¿ 99.6° was considered more open than the 3D5K X-ray structure. In each simulation, θ was monitored for each extracellular loop (figure 5.2A, first panel) and subsequently converted to a binary representation of "more open" than 3D5K (a value of 0) or "more closed" (assigned a value of 1) (figure 5.2A, second panel).

As a summary for each run, the number of loops in the closed conformation was determined (figure 5.2A, third panel) and converted to a frequency histogram (figure 5.2A, fourth panel) reflecting for each simulation the percentage occurrence of closed loopconformations (figure 5.2B). To obtain an overview of the extracellular loop conformations visited throughout all simulations, we calculated a histogram showing the θ distribution in all five runs (figure 5.2E). Similar to Schulz and Kleinekathöfer (2009) and Raunest and Kandt (2012b), the periplasmic OprM opening state was monitored through α-carbon triangular cross-sectional area (TCA) analysis (figure 5.3). The corresponding Asp416 and Val408 OprM residues representing the inner

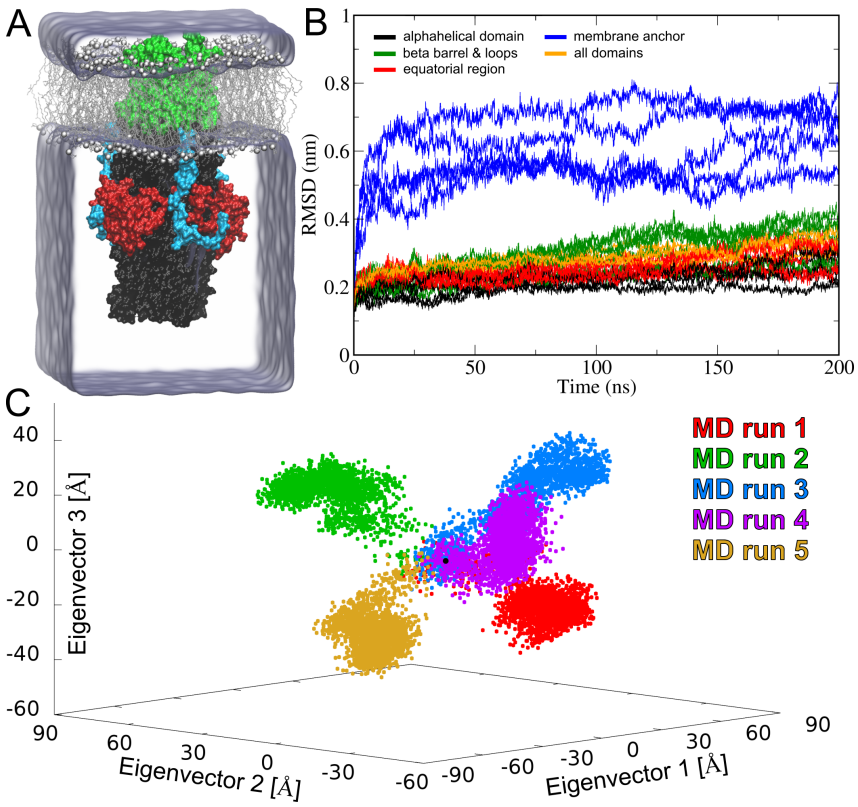

Figure 5.1: (A) Molecular dynamics simulations of wild-type 3D5K OprM were conducted in a POPE phospholipid / 150 mM NaCl water environment. The coloring scheme represents different OprM regions: β-barrel extracellular loops (green), α-helical domain (black), equatorial domain (red), membrane-anchoring N-terminus (blue). (B) Cα root-mean-square deviations after respective least- square fitting to the starting structure. While the largest conformational changes occur in the membrane-anchoring N-termini and the β-barrel / extracellular loops region, the overall RMSDs of the entire protein are still rising after 200 ns. (C) To assess the amount of conformational sampling during our simulations we carried out a principal component analysis mapping for each simulation OprM's path through conformational space down to the three dimensions spanned by the first three Eigenvectors.

(BNI) and outer periplasmic bottleneck (BNII) were selected based on their average minimum α-carbon distance to their TolC counterparts Asp374 (1.97 Å) and Gly365 (2.36 Å) (Andersen et al., 2002a,b) after superimposing the 3D5K and 1EK9 OprM and TolC crystal structures using PyMOL 1.5 (Delano, 2002) (figure 5.3A).

To analyze the distribution of sodium throughout the simulations, both one-dimensional (1D) Na^+ density profiles along the membrane normal and average spatial sodium distributions were computed. The former was done using the GROMACS tool g_density, and for the latter, we employed the VolMap function in VMD version 1.943 using a spatial resolution of 1 $Å^3$ to analyze the distribution of sodium

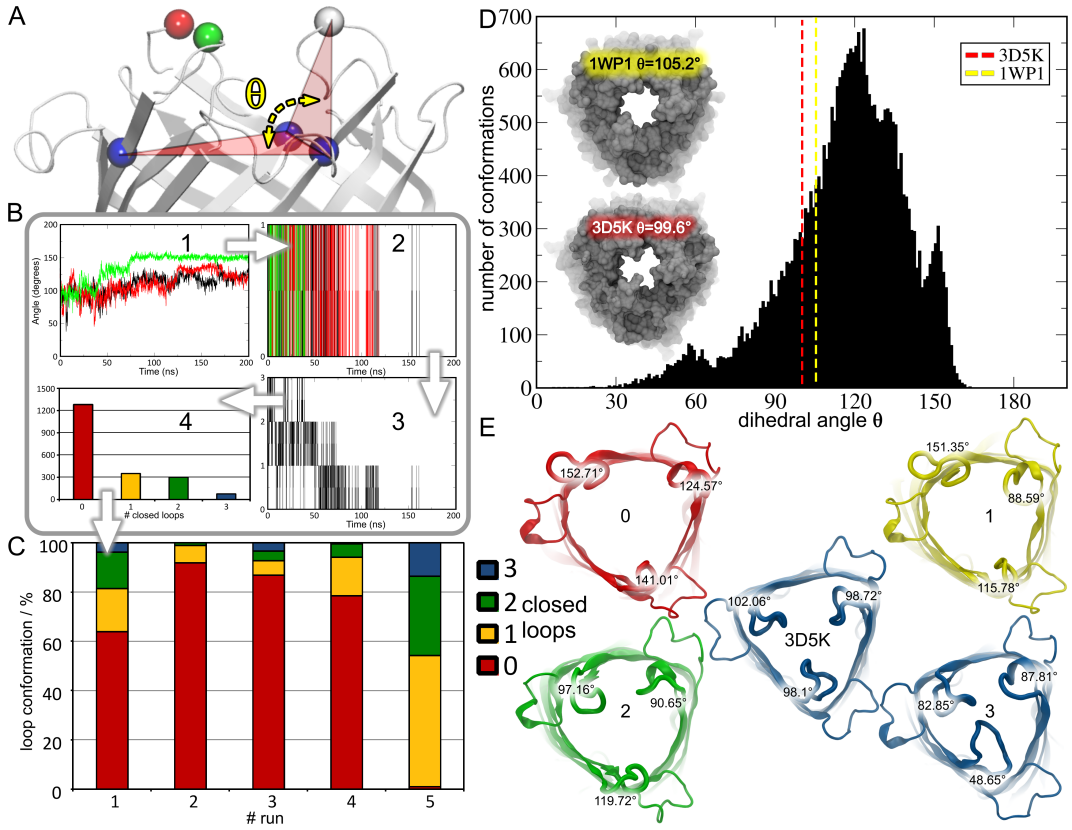

Figure 5.2: On the extracellular side the OprM opening state was monitored through the dihedral angle θ formed by the α-carbons of Ser113 in the β-barrel and Thr106 at the tip of each extracellular loop (A). As the outwardly closed 3D5K crystal structure displays an average θ value of 99.6°, an extracellular loop was regarded as closed as or more closed than 3D5K whenever $\theta \leq 99.6°$. In each simulation, θ was monitored for each extracellular loop (B1) and subsequently converted to a binary representation of "more open" than 3D5K (assigned a value of 0) or "more closed" (assigned a value of 1) (B2). As a summary for each run, the number of loops in the closed conformation was determined (B3) and converted to a frequency histogram (B4), reflecting for each simulation the number of closed extracellular loops and the percentage of simulation time spent in that conformation (C). As shown by the overall θ distribution calculated over all runs (D) and simulation snapshots and the 3D5K crystal structure (E), the extracellular loops adopt a broad range of conformations preferring the completely open state under the simulated conditions (C, D).

at four density levels ranging from 0.01 to 0.8 $Na^+/Å^3$. Sodium binding and unbinding were monitored analyzing the z-coordinate trajectory of each Na^+ ion and computing for each OprM monomer the percentage of simulation time a residue comes into contact with at least one sodium ion in at least one monomer. For this analysis, we employed a residue-Na distance cutoff of ≤ 3 Å. Both the sodium distributions and the residue contact analyses were performed for all five simulations.

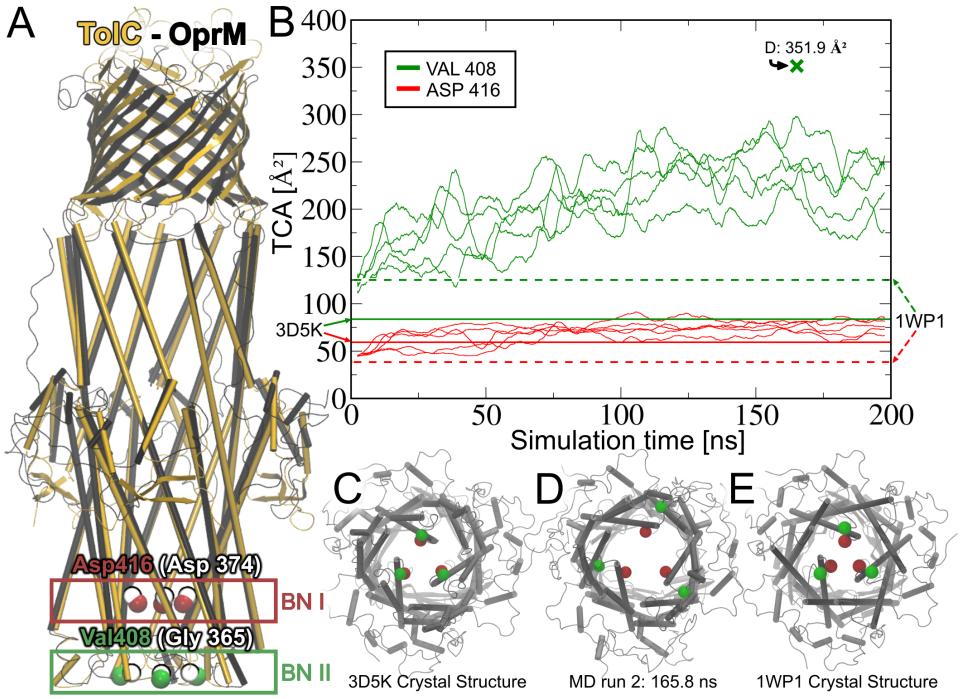

Figure 5.3: On the periplasmic side, the OprM opening state was monitored using the triangular cross- sectional area (TCA) spanned by the α-carbons of Asp416 and Val408, representing an inner (BNI) and outer periplasmic bottleneck (BNII) based on the residues' close proximity to their TolC counterparts Asp374 and Gly365 (A). Both bottlenecks open during the simulations, however in BNII the opening is stronger pronounced, leading to TCAs up to 3 times higher as in the 3D5K starting structure by the end of the simulations, whereas BNI is only 1.1 - 1.4 times more open after 200 ns (B). Conformational snapshots (C-E) illustrate periplasmic OprM conformations in the asymmetric (C) and symmetric crystal structure (E) as well as at maximum BNII opening (D). For the sake of clarity, the TCAs have been smoothed using a running average filter of 5ns. Bold crosses mark the unfiltered TCA maxima.

To quantify sodium occupation of the detected interaction site, we calculated the run and time-averaged percentage of simulation time one, two or three sodium ions are simultaneously present within 3 Å of Asp171 and Asp230. As described in Fischer and Kandt (2011), simulation average structures were calculated using an iterative scheme of calculating the average conformation and realigning the trajectory to that average structure before computing a new average structure. This procedure was repeated until the average structure stopped changing.

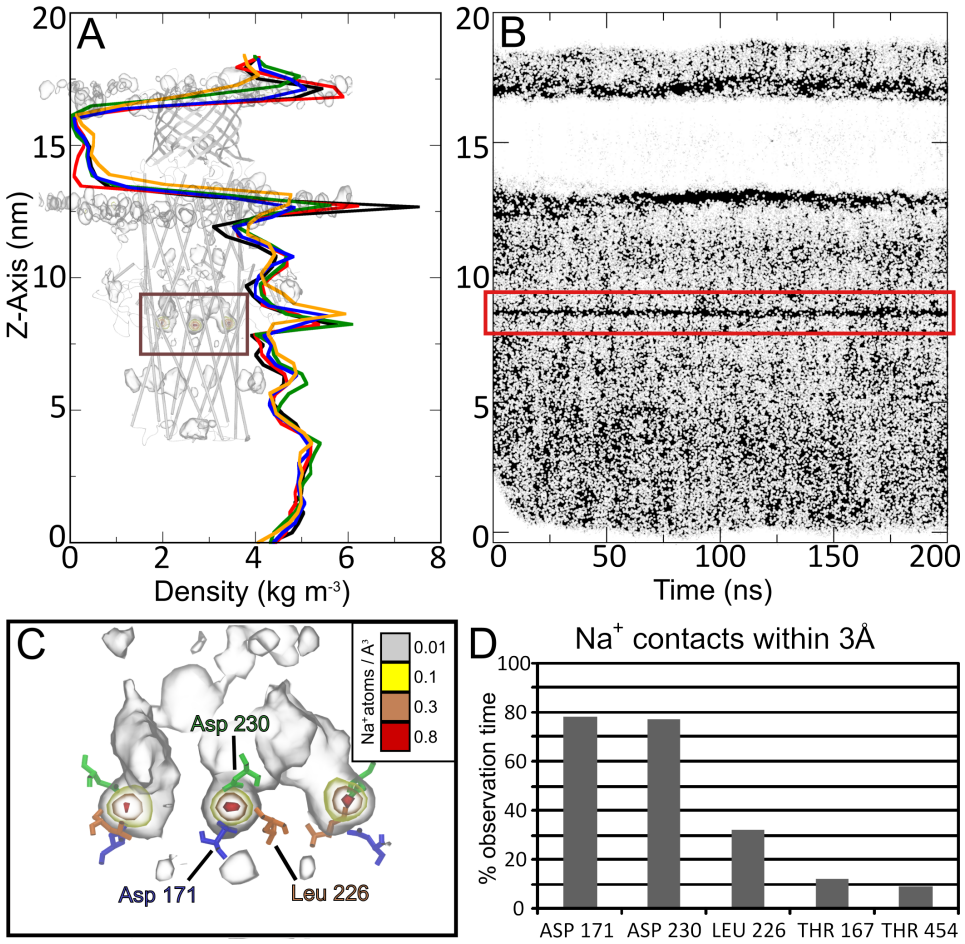

Figure 5.4: Sodium distribution and interaction sites. Partial sodium densities calculated for all simulations indicate the presence of two preferred sodium binding sites: at the lipid headgroups and in the lower equatorial region of the OprM channel (A). The protein site is continuously occupied for the entire simulation time as indicated by a stable sodium trace in a Na^+ z-trajectory plot shown representatively for run 1 (B). Run and time-averaged spatial sodium distributions in the lower equatorial region show for each monomer maxima in Na^+ density in the channel interior near Asp171, Asp230 and Leu226 (C) which we find interacting with sodium for 78%, 77% and 32% of the simulation time at a distance $\leq$ 3 Å (D).

5.4 Results

5.4.1 Protein Stability and Conformational Sampling

To monitor protein stability and conformational drift we computed for each simulation the Cα RMSDs for the whole protein, the β-barrel including the extracellular loops, α-helical domain, the equatorial domain and the membrane-anchoring N-terminus (figure 5.1B). With RMSDs ranging from 4 to 8 Å the largest confor-

mational changes occurred in the membrane-connecting N-terminal regions followed by the β-barrel / extracellular loops regions displaying RMSDs up to 4.5 Å.

For the equatorial and α-helical domain we observed RMSD ranges of 2 - 3.5 and 1.5 - 3 Å, respectively. As indicated by the continuously increasing RMSD curves of the entire protein, within the simulation time of 200 ns no stable RMSD plateaus are reached. Assessing the overall amount of conformational changes during our simulations we carried out a principal component analysis of the concatenated Cα-trajectories. Projecting OprM's pathway onto the first three Eigenvectors we find that in each run the protein explores different regions of conformational hyperspace around the X-ray structure (figure 5.1C).

5.4.2 Gating

Extracellular Side

To analyze the extracellular opening state of OprM, we monitored the dihedral angle θ formed by the Cα atoms of Ser113 in the β-barrel and Thr106 at the tip of each extracellular loop. Using the 3D5K X-ray structure's average θ dihedral of 99.6° as a reference, loop conformations were classified as "closed or more closed" than the crystal structure when $\theta \leq 99.6°$, whereas loop conformations with θ ¿ 99.6° were regarded as "more open" than the 3D5K X-ray structure.

Throughout the simulations we observe a variety of extracellular loop conformations ranging from 0 up to 3 closed loops (figure 5.2D), whereas the highest frequency was found to be in the range of 0 closed loops. Whereas the overall θ distribution calculated over all simulations ranges from 30° to 165°, displaying a maximum at 120 (figure 2D), the 1WP1 and 3D5K crystal structures show average θ values of $\sim 105°$ and $\sim 100°$ (figure 2D).

Periplasmic Side

Monitoring the OprM opening state on the periplasmic side, we calculated the triangular cross-sectional area (TCA) rendered by the Cα atoms of Asp415 and Val408 representing the inner (BNI) and outer periplasmic bottleneck (BNII) (figure 5.3A).

In all five simulations, the outer bottleneck BNII consistently opens up during the first 100 ns (figure 5.3C) reaching TCAs in the range of 206.1 ± 45.5 Å^2 with a maximum of 351.9 Å^2 after 165.2 ns in run 2 (figure 5.3D). The corresponding TCAs of the starting structure 3D5K (figure 5.3C) and the 1WP1 crystal structure (figure 5.3E) amount to 83.9 Å^2 and 125.3 Å^2, respectively. Starting from 59.4 Å^2 TCA in the 3D5K crystal structure, the inner bottleneck BNI also shows a slight opening trend leading to a TCA range from 60 to 84 Å^2 by the end of the simulations.

A summary of the average, minimum and maximum opening states of inner and outer periplasmic bottleneck compared to the 1WP1 and 3D5K crystal structures is given in table 5.1. Computing Connolly surfaces with different probe sphere radii

Table 5.1: OprM periplasmic opening states as observed in the crystal structures and our molecular dynamics simulations

	Outer bottleneck Val 408 (BNII)		Inner bottleneck Asp416 (BNI)	
OprM structure	avg TCA (Å^2)	avg Cα distance (Å)	avg TCA (Å^2)	avg Cα distance (Å)
3D5K	83.9	13.9	59.4	11.7
1WP1	125.3	17	38.7	9.5
Simulation	206.1 ± 45.5	21.9 ± 3	69.7 ± 11.3	12.7 ± 1.4
max	351.9	31.1	103.6	18.1
min	85.4	12	39.4	8.6

we find that the 3D5K X-ray structure is sterically permeable for molecules with a diameter of 1.2 Å or smaller, whereas the maximum BNI opening observed in one simulation at 107 ns permits the passage of molecules with a diameter of 6.4 Å or smaller.

5.4.3 Sodium Distributions and Binding Sites

Computing time-averaged 1D Na^+ density profiles along the z-axis, sodium z-trajectories as well as time and run-averaged spatial Na^+ density distributions, we searched for potential OprM sodium binding sites similar to the ones recently described for TolC (Raunest and Kandt, 2012b). Next to the lipid head groups the 1D Na^+ density profiles indicate a site of heightened sodium density in the lower equatorial region of the channel (figure 5.4A) continuously occupied throughout the entire simulation time as evident by a stable sodium trace in the Na^+ z-trajectory representatively shown for MD run 1 (figure 5.4B).

Zooming in on the individual sodium ions contributing to this trace, we find that once initial binding has occurred (during the first 5 - 40 ns), all three binding sites remain occupied but not by the same individual Na^+ ions as these exchange in all simulations (figure 5.5). Spatial sodium distributions in this region display maxima in sodium density exceeding 0.8 Na^+ / Å^3 near Asp171, Asp230 and Leu226 in each monomer (figure 5.4C). Counting the number of Na^+ - residue contacts within 3 Å we find that sodium predominantly interacts with Asp171 and Asp230 exhibiting run-averaged contact frequencies of 78 respectively 77% of the simulation time (figure 5.4D). Calculating the number of sodium ions simultaneously located within 3 Å of Asp171 and Asp230, we find single occupation of the Na^+ interaction site predominant, exhibiting run and time-averaged residence frequencies of 88 % in monomers A and B as well as 89% in monomer C (table 5.2).

To gain insight in the structural basis underlying the different sodium binding sites observed for OprM and TolC (Raunest and Kandt, 2012b) we compared the channel-internal distribution of Asp and Glu residues in the 3D5K OprM and 1EK9 TolC crystal structures (figure 5.6). In both cases in silico sodium binding is restricted to regions where two aspartates are located in close proximity to each other

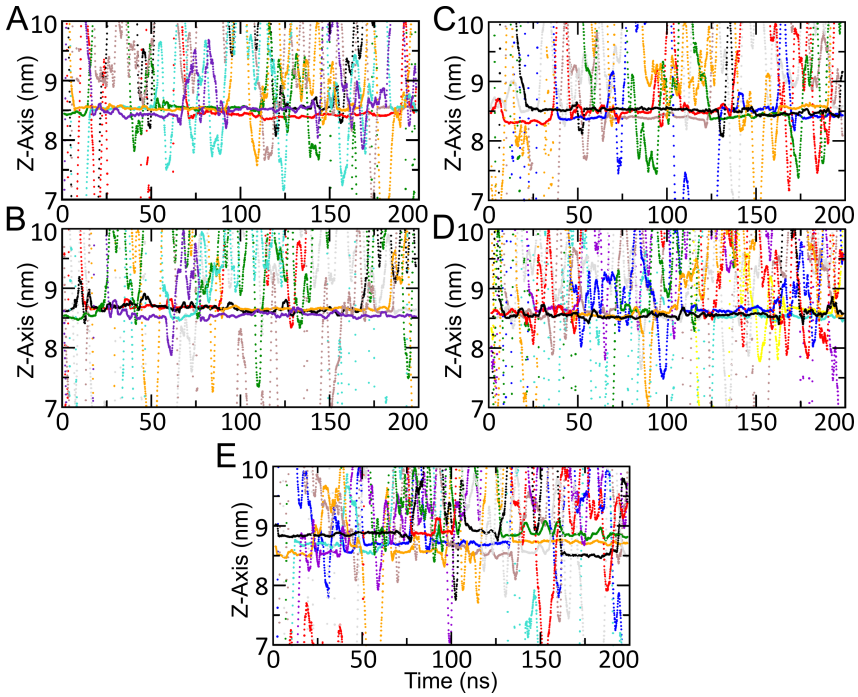

Figure 5.5: z-trajectories of the sodium ions occupying the observed Asp171/Asp230 Na^+ binding sites in each OprM monomer. Once initial sodium binding has occurred during the first 40 ns, in all five independent runs (A-E) the binding sites remain occupied for the rest of the simulation time albeit individual sodium ions do exchange. Vertical gaps between the sodium traces in (E) are due to a slightly tilted orientation OprM adopts in this simulation.

such as Asp374 and Asp371 in TolC's periplasmic bottleneck or Asp171 and Asp230 near the inward-facing side of the equatorial domain in OprM.

5.4.4 Simulation vs X-ray

Throughout our simulations, wild-type OprM adopts conformations on the extracellular (figure 5.2) and periplasmic side (figure 5.3) that have not been reported in the published crystal structures (Akama et al., 2004; Phan et al., 2010). To provide possible evidence explaining this discrepancy, we calculated the simulation average structure over all five simulations and compared it to the 3D5K X-ray structure (figure 5.7).

As indicated by superposition of the simulation average and X-ray structure after Cα least-squares fitting (figure 5.7A) and calculation of Cα displacements for each monomer (figure 5.7B), with an overall RMSD of 1.98 Å the conformational differences are small and the largest deviations occur in membrane-connecting N- terminus, the extracellular loop and the periplasmic tip regions. Particularly residues at the periplasmic tip region exceeding Cα displacements of 3 Å (highlighted in cyan in figure 5.7C) partially coincide with 4 Å crystal contacts (van der Waals

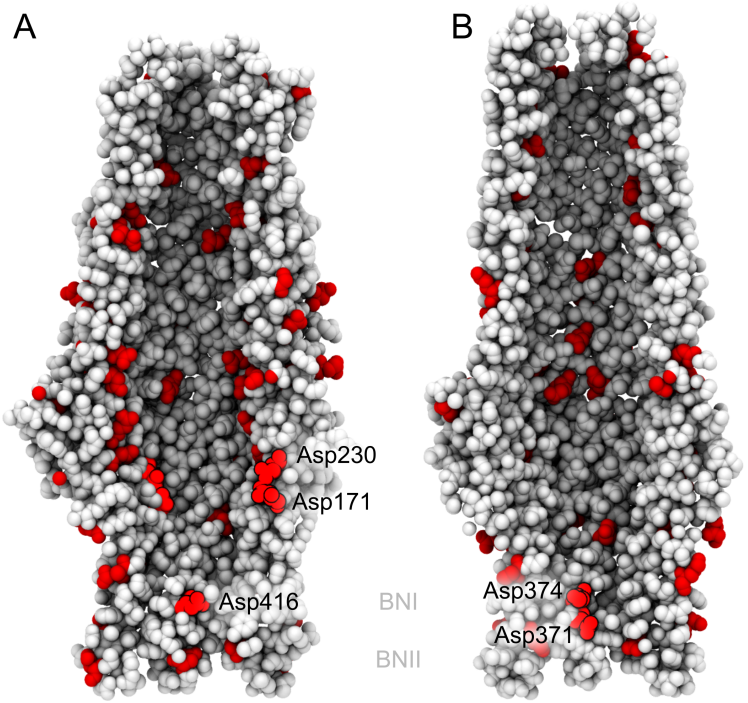

Figure 5.6: Aspartate and Glutamate (red) distribution in the 3D5K OprM (A) and 1EK9 TolC crystal structure (B). For a better view of the channel interior both proteins are shown in a cutaway representation. Aspartates constituting the inner bottleneck (BNI) (Asp416 in OprM, Asp374 in TolC) or involved in sodium binding (Asp171 and Asp230 in OprM, Asp374 and Asp371 in TolC) are highlighted in bright red.

Table 5.2: Run and time-averaged occupancy of the detected Asp171, Asp230 sodium interaction sites using a distance cutoff of 3 Å

Number of Na^+	Monomer A	Monomer B	Monomer C
0	10.8 %	10.6 %	10.1 %
1	88.3 %	88.4 %	88.9 %
2	0.9 %	1.0 %	1.0 %
3	0%	0%	0%

representation in figure 5.7C) in the 3D5K X-ray structure.

Additionally, we have analyzed the crystallographic B-factors of the OprM crystal structures plotting their Cα root-mean- square fluctuations along the membrane normal (figure 5.8). Both X-ray structures display flexibility patterns with maxima of root-mean-square fluctuations occurring at the extracellular and periplasmicend of the protein.

5.5 Discussion

In this study, we report a series of five independent, unbiased 200 ns MD simulations of wild-type OprM in a 150 mM NaCl phospholipid membrane / water environment, addressing the question of access regulation. On extracellular and periplasmic side OprM visits conformations throughout the simulations that have not been observed in the available OprM crystal structures. While opening and closing freely on ex-

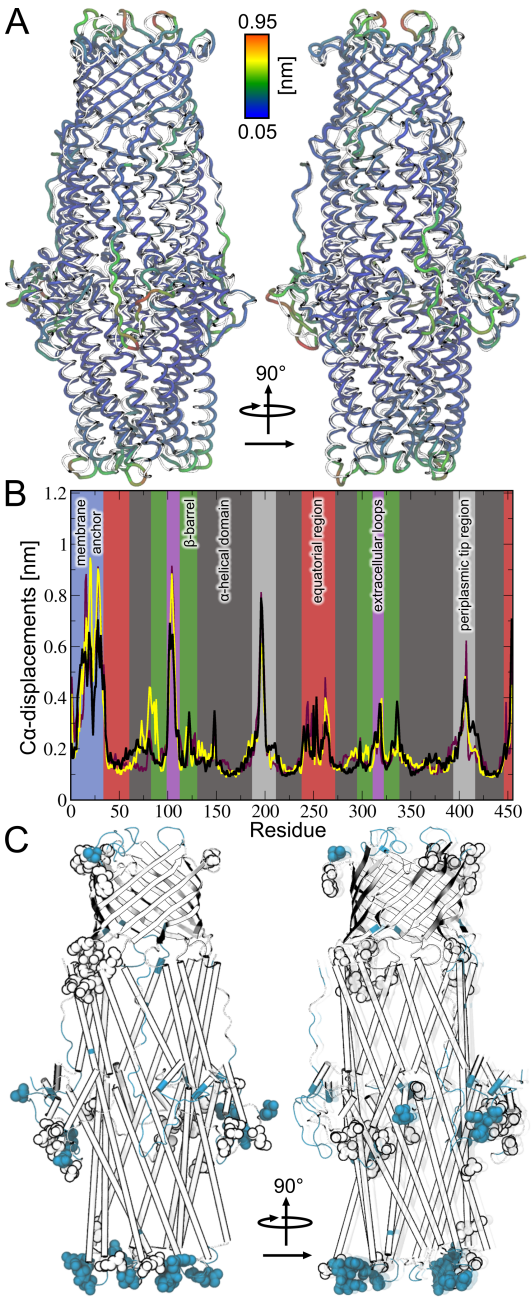

Figure 5.7: Superposition of the 3D5K OprM crystal structure (white) and the simulation average structure calculated over all simulations and colored by Cα displacement (A). Cα displacement graphs for each monomer indicate that the largest conformational differences occur in the extracellular loops, the periplasmic tips and the membrane anchor region (B). Residues with Cα displacements larger than 3 Å (dark cyan) partially coincide with 4 Å crystal contacts (van der Waals representation) in the 3D5K X-ray structure.

tracellular side we observe on periplasmic side an opening of the outer periplasmic bottleneck region (BNII) monitored at Val408. At the same time the inner periplasmic bottleneck (BNI) monitored through Asp416 shows a slighter opening trend restricting passage to molecules with a maximum diameter of 6.4 Å.

In addition to that we report a new sodium binding site at Asp171 and Asp230 located inside the OprM channel near the lower, periplasm-facing equatorial domain of the protein. We start this section discussing the limitations of our approach and then proceed to our findings and their biological implications.

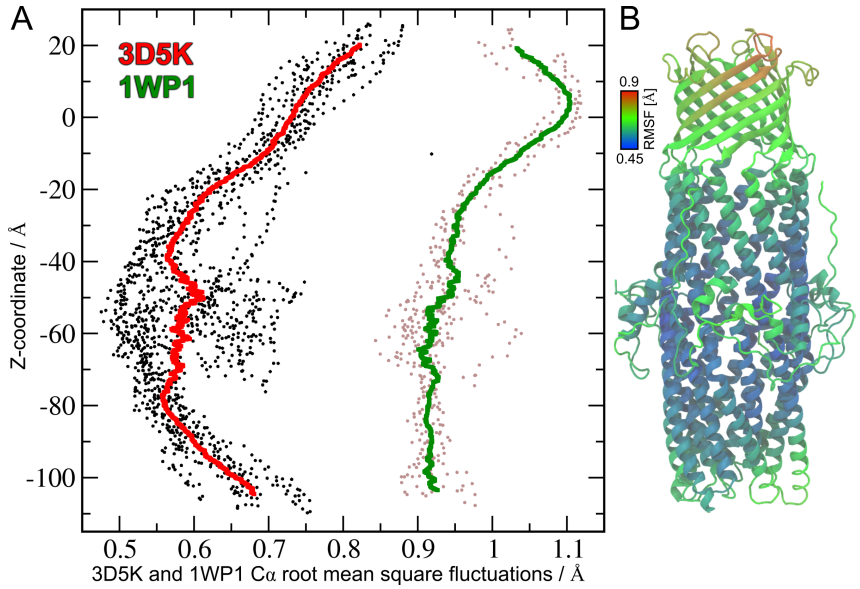

Figure 5.8: Crystal structure flexibility. Using the crystallographic B-factors, the backbone root-mean-square fluctuations (RMSF) of the OprM crystal structures is shown as profile along the membrane normal (A) as well as mapped onto the protein structure in case of the asymmetric 3D5K which we used as starting structure for our simulations (B).

5.5.1 Limitations of Our Approach

The microenvironment of OprM in the outer membrane of Pseudomonas aeruginosa is characterized by a variety of different lipids and membrane proteins (Stover et al., 2000), and whereas the extracellular side of OprM is exposed to lipopolysaccharides in the outer membrane leaflet, the periplasmic side is in contact with the peptidoglycan layer (Lambert, 2002). In contrast to these in vivo conditions, in silico we reduced the microenvironment to a simple POPE bilayer omitting a representation of the peptidoglycan layer.

While including both components in simulation studies would certainly be desirable, the development of the appropriate lipopolysaccharides and peptidoglycan molecular topologies is still in its infancy (Piggot et al., 2011, 2012; Straatsma and Soares, 2009) and at the time of writing such simulation parameters were not yet available for P. aeruginosa. On the other hand, our choice of a simple uniform membrane model excluding the peptidoglycan and other components is also justified by other simulation studies of outer membrane proteins making biologically relevant predictions using similar simplifications (Cuesta-Seijo et al., 2010; Gumbart et al., 2009; Hajjar et al., 2010; Raunest and Kandt, 2012b). Whereas one could argue that the OprM behavior we observe on the extracellular side could be an artifact of our simplified membrane model, it is encouraging that one of the first lipopolysaccharide simulations studying the outer membrane protein OprF (Straatsma and Soares,

2009) reported that the lipopolysaccharides have a stabilizing effect on the open conformation of the OprF extracellular loops that exhibit an architecture similar to that in OprM.

With any molecular dynamics study the question arises whether the simulation length is appropriate for the problem under investigation. In light of recent findings of atomistic membrane protein simulations showing that neither 300 ns TolC (Raunest and Kandt, 2012b) nor 1600 ns rhodopsin MD simulations (Grossfield and Zuckerman, 2009) were long enough for the protein structure to converge, our observation that OprM's Cα RMSDs still rise after 200 ns (figure 5.1B) is not surprising. With OprM sampling different regions of conformational space in each run (figure 1C), the protein structure has clearly not equilibrated yet, underscoring the need for long-time simulations.

However, our study's focus is on exploring wild-type OprM ground state dynamics near the X-ray structure and on the simulated 200 ns time scale the protein already displays unreported conformations (figure 5.2, 5.3) and ion interaction (figure 5.4) which might provide potential new insight into the OprM functional mechanism as discussed below. In that light, we consider the amount of conformational sampling achieved in our simulations adequate for the purpose of this investigation providing the first samples of membrane-embedded OprM dynamics at atomistic detail.

5.5.2 Simulation vs. X-ray

Undergoing opening and closing motions on extracellular side (figure 5.2) as well as opening on periplasmic side in the Val408 and to a lesser extent in the Asp416 region (figure 5.3), wild-type OprM adopts conformations throughout our simulations that have not been reported in the available crystal structures (Akama et al., 2004; Phan et al., 2010). A possible explanation for why these conformations have not been detected yet could be that the crystal environment sterically hinders the conformational changes we observe. This hypothesis is supported by the distribution of 4 Å crystal contacts in the 3D5K X-ray structure (figure 5.7C), coinciding particularly on periplasmic side with residues where the conformational differences between the simulation average and crystal structure are maximal (figure 5.7B, C).

Furthermore, with an overall Cα RMSD of 1.98 Å the conformational difference between simulation average and crystal structure is small, suggesting that similar to the results described in (Kandt et al., 2006; Raunest and Kandt, 2012b) the crystal might already contain the open and closed OprM conformations we observe in our simulations. In that case these conformations either constitute only a minority of the conformational ensemble in the 3D5K OprM crystal or on average are canceling each other out. Either way, the B-factor distributions in the 100 K (Phan et al., 2010) and 90 K (Akama et al., 2004) OprM crystal structures (figure 5.8), already imply a heightened degree of flexibility or crystal disorder on extracellular and periplasmic

side of the protein.

5.5.3 Extracellular Access

Whereas the available structural data (Akama et al., 2004; Phan et al., 2010) suggest that OprM access is restricted on both sides, our findings imply that OprM gating takes place only on the periplasmic side (figure 5.3) with OprM opening and closing freely on extracellular side (figure 5.2), suggesting here the absence of a gating mechanism in the isolated wild-type protein. Similar to our previous findings on the OprM homologue TolC (Raunest and Kandt, 2012b) the broad range of extracellular loop conformations visited throughout our simulations includes conformations both more closed and more open than the available X-ray structures (Akama et al., 2004; Phan et al., 2010) (figure 5.2C, D).

As indicated by the distribution of θ dihedral angles that we used to monitor the extracellular loops' opening state (figure 5.2C), with average θ dihedrals of 85° the X-ray conformations likely represent only a small section of possible loop conformations compared to the 30° - 160° θ range we observe in our MD simulations.

Whereas the OprM crystal structures limit extracellular access to compounds with diameters of up to 8 - 10 Å (which we estimated computing Connolly surfaces using different probe sphere radii), we find that OprM accessibility is sterically only limited by the inner diameter of the β-barrel itself when all three loops are open (figure 5.2D). If our simulations of the isolated protein are correct, these results could either indicate the presence of additional gating mechanisms in vivo or hint at the possibility of designing a novel group of OprM-directed drugs specifically targeting the protein interior.

Possible experimental scenarios for testing the hypothesis of extracellular OprM accessibility could involve double-spin-label electron spin resonance spectroscopy experiments measuring the distance between the tip regions of the extracellular approach. Another approach could be fluorescence spectroscopy experiments in which first a fluorescence dye-specific cross-linker is introduced into the OprM interior and then fluorescence activity is checked after the external application of dyes of different sizes followed by protein extraction and purification.

Beyond that, the extracellular accessibility behavior we observe for OprM bears a striking similarity to our previous findings on its *E. coli* homologue TolC (Raunest and Kandt, 2012b), displaying the same unrestricted opening and closing motions on extracellular side. In that light one could hypothesize that unregulated extracellular access might represent a general characteristic shared by all similarly structured efflux ducts. It will be interesting to see this possibility investigated further in future wet lab and computer experiments.

5.5.4 Periplasmic Access and Sodium Binding Sites

Whereas for the OprM *E. coli* homologue TolC combined mutagenesis - conductance experiments (Andersen et al., 2002a,b) as well as wild-type and mutant crystal structures (Bavro et al., 2008; Pei et al., 2011) indicated the existence of an inner (BNI) and outer periplasmic bottleneck (BNII), to the best of our knowledge corresponding investigations on OprM have not been reported yet. Given that (a) in TolC Asp374 and Gly365 have been used to monitor the opening states of the periplasmic bottlenecks (Bavro et al., 2008; Raunest and Kandt, 2012b; Schulz and Kleinekathöfer, 2009), and (b) the OprM structure is very similar to TolC (Phan et al., 2010) (figure 5.3A), we monitored OprM's periplasmic opening state using Asp416 (BNI) and Val208 (BNII) as these residues display the closest distance to their TolC counterparts after superimposing the OprM 3D5K and TolC 1EK9 wild-type crystal structures (figure 5.3A).

The following section is based on the assumptions that OprM too contains two periplasmic bottlenecks and Asp416 and Val408 are appropriate residues to monitor the bottlenecks' respective opening states. Whereas in TolC BNI and BNII display similar opening states in the wild type X-ray structures (Koronakis et al., 2000), inner and outer periplasmic bottleneck already differ by 21 (BNI) respectively 41 $Å^2$ TCA in the symmetric 1WP1 and asymmetric 3D5K OprM crystal structures (Akama et al., 2004; Phan et al., 2010). Throughout our simulations that difference becomes more pronounced in all runs (figure 5.3, table 5.1). BNII opens up in all simulations, reaching TCA values 2.08 - 2.98 times higher than in the 3D5K starting structure by the end of the simulations.

BNI on the other hand also displays an opening trend albeit much less pronounced, resulting in 1.08 - 1.4 times higher TCA values after 200 ns. At the same time we observe in all runs the occurrence of a previously unreported sodium binding site located inside the OprM channel near the equatorial domain where in each monomer the site is predominantly occupied by a single Na^+ ion coordinated by Asp171 and Asp230 (figure 5.4, table 5.2).

If our simulations are correct the observed BNII and BNI behavior could be interpreted in a similar way as our previous findings on TolC (Raunest and Kandt, 2012b) in so far that OprM gating on periplasmic side involves only the inner but not the outer bottleneck. Moreover the open BNII conformation could play a role in the formation of the functional complex, initiating interaction with the inner membrane transporter or the adaptor protein. Although not observed throughout the simulations, BNII closure in isolated OprM sodium-induced as seen for TolC (Raunest and Kandt, 2012b) or otherwise - cannot be excluded based on the available data.

Confining the BNII opening state by cross linking and subsequent OprM activity measurements could be one way to investigate this aspect experimentally. Whether the slight BNI opening trend of 5 - 24 $Å^2$ increase in TCA is negligible, a relaxation

effect or represents the beginning of a longer conformational transition inducing a periplasmic opening, cannot be decided based on our simulation data alone. However, the two available OprM crystal structures already displaying a much higher variation in the bottlenecks' opening state than the corresponding TolC structures could be interpreted as hinting at a principal functional difference between the two proteins.

If the observed BNI opening trend is negligible this would speak in favor of isolated OprM capable of gating when reconstituted in a POPE environment. If the BNI opening trend is not negligible it seems reasonable to assume that periplasmic OprM access regulation requires an additional component that is absent in our simulations. Electrophysiological conductance experiments on isolated OprM could shed some light onto this question. Either way, to gain deeper insight into the OprM gating and functional mechanism additional experiments are required which could involve similar approaches as carried out for TolC (Andersen et al., 2002a,b) complemented by double spin label EPR or fluorescence spectroscopy studies to monitor and quantify the OprM opening state on periplasmic side.

Partial and spatial density analysis indicates for each monomer the presence of a so far unreported sodium binding site coordinated by Asp171 and Asp230 (figure 5.4, 5.5). Unlike our previous findings on TolC (Raunest and Kandt, 2012b) our OprM data do not support an involvement of the sites in a Na^+ dependent periplasmic lock mechanism as only low Na^+ densities are observed in the bottlenecks region and the removal of all NaCl does not alter OprM's periplasmic opening behavior nor does it lead to changes in the protein dynamics or conformation on a 200 ns time scale (data not shown).

OprM's unsensitivity to NaCl also speaks against a structure stabilizing function of the three Na^+ sites as does the lack of electron density data in the OprM crystal structures in the region of the identified sites which could be interpreted as bound sodium (Phan et al., 2010). Two possible explanations are conceivable accounting for the different sodium interactions we observed for OprM and TolC. Either both proteins share a sodium-dependent lock mechanism or they employ different means of periplasmic access regulations. Given that each MD simulation is always limited by the amount of conformational sampling achieved, the instance that we did not observe any TolC-like sodium binding in OprM's periplasmic bottleneck region does not rule out the possibility that OprM and TolC employ a similar gating mechanism.

Longer simulation times might very well yield a different picture than our 200 ns samples. However, spontaneous sodium binding does occur in all our OprM runs too - albeit at a different position - which speaks for an adequate length of simulation time. Given the different distributions of negatively charged residues in the OprM and TolC channel interior and the observation that in silico sodium binding in OprM and TolC is restricted to neighboring aspartates (figure 5.6) of which TolC exhibits one near the BNI region and OprM near the inward-facing side of the equatorial domain, it appears reasonable to assume that despite their structural similarity

OprM and TolC are indeed different, employing different means of periplasmic access regulation.

In both cases more experimental work is needed to further elucidate the functional mechanism of efflux duct gating and the role of the potential sodium or comparable cation binding sites, for example by activity assays comparing wild-type and mutant OprM under different ionic conditions.

5.6 Conclusions

To gain insight into the OprM functional mechanism of gating and access regulation we computed 200 ns dynamics samples of the wild type protein in a phospholipid membrane / 150 mM NaCl water environment in a series of five independent and unbiased atomistic molecular dynamics simulations. On extracellular side we find OprM opening and closing freely in all runs, suggesting the absence of a gating mechanism on this side in the isolated protein. Assuming a similar architecture as in its *E. coli* homologue TolC, comprising an inner and outer bottleneck region, we monitored OprM's periplasmic opening state using Asp416 and Val408 which we selected based on their proximity to their TolC counterparts after superimposing OprM and TolC X-ray structures.

In all simulations an opening of both bottlenecks occurs albeit stronger pronounced in the outer bottleneck region reaching end conformations up to 3 times more open than the starting crystal structure, while the inner bottleneck is only 1.1 to 1.4 times more open by the end of the simulations. If our simulations are correct our findings imply that periplasmic gating occurs only in the inner bottleneck region at Asp416 and that in vivo additional components, absent in our simulations, might be required for gating if the observed opening trend of the inner bottleneck is not negligible. In addition to that we identified in each monomer a previously unreported sodium binding site in the channel interior coordinated by Asp171 and Asp230.

Chapter 6

dxTuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics

At the beginning of my Ph.D. all available molecular cavity detection tools were based on static structures and did not take protein and solvent dynamics into account. With dxTuber we introduced the first cavity detection tool relying on protein ensembles taken from molecular dynamics (MD) simulations. The voxel based routine has a graphical (GUI) and command line interface (CLI) and stores results in the common used PDB file format. dxTuber is licensed under the GPL v2[1] and can be downloaded at http://www.csb.bit.uni-bonn.de/dxtuber.html or http://code.google.com/p/dxtuber/.

Adapted from:
Raunest, M. and Kandt, C. (2011a). dxtuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J Mol Graph Model*, 29(7):895–905

[1] http://www.gnu.org/licenses/licenses.html#GPL

6.1 Abstract

Empty space in a protein structure can provide valuable insight into protein properties such as internal hydration, structure stabilization, substrate translocation, storage compartments or binding sites. This information can be visualized by means of cavity analysis. Numerous tools are available depicting cavities directly or identifying lining residues. So far, all available techniques base on a single conformation neglecting any form of protein and cavity dynamics.

Here we report a novel, grid-based cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics simulations to identify (I) internal cavities, (II) tunnels or (III) clefts on the protein surface. Driven by a graphical user interface, output can be exported in PDB format where cavities are described as individually selectable groups of adjacent voxels representing regions of high solvent residence probability. Cavities can be analyzed in terms of solvent density, cavity volume and cross-sectional area along a principal axis.

To assess dxTuber performance we performed test runs on a set of six example proteins representing the three main classes of protein cavities and compared our findings to results obtained with SURFNET, CAVER and PyMOL.

6.2 Introduction

The successful determination of a protein's high-resolution three-dimensional structure is a landmark on the way to understanding its function. Next to the actual atomic XYZ coordinates, 3D structures often contain additional information that can yield further insights into the protein in question. For example, empty space in a protein structure can provide valuable insight into protein properties such as internal hydration, structure stabilization, substrate translocation, storage compartments or substrate binding sites (Schmidtke et al., 2010; Volkamer et al., 2010). This information can be visualized by means of cavity analysis. Over the years numerous cavity detection tools have been developed including (Coleman and Sharp, 2009; Delano, 2002; Exner et al., 1998; Ho and Gruswitz, 2008; Kleywegt and Jones, 1994; Laskowski, 1995; Liang et al., 1998; Maeda and Kinoshita, 2009; Nicholls et al., 1991; Pellegrini-Calace et al., 2009; Petrek et al., 2007, 2006; Smart et al., 1996; Voss and Gerstein, 2010; Yaffe et al., 2008b) that depict cavities either directly (Coleman and Sharp, 2009; Delano, 2002; Exner et al., 1998; Ho and Gruswitz, 2008; Laskowski, 1995; Maeda and Kinoshita, 2009; Nicholls et al., 1991; Pellegrini-Calace et al., 2009; Petrek et al., 2007, 2006; Smart et al., 1996; Voss and Gerstein, 2010; Yaffe et al., 2008b) or indirectly by identifying lining residues (Liang et al., 1998) or filling a cavity with water molecules (Kleywegt and Jones, 1994). The main strategies used in these geometry-based algorithms (Schmidtke et al., 2010) can be grouped into four categories plus combinations of these.

One approach employs the rolling probe sphere method (Connolly, 1983a,b; Lee and Richards, 1971; Richards, 1977) where molecular surfaces are calculated combining different probe radii and surface types in order to (a) distinguish external from internal residues and (b) detect protein cavities. Examples of this approach include, GRASP (Nicholls et al., 1991), 3V (Voss and Gerstein, 2010) and PyMOL since version 1.3 (Delano, 2002).

Specifically aiming at pores and tunnel-like cavities, another strategy calculates pathways connecting protein interior and exterior. Most tools employing this approach begin with a user-defined starting point inside the protein from where an exit pathway is calculated while simultaneously recording a profile of the pore dimensions. Expressed as radius or diameter of a tunnel-filling sphere at given way points, such pore profiles allow a detailed characterization of the channel in question. Examples of this approach include HOLE (Smart et al., 1996), its successor CAVER (Petrek et al., 2006) and MoleAxis (Yaffe et al., 2008b). On the other hand, CAST (Liang et al., 1998) and the algorithm introduced by Maeda and coworkers (Maeda and Kinoshita, 2009) resort to alpha shape theory, Voronoi diagrams and Delaunay tessellation (Edelsbrunner and Mucke, 1994; Poupon, 2004; Richards, 1974) to identify protein cavities. Whereas the CAST algorithm is intended for general cavity detection, the program by Maeda et al. specifically focuses on interprotein cavities.

The fourth category of tools describes cavities as entities filled with (a) probe

spheres of parameterizable size as in SURFNET (Laskowski, 1995) and HOLLOW (Ho and Gruswitz, 2008), (b) water molecules as in VOIDOO (Kleywegt and Jones, 1994) or (c) voxels like in the algorithm introduced by Exner and coworkers (Exner et al., 1998) or dxTuber in this publication. Examples where combinations of these main strategies have been implemented include the CAVER successor MOLE (Petrek et al., 2007), PoreWalker (Pellegrini-Calace et al., 2009) and CHUNNEL (Coleman and Sharp, 2009). Whereas MOLE and PoreWalker employ the path finding approach in conjunction with Voronoi diagrams and alpha shape theory, CHUNNEL combines the path finding strategy with probe-sphere based molecular surfaces. However, unlike other path finding programs, CHUNNEL calculates multiple pathways leading into the protein without requiring a user-defined starting point.

All available techniques of cavity detection published to date have in common that they detect cavities based on a single protein conformation, neglecting any form of protein and cavity dynamics. Here we introduce a novel cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics (MD) simulations to detect cavities, tunnels and surface clefts. Using the VolMap plugin in VMD 1.8.7 (Humphrey et al., 1996), solvent and protein trajectories are converted to a voxel representation of OpenDX mass-weighed spatial density maps which serve as input files for our algorithm.

The general work flow of the analysis procedure is summarized in figure 6.1. Driven by a graphical user interface, dxTuber then separates protein-internal from protein-external solvent voxels, describing cavities as groups of adjacent voxels that represent protein-internal regions of high solvent residence probability. To this end three different search algorithms have been implemented aiming at the detection of three major types of protein cavities: (I) isolated protein internal cavities with no connection to the protein exterior - both intra- and inter-molecular, (II) tunnel-like cavities with at least one connection to the protein exterior and (III) clefts on the protein surface.

dxTuber results can be exported in PDB format where each cavity is an individually selectable object of coherent voxels written as pseudo atoms. For each voxel solvent residence probability information is stored as averaged and mass-weighed solvent density encoded as formal B-factors. Cavities and solvent density can be readily visualized and processed using common molecular viewers. Cavities can also be analyzed in terms of volume and profiles of cross-sectional area along a principal axis. The profiles are exportable in ASCII file format, and can be processed by common spreadsheet applications. dxTuber is available free of charge at `http://www.csb.bit.uni-bonn.de/downloads.html` and `http://code.google.com/p/dxtuber/`.

To assess dxTuber performance we performed test runs on a set of six example proteins (Buhrman et al., 2010; Karpowich et al., 2003; Koronakis et al., 2000; Schemberg et al., 2007; Schnell and Chou, 2008; Vreede et al., 2010) representing each of the three cavity classes by a large and small variant (table 1) and compared our findings to results obtained with SURFNET, CAVER and PyMOL. While an

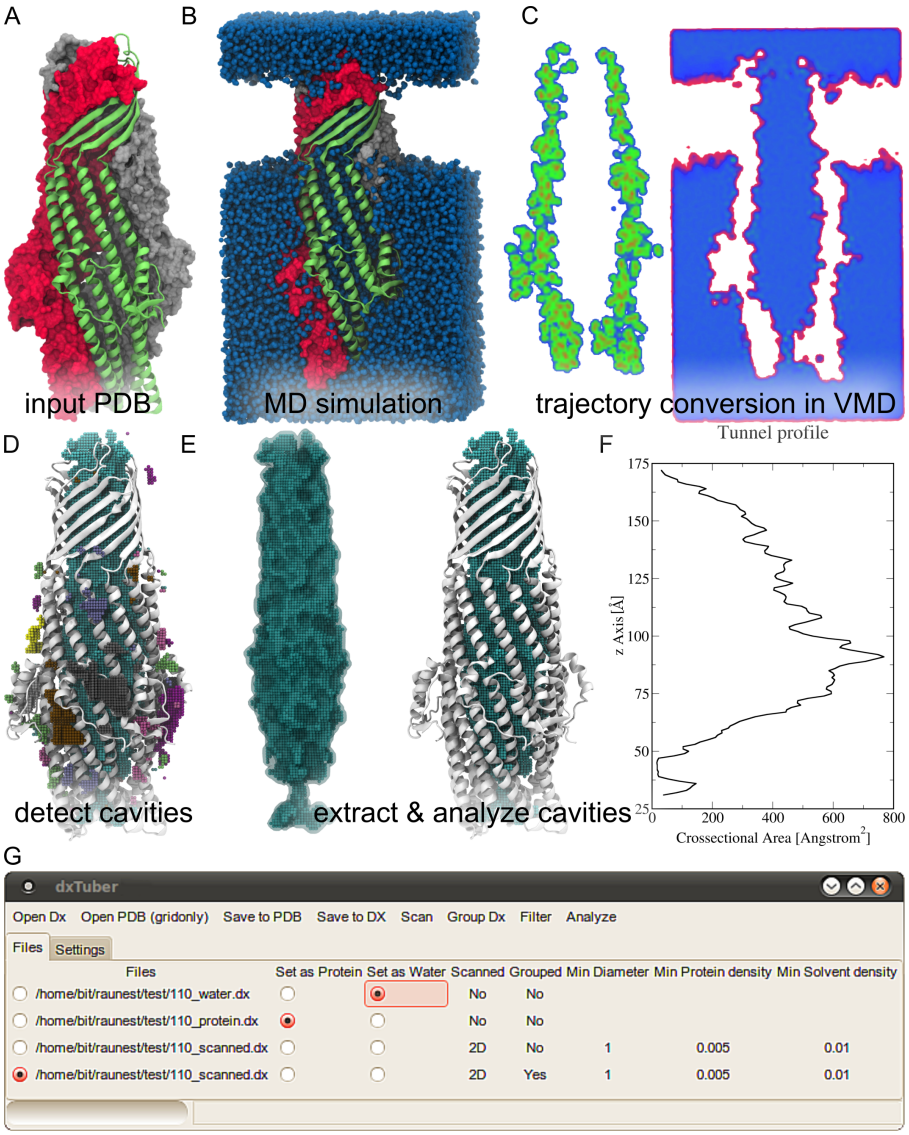

Figure 6.1: dxTuber work flow. Starting from an experimentally determined protein structure (A) a molecular dynamics simulation is performed to sample protein and solvent dynamics (B). Trajectories are converted to voxel-based residence probabilities using the VolMap function in VMD (Humphrey et al., 1996) (C). The resulting time-averaged and mass-weighted OpenDX protein and solvent densities constitute the dxTuber input files used to separate inside from outside solvent densities to determine cavities, channels or surface clefts (D). Cavities can be exported as OpenDX or PDB files where each cavity is stored as an individually selectable group of voxels (E). Further analysis can be done computing cavity volume or profiles of cross sectional area along a principal axis (F). dxTuber is controlled via a graphical user interface (G).

in-depth comparison of all cavity detection tools would certainly be desirable, such a review is beyond the scope of this paper. In our benchmark tests we therefore

restricted ourselves to these three cavity detection applications which we selected based on their wide spreading in the community and their frequent usage in our lab.

6.3 Metherials and Methods

6.3.1 Creating DX input Files

MD Simulations

To generate a sample of solvent and protein dynamics, MD simulations are carried out with the protein of interest solvated in, for example, water (figure 6.1A). The simulation time required depends on the diffusion velocity of the solvent and whether protein atoms are position-restrained. The usage of position restraints allows controlling the amount of protein dynamics to be probed. If the cavity structure close to the protein X-ray conformation is of interest, protein dynamics can be reduced or disabled using position restraints during the MD. Without position restraints, full protein and cavity dynamics will be considered.

For water as solvent, simulation times of 100 ps where found sufficient to sample cavities, clefts and tunnels when position restraints of 1000 $\frac{kJ}{mol\ nm^2}$ were used to keep the protein close to its X-ray conformation. Due to extended equilibration times of protein and environment, unbiased MD requires longer simulation time, depending on the system of interest.

VolMap Options

VolMap offers different options to convert MD trajectories into OpenDX density maps. Whereas “max” and “min” assign to each voxel the highest or lowest number of atoms detected at this position throughout the trajectory, “avg” yields a time-averaged value of the number of atoms per voxel (figure 6.2A).

For the use with dxTuber we applied two combinations of settings: (a) “average” for both water and protein and (b) “average” for water and “minimum” for protein atoms (table 6.2). The latter setting filters out regions of high protein flexibility, yielding the maximum volume accessible by water and thus the largest possible extension of a given cavity. To eliminate protein drift in the simulation box, trajectories should be aligned to a reference structure prior to VolMapping.

6.3.2 Cavity Detection

Scanning Density Voxels

To separate protein-interior solvent densities from exterior ones (figure 6.2), three search algorithms have been implemented in dxTuber, each aiming at the detection of a different type of cavity (figure 6.2B). Whereas 3D scanning is intended for isolated buried cavities in the protein interior, and 2D scanning primarily aims for

channel-like cavities that have a connection to the protein exterior, 1D scanning is recommended for clefts on the protein surface. In each method every solvent voxel - with a density exceeding 0.01 atoms/Å^3 - is tested if any protein voxels are located in both directions of the three principal axes. Voxel by voxel the algorithm proceeds away from the tested solvent voxel in positive and negative direction of the X, Y and Z principal axes until either a protein voxel is found or the end of the simulation box is reached. Depending on the scanning criterion used, different conditions of protein voxel distributions need to be fulfilled before a solvent voxel is considered protein-internal.

For 1D scanning at least one protein voxel in each direction of one principal axis is required (figure 6.2C). For 2D scanning at least one protein voxel in each direction of two principal axes is necessary (figure 6.2D), whereas 3D scanning expects at least one protein voxel in each direction of all three principal axes before a solvent voxel is classified as protein-internal (figure 6.2E). 2D scanning is the default method and recommended for an initial scan. Protein voxels are identified based on a minimum density cut-off of 0.005 protein atoms/Å^3. This is the standard setting that we found to give good results; however, protein and solvent density thresholds are freely adjustable by the user.

After scanning, results may contain false positives, i.e. incorrectly assigned internal solvent voxels (ISVs) on the protein surface. As long as these are not connected to true positives, dxTuber can separate them via subsequent grouping. However, if false positives are connected to true positives, a more restrictive scanning method should be applied. In terms of restrictiveness dxTuber's scanning algorithms are ranked 3D ¿ 2D ¿ 1D. Beyond that, results can be post-processed by filtering.

Grouping Cavities and Analysis

Detected ISVs can be summarized into groups of voxels representing protein cavities. To that end dxTuber loops over the coordinate system of the ISVs starting in Z direction followed by Y and then X (figure 6.3A). The first ISV detected will become group one. Subsequent ISVs are tested based on their already visited neighborhood to decide whether they are part of a known group or constitute a new one (figure 6.3B). If ISVs belong to different groups that are later found to nevertheless have common neighbors, all respective voxels will be merged into a single group that is assigned the smaller (= earlier) group identification number (figure 6.3C). Once completed, ISVs have been clustered into individual cavities (figures. 6.1D and 6.3B).

Once ISVs have been grouped into cavities, dxTuber can derive the cross-sectional area of a detected cavity along a selectable principal axis (X, Y or Z). To that end dxTuber loops over the chosen axis, counting in each step the number of ISVs in the plane perpendicular to the search axis. That way cross-sectional profiles can be generated which are exported as ASCII text files (figure 6.1F).

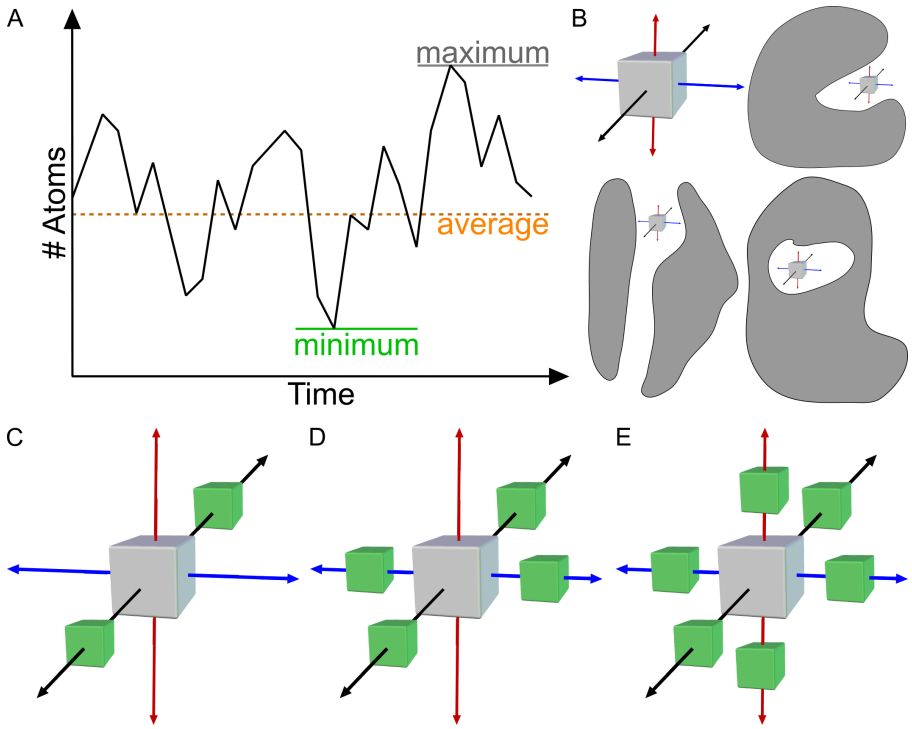

Figure 6.2: To convert an MD trajectory to a voxel representation VMD can calculate the average, minimum or maximum atom density per voxel (A). Depending on the type of cavity to be detected (B), dxTuber offers three search algorithms using different criteria to decide whether a solvent voxel (grey) is considered protein-internal (C-E). Aiming at surface clefts, 1D scanning requires at least one protein voxel in each direction of one principal axis (C). Intended for tunnel-like cavities, 2D scanning requires at least one protein voxel (dark grey) in each direction of two principal axes (D), whereas 3D scanning expects at least one protein voxel in each direction of three principal axes before a solvent voxel is regarded protein-internal.

Filtering

Next to grouping or employing a more restrictive scanning algorithm, dxTuber results can be filtered to reduce the amount of false positives. In particular if more restrictive scanning is not successful due to incomplete cavity detection or if false positive ISVs are still connected to true positives subsequent filtering is recommended.

By applying the neighbor filter each ISV is re-evaluated based on its 26 neighbor voxels and will be deleted if the amount of neighbor ISVs is smaller than a user-defined minimum. The default for this minimum threshold is ten. The number of neighboring ISVs for the filter option should be set carefully, since filtering always leads to a loss of data as true positives not fulfilling the minimum neighbor criterion are also deleted. Therefore the threshold should be set to a low value (¡ 10) to keep data loss to a minimum. The filter works best if two cavities of interest are connected by a thin line of voxels.

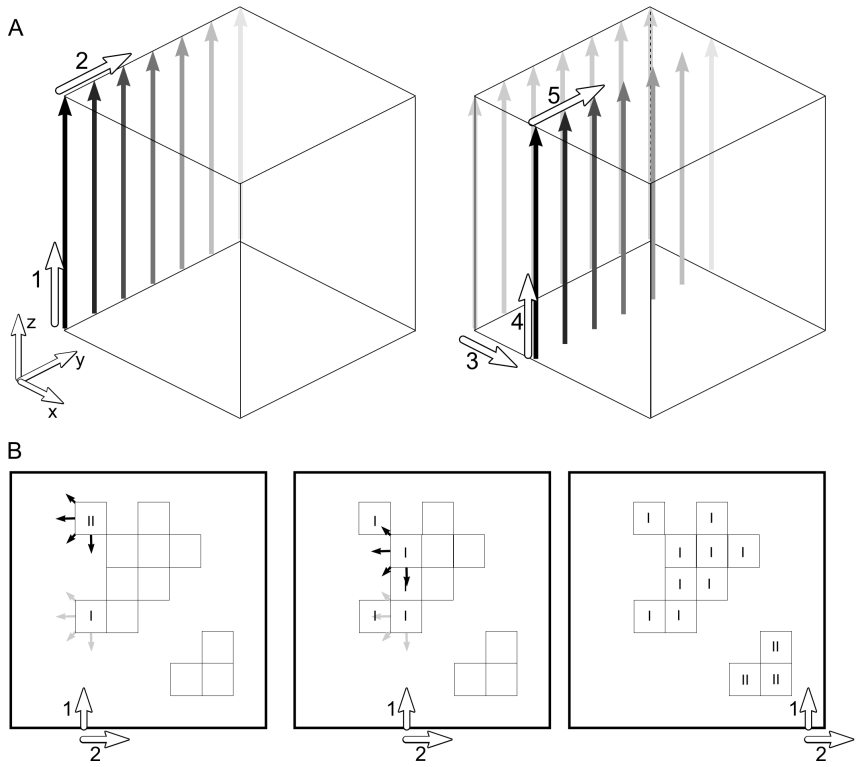

Figure 6.3: To group protein-internal solvent voxels (ISVs) into cavities, dxTuber loops over the coordinates space spanned by the ISVs starting in Z direction, followed by Y and then X (A). The first ISV detected will become group one. Subsequent ISVs are tested based on their already visited neighborhood to decide whether they are part of a known group or constitute a new one (B, left). If ISVs belong to different groups that are later found to nevertheless have common neighbors (B, middle), all respective voxels will be merged into a single group that is assigned the smaller (= earlier) group identification number (B, right).

6.3.3 Example Applications

Test Proteins

Each of the three main classes of protein cavities is represented by two proteins, each featuring a large and small variant of a given cavity type. Whereas isolated internal cavities are represented by 2ZOH photoactive yellow protein PYP after chromophore removal (Vreede et al., 2010) and the 2OGX molybdenum/tungsten storage protein MoSto (Schemberg et al., 2007), 1EK9 TolC (Koronakis et al., 2000) and the 2RLF Influenza M2 proton channel (Schnell and Chou, 2008) serve as examples for tunnel-like cavities. Clefts on the protein surface are represented by the 1N4D periplasmic vitamin B12 binding protein BtuF (Karpowich et al., 2003) and the 3K8Y GTPase H-Ras P21 (Buhrman et al., 2010).

Table 6.1: Test proteins

Protein		PDB	Cavity type
PYP	(Vreede et al., 2010)	2ZOH	Small intra-molecular cavity
MoSto	(Schemberg et al., 2007)	2OGX	Large inter-molecular cavity
TolC	(Koronakis et al., 2000)	1EK9	Wide continuous tunnel
M2	(Schnell and Chou, 2008)	2RLF	Narrow continuous/discontinuous tunnel(s)
BtuF	(Karpowich et al., 2003)	1N4D	Large ligand binding cleft
H-Ras	(Buhrman et al., 2010)	3K8Y	Small ligand binding cleft, partially occluded

Simulation Settings

Molecular dynamics simulations were performed using the GROMACS 4.0.3 package (Berendsen et al., 1995b; Hess et al., 2008) and GROMOS96-53a6 force field (Oostenbrink et al., 2004). In the simulations, all bond lengths were constrained by LINCS so that an integration time step of 2 fs could be chosen (Hess et al., 1997). Systems were simulated at a temperature of 310 K, maintained separately for protein, lipids (where present) and water by a Berendsen thermostat (Berendsen et al., 1984) with a time constant of T = 0.1 ps.

Pressure coupling was done using a Berendsen barostat (Berendsen et al., 1984) employing 1 bar reference pressure and a time constant of 4 ps. Whereas isotropic pressure coupling was employed for the water-soluble proteins PYP, MoSto, H-Ras P21 and BtuF, semiisotropic pressure coupling was used for the membrane proteins TolC and M2 to allow for bilayer fluctuations in the membrane plane. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation (Darden et al., 1993; Essmann et al., 1995) and twin range cut-offs of 1.0 nm and 1.4 nm were applied for computing the van-der-Waals interactions. TolC and M2 were inserted in pre-equilibrated 9.6 nm x 9.6 nm and 6.8 nm x 6.8 nm palmitoyloleoylphosphatidylethanolamine (POPE) bilayer patches based on (Tieleman and Berendsen, 1998) using INFLATEGRO (Kandt et al., 2007).

All systems were solvated with simple point charge water molecules (Berendsen et al., 1981). Standard protonation states were assumed for titratable residues except for M2 which was simulated with His37 in both protonated and deprotonated form. An overview of simulation times and position restraints applied in each system is given in table 6.2.

dxTuber, SURFNET, CAVER, and PyMOL Settings

All dxTuber analyses were run on average protein and water densities calculated at 1 $Å^3$ resolution via VolMap in VMD 1.8.7. For M2, additional minimum protein densities were computed whereas for PYP and TolC additional density map resolutions of 0.5 and 2.0 $Å^3$ were calculated to assess how voxel resolution affects calculation speed. dxTuber search algorithms and filters used are detailed in table

Table 6.2: Molecular dynamics simulation, VolMap and dxTuber settings

Protein	Simulation time	Used for cavity detection	Restraints on protein	Figure	dxTuber settings	Calculated Protein — Solvent densities
PYP	100 ps — 10 ns	All — all	Yes — no	6.4	3D	Avg — avg
MoSto	100 ps	All	Yes	6.4	3D	Avg — avg
TolC	20 ns	All	Yes	6.5	2D	Avg — avg
TolC	21ns	Last 10 ns	No	6.5	2D	Avg — avg
M2 †	20 ns	Last 10 ns	No	6.6	2D	Avg and min — avg
M2 ‡	20 ns	Last 10 ns	No	6.6	2D	Avg and min — avg
BtuF ♮	100 ps	All	Yes	6.7	1D (z-Axis)	Avg — avg
H-Ras	100 ps	All	Yes	6.7	2D	Avg — avg

† His37 deprotonated
‡ His37 protonated
♮ Neighbor 6 filter was applied for BtuF

6.2. SURFNET (Laskowski, 1995) has three main user-adjustable parameters affecting the cavity search: grid size, minimum gap sphere radius and maximum gap sphere radius. To keep the numer of false positives to a minimum these parameters were optimized for each protein (table 6.3). Main cavities were isolated using the SURFNET tool MASK. As CAVER (Petrek et al., 2006) is designed to find a path that leads out of the protein interior, cavity analysis requires a set of starting coordinates inside the protein as well as the number of how many different exit pathways to look for. Both parameters were optimized for each protein to describe the cavities of interest as accurate as possible (table 6.4).

For better comparison with dxTuber, CAVER's radius-based tunnel profiles were converted to cross-sectional area profiles. Implemented since version 1.3, cavity detection in PyMOL (Delano, 2002) is based on the MSMS molecular surface algorithm (Sanner et al., 1995, 1996) and controlled via the search parameters cavity detection radius and cavity detection cut-off. The settings used for each protein are listed in table 6.5. All cavity calculations were performed using a single Intel® Xeon® E5410 2.33 GHz CPU on a 4 GB RAM DELL Precision T5400 workstation running 32 bit Ubuntu Linux 10.04.

All results obtained with CAVER and SURFNET and PyMOL are based on the reference X-ray structures, except for M2: representative for the channel's open and closed state, simulation snapshots at 18.7 ns (open state) and 13.9 ns (closed state) were used. The conformations were selected based on their minimum Cα-RMSD to the simulation average structures calculated over the last 10 ns of the trajectories. All molecular illustrations were created using VMD (Humphrey et al., 1996) or PyMOL (Delano, 2002).

6.4 Results

We developed a novel cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics simulations to detect isolated internal

Table 6.3: SURFNET search parameters

Protein	Grid size [Å]	Minimum probe radius [Å]	Maximum probe radius [Å]
PYP	0.8	1.5	5.5
MoSto	2	2.5	10
TolC	2	3	10
M2 †	0.8	1.5	5.5
M2 ‡	0.8	1.6	5.5
BtuF	0.8	3	10
H-Ras	0.8	1	10

† His37 deprotonated
‡ His37 protonated

Table 6.4: CAVER search parameters

Protein	Start coordinates [distance to reference atoms]			Exit pathways
TolC	Thr163-CG2 chain Å: 11 Å, B: 10.3 Å, C: 9.9 Å			2
M2 †	Trp19-HZ2 chain Å: 6.6 Å, B: 4.6 Å, C: 4 Å			5
	His15-CG chain Å: 4.5 Å, B: 2.3 Å, C: 12.3 Å			2
M2 ‡	Trp19-CG chain Å: 5.8 Å, B: 7.5 Å, D: 3.5 Å			11
BtuF	Gln154-NE2: 1.7 Å			3
	Phe222-O: 2.9 Å			
	Glu223-O: 2 Å			
H-Ras		1st coordinate	2nd coordinate	1 from 1st coordinate
	GppNHp-PB:	2.0 Å	2.1 Å	2 from 2nd coordinate
	GppNHp-PG:	1.1 Å	1.0 Å	
	GppNHp-N3B:	1.6 Å	1.2 Å	

† His37 deprotonated
‡ His37 protonated

cavities, tunnels and surface clefts. To assess dxTuber performance we carried out test runs on six example proteins and compared our findings to results obtained with SURFNET, CAVER and PyMOL.

6.4.1 Isolated Internal Cavities

Specifically designed to find a path out of the protein interior CAVER does not support this type of cavities and has therefore been excluded from this section. PyMOL does not provide any quantitative volume information on detected cavities. The molybdenum storage protein (MoSto) (Schemberg et al., 2007) features an intermolecular cavity, large enough to accommodate up to 100 molybdenum (Mo) or tungsten (W) atoms (Schemberg et al., 2007).

Whereas SURFNET, dxTuber and PyMOL were all able to detect the cavity, there are differences in cavity shape and volume. In SURFNET the cavity is represented having a smooth shape and a volume of 10720 $Å^3$ (figure 6.4B), whereas in

Table 6.5: PyMOL search parameters

Protein	Detection radius [Å]	Detection cutoff radius [Å]
Pyp	7	3
MoSto	6	4
TolC	20	7
M2 †	7	4
M2 ‡	8	4
BtuF	20	4
H-Ras	20	3

† His37 deprotonated
‡ His37 protonated

dxTuber a volume of 9223 $Å^3$ is calculated and the cavity has an overall less regular shape revealing more structural detail (figure 6.4A). PyMOL gives a detailed view of the storage cavity but also finds false positives on the protein surface which are partly connected to the cavity. In this example restrained protein dynamics were used for the dxTuber analysis, yielding a Cα-RMSD of 0.3 Å in the simulation end conformation.

Photoactive yellow protein (PYP) is a 125 residue signaling bluelight photo receptor (Vreede et al., 2010) whose covalently bound chromophore (para-coumaric acid, pCA) was removed to create a cavity representative for small intra-molecular cavities. For dxTuber analysis both restrained (figure 6.4D) and unbiased protein dynamics were investigated (figure 6.4G-I) yielding Cα-RMSDs of 3.5 Å and 0.3 Å in the end conformations. For the unbiased simulation, cavity volume was monitored throughout the trajectory using time windows of 1 ns (figure 6.4I).

dxTuber (figure 4D), SURFNET (figure 6.4E), and PyMOL (figure 6.4F) find the cavity, and whereas PyMOL additionally detects five independent cavities with the main cavity of interest being similar to SURFNET. In dxTuber the cavity is smaller (50 $Å^3$ volume versus 152 $Å^3$ as reported by SURFNET) and also more extended by 2 Å towards the center of the protein. The innermost section of the dxTuber cavity coincides with a secondary cavity detected in PyMOL that is absent in SURFNET. Without position restraints on the protein, the cavity fluctuates in size and volume (figure 6.4G and H) and disappears after 6 ns (figure 6.4I).

6.4.2 Tunnel-like Cavities

In this section the calculation of cross-sectional profiles along the membrane normal was restricted to CAVER and dxTuber, as SURFNET and PyMOL do not include this feature. The outer membrane protein TolC functions as a major efflux duct in the outer membrane of Escherichia coli and at a length of 140 Å and 20 Å in diameter, the TolC channel is large enough to accommodate and transport entire proteins (Federici et al., 2005).

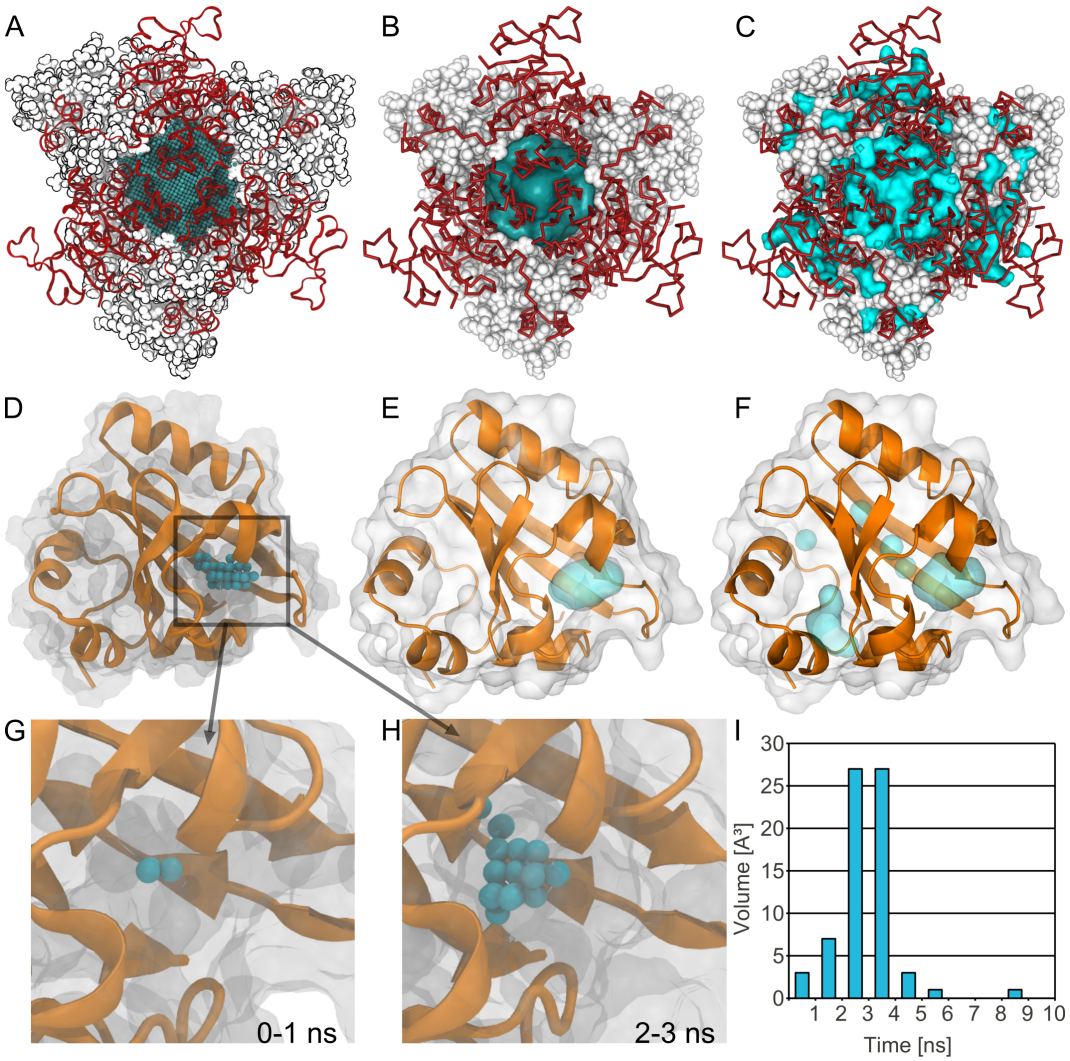

Figure 6.4: Isolated internal cavities with no connection to the protein exterior are represented by the molybdenum/tungsten storage protein MoSto (A-C) and photoactive yellow protein PYP (D-I). Whereas MoSto serves an example for a large intermolecular cavity, PYP after chromophore removal represents small intramolecular cavities. Each protein's main cavity is depicted as detected by dxTuber with (A and D) and without position restraints on the protein atoms (G and H), SURFNET (B and E) and PyMOL (C and F). As cavity separation is not implemented in PyMOL, (C) and (F) also show secondary cavities and surface clefts. When PYP dynamics are unrestrained, the main cavity fluctuates in size and volume and disappears after 6 ns (G-I).

dxTuber analysis was performed using both restrained (figure 6.5A) and unrestrained protein dynamics (figure 6.5B) yielding end Cα-RMSDs of 0.3 Å and 2.5 Å respectively. While all four applications detect the main cavity, the results differ in size and shape (figure 6.5A-E). In SURFNET the channel ends 10 Å before TolC's

lower (periplasmic) exit (figure 6.5C), whereas in dxTuber, the cavity is detected completely only when restrained protein dynamics are applied (figure 6.5A). When unrestrained protein dynamics are used the channel cavity ends 15 Å before the upper (extracellular) exit (figure 6.5A).

In CAVER the channel is narrowest of all five cases (figure 6.5D) and requires two tunnels set in the CAVER input parameters to be detected completely. If only a single tunnel is used, a similar effect as with SURFNET is seen with the lower 10 Å of the cavity missing. PyMOL (figure 6.5E) describes the main cavity in a similar way as SURNET, however the tunnel is shorter in the upper (extracellular) exit region but longer towards the lower (periplasmic) exit (approximately half way between the SURFNET and CAVER findings). Additionally PyMOL generates false positives on the protein surface at the TolC equatorial region. Compared to CAVER (figure 6.5G) the channel's profile of cross-sectional area is larger in dxTuber (figure 6.5F), exhibiting a continuous progression with a maximum peak at 85 Å. The CAVER profile is discontinuous displaying abrupt jumps and a maximum peak at 100 Å. While the dxTuber profiles are similar to each other, the ones obtained with restrained protein dynamics are larger than those derived from unrestrained MD (figure 6.5F).

With a trans-membrane channel 30 Å long and only a few water molecules wide, the pH-gated Influenza proton channel M2 (Schnell and Chou, 2008) serves as an example for a narrow tunnel-like cavity. M2 switches between two conformations where extracellular and intracellular water phases are either connected (protonated His37, open state) or separated (deprotonated His37, closed state) (figure 6.6). Whereas closed state M2 was captured in the NMR structure, a transition to the open state could be observed within 10 ns MD simulation when His37 was considered in protonated form. No position restraints were applied throughout the simulations. The M2 end conformations exhibit a total Cα-RMSD of 5.3 Å over all four monomers when His37 was protonated, and 4.4 Å when His37 was deprotonated.

All four applications identify separated cavities for closed state M2 - 2 in dxTuber (figure 6.6A and B), SURFNET (figure 6.6C) and CAVER (figure 6.6D) and 4 in PyMOL (figure 6.6E) - and a continuous tunnel cavity for open state M2 (figure 6.6F-J). Results differ in (a) channel narrowness ranging from dxTuber using average protein density (figure 6.6A and F) and CAVER (figure 6.6D and I) up to dxTuber run on minimum protein density (figure 6.6B and G) - and (b) the amount of false positives, i.e. cavity sections extruding to the protein exterior. This effect is most pronounced in SURFNET (figure 6.6C and H) and, to a lesser degree, in CAVER (figure 6.6D and J) which required seven tunnels for the closed state and eleven tunnels for open state M2 to find the cavities of interest. Apart from the exit regions the tunnel profiles are similar in CAVER (figure 6.6M) and dxTuber run on average protein density (figure 6.6K). When dxTuber is run on protein minimum density, the channel is wider, the difference between open and closed state is more pronounced and the bottle neck region in closed state M2 is 5 Å shorter (figure

6.6L).

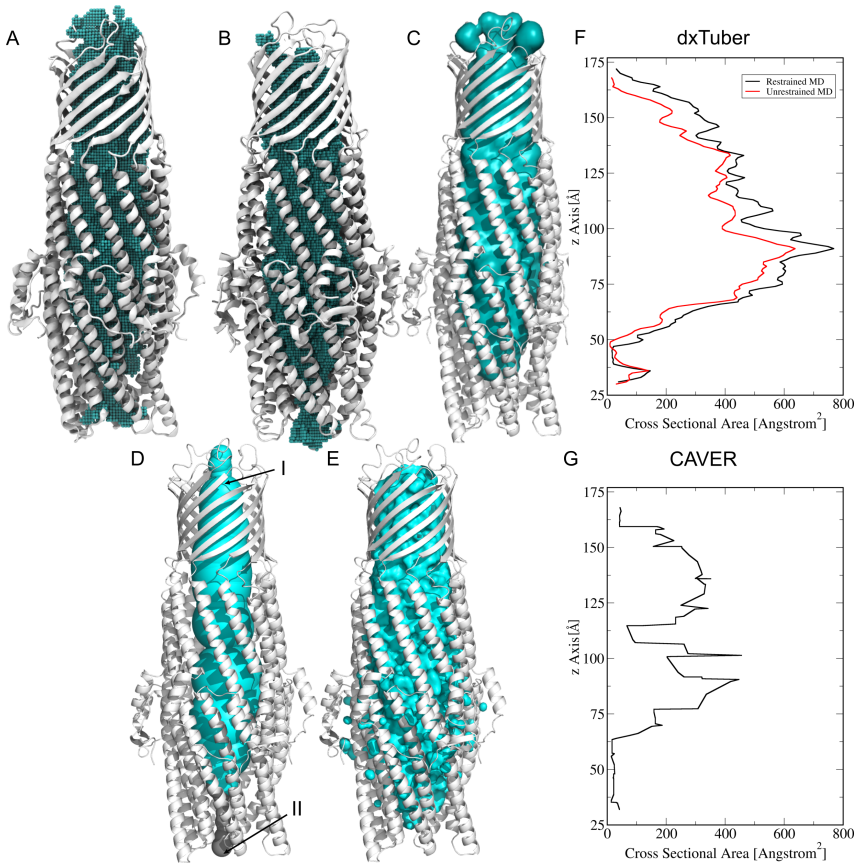

Figure 6.5: The bacterial efflux duct TolC is an example of a large tunnel-like cavity whose main channel is shown as depicted by dxTuber with (A and F, black tunnel profile) and without position restraints on the protein atoms (B and F, red tunnel profile), SURFNET (C), CAVER (D and G) and PyMOL (E). In CAVER two exit tunnels had to be calculated to represent the TolC channel completely. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6.4.3 Surface Clefts

In the periplasmic vitamin B12 binding protein BtuF (Karpowich et al., 2003) substrate binding occurs in a wedge-shaped, un-occluded binding cleft that is approximately 20 Å wide and 15 Å deep when the protein is in its closed state (Kandt et al., 2006). The cleft is detected by dxTuber (figure 6.7A), SURFNET (figure 6.7B) and partially by CAVER (figure 6.7C). PyMOL only detects the borders of the binding pocket; while additionally producing false positives lining the main binding cleft (figure 6.7D).

Although dxTuber produces the most detailed rendering of the binding cleft, it

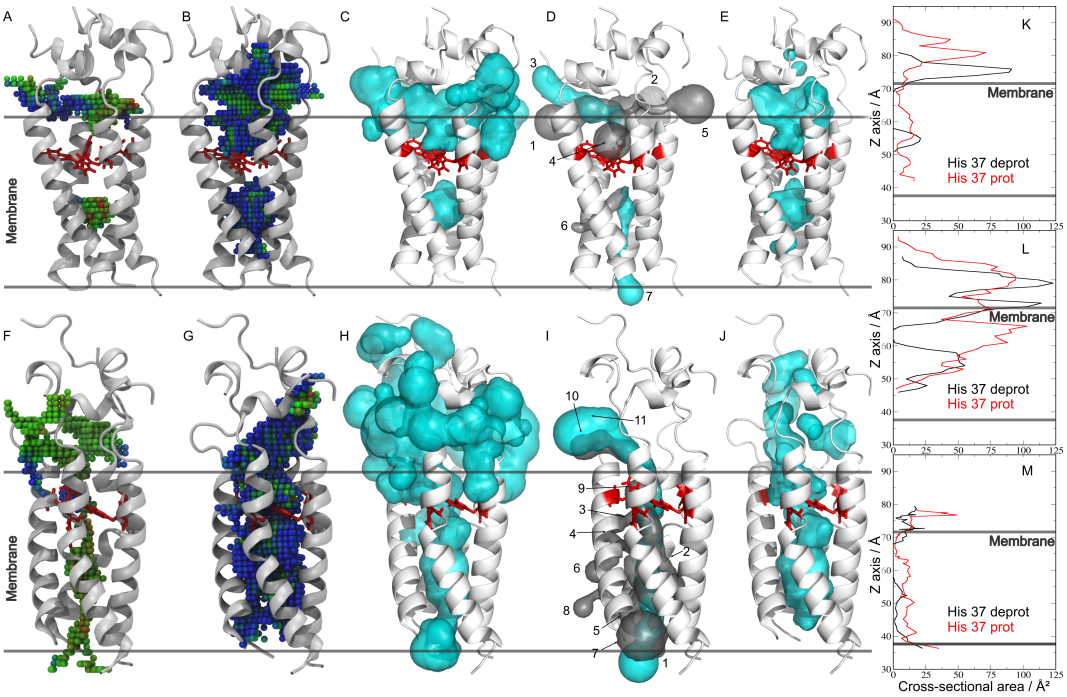

Figure 6.6: The pH-gated Influenza M2 proton channel serves as an example of a narrow tunnel-like cavity. Controlled by His37, the M2 tunnel either occurs as two separate half channels when His37 is deprotonated (A-E) or as single channel when His37 is protonated (F-J). The M2 channel is depicted as detected by dxTuber showing average (A, F and K) and maximum cavity extension (B, G and L) colored by solvent density (blue = low density, red = high density), SURFNET (C and H), CAVER (D, I, M) and PyMOL (E,J). Whereas closed state M2 required the calculation of seven CAVER exit tunnels (D), eleven exit tunnels were necessary for the open state (I). CAVER tunnels used for the profiles of cross-sectional area appear cyan.

incorrectly includes adjacent grooves in the surface of the N-terminal and C-terminal domains. In SURFNET the cleft appears hose-like protruding by up to 9 Å on both sides of the binding cleft whereas in CAVER, only a fraction of the cleft is detected calculating 3 exit pathways. Restrained protein dynamics were employed during the dxTuber MD leading to an end Cα-RMSD of 0.4 A. The partially occluded GTP binding niche of the activated signal transduction protein H-Ras P21 (Buhrman et al., 2010) concludes our set of example applications.

All four tools detect the entire binding cleft (figure 6.7E-H), with SURFNET (figure 6.7F) and CAVER (figure 6.7G) yielding almost identical results of hose-shaped cavities whereas dxTuber and PyMOL show more structural detail of the binding cleft (figure 6.7E and H). Three tunnels were computed in the CAVER analysis and restrained protein dynamics were employed for dxTuber resulting in an end Cα-RMSD of 0.3 Å.

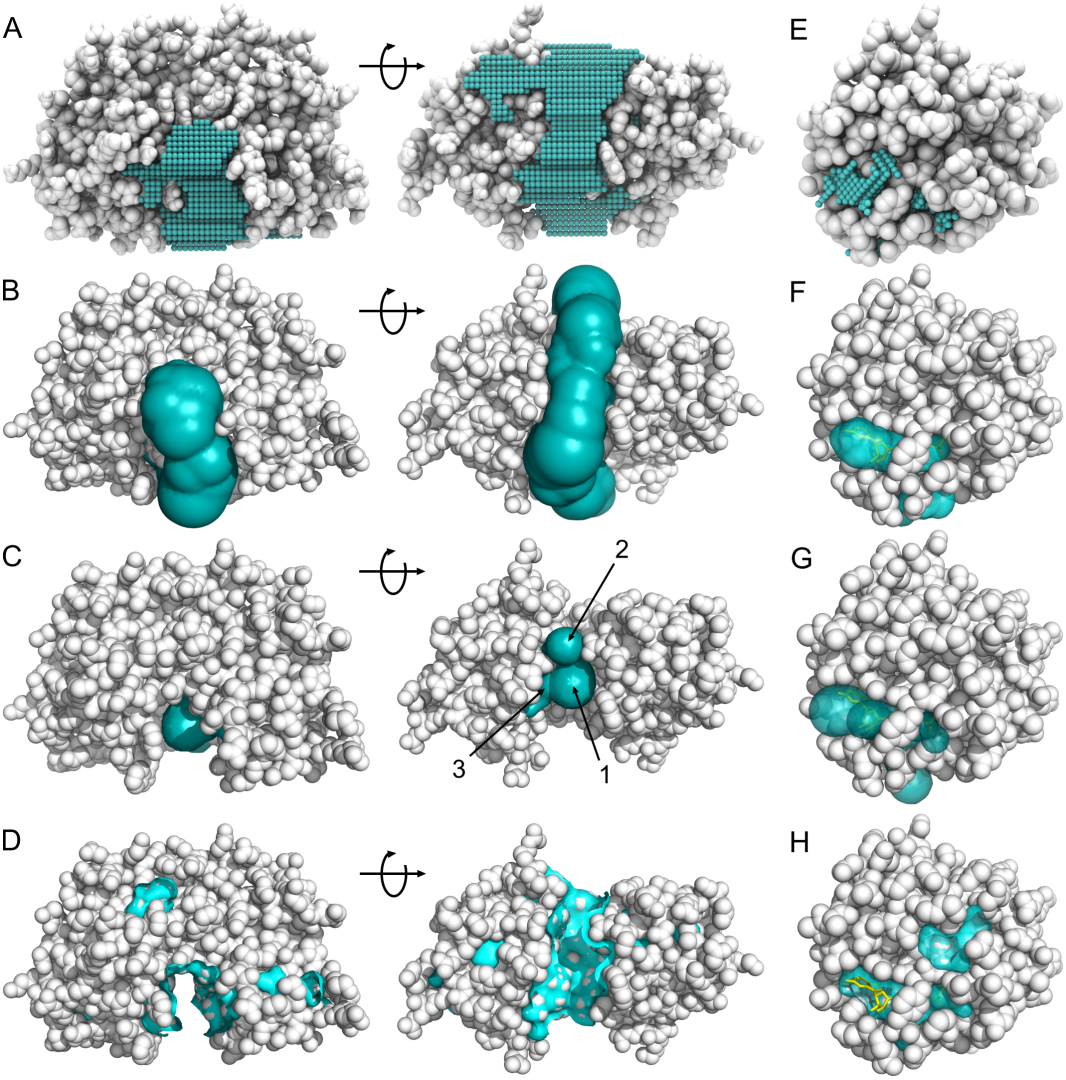

Figure 6.7: The large unoccluded vitamin B12 binding cleft in BtuF (A-D) and the small partially occluded GTP binding niche in H-Ras p21 (E-H) serve as examples for surface clefts. Clefts are shown as detected by dxTuber (A and E), SURFNET (B and F), CAVER (C and G) and PyMOL (D and H). Three CAVER tunnels were calculated for BtuF (C) and four for H-Ras p21 (H).

6.4.4 dxTuber Calculation Speed

Representing the smallest and the largest of our test proteins, the PYP and TolC analyses were additionally carried out at 0.5 and 2.0 $Å^3$ voxel size to assess how voxel resolution influences the calculation speed. At 0.5 $Å^3$ the PYP analysis required 3:53 min for trajectory conversion in VMD and 5:42 min for cavity detection in dxTuber. At 1.0 $Å^3$ resolution the analysis took 0:29 (VMD) plus 0:19 min (dxTuber) and 0:08 plus 0.02 min at 2.0 $Å^3$ voxel size. For TolC, the analysis was complete after 24:54 min in VMD plus 130:05 min in dxTuber when a voxel size of 0.5 $Å^3$ was used.

At 1.0 $Å^3$resolution the analysis required 3:03 min plus 6:17 min and 0:35 min plus 0:18 when a voxel size of 2.0 $Å^3$ was selected.

6.5 Discussion

Proteins are in a continuous state of motion that can easily exceed mere thermal fluctuations. To take into account the element of conformational dynamics which has so far been neglected by other cavity detection tools, we have developed a novel method employing protein and solvent residence probabilities obtained from molecular dynamics simulations to detect and analyze cavities, tunnels and surface clefts. After comparing dxTuber's performance to SURFNET, CAVER and PyMOL we proceed discussing our approach's methodological characteristics in context of the other three methods. We conclude proposing general proceeding recommendations for cavity detection.

6.5.1 Performance and Methodological Characteristics

Each of the three main classes of protein cavities presents particular challenges to cavity detection and the six example proteins we used as representatives might raise the question whether the number of test cases is sufficient. While it is naturally desirable to have as many test cases as possible, the six proteins used in this study are nevertheless likely to give a reasonable representation of the palette of protein cavities. Given that each class is exemplified by a small and a large variant, representing extreme sizes of each cavity type, it is reasonable to assume a tool's performance in these extreme cases is likely transferable to the range of intermediate cases in between. Future applications to other proteins of different sizes will show whether if this assumption is valid.

Whereas isolated cavities with no contact to the protein exterior are detected equally well by SURFNET, PyMOL and dxTuber (figure 6.4), the CAVER algorithm does not support this type of cavity but performs expectedly better with tunnel-like cavities as long as multiple pathways are calculated (figures 6.5 and 6.6). PyMOL and dxTuber also yield comparably good results here, whereas SURFNET did not detect the entire tunnel in case of TolC (figure 6.6) or produced a large amount of false positives on the protein surface in case of M2 (figure 6.7). Whereas H-Ras p21's small and partially occluded GTP binding cleft was detected by all four applications (figure 6.7E-H), the large, un-occluded vitamin B12 binding cleft in BtuF (figure 6.7A-D) was only detected by SURFNET and dxTuber although in both cases the detected cavity exceeds the actual binding cleft.

SURFNET

Each strategy to cavity detection bears its own limitations. Using gap spheres (Laskowski, 1995) placed on a 3D grid, SURFNET results depend on the orientation

of the protein of interest. Common to all grid-based algorithms, this limitation should be kept in mind particularly when comparing different conformations of the same protein - such as open and closed state M2 (figure 6.7). To minimize this artifact, structures should be aligned to the same reference conformation prior to cavity analysis. When larger cavities like the storage compartment in MoSto (figure 6.4) are of interest, SURFNET's hard-coded limit of 120,000 initial gap spheres must be compensated at the cost of a lower spatial resolution by making the search grid coarser. In case of MoSto we additionally increased the minimum gap sphere radius to ensure the entire cavity is detected completely. The overall smoothness of cavity shape, the general lack of fine structural detail as well as the protruding BtuF ligand binding cleft (figure 6.7B) are another consequence of SURFNET's gap sphere technique that is also seen with other probe-sphere employing methods like CAVER or PyMOL.

Although the SURFNET standard settings are a good starting point, grid size, minimum and maximum gap sphere radius should be optimized for each protein to achieve the best results (table 6.3). Often this requires a trade-off between completeness of cavity detection and the amount of false positives. For example, although a complete detection of the entire TolC channel (figure 6.5C) is possible using a smaller maximum gap sphere radius, the channel cavity then merges with the clefts on the outer protein surface into one inseparable entity under these conditions. On the other hand, for the narrow channel in M2 (figure 6.6C and H) a larger maximum gap sphere radius eliminates not only the false positives on the outer surface but also the tunnel cavity itself. Any SURFNET analysis is restricted to the 50 largest cavities and although initially handled as one object, individual cavities can be manually separated into individual isosurfaces using the MASK script of the SURFNET package. Although SURFNET calculates cavity volume, pore profiles cannot be computed.

CAVER

Specifically designed to find the shortest way out of the protein interior along a 3D grid from a user-defined starting point, the CAVER algorithm (Petrek et al., 2006) is not applicable to class I cavities (figure 6.4). Whereas the large, un-occluded surface cleft in BtuF was only detected partially (figure 6.7C), the smaller and partially occluded GTP binding niche in H-Ras p21 was rendered correctly (figure 6.7G). However, like with the TolC (figure 6.5D) and M2 channels (figure 6.6D and I), this requires calculating several exit tunnels and/or using multiple starting points (table 6.4). This is a consequence of (a) the 3D grid constructed from the user-defined starting point and (b) the cost function CAVER employs to calculate the exit pathway. Aimed at maximizing the tunnel radius as quickly as possible, the cost function favors an increase of the tunnel radius.

In each CAVER run several exit tunnels are calculated and the actual output

tunnel is selected based on the amount of favored steps of radius increase and disfavored steps of radius decrease. As seen with TolC and M2, this is problematic with bottleneck regions. One way to sample and output this disfavored regions is by placing the starting point into the bottleneck and always calculate more than one exit tunnel. While two tunnels were sufficient for TolC (figure 6.5D), eleven output tunnels had to be calculated for open state M2 (figure 6.6I). CAVER tunnel profiles approximate cross-sectional area by the radius of a tunnel-filling probe sphere. Common to all tools usingthat approach, this leads to inaccuracies with tunnel cavities whose cross-sectional profile is not circular (Maeda and Kinoshita, 2009). Furthermore, in case of TolC we observed jumps in the CAVER pore profile (figure 6.5G) suggesting the presence of a bottleneck when there is none. For reasons we do not understand yet, the CAVER path here temporarily deviates from the tunnel center and approaches the channel wall where only the placement of small probe spheres is sterically possible.

PyMOL

Cavity detection is a recent addition to the PyMOL molecular viewer (Delano, 2002) and is based on the MSMS molecular surface algorithm (Sanner et al., 1995, 1996). Cavities cannot be further analyzed nor separated into individually selectable objects, and the large BtuF surface cleft was not detected (figure 6.7D). Nevertheless, fast and robust results were yielded for all other example proteins (figures 6.4C and F, 6.5E, and 6.6E and J) when the search parameters cavity detection radius and cavity detection cut-off were manually optimized for each protein (table 6.5).

dxTuber

dxTuber is the first cavity detection tool that fully takes into account protein flexibility. This is done on the grounds of protein and solvent residence probability distributions derived from molecular dynamics simulations using solvent molecules to probe for cavities. Achieving a more realistic and accurate representation of protein and cavity structure dxTuber allows to study cavities within the framework of the dynamic protein.

Whereas statistics data on cavities can be difficult to obtain in other methods, the dxTuber approach offers an easy way of doing that. For each cavity its average and minimum or maximum extension can be calculated (figures 6.2A and 6.6A, B, F, and G) based on (a) how protein and solvent trajectories were converted into residence probability distributions or (b) which residence probability cut-off is used for cavity display using the averaged and mass-weighed solvent density per voxel encoded as formal B-factor in the exported PDB.

When multiple trajectory time windows are chosen for analysis (figure 6.4G-I) cavity dynamics can also be computed over time, monitoring profiles of cross-sectional area or cavity volume as in figure 6.4I. As the choice which molecules are

used for cavity probing is user-defined, dxTuber offers a high degree of application flexibility. For example, whereas assigning a prevalent solvent like water allows a more general cavity search, less widespread molecules such as bound ligands can be used for high-precision mapping of substrate binding sites or to determine preferred ligand-protein interaction sites.

Recent MD studies where the dxTuber approach could be useful include the mapping of protein-internal water distributions (Kandt et al., 2005, 2004; Yin et al., 2010), monitoring the opening state of transport proteins (Kandt and Tieleman, 2010; Sonne et al., 2007), investigating protein-internal gas diffusion pathways (Baron et al., 2009; Wang et al., 2010) or exploring ligand binding sites (Li et al., 2011). Unlike many other cavity detection tools specialized in one or two cavity classes, dxTuber is capable of reliably detecting all three main types of cavities where - apart from initial choices of protein orientation, probing solvent and scanning algorithm - no fine tuning of search parameters is required. Controlled by a graphical user interface, dxTuber also offers a high degree of accessibility and usability.

A common problem in cavity detection is the occurrence of false positives - i.e. additional cavities and surface grooves not in the current focus of analysis - and the necessity to separate these from the actual cavities of interest for later visualization or analysis. As detected cavities can be exported in PDB format where cavities are stored as individually selectable groups of voxels sharing the same atom name, dxTuber offers an easy and efficient way to identify and extract cavities of interest. Relying on molecular dynamics simulation trajectories to probe protein cavities, dxTuber's major limitation is the amount of effort necessary to perform a cavity analysis. Next to simulation length and system size, the total time required for a dxTuber analysis is determined by the voxel resolution used in trajectory conversion and cavity detection.

Whereas for a small system like PYP analysis times range from seconds to minutes using voxel resolution of 2.0 $Å^3$, 1.0 $Å^3$ and 0.5 $Å^3$, a large system like TolC requires seconds to hours depending on voxel sizes applied. While a coarser resolution might be beneficial for a quick assessment of large cavity structures, in our test runs we did not find the increased resolution of 0.5 $Å^3$ leading to considerable improvements over 1.0 $Å^3$ resolution which we recommend as standard setting. Another limitation is that per definition only those cavities will be detected that actually contain solvent molecules designated for cavity probing.

Whereas using water as the only solvent in our test runs was unproblematic, additional solvent types might become necessary to probe for, for example, hydrophobic compartments in a protein, micelle or a lipid bilayer. Pore profiles in dxTuber are not computed via the common tunnel-filling probe sphere approach used in (Pellegrini-Calace et al., 2009; Petrek et al., 2007, 2006; Smart et al., 1996). Instead the cross-sectional area is calculated as an area sum of protein-internal solvent voxels sharing the same Z coordinate.

While this leads to a higher accuracy and detail in tunnel profiles (figures 6.5F

and 6.6K and L) independent of a channel's cross-sectional shape, the approach is restricted to approximately linear tunnels running parallel to a principal axis. A potential way to overcome this limitation could be employing a slicing plane through the cavity voxels that is aligned perpendicular to the local tunnel normal. On the slicing plane local area, diameter or circumference of the tunnel could then be measured. Furthermore the exit regions of channels or surface clefts should be interpreted carefully as they might be represented incompletely due to the grid approach's dependence on the initial orientation of protein and cavity. When such cavities are of interest (figure 6.7A) and need to be represented completely, we recommend orienting the protein in such a way that the cavity is aligned to the X, Y or Z axes.

Although dxTuber generated a complete rendition of the BtuF substrate binding cleft that way (figure 6.7A), smaller adjacent surface clefts were also detected that could not be separated or eliminated automatically. This is a limitation of the employed 1D scanning algorithm which is currently the only way to represent large surface clefts completely. Since all cavities can be exported in PDB format, we suggest selecting and removing such artifact voxels manually. On the other hand, the different lengths of TolC tunnel profiles derived from restrained (figure 6.5F, black profile) or unrestrained protein dynamics (figure 6.5F, red profile) are not a limitation of the algorithm but instead result from a partial closure of the protein in the unrestrained MD simulation.

Similarly, the fluctuating and eventually disappearing cavity dxTuber detects in unrestrained PYP (figure 6.4G-I) reflects an adaption of the protein to deleted chromophore and an expulsion of initially cavity-internal water molecules. On the other hand, when PYP is position-restrained (figure 6.4D) the cavity is stable but smaller than in SURFNET (figure 6.4E) or PyMOL (figure 6.4F).

When protein position restraints are applied during the MD, one should keep in mind that force constants of 1000 $\frac{kJ}{mol\ nm^2}$ still allow modest protein fluctuation - as reflected by Cα-RMSDs between 0.3 and 0.4 Å that can limit water diffusion as seen with the MoSto or PYP dxTuber cavity volumes (figure 6.4A and D) which are 1497 $Å^3$ and 102 $Å^3$ smaller than their SURFNET counterparts (figure 6.4B and E). To completely eliminate any form of protein motion we recommend a force constant of 100,000 $\frac{kJ}{mol\ nm^2}$.

6.5.2 General Proceeding Recommendations

For a first impression of a protein with unfamiliar cavity structure, we found PyMOL a good starting point, yielding fast and robust results for the majority of cavity classes while requiring little effort in optimizing cavity search parameters. For a more in-depth analysis of a protein with familiar cavity structure there are two scenarios to be considered.

When molecular dynamics simulation data are not available, SURFNET gives

robust results for isolated internal cavities or surface clefts whereas CAVER is a reasonable choice for tunnel-like cavities. In both cases each method's individual limitations should be considered as discussed above. It is also beneficial to perform each analysis in combination with an initial PyMOL search. That way useful information is gained regarding (a) cavity types and choice of search algorithm, (b) cavity orientation and shape which determine the placement of starting points, grid size as well as minimum and maximum gap sphere radius and (c) validation of results and search parameters.

When molecular dynamics simulation data are available and conformational dynamics are to be taken into account, dxTuber is currently the only cavity detection tool available yielding excellent results for all cavity classes bearing in mind the specific limitations discussed above. Again, a prior PyMOL analysis can provide useful information regarding the choice of search algorithm or protein and cavity orientation, but is also useful to inspect a cavity's hydration state prior to MD simulation, and as a means for general cross-checking.

To explore cavity structure near the X-ray conformation, protein atoms should be positionrestrained (figures 6.4A and D, 6.5A, and 6.5A and E) whereas unbiased MD is recommended to investigate protein and cavity dynamics (figures 6.4G-I, 6.5B, and 6.6A and F). Depending on whether a cavity's average, minimum or maximum extension is of interest, trajectory to voxel conversion in VMD should be carried out using the respective setting for the calculation of protein and water densities.

6.6 Conclusions

We introduce a novel cavity detection method that fully takes into account cavity dynamics using protein and solvent residence probabilities derived from molecular dynamics simulations. Once MD trajectories have been converted to a voxel representation of mass-weighed spatial density maps using VMD, dxTuber identifies cavities as groups of coherent voxels representing protein-internal regions of high solvent density. Three different search algorithms have been implemented that yielded excellent results on six test proteins representing the three main classes of protein cavities: isolated internal cavities, tunnels and surface clefts.

dxTuber results can be exported in PDB format where each cavity is an individually selectable object of coherent voxels written as pseudo atoms, each holding the residence probability information in form of averaged and mass-weighed solvent density encoded as formal B-factors. Cavities can be (a) readily visualized and processed using common molecular viewers and (b) further analyzed in terms of volume and profiles of cross-sectional area along a principal axis. dxTuber is available free of charge for academic use at `http://www.csb.bit.uni-bonn.de/downloads.html` and at `http://code.google.com/p/dxtuber/`.

Chapter 7

Summary

7.1 Bacterial multidrug resistance

Antibiotics represent one of the main therapeutic tools both in human and veterinary medicine to cure bacterial infectious diseases. Over the past decades bacterial resistance against these chemical compounds arose due to the intensive treatment and today bacterial multidrug resistance is a serious issue in medical treatment. Multidrug efflux systems expel a wide range of harming compounds out of the bacterium and through theire over expression and theire ubiquitous spread along the bacterial membranes they play a major role in multidrug resistance of bacteria (Pos, 2009).

The efflux ducts (ED) TolC and OprM interact with several inner membrane transporters and facilitate drug export (Köhler et al., 1997; Masuda et al., 2000a,b; Moussatova et al., 2008; Nikaido, 2011; Poole et al., 1993; Saier and Paulsen, 2001). A blocked or disabled ED could inhibit substrate efflux across the outer membrane and therefor disturb or disable efflux systems. Without the drug efflux ability, bacteria would loose resistance to drugs. To target EDs with drugs the knowledge of their gating functionality is a prerequisite. In this thesis for both proteins unbiased MD simulations were computed in a POPE / NaCl / water environment with special focus on their gating mechanisms.

7.1.1 TolC

In both available wild-type crystal structures TolC access is limited by three loops on extracellular side and by twelve coiled-coil helices on periplasmic side (Higgins et al., 2004; Koronakis et al., 2000). On the latter side Andersen et al. (2002b) identified a double aspartate ring formed by Asp371 and Asp374 (bottleneck I, BNI) restricting substrate passage and found a network of hydrogen bonds and salt bridges between Tyr362–Asp153 stabilizing a second bottleneck (BNII) (Andersen et al., 2002a). In mutant "partially open" crystal structures introduced by Bavro et al. (2008); Pei et al. (2011) BNII was monitored by distance measurements of Gly365.

Simulations in Absence of the AcrB-DD

Five independent unbiased MD simulations were computed to gain insights into TolC gating. In all simulations the extracellular loops opened and closed freely hinting at the absence of a gating functionality on this side and the passage of substrates is only limited by the diameter of the transmembrane beta barrel on this side (figure 3.2). In one of the five simulations a closure of BNII was induced by the successive binding of two sodium ions as shown by heightened sodium densities in figure 3.4. BNII kept close for the remaining 180 ns of the simulation time, even if sodium ions near the hotspots were removed from the simulation system. However, the removal of all ions induced an reopening of BNII and BNI pointing to a sodium dependent lock on TolC's periplasmic side. By attacking this lock with antibiotics bacteria could loose resistances to drugs, which were previously exported via TolC.

Simulations in Presence of the AcrB-DD

To explore the effects of the AcrB docking domain (AcrB-DD) onto TolC's gating behavior the AcrB-DD has been added into the previously introduced TolC simulation system.

In the 1.05 μs simulation of TolC and AcrB-DD we observed an spontaneous docking of the AcrB-DD on TolC. At the end of the simulation we measured a tighter but asymmetric binding of the AcrB-DD on TolC. TolC preferred an open state of the extracellular loops hinting again for the absence of a gating functionality on this side. While on periplasmic side BNII opened up within the first 3 ns of simulation time, this open state was stabilized by the AcrB-DD indicating that BNII should be open before or at least during the assembly of TolC and AcrB. During the simulation BNI remained closed meaning that either AcrB-DD need more time to open BNI or other parts of the efflux system and / or ligands are needed to open TolC's BNI.

7.1.2 OprM

On periplasmic side of OprM a single aspartate ring composed of Asp416 (bottleneck I, BNI) might restrict substrate passage and a salt bridge between Tyr406 and Asp205 analogue to TolC's Tyr362–Asp153 salt bridge could stabilize OprM's closed conformation on this side (Phan et al., 2010).

Like TolC, OprM opened and closed freely on extracellular side indicating also the absence of an gating mechanism on extracellular side for OprM. Beside the opening of BNII also BNI opened up slightly, which is 1.1 to 1.4 more open by the end of the simulations in comparison to crystal structures (table 5.1, (Akama et al., 2004; Phan et al., 2010)). Both bottlenecks show low sodium densities in all simulations indicating the absence of an sodium dependant lock. OprM's periplasmic access is only limited by the inner aspartate ring formed by Asp416. However heightened sodium densities identified one sodium pocket for each monomer between the

equatorial domain and the inner bottleneck BNI managed by Asp171 and Asp230, respectively.

7.2 Cavity Detection

Biomolecular cavities can indicate functional regions like storage compartments, active sides, ligand binding- and ion-pockets inside and onside molecular structures. During the past decades lots of tools have been invented to discover these voids (see table 1.1). All of these are based on static structures and therefore only display cavities for single conformations.

With the release of dxTuber (chapter 6) the first cavity detection based on ensembles of protein conformations was introduced by Raunest and Kandt (2011a). Based on the assumption, that protein internal solvent molecules highlight cavities, the separation of internal solvent molecules from external ones can represent the distribution of cavities inside and onside molecular structures. For this purpose protein conformation and solvent distributions taken from MD trajectories are converted into gridded density maps by VolMap, a plugin from VMD (Humphrey et al., 1996).

Based on these grid or voxel representations of protein and solvent dxTuber separates internal solvent voxel (ISV) and external solvent voxel (ESV) as the first step and groups adjacent ISVs into cavities as the second step. Cavities can be stored in the well known PDB file format, where cavities are separated via the *atom name* field and solvent densities are stored as formal B-factors. After cavity detection dxTuber can calculate individual cavity volumes and the cross-sectional areas of single cavities along a principle axis and further more cavities can be easily post processed due to the well known PDB file format. Users can use either a graphical user interface or a command line version of dxTuber for cavity detection. Both are licensed under GPL v2[1] and can be downloaded at `http://www.csb.bit.uni-bonn.de/dxtuber.html` or `http://code.google.com/p/dxtuber/`.

[1] `http://www.gnu.org/licenses/licenses.html#GPL`

Publications

Peer-reviewed articles

- Fischer, N., Raunest, M., Schmidt, T., Koch, D., and Kandt, C. (2013). Efflux pump-mediated antibiotics resistance: Insights from computational structural biology. *Interdisciplinary Science: Computational life Sciences, CECAM proceedings special issue*, submitted

- Koch, D. C., Raunest, M., Harder, T., and Kandt, C. (2013). Unilateral access regulation: Ground state dynamics of the pseudomonas aeruginosa outer membrane efflux duct oprm. *Biochemistry*, 52(1):178–187

- Raunest, M. and Kandt, C. (2012b). Locked on one side only: ground state dynamics of the outer membrane efflux duct tolc. *Biochemistry*, 51(8):1719–1729

- Raunest, M. and Kandt, C. (2011a). dxtuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J Mol Graph Model*, 29(7):895–905

Invited talks

- Raunest, M. and Kandt, C. (2012c). Locked on one side only: Ground state dynamics of the outer membrane efflux duct tolc. In *Dipartimento di Fisica, Università di Cagliari*

Peer-reviewed abstracts

- Raunest, M. and Kandt, C. (2012a). Locked on one side only: Computer simulations of the outer membrane efflux duct tolc in absence and presence of the acrb-docking-domain. *Biophysical Journal*, 102:622a – 623a

- Fischer, N., Raunest, M., and Kandt, C. (2011a). Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase acrb. *Biophysical Journal*, 100(3, Supplement 1):155a –

- Raunest, M. and Kandt, C. (2011b). Dxtuber: Detecting protein cavities, tunnels, surface clefts based on protein, solvent dynamics. *Biophysical Journal*, 100(3, Supplement 1):48a –
- Raunest, M., Fischer, N., and Kandt, C. (2011). Computer simulation of tolc ground state dynamics and spontaneous binding of the acrb docking domain. *Biophysical Journal*, 100(3, Supplement 1):132a –
- Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2011a). Atomistic and coarse-grain computer simulations of the assembled acrab-tolc multidrug efflux pump. *Biophysical Journal*, 100(3, Supplement 1):146a –
- Fischer, N., Raunest, M., and Kandt, C. (2010a). Mapping the proton conduction pathways in the inner membrane multi-drug translocase acrb. *Biophysical Journal*, 98(3, Supplement 1):633a –
- Raunest, M., Nadine, F., and Kandt, C. (2010b). Molecular dynamics simulations reveal tolc flexibility in the acrb interface region. *Biophysical Journal*, 98(3, Supplement 1):374a –

Posters

- Fischer, N., Raunest, M., Schmidt, T. H., and Kandt, C. (2011b). Simulations of the multidrug efflux pump acrab-tolc; gordon research conference: Multidrug efflux systems. In *Multi-drug Efflux Systems; Les Diablerets, Switzerland; 12.-17.Aug*
- Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2011b). Atomistic and coarse-grain computer simulations of the assembled acrab-tolc multidrug efflux pump. In *Coarse-Grained Biomolecular Modeling Workshop Levi, Finland, March 7-12*
- Fischer, N., Raunest, M., and Kandt, C. (2010b). Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase acrb. In *BioScience 2010, Workshop on expanding the frontiers of biomolecular science, Forschungszentrum Jülich, Germany, November 15 - 17*
- Raunest, M. and Kandt, C. (2010). dxtuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. In *BioScience 2010, Workshop on expanding the frontiers of biomolecular science, Forschungszentrum Jülich, Germany, November 15 - 17*
- Raunest, M., Fischer, N., and Kandt, C. (2010a). Dynamics in the tolc gating region and cavity detection based on protein and water dynamics. In *462nd WE Heraeus Seminar,Transport across membranes: Multiple drug resistance, mechanisms and new tools, Jacobs University Bremen, 4. - 10. July*

- Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2010). Computer simulation of the assembled acrab-tolc multidrug efflux pump. In *462nd WE Heraeus Seminar,Transport across membranes: Multiple drug resistance, mechanisms and new tools Jacobs University Bremen, 4. - 10. July*

- Raunest, M., Bremer, C., Kott, N., Bung, C., Gerlich, W., Glebe, D., Cemic, F., and Hardt, M. (2009). Cholesterol depletion and replenishment of hepatitis b virions reversibly alter their ultrastructure and infectivity. In *Microscopy Conference 2009, 9th Multinational Conference on Microscopy 2009, Graz, 30. Aug - 4. Sep*

Bibliography

Actor, J. (2007). *Elsevier's integrated immunology and microbiology*. Elsevier's integrated series. Mosby.

Akama, H., Kanemaki, M., Yoshimura, M., Tsukihara, T., Kashiwagi, T., Yoneyama, H., ichiro Narita, S., Nakagawa, A., and Nakae, T. (2004). Crystal structure of the drug discharge outer membrane protein, oprm, of pseudomonas aeruginosa: dual modes of membrane anchoring and occluded cavity end. *J Biol Chem*, 279(51):52816–52819.

Alanis, A. J. (2005). Resistance to antibiotics: Are we in the post-antibiotic era? *Archives of Medical Research*, 36(6):697 – 705.

Alder, B. J. and Wainwright, T. E. (1957). Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics*, 27(5):1208–1209.

Andersen, C., Hughes, C., and Koronakis, V. (2001). Protein export and drug efflux through bacterial channel-tunnels. *Curr Opin Cell Biol*, 13(4):412–416.

Andersen, C., Koronakis, E., Bokma, E., Eswaran, J., Humphreys, D., Hughes, C., and Koronakis, V. (2002a). Transition to the open state of the tolc periplasmic tunnel entrance. *Proc Natl Acad Sci U S A*, 99(17):11103–11108.

Andersen, C., Koronakis, E., Hughes, C., and Koronakis, V. (2002b). An aspartate ring at the tolc tunnel entrance determines ion selectivity and presents a target for blocking by large cations. *Mol Microbiol*, 44(5):1131–1139.

Balasubramanian, D., Schneper, L., Kumari, H., and Mathee, K. (2013). A dynamic and intricate regulatory network determines pseudomonas aeruginosa virulence. *Nucleic Acids Res*, 41(1):1–20.

Baron, R., Riley, C., Chenprakhon, P., Thotsaporn, K., Winter, R. T., Alfieri, A., Forneris, F., van Berkel, W. J. H., Chaiyen, P., Fraaije, M. W., Mattevi, A., and McCammon, J. A. (2009). Multiple pathways guide oxygen diffusion into flavoenzyme active sites. *Proc Natl Acad Sci U S A*, 106(26):10603–10608.

Bavro, V. N., Pietras, Z., Furnham, N., Pérez-Cano, L., Fernández-Recio, J., Pei, X. Y., Misra, R., and Luisi, B. (2008). Assembly and channel opening in a bacterial drug efflux machine. *Mol Cell*, 30(1):114–121.

Benz, R., Maier, E., and Gentschev, I. (1993). Tolc of escherichia coli functions as an outer membrane channel. *Zentralbl Bakteriol*, 278(2-3):187–196.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981). Interaction models for water in relation to protein hydration. *Intermolecular Forces*, pages 331–342.

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995a). Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43 – 56.

Berendsen, H. J. C., Vanderspoel, D., and Vandrunen, R. (1995b). Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun Comput Phys Commun*, 91(1-3):43–56–.

Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide protein data bank. *Nat Struct Biol*, 10(12):980.

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

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112(3):535–542.

Blank, T. E., Nougayrède, J.-P., and Donnenberg, M. S. (2002). Chapter 3 - enteropathogenic escherichia coli. In Donnenberg, M. S., editor, *Escherichia Coli*, pages 81 – 118. Academic Press, San Diego.

Born, M. and Oppenheimer, R. (1927). Zur quantentheorie der molekelen. *Ann. Phys.*, 389(20):457–484.

Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu,

J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009). Charmm: the biomolecular simulation program. *J Comput Chem*, 30(10):1545–1614.

Buhrman, G., Holzapfel, G., Fetics, S., and Mattos, C. (2010). Allosteric modulation of ras positions q61 for a direct role in catalysis. *Proc Natl Acad Sci U S A*, 107(11):4931–4936.

Cascales, E., Buchanan, S. K., Duché, D., Kleanthous, C., Lloubès, R., Postle, K., Riley, M., Slatin, S., and Cavard, D. (2007). Colicin biology. *Microbiol Mol Biol Rev*, 71(1):158–229.

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005). The amber biomolecular simulation programs. *J Comput Chem*, 26(16):1668–1688.

Caves, L. S., Evanseck, J. D., and Karplus, M. (1998). Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci*, 7(3):649–66–.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., Biedermannova, L., Sochor, J., and Damborsky, J. (2012). Caver 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol*, 8(10):e1002708.

Clausius, R. (1870). Ueber einen auf die wärme anwendbaren mechanischen satz. *Annalen der Physik*, 217(9):124–130.

Coleman, R. G. and Sharp, K. A. (2009). Finding and characterizing tunnels in macromolecules with application to ion channels and pores. *Biophys J*, 96(2):632–645.

Connolly, M. L. (1983a). Analytical molecular surface calculation. *Journal of Applied Crystallography*, 16(5):548–558.

Connolly, M. L. (1983b). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713.

Cordero, B., Gómez, V., Platero-Prats, A. E., Revés, M., Echeverría, J., Cremades, E., Barragán, F., and Alvarez, S. (2008). Covalent radii revisited. *Dalton Trans*, (21):2832–2838.

Cuesta-Seijo, J. A., Neale, C., Khan, M. A., Moktar, J., Tran, C. D., Bishop, R. E., Pomès, R., and Privé, G. G. (2010). Pagp crystallized from sds/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure*, 18(9):1210–1219.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh ewald: An n [centerdot] log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092.

Das, B., Helms, V., Lounnas, V., and Wade, R. C. (2000). Multicopy molecular dynamics simulations suggest how to reconcile crystallographic and product formation data for camphor enantiomers bound to cytochrome p-450cam. *J Inorg Biochem*, 81(3):121–131.

Das, D., Xu, Q. S., Lee, J. Y., Ankoudinova, I., Huang, C., Lou, Y., DeGiovanni, A., Kim, R., and Kim, S.-H. (2007). Crystal structure of the multidrug efflux transporter acrb at 3.1a resolution reveals the n-terminal region with conserved amino acids. *J Struct Biol*, 158(3):494–502.

Day, W. A. and Maurelli, A. T. (2002). Chapter 7 - shigella and enteroinvasive escherichia coli: Paradigms for pathogen evolution and host-parasite interactions. In Donnenberg, M. S., editor, *Escherichia Coli*, pages 209 – VI. Academic Press, San Diego.

Delano, W. L. (2002). The pymol molecular graphics system.

Delgado, M. A., Solbiati, J. O., Chiuchiolo, M. J., Farías, R. N., and Salomón, R. A. (1999). Escherichia coli outer membrane protein tolc is involved in production of the peptide antibiotic microcin j25. *J Bacteriol*, 181(6):1968–1970.

Dinh, T., Paulsen, I. T., and Saier, M. H. (1994). A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J Bacteriol*, 176(13):3825–3831.

Domagk, G. (1939). The nobel prize in physiology or medicine 1939. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1939.

Duchesne, E. (1897). *La thèse prophétique d'Ernest Duchesne, 1897 sur l'antagonisme entre les moisissures et les microbes.* PhD thesis, L. Pariente, Paris.

Eastman, P. and Pande, V. S. (2010). Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J Comput Chem*, 31(6):1268–1272.

Edelsbrunner, H. and Mucke, E. P. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72.

Eicher, T., jea Cha, H., Seeger, M. A., Brandstätter, L., El-Delik, J., Bohnert, J. A., Kern, W. V., Verrey, F., Grütter, M. G., Diederichs, K., and Pos, K. M. (2012). Transport of drugs by the multidrug transporter acrb involves an access and a deep binding pocket that are separated by a switch-loop. *Proc Natl Acad Sci U S A*, 109(15):5687–5692.

Elsinghorst, E. A. (2002). Chapter 5 - enterotoxigenic escherichia coli. In Donnenberg, M. S., editor, *Escherichia Coli*, pages 155 – 187. Academic Press, San Diego.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593.

Eswaran, J., Hughes, C., and Koronakis, V. (2003). Locking tolc entrance helices to prevent protein translocation by the bacterial type i export apparatus. *J Mol Biol*, 327(2):309–315.

Exner, T., Keil, M., Moeckel, G., and Brickmann, J. (1998). Identifification of substrate channels and protein cavities. *Journal of Molecular Modeling*, 4:340–343. 10.1007/s008940050091.

Faust, A., Panjikar, S., Mueller, U., Parthasarathy, V., Schmidt, A., Lamzin, V. S., and Weiss, M. S. (2008). A tutorial for learning and teaching macromolecular crystallography. *Journal of Applied Crystallography*, 41(6):1161–1172.

Federici, L., Du, D., Walas, F., Matsumura, H., Fernandez-Recio, J., McKeegan, K. S., Borges-Walmsley, M. I., Luisi, B. F., and Walmsley, A. R. (2005). The crystal structure of the outer membrane protein vcec from the bacterial pathogen vibrio cholerae at 1.8 a resolution. *J Biol Chem*, 280(15):15307–15314.

Fischer, N. and Kandt, C. (2011). Three ways in, one way out: water dynamics in the trans-membrane domains of the inner membrane translocase acrb. *Proteins*, 79(10):2871–2885.

Fischer, N. and Kandt, C. (2013). Porter domain opening and closing motions in the multi-drug efflux transporter acrb. *Biochim Biophys Acta*, 1828(2):632–641.

Fischer, N., Raunest, M., and Kandt, C. (2010a). Mapping the proton conduction pathways in the inner membrane multi-drug translocase acrb. *Biophysical Journal*, 98(3, Supplement 1):633a –.

Fischer, N., Raunest, M., and Kandt, C. (2010b). Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase acrb. In *BioScience 2010, Workshop on expanding the frontiers of biomolecular science, Forschungszentrum Jülich, Germany, November 15 - 17.*

Fischer, N., Raunest, M., and Kandt, C. (2011a). Three ways in, one way out: Water dynamics in the trans-membrane domains of the inner membrane translocase acrb. *Biophysical Journal*, 100(3, Supplement 1):155a –.

Fischer, N., Raunest, M., Schmidt, T., Koch, D., and Kandt, C. (2013). Efflux pump-mediated antibiotics resistance: Insights from computational structural biology. *Interdisciplinary Science: Computational life Sciences, CECAM proceedings special issue*, submitted.

Fischer, N., Raunest, M., Schmidt, T. H., and Kandt, C. (2011b). Simulations of the multidrug efflux pump acrab-tolc; gordon research conference: Multi-drug efflux systems. In *Multi-drug Efflux Systems; Les Diablerets, Switzerland; 12.-17.Aug.*

Fleming (1945). The nobel prize in physiology or medicine 1945. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1945/.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of b. influenzae. *Br J Exp Pathol*, 3(3):226—236.

Fralick, J. A. (1996). Evidence that tolc is required for functioning of the mar/acrab efflux pump of escherichia coli. *J Bacteriol*, 178(19):5803–5805.

Fralick, J. A. and Burns-Keliher, L. L. (1994). Additive effect of tolc and rfa mutations on the hydrophobic barrier of the outer membrane of escherichia coli k-12. *J Bacteriol*, 176(20):6404–6406.

Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A., and Schulten, K. (2006). Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3):437–449.

Gabriel, E., Fagg, G. E., Bosilca, G., Angskun, T., Dongarra, J. J., Squyres, J. M., Sahay, V., Kambadur, P., Barrett, B., Lumsdaine, A., Castain, R. H., Daniel, D. J., Graham, R. L., and Woodall, T. S. (2004). Open MPI: Goals, concept, and design of a next generation MPI implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary.

Ge, Q., Yamada, Y., and Zgurskaya, H. (2009). The c-terminal domain of acra is essential for the assembly and function of the multidrug efflux pump acrab-tolc. *J Bacteriol*, 191(13):4365–4371.

Gotoh, N., Tsujimoto, H., Nomura, A., Okamoto, K., Tsuda, M., and Nishino, T. (1998). Functional replacement of oprj by oprm in the mexcd-oprj multidrug efflux system of pseudomonas aeruginosa. *FEMS Microbiol Lett*, 165(1):21–27.

Grossfield, A. and Zuckerman, D. M. (2009). Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu Rep Comput Chem*, 5:23–48.

Gumbart, J., Wang, Y., Aksimentiev, A., Tajkhorshid, E., and Schulten, K. (2005). Molecular dynamics simulations of proteins in lipid bilayers. *Curr Opin Struct Biol*, 15(4):423–431.

Gumbart, J., Wiener, M. C., and Tajkhorshid, E. (2009). Coupling of calcium and substrate binding through loop alignment in the outer-membrane transporter btub. *J Mol Biol*, 393(5):1129–1142.

Guvench, O. and MacKerell, A. D. (2009). Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput Biol*, 5(7):e1000435.

Hajjar, E., Bessonov, A., Molitor, A., Kumar, A., Mahendran, K. R., Winterhalter, M., Pagès, J.-M., Ruggerone, P., and Ceccarelli, M. (2010). Toward screening for antibiotics with enhanced permeation properties through bacterial porins. *Biochemistry*, 49(32):6928–6935.

Hancock, R. E. W. and Brinkman, F. S. L. (2002). Function of pseudomonas porins in uptake and efflux. *Annu Rev Microbiol*, 56:17–38.

Haynes, M. (2012). *CRC Handbook of Chemistry and Physics. 92nd Edition.*, volume 32. CRC Press.

Hendlich, M., Rippmann, F., and Barnickel, G. (1997). Ligsite: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model*, 15(6):359–63, 389.

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447.

Higgins, M. K., Eswaran, J., Edwards, P., Schertler, G. F. X., Hughes, C., and Koronakis, V. (2004). Structure of the ligand-blocked periplasmic entrance of the bacterial multidrug efflux protein tolc. *J Mol Biol*, 342(3):697–702.

Ho, B. K. and Gruswitz, F. (2008). Hollow: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol*, 8:49.

Hockney, R. (1970). The potential calculation and some applications. *Methods in Computational Physics*, 9:135–211.

Hubbard, S. J. and Thornton, J. M. (1993). 'NACCESS', computer program. Technical report, Department of Biochemistry Molecular Biology, University College London.

Humphrey, W., Dalke, A., and Schulten, K. (1996). Vmd: visual molecular dynamics. *J Mol Graph*, 14(1):33–8, 27–8.

Hwang, J., Zhong, X., and Tai, P. C. (1997). Interactions of dedicated export membrane proteins of the colicin v secretion system: Cvaa, a member of the membrane fusion protein family, interacts with cvab and tolc. *J Bacteriol*, 179(20):6264–6270.

Johnston, M. A., Galván, I. F., and Villà-Freixa, J. (2005). Framework-based design of a new all-purpose molecular simulation application: the adun simulator. *J Comput Chem*, 26(15):1647–1659.

Kandt, C., Ash, W. L., and Tieleman, D. P. (2007). Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, 41(4):475–488.

Kandt, C., Gerwert, K., and Schlitter, J. (2005). Water dynamics simulation as a tool for probing proton transfer pathways in a heptahelical membrane protein. *Proteins*, 58(3):528–537.

Kandt, C., Schlitter, J., and Gerwert, K. (2004). Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment. *Biophys J*, 86(2):705–717.

Kandt, C. and Tieleman, D. P. (2010). Holo-btuf stabilizes the open conformation of the vitamin b12 abc transporter btucd. *Proteins*, 78(3):738–753.

Kandt, C., Xu, Z., and Tieleman, D. P. (2006). Opening and closing motions in the periplasmic vitamin b12 binding protein btuf. *Biochemistry*, 45(44):13284–13292.

Karplus, M. and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 9(9):646–652.

Karpowich, N. K., Huang, H. H., Smith, P. C., and Hunt, J. F. (2003). Crystal structures of the btuf periplasmic-binding protein for vitamin b12 suggest a functionally important reduction in protein mobility upon ligand binding. *J Biol Chem*, 278(10):8429–8434.

Kim, D., Cho, C.-H., Cho, Y., Ryu, J., Bhak, J., and Kim, D.-S. (2008). Pocket extraction on proteins via the voronoi diagram of spheres. *J Mol Graph Model*, 26(7):1104–1112.

Kleywegt, G. J. and Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr*, 50(Pt 2):178–185.

Koch, D. C., Raunest, M., Harder, T., and Kandt, C. (2013). Unilateral access regulation: Ground state dynamics of the pseudomonas aeruginosa outer membrane efflux duct oprm. *Biochemistry*, 52(1):178–187.

Koronakis, V., Eswaran, J., and Hughes, C. (2004). Structure and function of tolc: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem*, 73:467–489.

Koronakis, V., Li, J., Koronakis, E., and Stauffer, K. (1997). Structure of tolc, the outer membrane component of the bacterial type i efflux system, derived from two-dimensional crystals. *Mol Microbiol*, 23(3):617–626.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000). Crystal structure of the bacterial membrane protein tolc central to multidrug efflux and protein export. *Nature*, 405(6789):914–919.

Kotra, L. P., Golemi, D., Amro, N. A., Liu, G.-Y., and Mobashery, S. (1999). Dynamics of the Lipopolysaccharide Assembly on the Surface of Escherichia coli. *Journal of the American Chemical Society*, 121(38):8707–8711.

Krishnamoorthy, G., Tikhonova, E. B., and Zgurskaya, H. I. (2008). Fitting periplasmic membrane fusion proteins to inner membrane transporters: mutations that enable escherichia coli acra to function with pseudomonas aeruginosa mexb. *J Bacteriol*, 190(2):691–698.

Köhler, T., Michéa-Hamzehpour, M., Henze, U., Gotoh, N., Curty, L. K., and Pechère, J. C. (1997). Characterization of mexe-mexf-oprn, a positively regulated multidrug efflux system of pseudomonas aeruginosa. *Mol Microbiol*, 23(2):345–354.

Lambert, P. A. (2002). Mechanisms of antibiotic resistance in pseudomonas aeruginosa. *J R Soc Med*, 95 Suppl 41:22–26.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati,

R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., Szustakowki, J., and Consortium, I. H. G. S. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Laskowski, R. A. (1995). Surfnet: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph*, 13(5):323–30, 307–8.

Laurie, A. T. R. and Jackson, R. M. (2005). Q-sitefinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, 21(9):1908–1916.

Lee, B. and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *Journal of molecular biology*, 55(3):379–400.

Lee, P.-H. and Helms, V. (2011). Identifying continuous pores in protein structures with propores by computational repositioning of gating residues. *Proteins*.

Lennard-Jones, J. E. (1931). Cohesion. *Proceedings of the Physical Society*, 43(5):461.

Li, J., Shi, R., Yang, C., and Zhu, X. (2011). Exploration of the binding of benzimidazole-biphenyl derivatives to hemoglobin using docking and molecular dynamics simulation. *Int J Biol Macromol*, 48(1):20–26.

Liang, J., Edelsbrunner, H., and Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci*, 7(9):1884–1897.

Lins, R. D. and Straatsma, T. P. (2001). Computer simulation of the rough lipopolysaccharide membrane of pseudomonas aeruginosa. *Biophys J*, 81(2):1037–1046.

Livermore, D. M. (2002). Multiple mechanisms of antimicrobial resistance in pseudomonas aeruginosa: our worst nightmare? *Clin Infect Dis*, 34(5):634–640.

Lobedanz, S., Bokma, E., Symmons, M. F., Koronakis, E., Hughes, C., and Koronakis, V. (2007). A periplasmic coiled-coil interface underlying tolc recruitment and the assembly of bacterial drug efflux pumps. *Proc Natl Acad Sci U S A*, 104(11):4612–4617.

Lomovskaya, O., Warren, M. S., Lee, A., Galazzo, J., Fronko, R., Lee, M., Blais, J., Cho, D., Chamberland, S., Renau, T., Leger, R., Hecker, S., Watkins, W., Hoshino, K., Ishida, H., and Lee, V. J. (2001). Identification and characterization of inhibitors of multidrug resistance efflux pumps in pseudomonas aeruginosa: novel agents for combination therapy. *Antimicrob Agents Chemother*, 45(1):105–116.

Lugtenberg, B. and Alphen, L. V. (1983). Molecular architecture and functioning of the outer membrane of escherichia coli and other gram-negative bacteria. *Biochim Biophys Acta*, 737(1):51–115.

Lyczak, J. B., Cannon, C. L., and Pier, G. B. (2000). Establishment of pseudomonas aeruginosa infection: lessons from a versatile opportunist. *Microbes Infect*, 2(9):1051–1060.

Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., and Hearst, J. E. (1993). Molecular cloning and characterization of acra and acre genes of escherichia coli. *J Bacteriol*, 175(19):6299–6313.

Maeda, M. H. and Kinoshita, K. (2009). Development of new indices to evaluate protein-protein interfaces: assembling space volume, assembling space distance, and global shape descriptor. *J Mol Graph Model*, 27(6):706–711.

Masuda, N., Sakagawa, E., Ohya, S., Gotoh, N., Tsujimoto, H., and Nishino, T. (2000a). Contribution of the mexx-mexy-oprm efflux system to intrinsic resistance in pseudomonas aeruginosa. *Antimicrob Agents Chemother*, 44(9):2242–2246.

Masuda, N., Sakagawa, E., Ohya, S., Gotoh, N., Tsujimoto, H., and Nishino, T. (2000b). Substrate specificities of mexab-oprm, mexcd-oprj, and mexxy-oprm efflux pumps in pseudomonas aeruginosa. *Antimicrob Agents Chemother*, 44(12):3322–3327.

Matias, V. R. F., Al-Amoudi, A., Dubochet, J., and Beveridge, T. J. (2003). Cryo-transmission electron microscopy of frozen-hydrated sections of escherichia coli and pseudomonas aeruginosa. *J Bacteriol*, 185(20):6112–6118.

Meroueh, S. O., Bencze, K. Z., Hesek, D., Lee, M., Fisher, J. F., Stemmler, T. L., and Mobashery, S. (2006). Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci U S A*, 103(12):4404–4409.

Mesaros, N., Nordmann, P., Plésiat, P., Roussel-Delvallez, M., Eldere, J. V., Glupczynski, Y., Laethem, Y. V., Jacobs, F., Lebecque, P., Malfroot, A., Tulkens, P. M., and Bambeke, F. V. (2007). Pseudomonas aeruginosa: resistance and therapeutic options at the turn of the new millennium. *Clin Microbiol Infect*, 13(6):560–578.

Mikolosko, J., Bobyk, K., Zgurskaya, H. I., and Ghosh, P. (2006). Conformational flexibility in the multidrug efflux system protein acra. *Structure*, 14(3):577–587.

Moussatova, A., Kandt, C., O'Mara, M. L., and Tieleman, D. P. (2008). Atp-binding cassette transporters in escherichia coli. *Biochim Biophys Acta*, 1778(9):1757–1771.

Murakami, S. (2008). Multidrug efflux transporter, acrb–the pumping mechanism. *Curr Opin Struct Biol*, 18(4):459–465.

Murakami, S., Nakashima, R., Yamashita, E., Matsumoto, T., and Yamaguchi, A. (2006). Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. *Nature*, 443(7108):173–179.

Murakami, S., Nakashima, R., Yamashita, E., and Yamaguchi, A. (2002). Crystal structure of bacterial multidrug efflux transporter acrb. *Nature*, 419(6907):587–593.

Nakajima, A., Sugimoto, Y., Yoneyama, H., and Nakae, T. (2002). High-level fluoroquinolone resistance in pseudomonas aeruginosa due to interplay of the mexab-oprm efflux pump and the dna gyrase mutation. *Microbiol Immunol*, 46(6):391–395.

Nakajima, H., Kobayashi, K., Kobayashi, M., Asako, H., and Aono, R. (1995). Overexpression of the roba gene increases organic solvent tolerance and multiple antibiotic and heavy metal ion resistance in escherichia coli. *Appl Environ Microbiol*, 61(6):2302–2307.

Nakashima, R., Sakurai, K., Yamasaki, S., Nishino, K., and Yamaguchi, A. (2011). Structures of the multidrug exporter acrb reveal a proximal multisite drug-binding pocket. *Nature*, 480(7378):565–569.

Nataro, J. P. and Steiner, T. (2002). Chapter 6 - enteroaggregative and diffusely adherent escherichia coli. In Donnenberg, M. S., editor, *Escherichia Coli*, pages 189 – 207. Academic Press, San Diego.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, 11(4):281–296.

Nightingale, E. R. (1959). Phenomenological Theory of Ion Solvation. Effective Radii of Hydrated Ions. *The Journal of Physical Chemistry*, 63(9):1381–1387.

Nikaido, H. (2009). Multidrug resistance in bacteria. *Annu Rev Biochem*.

Nikaido, H. (2011). Structure and mechanism of rnd-type multidrug efflux pumps. *Adv Enzymol Relat Areas Mol Biol*, 77:1–60.

Nomura, M. and Witten, C. (1967). Interaction of colicins with bacterial cells. 3. colicin-tolerant mutations in escherichia coli. *J Bacteriol*, 94(4):1093–1111.

Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *J Comput Chem*, 25(13):1656–1676.

Osborn, M. J. (1969). Structure and biosynthesis of the bacterial cell wall. *Annu Rev Biochem*, 38:501–538.

Osborn, M. J., Rick, P. D., Lehmann, V., Rupprecht, E., and Singh, M. (1974). Structure and biogenesis of the cell envelope of gram-negative bacteria. *Ann N Y Acad Sci*, 235(0):52–65.

Pei, X.-Y., Hinchliffe, P., Symmons, M. F., Koronakis, E., Benz, R., Hughes, C., and Koronakis, V. (2011). Structures of sequential open states in a symmetrical opening transition of the tolc exit duct. *Proc Natl Acad Sci U S A*, 108(5):2112–2117.

Pellegrini-Calace, M., Maiwald, T., and Thornton, J. M. (2009). Porewalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comput Biol*, 5(7):e1000440.

Petrek, M., Kosinová, P., Koca, J., and Otyepka, M. (2007). Mole: a voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, 15(11):1357–1363.

Petrek, M., Otyepka, M., Banás, P., Kosinová, P., Koca, J., and Damborský, J. (2006). Caver: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, 7:316.

Petrova, S. S. and Solov'ev, A. D. (1997). The origin of the method of steepest descent. *Historia Mathematica*, 24(4):361 – 375.

Phan, G., Benabdelhak, H., Lascombe, M.-B., Benas, P., Rety, S., Picard, M., Ducruix, A., Etchebest, C., and Broutin, I. (2010). Structural and dynamical insights into the opening mechanism of p. aeruginosa oprm channel. *Structure*, 18(4):507–517.

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

Piggot, T. J., Holdbrook, D. A., and Khalid, S. (2011). Electroporation of the e. coli and s. aureus membranes: molecular dynamics simulations of complex bacterial membranes. *J Phys Chem B*, 115(45):13381–13388.

Piggot, T. J., Holdbrook, D. A., and Khalid, S. (2012). Conformational dynamics and membrane interactions of the e. coli outer membrane protein feca: A molecular dynamics simulation study. *Biochim Biophys Acta*.

Poole, K., Krebes, K., McNally, C., and Neshat, S. (1993). Multiple antibiotic resistance in pseudomonas aeruginosa: evidence for involvement of an efflux operon. *J Bacteriol*, 175(22):7363–7372.

Pos, K. M. (2009). Drug transport mechanism of the acrb efflux pump. *Biochim Biophys Acta*, 1794(5):782–793.

Poupon, A. (2004). Voronoi and voronoi-related tessellations in studies of protein structure and interaction. *Curr Opin Struct Biol*, 14(2):233–241.

Putman, M., van Veen, H. W., and Konings, W. N. (2000). Molecular properties of bacterial multidrug transporters. *Microbiol Mol Biol Rev*, 64(4):672–693.

Ramos, J. L. (2010). *Pseudomonas: Volume 6: Molecular Microbiology, Infection and Biodiversity*. Springer.

Raunest, M., Bremer, C., Kott, N., Bung, C., Gerlich, W., Glebe, D., Cemic, F., and Hardt, M. (2009). Cholesterol depletion and replenishment of hepatitis b virions reversibly alter their ultrastructure and infectivity. In *Microscopy Conference 2009, 9th Multinational Conference on Microscopy 2009, Graz, 30. Aug - 4. Sep.*

Raunest, M., Fischer, N., and Kandt, C. (2010a). Dynamics in the tolc gating region and cavity detection based on protein and water dynamics. In *462nd WE Heraeus Seminar,Transport across membranes: Multiple drug resistance, mechanisms and new tools, Jacobs University Bremen, 4. - 10. July.*

Raunest, M., Fischer, N., and Kandt, C. (2011). Computer simulation of tolc ground state dynamics and spontaneous binding of the acrb docking domain. *Biophysical Journal*, 100(3, Supplement 1):132a –.

Raunest, M. and Kandt, C. (2010). dxtuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. In *BioScience 2010, Workshop on expanding the frontiers of biomolecular science, Forschungszentrum Jülich, Germany, November 15 - 17*.

Raunest, M. and Kandt, C. (2011a). dxtuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J Mol Graph Model*, 29(7):895–905.

Raunest, M. and Kandt, C. (2011b). Dxtuber: Detecting protein cavities, tunnels, surface clefts based on protein, solvent dynamics. *Biophysical Journal*, 100(3, Supplement 1):48a –.

Raunest, M. and Kandt, C. (2012a). Locked on one side only: Computer simulations of the outer membrane efflux duct tolc in absence and presence of the acrb-docking-domain. *Biophysical Journal*, 102:622a – 623a.

Raunest, M. and Kandt, C. (2012b). Locked on one side only: ground state dynamics of the outer membrane efflux duct tolc. *Biochemistry*, 51(8):1719–1729.

Raunest, M. and Kandt, C. (2012c). Locked on one side only: Ground state dynamics of the outer membrane efflux duct tolc. In *Dipartimento di Fisica, Università di Cagliari*.

Raunest, M., Nadine, F., and Kandt, C. (2010b). Molecular dynamics simulations reveal tolc flexibility in the acrb interface region. *Biophysical Journal*, 98(3, Supplement 1):374a –.

Reffay, M., Gambin, Y., Benabdelhak, H., Phan, G., Taulier, N., Ducruix, A., Hodges, R. S., and Urbach, W. (2009). Tracking membrane protein association in model membranes. *PLoS One*, 4(4):e5035.

Richards, F. M. (1974). The interpretation of protein structures: total volume, group volume distributions and packing density. *J Mol Biol*, 82(1):1–14.

Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu Rev Biophys Bioeng*, 6:151–176.

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

Römer, A., Wieler, L. H., and Schierack, P. (2012). Analyses of intestinal commensal escherichia coli strains from wild boars suggest adaptation to conventional pig production conditions. *Vet Microbiol*, 161(1-2):122–129.

Saier, M. H. and Paulsen, I. T. (2001). Phylogeny of multidrug transporters. *Semin Cell Dev Biol*, 12(3):205–213.

Sanner, M. F., Olson, A. J., and Spehner, J. C. (1995). Fast and robust computation of molecular surfaces. pages 406–407.

Sanner, M. F., Olson, A. J., and Spehner, J. C. (1996). Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320.

Schemberg, J., Schneider, K., Demmer, U., Warkentin, E., Müller, A., and Ermler, U. (2007). Towards biological supramolecular chemistry: a variety of pocket-templated, individual metal oxide cluster nucleations in the cavity of a mo/w-storage protein. *Angew Chem Int Ed Engl*, 46(14):2408–2413.

Schmidt, B., Ho, L., and Hogg, P. J. (2006). Allosteric disulfide bonds. *Biochemistry*, 45(24):7429–7433.

Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2010). Computer simulation of the assembled acrab-tolc multidrug efflux pump. In *462nd WE Heraeus Seminar,Transport across membranes: Multiple drug resistance, mechanisms and new tools Jacobs University Bremen, 4. - 10. July.*

Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2011a). Atomistic and coarse-grain computer simulations of the assembled acrab-tolc multidrug efflux pump. *Biophysical Journal*, 100(3, Supplement 1):146a –.

Schmidt, T. H., Fischer, N., Raunest, M., and Kandt, C. (2011b). Atomistic and coarse-grain computer simulations of the assembled acrab-tolc multidrug efflux pump. In *Coarse-Grained Biomolecular Modeling Workshop Levi, Finland, March 7-12.*

Schmidtke, P., Souaille, C., Estienne, F., Baurin, N., and Kroemer, R. T. (2010). Large-scale comparison of four binding site detection algorithms. *J Chem Inf Model*, 50(12):2191–2200.

Schnell, J. R. and Chou, J. J. (2008). Structure and mechanism of the m2 proton channel of influenza a virus. *Nature*, 451(7178):591–595.

Schofield, P. (1973). Computer simulation studies of the liquid state. *Computer Physics Communications*, 5(1):17 – 23.

Schulz, R. and Kleinekathöfer, U. (2009). Transitions between closed and open conformations of tolc: the effects of ions in simulations. *Biophys J*, 96(8):3116–3125.

Schulz, R., Vargiu, A. V., Ruggerone, P., and Kleinekathöfer, U. (2011). Role of water during the extrusion of substrates by the efflux transporter acrb. *J Phys Chem B*, 115(25):8278–8287.

Seeger, M. A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., and Pos, K. M. (2006). Structural asymmetry of acrb trimer suggests a peristaltic pump mechanism. *Science*, 313(5791):1295–8–.

Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O., and Grütter, M. G. (2007). Drug export pathway of multidrug exporter acrb revealed by darpin inhibitors. *PLoS Biol*, 5(1):e7.

Shannon, R. D. (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A*, 32(5):751–767.

Shaw, D. E., Dror, R. O., Salmon, J. K., Grossman, J. P., Mackenzie, K. M., Bank, J. A., Young, C., Deneroff, M. M., Batson, B., Bowers, K. J., Chow, E., Eastwood, M. P., Ierardi, D. J., Klepeis, J. L., Kuskin, J. S., Larson, R. H., Larsen, K. L., Maragakis, P., Moraes, M. A., Piana, S., Shan, Y., and Towles, B. (2009). Millisecond-scale molecular dynamics simulations on anton. In *SC '09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pages 1–11, New York, NY, USA. ACM.

Shlaes, D. (2011). 3.23 - antibiotics: The miracle menaced. In Moo-Young, M., editor, *Comprehensive Biotechnology (Second Edition)*, pages 243 – 254. Academic Press, Burlington, second edition edition.

Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A., and Sansom, M. S. (1996). Hole: a program for the analysis of the pore dimensions of ion channel structural models. *J Mol Graph*, 14(6):354–60, 376.

Sonne, J., Kandt, C., Peters, G. H., Hansen, F. Y., Jensen, M., and Tieleman, D. P. (2007). Simulation of the coupling between nucleotide binding and transmembrane domains in the atp binding cassette transporter btucd. *Biophys J*, 92(8):2727–2734.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., and Olson, M. V. (2000). Complete genome sequence of pseudomonas aeruginosa pao1, an opportunistic pathogen. *Nature*, 406(6799):959–964.

Straatsma, T. P. and Soares, T. A. (2009). Characterization of the outer membrane protein oprf of pseudomonas aeruginosa in a lipopolysaccharide membrane by computer simulation. *Proteins*, 74(2):475–488.

Strateva, T. and Yordanov, D. (2009). Pseudomonas aeruginosa - a phenomenon of bacterial resistance. *J Med Microbiol*, 58(Pt 9):1133–1148.

Su, C.-C., Li, M., Gu, R., Takatsuka, Y., McDermott, G., Nikaido, H., and Yu, E. W. (2006). Conformation of the acrb multidrug efflux pump in mutants of the putative proton relay pathway. *J Bacteriol*, 188(20):7290–7296.

Su, C.-C., Long, F., Zimmermann, M. T., Rajashankar, K. R., Jernigan, R. L., and Yu, E. W. (2011). Crystal structure of the cusba heavy-metal efflux complex of escherichia coli. *Nature*, 470(7335):558–562.

Sulavik, M. C., Houseweart, C., Cramer, C., Jiwani, N., Murgolo, N., Greene, J., DiDomenico, B., Shaw, K. J., Miller, G. H., Hare, R., and Shimer, G. (2001). Antibiotic susceptibility profiles of escherichia coli strains lacking multidrug efflux pump genes. *Antimicrob Agents Chemother*, 45(4):1126–1136.

Symmons, M. F., Bokma, E., Koronakis, E., Hughes, C., and Koronakis, V. (2009). The assembled structure of a complete tripartite bacterial multidrug efflux pump. *Proc Natl Acad Sci U S A*, 106:7173–7178.

Tamura, N., Murakami, S., Oyama, Y., Ishiguro, M., and Yamaguchi, A. (2005). Direct interaction of multidrug efflux transporter acrb and outer membrane channel tolc detected via site-directed disulfide cross-linking. *Biochemistry*, 44(33):11115–11121.

Thorpe, C. M., Ritchie, J. M., and Acheson, D. W. (2002). Chapter 4 - enterohemorrhagic and other shiga toxin-producing escherichia coli. In Donnenberg, M. S., editor, *Escherichia Coli*, pages 119 – V. Academic Press, San Diego.

Tieleman, D. P. and Berendsen, H. J. (1998). A molecular dynamics study of the pores formed by escherichia coli ompf porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys J*, 74(6):2786–2801.

Tikhonova, E. B., Yamada, Y., and Zgurskaya, H. I. (2011). Sequential mechanism of assembly of multidrug efflux pump acrab-tolc. *Chem Biol*, 18(4):454–463.

Tikhonova, E. B. and Zgurskaya, H. I. (2004). Acra, acrb, and tolc of escherichia coli form a stable intermembrane multidrug efflux complex. *J Biol Chem*, 279(31):32116–32124.

Touzé, T., Eswaran, J., Bokma, E., Koronakis, E., Hughes, C., and Koronakis, V. (2004). Interactions underlying assembly of the escherichia coli acrab-tolc multidrug efflux system. *Mol Microbiol*, 53(2):697–706.

Typas, A., Banzhaf, M., van den Berg van Saparoea, B., Verheul, J., Biboy, J., Nichols, R. J., Zietek, M., Beilharz, K., Kannenberg, K., von Rechenberg, M., Breukink, E., den Blaauwen, T., Gross, C. A., and Vollmer, W. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143(7):1097 – 1109.

Törnroth-Horsefield, S., Gourdon, P., Horsefield, R., Brive, L., Yamamoto, N., Mori, H., Snijder, A., and Neutze, R. (2007). Crystal structure of acrb in complex with a single transmembrane subunit reveals another twist. *Structure*, 15(12):1663–1673.

Vaccaro, L., Koronakis, V., and Sansom, M. S. P. (2006). Flexibility in a drug transport accessory protein: molecular dynamics simulations of mexa. *Biophys J*, 91(2):558–564.

Vaccaro, L., Scott, K. A., and Sansom, M. S. P. (2008). Gating at both ends and breathing in the middle: conformational dynamics of tolc. *Biophys J*, 95(12):5681–5691.

Vargiu, A. V., Collu, F., Schulz, R., Pos, K. M., Zacharias, M., Kleinekathöfer, U., and Ruggerone, P. (2011). Effect of the f610a mutation on substrate extrusion in the acrb transporter: explanation and rationale by molecular dynamics simulations. *J Am Chem Soc*, 133(28):10704–10707.

Verlet, L. (1967). Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103.

Volkamer, A., Griewel, A., Grombacher, T., and Rarey, M. (2010). Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model*, 50(11):2041–2052.

Vollmer, W. and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in escherichia coli. *Biochim Biophys Acta*, 1778(9):1714–1734.

Vollmer, W., Blanot, D., and de Pedro, M. A. (2008a). Peptidoglycan structure and architecture. *FEMS Microbiol Rev*, 32(2):149–167.

Vollmer, W., Joris, B., Charlier, P., and Foster, S. (2008b). Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev*, 32(2):259–286.

Voss, N. R. and Gerstein, M. (2010). 3v: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res*, 38(Web Server issue):W555–W562.

Vreede, J., Juraszek, J., and Bolhuis, P. G. (2010). Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc Natl Acad Sci U S A*, 107(6):2397–2402.

Vuillemin, P. (1889). Antibiose et symbiose. *Assoc. franc. pour l'Avanc. des Sciences*, 2:525—-542.

Wandersman, C. and Delepelaire, P. (1990). Tolc, an escherichia coli outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci U S A*, 87(12):4776–4780.

Wang, Y., Shaikh, S. A., and Tajkhorshid, E. (2010). Exploring transmembrane diffusion pathways with molecular dynamics. *Physiology (Bethesda)*, 25(3):142–154.

White, D. G., Goldman, J. D., Demple, B., and Levy, S. B. (1997). Role of the acrab locus in organic solvent tolerance mediated by expression of mara, soxs, or roba in escherichia coli. *J Bacteriol*, 179(19):6122–6126.

Wong, K. K., Brinkman, F. S., Benz, R. S., and Hancock, R. E. (2001). Evaluation of a structural model of pseudomonas aeruginosa outer membrane protein oprm, an efflux component involved in intrinsic antibiotic resistance. *J Bacteriol*, 183(1):367–374.

Xu, Y., Moeller, A., Jun, S.-Y., Le, M., Yoon, B.-Y., Kim, J.-S., Lee, K., and Ha, N.-C. (2012). Assembly and channel opening of outer membrane protein in tripartite drug efflux pumps of gram-negative bacteria. *J Biol Chem*, 287(15):11740–11750.

Yaffe, E., Fishelovitch, D., Wolfson, H. J., Halperin, D., and Nussinov, R. (2008a). Molaxis: a server for identification of channels in macromolecules. *Nucleic Acids Res*, 36(Web Server issue):W210–W215.

Yaffe, E., Fishelovitch, D., Wolfson, H. J., Halperin, D., and Nussinov, R. (2008b). Molaxis: efficient and accurate identification of channels in macromolecules. *Proteins*, 73(1):72–86.

Yao, X.-Q., Kenzaki, H., Murakami, S., and Takada, S. (2010). Drug export and allosteric coupling in a multidrug transporter revealed by molecular simulations. *Nat Commun*, 1:117.

Yin, H., Feng, G., Clore, G. M., Hummer, G., and Rasaiah, J. C. (2010). Water in the polar and nonpolar cavities of the protein interleukin-1β. *J Phys Chem B*, 114(49):16290–16297.

Yu, E. W., Aires, J. R., McDermott, G., and Nikaido, H. (2005). A periplasmic drug-binding site of the acrb multidrug efflux pump: a crystallographic and site-directed mutagenesis study. *J Bacteriol*, 187(19):6804–6815.

Yu, E. W., Aires, J. R., and Nikaido, H. (2003a). Acrb multidrug efflux pump of escherichia coli: composite substrate-binding cavity of exceptional flexibility generates its extremely wide substrate specificity. *J Bacteriol*, 185(19):5657–5664.

Yu, E. W., McDermott, G., Zgurskaya, H. I., Nikaido, H., and Koshland, D. E. (2003b). Structural basis of multiple drug-binding capacity of the acrb multidrug efflux pump. *Science*, 300(5621):976–980.

Zakharov, S. D., Eroukova, V. Y., Rokitskaya, T. I., Zhalnina, M. V., Sharma, O., Loll, P. J., Zgurskaya, H. I., Antonenko, Y. N., and Cramer, W. A. (2004). Colicin occlusion of ompf and tolc channels: outer membrane translocons for colicin import. *Biophys J*, 87(6):3901–3911.

Zgurskaya, H. I. (2009). Multicomponent drug efflux complexes: architecture and mechanism of assembly. *Future Microbiol*, 4(7):919–932.

Zgurskaya, H. I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and function of the outer membrane channel tolc in multidrug resistance and physiology of enterobacteria. *Front Microbiol*, 2:189.

Zgurskaya, H. I. and Nikaido, H. (2000). Multidrug resistance mechanisms: drug efflux across two membranes. *Mol Microbiol*, 37(2):219–225.

Zgurskaya, H. I., Yamada, Y., Tikhonova, E. B., Ge, Q., and Krishnamoorthy, G. (2009). Structural and functional diversity of bacterial membrane fusion proteins. *Biochim Biophys Acta*, 1794(5):794–807.

Zhao, Q., Li, X. Z., Srikumar, R., and Poole, K. (1998). Contribution of outer membrane efflux protein oprm to antibiotic resistance in pseudomonas aeruginosa independent of mexab. *Antimicrob Agents Chemother*, 42(7):1682–1688.