HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

The Haemolysin A Type 1 Secretion System of *Escherichia coli*

–

Characterization and Biotechnological Application

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Christian K. W. Schwarz

aus Solingen

Düsseldorf, August 2012

aus dem Institut für Biochemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Lutz Schmitt

Korreferent: Prof. Dr. Karl-Erich Jaeger

Tag der mündlichen Prüfung: 26.10.2012

Table of Contents

Table of contents I

1 Summary 1

2 Zusammenfassung 2

3 Introduction 3

3.1 Protein-mediated transport across membranes 3
3.2 ABC transporters 4
3.2.1 Nucleotide-binding domains 5
3.2.2 Transmembrane domains 7
3.2.3 Mechanism of ABC transporters 8
3.3 The cell wall of prokaryotes 9
3.4 Secretion systems in Gram-negative bacteria 10
3.5 Type 1 secretion systems (T1SS) 12
3.6 The haemolysin A (HlyA) T1SS of *E. coli* 12
3.6.1 The toxin HlyA 13
3.6.2 The outer membrane protein TolC 15
3.6.3 The membrane fusion protein HlyD 16
3.6.4 The ABC transporter HlyB 18
3.6.5 The C39 peptidase-like domain of HlyB 19
3.7 Application of the HlyA T1SS in biotechnology 20
3.8 Production strategies for peptides 21

4 Aims 23

4.1 Characterization of the HlyA T1SS 23
4.2 Biotechnological Application of the HlyA T1SS 23

5 Publications 25

5.1 Chapter I 25
5.2 Chapter II 77
5.3 Chapter III 82
5.4 Chapter IV 86
5.5 Chapter V 119
5.6 Chapter VI 127

6 Discussion 132

6.1 Characterization of the HlyA T1SS of *E. coli* 132
6.1.1 The CLD is pivotal for the secretion of HlyA 132
6.1.2 Function of the C-terminal secretion signal 136
6.1.2.1 The secretion signal represents the substrate for HlyB 137
6.1.2.2 The activating model 138
6.1.2.3 The terminating model 139
6.1.2.4 Does the secretion signal activate or terminate the transport? 139
6.1.3 Energization of T1SS transport 142
6.1.4 Function of HlyD during transport 148
6.1.5 The activating alternating access model 149
6.2 Biotechnological Application of the HlyA T1SS of *E. coli* 151
6.2.1 Biotechnological production of proteins 151
6.2.2 Biotechnological production of peptides 152

7 Literature 153

8 Acknowledgment/Danksagung 165

9 Curriculum vitae 168

10 List of publications 170

11 List of patent applications 171

12 Participation at scientific conferences 172

13 Erklärung zur Promotion 174

1 Summary

Haemolysin A (HlyA) is a major virulence factor of uropathogenic *Escherichia coli (E. coli)* strains that cause severe diseases in humans. The toxin is transported by a dedicated Type 1 secretion system (T1SS) from the bacterial cytoplasm to the surrounding, where it lyses host cells. Three membrane proteins assemble the HlyA T1SS secretion complex and are essential to catalyze the one-step transport of the 110 kDa HlyA across the Gram-negative cell envelope. TolC persists in the outer membrane and represents a pore for the passage of HlyA. The membrane fusion protein HlyD bridges TolC to the ABC transporter HlyB at the inner membrane. Although important insights into the coordinated secretion process were gained in the last decades, many questions remained unanswered. For example, an N-terminal domain of HlyB represents about 20 % of the primary sequence, however, its function was unknown.

As part of this doctoral research, this C39 peptidase-like domain (CLD) was characterized in great detail by structural and biochemical approaches. These studies revealed the pivotal function of the CLD in the secretion process. It interacts with the RTX domain of unfolded HlyA to protect the secretion-competent state from being aggregated and/or degraded.

The HlyA T1SS has also been analyzed for its applicability as production platform for proteins. Unfortunately, many proteins were secreted barely or not at all. This doctoral research provides an explanation for this failure and demonstrates that the secretion efficiency of proteins correlates inversely with the folding velocity of the passenger protein. Moreover, non-secreted passenger proteins could be secreted after the introduction of slow-folding mutations and the secreted protein was shown to be functional.

Despite this "slow-folding concept", the requirement of slow-folding proteins hampers the industrial application of the HlyA T1SS as production platform. Therefore, another focus of this doctoral research was the improvement of the secretion system to circumvent this limitation. A new secretion system could be established that enables the secretion of all proteins tested so far with good yields and in a functional form. Importantly, the optimized system also represents the first recombinant system for the production of peptides.

2 Zusammenfassung

Haemolysin A (HlyA) ist ein maßgeblicher Virulenzfaktor uropathogener *E. coli* Stämme, die schwere Erkrankungen in Menschen hervorrufen. Das Toxin wird von den Bakterien mit einem Typ 1 Sekretionssystem (T1SS) in die Umgebung sezerniert, wo es Wirtszellen lysiert. Drei Membranproteine bilden den T1SS Komplex, der den einstufigen Transport über die gesamte Gram-negative Zellhülle bewerkstelligt. TolC befindet sich in der äußeren Membran und bildet dort eine Pore, die von einem Membranfusionsprotein (Haemolysin D, HlyD) mit dem ABC Transporter Haemolysin B (HlyB) in der inneren Membran verbunden wird. Obwohl viele Details über den Mechanismus der Sezernierung bekannt sind, bleiben wichtige Fragen bislang unbeantwortet. Beispielsweise ist die Funktion einer N-terminalen Domäne von HlyB unbekannt, die circa 20% der gesamten Primärsequenz ausmacht. Diese Domäne, die C39 peptidase-like domain (CLD), wurde im Rahmen dieser Doktorarbeit strukturell und biochemisch charakterisiert. Die erzielten Ergebnisse gewähren Einblicke in die essentielle Funktion der CLD für die Sezernierung von HlyA. Die CLD interagiert mit der RTX Domäne von HlyA und schützt so vermutlich dessen ungefalteten, sezernierbaren Zustand innerhalb der Zelle vor Aggregation und/oder proteolytischem Abbau.

Das HlyA T1SS wurde außerdem auf dessen Eignung zur Produktion von Proteinen hin untersucht. Zahlreiche Proteine wurden jedoch schlecht oder gar nicht sezerniert. In dieser Doktorarbeit wird eine Erklärung für die fehlgeschlagenen Produktionsversuche geliefert. Demnach korrelieren die Sezernierungseffizienzen umgekehrt mit der Faltungsgeschwindigkeit der zu transportierenden Proteine. Außerdem konnte ein ursprünglich nicht-sezerniertes Protein durch die Insertion einer spezifischen Mutation, die zu einer Reduktion der Faltungsgeschwindigkeit führt, in guten Ausbeuten und in einem funktionalen Zustand sezerniert werden.

Trotz dieses "*slow-folding*" Konzepts ist die Nutzung des HlyA T1SS für die Proteinproduktion stark eingeschränkt, da viele Protein zu schnell falten. Im Rahmen dieser Doktorarbeit konnte nun das Sekretionssystem optimiert werden. Dabei gelang die Entwicklung eines universellen Sekretionssystems, das bislang die Produktion jedes Proteins in einem funktionalen Zustand ermöglicht. Die entwickelte Technologie stellt zugleich das erste biotechnologische Verfahren zur Herstellung beliebiger Peptide dar.

3 Introduction

All living cells are characterized by the presence of a plasma membrane, which builds up a barrier to separate the extracellular environment from the cellular space. The plasma membrane consists of a lipid bilayer that is impermeable for the passage of ions and most water-soluble, hydrophilic molecules [1,2]. Despite being important as a barrier, mechanisms had to be evolved in living cells to exchange specific molecules with the environment, for example to extract nutrients from the outside or to extrude toxic metabolites [3]. This selective permeability of cells is conferred by membrane proteins that are embedded within the lipid bilayer. Membrane proteins are involved in many vital functions of cells, for example signaling, enzymatic processes, cell adhesion or ion transport [3,4]. Both components together, membrane proteins and the lipid bilayer, build up biological membranes: oriented, two-dimensional barriers with selective permeability [1]. The importance of membrane proteins is furthermore reflected by the number of genes encoding them, which account for up to 30 % of all open-reading frames (ORFs) in sequenced genomes [5].

3.1 Protein-mediated transport across membranes

Membrane proteins assist the transport of ions and many hydrophilic molecules across biological membranes by two fundamental different mechanisms. Membrane channels catalyze the passive diffusion of substrates along a gradient (downhill), whereas membrane transporters move cargo against a gradient (uphill) [2] (Figure 3-1). Uphill transport requires external energy, which can be provided either by coupling the transport to the downhill movement of a second substrate (secondary active transport) [6] or by primary energy sources, for example ATP or light (primary active transport) [7,8]. A large class of these primary active transporters is the ATP-driven ATP-binding cassette (ABC) transporter family.

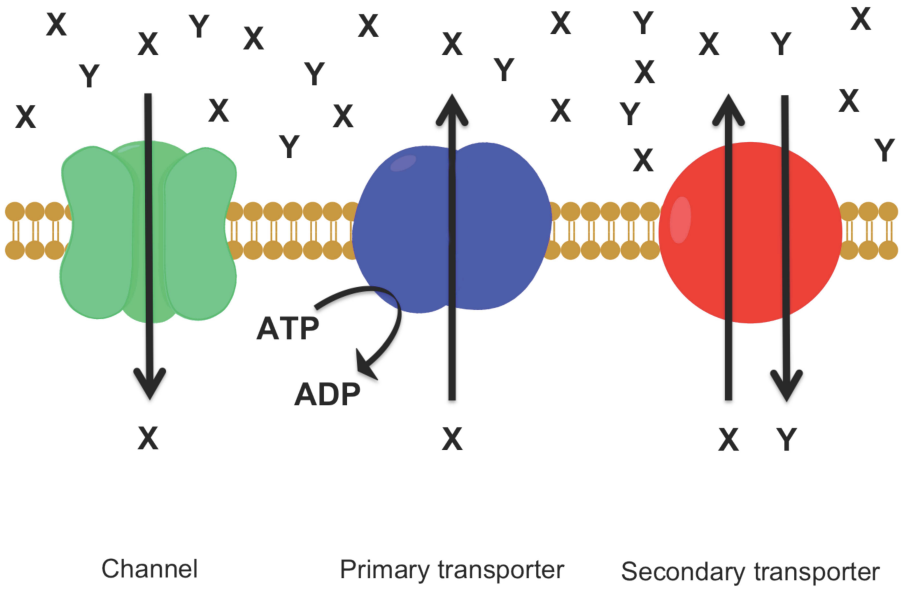


Figure 3-1: Representation of three classes of membrane proteins involved in the transport across biological membranes. Channels (left, green) mediate the passive diffusion of substrates (X) across biological membranes. Primary active transporters (middle, blue) utilize primary energy sources for the transport of substrates (e. g. hydrolysis of ATP to ADP and P_i), whereas secondary active transporters (right, red) couple the transport of substrates with the downhill movement of a second substrate (Y). Brown: lipid bilayer.

3.2 ABC transporters

According to the alternating access model of O. Jardetzky from 1966, membrane transporters cycle between two conformations with different affinities for the transported substrate, which is sufficient for transport. The inward-facing state is characterized by a high-affinity binding site for the substrate and the outward-facing state has a binding site with a lower affinity for the substrate [9]. Whereas in secondary active transporters these conformational changes are energized by the downhill translocation of a second substrate, the ATP-driven transport is a primary active process and the conformational changes are powered by ATP binding and hydrolysis [10]. ABC transporters form a huge family of ATP-dependent transporters found in all three kingdoms of life: bacteria, archaea and eukaryotes [11]. ABC transporters catalyze the transport of various substrates ranging from organic ions, amino acids, polysaccharides, peptides and proteins. In general, eukaryotes harbour ABC exporters and the human genome, for example, encodes 48 genes for ABC transporters [12]. Prokaryotes, in contrast, contain both importers and exporters [13] and in *E. coli* K12, 50 ABC importers and 15 ABC exporters are present [14]. In each case, ABC transporters have four core domains that

build up the minimal functional unit, two nucleotide-binding domains (NBDs) and two transmembrane domains (TMDs). Substrate binding proteins (SBP) are a fifth domain of the functional unit of importers. SBPs bind the substrate with high affinity and deliver it to the TMDs of the ABC transporter [15]. Worth mentioning is the presence of another class of ABC-dependent transport systems, which was recently discovered, the energy coupling factor (ECF) transporters. Here, the import depends on the presence of membrane-embedded proteins unrelated to SBPs [16-18]. However, here we focus on the ABC transporter class.

3.2.1 Nucleotide-binding domains

In ABC transporters, the hydrolysis of ATP is catalyzed by the highly conserved NBDs. As members of the 'P-loop' NTPases superfamily [19], NBDs contain the Walker A (consensus sequence: Gxx(G)xGKST) and Walker B (consensus sequence: ΦΦΦΦD, where Φ can be any hydrophobic amino acid) motifs, which were discovered in 1982 by Sir J. E. Walker (Nobel Prize in Chemistry, 1997) (Figure 3-2, A) [20]. The signature motif of ABC transporters, also called C-loop (LSGGQ), and a region around a conserved glutamine (Q-loop) are also characteristic motifs of NBDs [21-23]. Biochemical and structural studies of a related ATPase suggested the first mechanism for the catalytic cycle of NBDs [22], which was confirmed for isolated NBDs of ABC transporters in great detail [23-33]. The binding of ATP induces a dimerization of the NBDs in a 'head-to-tail' arrangement as observed in the HlyB-NBDs [34] (Figure 3-2, B and C). The nucleotide-binding sites are located at the dimer interface and are formed by the Walker A motif of monomer 1 and the C-loop of monomer 2. A water molecule in close proximity to the ATP is activated, either by deprotonation or polarization, and the activated water molecule attacks the γ-phosphate of ATP to initiate the cleavage of the β-γ-phosphodiester bond of ATP [28,35,36]. The aspartate of the Walker B motif coordinates an essential Mg^{2+} ion, which stabilizes the negative charge that develops on the γ-phosphate group. This lowers the energy of the transition state and facilitates the nucleophilic attack of the activated water molecule on the ATP γ-phosphate group [37-39]. After hydrolysis, electrostatic repulsion between ADP and inorganic phosphate (P_i) might destabilize the NBD dimer to restore the basal state as suggested for the NBD MJ0796 [26,34]. However, other mechanisms were postulated, for example for the NBDs of HlyB [28], and the precise process remains elusive.

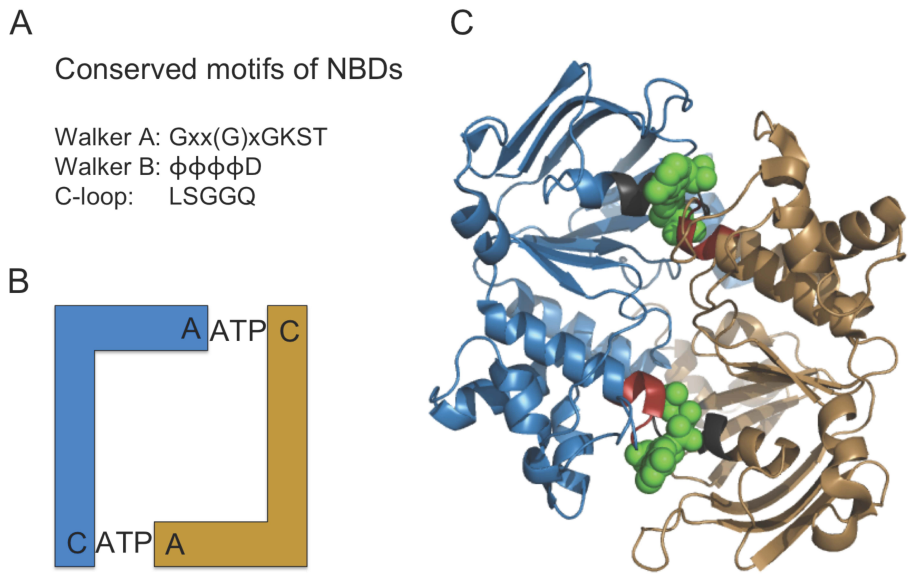


Figure 3-2: Conserved motifs of the NBDs are important for the ATP-dependent dimerization of NBDs. A: The Walker A, Walker B and C-loop of NBDs are highly, conserved motifs with the shown consensus sequences (G: glycine, x: any amino acid, K: lysine, S: serine, T: threonine, ϕ: any hydrophobic amino acid, D: aspartate, L: lysine, Q: glutamine). **B:** Schematic representation of the head-to-tail dimer of NBDs with bound ATP. A highlights the Walker A and C the C-loop motif. One monomer is shown in blue, the other in light brown. **C:** Two ATP molecules (green spheres) are bound at the dimer interface between the Walker A motif of one and the C-loop motif of the opposing monomer. The monomers of the HlyB-NBD dimer (PDB code 1XEF) are shown in blue and light brown. The Walker A and the C-loop motifs of the NBDs are highlighted in black and red, respectively [28]. The figure was prepared using PyMol (www.pymol.org).

The stoichiometry of ATP consumption to transported substrate was determined to be two, for example, for the glycine-betaine ABC importer OpuA from *Lactococcus lactis* [40] or the maltose and glycine-betaine import systems of *E. coli* [41]. These data are consistent with structural and biochemical data [42-46]. For other ABC exporters such as the peptide transporter TAP1/2 or the ABC-dependent ion channel CFTR, the hydrolysis of one ATP per transport cycle was suggested [47,48]. Therefore, it might be possible that one or two ATP molecules are consumed per catalytic cycle depending on the energy amount required to destabilize dimeric NBDs [49].

Although the coordinated process of coupling the catalytic cycle of the NBDs to the transport of a substrate is barely understood, recent studies of the maltose transporter of *E. coli* indicate that ATP binding represents the power-stroke of transport and that ATP hydrolysis is a chemical reaction to separate dimeric NBDs [35]. The Q-loop of the NBDs is known to be an important feature for this coupling [50-52]. The Q-loop is located at

the interface between the NBDs and TMDs [53] and transmits conformational changes to the TMDs, which is mandatory for transport [50-52,54].

3.2.2 Transmembrane domains

In addition to the NBDs, another canonical component of transporters is essential to drive the membrane transport of substrates, the TMDs [55]. TMDs are closed at any time on one side of the membrane to prevent passive diffusion of ions or molecules [13]. TMDs of different ABC transporters share little primary sequence homology, which represents an evolutionary adaption to the recognition and transport of various unrelated substrates [13,34,56]. Typically, TMDs are composed of two sets of six transmembrane helices, although the number of helices for importers varies from 2x5 to 2x10 [49].

The TMDs of importers contain a conserved sequence being located about 90– 100 amino acids away from their C-terminus, the EAA motif [57,58]. The short α-helical motif contacts and interacts directly with the adjacent NBD [13,52] and is called 'coupling helix 1' (CH1) [59] (Figure 3-3, A and B).

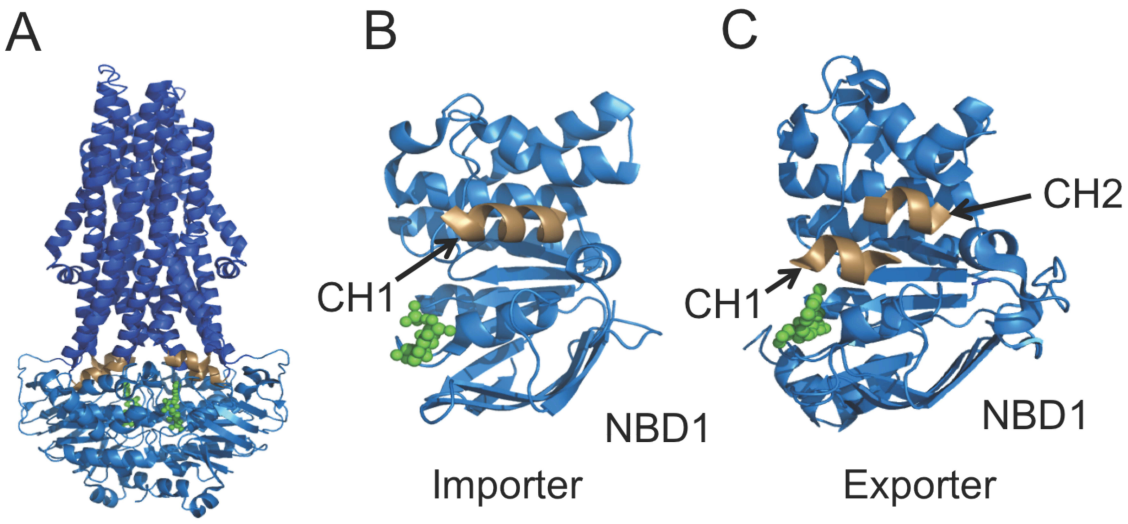


Figure 3-3: Arrangement of coupling helices of ABC importers and exporters at the NBDs. Type I and Type II ABC transporters are importers with one coupling helix (CH1). In contrast, Type III ABC transporters are exporters and contain two coupling helices (CH1 and CH2). Structurally, CH2 of exporters equals to CH1 of importers and interacts with residues in and around the Q-loop of the NBDs, however, CH2 of exporters originates from the opposite TMD of the transporter. CH1 represents a unique feature of exporters and interacts with both NBDs. **A:** Type III exporter Sav1866 of *Staphylococcus aureus* in an ADP-bound closed state (PDB code 2HYD). The coupling helices CH1 and CH2 are parts of the TMDs and are located at the interface of the TMDs and NBDs. **B:** Top-down view onto the single, ATP-bound NBD MalK from the maltose importer $MalFGK_2$ seen from the plasma membrane (PDB code 2R6G). Only the CH1 of the TMD is shown for clarity. **C:** Top-down view onto a single NBD of Sav1866, CH1 and CH2 of the TMDs and bound ADP seen from the plasma membrane (PDB code 2HYD). ADP and ATP: green spheres; coupling helices: light brown; NBDs: light blue; TMDs: dark blue. Figures were generated using PyMol (www.pymol.org).

The interaction sites of CH1 at the NBD are located in and around the Q-loop [50-52]. Exporters feature two major architectural differences compared to ABC importers. First, ABC exporters contain a second coupling helix, CH2 (Figure 3-3, C) that, in contrast to the CH1 of ABC importers, crosses over to contact the opposite NBD [49,59]. Despite this cross-over, CH2 of exporters is located at an equivalent position to CH1 of importers in and around the Q-loop of the NBDs [59]. CH1 of exporters is unique and interacts with both, the adjacent and the opposite NBD [59]. Secondly, the NBDs of exporters contain another conserved sequence, the X-loop, that interacts with the coupling helices and that is absent in importers [51,59]. Therefore, each TMD of ABC importers interacts solely with the adjacent NBD, whereas the TMDs of ABC exporters interact with both NBDs [49].
In importers and exporters, the interaction of the coupling helices with the NBDs seems to transmit a conformational change to the TMDs. These interactions are considered to couple the catalytic cylce of the NBDs to a movement of the TMDs that induces the conversion of the inward- to the outward-facing state of ABC transporters, and *vice versa* [59].

3.2.3 Mechanism of ABC transporters

Substrate binding to the inward-facing state of an ABC transporter inititates the transport by stimulating the dimerization of the ATP-bound NBDs [34,60-66]. For exporters, the substrate binds directly to the TMDs, whereas the substrate-loaded SBP binds to the TMDs of importers [13]. Dimerization of the NBDs generates the outward-facing state of the transporter [67]. The precise mechanism of coupling substrate binding and NBD dimerization to the transport across the TMDs is not completely understood [13]. However, evidences are emerging that the coupling helices of the TMDs fulfill this cross-talk between the NBDs and TMDs [42-44,68]. Via the interaction of the TMDs with the NBDs, the NBD dimerization is sensed by the TMDs and the transporter changes from the inward-facing to the outward-facing state. Subsequently, ABC exporters can extrude their substrate to the extracellular side of the membrane, whereas ABC importers now accept their substrates from the dedicated SBP [13,42-45]. In the next step, ATP is hydrolyzed, the NBD dimer dissociates and the TMDs are flipped from the outward-facing to the inward-facing conformation [22,34,35]. In import systems, substrates are presumably released to the cytoplasm in this step. The release of P_i and ADP from the NBDs restores the basal state of the transporter and completes one transport cycle [49].

Figure 3-4 illustrates the structural differences of the inward- and outward-facing conformations of ABC transporters.

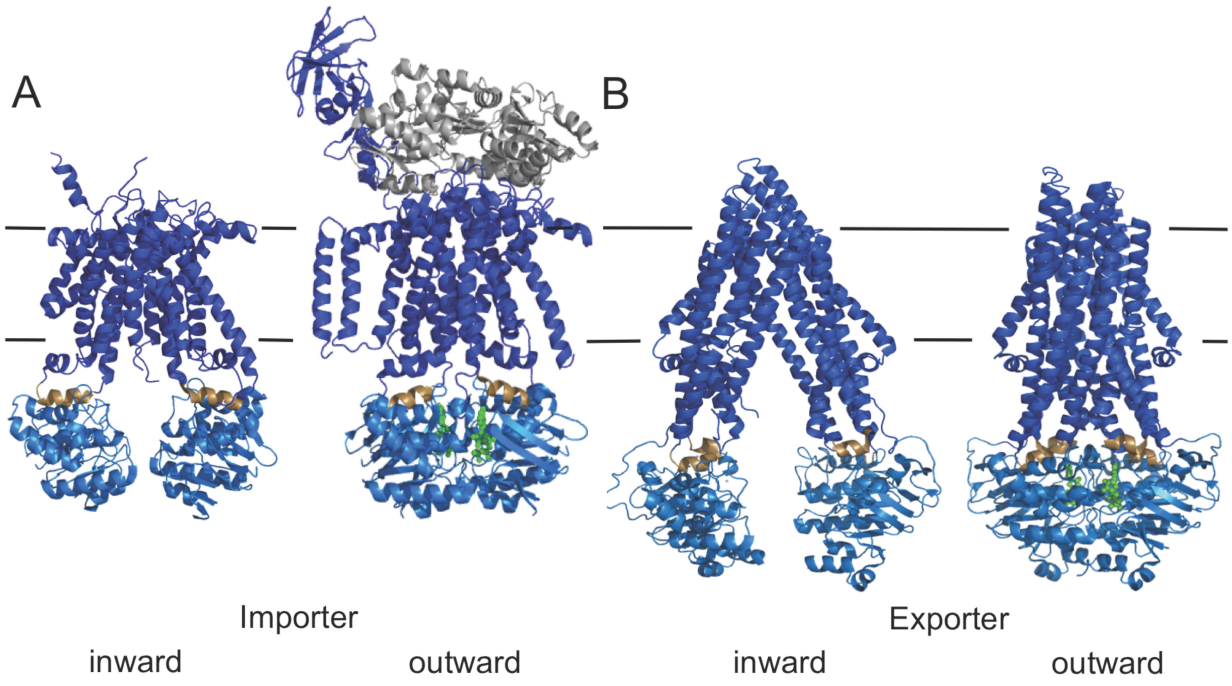


Figure 3-4: X-ray structures of ABC transporters in the inward- and outward-facing states. A: Inward-facing structure (left) and outward-facing structure (right) of the ABC importer maltose permease $MalFGK_2$ of *E. coli* (PDB codes 3FH6 and 2R6G). The outward-facing conformation was co-crystallized in complex with the SBP MalE and ATP. The regulatory domain of MalK was omitted for clarity. **B:** The multidrug resistance exporter P-glycoprotein in the inward-facing state (PDB code 3G5U, left) and the putative multidrug exporter Sav1866 of *Staphylococcus aureus* in an ADP-bound outward-facing state (PDB code 2HYD, right). ATP and ADP: green spheres; coupling helices: light brown; MalE: gray; NBDs: light blue; TMDs: dark blue. Black lines indicate the membrane. The figure was prepared using PyMol (www.pymol.org).

3.3 The cell wall of prokaryotes

Being integral components of the plasma membrane in prokaryotes and eukaryotes, ABC transporters catalyze the transport of various substrates across this barrier. In prokaryotes, the cell wall represents a second barrier that surrounds the plasma membrane and both together, the plasma membrane and the cell wall, build up the cell envelope of prokaryotes [69]. Based on the architecture of the cell wall, all bacteria are classified into two groups: Gram-positive and Gram-negative bacteria [69]. The cell wall of Gram-positive bacteria is a thick and undifferentiated layer consisting of peptidoglycan, (lipo)teichoic acids and surface proteins [70] (Figure 3-5).

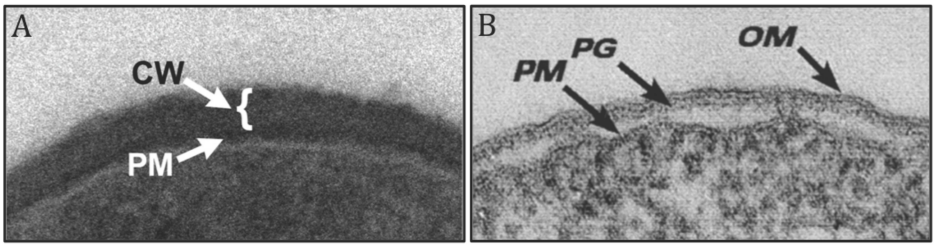


Figure 3-5: Thin sections by electron microscopy of Gram-positive and Gram-negative bacteria. Arrows indicate the components of the cell envelopes. **A:** Thin section of the Gram-positive bacterium *Staphylococcus aureus* (modified from Matias *et al*., 2006 [70]). **B:** Thin section of the Gram-negative bacterium *E. coli* K-12 (modified from Beveridge, 1999 [69]). CW: cell wall; PG: peptidoglycan layer; PM: plasma membrane; OM: outer membrane.

Gram-negative bacteria, in contrast, have a multi-layered cell wall with an outer membrane, a thin peptidoglycan layer and a periplasm. This cell wall and the plasma membrane (also called inner membrane) build up the cell envelope of Gram-negative bacteria. Several transport systems were evolved for the passage of substrates across the cell envelopes of bacteria. As focus of this doctoral thesis, the protein transport from the cytoplasm to the exterior of Gram-negative bacteria was investigated and this process is called secretion.

3.4 Secretion systems in Gram-negative bacteria

In addition to the transport across the plasma membrane, secretion systems of Gram-negative bacteria mediate the passage of substrates across the entire cell envelope including the cell wall [69]. In the Gram-negative bacterium *E. coli,* several different secretion systems are known, which are named Type 1 up to Type 6 secretion systems (Figure 3-6) [71,72]. Some of the systems export the secreted proteins in one step from the cytoplasm to the exterior, namely the Type 1, Type 3, Type 4 and Type 6 pathways. In other systems, proteins are first translocated into the periplasm by the universal Sec or the twin-arginine (Tat) pathway and then the Type 2 or Type 5 systems complete the secretion process [72]. Each of the six secretion systems is briefly described below and schematically shown in Figure 3-6.

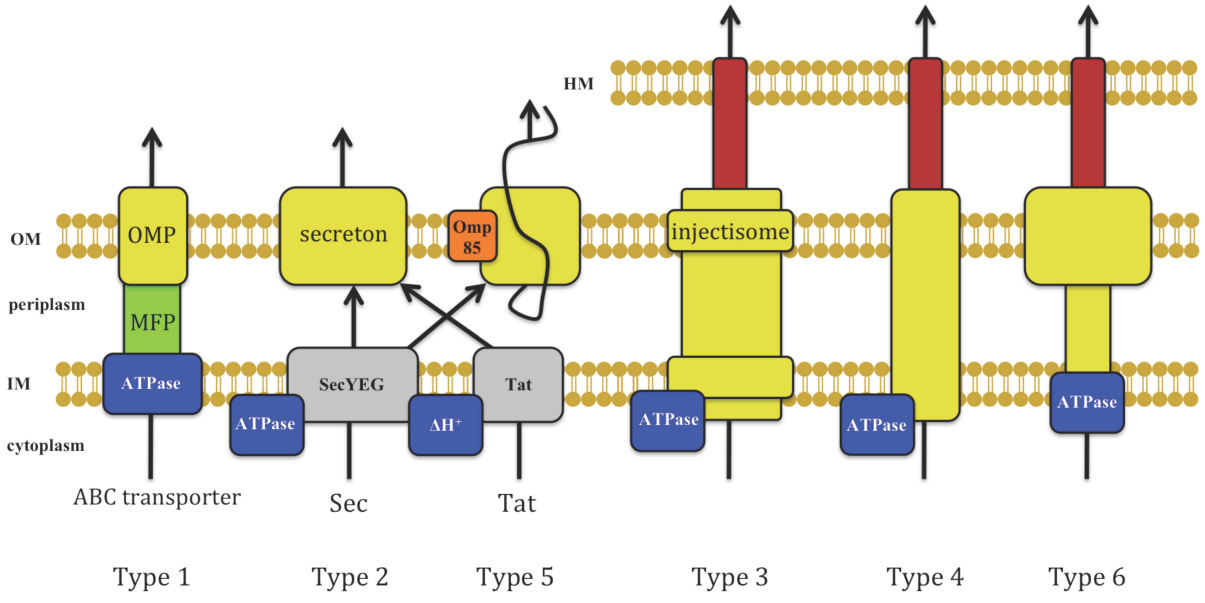


Figure 3-6: Scheme of known secretion systems in Gram-negative bacteria. Details are described in section 3.4. ATPases are labeled, ΔH^+ indicates an involvement of the proton motive force for transport. IM: inner membrane; MFP: membrane fusion protein; OM: outer membrane; OMP: outer membrane protein; HM: host membrane [72].

Type 1 secretion systems (T1SS) are composed of an ABC transporter, a membrane fusion protein (MFP) and an outer membrane protein (OMP) that build up a tripartite complex spanning the entire cell envelope. The ABC transporter energizes transport by ATP hydrolysis. The secretion signal of substrates is located at their C-terminus and is not cleaved during or after the transport [73].

Type 2 secretion systems (T2SS) involve a set of 12– 16 different proteins that constitute a large protein complex called secreton. After transport of the proteins into the periplasm by the Sec or Tat pathway, the T2SS translocates the periplasmic intermediates across the outer membrane. The energy source for the transport is not completely known, however, the proton motif force is suggested to be involved [74,75].

The Type 3 secretion system (T3SS) is a huge protein complex, which is termed injectisome. This machinery enables the injection of proteins directly into the cytoplasm of host cells. Up to 25 proteins are needed to assemble the functional injectisome and the secretion process is energized by ATP hydrolysis [76].

Type 4 secretion systems (T4SS) are composed of 11 or more subunits and use ATP to drive the transport. In addition to proteins, T4SS transfer nucleic acids into plant, animal and bacteria cells, which is unique for secretion systems [77].

Substrates of Type 5 secretion systems (T5SS) are translocated via the ATP-dependent

Sec or the secondary active Tat pathway into the periplasm. Here, their C-terminal β-domain inserts autocatalytically inside the outer membrane and facilitates the transport of the N-terminal passenger domain into the extracellular space. Because of this self-proceeding mechanism, T5SS substrates have been called autotransporters, however, recent evidences suggest an involvement of other proteins in the transport process, for example Omp85 [78].

Type 6 secretion systems (T6SS) facilitate the ATP-dependent transport of proteins directly into the cytoplasm of host cells, similar to T3SS or T4SS [79]. T6SS are suggested to mimic bacteriophage machineries to puncture target cells for the release of the cargo, however, various essential components have not been characterized at all and the mechanism is barely understood [80].

3.5 Type 1 secretion systems (T1SS)

Compared to other secretion systems of Gram-negative bateria that are assembled of numerous subunits and components, only three proteins build up a functional T1SS machinery. These three proteins are sufficient to catalyze the transport of proteins up to 900 kDa across the entire cell envelope [81]. T1SS are widespread among Gram-negative bacteria and catalyze the transport of unrelated proteins with different functions and molecular masses. The HasA T1SS of *Serratia marcescens*, for example, secretes a hemophore (HasA, 19 kDa) under iron starvation conditions to acquire heme [82], whereas a 900 kDa protein (LapA) is secreted by a T1SS of *Pseudomonas fluorescence*, which is involved in biofilm formation [81].

The secretion cycle of T1SS starts upon interaction of the substrate with the inner membrane complex consisting of the ABC transporter and the MFP [73,83,84]. This interaction recruits the OMP to complete the functional secretion apparatus and proteins are exported without any periplasmic intermediates to the exterior [73,83,84]. After transport, the substrate is released, the secretion machinery dissociates and the next secretion cycle can be initiated [83].

3.6 The haemolysin A (HlyA) T1SS of *E. coli*

Uropathogenic *E. coli* strains (UPEC) cause severe diseases in humans, for example acute and chronic urinary tract infections [85-87]. A major virulence factor of these strains is haemolysin A (HlyA), a haemolytic toxin that lysis cells of the host [88-90]. This toxin is

secreted by the HlyA T1SS (Figure 3-7) that represents the prototypical and currently best characterized T1SS [91]. Here, the ABC transporter haemolysin B (HlyB) and the MFP haemolysin D (HlyD) form a stable complex located in the inner membrane [83]. Upon interaction of the substrate with this complex, the OMP TolC is recruited and the functional secretion machinery is formed [92]. The deletion of each of the three membrane proteins totally abolishes the transport process [73,83,84]. HlyA is secreted in one step to the exterior and without the presence of any periplasmic intermediates. The dissociation of the secretion complex resets the system and completes one secretion cycle [73,83]. Except for the gene of TolC, the determinants of the HlyA T1SS are encoded on the *hly* operon. This operon is either located on the chromosome or on a conjugative plasmid [93] and its expression is tightly regulated [94-100].

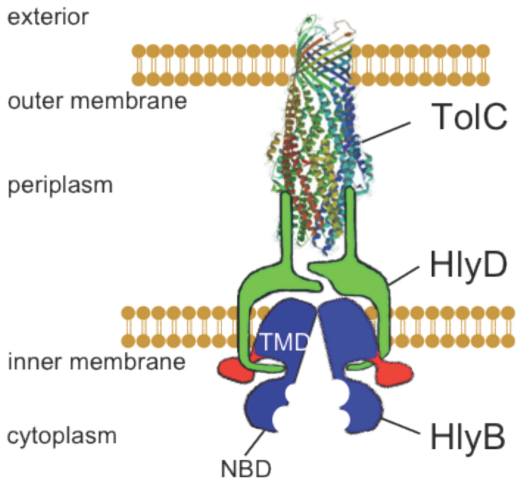


Figure 3-7: Representation of the assembled HlyA T1SS machinery. The dimeric ABC transporter HlyB with an N-terminal domain of unknown function (red), the TMD and the C-terminal NBD is shown; the membrane fusion protein HlyD with a large periplasmic domain, one transmembrane segment and an N-terminal cytoplasmic tail of 60 amino acids; TolC: outer membrane protein (PDB code 1EK9).

3.6.1 The toxin HlyA

HlyA is a major virulence factor of UPEC strains [86,101,102] that stimulates a number of responses in mammalian target cells [103], for example superoxide generation [104], Ca^{2+} oscillation [85], alteration of histone acetylation and phosphorylation patterns [105] and activation of MAP-kinase signaling [106]. HlyA also forms pores in the plasma membrane, which causes cell lysis and cell death [107,108]. The major region for the insertion into the membrane is located within the N-terminal, hydrophobic domain of HlyA (Figure 3-8).

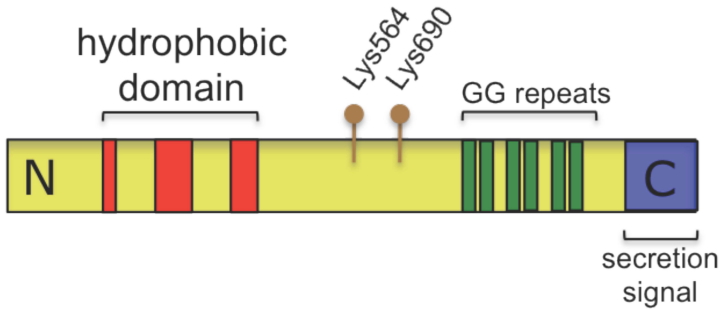


Figure 3-8: Schematic representation of the pore-forming toxin HlyA. HlyA has a hydrophobic N-terminal domain, which is important for the insertion into the host membrane. The C-terminal part contains the RTX domain of HlyA with several GG repeats and the secretion signal, which is located within the last C-terminal ~ 60 amino acids. HlyC acylates the two indicated lysine residues at position 564 and 690 prior to secretion to convert the non-toxic form into haemolytic HlyA. N: N-terminus; C: C-terminus.

Two posttranslational events are essential for the cytotoxicity of HlyA. First, the acyl transferase haemolysin C (HlyC) modifies intracellularly two specific lysine residues of HlyA (Lys564 and Lys690) and attaches fatty acids, for example myristic or palmitic acid [109-111]. Although essential for the cytotoxicity of HlyA, this modification does not influence the secretion process [110,112,113]. Secondly, the secreted HlyA binds Ca^{2+} ions at the C-terminal repeats-in-toxin (RTX) domain that is characterized by specific nona-peptide repeats of the consensus sequence GGxGxDxUx (G: glycine, D: aspartate, x: any amino acid residue, U: bulky hydrophobic residue), so-called GG repeats [114,115]. This RTX domain classifies HlyA as member of the RTX protein superfamily [116-118]. The GG repeats bind Ca^{2+} ions (dissociation constant, $K_D \approx 100$ µM [115]), which triggers the folding of HlyA into the biological active conformation [90,119,120]. Since the cytoplasmic Ca^{2+} concentration in *E. coli* is low (high nM range) [121], the GG repeats do not bind Ca^{2+} ions prior to secretion and RTX proteins are transported presumably in an unfolded state from the cytoplasm to the exterior [122-125]. Here, the Ca^{2+} concentration is high (up to 10 mM) [121] and binding of Ca^{2+} ions to the GG repeats induces folding [90,115,119].

Another characteristic of HlyA and all T1SS substrates is the presence of a C-terminal signal that is essential for secretion. This secretion signal is not cleaved during or after transport [126,127]. The precise role of the secretion signal during secretion is not understood, however, an interaction with the isolated NBDs of HlyB was demonstrated [128]. Since the secretion signal is located within the last C-terminal ~ 60 amino acids, T1SS substrates are suggested to be fully translated before the secretion starts and transport

occurs post-translationally [129-131]. How the unfolded, secretion-competent state of RTX proteins is maintained in the cells prior to secretion is unknown and, so far, no involvement of any chaperone during secretion was observed for HlyA [73]. In contrast, the chaperone SecA was shown to be essential for the secretion of the hemophore HasA by the dedicated T1SS of *Serratia marcescens* [132].

3.6.2 The outer membrane protein TolC

TolC is a member of the OMP family and has an important role in many processes of *E. coli*. Mutants lacking *tolC* have pleiotropic phenotypes including sensitivity to acids, increased susceptibility to antibiotics and decreased virulence [133]. As most important function, TolC is an integral part of several transport systems. Here, it interacts with several transport complexes of the inner membrane and forms a pore in the outer membrane to facilitate the export of various substrates, from small molecules to large proteins [133]. The constitutive expression of the *tolC* gene yields ~ 1500 protein copies per cell [134], however, the expression can be upregulated in response to the environment [135]. The crystal structure of TolC was solved in 2000 demonstrating that trimeric TolC assembles into a 140 Å long channel-tunnel with a diameter of about 35 Å [123] (Figure 3-9, A).

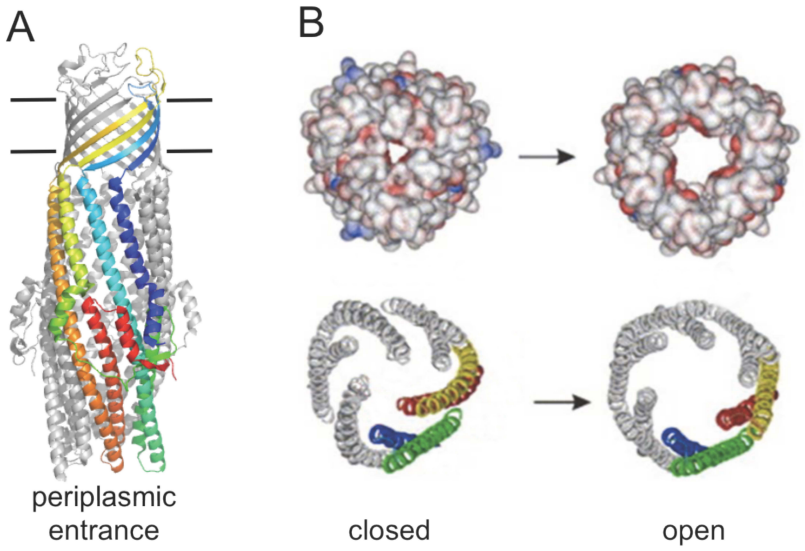


Figure 3-9: Structure and iris-like opening of TolC. A: TolC (PDB code 1EK9) assembles into a trimer and builds a pore in the outer membrane that is closed in the resting state. One protomer in the cartoon depiction is coloured in rainbow from blue (N-terminus) to red (C-terminus). The outer membrane is indicated with black lines. **B:** The closed state of TolC is opened upon interaction with a transport complex in the inner membrane. The conversion of the closed to the open state was modeled and is shown from the periplasmic site. Space-filled (upper) and ribbon (lower) depictions are shown. Helices of one protomer of trimeric TolC are coloured. The picture was modified from Koronakis, 2003 [136].

The overall structure of TolC can be subdivided into a β-barrel domain that spans the outer membrane and a periplasmic domain. The periplasmic domain is mainly build of α-helices and forms a cylinder spanning nearly 100 Å into the periplasm [123]. Whereas TolC is closed in the resting state, the interaction of TolC with a transport complex in the inner membrane is supposed to induce the opening by an iris-like mechanism. The opening of TolC could form a tunnel with a diameter of as much as 30 Å and substrates can be exported across the outer membrane [123,137,138] (Figure 3-9, B).

Members of the TolC family act in concert with primary active and secondary active transporters [123,139,140]. A prominent example of a secondary active transporter is AcrB from *E. coli*, which persists in the inner membrane and is driven by the proton motive force [141,142]. AcrB is part of a multidrug efflux system that extrudes cationic, neutral and anionic substances and confers an intrinsic resistance of *E. coli* against dyes, detergents and most lipophilic antibiotics [143]. In T1SS, primary active ABC transporters like HlyB energize the TolC-dependent secretion in an ATP-driven manner [10].

The connection of TolC to the transporter in the inner membrane is mediated by a MFP, which is an essential component of the functional secretion complex [133]. The MFP AcrA is part of the AcrAB-TolC system and HlyD completes the HlyA T1SS [83,134].

3.6.3 The membrane fusion protein HlyD

Members of the MFP family are found almost exclusively in Gram-negative bacteria [144]. MFPs are located predominantly in the periplasm and are essential components of various export systems in Gram-negative bacteria, for example multidrug, heavy-atom efflux systems or T1SS. For efflux systems, structural data are available indicating an oligomeric assembly of MFPs [140,145-147], however, the precise stoichiometry and the architecture of the MFPs remain elusive. The MFP AcrA of the AcrAB-TolC system, for example, assembles as a dimer of dimers in the asymmetric unit of the crystal, whereas soluble AcrA has been found to be monomeric [146,148]. Contradictorily, cross-linking experiments of AcrA *in vivo* suggest a trimeric form [149].

Although no structures are available for MFPs of T1SS, *in vivo* cross-linking experiments with HlyD suggest a trimeric assembly, which mirrors the symmetry of TolC [83,92]. Structural analysis with the related MFP of the ATP-dependent macrolide [150] efflux system MacAB-TolC of *E. coli* suggest a hexameric assembly [147], however, this is not a T1SS. Here, six protomers build a funnel-like structure with a central channel (Figure

3-10). The internal diameter of 30 Å equals the maximal diameter of the TolC tunnel. Thus, the OMP and the MFP of secretion systems physically define the diameter of the transport channel and limit the maximal diameter of potential transport substrates.

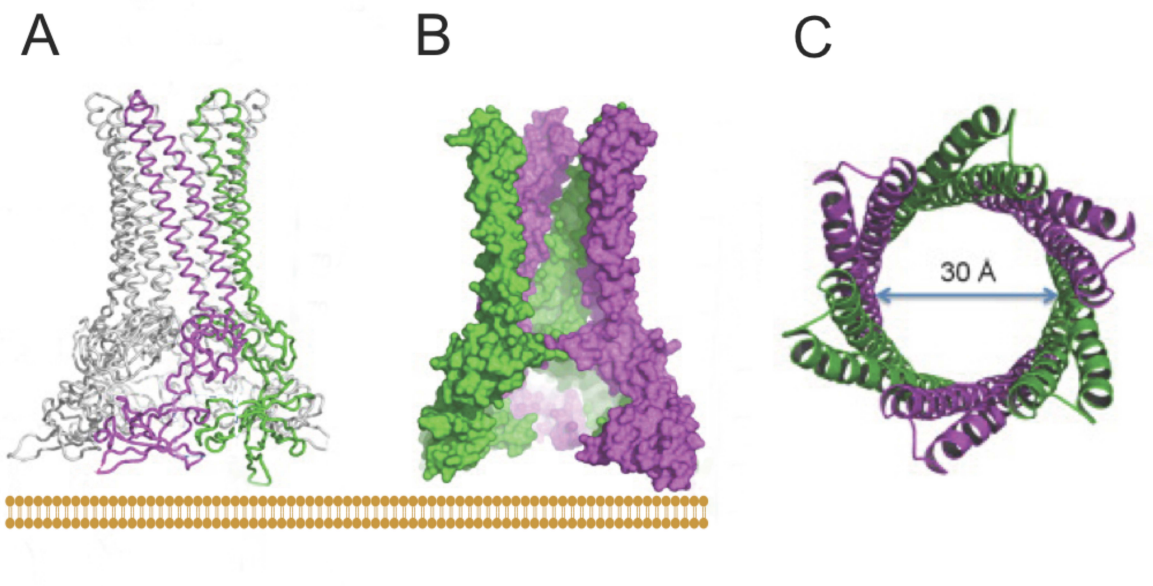


Figure 3-10: Structure of the hexameric MFP MacA of *E. coli*. A: Two MacA protomers are in the asymmetric unit (shown in green and magenta) and form by crystal symmetry a hexameric funnel-like structure with 2 additional MacA dimers (gray). **B:** The surface presentation of the MacA hexamer indicates the formed channel (the two front protomers were omitted for clarity). **C:** Hexameric MacA is viewed from the top and the internal diameter of 30 Å is indicated. For clearity, only the α-helical, upper part of MacA is shown. The inner membrane relative to MacA is indicated in light brown. The figure was modified from Yum *et al.*, 2009 [147].

Despite limited knowledge about the function, MFPs are known to connect the OMP to the transporter in the inner membrane. The formed complex enables the transport of substrates across the whole cell envelope of Gram-negative bacteria [10,83]. In addition to a periplasmic domain, HlyD carries a transmembrane segment at the N-terminus, which spans the inner membrane and enters the cytoplasm of the cell [10,151] (Figure 3-7). The cytoplasmic N-terminal tail of HlyD consists of 60 amino acids and is essential and specific for Type 1 export, whereby it fulfills a dual function: binding of the substrate HlyA and recruitment of TolC to the substrate-engaged inner membrane complex [92]. Independently of this tail, HlyD forms a stable complex with HlyB in the inner membrane [83,92]. For the MFP MacA, the transmembrane segment was shown to stimulate the ATPase activity of the ABC transporter MacB *in vitro* and this stimulation was

suggested to be essential for transport [152,153]. For HlyD, no comparable data are available since the overexpression of HlyD failed in the past [154,155].

3.6.4 The ABC transporter HlyB

All ABC transporters contain two essential modules for the transport, the NBDs and the TMDs. The NBDs bind and hydrolyze ATP, whereas the TMDs are forming the translocation pore and mediate transport specificity [27,34]. Detailed studies of the isolated NBDs of HlyB and other ABC transporters demonstrated that dimerization of the NBDs is essential for their ATPase activity [23-33] (see section 3.2). Since HlyB has a single NBD fused C-terminal to one TMD [73], the functional species of HlyB is suggested to be an oligomer, likely a homodimer. A detailed characterization of HlyB was hindered in the past mainly due to the abscence of a robust overexpression protocol for HlyB [155,156]. *In vivo* cross-linking experiments with a reconstituted HlyA export system demonstrated that HlyB forms a stable inner membrane complex with HlyD and that HlyB interacts with the substrate HlyA [83]. An interaction of HlyB and HlyA was further supported by secretion analysis with mutated versions of HlyB. Here, specific mutations in HlyB compensated for secretion defects that were caused by mutations in the secretion signal of HlyA [157,158]. Surface plasmon resonance (SPR) spectroscopy experiments revealed an interaction of the isolated NBD of HlyB with a C-terminal fragment of HlyA, called HlyA1. Importantly, this interaction was strictly dependend on the presence of the C-terminal secretion signal of HlyA and the addition of ATP accelerated the disruption of the NBD-HlyA1 complex. The disruption was not observed in the presence of the non-hydrolyzable ATP analogon AMP-PNP suggesting that ATP hydrolysis is essential for the dissociation of the NBD-HlyA1 complex [128]. The interaction of HlyA with HlyB and HlyD is essential to recruit TolC, which completes the functional secretion complex [92,159]. Topology studies with HlyB indicated the presence of eight transmembrane segments, however, the two most proximal, N-terminal transmembrane segments are poorly predicted and are located within a non-canonical region of ABC transporters [160,161]. The function of this N-terminal extension of around 150 amino acids – corresponding to about 20 % of the entire primary sequence of HlyB – is unknown [73,156].

3.6.5 The C39 peptidase-like domain of HlyB

A sequence alignment of the N-terminal region of HlyB revealed a homology to C39 peptidases, which are cysteine proteases of the papain superfamily [162]. C39 peptidases are only found as N-terminal domains in ABC transporters (Figure 3-11) that are involved in the secretion of antibacterial peptides called bacteriocins [163].

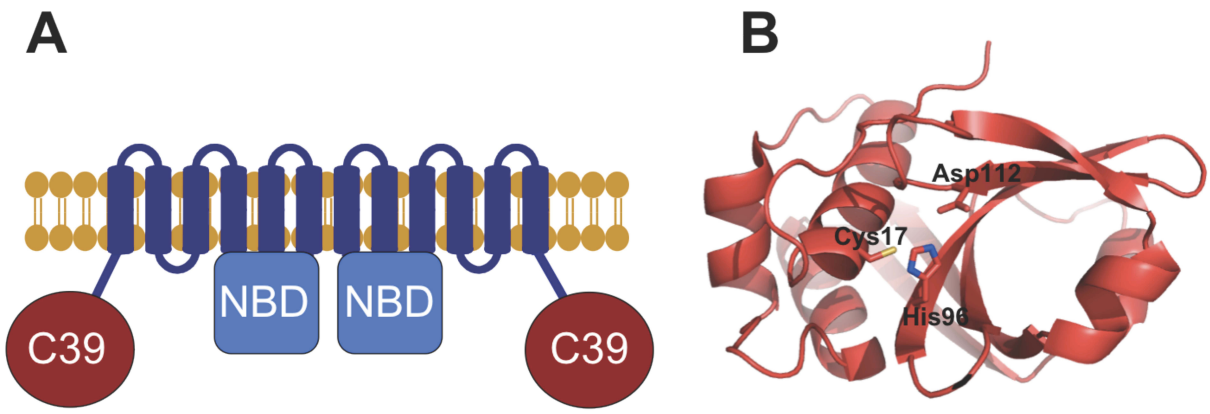


Figure 3-11: Schematic representation of bacteriocin ABC transporters and the X-ray structure of the C39 peptidase ComA-PEP. A: In addition to the canonical TMD and NBD, bacteriocin ABC transporters carry an extra N-terminal C39 peptidase domain. This C39 peptidase recognizes a conserved stretch of 15 amino acids, the double-glycine motif of the bacteriocin, and cleaves the pro-peptide proteolytically to produce the mature bacteriocin. **B:** X-ray structure of the C39 peptidase ComA-PEP (PDB code 3K8U). The overall structure is shown as cartoon, the catalytic triad consisting of Cys17, His96 and Asp112 is shown as sticks with sulfur, oxygen and nitrogen highlighted in yellow, red and blue, respectively. The picture was prepared with PyMol (www.pymol.org). C39: C39 peptidase; dark blue: TMDs; light brown: plasma membrane; NBD: nucleotide-binding domain.

The N-terminal C39 peptidase cleaves the precursor peptide C-terminal to a conserved stretch of 15 amino acids, the double-glycine motif, which produces mature bacteriocin [163-166]. As member of the cysteine protease superfamily, the active site of C39 peptidases is a catalytic triade consisting of cysteine, histidine and aspartate [162,167] (Figure 3-11, B). The imidazole group of the histidine facilitates the deprotonation of the sulfhydryl group of the cysteine. The protonated imidazole group is stabilized by an ionic bond with the carboxylate side chain of the aspartate. The thiolate anion of the cysteine subsequently attacks the carbonyl group of the cleavage site and initiates proteolysis [168]. Sequence alignments of the N-terminal domain of HlyB indicated the abscence of the catalytic essential cysteine of such proteases [164]. Furthermore, HlyA does not contain the conserved double-glycine motif of bacteriocins and is not cleaved during the secretion process [126,127]. Obviously, the N-terminal extension of HlyB renders a different, unknown

function that differs from proteolysis. Therefore, the domain was named "C39 peptidase-like domain" (CLD).

3.7 Application of the HlyA T1SS in biotechnology

The secretion of recombinant proteins into the exterior of cells offers several advantages compared to the production in the cytoplasm of cells [169,170], for example, increased biological activity [171] or product stability [125,172,173]. The initial product purity is also very high as many bacteria, for example *E. coli* K-12 or B, do not secrete proteins endogenously [170,174]. This simplifies downstream processing that currently accounts for up to 80 % of the entire production costs [175]. Since decades, the HlyA T1SS of *E. coli* was analyzed for its ability to secrete recombinant proteins [176]. Proteins-of-interest were fused in front of C-terminal fragments of HlyA and co-expressed with the ABC transporter HlyB and the MFP HlyD. The OMP TolC was expressed constitutively by the used *E. coli* strains. Several fusion proteins could be secreted with this rational, however, many were secreted poorly or failed to be secreted [125,169,177,178]. In a previous study, the secretion efficiency correlated inversely with the size of the passenger moiety. This suggested that the size or the folding state of a passenger protein influences the secretion process [122]. This suggestion was analyzed in more detail with the maltose binding protein (MBP) of *E. coli*. Wild-type MBP was initially not secreted, however, MBP mutants with reduced folding kinetics were secreted efficiently [124]. This study demonstrated an inverse correlation of the folding velocity of passenger proteins and the secretion efficiencies and offered a concept to secrete MBP and likely other proteins, the "slow-folding concept". Structural studies with the outer membrane protein TolC provide an explanation for this finding [123]. Trimeric TolC forms a channel with an internal diameter of maximal 30 Å in the open state [123], which physically limits the size of the secretion channel and accomodates only the secretion of small, unfolded or partially folded proteins [82]. The T1SS substrate HasA, for example, consists of 188 amino acids and is secreted efficiently by the dedicated T1SS in its unfolded state, whereas folded HasA is not secreted [179]. Due to these limitations, the most important application of T1SS is currently the presentation of heterologous antigens by live-vaccines [91]. Here, proteins of a pathogen (antigen) are directed to the HlyA secretion system as fusion protein with a C-terminal part of HlyA. Gram-negative bacteria that secrete the fusion protein, for example non-pathogenic *Salmonella* strains, are subsequently injected as

vaccines. The secreted fusion protein induces a protective immunity against the original, pathogen host of the antigen. If fusion proteins are not secreted by the HlyA T1SS, domains of the antigen might be used that can be secreted and that subsequently induce an immune response [180]. Since domains of proteins are in principle secretable and the size of a passenger protein limits the secretion efficiency, the HlyA T1SS might also accomodate the secretion of peptides, whose production with recombinant techniques currently fail in most cases.

3.8 Production strategies for peptides

Peptides are composed of amino acids, conventionally of less than 75– 100 of these building blocks [181,182]. Currently, the recombinant production of peptides in large quantities fails in many cases because of some inherent properties of peptides: 1. they are characterized by the abscence of tertiary structures, which induces an intrinsic lability and instability that directs them to proteolytic degradation, 2. they tend to form insoluble aggregates and 3. they often exhibit a biological function, which might be harmful to the production host [183-185]. Despite the development of strategies for the recombinant production of some specific peptides (see for example [183,184,186,187]), the great majority is currently produced by chemical synthesis (Figure 3-12). Only this method permits the production of large quantities of pure peptides and also the incorporation of, for example, unnatural amino acids [188]. The most common strategy is the solid-phase peptide synthesis (SPPS) invented in 1963 by R. B. Merrifield (Nobel Prize in Chemistry, 1984). A first amino acid is covalently linked to a solid support, protected amino acids are coupled stepwise until the peptide of interest is assembled, and the product is released [189]. In many cases, SPPS is a very fast and efficient method for the synthesis of short peptides up to 15 amino acids, however, it has limitations for the production of larger peptides [181]. Especially the preparation of highly pure peptides ≥ 15 amino acids is hampered due to incomplete coupling reactions and side-reactions like additions, deletions and epimerization during the coupling and deprotection reactions [188,190]. Moreover, the production yields drop substantially with the length of the peptide [181]. Assuming an efficiency of 99 % for each reaction cycle, the synthesis of a peptide with 15 amino acids would yield 86 % (0.99^{15}) of the desired product, whereas the yield drops to 75 % for a peptide of 30 amino acids (0.99^{30}). Furthermore, the purification process becomes increasingly difficult for larger products [181]. Since 1994,

the 'native chemical ligation' offers a new avenue to produce larger peptides and even proteins by chemistry [191]. Unprotected peptides, still produced by SPPS, are fused to each other in aqueous solution by a chemoselective reaction, for example, an N-terminal cysteine reacts with a C-terminal thioester moiety and forms a peptide bond at the ligation site [188]. The preparation of the 203 amino acid protein HIV-1 protease indicates the potential of this approach [192].

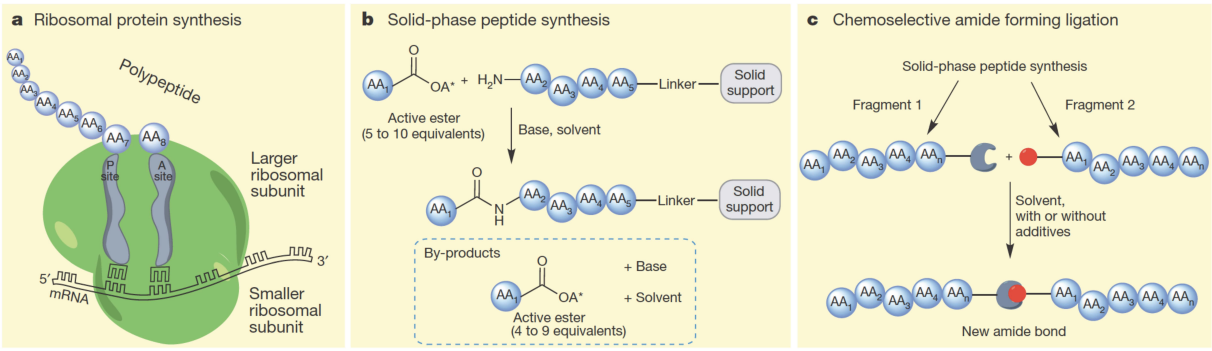


Figure 3-12: Representation of biochemical and chemical strategies for the formation of peptide bonds. A: During the ribosomal synthesis of polypeptides, an aminoacyl tRNA with an unprotected amino acid (blue spheres, AA_x) enters to the A site (gray, right) and the peptide bond is formed with the C-terminal carboxy group of the nascent polypeptide at the P site (gray, left). mRNA: messanger RNA. **B:** In solid-phase peptide synthesis (SPPS), an amino acid is immobilized on a solid support. Active esters of fully protected amino acid monomers in combination with coupling reagents are necessary for the amide linkage to the free amino group of the growing peptide. **C:** In the chemoselective amide forming ligation, two separated peptide fragments are typically generated by SPPS. Both fragments are fused subsequently via two specific reactive functional groups (filled red circle and matching gray shape), which assembles larger peptides or even proteins. The figure was taken from Pattabiraman and Bode, 2011 [190].

Despite the progress in peptide chemistry during the last century, all current synthetic methods are regarded as rather expensive, inelegant and generate huge amounts of chemical waste [190]. For example, the large-scale manufacture of the HIV drug Fuzeon of 36 amino acids is currently the best-established production process for a large peptide and still consumes 45,000 kg raw materials for the production of 1,000 kg Fuzeon, not including the solvents used during synthesis and purification [181,190,193]. According to the "American Chemical Society Green Chemistry Institute" in 2007, the top challenge of organic chemistry is the establishment of methods to avoid poor atom economy reagents for amide formation [194], which illustrates the increasing demand for new production strategies.

4 Aims

4.1 Characterization of the HlyA T1SS

As major virulence factor of pathogenic UPEC strains the toxin HlyA causes severe diseases in humans [88]. These bacteria produce HlyA inside the cells and release it into the surrounding, where its cytotoxic activity harms various cell types [88]. The transport of the 110 kDa HlyA across the Gram-negative cell envelope of *E. coli* is catalyzed by the HlyA transport system, which represents the paradigm of T1SS. The functional T1SS is a tripartite complex of the ABC transporter HlyB, the MFP HlyD and the OMP TolC. HlyB and HlyD form a stable inner membrane complex that interacts with HlyA. This interaction recruits TolC and induces the assembly of the functional secretion apparatus [83]. Unfolded HlyA is transported in one step into the exterior and the binding of Ca^{2+} ions to the RTX domain induces the folding of HlyA into its functional conformation [90,119,120]. After the transport of HlyA, TolC dissociates from the inner membrane complex and the T1SS machinery is reset. Despite the current knowledge about the secretion process, many aspects of the transport process are not understood and the precise mechanism for the secretion of proteins up to 900 kDa remains elusive.

A first target of this doctoral thesis was the N-terminal extension of the ABC transporter HlyB with unknown function. For the assignment of a function, this domain should be produced in large quantities and in a homogeneous state. Subsequently, this domain should be characterized in great detail by structural and biochemical approaches.

4.2 Biotechnological Application of the HlyA T1SS

Currently, the application of the HlyA T1SS as production platform for proteins is hampered since many fusion proteins are secreted in low amounts or not at all [125]. A previous study suggested a concept to circumvent this limitation, the "slow-folding concept" [124]. As MBP, which was used for the mentioned study, is a periplasmic protein of *E. coli* that is normally translocated by the Sec pathway, it endows intrinsic properties to pass biological membranes.

Therefore, a second task of this doctoral thesis was the validation of the "slow-folding concept" for a cytoplasmic protein with eukaryotic origin. In addition, the folding state

and the biological activity of the secreted protein should be analyzed during this doctoral thesis.

Even if the "slow-folding concept" can be applied successfully for the production of all proteins, slow-folding kinetics of the proteins are a prerequisite still impairing the industrial application of the HlyA T1SS. Many proteins simply do not fold slowly and the generation of proteins with decreased folding kinetics, for example by directed evolution strategies, is lab-, time- and cost-intensive. Other available secretion systems do not offer an universal alternative, since the secretion of arbitrary target proteins is always a trial-and-error approach and the secretion efficiency depends on several parameters, for example the used signal peptide or the expression host [195].

Therefore, a third task of this doctoral thesis was the improvement of the HlyA T1SS to allow the secretory production of, in the optimal case, all proteins and without the insertion of slow-folding mutations. The applicability of the HlyA T1SS for the recombinant production of peptides should also be evaluated as a fourth task.

5 Publications

5.1 Chapter I

Title:	RTX toxin transporters tether its substrate prior to secretion via the unique function of its N-terminal domain
Authors:	Lecher J.*, Schwarz C. K. W.*, Stoldt M., Smits S. H. J., Willbold D. and Schmitt L.
	*: These authors contributed equally.
Published in:	Structure (submitted)
Impact Factor:	6.337
Own proportion of this work:	40 %
	Secretion analysis.
	Production, purification and biochemical characterization of the CLD.
	Production, purification and biochemical characterization of HlyA, HlyA1 and HlyA2.
	Bioinformatical analysis.
	Writing of the manuscript.

From: Lutz Schmitt <Lutz.Schmitt@hhu.de>
Date: July 30, 2012 2:59:11 PM GMT+02:00
To: Christian Schwarz <christianschwarz2@gmx.de>
Subject: **Fwd: Structure Decision**

Anfang der weitergeleiteten E-Mail:
Von: Structure Editorial Office <structure@cell.com>
Datum: 18. Juli 2012 20:15:12 MESZ
An: Lutz.Schmitt@uni-duesseldorf.de
Betreff: **Structure Decision**

Lutz Schmitt, Ph.D.
Heinrich Heine University
Institute of Biochemistry
Universitaetsstr. 1
Dueseldorf 40225
GERMANY

Jul 18, 2012

Manuscript: STRUCTURE-D-12-00221
Title: RTX toxin transporters tether its substrate prior to secretion via the unique function of its N- terminal domain

Dear Dr. Schmitt,

Thank you for submitting this paper to Structure. We have now received comments on it from two referees, which I am including with this letter. As you will see, both referees support the publication of the paper, although there are some comments that require attention. We would therefore ask you to revise your manuscript in the light of these comments, and we would expect a suitably revised manuscript to be accepted for publication. Please be sure to address all the points in full.

While you are revising your manuscript, please attend to the following editorial points to help expedite the publication of your manuscript.

Abstract: Please make sure that the abstract does not exceed 150 words.

Introduction: Please include a brief summary of your key findings in the final paragraph of the Introduction section.

Figures and Tables: Structure allows a maximum of ten display items (figures and tables), so please do not exceed this number. For papers describing new structures to atomic resolution, one figure should show the whole molecule or a set of fragments in sufficient detail so that the locations of individual residues can be identified. The Table containing X-Ray data collection, phasing, and refinement

statistics, or the Table containing NMR and refinement statistics MUST be included in the main text.

Also note that all figures' and tables' legends in the main text need to include the information about what supplementary information display items they are supported by. The same applies for the supplementary figures' and tables' legends that need to specify which main text item they are supporting. More details about the organization of Supplementary Information can be found below. Please keep in mind this information while preparing the final version of your article.

Data deposition: Atomic coordinates, EM data, NMR assignments, nucleic acid sequences, and protein sequences should be deposited in an appropriate databank, and the entry code or accession number must be cited.

References: All in-text citations and listed references should be set in Cell Press style: in-text citations should include first author name and year of publication (Smith et al., 2003), and the reference list should appear in alphabetical order. For examples of the new format, please visit www.cell.com/structure/authors#specific.

Hightlights & Graphical Abstracts: From January 2010, Cell Press journals will include two new features in the online Summary section of articles - Graphical Abstracts and Highlights. Highlights are required for articles published from January 2010 onward. Graphical abstracts are optional. Examples of these new features can be seen in the prototypes at http://beta.cell.com/index.php/2009/07/article-of-the-future/. Details about preparing these items are available here: www.cell.com/structure/highlights.

Supplemental Information: Cell Press is implementing a new approach to Supplemental Information. Please read our new Supplemental Information Guide at www.cell.com/structure/supplemental_information_guide.

Overall length: Papers should not exceed 55,000 characters, including spaces. Failure to meet the length limit may result in the return of your manuscript for further revisions, delaying publication.

Please send your revised manuscript via EES as quickly as possible with a cover letter stating clearly how the text has been changed in response to the comments or giving your reasons for not making some of the suggested changes. If your paper is accepted for publication, you will be asked to upload a final submission via EES.

Please do not hesitate to contact us if you have any problems or questions.

Kind Regards,

Andrej Sali, Editor
sali@salilab.org

On behalf of the editors

Alana Melton
Editorial Assistant, Structure
600 Technology Square, 5th Floor
Cambridge, MA 02139
Phone: 617-386-2128
Fax: 617-397-2810
structure@cell.com

DID YOU KNOW?
Structure now publishes all accepted papers on-line ahead of print.

Reviewer #1: Lecher et al. describe in their manuscript the structural and functional characterization of the C39-like domain of HylB. These domains are usually used to cleave the sinalling peptide, but the C39-like domain of HylB is proteolytically inactive. Instead it seems to bind to the unfolded form of the transport substrate HylA. The authors determine the NMR structure and map the interaction site with the substrate.
Overall, these are interesting results with relevance for the entire (ABC) transporter field. The paper is clearly written ans should be accepted for publication.
One minor comment:
In figure 2 the protein band of DeltaCLD HylB runs below the band of HylD, although it is larger. Is there any explanation?

Reviewer #2: This is a very nice and pathbreaking work on hemolysin secretion by T1SS. It shows that an N-terminal domain named CLD of the ABC protein ,part of the T1SS apparatus, is required for hemolysin secretion. The NMR structure of CLD is produced. The work shows that CLD interacts with two peptides of hemolysin : one (HlyA1) corresponding to AA 807 to 1023 and the second (HlyA2) corresponding to AA 807-966 that lacks the C-terminal so called secretion signal. Both contain 3 GGXGXD repeats, a motif conserved in this family of substrates secreted by T1SS. Site directed mutagenesis of the CLD shows the residues involved in hemolysin secretion.
I have only few comments/
1) The title : « RTX Toxin transporters tether its substrate prior to secretion »
And the highlights : »Substrate recognition by the C39 like domain is likely the first step of T1SS « do not correspond neither to the results nor even to a discussion :
I don't se in the manuscript any experiment showing that this is a first step. As it is presented it could be two independent concommittent interactions.
If the authors want to keep this statement, it would be good to show for example that
the LCD domain HlyA2 interaction induces the TolC recruitment in the T1SS.

2) In the structure of the ComA-PEP, the C39 peptidase domain interacts with the GG motif of the ComC substrate (Ishii JBC 2010). Is there any indication whether

the GGXGXG motif is involved in interaction with the CLD ? At least this point could be discussed.

3) In the introduction, it would be helpfull to give some information about the C39 peptidase ABC transporters. The description is given in the discussion, but for a reader who is not familiar with the system, it might be confusing.

4) The experiment with folded and unfolded HlyA is crucial. Yet, there is no information on the procedure of purification or on the state of the folded protein, besides the fact that the folded form does not bind calcium. The folded form could simply be agregated.
May be it would be nice to take the unfolded form incubated with Ca and show that after incubation it binds CLD or the opposite incubate the folded form with EGTA and test the binding.

5) The stability of the Y20 and F99 HlyB mutants which do not secrete hemolysin should be tested by WB with anti-HlyB antibodies.

6) The S1 figure should be in the paper and should be revealed by WB with anti-HlyA the pellets also.

RTX toxin transporters tether its substrate prior to secretion via the unique function of its N- terminal domain

Justin Lecher[1,2]*, Christian KW Schwarz[3]*, Matthias Stoldt[1,2], Sander HJ Smits[3], Dieter Willbold[1,2] and Lutz Schmitt[3#]

[1]Institute of Physical Biology, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany;

[2]Institute of Structural Biochemistry (ICS-6), Research Centre Jülich, 52425 Jülich, Germany;

[3]Institute of Biochemistry, Heinrich-Heine-Universität, Universitätsstr. 1, 40225 Düsseldorf, Germany;

- These authors contributed equally

To whom correspondence should be addressed:

Dieter Willbold, E-mail: d.willbold@fz-juelich.de

Lutz Schmitt, E-mail: lutz.schmitt@hhu.de

Running title: An ABC transporter with dual function

Summary

Type 1 secretion systems (T1SS) catalyze the one-step protein transport across the membranes of Gram-negative bacteria and are composed of an outer membrane protein, a membrane fusion protein and an ABC transporter. The homodimeric ABC transporter consist of the canonical nucleotide binding and transmembrane domains. For the toxin haemolysin A (HlyA), the ABC transporter HlyB carries an additional, N-terminal domain sharing about 40% homology to C39 peptidases, but this 'C39-like domain' (CLD) is suggested to feature another, yet unknown function. Our functional and structural analysis demonstrates that the CLD is essential for secretion and that it specifically interacts with the unfolded state of HlyA. We determined the NMR structure of the CLD as well as the substrate-binding region within the CLD. This mode of action represents a novel mechanism within T1SS and answers the question, how a large and unfolded substrate is protected inside the cells prior to secretion.

Keywords

Bacteriocins, GG repeats, Haemolysin B, RTX toxin, TolC

Abbreviations

ABC: ATP binding cassette; CLD: C39 peptidase-like domain; CSP: chemical shift perturbation; GG motif: double glycine motif; Hly: Haemolysin; HSQC: heteronuclear single quantum coherence; MFP: membrane fusion protein; NOE: nuclear Overhauser effect; RTX: repeat-in-toxins; T1SS: Type 1 secretion system; wt: wild-type.

Introduction

Type 1 secretion systems (T1SS) are widespread among Gram-negative bacteria (Delepelaire, 2004; Holland et al., 2005) and catalyze the one-step translocation of polypeptides across both, the inner and outer membranes, directly to the exterior. The transported substrates include toxins, lipases, haem-binding and S-layer proteins, which are presumably transported in an unfolded state (Bakkes et al., 2010; Holland et al., 2005). A paradigm of such substrates is haemolysin A (HlyA), a 110 kDa large toxin of the **r**epeat-in-**tox**ins (RTX) superfamily that lyses, for example, mammalian erythrocytes (Linggood and Ingram, 1982). The hallmark of HlyA and other RTX toxins are the RTX domains that contain glycine- and aspartate-rich nonapeptide repeats with the consensus sequence GGxGxDxUx (x: any amino acid, U: large, hydrophobic amino acid), the so-called GG repeats. Ca^{2+} ions are bound by these GG repeats with high affinity, which triggers folding *in vitro*, most likely after secretion of the T1SS substrate into the environment (Baumann et al., 1993). Depending on the substrate, which is transported, the number of GG repeats varies from less than 10 to more than 40 (Linhartova et al., 2010). Another characteristic of all T1SS substrates is the C-terminal signal that is essential for secretion and is not cleaved during or after secretion. In the case of HlyA, it is encoded within the last 50– 60 C-terminal amino acids (Ghigo and Wandersman, 1994; Kenny et al., 1994; Kenny et al., 1992). In the past, a C-terminal fragment of HlyA, HlyA1, comprising residues 807- 1024 has been used as a surrogate (Benabdelhak et al., 2003). HlyA1 contains the C-terminal secretion sequence, three conserved GG repeats and is secreted as efficiently as HlyA. Furthermore, its folding is strictly Ca^{2+}-dependent (Sanchez-Magraner et al., 2010; Sanchez-Magraner et al., 2007).

In general, T1SS are composed of three membrane proteins, an ATP-binding cassette (ABC) transporter, a membrane fusion protein (MFP), and an outer membrane protein (Holland et al., 2005). For the HlyA T1SS, these proteins are the ABC transporter HlyB, the MFP HlyD and the outer membrane protein TolC (Koronakis et al., 2000; Letoffe et al., 1990). HlyB and HlyD form a stable complex in the inner membrane and recruit TolC only in the presence of the substrate HlyA (Thanabalu et al., 1998). The functional complex forms a tunnel across the periplasm, which guarantees the continuous translocation of likely unfolded HlyA without any periplasmic intermediate (Bakkes et al., 2010; Kenny et al., 1992; Koronakis et al., 2000; Thanabalu et al., 1998). In the presence of the acyl carrier protein HlyC, two specific lysine residues of HlyA are acylated prior to secretion, converting non-toxic HlyA into the active HlyA (Stanley et al., 1994). However, this posttranslational modification does not influence the secretion process (Nicaud et al., 1985).

Canonical ABC transporters are composed of four modules, two transmembrane domains (TMD) and two nucleotide-binding domains (NBD). HlyB, however, contains an additional N-terminal extension of 123 amino acids. Sequence alignments revealed a homology of 42% to C39 peptidase domains of ABC transporters that cleave dedicated substrates after an N-terminal leader peptide, called the GG motif (Havarstein et al., 1995). However, the catalytically essential cysteine of C39 peptidases is absent. Moreover, HlyA does not contain an N-terminal leader peptide and is not cleaved during or after the transport. Obviously, the N-terminal domain of HlyB cannot fulfill an equivalent function and is therefore called ‘C39-like domain’ (CLD). However, the exact role of the CLD is unknown.

Here, the CLD of HlyB was investigated in great detail by *in vivo*, biochemical and structural analysis. We demonstrate that the presence of the CLD is essential to secrete HlyA and that the CLD binds the RTX domain of HlyA1 exclusively in the unfolded state. The solution

structure of the CLD was solved by NMR spectroscopy, which also allowed a precise mapping of the substrate-binding region. Taken together, our results suggest that the CLD shields the substrate to prevent for example aggregation of the secretion-competent, unfolded state. Obviously, the ABC transporter HlyB carries an additional function besides energizing substrate transport. It binds the substrate HlyA through the CLD, a mechanism that recognizes the unfolded substrate in the cytosol. This might represent a general mechanism for ABC transporters involved in the secretion of RTX proteins and finally shows how large, unfolded RTX toxins are protected inside the cells. The function of the CLD might also explain the lack of any known chaperone involved in HlyA secretion and significantly enhances our understanding of T1SS.

Results

Purification of the CLD, HlyA, HlyA1 and HlyA2

The CLD was cloned, expressed and purified as previously described (Schwarz et al., 2011). HlyA and the C-terminal fragments, HlyA1 (amino acids 807–1024) containing three highly conserved RTX repeat and HlyA2 (amino acids 807–966), were expressed and purified as previously described (Benabdelhak et al., 2003). HlyA2 (18 kDa) corresponds to the C-terminal fragment HlyA1, but lacks the secretion signal. For a detailed scheme of the constructs see Figure 1.

Secretion analysis of HlyA in the presence of HlyB or ΔCLD-HlyB

Full-length HlyA (amino acids 1-1024) was produced in *E. coli* BL21 (DE3) (Figure S1, lane 2). Together with wild-type (wt) HlyB and HlyD, HlyA was secreted to the supernatant (Figure S1, lane 5) and virtually none remained inside the cells (Figure S1, lane 3). In cells producing HlyA in combination with ΔCLD-HlyB and HlyD, however, no HlyA could be detected in the supernatant (Figure S1, lane 6). Instead, HlyA accumulated inside the cells (Figure S1, lane 4). In the wt and ΔCLD backgrounds, all components of the secretion complex (HlyB, HlyD and TolC) were expressed in similar amounts and localized to the membrane fraction of the cells (Figure 2). Thus, neither decreased expression levels nor mis-localization of ΔCLD-HlyB were responsible for the secretion defect. Taken together, these results demonstrate that the CLD of HlyB is essential for the secretion of HlyA.

Intrinsic tryptophan fluorescence to determine the folding state of HlyA1

The folding states of HlyA1, refolded in the presence of $CaCl_2$ or EDTA, were investigated using intrinsic tryptophan fluorescence spectroscopy (Bakas et al., 1998; Sanchez-Magraner et al., 2010; Sanchez-Magraner et al., 2007). HlyA1 contains a single tryptophan residue (amino acid 914), which was used as intramolecular probe. The fluorescence signal of "unfolded" HlyA1 increases approximately threefold after the addition of 8 mM Ca^{2+} (Figure S2). In contrast, the fluorescence intensity of "folded" HlyA1 did not change after the addition of Ca^{2+}. Titration experiments revealed a K_D for the binding of Ca^{2+} to unfolded HlyA1 of 230 µM ± 24 µM with a sigmoidal binding curve (h = 2.14 ± 0.40) (Figure S2). These results are in agreement with previous studies (Sanchez-Magraner et al., 2010; Sanchez-Magraner et al., 2007) and indicate that the two purification protocols specifically yielded the unfolded and folded HlyA1 species.

Pull-down assays of the CLD with HlyA, HlyA1 and HlyA2

Since HlyA was expressed in similar amounts in the wt and the ΔCLD background (Figure S1, lane 2 and 4), one can exclude that the CLD influences expression levels of the protein. Therefore, a potential interaction of the CLD with HlyA was investigated by pull-down assays (Figure S3). His_{10}-CLD was immobilized and incubated with unfolded HlyA. After extensive washing, both HlyA and the CLD co-eluted (Figure S3, lane 6), indicating a specific interaction between both proteins. As the characteristic feature of RTX toxins are the GG repeats, we suspected an interaction site within the RTX domain and purified HlyA1 (24.7 kDa), which contains three conserved GG repeats (Figure 1), in either an unfolded or folded state (Benabdelhak et al., 2003; Holland et al., 2005) (for a description of the folding experiments, see Supplementary Information). Unfolded His_6-HlyA1 was immobilized and the pull-down assay was repeated with the CLD lacking the His_{10}-tag. The CLD co-eluted with unfolded HlyA1 (Figure 3, lane 6) indicating an interaction. Since HlyA is presumably transported in an

unfolded state (Bakkes et al., 2010), a possible influence of the folding state on the interaction with the CLD was investigated. Pull-down experiments were repeated with folded HlyA1 and, in contrast to the results with unfolded HlyA1, the CLD did not co-elute (Figure 3, lane 7). This indicates an interaction of the CLD with the unfolded state of the substrate. To confirm the RTX domain as the region containing the binding site of the CLD, pull-down experiments were repeated with HlyA2 (HlyA1 lacking the secretion signal). As in the case of HlyA1, HlyA2 and the CLD co-eluted (Figure 4, lane 5). Since non-specific binding of either HlyA, HlyA1, HlyA2 or the CLD to the Ni-NTA beads was not observed in the corresponding control experiments, these data demonstrate an interaction of the CLD with the unfolded domain of HlyA1 upstream of secretion sequence and containing three RTX repeats.

NMR structure of the CLD

Sequence alignments of the CLD indicated a similarity to C39 peptidases. Since the CLD interacts with the unfolded RTX domain of HlyA and HlyA1, the biological function seems to differ from that of C39 peptidases, which cleave an N-terminal leader peptide from dedicated substrates. To understand this difference on a structural level, we determined the solution structure of the isolated CLD by NMR spectroscopy (PDB entry 3zua). The overall structure features a α+β-fold and is similar to the C39 peptidase ComA-PEP (Ishii et al., 2010) displaying an rmsd of 1.8 Å over 137 C_α atoms as calculated by the SSM routine implemented in coot (Emsley et al., 2010). It consists of a central, six-stranded, antiparallel β-sheet core surrounded by α-helices (Figure 5A). The fold can be divided in two subdomains (Figure S4, A). The major domain contains the β-sheet core (β_1–β_6) supplemented by two small α-helices (α_4 and α_5). The smaller domain comprises three consecutive N-terminal α-helices (α_1–α_3) sitting on top of the major domain. A structural superimposition of the CLD

and ComA-PEP confirmed the replacement of the catalytically important cysteine (Wu and Tai, 2004) by a tyrosine (Tyr^{9}) (Figure 5B). Compared to the catalytically histidine of ComA-PEP (His^{96}), the equivalent histidine of the CLD (His^{83}) is flipped by 180°. The side chain of Tyr^{9} fills the resulting gap and Arg^{82} occupies the structurally equivalent position of the histidine of the ComA-PEP (His^{96}).

The dislocation of His^{83} in the CLD is stabilized via a π–π stacking of the imidazole ring with the indole group of Trp^{77} (Figure 5B). His-Trp stacks are common motifs used to stabilize proteins with an energy gain of up to 4 kJ/mol (Samanta et al., 1999). However, this kind of arrangement has not been described for members of the C39 peptidase or papain family. Interestingly, in all CLDs of ABC transporters this tryptophan residue is conserved (Figure S5), suggesting that the observed His-Trp stacking represents a general feature of CLDs.

Mapping of the substrate-binding region of HlyA

The pull-down experiments indicated an interaction of the CLD with the unfolded RTX domain of HlyA or HlyA1. We used the advantage of NMR spectroscopy to further characterize this interaction by chemical shift perturbation (CSP) experiments with (^{1}H-^{15}N)-HSQC spectra using HlyA2. At the lowest molar ratios between HlyA2 and the CLD (0.03 and 0.1:1), only slight effects on the chemical shifts were observed, whereas significant changes in the ^{1}H and ^{15}N frequencies of distinct correlation signals were detected at the higher molar ratios (0.3 and 1:1) (Figure 6). These changes cluster in sheets β_3 – β_5 and helix α_1. The central part of the interaction surface is comprised of hydrophobic amino acid residues (around Ile^{15} and Leu^{96}), surrounded by a hydrophilic rim located on sheet β_5, on helix α_1 and in loop l_{34} (Figure 6). The chemical shifts of the other residues of the CLD did not undergo such changes, demonstrating the specificity of the interaction. The CSP experiments were

additionally carried out at different Ca^{2+} concentrations (1–5 mM), starting from the HlyA2-CLD complex in the absence of Ca^{2+}. With increasing Ca^{2+} concentrations, the cross peaks clearly shifted back towards the reference spectrum of unbound CLD (Figure S6). These CSP experiments were consistent with the pull-down assays described above and allowed the mapping of the binding region on the CLD for HlyA2.

Mutational analysis of the CLD of HlyB

The above-described experiments indicated that the HlyA2 binding-region within the CLD is located on the opposite site of the ligand-binding region ComA-PEP. To further support the NMR data, mutations were introduced into HlyB to analyze whether residues corresponding to the binding-site of Com-PEP (Ishii et al., 2010) or corresponding to those residues identified by the CSP experiments influence substrate binding. To investigate the contribution of residues corresponding to ComA-PEP, we mutated L43 and Q59 of HlyB (corresponding to A51 and A67 of ComA-PEP) to Trp (Figure 7). The Trp residues were introduced in ComA-PEP to demonstrate te importance of both residues in substrate binding and hydrolysis. In parallel, we constructed the mutants E15A, I16A, Y20A and F99A of HlyB. Secretion analysis of these mutations demonstrated that none of the residues implicated in substrate recognition of ComA-PEP influenced HlyA secretion (Figure 7B). In clear contrast, mutation of Y20 or F99 drastically reduced HlyA secretion. Since none of the mutations at the "ComA-PEP" interface influenced substrate secretion of this T1SS, our mutational analysis strongly supports the results of the CSP experiments and suggests that C39 peptidase domains and CLDs use a different binding mode of their cognate substrates.

Discussion

The crystal structures of two members of the RTX toxin superfamily demonstrated that GG repeats bind Ca^{2+} ions (Baumann et al., 1993; Hamada et al., 1996), which are essential for activity (Ostolaza and Goni, 1995; Ostolaza et al., 1995; Rose et al., 1995). Recent studies also showed that C-terminal fragments of HlyA bind Ca^{2+} ions with similar affinities as full-length HlyA ($K_{0.5}$ ≈ 100 µM), which induce their folding (Sanchez-Magraner et al., 2010; Sanchez-Magraner et al., 2007). Since the cytoplasmic Ca^{2+} concentration in *E. coli* is very low (low µM range), whereas the extracellular concentration is typically in the mM range (Holland et al., 1999; Jones et al., 1999), HlyA is suggested to be transported in an unfolded state before Ca^{2+} ions trigger the folding at the exterior (Chang et al., 1986; Holland et al., 1999; Jones et al., 1999). As the CLD interacts exclusively with the unfolded substrate, a cytoplasmical localization of the CLD can be assumed. This localization is supported by recent studies of the ABC transporter NukT, which transports the bacteriocin nukacin. Here, the C39 peptidase domain was also shown to be localizaed at the cytoplasmical site (Nishie et al., 2011).

So far, no structural information is available for CLDs and only the crystal structure of one C39 peptidase, ComA-PEP, has described in the literature (Ishii et al., 2010). Here, we report the solution structure of the HlyB-CLD determined by NMR spectroscopy. Despite a similar overall fold compared to the C39 peptidase (Ishii et al., 2010), the CLD showed a degenerated proteolytic site and revealed a π–π stacking, which might be a characteristic feature of CLDs (Figure 8). This π–π stacking also explains why a “simple” re-introduction of Cys at position 9 of the CLD (Y9C) does not restore proteolytic activity (data not shown). Importantly, the binding site of HlyA2 is located on the opposite side compared to the substrate-binding site of ComA-PEP (Figures 7 and S7). Even though the homology between

ComA-PEP and the CLD of HlyB and a resulting similar overall structure, the CLD of HlyB represents a particular protein class with a distinct binding site imposing specificity for the RTX toxin HlyA.

Based on these findings, a sequence analysis of ABC transporters containing either a C39 peptidase domain or a CLD was performed and related to their dedicated substrates. Placed in a phylogenetic tree, an unambiguous separation based on the nature of the N-terminal domain emerged (Figure 8). All transporters containing a C39 peptidase domain have rather small substrates (< 10 kDa). These belong to the bacteriocin family that possess an N-terminal, cleavable leader peptide. In contrast, ABC transporters with CLDs transport large substrates (≥ 55 kDa) that contain a C-terminal secretion signal and, importantly, are all members of the RTX toxin family. Thus, clearly the existence of an RTX domain in the transported substrate is linked to the presence of a CLD in the cognate ABC transporter (Figure 8). Moreover, this analysis allows now prediction of substrates for ABC transporters, and *vice versa*, in newly sequenced genomes. For example, the ABC transporter ClyB from *Enterococcus faecalis* with unknown function contains a C39 peptidase, and likely transports a bacteriocin with an N-terminal leader peptide, whereas AqxB from *Actinobacillus equuli* presumably transports a RTX toxin, as it harbors a CLD (Figure 8).

Some ABC transporters dedicated to protein secretion in Gram-negative bacteria are devoid of either a C39 peptidase or a CLD and, therefore, do not match the described classes. One such example is HasD, which is part of the HasA T1SS of *Serratia marcescens* (Letoffe et al., 1994; Masi and Wandersman, 2010). In accordance with our classification system, HasA is not cleaved during the transport and contains no RTX domain. Nevertheless, HasA contains primary recognition sites that interact with HasD prior to the secretion. These sites, however, do not correspond to any specific motif and their precise identity remains to be

defined (Masi and Wandersman, 2010). Interestingly, the chaperone SecB was shown to be essential for the HasA T1SS by inhibiting the folding and maintaining the unfolded, secretion-competent state of HasA (Delepelaire and Wandersman, 1998). Previously, it was suggested that a cytoplasmic chaperone of some kind is also required to protect unfolded HlyA inside the cell (Holland et al., 2005), however, no chaperone could be identified so far (Holland et al., 1990). Here, we provide strong evidence that the CLD of HlyB fulfills this role. The CLD shields HlyA to prevent its aggregation and/or degradation prior to secretion. Therefore, HlyB carries a dual function in the secretion process by energizing the transport and shielding the RTX toxin. To our knowledge, this mechanism is novel for T1SS and significantly expands the understanding of protein transport across membranes.

Experimental Procedures

Cloning, expression and purification of the CLD

The plasmid encoding the CLD was constructed from the parental plasmid pET28b-HlyB by the introduction of a stop codon (TAA) at amino acid position 140 of wild-type HlyB. *E. coli* BL21 (DE3) pLysS was grown at 37 °C in medium containing kanamycin (30 µg/ml) and chloramphenicol (30 µg/ml) under vigorously shaking at 160 rpm. For biochemical experiments, 2xYT medium or for isotopic enrichment M9 medium supplemented with $^{13}C_6$-glucose (2 g/L; 99.9% CIL) and/or ^{15}N-ammonium chloride (0.5 g/L; 99.9% CIL) was used. The bacteria were grown to an OD_{600} of 0.6 before the expression of the CLD was induced with 0.1 mM IPTG. Cells of a 4 L culture were harvested 4 h after induction by centrifugation and resuspended in 60 ml buffer A (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 10 mM imidazole, Complete Protease Inhibitor Cocktail tablet; Roche). All subsequent steps were performed at 4 °C. After the addition of DNAse (0.1 mg/ml) and lysozyme (1.0 mg/ml), cells were incubated for 20 min on ice and disrupted by three passages through a cell disrupter (Constant Systems) at 2.5 kbar. The homogenate was centrifuged for 30 min at 14,000 x g and the supernatant was centrifuged again for 1 h at 120,000 x g. The supernatant was applied to an immobilized metal-ion affinity chromatography (IMAC) column (5 ml HiTrap Chelating HP column, GE Healthcare), loaded with Ni_2SO_4 and equilibrated in buffer A. The CLD was eluted with a linear gradient from 10 mM to 500 mM imidazole. Elution fractions were immediately diluted 1:1 with buffer C (100 mM citrate, 50 mM LiCl, 2 mM DTT, pH 6.0) and concentrated by ultrafiltration using an Amicon Ultra Centrifugal Filter device (10 kDa MWCO, Millipore). When required, His_6-tagged TEV-protease (Kapust et al., 2001) was used to cleave the N-terminal His_{10}-tag of the CLD, resulting in an N-terminal overhang of Gly-Ala. TEV protease was purified as described previously (Gluck et al., 2009) and the digestion was

performed during dialysis against digestion buffer (50 mM Tris-HCl, 150 mM NaCl, 2 mM DTT, pH 8.0). After a second dialysis against digestion buffer without DTT, the His_6-TEV protease and the released His_{10}-tag were removed by an additional IMAC step and the flow through, containing the untagged CLD, was collected and concentrated. A homogenous sample of the CLD was prepared by size-exclusion chromatography (SEC) using a Superdex 75 16/60 column (GE Healthcare) in buffer D (50 mM citrate, 50 mM LiCl, 5 mM DTT, pH 6.0). Elution fractions containing the CLD were pooled, concentrated and stored at 4 °C. For subsequent experiments, the CLD was dialyzed against the appropriate buffer. The protein concentration was determined with a nanodrop device (PeqLab) and the calculated molecular mass and extinction coefficient.

Expression and purification of HlyA

Non-acylated HlyA was purified as described previously from inclusion bodies (Jumpertz et al., 2010). In brief, after washing of the IBs, HlyA was dissolved in denaturation buffer (50 mM Tris-HCl, 8 M urea, pH 7.0) and purified by anion-exchange chromatography (DEAD fast flow sepharose, GE Healthcare). HlyA was eluted with a linear gradient from 0 mM to 300 mM NaCl in denaturation buffer. Further purification steps were omitted. HlyA could be stored at 4 °C for several months. Pull-down assays were performed with HlyA dialyzed against denaturation buffer with 1.5 M Urea and 10 mM EDTA. The protein concentration was determined with a nanodrop device (PeqLab) and the calculated molecular mass and extinction coefficient (Figure 1).

Expression and purification of HlyA1 and HlyA2

The C-terminal fragments of HlyA, HlyA1 (amino acid residues 807–1024, 24.9 kDa) and HlyA2 (amino acid residues 807–966, 17.7 kDa) were both purified as described previously

(Benabdelhak et al., 2003). HlyA1 forms inclusion bodies, which were purified, whereas HlyA2 represents a soluble degradation product of HlyA1 devoid of the last 57 amino acid residues and, therefore, the secretion signal. The introduction of an N-terminal His_6-tag allowed the preparation of pure proteins after an IMAC for HlyA1, and after an IMAC in combination with an anion-exchange chromatography for HlyA2. *E. coli* BL21 (DE3) was grown in 2xYT medium at 37 °C to an OD_{600} of 1.0, induced with 4 mM arabinose and harvested after 90 min of cell growth. Cells were resuspended in lysis buffer (10 mM Tris-HCl, pH 8.0, 100 mM KCl, 1 mM PMSF, 0.2 mM EDTA, 10% glycerol) and broken by three passages through a cell disrupter (Constant Systems) at 2.5 kbar. The insoluble fraction was collected, dissolved in 25 ml lysis buffer with 6 M guanidine-HCl and dialyzed exhaustively against lysis buffer containing 1 mM PMSF and either 0.2 mM EDTA ("unfolded" HlyA1) or 10 mM $CaCl_2$ ("folded" HlyA1), respectively. The soluble fraction after cell disruption was centrifuged for 1 h at 120.000 x g at 4 °C to remove debris. Both HlyA fragments were loaded onto HiTrap Chelating HP columns (GE Healthcare) loaded with Ni_2SO_4 and equilibrated in lysis buffer containing 10 mM imidazole. Proteins were eluted with a linear gradient from 10 mM to 500 mM imidazole. Fractions with the HlyA fragments were pooled and concentrated. HlyA1 was stored in lysis buffer with 25% glycerol at – 80 °C. In addition to the described purification protocol (Benabdelhak et al., 2003), HlyA2 was further purified via anion-exchange chromatography using a HiTrap Q HP column (GE Healthcare), equilibrated in lysis buffer without KCl. HlyA2 was eluted with a linear gradient from 0 mM to 500 mM KCl. Fractions containing pure HlyA2 were pooled, concentrated and stored in lysis buffer with 25% glycerol at – 80°C. Protein concentrations were determined with a nanodrop device (PeqLab) and the calculated molecular masses and extinction coefficients (Figure 1).

Secretion analysis of HlyA in the presence of wild-type HlyB, ΔCLD-HlyB or HlyB with mutants within the CLD of HlyB

Plasmid pK184-HlyBD was mutated by site-directed mutagenesis and the appropriate oligonucleotides at the positions E15A, I16A, Y20A and F99A and two point mutations were inserted iteratively at position L43W and Q59W. *E. coli* BL21 (DE3) harboring plasmids pSU2726 (encoding for HlyA) and either pK184-HlyBD, pK184-ΔCLD-HlyBD or mutated pK184-HlyBD versions was grown at 37 °C in the presence of ampicillin (100 µg/mL) and kanamycin (30 µg/mL) (Bakkes et al., 2010; Jumpertz et al., 2010). The expression was induced at an OD_{600} of 0.5 with 1 mM β-D-thiogalactopyranoside. After 4 h growth, cultures were centrifuged for 10 min at 13.200 x g, 4 °C. Supernatant samples were taken for SDS-PAGE analysis and subsequent CBB staining or immunoblotting using antibodies raised against HlyA. Cells were broken with a cell disrupter (Constant Systems) and cell debris were removed by centrifugation for 20 min at 14,000 x g, 4 °C. The soluble fraction was centrifuged for 90 min at 120,000 x g, 4 °C and the pellet was analyzed by immunoblotting with antibodies raised against HlyB, HlyD and TolC, respectively.

Intrinsic tryptophan fluorescence of HlyA1

Measurements were performed with 1 µM HlyA1, refolded either in the presence of 0.2 mM EDTA ("unfolded" HlyA1) or 20 mM $CaCl_2$ ("folded" HlyA1) at a Fluorolog III spectrometer (Horiba). After excitation (295 nm, 2 nm slit width), the emission spectrum was recorded (310–360 nm, 6 nm slit width, 0.3 sec integration time) before and after the addition of 8 mM $CaCl_2$. The normalized fluorescence was plotted against the wavelength. For titration experiments, Ca^{2+} was added gradually, fluorescence signals from 325–333 nm were averaged, normalized and plotted against the Ca^{2+} concentration. Data evaluation was done

with the software Prism (GraphPad Software, Inc.) using a non-linear regression (*one site – Specific binding with Hill slope*).

Pull-Down assays

His_{10}-CLD and HlyA

50 µl magnetic Ni-NTA beads (Qiagen, Hilden) were equilibrated three times with interaction buffer (0.5 M urea, 150 mM NaCl, 20 mM imidazole, pH 8.0) and 30 µg His_{10}-CLD were immobilized. The beads were washed three times with interaction buffer to remove unbound His_{10}-CLD. Unfolded HlyA was added to the beads (50 µl, 5.3 mg/ml) and gently mixed for 15 min at room temperature before washing of the beads seven times with 100 µl of interaction buffer. His_{10}-CLD was eluted with elution buffer (interaction buffer supplemented with 500 mM imidazole). The same procedure was performed in the absence of the His_{10}-CLD to investigate non-specific binding of HlyA to the beads.

His_6-HlyA1/A2 and the CLD

50 µl magnetic Ni-NTA beads (Qiagen, Hilden) were equilibrated with binding buffer (20 mM HEPES, 150 mM NaCl, 20 mM imidazole, pH 7.0) and 30 µg His_6-HlyA1 were immobilized. In pull-down assays with folded His_6-HlyA1, buffers were supplemented with 20 mM $CaCl_2$. Beads were washed three times with binding buffer to remove unbound His_6-HlyA1. The CLD was added to the beads (40 µl, 2.0 mg/ml) and gently mixed for 15 min at room temperature, before the beads were washed seven times with 100 µl of binding buffer. His_6-HlyA1 was eluted with 30 µl of elution buffer (binding buffer supplemented with 300 mM imidazole). The same procedure was performed in the absence of His_6-HlyA1 to investigate non-specific binding of the CLD to the beads. Pull-down assays with His_6-HlyA2 and the CLD were performed identically.

NMR spectroscopy

NMR experiments were performed at 303 K on a Varian VNMRS spectrometer at a proton frequency of 900 MHZ with a cryogenically cooled triple resonance probe. Protein samples contained [*U*-^{15}N]- or [*U*-^{15}N,^{13}C]-CLD in a buffer system of 100 mM citrate, pH 6.0, 100 mM LiCl, 10 mM dithiotreitol, 0.5 mM EDTA, 0.05% (w/v) sodium azide and 7% (v/v) D_2O, at a protein concentration of 2.5 mM. Precise procedures applied for NMR experiments, data evaluation and structure calculation are described in greater detail in Supplementary Information.

NMR titrations

CSP experiments were carried out to characterize the interaction between HlyA2 and the CLD. Briefly, a series of (^{1}H-^{15}N)-HSQC spectra was recorded, keeping the CLD concentration constant, while HlyA2 was diluted into the sample. Starting from an equimolar sample of [*U*-^{15}N]-CLD and [*NA*]-HlyA2 at 75 µM in titration buffer (25 mM Tris-HCl, pH 7.0, 125 mM NaCl, 2 mM dithiotreitol, 0.5 mM EDTA, 7% (v/v) D_2O), a three step dilution series with a solution of 75 µM [*U*-^{15}N]-CLD in titration buffer was set up, each step reducing the HlyA2 concentration to a third. This resulted in protein-protein ratios of 1:1, 1:0.3, 1:0.1 and 1:0.03 with a constant CLD concentration of 75 µM. In order to evaluate the effect of the folding state of HlyA2 in NMR experiments, equimolar samples of the two proteins at 75 µM containing 0, 1 and 5 mM $CaCl_2$ were prepared and (^{1}H-^{15}N)-HSQC spectra recorded. The CSP data was analyzed with CcpNmr analysis (Vranken et al., 2005). Cross peaks were assigned using the deposited chemical shifts (BMRB entry 17403) (Lecher et al., 2011). Deviations due to differences in buffer conditions were corrected by selecting to the closest neighboring

cross peak, if the selection was unambiguous. Trajectories following the cross peak movement were calculated and plotted against the residue numbers.

Acknowledgements

We thank Stefan Jenewein and Thorsten Jumpertz for important contributions in the initial stages of the project. We thank Diana Clausnitzer, Protein Production Facility, of the Heinrich Heine University, for the mutagenesis procedures and valuable discussion. We are indebted to I. Barry Holland, University of Paris-Sud, for an on-going, stimulating and very fruitful collaboration. We thank the Ministry of Innovation, Science and Research of the German Federal State North Rhine-Westphalia (NRW) (scholarship from the CLIB Graduate Cluster Industrial Biotechnology to CKWS), the International Helmholtz-Research School (IHRS) Biosoft (scholarship to JL) for financial support and the Heinrich-Heine-Universität Düsseldorf.

References

Bakas, L., Veiga, M.P., Soloaga, A., Ostolaza, H., and Goni, F.M. (1998). Calcium-dependent conformation of E. coli alpha-haemolysin. Implications for the mechanism of membrane insertion and lysis. Biochim Biophys Acta *1368*, 225-234.

Bakkes, P.J., Jenewein, S., Smits, S.H., Holland, I.B., and Schmitt, L. (2010). The rate of folding dictates substrate secretion by the Escherichia coli hemolysin type 1 secretion system. J Biol Chem *285*, 40573-40580.

Barrett, A.J., and Rawlings, N.D. (2001). Evolutionary lines of cysteine peptidases. Biol Chem *382*, 727-733.

Baumann, U., Wu, S., Flaherty, K.M., and McKay, D.B. (1993). Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. EMBO J *12*, 3357-3364.

Benabdelhak, H., Kiontke, S., Horn, C., Ernst, R., Blight, M.A., Holland, I.B., and Schmitt, L. (2003). A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. J Mol Biol *327*, 1169-1179.

Chang, C.F., Shuman, H., and Somlyo, A.P. (1986). Electron probe analysis, X-ray mapping, and electron energy-loss spectroscopy of calcium, magnesium, and monovalent ions in log-phase and in dividing Escherichia coli B cells. J Bacteriol *167*, 935-939.

Delepelaire, P. (2004). Type I secretion in gram-negative bacteria. Biochim Biophys Acta *1694*, 149-161.

Delepelaire, P., and Wandersman, C. (1998). The SecB chaperone is involved in the secretion of the Serratia marcescens HasA protein through an ABC transporter. EMBO J *17*, 936-944.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Ghigo, J.M., and Wandersman, C. (1994). A carboxyl-terminal four-amino acid motif is required for secretion of the metalloprotease PrtG through the Erwinia chrysanthemi protease secretion pathway. J Biol Chem *269*, 8979-8985.

Gluck, J.M., Wittlich, M., Feuerstein, S., Hoffmann, S., Willbold, D., and Koenig, B.W. (2009). Integral membrane proteins in nanodiscs can be studied by solution NMR spectroscopy. Journal of the American Chemical Society *131*, 12060-12061.

Hamada, K., Hata, Y., Katsuya, Y., Hiramatsu, H., Fujiwara, T., and Katsube, Y. (1996). Crystal structure of Serratia protease, a zinc-dependent proteinase from Serratia sp. E-15, containing a beta-sheet coil motif at 2.0 A resolution. J Biochem *119*, 844-851.

Havarstein, L.S., Diep, D.B., and Nes, I.F. (1995). A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol *16*, 229-240.

Holland, I.B., Jones, H.E., Campbell, A.K., and Jacq, A. (1999). An assessment of the role of intracellular free Ca2+ in E. coli. Biochimie *81*, 901-907.

Holland, I.B., Kenny, B., and Blight, M. (1990). Haemolysin secretion from E coli. Biochimie *72*, 131-141.

Holland, I.B., Schmitt, L., and Young, J. (2005). Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). Mol Membr Biol *22*, 29-39.

Ishii, S., Yano, T., Ebihara, A., Okamoto, A., Manzoku, M., and Hayashi, H. (2010). Crystal structure of the peptidase domain of Streptococcus ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. J Biol Chem *285*, 10777-10785.

Jones, H.E., Holland, I.B., Baker, H.L., and Campbell, A.K. (1999). Slow changes in cytosolic free Ca2+ in Escherichia coli highlight two putative influx mechanisms in response to changes in extracellular calcium. Cell Calcium *25*, 265-274.

Jumpertz, T., Chervaux, C., Racher, K., Zouhair, M., Blight, M.A., Holland, I.B., and Schmitt, L. (2010). Mutations affecting the extreme C terminus of Escherichia coli haemolysin A reduce haemolytic activity by altering the folding of the toxin. Microbiology *156*, 2495-2505.

Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein engineering *14*, 993-1000.

Kenny, B., Chervaux, C., and Holland, I.B. (1994). Evidence that residues -15 to -46 of the haemolysin secretion signal are involved in early steps in secretion, leading to recognition of the translocator. Mol Microbiol *11*, 99-109.

Kenny, B., Taylor, S., and Holland, I.B. (1992). Identification of individual amino acids required for secretion within the haemolysin (HlyA) C-terminal targeting region. Mol Microbiol *6*, 1477-1489.

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*, 914-919.

Lecher, J., Stoldt, M., Schwarz, C.K., Smits, S.H., Schmitt, L., and Willbold, D. (2011). 1H, 15N and 13C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB). Biomol NMR Assign *5*, 199-201.

Letoffe, S., Delepelaire, P., and Wandersman, C. (1990). Protease secretion by Erwinia chrysanthemi: the specific secretion functions are analogous to those of Escherichia coli alpha-haemolysin. EMBO J *9*, 1375-1382.

Letoffe, S., Ghigo, J.M., and Wandersman, C. (1994). Secretion of the Serratia marcescens HasA protein by an ABC transporter. J Bacteriol *176*, 5372-5377.

Linggood, M.A., and Ingram, P.L. (1982). The role of alpha haemolysin in the virulence of Escherichia coli for mice. J Med Microbiol *15*, 23-30.

Linhartova, I., Bumba, L., Masin, J., Basler, M., Osicka, R., Kamanova, J., Prochazkova, K., Adkins, I., Hejnova-Holubova, J., Sadilkova, L., *et al.* (2010). RTX proteins: a highly diverse family secreted by a common mechanism. FEMS Microbiol Rev *34*, 1076-1112.

Masi, M., and Wandersman, C. (2010). Multiple signals direct the assembly and function of a type 1 secretion system. J Bacteriol *192*, 3861-3869.

Nicaud, J.M., Mackman, N., Gray, L., and Holland, I.B. (1985). Characterisation of HlyC and mechanism of activation and secretion of haemolysin from E. coli 2001. FEBS Lett *187*, 339-344.

Nishie, M., Sasaki, M., Nagao, J., Zendo, T., Nakayama, J., and Sonomoto, K. (2011). Lantibiotic transporter requires cooperative functioning of the peptidase domain and the ATP binding domain. J Biol Chem *286*, 11163-11169.

Ostolaza, H., and Goni, F.M. (1995). Interaction of the bacterial protein toxin alpha-haemolysin with model membranes: protein binding does not always lead to lytic activity. FEBS Lett *371*, 303-306.

Ostolaza, H., Soloaga, A., and Goni, F.M. (1995). The binding of divalent cations to Escherichia coli alpha-haemolysin. Eur J Biochem *228*, 39-44.

Rose, T., Sebo, P., Bellalou, J., and Ladant, D. (1995). Interaction of calcium with Bordetella pertussis adenylate cyclase toxin. Characterization of multiple calcium-binding sites and calcium-induced conformational changes. J Biol Chem *270*, 26370-26376.

Samanta, U., Pal, D., and Chakrabarti, P. (1999). Packing of aromatic rings against tryptophan residues in proteins. Acta Crystallogr D Biol Crystallogr *55*, 1421-1427.

Sanchez-Magraner, L., Cortajarena, A.L., Garcia-Pacios, M., Arrondo, J.L., Agirre, J., Guerin, D.M., Goni, F.M., and Ostolaza, H. (2010). Interdomain Ca(2+) effects in Escherichia coli alpha-haemolysin: Ca(2+) binding to the C-terminal domain stabilizes both C- and N-terminal domains. Biochim Biophys Acta *1798*, 1225-1233.

Sanchez-Magraner, L., Viguera, A.R., Garcia-Pacios, M., Garcillan, M.P., Arrondo, J.L., de la Cruz, F., Goni, F.M., and Ostolaza, H. (2007). The calcium-binding C-terminal domain of Escherichia coli alpha-hemolysin is a major determinant in the surface-active properties of the protein. J Biol Chem *282*, 11827-11835.

Schwarz, C.K., Tschapek, B., Jumpertz, T., Jenewein, S., Lecher, J., Willbold, D., Panjikar, S., Holland, I.B., Smits, S.H., and Schmitt, L. (2011). Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the Escherichia coli ABC transporter haemolysin B. Acta Crystallogr Sect F Struct Biol Cryst Commun *67*, 630-633.

Stanley, P., Packman, L.C., Koronakis, V., and Hughes, C. (1994). Fatty acylation of two internal lysine residues required for the toxic activity of Escherichia coli hemolysin. Science *266*, 1992-1996.

Thanabalu, T., Koronakis, E., Hughes, C., and Koronakis, V. (1998). Substrate-induced assembly of a contiguous channel for protein export from E.coli: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. EMBO J *17*, 6487-6496.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687-696.

Wu, K.H., and Tai, P.C. (2004). Cys32 and His105 are the critical residues for the calcium-dependent cysteine proteolytic activity of CvaB, an ATP-binding cassette transporter. J Biol Chem *279*, 901-909.

Figure legends

Figure 1:

Schematic view of the HlyA constructs used in this study with corresponding, calculated molecular masses (in kDa) and extinction coefficients (ε, in M^{-1} cm^{-1}). (**A**) HlyA (UniProtKB entry Q1R2T5) contains an N-terminal, putative hydrophobic membrane insertion domain (M), a C-terminal secretion signal (blue) and six RTX repeats of the consensus sequence GGxGxDxUx (x, any amino acid residue, U, large, hydrophobic amino acid residue). The two indicated lysine residues are acylated in toxic form of HlyA, (**B**) the C-terminal fragment of HlyA, HlyA1, consists of the secretion signal, three conserved RTX repeats and a modified RTX repeat (GGKDDKLSL), (**C**) HlyA2 corresponds to HlyA1 devoid of the last 57 amino acid residues and therefore the secretion signal.

Figure 2:

Western blots of membrane fractions after secretion analysis of HlyA with HlyB and ΔCLD-HlyB, respectively. HlyB (79.8 kDa), ΔCLD-HlyB (63.2 kDa), HlyD (56.5 kDa) and TolC (53.7 kDa) are probed with polyclonal antibodies raised against each protein and detected by immunofluorescence. TolC is expressed constitutively in *E. coli*. Lane 1, 3 and 6: HlyA T1SS components in the wild-type HlyB background, lane 2, 4 and 7: HlyA T1SS components in the ΔCLD-HlyB background, lane 5: molecular weight markers (in kDa). Note, the higher molecular weight form of HlyB may represent the dimer.

Figure 3:

CBB stained SDS-PAGE (15%) of pull-down experiments with Ni-NTA beads between purified HlyA1 and the CLD. Lane 1: Molecular weight markers (in kDa), lane 2: unfolded His_6-HlyA1,

lane 3: CLD, lane 4: supernatant after incubation of the immobilized, unfolded His_6-HlyA1 and the CLD in binding buffer, lane 5: supernatant after repeated washing with binding buffer, lane 6: eluted fraction after incubation with binding buffer supplemented with 300 mM imidazole, lane 7: eluted fraction of an identical pull-down experiment, but with folded instead of unfolded His_6-HlyA1. Note, equal volumes of the CLD sample were loaded on lanes 3 and 4.

Figure 4:

CBB stained SDS-PAGE (15%) of a pull-down experiment with Ni-NTA beads between purified HlyA2 and the CLD. Lane 1: His_6-HlyA2, lane 2: CLD, lane 3: supernatant after incubation of the immobilized, unfolded His_6-HlyA2 and the CLD in binding buffer, lane 4: supernatant after repeated washing with binding buffer, lane 5: eluted fraction after incubation with binding buffer supplemented with 300 mM imidazole, lane 6: molecular weight markers (in kDa). Note, equal volumes of the CLD sample were loaded on lanes 2 and 3.

Figure 5:

(A) Solution structure of the HlyB-CLD solved with NMR spectroscopy (PDB entry 3zua). The protein contains a six-stranded, antiparallel β-sheet core (yellow) and five flanking α-helices (blue). The overall structure of the HlyB-CLD is very similar to that of ComA-PEP and adopts the characteristic fold of papain-like proteases (Barrett and Rawlings, 2001). The N- and C-termini are indicated. Secondary structure elements are labeled. **(B)** Structural superimposition of the catalytic site of ComA-PEP (PDB entry 3k8u, green) and the equivalent site in the HlyB-CLD (orange). Both structures superimpose with a RMSD of 1.8 Å

over 137 C^{α} atoms. Amino acid residues of the active center are shown in ball-and-stick representation and are labeled.

Figure 6:

Interaction of HlyA2 with tin gray. The ComA-PEPhe CLD characterized by NMR chemical shift perturbation (CSP) experiments. (**A**) Part of the (^{1}H-^{15}N)-HSQC spectra illustrating the CSP. Starting with an equimolar mixture of [*NA*]-HlyA2 and [*U*-^{15}N]-CLD (75 µM), the CLD concentration was kept constant, while HlyA2 was diluted in a three step series, each step reducing the HlyA2 concentration to a third. Arrows highlight the trajectory of chemical shift changes. (**B**) The chemical shift changes were mapped onto the CLD structure. Interacting regions are highlighted in yellow, while non-interacting regions are shown in blue. (**C**) The chemical shift changes were mapped onto the primary sequence of the CLD. The main interface is formed by α_1 and the complete β_4–β_6 segment (the red and green lines indicate a cut-off of 1 or 2 σ, respectively).

Figure 7:

(A) Cartoon representation of the NMR structure of the CLD (grey). Residues forming the substrate-binding interface of ComA-PEP are shown as a magenta surface. Leu43 corresponds to Ala51 and Gln59 to Ala61 in ComA-PEP (Ishii et al., 2010). Residues that displayed the largest shifts in the CSP experiments and are supposed to interact with HlyA2 are shown as a green surface. The amino acids that were mutated in this study are labeled. . The amino acids displaying are labeled. (B) CBB-stained SDS page of the supernatant of HlyA secreting *E. coli* cells. Left panel: protein markers. The variants of HlyB used in these secretion experiments, wt, ΔCLD or mutants, are indicated.

Figure 8:

Phylogenetic tree of ABC transporters with C39 peptidases or CLDs. ABC transporters with C39 peptidases are shown in the continuous ellipse, ABC transporters with CLDs are encircled with a dashed ellipse. All ABC transporters with CLDs transport substrates of the RTX toxin family, whereas ABC transporters with C39 peptidases transport bacteriocins. The dedicated ABC transporter-substrate pairs are: ApxIIIB (UniProtKB entry Q04473) and ApxIIID, NukT (UniProtKB entry Q75V16) and Nukacin, CvaB (UniProtKB entry P22520) and Colicin V, ClyB (UniProtKB entry Q57403) and unknown, LagD (UniProtKB entry P59852) and LagA, LtnT (UniProtKB entry O87239) and cytolysin B, ComA (UniProtKB entry P59653) and ComC, HlyB (UniProtKB entry Q1R2T6) and HlyA, LktB (UniProtKB entry P23702) and LktA, PaxB (UniProtKB entry Q9RCG7) and PaxA, RtxB (UniProtKB entry A1YKX0) and RtxA, AqxB (UniProtKB entry Q8KWZ8) and unknown.

Figure 1:

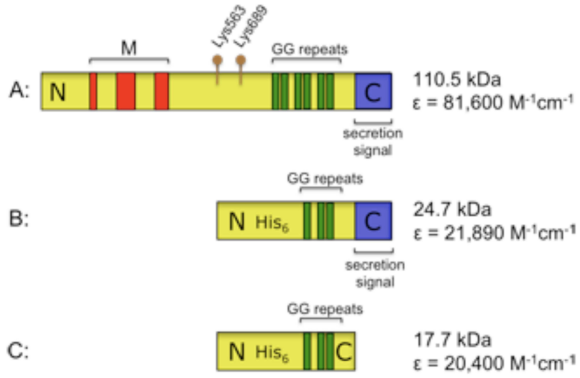


Figure 2:

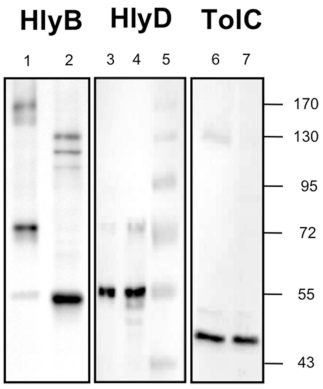


Figure 3:

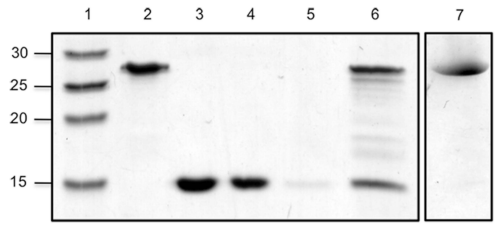


Figure 4:

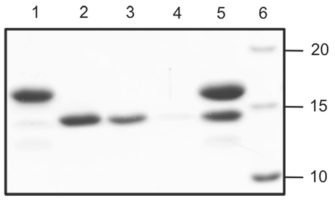


Figure 5:

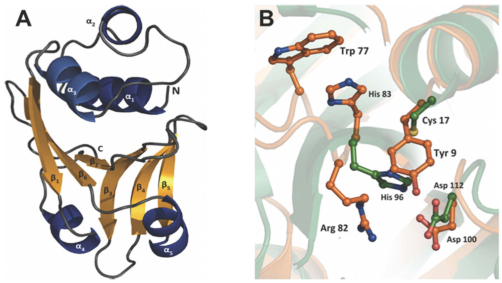


Figure 6:

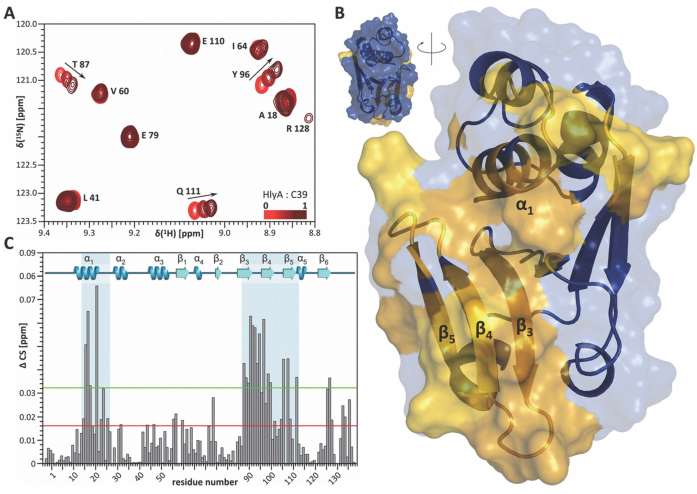


Figure 7:

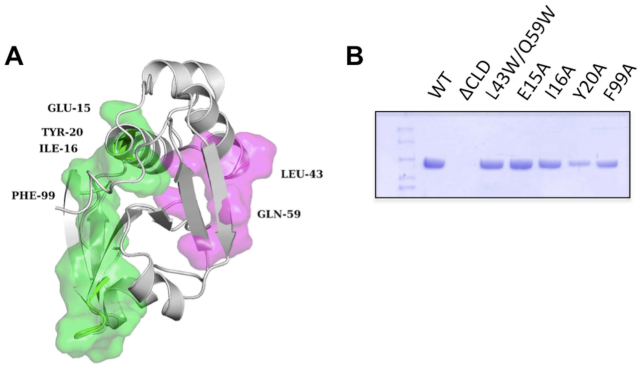
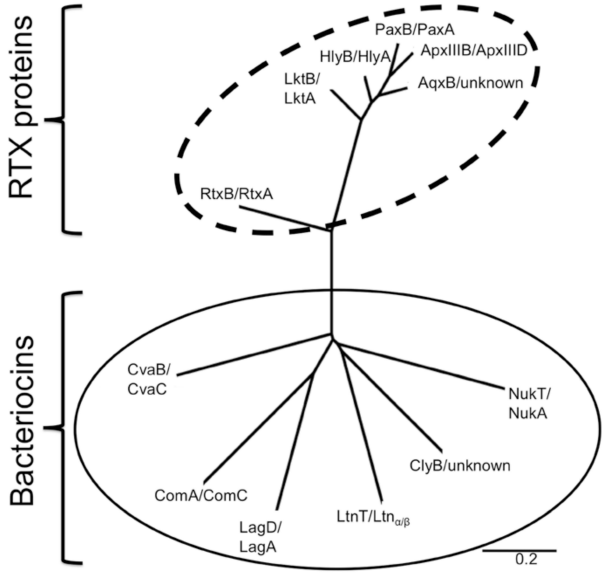


Figure 8:



Supplementary Information

Material and Methods

NMR Spectroscopy

NMR spectroscopic experiments were performed at 303 K on a Varian VNMRS spectrometer at a proton frequency of 900 MHz with a cryogenically cooled Z-axis pulse-field-gradient (PFG) triple resonance probe, unless stated differently. Protein samples contained [*U*-^{15}N]- or [*U*-^{15}N,^{13}C]-CLD in aqueous solution (100 mM citrate, pH 6.0, 100 mM LiCl, 10 mM dithiotreitol, 0.5 mM EDTA, 0.05% (w/v) sodium azide, 7% $^{2}H_2O$). In samples with [*U*-^{15}N,^{13}C]-protein, deuterated citrate (D4, 98% CIL), DTT (D6, 99% Sigma Aldrich) and EDTA (D12, 98% CIL) were used. All NMR data were processed using NMRPipe v. 5.5.2010 (Delaglio et al., 1995) or Azara v. 2.8 (provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge) software packages and analyzed with CcpNmr v. 2.1.5 (Vranken et al., 2005). Backbone and side-chain assignments were obtained using standard triple resonance heteronuclear NMR techniques as reported previously (Lecher et al., 2011). In order to derive Nuclear Overhauser enhancement (NOE) based distance restraints for structure calculation 3D-(^{1}H-^{1}H-^{15}N)-NOESY-HSQC (150 ms mixing time), (^{1}H-$^{13}C_{ali}$-^{1}H)-HSQC-NOESY (100 ms mixing time) and (^{1}H-$^{13}C_{aro}$-^{1}H)-HSQC-NOESY (120 ms mixing time) experiments were recorded using [*U*-^{15}N]- or [*U*-^{15}N,^{13}C]-CLD at 2.5 or 2 mM, respectively. Suppression of water was performed by application of WATERGATE technique. Decoupling of ^{15}N and ^{13}C nuclei during ^{1}H acquisition was achieved using GARP and adiabatically by WURST sequences. NOE cross peaks were identified interactively and the peak volume were integrated automatically using truncated box sum methods.

ARIA 2.3.1 (Rieping et al., 2007) was used for distance calibration of NOE cross peak volumes (Linge et al., 2001) and automated NOE cross peak assignment based on the previously published resonance assignment of the CLD. A set of 1669 ambiguous and 5879 unambiguous distance restrains (see supp. table TI) could be derived. Backbone dihedral ϕ and ψ angles were estimated from secondary chemical shifts of backbone atoms (H_N, H_α, N, C', C_α, C_β) using TALOS+ (Shen et al., 2009). Structure calculations based on NOE distance

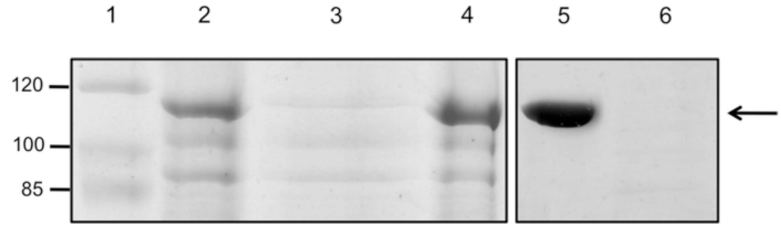
restraints and dihedral angle restraints were accomplished with a modified version of CNS v. 1.2.1 (Brunger, 2007; Rieping et al., 2007) using an optimized version of the PARALLHDG force field (v.5.3) (Linge and Nilges, 1999). To account for special protonation states of histidine side chains and *cis*-isomeric prolines, the HIS-PATCH and PRO-PATCH function of the force field were used. The MD protocol contained 45 ns high-temperature torsion angle dynamics (10,000 K) and 20,000 steps during the two cooling steps (2,000 K and 50 K). Experimental data terms were weighted with a log-harmonic-potential (Nilges et al., 2008). Refinement of the NOE cross peak assignment was carried out by iterative cycles of manual resonance assignment using CcpNmr and automated cycles using ARIAs approach.

10 lowest energy structures (overall CNS energy) out of 100 calculated were selected and validated using a combination of Molprobity (Chen et al., 2010), PROCHECK-NMR (Laskowski et al., 1996) and Cing (http://nmr.cmbi.ru.nl/cing/). The ensemble was deposited in the PDB (Berman et al., 2000) under entry 3ZUA. Chemical shift assignments were deposited previously in the BMRB (Ulrich et al., 2008) under accession number 17403. Structure figures were prepared with the program PyMol (Schrödinger, 2010). To characterize the dynamic properties of the CLD in solution, longitudinal (R_1) and transversal (R_2) relaxation of ^{15}N, as well as steady-state $^{15}N\{^1H\}$-NOE were recorded using standard methods (Farrow et al., 1994; Pervushin et al., 1997; Zhu et al., 2000). Conditions were used as described above using 1 mM [*U*-^{15}N]-CLD. For measurements of R_1 rates, relaxation delay values of 63, 131, 243, 369, 528, 739, 1161 and 1583 ms, for R_2 rates delay values of 10, 30, 50, 70, 90, 110 and 130 ms were used. 2.5 s ^1H-saturation was applied during ^{15}N-$\{^1H\}$-NOE experiments. Data evaluation, including cross peak integration and curve fitting, was accomplished with internal routines of CcpNmr v. 2.1.5.

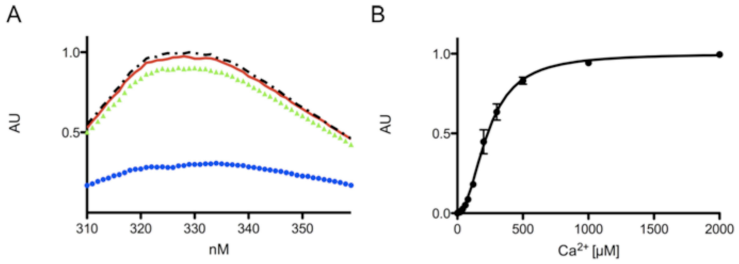
Supplementary figures

Supplementary figure S1:

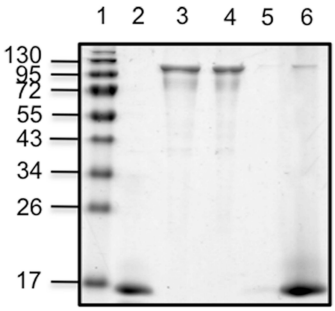
Secretion analysis with pro-HlyA. Crude cell extracts were taken for SDS-PAGE analysis and subsequently stained with CBB (lane 1–4) or supernatant samples were used for western blotting using an antibody raised against HlyA (lane 5–6). Lane 1: Molecular weight markers (in kDa), lane 2: cells expressing only *hlyA* (110.5 kDa), lane 3: cells expressing *hlyA, hlyB* and *hlyD*, lane 4: cells expressing *hlyA*, *ΔCLD-hlyB* and *hlyD*, lane 5: supernatant sample of lane 3, lane 6: supernatant sample of lane 4. The arrow indicates the position of pro-HlyA.



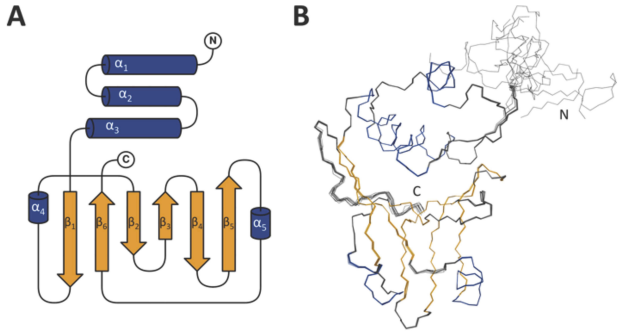
Supplementary figure S2:
Intrinsic tryptophan fluorescence spectroscopy with unfolded and folded HlyA1. (**A**) Normalized fluorescence spectrum from 310–360 nm with unfolded HlyA1 (blue), unfolded HlyA1 after the addition of 8 mM $CaCl_2$ (red), folded HlyA1 (green), and folded HlyA1 after the addition of 8 mM $CaCl_2$ (black); (**B**) Titration of $CaCl_2$ with unfolded HlyA1. The normalized fluorescence is plotted against the Ca^{2+} concentration; data were fitted with the software Prism (GraphPad Software).



Supplementary figure S3:
CBB stained SDS-PAGE (10%) of a pull-down experiment with Ni-NTA beads and the purified His_{10}-CLD and unfolded pro-HlyA. Lane 1: molecular weight markers (in kDa), lane 2: His_{10}-CLD, lane 3: pro-HlyA, lane 4: supernatant sample after incubation of pro-HlyA and the His_{10}-CLD in interaction buffer, lane 5: supernatant sample after repeated washing with interaction buffer, lane 6: eluted fraction after incubation with interaction buffer supplemented with 500 mM imidazole.

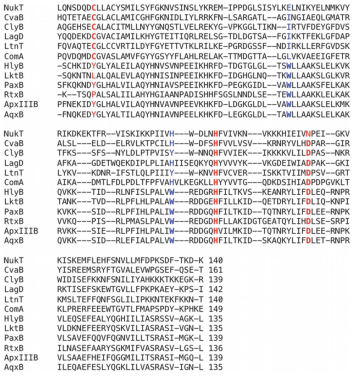


Supplementary figure S4:
(**A**) Topology diagram of the CLD; (**B**) Superimposition of 10 lowest energy NMR structures (CNS overall energy). The protein shows a very well-defined fold in the core region (amino acids 9–127), which was determined with a backbone atom deviation of 0.234 Å. In contrast, the N-terminal residues (amino acids 4–9) exhibit a high degree of flexibility.

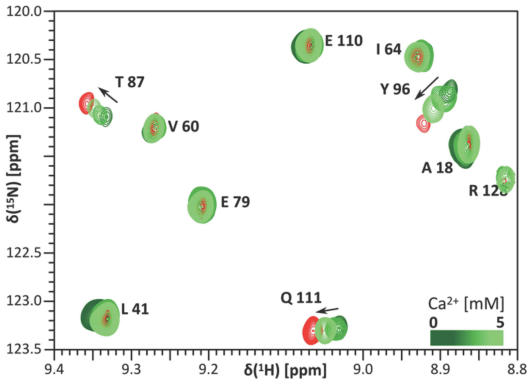


Supplementary figure S5:

Sequence alignment of ABC transporters with C39 peptidases and CLDs. Residues of the catalytic site are highlighted in red, Trp[77] of the HlyB-CLD and corresponding amino acid residues are highlighted in blue.

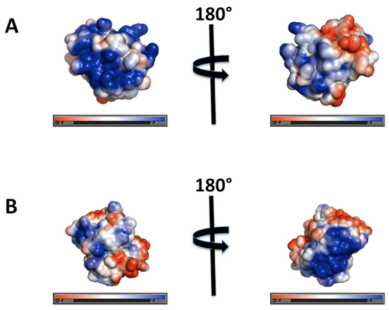


Supplementary figure S6:
CSP after addition of Ca^{2+} to the HlyA2-CLD complex. 0, 1 and 5 mM $CaCl_2$ were added to an equimolar solution of HlyA2 and the CLD (both 75 µM) in titration buffer, and (^{1}H-^{15}N)-HSQC spectra were recorded as described (see Materials and Methods). With an increasing concentration of Ca^{2+} ions (dark to light green), the resonance signals shifted towards the position of the CLD in the presence of 5 mM Ca^{2+} without HlyA2 (red).



Supplementary figure S7:

Surface electrostatic potential of the ComA-PEP (A) and the CLD (B), calculated with the APBS tool (Baker et al., 2001) implemented in Pymol (Schroedinger, 2010). Positive electrostatic surface potentials are coloured in blue, negative potentials in red contoured at 2kT.



Supplementary table I:

Experimental restraints and ensemble statistics of HlyB-CLD.

Experimental restraints	
Total NOE distance restraints	
Unambiguous	5879
Intraresidue	1669
Interresidue	
Sequential	1180
Medium range	1398
Long range	1632
Ambiguous	1494
Hydrogen bond restraints	1
Proline isomer restraints	1
Number of NOE upper bound violations	
Violations > 0.5 Å	15.5 (±5.4)
Violations > 0.3 Å	26.6 (±8.4)
Violations > 0.1 Å	93.5 (±14.9)
Deviations from idealized covalent geometry	
RMSD	
Bonds (Å)	0.0023 (±0.0000)
Angles (deg)	0.3539 (±0.0029)
Impropers (deg)	0.2742 (±0.0053)
Deviations from experimental restraints	
RMSD	
NOE restraints (Å)	0.0581 (±0.014)
Dihedral angle restraints (deg)	0.253 (±0.0176)
Hbond restraints (Å)	0
Ensemble RMSD (Å)	
Core residues (9-127)	
Backbone	0.23 (±0.04)
Heavy atoms	0.63 (±0.05)

References

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.
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 Research *28*, 235-242.
Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nature Protocols *2*, 2728-2733.
Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.
Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.
Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry *33*, 5984-6003.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR *8*, 477-486.
Lecher, J., Stoldt, M., Schwarz, C.K., Smits, S.H., Schmitt, L., and Willbold, D. (2011). 1H, 15N and 13C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB). Biomol NMR Assign *5*, 199-201.
Linge, J.P., and Nilges, M. (1999). Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. J Biomol NMR *13*, 51-59.
Linge, J.P., O'Donoghue, S.I., and Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods in enzymology *339*, 71-90.
Nilges, M., Bernard, A., Bardiaux, B., Malliavin, T., Habeck, M., and Rieping, W. (2008). Accurate NMR structures through minimization of an extended hybrid energy. Structure *16*, 1305-1312.
Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A *94*, 12366-12371.
Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T.E., and Nilges, M. (2007). ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics *23*, 381-382.
Schrödinger, L. (2010). The PyMOL Molecular Graphics System.
Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR *44*, 213-223.
Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., *et al.* (2008). BioMagResBank. Nucleic Acids Research *36*, D402-408.
Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687-696.

Zhu, G., Xia, Y., Nicholson, L.K., and Sze, K.H. (2000). Protein dynamics measurements by TROSY-based NMR experiments. Journal of Magnetic Resonance *143*, 423-426.

5.2 Chapter II

Title:	Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the *Escherichia coli* ABC transporter haemolysin B.
Authors:	Schwarz C. K. W., Tschapek B., Jumpertz T., Jenewein S., Lecher J., Willbold D., Panjikar S., Holland I. B., Smits S. H. J., Schmitt L.
Published in:	Acta Crystallogr Sect F Struct Biol Cryst Commun
Impact Factor:	0.563
Own proportion of this work:	60 % Production and purification of the CLD. Buffer screen, crystallization and preliminary X-ray analysis with the CLD. Writing of the manuscript.

crystallization communications

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

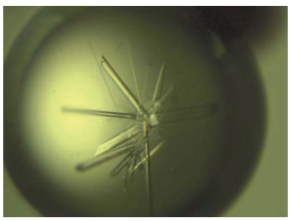
Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the *Escherichia coli* ABC transporter haemolysin B

Christian K. W. Schwarz,[a] Britta Tschapek,[a] Thorsten Jumpertz,[a] Stefan Jenewein,[a] Justin Lecher,[b,c] Dieter Willbold,[b,c] Santosh Panjikar,[d] I. Barry Holland,[e] Sander H. J. Smits[a] and Lutz Schmitt[a]*

[a]Institute of Biochemistry, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany, [b]Institute of Physical Biology, Heinrich Heine University Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany, [c]Institute of Structural Biochemistry (ICS-6), Research Centre Jülich GmbH, Jülich, Germany, [d]EMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and [e]Institute of Microbiology and Genetics, University of Paris-Sud, Orsay, France

Correspondence e-mail: lutz.schmitt@hhu.de

Received 28 January 2011
Accepted 23 March 2011



© 2011 International Union of Crystallography
All rights reserved

The ABC transporter haemolysin B (HlyB) from *Escherichia coli* is part of a type I secretion system that translocates a 110 kDa toxin in one step across both membranes of this Gram-negative bacterium in an ATP-dependent manner. Sequence analysis indicates that HlyB contains a C39 peptidase-like domain at its N-terminus. C39 domains are thiol-dependent peptidases that cleave their substrates after a GG motif. Interestingly, the catalytically invariant cysteine is replaced by a tyrosine in the C39-like domain of HlyB. Here, the overexpression, purification and crystallization of the isolated C39-like domain are described as a first step towards obtaining structural insights into this domain and eventually answering the question concerning the function of a degenerated C39 domain in the ABC transporter HlyB.

1. Introduction

The ATP-binding cassette (ABC) transporter HlyB is a central element of the type I secretion machinery of *Escherichia coli*. Together with the inner membrane protein HlyD, which interacts with the outer membrane protein TolC, HlyB forms a continuous tunnel from the cytoplasm to the exterior to secrete the 110 kDa toxin HlyA without any periplasmic intermediates (Balakrishnan *et al.*, 2001; Mackman *et al.*, 1986, 1987; Benabdelhak *et al.*, 2003; Gray *et al.*, 1986).

In general, a functional ABC transporter consists of two nucleotide binding domains (NBDs) and two transmembrane domains (TMDs). However, and in contrast to this general blueprint, HlyB contains an additional N-terminal domain of about 150 amino acids that is conserved in bacteriocin ABC transporters (Wu & Tai, 2004; Ishii *et al.*, 2010). A sequence alignment classifies the N-terminal region of HlyB as a C39 domain. This family of cysteine peptidases cleave their substrates after a GG motif using a catalytic dyad consisting of a conserved cysteine and histidine. In HlyB, however, the catalytic dyad is degenerated: the highly conserved cysteine residue is mutated to tyrosine (Ishii *et al.*, 2010; Wu & Tai, 2004). Thus, this domain is very likely to be inactive, giving rise to the name 'C39-like domain'.

Curiously, the C39-like domain of HlyB appears to be essential for secretion of the toxin HlyA. In a system in which HlyB lacked the C39-like domain, secretion of HlyA was completely abolished (data not shown) and therefore the question arises as to the specific role that this defective protease may play during substrate secretion. We expressed and purified the C39-like domain in *E. coli* with the objective of determining its structure using X-ray crystallography. Here, we describe the preliminary results of the crystallization of the C39-like domain of HlyB from *E. coli*.

2. Materials and methods

2.1. Cloning, expression and purification

E. coli strain BL21 (DE3) pLysS was transformed with the vector pET28b harbouring the HlyB-C39 gene (gene ID Q1R2T6) for overexpression of the C39-like domain (residues 2–144 of HlyB). Since the original sequence of the C39-like domain contained no

electronic reprint

methionine, two isoleucines (Ile16 and Ile30) were exchanged for methionines in order to facilitate the possibility of introducing selenomethionine and providing a tool for structure determination by anomalous diffraction. Additionally, the construct contained an N-terminal His_{10} tag and a linker (primary sequence SENLYFQ-GAMANS). Cells were grown at 310 K and 180 rev min^{-1} in rich medium (2×YT) to an OD_{600} of 0.6 before expression of the C39-like domain was induced by the addition of 1 m*M* isopropyl *β*-D-1-thio-galactopyranoside (IPTG). All subsequent steps were performed at 277 K. Cells from 4 l bacterial cell culture were harvested by centrifugation 5 h after induction and resuspended in 60 ml 50 m*M* HEPES pH 7.0, 150 m*M* NaCl, 10%(*v*/*v*) glycerol. After addition of DNAse (0.1 mg ml^{-1}), the cells were incubated for 20 min and disrupted three times at 2.5 MPa using a cell disrupter (Constant Systems). The homogenate was centrifuged for 30 min at 30 000*g*. Expression of the C39-like domain resulted in the formation of inclusion bodies. The pellet of inclusion bodies was washed once with 50 m*M* Tris–HCl pH 7.0, 0.5% Triton X-100, 1 m*M* dithiothreitol (DTT), 0.05% NaN_3 and once with 50 m*M* Tris–HCl pH 7.0, 1 m*M* EDTA, 0.05% NaN_3. Following the final centrifugation step, the inclusion bodies were resuspended in 20 ml buffer (6 *M* guanidinium hydrochloride, 10 m*M* Tris–HCl pH 7.0, 100 m*M* KCl, 10 m*M* DTT) until the pellet was dissolved. The protein suspension was centrifuged for 1 h at 50 000*g* and the protein concentration of the supernatant was determined using a NanoDrop device (PeqLab Biotechnology GmbH) using the molecular mass of the C39-like domain (19.46 kDa including the His_{10} tag) and its calculated extinction coefficient (25 440 M^{-1} cm^{-1}; http://expasy.org/tools/protparam.html). The resulting supernatant was diluted to 2 mg ml^{-1} in resuspension buffer and dialyzed overnight in a 1:125 ratio against renaturation buffer consisting of 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 1 m*M* EDTA, 10%(*v*/*v*) glycerol, 2 m*M* DTT, 0.5%(*w*/*v*) PEG 3350 using dialysis tubing with a molecular-weight cutoff (MWCO) of 6–8 kDa (Spektra/Por, Spectrum Laboratories, California, USA). After dialysis, any precipitate was separated from soluble protein by centrifugation for 10 min at 21 000*g*. The supernatant was applied onto an immobilized metal-ion affinity chromatography (IMAC) column for purification of the C39-like domain. However, the C39-like domain precipitated after elution. Therefore, we established a purification procedure without IMAC. The supernatant was applied onto a cation-exchange (cIEX) chromatography column (5 ml SP HP column, GE Healthcare) pre-equilibrated with buffer *A* (10 m*M* HEPES pH 7.0, 1 m*M* DTT). After washing with five column volumes (CV) of buffer *A*, proteins were eluted with a linear gradient of 40 CV to 100% buffer *B* (buffer *A* containing 500 m*M* NaCl). Proteins eluting at 190 m*M* NaCl (38% buffer *B*) and 375 m*M* NaCl (75% buffer *B*) were concentrated using an Amicon Ultra Centrifugal Filter device (10 kDa molecular-weight cutoff, Millipore) and analyzed by SDS–PAGE and Western blotting with an antibody recognizing the His_{10} tag (according to the protocol of the manufacturer, Qiagen). However, the protein fraction that eluted from the column at 375 m*M* NaCl could not be concentrated above 2 mg ml^{-1}. Nevertheless, both species were applied onto a size-exclusion chromatography (SEC) column (Superdex 200 16/60, GE Healthcare) equilibrated with 50 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 1 m*M* DTT.

2.2. Buffer screen

A buffer screen was performed with the C39-like domain fraction that eluted at 375 m*M* NaCl during cIEX chromatography. The protein (2 mg ml^{-1}) was mixed in a 1:1 ratio with commercially available crystallization buffers (Basic Crystallography Kit, Sigma–Aldrich). Buffer conditions with no visible precipitation were analyzed further. 1 ml of the C39-like domain (2 mg ml^{-1}) was dialyzed against 1 l of each buffer positively evaluated in the 96-well format using Slide-A-Lyzer dialysis cassettes (10 kDa molecular-weight cutoff, Thermo Scientific) and subsequently concentrated using filter devices (see above) until precipitation occurred. After purification and dialysis against 50 m*M* citrate pH 6.5 and 20 m*M* LiCl, the C39-like domain could be concentrated to 20 mg ml^{-1}.

2.3. SEC–MALS

SEC–MALS (SEC–multi-angle light scattering) experiments were performed using an analytical Superdex 200 10/300 column (GE Healthcare) equilibrated with 25 m*M* Tris pH 8.0 and 150 m*M* NaCl on an ÄKTA Purifier (GE Healthcare) connected to a triple-angle light-scattering detector (miniDAWN TREOS, Wyatt Technology) followed by a differential refractive-index detector (Optilab rEX, Wyatt Technology). 100 µl purified C39-like domain at a protein concentration of 1 mg ml^{-1} was loaded onto the analytical Superdex 200 10/300 column. Data were analyzed with the *ASTRA* software package (Wyatt Technology).

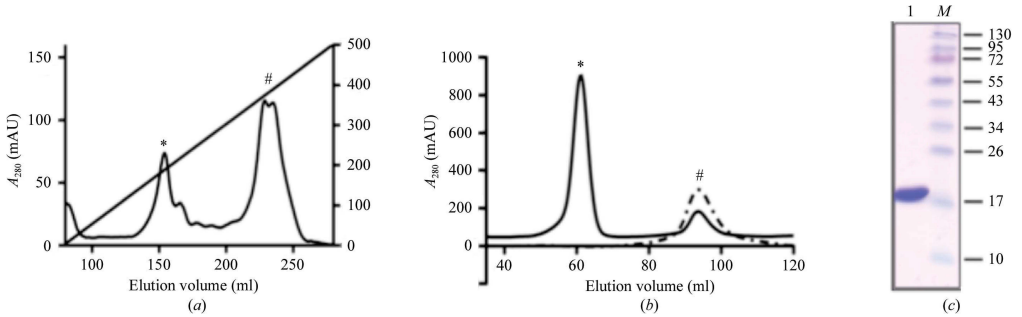


Figure 1
(*a*) cIEX column elution profile of the refolded C39-like domain. One low-salt peak (*) and one high-salt peak (#) were observed. (*b*) Size-exclusion elution profile (Superdex 200 16/60, GE Healthcare) of the low-salt (black line) and high-salt (dashed line) species. The peaks corresponding to oligomeric (*) and monomeric (#) C39-like domain are marked. (*c*) A Coomassie-stained SDS–PAGE gel of the C39-like domain after purification.

electronic reprint

2.4. Crystallization and preliminary X-ray analysis

Crystallization conditions were screened using the sitting-drop vapour-diffusion method in combination with commercially available buffers (Sigma Basic Screen I and II) using the monomeric as well as the oligomeric form of the C39-like domain at a protein concentration of 20 mg ml^{-1}. Crystals appeared when samples of protein were mixed with an equal volume of precipitant solution containing 3.6–4 *M* formate. Using the 'no-fail' cryoprotection method (http://capsicum.colgate.edu/chwiki/), 37.5% glucose in mother liquor was added gradually to the crystal solutions to give a final concentration of 20% glucose before flash-cooling the crystals in liquid nitrogen. A high-resolution data set was collected from a single crystal at 100 K at a wavelength of 0.87260 Å using a MAR 225 detector (MAR Research) on ESRF beamline ID23eh1 using 0.75° rotation per frame. The data set was processed using the *XDS* package (Kabsch, 2010) and scaled using *XSCALE*.

3. Results and discussion

The C39-like domain of HlyB was expressed as inclusion bodies in *E. coli*. After preparation and crude purification of the inclusion bodies, proteins were denatured and refolded. A cIEX step was used to purify the refolded C39-like domain (see §2). During elution with a salt gradient, two species eluted at 190 m*M* NaCl (38% buffer *B*) and 375 m*M* NaCl (75% buffer *B*) (Fig. 1). SDS–PAGE and Western-blot analysis using an anti-His-tag antibody showed that both eluted fractions corresponded to the C39-like domain (data not shown). Therefore, both species were concentrated and applied onto an SEC column. However, the resulting elution chromatograms were different. The low-salt species (190 m*M* NaCl) gave two peaks (61 and 93 ml) corresponding to an oligomeric and a monomeric species, respectively. Based on a calibration curve, the molecular weights were calculated as approximately 200 and 20 kDa, respectively. The higher oligomeric fraction of the C39-like domain yielded approximately six times more protein: 9 mg per litre of cells for the oligomer compared with 1.5 mg per litre of cells for the monomeric species. For further investigation of the molecular weight of the low-salt oligomeric species, more precise MALS experiments were performed and the molecular weight was determined to be 233.4 ± 1.4 kDa, corresponding to a dodecamer of the C39-like domain (Fig. 2). Analysis of the high-salt species (375 m*M* NaCl) by size-exclusion chromatography revealed that the species corresponded to a monomeric C39-like domain only. After purification, the C39-like domain was obtained without any visible impurities on an SDS–PAGE gel (Fig. 1*c*).

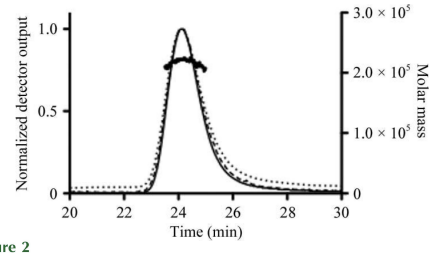


Figure 2
The oligomeric state of the C39-like domain was analyzed by SEC–MALS. The elution profile from a Superdex 200 10/300 column is shown. Straight line, normalized UV signal; dashed line: normalized LS (90°) signal; dotted line, refractive index signal. Filled circles indicate the corresponding molar masses.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

No. of crystals	1
Beamline	ID23eh1, ESRF
Wavelength (Å)	0.87260
Detector	MAR Mosaic 225 mm
Crystal-to-detector distance (mm)	200
Rotation range per image (°)	0.75
Total rotation range (°)	230
Exposure time per image (s)	1
Resolution range (Å)	20–2.1 (2.2–2.1)
Space group	$P2_12_12$ or $P2_12_12_1$
Unit-cell parameters	
a (Å)	117.6
b (Å)	134.2
c (Å)	139.4
$\alpha = \beta = \gamma$ (°)	90.0
Mosaicity (°)	0.65
Total No. of measured intensities	720832
Unique reflections	240680
Multiplicity	2.9
Completeness (%)	95.6 (97.3)
$R_{\rm merge}$ (%)†	9.0 (21.9)
$\langle I/\sigma(I)\rangle$	9.4 (4.6)
Overall B factor from Wilson plot (Å^2)	24.5

† $R_{\rm merge}$ is defined as $\sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.

The monomeric species precipitated at the relatively low protein concentration of 2 mg ml^{-1}. In our hands, this concentration was too low to obtain any protein crystals since more than 90% of the drops in the crystallization experiments did not show any precipitation at such a low protein concentration. The buffer screen performed revealed a buffer condition (20 m*M* citrate pH 6.5 and 20 m*M* $NiCl_2$) in which the protein was very stable (to above 20 mg ml^{-1}). However, despite extensive efforts no well diffracting crystals of the C39-like domain could be obtained.

Therefore, we went back to the initial oligomeric species and crystallized this form of the C39-like domain containing a His_{10} tag. Crystals were observed in various screening conditions. The best diffracting crystals (Fig. 3) were obtained after 30 d at 285 K in 3.8 *M* sodium formate. For diffraction analysis, crystals were cryoprotected using the 'no-fail' cryoprotection method (http://capsicum.colgate.edu/chwiki/). Here, crystal-containing drops were substituted step by step with small volumes of 37% glucose dissolved in mother liquor to give a final glucose concentration of 20%. Using this method, well diffracting crystals could be obtained and a native data set was collected.

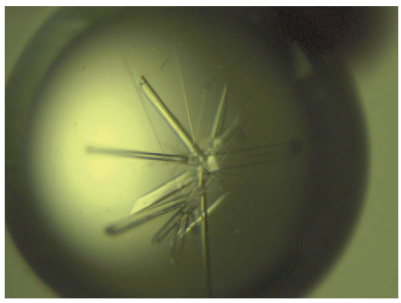


Figure 3
Crystal of the C39-like domain. The crystal has dimensions of 80 × 80 × 100 µm.

electronic reprint

Preliminary analysis of the data set using the *XDS* package revealed that the crystals belonged to a primitive orthorhombic space group (either $P2_12_12$ or $P2_12_12_1$) with large unit-cell parameters (a = 117.6, b = 134.2, c = 139.4 Å) for a protein of this size. At the moment we cannot distinguish between the two space groups since the systematic absences observed are not conclusive. A self-rotation function was calculated and revealed three twofold axes perpendicular to each other, indicating an orthorhombic space group. The presence of substantial peaks on the $\chi = 180^\circ$ and $\chi = 120^\circ$ sections of the self-rotation function indicated the presence of twofold and threefold rotational symmetry in the expected dodecamer of the asymmetric unit. Additionally, no significant peaks could be detected in the $\chi = 60^\circ$ or $\chi = 90^\circ$ sections, indicating a lack of fourfold or sixfold symmetry. The overall data set was cut off at a resolution of 2.1 Å with an $R_{\rm merge}$ value in the highest resolution shell of below 30% (see Table 1 for data statistics). Consideration of the unit-cell volume (2.2×10^6 Å^3) and the molecular weight of 19 460 Da suggests that the asymmetric unit contains one copy of the dodecamer, corresponding to a $V_{\rm M}$ value of 2.36 Å^3 Da^{-1} with an estimated solvent content of 47.61% (Matthews, 1968).

This suggests that the initially observed oligomeric species is likely to be the state in which the C39-like domain is sufficiently stable and homogenous to crystallize under the conditions used in this study.

An attempt to solve the structure of the C39-like domain taking advantage of the introduced selenomethionine residues by using multiple anomalous dispersion (MAD) phasing is currently under way. The result of this structural analysis should provide an insight into the function of the C39-like domain of HlyB and its role in the secretion of HlyA.

We would like to acknowledge Solvej Siedler for important assistance during the initial stages of this project. This work was funded by the EDICT (European Initiative on Channels and Transporters) consortium FP7 Theme (Health-2007-2.1.1-5) to BT and LS. We thank the Ministry of Innovation, Science and Research of the German Federal State North Rhine-Westphalia (NRW) and the Heinrich Heine University Düsseldorf (scholarship from the CLIB-Graduate Cluster Industrial Biotechnology to CKWS).

References

Balakrishnan, L., Hughes, C. & Koronakis, V. (2001). *J. Mol. Biol.* **313**, 501–510.
Benabdelhak, H., Kiontke, S., Horn, C., Ernst, R., Blight, M. A., Holland, I. B. & Schmitt, L. (2003). *J. Mol. Biol.* **327**, 1169–1179.
Gray, L., Mackman, N., Nicaud, J. M. & Holland, I. B. (1986). *Mol. Gen. Genet.* **205**, 127–133.
Ishii, S., Yano, T., Ebihara, A., Okamoto, A., Manzoku, M. & Hayashi, H. (2010). *J. Biol. Chem.* **285**, 10777–10785.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Mackman, N., Baker, K., Gray, L., Haigh, R., Nicaud, J. M. & Holland, I. B. (1987). *EMBO J.* **6**, 2835–2841.
Mackman, N., Nicaud, J. M., Gray, L. & Holland, I. B. (1986). *Curr. Top. Microbiol. Immunol.* **125**, 159–181.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Wu, K.-H. & Tai, P. C. (2004). *J. Biol. Chem.* **279**, 901–909.

electronic reprint

5.3 Chapter III

Title:	^{1}H, ^{15}N and ^{13}C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB).
Authors:	Lecher J., Stoldt M., Schwarz C. K. W., Smits S. H. J., Schmitt L., Willbold D.
Published in:	Biomol NMR Assign
Impact Factor:	0.707
Own proportion of this work:	10 % Production and purification protocol of the CLD.

Biomol NMR Assign (2011) 5:199–201
DOI 10.1007/s12104-011-9299-0

ARTICLE

^{1}H, ^{15}N and ^{13}C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB)

Justin Lecher · Matthias Stoldt · Christian K. W. Schwarz · Sander H. J. Smits · Lutz Schmitt · Dieter Willbold

Received: 17 January 2011 / Accepted: 10 February 2011 / Published online: 24 February 2011
© Springer Science+Business Media B.V. 2011

Abstract ATP-binding cassette (ABC) transporters are ubiquitous integral membrane proteins, which catalyze the translocation of molecules across biological membranes in an ATP-dependent manner. Despite the diversity in the transported substrates, they all share the same architecture, comprised of two transmembrane (TMD) and two nucleotide-binding domains (NBD). Members of the bacteriocin ABC transporter subfamily feature a special domain, belonging to the C39 (cystein protease family 39) peptidase protein family. These domains are assumed to cleave a C-terminal signal sequence from the protein or peptide substrate before or during the transport process. Although the C39 peptidase-like domain of the ABC transporter haemolysin B from *E. coli* shows no proteolytic activity, it is essential for the function of this transporter. In order to elucidate the contribution of the isolated C39 peptidase-like domain in the whole transport process, the backbone and side chain ^{1}H, ^{15}N and ^{13}C-NMR chemical shifts have been assigned.

Keywords ABC transporter · C39 peptidase-like · Heteronuclear NMR · Resonance assignment

J. Lecher · M. Stoldt · D. Willbold (✉)
Institute of Structural Biochemistry (ICS-6), Research Centre Juelich, Juelich, Germany
e-mail: d.willbold@fz-juelich.de

M. Stoldt · D. Willbold
Institut für Physikalische Biologie, Heinrich-Heine-Universitaet Duesseldorf, Duesseldorf, Germany

C. K. W. Schwarz · S. H. J. Smits · L. Schmitt
Institut für Biochemie, Heinrich-Heine-Universitaet Duesseldorf, Duesseldorf, Germany

Biological context

The ATP-binding cassette (ABC) transporter haemolysin B (HlyB) is the central element of a Type I transport machinery in *E. coli* (Holland et al. 2005). Together with the outer membrane protein TolC and the membrane fusion protein (HlyD), a continuous transport complex is formed, catalyzing the translocation of the 110 kDa toxin HlyA. HlyA is secreted in one step from the cytoplasm to the exterior without any periplasmic intermediates (Gray et al. 1986).

HlyB belongs to the family of bacteriocin ABC transporters, which possess two transmembrane domains (TMD), and two nucleotide-binding domains (NBD), which fuel the transport via the energy released upon ATP binding and/or hydrolysis. Bacteriocin ABC transporters feature an additional domain, which belongs to the C39 peptidase family. Peptidases of this family are cysteine proteases, which are using a catalytic triad consisting of a cysteine, a histidine and an aspartic acid residue, to hydrolyze peptide bonds. As part of ABC transporters their function is assumed to cleave the C-terminal located signal sequence C-terminal of a conserved GG-motif during or upon substrate transport (Havarstein et al. 1995). In comparison to proteolytically active C39 peptidase domains, the C39 peptidase-like domain of HlyB amino acid sequence contains a tyrosine at the position of the cysteine residue of catalytic triad, rendering it proteolytically inactive. Therefore this domain in HlyB is called C39 peptidase-like domain.

To gain a better understanding of the function of the C39 peptidase-like domain in the whole Type 1 machinery and its role during the transport of HlyA, we want to reveal the structure and dynamics of the isolated C39 peptidase-like domain of the ABC transporter HlyB. Although the

catalytic triad is degenerated, it is assumed to still bind, but not cleave a peptide bond, likely containing the GG motif (Wu and Tai 2004). Here we report the ^{1}H, ^{15}N and ^{13}C sequence-specific backbone and side chain resonance chemical shift assignment of the isolated C39 peptidase-like domain of HlyB.

Methods and experiments

Expression and purification of isotope labeled protein

The gene coding for the N-terminal 137 amino acid residues of the ABC transporter HlyB (uniprot ID Q47258) containing its C39 peptidase-like domain has been cloned into the pET22b (Novagen, Darmstadt, Germany), and the sequence has been verified by DNA sequencing. *E. coli* BL21 (DE3)-pLysS (Novagen, Darmstadt, Germany) were transformed with the plasmid DNA and grown in LB medium at 37°C up to an OD_{600} of 0.8. Cells were harvested and diluted 3-fold in M9 medium supplemented with ^{15}N-enriched ammonium chloride and ^{13}C enriched glucose. While growing at 37°C with 160 rpm shaking, protein expression was induced at an OD_{600} of 0.8 with 0.4 mM IPTG for 4 h. The N-terminal His_{10}-tagged C39 peptidase-like domain was purified from the soluble fraction via affinity chromatography using NTA-Sepharose loaded with $Ni(II)^{2+}$. Subsequently the His-tag was cleaved off of the C39 domain using a His-tagged TEV protease (Kapust et al. 2001) leaving 4 additional N-terminal amino acid residues that do not exist in the wildtype-sequence. The released His-tag and the protease were separated from the C39 peptidase-like domain using a second affinity chromatographic (IMAC) step. Here the C39 peptidase-like domain run through the column and ended up in the flowthrough. Final purity was achieved using a size exclusion chromatography column (HighLoad SD75 16/60 (GE Healthcare)). For NMR experiments the pooled fractions were concentrated by ultrafiltration with a VivaSpin 2 (Sartosius Stedim, Goettingen) device up to 2.5 mM.

NMR spectroscopy

The NMR samples contained [U-^{15}N] or [U-^{13}C,^{15}N] labeled C39 peptidase-like domain at concentrations between 1.5 and 2.5 mM in aqueous solution with 100 mM lithium chloride, 100 mM deuterated citrate (D4, 98%, CIL), 10 mM deuterated dithiothreritrol (D6, 99% Sigma–Aldrich), 0.5 mM EDTA, 7% (v/v) $^{2}H_2O$ and 0.05% (w/v) sodium azide.

NMR spectroscopic experiments were performed at 303 K on Varian VNMRS instruments at proton frequencies of 800 and 900 MHz, each equipped with a cryogenically cooled Z-axis pulse-field-gradient (PFG) triple resonance probe.

Resonance assignment was accomplished using the following experiments:

2D (^{1}H-^{15}N)-HSQC, 2D ct-(^{1}H-^{13}C)-HSQC, 3D HNCACB, CBCA(CO)NH, HBHA(CO)NH, HNHA, HNCO, H(CCO)NH, C(CO)NH, H(C)CH-COSY (20 ms mixing time), (^{1}H-^{1}H-^{15}N)-NOESY-HSQC (150 ms mixing time) and (^{1}H-^{13}C-^{1}H)-HSQC-NOESY (100 ms mixing time). Aromatic side chain resonances were assigned using 2D ct-(^{1}H-^{13}C)-HSQC and 3D (^{1}H-^{13}C-^{1}H)-HSQC-NOESY (120 ms mixing time). Suppression of water was performed by application of WATERGATE technique. Decoupling of ^{15}N and ^{13}C nuclei during ^{1}H acquisition was achieved using GARP and adiabatically by WURST sequences. ^{1}H chemical shifts were referenced directly to DSS at 0 ppm and ^{13}C and ^{15}N shifts were referenced indirectly to DSS, using the absolute frequency ratios.

All NMR data was processed using NMRPipe v.5.5.2010 (Delaglio et al. 1995) or Azara v.2.8 (provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge) and evaluated with CcpNmr v.2.1.5 (Vranken et al. 2005) software packages.

Assignments and data deposition

The interpretation of triple resonance spectra led to a sequence-specific backbone resonance assignment of 99%. Backbone N of the three prolines and C' of their preceding amino acid resonances have not been assigned.

Aliphatic and aromatic side chain proton, nitrogen and carbon resonances assignments are completed to 97% (excluding ^{1}H-^{15}N and ^{15}N of lysine, OH, SH, side chain $^{13}C'$, $^{13}C^{\zeta}$, and aromatic quaternary ^{13}C).

Several unusual chemical shifts have been observed around histidine residue 83, which is annotated to be of catalytic importance. The chemical shifts of L43 $H^{\delta 2}$, V76 H^{β}, R78 H and N and H83 $C^{\varepsilon 1}$, $H^{\varepsilon 1}$, $N^{\delta 1}$ and $N^{\varepsilon 2}$ deviate significantly from the statistical mean, but have been confirmed in several spectra. The side chain H^{ε} of H83 was observed in the (^{1}H-^{15}N)-HSQC (see Fig. 1 inlay) as well as in 3D NOESY spectra. To distinguish between both tautomeric forms of the histidine side chain, a long-range (^{1}H-^{15}N)-HMQC (Pelton et al. 1993) was used.

Secondary chemical shifts of the backbone and C^{β} resonances were analyzed using Talos+ (Shen et al. 2009), predicting 6 α-helical and 6 β-stranded segments. The α-helices comprise residues Y9-Y20, E28-R33, L43-S52, I64-L67, E101-Q103 and Q111-L117, whereas the β-strands lie between V57-K62, P73-W77, F84-V89, Y96-D100, R107-L109 and H121-A126.

The data for the ^{1}H, ^{15}N and ^{13}C chemical shifts of the C39 domain have been deposited in the BioMagResBank

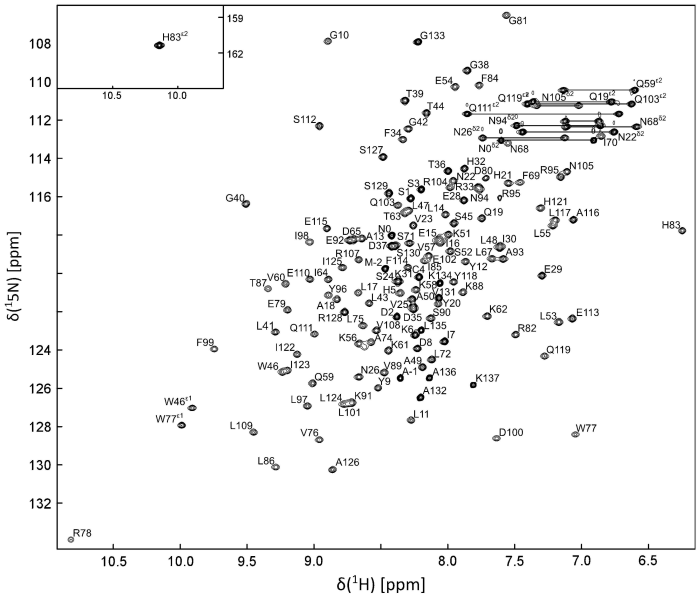


Fig. 1 2D (^{1}H-^{15}N)-HSQC spectrum of a 2.5 mM sample of [U-^{15}N]-labeled C39 peptidase like domain of HlyB from *E. coli* recorded at 900 MHz (T = 303 K, pH = 6.0, 7% $^{2}H_2O$). Backbone resonance assignments are indicated by *one-letter* amino acid code and sequence number. Positions of side chain resonances are indicated. The side chain amide groups of asparagine and glutamine residues are connected by *horizontal lines*. The *inlay* shows the side chain resonance $H^{\varepsilon 2}/N^{\varepsilon 2}$ of histidine 83

(http://www.bmrb.wisc.edu) under the accession number 17403.

Acknowledgments We would like to thank Melanie Schwarten for helpful discussions. The research has been funded by a fellowship from the International Helmholtz Research School on Biophysics and Soft Matter ("Biosoft") to Justin Lecher. We thank the Ministry of Innovation, Science, and Research of the German Federal State North Rhine-Westphalia (NRW) and the Heinrich-Heine-Universitaet Duesseldorf (scholarship from the CLIB-Graduate Cluster Industrial Biotechnology for Christian Schwarz).

References

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293

Gray L, Mackman N, Nicaud JM, Holland IB (1986) The carboxy-terminal region of haemolysin 2001 is required for secretion of the toxin from Escherichia coli. Mol Gen Genet 205(1):127–133

Havarstein LS, Diep DB, Nes IF (1995) A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. Mol Microbiol 16(2):229–240

Holland IB, Schmitt L, Young J (2005) Type 1 protein secretion in bacteria, the ABC transporter dependent pathway (review). Mol Membr Biol 22(1–2):29–39

Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14(12):993–1000

Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from Escherichia coli, using two-dimensional heteronuclear NMR techniques. Protein Sci 2(4):543–558. doi:10.1002/pro.5560020406

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44(4):213–223. doi:10.1007/s10858-009-9333-z

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59(4):687–696. doi:10.1002/prot.20449

Wu KH, Tai PC (2004) Cys32 and His105 are the critical residues for the calcium-dependent cysteine proteolytic activity of CvaB, an ATP-binding cassette transporter. J Biol Chem 279(2):901–909. doi:10.1074/jbc.M308296200

5.4 Chapter IV

Title:	Rational and Irrational Approaches to Convince a Protein to Crystallize
Authors:	Abts A., Schwarz C. K. W., Tschapek B., Smits S. H. J., Schmitt L.
Published in:	InTech
Impact Factor:	not available
Own proportion of this work:	30 % Writing of the manuscript.

22

Rational and Irrational Approaches to Convince a Protein to Crystallize

André Abts, Christian K. W. Schwarz, Britta Tschapek,
Sander H. J. Smits and Lutz Schmitt
Institute of Biochemistry, Heinrich-Heine University, Düsseldorf, Germany

1. Introduction

The importance of structural biology has been highlighted in the past few years not only as part of drug discovery programs in the pharmaceutical industry but also by structural genomics programs. Mutations of human proteins have been long recognized as the source of severe diseases and a structural knowledge of the consequences of a mutation might open up new approaches of drugs and cure. Although the function of a protein can be studied by several biochemical and/or biophysical techniques, a detailed molecular understanding of the protein of interest can only be obtained by combining functional data with the knowledge of the three-dimensional structure. In principle three techniques exist to determine a protein structure, namely X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and electron microscopy (EM). According to the protein data bank (pdb; http://www.rcsb.org) that provides a general and open-access platform for structures of biomolecules, X-ray crystallography contributes more than 90% of all structures in the pdb, a clear emphasis of the importance of this technique.
To perform X-ray crystallography it is essential to have large amounts of pure and homogenous protein to perform an even today still "trail and error"-based screening matrix to obtain well diffracting protein crystals. Therefore, successful protein crystallization requires three major and crucial steps, all of them associate with specific problems and challenges that need to be overcome and solved. These steps are (I) protein expression, (II) protein purification and (III) the empirical search for crystallization conditions. As summarized in Figure 1, every single step needs to be optimized along the long and stoney road to obtain protein crystals suitable for structure determination of your "most-beloved" protein via X-ray crystallography. This chapter will focus on these three steps and suggests strategies how to perform and optimize each of these three steps on the road of protein structure determination.

2. Protein expression (I)

To crystallize a protein, the first requirement is the expression of your protein in high amounts and most importantly on a regular basis. This implies that it is possible to obtain a freshly purified protein at least weekly. In general, it is possible to express a protein either homologously or heterologously (see Figure 1 – (I) expression). Especially for large proteins,

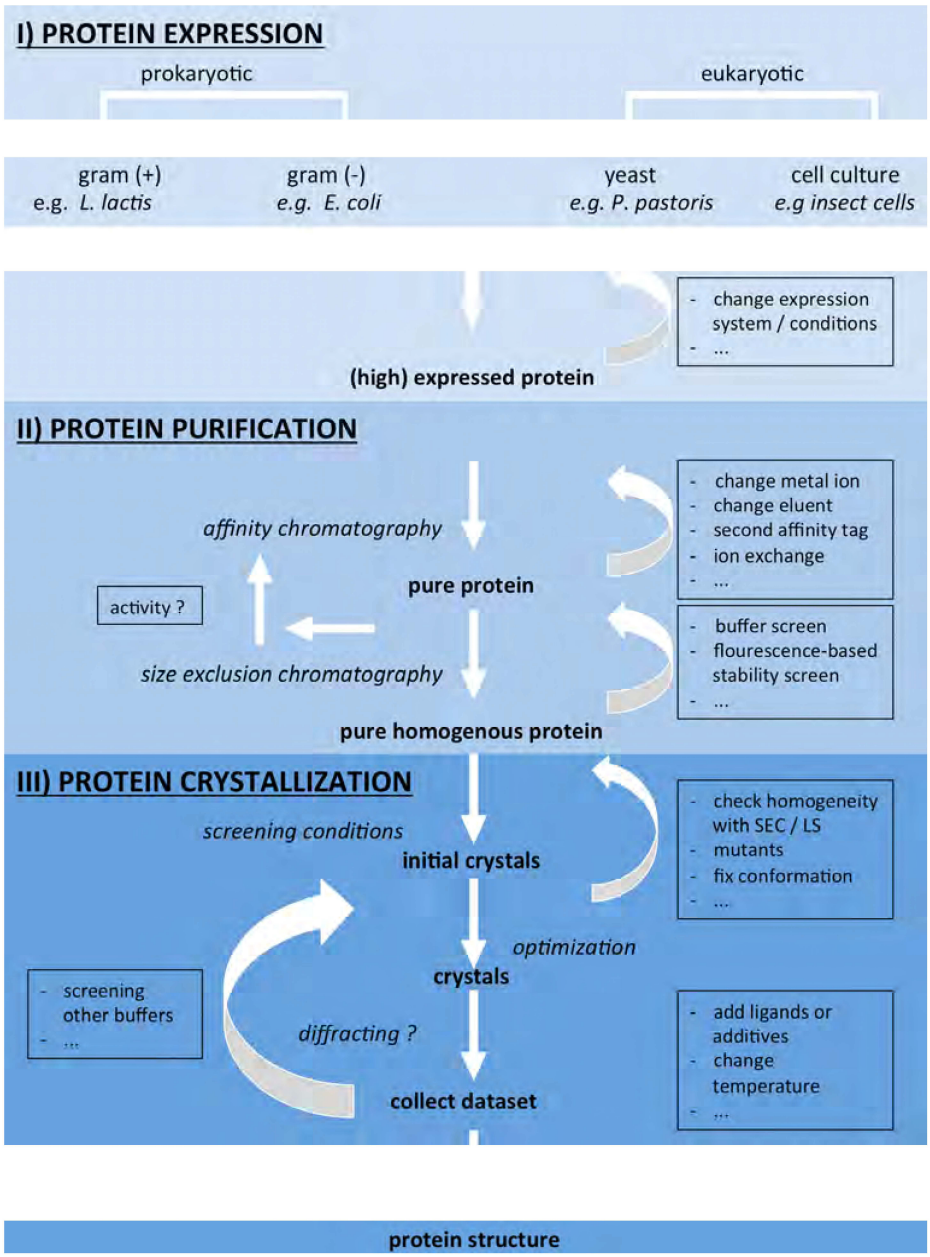


Fig. 1. Schema highlighting the three steps towards a protein crystal. (I) Expression (II) Purification (III) Protein crystallization.

proteins containing a co-factor or a ligand, the natural habitat is likely the best choice to express the protein. However, often the natural host, for example humans, produce only low amounts of protein and suitable overexpression protocolls are not available. To circumvent this problem, several expression strains, cell lines as well as a large number of expression vectors have been developed to allow expression of any protein in a different host (heterologous expression). In general, the used organisms for protein expression can be divided into two different groups: prokaryotic and eukaryotic expression hosts. The natural organism of the protein of interest mainly dictates the choice, which expression system to use. If working with a bacterial protein, it is very likely that also a prokaryotic host is able to express the protein in high amounts. The same holds true for proteins originating from a eukaryotic host, which is likely best overexpressed in a eukaryotic host. These proteins often require posttranslational modification such as glycosylation or disulfide bond formation, which are possible in eukaryotic expression hosts. The most common used heterologous expression host is the gram negative bacteria *Escherichia coli* since it is commercially available and a large number of expression cassettes have been developed. Thus, it is the most widely used expression system with expression rates of several mg/L of culture. The best characterised and understood expression hosts are described in more detail below and the commecially available systems are listed in table 1.

2.1 Expression hosts – Prokaryotic

2.1.1 Gram negative - *E. coli*

As mentioned above, *E. coli* is the most common used expression system (Figure 1 – (I) expression, left side). This is further highlighted by the fact that 80% of all protein structures deposited in the protein data bank were overexpressed in *E. coli* (Sorensen and Mortensen 2005). There are several advantages promoting *E. coli* as expression host: (A) Cultivation of *E. coli* is simple and a doubling time of 30 minutes is rather quick allowing the fast generation of biomass, (B) genetics are well understood and any genetical manipulation is well established, (C) expression levels of up to 60 % of the total protein mass within the cell make the next step, protein purification rather straight forward and finally (D) the cultivation does require only standard equipment normally present in every biochemical laboratory and therefore expression using *E.coli* is relatively cheap. In the last decades, many different plasmid based expression systems have been developed such as the pET vector systems, which contain several different expression plasmids with a choice for the affinity tag on either termini of the protein as well as the possiblity to use a dual cassette when expressing two or more proteins at once. The selection pressure derived from different antibiotics, and the resistance genes encoded on these plasmids further simplify laboratory practice. Only cells harbouring the right plasmids are able to grow and therefore expenditure on sterility is low.

The typical *E. coli* expression system is plasmid-based, which can be transferred to different *E. coli* strains ((Sorensen and Mortensen 2005), Novagen pET vector table). An *E. coli* expression vector consists mainly of five important parts: the replicon, a resistance marker, a promotor and a so-called multiple cloning site (MCS) (Baneyx 1999; Jonasson, Liljeqvist et al. 2002). The replicon is the crucial part of a plasmid to maintain it inside a cell. It is recognized and duplicated by the replication machinery (Baneyx 1999). The selection marker allows the identification of cells carrying a plasmid as it encodes for a resistance, e.g. against antibiotics (see above)(Sorensen and Mortensen 2005). The promotor sequence is the recognition site for the RNA polymerase, however it is inactive under initial cultivation conditions. The addition of an inducer (sometimes also a temperature change) switches the

promotor from 'off' to 'on' whereby the expression is initiated (Jana and Deb 2005). Common inducers are isopropyl-β-D-thiogalactopyranosid (IPTG) in the pET system or arabinose for pBAD vectors (Invitrogen™). The multiple cloning site (MCS) is a short DNA segment combining many (up to 20) restriction sites. This feature simplified the insertion of genes into the plasmid enormously and made cloning procedures very convenient. However, new cloning strategies, which are independent of restriction enzymes and ligases, are emerging and will replace the standard approaches some day (see for example Li and Elledge 2007).

2.1.2 Gram positive - *L. lactis*

Within the prokaryotic expression system also some gram (+) bacteria are used for protein overexpression (Figure 1 – (I) expression left side). Here, lactic acid bacteria play a privotal role. They are used in food industry and are known since 1873 when Joseph Lister isolated the first strain (Teuber 1995; Mierau and Kleerebezem 2005). For the overexpression of recombinant proteins, there are many lactic acid bacteria around, *lactococci, lactobacilli, streptococci* and *leuconostocs*. (Gasson 1983; van de Guchte, Kok et al. 1992; de Vos and Vaughan 1994). The best characterized and most widely used host is *Lactococcus lactis*, which is famous for its usage within food fermentation and like for *E. coli* the genome, metabolisms and molecular modifications are well known and established (Bolotin, Wincker et al. 2001; Guillot, Gitton et al. 2003). Thus, it has been called the 'bug of the next millennium'(Konings, Kok et al. 2000). All established gram (+) bacteria expression hosts are able to overexpress proteins homologously or heterologously. Since only one, namely the cyctoplasmic membrane is present (Kunji, Slotboom et al. 2003), this host is in comparison to the two-membrane system of gram (-) bacteria, a good choice to express eukaryotic as well as prokaryotic membrane proteins or proteins with membrane anchors (Kunji, Slotboom et al. 2003). The promoter used for expression in *L. lactis* is induced by the external addition of nisin. Nisin is an antimicrobial active peptide, which interacts with lipid II in the cytoplasmic membrane of gram (+) bacteria and causes cell lysis. Interestingly, nisin is produced by *L. lactis* itself. The expression strain is deleted of the nisin producing genes and therefore external nisin can be used as inductor. Nisin binds to NisK, which as part of a two-component system, phosphorylates NisR, which in turn binds to the promotor P_{nisA} thereby allowing synthesis of the protein located downstream on the plasmid. Since nisin is also active against *L. lactis* itself, the concentration range of nisin used in such expression studies is relatively narrow to circumvent killing of the *L. lactis* expression strain. This is clearly a draw back of this expression system since expression can basically only turned on with a certain nisin concentration. Between the inducer concentration and the expressed protein a linear behaviour is observed. Unfortunately the nisin concentration range between the minimal and maximum nisin concentration is very small, nisin concentration higher than 25 ng per liter of cells, cause cell death.

However, *L. lactis* has been proven to be a very efficient expression system. Kuipers et al. created many expression hosts and plasmids to produce any protein of interest by cloning it downstream of the P_{nisA} promotor. With this nisin inducible (NICE) expression system, it is now possible to induce the protein production with minimal concentration (0.1 – 5 ng) of nisin (de Ruyter, Kuipers et al. 1996; Kuipers, de Ruyter et al. 1998). The amount of produced recombinant protein can reach up to 50 % of the total intracellular proteins (Kuipers, Beerthuyzen et al. 1995; de Ruyter, Kuipers et al. 1996). Following a few examples for expressed proteins in *L. lactis* whose structures have been solved: an ECF-type ABC transporter (PDB:3RLB)(Erkens, Berntsson et al. 2011), a peptide binding protein OppA

(PDB:3RYA)(Berntsson, Doeven et al. 2009) and the multidrug binding transcriptional regulator LmrR (PDB:3F8B)(Madoori, Agustiandari et al. 2009).

2.2 Eukaryotic expression hosts

The great benefit of choosing a eukaryotic host for overexpression of a protein of interest are the availability of a posttranslational modification system as well as the frequently enhanced protein folding (Midgett and Madden 2007). Eukaryotic proteins tend to misfold or lack biological activity when expressed in prokaryotic expression systems such as *E. coli* (Cregg, Cereghino et al. 2000; Midgett and Madden 2007). To overexpress these proteins, different yeast strains, insect cells or even mammalian cell lines have been developed as expression hosts (Figure 1 - (I) expression, right side). Eukaryotic expression systems are often more expensive, provide low expression levels and are sometimes hard to handle, when compared to bacterial systems. However, the genetic and cellular contexts are more similar to the original protein-expressing organism (Midgett and Madden 2007). In the following sections, some of the commonly employed eukaryotic expression systems will be described.

2.2.1 Yeast expression systems - *Saccharomyces cerevisiae* and *Pichia pastoris*

The most widely used yeast strains to express protein are *Saccharomyces cerevisiae* and *Pichia pastoris*, which offer the major advantage of a posttranslational modification system for glycosylation, proteolytic processing as well as disulfide bond formation, which for some proteins are essential for the function and/or correct folding (Cregg, Cereghino et al. 2000; Midgett and Madden 2007). The handling of yeast expression systems is similar to prokaryotic systems with respect to the genetic background and cultivation. Similar to the bacterial vector systems, expression in yeast starts with a plasmid-based cloning part which can be performed in *E. coli* (Cregg 2007). Afterwards the expression cassette gets integrated into the genome by simple homologous recombination in the yeast. One major advantage in *P. pastoris* is the insertion of multiple copies of the protein DNA-sequence into genomic DNA, which increases expression yield.

The biggest advantage of yeast as expression system is that well established protocols for fermentation are available. Optimal fermentation of *P. pastoris* can end up with more than 130 gram of cells per liter of culture. Even if expression levels in the cell are not that high the mass of cells easily compensates for this disadvantage (Wegner 1990; Cregg, Cereghino et al. 2000; Hunt 2005; Cregg 2007; Midgett and Madden 2007). Examples of crystal structures from proteins expressed in *P. pastoris* are a human monoamine oxidase B (PDB:3PO7) (Binda, Aldeco et al. 2010) and a protein involved in cell adhesion NCAM2 IG3-4 (PDB:2XY1)(Kulahin, Kristensen et al. 2011).

2.2.2 Insect cells

The expression system in insect cells is beside yeast a well-characterised alternative to express eukaryotic proteins (Midgett and Madden 2007). As insect cells are higher eukaryotic systems their posttranslational modification machinery can carry out more complex alterations than yeast strains. They also have a machinery for the folding of mammalian proteins. The most commonly used vector system for recombinant protein expression in insect cells is baculovirus, which can also be used for gene transfer and expression in mammalian cells (Smith, Summers et al. 1983; D., L.K. et al. 1992; Altmann, Staudacher et al. 1999). A few examples of proteins expressed in insect cells that resulted in

crystal structures are the transferase Ack1 (PDB:3EQP)(Kopecky, Hao et al. 2008), a human hydrolase (PDB: 2PMS)(Senkovich, Cook et al. 2007) and myosin VI (PDB:2BKI)(Menetrey, Bahloul et al. 2005).

2.2.3 Mammalian cell lines

The expression of proteins in mammalian cell lines is the most expensive and complex alternative. Especially for human membrane proteins this expression system has been proven to express the most active protein (Tate, Haase et al. 2003; Lundstrom 2006; Lundstrom, Wagner et al. 2006; Eifler, Duckely et al. 2007). The resulting protein amount, however, obtained from mammalian cell lines is mostly only sufficient for functional studies. Using mammalian cells lines is the most challenging variant of protein overexpression and therefore only choosen if any of the other expression system described failed. Some examples of protein structures expressed in mammalian cell lines are the hydrolase PCSK9 (PDB:2QTW)(Hampton, Knuth et al. 2007) and the acetylcholine receptor AChBP (PDB:2BYQ)(Hansen, Sulzenbacher et al. 2006).
Table 1 sums up advantages and disadvantages of the above mentioned overexpression systems used for protein crystallography.

Expression system	Pros	Cons
Prokaryotic		
Gram negative *E. coli*	- costs - simplicity (genetic/culture) - yield	- inclusion bodies - protein folding - posttranslational modifications - protein secretion
Gram positive *L. lactis*	- costs - protein secretion - one membrane	- posttranslational modifications
Eukaryotic		
Yeast	- high cell densities - costs - simple cultivation - posttranslational modifications	- hyper glycosylation, non-native - lipid composition is different to mammalian cells
Insect cells	- more native lipid environment - good track record of functional proteins	- costs - non-native lipid environment - glycosylation pattern different to mammalian cells - protein amount
Mammalian cells	- native conditions for human membrane proteins to investigate diseases - posttranslational modifications - lipid environment - good track record of functional protein	- costs - difficult to establish - protein amount

Table 1. Overview of expression systems. Summarized are the advantages and disadvantages.

3. Purification

After having expressed your protein of interest, the race for crystals is by no means finished. The next step on the long road to structure determination is to isolate the protein or - phrasing it differently - to remove all other proteins present in the cell (Figure 1 – (II) purification). An elegant method to do so is the genetic attachment of an affinity tag on either site of the protein or in some cases on both sides (Waugh 2005). This affinity tag has the possibility to bind high affine to a immobilized ligand on a matrix, while all other proteins have a much more reduced binding affinity and therefore flow through the matrix (Figure 1 – (II) purification 1st step). This allows a one-step purification, which in almost all cases is relatively harmless for the protein and likely does not interfere with folding and/or overall structure of the protein. There are a lot of affinity tags available as well as matrix materials (Terpe 2003). The well known and most often used affinity tag is the poly-histidine tag (Porath, Carlsson et al. 1975; Gaberc-Porekar and Menart 2001), which can vary in length as well as in position but the overall purification strategy is the same. From all the structures solved nowadays, almost 60 % of the proteins are purified via a histidine tag; mainly due to the great purification efficiency, which can be as large as 90% after a single purification step (Gaberc-Porekar and Menart 2001; Arnau, Lauritzen et al. 2006). Therefore, most commercially available expression systems and methods contain a his-tag encoded on the plasmid. Besides the his-tag, there are other tags avaible and used for protein purification, of which the Strep-, CBP-, GST-, MBP-tag are described below.

3.1 Choice of the right tag

3.1.1 Polyhistidine-tag (his-tag)

As mentioned above the polyhistidine-tag is the most common affinity tag and the required affinity resins and chemicals are relatively inexpensive. The purification step is a so-called immobilized metal ion affinity chromatography (IMAC) (Porath, Carlsson et al. 1975). Here, a matrix is able to bind bivalent metal ions. For example nitrilotriactetic acid (NTA), which is a chelator and binds metal ions like Ni^{2+}, Zn^{2+}, Co^{2+} or Cu^{2+} (Hochuli, Dobeli et al. 1987). These metal ions have a high affinity to the imidazole group of the amino acid histidine. A stretch of histindines in a row with for example an *E. coli* protein is very unusual. Thus, **t**he genetical introduction of several, in most cases 6- 10 histidines in a row selects for specific binding of this protein. As eluant very elegantly imidazole can be used, which competes with the histidine tag and elutes the protein of interest. When used in low concentrations, imidazole can also be used to remove unspecifically bound proteins, which bind with low affinity to the matrix (Hefti, Van Vugt-Van der Toorn et al. 2001). Normally, a protein with a 6-10 histidine tag should be bound to the matrix relative strongly and 100-250 mM imidazole in the buffer is required to elute the protein from the resin. In contrast, proteins with a low affinity to the matrix can already be eluted with 10-50 mM imidazole (the "impurities" of *E. coli*). Therefore, a linear imidazol gradient, for example, separates the protein of interest and impurities (Hochuli, Dobeli et al. 1987; Gaberc-Porekar and Menart 2001). Although the polyhistidine-tag is the most common and mostly an efficient variant, there are a few applications where the his-tag can cause problems. Metalloproteins can interact either directly with the his-tag or with the ions immobilized on the matrix. In comparison to some other affinity-tags, the specificity of the his-tag is not that high and in some cases this results in the co-purification of other proteins (Waugh 2005).

3.1.2 Strep-tag

In comparison to the his-tag, which binds to immobilized metal ions, the strep-tag II constists of a small octapeptide (WSHPQFEK), which binds to the protein streptavidin (Schmidt, Koepke et al. 1996). The commercial available matrix is a streptavidin variant and is called Strep-Tactin. This variant is able to bind the Strep-tag II octapeptide under mild buffer conditions and can be gently eluted with biotin derivates such as desthiobiotin (Schmidt, Koepke et al. 1996; Voss and Skerra 1997). Especially for metal-ion containing enzymes it is a promising alternative to the his-tag (Groß, Pisa et al. 2002). However, as chemicals are more expensive and the matrix has a lower binding capacity, compared to NTA resins, it is often not the first option choosen. Moreover, it cannot be used under denaturating conditions since Strep-Tactin denatures and will not bind the tag anymore (Terpe 2003; Waugh 2005). Examples of proteins crystallized after a Strep-tag purification are OpuBC (PDB:3R6U)(Pittelkow, Tschapek et al. 2011) and *Af*ProX (PDB:3MAM)(Tschapek, Pittelkow et al. 2011) as well as the sodium dependent glycine betain transporter BetP from *Corynebacterium glutamicum* (PDB:2WIT)(Ressl, Terwisscha van Scheltinga et al. 2009).

3.1.3 CBP-tag

Another peptide tag, is the calmodulin binding peptide, first described in 1992 (Stofko-Hahn, Carr et al. 1992). This peptide is prolonged compared to the Strep-tag II, consisting of 26 amino acids and binds with nanomolar affinity to calmodulin in the presence of Ca^{2+} (Blumenthal, Takio et al. 1985). It is derived from the C-terminus of the skeletal-muscle myosin light-chain kinase, which makes the system an excellent choice for proteins expressed using a prokaryotic expression system, since in prokaryotic systems nearly no protein interacts with calmodulin. This allows extensive washing to remove impurities and elution with EGTA, which complexes specifically Ca^{2+}, and a protein recovery around 90 % can be achieved (Terpe 2003). A drawback of this tag however is that the CBP tag can only be fused to the C-terminus of the protein since it has been shown that CBP on the N-terminus negatively influences the translation and thereby the expression rate (Zheng, Simcox et al. 1997).

3.1.4 GST-tag

With respect to the length of the tags, the his-tag contains only a few amino acids, the Strep-tag II and the CBP-tag already contain 8 – 26 amino acids, but it is possible to fuse whole proteins with 26 – 40 kDa to a recombinant protein. Here, the high affinity binding of the protein to their substrate is used to purify the protein of interest (Smith and Johnson 1988). In the case of the glutathione S-transferase (GST, 26 kDa) the protein specifically binds to immobilized glutathione. To elute the fusion protein from the resin, non-denaturating buffer conditions employing reduced glutathione are used (Terpe 2003). The tag can help to protect the recombinant protein from degradation by cellular proteases. It is recommended to cleave off the GST-tag after purification with a specific protease like thrombin or TEV (Tobacco Etch Virus) protease (Terpe 2003).

3.1.5 MBP-tag

Another affnitiy tag, which can be fused to the protein of interest, is the maltose binding protein (MBP) from *E. coli*. This protein has a molecular weight of 40 kDa and has the ability to bind to a cross-linked amylose matrix. The binding affinity is in the micro molar range

and the tag can be used in a pH range from 7.0 – 8.0, however, denaturating buffer conditions are not possible (di Guan, Li et al. 1988). The elution of the recombinant protein is recommended with 10 mM maltose. A great opportunity of the MBP-tag is the increasing solubility effect of the recombinant protein in prokaryotic expression systems and even more pronounced in eukaryotic systems (Sachdev and Chirgwin 1999). Like the CBP-tag, a fusion at the N-terminal side might influence translation and expression rates (Sachdev and Chirgwin 1999).

3.1.6 Tag position and double tags

As described above, the position of the tag either at the N- or C-terminus has a considerable influence on translation and expression rate as well as on the biological function (Arnau, Lauritzen et al. 2006). If information regarding activity of the protein is already available especially about the location of interaction sites, this should be included in the protein design, meaning tag position etc. In general, the tag should be placed at the position of the protein, which is less important for interactions and/or expression. To minimize the influence of the tag on folding and/or activity in some cases it helps to create a linker region of a few amino acids between the tag and the protein (Gingras, Aebersold et al. 2005). A very efficient and sophisticated solution is, the addition of amino acids between tag and protein of interest, which functions not only as an accessibility increasing factor, but, also encodes for a recognition site for proteases like thrombin or TEV. Due to this arrangement the tag – protein interaction is minimized and the tag can be cleaved off if necessary (Arnau, Lauritzen et al. 2006). In some special cases a combination of two affinity tags results in enhanced solubility and more efficient purification. To enhance the purity of a protein, often a construct of two different short affinity tags like his-tag and Strep-tag or CBP-tag can be engineered (Rubio, Shen et al. 2005). Also a combination of two his-tag or two strep-tag kept apart by a linker region enhances the binding affinity extremely. This allows more stringent washing steps prior to elution of the protein (Fischer, Leech et al. 2011).

3.2 Size exclusion chromatography and ion exchange chromatography

Despite the usage of affinity tags a second purification step is sometimes required (Figure 1 – (II) purification). Which kind of purification procedure is required depends on the nature of impurities. If these impurities differ in molar mass compared to the protein of interest, a method based on size separation can be applied. Size exclusion chromatography (SEC) also separates different oligomeric species of the protein from each other, which otherwise would strongly inhibit crystallization and also allows analysis of stability and monodispersity of the protein (Regnier 1983a; Regnier 1983b).

However, in many cases, SEC is not sufficient to remove all impurities. Then separation by overall charge of the protein might be an option. Depending on the isoelectric point of the protein either anion or cation exchange chromatography can be performed. The protein binds to a matrix under very low ionic strength and is eluted afterwards either by increasing the ionic strength or by pH variation. Similar results can be achieved by hydrophobic interaction chromatography. Here, proteins with different surface properties show differences in their binding strength and binding of the protein is done inversely as during ion exchange chromatography. High ionic strength favors protein binding to a hydrophobic matrix and elution takes place when reducing the ionic strength. Although there are many other possibilities to increase the purity of a protein, the above mentioned techniques are without any doubt the most widely used and general applicable methods.

3.3 How to get a homogenous protein solution

In some cases isolated proteins are stable and homogenous at high concentrations after the purification and can be directly used for crystallization experiments. Often, however, the protein does not behave ideal and precipitates at high concentrations or forms aggregates or inhomogenous, oligomeric species; all of them prohibit crystal growth. SEC is a very elegant method to visualize the stability and oligomeric state of a protein. If the stability or the homogeneity of a protein sample is critical, you need to adapt your purification protocol and search for an optimized procedure. Different approaches are summarized below, for example a buffer screen to enhance protein solubility, multi-angle light scattering experiments to determine the absolute mass and the oligomeric state of the protein sample or fluorescence-based experiments to investigate the stability of the protein of interest.

3.3.1 Purified proteins – An *in vitro* system

After a protein is expressed in a soluble form, the subsequent purification procedure changes the environment of the protein dramatically. The cytoplasm of the cells, where the overexpression takes place, is packed with macromolecules. In *E. coli*, for example, the concentrations of proteins, RNAs and DNAs are about 320 mg/mL, 120 mg/mL, and 18 mg/mL, respectively (Cayley, Lewis et al. 1991; Zimmerman and Trach 1991; Elowitz, Surette et al. 1999) resulting in an overall concentration of macromolecules of above 450 mg/mL. During cell lysis and the first purification step, likely an IMAC (see above), the protein is separated from almost all other cell components. This rigorous procedure is accompanied with a severe change of the environment into an *in vitro* system. As a result proteins often tend to aggregate, precipitate or form inhomogeneous oligomeric states that prevent the formation of crystals in further experiments. Therefore one of the biggest challenges in structural studies is the preparation of protein solutions with high concentrations (as a rule of thumb 10-20 mg/mL) in a homogenous state. To fulfill these requirements, the *in vitro* system needs to be optimized with respect to different parameters as highlighted in *Figure 1 – (II) purification*. If a sufficient protein sample cannot be obtained, different strategies are available to increase the important characteristics of the protein: purity and homogeneity. As mentioned above, the usage of different metal ions during IMAC, ion exchange, a second affinity chromatography etc. can be sufficient to enhance purity. This might also lead to an increased stability. However, if the stability and/or homogeneity of a protein is still a problem, screening for a new buffer composition is essential to succeed during crystallization trials.

3.3.2 Buffer composition

Many examples illustrate the importance of an adequate buffer composition for protein stability, homogeneity, conformation, and activity (Urh, York et al. 1995; Holm and Hansen 2001; Jancarik, Pufan et al. 2004; Collins, Stevens et al. 2005). Some buffers are very frequently used and recommended by manufactures (see for example Qiagen, Roche, New England BioLabs, Fermentas, etc.). All of them contain a buffer reagent that keeps the pH constant in a well-defined range. Well-known examples are phosphate, tris (hydroxymethyl) aminomethane (Tris), or HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) that buffer at the physiological relevant pH range of 6- 9 (Durst and Staples 1972; Chagnon and Corbeil 1973; Tornquist, Paallysaho et al. 1995). In recent years, the development of other buffer systems has been quite successful (Taha 2005) (for a list of buffers and corresponding

pH ranges, see for example: http://delloyd.50megs.com/moreinfo/buffers2.html). Next to the well-defined pH, the stability and homogeneity of proteins depend on many other parameters, for example ionic strength, the presence of ligands and/or co-factors, divalent ions, glycerol, etc. The appropriate buffer composition cannot be predicted so far and needs to be identified by trial-and-error approaches.

3.4 Protein purification – How to overcome problems

In this part we would like to present some pitfalls that might occur during protein purification and provide some ‚rationales' to overcome these problems. As usual, the crucial step of solving a problem is its identification. Here, we are trying to sensitize the reader to indications, which might point towards problems related to instability and/or inhomogeneity of the protein sample. Moreover, such problems cannot always be recognized without the adequate technique(s). Therefore, we are introducing techniques that are capable to visualize the state of proteins.

3.4.1 Visible protein precipitations during IMAC

A very obvious stability problem is the formation of precipitations in the elution fractions of a chromatography step (see Figure 2). In this example, the his-tagged protein was eluted with a linear imidazole gradient from 10 to 500 mM imidazole and eluted at about 250 mM imidazole. Protein precipitation occured immediately after elution (Figure 2A and B) and continued (Figure 2C) resulting in a low amount of soluble protein. This aggregation can be reduced by dilution with a IMAC buffer (typically lacking imidazole) immediatly after the elution. Thereby, dilution hinders the concentration-dependent aggregation. In many cases, this rational is not sufficient to prevent precipitation. After applying, for example, a *buffer screen* (see Figure 1 – (II) purification) the new defined buffer is used for the chromatography or the eluting protein is diluted into the new buffer (see Figure 2D).
Other elution strategies of his-tagged proteins from an IMAC column are available. As described before, competing the poly-histidine from the IMAC column by imidazole is the most common elution strategy, however, for some proteins other strategies are superior, for example, replacing imidazole by histidine. Imidazole is only a mimic for histidine. If one uses histidine instead of imidazole aggregation can be avoided as concentration of the eluent can be reduced by a factor of ten. An example for a protein sensitive to imidazole concentration is shown in Figure 3B. Here a comparative SEC chromatogram is shown. After elution from the IMAC column with imidazole only a very small amount of the protein elutes at the volume corresponding to the size of a monomer or the dimer, respectively (Figure 3B, continuous line). Most of the protein passes the column very fast and elutes at the void volume indicating large radii meaning aggregated protein. Yields of dimeric (at about 150 mL) and monomeric (at about 180 mL) proteins are strongly increased after an elution with histidine (dashed line) compared to an elution with imidazole (continous line) and only the monomeric species could be crystallized (data not shown). The choice of the eluent in IMAC might therefore be an important step in a purification protocol. Another elution strategy of his-tagged proteins is a pH change from 8 to 4. In an acidic environment, histidines become positively charged and are therefore released from the column matrix. This strategy results in a sharp elution from the matrix and the protein is eluted highly concentrated. Although this strategy is recommended by the manufacturers (see GE Healthcare, Qiagen, etc.) the desired protein needs to retain activity at acidic pHs. The

bivalent metal ions (Ni^{2+}, Co^{2+}, Zn^{2+},…, see above) that complex the his-tag can be removed from the matrix by chelating reagents as ethylenediaminetetraacetic acid (EDTA) as another elution strategy (Muller, Arndt et al. 1998)

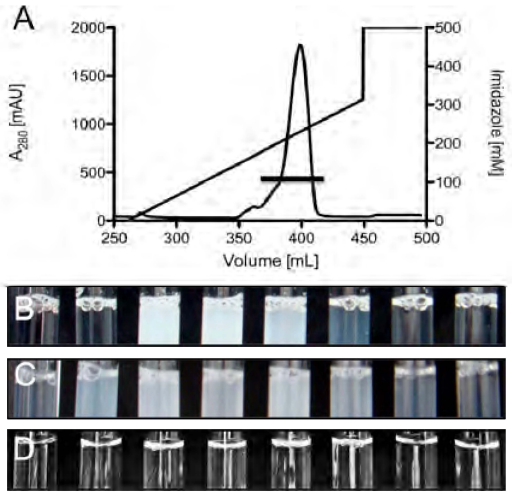


Fig. 2. Elution fractions of an IMAC. The protein was eluted via a linear imidazole gradient from 10 to 500 mM and the absorption at 280 nm was recorded. The elution fractions were collected and photographed. A: IMAC chromatogram of the his-tagged protein. Elution fractions containing the desired protein (indicated by a bar) are collected and shown in B – D. B and C: Elution fractions of the protein in 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 immediately and 10 min after the elution, respectively. D: The elution fractions were immediately mixed in a 1 to 1 ratio with a buffer that enhances the protein stability (50 mM citrate, 50 mM LiCl, pH 6.00) evaluated during a solubility screen.

3.4.2 Invisible aggregations

Sometimes aggregation of proteins in solution can not be detected directly by eye. This inhomogeneity of protein samples can be visualized SEC, a method that separates proteins by their hydrodynamical radius (see above). Protein aggregates are eluting at the void volume, since they are clumbed together resulting in a big hydrodynamical radius (see Figure 3A and B). If invisible aggregation is detected the buffer composition needs to be adjusted. In one case we applied this technique to visualize the state of a protein after an IMAC, and the resulting elution profile is shown in Figure 3A (continous line). Comparable to the imidazole-induced precipitation described above, the protein aggregated and elutes within the void volume of the column (about 40 mL). Moreover, several other protein species elute from 55 to 80 mL indicating a highly inhomogeneous protein sample. The running buffer of the SEC was 50 mM Tris-HCl, pH 8.0 and 150 mM NaCl. Remarkably, a simple change to a new buffer (20 mM HEPES, 150 mM NaCl, pH 7.0) Resulted in a stable and homogenous protein sample (Figure 3A, dotted line), which was suitable for

crystallization trials. Next to the rigorous change in the homogeneity of the protein, the biological activity of the protein could only be determined in the new buffer system. The influence of the buffer composition for the protein activity is a well-known phenomenon (Urh, York et al. 1995; Holm and Hansen 2001; Zaitseva, Jenewein et al. 2005) and in many cases the activity goes hand in hand with an optimal buffer for the purification. Mentionable, the new buffer was not found by trial-and-error approaches. We searched for literature dealing with homologous proteins, especially for established purification protocols. This literature search revealed the new buffer, illustrating that not every step towards a protein structure determination must be a trial-and-error process.

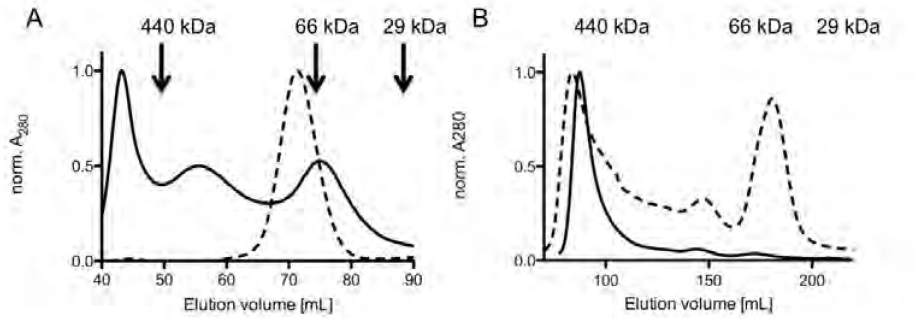


Fig. 3. Size exclusion chromatograms (UV 280nm) of proteins in different buffers. A: The homogeneity of a protein was analyzed in two different buffers; continuous line: 50 mM Tris-HCl, 150 mM NaCl, pH 8.00; dotted line: 20 mM Hepes, 150 mM NaCl, pH 7.00. B: The protein was eluted of the IMAC column either with imidazole (continous line) or with histidine (dotted line), concentrated and applied to the SEC.

Another example for the influence of the buffer composition was published bei Mavaro et al. (Mavaro, Abts et al. 2011). Instead of the buffer agent, the ionic strength of the buffer was the crucial determinant. Purification of the protein in low-salt buffer resulted in an inhomogenous protein sample containing a mixture of aggregates, dimers and monomers without biological activity. However, a simple change to high-salt buffer allowed the purification of homogenous dimeric protein, that was able to bind its substrate.

3.4.3 Overcoming protein instability

In the previous sections different strategies were mentioned to enhance the stability and the homogeneity of purified proteins and in all cases the buffer composition was the solution. Still, the essential question how to determine the optimal buffer to make a protein feel happy *in solution* is not answered? Some rationales and experiences are listed above: different elution strategies for IMAC purifications, the usage of frequently used buffer agents and a literature research for established purification protocols of related proteins. However, in many cases these approaches do not solve the problems occuring during the purification. But, is there a general methodology to overcome the problems? Unfortunately, the answer is as frustrating as challenging - there is not a general panacea around for the right buffer composition of a protein. If a new buffer needs to be found, trial-and-error

approaches have to be applied. A lot of different parameters are influencing the state of a protein, i. e. the buffer agent, the salt concentration, presence of metal ions with different valences, the hydrophobicity, and even the temperature of the buffer. The analysis of the protein in different buffers can be done by SEC and/or light scattering experiments. However, screening of all the different variables is very labor- and cost-intensive, and time-consuming, moreover only combinations of two or more additives might be sufficient to enhance the solubility and homogeneity of the protein. Therefore high-throughput methods are needed that handle a lot of different conditions simultaneously using as few as possible protein sample.

3.4.4 Buffer screen – Enhancing the solubility

Many publications are available suggesting methods for a solubility screening to allow the crystallization of initially inhomogeneous, aggregating protein samples (Jancarik, Pufan et al. 2004; Zaitseva, Holland et al. 2004; Collins, Stevens et al. 2005; Sala and de Marco 2010; Schwarz, Tschapek et al. 2011). In all of these methods aggregating protein samples are mixed with commercially available crystallization screens incubated for a period of time, and analyzed for precipitation visually using a light microscope. Screening conditions resulting in no precipitations are analyzed upon their composition, and protein samples are further examined with respect to their solubility and homogeneity under these conditions by SEC or light scattering experiments. This technique allows high-throughput screening in a 96-well format, where an automated pipetting system mixes only 50-200 nL of protein solution with 50-200 nL of buffer solutions to minimize the needed protein sample and increase the screening efficiency. Several buffer screens are commercially available that cover many different buffer agents, salt concentrations and other buffer parameters (i. e. from Hampton Research, Molecular Dimensions, Sigma, Jena Bioscience, Qiagen). After a solubility screening was applied, we were able to stabilize a previously unstable protein at concentrations above 3 mg/mL (see above "Protein precipitations during IMAC" and Figure 2D) at concentrations of up to 100 mg/mL for weeks (Schwarz, Tschapek et al. 2011). Typically, the new buffer (50 mM citrate, 50 mM $LiCl_2$, pH 6,00) should be used during the entire purification procedure starting with cell lysis. In the described case, the new buffer contains citrate, which is incompatible with an IMAC purification. Therefore the protein was immediately mixed with the new buffer after the elution of the IMAC column.

3.4.5 Size-exclusion chromatography versus light scattering experiments

Size-exclusion chromatography (SEC) and light scattering experiments (LS) are very helpful tools to analyze the homogeneity (Collins, Stevens et al. 2005) and the molecular mass of proteins; however both of them have advantages and disadvantages compared to each other. In SEC experiments proteins are separated based on their hydrodynamic radius by partitioning between a mobile phase and a stationary liquid within the pores of a matrix. All SEC columns are characterized by the volumes V_0, the liquid volume in the interstitial space between particles, V_i, the volume contained in the matrix pores and V_T, the total diffusion volume ($V_0 + V_i$) (Regnier 1983a; Regnier 1983b). In dependency of the hydrodynamic radius molecules are eluting at specific retention volumes in between V_0 and V_T with big molecules eluting first. After a calibration of a SEC column with proteins of known

molecular weight (i. e. Sigma-Aldrich, "Kit for Molecular Weights") the molecular mass of the protein of interest can be roughly estimated; the elution volume is correlated to the $\log_{10}$ of the molecular weight (therefore, the hydrodynamic radius is considered to be proportional to the molecular weight). However, many extraneous mechanisms such as adsorptive, hydrophobic and ionic effects are further limiting the correlation between the retention volume and the molecular mass giving sometimes rise to wrong estimations.
Light scattering (LS) experiments can be applied to overcome these disadvantages and investigate the exact molecular weight of the protein sample. The rayleigh scattering of particles of monochromatic light depends directly on the molar mass of the particle. If you know the exact number of particles you can calculate the average molar mass of these particles. This technique is very powerful when used online after separation of the protein depending on their hydrodynamic radius, meaning SEC. This technique is always superior to normal SEC but requires special equipment and especially more time. However, if the protein fold is not really globular or other effects occur (see above: ionic, hydrophobic, etc.) assumption on size and oligomeric state based on SEC is not possible at all. For protein crystallization information about monodispersity, which can be provided by such an experiment, is an additional benefit.

3.4.6 Analysis of the homogeneity – High-throughput methods

Despite the development of various sophisticated methods, a bottleneck of homogeneity screening is high-throughput analysis. As mentioned above, proteins need to be analyzed by SEC and/or LS experiments after visual read-out of the protein-buffer droplets. Therefore, fluorescence-based solubility screens were developed that allow the high-throughput analyzes of many samples in a 96-well format (Ericsson, Hallberg et al. 2006; Alexandrov, Mileni et al. 2008; Kean, Cleverley et al. 2008). All these assays use fluorophores as reporters of the protein state. A suitable fluorophore is, for example, Sypro Orange, which exhibits different fluorescence properties as a function of its environment. This dye is almost dark in hydrophilic environment, however, after binding to hydrophobic molecules, it emits light at 570 nm. In inhomogenous and unfolded protein samples hydrophobic amino acids are exposed to the surface of proteins (Murphy, Privalov et al. 1990). An increase in the fluorescence signal of Sypro Orange correlates therefore with unfolding events of proteins. The homogeneity screening can be performed in basically two ways: temperature- or time-dependent. For the first setup the protein sample is heated gradually in distinct steps (i. e. 1 °C) and the emission is monitored at 570 nm. Hereby, a "melting" temperature is determined, which is characterized by 50% fluorescence of the maximal fluorescence at the highest temperature; the higher the melting temperature, the higher the stability of the protein (Ericsson, Hallberg et al. 2006). Secondly, the protein sample is incubated at a specific temperature (i. e. 40°C) and the fluorescence is measured for a period of time. The "half-life" time, at which 50 % fluorescence of the maximum fluorescence in one sample is detected, can be compared to all buffer conditions. In Figure 4 an example of the time-dependent approach is shown. Here, the protein is incubated in different buffers with various salt concentrations. The emission of Sypro Orange is recorded each minute at 570 nm. An analysis of all time-dependent fluorescence plots indicates that the protein is most stable in buffers containing 125 mM NaCl but unfolds fast in 1 M ammonium sulfate. These assays result in qualitative indications about a favourable environment of proteins that enhance the stability. Ericsson et al. proved the concept of this method by applying it to

different proteins (Ericsson, Hallberg et al. 2006). The stability optimization yielded a twofold increase in initial crystallization leads. Moreover these assays enable the search for putative ligands of the protein. Upon binding of a substrate in the binding pocket or an inhibitor, the stability of the protein increases, which can be detected experimentally.

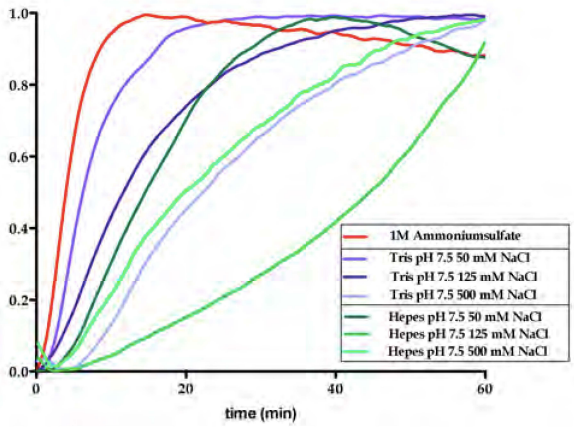


Fig. 4. Time-dependent stability optimization screen using Sypro Orange as reporter. The protein is diluted 1:50 into each test buffer containing Sypro Orange, excited with 490 nm and the fluorescence at 570 nm is measured for 60 minutes automatically with a PLATE READER (Fluorostar, BMG Labtech). Normalized fluorescence is plotted against the time.

4. Protein crystallization: Introduction

Protein crystals suitable for X-ray diffraction experiments and usable for subsequent structure determination are normally relatively large with a size of at least 10 to 100 µm. In contrast to crystals of mineral compounds, protein crystals are rather soft and sensitive to mechanical stress and temperature fluctuations. These properties are due to weak interactions between single proteins within the crystal, their high flexibility as well as the size of the macromolecules. The periodic network of building blocks is held together by dipole-dipole interactions, hydrogen bonds, salt bridges, van der Waals contacts or hydrophobic interactions. All of them have binding energies in the low kcal/mol range.
Especially the limited number of crystal contacts and their directionality are the largest difference to the high interactions generally observed in salt crystals. An example of the interactions within a protein crystal is shown in Figure 5. This picture highlights the main pitfalls in protein crystallization. A protein is a highly irregular shaped and flexible macromolecule which allows weak and stinted interactions at very specific locations of its surface. All vacuity is filled with buffer, in general not contributing to any kind of interactions between the protein molecules. Figure 5A shows a protein of around 30 kDa, which crystallizes in a rather small unit cell (shown in black). Only one protein monomer is located in the asymmetric unit of the unit cell, the other shown monomers represent symmetry related proteins. Figure 5B highlights the three-dimensional packing of protein molecules within a crystal.

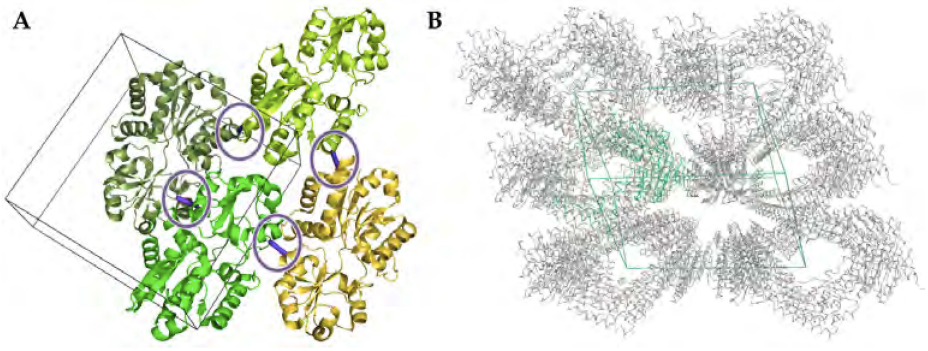


Fig. 5. Example of the packing within a crystal. A: The unit cell is shown in black, crystal contacts are highlighted with purple circles and lines. B: Three-dimensional crystal packing of a different protein. The unit cell as well as one protein monomer are depicted in green.

The flexibility as well as the other mentioned characteristics of proteins are responsible for the problems occuring during crystallization trials and despite extensive efforts not every protein is suitable for crystallization. If one cannot generate crystals one has to move back several steps and change the properties of the protein, e. g. surface properties by mutation of single amino acids, truncation of the protein or sometimes only changing buffer compositions that result in a more suitable protein for crystallization (see Figure 1 and also below). There are several prediction servers available that help choosing the 'right' protein and modification (Linding, Jensen et al. 2003; Goldschmidt, Cooper et al. 2007). However, protein crystallization still remains an empirical approach, sometimes called voodoo, while crystallography is science.

4.1 Phase diagram

The conditions or protocols for obtaining good crystals are still poorly understood and despite all progress and efforts protein crystallization is a trail-and-error approach. However, a step towards a better understanding of crystal growth can be achieved by analyzing the phase diagram of a protein-water mixture. The phase diagram is a simple illustration to help understanding how protein crystals are formed. Mostly, it is shown as a function of two ambient conditions that can be manipulated, i. e. the temperature and the concentration. Three-dimensional diagrams (two dependent parameters) have also been reported (Sauter, Lorber et al. 1999) and even a few more complex ones have been determined as well (Ewing, Forsythe et al. 1994). Figure 6 shows a schematic phase diagram for a protein solution as a function of protein concentration and precipitant concentration. The phase diagram is broken down into four distinct zones (Rosenbaum and Zukoski 1996; Haas and Drenth 1999; Asherie 2004):

1. Undersaturated zone: Under these condition the protein will stay in solution as neither the concentration of the protein nor of the precipitant is high enough to reach supersaturation.
2. Precipitation zone: Is the protein concentration or the precipitant concentration too high, the protein precipitates out of solution; this kind of solid material is not useful for crystallographic studies.

3. Labil zone: This is the most important configuration of the two parameters, as nucleation and initial crystal growth take place under these conditions.
4. Metastable zone: After initial crystals are formed and start growing in the labil zone, protein concentration decreases in the drop and the metastable zone will be reached. Here the crystal can grow further to its final maximum size.

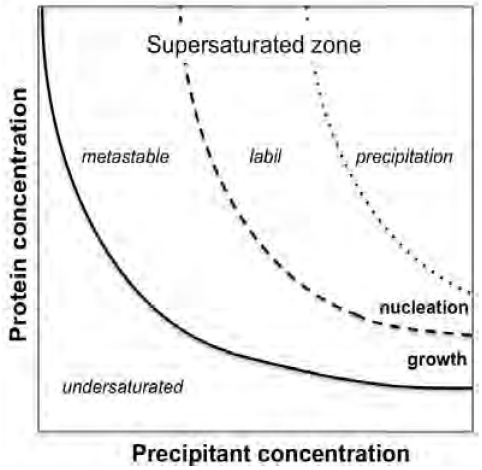


Fig. 6. A basic solubility phase diagram for a given temperature (adapted from (Rupp 2007).

The curve separating the undersaturated zone from the supersaturated one is called *solubility curve*. If conditions are chosen below the solubility curve, the protein will stay in solution and never crystallize. This means when a protein crystal is placed in a solvent, which is free of protein, it will start to dissolve. If the volume of the droplet is small enough it will not dissolve completely: it will stop dissolving when the concentration of the protein in the droplet reaches a certain level. At this concentration the crystal loses protein molecules at the same rate at which protein associate to the crystal – the system is at equilibrium. Determination of the solubility of the protein of interest might be a helpful information at the beginning of crystallization experiments. This can be done in a two-dimensional screen varying for example ammonium sulfate concentrations as well as the protein concentration.

4.2 Crystallization techniques

Crystallization is a phase transition phenomenon. Protein crystals grow from a supersaturated aqueous protein solution. Varying the concentration of precipitant, protein and additives, pH, temperature and other parameters induce the supersaturation. However, as mentioned before, prediction of this kind of phase diagrams is *a priori* impossible.
Protein crystallization can be divided into two main steps:

1. Generating initial crystals: 'Searching the needle in a haystack'
2. Empirical optimization of these crystallization condition

The first step is mostly based on experiences from other crystallization trials with different proteins. Nowadays several supplier offer crystallization screens that contain solutions for

initial experiments that were used successfully in the past for crystallization trials (Jancarik and Kim 1991), so-called "sparse matrix screens". There are also some trials around to use more systematic approaches (Brzozowski and Walton 2000) to get more information about solubility prior and simultaneous to crystallization (incomplete factorials, solubility assays). Both kinds of screens can be applied to different crystallization techniques.

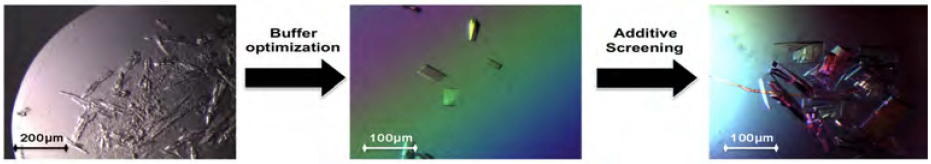


Fig. 7. Crystal optimization. First steps in crystal optimization are shown. Initial protein crystals look weak and fragile, after screening around this initial buffer composition crystal evaluation by eye results in less fragile, homogeneous looking crystals. However, diffraction quality was poor. Therefore an additive screening was performed that resulted in a different crystal form. These crystals finally were able to diffract X-rays to a reasonable resolution.

A lead/hit in that initial step might not be a 'real' crystal rather than a crystalline precipitate or just phase separation. In the next step, fine-tuning the buffer composition further optimizes this hit. Varying pH, salt concentration, type and concentration of precipitant and protein concentration are expected to yield larger and hopefully also better-diffracting crystals. In this step, the chemicals used are much more defined and therefore it is a more systematic than empirical screening (see Figure 7).

4.2.1 Vapor diffusion

The most popular and simplest technique to obtain protein crystals is the vapor diffusion method either in the sitting or hanging drop variant (see Figure 8). For both a defined volume (mostly < 1µl) of protein solution is mixed with an equivalent volume of screening solution and then equilibrated against the original precipitant/screening concentration. During this equilibration, the vapor pressure of the solution rises as the protein crystallizes (protein in solution lowers water activity) while the water evaporates to maintain equilibrium, which causes the precipitant concentration to rise. Therefore, if the crystal growth is sensitive to the precipitant concentration, vapor diffusion can rapidly force the mixture to unstable conditions where growth and nucleation are too rapid. This is the main disadvantage of vapor diffusion: Growing large crystals might be problematic!

4.2.2 Micro batch method

In this set-up the protein solution is mixed with screening solution at concentrations required for supersaturation right at the beginning of the experiment. Typical drop sizes of micro batch experiments ranges from 1-2 µl. The drop is then covered with oil, which acts as an inert sealing to protect the drops during incubation from evaporation (see Figure 8).

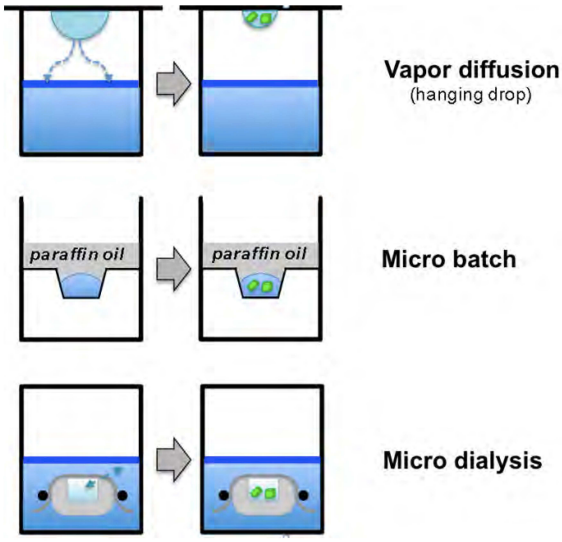


Fig. 8. Protein crystallization techniques. Schematic representation of a) vapor diffusion, b) micro batch and c) micro dialysis crystallization techniques widely used for crystal growth (adapted from (Drenth 2006)).

4.2.3 Micro dialysis

Dialysis is another way to change the buffer composition and increase its concentration in the crystallization experiment gradually (see Figure 8). Micro-dialysis buttons are exposed to different screening buffers. This method requires rather high amounts of protein but might yield large crystals.

After obtaining initial crystal hits in a commercial screen the tough part of crystal optimization starts. By varying pH, salt concentration, temperature, precipitant concentration or protein concentration these initial crystals should be reproduced and become larger, more regular shaped or are simply growing faster. A further improvement of crystal quality might be achieved by the addition of small amounts of so called 'additives'. At this point basically each chemical compound might be sufficient to improve the crystal quality. Luckily, there are some preferable working additives, which have been proven to produce better crystal in more than one case. Especially compounds that are known to reduce undirected interactions in proteins like organic solvents, i. e. DMSO or phenol, or detergents and reducing agents are very often used at this stage and helpful to force more homogeneous well diffracting crystals.

4.3 Crystal nucleation

There are two fundamental steps during protein crystallization: Nucleation and crystal growth. If one cannot obtain single crystals of adequate quality for analysis, this is generally a consequence of problems associated with the growth phase (see above). But failure to obtain any crystals at all or failure to obtain single, supportable nuclei reflects difficulties in the nucleation step. Therefore control of nucleation is a powerful tool to optimize protein crystals

or sometimes it is the only way to get crystals at all. Nucleation can take place either homogeneous meaning in the bulk of the solution, when the supersaturation is high enough for the free-energy barrier to nucleus formation to be overcome or heterogeneous mostly by solid material in the crystallization solution. This can also occur even when the supersaturation is not achieved. Therefore in order to control nucleation one has to work with highly clean solutions to avoid nucleation by the second mentioned possibility. The nucleation zone can be bypassed by insertion of crystals, crystal seeds or other nucleants to the protein/precipitant mixture. Addition of crystals or tiny fragments of crystals is called seeding. This method is then subdivided into macro- and micro-seeding dependent on the size of the nucleant added. In macro-seeding experiments one single, already well-formed but small crystal is placed into a new crystallization solution at lower saturation. Microseeding in contrast requires small fragments of a crystal or almost invisible microcrystalline precipitate. These 'seeds' are then transferred into a fresh crystallization solution either by a seeding wand which is dipped into the microseed mixture to pick up seeds and then touched across the surface of the new drop or by a animal whisker or hair that is stroked over the surface of the parent crystal to trap the nuclei and then is drawn through the new drop. As this method also enhances the speed of crystal growth it can be used with sensitive substrate that undergo decomposition over time. Oswald et al. proved this in 2008 by solving the structure of ChoX from *Sinorhizobium meliloti* in complex with a highly hydrolyzing substrate, acetylcholine (Oswald, Smits et al. 2008). In classical vapor diffusion experiments crystals appear after four weeks but data showed only little electron density in the ligand-binding site and turned out to result from a choline bound instead of acetylcholine. Hydroxylation was favored due to the relatively long time for crystal growth but also because of an acetic pH in the crystallization set-up. To circumvent these problems accelerated crystal growth was required. In this case micro-seeding results in crystals suitable for data collection in less than 24 hours.

Recent years more effort in nucleation control yielded in fancy materials that can be used as nuclei for crystals. These methods use the second way of nuclei formation, as a solid material is introduce into the crystallization solution as an 'universal' nucleant (Chayen, Saridakis et al. 2006). There have been several substances that have been tried more or less successful. Some have been useful for individual proteins, but mostly they were not applicable in general (McPherson and Shlichta 1988; Chayen, Radcliffe et al. 1993; Blow, Chayen et al. 1994). In 2001, Chayen et al. proposed the idea of using porous silicon whose pore size is comparable with the size of a protein molecule. In theory such pores may confine and concentrate the protein molecules at the surface of the silicone and thereby encourage them to form crystal nuclei (Chayen, Saridakis et al. 2001). These nucleants have made it to commercial availability (www.moleculardimensions.com) and have proven to be suitable for different kinds of proteins and even membrane proteins that have not been possible to crystallize before formed nice crystals in the presence of these nucleants.

4.4 Cryoprotection

Exposure of a protein crystal at room temperature results in dramatic radiation damage due to radicals formed by the ionizing X-ray photons. To reduce that harmful disintegration of the protein crystal the crystal is cooled to ≈100K with the help of liquid nitrogen (Low, Chen et al. 1966; Hope 1988; Rodgers 1994; Garman 1999). However, it is common for the cooling process to disrupt the crystal order and decrease diffraction quality. Thus, the crystal must be cooled fast so that the water in the solvent channels is in the vitreous rather than in the

crystalline state at the end of this procedure. As for pure water this cooling has to take place very quick (10^{-5}s, (Johari, Hallbrucker et al. 1987), some water molecules can be replaced by a cryoprotective solution prior to cooling (Juers and Matthews 2004). This exceeds the time window for freezing up to 1-2s (Garman and Owen 2006) however, finding a good 'cryoprotectant' for a special protein crystal again involves substantial screening. Once flash frozen in liquid nitrogen, the crystal must be kept below the glass transition temperature of the cryobuffer at or below 155K at all times (Weik, Kryger et al. 2001).

4.5 What can you do when all efforts did not succeed in crystals?

4.5.1 Buffer composition – Again!

The choice of the right buffer used for crystallization experiments is very crucial. As shown above, every protein needs its own buffer composition to feel kind of happy in this aqueous artificial environment. Especially as high protein concentrations (>10mg/ml) are required for crystallization, one might has to test several buffer compositions again (see also Figure 1). As a rule of thumb you should obtain around 50% of clear drops immediately after mixing protein and buffer solution. If you detect drastically more precipitation in your drops you should think first about lower protein concentration but of course secondly about changing your buffer system again.

4.5.2 How to obtain a rigid protein suitable for crystallization?

To overcome the problem of flexibility of some regions in the protein addition of ligands is often a very powerful tool to fix the protein in a single conformation that is more favorable for crystallization. A good example for this strategy is the crystallization of so-called substrate binding proteins (for a recent review see (Berntsson, Smits et al. 2010)). These proteins catch their substrate in the periplasm of bacteria or on the outer membrane of archaea and then deliver it to their cognate transport system located in the membrane. The mechanism of substrate binding is quite well understood. These binding proteins all consist of two domains, which rotate towards each other during the binding event. In solution without substrate they are quite flexible and NMR-studies proved a equilibrium between open and closed conformation (Tang, Schwieters et al. 2007). Analysis of all available structures for this class of proteins showed that more than 95% were crystallized with a ligand bound (Berntsson, Smits et al. 2010). Thus, a stabilization of the two domains seems to simplify crystal contact formation dramatically. Although people always want to obtain a functional conformation of the protein in their crystal structure, it is sometimes helpful to think about how to stop the protein from doing its job. A non-functional protein is in general less flexible and fixed in one conformation. One example for successful implementation of this strategy is the crystal structure of NhaA from *Escherichia coli* solved in 2005 (Hunte, Screpanti et al. 2005). Here, Hunte et al., downregulated the protein activity by working at an acidic pH of 4. Although the protein shows almost no activity at this pH the structure reveals the basis for mechanism of Na+/H+ exchange and also its regulation by pH could be understood.

4.5.3 Rational protein design for crystallization: Surface engineering

The first example of rational protein design that yielded a good diffracting protein crystal is given by Lawson et al. in 1991 (Lawson, Artymiuk et al. 1991). They compared amino acids

involved in crystal contact formation of the rat ferritin protein L. (which is highly homologous to human ferritin H, the target protein) with the amino acids present at that position in human ferritin H. A replacement of Lys86, found in the human sequence, with Glu, which occurs in rat, recreated a Ca^{2+} binding bridge that mediates crystal contacts in the rat ortholog. As this method was successful for several other proteins (McElroy, Sissom et al. 1992; Braig, Otwinowski et al. 1994; Horwich 2000), a general protocol was required. The concept Derewenda et al. proposed in 2004 is based on the general equation for the free energy that drives protein crystallization:

$$\Delta G = \Delta H - T(\Delta S_{protein} + \Delta S_{solvent})$$

As the enthalpy values of intermolecular interactions in a crystal lattice are rather small (see above), crystallization is very sensitive to entropy changes of both protein and solvent. The formation of ordered protein aggregates carries a negative entropy term. This can only be overcome by positive entropy from the release of water bound to the protein. However, large hydrophilic residues (e.g. lysines, arginies, glutamates, glutamines) exposed on the protein surface need to be ordered. Since they are rather flexible this can cause problems. This can be overcome by mutating large amino acids into smaller ones, for example alanines. Among these large amino acids lysines and glutamates play a particular role, as they are always (with only very few exeptions) located on the protein surface (Baud and Karlin 1999). Both lysines and glutamates are typically disfavored at interfaces in protein protein complexes (Lo Conte, Chothia et al. 1999), therefore it is rather straight forward to assume that lysine and glutamate to alanine mutants are good targets for protein crystallization if wildtype protein hardly forms crystals. However this also means that you have to go several steps backwards on road to a protein structure determination (see Figure 1).

4.5.4 Affinity tag removal: Philosophic question???

Another variant in protein crystallization nowadays is the affinity tag used for purification of the desired protein. The decision about position and choice of the affinity tag are mostly made at the beginning of the long way to a crystal structure (see Figure 1). However, it becomes crucial again at the crystallization step. In general most people like to remove the tag before crystallization to prove a physiological conformation. But, there are examples where the tag played a pivotal role in crystallization (Smits, Mueller et al. 2008a). The crystal structure of the octopine dehydrogenase from *Pecten maximus* is shown in Figure 9 (Smits, Mueller et al. 2008b), with the interactions sides/crystal contacts highlighted in green.
In Figure 9A contacts look quite similar to that presented in Figure 5. However when having a closer look on the his-tag, you recognize that it is located in a cavity formed by another monomer of that protein. In that cavity it can perform several hydrogen bonds with amino acids from the other monomer resulting in a very strong interaction which yields good quality crystals.

4.5.5 Crystallization using antibody fragments

A number of ways to stabilize proteins for crystallography have been developed, for example genetic engineering, co-crystallization with natural ligands and reducing surface

entropy (see above). Recently, crystallization mediated by antibody fragments has moved into the focus of crystallographers especially to obtain crystals of membrane proteins (Ostermeier, Iwata et al. 1995; Hunte and Michel 2002). Membrane protein crystallization is even tougher compared to soluble proteins, because of the amphipathic surface of the molecules. As they are located in the lipid bilayer most of their surface is hydrophobic and must be covered to keep them in solution. This is maintained by detergents. The detergent micelles cover the hydrophobic surface and therefore this area is no longer available to form crystal contacts. Crystal contacts can only be formed by the polar surfaces of these proteins. As many membrane proteins contain only relatively small hydrophilic domains, a strategy to increase the probability of getting well-ordered crystals is required. Antibody fragments can play this role. They can be designed for binding at specific regions in the protein and then function as additional polar domain in the membrane protein complex (for example see (Ostermeier, Iwata et al. 1995; Huber, Steiner et al. 2007).

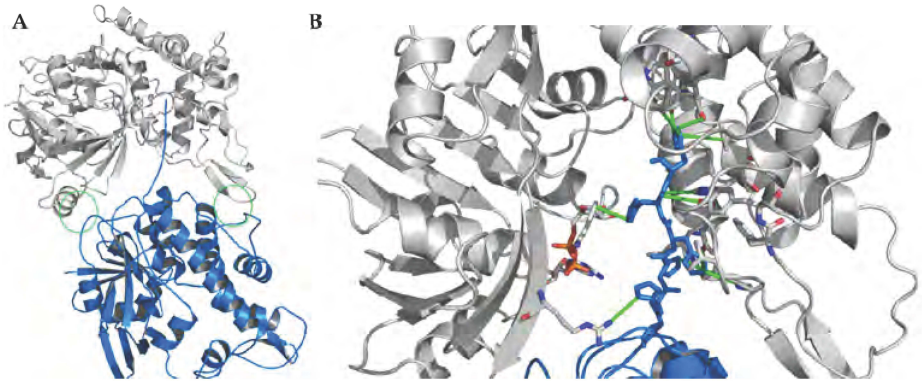


Fig. 9. Crystal contacts in OcDH protein. A: Overall view on two monomers. Surface Crystal contacts are highlighted in the green circles. B: Zoom in on the His-tag of one monomer. The his-tag of one monomer in the crystal structure is located near the binding site in a deep cavity formed by the other monomer. Therefore it is able to form several hydrogen bonds (highlighted in green) with side and main chains of the other protein but also with the ligand bound in this binding site (orange).

5. Conclusion

For what reason do we effort so much work on good quality crystals?

Single good quality crystals constitute an essential prerequisite for structural investigations of biological macromolecules using X-ray diffraction. The harder one works on crystal quality the easier the determination of a reasonable atomic model of the molecule of interest becomes. The vast majority of problems encountered in crystal structure determination can typically be traced back to data-quality issues caused by crystal imperfections. Consequently, although primary focus of structural biology is on the macromolecule that makes up a crystal, there is also considerable interest in the physical properties, nucleation

and growth of the crystals themselves. Statistics of various Structural Genomics Centers proved that protein crystallization is still despite all the progress in the technology of crystallization robotics is still a rather tough field in biological science. Success rate ranges for small prokaryotic proteins from 10-30% and decreases dramatically to a few percent for human proteins. The struggle obtaining crystals for protein structure determination is justified. After all efforst looking at electron density and subsequent the protein structure is still one of the most intriguing as well auspicious parts in structural biology

6. References

Alexandrov, A. I., M. Mileni, et al. (2008). "Microscale fluorescent thermal stability assay for membrane proteins." Structure 16(3): 351-359.

Altmann, F., E. Staudacher, et al. (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." Glycoconj J 16(2): 109-123.

Arnau, J., C. Lauritzen, et al. (2006). "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins." Protein expression and purification 48(1): 1-13.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods 34(3): 266-272.

Baneyx, F. (1999). "Recombinant protein expression in *Escherichia coli*." Current opinion in biotechnology 10(5): 411-421.

Baud, F. and S. Karlin (1999). "Measures of residue density in protein structures." Proc Natl Acad Sci U S A 96(22): 12494-12499.

Berntsson, R. P., M. K. Doeven, et al. (2009). "The structural basis for peptide selection by the transport receptor OppA." EMBO J 28(9): 1332-1340.

Berntsson, R. P., S. H. Smits, et al. (2010). "A structural classification of substrate-binding proteins." FEBS Lett 584(12): 2606-2617.

Binda, C., M. Aldeco, et al. (2010). "Interactions of Monoamine Oxidases with the Antiepileptic Drug Zonisamide: Specificity of Inhibition and Structure of the Human Monoamine Oxidase B Complex." J Med Chem.

Blow, D. M., N. E. Chayen, et al. (1994). "Control of nucleation of protein crystals." Protein Sci 3(10): 1638-1643.

Blumenthal, D. K., K. Takio, et al. (1985). "Identification of the calmodulin-binding domain of skeletal muscle myosin light chain kinase." Proc Natl Acad Sci U S A 82(10): 3187-3191.

Bolotin, A., P. Wincker, et al. (2001). "The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. lactis IL1403." Genome research 11(5): 731-753.

Braig, K., Z. Otwinowski, et al. (1994). "The crystal structure of the bacterial chaperonin GroEL at 2.8 A." Nature 371(6498): 578-586.

Brzozowski, A. M. and J. Walton (2000). "Clear strategy screens for macromolecular crystallization." J Appl Cryst 34: 97-101.

Cayley, S., B. A. Lewis, et al. (1991). "Characterization of the cytoplasm of *Escherichia coli* K-12 as a function of external osmolarity. Implications for protein-DNA interactions in vivo." J Mol Biol 222(2): 281-300.

Chagnon, A. and M. Corbeil (1973). "Use of an organic buffer (HEPES) in human lymphocytoid cell line cultures." In Vitro 8(4): 283-287.

Chayen, N. E., J. W. Radcliffe, et al. (1993). "Control of nucleation in the crystallization of lysozyme." Protein Sci 2(1): 113-118.

Chayen, N. E., E. Saridakis, et al. (2001). "Porous silicon: an effective nucleation-inducing material for protein crystallization." J Mol Biol 312(4): 591-595.

Chayen, N. E., E. Saridakis, et al. (2006). "Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium." Proc Natl Acad Sci U S A 103(3): 597-601.

Collins, B., R. C. Stevens, et al. (2005). "Crystallization Optimum Solubility Screening: using crystallization results to identify the optimal buffer for protein crystal formation." Acta Crystallogr Sect F Struct Biol Cryst Commun 61(Pt 12): 1035-1038.

Cregg, J. M. (2007). Pichia Protocols, Humana Press.

Cregg, J. M., J. L. Cereghino, et al. (2000). "Recombinant protein expression in Pichia pastoris." Molecular biotechnology 16(1): 23-52.

D., O. R., M. L.K., et al. (1992). Baculovirus Expression Vectors A Laboratory Manual. New York.

de Ruyter, P. G., O. P. Kuipers, et al. (1996). "Controlled gene expression systems for Lactococcus lactis with the food-grade inducer nisin." Applied and environmental microbiology 62(10): 3662-3667.

de Vos, W. M. and E. E. Vaughan (1994). "Genetics of lactose utilization in lactic acid bacteria." FEMS microbiology reviews 15(2-3): 217-237.

di Guan, C., P. Li, et al. (1988). "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein." Gene 67(1): 21-30.

Drenth, J. (2006). "Principles of Protein X-Ray Crystallography." Springer 3.

Durst, R. A. and B. R. Staples (1972). "Tris-tris-HCl: a standard buffer for use in the physiologic pH range." Clin Chem 18(3): 206-208.

Eifler, N., M. Duckely, et al. (2007). "Functional expression of mammalian receptors and membrane channels in different cells." Journal of structural biology 159(2): 179-193.

Elowitz, M. B., M. G. Surette, et al. (1999). "Protein mobility in the cytoplasm of *Escherichia coli*." J Bacteriol 181(1): 197-203.

Ericsson, U. B., B. M. Hallberg, et al. (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem 357(2): 289-298.

Erkens, G. B., R. P. Berntsson, et al. (2011). "The structural basis of modularity in ECF-type ABC transporters." Nat Struct Mol Biol 18(7): 755-760.

Ewing, F., E. Forsythe, et al. (1994). "Orthorhombic lysozyme solubility." Acta Crystallogr D Biol Crystallogr 50(Pt 4): 424-428.

Fischer, M., A. P. Leech, et al. (2011). "Comparative Assessment of Different Histidine-Tags for Immobilization of Protein onto Surface Plasmon Resonance Sensorchips." Analytical chemistry.

Gaberc-Porekar, V. and V. Menart (2001). "Perspectives of immobilized-metal affinity chromatography." Journal of biochemical and biophysical methods 49(1-3): 335-360.

Garman, E. (1999). "Cool data: quantity AND quality." Acta Crystallogr D Biol Crystallogr 55(Pt 10): 1641-1653.

Garman, E. F. and R. L. Owen (2006). "Cryocooling and radiation damage in macromolecular crystallography." Acta Crystallogr D Biol Crystallogr 62(Pt 1): 32-47.

Gasson, M. J. (1983). "Plasmid complements of *Streptococcus lactis* NCDO 712 and other lactic streptococci after protoplast-induced curing." Journal of bacteriology 154(1): 1-9.

Gingras, A. C., R. Aebersold, et al. (2005). "Advances in protein complex analysis using mass spectrometry." The Journal of physiology 563(Pt 1): 11-21.

Goldschmidt, L., D. R. Cooper, et al. (2007). "Toward rational protein crystallization: A Web server for the design of crystallizable protein variants." Protein Sci 16(8): 1569-1576.

Groß, R., R. Pisa, et al. (2002). "Isolierung der trimeren Hydrgenase aus Wollinella succinogenes durch StrepTactin-Affinitätschromatographie." Biospektrum 1: 101.

Guillot, A., C. Gitton, et al. (2003). "Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium." Proteomics 3(3): 337-354.

Haas, C. and J. Drenth (1999). "Understanding protein crystallization on the basis of the phase diagram." Journal of Crystal Growth 196(2-4): 388-394.

Hampton, E. N., M. W. Knuth, et al. (2007). "The self-inhibited structure of full-length PCSK9 at 1.9 A reveals structural homology with resistin within the C-terminal domain." Proc Natl Acad Sci U S A 104(37): 14604-14609.

Hansen, S. B., G. Sulzenbacher, et al. (2006). "Structural characterization of agonist and antagonist-bound acetylcholine-binding protein from *Aplysia californica*." J Mol Neurosci 30(1-2): 101-102.

Hefti, M. H., C. J. Van Vugt-Van der Toorn, et al. (2001). "A novel purification method for histidine-tagged proteins containing a thrombin cleavage site." Analytical biochemistry 295(2): 180-185.

Hochuli, E., H. Dobeli, et al. (1987). "New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues." Journal of chromatography 411: 177-184.

Holm, J. and S. I. Hansen (2001). "Effect of hydrogen ion concentration and buffer composition on ligand binding characteristics and polymerization of cow's milk folate binding protein." Biosci Rep 21(6): 745-753.

Hope, H. (1988). "Cryocrystallography of biological macromolecules: a generally applicable method." Acta Crystallogr B 44 (Pt 1): 22-26.

Horwich, A. (2000). "Working with Paul Sigler." Nat Struct Biol 7(4): 269-270.

Huber, T., D. Steiner, et al. (2007). "In vitro selection and characterization of DARPins and Fab fragments for the co-crystallization of membrane proteins: The Na(+)-citrate symporter CitS as an example." J Struct Biol 159(2): 206-221.

Hunt, I. (2005). "From gene to protein: a review of new and enabling technologies for multi-parallel protein expression." Protein expression and purification 40(1): 1-22.

Hunte, C. and H. Michel (2002). "Crystallisation of membrane proteins mediated by antibody fragments." Curr Opin Struct Biol 12(4): 503-508.

Hunte, C., E. Screpanti, et al. (2005). "Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH." Nature 435(7046): 1197-1202.

Jana, S. and J. K. Deb (2005). "Strategies for efficient production of heterologous proteins in Escherichia coli." Applied microbiology and biotechnology 67(3): 289-298.

Jancarik, J. and S. H. Kim (1991). "Sparse matrix sampling: a screening method for crystallization of proteins." J Appl Cryst 24: 409-411.

Jancarik, J., R. Pufan, et al. (2004). "Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins." Acta Crystallogr D Biol Crystallogr 60(Pt 9): 1670-1673.

Johari, G. P., A. Hallbrucker, et al. (1987). "The glass-liquid transition of hyperquenched water." Nature 330: 552-553.

Jonasson, P., S. Liljeqvist, et al. (2002). "Genetic design for facilitated production and recovery of recombinant proteins in Escherichia coli." Biotechnology and applied biochemistry 35(Pt 2): 91-105.

Juers, D. H. and B. W. Matthews (2004). "Cryo-cooling in macromolecular crystallography: advantages, disadvantages and optimization." Q Rev Biophys 37(2): 105-119.

Kean, J., R. M. Cleverley, et al. (2008). "Characterization of a CorA Mg2+ transport channel from Methanococcus jannaschii using a Thermofluor-based stability assay." Mol Membr Biol 25(8): 653-663.

Konings, W. N., J. Kok, et al. (2000). "Lactic acid bacteria: the bugs of the new millennium." Curr Opin Microbiol 3(3): 276-282.

Kopecky, D. J., X. Hao, et al. (2008). "Identification and optimization of N3,N6-diaryl-1H-pyrazolo[3,4-d]pyrimidine-3,6-diamines as a novel class of ACK1 inhibitors." Bioorg Med Chem Lett 18(24): 6352-6356.

Kuipers, O. P., M. M. Beerthuyzen, et al. (1995). "Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction." The Journal of biological chemistry 270(45): 27299-27304.

Kuipers, O. P., P. G. de Ruyter, et al. (1998). "Quorum sensing-controlled gene expression in lactic acid bacteria." J Biotechnol 64: 15-21.

Kulahin, N., O. Kristensen, et al. (2011). "Structural model and trans-interaction of the entire ectodomain of the olfactory cell adhesion molecule." Structure 19(2): 203-211.

Kunji, E. R., D. J. Slotboom, et al. (2003). "*Lactococcus lactis* as host for overproduction of functional membrane proteins." Biochimica et biophysica acta 1610(1): 97-108.

Lawson, D. M., P. J. Artymiuk, et al. (1991). "Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts." Nature 349(6309): 541-544.

Li, M. Z. and S. J. Elledge (2007). "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC." Nat Methods 4(3): 251-256.

Linding, R., L. J. Jensen, et al. (2003). "Protein disorder prediction: implications for structural proteomics." Structure 11(11): 1453-1459.

Lo Conte, L., C. Chothia, et al. (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol 285(5): 2177-2198.

Low, B. W., C. C. Chen, et al. (1966). "Studies of insulin crystals at low temperatures: effects on mosaic character and radiation sensitivity." Proc Natl Acad Sci U S A 56(6): 1746-1750.

Lundstrom, K. (2006a). "Structural genomics for membrane proteins." Cellular and molecular life sciences : CMLS 63(22): 2597-2607.

Lundstrom, K., R. Wagner, et al. (2006b). "Structural genomics on membrane proteins: comparison of more than 100 GPCRs in 3 expression systems." Journal of structural and functional genomics 7(2): 77-91.

Madoori, P. K., H. Agustiandari, et al. (2009). "Structure of the transcriptional regulator LmrR and its mechanism of multidrug recognition." EMBO J 28(2): 156-166.

Mavaro, A., A. Abts, et al. (2011). "Substrate recognition and specificity of NISB, the lantibiotic dehydratase involved in nisin biosynthesis." J Biol Chem.

McElroy, H. H., G. W. Sissom, et al. (1992). "Studies on engineering crystallizability by muttion of surface residues of human thymidylate synthase." J. Cryst. Growth 122: 265-272.

McPherson, A. and P. Shlichta (1988). "Heterogeneous and epitaxial nucleation of protein crystals on mineral surfaces." Science 239(4838): 385-387.

Menetrey, J., A. Bahloul, et al. (2005). "The structure of the myosin VI motor reveals the mechanism of directionality reversal." Nature 435(7043): 779-785.

Midgett, C. R. and D. R. Madden (2007). "Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies." Journal of structural biology 160(3): 265-274.

Mierau, I. and M. Kleerebezem (2005). "10 years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis." Applied microbiology and biotechnology 68(6): 705-717.

Muller, K. M., K. M. Arndt, et al. (1998). "Tandem immobilized metal-ion affinity chromatography/immunoaffinity purification of His-tagged proteins--evaluation of two anti-His-tag monoclonal antibodies." Anal Biochem 259(1): 54-61.

Murphy, K. P., P. L. Privalov, et al. (1990). "Common features of protein unfolding and dissolution of hydrophobic compounds." Science 247(4942): 559-561.

Ostermeier, C., S. Iwata, et al. (1995). "Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase." Nat Struct Biol 2(10): 842-846.

Oswald, C., S. H. Smits, et al. (2008). "Microseeding - a powerful tool for crystallizing proteins complexed with hydrolyzable substrates." Int J Mol Sci 9(7): 1131-1141.

Pittelkow, M., B. Tschapek, et al. (2011). "The Crystal Structure of the Substrate-Binding Protein OpuBC from *Bacillus subtilis* in Complex with Choline." Journal of molecular biology 411(1): 53-67.

Porath, J., J. Carlsson, et al. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature 258(5536): 598-599.

Regnier, F. E. (1983a). "High-performance liquid chromatography of biopolymers." Science 222(4621): 245-252.

Regnier, F. E. (1983b). "High-performance liquid chromatography of proteins." Methods Enzymol 91: 137-190.

Ressl, S., A. C. Terwisscha van Scheltinga, et al. (2009). "Molecular basis of transport and regulation in the Na(+)/betaine symporter BetP." Nature 458(7234): 47-52.

Rodgers, D. W. (1994). "Cryocrystallography." Structure 2(12): 1135-1140.

Rosenbaum, D. F. and C. F. Zukoski (1996). "Protein interactions and crystallization." Journal of Crystal Growth 169(4): 752-758.

Rubio, V., Y. Shen, et al. (2005). "An alternative tandem affinity purification strategy applied to Arabidopsis protein complex isolation." The Plant journal : for cell and molecular biology 41(5): 767-778.

Rupp, B. (2007). "Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology." Taylor & Francis Ltd 1.

Sachdev, D. and J. M. Chirgwin (1999). "Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein." Journal of protein chemistry 18(1): 127-136.

Sala, E. and A. de Marco (2010). "Screening optimized protein purification protocols by coupling small-scale expression and mini-size exclusion chromatography." Protein Expr Purif 74(2): 231-235.

Sauter, C., B. Lorber, et al. (1999). "Crystallogenesis studies on yeast aspartyl-tRNA synthetase: use of phase diagram to improve crystal quality." Acta Crystallogr D Biol Crystallogr 55(Pt 1): 149-156.

Schmidt, T. G., J. Koepke, et al. (1996). "Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin." Journal of molecular biology 255(5): 753-766.

Schwarz, C. K., B. Tschapek, et al. (2011). "Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the *Escherichia coli* ABC transporter haemolysin B." Acta Crystallogr Sect F Struct Biol Cryst Commun 67(Pt 5): 630-633.

Senkovich, O., W. J. Cook, et al. (2007). "Structure of a complex of human lactoferrin N-lobe with pneumococcal surface protein a provides insight into microbial defense mechanism." Journal of molecular biology 370(4): 701-713.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase." Gene 67(1): 31-40.

Smith, G. E., M. D. Summers, et al. (1983). "Production of human beta interferon in insect cells infected with a baculovirus expression vector." Molecular and cellular biology 3(12): 2156-2165.

Smits, S. H., A. Mueller, et al. (2008a). "Coenzyme- and His-tag-induced crystallization of octopine dehydrogenase." Acta Crystallogr Sect F Struct Biol Cryst Commun 64(Pt 9): 836-839.

Smits, S. H., A. Mueller, et al. (2008b). "A structural basis for substrate selectivity and stereoselectivity in octopine dehydrogenase from *Pecten maximus*." J Mol Biol 381(1): 200-211.

Sorensen, H. P. and K. K. Mortensen (2005). "Advanced genetic strategies for recombinant protein expression in *Escherichia coli*." Journal of biotechnology 115(2): 113-128.

Stofko-Hahn, R. E., D. W. Carr, et al. (1992). "A single step purification for recombinant proteins. Characterization of a microtubule associated protein (MAP 2) fragment which associates with the type II cAMP-dependent protein kinase." FEBS letters 302(3): 274-278.

Taha, M. (2005). "Buffers for the physiological pH range: acidic dissociation constants of zwitterionic compounds in various hydroorganic media." Ann Chim 95(1-2): 105-109.

Tang, C., C. D. Schwieters, et al. (2007). "Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR." Nature 449(7165): 1078-1082.

Tate, C. G., J. Haase, et al. (2003). "Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter." Biochimica et biophysica acta 1610(1): 141-153.

Terpe, K. (2003). "Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems." Applied microbiology and biotechnology 60(5): 523-533.

Teuber, M. (1995). "The genus Lactococcus." The genera of lactic acid bacteria 2.

Tornquist, K., J. Paallysaho, et al. (1995). "Influence of Hepes- and CO2/HCO(3-)-buffer on Ca2+ transients induced by TRH and elevated K+ in rat pituitary GH4C1 cells." Mol Cell Endocrinol 112(1): 77-82.

Tschapek, B., M. Pittelkow, et al. (2011). "Arg149 Is Involved in Switching the Low Affinity, Open State of the Binding Protein AfProX into Its High Affinity, Closed State." Journal of molecular biology 411(1): 36-52.

Urh, M., D. York, et al. (1995). "Buffer composition mediates a switch between cooperative and independent binding of an initiator protein to DNA." Gene 164(1): 1-7.

van de Guchte, M., J. Kok, et al. (1992). "Gene expression in *Lactococcus lactis*." FEMS microbiology reviews 8(2): 73-92.

Voss, S. and A. Skerra (1997). "Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification." Protein engineering 10(8): 975-982.

Waugh, D. S. (2005). "Making the most of affinity tags." Trends in biotechnology 23(6): 316-320.

Wegner, G. H. (1990). "Emerging applications of the methylotrophic yeasts." FEMS microbiology reviews 7(3-4): 279-283.

Weik, M., G. Kryger, et al. (2001). "Solvent behaviour in flash-cooled protein crystals at cryogenic temperatures." Acta Crystallogr D Biol Crystallogr 57(Pt 4): 566-573.

Zaitseva, J., I. B. Holland, et al. (2004). "The role of CAPS buffer in expanding the crystallization space of the nucleotide-binding domain of the ABC transporter haemolysin B from *Escherichia coli*." Acta Crystallogr D Biol Crystallogr 60(Pt 6): 1076-1084.

Zaitseva, J., S. Jenewein, et al. (2005). "Functional characterization and ATP-induced dimerization of the isolated ABC-domain of the haemolysin B transporter." Biochemistry 44(28): 9680-9690.

Zheng, C. F., T. Simcox, et al. (1997). "A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins." Gene 186(1): 55-60.

Zimmerman, S. B. and S. O. Trach (1991). "Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*." J Mol Biol 222(3): 599-620.

5.5 Chapter V

Title:	Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form.
Authors:	Schwarz C. K. W., Landsberg C. D., Lenders M. H., Smits S. H. J., Schmitt L.
Published in:	J Biotechnol
Impact Factor:	2.970
Own proportion of this work:	65 % Construction of the used plasmids. Secretion analysis. Purification of the used proteins. Biochemical analysis of the proteins. Writing of the manuscript.

Journal of Biotechnology 159 (2012) 155–161

Contents lists available at SciVerse ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form

Christian K.W. Schwarz, Christine D. Landsberg, Michael H.H. Lenders, Sander H.J. Smits, Lutz Schmitt*

Institute of Biochemistry, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 4 October 2011
Received in revised form 7 February 2012
Accepted 9 February 2012
Available online 17 February 2012

Keywords:
Type 1 secretion
Haemolysin A
IFABP
Inclusion bodies
FKBP
Purification protocol

ABSTRACT

A biotechnological production of proteins through protein secretion systems might be superior to the conventional cytoplasmic production, because of the absence of large amounts of proteases present in the extracellular space and the ease of purification or downstream processing. However, secretion of proteins is still a trial-and-error approach and many proteins fail to be secreted. Recently, a study of a Type 1 secretion system revealed that the folding rate of the passenger protein dictates secretion efficiency. Here, the well-known MalE failed to be secreted when fused to a C-terminal fragment of the natural substrate haemolysin A. In contrast, slow-folding mutants of MalE were secreted in high yields. However, MalE is a bacterial protein that is targeted to the periplasmic space of *E. coli* and possesses the intrinsic capability to cross a membrane. Therefore, we applied the same approach for another eukaryotic protein that resides in the cytoplasm. As an example, we chose the intestinal fatty acid binding protein (IFABP) and highlight the universal potential of this Type 1 secretion system to secrete proteins with slow-folding kinetics (here the G121V mutant). Finally, a one-step purification protocol was established yielding 1 mg of pure IFABP G121V per liter culture supernatant. Moreover, secreted IFABP G121V was shown to reach a folded state, which is biologically active.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Substrates of Type 1 secretion systems (T1SS) are translocated in one step from the cytoplasm to the exterior via dedicated transport systems in Gram-negative bacteria. The substrate spectrum of these T1SS contains structurally and functionally unrelated proteins and, more importantly, proteins of different molecular weights ranging from, for example, the hemophore HasA (19 kDa) (Letoffe et al., 1994) or the lipase LipA (65 kDa) of *Serratia marcescens* (Meier et al., 2007) to the 800 kDa protein LapA that is involved in the biofilm formation by *Pseudomonas fluorescence* (Hinsa et al., 2003).

In general, T1SS mediate the passage across the inner and outer membrane of Gram-negative bacteria in one step, through the concerted interaction of two inner membrane proteins, a membrane fusion protein (MFP) and an ATP binding cassette (ABC) transporter, and an outer membrane factor (OMF). These proteins form a complex that catalyzes the secretion without the appearance of a periplasmic intermediate (Holland et al., 2005; Thanabalu et al., 1998).

The paradigm of such a T1SS is the haemolysin A (HlyA) secretion system of *Escherichia coli* (*E. coli*), which is composed of the ABC transporter haemolysin B (HlyB), the MFP haemolysin D (HlyD), and the multi-functional TolC (Jarchau et al., 1994). Upon interaction of the substrate, HlyA, with HlyB and/or HlyD, TolC is recruited and a continuous export channel is formed from the cytosol directly to the exterior (Benabdelhak et al., 2003; Thanabalu et al., 1998). HlyA is a 110 kDa toxin, which lyses, for example, erythrocytes (Linggood and Ingram, 1982). The non-cleavable secretion signal of HlyA is located at the C-terminal 50–60 amino acids and is essential and, by itself, sufficient for secretion (Gray et al., 1986; Hess et al., 1990; Jarchau et al., 1994; Kenny et al., 1994; Stanley et al., 1991).

The common denominator of HlyA and the vast majority of T1SS substrates are glycine-rich nonapeptide repeats (GG repeats) that contain the consensus sequence GGxGxDxUx (x: any amino acid and U: large or hydrophobic amino acid), which are located upstream of the secretion signal. These GG repeats increase the secretion levels as demonstrated by the higher yields of secreted proteins, when compared to the secretion signal alone (Kenny et al., 1991; Sanchez-Magraner et al., 2007). Therefore, a 23 kDa C-terminal fragment of HlyA consisting of 218 amino acids and three GG repeats was used for the secretion of a couple of fusion proteins, such as β-lactamase or scFv. Here, the gene of interest was fused 5′ to the gene encoding the 23 kDa C-terminal fragment of

Abbreviations: DAUDA, 11-(dansylamino)undecanoic acid; ddFKBP, destabilizing domain FKBP12; FKBP12, human FK506- or rapamycin-binding protein; GG repeats, glycine-rich, nonapeptide repeats; HlyA, haemolysin A; HlyAc, C-terminal fragment of HlyA; IFABP, rat intestinal fatty acid binding protein; IMAC, immobilized metal-ion affinity chromatography; MalE, maltose binding protein; T1SS, Type 1 secretion system.

* Corresponding author. Tel.: +49 211 81 10773; fax: +49 211 81 15310.
E-mail address: Lutz.Schmitt@hhu.de (L. Schmitt).

0168-1656/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2012.02.005

HlyA (Blight and Holland, 1994; Chervaux et al., 1995; Fernandez et al., 2000; Mackman et al., 1987). Despite promising initial results, many of the chimeras employed could not be secreted, which limited the biotechnological exploitation of this T1SS (Blight and Holland, 1994). Recently, Bakkes et al. (2010) showed that a fusion protein composed of the maltose binding protein (MalE) and the above-mentioned 23 kDa fragment of HlyA (HlyAc), which was initially not secreted, could be secreted by the introduction of slow-folding mutations in MalE (Bakkes et al., 2010). However, MalE endows intrinsic properties to pass membranes as it is normally translocated via the Sec pathway into the periplasm of *E. coli* (Cover et al., 1987). Although the slow-folding strategy was successfully applied for MalE, several questions remained such as if the described T1SS is capable of secreting: (1) proteins of heterologous origin, for example, from eukaryotic origin, (2) proteins, which are normally not transported across membranes, (3) proteins that are normally expressed as inclusion bodies or (4) biologically active proteins?

To answer these questions, the eukaryotic, cytoplasmic intestinal fatty acid binding protein (IFABP) from rats was fused to HlyAc. The protein consists of 132 amino acids (15.1 kDa), binds a single molecule of a long-chain fatty acid and is important in the fatty acid metabolism of humans as well as other mammals (Rowland et al., 2009; Sweetser et al., 1987). In vitro, it is also used for binding analysis of lipophilic drugs and the prediction of drug pharmacokinetic parameters (Chuang et al., 2008; Rowland et al., 2009).

The slow-folding mutant G121 V of IFABP binds fatty acids with high affinity (K_D = 120 nM ± 25 nM) but forms, in contrast to the wild-type, inclusion bodies during its expression in *E. coli* (Kim and Frieden, 1998). Therefore, we chose IFABP G121V to investigate the general applicability of the HlyA T1SS for the secretion of heterologous, slow-folding substrates, the impact of inclusion body formation onto the secretion system and to compare the biological activity of secreted IFABP with the biological activity of IFABP G121V purified from inclusion bodies.

If slow-folding mutations of a desired protein are not available, directed evolution experiments might be used (Crameri et al., 1996, 1998). However, these methods are based on trial-and-error approaches and high-throughput screening methods are necessary to identify high-secreting clones (Farinas et al., 2001). Therefore, this approach is time-consuming and cost-intensive. A solution might be the incorporation of a "destabilizing domain" in front of the passanger protein. Such domains confer instability to the fused proteins and induce a subsequent degradation of the fusion proteins. The destabilizing domain FKBP (ddFKBP), an engineered version of the human FK506- or the rapamycin-binding protein (FKBP12), is the best-characterized destabilization domain in eukaryotes (Armstrong and Goldberg, 2007; Banaszynski et al., 2006; Herm-Gotz et al., 2007; Park et al., 1992). However, no studies in bacteria are available. Thus, the fusion of the protein of interest to a destablizing domain might provide a general tool to decrease its folding kinetics. This, in principle, should facilitate the secretion of these chimeras through the T1SS.

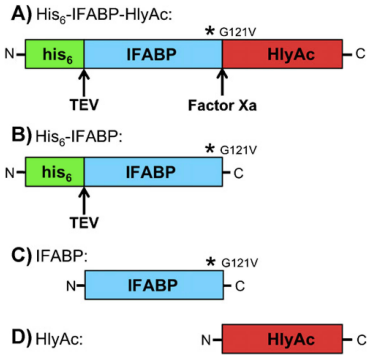


Fig. 1. Schematic view of the constructs used in this study. (A) Rat IFABP(wt) or IFABP(G121V) (blue) with an N-terminal his_6-tag (green) were cloned in front of the secretion module HlyAc (red); 42.1 kDa. (B) After the factor Xa digest, his_6-IFABP was separated from HlyAc; 18.3 kDa. (C) The his_6-tag was removed with the TEV protease; 15.3 kDa. (D) HlyAc, used as secretion module, is shown; 27.2 kDa.

Accordingly, we compared the potential of the destabilizing domain with the introduction of slow-folding mutants on the secretability of IFABP. Our results demonstrated that the ddFKBP is not applicable to increase the secretion levels, but that the introduction of a slow-folding mutation into IFABP converts a non-secreted protein into a well-accepted substrate of the HlyA T1SS. Furthermore, we show that secreted IFABP is correctly folded and biologicl active, suggesting that this secretion system is an attractive candidate for a biotechnological protein production system in appropiate yields.

2. Materials and methods

2.1. Construction of the vectors

The genes of rat IFABP WT and rat IFABP G121V were encoded on plasmids pQE IFABP WT and pQE IFABP G121V, respectively (kind gift of Dr. Carl Frieden, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, USA) (Kim and Frieden, 1998). The plasmids pSOI-IFABP(wt), pSOI-ddFKBP IFABP(wt), pSOI-IFABP(G121V) and pSOI-ddFKBP IFABP(G121V) were prepared using the *In-Fusion® Advantage PCR Cloning Kit* (Clontech). Employing this strategy, IFABP–HlyAc and ddFKBP–IFABP–HlyAc fusion proteins were generated according to Bakkes et al. (2010) including a factor Xa cleavage site located between IFABP and HlyAc and an N-terminal his_6-tag followed by a TEV protease cleavage site (see Fig. 1). The sequences of all used constructs (see Table 1) were verified by DNA sequencing.

Table 1
Plasmids used in this study.

Plasmids		
Name	Description	Reference
pK184-HlyBD	Plasmid encoding for the genes *hlyB* and *hlyD*	Bakkes et al. (2010)
pSOI-HlyAc	Plasmid encoding for the 23 kDa C-terminal fragment of HlyA (HlyAc)	Bakkes et al. (2010)
pSOI-IFABP(wt)	IFABP(wt) inserted in pSOI-HlyAc (his_6–IFABP(wt)–HlyAc)	This study
pSOI-IFABP(G121V)	IFABP(G121V) inserted in pSOI-HlyAc (his_6–IFABP(G121V)–HlyAc)	This study
pSOI-ddFKBP IFABP(wt)	ddFKBP inserted in pSOI-IFABP(wt) N-terminally of IFABP(wt) (ddFKBP–IFABP(wt)–HlyAc)	This study
pSOI-ddFKBP IFABP(G121V)	ddFKBP inserted in pSOI-IFABP(G121V) N-terminally of IFABP(G121V) (ddFKBP–IFABP(G121V)–HlyAc)	This study

2.2. Protein expression, secretion and purification

2.2.1. Secretion experiments

All secretion experiments were performed with small variations as described (Bakkes et al., 2010). In brief, *E. coli* WM2429 carrying both, pK184-HlyBD and a pSOI construct, was grown in 2xYT medium supplemented with kanamycin (30 μg mL^{-1}) and ampicillin (100 μg mL^{-1}) at 37 °C with agitation. At an OD_{600} of 0.5, expression of *hlyB* and *hlyD* and the fusion genes, encoded by the pSOI plasmids, were induced with 1.5 mM IPTG and 10 mM arabinose. The pSOI plasmids carry an arabinose promoter, while a *lac* promotor controls the expression of HlyB/D. After 4 h growth, cell cultures were centrifuged for 20 min at 14,000 × *g*, 4 °C and cells (OD_{eq} = 0.1) and supernatant samples (16 μL) were analyzed by SDS-PAGE using Coomassie Brilliant Blue (CBB) staining and/or Western blotting using a HlyA specific, polyclonal antibody or an anti-his-tag antibody (Qiagen, Hilden). An HRP-conjugated, secondary antibody in combination with the ECL advance kit (GE Healthcare) was used for the visualization of the fusion bands. If indicated, 160 μL culture supernatant was precipitated for 30 min on ice with 20% trichloroacedic acid (TCA). After centrifugation for 20 min at 4 °C, the pellet was resolved and used for SDS-PAGE analysis.

2.2.2. Purification of the secreted proteins HlyAc or his_6–IFABP(G121V)–HlyAc

Cells secreting either HlyAc or his_6–IFABP(G121V)–HlyAc were grown for 4 h at 37 °C with agitation. After centrifugation for 20 min and 14,000 × *g*, 4 °C, the supernatant was supplemented with 10 mM imidazole and 20 mM $MgCl_2$. For the subsequent immobilized metal-ion affinity chromatography (IMAC), the supernatant was loaded on a HiTrap IMAC HP column (5 mL, GE Healthcare) preloaded with Ni^{2+} and pre-equilibrated in buffer (10 mM Tris–HCl, 150 mM NaCl, 10 mM imidazole, pH 8.0) using a FPLC system (ÄktaSystem, GE Healthcare). After washing the column with this buffer, HlyAc or his_6–IFABP(G121V)–HlyAc was eluted with the buffer supplemented with 250 mM imidazole and a linear gradient of 10 column volumes ranging from 10 to 250 mM imidazole. Protein fractions were pooled and concentrated by ultrafiltration (Amicon Filter devices, 10 kDa MWCO). The protein concentration was determined with a NanoDrop device (PeqLab) using the calculated extinction coefficients and molecular masses (HlyAc: 23,380 M^{-1} cm^{-1}, 27.2 kDa; his_6–IFABP(G121V)–HlyAc: 40,340 M^{-1} cm^{-1}, 42.1 kDa; his_6–IFABP(G121V): 18,450 M^{-1} cm^{-1}, 18.3 kDa) (http://expasy.org/tools/protparam.html).

2.2.3. Factor Xa digestion and anion-exchange chromatography

50 μg his_6–IFABP(G121V)–HlyAc was cleaved with 1 μg factor Xa (NEB) overnight at 25 °C. An anion-exchange chromatography was performed to separate cleaved his_6–IFABP(G121V) and uncleaved his_6–IFABP(G121V)–HlyAc, HlyAc and factor Xa. The digestion mixture was diluted tenfold in 10 mM NaH_2PO_4, 20 mM NaCl, pH 7.3 and loaded on a Q HP column (GE Healthcare) with a FPLC system (GE Healthcare). The his_6–IFABP(G121V) eluted in the flow through.

2.2.4. Removal of the his_6-tag with the TEV protease

4 μg his_6–IFABP(G121V) was digested with 1 U of TEV protease (ProSpec) overnight at 25 °C in buffer supplemented with 10 mM dithiothreitol (DTT). After cleavage, the his_6-tag and the TEV protease (containing a his-tag) were removed by incubation of the mixture with magnetic Ni^{2+} beads (Qiagen) and the supernatant was stored until further analysis.

2.2.5. Expression and purification of rat IFABP G121V from E. coli

Chemically competent *E. coli* BL21 (DE3) was transformed with pQE IFABP G121V and grown on LB agar plates supplemented with 100 μg mL^{-1} ampicillin. Overnight cultures of single colonies were used to inoculate 2 L 2xYT medium containing 100 μg mL^{-1} ampicillin with a start OD_{600} of 0.1 at 37 °C and gene expression was induced with 1 mM IPTG at OD_{600} of 0.5. Cells were grown for 4 h at 160 rpm and agitation, harvested by centrifugation and directly used or stored at −20 °C. Rat IFABP G121V was purified from inclusion bodies with minor changes as described (Kim and Frieden, 1998). Briefly, cells were broken with a cell disruptor (Constant Systems) and the lysate was centrifuged (20 min, 13,500 × *g*). The pellet was washed twice with 20 mM Tris–HCl, 0.25 mM EDTA, 0.5% Triton X-100, pH 8.0 and twice with 20 mM Tris–HCl, 0.25 mM EDTA, pH 8.0. Crude inclusion bodies were dissolved in 20 mM Tris–HCl, 0.25 mM EDTA, 6 M guanidinium-hydrochloride, pH 8.0 and the protein concentration was determined with a NanoDrop device (PeqLab) as above. For refolding, IFABP G121V was diluted to 0.2 mg mL^{-1} in 20 mM Tris–HCl, 0.25 mM EDTA, pH 8.0 and residual guanidinium-hydrochloride was removed by extensive dialysis overnight at 4 °C. Dialyzed, refolded IFABP G121V was centrifuged (20 min, 50,000 × g, 4 °C) and the supernatant was purified by anion-exchange chromatography (5 mL HP Q column, GE Healthcare). IFABP G121V eluted in the flow through. IFABP G121V was concentrated via ultrafiltration (Amicon filter device, MWCO: 10 kDa) and applied to a size-exclusion chromatography (Superdex 75 16/60, GE Healthcare) pre-equilibrated in 20 mM K_2HPO_4, 0.25 mM EDTA, pH 7.3. Protein containing fractions were concentrated by ultrafiltration (Amicon filter device, MWCO: 10 kDa) and stored at −80 °C.

2.3. Folding state of secreted IFABP (G121V)

Intrinsic tryptophane fluorescence measurements were performed using 0.5 μM protein at 20 °C and a Fluorolog®-3 (Horiba) in buffer (20 mM K_2HPO_4, 0.25 mM EDTA, pH 7.3) with or without 6 M urea. 290 nm was used as excitation wavelength and the emission spectrum was recorded between 300 nm and 440 nm (slit width 5 nm for both, excitation and emission).

2.4. Binding affinity determination of DAUDA to IFABP G121V

11-(Dansylamino)undecanoic acid (DAUDA, Cayman Chemical) was prepared as a 1 mM stock solution in 50% isopropanol and 50% buffer (20 mM KH_2PO_4, 0.25 mM EDTA, pH 7.3) and further diluted to the appropriate concentrations with buffer. 100 nM IFABP was used in buffer and a 1 mL silica glass cuvette. The fluorescence signal was recorded at a wavelength of 500 nm after excitation of DAUDA at 350 nm using a Fluorolog®-3 (Horiba) (Kim and Frieden, 1998). The slit width was adjusted to 3.5 nm for both, the excitation and emission, and the signal was integrated for 0.5 s. Fluorescence signals of DAUDA in buffer in the absence of protein was subtracted as background. The corrected and normalized fluorescence signals were plotted against the DAUDA concentration. Assuming one binding site per monomer in IFABP, the K_D for substrate binding was determined by non-linear, least-squares fitting of the data to the following equation: $F = F_0 + (\Delta F/2P_0)[(K_D + P_0 + L_0) - ((K_D + P_0 + L_0)^2 - 4L_0P_0)^{1/2}]$, which corrected the data for the concentration of the receptor. F is the measured fluorescence; F_0 the fluorescence of IFABP in the absence of DAUDA; ΔF is the change of fluorescence; P_0 and L_0 are the total concentrations of protein and substrate. Data were fitted using the program Prism 5.

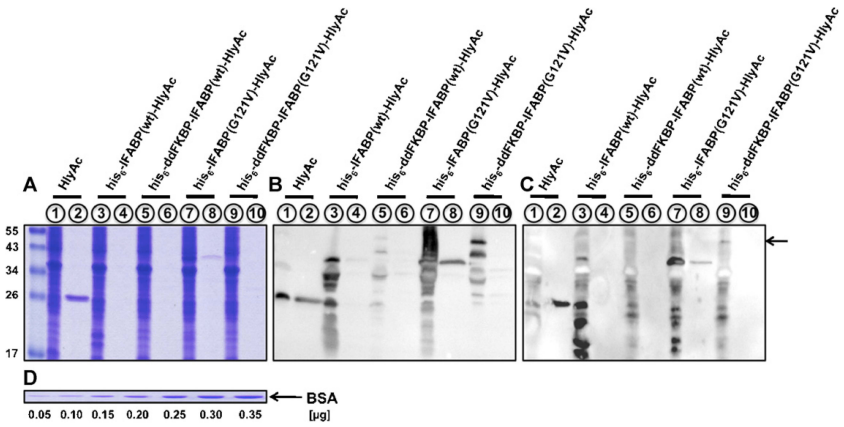


Fig. 2. Secretion experiments and analysis of whole cell extracts and supernatant samples. The arrow indicates the position of his_6–ddFKBP–IFABP(wt)–HlyAc and his_6–ddFKBP–IFABP(G121V)–HlyAc (54.0 kDa). (A) CBB staining, (B) and (C) western blott using a polyclonal antibody raised against HlyA or an antibody against a his-tag (Qiagen, Hilden). (D) SDS-PAGE gel of BSA with indicated amounts. Molecular weight markers with corresponding sizes are shown in (A). For pictures A–C, lanes are numbered from 1 to 10. Odd numbers: whole cell extracts ($OD_{eq} \approx 0.1$), even numbers: 16 µL cell-free supernatant.

3. Results

3.1. Slow-folding IFABP is secreted via the HlyA system

In a previous study, a fusion protein of wild-type MBP and HlyAc was not secreted, while slow-folding mutants of MalE were secretable with different efficiencies, indicating an inverse correlation between the folding rate and the secretion efficiency (Bakkes et al., 2010). However, MalE originates from *E. coli* and is a protein that is endowed with intrinsic properties for the translocation across membranes (Cover et al., 1987). Therefore, we challenged the HlyA secretion system with the eukaryotic rat intestinal fatty acid binding protein (IFABP), which is naturally localized in the cytoplasm (Hardy and Randall, 1991; Sweetser et al., 1987; Weiss et al., 1988). The pSOI-IFABP plasmids, encoding fusions of IFABP(wt) or IFABP(G121V) with HlyAc, were designed as described in Bakkes et al. (2010). His_6-IFABP(wt)-HlyAc and his_6–IFABP(G121V)–HlyAc (42.1 kDa) were produced intracellularly as shown by SDS-PAGE analysis and CBB staining. This was confirmed by western blotting using a polyclonal antibody against HlyA or the N-terminal his-tag of the proteins (see Fig. 2B and C, lanes 3 and 7). However, the wild-type IFABP fusion protein was not detected in the culture supernatant by CBB staining or western blots with the anti-his antibody, and the anti-HlyA antibody (Fig. 2A–C, lane 4 and Fig. 3). In striking contrast, the fusion protein of slow-folding IFABP(G121V) with HlyAc was secreted with good yields (Fig. 2A, lane 8 and Fig. 3). Remarkably, the secreted fusion protein could already be visualized by CBB staining without previous concentration of the sample. Nevertheless, TCA precipitated supernatants were analyzed by SDS-PAGE and CBB staining to highlight the purity of secreted IFABP(G121V) (see Fig. 3). Based on a comparison with known amounts of BSA (Fig. 2A, lane 8 with D), the amount of secreted his_6–IFABP(G121V)–HlyAc was quantified to be ~6 mg L^{-1} of culture corresponding to 2 mg/OD_{600}. Importantly, IFABP(G121V) was secreted as a soluble protein and no aggregation could be observed, which is in contrast to the cytoplasmic expression, which leads to insoluble inclusion bodies (Kim and Frieden, 1998).

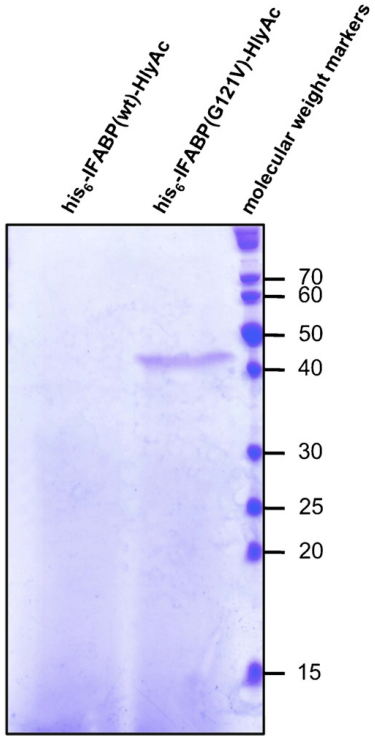


Fig. 3. TCA precipitation of supernatant samples. Cells carrying pK184-HlyBD and either pSOI-IFABP(wt) or pSOI-IFABP(G121V) were grown as described in Section 2 160 µL cell-free supernatant was TCA precipitated, the pellets were resolved, loaded on a SDS-PAGE gel and visualized with CBB. Molecular weight standards with corresponding masses are shown in the right panel.

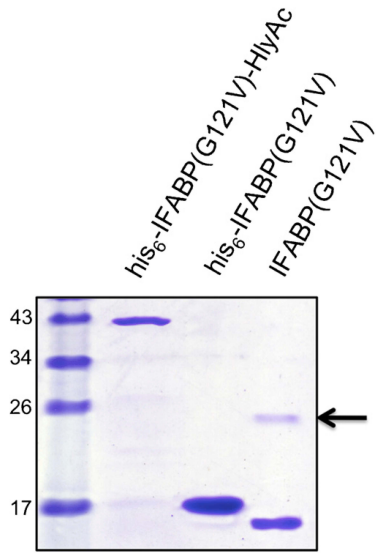


Fig. 4. SDS-PAGE analysis of the purification and the proteolytic digestion of secreted his_6–IFABP(G121V)–HlyAc fusion protein. His_6–IFABP(G121V)–HlyAc was eluted from an IMAC column (left lane), digested with Factor Xa to remove HlyAc and produce his_6–IFABP(G121V) (middle lane), and subsequently with the TEV protease (ProSpec) to remove the his_6-tag from secreted IFABP(G121V) (right lane). The arrow indicates the TEV protease (27 kDa), which was not entirely removed. Molecular weight markers with corresponding sizes are shown on the left.

3.2. ddFKBP does not promote secretion

The above results indicated that the folding rate of a protein dictates the secretion efficiency. Therefore, we were intrigued whether a "folding-destabilizer", such as ddFKBP, would result in secretion of wild-type IFABP-HlyAc in *E. coli*. Thus, ddFKBP was inserted behind the TEV protease cleavage site, at the N-terminus of IFABP(wt)–HlyAc and IFABP(G121V)–HlyAc, respectively. Full-length proteins (54.0 kDa) were expressed inside the cells since the N-terminal his_6-tag and HlyAc were both detectable by Western blot analysis of whole cell extracts (see Fig. 2B and C, lanes 5 and 9). However, the overall protein levels were reduced in the ddFKBP contexts when compared to the levels of proteins without ddFKBP. This suggests either a reduced expression rate or a ddFKBP-induced proteolysis. Also, IFABP WT was still not secreted (see Fig. 2, lane 6) and the secretion levels of IFABP(G121V)–HlyAc with the ddFKBP decreased (compare Fig. 2, lanes 8–10).

3.3. Purification and concentration of his_6–IFABP(G121V)–HlyAc in one step

For the separation of the his-tagged IFABP from the few impurities present in the culture medium (Mergulhao et al., 2005; Sandkvist and Bagdasarian, 1996), a one-step purification protocol was established. SDS-PAGE analysis (see Fig. 4) indicated the high purity of the eluted protein. In addition, the 260/280 ratio between 0.57 and 0.60 (data not shown) is in good agreement with the theoretical ratio of 0.57 for pure proteins (Glasel, 1995) and suggests the absence of RNA or DNA impurities in the protein sample. In summary, 1 mg of pure and homogenous his_6–IFABP(G121V)–HlyAc could be obtained from a 1 L bacterial cell culture.

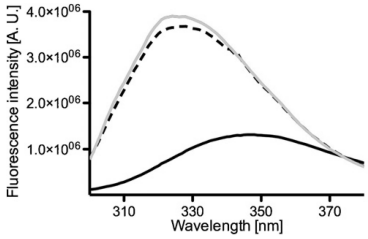


Fig. 5. Intrinsic tryptophan fluorescence measurements of IFABP G121V. The proteins (0.5 μM) were excited with a wavelength of 290 nm and the fluorescence was recorded between 300 and 380 nm. The fluorescence maximum of IFABP G121V in 6 M urea is at 346 nm, whereas the maximum shifts towards 326 nm in buffer without urea. The emission spectrum of secreted his_6–IFABP(G121V) superimposes with IFABP G121V in buffer without urea, indicating a folded state. Dashed line: IFABP G121V purified from inclusion bodies in buffer without urea; black line: IFABP G121 V purified from inclusion bodies in buffer with 6 M urea; grey line: his_6–IFABP(G121V) in buffer without urea.

3.4. Removal of HlyAc and the his_6-tag to generate IFABP(G121V)

In addition to the proteins of interest (wild-type IFABP or IFABP (G121V)), the constructed plasmids encode the C-terminal fragment HlyAc (23 kDa), an N-terminal his_6-tag and two specific protease cleavage sites for the TEV protease and factor Xa (see Fig. 1). Therefore, passenger proteins can be separated from HlyAc by a proteolytic digestion with factor Xa. His_6–IFABP(G121V)–HlyAc was incubated with factor Xa (NEB) and the flow-through of an anion-exchange chromatography contained pure, cleaved his_6–IFABP(G121V) (see Fig. 4). The his_6-tag was removed with the TEV protease. However, the TEV protease could not be removed quantitatively (see Fig. 4).

3.5. Folding and substrate binding activity of his_6–IFABP(G121V)–HlyAc, his_6–IFABP(G121V) and IFABP(G121V)

In this study, we investigated the folding state of secreted IFABP(G121V) by intrinsic tryptophan fluorescence spectroscopy and compared it with the IFABP G121V purified from inclusion bodies. The emission spectrum of IFABP G121V in 6 M urea showed a maximum at 346 nm. In buffer without urea, the emission intensity significantly increased and the maximum shifted to 326 nm, which is in line with previous studies (see Fig. 5) (Li and Frieden, 2007). Secreted his_6–IFABP(G121V) exhibited the same emission spectrum as IFABP G121V from inclusion bodies without urea. Similar results were obtained for his_6–IFABP(G121V)–HlyAc. However, the emission spectrum was slightly influenced by the presence of HlyAc (data not shown). These data indicate that secreted IFABP(G121V) is folded comparable to the protein purified from inclusion bodies.

Next, we investigated the binding activity of secreted IFABP(G121V) in titration experiments with the fatty acid analog 11-(dansylamino)undecanoic acid (DAUDA). The fluorescence increase at 500 nm upon binding was monitored as described before (see Fig. 6) (Kim and Frieden, 1998; Wilkinson and Wilton, 1986). Fluorescence was corrected for the IFABP concentrations (100 nM) with the "receptor correction equation" (see Section 2). For a validation of the experimental setup, IFABP G121V was purified directly from inclusion bodies and the binding of DAUDA was measured. The determined dissociation constant (K_D) of 126 nM ± 7 nM is, within experimental errors, identical to the published data for IFABP G121V (120 nM ± 25 nM; see Table 2) (Kim and Frieden, 1998). The K_D value of secreted his_6–IFABP(G121V)–HlyAc (see Fig. 6 and Table 2) was determined to be 195 nM ± 13 nM, which is in the same range. As a control, HlyAc without fused

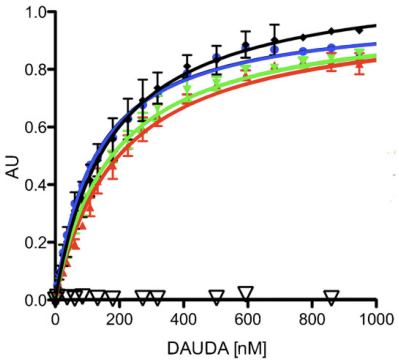


Fig. 6. Fluorescence titration plots using DAUDA, a fluorescence fatty acid analog. All of the IFABP variants share a similar binding curve and similar K_D values for the DAUDA binding, whereas HlyAc does not bind (see also Table 2). Corrected and normalized signals were plotted against the DAUDA concentration, and the solid line represents the fit employing a 1:1 binding-site model corecting for the protein concentration (see Section 2) with the program Prism 5. All plots represent two independent measurements with the error representing the standard deviation. Blue: IFABP G121V purified, refolded from inclusion bodies, red: his_6–IFABP(G121V)–HlyAc, green: his_6–IFABP(G121V), black: IFABP(G121V), open-triangles: HlyAc.

Table 2
Summary of the dissociation constants of the DAUDA/IFAB complexes used in this study. Data were analyzed as described in Section 2.

Protein	Dissociation constant (K_D)
IFABP G121V (Kim and Frieden, 1998)	120 ± 25 nM
IFABP G121V (out of IB)	126 ± 7 nM
His_6–IFABP(G121V)–HlyAc	195 ± 13 nM
His_6–IFABP(G121V)	174 ± 9 nM
IFABP(G121V)	138 ± 11 nM
HlyAc	No binding

IFABP was used under identical experimental conditions, and no binding of DAUDA was detected (see Fig. 6, open triangles). Nevertheless, a possible influence of HlyAc and the his_6-tag onto the binding behavior of IFABP was investigated. After proteolysis and subsequent purification of his_6–IFABP(G121V) and IFABP(G121V), respectively (see Fig. 4), the titration experiments were repeated. The K_D values of his_6–IFABP(G121V) without HlyAc and the K_D value of IFABP(G121V) were measured to be 174 nM ± 9 nM and 138 nM ± 11 nM, respectively. Therefore, secreted IFABP(G121V) exhibits the same binding affinity for DAUDA as the mutant protein purified directly from inclusion bodies.

4. Discussion

Secretion of proteins to the culture medium exhibits many advantages compared to a conventional, cytoplasmic protein location. For example, proteolytic degradation by intracellular proteases is prevented, the formation of disulfide bonds is enhanced in the oxidative environment of the medium and the subsequent purification protocols are simplified due to less proteinogenic impurities in the medium. These features allow faster downstream processing in combination with lower production costs. Many improvements in the protein production via secretion have been made in the recent decades and numerous systems are already commercially exploited. However, the secretion efficiency of these systems depends on the combination of the host strain, the signal peptide and the target protein and needs to be optimized by trial-and-error approaches for each specific task (Yoon et al., 2010).

With the HlyA T1SS of Gram-negative bacteria, high yields of secreted proteins can be produced in principle, and only three proteins are required to catalyze the secretion. The efficiency in combination with the simplicity of the secretion apparatus make this system an attractive candidate for the biotechnological production or proteins (Blight and Holland, 1994). However, the most important application so far has been the presentation of heterologous antigens in live-attenuated bacterial vaccines (Gentschev et al., 2002). The biotechnological usage is still hampered by the fact that many fusion proteins are not or only poorly secreted (Blight and Holland, 1994; Holland et al., 1990). Bakkes et al. (2010) demonstrated that the rate of folding of the passenger MalE dictates the efficiency of the secretion with the HlyA T1SS. While wild-type MalE was virtually not secreted, slow-folding mutants of MalE were secreted in yields that directly correlated with the folding rate of the corresponding mutant. Although, such mutations are an interesting approach to convert a protein into a secretion-competent substrate, the identification of mutations, which affect the folding rate without interfering with the function, is likely to be a very time-consuming process. As an alternative to this approach, the destabilizing domain ddFKBP was investigated (Armstrong and Goldberg, 2007; Banaszynski et al., 2006; Herm-Gotz et al., 2007). Despite the intriguing rational behind this approach, the secretion of wild-type IFABP failed completely, and the secretion levels of slow-folding IFABP(G121V) decreased, ruling out the applicability of the destabilizing domain.

MalE is an intrinsic *E. coli* protein designed for crossing a membrane. In this study, we wanted to expand the applicability of the system to a broader range of proteins. Indeed, we demonstrated that cytoplasmic, eukaryotic IFABP was not secreted when fused to HlyAc, whereas the fusion protein of the slow-folding IFABP G121V was secreted in yields of up to 6 mg L^{-1} cell culture. Nevertheless, the maximal secretion capacity of fusion proteins is not reached yet, as HlyAc alone without a passenger protein is secreted in much higher yields (see Fig. 2, lane 2).

Importantly, his_6–IFABP(G121V)–HlyAc was secreted as a soluble protein and no aggregation was observed, which is in clear contrast to IFABP G121V that forms inclusion bodies in *E. coli* (Kim and Frieden, 1998).

Moreover, we demonstrate that secreted IFABP(G121V) is capable of binding DAUDA with a binding affinity identical, within experimental error, to the protein produced from inclusion bodies (Kim and Frieden, 1998). This emphasizes the advantage of the T1SS, since a functional re-folding of proteins from inclusion bodies is often problematic and accounts for the major costs in the production of recombinant proteins in *E. coli* (Blight and Holland, 1994; Singh and Panda, 2005). Beneficially, the covalently linked HlyAc of the fusion protein has only a minor effect on the activity of IFABP(G121V), as indicated by a K_D value in the same range (see Table 2). In this study, wild-type IFABP could not be secreted as a fusion protein with HlyAc, indicating that the folding is too fast. In contrast, a single point mutation (G121V), which displays an approximately 100-fold reduced refolding rate (Chattopadhyay et al., 2002; Kim and Frieden, 1998), was secreted efficiently.

5. Conclusions

The results of this study further support our initial conclusion that the folding rate of a fusion protein dictates its secretion efficiency in T1SS. Not only bacterial proteins, targeted to the periplasmic space, can be designed to become a substrate of a T1SS, but also cytosolic, eukaryotic proteins. Furthermore, our data show that the described method results in the production of correctly folded, active protein in sufficient yields for biochemical studies. IFABP G121V is produced in inclusion bodies when expressed in

E. coli. The re-folding of denatured inclusion bodies, which accounts for the major costs in the production of recombinant proteins (Singh and Panda, 2005), can therefore be circumvented with this T1SS. In summary, our T1SS is a very attractive candidate for the biotechnological production of proteins via secretion into the extracellular space.

Acknowledgements

We acknowledge the assistance of Antonia Konczwald during the initial stages of the project. We thank Dr. Carl Frieden (Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, USA) for providing us with the genes IFABP WT and IFABP G121V. We thank the Ministry of Innovation, Science and Research of the German Federal State North Rhine-Westphalia (NRW) and the Heinrich Heine University Düsseldorf (scholarship from the CLIB Graduate Cluster Industrial Biotechnology to CKWS).

References

Armstrong, C.M., Goldberg, D.E., 2007. An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*. Nat. Methods 4, 1007–1009.
Bakkes, P.J., Jenewein, S., Smits, S.H., Holland, I.B., Schmitt, L., 2010. The rate of folding dictates substrate secretion by the *Escherichia coli* haemolysin type 1 secretion system. J. Biol. Chem. 285, 40573–40580.
Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., Wandless, T.J., 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell 126, 995–1004.
Benabdelhak, H., Kiontke, S., Horn, C., Ernst, R., Blight, M.A., Holland, I.B., Schmitt, L., 2003. A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. J. Mol. Biol. 327, 1169–1179.
Blight, M.A., Holland, I.B., 1994. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. Trends Biotechnol. 12, 450–455.
Chattopadhyay, K., Zhong, S., Yeh, S.R., Rousseau, D.L., Frieden, C., 2002. The intestinal fatty acid binding protein: the role of turns in fast and slow folding processes. Biochemistry 41, 4040–4047.
Chervaux, C., Sauvonnet, N., Le Clainche, A., Kenny, B., Hung, A.L., Broome-Smith, J.K., Holland, I.B., 1995. Secretion of active beta-lactamase to the medium mediated by the *Escherichia coli* haemolysin transport pathway. Mol. Gen. Genet. 249, 237–245.
Chuang, S., Velkov, T., Horne, J., Porter, C.J., Scanlon, M.J., 2008. Characterization of the drug binding specificity of rat liver fatty acid binding protein. J. Med. Chem. 51, 3755–3764.
Cover, W.H., Ryan, J.P., Bassford Jr., P.J., Walsh, K.A., Bollinger, J., Randall, L.L., 1987. Suppression of a signal sequence mutation by an amino acid substitution in the mature portion of the maltose-binding protein. J. Bacteriol. 169, 1794–1800.
Crameri, A., Raillard, S.A., Bermudez, E., Stemmer, W.P., 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391, 288–291.
Crameri, A., Whitehorn, E.A., Tate, E., Stemmer, W.P., 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat. Biotechnol. 14, 315–319.
Farinas, E.T., Bulter, T., Arnold, F.H., 2001. Directed enzyme evolution. Curr. Opin. Biotechnol. 12, 545–551.
Fernandez, L.A., Sola, I., Enjuanes, L., de Lorenzo, V., 2000. Specific secretion of active single-chain Fv antibodies into the supernatants of *Escherichia coli* cultures by use of the hemolysin system. Appl. Environ. Microbiol. 66, 5024–5029.
Gentschev, I., Dietrich, G., Goebel, W., 2002. The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. Trends Microbiol. 10, 39–45.
Glasel, J.A., 1995. Validity of nucleic acid purities monitored by 260 nm/280 nm absorbance ratios. Biotechniques 18, 62–63.
Gray, L., Mackman, N., Nicaud, J.M., Holland, I.B., 1986. The carboxy-terminal region of haemolysin 2001 is required for secretion of the toxin from *Escherichia coli*. Mol. Gen. Genet. 205, 127–133.
Hardy, S.J., Randall, L.L., 1991. A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. Science 251, 439–443.
Herm-Gotz, A., Agop-Nersesian, C., Munter, S., Grimley, J.S., Wandless, T.J., Frischknecht, F., Meissner, M., 2007. Rapid control of protein level in the apicomplexan *Toxoplasma gondii*. Nat. Methods 4, 1003–1005.
Hess, J., Gentschev, I., Goebel, W., Jarchau, T., 1990. Analysis of the haemolysin secretion system by PhoA-HlyA fusion proteins. Mol. Gen. Genet. 224, 201–208.
Hinsa, S.M., Espinosa-Urgel, M., Ramos, J.L., O'Toole, G.A., 2003. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. Mol. Microbiol. 49, 905–918.
Holland, I.B., Kenny, B., Steipe, B., Pluckthun, A., 1990. Secretion of heterologous proteins in *Escherichia coli*. Methods Enzymol. 182, 132–143.
Holland, I.B., Schmitt, L., Young, J., 2005. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway. Mol. Membr. Biol. 22, 29–39.
Jarchau, T., Chakraborty, T., Garcia, F., Goebel, W., 1994. Selection for transport competence of C-terminal polypeptides derived from *Escherichia coli* hemolysin: the shortest peptide capable of autonomous HlyB/HlyD-dependent secretion comprises the C-terminal 62 amino acids of HlyA. Mol. Gen. Genet. 245, 53–60.
Kenny, B., Chervaux, C., Holland, I.B., 1994. Evidence that residues −15 to −46 of the haemolysin secretion signal are involved in early steps in secretion, leading to recognition of the translocator. Mol. Microbiol. 11, 99–109.
Kenny, B., Haigh, R., Holland, I.B., 1991. Analysis of the haemolysin transport process through the secretion from *Escherichia coli* of PCM, CAT or beta-galactosidase fused to the Hly C-terminal signal domain. Mol. Microbiol. 5, 2557–2568.
Kim, K., Frieden, C., 1998. Turn scanning by site-directed mutagenesis: application to the protein folding problem using the intestinal fatty acid binding protein. Protein Sci. 7, 1821–1828.
Letoffe, S., Ghigo, J.M., Wandersman, C., 1994. Secretion of the *Serratia marcescens* HasA protein by an ABC transporter. J. Bacteriol. 176, 5372–5377.
Li, H., Frieden, C., 2007. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and ^{19}F NMR. Proc. Natl. Acad. Sci. U.S.A. 104, 11993–11998.
Linggood, M.A., Ingram, P.L., 1982. The role of alpha haemolysin in the virulence of *Escherichia coli* for mice. J. Med. Microbiol. 15, 23–30.
Mackman, N., Baker, K., Gray, L., Haigh, R., Nicaud, J.M., Holland, I.B., 1987. Release of a chimeric protein into the medium from *Escherichia coli* using the C-terminal secretion signal of haemolysin. EMBO J. 6, 2835–2841.
Meier, R., Drepper, T., Svensson, V., Jaeger, K.E., Baumann, U., 2007. A calcium-gated lid and a large beta-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. J. Biol. Chem. 282, 31477–31483.
Mergulhao, F.J., Summers, D.K., Monteiro, G.A., 2005. Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. 23, 177–202.
Park, E.C., Finley, D., Szostak, J.W., 1992. A strategy for the generation of conditional mutations by protein destabilization. Proc. Natl. Acad. Sci. U.S.A. 89, 1249–1252.
Rowland, A., Knights, K.M., Mackenzie, P.I., Miners, J.O., 2009. Characterization of the binding of drugs to human intestinal fatty acid binding protein (IFABP): potential role of IFABP as an alternative to albumin for in vitro–in vivo extrapolation of drug kinetic parameters. Drug Metab. Dispos. 37, 1395–1403.
Sanchez-Magraner, L., Viguera, A.R., Garcia-Pacios, M., Garcillan, M.P., Arrondo, J.L., de la Cruz, F., Goni, F.M., Ostolaza, H., 2007. The calcium-binding C-terminal domain of *Escherichia coli* alpha-hemolysin is a major determinant in the surface-active properties of the protein. J. Biol. Chem. 282, 11827–11835.
Sandkvist, M., Bagdasarian, M., 1996. Secretion of recombinant proteins by Gram-negative bacteria. Curr. Opin. Biotechnol. 7, 505–511.
Singh, S.M., Panda, A.K., 2005. Solubilization and refolding of bacterial inclusion body proteins. J. Biosci. Bioeng. 99, 303–310.
Stanley, P., Koronakis, V., Hughes, C., 1991. Mutational analysis supports a role for multiple structural features in the C-terminal secretion signal of *Escherichia coli* haemolysin. Mol. Microbiol. 5, 2391–2403.
Sweetser, D.A., Heuckeroth, R.O., Gordon, J.I., 1987. The metabolic significance of mammalian fatty-acid-binding proteins: abundant proteins in search of a function. Annu. Rev. Nutr. 7, 337–359.
Thanabalu, T., Koronakis, E., Hughes, C., Koronakis, V., 1998. Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. EMBO J. 17, 6487–6496.
Weiss, J.B., Ray, P.H., Bassford Jr., P.J., 1988. Purified secB protein of *Escherichia coli* retards folding and promotes membrane translocation of the maltose-binding protein in vitro. Proc. Natl. Acad. Sci. U.S.A. 85, 8978–8982.
Wilkinson, T.C., Wilton, D.C., 1986. Studies on fatty acid-binding proteins. The detection and quantification of the protein from rat liver by using a fluorescent fatty acid analogue. Biochem. J. 238, 419–424.
Yoon, S.H., Kim, S.K., Kim, J.F., 2010. Secretory production of recombinant proteins in *Escherichia coli*. Recent Pat. Biotechnol. 4, 23–29.

5.6 Chapter VI

Title:	Secretion of slow-folding proteins by a Type 1 secretion system.
Authors:	Schwarz C. K. W., Lenders M. H., Smits S. H. J., Schmitt L.
Published in:	Bioengineered
Impact Factor:	to be released in 2012
Own proportion of this work:	60 % All experimental data. Writing of the manuscript.

ARTICLE ADDENDUM

Bioengineered 3:5, 1-4; September/October 2012; © 2012 Landes Bioscience

Secretion of slow-folding proteins by a Type 1 secretion system

Christian K.W. Schwarz, Michael H.H. Lenders, Sander H.J. Smits and Lutz Schmitt*

Institute of Biochemistry; Heinrich-Heine-Universität Düsseldorf; Düsseldorf, Germany

This manuscript has been published online, prior to printing. Once the issue is complete and page numbers have been assigned, the citation will change accordingly.

Protein production through dedicated secretion systems might offer an potential alternative to the conventional cytoplasmical expression. The application of Type 1 secretion systems of Gram-negative bacteria, however, where often not successful in the past for a wide range of proteins. Recently, two studies using the *E. coli* maltose binding protein (MalE) and the rat intestinal fatty acid binding protein (IFABP) revealed a rational to circumvent these limitations. Here, wild-type passenger proteins were not secreted, while folding mutants with decreased folding kinetics were efficiently exported to the extracellular space. Subsequently, an one-step purification protocol yielded homogeneous and active protein. Taken together, theses two studies suggest that the introduction of slow-folding mutations into a protein sequence might be the key to use Type 1 secretion systems for the biotechnological production of proteins.

Keywords: type 1 secretion, HlyB, protein folding, mutation, relaxation, IFABP, maltose, western blot

Submitted: 03/27/12

Revised: 05/09/12

Accepted: 05/10/12

http://dx.doi.org/10.4161/bioe.20712

*Correspondence to: Lutz Schmitt; Email: lutz.schmitt@hhu.de

Addendum to: Schwarz CK, Landsberg CD, Lenders MH, Smits SH, Schmitt L. Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form. J Biotechnol 2012; 159:155–61; PMID:22366493; http://dx.doi.org/10.1016/j.jbiotec.2012.02.005.

Introduction

Type 1 secretion systems (T1SS) of Gram-negative bacteria transport substrates directly form the cytoplasm to the exterior without the formation of any periplasmic intermediate. The substrate spectrum contains functional unrelated and different-sized proteins, i.e., the hemophore HasA (19 kDa) and the lipase LipA (65 kDa) of *Serratia marcescens*[1,2] or the 900 kDa protein LapA that is involved in biofilm formation of *Pseudomonas fluorescence*.[3] A transport machinery composed of two inner membrane proteins, a membrane fusion protein (MFP) and an ATP-binding-cassette (ABC) transporter, as well as an outer membrane factor (OMF) catalyzes the secretion process.[4,5] The paradigm of T1SS is the hemolysin A (HlyA) secretion system of *Escherichia coli* (*E. coli*), which consists of the ABC transporter hemolysin B (HlyB), the MFP hemolysin D (HlyD) and the multi-functional TolC (OMF).[6] TolC is recruited, upon interaction of the substrate (HlyA) with HlyB and HlyD, and a continuous export channel is formed bridging the cytoplasm directly to the exterior.[4,7]

HlyA and other T1SS substrates are suggested to be transported in an unfolded state[8] and folding is triggered by binding of Ca^{2+} ions to so-called GG repeats in the extracellular space, directly after transport. Previously, several proteins were successfully secreted as fusion proteins with a 23 kDa C-terminal fragment of HlyA (HlyAc), for example β-lactamase or antibody fragments.[9-12] However, many proteins could not be secreted limiting the exploitation of the T1SS pathway for biotechnological purposes.[9] Recently, two publications investigated these limitations.[8,13] Here, the maltose binding protein (MalE) and the intestinal fatty acid binding protein (IFABP) could initially not be secreted when fused to HlyAc. However, the use of slow-folding proteins with decreased folding kinetics allowed the secretion into the culture medium.[8,13] Interestingly, the mutant IFABP(G121V), accumulates in *E. coli* as inclusion bodies, whereas secreted IFABP(G121V) was soluble and active. Taken together, both studies suggest a rational to secrete proteins of interest in good yields and in a soluble, functional state with an *E. coli* T1SS. However, the successful application requires proteins with slow-folding kinetics and the identification of such mutants might hamper the general applicability

of the approach. In other words, the general application of this approach requires screening of protein libraries that contain random mutations generated by, for example, error-prone PCR. This should be performed with suitable activity assays that allow a simple read-out to identifiy secreted and moreover, active proteins.

Secretion of a Functional Cytoplasmic Protein

The HlyA T1SS was investigated for its applicability to secrete an eukaryotic cytoplasmic protein, the rat intestinal fatty acid binding protein (IFABP).[15] IFABP is a 131 residue (15 kDa) cytoplasmical protein, which belongs to the family of fatty acid binding proteins in mammals and binds a single molecule of fatty acids.[16] The binding to lipophilic drugs, for example, is used for the prediction of drug pharmacokinetic parameters in vitro.[17,18] The slow-folding mutant IFBAP G121V still binds fatty acids with high affinity (K_D = 120 nM ± 25 nM) and is currently produced in *E. coli* out of inclusion bodies.[19]

Both, wild-type IFABP and the slow-folding variant IFABP(G121V), were cloned inside a secretion vector in front of HlyAc. In the background of the T1SS, wild-type IFABP(wt)-HlyAc was not secreted into the medium of the *E. coli* culture, whereas IFABP(G121V)-HlyAc was secreted in good yields of up to 6 mg per liter cell culture (corresponding to 2 mg/OD). The secreted fusion protein could even be visualized by Coommassie brilliant blue (CBB) staining in the supernatant without concentrating the sample. Interestingly, the secreted fusion protein was soluble and stable in the culture medium and during the subsequent purification steps, which displays an advantage compared with its currently expressed form being insoluble inclusion bodies.[19]

Bakkes et al. demonstrated the secretion of a slow-folding mutant of MalE. The passenger MalE with two introduced slow-folding mutations was secreted and shown to be active in binding to amylose, although the major fraction did not bind to the ligand.[14] The reason for this behavior, however, was not investigated. Therefore, Schwarz et al. investigated the folding state and the activity of secreted IFABP in more detail and compared it with published results. The folding state of secreted IFABP(G121V) was studied by intrinsic tryptophane fluorescence spectroscopy and compared with IFABP(G121V), which was purified out of cytosolic inclusion bodies and refolded.[20] These studies indicated that secreted IFABP(G121V) was folded comparable to the reference protein. Since the folding state of a protein does not directly imply that the protein is active, the activity of secreted IFABP(G121V) was analyzed. The fatty acid analog 11-(dansylamino) undecanoic acid (DAUDA) was used for titration experiments to determine the binding activity of IFABP(G121V). For the validation of the experimental setup, the binding affinity of the reference protein IFABP(G121V) (non-secreted control) was examined by fluorescence spectroscopy. The determined dissociation constant (K_D) of 126 nM ± 7 nM is in line with the published data (120 nM ± 25 nM).[19] The K_D value of secreted IFABP(G121V)-HlyAc was determined to be 195 nM ± 13 nM. This is in the same range as the K_D for the non-secreted control demonstrating that secreted IFABP G121V exhibits similar binding parameters.

Unfolding Domains—Useful Tools for Secretion?

The major bottleneck of this secretion technology is its dependency on slow-folding proteins. As an alternative route to directed evolution appoaches (see below), a destabilizing domain was analyzed to transfer instability to the fusion proteins. The destabilizing domain ddFKBP, an engineered version of the human FKBP12, was used, which was already applied successfully in eukaryotes to destabilize fused proteins,[21-24] however, no studies were performed in bacteria so far. Schwarz et al. investigated, whether the ddFKBP destabilizes fusion proteins in *E. coli* and if this domain allows the secretion of proteins via the T1SS without the introduction of slow-folding mutations. The ddFKBP was fused N-terminally to HlyAc, IFABP(wt)-HlyAc and IFABP(G121V)-HlyAc, respectively. All protein constructs were expressed as indicated by CBB staining and western blot analysis of whole cell extracts. Obviously, ddFKBP did not increase secretion levels of the fusion constructs. These results indicate that the intrinsic folding state of HlyAc, IFABP(wt) and IFABP(G121V) is not changed drastically and that the ddFKBP does not facilitate the secretion of a protein of interest with the T1SS. This raises the question whether the ddFKBP transmits instability to fused proteins or induces their degradation in bacteria?

Engineering of Proteins with Slow Folding Characteristics

The successful secretion of proteins via the HlyA T1SS depends on the availability of slow-folding mutants. Besides wild-type proteins with slow-folding kinetics, many proteins fold too quickly inside cells, which hampers the general applicability of T1SS. Up to now, the a priori prediction and alteration of folding properties is impossible with biochemical, computational or bioinformatical approaches. Nevertheless, directed evolution methods are available to generate proteins with desired characteristics, i.e., altered folding characteristics.[25,26] Basically, this can be performed within three relatively easy experimental steps (see **Fig. 1**).

Step 1 represents the random mutagenesis of the wild-type protein of interest (POI_*wt*). Various methods are available to mutate the gene, which are summarized in, for example Wong et al.[27] The mutagenesis products are inserted inside the secretion plasmid in front of the secretion signal (HlyAc) DNA sequence. This results in a library of mutated POI (POI_*mut*) fusion proteins. *E. coli* cells are transformed with the plasmid library together with the plasmid encoding the T1SS complex (**Step 2**) and obtained colonies are analyzed in secretion experiments (**Step 3**). Since the secretion signal is localized at the C-terminal end of the fusion protein, only full-length fusion proteins will be secreted; fusion proteins with artificial stop codons introduced via the error prone PCR reaction are not secreted and thereby automatically eliminated from the screen. Mutations that do not alter the folding kinetics will not be secreted either.

However, proteins with slow-folding kinetics can be detected, by SDS-PAGE analysis of the culture supernatants. Staining of the gels or immunochemical detection visualizes the secreted proteins and allows the densometrical quantification of secreted proteins. A great advantage is that the amount of secreted protein is inversely correlated with the folding rate of the protein. In other words, the protein secreted most efficiently will have the slowest folding kinetics. Furthermore, supernatant samples can be used immediately for enzyme-linked activity or immunosorbent assays (ELISA). If such an activity of the POI can be visualized, the secretion system enables high-throughput screening of a vast pool of mutants. For example, the lipolytic activity of secreted lipases can be visualized on agar plates, which contain the appropiate substrate of the lipase.[28] In summary, the technology described in Schwarz et al. represents a valuable and novel tool, usuable for various experimental setups, for example, the generation of folding mutants allowing secretion and subsequently the screening of secreted proteins with improved characteristics which are interesting within the biotechnological area.

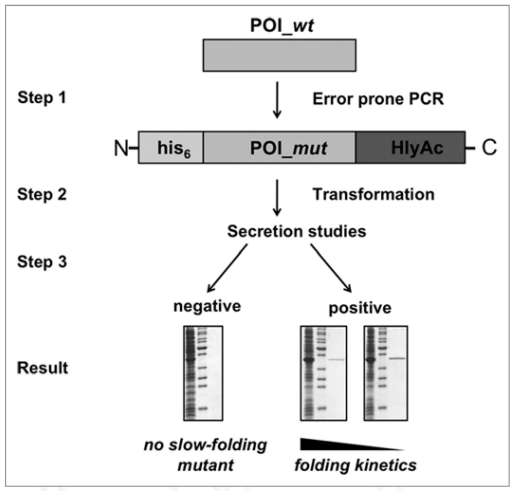


Figure 1. Identification of slow-folding mutants employing an *E. coli* T1SS. Scheme of the three-step protocol for the generation of slow-folding mutants of a protein of interest. Step 1: Random mutation of a POI-wt sequence and insertion of the resulting POI-mut in front of HlyAc. Step 2: Transformation of *E. coli* cells with the generated plasmid library in combination with the plasmid encoding the T1SS components. Step 3: Small-scale secretion studies to identify clones that secrete fusion proteins. The plasmids of secretion-positive clones are subsequently sequenced to identify the corresponding slow-folding mutations. If available, agar plate assays might be used to test the activity of these secreted proteins (see text).

Acknowledgements

We are indebted to Diana Clausnitzer, Protein Production Facility, of the Heinrich Heine University, for valuable discussion during the cloning of the constructs. We thank the Ministry of Innovation, Science and Research of the German Federal State North Rhine-Westphalia (NRW) and the Heinrich Heine University Düsseldorf (scholarship from the CLIB Graduate Cluster Industrial Biotechnology to CKWS and scholarship from the Biostruct NRW Graduate School to M.H.H.L.).

References

1. Meier R, Drepper T, Svensson V, Jaeger KE, Baumann U. A calcium-gated lid and a large beta-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. J Biol Chem 2007; 282:31477-83; PMID:17728256; http://dx.doi.org/10.1074/jbc.M704942200.
2. Létoffé S, Ghigo JM, Wandersman C. Secretion of the *Serratia marcescens* HasA protein by an ABC transporter. J Bacteriol 1994; 176:5372-7; PMID:8071214.
3. Hinsa SM, Espinosa-Urgel M, Ramos JL, O'Toole GA. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. Mol Microbiol 2003; 49:905-18; PMID:12890017; http://dx.doi.org/10.1046/j.1365-2958.2003.03615.x.
4. Thanabalu T, Koronakis E, Hughes C, Koronakis V. Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. EMBO J 1998; 17:6487-96; PMID:9822594; http://dx.doi.org/10.1093/emboj/17.22.6487.
5. Holland IB, Schmitt L, Young J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). Mol Membr Biol 2005; 22:29-39; PMID:16092522; http://dx.doi.org/10.1080/09687860500042013.
6. Jarchau T, Chakraborty T, Garcia F, Goebel W. Selection for transport competence of C-terminal polypeptides derived from *Escherichia coli* hemolysin: the shortest peptide capable of autonomous HlyB/HlyD-dependent secretion comprises the C-terminal 62 amino acids of HlyA. Mol Gen Genet 1994; 245:53-60; PMID:7531275; http://dx.doi.org/10.1007/BF00279750.
7. Benabdelhak H, Kiontke S, Horn C, Ernst R, Blight MA, Holland IB, et al. A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. J Mol Biol 2003; 327:1169-79; PMID:12662939; http://dx.doi.org/10.1016/S0022-2836(03)00204-3.
8. Bakkes PJ, Jenewein S, Smits SH, Holland IB, Schmitt L. The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system. J Biol Chem 2010; 285:40573-80; PMID:20971850; http://dx.doi.org/10.1074/jbc.M110.173658.
9. Blight MA, Holland IB. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. Trends Biotechnol 1994; 12:450-5; PMID:7765544; http://dx.doi.org/10.1016/0167-7799(94)90020-5.
10. Fernández LA, Sola I, Enjuanes L, de Lorenzo V. Specific secretion of active single-chain Fv antibodies into the supernatants of *Escherichia coli* cultures by use of the hemolysin system. Appl Environ Microbiol 2000; 66:5024-9; PMID:11055959; http://dx.doi.org/10.1128/AEM.66.11.5024-9.2000.
11. Chervaux C, Sauvonnet N, Le Clainche A, Kenny B, Hung AL, Broome-Smith JK, et al. Secretion of active beta-lactamase to the medium mediated by the *Escherichia coli* haemolysin transport pathway. Mol Gen Genet 1995; 249:237-45; PMID:7500946; http://dx.doi.org/10.1007/BF00290371.
12. Mackman N, Baker K, Gray L, Haigh R, Nicaud JM, Holland IB. Release of a chimeric protein into the medium from *Escherichia coli* using the C-terminal secretion signal of haemolysin. EMBO J 1987; 6:2835-41; PMID:3119330.
13. Schwarz CK, Landsberg CD, Lenders MH, Smits SH, Schmitt L. Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form. J Biotechnol 2012; 159:155-61; PMID:22366493; http://dx.doi.org/10.1016/j.jbiotec.2012.02.005.

14. Bakkes PJ, Jenewein S, Smits SH, Holland IB, Schmitt L. The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system. J Biol Chem 2010; 285:40573-80; PMID:20971850; http://dx.doi.org/10.1074/jbc.M110.173658.
15. Sweetser DA, Heuckeroth RO, Gordon JI. The metabolic significance of mammalian fatty-acid-binding proteins: abundant proteins in search of a function. Annu Rev Nutr 1987; 7:337-59; PMID:3300738; http://dx.doi.org/10.1146/annurev.nu.07.070187.002005.
16. Veerkamp JH, Peeters RA, Maatman RG. Structural and functional features of different types of cytoplasmic fatty acid-binding proteins. Biochim Biophys Acta 1991; 1081:1-24; PMID:1991151.
17. Chuang S, Velkov T, Horne J, Porter CJ, Scanlon MJ. Characterization of the drug binding specificity of rat liver fatty acid binding protein. J Med Chem 2008; 51:3755-64; PMID:18533710; http://dx.doi.org/10.1021/jm701192w.
18. Rowland A, Knights KM, Mackenzie PI, Miners JO. Characterization of the binding of drugs to human intestinal fatty acid binding protein (IFABP): potential role of IFABP as an alternative to albumin for in vitro-in vivo extrapolation of drug kinetic parameters. Drug Metab Dispos 2009; 37:1395-403; PMID:19398502; http://dx.doi.org/10.1124/dmd.109.027656.
19. Kim K, Frieden C. Turn scanning by site-directed mutagenesis: application to the protein folding problem using the intestinal fatty acid binding protein. Protein Sci 1998; 7:1821-8; PMID:10082380; http://dx.doi.org/10.1002/pro.5560070818.
20. Li H, Frieden C. Observation of sequential steps in the folding of intestinal fatty acid binding protein using a slow folding mutant and 19F NMR. Proc Natl Acad Sci USA 2007; 104:11993-8; PMID:17615232; http://dx.doi.org/10.1073/pnas.0705253104.
21. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible and tunable method to regulate protein function in living cells using synthetic small molecules. Cell 2006; 126:995-1004; PMID:16959577; http://dx.doi.org/10.1016/j.cell.2006.07.025.
22. Armstrong CM, Goldberg DE. An FKBP destabilization domain modulates protein levels in *Plasmodium falciparum*. Nat Methods 2007; 4:1007-9; PMID:17994030; http://dx.doi.org/10.1038/nmeth1132.
23. Herm-Götz A, Agop-Nersesian C, Münter S, Grimley JS, Wandless TJ, Frischknecht F, et al. Rapid control of protein level in the apicomplexan *Toxoplasma gondii*. Nat Methods 2007; 4:1003-5; PMID:17994029; http://dx.doi.org/10.1038/nmeth1134.
24. Park EC, Finley D, Szostak JW. A strategy for the generation of conditional mutations by protein destabilization. Proc Natl Acad Sci USA 1992; 89:1249-52; PMID:1311090; http://dx.doi.org/10.1073/pnas.89.4.1249.
25. Crameri A, Whitehorn EA, Tate E, Stemmer WP. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nat Biotechnol 1996; 14:315-9; PMID:9630892; http://dx.doi.org/10.1038/nbt0396-315.
26. Crameri A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 1998; 391:288-91; PMID:9440693; http://dx.doi.org/10.1038/34663.
27. Wong TS, Roccatano D, Zacharias M, Schwaneberg U. A statistical analysis of random mutagenesis methods used for directed protein evolution. J Mol Biol 2006; 355:858-71; PMID:16325201; http://dx.doi.org/10.1016/j.jmb.2005.10.082.
28. Eom GT, Song JK, Ahn JH, Seo YS, Rhee JS. Enhancement of the efficiency of secretion of heterologous lipase in *Escherichia coli* by directed evolution of the ABC transporter system. Appl Environ Microbiol 2005; 71:3468-74; PMID:16000750; http://dx.doi.org/10.1128/AEM.71.7.3468-74.2005.
29. Farinas ET, Bulter T, Arnold FH. Directed enzyme evolution. Curr Opin Biotechnol 2001; 12:545-51; PMID:11849936; http://dx.doi.org/10.1016/S0958-1669(01)00261-0.

©2010 Landes Bioscience. Do not distribute.

6 Discussion

Four major projects were the focus of this doctoral thesis and all were dealing with the HlyA T1SS. First, the transport process of HlyA by the dedicated T1SS was investigated, especially the function of the N-terminal domain of HlyB with unknown function. Second, the "slow-folding concept" for the production of proteins with the HlyA T1SS should be validated for a cytoplasmical protein with eukaryotic origin and its biological functionality should be determined. Third, the secretion system should be optimized to facilitate the secretory production of wild-type proteins without the introduction of slow-folding mutations. Fourth, the applicability of the HlyA T1SS for the recombinant production of peptides should be evaluated.

The obtained results of this doctoral thesis are summarized below. Based on these results in combination with published data, an advanced model for the mechanism of T1SS is suggested: the activating alternating access model.

6.1 Characterization of the HlyA T1SS of *E. coli*

6.1.1 The CLD is pivotal for the secretion of HlyA

Despite accounting for about 20 % of the entire primary sequence of HlyB, the function of the N-terminal domain of HlyB was unknown. Sequence alignments indicated a homology to C39 peptidases, but some indications excluded a potential proteolytic activity of the CLD. The catalytical essential cysteine of C39 peptidases is absent in the HlyB-CLD (Figure 6-1) and HlyA is not proteolytically processed during transport.

```
ComA        10 PQVDQMDCGVASLAMVFGYYGSYYFLAHLRELAK-TTMDGTTA--LGLVKVAEEIGFETR
HlyB        3  -SCHKIDYGLYALEILAQYHNVSVNPEEIKHRFD-TDGTGLGL--TSWLLAAKSLELKVK

ComA        AIKA---DMTLFDLPDLTFPFVAHVLKEGKLLHYYVVTG---QDKDSIHIADPDPGVKLT
HlyB        QVKK---TID--RLNFISLPALVW---REDGRHFILTKVS--KEANRYLIFDLEQ-RNPR

ComA        KLPRERFEEEWTGVTLFMAPSPDY-KPHKE 149
HlyB        VLEQSEFEALYQGHIILIASRSSV-AGK-L 135
```

Figure 6-1: Sequence alignment of the C39 peptidase of ComA and the CLD of HlyB. The ABC transporter ComA of *Streptococcus pneumoniae* is involved in the secretion of the quorum-sensing signal ComC and carries a C39 peptidase (ComA-PEP). The amino acids of the catalytic triad (Cys17, His96 and Asp112) are highlighted in red. The sequence alignment indicates the absence of the catalytic cysteine in the CLD, whereas the catalytic histidine and aspartate are present. Two tryptophans (W) are present in the CLD of HlyB (W56 and W77) and absent in the C39 peptidase of the ABC transporter (blue). These tryptophans represent a characteristic feature of all CLDs.

Moreover, HlyA does not contain the conserved binding motif for C39 peptidases, the double-glycine motif of bacteriocins [163].

To gain insights into its function, the CLD was analyzed in great detail by using biochemical and biophysical techniques. First, an overexpression and purification protocol was established for the production of the pure and homogeneous CLD in sufficient amounts for structural and biochemical approaches. Many technical and methodical aspects of these protocols were included in a book chapter for the production of pure protein for structural purposes (Chapter IV). Subsequently, the essential role of the CLD for the secretion of HlyA was demonstrated by secretion analysis with HlyA in combination with wild-type HlyB, HlyB lacking the CLD or HlyB carrying selected mutations within the CLD. The function of the CLD was further characterized by various biochemical techniques, for example pull-down assays (Chapter I) or SPR spectroscopy (Figure 6-2). The SPR spectroscopy analysis revealed an interaction between the C-terminal fragment of HlyA (HlyA1) and the isolated CLD with a K_D value of about 5 µM. Further experiments are currently ongoing to validate the data.

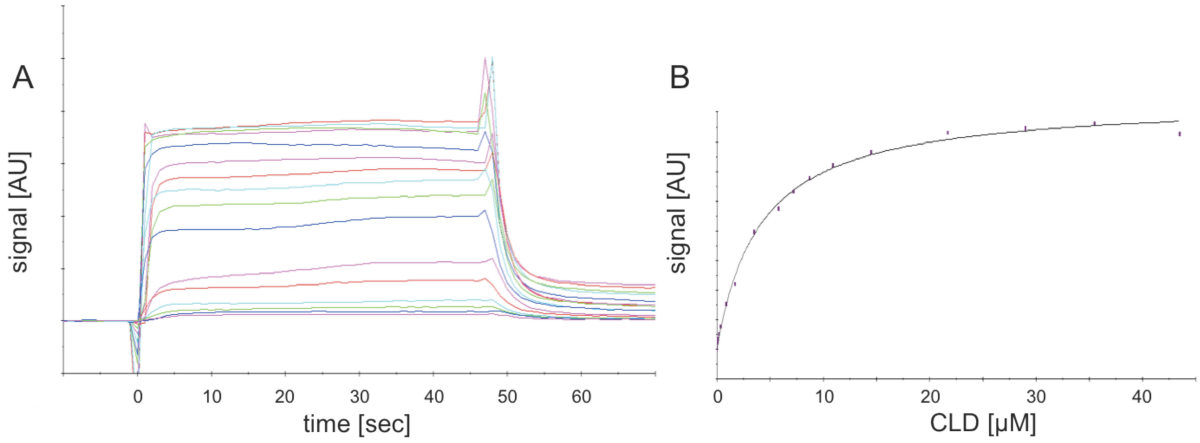


Figure 6-2: SPR spectroscopy analysis of the interaction between the C-terminal fragment of HlyA (HlyA1) and the isolated CLD. A: Sensorgram showing the interaction of HlyA1 and the isolated CLD. HlyA1 was immobilized via a his-tag and the CLD was injected at various concentrations. The interaction proceeded for 45 sec before the flow cell was washed with buffer **B:** Steady-state signals (averaged at 30–40 sec) were plotted against the CLD concentrations. The K_D value for the binding between HlyA1 and the CLD was determined to be ≈ 5 µM. Further SPR spectroscopy experiments are currently ongoing to validate the obtained data. AU: arbitrary units. Figures were made and data were analyzed with the software BIAevaluation 4.1 (GE Healthcare).

X-ray crystallography and NMR spectroscopy were applied for the determination of the high-resolution structure of the isolated CLD. Despite the growth of well-diffracting

crystals of the CLD (Chapter II), the X-ray structure could not be solved since the phases of the diffraction pattern could not be extracted. Molecular replacement using the ComA-PEP structure (PDB code 3K8U) as template did not work and no solution was found. Moreover, the incorporation of heavy atoms failed since no well-diffracting crystals could be obtained.

Instead, the structure was solved in cooperation with the "Institute of Complex Systems" (ICS-6) of Prof. Dr. Dieter Willbold in the "Forschungszentrum Jülich" by NMR spectroscopy (Chapter I and III). The combination of biochemical experiments and NMR spectroscopy enabled the determination of the precise binding region of the CLD for its binding partner, the unfolded RTX domain of HlyA (Chapter I).

For the first time, a structural and biochemical characterization of a C39 peptidase-like domain was performed. The CLD binds the unfolded RTX domain of HlyA prior to and/or during the secretion, which protects the secretion-competent HlyA inside the cell from being aggregated and/or degradaded. This mechanism seems to have evolved for the secretion of all RTX proteins and might explain why no external chaperones are identified being involved in the secretion of HlyA [177]. The RTX domain of HlyA is located N-terminal to and synthesized prior to the secretion signal. Therefore, the interaction of the RTX domain with the CLD likely occurs co-translationally and might recruit TolC to pre-settle the entire secretion complex (Figure 6-3). Since functional HlyB is suggested to be a homodimer, two CLDs are present per T1SS machinery. Thus, two (or more) binding sites within the RTX domain for the CLD are likely to be present in HlyA. An assembly of the entire HlyA T1SS independently of the secretion signal was also suggested by Masi *et al*., 2010 [129]. Such a coordinated assembly was also demonstrated for the HasA T1SS, where the substrate HasA pre-settles the T1SS machinery before the secretion signal initiates the transport process [129].

Sequence analysis were used to suggest a new classification system for ABC-dependent protein transport in Gram-negative bacteria. A first class, the RTX class, is characterized by ABC transporters with CLDs and transport substrates with RTX domains (Figure 6-4). The CLD binds the unfolded RTX protein via the RTX domain, which protects the secretion-competent state prior to secretion in the cell. The presence of two tryptophans in the primary sequence characterizes CLDs (W56 and W77 in the CLD of HlyB) and distinguishes this class from the second class, the bacteriocin class. Bacteriocin ABC transporters carry C39 peptidases that cleave the dedicated substrates, the bacteriocins,

during the transport process. This proteolytic processing matures the pro-peptide to the bioactive bacteriocin.

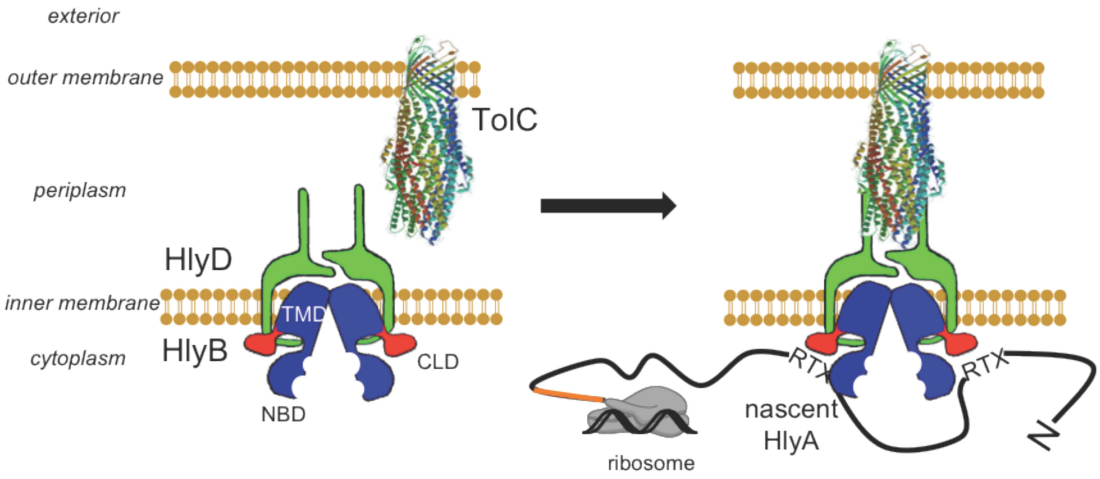


Figure 6-3: The unfolded RTX domain of HlyA interacts with the CLD of HlyB. Dimeric HlyB and oligomeric HlyD form a stable inner membrane complex and TolC is located in the outer membrane. Ribosomally synthesized HlyA interacts with the CLDs of HlyB via the unfolded RTX domain, presumably co-translationally, and recruits TolC to assemble the complete T1SS. RTX: RTX domain; N: N-terminus.

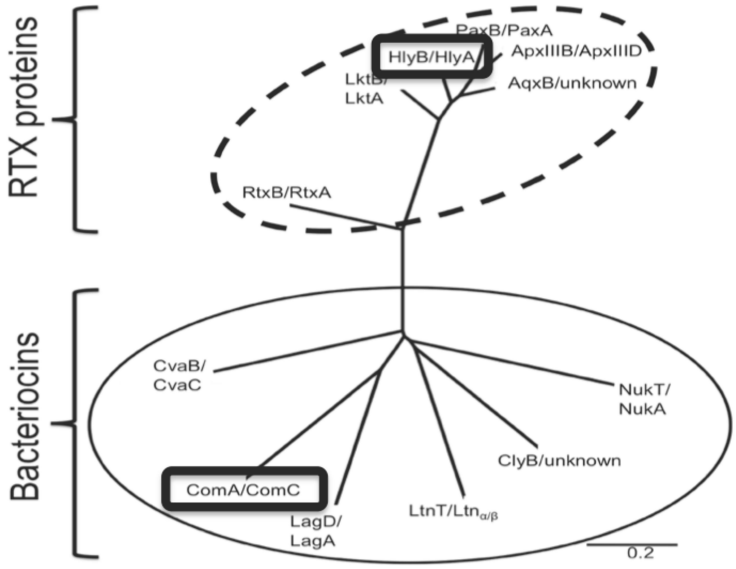


Figure 6-4: Phylogenetic tree of ABC transporters with C39 peptidases and CLDs. ABC transporters with C39 peptidases are shown in the continuous ellipse, ABC transporters with CLDs are encircled with a dashed ellipse. The bacteriocin transporter ComA and the RTX protein transporter HlyB with the dedicated substrates ComC and HlyA, respectively, are encircled in black. All ABC transporters with CLDs transport substrates of the RTX toxin family, whereas ABC transporters with C39 peptidases transport bacteriocins. The dedicated ABC transporter-substrate pairs are: ApxIIIB (UniProtKB entry Q04473) and ApxIIID, NukT (UniProtKB entry Q75V16) and Nukacin, CvaB (UniProtKB entry P22520) and Colicin V, ClyB (UniProtKB entry Q57403) and unknown, LagD (UniProtKB entry P59852) and LagA, LtnT (UniProtKB entry O87239) and cytolysin B, ComA (UniProtKB entry P59653) and ComC, HlyB (UniProtKB entry Q1R2T6) and HlyA, LktB (UniProtKB entry P23702) and LktA, PaxB (UniProtKB entry Q9RCG7) and PaxA, RtxB (UniProtKB entry A1YKX0) and RtxA, AqxB (UniProtKB entry Q8KWZ8) and unknown.

At least one third class comprises ABC transporters without extra domains in addition to the canonical NBDs and TMDs. Here, the substrates are protected by external chaperones and are not processed during the transport. The HasA T1SS of *Serratia marcescens* is a prominent member of this group [132]. This class is not included in Figure 3-1.

Interestingly, *in vitro* studies with the bacteriocin transporter NukT demonstrated that its C39 peptidase is proteolytically active in the presence of ATP, whereas the substitution of ATP with ATP-γ-S (a non-hydrolyzable ATP analogue) completely abolished its activity. The same results were obtained, if catalytical essential amino acids of the NBDs in the Walker A or Walker B motifs were mutated [196,197]. Therefore, ATP binding and hydrolysis is essential for the proteolytic activity of the C39 peptidase of NukT. Whether the ATPase activity of RTX transporters also influences the binding of the CLDs to the substrate is currently unknown.

6.1.2 Function of the C-terminal secretion signal

Several high-resolution X-ray structures of full-length ABC transporters in various catalytic stages have been solved since 2006 [35,42-44,53,59,198-205]. These structural data suggest a mechanism of the transport process [42-45], which is in line with the alternating access model [9] (see section 3.2). A substrate interacts with the ABC transporter, which induces the ATP-dependent dimerization of the NBDs. The conformational changes are transmitted to the TMDs via the coupling helices, which converts the orientation of the substrate binding pocket from the inward- to the outward-facing conformation. Subsequently, the substrate is released on the opposite site of the membrane. If one or two ATP molecules are hydrolyzed per catalytic cycle to energize the transport is not precisely known and might depend on the transport system (3.2.1). For clarity, however, the hydrolysis of two ATP molecules per catalytic cycle is suggested from now on. ATP binding represents the power-stroke of transport and ATP hydrolysis induces the dissociation of the NBD dimer to reset the transporter. All available structural information of full-length ABC transporters were obtained for transport systems that are involved in the transport of rather small substrates, in example drugs [43,59,200,203], nutrients [35,44,198,199], lipids [204] or vitamins [201,202,205], and no structure has been solved for an ABC transporter involved in protein transport. The mechanism of ABC-dependent protein transport is currently barely understood, but as members of the ABC

transporter family, a similar mechanism is to be expected. In the next sections, experimental data of this doctoral thesis and published results of ABC-dependent transport are summarized and the implications on the transport mechanism are considered.

6.1.2.1 The secretion signal represents the substrate for HlyB

The hemophore HasA (188 amino acids, 19 kDa) carries a characteristical C-terminal secretion signal of T1SS substrates and is secreted in *E. coli* by a dedicated system consisting of the ABC transporter HasD, the MFP HasE and the OMP TolC. Secretion studies showed that the last C-terminal 56 amino acids of HasA (HasAC) were sufficient for secretion, however, HasA lacking the C-terminal 14 amino acids (HasA(1-174)) failed to be secreted [206,207]. HasA(1-174) still induces the association of the entire secretion complex, however, the complex is trapped and not able to dissociate. A mutation in the ABC transporter HasD, which abolishes ATP hydrolysis, causes the same effect and prevents the dissociation of the assembled secretion complex [129]. These results suggest an impact of the last 14 C-terminal amino acids of HasA on the catalytic cycle of the NBDs and a requirement of the catalytic cycle for the dissociation of the secretion complex. If both, HasAC and HasA(1-174), were co-expressed, none of them were secreted, but the secretion complex was able to dissociate [207]. Therefore, the C-terminus of HasA promotes the disassembly of the secretion complex, obviously by effecting the catalytic cycle of the NBDs. Previous studies with the related metalloprotease T1SS of *Erwinia chrysanthemi* suggested an inhibitory effect of the secretion signal for the ATPase activity of the dedicated ABC transporter PrtD [208].

During this doctoral research and in cooperation with Dr. Britta Ries [209], an overexpression and purification protocol was established to produce purified detergent-stabilized HlyB in sufficient amounts for biochemical experiments [209]. Usually, the procedure yields 5 mg pure HlyB from a 4 L bacterial culture. Based on these protocols, the ATPase activity of the ABC transporter HlyB could be investigated *in vitro* [209]. Purified HlyB alone displays a basal ATPase activity, which, in contrast to PrtD [208], was stimulated about fivefold by the addition of the substrate HlyA. A C-terminal fragment of HlyA, HlyA1, also stimulated the ATPase activity of HlyB about threefold. Importantly, this stimulation was strictly dependent on the presence of the C-terminal secretion signal of HlyA [209]. These data confirm previous studies that demonstrated an interaction

of the secretion signal of HlyA with the isolated NBDs of HlyB [128]. Stimulation of the ATPase activity by transport substrates were also demonstrated for other ABC transporters, for example TAP, a peptide transporter [210], the multi-drug transporter P-glycoprotein (P-gp) [62,63,211] or $MalFGK_2$, where maltose, bound to the substrate binding protein MalE, stimulated the ATPase activity [212].

Another study demonstrated the autonomous secretion of the last C-terminal 62 amino acids of HlyA encoding the secretion signal [131]. Therefore, the secretion signal of HlyA contains all information necessary and required for secretion. The secretion signal interacts with the NBDs, stimulates the ATPase activity of the NBDs and represents the substrate for HlyB. Whether the interaction of the secretion signal with HlyB occurs solely with the NBDs or also with the TMDs is not known. However, the last 62 amino acids of HlyA are autonomously secreted. Since TMDs build the conduct of ABC transporters through the membrane and substrates of ABC transporters typically interact with the TMDs [34,60-67,213], (an) additional interaction site(s) for the secretion signal at the TMDs of HlyB is likely.

The precise physiologic effect of the ATPase stimulation by the secretion signal for T1SS transport is not known, however, two scenarios with fundamental different mechanisms are possible. Either the stimulation of the secretion signal via the NBDs induces the dimerization of the NBDs, which converts the inward- to the outward-facing conformation of HlyB and initiates secretion (activating model) or the stimulation induces ATP hydrolysis, release of the substrate and the dissociation of the secretion complex (terminating model).

6.1.2.2 The activating model

In the activating model, the interaction of the secretion signal with the NBDs stimulates their dimerization, which initiates the secretion process. The inward-facing state is converted to the outward-facing state and the substrate (the secretion signal) is released from HlyB and inserted into the transport channel. This mechanism implies a C-terminal secretion of the substrate (secretion with the C-terminus ahead). The substrate reaches the exterior and the release of the substrate from the transport channel is transmitted to or recognized by the transporter. This release induces the dissociation of the NBDs and concomitantly the disassembly of the secretion complex to reset the system.

6.1.2.3 The terminating model

In the terminating model, the secretion signal terminates the transport process. After the initiation of the transport, the N-terminus of the substrate is transported ahead (N-terminal secretion) and the secretion signal remains bound to the ABC transporter during transport. Finally, the release of the substrate, accompanied with the dissociation of the secretion signal from the transporter, induces ATP hydrolysis, dissociation of the NBD dimer and re-setting of the T1SS machinery.

6.1.2.4 Does the secretion signal activate or terminate the transport?

The fundamental difference between the activating and the terminating model is the orientation of the protein substrate engaged inside the transport complex during transport. In the activating model, the protein substrate is secreted with its C-terminus ahead, whereas the terminating model assumes an N-terminal secretion.

Currently, no structural data for ABC transporters involved in the transport of proteins are available and, concomitantly, the orientation of the incorporated substrate is unknown. However, various experimental data suggest that the protein substrate of T1SS activates the transport process:

First, the secretion of chimeric proteins with the HlyA T1SS is well-known [91,122,124,176,177,180,214,215]. Here, arbitrary passenger proteins are fused N-terminal to fragments of HlyA, which contain at least the C-terminal secretion signal of HlyA. The the arbitrary N-terminus of the fusion protein cannot mediate any specificity for the HlyA T1SS. Consequently, the secretion process starts after the synthesis of the secretion signal and therefore post-translationally. According to the terminating model, the N-terminus of the fusion protein is secreted ahead. A denominator for the specific insertion of arbitrary N-termini is absent. Moreover, it is rather unlikely that an arbitrary N-terminal domain persists at the transporter and can be inserted into the secretion complex after the secretion signal is synthesized. In the activating model, in contrast, the C-terminal secretion signal represents a common denominator of each fusion protein that can be recognized and inserted into the transport complex by the ABC transporter and/or the MFP to initiate the transport.

Secondly, initial results indicate that the entire HlyA T1SS complex might be trapped with a fusion protein of MalE and a C-terminal fragment of HlyA, HlyA1, that contains parts of the RTX domain and the secretion signal (unpublished data). This fusion protein

cannot be secreted with the HlyA T1SS and previous studies demonstrated that the folding of MalE prohibits the secretion, presumably by blocking the secretion channel [124]. Nevertheless, the T1SS complex assembles and can be stably purified *in vitro*. Therefore, HlyA1 stabilizes the complex, presumably by being incorporated in the secretion channel, while folded MalE represents a plug at the entrance of the complex. This result indicates a C-terminal insertion of the fusion protein and supports the activating model.

Thirdly, HasA was shown to interact with the ABC transporter HasD via multiple regions that are scattered throughout the entire primary sequence [129]. If the secretion occurs N-terminally, one or a few interaction sites at the N-terminus of HasA would be sufficient for its insertion into the transport complex and additional interaction sites for HasD would even hinder the passage by the repetitive interactions between HasA and HasD. Therefore, multiple scattered interaction sites suggest a pre-settling of the substrate at the transporter to assemble the complete T1SS machinery before the secretion signal initiaties the secretion.

Fourthly, previous studies demonstrated that the HlyA secretion signal interacts with the isolated NBDs of HlyB, thus being a substrate of the ABC transporter [128]. In line with many other studies, the substrate should induce the ATP-dependent dimerization of the NBDs and thereby initiate the transport process [34,35,60-65]. The dimerization converts the inward- to the outward-facing conformation of the transporter and the substrate is released from the transporter. As ATP hydrolysis occurs autocatalytically at the dimeric NBDs [25,34,37], the complete catalytic cycle of the NBDs is possible: substrate binding, NBD dimerization, release of the substrate, ATP hydrolysis, NBD dissociation. Therefore, the substrate of an ABC transporter increases ATPase activity, which was previously demonstrated [34,60-65,210]. In the terminating model, in contrast, the secretion signal has to be adjacent to the transport complex until the end of the transport. The release of the substrate terminates the transport process, which would, for example, pull the secretion signal away from the transport machinery and trigger ATP hydrolysis and dissociation of the NBDs. The trigger of the dissociation – pulling or another signal - is absent during *in vitro* ATPase experiments with isolated ABC transporters. The substrate/secretion signal should bind to the NBDs, comparable to the expected *in vivo* situation during the transport. This binding actually would trap the catalytic cycle of the NBDs, at least partially, in the substrate- and ATP-bound state. Therefore, the addition of substrates to isolated ABC transporters should inhibit rather than increase the ATPase activity. Such

an inhibition was demonstrated for the ABC transporter PrtD of the metalloprotease T1SS of *Erwinia chrysanthemi* [208]. The terminating model also assumes a continous interaction of the secretion signal with the ABC transporter during the transport. Usually, substrates bind to the inward-facing state of ABC transporters, which converts the transporter to the outward-facing state [42-45]. Subsequently, the secretion signal should dissociate from the ABC transporter. Previous studies demonstrated an interaction of HlyA1 with the isolated NBDs of HlyB [128]. Importantly, this interaction was dependend on the presence of the C-terminal secretion signal and was disrupted by the addition of ATP .

The ATPase activity of the purified ABC transporter HlyB supports the activating model (6.1.2.1). Here, the expected stimulation of the ATPase activity by the secretion signal was shown [209]. Similar results were obtained with the purified ABC transporter TAP, a peptide exporter, which shows a strict coupling of the ATPase activity to the presence of the substrate [210].

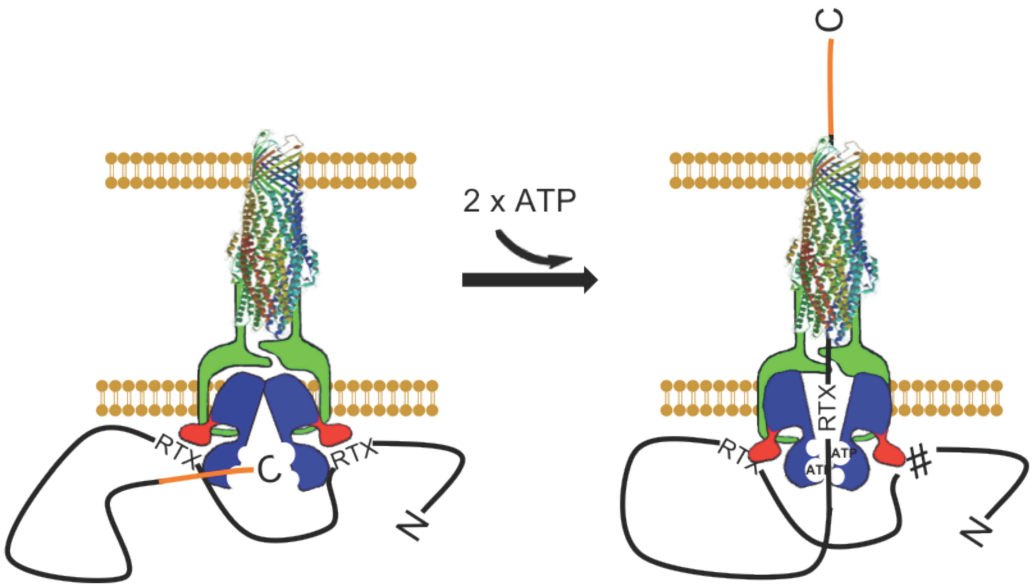


Figure 6-5: Activating model of the HlyA secretion process. After its synthesis, the C-terminal secretion signal of HlyA (orange) interacts with the NBDs of HlyB (blue), which forces the ATP-dependent dimerization of the NBDs. This converts the inward-facing conformation of HlyB to the outward-facing state, opens a conduct-like structure of HlyB and enables the insertion of the secretion signal into the transport complex. ATP hydrolysis occurs autocatalytically and the hydrolysis products ADP and P_i are formed. The stoichiometry of transported substrate to hydrolyzed ATP is unknown and repeated catalytic cycles of the NBDs might occur during transport, however, the secretion signal is the only known stimulus for the ATPase activity of HlyB. Whether HlyA contains additional interaction sites for the inner membrane complex N-terminal to the RTX domain is unknown, however, further unknown interaction sites of HlyA (#) for the inner membrane complex are suggested in respect to studies with the HasA T1SS [129]. Here, the substrate HasA was shown to carry interaction sites for the ABC transporter HasD, which are scattered across the entire primary sequence of HasA. C: C-terminus; green: HlyD N: N-terminus; RTX: RTX domain; rainbow: TolC (PDB code 1EK9); #: unknown, hypothetical sequence motif.

Fifthly, the secretion signal of HlyA alone is autonomously secreted via the T1SS [131]. Therefore, the C-terminal amino acids provide all informations for the secretion, including the activation.

Taken together, these data support the transport of T1SS substrates by the activating model. The interaction of the CLD with the unfolded RTX domain of HlyA presumably occurs co-translationally, which might assemble the T1SS complex before the secretion signal inititates the transport (Figure 6-5). The substrate passes the translocon with its C-terminus first and the release of the substrate induces re-setting of the transport machinery. Whether the secretion signal of HlyA is the only motif that stimulates the ATPase activity of HlyB is unknown, however, no other data are available and the current data support this view. Since the secretion signal is the only known substrate, it remains elusive, how iterative catalytic cylces of the NBDs might be stimulated to energize the transport of HlyA. This raises the question how the transport of T1SS is energized.

6.1.3 Energization of T1SS transport

According to the alternating access model, a substrate binds to the inward-facing conformation of the ABC transporter. Subsequently, the ABC transporter switches to the outward-facing conformation and the substrate is released on the opposite site of the membrane. The conversion of the outward- to the inward-facing state resets the transporter. Two ATP molecules are hydrolyzed during one catalytic cycle (see section 3.2.1). T1SS facilitate the secretion of proteins with masses up to 900 kDa in one step across the Gram-negative cell envelope of bacteria [81] and it seems unlikely that two ATP molecules are sufficient to energize this process. The Sec pathway of bacteria, for example, transports proteins across one membrane, the plasma membrane, and consumes one ATP molecule per 20– 30 transported amino acids [216,217]. This transport system, however, is not driven by ABC transporters. Whether iterative catalytic cycles of the NBDs energize the ABC-driven transport of T1SS substrates is not known. In this model, however, the substrate is partially located in the translocon during the repetitive ATP hydrolysis cycles. Consequently, the substrate occupies parts of the transporter during transport in a conduct-like structure. This represents a difference to the known alternating access model.

ABC transporters are primary active transporters that use ATP binding and hydrolysis to energize the transport process, but other energy sources were also suggested to be involved for the secretion of HlyA. For example, an involvement of the PMF in early stages of the HlyA secretion process was shown [130]. An energetically favourable transfer of HlyA requiring no additional energy was also suggested [130], but the authors of this study did not mention a potential driving force.

One catalytic cycle of the NBDs is induced by the substrate of the ABC transporter, for HlyB by the secretion signal, which is energized by the hydrolysis of two ATP molecules. The next step of the transport could be theoretically driven without external energy. Inside cells, the synthesis of HlyA increases its concentration, whereas the HlyA concentration outside of the cell is low at the onset of the secretion. This situation would enable diffusion across the cell wall through a conduct to the exterior. HlyB persists in the inward-facing conformation and the secretion signal of HlyA binds to the NBDs (and presumably to the TMDs). The interaction converts the inward- to the outward facing site and the conduct-like structure of HlyB is opened. Subsequently, the secretion signal is released and diffuses through this conduct and the entire T1SS translocon to the exterior.

The directionality of ABC-driven transport is conferred by the preferential release of the substrate at the outward-facing site with a reduced binding affinity for the substrate compared to the inward-facing site. This model also enables transport against a concentration gradient. According to this model, high concentrations of the substrate at the outward-facing site of the transporter should disturb or inhibit transport, because the substrate cannot dissociate from the outward-facing state. Exactly such an inhibition of transport was shown for the ABC transporter P-gp. Moreover, the high concentration of substrate on the opposite was suggested to inhibit the dissociation of transported substrate from the outward-facing site [218]. Studies with the drug resistance ABC transporter Pdr5 of the yeast *Saccaromyces cerevisiae* indicated similar results. Here, the substrates inhibited the ATPase activity of the transporter, most likely "...due to a substrate-mediated lock of the outward facing conformation" [219]. Previous studies even demonstrated the reversed transport of ethidium by the ABC transporter LmrA of *Lactococcus lactis*, a prokaryotic homolog of P-gp. Remarkably, the reversed transport was associated with ATP synthesis [220]. Whether the transport direction of ABC transporters can be reversed, however, is controversely discussed [220-222].

These mentioned studies indicate that ABC transporters must not operate in an unidirectional fashion. The observed physiological unidirectionality of ABC transporters is generated by a kinetic inhibition at physiological conditions rather than by a mechanical irreversibility [220]. In the cytoplasm, the concentration of ATP (typically 1– 10 mM [223]) and ADP (0.1– 0.3 mM [224]) favors ATP hydrolysis, which determines the direction of transport; export of substrates by ABC exporters and import of substrates by ABC importers [220]. Similar results were obtained for the Na^+/K^+-ATPase that couple the extrusion of three Na^+ ions to the uptake of two K^+ ions. This membrane protein belongs to the P-type ATPase superfamily, in which phosphorylation and dephosphorylation of the transporter is crucial for transport [225]. Nevertheless, the transport direction of this ATPase can be inverted [226,227]. Studies with the F_0F_1-ATPase, which couples downhill transport of protons with ATP synthesis in the oxidative phosphorylation, also demonstrated the reversibility of the catalyzed reaction. The synthesis reaction can be reversed and ATP hydrolysis energizes the transport of protons against a gradient [228,229] These studies suggest that ATP-driven transport is not unidirectional, however, the transport is directional under phyiological conditions.

The mentioned data support a model, in which ABC transporters represent a shuttle that enables the directed transport of substrates. ABC transporters posses the inward-facing conformation until the substrate induces the conversion to the outward-facing conformation. Subsequently, the substrate is released and diffuses to the exterior, for protein substrates through a conduct-like structure of the ABC transporter. The conversion results from the substrate-induced ATP-dependent dimerization of the NBDs, which represents the power-stroke of transport. ATP hydrolysis is necessary to disrupt the NBD dimer and to restore the inward-facing conformation of the transporter. The reverse transport is prevented by a reduced binding affinity of the substrate to the outward-facing state, the conversion of the outward- to the inward-facing state of the transporter after the release of the substrate and the physiological concentrations of ATP and ADP. For the secretion of HlyA and other RTX proteins, the Ca^{2+}-induced folding further prevents the reversed transport since folded proteins do physically not fit into the secretion complex. The Ca^{2+}-induced folding of RTX proteins will be discussed in more detail below.

According to the described model, ABC transporters are able to catalyze the transport of a single substrate without ATP hydrolysis. Substrate binding to the transporter induces

ATP-dependent dimerization of the NBDs, the conversion of the inward- to the outward-facing conformation and the release of the substrate. Subsequently, the transporter is trapped in the outward-facing conformation, which prevents further transport. This trapping was observed in the HasA T1SS [129]. Here, a catalytical essential glutamate for the ATPase activity of the ABC transporter HasD was mutated, which trapped the assembled T1SS complex. Whether traces of the substrate HasA were secreted in the supernatant of the culture was not shown in the study. Moreover, it is not known, if the small amounts of secreted proteins are detectable (one substrate per one ABC transporter). Therefore, it is currently unknown if ABC transporters are competent to catalyze the transport of a single substrate without ATP hydrolysis. Comparable studies with the HlyA T1SS should be performed.

For the extremely large RTX proteins [117,230] the suggested mechanism might not facilitate an efficient secretion and further action is required for the transport. Here, the RTX domain of HlyA and other RTX proteins might catalyze the directed and efficient transport. RTX domains are characterized by the presence of GG repeats. They are located at the C-terminal part of RTX proteins and N-terminal to the secretion signal [114-118] (Figure 3-8). According to the activating model (see section 6.1.2.2), the C-terminal secretion signal of HlyA is inserted into the secretion machinery and the N-terminal RTX domain follows during transport. Being unfolded inside the cytoplasm, folding of HlyA is induced at the exterior by binding of Ca^{2+} ions to the GG repeats. This folding confers directionality to the transport as the reverse transport of the folded protein is prevented. Moreover, folding represents a potential driving force to pull the protein out of the transport channel into the surrounding. A similar mechanism was also suggested by Linhartova *et al*. in 2010 [117]. The pulling mechanism is supported by various experimental data.

First, the structures of RTX proteins clearly show that Ca^{2+} ions are bound to the GG repeats [114,231]. Biochemical data demonstrated that this binding induces the folding of HlyA [232] and other RTX proteins, for example CyaA of *Bordetella pertussis* [120]. Moreover, increasing Ca^{2+} concentrations in the growth medium increase the secretion levels of HlyA [233]. Similar results were obtained during this doctoral thesis for the C-terminal fragment of HlyA, HlyA1. Addition of Ca^{2+} to the growth medium increased the secretion levels (Figure 6-6). Whereas the addition of 0.05 mM Ca^{2+} did visually not influence the secretion levels, the addition of 0.15 mM Ca^{2+} increased the secretion levels. The maximal secretion levels were achieved above 0.3 mM Ca^{2+}. In previous studies K_D

values of 0.1– 0.15 mM were determined for the binding of Ca^{2+} to HlyA and C-terminal fragments of HlyA, which induced the folding [115,234,235]. Therefore, the secretion analysis indicates that the Ca^{2+}-induced folding increases secretion efficiency.

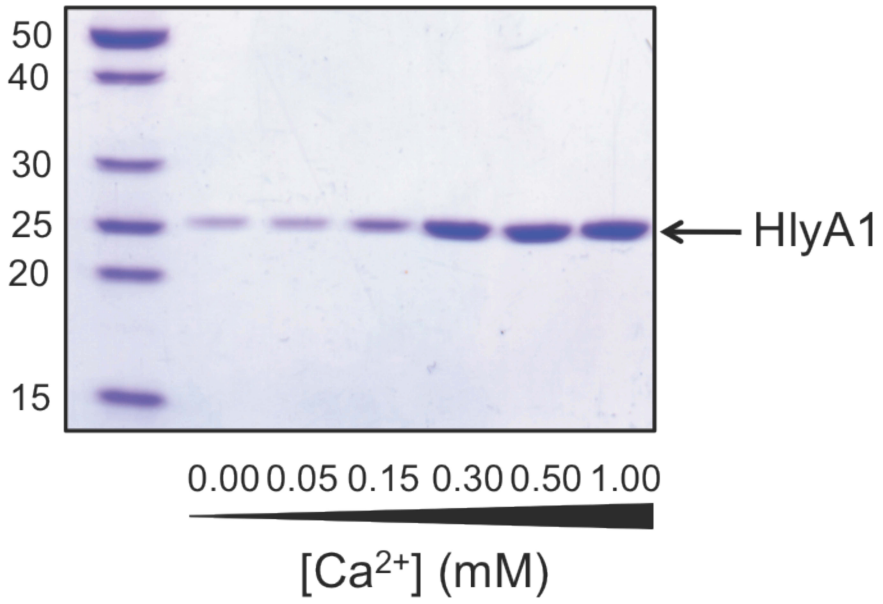


Figure 6-6: Secretion analysis of HlyA1 at various Ca^{2+} concentrations. Coomassie Brilliant Blue stained SDS-PAGE gel (15%) of supernatant samples from *E. coli* cultures secreting HlyA1 (23.8 kDa) at indicated Ca^{2+} concentrations.

Coherent with these data, fusion constructs of a protein-of-interest and C-terminal fragments of HlyA were secreted less efficiently, if parts of the RTX domains were deleted [122]. These results, however, could also be explained by an impaired interaction of the secretion constructs with the CLD via the RTX domain (Chapter I). Interestingly, rather small RTX proteins like proteases (around 50 kDa) contain seven to eight GG repeats, whereas large RTX proteins (> 500 kDa) contain > 40 GG repeats. Furthermore, some of the large RTX proteins contain an additional N-terminal RTX domain [117]. Obviously, the number of GG repeats correlates with the molecular mass of the RTX protein. The proportion of GG repeats to protein mass is about one GG repeat per 6– 12 kDa, which might reflect the capacity of one GG repeat to energize the transport by Ca^{2+}-induced folding.

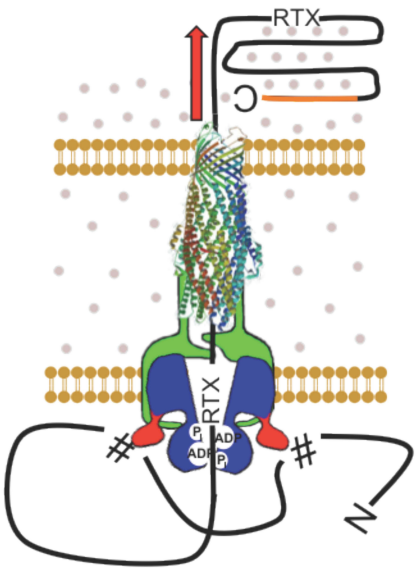


Figure 6-7: Pulling mechanism for the secretion of RTX proteins. In accordance to the activating model (C-terminal secretion), the unfolded RTX domain of RTX proteins reaches the exterior directly after the C-terminal secretion signal. The binding of Ca^{2+} ions to the GG repeats induces the folding of the RTX protein. This folding prevents the reverse diffusion of HlyA and represents a potential driving force (indicated by a red arrow) to accelerate the secretion process by pulling the RTX protein out of the cell and the cell envelope. C: C-terminus; gray speres: Ca^{2+}; green: HlyD N: N-terminus; RTX: RTX domain; rainbow: TolC (PDB code 1EK9); #: unknown, hypothetical sequence motif.

If folding of the RTX domain "pulls" the N-terminal domain out of the cell, this would further support the C-terminal secretion and thereby the activating model for T1SS transport (see section 6.1.2.2). In the terminating model (N-terminal secretion), the C-terminal RTX domain would reach the exterior at the end of the transport process unable to accelerate the secretion process. This "pulling mechanism" might represent an evolutionary adaption to afford the efficient secretion of rather large RTX proteins and might be not required for the transport of smaller proteins. However, the majority of proteins is secreted in an unfolded state with the dedicated transport systems. In eukaryotes, for example, the post-translational transport of unfolded or loosely folded proteins from the cytoplasm into the lumen of the endoplasmic reticulum is catalyzed by a tetrameric complex, the Sec62/Sec63 complex. The proteins are inserted into the channel and the transport is accomodated by brownian motion. In this system, the protein can slide in either direction. Therefore, the chaperone BiP binds to the unfolded protein inside the lumen of the endoplasmic reticulum and prevents the movement back into the cytosol. This ratcheting mechanism results in a net forward transport. Since the binding of BiP to the protein depends on ATP binding and hydrolysis, the transport consumes ATP [236]. Importantly, ATP is not consumed for the transport *per se*, but to

generate directionality of transport.

If proteins would specifically fold after transport, back movement is prevented even without any ATP consumption. Therefore, folding of proteins after their transport might represent an elegant strategy to drive transport and generate unidirectionality without the consumption of external energy. For RTX proteins, the Ca^{2+}-induced binding might represent such a mechanism that could even accelerate the transport by pulling the protein out of the cell and enable an efficient transport. That folding properties of the transported protein might provide a molecular driving force for transport was also recently suggested for Type V secretion systems of Gram-negative bacteria [237].

6.1.4 Function of HlyD during transport

The cytoplasmic tail of HlyD is essential for the secretion of HlyA and interacts with HlyA to recruit TolC [92]. Another study focussed on the periplasmic domain of HlyD and demonstrated that single point mutations in this domain decrease the secretion levels of HlyA and reduce the specific activity of secreted HlyA drastically (20– 250 fold), thus being important for the release of functional HlyA [233]. The authors stated that HlyD is "...important not only for the transport of HlyA itself, but also to ensure a correct folding step for the translocated toxin", however, a mechanism, how HlyD assists in the folding of HlyA, was not suggested. One way to interpret the data of Pimenta *et al.* [233] is that HlyD prevents folding of HlyA inside the translocon rather than actively assisting in the folding of HlyA. HlyD might function as a seal that covers the T1SS tunnel to prevent the influx of Ca^{2+} ions into the secretion complex, whose concentration is high inside the periplasm [238]. Thereby, HlyA is maintained in an unfolded state inside the T1SS machinery, which enables the passage through the complete translocon. Defects in the integrity of HlyD allow the influx of Ca^{2+} ions, which induces folding (at least partially) of complex-engaged HlyA. As a consequence (partially) folded HlyA is transported less efficiently and/or secreted HlyA takes a different folding pathway yielding protein with an altered specific activity. The study of Pimenta *et al.* [233] showed exactly the expected results by the secretion analysis with HlyD mutants: decreased secretion levels and a changed specific activity of secreted HlyA, in this case a reduced specific activity. Importantly, an increase in the extracellular Ca^{2+} concentration during the secretion experiments increased the observed effects, which suggests the entrance of Ca^{2+} ions into the secretion channel during transport. Beside the crucial involvement in the

initiation of the transport process [92], these data support a further function of HlyD for the efficient secretion of HlyA: shielding the transport channel from Ca^{2+} ions to maintain unfolded HlyA within the transport complex.

6.1.5 The activating alternating access model

Based on the described observations in this chapter, an "activating alternating access" model for the ABC-driven transport of T1SS substrates is suggested (Figure 6-8). The co-translational interaction of the substrate with the inner membrane complex recruits the OMP (step 1) [92,129]. During this doctoral reseach, an interaction of the RTX domain of unfolded HlyA with the CLD of HlyB was demonstrated. Whether additional interaction sites are present at the N-terminus of HlyA for HlyB or HlyD is currently unknown. However, additional sites at the N-terminus of HlyA would further protect the unfolded, secretion-competent state of HlyA in the cell prior to secretion. The NBDs of the ABC transporter are separated in the inward-facing conformation, which enables binding of the substrate (the secretion signal of HlyA) (step 2). The binding induces two events: 1. ATP-dependent dimerization of the NBDs, the power stroke of transport, which 2. converts the inward- into the outward-facing state (step 3). The substrate/secretion signal is released through a conduct-like structure of the ABC transporter into the transport complex and reaches the exterior (step 4). Whether the transport is energized by iterative catalytic cycles of the NBDs is unknown. Theoretically, the transport of HlyA and other RTX proteins might be driven by brownian motion and the directionality of transport might be conferred by folding of the protein in the exterior. This Ca^{2+}-induced folding of the RTX domains might also accelerate the secretion by pulling the N-terminal domain out of the cell (step 5). The hydrolysis of ATP and the sequential release of P_i and ADP completes one catalytic cycle of the NBDs. The release of the substrate to the exterior restores the inward-facing conformation of the transporter and the secretion complex dissociates (step 6).

Despite the results of this doctoral thesis, the precise mechanism of protein transport by T1SS is far away from being understood. One major breakthrough of this doctoral thesis was the establishment of an overexpression system for the ABC transporter HlyB and the membrane fusion protein HlyD in *E. coli*, however these data were not the focus of this doctoral thesis. Now, the detailed structural and biochemical characterization of

both membrane proteins are possible, which might answer the remaining questions of protein transport with T1SS and enhance the knowledge of membrane transport significantly in the future.

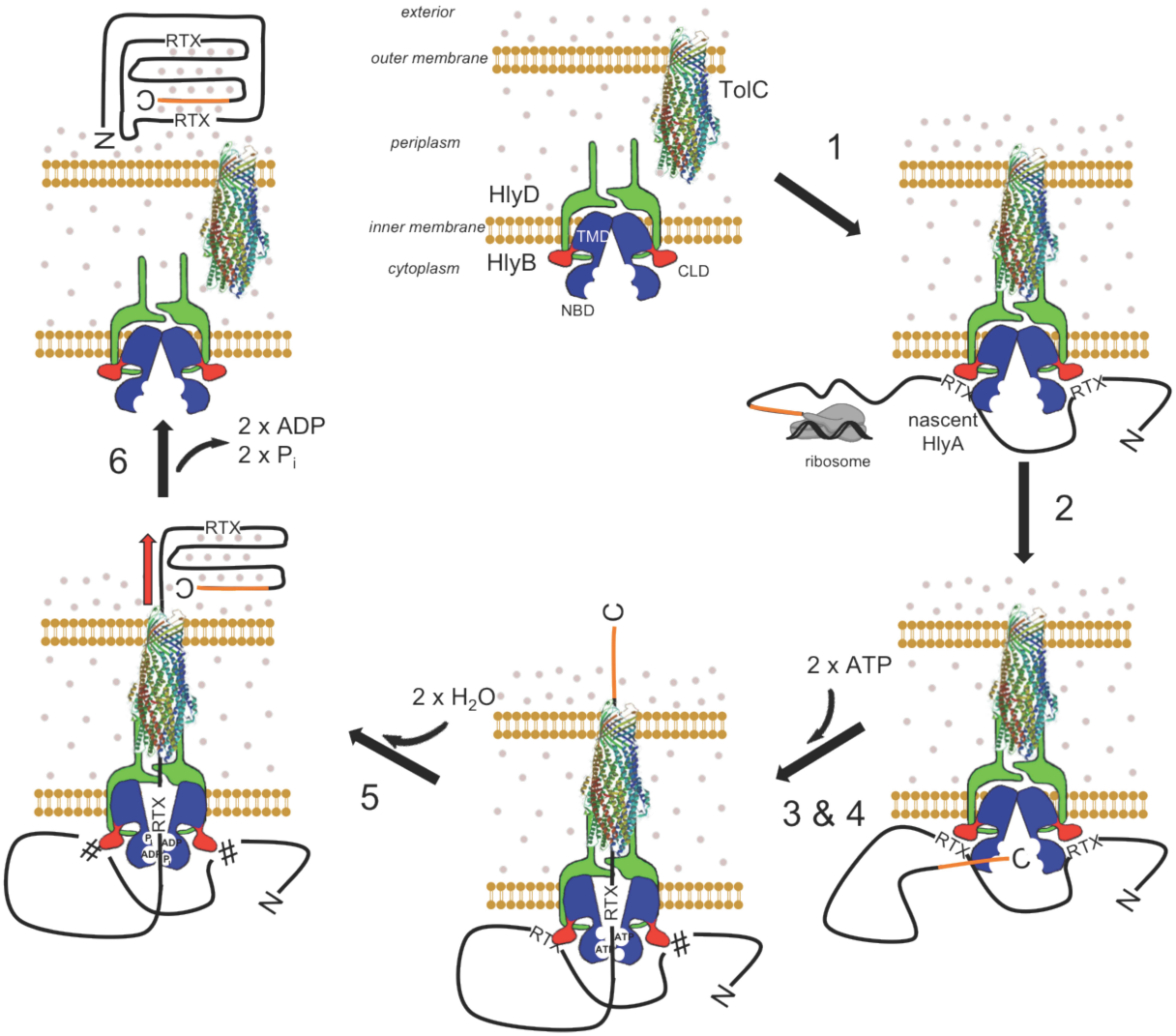


Figure 6-8: Representation of the *activating alternating access model* for the HlyA T1SS. Dimeric HlyB and oligomeric HlyD form a stable inner membrane complex and TolC is located in the outer membrane. Ribosomally synthesized HlyA interacts with the CLD of HlyB via the RTX domain, presumably co-translationally, and recruits TolC to assemble the complete HlyA T1SS (**1**). After synthesis, the C-terminal secretion signal of HlyA (orange) interacts with the NBDs of HlyB (**2**), which forces the ATP-dependent dimerization of the NBDs (**3**). The conversion of the inward-facing to the outward-facing state opens a conduct-like structure of HlyB and enables the release of the secretion signal from HlyB and the C-terminal transport to the cell surface (**4**). After the secretion signal, the RTX domain reaches the exterior. Ca^{2+}-mediated folding of the RTX domain directs transport to the exterior and might pull the N-terminal moiety of HlyA out of the cell (represented by the red arrow) (**5**). If HlyA contains additional interaction sites for the inner membrane complex N-terminal to the RTX domain (represented by a #) is unknown, however, it could further protect the N-terminal, hydrophobic domain of HlyA prior to secretion. The release of HlyA induces the dissociation of the NBD dimer and re-setting of the transport system (**6**). How many ATP molecules are fueled for the transport of one HlyA molecule is unknown, however, two ATP molecules are hydrolyzed for one catalytic cycle of the NBDs. In principle, this cycle would be sufficient for the transport of HlyA. Brownian motion in combination with the Ca^{2+}-induced folding of the RTX domain might drive the transfer of HlyA to the exterior. Blue: NBD and TMD of HlyB; C: C-terminus; gray speres: Ca^{2+}; green: HlyD N: N-terminus; RTX: RTX domain; rainbow: TolC (PDB code 1EK9); red: CLD; #: unknown, hypothetical sequence motif.

6.2 Biotechnological Application of the HlyA T1SS of *E. coli*

6.2.1 Biotechnological production of proteins

As the secretion of proteins features several advantages compared to the conventional production inside cells [169], the applicability of the HlyA T1SS of *E. coli* to secrete desired proteins was investigated. We have demonstrated in great detail that folding of the passenger protein competes with secretion and that slow-folding variants of proteins can be secreted in good yields. Moreover, the secreted proteins were shown to be biologically active. These studies represent a rational for the biotechnological production of proteins with the HlyA T1SS by the introduction of slow-folding mutations (Chapter V). In addition, the HlyA T1SS might also be a valuable tool for the generation of mutated proteins, for example proteins with improved lipolytic activity (Chapter VI). Despite these efforts, the usage of the secretion system is impaired by the prerequisite of slow-folding proteins. We analyzed the destabilizing domain ddFKBP to transfer instability to the secretion constructs and thereby maintaining the unfolded state. In eukaryotes, the ddFKBP was already applied successfully to destabilize fused proteins [239-241], whereas no studies were available for prokaryotes. Although the fusion constructs were expressed in the cells, the secretion levels were not increased. Obviously, the ddFKBP did not change the folding state of the fused proteins drastically and does not represent an useful tool for T1SS (Chapter V).

In parallel to these studies a new secretion system was invented that allows the secretion of any protein without the insertion of slow-folding mutations. The secretion system was not optimized in terms of the used secretion constructs and conditions, yet, nevertheless the initial secretion levels in batch cultures were already high and protein titers of up to 1 g per liter of cell culture were reached. Until now, this system succeeded in the secretory production of all analyzed fusion proteins including wild-type proteins like MBP and IFABP, which could not be secreted with the former system [124,125,156]. Moreover, the secreted proteins were shown to be functional. Therefore, the new secretion system might boost the secretory production of proteins in biotechnology and enable the production of proteins, which cannot be produced with current strategies.

6.2.2 Biotechnological production of peptides

The invented secretion system was also successfully employed in the secretion of peptides with good yields, without any visible degradation products and as soluble fusion proteins. These data suggest that the system represents the first universal biotechnological system for the production of peptides circumventing the known problems with recombinant techniques: instability, aggregation and cytotoxicity. In many cases, the separation of the secreted peptide from the secretion signal might be important for the quality of the product. The insertion of a short recognition sequence for a site-specific protease in between the peptide-of-interest and the C-terminal secretion signal represents a well-established procedure for this purpose. With the secretion system some residual amino acids are left attached to the C-terminus of the peptides or proteins (for example the tetrapeptide IDGR, which represents the recognition site for the protease Factor Xa). Moreover, an undesired methionine of the start codon might be present at the N-terminus of the peptide. Since the secretion of the peptides succeeded, obviously those peptides can be expressed in *E. coli,* and presumably other organisms, in full-length without being proteolytically degraded. Motivated by those results, another strategy was invented to produce those peptides without any artificial amino acids. This technique produces even higher yields than the first secretory system. In respect to the intellectual property status of both technologies, the detailed description of the systems will be published in the future. The technology of the new secretion system was submitted as patent application to the "Europäisches Patentamt" and is pending. The second strategy for the recombinant production of peptides will be submitted in the near future as patent application.

7 Literature

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

2 Gouaux, E. & Mackinnon, R. Principles of selective ion transport in channels and pumps. *Science* **310**, 1461-1465, doi:310/5753/1461 [pii] 10.1126/science.1113666 (2005).

3 Jeremy Berg, J. T., Lubert Stryer. *Biochemistry*. Vol. 5th 319 (W. H. Freeman and Company, 2002).

4 Bruce Alberts, A. J., Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Vol. 4th 583 (Garland Science, 2002).

5 Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-1038, doi:10.1002/pro.5560070420 (1998).

6 Abramson, J., Iwata, S. & Kaback, H. R. Lactose permease as a paradigm for membrane transport proteins (Review). *Mol Membr Biol* **21**, 227-236, doi:10.1080/09687680410001716862 TXRHTLY6QEGUC21X [pii] (2004).

7 Boyer, P. D. The ATP synthase--a splendid molecular machine. *Annu Rev Biochem* **66**, 717-749, doi:10.1146/annurev.biochem.66.1.717 (1997).

8 Lanyi, J. K. Bacteriorhodopsin. *Annu Rev Physiol* **66**, 665-688, doi:10.1146/annurev.physiol.66.032102.150049 (2004).

9 Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969-970 (1966).

10 Zgurskaya, H. I. Multicomponent drug efflux complexes: architecture and mechanism of assembly. *Future Microbiol* **4**, 919-932, doi:10.2217/fmb.09.62 (2009).

11 Albers, S. V., Koning, S. M., Konings, W. N. & Driessen, A. J. Insights into ABC transport in archaea. *J Bioenerg Biomembr* **36**, 5-15 (2004).

12 Dean, M. & Annilo, T. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. *Annu Rev Genomics Hum Genet* **6**, 123-142, doi:10.1146/annurev.genom.6.080604.162122 (2005).

13 Jones, P. M., O'Mara, M. L. & George, A. M. ABC transporters: a riddle wrapped in a mystery inside an enigma. *Trends Biochem Sci* **34**, 520-531, doi:S0968-0004(09)00145-5 [pii] 10.1016/j.tibs.2009.06.004 (2009).

14 Moussatova, A., Kandt, C., O'Mara, M. L. & Tieleman, D. P. ATP-binding cassette transporters in *Escherichia coli*. *Biochim Biophys Acta* **1778**, 1757-1771, doi:S0005-2736(08)00194-6 [pii] 10.1016/j.bbamem.2008.06.009 (2008).

15 Berntsson, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J. & Poolman, B. A structural classification of substrate-binding proteins. *FEBS Lett* **584**, 2606-2617, doi:S0014-5793(10)00327-3 [pii] 10.1016/j.febslet.2010.04.043 (2010).

16 Duurkens, R. H., Tol, M. B., Geertsma, E. R., Permentier, H. P. & Slotboom, D. J. Flavin binding to the high affinity riboflavin transporter RibU. *J Biol Chem* **282**, 10380-10386, doi:M608583200 [pii] 10.1074/jbc.M608583200 (2007).

17 Zhang, P., Wang, J. & Shi, Y. Structure and mechanism of the S component of a bacterial ECF transporter. *Nature* **468**, 717-720, doi:nature09488 [pii] 10.1038/nature09488 (2010).

18 Erkens, G. B. *et al.* The structural basis of modularity in ECF-type ABC transporters. *Nat Struct Mol Biol* **18**, 755-760, doi:nsmb.2073 [pii] 10.1038/nsmb.2073 (2011).

19 Saraste, M., Sibbald, P. R. & Wittinghofer, A. The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* **15**, 430-434 (1990).

20 Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**, 945-951 (1982).

21 Schmitt, L. & Tampe, R. Structure and mechanism of ABC transporters. *Curr Opin Struct Biol* **12**, 754-760, doi:S0959440X02003998 [pii] (2002).

22 Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101**, 789-800, doi:S0092-8674(00)80890-9 [pii] (2000).

23 Schmitt, L., Benabdelhak, H., Blight, M. A., Holland, I. B. & Stubbs, M. T. Crystal structure of the nucleotide-binding domain of the ABC-transporter haemolysin B: identification of a variable region within ABC helical domains. *J Mol Biol* **330**, 333-342, doi:S0022283603005928 [pii] (2003).

24 Benabdelhak, H. *et al.* Positive co-operative activity and dimerization of the isolated ABC ATPase domain of HlyB from *Escherichia coli. Biochem J* **386**, 489-495, doi:BJ20041282 [pii] 10.1042/BJ20041282 (2005).

25 Nikaido, K., Liu, P. Q. & Ames, G. F. Purification and characterization of HisP, the ATP-binding subunit of a traffic ATPase (ABC transporter), the histidine permease of *Salmonella typhimurium*. Solubility, dimerization, and ATPase activity. *J Biol Chem* **272**, 27745-27752 (1997).

26 Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* **10**, 139-149, doi:S1097276502005762 [pii] (2002).

27 Zaitseva, J. *et al.* A molecular understanding of the catalytic cycle of the nucleotide-binding domain of the ABC transporter HlyB. *Biochem Soc Trans* **33**, 990-995, doi:BST20050990 [pii] 10.1042/BST20050990 (2005).

28 Zaitseva, J., Jenewein, S., Jumpertz, T., Holland, I. B. & Schmitt, L. H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J* **24**, 1901-1910, doi:7600657 [pii] 10.1038/sj.emboj.7600657 (2005).

29 Diederichs, K. *et al.* Crystal structure of MalK, the ATPase subunit of the trehalose/maltose ABC transporter of the archaeon *Thermococcus litoralis*. *EMBO J* **19**, 5951-5961, doi:10.1093/emboj/19.22.5951 (2000).

30 Yuan, Y. R. *et al.* The crystal structure of the MJ0796 ATP-binding cassette. Implications for the structural consequences of ATP hydrolysis in the active site of an ABC transporter. *J Biol Chem* **276**, 32313-32321, doi:10.1074/jbc.M100758200 M100758200 [pii] (2001).

31 Karpowich, N. *et al.* Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure* **9**, 571-586, doi:S0969212601006177 [pii] (2001).

32 Oswald, C., Holland, I. B. & Schmitt, L. The motor domains of ABC-transporters. What can structures tell us? *Naunyn Schmiedebergs Arch Pharmacol* **372**, 385-399, doi:10.1007/s00210-005-0031-4 (2006).

33 Fetsch, E. E. & Davidson, A. L. Vanadate-catalyzed photocleavage of the signature motif of an ATP-binding cassette (ABC) transporter. *Proc Natl Acad Sci U S A* **99**, 9685-9690, doi:10.1073/pnas.152204499 152204499 [pii] (2002).

34 Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. *Nat Struct Mol Biol* **11**, 918-926, doi:10.1038/nsmb836/nsmb836 [pii] (2004).

35 Oldham, M. L. & Chen, J. Snapshots of the maltose transporter during ATP hydrolysis. *Proc Natl Acad Sci U S A* **108**, 15152-15156, doi:1108858108 [pii] 10.1073/pnas.1108858108 (2011).

36 Orelle, C., Dalmas, O., Gros, P., Di Pietro, A. & Jault, J. M. The conserved glutamate residue adjacent to the Walker-B motif is the catalytic base for ATP hydrolysis in the ATP-binding cassette transporter BmrA. *J Biol Chem* **278**, 47002-47008, doi:10.1074/jbc.M308268200 M308268200 [pii] (2003).

37 Verdon, G. *et al.* Formation of the productive ATP-Mg2+-bound dimer of GlcV, an ABC-ATPase from *Sulfolobus solfataricus*. *J Mol Biol* **334**, 255-267, doi:S0022283603012002 [pii] (2003).

38 Geourjon, C. *et al.* A common mechanism for ATP hydrolysis in ABC transporter and helicase superfamilies. *Trends Biochem Sci* **26**, 539-544, doi:S0968-0004(01)01907-7 [pii] (2001).

39 Story, R. M. & Steitz, T. A. Structure of the recA protein-ADP complex. *Nature* **355**, 374-376, doi:10.1038/355374a0 (1992).

40 Patzlaff, J. S., van der Heide, T. & Poolman, B. The ATP/substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J Biol Chem* **278**, 29546-29551, doi:10.1074/jbc.M304796200 M304796200 [pii] (2003).

41 Mimmack, M. L. *et al.* Energy coupling to periplasmic binding protein-dependent transport systems: stoichiometry of ATP hydrolysis during transport in vivo. *Proc Natl Acad Sci U S A* **86**, 8257-8261 (1989).
42 Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213-216, doi:nature05626 [pii] 10.1038/nature05626 (2007).
43 Hohl, M., Briand, C., Grutter, M. G. & Seeger, M. A. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. *Nat Struct Mol Biol* **19**, 395-402, doi:nsmb.2267 [pii] 10.1038/nsmb.2267 (2012).
44 Oldham, M. L. & Chen, J. Crystal structure of the maltose transporter in a pretranslocation intermediate state. *Science* **332**, 1202-1205, doi:science.1200767 [pii] 10.1126/science.1200767 (2011).
45 Loo, T. W. & Clarke, D. M. Recent progress in understanding the mechanism of P-glycoprotein-mediated drug efflux. *J Membr Biol* **206**, 173-185, doi:10.1007/s00232-005-0792-1 (2005).
46 Sauna, Z. E. & Ambudkar, S. V. Evidence for a requirement for ATP hydrolysis at two distinct steps during a single turnover of the catalytic cycle of human P-glycoprotein. *Proc Natl Acad Sci U S A* **97**, 2515-2520, doi:97/6/2515 [pii] (2000).
47 Zhou, Z. *et al.* The two ATP binding sites of cystic fibrosis transmembrane conductance regulator (CFTR) play distinct roles in gating kinetics and energetics. *J Gen Physiol* **128**, 413-422, doi:jgp.200609622 [pii] 10.1085/jgp.200609622 (2006).
48 Chen, M., Abele, R. & Tampe, R. Functional non-equivalence of ATP-binding cassette signature motifs in the transporter associated with antigen processing (TAP). *J Biol Chem* **279**, 46073-46081, doi:10.1074/jbc.M404042200 M404042200 [pii] (2004).
49 Zolnerciks, J. K., Andress, E. J., Nicolaou, M. & Linton, K. J. Structure of ABC transporters. *Essays Biochem* **50**, 43-61, doi:bse0500043 [pii] 10.1042/bse0500043 (2011).
50 Pagant, S., Brovman, E. Y., Halliday, J. J. & Miller, E. A. Mapping of interdomain interfaces required for the functional architecture of Yor1p, a eukaryotic ATP-binding cassette (ABC) transporter. *J Biol Chem* **283**, 26444-26451, doi:M803912200 [pii] 10.1074/jbc.M803912200 (2008).
51 Oancea, G. *et al.* Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. *Proc Natl Acad Sci U S A* **106**, 5551-5556, doi:0811260106 [pii] 10.1073/pnas.0811260106 (2009).
52 Dalmas, O. *et al.* The Q-loop disengages from the first intracellular loop during the catalytic cycle of the multidrug ABC transporter BmrA. *J Biol Chem* **280**, 36857-36864, doi:M503266200 [pii] 10.1074/jbc.M503266200 (2005).
53 Locher, K. P., Lee, A. T. & Rees, D. C. The E. coli BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296**, 1091-1098, doi:10.1126/science.1071142 296/5570/1091 [pii] (2002).
54 Zolnerciks, J. K., Wooding, C. & Linton, K. J. Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein. *FASEB J* **21**, 3937-3948, doi:fj.07-8610com [pii] 10.1096/fj.07-8610com (2007).
55 Doige, C. A. & Ames, G. F. ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. *Annu Rev Microbiol* **47**, 291-319, doi:10.1146/annurev.mi.47.100193.001451 (1993).
56 Chang, G. Multidrug resistance ABC transporters. *FEBS Lett* **555**, 102-105, doi:S0014579303010858 [pii] (2003).
57 Dassa, E. & Hofnung, M. Sequence of gene malG in *E. coli* K12: homologies between integral membrane components from binding protein-dependent transport systems. *EMBO J* **4**, 2287-2293 (1985).
58 Saurin, W., Koster, W. & Dassa, E. Bacterial binding protein-dependent permeases: characterization of distinctive signatures for functionally related integral cytoplasmic membrane proteins. *Mol Microbiol* **12**, 993-1004 (1994).
59 Dawson, R. J. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180-185, doi:nature05155 [pii] 10.1038/nature05155 (2006).

60 Davidson, A. L., Shuman, H. A. & Nikaido, H. Mechanism of maltose transport in *Escherichia coli*: transmembrane signaling by periplasmic binding proteins. *Proc Natl Acad Sci U S A* **89**, 2360-2364 (1992).

61 Mannering, D. E., Sharma, S. & Davidson, A. L. Demonstration of conformational changes associated with activation of the maltose transport complex. *J Biol Chem* **276**, 12362-12368, doi:10.1074/jbc.M011686200
M011686200 [pii] (2001).

62 Gao, M. *et al.* Comparison of the functional characteristics of the nucleotide binding domains of multidrug resistance protein 1. *J Biol Chem* **275**, 13098-13108 (2000).

63 Qu, Q., Russell, P. L. & Sharom, F. J. Stoichiometry and affinity of nucleotide binding to P-glycoprotein during the catalytic cycle. *Biochemistry* **42**, 1170-1177, doi:10.1021/bi026555j (2003).

64 Gradia, S. *et al.* hMSH2-hMSH6 forms a hydrolysis-independent sliding clamp on mismatched DNA. *Mol Cell* **3**, 255-261, doi:S1097-2765(00)80316-0 [pii] (1999).

65 Lamers, M. H., Winterwerp, H. H. & Sixma, T. K. The alternating ATPase domains of MutS control DNA mismatch repair. *EMBO J* **22**, 746-756, doi:10.1093/emboj/cdg064 (2003).

66 Austermuhle, M. I., Hall, J. A., Klug, C. S. & Davidson, A. L. Maltose-binding protein is open in the catalytic transition state for ATP hydrolysis during maltose transport. *J Biol Chem* **279**, 28243-28250, doi:10.1074/jbc.M403508200
M403508200 [pii] (2004).

67 Martin, C. *et al.* Drug binding sites on P-glycoprotein are altered by ATP binding prior to nucleotide hydrolysis. *Biochemistry* **39**, 11901-11906, doi:bi000559b [pii] (2000).

68 Mourez, M., Hofnung, M. & Dassa, E. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J* **16**, 3066-3077, doi:10.1093/emboj/16.11.3066 (1997).

69 Beveridge, T. J. Structures of gram-negative cell walls and their derived membrane vesicles. *J Bacteriol* **181**, 4725-4733 (1999).

70 Matias, V. R. & Beveridge, T. J. Native cell wall organization shown by cryo-electron microscopy confirms the existence of a periplasmic space in *Staphylococcus aureus*. *J Bacteriol* **188**, 1011-1021, doi:188/3/1011 [pii]
10.1128/JB.188.3.1011-1021.2006 (2006).

71 Bleves, S. *et al.* Protein secretion systems in *Pseudomonas aeruginosa*: A wealth of pathogenic weapons. *Int J Med Microbiol* **300**, 534-543, doi:S1438-4221(10)00085-8 [pii]
10.1016/j.ijmm.2010.08.005 (2010).

72 Tseng, T. T., Tyler, B. M. & Setubal, J. C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. *BMC Microbiol* **9 Suppl 1**, S2, doi:1471-2180-9-S1-S2 [pii]
10.1186/1471-2180-9-S1-S2 (2009).

73 Holland, I. B., Schmitt, L. & Young, J. Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol* **22**, 29-39 (2005).

74 Cianciotto, N. P. Type II secretion: a protein secretion system for all seasons. *Trends Microbiol* **13**, 581-588, doi:S0966-842X(05)00256-8 [pii]
10.1016/j.tim.2005.09.005 (2005).

75 Filloux, A. The underlying mechanisms of type II protein secretion. *Biochim Biophys Acta* **1694**, 163-179, doi:S0167488904001181 [pii]
10.1016/j.bbamcr.2004.05.003 (2004).

76 Cornelis, G. R. The type III secretion injectisome. *Nat Rev Microbiol* **4**, 811-825, doi:nrmicro1526 [pii]
10.1038/nrmicro1526 (2006).

77 Christie, P. J. & Cascales, E. Structural and dynamic properties of bacterial type IV secretion systems (review). *Mol Membr Biol* **22**, 51-61 (2005).

78 Dautin, N. & Bernstein, H. D. Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu Rev Microbiol* **61**, 89-112, doi:10.1146/annurev.micro.61.080706.093233 (2007).

79 Bingle, L. E., Bailey, C. M. & Pallen, M. J. Type VI secretion: a beginner's guide. *Curr Opin Microbiol* **11**, 3-8, doi:S1369-5274(08)00007-6 [pii]
10.1016/j.mib.2008.01.006 (2008).

80 Bonemann, G., Pietrosiuk, A. & Mogk, A. Tubules and donuts: a type VI secretion story. *Mol Microbiol* **76**, 815-821, doi:MMI7171 [pii]
10.1111/j.1365-2958.2010.07171.x (2010).

81 Hinsa, S. M., Espinosa-Urgel, M., Ramos, J. L. & O'Toole, G. A. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Mol Microbiol* **49**, 905-918, doi:3615 [pii] (2003).

82 Sapriel, G., Wandersman, C. & Delepelaire, P. The SecB chaperone is bifunctional in *Serratia marcescens*: SecB is involved in the Sec pathway and required for HasA secretion by the ABC transporter. *J Bacteriol* **185**, 80-88 (2003).

83 Thanabalu, T., Koronakis, E., Hughes, C. & Koronakis, V. Substrate-induced assembly of a contiguous channel for protein export from *E.coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J* **17**, 6487-6496, doi:10.1093/emboj/17.22.6487 (1998).

84 Letoffe, S., Delepelaire, P. & Wandersman, C. Protein secretion in gram-negative bacteria: assembly of the three components of ABC protein-mediated exporters is ordered and promoted by substrate binding. *EMBO J* **15**, 5804-5811 (1996).

85 Uhlen, P. *et al.* Alpha-haemolysin of uropathogenic *E. coli* induces Ca2+ oscillations in renal epithelial cells. *Nature* **405**, 694-697, doi:10.1038/35015091 (2000).

86 Dhakal, B. K. & Mulvey, M. A. The UPEC pore-forming toxin alpha-hemolysin triggers proteolysis of host proteins to disrupt cell adhesion, inflammatory, and survival pathways. *Cell Host Microbe* **11**, 58-69, doi:S1931-3128(11)00401-X [pii] 10.1016/j.chom.2011.12.003 (2012).

87 Bakas, L., Mate, S., Vazquez, R., Herlax, V. in *Biochemistry* 452 (InTech, 2012).

88 Menestrina, G., Moser, C., Pellet, S. & Welch, R. Pore-formation by *Escherichia coli* hemolysin (HlyA) and other members of the RTX toxins family. *Toxicology* **87**, 249-267 (1994).

89 Suttorp, N., Floer, B., Schnittler, H., Seeger, W. & Bhakdi, S. Effects of *Escherichia coli* hemolysin on endothelial cell function. *Infect Immun* **58**, 3796-3801 (1990).

90 Ludwig, A., Jarchau, T., Benz, R. & Goebel, W. The repeat domain of *Escherichia coli* haemolysin (HlyA) is responsible for its Ca2+-dependent binding to erythrocytes. *Mol Gen Genet* **214**, 553-561 (1988).

91 Gentschev, I., Dietrich, G. & Goebel, W. The *E. coli* alpha-hemolysin secretion system and its use in vaccine development. *Trends Microbiol* **10**, 39-45, doi:S0966842X01022594 [pii] (2002).

92 Balakrishnan, L., Hughes, C. & Koronakis, V. Substrate-triggered recruitment of the TolC channel-tunnel during type I export of hemolysin by *Escherichia coli*. *J Mol Biol* **313**, 501-510, doi:10.1006/jmbi.2001.5038 S0022-2836(01)95038-7 [pii] (2001).

93 Goebel, W. & Schrempf, H. Isolation and characterization of supercoiled circular deoxyribonucleic acid from beta-hemolytic strains of *Escherichia coli*. *J Bacteriol* **106**, 311-317 (1971).

94 Nieto, J. M. *et al.* Expression of the hemolysin operon in *Escherichia coli* is modulated by a nucleoid-protein complex that includes the proteins Hha and H-NS. *Mol Gen Genet* **263**, 349-358 (2000).

95 Koronakis, V., Cross, M. & Hughes, C. Expression of the *E.coli* hemolysin secretion gene hlyB involves transcript anti-termination within the hly operon. *Nucleic Acids Res* **16**, 4789-4800 (1988).

96 Welch, R. A. & Pellett, S. Transcriptional organization of the *Escherichia coli* hemolysin genes. *J Bacteriol* **170**, 1622-1630 (1988).

97 Hess, J., Wels, W., Vogel, M. & Goebel, W. Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with a corresponding chromosomal hemolysin sequence. *FEMS Microbiol Lett* **34**, 1-11 (1986).

98 Koronakis, V., Cross, M. & Hughes, C. Transcription antitermination in an *Escherichia coli* haemolysin operon is directed progressively by cis-acting DNA sequences upstream of the promoter region. *Mol Microbiol* **3**, 1397-1404 (1989).

99 Vogel, M., Hess, J., Then, I., Juarez, A. & Goebel, W. Characterization of a sequence (hlyR) which enhances synthesis and secretion of hemolysin in *Escherichia coli*. *Mol Gen Genet* **212**, 76-84 (1988).

100 Jubete, Y., Zabala, J. C., Juarez, A. & de la Cruz, F. hlyM, a transcriptional silencer downstream of the promoter in the hly operon of *Escherichia coli*. *J Bacteriol* **177**, 242-246 (1995).

101 Hilbert, D. W. *et al.* Clinical *Escherichia coli* isolates utilize alpha-hemolysin to inhibit in vitro epithelial cytokine production. *Microbes Infect*, doi:S1286-4579(12)00039-1 [pii] 10.1016/j.micinf.2012.01.010 (2012).

102 Colin Hughes, J. H., Anthony Roberts, Werner Goebel. Hemolysin Production as a Virulence Marker in Symptomatic and Asymptomatic Urinary Tract Infections Caused by *Escherichia coli. Infection and Immunity* **39**, 546–551 (1982).

103 Hyland, C., Vuillard, L., Hughes, C. & Koronakis, V. Membrane interaction of *Escherichia coli* hemolysin: flotation and insertion-dependent labeling by phospholipid vesicles. *J Bacteriol* **183**, 5364-5370 (2001).

104 Bhakdi, S. & Martin, E. Superoxide generation by human neutrophils induced by low doses of *Escherichia coli* hemolysin. *Infect Immun* **59**, 2955-2962 (1991).

105 Hamon, M. A. *et al.* Histone modifications induced by a family of bacterial toxins. *Proc Natl Acad Sci U S A* **104**, 13467-13472, doi:0702729104 [pii] 10.1073/pnas.0702729104 (2007).

106 Ratner, A. J. *et al.* Epithelial cells are sensitive detectors of bacterial pore-forming toxins. *J Biol Chem* **281**, 12994-12998, doi:M511431200 [pii] 10.1074/jbc.M511431200 (2006).

107 Bhakdi, S., Mackman, N., Nicaud, J. M. & Holland, I. B. *Escherichia coli* hemolysin may damage target cell membranes by generating transmembrane pores. *Infect Immun* **52**, 63-69 (1986).

108 Gadeberg, O. V. & Orskov, I. In vitro cytotoxic effect of alpha-hemolytic *Escherichia coli* on human blood granulocytes. *Infect Immun* **45**, 255-260 (1984).

109 Stanley, P., Packman, L. C., Koronakis, V. & Hughes, C. Fatty acylation of two internal lysine residues required for the toxic activity of *Escherichia coli* hemolysin. *Science* **266**, 1992-1996 (1994).

110 Hughes, C., Stanley, P. & Koronakis, V. *E. coli* hemolysin interactions with prokaryotic and eukaryotic cell membranes. *Bioessays* **14**, 519-525, doi:10.1002/bies.950140804 (1992).

111 Issartel, J. P., Koronakis, V. & Hughes, C. Activation of *Escherichia coli* prohaemolysin to the mature toxin by acyl carrier protein-dependent fatty acylation. *Nature* **351**, 759-761, doi:10.1038/351759a0 (1991).

112 Hardie, K. R., Issartel, J. P., Koronakis, E., Hughes, C. & Koronakis, V. In vitro activation of *Escherichia coli* prohaemolysin to the mature membrane-targeted toxin requires HlyC and a low molecular-weight cytosolic polypeptide. *Mol Microbiol* **5**, 1669-1679 (1991).

113 Nicaud, J. M., Mackman, N., Gray, L. & Holland, I. B. Characterisation of HlyC and mechanism of activation and secretion of haemolysin from *E. coli* 2001. *FEBS Lett* **187**, 339-344, doi:0014-5793(85)81272-2 [pii] (1985).

114 Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J* **12**, 3357-3364 (1993).

115 Ostolaza, H., Soloaga, A. & Goni, F. M. The binding of divalent cations to *Escherichia coli* alpha-haemolysin. *Eur J Biochem* **228**, 39-44 (1995).

116 Lally, E. T., Hill, R. B., Kieba, I. R. & Korostoff, J. The interaction between RTX toxins and target cells. *Trends Microbiol* **7**, 356-361, doi:S0966-842X(99)01530-9 [pii] (1999).

117 Linhartova, I. *et al.* RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* **34**, 1076-1112, doi:FMR231 [pii] 10.1111/j.1574-6976.2010.00231.x (2010).

118 Welch, R. A. Pore-forming cytolysins of gram-negative bacteria. *Mol Microbiol* **5**, 521-528 (1991).

119 Boehm, D. F., Welch, R. A. & Snyder, I. S. Domains of *Escherichia coli* hemolysin (HlyA) involved in binding of calcium and erythrocyte membranes. *Infect Immun* **58**, 1959-1964 (1990).

120 Rose, T., Sebo, P., Bellalou, J. & Ladant, D. Interaction of calcium with *Bordetella pertussis* adenylate cyclase toxin. Characterization of multiple calcium-binding sites and calcium-induced conformational changes. *J Biol Chem* **270**, 26370-26376 (1995).

121 Holland, I. B., Jones, H. E., Campbell, A. K. & Jacq, A. An assessment of the role of intracellular free Ca2+ in *E. coli*. *Biochimie* **81**, 901-907, doi:S0300-9084(99)00205-9 [pii] (1999).

122 Kenny, B., Haigh, R. & Holland, I. B. Analysis of the haemolysin transport process through the secretion from *Escherichia coli* of PCM, CAT or beta-galactosidase fused to the Hly C-terminal signal domain. *Mol Microbiol* **5**, 2557-2568 (1991).

123 Koronakis, V., Sharff, A., Koronakis, E., Luisi, B. & Hughes, C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, 914-919, doi:10.1038/35016007 (2000).

124 Bakkes, P. J., Jenewein, S., Smits, S. H., Holland, I. B. & Schmitt, L. The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system. *J Biol Chem* **285**, 40573-40580, doi:M110.173658 [pii] 10.1074/jbc.M110.173658 (2010).

125 Schwarz, C. K., Landsberg, C. D., Lenders, M. H., Smits, S. H. & Schmitt, L. Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form. *J Biotechnol*, doi:S0168-1656(12)00127-7 [pii] 10.1016/j.jbiotec.2012.02.005 (2012).

126 Koronakis, V., Koronakis, E. & Hughes, C. Isolation and analysis of the C-terminal signal directing export of *Escherichia coli* hemolysin protein across both bacterial membranes. *EMBO J* **8**, 595-605 (1989).

127 Felmlee, T., Pellett, S., Lee, E. Y. & Welch, R. A. *Escherichia coli* hemolysin is released extracellularly without cleavage of a signal peptide. *J Bacteriol* **163**, 88-93 (1985).

128 Benabdelhak, H. *et al.* A specific interaction between the NBD of the ABC-transporter HlyB and a C-terminal fragment of its transport substrate haemolysin A. *J Mol Biol* **327**, 1169-1179, doi:S0022283603002043 [pii] (2003).

129 Masi, M. & Wandersman, C. Multiple signals direct the assembly and function of a type 1 secretion system. *J Bacteriol* **192**, 3861-3869, doi:JB.00178-10 [pii] 10.1128/JB.00178-10 (2010).

130 Koronakis, V., Hughes, C. & Koronakis, E. Energetically distinct early and late stages of HlyB/HlyD-dependent secretion across both *Escherichia coli* membranes. *EMBO J* **10**, 3263-3272 (1991).

131 Jarchau, T., Chakraborty, T., Garcia, F. & Goebel, W. Selection for transport competence of C-terminal polypeptides derived from *Escherichia coli* hemolysin: the shortest peptide capable of autonomous HlyB/HlyD-dependent secretion comprises the C-terminal 62 amino acids of HlyA. *Mol Gen Genet* **245**, 53-60 (1994).

132 Delepelaire, P. & Wandersman, C. The SecB chaperone is involved in the secretion of the *Serratia marcescens* HasA protein through an ABC transporter. *EMBO J* **17**, 936-944, doi:10.1093/emboj/17.4.936 (1998).

133 Zgurskaya, H. I., Krishnamoorthy, G., Ntreh, A. & Lu, S. Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. *Front Microbiol* **2**, 189, doi:10.3389/fmicb.2011.00189 (2011).

134 Tikhonova, E. B. & Zgurskaya, H. I. AcrA, AcrB, and TolC of *Escherichia coli* Form a Stable Intermembrane Multidrug Efflux Complex. *J Biol Chem* **279**, 32116-32124, doi:10.1074/jbc.M402230200 M402230200 [pii] (2004).

135 Zhang, A., Rosner, J. L. & Martin, R. G. Transcriptional activation by MarA, SoxS and Rob of two tolC promoters using one binding site: a complex promoter configuration for tolC in *Escherichia coli*. *Mol Microbiol* **69**, 1450-1455, doi:MMI6371 [pii] 10.1111/j.1365-2958.2008.06371.x (2008).

136 Koronakis, V. TolC--the bacterial exit duct for proteins and drugs. *FEBS Lett* **555**, 66-71, doi:S0014579303011256 [pii] (2003).

137 Andersen, C. *et al.* Transition to the open state of the TolC periplasmic tunnel entrance. *Proc Natl Acad Sci U S A* **99**, 11103-11108, doi:10.1073/pnas.162039399 162039399 [pii] (2002).

138 Pei, X. Y. *et al.* Structures of sequential open states in a symmetrical opening transition of the TolC exit duct. *Proc Natl Acad Sci U S A*, doi:1012588108 [pii] 10.1073/pnas.1012588108 (2011).

139 Paulsen, I. T., Park, J. H., Choi, P. S. & Saier, M. H., Jr. A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria. *FEMS Microbiol Lett* **156**, 1-8, doi:S0378-1097(97)00379-0 [pii] (1997).

140 Su, C. C. *et al.* Crystal structure of the membrane fusion protein CusB from *Escherichia coli*. *J Mol Biol* **393**, 342-355, doi:S0022-2836(09)01025-0 [pii] 10.1016/j.jmb.2009.08.029 (2009).

141 Thanassi, D. G., Cheng, L. W. & Nikaido, H. Active efflux of bile salts by *Escherichia coli*. *J Bacteriol* **179**, 2512-2518 (1997).

142 Zgurskaya, H. I. & Nikaido, H. Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc Natl Acad Sci U S A* **96**, 7190-7195 (1999).

143 Sulavik, M. C. *et al.* Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob Agents Chemother* **45**, 1126-1136, doi:10.1128/AAC.45.4.1126-1136.2001 (2001).

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

145 Higgins, M. K., Bokma, E., Koronakis, E., Hughes, C. & Koronakis, V. Structure of the periplasmic component of a bacterial drug efflux pump. *Proc Natl Acad Sci U S A* **101**, 9994-9999, doi:10.1073/pnas.0400375101 0400375101 [pii] (2004).

146 Mikolosko, J., Bobyk, K., Zgurskaya, H. I. & Ghosh, P. Conformational flexibility in the multidrug efflux system protein AcrA. *Structure* **14**, 577-587, doi:S0969-2126(06)00063-3 [pii] 10.1016/j.str.2005.11.015 (2006).

147 Yum, S. *et al.* Crystal structure of the periplasmic component of a tripartite macrolide-specific efflux pump. *J Mol Biol* **387**, 1286-1297, doi:S0022-2836(09)00216-2 [pii] 10.1016/j.jmb.2009.02.048 (2009).

148 Zgurskaya, H. I. & Nikaido, H. AcrA is a highly asymmetric protein capable of spanning the periplasm. *J Mol Biol* **285**, 409-420, doi:S0022-2836(98)92313-0 [pii] 10.1006/jmbi.1998.2313 (1999).

149 Zgurskaya, H. I. & Nikaido, H. Cross-linked complex between oligomeric periplasmic lipoprotein AcrA and the inner-membrane-associated multidrug efflux pump AcrB from *Escherichia coli*. *J Bacteriol* **182**, 4264-4267 (2000).

150 Tenson, T., Lovmar, M. & Ehrenberg, M. The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J Mol Biol* **330**, 1005-1014, doi:S0022283603006624 [pii] (2003).

151 Schulein, R., Gentschev, I., Mollenkopf, H. J. & Goebel, W. A topological model for the haemolysin translocator protein HlyD. *Mol Gen Genet* **234**, 155-163 (1992).

152 Tikhonova, E. B., Devroy, V. K., Lau, S. Y. & Zgurskaya, H. I. Reconstitution of the *Escherichia coli* macrolide transporter: the periplasmic membrane fusion protein MacA stimulates the ATPase activity of MacB. *Mol Microbiol* **63**, 895-910, doi:MMI5549 [pii] 10.1111/j.1365-2958.2006.05549.x (2007).

153 Lin, H. T. *et al.* MacB ABC transporter is a dimer whose ATPase activity and macrolide-binding capacity are regulated by the membrane fusion protein MacA. *J Biol Chem* **284**, 1145-1154, doi:M806964200 [pii] 10.1074/jbc.M806964200 (2009).

154 Pimenta, A. L., Young, J., Holland, I. B. & Blight, M. A. Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Mol Gen Genet* **261**, 122-132 (1999).

155 Blight, M. A., Menichi, B. & Holland, I. B. Evidence for post-transcriptional regulation of the synthesis of the *Escherichia coli* HlyB haemolysin translocator and production of polyclonal anti-HlyB antibody. *Mol Gen Genet* **247**, 73-85 (1995).

156 Jenewein, S. *The Escherichia coli haemolysin transporter - A paradigm for Type I secretion* Dr. rer. nat thesis, Heinrich-Heine-Universität Düsseldorf, (2008).

157 Sheps, J. A., Cheung, I. & Ling, V. Hemolysin transport in *Escherichia coli*. Point mutants in HlyB compensate for a deletion in the predicted amphiphilic helix region of the HlyA signal. *J Biol Chem* **270**, 14829-14834 (1995).

158 Zhang, F., Sheps, J. A. & Ling, V. Complementation of transport-deficient mutants of *Escherichia coli* alpha-hemolysin by second-site mutations in the transporter hemolysin B. *J Biol Chem* **268**, 19889-19895 (1993).

159 Wandersman, C. & Delepelaire, P. TolC, an *Escherichia coli* outer membrane protein required for hemolysin secretion. *Proc Natl Acad Sci U S A* **87**, 4776-4780 (1990).

160 Wang, R. C. *et al.* Analysis of the membrane organization of an *Escherichia coli* protein translocator, HlyB, a member of a large family of prokaryote and eukaryote surface transport proteins. *J Mol Biol* **217**, 441-454 (1991).

161 Gentschev, I. & Goebel, W. Topological and functional studies on HlyB of *Escherichia coli*. *Mol Gen Genet* **232**, 40-48 (1992).

162 Ishii, S. *et al.* Crystal structure of the peptidase domain of *Streptococcus* ComA, a bifunctional ATP-binding cassette transporter involved in the quorum-sensing pathway. *J Biol Chem* **285**, 10777-10785, doi:M109.093781 [pii] 10.1074/jbc.M109.093781 (2010).

163 Havarstein, L. S., Diep, D. B. & Nes, I. F. A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol* **16**, 229-240 (1995).

164 Wu, K. H. & Tai, P. C. Cys32 and His105 are the critical residues for the calcium-dependent cysteine proteolytic activity of CvaB, an ATP-binding cassette transporter. *J Biol Chem* **279**, 901-909, doi:10.1074/jbc.M308296200 M308296200 [pii] (2004).

165 Dirix, G. *et al.* Peptide signal molecules and bacteriocins in Gram-negative bacteria: a genome-wide in silico screening for peptides containing a double-glycine leader sequence and their cognate transporters. *Peptides* **25**, 1425-1440, doi:10.1016/j.peptides.2003.10.028 S0196-9781(04)00293-1 [pii] (2004).

166 Kotake, Y., Ishii, S., Yano, T., Katsuoka, Y. & Hayashi, H. Substrate recognition mechanism of the peptidase domain of the quorum-sensing-signal-producing ABC transporter ComA from Streptococcus. *Biochemistry* **47**, 2531-2538, doi:10.1021/bi702253n (2008).

167 Drenth, J., Jansonius, J. N., Koekoek, R. & Wolthers, B. G. The structure of papain. *Adv Protein Chem* **25**, 79-115 (1971).

168 Angelides, K. J. & Fink, A. L. Cryoenzymology of papain: reaction mechanism with an ester substrate. *Biochemistry* **17**, 2659-2668 (1978).

169 Mergulhao, F. J., Summers, D. K. & Monteiro, G. A. Recombinant protein secretion in *Escherichia coli. Biotechnol Adv* **23**, 177-202, doi:S0734-9750(04)00117-X [pii] 10.1016/j.biotechadv.2004.11.003 (2005).

170 Sandkvist, M. & Bagdasarian, M. Secretion of recombinant proteins by Gram-negative bacteria. *Curr Opin Biotechnol* **7**, 505-511, doi:S0958-1669(96)80053-X [pii] (1996).

171 Maskos, K., Huber-Wunderlich, M. & Glockshuber, R. DsbA and DsbC-catalyzed oxidative folding of proteins with complex disulfide bridge patterns in vitro and in vivo. *J Mol Biol* **325**, 495-513, doi:S0022283602012482 [pii] (2003).

172 Jeong, K. J. & Lee, S. Y. Secretory production of human granulocyte colony-stimulating factor in *Escherichia coli*. *Protein Expr Purif* **23**, 311-318, doi:10.1006/prep.2001.1508 S1046-5928(01)91508-6 [pii] (2001).

173 Loo, T., Patchett, M. L., Norris, G. E. & Lott, J. S. Using secretion to solve a solubility problem: high-yield expression in *Escherichia coli* and purification of the bacterial glycoamidase PNGase F. *Protein Expr Purif* **24**, 90-98, doi:10.1006/prep.2001.1555 S1046592801915554 [pii] (2002).

174 Francetic, O., Belin, D., Badaut, C. & Pugsley, A. P. Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion. *EMBO J* **19**, 6697-6703, doi:10.1093/emboj/19.24.6697 (2000).

175 Hellwig, S., Drossard, J., Twyman, R. M. & Fischer, R. Plant cell cultures for the production of recombinant proteins. *Nat Biotechnol* **22**, 1415-1422, doi:nbt1027 [pii] 10.1038/nbt1027 (2004).

176 Mackman, N. *et al.* Release of a chimeric protein into the medium from *Escherichia coli* using the C-terminal secretion signal of haemolysin. *EMBO J* **6**, 2835-2841 (1987).

177 Blight, M. A. & Holland, I. B. Heterologous protein secretion and the versatile *Escherichia coli* haemolysin translocator. *Trends Biotechnol* **12**, 450-455, doi:0167-7799(94)90020-5 [pii] (1994).

178 Nakano, H., Kawakami, Y. & Nishimura, H. Secretion of genetically-engineered dihydrofolate reductase from *Escherichia coli* using an *E. coli* alpha-hemolysin membrane translocation system. *Appl Microbiol Biotechnol* **37**, 765-771 (1992).

179 Debarbieux, L. & Wandersman, C. Folded HasA inhibits its own secretion through its ABC exporter. *EMBO J* **20**, 4657-4663, doi:10.1093/emboj/20.17.4657 (2001).

180 Gentschev, I. *et al.* Development of antigen-delivery systems, based on the *Escherichia coli* hemolysin secretion pathway. *Gene* **179**, 133-140, doi:S0378-1119(96)00424-6 [pii] (1996).

181 Bray, B. L. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat Rev Drug Discov* **2**, 587-593, doi:10.1038/nrd1133 nrd1133 [pii] (2003).

182 Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov Today* **15**, 40-56, doi:S1359-6446(09)00370-5 [pii] 10.1016/j.drudis.2009.10.009 (2010).

183 Wu, G. Q., Li, L. X., Ding, J. X., Wen, L. Z. & Shen, Z. L. High-level expression and novel purification strategy of recombinant thanatin analog in *Escherichia coli*. *Curr Microbiol* **57**, 95-101, doi:10.1007/s00284-008-9106-z (2008).

184 Macao, B. *et al.* Recombinant amyloid beta-peptide production by coexpression with an affibody ligand. *BMC Biotechnol* **8**, 82, doi:1472-6750-8-82 [pii] 10.1186/1472-6750-8-82 (2008).

185 Jenssen, H. & Aspmo, S. I. Serum stability of peptides. *Methods Mol Biol* **494**, 177-186, doi:10.1007/978-1-59745-419-3_10 (2008).

186 Andersons, D., Engstrom, A., Josephson, S., Hansson, L. & Steiner, H. Biologically active and amidated cecropin produced in a baculovirus expression system from a fusion construct containing the antibody-binding part of protein A. *Biochem J* **280 (Pt 1)**, 219-224 (1991).

187 Piers, K. L., Brown, M. H. & Hancock, R. E. Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. *Gene* **134**, 7-13, doi:0378-1119(93)90168-3 [pii] (1993).

188 Kimmerlin, T. & Seebach, D. '100 years of peptide synthesis': ligation methods for peptide and protein synthesis with applications to beta-peptide assemblies. *J Pept Res* **65**, 229-260, doi:JPP214 [pii]
10.1111/j.1399-3011.2005.00214.x (2005).

189 Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 6, doi:10.1021/ja00897a025 (1963).

190 Pattabiraman, V. R. & Bode, J. W. Rethinking amide bond synthesis. *Nature* **480**, 471-479, doi:nature10702 [pii]
10.1038/nature10702 (2011).

191 Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776-779 (1994).

192 Torbeev, V. Y. & Kent, S. B. Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. *Angew Chem Int Ed Engl* **46**, 1667-1670, doi:10.1002/anie.200604087 (2007).

193 Marx, V. Roche's fuzeon challenge. *Chem. Eng. News* **84**, 2 (2005).

194 Constable, D. J. *et al.* Key green chemistry research areas - a perspective from pharmaceutical manufacturers. *Green Chemistry* **9**, 10 (2007).

195 Yoon, S. H., Kim, S. K. & Kim, J. F. Secretory production of recombinant proteins in Escherichia coli. *Recent Pat Biotechnol* **4**, 23-29, doi:BIOT- ABSTRACT 10 [pii] (2010).

196 Nishie, M., Shioya, K., Nagao, J., Jikuya, H. & Sonomoto, K. ATP-dependent leader peptide cleavage by NukT, a bifunctional ABC transporter, during lantibiotic biosynthesis. *J Biosci Bioeng* **108**, 460-464, doi:S1389-1723(09)00280-1 [pii]
10.1016/j.jbiosc.2009.06.002 (2009).

197 Nishie, M. *et al.* Lantibiotic transporter requires cooperative functioning of the peptidase domain and the ATP binding domain. *J Biol Chem* **286**, 11163-11169, doi:M110.212704 [pii]
10.1074/jbc.M110.212704 (2011).

198 Kadaba, N. S., Kaiser, J. T., Johnson, E., Lee, A. & Rees, D. C. The high-affinity *E. coli* methionine ABC transporter: structure and allosteric regulation. *Science* **321**, 250-253, doi:321/5886/250 [pii]
10.1126/science.1157987 (2008).

199 Oldham, M. L., Khare, D., Quiocho, F. A., Davidson, A. L. & Chen, J. Crystal structure of a catalytic intermediate of the maltose transporter. *Nature* **450**, 515-521, doi:nature06264 [pii]
10.1038/nature06264 (2007).

200 Dawson, R. J. & Locher, K. P. Structure of the multidrug ABC transporter Sav1866 from *Staphylococcus aureus* in complex with AMP-PNP. *FEBS Lett* **581**, 935-938, doi:S0014-5793(07)00122-6 [pii]
10.1016/j.febslet.2007.01.073 (2007).

201 Hvorup, R. N. *et al.* Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. *Science* **317**, 1387-1390, doi:1145950 [pii]
10.1126/science.1145950 (2007).

202 Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315**, 373-377, doi:1133488 [pii]
10.1126/science.1133488 (2007).

203 Aller, S. G. *et al.* Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science* **323**, 1718-1722, doi:323/5922/1718 [pii]
10.1126/science.1168750 (2009).

204 Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, 19005-19010, doi:0709388104 [pii]
10.1073/pnas.0709388104 (2007).

205 Korkhov, V. M., Mireku, S. A., Hvorup, R. N. & Locher, K. P. Asymmetric states of vitamin B12 transporter BtuCD are not discriminated by its cognate substrate binding protein BtuF. *FEBS Lett* **586**, 972-976, doi:S0014-5793(12)00165-2 [pii]
10.1016/j.febslet.2012.02.042 (2012).

206 Letoffe, S., Ghigo, J. M. & Wandersman, C. Secretion of the *Serratia marcescens* HasA protein by an ABC transporter. *J Bacteriol* **176**, 5372-5377 (1994).

207 Cescau, S., Debarbieux, L. & Wandersman, C. Probing the in vivo dynamics of type I protein secretion complex association through sensitivity to detergents. *J Bacteriol* **189**, 1496-1504, doi:JB.01480-06 [pii]
10.1128/JB.01480-06 (2007).
208 Delepelaire, P. PrtD, the integral membrane ATP-binding cassette component of the *Erwinia chrysanthemi* metalloprotease secretion system, exhibits a secretion signal-regulated ATPase activity. *J Biol Chem* **269**, 27952-27957 (1994).
209 Tschapek, B. *Biophysical characterization of ABC transporters catalyzing import and export function* Dr. rer. nat. thesis, Heinreich-Heine-Universität, (2011).
210 Herget, M. *et al.* Purification and reconstitution of the antigen transport complex TAP: a prerequisite for determination of peptide stoichiometry and ATP hydrolysis. *J Biol Chem* **284**, 33740-33749, doi:M109.047779 [pii]
10.1074/jbc.M109.047779 (2009).
211 Ambudkar, S. V. *et al.* Partial purification and reconstitution of the human multidrug-resistance pump: characterization of the drug-stimulatable ATP hydrolysis. *Proc Natl Acad Sci U S A* **89**, 8472-8476 (1992).
212 Davidson, A. L. & Nikaido, H. Purification and characterization of the membrane-associated components of the maltose transport system from *Escherichia coli*. *J Biol Chem* **266**, 8946-8951 (1991).
213 van Veen, H. W., Margolles, A., Muller, M., Higgins, C. F. & Konings, W. N. The homodimeric ATP-binding cassette transporter LmrA mediates multidrug transport by an alternating two-site (two-cylinder engine) mechanism. *EMBO J* **19**, 2503-2514, doi:10.1093/emboj/19.11.2503 (2000).
214 Holland, I. B., Kenny, B., Steipe, B. & Pluckthun, A. Secretion of heterologous proteins in *Escherichia coli*. *Methods Enzymol* **182**, 132-143 (1990).
215 Schwarz, C. K., Landsberg, C. D., Lenders, M. H., Smits, S. H. & Schmitt, L. Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form. *J Biotechnol* **159**, 155-161, doi:S0168-1656(12)00127-7 [pii]
10.1016/j.jbiotec.2012.02.005 (2012).
216 Uchida, K., Mori, H. & Mizushima, S. Stepwise movement of preproteins in the process of translocation across the cytoplasmic membrane of *Escherichia coli*. *J Biol Chem* **270**, 30862-30868 (1995).
217 Schiebel, E., Driessen, A. J., Hartl, F. U. & Wickner, W. Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64**, 927-939, doi:0092-8674(91)90317-R [pii] (1991).
218 Omote, H. & Al-Shawi, M. K. A novel electron paramagnetic resonance approach to determine the mechanism of drug transport by P-glycoprotein. *J Biol Chem* **277**, 45688-45694, doi:10.1074/jbc.M206479200
M206479200 [pii] (2002).
219 Ernst, R. *et al.* A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proc Natl Acad Sci U S A* **105**, 5069-5074, doi:0800191105 [pii]
10.1073/pnas.0800191105 (2008).
220 Balakrishnan, L., Venter, H., Shilling, R. A. & van Veen, H. W. Reversible transport by the ATP-binding cassette multidrug export pump LmrA: ATP synthesis at the expense of downhill ethidium uptake. *J Biol Chem* **279**, 11273-11280, doi:10.1074/jbc.M308494200
M308494200 [pii] (2004).
221 Venter, H., Shilling, R. A., Velamakanni, S., Balakrishnan, L. & Van Veen, H. W. An ABC transporter with a secondary-active multidrug translocator domain. *Nature* **426**, 866-870, doi:10.1038/nature02173
nature02173 [pii] (2003).
222 van den Berg van Saparoea, H. B., Lubelski, J., van Merkerk, R., Mazurkiewicz, P. S. & Driessen, A. J. Proton motive force-dependent Hoechst 33342 transport by the ABC transporter LmrA of *Lactococcus lactis*. *Biochemistry* **44**, 16931-16938, doi:10.1021/bi051497y (2005).
223 Beis, I. & Newsholme, E. A. The contents of adenine nucleotides, phosphagens and some glycolytic intermediates in resting muscles from vertebrates and invertebrates. *Biochem J* **152**, 23-32 (1975).
224 Williams, S. P., Fulton, A. M. & Brindle, K. M. Estimation of the intracellular free ADP concentration by 19F NMR studies of fluorine-labeled yeast phosphoglycerate kinase in vivo. *Biochemistry* **32**, 4895-4902 (1993).

225 Jorgensen, P. L., Hakansson, K. O. & Karlish, S. J. Structure and mechanism of Na,K-ATPase: functional sites and their interactions. *Annu Rev Physiol* **65**, 817-849, doi:10.1146/annurev.physiol.65.092101.142558 092101.142558 [pii] (2003).
226 Garrahan, P. J. & Glynn, I. M. The incorporation of inorganic phosphate into adenosine triphosphate by reversal of the sodium pump. *J Physiol* **192**, 237-256 (1967).
227 Lew, V. L., Glynn, I. M. & Ellory, J. C. Net synthesis of ATP by reversal of the sodium pump. *Nature* **225**, 865-866 (1970).
228 Rouslin, W., Erickson, J. L. & Solaro, R. J. Effects of oligomycin and acidosis on rates of ATP depletion in ischemic heart muscle. *Am J Physiol* **250**, H503-508 (1986).
229 Chinopoulos, C. *et al.* Forward operation of adenine nucleotide translocase during F0F1-ATPase reversal: critical role of matrix substrate-level phosphorylation. *FASEB J* **24**, 2405-2416, doi:fj.09-149898 [pii]10.1096/fj.09-149898 (2010).
230 Chen, C. Y. *et al.* Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome Res* **13**, 2577-2587, doi:10.1101/gr.1295503 13/12/2577 [pii] (2003).
231 Meier, R., Drepper, T., Svensson, V., Jaeger, K. E. & Baumann, U. A calcium-gated lid and a large beta-roll sandwich are revealed by the crystal structure of extracellular lipase from *Serratia marcescens*. *J Biol Chem* **282**, 31477-31483, doi:M704942200 [pii] 10.1074/jbc.M704942200 (2007).
232 Soloaga, A., Ramirez, J. M. & Goni, F. M. Reversible denaturation, self-aggregation, and membrane activity of *Escherichia coli* alpha-hemolysin, a protein stable in 6 M urea. *Biochemistry* **37**, 6387-6393, doi:10.1021/bi9730994 bi9730994 [pii] (1998).
233 Pimenta, A. L., Racher, K., Jamieson, L., Blight, M. A. & Holland, I. B. Mutations in HlyD, part of the type 1 translocator for hemolysin secretion, affect the folding of the secreted toxin. *J Bacteriol* **187**, 7471-7480, doi:187/21/7471 [pii] 10.1128/JB.187.21.7471-7480.2005 (2005).
234 Sanchez-Magraner, L. *et al.* The calcium-binding C-terminal domain of *Escherichia coli* alpha-hemolysin is a major determinant in the surface-active properties of the protein. *J Biol Chem* **282**, 11827-11835, doi:M700547200 [pii] 10.1074/jbc.M700547200 (2007).
235 Sanchez-Magraner, L. *et al.* Interdomain Ca(2+) effects in *Escherichia coli* alpha-haemolysin: Ca(2+) binding to the C-terminal domain stabilizes both C- and N-terminal domains. *Biochim Biophys Acta* **1798**, 1225-1233, doi:S0005-2736(10)00104-5 [pii] 10.1016/j.bbamem.2010.03.007 (2010).
236 Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663-669, doi:nature06384 [pii] 10.1038/nature06384 (2007).
237 Renn, J. P., Junker, M., Besingi, R. N., Braselmann, E. & Clark, P. L. ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. *Chem Biol* **19**, 287-296, doi:S1074-5521(11)00440-6 [pii] 10.1016/j.chembiol.2011.11.009 (2012).
238 Chang, C. F., Shuman, H. & Somlyo, A. P. Electron probe analysis, X-ray mapping, and electron energy-loss spectroscopy of calcium, magnesium, and monovalent ions in log-phase and in dividing *Escherichia coli* B cells. *J Bacteriol* **167**, 935-939 (1986).
239 Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004, doi:S0092-8674(06)01013-0 [pii] 10.1016/j.cell.2006.07.025 (2006).
240 Armstrong, C. M. & Goldberg, D. E. An FKBP destabilization domain modulates protein levels in Plasmodium falciparum. *Nat Methods* **4**, 1007-1009, doi:nmeth1132 [pii] 10.1038/nmeth1132 (2007).
241 Herm-Gotz, A. *et al.* Rapid control of protein level in the apicomplexan Toxoplasma gondii. *Nat Methods* **4**, 1003-1005, doi:nmeth1134 [pii] 10.1038/nmeth1134 (2007).

8 Acknowledgment/Danksagung

Mein erster Dank gilt Lutz. Für die tolle Doktorandenzeit, für die unglaublichen Arbeitsbedingungen, für die Unterstützung bei meiner Arbeit, für Ideen, Einsatz und Hingabe. Trotz Deines vollen Terminkalenders bist Du immer für mich da. Danke.

Außerdem danke ich Sander. Für seine Betreuung, für seine Unterstützung, für zahlreiche Abende am Computer, für meine Kaffeesucht. ☺ Danke für alles.

Außerdem danke ich meinen traumhaften Arbeitskollegen, die einen jederzeit mit Rat und Tat unterstützt haben. Hervorheben möchte ich Andre, Britta und Philipp. Ohne Euch könnte ich nicht Äkta, HPLC oder lineares Plasmid (und natürlich vieles mehr). ☺ Es wird mir immer eine Freude sein mit Euch Wissenschaft zu betreiben. Außerdem danke ich Euch für den gemeinsamen Kaffeeklatsch, die Mittagspausen, Grillabende, Karaoke-Shows, Hochzeiten…

Uli, danke für die grandiose Ausbildung während des Studiums. Danke Stefan, Deine Arbeit hat mich stets motiviert. Thorsten, Du hast mir gezeigt wo der Hase im Institut lang läuft, danke dafür. Danke Nils, für Wissen ohne Blättern. Danke Jan und Ricarda, ohne Euch wären viele Tage langweilig gewesen.
Für die Betreuung meiner Bachelorarbeit bedanke ich mich bei Prof. Dr. C. Seidel und insbesondere bei Heike. Ich danke den Damen an der DNA-Front Diana und Iris. In diesem Zusammenhang sollte auch einmal In-Fusion Cloning fallen; geiles Zeug. Nicht zu vergessen sind Isa, Martina, Silke, Frau Rasid und Frau Blum; danke für all die Hilfe, die häufig übersehen wird. Vielen Dank Fr. Simons, ich bin ja eigentlich schlecht in Zettelwirtschaft.
Danke an alle anderen, die das Beste für mich wollen.

Ich möchte mich außerdem bei meinen ehemaligen Studenten Michael, Antonia und Tine bedanken. Es hat mir sehr viel Spaß gemacht Euch zu betreuen. Ich wünsche Euch weiterhin viel Glück und Erfolg.

Vielen Dank Justin, Matthias, Dieter, Mirko, Khana, Ümit, Reine, Alex, Filip, Susanne und Prof. Jaeger für die gute Zusammenarbeit.
Thank you Barry. For your help, your enthusiasm and your amazing input.

Danke an das Foschungsmanagement der HHU und die DIWA.

Mein Dank gilt dem Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, sowie der Heinrich-Heine-Universität Düsseldorf für das Stipendium innerhalb des Graduiertenclusters für industrielle Biotechnologie (CLIB).
Danke Steffi, Janpeter und Alex für die gemeinsamen Tage.

Danke an alle, die meine Dissertation gegengelesen haben.

Obgleich diese Arbeit der Wissenschaft gewidmet ist, wurde sie erst durch viele Menschen ermöglicht, die nicht wissen wie "man eine Kultur schüttelt", warum eine DNA kein Protein ist oder weshalb eine Eismaschine im Labor steht (wieso kennt jemand eigentlich unsere Eismaschine?!). Auch bei diesen Menschen möchte ich mich ausdrücklich bedanken.

Zunächst möchte ich meinen Vater hervorheben. Vielen Dank für Deine Unterstützung und dafür, dass Du auf so vieles verzichtet hast.

Außerdem möchte ich meiner Schwester danken. Besonders für die Urlaubstage im Düsseldorfer Zoo. ☺

Meiner Schwester Katharina, meinem Schwager mit dem besonderen Namen, Felix mit dem besonderen Zweitnamen und Lea (ja wofür eigentlich schon? ☺) danke ich für viele lustige Momente und wünsche Euch alles Glück und noch mehr Gesundheit für die Zukunft.

Zu ganz besonderem Dank bin ich Marian gegenüber verpflichtet, ohne den ich heute den ehrenwerten Beruf des Schreiners ausüben würde. Es wären aber fantastische Tische, Schränke etc. geworden. Hätte er mich vielleicht etwas seltener getunnelt, wer weiß, vielleicht würde ich auch noch der Fußballkarriere nacheifern. Am meisten danke ich Dir aber für mehr als 15 Jahre beste Freundschaft.
In diesem Zusammenhang danke ich auch Lea, Pu und Ecki, die mich immer wie ein Familienmitglied behandelt haben. Außerdem danke ich Lucy für die gemeinsame Zeit in den letzten Jahren.

Ich danke meinen ehemaligen Fußballkollegen. Danke Sportring, danke Britannia. Es war eine super Zeit. Danke Karl, wir geben immer alles für Dich.

Ich möchte Jogi für die Studienzeit und die Zeit danach danken. Auch in Zukunft wird es mir mit niemandem mehr Spaß machen über Wissenschaft zu debattieren und zeitgleich verköstigt zu werden. ☺ Ich wünsche Dir, Nina und Deiner ganzen Familie nur das Beste von allem! Danke.

Ich danke Silja. Für zahlreiche Fachgespräche, für eine tolle Zeit in Düsseldorf, für Deine Freundschaft. Rückwirkend betrachtet ist wohl Dein größter Verdienst die etablierte Doppelkopfrunde (wenn auch mit falscher Intention ☺). Die ursprüngliche Melkkuh dankt den weiteren Teilnehmern Kathy und Christian, auch für alle anderen gemeinsamen Unternehmungen und die Freundschaft.

Inga, Du machst mich verrückt und motivierst mich. Du hörst mir zu und hältst mich aus. Danke für alles, Du bist eine mega Braut.

9 Curriculum vitae

E-MAIL: CHRISTIAN.SCHWARZ@HHU.DE

CHRISTIAN K. W. SCHWARZ

DATE OF BIRTH: 30TH JANUARY 1984
NATIONALITY: GERMAN

PROFILE

- Have gained excellent grades during study and academic apprenticeship, now applying this in a fast-track PhD program.
- Extensive experiences in R&D by doctoral research, internships and as scientific assistant.
- Several publications in scientific journals and presentations at scientific conferences.

EDUCATION

Oct 2008 - present	**Fast-track PhD programme**, Heinrich Heine University Duesseldorf
Oct 2005 – Sep 2008	**BSc Biochemistry**, Heinreich Heine University Duesseldorf Overall grade: 1.4
Sep 2003 – Jul 2005	**State examined Biological Technical Assistant**, Berufskolleg Hilden Overall grade: 1.4
Sep 1994 - Jun 2003	**Abitur**, August-Dicke-Gymnasium Solingen Overall grade: 3.1

WORK EXPERIENCE

1st Aug 2009 - present	**PhD program, Institute for Biochemistry, Heinrich Heine University Duesseldorf**
23th Jun 2008 – 31th August 2008	**Heinrich Heine University, Duesseldorf** **Institute of Molecular Physical Chemistry** *Scientific assistant*
6th Aug 2007 – 16th Oct 2007	**University of Michigan, Ann Arbor** *Internship*
4th Jul 2005 – 30th Sep 2005	**Henkel KgaA, Duesseldorf, Skin/Hair Physiology** *Internship*
30th Aug 2004 – 22th Oct 2004	**Microscience Technologies Ltd., Edinburgh** *Internship*

PUBLICATIONS

Secretion of slow-folding proteins by a Type 1 Secretion System; C. K. W. Schwarz, M. H. H. Lenders, S. H. J. Smit and L. Schmitt; Bioengineered Bugs 3, (2012).

Using an E. coli Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form; C. K. W. Schwarz, C. D. Landsberg, M. H. Lenders, S. H. J. Smits and L. Schmitt, J Biotechnol; (2012); 159, 155-161

Rational and irrational approaches to convince a protein to crystallize; A. Abts, C. K. W. Schwarz, B. Tschapek, S. H. J. Smits and L. Schmitt; Intech; (2012); ISBN 978-953-307-610-2

Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the Escherichia coli ABC transporter haemolysin B; C. K. W. Schwarz, B. Tschapek, T. Jumpertz, S. Jenewein, J. Lecher, D. Willbold, S. Panjikar, I. B. Holland, S. H. J. Smits and L. Schmitt; Acta Crystallogr Sect F Struct Biol Cryst Commun; (2011); F67, 630–633

^{1}H, ^{15}N and ^{13}C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB); J. Lecher, M. Stoldt, C. K. W. Schwarz, S. H. J. Smits, L. Schmitt and D. Willbold, Biomol NMR Assign; (2011); 3, 199-201

PATENT APPLICATION

21th Oct 2011 — **Agents and Methods for the Expression and Secretion of Peptides and Proteins; C. K. W. Schwarz, L. Schmitt, S. H. J. Smits, European patent application (2011). EP11186225.6.**

10 List of publications

1. Lecher, J.*, Schwarz, C. K. W.*, Stoldt, M., Smits, S. H. J., Willbold, D., Schmitt L. RTX toxin transporters tether its substrate prior to secretion via the unique function of its N-terminal domain. *Structure* (submitted, 2012).

* equal contribution

2. Schwarz, C. K. W., Lenders M. H., Smits S. H. J., Schmitt L. Secretion of slow-folding proteins by a Type 1 secretion system. *Bioengineered* **3**, doi:20712 [pii] (2012).

3. Schwarz, C. K. W., Landsberg, C. D., Lenders, M. H., Smits, S. H., Schmitt, L. Using an *E. coli* Type 1 secretion system to secrete the mammalian, intracellular protein IFABP in its active form. *J Biotechnol* **159**, 155-161, doi:S0168-1656(12)00127-7 [pii] 10.1016/j.jbiotec.2012.02.005 (2012).

4. Abts, A., Schwarz, C. K. W., Tschapek, B., Smits, S. H., Schmitt, L. in *Modern Aspects of Bulk Crystal and Thin Film Preparation*, editors N. Kolesnikov & E. Borisenko, Ch. 22, 497-528, InTech, ISBN 978-953-307-610-2, (2012).

5. Lecher J., Stoldt M., Schwarz C. K. W., Smits S. H. J., Schmitt L., Willbold D. ^{1}H, ^{15}N and ^{13}C resonance assignment of the N-terminal C39 peptidase-like domain of the ABC transporter Haemolysin B (HlyB). *Biomol NMR Assign* **5**, 199-201, doi:10.1007/s12104-011-9299-0 (2011).

6. Schwarz C. K. W., Tschapek B., Jumpertz T., Jenewein S., Lecher J., Willbold D., Panjikar S., Holland I. B., Smits S. H. J., Schmitt L. Crystallization and preliminary X-ray crystallographic studies of an oligomeric species of a refolded C39 peptidase-like domain of the *Escherichia coli* ABC transporter haemolysin B. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 630-633, doi:S1744309111010876 [pii]10.1107/S1744309111010876 (2011).

11 List of patent applications

1. Schwarz, C. K. W., Schmitt, L., Smits, S. H. J. Agents and Methods for the Expression and Secretion of Peptides and Proteins. European patent application (2011). EP11186225.6.

12 Participation at scientific conferences

1. CLIB International Conference 2012, CLIB2021 e. V., 23.04.2012 – 24.04.2012, Düsseldorf, Germany

2. Molecular Life Sciences 2011, GBM, 24.09.2011 – 28.09.2011, Frankfurt am Main, Germany

3. 61st Lindau Nobel Laureate Meeting, 26.06.2011 – 01.07.2011, Lindau am Bodensee, Germany

4. CLIB International Conference 2011, CLIB2021 e. V., 05.04.2011 – 06.04.2011, Düsseldorf, Germany

> *Oral Presentation*: Type I secretion as a transport vehicle to express large amounts of biotechnologically interesting proteins
>
> *Poster Presentation*: Type I secretion as a transport vehicle to express large amounts of biotechnologically interesting proteins, Christian K. W. Schwarz, P. J. Bakkes, S. Jenewein, S. H. Smits, I. B. Holland, L. Schmitt

5. The Modern RNA World, Murnau Conference on Structural Biology 2010, 12.10.2010 – 16.10.2010, Murnau, Germany

> *Poster Presentation*: Expression, Purification and Structural Studies of CRISPR associated protein Csn2 from *Streptococcus agalactiae,* P. Ellinger, C. K. W. Schwarz, S. Zobel, R. Wurm, Ü. Pul, R. Wagner, S. Smits, L. Schmitt

6. Transport across membranes: Multiple drug resistance, mechanisms and new tools, 462nd WE Heraeus Seminar, 04.07.2010 – 10.07.2010, Bremen, Germany

> *Poster Presentation*: The ABC transporter HlyB and its C39 peptidase-like domain – an appendix without function?, Christian K. W. Schwarz, J. Lecher,T. Jumpertz, M. Stoldt, D. Willbold, S. H. J. Smits and L. Schmitt

7. CLIB International Conference 2010, CLIB2021 e. V., 02.04.2010 – 05.04.2010, Düsseldorf, Germany

> *Poster Presentation*: Rational design of Type I secretion substrates, Christian K. W. Schwarz, S. Smits, L. Schmitt

8. Protein Transport Across Cell Membranes, Gordon Research Conference, 07.03.2010 – 12.03.2010, Galveston, Texas, USA

Poster Presentation: The ABC transporter HlyB and its C39 peptidase-like domain – an appendix without function?, Christian K. W. Schwarz, J. Lecher, T. Jumpertz, M. Stoldt, D. Willbold, S. H. J. Smits and L. Schmitt

13 Erklärung zur Promotion

Die hier vorliegende Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Christian K. W. Schwarz

Düsseldorf, den 03.08.2012