Assisted Secretion of a Trimeric Autotransporter Adhesin from *Salmonella*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Iwan Grin

aus Lwiw / Ukraine

Tübingen

2014

Tag der mündlichen Prüfung: 29. 04. 2015

Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter	Prof. Dr. Dirk Linke
2. Berichterstatter	Prof. Dr. Doron Rapaport

Table of Contents

Figures

Abbreviations

2D, 3D	two dimensional, three dimensional
Å	Angström
ABC	ATP-binding cassette
ATP	Adenosine triphosphate
Bam	β-barrel assembly machinery
C-terminus	carboxy terminus
CU	Chaperone-Usher
DNA	desoxyribonucleic acid
E. coli	*Escherichia coli*
GSP	general secretory pathway
IM	inner membrane
MATH	Meprin and TRAF homology
MFP	membrane fusion protein
N-terminus	amino terminus
OM	outer membrane
OMP	outer membrane protein
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
r.m.s.d	root mean square deviation
RNA	ribonucleic acid
SDS	sodium docecyl sulfate
SLS	Swiss Light Source
SRP	signal recognition particle
T(1-9)SS	Type (I-IX) secretion system
TAA	trimeric autotransporter adhesin
Tat	Two arginine translocation
TF	trigger factor
TNF	tumor necrosis factor
TPSS	Two-partner secretion system
TRAF2	TNF receptor associated factor 2

1 Summary

Type Vc secretion systems, also known as Trimeric Autotransporter Adhesins (TAAs) are important virulence factors of Gram-negative bacteria. This subclass of bacterial autotransporters forms obligate homotrimers on the surface of bacterial cells, anchored in the outer membrane by the translocator domain at the C-terminus of the protein, through which the rest of the polypeptide is threaded during their biogenesis. The mechanism of this autotransport has been a matter of some debate over the last two decades.

During our investigation of the *Salmonella* adhesin SadA we have discovered that it forms an operon with a small predicted lipoprotein, which we named SadB. The operon is conserved in *Salmonella*, *Shigella* and *Escherichia*. Structure and composition of operons can only be analyzed by looking at the operon as a whole, and not just comparing the constituent gene or protein sequences. Not satisfied by the functionality of bioinformatics tools and databases available for this purpose, I have developed GCView - a genomic context viewer for protein homology searches. Starting from a protein sequence, GCView runs a PSI-BLAST search, retrieves the genomic context for each of the hits, and displays the results as a clear, interactive graphical output. Additional queries can be added to the result to iteratively refine the output. Results can be mapped onto a taxonomy tree to survey the distribution of genes or operons of interest in different organisms; or grouped by the number and order of hits contained therein, allowing to quickly gain an overview of the compositions and structures of operons occurring in sequenced genomes. GCView is embedded in the Max-Planck Institute Bioinformatics Toolkit (MPI Toolkit). Within the MPI Toolkit, results from GCView can be forwarded to other tools for further analysis. Results from other homology search tools within the MPI toolkit can also be forwarded to GCView, allowing for complete flexibility and transparency in the search process. GCView is freely available at http://toolkit.tuebingen.mpg.de/gcview.

As autotransporters are not typically associated with lipoproteins, I was interested in the function of SadB and its interactions with SadA. In this work, I confirmed the prediction that SadB is indeed a lipoprotein, attached to the periplasmic side of the inner membrane. I was able to demonstrate that co-expression of SadB with SadA leads to increased quantity of SadA on the bacterial cell surface, as well as improved folding of the adhesin, which is evidenced by increased protease resistance, as compared to the expression of SadA alone. Additionally, I was able to purify a soluble variant of the protein and obtain a crystal structure with a resolution of 2.45Å. The crystal structure shows, that SadB forms a homotrimer, just as SadA. A long N-terminal coiled coil holds the trimer together and projects the Ig-like C-terminal domain away from the inner membrane. The C-terminal domain consists of two antiparallel β-sheets in a β-sandwich topology. This overall structure is similar to eukaryotic Meprin and TRAF homology (MATH) domains, as found e.g in the TNF receptor associated factor 2 (TRAF2) protein. Based on the mode of action of TRAF2, which preferentially binds peptides from trimeric TNF receptor over the monomeric form, I propose that SadB transiently binds three SadA polypeptides during their passage through the periplasm, and supports the correct export of their passenger domains.

2 Introduction

There is an old joke among biologists that goes something like this: “Support microbes – they are the only culture some people have”. But it is, in fact, the microbes in our bodies that support our health and wellbeing in our daily lives. Only in recent years have we started to understand and appreciate the importance and complexity of the human microbiota. Be it in the gut, in the mouth or on the skin, everywhere in our bodies we find distinct and well balanced microbial communities. Disturbances of this balance have been linked to various conditions related to health ranging from obesity to asthma (Redinbo 2014; Nwosu et al. 2014; Moran & Shanahan 2014).

To uphold the balance between themselves and the host, but also among their peers, bacteria must at all times be able to sense their surroundings and to exchange signals with their neighbors. These functions, as well as nutrient uptake, locomotion, and adhesion, are localized at the bacterial cell envelope. In Gram-positive bacteria, the cell envelope consists of a plasma membrane containing phospholipids and a thick peptidoglycan layer. In Gram-negative bacteria the peptidoglycan layer is much thinner, and the plasma membrane corresponds to the inner membrane (IM). In addition to that, Gram-negative bacteria also have an outer membrane (OM), which is an asymmetric bilayer consisting of phospholipids in the inner leaflet and lipopolysaccharides in the outer leaflet. All of these layers have to be overcome in order to get information or molecules from the extracellular space into the cytoplasm, where they are processed and a response is generated. The response, often in the form of proteins which are secreted to the cell surface or the extracellular space, again has to be transduced through the cell envelope to reach its destination. To be able to export signal molecules, toxins, proteins, but also DNA and RNA in a regulated and orderly fashion, bacteria have evolved a plethora of translocation and secretion systems. The multitude of unique secretion mechanisms, complex compositions and structures observed in secretion systems suggests that multiple such systems arose independently during evolution, each filling a particular niche.

2.1 The Sec, Tat and Bam Machineries – Membrane protein housekeeping

The focus of this work will be on Gram-negative bacteria, with Gram-positive secretion systems mentioned where appropriate. In Gram-negative bacteria, as described above, exported molecules have to cross two major barriers: the inner and outer membranes. For general translocation of proteins across and insertion into the IM two main mechanisms have been described: The Sec translocon and the twin arginine translocation (Tat) pathway.

The Sec system is the general system for translocation of unfolded proteins across the IM (Lycklama A Nijeholt & Driessen 2012). During translation at the ribosome, a N-terminal signal peptide is recognized by either the signal recognition particle (SRP) or the trigger factor (TF), depending on its sequence (von Heijne 1990). The translation complex is transferred to the SecYEG translocon in the IM. There, if the signal peptide was bound by the SRP, the nascent protein is co-translationally inserted into the IM, and no signal peptide is cleaved. TF-bound proteins, on the other hand, take a different path through the Sec machinery and are translocated into the periplasm post-translationally. Upon completion of translocation the signal peptide is cleaved off by a

specialized signal peptidase (Leader peptidase or type I signal peptidase) (Paetzel 2014).

Lipoproteins, i.e. proteins with covalently attached lipid moieties, also use this pathway. They carry a slightly different signal peptide, which has a cysteine residue at the +1 position after the signal peptidase cleavage site. After translocation, a tri-palmityl lipid anchor is attached to that cysteine, anchoring the protein to the membrane, and the signal peptide is cleaved off by a lipoprotein signal peptidase (Type II signal peptidase) (Zückert 2014).

Folded proteins, protein-cofactor, and protein-protein complexes are exported by the twin arginine translocation pathway (Fröbel et al. 2012). Once again, a N-terminal signal sequence, similar to the Sec translocation signal, but containing the Tat consensus motif S-R-R-x-F-L-K, must be present for targeting of the substrate protein to the Tat translocation complex. The arginine moieties are essential for efficient recognition and translocation of the substrate, hence the name of the pathway. The exact mechanism of translocation of folded substrates of different shapes and sizes by the Tat translocase remains unclear. Similar to the Sec system the signal peptide is cleaved off after translocation and the substrate is released into the periplasm or lipid-modified and retained at the membrane.

In the OM the situation is very different. While the IM must maintain a tight seal to prevent leakage of ions and uphold the gradient of the proton motive force, the OM is permeable for ions and small molecules through outer membrane β-barrel proteins (OMPs). Additionally, no easily accessible energy source is available, as no nucleoside triphosphates are present in the periplasm. No general systems for protein translocation, akin to those in the IM, have been described for the OM. The insertion of OMPs into the membrane bilayer is catalyzed by the β-barrel assembly machinery (Bam) complex (Sauri et al. 2009; Selkrig et al. 2014), although exceptions may exist, e.g. see (Selkrig et al. 2012).

OMPs are synthesized in the cytoplasm and translocated into the periplasm by the Sec machinery. There, periplasmic chaperones such as SurA, Skp, FkpA, and the chaperone/protease DegP prevent premature (mis-)folding of the polypeptide while it makes its way to the outer membrane. At the OM a C-terminal motif is recognized by the Bam complex (Robert et al. 2006; Paramasivam et al. 2012) and the substrate protein is inserted into the membrane. The Bam complex of *E. coli* consists of BamA, a 16-stranded outer membrane β-barrel, and the accessory lipoproteins BamBCDE. Complex compositions vary in other organisms. BamA and BamD are essential for cell viability in *E. coli*, while deletions of *bamB*, *bamC* and *bamE* merely reduce the efficiency of OMP assembly and insertion (Malinverni & Silhavy 2011).

2.2 Secretion Systems

The systems described above mainly function in general housekeeping of the bacterial cell, to target proteins to the membranes or the periplasmic space. For efficient delivery of substrates to targets outside the bacterial cell more specialized secretion systems have evolved.

Secretion systems are large, macromolecular machines which create a conduit through one, or both membranes of Gram-negative bacteria. They facilitate the passage of small molecules, proteins, DNA, or even DNA-protein complexes from the cytoplasm of a bacterial cell to the extracellular

space or directly into another cell. Several distinct secretion system types have been described, grouped by structural and functional characteristics. They have traditionally been designated by roman numerals. At present, secretion systems of type I through type IX are described in literature, as well as the chaperone-usher (CU) system for assembly of pili on the surface of Gram-negative bacteria (Kuhn 2014; Waksman 2012). It should be noted that several of these secretion systems utilize the Sec translocon for crossing the IM and at least type V secretion is dependent on the Bam system for OM insertion.

In the following section a brief overview over each of the secretion systems will be given, highlighting their specific functional and structural properties, as well as the functional niche they occupy.

2.2.1 The big guns – Type I, III, IV, VI, VII and IX Secretion Systems

Secretion systems of types I, III, IV, VI, VII, and IX do not depend on the Sec translocon for secretion of substrates as they extend from the cytoplasm to the extracellular space themselves. However, assembly of these secretion systems often requires the Sec machinery.

Type I secretion systems (T1SS) were first discovered in uropathogenic *E. coli* strains and consist of three components: an ATP-binding cassette (ABC) transporter, which is an integral IM protein; a “membrane fusion” protein (MFP), spanning the periplasmic space, which is also anchored in the IM; and lastly an outer membrane protein, which interacts with the periplasmic part of the MFP. Type I substrates are characterized by a C-terminal signal which is not cleaved upon secretion. Substrates are recognized by the ABC transporter, which also provides the energy for the transport process from ATP hydrolysis, and secreted across both membranes in one step. Based on the channel diameter it is believed that substrates are not fully folded during transfer. Substrates include cytotoxins like hemolysin, bacteriocins, cell surface layer proteins, proteases, and lipases (Thomas et al. 2014; Long et al. 2012).

The type III secretion system (T3SS), or Injectisome, is a molecular syringe which allows direct injection of effector proteins into the cytoplasm of a host cell. It is a major factor in bacterial virulence. Evolutionarily related to bacterial flagella, it consists of more than 20 components spanning both bacterial membranes. Type III secretion is a precisely orchestrated sequence of different classes of substrates. Following formation of the base complex, subunits of the “needle” are secreted and assemble into a hollow channel extending outwards from the bacterial cell envelope. After the needle has reached a defined length, it is capped by the needle tip and the system is primed for injection. Upon contact of the needle tip with a host cell, secretion of type III translocators commences, which form a pore in the host cell membrane, followed by secretion of effectors which manipulate various processes in the host cell, such as actin polymerization eventually leading to the uptake of the bacterium into the host. Substrates are recognized by a signal sequence at their N-terminus, unfolded by an ATPase at the needle base, and threaded through the needle (Burkinshaw & Strynadka 2014; Diepold & Wagner 2014).

Type IV secretion systems (T4SS) encompass conjugation machines for delivery of DNA into bacterial cells as well as effector translocator systems for injection of proteins and protein complexes into eukaryotic hosts. Conjugation is responsible for rapid horizontal transfer of

antibiotic resistance genes to other bacteria, even across species borders, and is also an important tool in molecular biology. The secretion machinery consists of up to 30 different proteins forming subassemblies in the IM and OM. Also, long extracellular pili are extruded from T4SS and used for contacting other cells during conjugation. Substrates can either be acquired from the cytoplasm, bearing a specific type IV signal at the C-terminus and at other positions, or from the periplasm after being translocated by the Sec machinery (Christie et al. 2014).

Type VI secretion systems (T6SS) are homologous to the tails of contractile phages. They form “nano-crossbows” which allow bacteria to physically penetrate the cell envelope of other cells and inject effectors. Structurally T6SS consist of an outer sheath, which encloses an inner tube tipped with a spike complex. When the system is triggered, the sheath contracts to propel the inner tube toward the target cell, penetrating its membrane and allowing injection of effectors. This mechanism is not only used by bacteria to attack cells, but appears also to be a countermeasure against attacks from other bacteria. In such cases T6SS are assembled targeted at the agressor and effectors aim at counteracting its attack (Schwarz et al. 2010; Zoued et al. 2014).

Type VII secretion systems (T7SS) have mainly been described in *Mycobacteria*, where they are a major determinant of virulence. Other organisms with T7SS are *Actinobacteria* and Gram-positive bacteria such as *Staphylococcus*. Substrates are secreted in a folded, homo- or heterodimeric state. The secretion signal involves a YxxxD/E motif in a flexible region of one of the dimer subunits. Nucleotide binding domains in the cytosolic part of the inner membrane channel complex presumably provide the energy for the transport process. Whether the transport occurs in one step across both membranes or in two distinct steps remains unclear. Also, no outer membrane components of the system have been identified so far (Houben et al. 2014).

The type IX secretion system has only recently been described and is widespread in the *Bacteroidetes*. It is associated with gliding motility, secretion of digestive enzymes to break down environmental macromolecules, and also secretion of gingipain protease virulence factors (Sato et al. 2010). The secretion apparatus consists of at least 7 distinct proteins forming a complex with a molecular mass of at least 1.2 MDa. Substrates have a conserved C-terminal domain, presumably involved in recognition.

In contrast, type II, V and VIII secretion systems as well as the CU pathway have no dedicated components for crossing the IM and use the Sec or Tat machineries for this step.

2.2.2 Type II and VIII Secretion Systems and the CU pathway

The type II secretion system (T2SS), also referred to as General Secretory Pathway (GSP), secretes fully folded proteins from the periplasmic space. Type II substrates mainly play roles in nutrient acquisition by degradation of extracellular biopolymers, but also include toxins, adhesins and other classes of proteins. Substrates are translocated into the periplasm by the Sec or Tat pathways. As the substrates are fully folded, the recognition signal for the T2SS is assumed to be one or several structural motifs, the exact nature of which still remains elusive. Export of substrates through the outer membrane secretin is driven by the formation of a pseudopilus, to which exoproteins are attached and then pushed out. The energy for this process is gained from ATP hydrolysis and proton motive force at the inner membrane (Nivaskumar & Francetic 2014; Douzi et al. 2012).

The nucleation-precipitation mechanism of assembly of curli, amyloid fibers on the surface of enterobacteria has been termed type VIII secretion system. It plays a major role in the formation of biofilms of *E. coli* and *Salmonella*, which consist of curli and carbohydrate polymers such as cellulose. Curli subunits are translocated into the periplasm by the Sec machinery, where they remain in an unfolded soluble state. The second translocation step across the OM is mediated by a secretion complex consisting of at least three proteins. On the cell surface curli polymerization is initiated by a nucleator subunit and then proceeds in a self-templating manner (Evans & Chapman 2014).

The chaperone-usher system of Gram-negative bacteria is used for the assembly of pili on the cell surface. Substrates are translocated into the periplasm by the Sec system, where they are recognized by specific chaperones. The chaperone-substrate complex then docks at the outer membrane usher pore, through which the substrates are exported and polymerized into the pilus. Recent structural studies have shed light on the molecular mechanism of this secretion system. Pilus subunits have an Ig-like fold which is missing the last β-strand, destabilizing the protein. During polymerization, the missing secondary structure element of one subunit is provided *in trans* by an N-terminal peptide of the subsequent subunit, linking them together in a process termed “donor strand exchange”. This also allows for tight control of subunit ordering, as the N-terminal extension peptide of each subunit only has high affinity to the specific type of subunit preceding it in assembly order (Remaut et al. 2008; Busch & Waksman 2012; Geibel & Waksman 2014).

2.3 Autotransporters – the Type V Secretion System

Outer membrane autotransporters and two-partner secretion systems (TPSS), collectively referred to as type V secretion system (T5SS) are the main focus of this work. The term “autotransporter” was first defined by Meyer and coworkers (Klauser et al. 1992), referring to systems in which the translocation function and the secretion substrate reside on the same polypeptide chain. Later, Henderson and coworkers (Henderson et al. 2004) proposed that TPSS should be a subgroup of T5SS based on an identical mode of translocation, sequence similarity and phylogenetic evidence of domain exchange between classical autotransporters and TPSS passengers of equivalent function. Extending the classical definition to accommodate for TPSS, our group defines autotransporters as OM translocation systems which are autonomous from cytosolic energy sources (Leo et al. 2012).

Type V secretion systems are classified in 5 subtypes (Va – Ve) based on structural features. Type Va are monomeric autotransporters, type Vb encompasses the TPSS, type Vc are trimeric autotransporter adhesins (TAAs), type Vd are a recently described group of “fused TPSSs”, and type Ve are monomeric autotransporters with the domain order reversed (Fig. 1).

The general organization of a T5SS consists of two parts: an outer membrane β-barrel pore, referred to as anchor domain, translocator domain, or β-domain; and the translocated extracellular part - the passenger domain, extracellular domain, or α-domain. Historically, the term “passenger domain” refers to the whole exported portion of the protein which can in fact consist of multiple protein domains. In TPSSs, the membrane anchor and passenger domains reside on separate polypeptide chains. In autotransporters the passenger is usually at the N-terminus and the membrane anchor at the C-terminus, but the domain order can also be inverted (Type Ve).

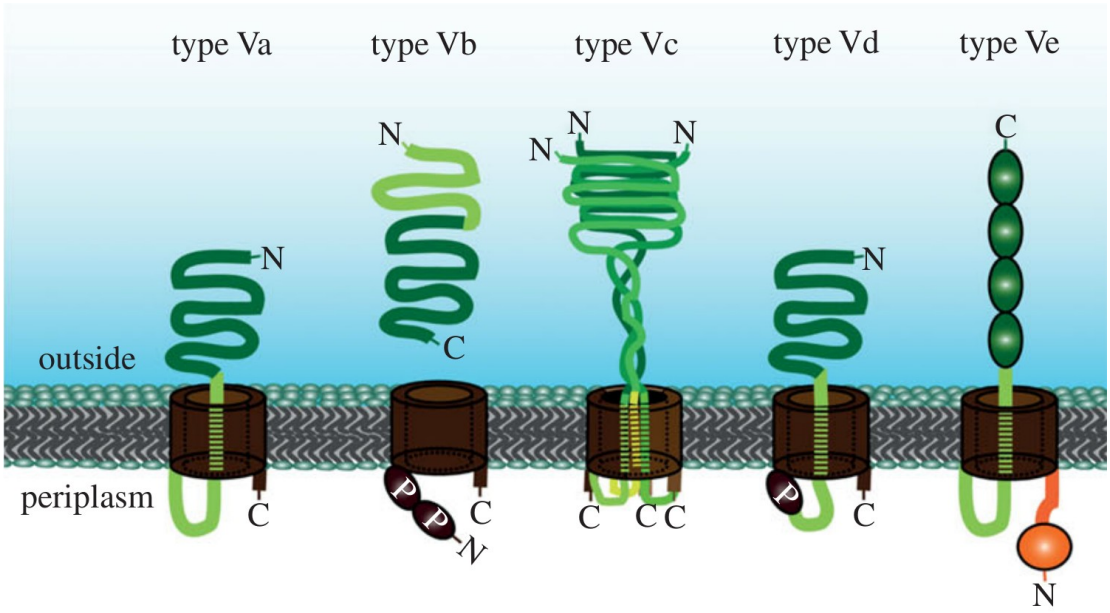

Figure 1: Overview of the type V secretion systems. *The membrane anchor domain is displayed in brown, linker/Tps regions in light green, passengers in dark green, and periplasmic domains in orange. POTRA domains are labeled (P). (From (Leo et al. 2012))*

Being outer membrane proteins, all T5SS proteins have a N-terminal signal peptide targeting them to the periplasm *via* the Sec translocon. For some very large autotransporters an unusually long signal peptide of 42 or more amino acids has been described. Szabady and coworkers (Szabady et al. 2005) have proposed that this extended signal peptide temporarily anchors the unfolded autotransporter polypeptide chain at the Sec machinery, allowing downstream processes such as OM insertion of the β-domain to take place while the passenger domain is prevented from folding and aggregating by giving periplasmic chaperones more time to bind to it. In the same vein, for some autotransporters, most prominently NalP from *Neisseria meningitidis*, a lipoprotein signal peptide was reported, and lipidation could be demonstrated (Roussel-Jazédé et al. 2013). In the case of NalP this led to delayed autocatalytic processing of the protein, allowing it to remain at the cell surface for an extended amount of time. Both examples suggest that timing is of importance during the biogenesis of autotransporters.

The membrane anchor domain is the defining feature of T5SS. While the sequences of the passenger domains of different T5SS may vary widely, all anchor domains are homologous and unknown T5SS proteins can be identified by the presence of an intact anchor domain (Remmert et al. 2009; Remmert et al. 2010; Arnold et al. 2007).

Typically, the β-barrel consists of 12 β-strands in monomeric autotransporters (type Va, Ve), 3x4 β-strands in trimeric autotransporters (type Vc) and 16 β-strands in TPSS and “fused TPSS” (type Vb, Vd). Efficient insertion of the anchor into the OM requires the Bam complex (Knowles et al. 2009). The function of the anchor domain is to translocate the passenger domain across the outer membrane and to serve as attachment point in the cell envelope.

The mode of translocation is a matter of some controversy. Several models have been proposed. The predominant model, put forward by Meyer and coworkers (Pohlner et al. 1987) for monomeric

autotransporters, suggests that after the membrane anchor has been inserted into the OM a hairpin is formed at the C-terminus of the passenger, which loops out through the membrane anchor pore. Folding of the exported parts of the passenger domain then pulls the rest of the protein out. It is attractive to assume that such a mechanism is conserved across all T5SS.

Structural studies of autotransporter membrane anchor domains using X-ray crystallography have shown that in the final folded state, the membrane pore is occluded by the α-helix (the three α-helices in TAAs, or an extended peptide in type Ve secretion systems, respectively) connecting the passenger to the membrane anchor (Oomen et al. 2004; Clantin et al. 2007; Meng et al. 2006; Fairman et al. 2012). This is in line with the hairpin model and supports the notion that autotransport happens through translocator domain pore. No structural data is available for intermediate states of the translocation process.

The hairpin model is also well supported by biochemical data. When autotransport is stalled near the C-terminus of the passenger domain, the N-terminus of the passenger remains protected from Proteinase K degradation while the C-terminus is readily degraded, indicating a C-to-N direction of export (Junker et al. 2009). Also, residues in the C-terminal part of the passenger domain could be successfully crosslinked to BamA, suggesting an active role of the Bam complex in passenger translocation. However, the exact nature of this role remains speculative. Another clue is that the folding core of the passenger domain appears to be near its C-terminus, with often little autonomous folding propensity at the N-terminal end (Peterson et al. 2010; Soprova et al. 2010; Junker et al. 2006). This suggests that folding nucleates at the C-terminus and proceeds to the N-terminus, folding up the polypeptide as soon as it is exported, and thus preventing it from sliding back into the periplasm, a mechanism known as “Brownian ratchet” (Klauser et al. 1992). As it has been shown that an autotransporter (the TAA YadA from *Yersinia*) can be heterologously expressed and successfully integrated into the outer membrane of yeast mitochondia (Müller et al. 2011; Ulrich et al. 2014), it is unlikely that external factors beyond the very basally conserved ones, like the Bam complex, are involved in autotransporter folding.

While the membrane anchor's role in T5SS is a mainly mechanistical one, the biological “business end” of T5SSs is the passenger domain. A broad spectrum of enzymatic activities such as proteolysis and lipolysis as well as adhesion to biotic and abiotic surfaces have been reported in the literature (Henderson & Nataro 2001). After transport the passenger can be retained at the membrane tethered to the anchor, or released into the extracellular milieu. In the following chapters I will give an overview on the different groups of autotransporter proteins, their properties, and the similarities and differences between them.

2.3.1 Type Va and Ve – Monomeric Autotransporters

The first type V secretions systems described were monomeric autotransporters. This group, collectively referred to as type Va, comprises many important virulence factors such as the adhesin Pertactin from *Bordetella pertussis* (Leininger et al. 1991), proteases IgA1 as well as NalP from *Neisseria* spp. (Plaut et al. 1975; van Ulsen et al. 2003), and AIDA-I from *Escherichia coli* (Benz & Schmidt 1992). In some cases the passenger domain of monomeric autotransporters can be released into the extracellular medium by (auto-)proteolytic cleavage of the α-helix attaching the passenger

to the anchor domain, e.g. App and IgA1 protease in *Neisseria* which can either be autoproteolytically cleaved or processed by NalP, resulting in different forms of those proteins (van Ulsen et al. 2003).

Recently, the proteins Intimin from *E. coli* and Invasin from *Yersinia* spp. have been discovered to belong to the autotransporter family (Tsai et al. 2010; Oberhettinger et al. 2012). The curious thing about the topology of these monomeric autotransporters is that the membrane anchor domain is N-terminal of the passenger domain. This has repercussions on the autotransport mechanism: Applying the hairpin model, the passenger would loop out in N-to-C order, contrary to the C-to-N order of type Va autotransporters. To acknowledge this significant difference, these proteins were assigned to the type Ve subgroup of T5SSs.

2.3.2 Type Vb and Vd – Two-Partner Secretion Systems

Type Vb secretion systems, or Two-Partner Secretion Systems, as the name implies, consist of two separate polypeptide chains, the passenger protein and the translocator OM β-barrel protein. Systematically they are referred to as TpsA and TpsB, respectively. These usually form operons, with the *tpsB* gene upstream of the *tpsA* gene, although the order can be reversed. Most TpsB proteins exclusively transport their respective TpsA partner, but examples of multiple substrates using the same transporter, or multiple transporter transporting the same substrate exist (ur Rahman et al. 2014; Julio & Cotter 2005). This specificity is confered by the polypeptide transport-associated (POTRA) domains, two of which are found at the N-terminus of all TpsB proteins.

The membrane pore of TpsB proteins is a 16-stranded β-barrel homologous to BamA. A conserved loop blocks the barrel prior to substrate recognition. Upon recognition of a conserved N-terminal TPS domain of the TpsA protein by the POTRA domains of the TpsB protein, a conformational change is induced by which the loop is displaced from the barrel, and the substrate is exported in a hairpin-like fashion, similar to the model predicted for autotransporters (As proposed in (Mazar & Cotter 2006; Clantin et al. 2007)).

Type Vd secretions systems were discovered only recently (Salacha et al. 2010). The first described protein of the family is PlpD from *Pseudomonas aeruginosa*. The N-terminal part of the protein is a lipase passenger domain, followed by a POTRA domain and a TpsB-like domain at the C-terminus – overall resembling a fused TPSS. This group might represent an intermediate step between type Va and type Vb secretion systems and thus support the notion of an evolutionary relation between the two groups.

2.3.3 Type Vc – Trimeric Autotransporter Adhesins

Trimeric autotransporters, or type Vc secretions systems, share the same basic structure as monomeric autotransporters, having a C-terminal β-barrel translocator domain and a N-terminal passenger. The most striking difference, of course, is that they form obligate homotrimers on the cell surface. Furthermore, the predominant function described for type Vc secretion system proteins is adhesion (therefor the name Trimeric Autotransporter Adhesins (TAAs)); no enzymatic activities have been described to date, and they are not released from the surface by (auto-)proteolysis.

TAA passengers have a modular structure and are composed of a limited number of domain

building blocks. The structural space available to them is mainly constrained by the necessity to form a homotrimer. Most autotransporters alternate between stretches of coiled coils and globular domains, with adapters in between. The complexity of domain composition and ratio of coiled coils to globular domains can vary widely from autotransporters containing mostly coiled coils such as UspA2 from *Moraxella catarrhalis* all the way to TAAs composed almost exclusively of globular domains such as XadA from *Xanthomonas oryzae*; from simple head-stalk-anchor topologies such as the prototypic autotransporter adhesin YadA from *Yersinia enterocolitica* to the complex topologies observed in e.g. SadA from *Salmonella* enterica and BadA from *Bartonella henselae*. Presumably this repetitive modular structure allows pathogens to exchange their arsenal of adhesins by recombination, and thus evade the host immune defense. It also facilitates structural studies by allowing to apply a domain dictionary approach and reconstruct a model of the whole adhesin fiber (Fig. 2), which in many cases is not amenable for structure determination due to the large size of the molecule, from fragment structures encompassing single domains (Hartmann et al. 2012).

The 12-stranded β-barrel membrane anchor of TAAs is assembled from the three 4-stranded β-sheet fragments contributed by each of the monomers. The Bam complex plays a crucial role in efficient biogenesis of TAAs, as was demonstrated for YadA. Furthermore the periplasmic chaperones such as SurA, Skp, and SecB are involved in stabilizing the long unfolded polypeptide chain during its passage through periplasm, and the protease/chaperone DegP is involved in quality control and degradation of misfolded TAAs.

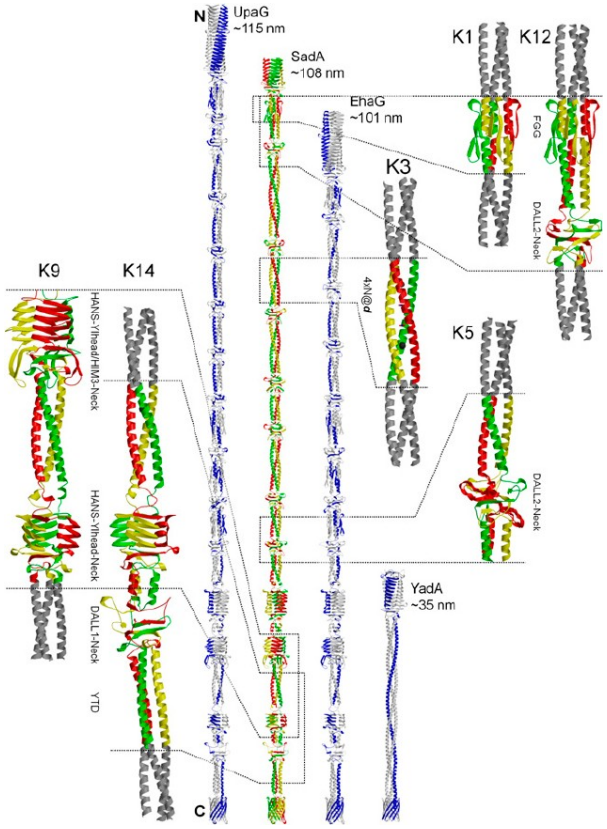

Figure 2: Crystal structures of fragments and reconstructed full fibers of the TAAs SadA, UpaG and EhaG. *The three chains of the SadA trimers are colored individually; the coiled-coil adaptors fused to the fragments are shown in gray. The reconstructed fibers are compared with a prototypical simple TAA, YadA. (From (Hartmann et al. 2012))*

3 Aims of this work

The focus of the presented work is to gain insights into the biogenesis of trimeric autotransporter adhesins in enterobacteria using the *Salmonella* trimeric autotransporter adhesin SadA as model system. In particular, the role of the periplasmic inner membrane lipoprotein SadB, which presumably forms an operon with SadA, will be examined. To facilitate the investigation of operons using bioinformatics, it is our goal to create a specialized bioinformatics tool which enables the expedient and efficient analysis of operon structure and composition.

4 Publications and contribution of the candidate

4.1 Publications used in this thesis

Grin, I. & Linke, D., 2011. GCView: the genomic context viewer for protein homology searches. *Nucleic acids research*, 39(Web Server issue), pp.W353–6.

The idea for a bioinformatics tool developed in discussions between Dirk Linke and myself. I implemented of the backend logic (in Python) as well as the web frontend (HTML/JavaScript). Integration of GCView into the MPI Toolkit (Ruby/RHTML) was also written by me with help from Christina Wassermann and Vikram Alva.

I wrote the paper draft, and together with Dirk Linke participated in the revision of the manuscript including the published version. Changes to the GCView code requested by the reviewers were implemented by me.

Grin, I., Hartmann, M.D., Sauer, G., Hernandez Alvarez, B., Schütz, M., Wagner, S., Madlung, J., Macek, B., Felipe-Lopez, A., Hensel, M., Lupas, A. & Linke, D., 2014. A trimeric lipoprotein assists in trimeric autotransporter biogenesis in enterobacteria. *The Journal of biological chemistry*, 289(11), pp.7388–98.

Constructs were designed by Dirk Linke. Microscopy, flow cytometry, cross-linking, phage display and cell shaving experiments were designed by me. I also contributed large parts of the practical work and data analysis. The crystal screen pipetting robot was operated by Kerstin Bär and Reinhold Albrecht. I participated in the acquisition of the X-ray diffraction data at the SLS, Villigen, Switzerland. Processing of X-ray data was done by Marcus D. Hartmann. Model building and refinement was done by me with support from Marcus D. Hartmann. Mass spectrometry of the full-length construct was done by Guido Sauer, subsequent data analysis was done by Dirk Linke and me. Mass spectrometry for the cell shaving experiments was done by Johannes Madlung at the Proteome Center Tübingen, subsequent data analysis was done by me. *In vivo Salmonella* imaging was contributed by Alfonso Felipe-Lopez.

I wrote large parts of the paper draft together with Dirk Linke. The section concerning *in vivo* data from *Salmonella* was contributed by Michael Hensel. Revisions of the manuscript were done by Dirk Linke and me, with input from the other co-authors. Additional experiments requested by the reviewers were done by me in the lab of Samuel Wagner.

4.2 Additional publications

Thein, M., Sauer, G., Paramasivam, N., Grin, I. & Linke, D., 2010. Efficient subfractionation of Gram-negative bacteria for proteomics studies. *Journal of proteome research*.

I contributed in the analysis of the copious amounts of mass spectrometry data and participated in the revision of the manuscript.

Müller, J.E.N., Papic, D., Ulrich, T., Grin, I., Schütz, M., Oberhettinger, P., Tommassen, J., Linke, D., Dimmer, K.S., Autenrieth, I.B. & Rapaport, D., 2011. Mitochondria can recognize and assemble fragments of a beta-barrel structure. *Molecular biology of the cell*, 22(10), pp.1638–47.

I contributed a YadA MA construct to this publication.

Grin, I., Schwarz, H. & Linke, D., 2011. Electron microscopy techniques to study bacterial adhesion. *Advances in experimental medicine and biology*, 715, pp.257–69.

I wrote the manuscript draft with Heinz Schwarz and Dirk Linke, and participated in the revision of the manuscript including the published version.

Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N. & Hernandez Alvarez, B., 2012. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51), pp.20907–12.

I contributed samples for electron micrographs of bacteria expressing SadA (Figure 1 in the paper). Furthermore I took over the modeling of full-length adhesin fibers after Stanislav Dunin-Horkawicz, designed the supplementary figures and participated in the revision of the manuscript including the published version.

Leo, J.C., Grin, I. & Linke, D., 2012. Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1088–101.

I provided bioinformatics data on the genomic organization of type Vb and type Vd secretion systems (Figure 3a-c in the paper), and participated in the revision of the manuscript including the published version.

5 Results

5.1 GCView - the genomic context viewer for protein homology searches

The investigation of operon structure and genomic neighborhood of a gene of interest can reveal important insights into the evolution of a biological system. A significant portion of bacterial genes are organized in operons (Price et al. 2006), and even if this is not the case, are found in similar locations in the genomes of related species (Korbel et al. 2004). Expanding the field of view and looking at a larger region of a chromosome oftentimes reveals information on gene duplications, insertions, deletions, or translocations. With an increasing number of well annotated fully sequenced genomes publicly available in databases such as UNIPROT and the NCBI databases, retrieving and analyzing the genes upstream and downstream of a gene of interest has become almost trivial. Specialized databases such as BioCyc (Karp et al. 2005), STRING (Szklarczyk et al. 2011), The SEED (Overbeek et al. 2005) and Ensembl Bacteria (Kersey et al. 2010) use this data to provide precomputed and curated information on the operon structure and genomic location of well investigated proteins and operons. Still, when the gene of interest or the target organism are not included in the operon databases, the intermediate level between fully curated database and raw genomic data consists largely of manual work.

We developed GCView - a tool for viewing the genomic context for protein homology searches to ease and automate extracting and comparing the genomic regions upstream and downstream of protein-coding genes of interest. It is integrated into the Bioinformatics Toolkit of the Max-Planck Institute for Developmental Biology (MPI Toolkit) (Biegert et al. 2006). The tool can be accessed through a web interface at http://toolkit.tuebingen.mpg.de/gcview. This website is free and open to all users without login requirement. However, an account can be created at the MPI Toolkit website, which allows storing results for an extended period of time. Account creation is also free and open to all users.

GCView uses the protein homology search tools integrated in the MPI Toolkit, such as PSI-BLAST (Altschul et al. 1997) to generate a list of proteins similar to the protein of interest. Hits with an E-value below a threshold specified in the input parameters for the tool, as well as hits for proteins from organisms for which no fully sequenced and assembled genomes are available, are removed from the list.

We chose protein homology over DNA sequence similarity due to the higher sensitivity of protein searches. In consequence, only protein-coding genes can be analyzed using GCView.

For all protein-coding genes from fully sequenced and assembled genomes in the NCBI Bacteria repository, we have compiled a database of the corresponding protein identifiers and genomic locations to use for lookups with GCView. Using this database, the entries from the genomic region upstream and downstream of each protein on the BLAST hit list are collected, resulting in one genome region for each hit. Overlapping regions from the same genome are merged.

Multiple queries can be integrated by GCView to analyze multiple proteins at once, e.g. all known

components of an operon. In this case a separate list is generated for each of the inputs and the processing steps described above are performed on each of the lists. Finally, the resulting regions are grouped by the number and order of proteins of interest contained therein and displayed. This workflow is summarized in Figure 3.

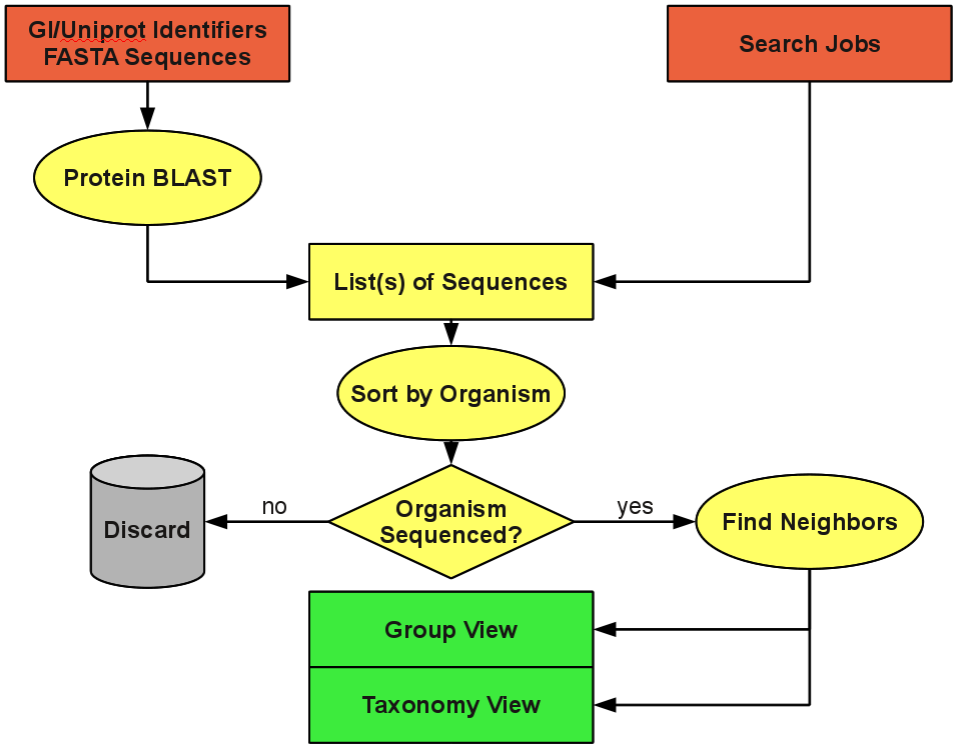

Figure 3: Workflow of GCView. *Red boxes represent inputs to the tool. Yellow boxes stand for processing steps. Green boxes are outputs from GCView. (From (Grin & Linke 2011))*

The results of GCView are presented in two different views. The Group View provides an overview of the results (Fig. 4A). All genomes that have a certain operon composition, or more generally, the same number and order of the protein-coding genes of interest, are grouped together. An overview image for each group shows a schematic representation of the number and arrangement of the genes in this group. Each gene from the input query is represented as a colored arrow. A legend on the top of the page explains the color code. It should be noted that in the overview image the arrows are not to scale with the corresponding genes and the color intensity does not indicate the degree of similarity between query and hit.

If two queries map onto the same gene, this is indicated by a fused arrow, suggesting a possible gene fusion. It is also possible to use single domains of proteins as query in GCView and analyze domain composition and domain rearrangements within a protein of interest.

Finally, genes which are located between genes of interest, yet are not homologous to any of the query sequences, are displayed as a gray box. This allows the user to quickly spot insertions between genes of interest. Each group can be expanded to get a detailed view of the genomic regions contained therein.

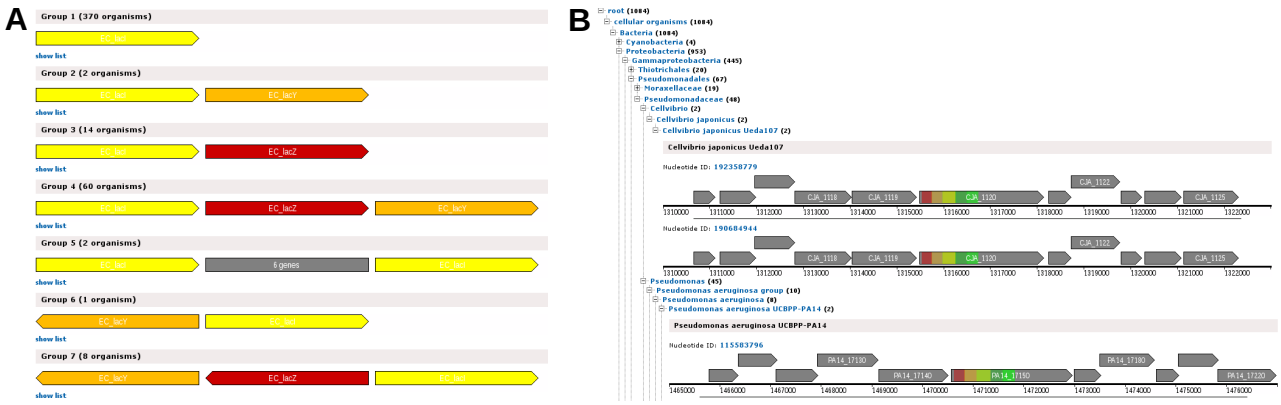

Figure 4: Output from GCView.
A: *Group View. The components of the lac operon were used as input.*
B: *Example of the Taxonomy view, showing POTRA domains from Omp85 and related proteins.*

The second available view of the GCView results is the Taxonomy view (Fig. 4B). Here the results are mapped onto a taxonomy tree. A number on each branch of the tree indicates the number of hits in this taxon and its subgroups. At the leaves of the tree the detailed view of the corresponding hits can be found. Empty branches are omitted for clarity.

The detail view of the genomic regions is identical for both the Group and Taxonomy Views (Fig. 5). A schematic image shows the protein-coding genes in the region of interest. A ruler is included on the bottom of each image for scale. Each gene is represented as an arrow. The length and orientation of the arrows correlate with the length of the gene, and with its location on the forward or reverse strand. Regions of homology are displayed as colored boxes of appropriate width on the arrows. The intensity of the color correlates with the degree of similarity between hit and query sequences. On hovering the mouse pointer over an arrow a tooltip window with details about the gene is displayed. The information includes the precise location and length of the gene, as well as the annotation and distances to the neighboring genes.

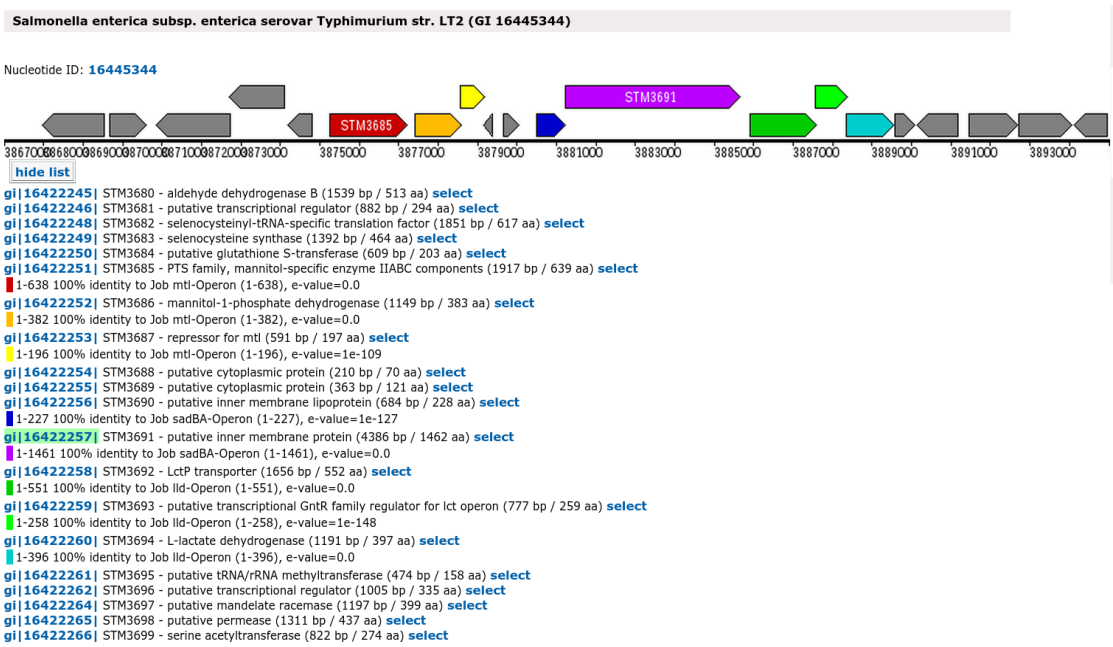

Figure 5: The Salmonella sadBA operon and its genomic environment, shown in Detail View

Below the each image the protein-coding genes of the region are also compiled as a list, with the search hits for each of them highlighted. Entries from this list can be selected to be added to the GCView search to iteratively expand the set of analyzed sequences, or for submission to other tools within the MPI Toolkit for further analysis.

5.2 A Trimeric Lipoprotein Assists in Trimeric Autotransporter Biogenesis in Enterobacteria

Trimeric Autotransporter Adhesins are very large molecules in the outer membrane of Gram-negative bacteria. Their biogenesis is to large extent similar to the biogenesis of all other outer membrane proteins, in that they possess a N-terminal signal peptide for Sec translocation into the periplasm, and that the Bam complex is required for correct insertion into the outer membrane. Yet, therein lies the first challenge; the β-domain lies at the very C-terminus of the polypeptide chain, and three chains must associate to form a 12-stranded barrel in the outer membrane. Furthermore, the rest of the protein must be kept unfolded until it can then, upon or during formation of the barrel, be threaded through in an orderly fashion. The details of this process are not yet fully understood.

While working on the *Salmonella* trimeric autotransporter adhesin SadA, we observed using GCView, that the genomic location of *sadA* and its homologs is conserved in *Salmonella*, *Shigella* and *Escherichia* (Hartmann et al. 2012). The gene is always located between the *mtl* and *lld* operons for mannitol and L-lactate metabolism, respectively (Fig. 6A). Interestingly, in all genomes which had *sadA* we also observed a gene coding for a small predicted lipoprotein directly upstream of the autotransporter gene. The short intergenic distance led us to assume that the two genes form an operon, whereupon we named the lipoprotein protein SadB.

Using HHPred (Söding et al. 2005) we searched the Protein Data Bank for proteins with known structure homologous to SadB. Our search returned only one known structure of a paralog, YajI from *E. coli*, which is conserved in enterobacteria. The gene *yajI* does not appear to be part of an operon, although its genomic location is also conserved (Fig. 6B). In the available NMR structure (2JWY) only the C-terminal domain is resolved. A biological function was not described for either YajI or SadB.

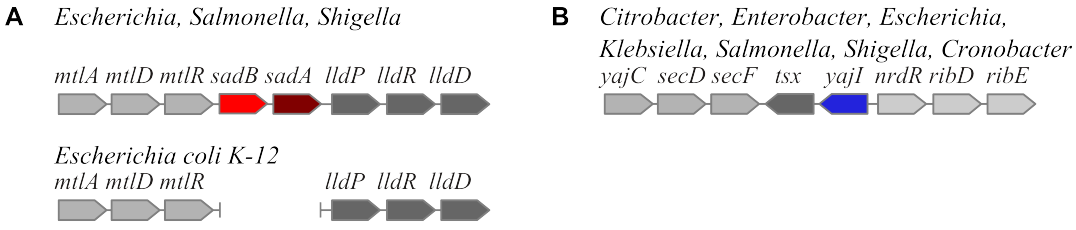

Figure 6: Genomic context of sadB and yajI is conserved in Enterobacteria.
A: *sadBA operon is located between the mtl operon and the lld operon. E. coli K12 has a 5-kb deletion encompassing the sadBA operon.*
B: *sadB paralog yajI is located between the yajC-secDF operon and the putative nrdR-ribDE-nusB operon. Note that although sadB and sadA are linked in an operon, yaiJ and tsx are not. (From (Grin et al. 2014))*

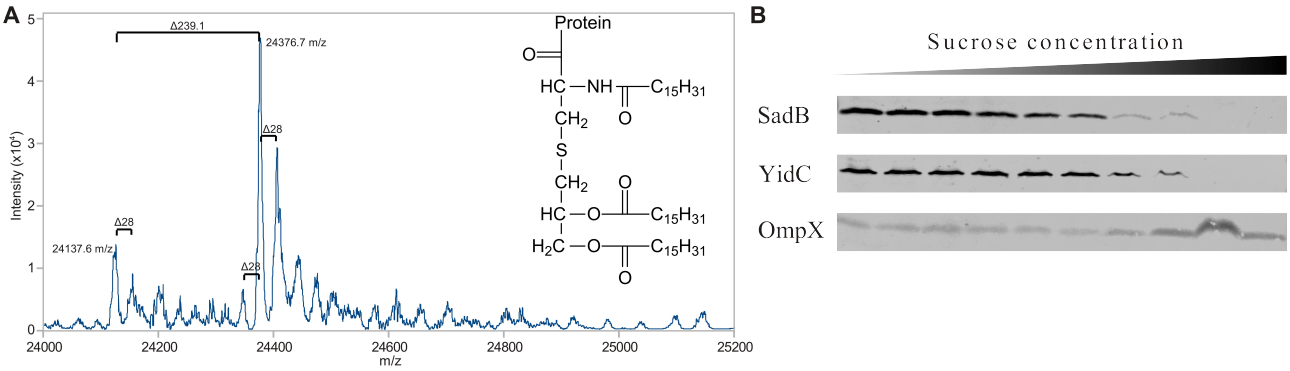

Figure 7: SadB is a periplasmic inner membrane lipoprotein.
A: *mass spectrometry analysis of full-length SadB with native lipid anchor. The main mass peak at 24376.7 Da corresponds to the expected molecular weight of the protein with a tri-palmityl anchor (see inset). The lighter mass at 24137.6 Da is the result of the loss of one of the palmityl chains by hydrolysis. Mass differences of 28 Da (-C_2H_4-) result from the incorporation of stearic acid (C18:0) or myristic acid (C14:0) instead of palmitic acid (C16:0) in the lipid anchor.*
B: *subcellular localization of SadB by density gradient centrifugation. In a sucrose gradient SadB is found in the lower density band corresponding to the inner membrane. As control, OmpX (an outer membrane β-barrel protein) is only found in the higher density band corresponding to the outer membrane. (From (Grin et al. 2014))*

SadB was predicted to be a periplasmic inner membrane lipoprotein by LipoP based on the aspartic acid at the +2 position of the predicted signal protease cleavage site. We verified the prediction by subcellular fractionation of *E. coli* heterologously expressing *sadB* from a plasmid (Thein et al. 2010). SadB is found in the lower density membrane band of the sucrose gradient , colocalized with YidC - an inner membrane protein chaperone/insertase (Fig. 7B). Mass spectrometry analysis of full length SadB extracted from membrane preparations confirmed successful cleavage of the signal peptide at the predicted position and attachment of a canonical tripalmityl lipid anchor to the N-terminal cysteine residue (Fig. 7A). We conclude from these findings that SadB is indeed a lipoprotein of the inner membrane, protruding into the periplasmic space.

Based on the functional linkage suggested by the genomic association and the fact that SadA resides in the periplasmic space prior to its export to the cell surface, we hypothesized that SadB is involved in the biogenesis of SadA. We created inducible expression constructs of either the full *sadB*A operon or just *sadA* alone to elucidate the effect of SadB on SadA surface display. It should be noted that the native expression conditions of *sadB*A are not known, and thus the native promotor of the operon could not be used (Humphries et al. 2003).

Immunofluorescence microscopy of whole bacteria stained with α-SadA antibody (Hartmann et al. 2012) showed a strong fluorescence signal uniformly distributed over the whole surface of bacteria expressing *sadB*A. Bacteria expressing only *sadA* hat a weaker signal overall, with isolated foci of fluorescence intensity (Fig. 8A). Quantification by flow cytometry showed that bacteria expressing *sadB*A had a 4-6 fold higher mean fluorescence intensity compared to cells expressing only the adhesin, suggesting a significantly higher amount of SadA on the cell surface (Fig. 8B). An explanation for this observation could be increased export efficiency of SadA in the presence of SadB. Reduced transport efficacy of SadA in absence of the lipoprotein would lead to accumulation

of the adhesin in the periplasm and could trigger degradation by the periplasmic protease DegP, leading to reduced amount of detectable surface localized protein (Grosskinsky et al. 2007).

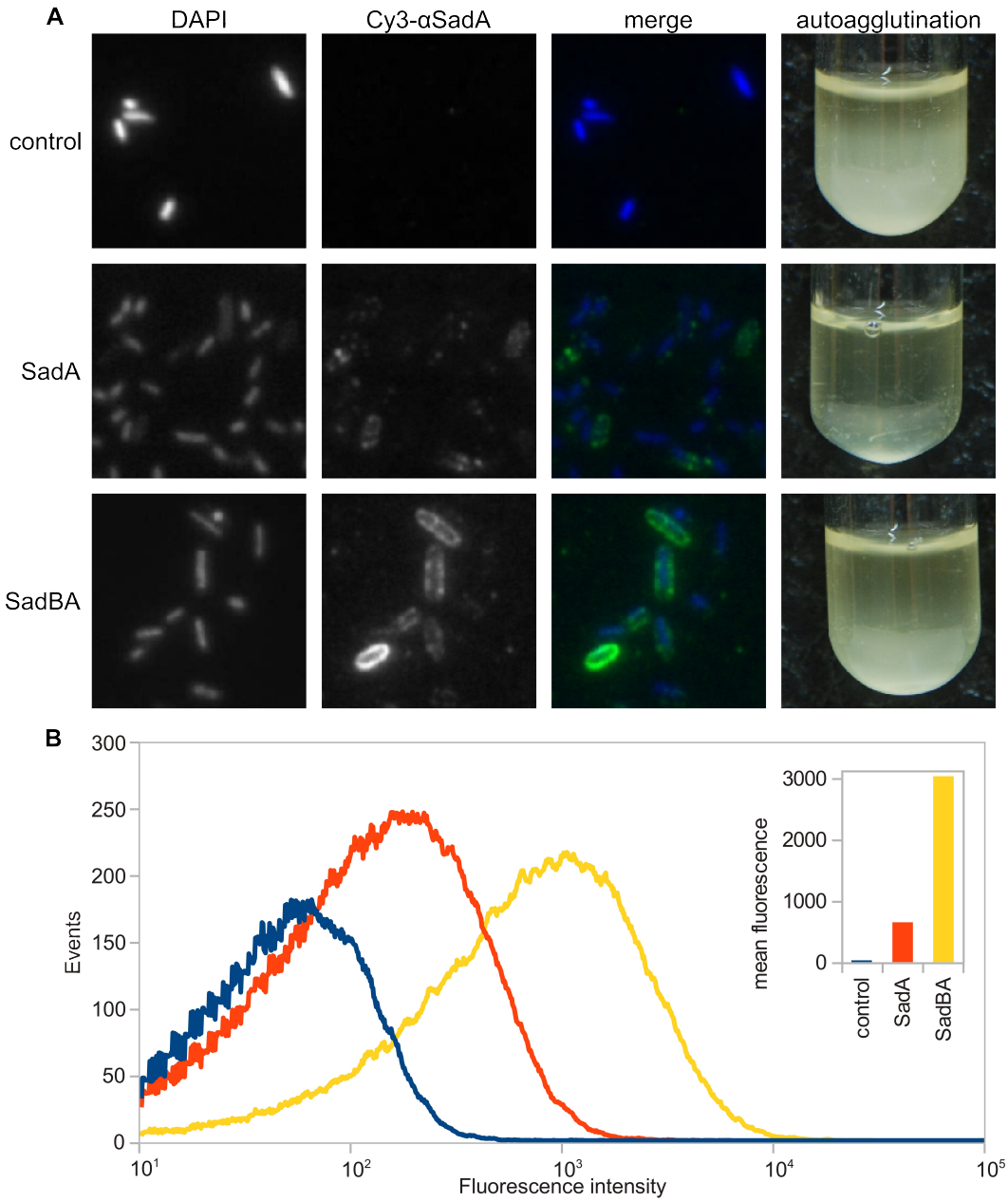

Figure 8: SadB enhances the surface display of SadA and reduces autoagglutination.
A: *immunofluorescence microscopy and cell aggregation assay of bacteria expressing sadA, sadBA, or an empty vector control. Surface-localized SadA was stained using a specific antibody.*
B: *flow cytometry analysis of bacteria expressing sadA, sadBA, or an empty vector control. Surface-localized SadA was stained using a specific antibody. The inset shows the mean fluorescence intensity for each sample. (From (Grin et al. 2014))*

During our experiments with cells expressing *sadA* alone, we observed that the bacteria formed dense, sticky pellets after centrifugation, which were difficult to resuspend. This observation was confirmed in an autoagglutination assay (Fig. 8A, right column). Bacteria were left to settle after induction of either *sadB*A or just *sadA* expression. After 8-12 h the supernatant of the culture

expressing *sadA* was clear and the bacteria had all settled to the bottom of the tube. In contrast, the culture expressing the full operon was still turbid, as was the empty vector control. Viability of the bacteria was unaffected. An explanation for this effect could be misfolding of SadA on the surface of the bacteria, leading to exposed hydrophobic surfaces in the protein by which cells then aggregate. This suggests that SadB is required for the biogenesis of well formed SadA trimers on the surface of bacteria.

Based on the previous finding we went on to investigate if SadB does indeed improve the folding and thus protease stability of SadA. We performed cell shaving with Proteinase K and analyzed the resulting proteolytic fragments by SDS PAGE and subsequent mass spectrometry. After 10 min of incubation of the bacteria with protease, we could detect significantly more peptides from high molecular weight fragments in the SadBA sample compared to the sample containing only SadA, after normalization for total protein content. This means that SadA was more readily digested by Proteinase K when SadB was not present during the passage of SadA through the periplasm. This finding further supports the hypothesis that SadB has a direct effect on the final folded state of the SadA fiber. Based on the knowledge that SadA contains highly repetitive sequence motifs (Hartmann et al. 2012), especially six consecutive repeats of 70-120 amino acids in the N-terminal part of the stalk of SadA, which have between 55% and >90% identity, it is tempting to assume that SadB could help to define the register of the three exporting SadA chains that form the final trimer by synchronizing the export.

To show that the results obtained so far are relevant in a more native setting, we investigated the role of SadB in *Salmonella* enterica. As discussed above, the native expression conditions of the *sadB*A operon are not known. Therefore *Salmonella* strains were generated in which *sadB*A or only *sadA* is under control of the arabinose inducible P_{BAD} promotor. After induction, in both the P_{BAD}::*sadB*A and the P_{BAD}::*sadA* strains production of SadA was detectable. The strain expressing only *sadA* showed significantly lower surface signal for SadA compared to the *sadB*A strain. The surface localization was diffuse in the strain expressing *sadA* only, whereas the expression of *sadB*A resulted in a clustered distribution with several high intensity foci per cell. The localization of SadA in clusters on the cell surface might be required for the function as an adhesin. However the exact role of SadA in *Salmonella* adhesion and pathogenesis is unknown.

We attempted several approaches to show direct interaction of SadB with SadA. A phage display screen against a library of 7-mer or 12-mer peptides using SadB as bait did not converge on a specific sequence. The sequences recovered did however show an alternating pattern of hydrophilic and hydrophobic residues, somewhat similar but not strikingly identical to sequence motifs recurrently found in the sequences of TAAs. Furthermore, chemical cross-linking experiments using membrane permeant cross-linking agents failed to show an interaction of SadB with SadA *in vivo*. This could indicate that the interaction does not rely on a defined binding site, but rather on transient binding at multiple binding motifs, which can be easily released as the passenger domain is exported.

To obtain a structure of SadB, we created a construct without the N-terminal signal peptide, and in which the lipid-modifiable Cys 22 was replaced by a serine. The resulting protein was solubly

expressed in large amounts in the cytosol of *E. coli* and could successfully be purified. Crystallization trials yielded crystals diffracting to 2.45 Å. The structure was solved via single isomorphous replacement using a platinum derivative.

SadB is, in striking analogy to SadA, a homotrimer, held together by an extended coiled coil at its N-terminus (Fig. 9A and 9B). The C-terminal domains of each chain display a β-sandwich topology, composed of two antiparallel four-stranded β-sheets. The topology of this variant of the Ig fold is identical between SadB and the structure of its paralog YajI (2JWY), which was very helpful in tracing the electron density of the domain (Fig. 9C and 9D).

No trimerization was observed during protein purification of SadB, including size exclusion chromatography. This suggests a low trimerization propensity of the coiled coil. Interestingly the solution NMR structure of YajI is also in its monomeric state, despite the fact that sequence analysis suggests the presence of a N-terminal coiled coil similar to SadB, indicating a similarly low trimerization rate in solution. Our explanation for this is the high number of polar residues in core positions of the coiled coil. Of the 18 core residues of the nine heptad repeats of SadB, six are glutamines. Four of them occupy position *a* of the heptad repeat, while two are in position *d*. In total, four of the glutamines are arranged within two closely spaced segments with 10 consecutive polar residues each, providing very little local hydrophobicity for coiled coil assembly. Further, in the N-terminal half of the coiled coil domain, which is closest to the bacterial inner membrane, 22 of 28 residues are polar, and 14 are charged. Polar core residues are known to lower the stability and thus the oligomerization propensity of coiled coils and play a prominent role in the coiled coil stalk segments of SadA (Eckert et al. 1998; Hartmann et al. 2009). We expect that the trimerization of SadB requires elevated local concentration of the protein, such as would exist *in vivo* on the outside of the inner membrane where the protein is localized.

We used the DALI server (Holm & Rosenström 2010) to search the PDB for structures similar to the C-terminal domain of SadB. Using the N-terminal domain resulted in unspecific hits to coiled coil containing proteins. Besides the similarity to YajI, which was mentioned above, several bacterial proteins of often unknown function were found which display high structural similarity to the Ig fold of SadB. However, the strand topology deviates from that of SadB, and in most cases no N-terminal coiled coil was present. We also found high structural similarity to the C-terminal Meprin and TRAF homology (MATH) domain of human TNF receptor associated factor 2 (TRAF2) (See Fig. 9E-H). No similarity between SadB and MATH domains is detectable on the sequence level, yet the strand topology is identical between SadB and all known MATH domains. The described function of MATH domains is to bind peptides from the cytoplasmic domain of TNF receptors. The binding site is located on the outer β-sheet of the domain. The TNF receptor trimerizes upon ligand binding, positioning the cytoplasmic domains in a triangle fit for binding to the MATH domains of a TRAF2 trimer. This binding mode prefers trimeric, ligand-bound receptor complexes over single receptor molecules by means of increased avidity of three binding sites over a single one (McWhirter et al. 1999; Elgueta et al. 2009).

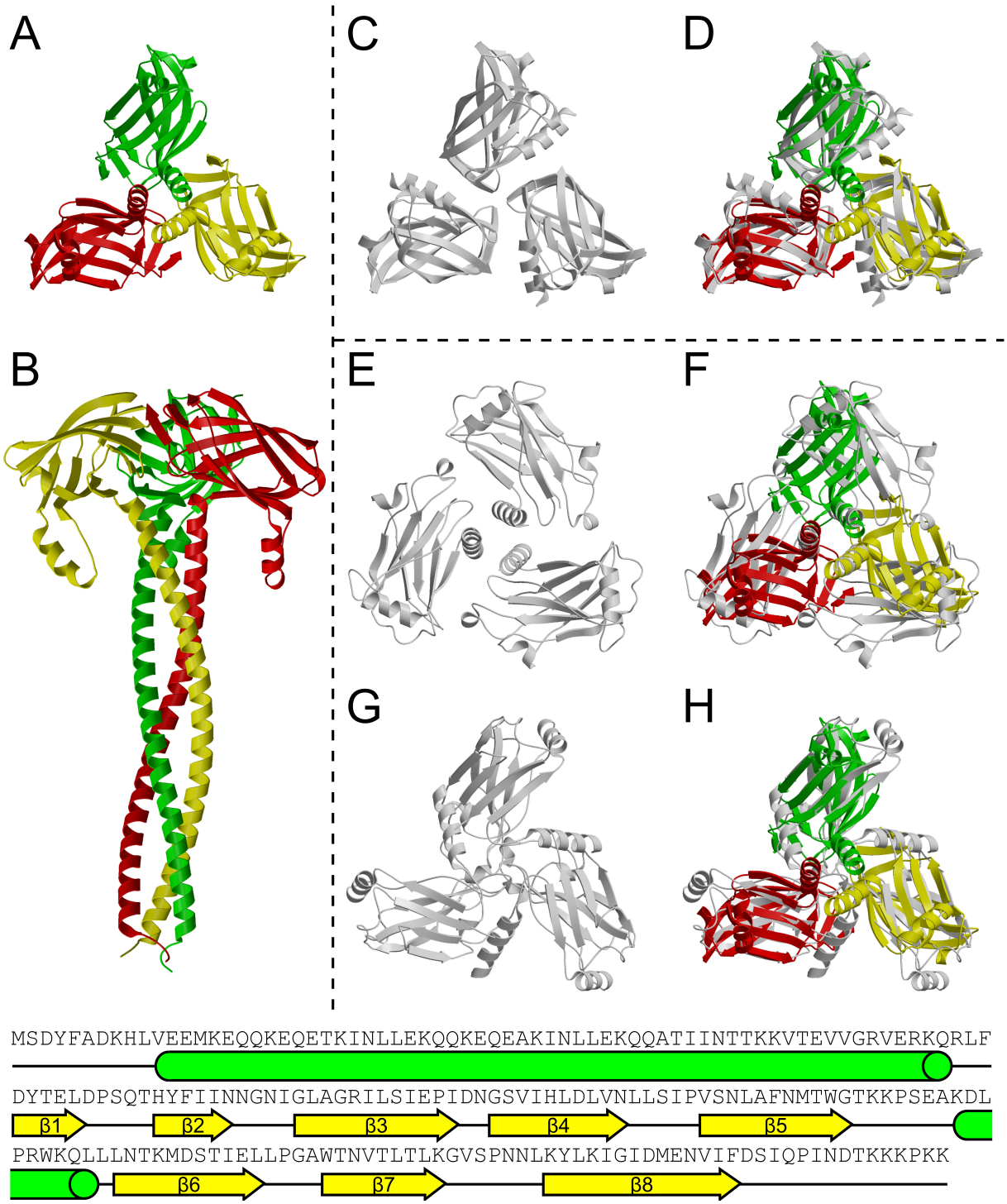

Figure 9: Structure of SadB compared with the structures of TRAF2 and YajI.
A and B: *structure of SadB viewed alongside and perpendicular to the axis of the coiled coil, colored by monomer.*
C and D: *superposition of SadB (colored) with the YajI structure from PDB entry 2JWY (gray). The YajI C domain is superimposed on the three C domains of SadB (r.m.s.d. 1.9, 2.0, and 2.0 Å for 101, 102, and 102 aligned CA positions, respectively).*
E–H: *superposition of SadB(colored) with the TRAF2 structure from PDB entry 1CZY (gray). E and F, superposition is based on the coiled coil, and in G and H, individual monomers of the Traf2 C domains are superimposed on the individual C domains of SadB (r.m.s.d. 3.2, 3.2, and 3.0 Å for 99, 105, and 99 aligned CA positions, respectively). It is apparent that the orientation of the C domains with respect to the coiled coil is different in the two proteins. (From (Grin et al. 2014))*

Although no evolutionary link can be established between SadB and MATH domains of eukaryotic TRAF proteins based on sequence similarity, the obvious structural similarity as well as the fact that both act as a trimer on trimeric membrane proteins leads us to speculate that the mode of action is similar: To bind and/or bring together three unfolded protein chains of a second trimeric protein.

5.3 Conclusion and Outlook

GCView was developed to address the need for efficient analysis of the genomic context of proteins, while investigating operons or shared synteny. Available databases either did not offer the needed information for my proteins of interest, presented too much information around the query which was not relevant for me, or did not support the level of interactivity desired.

Harnessing the power of the MPI Toolkit, I built GCView to fill the gap between such curated operon databases, and the “raw” annotated genome and proteome data from sequencing projects. GCView uses the protein homology search tools for sensitive retrieval of sequences similar to the input query. While a larger genomic region is displayed in the output of GCView, only the query and its homologs are highlighted, providing a greater focus on the query. Additionally, the degree of similarity is easily recognizable in the output from the color intensity of the homology region. The query can be expanded interactively by dynamically adding more proteins to the search in an iterative manner. Users can have full control over the parameters of the homology search by running it separately within the MPI Toolkit and then forwarding the results to GCView, providing transparency of the process by which the results are generated and allowing any kind of specialized search possible with the BLAST suite to be used with GCView. This includes the use of single domains as query to reveal different domain contexts and compositions.

The embedding of GCView in the framework of the MPI Toolkit allows the tool to benefit from the established environment. It is possible to run multiple GCView jobs concurrently and the results can be saved for later inspection. Furthermore, the results of GCView can be forwarded to other tools for in-depth analysis, or, as mentioned above, *vice versa* results from homology search tools can be forwarded to GCView.

With the help of GCView I established that the *Salmonella* trimeric autotransporter adhesin SadA in forms an operon with SadB, a protein of unknown function.

I was able to demonstrate that SadB is a periplasmic inner membrane lipoprotein of *Salmonella* spp. Homologous proteins exist in other enterobacteria, but could not be detected outside this family.

Despite the fact that no direct interaction between SadB and its cognate autotransporter SadA could be demonstrated, I have accumulated evidence that SadB has direct influence on export and folding of SadA. The absence of SadB leads to reduced surface display and less stable folding of SadA on the surface of bacteria, which is evidenced by lowered resistance against proteolysis.

I propose that SadB binds three nascent SadA polypeptide chains during or after translocation by the Sec machinery. This would presumably support synchronization of the export of the three long, highly repetitive and at that stage unfolded autotransporter subunits by reducing out-of-register interactions between them. Such a binding must necessarily be rather weak and easily broken and

reformed as the chains are threaded out through the outer membrane. A longer retention of unfolded passenger domains at the bacterial inner membrane to achieve productive autotransport has also been suggested for the unusually long signal peptides frequently found in autotransporters (Szabady et al. 2005). SadB could take an analogous supporting role during the export of SadA. The fact that no homologs of SadB could be found outside of enterobacteria, and no similar lipoproteins were detected in autotransporter-containing operons in other species suggests that it is a specific development of enterobacteria to support the export of a specific class of TAAs. Why homologs of SadA require a helper protein for proper surface display, whereas other complex autotransporters such as *Bartonella* BadA do not appear to have one, or whether analogous system exist for other autotransporters will be an interesting subject for future research.

6 Literature

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp.3389–402.

Arnold, T., Poynor, M., Nussberger, S., Lupas, A.N. & Linke, D., 2007. Gene duplication of the eight-stranded beta-barrel OmpX produces a functional pore: a scenario for the evolution of transmembrane beta-barrels. *Journal of molecular biology*, 366(4), pp.1174–84.

Benz, I. & Schmidt, M.A., 1992. AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is synthesized via a precursor molecule. *Molecular microbiology*, 6(11), pp.1539–46.

Biegert, A., Mayer, C., Remmert, M., Söding, J. & Lupas, A.N., 2006. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic acids research*, 34(Web Server issue), pp.W335–9.

Burkinshaw, B.J. & Strynadka, N.C.J., 2014. Assembly and structure of the T3SS. *Biochimica et biophysica acta*, 1843(8), pp.1649–63.

Busch, A. & Waksman, G., 2012. Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1112–22.

Christie, P.J., Whitaker, N. & González-Rivera, C., 2014. Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et biophysica acta*, 1843(8), pp.1578–91.

Clantin, B., Delattre, A.-S., Rucktooa, P., Saint, N., Méli, A.C., Locht, C., Jacob-Dubuisson, F. & Villeret, V., 2007. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science (New York, N.Y.)*, 317(5840), pp.957–61.

Diepold, A. & Wagner, S., 2014. Assembly of the bacterial type III secretion machinery. *FEMS microbiology reviews*, 38(4), pp.802–22.

Douzi, B., Filloux, A. & Voulhoux, R., 2012. On the path to uncover the bacterial type II secretion system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1059–72.

Eckert, D.M., Malashkevich, V.N. & Kim, P.S., 1998. Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *Journal of molecular biology*, 284(4), pp.859–65.

Elgueta, R., Benson, M.J., de Vries, V.C., Wasiuk, A., Guo, Y. & Noelle, R.J., 2009. Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunological reviews*, 229(1), pp.152–72.

Evans, M.L. & Chapman, M.R., 2014. Curli biogenesis: order out of disorder. *Biochimica et biophysica acta*, 1843(8), pp.1551–8.

Fairman, J.W., Dautin, N., Wojtowicz, D., Liu, W., Noinaj, N., Barnard, T.J., Udho, E., Przytycka, T.M., Cherezov, V. & Buchanan, S.K., 2012. Crystal structures of the outer membrane domain of intimin and invasin from enterohemorrhagic E. coli and enteropathogenic Y. pseudotuberculosis. *Structure (London, England : 1993)*, 20(7), pp.1233–43.

Fröbel, J., Rose, P. & Müller, M., 2012. Twin-arginine-dependent translocation of folded proteins. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1029–46.

Geibel, S. & Waksman, G., 2014. The molecular dissection of the chaperone-usher pathway. *Biochimica et biophysica acta*, 1843(8), pp.1559–67.

Grin, I., Hartmann, M.D., Sauer, G., Hernandez Alvarez, B., Schütz, M., Wagner, S., Madlung, J., Macek, B., Felipe-Lopez, A., Hensel, M., Lupas, A. & Linke, D., 2014. A trimeric lipoprotein assists in trimeric autotransporter biogenesis in enterobacteria. *The Journal of biological chemistry*, 289(11), pp.7388–98.

Grin, I. & Linke, D., 2011. GCView: the genomic context viewer for protein homology searches. *Nucleic acids research*, 39(Web Server issue), pp.W353–6.

Grosskinsky, U., Schütz, M., Fritz, M., Schmid, Y., Lamparter, M.C., Szczesny, P., Lupas, A.N., Autenrieth, I.B. & Linke, D., 2007. A conserved glycine residue of trimeric autotransporter domains plays a key role in Yersinia adhesin A autotransport. *Journal of bacteriology*, 189(24), pp.9011–9.

Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N. & Hernandez Alvarez, B., 2012. Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51), pp.20907–12.

Hartmann, M.D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D.N., Sauer, G., Dunin-Horkawicz, S., Lupas, A.N. & Alvarez, B.H., 2009. A coiled-coil motif that sequesters ions to the hydrophobic core. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), pp.16950–5.

Von Heijne, G., 1990. The signal peptide. *The Journal of Membrane Biology*, 115(3), pp.195–201.

Henderson, I.R. & Nataro, J.P., 2001. Virulence functions of autotransporter proteins. *Infection and immunity*, 69(3), pp.1231–43.

Henderson, I.R., Navarro-Garcia, F., Desvaux, M., Fernandez, R.C. & Ala'Aldeen, D., 2004. Type V protein secretion pathway: the autotransporter story. *Microbiology and molecular biology reviews : MMBR*, 68(4), pp.692–744.

Holm, L. & Rosenström, P., 2010. Dali server: conservation mapping in 3D. *Nucleic acids research*, 38(Web Server issue), pp.W545–9.

Houben, E.N.G., Korotkov, K. V & Bitter, W., 2014. Take five - Type VII secretion systems of Mycobacteria. *Biochimica et biophysica acta*, 1843(8), pp.1707–16.

Humphries, A.D., Raffatellu, M., Winter, S., Weening, E.H., Kingsley, R.A., Droleskey, R., Zhang, S., Figueiredo, J., Khare, S., Nunes, J., Adams, L.G., Tsolis, R.M. & Bäumler, A.J., 2003. The use of flow cytometry to detect expression of subunits encoded by 11 Salmonella enterica serotype Typhimurium fimbrial operons. *Molecular microbiology*, 48(5), pp.1357–76.

Julio, S.M. & Cotter, P.A., 2005. Characterization of the filamentous hemagglutinin-like protein FhaS in Bordetella bronchiseptica. *Infection and immunity*, 73(8), pp.4960–71.

Junker, M., Besingi, R.N. & Clark, P.L., 2009. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Molecular microbiology*, 71(5), pp.1323–32.

Junker, M., Schuster, C.C., McDonnell, A. V, Sorg, K.A., Finn, M.C., Berger, B. & Clark, P.L., 2006. Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 103(13), pp.4918–23.

Karp, P.D., Ouzounis, C.A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahrén, D., Tsoka, S., Darzentas, N., Kunin, V. & López-Bigas, N., 2005. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic acids research*, 33(19), pp.6083–9.

Kersey, P.J., Lawson, D., Birney, E., Derwent, P.S., Haimel, M., Herrero, J., Keenan, S., Kerhornou, A., Koscielny, G., Kähäri, A., Kinsella, R.J., Kulesha, E., Maheswari, U., Megy, K., Nuhn, M., Proctor, G., Staines, D., Valentin, F., Vilella, A.J. & Yates, A., 2010. Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic acids research*, 38(Database issue), pp.D563–9.

Klauser, T., Pohlner, J. & Meyer, T.F., 1992. Selective extracellular release of cholera toxin B subunit by Escherichia coli: dissection of Neisseria Iga beta-mediated outer membrane transport. *The EMBO journal*, 11(6), pp.2327–35.

Knowles, T.J., Scott-Tucker, A., Overduin, M. & Henderson, I.R., 2009. Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nature reviews. Microbiology*, 7(3), pp.206–14.

Korbel, J.O., Jensen, L.J., von Mering, C. & Bork, P., 2004. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nature biotechnology*, 22(7), pp.911–7.

Kuhn, A., 2014. Introduction to special issue on protein trafficking and secretion. *Biochimica et biophysica acta*, 1843(8), pp.1429–32.

Leininger, E., Roberts, M., Kenimer, J.G., Charles, I.G., Fairweather, N., Novotny, P. & Brennan, M.J., 1991. Pertactin, an Arg-Gly-Asp-containing Bordetella pertussis surface protein that promotes adherence of mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 88(2), pp.345–9.

Leo, J.C., Grin, I. & Linke, D., 2012. Type V secretion: mechanism(s) of autotransport through the

bacterial outer membrane. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1088–101.

Long, F., Su, C.-C., Lei, H.-T., Bolla, J.R., Do, S. V & Yu, E.W., 2012. Structure and mechanism of the tripartite CusCBA heavy-metal efflux complex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1047–58.

Lycklama A Nijeholt, J.A. & Driessen, A.J.M., 2012. The bacterial Sec-translocase: structure and mechanism. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1016–28.

Malinverni, J.C. & Silhavy, T.J., 2011. Assembly of Outer Membrane β-Barrel Proteins: the Bam Complex. *EcoSal Plus*, 2011.

Mazar, J. & Cotter, P.A., 2006. Topology and maturation of filamentous haemagglutinin suggest a new model for two-partner secretion. *Molecular microbiology*, 62(3), pp.641–54.

McWhirter, S.M., Pullen, S.S., Holton, J.M., Crute, J.J., Kehry, M.R. & Alber, T., 1999. Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proceedings of the National Academy of Sciences of the United States of America*, 96(15), pp.8408–13.

Meng, G., Surana, N.K., St Geme, J.W. & Waksman, G., 2006. Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *The EMBO journal*, 25(11), pp.2297–304.

Moran, C.P. & Shanahan, F., 2014. Gut microbiota and obesity: role in aetiology and potential therapeutic target. *Best practice & research. Clinical gastroenterology*, 28(4), pp.585–97.

Müller, J.E.N., Papic, D., Ulrich, T., Grin, I., Schütz, M., Oberhettinger, P., Tommassen, J., Linke, D., Dimmer, K.S., Autenrieth, I.B. & Rapaport, D., 2011. Mitochondria can recognize and assemble fragments of a beta-barrel structure. *Molecular biology of the cell*, 22(10), pp.1638–47.

Nivaskumar, M. & Francetic, O., 2014. Type II secretion system: a magic beanstalk or a protein escalator. *Biochimica et biophysica acta*, 1843(8), pp.1568–77.

Nwosu, F.C., Avershina, E., Wilson, R. & Rudi, K., 2014. Gut Microbiota in HIV Infection: Implication for Disease Progression and Management. *Gastroenterology research and practice*, 2014, p.803185.

Oberhettinger, P., Schütz, M., Leo, J.C., Heinz, N., Berger, J., Autenrieth, I.B. & Linke, D., 2012. Intimin and invasin export their C-terminus to the bacterial cell surface using an inverse mechanism compared to classical autotransport. *PloS one*, 7(10), p.e47069.

Oomen, C.J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. & Gros, P., 2004. Structure of the translocator domain of a bacterial autotransporter. *The EMBO journal*, 23(6), pp.1257–66.

Overbeek, R. et al., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic acids research*, 33(17), pp.5691–702.

Paetzel, M., 2014. Structure and mechanism of Escherichia coli type I signal peptidase. *Biochimica et biophysica acta*, 1843(8), pp.1497–508.

Paramasivam, N., Habeck, M. & Linke, D., 2012. Is the C-terminal insertional signal in Gram-negative bacterial outer membrane proteins species-specific or not? *BMC genomics*, 13, p.510.

Peterson, J.H., Tian, P., Ieva, R., Dautin, N. & Bernstein, H.D., 2010. Secretion of a bacterial virulence factor is driven by the folding of a C-terminal segment. *Proceedings of the National Academy of Sciences of the United States of America*, 107(41), pp.17739–44.

Plaut, A.G., Gilbert, J. V, Artenstein, M.S. & Capra, J.D., 1975. Neisseria gonorrhoeae and neisseria meningitidis: extracellular enzyme cleaves human immunoglobulin A. *Science (New York, N.Y.)*, 190(4219), pp.1103–5.

Pohlner, J., Halter, R., Beyreuther, K. & Meyer, T.F., 1987. Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA protease. *Nature*, 325(6103), pp.458–62.

Price, M.N., Arkin, A.P. & Alm, E.J., 2006. The life-cycle of operons. *PLoS genetics*, 2(6), p.e96.

Redinbo, M.R., 2014. The Microbiota, Chemical Symbiosis, and Human Disease. *Journal of molecular biology*.

Remaut, H., Tang, C., Henderson, N.S., Pinkner, J.S., Wang, T., Hultgren, S.J., Thanassi, D.G., Waksman, G. & Li, H., 2008. Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell*, 133(4), pp.640–52.

Remmert, M., Biegert, A., Linke, D., Lupas, A.N. & Söding, J., 2010. Evolution of outer membrane beta-barrels from an ancestral beta beta hairpin. *Molecular biology and evolution*, 27(6), pp.1348–58.

Remmert, M., Linke, D., Lupas, A.N. & Söding, J., 2009. HHomp--prediction and classification of outer membrane proteins. *Nucleic acids research*, 37(Web Server issue), pp.W446–51.

Robert, V., Volokhina, E.B., Senf, F., Bos, M.P., Van Gelder, P. & Tommassen, J., 2006. Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS biology*, 4(11), p.e377.

Roussel-Jazédé, V., Grijpstra, J., van Dam, V., Tommassen, J. & van Ulsen, P., 2013. Lipidation of the autotransporter NalP of Neisseria meningitidis is required for its function in the release of cell-surface-exposed proteins. *Microbiology (Reading, England)*, 159(Pt 2), pp.286–95.

Salacha, R., Kovacić, F., Brochier-Armanet, C., Wilhelm, S., Tommassen, J., Filloux, A., Voulhoux, R. & Bleves, S., 2010. The Pseudomonas aeruginosa patatin-like protein PlpD is the archetype of a novel Type V secretion system. *Environmental microbiology*, 12(6), pp.1498–512.

Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M.J., Rhodes, R.G. & Nakayama, K., 2010. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(1), pp.276–81.

Sauri, A., Soprova, Z., Wickström, D., de Gier, J.-W., Van der Schors, R.C., Smit, A.B., Jong, W.S.P. & Luirink, J., 2009. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology (Reading, England)*, 155(Pt 12), pp.3982–91.

Schwarz, S., Hood, R.D. & Mougous, J.D., 2010. What is type VI secretion doing in all those bugs? *Trends in microbiology*, 18(12), pp.531–7.

Selkrig, J., Leyton, D.L., Webb, C.T. & Lithgow, T., 2014. Assembly of β-barrel proteins into bacterial outer membranes. *Biochimica et biophysica acta*, 1843(8), pp.1542–50.

Selkrig, J., Mosbahi, K., Webb, C.T., Belousoff, M.J., Perry, A.J., Wells, T.J., Morris, F., Leyton, D.L., Totsika, M., Phan, M.-D., Celik, N., Kelly, M., Oates, C., Hartland, E.L., Robins-Browne, R.M., Ramarathinam, S.H., Purcell, A.W., Schembri, M.A., Strugnell, R.A., Henderson, I.R., Walker, D. & Lithgow, T., 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature structural & molecular biology*, 19(5), pp.506–10, S1.

Söding, J., Biegert, A. & Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic acids research*, 33(Web Server issue), pp.W244–8.

Soprova, Z., Sauri, A., van Ulsen, P., Tame, J.R.H., den Blaauwen, T., Jong, W.S.P. & Luirink, J., 2010. A conserved aromatic residue in the autochaperone domain of the autotransporter Hbp is critical for initiation of outer membrane translocation. *The Journal of biological chemistry*, 285(49), pp.38224–33.

Szabady, R.L., Peterson, J.H., Skillman, K.M. & Bernstein, H.D., 2005. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proceedings of the National Academy of Sciences of the United States of America*, 102(1), pp.221–6.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J. & von Mering, C., 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*, 39(Database issue), pp.D561–8.

Thein, M., Sauer, G., Paramasivam, N., Grin, I. & Linke, D., 2010. Efficient subfractionation of Gram-negative bacteria for proteomics studies. *Journal of proteome research*.

Thomas, S., Holland, I.B. & Schmitt, L., 2014. The Type 1 secretion pathway - the hemolysin system and beyond. *Biochimica et biophysica acta*, 1843(8), pp.1629–41.

Tsai, J.C., Yen, M.-R., Castillo, R., Leyton, D.L., Henderson, I.R. & Saier, M.H., 2010. The bacterial intimins and invasins: a large and novel family of secreted proteins. *PloS one*, 5(12), p.e14403.

Ulrich, T., Oberhettinger, P., Schütz, M., Holzer, K., Ramms, A.S., Linke, D., Autenrieth, I.B. & Rapaport, D., 2014. Evolutionary Conservation in Biogenesis of β-Barrel Proteins Allows Mitochondria to Assemble a Functional Bacterial Trimeric Autotransporter Protein. *The Journal of biological chemistry*, 289, pp.29457–29470.

Van Ulsen, P., van Alphen, L., ten Hove, J., Fransen, F., van der Ley, P. & Tommassen, J., 2003. A Neisserial autotransporter NalP modulating the processing of other autotransporters. *Molecular microbiology*, 50(3), pp.1017–30.

ur Rahman, S., Arenas, J., Öztürk, H., Dekker, N. & van Ulsen, P., 2014. The polypeptide transport-associated (POTRA) domains of TpsB transporters determine the system specificity of two-partner secretion systems. *The Journal of biological chemistry*, 289(28), pp.19799–809.

Waksman, G., 2012. Bacterial secretion comes of age. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1014–5.

Zoued, A., Brunet, Y.R., Durand, E., Aschtgen, M.-S., Logger, L., Douzi, B., Journet, L., Cambillau, C. & Cascales, E., 2014. Architecture and assembly of the Type VI secretion system. *Biochimica et biophysica acta*, 1843(8), pp.1664–73.

Zückert, W.R., 2014. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochimica et biophysica acta*, 1843(8), pp.1509–16.

7 Published research articles

GCView: the genomic context viewer for protein homology searches

Grin, I. & Linke, D.

2011

Nucleic acids research, 39(Web Server issue), pp.W353–6.

Page 33

A trimeric lipoprotein assists in trimeric autotransporter biogenesis in enterobacteria

Grin, I., Hartmann, M.D., Sauer, G., Hernandez Alvarez, B., Schütz, M., Wagner, S., Madlung, J., Macek, B., Felipe-Lopez, A., Hensel, M., Lupas, A. & Linke, D.

2014

The Journal of biological chemistry, 289(11), pp.7388–98.

Page 37

Published online 23 May 2011 *Nucleic Acids Research, 2011, Vol. 39, Web Server issue* ***W353–W356***
doi:10.1093/nar/gkr364

GCView: the genomic context viewer for protein homology searches

Iwan Grin and Dirk Linke*

Max Planck Institute for Developmental Biology, Department I, Protein Evolution, Spemannstr. 35, 72076 Tübingen, Germany

Received February 18, 2011; Revised April 18, 2011; Accepted April 27, 2011

ABSTRACT

Genomic neighborhood can provide important insights into evolution and function of a protein or gene. When looking at operons, changes in operon structure and composition can only be revealed by looking at the operon as a whole. To facilitate the analysis of the genomic context of a query in multiple organisms we have developed Genomic Context Viewer (GCView). GCView accepts results from one or multiple protein homology searches such as BLASTp as input. For each hit, the neighboring protein-coding genes are extracted, the regions of homology are labeled for each input and the results are presented as a clear, interactive graphical output. It is also possible to add more searches to iteratively refine the output. GCView groups outputs by the hits for different proteins. This allows for easy comparison of different operon compositions and structures. The tool is embedded in the framework of the Bioinformatics Toolkit of the Max-Planck Institute for Developmental Biology (MPI Toolkit). Job results from the homology search tools inside the MPI Toolkit can be forwarded to GCView and results can be subsequently analyzed by sequence analysis tools. Results are stored online, allowing for later reinspection. GCView is freely available at http://toolkit.tuebingen.mpg.de/gcview.

INTRODUCTION

In bacterial and archaeal genomes, about one half of all protein-coding genes are organized into operons. (1). But even for the other half, conservation of the genomic context i.e. the genes upstream and downstream on the chromosome, is observable between related species (2). The genomic context can provide important information about duplication, insertion, translocation or deletion events. While the past decades have equipped scientists with a broad range of excellent bioinformatics tools for analysis and comparison of single protein sequences, taking a step back and looking at the bigger genomic picture and comparing it between different organisms is still largely manual work. For many well annotated proteins and operons, databases like BioCyc (3), STRING (4), The SEED (5) or Ensembl Bacteria (6) can provide important information. However, looking beyond the content of those databases to extend the search into more genomes or investigating less well-characterized proteins can be challenging.

GCView, the Genomic Context Viewer for protein homology searches aims to ease and automate the manual process of extracting and comparing genomic regions of interest. It is integrated into the Bioinformatics Toolkit of the Max-Planck Institute for Developmental Biology (MPI Toolkit) (7) and can be accessed through a user-friendly web interface at http://toolkit.tuebingen.mpg.de/gcview. This website is free and open to all users and there is no login requirement.

GCView uses protein homology to assign corresponding genes. The underlying homology information is taken from standard protein homology search tools like BLASTp or PSI-BLAST (8). In contrast to the above mentioned databases such as STRING, the homology searches are not precomputed, giving the user full control over and insight into the processes leading to the final result.

GCView can integrate multiple searches (e.g. one for each component of an operon) and compile a comprehensive overview of the combinatorial variants found in different genomes. Genomes featuring the same number and order of genes of interest are grouped together.

The results can be mapped onto a taxonomy tree for a quick overview of the distribution of operon structures throughout all sequenced procaryotic organisms.

The output is a series of images showing the genomic regions that contain the genes of interest. Additionally, for each image a list of the encoded proteins is provided that contains additional information such as descriptions and database links. Hits from the underlying searches are colored in the output for easy identification.

*To whom correspondence should be addressed. Tel: +49 7071 601357; Fax: +49 7071 601349; Email: dirk.linke@tuebingen.mpg.de

The integration into the MPI Toolkit allows users to run homology search jobs independent of GCView, providing maximum control over the input parameters, and then to internally forward the results to GCView for integration. Consequently, the results from GCView can also be forwarded to other specialized tools for a more detailed analysis of subsets of proteins or genes. All results are stored on the server for 2 weeks and can be revisited and reviewed at a later time point. It is possible to create an account on the MPI Toolkit, which allows jobs to be bound to the account and saved for extended periods of time.

FUNCTIONALITY

The design goal for GCView was to provide a quick and accurate overview of the combinatorial variants of operons in different genomes based on well established homology search methods accessible through a user-friendly straightforward web interface. The workflow of the tool is summarized in Figure 1.

Input

GCView accepts several different types of input: FASTA protein sequences, protein GI or UniProt identifiers and forwarded homology search jobs. Currently GCView is limited to protein homology searches or protein sequences as input, mostly due to the higher sensitivity of protein searches compared to DNA searches. The inclusion of DNA searches (BLASTn) is planned for a future version. It is possible to use not only full protein sequences, but also single domains as query for the search. Genes containing multiple domains will be labeled accordingly in the output.

Primarily, homology search jobs can be forwarded to GCView within the MPI Toolkit. If, alternatively, FASTA sequences or protein identifiers are provided, GCView internally executes a PSI-BLAST run for each sequence or identifier provided and analyzes the results. Additional input parameters are the size of the genomic region to be displayed and the *E*-value cutoffs for the results to be included in the output. The size of the genomic region is interpreted as the number of genes to be extracted before the first hit and after the last hit in any genome.

Note that the quality of the GCView results strongly depends on the underlying homology search being exhaustive, i.e. containing results at least up to the *E*-value cutoff specified for GCView. This is especially important in Group View: only exhaustive searches lead to a maximum of labeled operon components. Operons with unlabeled components lead to additional groups, which would not be observed after an exhaustive search. For the same reason, caution is advised when using BLAST databases prefiltered at a certain sequence similarity cutoff.

For technical reasons, it is only possible to use BLAST databases, which contain GI or UniProt identifiers. Using a database which does not provide appropriate identifiers in the output will not give any results in GCView.

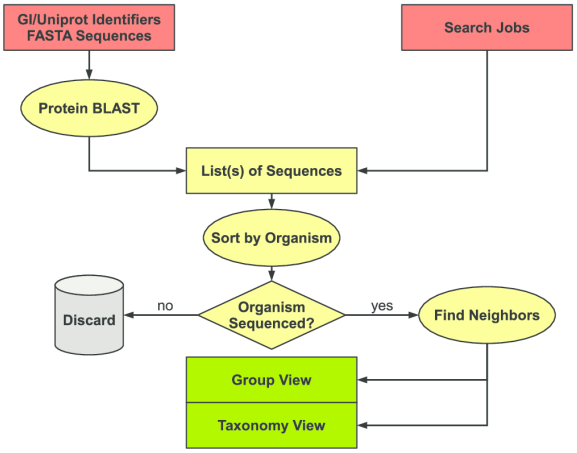

Figure 1. GCView workflow. Input: red; processing: yellow and results: green.

Processing

From each input homology search, a list of protein GI numbers is extracted along with the exact region and degree of similarity. The lists are filtered for proteins with *E*-values below the threshold specified in the input and for proteins from organisms which have not been fully sequenced. The backend database of sequenced genome data is built from the genomes found in NCBI GenBank (9) (ftp://ftp.ncbi.nih.gov/genomes/Bacteria) and comprises fully sequenced bacterial and archaeal genomes.

For each hit the genes upstream and downstream of the hit are extracted from the database, resulting in one genome chunk for each hit. The number of genes extracted depends on the range set in the input parameters. Overlapping regions from the same genome are subsequently merged. This implies that an operon which has been duplicated in a genome can show up as one or two chunks, depending on the distance between the duplicates and the range settings. After merging, the resulting regions are grouped by the number and order of genes of interest.

OUTPUT

GCView generates two different views for the results: the Group View and the Taxonomy View. Both views contain the same information the difference is in the sorting. Figure 2 shows example outputs for both views for two different runs of GCView.

The Group View presents an overview of the results. A group comprises all organisms which contain a specific number and order of the genes of interest.

A schematic image of each group summarizes which of the genes of interest can be found in the group and in which order they appear in the genome. Each query gene is represented by a colored arrow. The colors are explained in the legend, which is displayed on the top of the page. Additionally, the identifier of the input query is indicated on each arrow. The arrows in the Group View

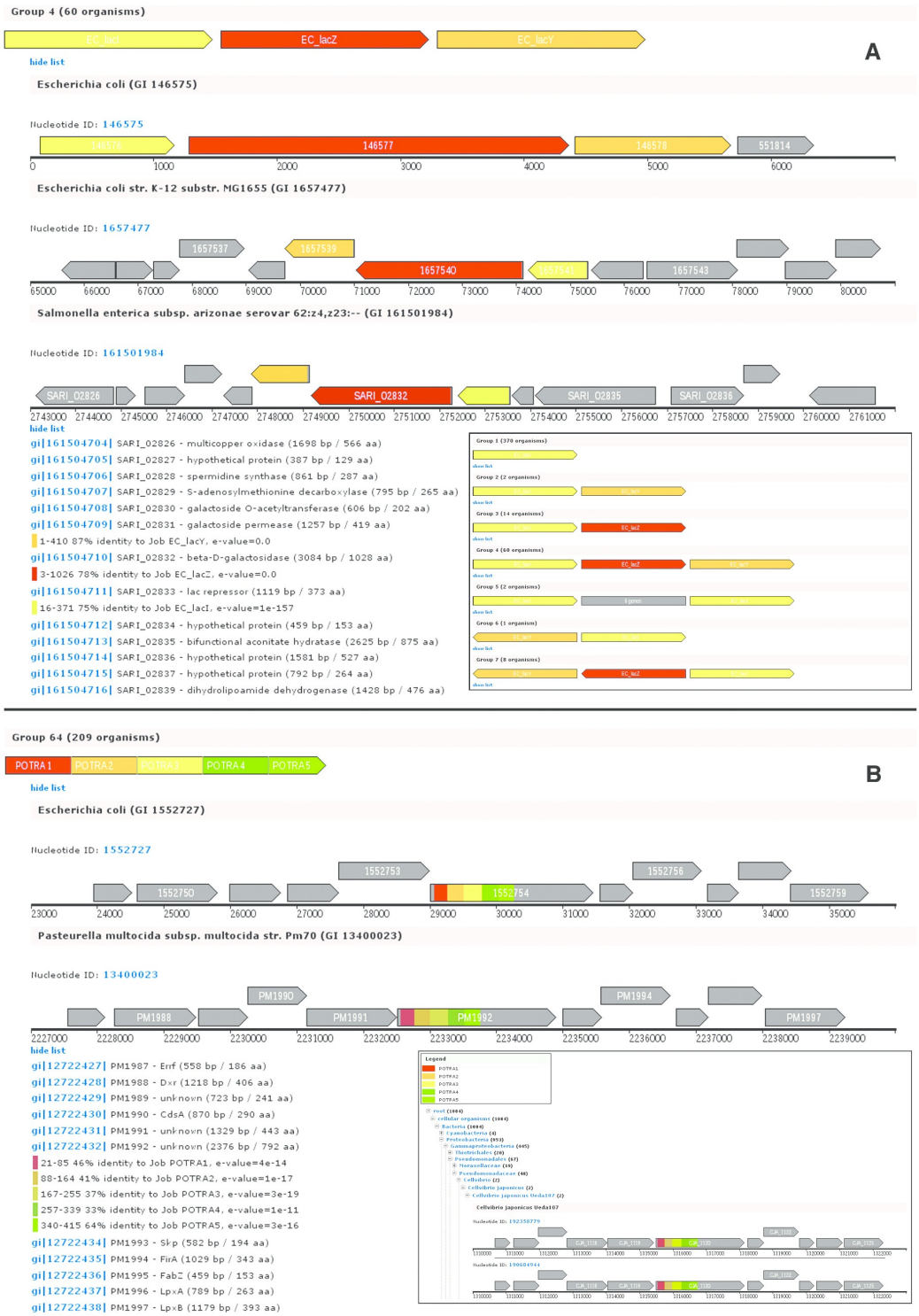

Figure 2. Example output. (**A**) Using GCView to look at different operon components. The lac Operon (Demo Data) is shown in Group View with one group expanded. Insert: Group View Overview for the same run. (**B**) Using GCView to look at single domains in different contexts. POTRA domains from Omp85 and related proteins (10) in different organisms shown in Group View. Insert: Taxonomy View for the same run.

are not to scale and the colors do not indicate the degree of identity between query and hit sequences. Fused arrows indicate that multiple query sequences were mapped onto one gene.

Gray boxes represent one or multiple genes that are not homologous to any of the query sequences but located between genes of interest. A number indicates how many genes are represented by the corresponding box. The groups can be expanded to view the detailed genomic context for each organism in the respective group.

The Taxonomy View maps all results onto a taxonomy tree. The numbers next to the organism names represent the number of hits in this taxon and its sub-groups. Branches of the tree can be collapsed or expanded as required. The detailed information for each hit can be viewed at the leaves of the tree.

The detail representation of every genomic region is identical in both views. Each representation contains a genome ID, indicating the nucleotide GI number of the genome from which the corresponding region was extracted. In the case that genes of interest are located in several non-overlapping regions of the same genome (e.g. due to operon duplication), multiple representations with the same nucleotide ID are shown, one for each region.

A schematic image of the region shows the genetic neighborhood of the genes of interest. Protein-coding genes are shown as arrows. Regions of homology to the genes of interest are highlighted in the corresponding colors, which are indicated in the legend. In contrast to the Group View, the intensity of the color corresponds to the identity score of the hit and the arrow length correlates with the length of the gene. Please note that the scale may differ between different images. The ruler at the bottom of each image shows the position in the genome. Each section of the ruler corresponds to 1000 bp. Various details for each gene (description, precise location, length, distance to neighboring genes) can be viewed by hovering the mouse over the arrows.

Clicking on an arrow expands a detailed list of the genes in the image and the search hits therein. The selected gene is highlighted in the list. A clipboard widget located in the right corner of the screen can be used to pick genes from the output. These genes can be forwarded to sequence retrieval tools for further analysis or used in another GCView run for an iterative expansion of the set of analyzed genes.

CONCLUSIONS

We present GCView, an interactive web tool for automated retrieval and comparison of the genomic context of protein-coding genes. The underlying homology searches use protein sequences instead of DNA for higher sensitivity. Compared to classical databases like The SEED or BioCyc, the advantages of GCView are: (i) a greater focus on the query, as only the homologs of the input proteins are highlighted, and the degree of similarity is easily visible from the output; (ii) interactivity, as the query can iteratively be extended by more proteins of interest; (iii) transparency, as the user can have full control over the parameters of the underlying homology search; and (iv) flexibility, as e.g. single domains can be used as query, revealing different domain contexts. GCView is embedded into the MPI Toolkit, which allows users to save their GCView runs for later reinspection and directly analyze the genes found by GCView using a broad range of sequence and structure analysis tools.

ACKNOWLEDGEMENTS

The authors wish to thank the people involved in the maintenance of the MPI Toolkit, especially Christina Wassermann, Vikram Alva, and Andé Noll, and furthermore Andrei Lupas for continuing support.

FUNDING

Funding for the project as well as for open access charge: Departmental funding of the Max Planck Society.

Conflict of interest statement. None declared.

REFERENCES

1. Price,M.N., Arkin,A.P. and Alm,E.J. (2006) The life-cycle of operons. *PLoS Genet.*, **2**, e96.
2. Korbel,J.O., Jensen,L.J., von Mering,C. and Bork,P. (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nature Biotechnol.*, **22**, 911–917.
3. Karp,P.D., Ouzounis,C.A., Moore-Kochlacs,C., Goldovsky,L., Kaipa,P., Ahren,D., Tsoka,S., Darzentas,N., Kunin,V. and Lopez-Bigas,N. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.*, **33**, 6083–6089.
4. Szklarczyk,D., Franceschini,A., Kuhn,M., Simonovic,M., Roth,A., Minguez,P., Doerks,T., Stark,M., Muller,J., Bork,P. *et al.* (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.
5. Overbeek,R., Begley,T., Butler,R.M., Choudhuri,J.V., Chuang,H.Y., Cohoon,M., de Crecy-Lagard,V., Diaz,N., Disz,T., Edwards,R. *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.*, **33**, 5691–5702.
6. Kersey,P.J., Lawson,D., Birney,E., Derwent,P.S., Haimel,M., Herrero,J., Keenan,S., Kerhornou,A., Koscielny,G., Kahari,A. *et al.* (2010) Ensembl Genomes: extending Ensembl across the taxonomic space. *Nucleic Acids Res.*, **38**, D563–D569.
7. Biegert,A., Mayer,C., Remmert,M., Soding,J. and Lupas,A.N. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.*, **34**, W335–W339.
8. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
9. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2011) GenBank. *Nucleic Acids Res.*, **39**, D32–D37.
10. Arnold,T., Zeth,K. and Linke,D. (2010) Omp85 from the thermophilic cyanobacterium Thermosynechococcus elongatus differs from proteobacterial Omp85 in structure and domain composition. *J. Biol. Chem.*, **285**, 18003–18015.

⌘ Author's Choice

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 289, NO. 11, pp. 7388–7398, March 14, 2014

A Trimeric Lipoprotein Assists in Trimeric Autotransporter Biogenesis in Enterobacteria*[S]

Received for publication, September 6, 2013, and in revised form, December 16, 2013 Published, JBC Papers in Press, December 25, 2013, DOI 10.1074/jbc.M113.513275

Iwan Grin[‡§], Marcus D. Hartmann[‡1], Guido Sauer[‡2], Birte Hernandez Alvarez[‡], Monika Sch tz[§], Samuel Wagner[§3], Johannes Madlung[¶], Boris Macek[¶], Alfonso Felipe-Lopez[‖4], Michael Hensel[‖], Andrei Lupas[‡], and Dirk Linke[‡5]

From the [‡]Max Planck Institut f r Entwicklungsbiologie, 72076 T bingen, the [§]Institut f r Medizinische Mikrobiologie and Hygiene and the [¶]Proteome Center T bingen, Universit t T bingen, 72076 T bingen, and the [‖]Abteilung Mikrobiologie, Fachbereich Biologie/Chemie, Universit t Osnabr ck, 49076 Osnabr ck, Germany

Background: Autotransporter adhesins reach the bacterial cell surface by a complex mechanism.
Results: In the case of the autotransporter SadA from *Salmonella*, a lipoprotein assists in surface display.
Conclusion: The similarity to eukaryotic MATH domains suggests that the lipoprotein assists in trimerization of SadA.
Significance: Understanding the similarities between autotransport systems might lead to new ways of inhibiting bacterial adhesion.

Trimeric autotransporter adhesins (TAAs) are important virulence factors of many Gram-negative bacterial pathogens. TAAs form fibrous, adhesive structures on the bacterial cell surface. Their N-terminal extracellular domains are exported through a C-terminal membrane pore; the insertion of the pore domain into the bacterial outer membrane follows the rules of β-barrel transmembrane protein biogenesis and is dependent on the essential Bam complex. We have recently described the full fiber structure of SadA, a TAA of unknown function in *Salmonella* and other enterobacteria. In this work, we describe the structure and function of SadB, a small inner membrane lipoprotein. The *sadB* gene is located in an operon with *sadA*; orthologous operons are only found in enterobacteria, whereas other TAAs are not typically associated with lipoproteins. Strikingly, SadB is also a trimer, and its co-expression with SadA has a direct influence on SadA structural integrity. This is the first report of a specific export factor of a TAA, suggesting that at least in some cases TAA autotransport is assisted by additional periplasmic proteins.

The outer membrane of Gram-negative bacteria forms the outermost barrier between the bacterial cell and the outside world. As such, the role of the outer membrane is astoundingly complex. It acts as a protective barrier against harmful substances such as antibiotics, bacteriocins, and, especially for pathogenic bacteria, also against factors of the host immune system. At the same time it permits interaction with the outside, such as uptake of nutrients, export of secreted factors, as well as sensing and adhesion. These very different functions are typically performed by transmembrane β-barrel proteins, the major protein family present in the outer membrane. This diverse group of evolutionarily related integral membrane proteins (1, 2) is characterized by its three-dimensional structure, a cylinder formed by antiparallel β-strands spanning the outer membrane and connected by loops on either side. The pore inside of a β-barrel allows the passage of small molecules such as ions or nutrients but also protein domains or whole proteins (3).

A special class of β-barrel proteins are monomeric and trimeric autotransporter proteins, commonly referred to as type Va and type Vc secretion systems, respectively (4, 5). These are large proteins (often of 3000 residues or more) consisting of a translocator domain, a typically 12-stranded barrel that is inserted into the outer membrane, and a passenger part, which is exported to the bacterial surface through the pore formed by the translocator domain. The translocator domain acts as an export pore as well as an anchor, tethering the exported passenger to the bacterial surface; in several cases of monomeric autotransporters, the passenger domain is proteolytically cleaved and released into the extracellular space post-translocationally (6). In the case of trimeric autotransporters, each monomer of the homotrimer contributes four strands to the barrel, through which then all three passengers are exported.

Although the passengers of monomeric autotransporters can be structurally and functionally rather diverse, trimeric autotransporter passengers are trimeric coiled-coil structures interspersed with a limited number of domains, which are thought to modulate the flexibility of the otherwise rigid fiber, or provide adhesion to abiotic surfaces, biopolymers (*e.g.* collagen or fibronectin), and host cell surface structures. The ratio of coiled-coil segments to other domains varies dramatically between different trimeric autotransporters. *Yersinia* YadA consists of a single extended coiled coil with only one head domain at its end, and others such as *Haemophilus* Hia have a low content in coiled-coil segments (7). With adhesion being their main function, this group of proteins is commonly referred to as trimeric autotransporter adhesins (TAAs)[6] (8).

* This work was supported in part by Collaborative Research Center Grant SFB766 (German Science Foundation) (to D. L.) and by institutional funds from the Max Planck Society.
⌘ *Author's Choice*—Final version full access.
[S] This article contains supplemental material 1 and 2.
[1] Supported by Collaborative Research Center Grant SFB944 from the German Science Foundation.
[2] Present address: SGS M-SCAN GmbH, Engesserstrasse 4a, 79108 Freiburg, Germany.
[3] Supported by the Alexander von Humboldt Foundation through a Sofja Kovalevskaja Award.
[4] Supported by a Deutscher Akademischer Austauschdienst fellowship.
[5] To whom correspondence should be addressed: Dept. of Biosciences, University of Oslo, 0316 Oslo, Norway. E-mail: dirk.linke@ibv.uio.no.

[6] The abbreviations used are: TAA, trimeric autotransporter adhesin; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; PDB,

The biogenesis of monomeric and trimeric autotransporter adhesins is in large part similar to that of other outer membrane proteins. Both groups have an N-terminal signal peptide targeting them for export into the periplasm by the Sec machinery. A first challenge arises from the fact that the β-barrel domain is typically at the very C terminus of the polypeptide, the part that is translocated last. During this time, the passenger must be kept from (mis-)folding or aggregating in the periplasm. This role is performed by periplasmic chaperones such as Skp and SurA as well as the chaperone/protease DegP. Additionally, in the case of TAAs, which are translocated into the periplasm as monomers, trimerization of the β-barrel segments must occur. The Bam complex then catalyzes the insertion of the β-barrel domain into the outer membrane, upon or during which the passenger is exported, yielding the mature protein (4).

During previous work on the *Salmonella* trimeric autotransporter adhesin SadA, we noted that in the enterobacterial genera *Escherichia*, *Salmonella,* and *Shigella,* the chromosomal location of *sadA* is conserved between the *mtl* operon for mannitol metabolism and the *lld* operon for L-lactate metabolism (9). Further investigation revealed that the adhesin forms an operon with a small predicted lipoprotein (STM3690 in *Salmonella enterica*) encoded directly upstream of the adhesin gene. We named this lipoprotein SadB in *S. enterica*. In this study, we determined the three-dimensional structure of SadB using x-ray crystallography and show that SadB enhances the surface display of SadA, suggesting a direct involvement of SadB in the autotransport mechanism of the trimeric autotransporter adhesin SadA.

MATERIALS AND METHODS

Strains, Plasmids, Primers, and Sequence Data—Sequence information relevant for this work was retrieved from NCBI as follows: SadB, AAL22549.1/gi|16422256; YajI, YP_488704.1/gi|388476518. Strains (Table 1), primers (Table 2), and plasmids (Table 3) used in this study are listed below.

Bioinformatics—The genomic context of *sadB* and its paralogs was investigated using GCView (10) in the MPI Bioinformatics Toolkit (11). Sequences homologous to SadB or YajI were collected from up to three rounds of PSI-BLAST (12) and forwarded to GCView for genomic context lookup and inspection. Genes upstream and downstream of the gene of interest were selected for further iterations of GCView to verify the conservation of the genomic context.

DALI—Upon determination of the three-dimensional structure of SadB, the model was submitted to the Dali server (13) to search for structurally similar proteins. The query consisted of either the full model or just the C-terminal domain (residues 90–213).

Cloning—All primers for pASK IBA vectors were designed using Primer D'Signer 1.1 software.

The *sadB* gene from *S. typhimurium* was cloned into pASK-IBA3 using primers Fwdlipo and Revlipostop. For cytosolic expression, a construct of *sadB* without the N-terminal signal peptide was created using primers FwdlipoSol and Revlipostop. The *sadBA* operon was cloned into pASK-IBA3 using primers Fwdlipo and RevSadAStop.

Protein Expression and Purification—Cultures were grown at 37 °C in LB medium supplemented with 0.1 mg/ml ampicillin to an A_{600} = 0.6–0.8. Protein expression was subsequently induced by adding 0.2 μg/ml anhydrotetracycline (AHTC). After 4 h, cells were harvested by centrifugation. Cell pellets were resuspended in 20 mM Mops/NaOH, pH 6.5, 50 mM NaCl,

Protein Data Bank; AHTC, anhydrotetracycline.

TABLE 1
Strains used in this study

Designation	Relevant characteristics	Source or Ref.
E. coli TOP 10	Cloning strain	Invitrogen
S. enterica serovar *typhimurium* NCTC12023	Wild type	NCTC
MvP681	ΔSTM3691::aph	This study
MvP682	ΔSTM3691::FRT	This study
MvP788	aph araC PBAD::STM3690	This study
MvP789	FRT araC PBAD::STM3690	This study
MvP790	aph araC PBAD::STM3691	This study
MvP791	FRT araC PBAD::STM3691	This study

TABLE 3
Plasmids used in this study

pASK IBA3	Expression vector	IBA, Germany
pASK IBA3 SadB	*sadB* in pASK-IBA3	This study
pASK IBA3 SadB ΔSignal peptide	*sadBΔ1–22* in pASK-IBA3	This study
pASK IBA2-SadA	*sadA* in pASK-IBA2	40
pASK IBA3 SadBA	*sadBA* in pASK-IBA3	This study
pKD46	Expression of Red recombinase	19
pCP20	Expression of FLP recombinase	19
pBAD-Myc-HisB	Expression vector *araC* P_{BAD}	Invitrogen
p2795	Basis vector for *aph* cassette	21
p3253	*araC* P_{BAD} in p2795	This study
pWRG435	pFPV.25.1 mTagRFP	Roman G. Gerlach

TABLE 2
Primers used in this study

Designation	Sequence 5′ to 3′
Fwdlipo	ATACACGGTCTCAAATGCACAAAAATGGAAAATTTATCCC
Revlipostop	ATACACGGTCTCAGCGCTTTATTTTTTTGGCTTCTTTTTTGTATCG
RevSadAStop	ATACACGGTCTCAGCGCTTTACCACTGGAAGCCCGC
FwdlipoSol	ATGGTAGGTCTCAAATGAGTGATTACTTCGCAGATAAACACC
STM3691-Del-For	ATAATAGCCATTGATACAATTATTTTAGAAAAGGAAATTAGTGTAGGCTGGAGCTGCTTC
STM3691-Del-Rev	TGCCATTGCCTTTGATGTCGGGAGTGTTGTTACTTCATTACATATGAATATCCTCCTTAG
STM3960-Red-BAD-For	TATTTTTATAAAGCATTGCTATGAGCAATTGATAAATAACGTGTAGGCTGGAGCTGCTTCG
STM3690-Red-BAD-Rev	CCAACGCAAGCAGGGGGATAAATTTTCCATTTTTGTGCATGGTTAATTCCTCCTGTTAGC
STM3691-Red-BAD-For	TAACGATACAAAAAAGAAGCCAAAAAAATAATAGCCATTGTGTAGGCTGGAGCTGCTTCG
STM3691-Red-BAD-Rev	CCGTAGCGGCATTCCAGAGGACTTTAAATATTCTATTCATGGTTAATTCCTCCTGTTAGC
STM3690-RedBAD-Check	GGTTGTGTTGATAATCGTAGC
STM-Ctrl-For	AAAGGTCACCGAAGTCGTTG
STM-Ctrl-Rev	GGCAACATAGCCTTTCAGC

10 mM $MgCl_2$ containing protease inhibitor mix (Roche Applied Science) and a pinch of DNase I (Applichem). Bacteria were lysed using a French press. Cell debris and membranes were pelleted by ultracentrifugation at 200,000 × *g* for 45 min at 8 °C. The supernatant was diluted with Buffer A (20 mM Mops/NaOH, pH 6.5, 1 mM EDTA) and loaded onto a cation exchange column (Source S, GE Healthcare). In the case of the SadB construct containing the native lipid anchor, 1% *n*-octyl-polyoxyethylene (Bachem, Buchs, Switzerland) was added to all chromatography buffers. Bound protein was eluted using a linear salt gradient of 0–1 M NaCl in Buffer A. Fractions containing SadB were identified using SDS-PAGE and pooled. The pooled fractions were purified to homogeneity on a Sephadex S75 size exclusion column (GE Healthcare) equilibrated in 20 mM Mops/NaOH, pH 6.5, 50 mM NaCl. Purified protein was stored at 4 °C.

Antibody Purification—Rabbit anti-SadA antibody was described before (9). Rabbit anti-SadB antibody was raised using purified SadB in an in-house facility. The obtained polyclonal serum was affinity purified on a 1-ml HiTrap NHS-activated HP Column (GE Healthcare) according to the manufacturer's manual. Purified SadB was coupled to the column to be used as bait. Antibodies were eluted with 1 M NaCl, 1 M $MgCl_2$, and 4 M $MgCl_2$. Only the 4 M $MgCl_2$ fraction was used for all subsequent experiments.

Cell Shaving—Cells were grown in LB medium supplemented with 0.1 mg/ml ampicillin at 37 °C to an A_{600} = 0.6. Protein expression was induced by addition of 0.2 μg/ml AHTC for 2 h. 1 × 10^9 cells were harvested by centrifugation in a tabletop centrifuge, washed once with PBS, and then resuspended in 100 μl of PBS. 0.2 units of proteinase K were added, and samples were incubated for 10, 30, or 60 min at 37 °C. After incubation, the reaction was stopped by addition of protease inhibitor mix (Roche) and vigorous mixing. Whole cells were spun down in a tabletop centrifuge. The supernatant was analyzed by mass spectrometry as described below.

SDS-PAGE and In-gel Digestion—The supernatants from cell shaving with proteinase K were submitted to a gel run on a one-dimensional SDS-PAGE (NuPAGE 12% precast BisTris gels, Invitrogen). Each gel lane was cut in eight equally sized slices (vertical axis) for in-gel digestion. The proteins were subjected to tryptic in-gel digestion as described previously (14). The resulting peptide mixtures were desalted with C18 Stage Tips before LC/MS measurement.

Liquid Chromatography-Mass Spectrometry (MS) Analysis—Mass spectrometry analysis of the complete, lipidated SadB lipoprotein was performed using an ion trap (HCTultra PTM Discovery, Bruker Daltonics) equipped with a nano-ESI source (Proxeon Biosystems). LC-MS analysis of in-gel digests was performed on a nano-LC (Easy-nLC, Thermo Fisher Scientific) coupled to an LTQ-Orbitrap-XL (Thermo Fisher Scientific) through a nano-LC-MS interface (Proxeon Biosystems), as described previously (14). Peptides were eluted using a segmented gradient of 5–90% HPLC solvent B (80% acetonitrile in 0.5% acetic acid) at a flow rate of 200 nl/min over 57 min. MS data acquisition was conducted in the positive ion mode. The mass spectrometer was operated in the data-dependent mode to automatically switch between MS and MS/MS acquisition. Survey full-scan MS spectra were acquired in the mass range of *m/z* 300–2000 in the Orbitrap mass analyzer at a resolution of 60,000. An accumulation target value of 10^6 charges was set, and the lock mass option was used for internal calibration (15). The 10 most intense ions were sequentially isolated and fragmented in the linear ion trap using collision-induced dissociation at the ion accumulation target value of 5000 and default collision-induced dissociation settings. The ions already selected for MS/MS were dynamically excluded for 90 s. The resulting peptide fragment ions were recorded in the linear ion trap.

Data Processing and Analysis—Raw files were processed using the MaxQuant software (version 1.2.2.9) (16). Raw MS spectra were first processed by the Quant module to generate peak lists. To retrieve peptide sequences from the processed spectra, the integrated Andromeda peptide search engine (17) was utilized. The processed MS spectra were searched against a decoy *Salmonella enterica* subsp. *typhimurium* LT2 database (Uniprot organism 99287 reference proteome as of December 2, 2013) containing 4536 forward protein entries plus the sequences of 248 commonly observed contaminants.

In the database search, carbamidomethylation (Cys) was set as fixed modification, whereas oxidation (Met) and acetylation (protein N termini) were set as variable modifications. The mass tolerances for precursor and fragment ions were set to 6 ppm and 0.5 Da, respectively. A false discovery rate of 1% was set at the peptide, protein level.

Subcellular Localization—To determine the subcellular localization of SadB density gradient centrifugation, method 5 described by Thein *et al.* (18) was used. Briefly, cells expressing SadB with its native signal peptide were grown overnight at 30 °C without addition of AHTC, harvested, and then lysed as described above. The leakiness of the expression system was sufficient to generate usable amounts of natively localized SadB. After ultracentrifugation, the supernatant was discarded, and the membrane pellet was resuspended in 1 ml of 10 mM Tris/HCl, pH 7.0, 15% (w/w) sucrose, 5 mM EDTA. A 30–55% (w/w) continuous sucrose gradient was prepared on a Biocomp Gradient Station (Fredericton, New Brunswick, Canada) in 13-ml centrifuge tubes (SW 41 Ti, Beckman Instruments). All sucrose solutions contained 10 mM Tris/HCl, pH 7.0, 5 mM EDTA. The sample was carefully layered on top of the gradient and centrifuged at 250,000 × *g* for 12–16 h. After centrifugation, the gradient was split into 1-ml fractions that were analyzed by SDS-PAGE and subsequent Western blotting with αSadB antibody. αOmpX (2) and αYidC were used as markers for outer and inner membrane fractions, respectively. α-Rabbit DyLight 800-conjugated antibodies (Pierce) were used as secondary antibody. The membranes were scanned on an Odyssey infrared imaging system and analyzed using Image Studio 2 (LI-COR Biosciences, Lincoln, NE).

Flow Cytometry—Cells were grown in LB medium supplemented with 0.1 mg/ml ampicillin at 37 °C to an A_{600} = 0.6. Protein expression was induced by addition of 0.2 μg/ml AHTC for 2 h. 1 × 10^9 cells were harvested by centrifugation in a tabletop centrifuge, washed with 1% BSA in PBS, and stained with affinity-purified rabbit αSadA (see above) in 1% BSA/PBS for 1 h at 4 °C and subsequently with allophycocyanin-conjugated secondary antibody (1:200, Jackson ImmunoResearch) in 1% BSA/PBS for 1 h at 4 °C in the dark. Surface localization of

SadA was measured by flow cytometry in a BD Biosciences LSR II. Measurements were analyzed using WinMDI (J. Trotter) software. Data are means for three independent experiments.

Fluorescence Microscopy—Samples for immunofluorescence microscopy were prepared using the protocol for FACS (above). Cy3-conjugated goat IgG anti-rabbit IgG (Jackson ImmunoResearch) was used as secondary antibody. After immunolabeling, cells were immobilized on poly-L-lysine-coated coverslips, stained with 0.4 μg/ml DAPI for 10 min in the dark, embedded in Mowiol-DABCO, and examined under a Zeiss Axioplan microscope with an EXFO X-Cite 120 excitation light source.

Generation of S. enterica Strains—*S. enterica* serovar *typhimurium* strain NCTC 12023 was used as wild-type strain, and other *Salmonella* strains used in this study are isogenic (Table 1). A deletion strain in *sadA* (STM3691) was generated by λ Red recombinase-mediated recombination basically as described before (19). For the generation of strains with expression of *sadBA* or only *sadA* under control of the P_{BAD} promoter of the arabinose operon, we used Red recombineering (20, 21). pBAD-myc HisB was digested with NdeI and SacI, and a fragment containing *araC* and P_{BAD} was recovered and subcloned in p2795 (21). The resulting plasmid p3253 served as template vector for generation of a promoter cassette consisting of *aph* flanked by FRT sites, *araC* and P_{BAD}. p3253 was amplified by PCR using STM3960-Red-BAD-For/Rev or STM3961-Red-BAD-For/Rev for chromosomal integration of the P_{BAD} promoter cassette upstream of *sadBA* or *sadA*, respectively.

The proper insertion of the promoter cassette was controlled by PCR using check primers listed in Table 2. If required, the *aph* resistance gene was removed by FLP-mediated recombination. The functionality of the P_{BAD} promoter cassette was confirmed by chromosomal integration upstream of *phoN* and determination of phosphatase activity in response to induction by arabinose (data not shown).

Analysis of SadA Surface Expression in Salmonella—*Salmonella* strains harboring the P_{BAD} promoter cassette upstream of *sadA* or *sadBA* were grown in LB medium overnight and subcultured by 1:31 dilution in LB supplemented with 0.4% glucose or 0.4% arabinose for repression or induction of P_{BAD}, respectively. Cultures were grown at 37 °C with aeration for 4 h. The absorbance of bacterial cultures was adjusted to A_{600} of 0.2 in PBS, and 50 μl of this suspension were dropped on coverslips with 0.02% poly-L-lysine and allowed to dry at room temperature. Subsequently, bacteria were fixed with 3% paraformaldehyde in PBS for 60 min at 37 °C. Fixed samples were washed three times with PBS, and washing was repeated after each incubation step. Samples were stained with a 1:100 dilution of rabbit-α-SadA in blocking solution (PBS, goat serum 2%, BSA 2%) and incubated for 1 h at RT. Bound antibody was labeled with α-goat-α-rabbit Alexa488. Bacteria harbored pWRG435 for constitutive expression of mTagRFP. Finally, samples were mounted on glass slides with Fluoroprep (BioMerieux) and sealed with Entellan (Merck). Stained samples were kept at 4 °C until observation.

Quantification of SadA Surface Expression—Stained samples of *Salmonella* were observed with a Cell Observer® Zeiss microscope with an ×100 objective and N.A. of 1.54. Excitation of Alexa488 was performed by LED illumination at 10% intensity; DAPI was exited with a mercury lamp and the respective excitation/emission filter. Z stacks were acquired from various fields of view with an Axiocam® Zeiss. Images were then processed and analyzed with ZEN 2012. For quantification of the signal of SadA, regions of interest were scored for at least 100 individual bacteria.

TABLE 4
Data collection and refinement statistics

Values in parentheses refer to the highest resolution shell. The Ramachandran plot statistics show the percentage of residues in the most favored/additionally allowed/generously allowed/disallowed regions, respectively, as defined and determined using the program Procheck (27).

	Native	K_2PtCl_4 derivative
Wavelength	1.0 Å	1.071 Å
Space group	H3	H3
Cell dimensions	$a = b = 118.45$, $c = 159.18$	$a = b = 118.53$, $c = 158.84$
Resolution	40 to 2.45 Å (2.60 to 2.45 Å)	40 to 3.0 Å (3.18 to 3.0 Å)
Completeness	99.7% (99.2%)	99.8% (98.7%)
Redundancy	4.0 (3.9)	20.9 (19.9)
$I/\sigma(I)$	19.20 (1.86)	34.95 (3.92)
R_{merge}	4.2 (67.9)	7.2 (87.3)
R_{cryst}/R_{free}	20.8/25.0%	
Ramachandran plot statistics	93.8/5.8/0.2/0.2%	

Phage Display—Phage display experiments were performed using the Ph.D.-7 and Ph.D.-12 Phage Display Libraries from New England Biolabs following the manufacturer's manual. Purified SadB (cytosolic construct without signal peptide) was used as bait. After three rounds of panning, phage DNA from up to 50 clones was isolated (QIAprep Spin M13 kit, Qiagen) and sequenced.

X-ray Crystallography—For crystallization, the protein was concentrated to 8 mg/ml in 20 mM MOPS/HCl, pH 7, 50 mM NaCl. Crystallization trials were performed at 20 °C in 96-well sitting-drop vapor diffusion plates with 50 μl of reservoir solution and drops containing 400 nl of protein solution in addition to 400 nl of reservoir solution. The best diffracting crystals were obtained within 1 week with a reservoir solution containing 0.2 M lithium sulfate, 0.1 M Tris, pH 8.5, and 20%(w/v) PEG 4000. Single crystals were transferred into a cryoprotectant drop containing reservoir solution supplemented with 10% (v/v) PEG 400 before flash-cooling in liquid nitrogen. For experimental phasing, crystals were soaked overnight in a drop containing reservoir solution supplemented with 5 mM K_2PtCl_4 prior to cryoprotection and flash-cooling. A native and a derivative dataset were collected at beamline X10SA (PXII) at the SLS (Paul Scherrer Institute, Villigen, Switzerland) at 100 K using a PILATUS 6M detector (DECTRIS). Diffraction images were processed and scaled using the XDS program suite (22). Using SHELXD (23), four strong platinum sites were identified. After density modification with SHELXE, the resulting electron density map could be traced by Buccaneer (24) to a large extent. The model was completed by cyclic manual modeling with Coot (25) and refinement with Phenix (26). Analysis with Procheck (27) showed excellent geometries for the final structure. Data collection and refinement statistics are summarized in Table 4.

For visual comparison against the structures of homologous proteins (PDB codes 1CZY and 2JWY) the secondary structure matching algorithm (28) as implemented in Coot (25) was used. For 1CZY, the three β domains were individually superimposed on the three β domains of SadB. In 2JWY, the first model of the

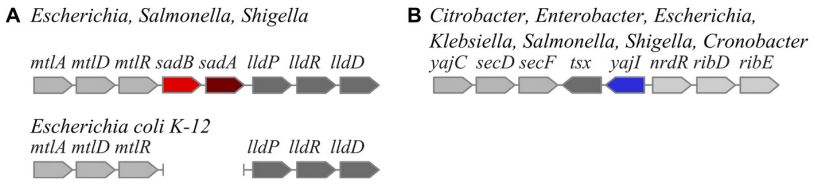

FIGURE 1. **Genomic context of *sadB* and *yajI* is conserved in Enterobacteria.** *A, sadBA* operon is located between the *mtl* operon and the *lld* operon. *E. coli* K12 has a 5-kb deletion encompassing the *sadBA* operon. *B, sadB* paralog *yajI* is located between the *yajC-secDF* operon and the putative *nrdR-ribDE-nusB* operon. Note that although *sadB* and *sadA* are linked in an operon, *yaiJ* and *tsx* are not.

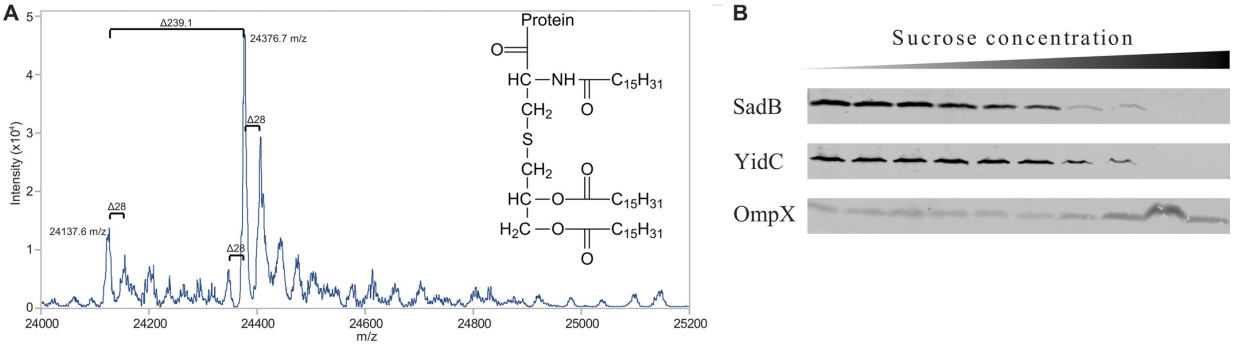

FIGURE 2. **SadB is a periplasmic inner membrane lipoprotein.** *A,* mass spectrometry analysis of full-length SadB with native lipid anchor. The main mass peak at 24376.7 Da corresponds to the expected molecular weight of the protein with an tri-palmityl anchor (see *inset*). The lighter mass at 24,137.6 Da is the result of the loss of one of the palmityl chains by hydrolysis. Mass differences of 28 Da (-C2H4-) result from the incorporation of stearic acid (C18:0) or myristic acid (C14:0) instead of palmitic acid (C16:0) in the lipid anchor. *B,* subcellular localization of SadB by density gradient centrifugation. In a sucrose gradient SadB is found in the lower density band corresponding to the inner membrane. As control, OmpX (an outer membrane β-barrel protein) is only found in the higher density band corresponding to the outer membrane.

NMR ensemble was superimposed on each of the three β domains of SadB.

RESULTS AND DISCUSSION

Bioinformatics—Detailed analysis of the neighboring genes of *Salmonella* trimeric autotransporter adhesin *sadA* revealed an open reading frame encoding a small predicted periplasmic lipoprotein upstream of *sadA*, which is conserved in *Salmonella, Shigella,* and *Escherichia*. Markedly, in enterobacterial strains in which the adhesin was lost, such as the laboratory strain *E. coli* K12, the deletion also encompasses the lipoprotein gene, such that the *mtl* operon is directly followed by the *lld* operon, with both operons remaining intact (Fig. 1*A*). This, together with the short intergenic distance of 44 bp in *Salmonella,* leads us to hypothesize that the gene coding for the lipoprotein, which we subsequently call *sadB* and *sadA,* forms an operon.

Sensitive sequence homology searches using HHPred (29) against the PDB revealed only one known structure of a similar protein, a paralog that exists in Enterobacteria (YajI in *E. coli*, 14% sequence identity, 41% sequence similarity). The gene *yajI* is also found in a conserved genomic location, between the *yajC-secDF* operon and the putative *nrdR-ribDE-nusB* operon. Many species also have the *tsx* gene in the same location; this codes for an outer membrane β-barrel nucleotide transporter (which is not an autotransporter, Fig. 1*B*). The two genes do not appear to form an operon based on their intergenic distance of ~300 bp. The available NMR structure of YajI (2JWY) only covers the C-terminal domain of the lipoprotein. No functional data are available for either of the paralogous proteins.

SadB Is an Inner Membrane Lipoprotein—SadB was predicted to be an inner membrane lipoprotein by LipoP and ClubSub-P (30, 31) due to an aspartic acid in the +2 position of the predicted cleavage site. To verify the prediction, we overexpressed *sadB* from a plasmid. Very mild induction was used to avoid overloading cellular secretion and signal peptide processing mechanisms and to ensure native localization of SadB; in our hands, massive overexpression led to partial inclusion body formation, improper processing of the signal peptide, and mislocalization. For the experiments shown in Fig. 2, the very low leaky expression of the vector was used (no extra inducing agent was added to the culture medium). Membrane fractionation (18) showed that SadB is found in the lower density membrane band of the sucrose gradient, corresponding to localization in the inner membrane (Fig. 2*B*). Mass spectrometry of SadB isolated from membrane preparations further confirmed the cleavage of the signal peptide and attachment of a canonical tripalmityl (C16:0) anchor to the N-terminal cysteine residue of the protein, with minor mass peaks originating from stearic acid (C18:0, mass increase by 28 Da) or from myristic acid (C14:0, mass decrease by 28 Da) replacing individual palmityl residues (Fig. 2*A*). We can conclude from these results that SadB is, as predicted, a lipoprotein of the inner membrane, protruding into the periplasmic space of the bacteria.

SadB Enhances the Surface Display of SadA in E. coli—As noted above, SadB has no homologs of known function. The genomic association with *sadA* suggests a functional linkage, which is supported by the localization in the periplasm, where SadA is transiently localized on its way to the cell surface. We

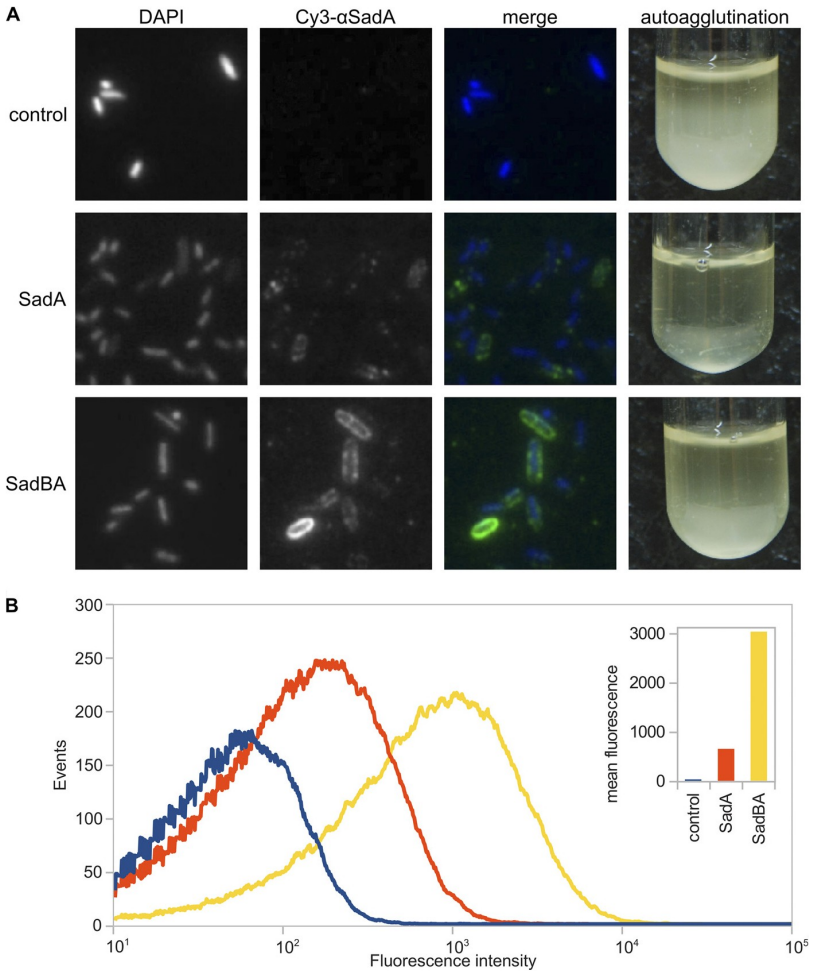

FIGURE 3. **SadB enhances the surface display of SadA and reduces autoagglutination.** *A*, immunofluorescence microscopy and cell aggregation assay of bacteria expressing *sadA*, *sadBA*, or an empty vector control. Surface-localized SadA was stained using a specific antibody. *B*, flow cytometry analysis of bacteria expressing *sadA*, *sadBAm*, or an empty vector control. Surface-localized SadA was stained using a specific antibody. The *inset* shows the mean fluorescence intensity for each sample.

therefore hypothesized that SadB is involved in the biogenesis of SadA by an unknown mechanism. To investigate whether SadB has any effect on the export of SadA to the surface, we created inducible overexpression constructs for the whole operon or for *sadA* alone. As the native expression conditions of the *sadBA* operon are unknown, we did not use the original promoter.

After staining of whole bacteria with αSadA antibody, immunofluorescence microscopy showed a strong signal over the whole surface of bacteria expressing *sadBA*. Bacteria lacking SadB displayed a much weaker fluorescence signal (Fig. 3). Flow cytometry using cells stained with an αSadA antibody confirmed that bacteria expressing *sadBA* have a 4–6-fold higher mean fluorescence over cells expressing only *sadA*, suggesting a significantly higher amount of SadA on the cell surface (Fig. 3*B*). In line with our hypothesis, this observation can be explained by accumulation of SadA in the periplasm in the absence of the SadB or degradation of incompletely translocated SadA by periplasmic proteases such as DegP.

When expressing *sadA* alone, we noticed that the cells formed denser, stickier pellets after centrifugation, which were hard to resuspend. This was confirmed in an autoagglutination assay, where cells were left to settle down after induction of protein expression. After 8–12 h the supernatant of cultures expressing *sadA* was clear, with all bacteria on the bottom of the tube, whereas cultures expressing *sadBA* and controls were still turbid (Fig. 3). Cell viability was unaffected. A possible explanation for this effect is misfolding of SadA on the cell surface, which leads to exposed hydrophobic surfaces in the protein, by which cells then aggregate. This would suggest that SadB supports the biogenesis of well formed SadA trimers.

To analyze if SadB leads to improved folding and thus protease stability of SadA, we performed a cell shaving assay using proteinase K and analyzed the resulting fragments by mass spectrometry (Fig. 4 and details for detected peptides and mapping onto the SadA sequence can be found in supplemental S1 and S2). After 10 min of incubation with the protease, we found a significantly higher fraction of peptides from high molecular

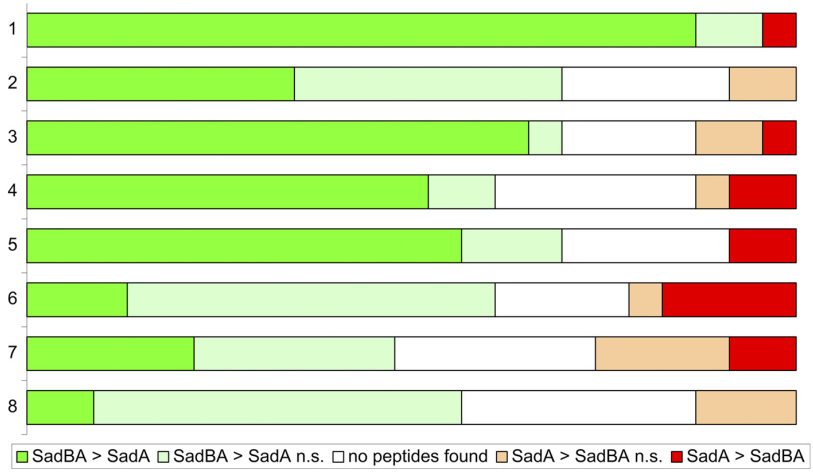

FIGURE 4. **Mass spectrometry of SadA after cell shaving with proteinase K (partial digest).** Most peptides of SadA are more abundant in samples where the whole operon was expressed, compared with samples where only *sadA* was expressed, in line with previous findings that SadB enhances SadA surface display. Furthermore, SadA is more susceptible to proteolysis in the absence of SadB based on the observation that the largest fragments are only found in SadBA samples. Supernatants from cell shaving with proteinase K were separated on SDS-PAGE. Each gel lane was cut into eight equally sized slices (vertical axis, where *1* represents the slice with the highest molecular weight of fragments, and *8* represents the lowest), and the peptides in each slice were analyzed by mass spectrometry. In total, 23 peptides specific to SadA were identified. For each slice, peptide intensities were compared between samples with SadBA or SadA after normalizing for total protein amount. Peptides were grouped depending on which sample the intensity of the peptide was higher. A 4-fold difference was considered significant (*dark red* and *green bars*). Peptides with a lower difference (below the arbitrary significance cutoff level) were considered not significant (*n.s.*) and are shown as *light red* and *green bars*, for reference. Sometimes the individual peptides were not detected in a particular lane (*white bar*).

weight fragments in the SadBA sample compared with the sample where only SadA was expressed after normalization for total protein amount; in other words, SadA was partially digested in both samples but was more easily and quickly broken down to smaller fragments in the absence of SadB. This finding supports the notion that SadB indeed improves the protease resistance of SadA and suggests a direct effect of SadB on the folded state of the SadA fiber. Based on the knowledge that SadA contains highly repetitive sequence motifs (9), especially six consecutive repeats of 70–120 amino acids in the stalk of SadA, which have between 55 and >90% identity, it is tempting to assume that SadB could help to define the register of the three exporting SadA chains that form the final trimer by synchronizing the export.

SadB Is Required for Proper Surface Expression of SadA in S. enterica—To show that the observed SadB effect is relevant also in a more native setting, we investigated the role of SadB in *S. enterica*. Various growth conditions were used, but none of these conditions resulted in expression of SadA detectable by Western blotting or immunofluorescence of bacterial cells. This observation is in line with the previous studies indicating that most *Salmonella* adhesins are not expressed under culture conditions (32). To obtain an experimental system that allows analysis of SadB function in *Salmonella*, we generated strains with *sadBA* or only *sadA* under control of the inducible promoter P_{BAD}. Upon induction with arabinose, synthesis of SadA was observed for both P_{BAD}::*sadBA* and P_{BAD}::*sadA* strains. However, the analysis of surface expression showed that signals for SadA were reduced if only *sadA* was expressed in *Salmonella* (Fig. 5). Although the surface localization of SadA was diffuse for the P_{BAD}::*sadA* strain, the P_{BAD}::*sadBA* strain showed that a clustered distribution of SadA with several foci per cell were found with high signal levels. Neither *sadA* nor *sadBA* expression in *Salmonella* resulted in macroscopic autoaggregation. However, microscopic inspection indicated that *sadBA* expression, but not expression of *sadA* alone, resulted in formation of small clusters of SadA-positive cells (Fig. 5). This again suggests a direct involvement of SadB in the biogenesis of functionally active SadA.

We anticipate that the localization of SadA in clusters might be required for the proposed function as an adhesin. However, the exact role of SadA in *Salmonella* adhesion and pathogenesis is still unknown, although it has been reported to promote weak adhesion to eukaryotic cells and biofilm formation (33).

Direct Interaction of SadB with SadA Cannot Be Shown—To assay whether SadB could recognize and specifically bind unfolded peptides with similarity to the SadA sequence, we used a phage display assay with SadB as bait with a library of random 7-mer or 12-mer peptides. The phage display experiment did not converge on a specific sequence. Rather, the recovered sequence motifs in general showed an alternating pattern of hydrophilic and hydrophobic residues (for examples see Fig. 6), somewhat similar but not strikingly identical to sequence motifs from the head, coiled-coil stalk, and membrane anchor of SadA and other trimeric autotransporter proteins. Also, chemical cross-linking experiments coupled to antibody pulldown assays failed to show an interaction of SadB with SadA *in vivo* (data not shown), suggesting that the interaction is weak and transient.

Structure of SadB—To obtain the structure of SadB, a construct lacking the N-terminal signal peptide and replacing the lipid-modifiable cysteine residue Cys-22 by a serine residue was created. The resulting protein was solubly expressed in the cytosol without the lipid anchor. Crystallization trials yielded crystals diffracting to 2.45 Å. After initial attempts to solve the structure via molecular replacement with the NMR structure of

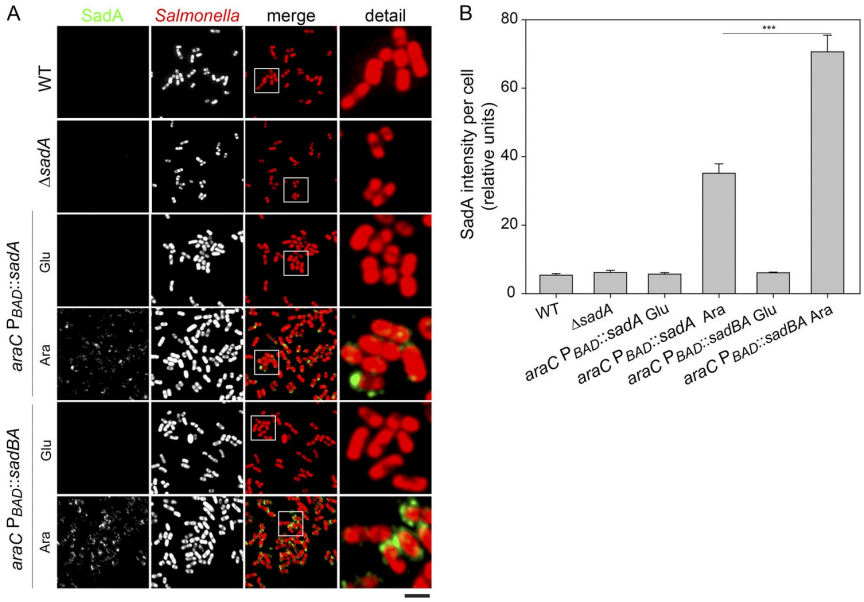

FIGURE 5. **Role of SadB in surface expression of SadA by *S. enterica*.** *A*, inducible expression of *sadBA* or *sadA* in *Salmonella*. WT, Δ*sadA*, and recombinant strains bearing the inducible arabinose promoter P_{BAD}::*sadBA* or P_{BAD}::*sadA* were grown for 4 h in LB with 0.4% glucose or arabinose as indicated, immobilized on coverslips with 0.02% poly-L-lysine, immunostained, and observed by confocal laser scanning microscopy. SadA was detected by immunostaining (*green*). Bacterial cells were labeled mTagRFP. *Scale bar*, 5 μm (overview) and 1 μm (detail). *B*, surface expression of SadA is reduced in absence of SadB. The immunofluorescence signals of SadA from each strain shown in *A* were quantified using ZEN 2012. At least 100 bacteria were scored. Statistical analysis was performed with the Student's *t* test (***, $p < 0.001$).

FIGURE 6. **Representative peptides selected by phage display against SadB.**

YajI (2JWY) were not fruitful, we were able to solve it via single isomorphous replacement using a platinum derivative. Like SadA, SadB is a homotrimer (Fig. 7, *A* and *B*); it is held together by an extended N-terminal coiled coil of nine heptads, which leads into three separate globular C-terminal domains of β-sandwich topology, each composed of two antiparallel β-sheets. In this variant of the Ig fold, the first sheet has the strand order β-1, β-8, β-5, and β-6, whereas the loop connecting β-5 and β-6 is especially long and includes a short α-helix protruding toward the N terminus. The second sheet has strand order β-2, β-3, β-4, and β-7. In tracing the electron density of this domain, the YajI structure was very helpful as it has the identical topology (Fig. 7, *C* and *D*).

Trimerization of the protein was not observed during protein purification, including size exclusion chromatography, suggesting a low trimerization propensity of the coiled coil. Notably, the paralogous protein YajI was described as a monomer in the PDB database (2JWY), although sequence analysis suggests the presence of an N-terminal coiled coil, probably because it does not readily trimerize in solution either. We ascribe this to the high number of polar residues in core positions of the coiled-coil domain. Of the 18 core residues of each SadB protomer, six are glutamine, four in position *a* of the heptad repeat and two in position *d*. Of these, four are arranged within two closely spaced segments with 10 consecutive polar residues each, which provide very little local hydrophobicity for coiled-coil assembly. Indeed, in the N-terminal half of the coiled coil, 22 of 28 residues are polar and 14 are charged. Polar core residues are known to lower the stability and thus the oligomerization propensity of coiled coils (34) and to play, for example, a prominent role in the coiled-coil segments of SadA (35). Trimerization of SadB is therefore expected to be dependent on elevated local concentrations of the protein as would happen *in vivo* on the outside of the inner membrane, to which the protein is tethered by its lipid anchor.

SadB Is Topologically Similar to the Eukaryotic MATH Domain—A DALI search using the C-terminal domain of SadB (residues 90–213) revealed that, beside the already known similarity and clear homology to YajI (with a Z-score of 10.6 and an r.m.s.d. of 2.9 Å over 116 aligned residues), SadB shows high structural similarity to a number of bacterial proteins that are also variants of the Ig fold. All of these proteins display a different β-strand topology as shown by the shorter alignable region (typically ~80–90 residues), and in most cases they do not have an N-terminal coiled coil and are typically of unknown function. Interestingly, however, the structure is also highly similar to the C-terminal domain of human TNF receptor associated factor 2 (TRAF2) (PDB codes 1CZY and 1QSC and similar to Fig. 7, *E–H*), which is called the Meprin and TRAF homology (MATH) domain. This similarity between SadB and MATH domains is undetectable on the sequence level, yet the topology of β-strands in the C domain is identical between SadB, YaiI, and all known MATH domain structures. The DALI Z-scores range from 5.8 to 6.2, and the r.m.s.d. from 3.1 to 3.2 Å, over

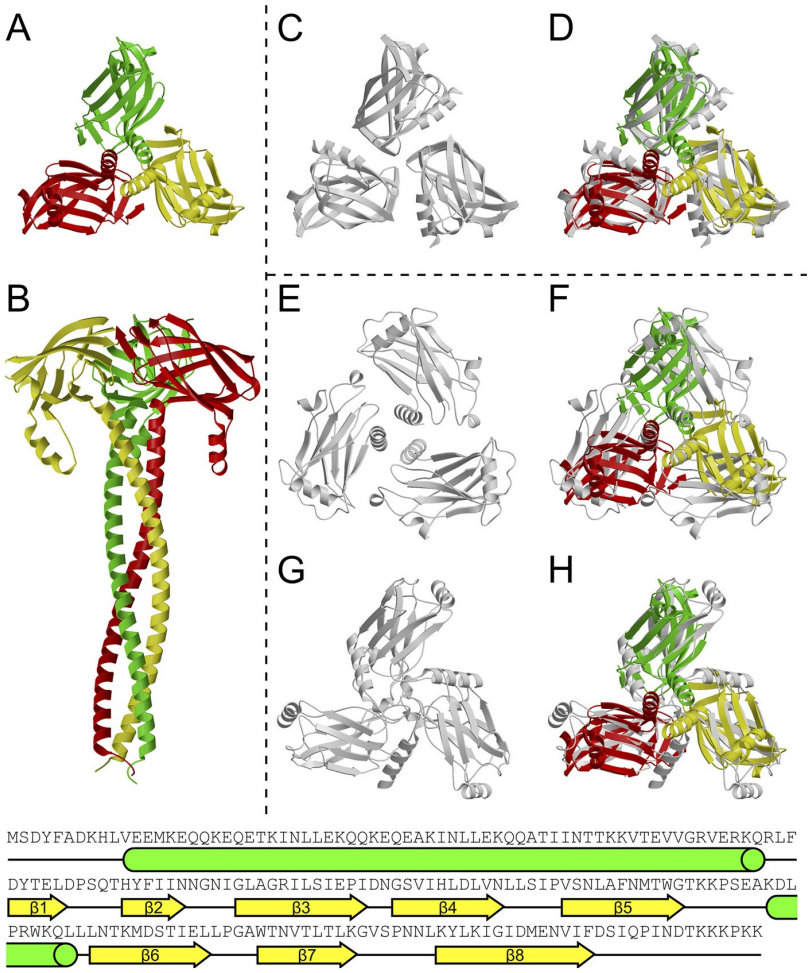

FIGURE 7. **Structure of SadB compared with the structures of TRAF2 and YajI.** *A* and *B*, structure of SadB viewed alongside and perpendicular to the axis of the coiled coil, colored by monomer. *C* and *D*, superposition of SadB (*colored*) with the YajI structure from PDB entry 2JWY (*gray*). The YajI C domain is superimposed on the three C domains of SadB (r.m.s.d. 1.9, 2.0, and 2.0 Å for 101, 102, and 102 aligned CA positions, respectively). *E–H*, superposition of SadB (*colored*) with the TRAF2 structure from PDB entry 1CZY (*gray*). *C* and *D*, superposition is based on the coiled coil, and in *E* and *F*, individual monomers of the Traf2 C domains are superimposed on the individual C domains of SadB (r.m.s.d. 3.2, 3.2, and 3.0 Å for 99, 105, and 99 aligned CA positions, respectively). It is apparent that the orientation of the C domains with respect to the coiled coil is different in the two proteins.

101–103 residues. MATH domains are known to bind peptides from the cytoplasmic domain of TNF receptors across an interface on the outer β-sheet. The receptors trimerize upon binding to trimeric TNF, which positions the cytoplasmic domains in a triangle fit for binding to the MATH domains of a TRAF2 trimer. This mode of binding prefers trimeric, ligand-bound receptor complexes over single receptor molecules by means of increased avidity of three binding sites over one (36, 37). Even though an evolutionary link between SadB and the MATH domains of Traf proteins cannot be established based on sequence analysis, the striking structural similarity and the fact that both act as a trimer on trimeric membrane-bound proteins suggest a comparable mode of action to bring together or bind three unstructured protein chains of a second trimeric protein (Fig. 8).

Conclusion—SadB is a trimeric lipoprotein located in the inner membrane of *Salmonella* spp., with homologs in other enterobacteria, including pathogenic *E. coli* species. The *sadAB* operon is conserved in Enterobacteriaceae. A second paralogous protein of unknown function (YajI) exists in almost all enterobacteria but is not linked to an operon or autotransporter.

The reduced amount and increased stickiness of surface-localized SadA in the absence of SadB and the decreased resistance of SadA to proteinase K in the absence of SadB all suggest a direct influence of SadB on SadA export and folding. If indeed trimeric SadB can bind three nascent SadA polypeptide chains in the periplasm after or during Sec-dependent secretion, it might directly influence the (trimeric) autotransport process. We assume that such an interaction would be weak and transient, also from the fact that cross-linking, pulldown, and phage display experiments performed in this study were not conclusive. A weak interaction close to the inner membrane, and thus to the Sec machinery where the SadA chain is extruded to the periplasm, would presumably help to synchronize the export of three SadA chains, avoiding out-of-register interactions of the highly repetitive, long, and at that stage unfolded polypeptides,

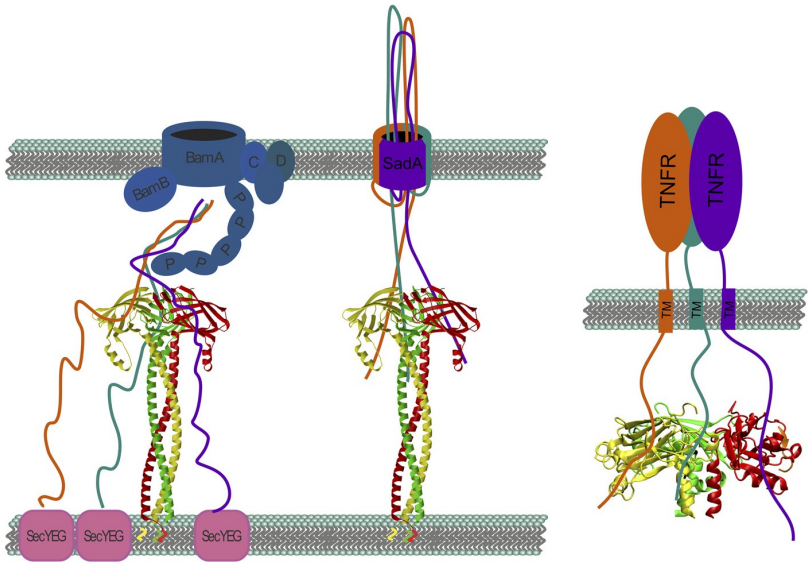

FIGURE 8. *Left*, proposed mode of action for SadB. SadB could be involved in initial stages (*left side*) and/or later stages (*right side*) of SadA membrane insertion and autotransport. *Right*, TRAF2 binding to TNF receptor (adapted from McWhirter *et al.* (36)). The trimeric MATH domain interacts with the three unfolded peptide chains of the TNF receptor. We propose an analogous mechanism where SadB supports the trimerization of unfolded SadA.

as displayed in Fig. 8. A longer retention of unfolded passenger domains at the bacterial inner membrane to achieve productive autotransport has also been suggested for the unusually long signal peptides frequently found in autotransporters (38, 39). SadB could take an analogous role in the case of SadA. It seems that SadB is a specific invention of enterobacteria to help in the autotransport of a specific class of TAAs, and whether other analogous "helper" systems exist for other autotransporter systems will be an interesting subject for future research.

Acknowledgments—We thank Reinhard Albrecht and Kerstin Baer for setting up the crystallization experiments; the staff of beamline PXII/Swiss Light Source for excellent technical support; Silvia Deiss for technical assistance; Roman Gerlach and Valerie Marquard for the construction of Salmonella strains; Annika Hopkins for phage display; Heinz Schwarz for help with SadB antibody production, and Jan-Willem de Gier for the YidC antibody.

REFERENCES

1. Remmert, M., Biegert, A., Linke, D., Lupas, A. N., and Söding, J. (2010) Evolution of outer membrane β-barrels from an ancestral β hairpin. *Mol. Biol. Evol.* **27,** 1348–1358
2. Arnold, T., Poynor, M., Nussberger, S., Lupas, A. N., and Linke, D. (2007) Gene duplication of the eight-stranded β-barrel OmpX produces a functional pore: a scenario for the evolution of transmembrane β-barrels. *J. Mol. Biol.* **366,** 1174–1184
3. Schulz, G. E. (2003) Transmembrane β-barrel proteins. *Adv. Protein Chem.* **63,** 47–70
4. Leo, J. C., Grin, I., and Linke, D. (2012) Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367,** 1088–1101
5. Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., and Ala'Aldeen, D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.* **68,** 692–744
6. Dautin, N. (2010) Serine protease autotransporters of enterobacteriaceae (SPATEs): biogenesis and function. *Toxins* **2,** 1179–1206
7. Szczesny, P., and Lupas, A. (2008) Domain annotation of trimeric autotransporter adhesins– daTAA. *Bioinformatics* **24,** 1251–1256
8. Linke, D., Riess, T., Autenrieth, I. B., Lupas, A., and Kempf, V. A. (2006) Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* **14,** 264–270
9. Hartmann, M. D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A. N., and Hernandez Alvarez, B. (2012) Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 20907–20912
10. Grin, I., and Linke, D. (2011) GCView: the genomic context viewer for protein homology searches. *Nucleic Acids Res.* **39,** W353–W356
11. Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A. N. (2006) The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.* **34,** W335–W339
12. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402
13. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–W549
14. Borchert, N., Dieterich, C., Krug, K., Schütz, W., Jung, S., Nordheim, A., Sommer, R. J., and Macek, B. (2010) Proteogenomics of *Pristionchus pacificus* reveals distinct proteome structure of nematode models. *Genome Res.* **20,** 837–846
15. Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4,** 2010–2021
16. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.–range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26,** 1367–1372
17. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10,** 1794–1805
18. Thein, M., Sauer, G., Paramasivam, N., Grin, I., and Linke, D. (2010) Efficient subfractionation of Gram-negative bacteria for proteomics studies. *J. Proteome Res.* **9,** 6135–6147
19. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 6640–6645
20. Gerlach, R. G., Hölzer, S. U., Jäckel, D., and Hensel, M. (2007) Rapid

engineering of bacterial reporter gene fusions by using Red recombination. *Appl. Environ. Microbiol.* **73,** 4234–4242
21. Husseiny, M. I., and Hensel, M. (2005) Rapid method for the construction of *Salmonella enterica* serovar Typhimurium vaccine carrier strains. *Infect. Immun.* **73,** 1598–1605
22. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800
23. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122
24. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 1002–1011
25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
26. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
27. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291
28. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2256–2268
29. Söding, J., Biegert, A., and Lupas, A. N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33,** W244–W248
30. Rahman, O., Cummings, S. P., Harrington, D. J., and Sutcliffe, I. C. (2008) Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World J. Microbiol. & Biotechnol.* **24,** 2377–2382
31. Paramasivam, N., and Linke, D. (2011) ClubSub-P: cluster-based subcellular localization prediction for Gram-negative bacteria and archaea. *Front. Microbiol.* **2,** 218
32. Humphries, A. D., Raffatellu, M., Winter, S., Weening, E. H., Kingsley, R. A., Droleskey, R., Zhang, S., Figueiredo, J., Khare, S., Nunes, J., Adams, L. G., Tsolis, R. M., and Bäumler, A. J. (2003) The use of flow cytometry to detect expression of subunits encoded by 11 *Salmonella enterica* serotype Typhimurium fimbrial operons. *Mol. Microbiol.* **48,** 1357–1376
33. Raghunathan, D., Wells, T. J., Morris, F. C., Shaw, R. K., Bobat, S., Peters, S. E., Paterson, G. K., Jensen, K. T., Leyton, D. L., Blair, J. M., Browning, D. F., Pravin, J., Flores-Langarica, A., Hitchcock, J. R., Moraes, C. T., Piazza, R. M., Maskell, D. J., Webber, M. A., May, R. C., MacLennan, C. A., Piddock, L. J., Cunningham, A. F., and Henderson, I. R. (2011) SadA, a trimeric autotransporter from *Salmonella enterica* serovar Typhimurium, can promote biofilm formation and provides limited protection against infection. *Infect. Immun.* **79,** 4342–4352
34. Eckert, D. M., Malashkevich, V. N., and Kim, P. S. (1998) Crystal structure of GCN4-pIQI, a trimeric coiled coil with buried polar residues. *J. Mol. Biol.* **284,** 859–865
35. Hartmann, M. D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D. N., Sauer, G., Dunin-Horkawicz, S., Lupas, A. N., and Alvarez, B. H. (2009) A coiled-coil motif that sequesters ions to the hydrophobic core. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 16950–16955
36. McWhirter, S. M., Pullen, S. S., Holton, J. M., Crute, J. J., Kehry, M. R., and Alber, T. (1999) Crystallographic analysis of CD40 recognition and signaling by human TRAF2. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 8408–8413
37. Elgueta, R., Benson, M. J., de Vries, V. C., Wasiuk, A., Guo, Y., and Noelle, R. J. (2009) Molecular mechanism and function of CD40/CD40L engagement in the immune system. *Immunol. Rev.* **229,** 152–172
38. Szabady, R. L., Peterson, J. H., Skillman, K. M., and Bernstein, H. D. (2005) An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 221–226
39. Hiss, J. A., and Schneider, G. (2009) Architecture, function and prediction of long signal peptides. *Brief. Bioinformatics* **10**, 569–578
40. Hernandez Alvarez, B., Hartmann, M. D., Albrecht, R., Lupas, A. N., Zeth, K., and Linke, D. (2008) A new expression system for protein crystallization using trimeric coiled-coil adaptors. *Protein Eng. Des. Sel.* **21,** 11–18

8 Acknowledgements

It's been a long road, getting from there to here. During these past years I had the privilege and pleasure to meet many great people, who all touched my life in some way. These words are for You.

I wish to thank Andrei Lupas for giving me the opportunity to work in his department and to expand my knowledge of bioinformatics and protein evolution.

My deepest thanks go to Dirk Linke, for patiently supervising my thesis, for giving me the opportunity to attend the FEMS conferences in 2009, 2011 and 2013, and for everything else.

Furthermore, I would like to thank Doron Rapaport for being my official university supervisor and for a very timely reminder.

All the members of the Linke group have my thanks for making this ride an enjoyable one: Thomas Arnold, Dorothea Röhrich-Dönitz, Marcus “Mö” Thein, Nagarajan Paramasivam, Shakeel Shahid, and Jack Leo (in approximate chronological order)

Also, I wish to thank all the colleagues from Department I and the whole MPI for the good atmosphere and the everpresent support. Especially Carolin Ewers, Martin Schückel, Silvia Würtenberger, Iuliia Boichenko, Astrid Ursinus, Kerstin Bär, Reinhard Albreicht, Dara Fourouzan, Vikram Alva, Heinz Schwarz, Jürgen Berger, Thomas and Marlene Holder, Klaus Kopec, Martin Mechelke and Christina Wassermann. And Bijan...

Finally, I thank Samuel Wagner and the members of the Wagner Lab, for the past two years: Susann Zilkenat, Tobias Dietsche, Melanie Riess, Andrea Eipper, Lea Krampen, Julia Victoria Monjaras-Feria, Mehari Tesfazgi, Claudia Edith Torres-Vargas, Simon Krodel, and the incredible Thomas Trunk.

To my family – Juri, Christina and Yuliya – Thank You for all the inspiration, support, understanding, patience and love!

And to Elena. I could not have done this without You.