Elucidating the architecture of the type III secretion system export apparatus

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Susann Zilkenat
aus Frankfurt (Oder)

Tübingen,
2017

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 20.07.2017
Dekan: Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter: Prof. Samuel Wagner, PhD
2. Berichterstatter: Prof. Dr. Andreas Peschel

Contents

List of figures

Abbreviations

AP	affinity purification
BAM	β-barrel assembly machinery
BN	Blue native
CsCl-GCP	CsCl gradient centrifugation-purification
DDM	n-Dodecyl-β-maltoside
EM	electron microscopy
fT3SS	flagellar type III secretion system
IM	inner membrane
IP	immunoprecipitation
MBP	maltose binding protein
MS	mass spectrometry
OM	outer membrane
PCS	peptide-concatenated standard
SCAM	substituted cysteine accessibility method
SCV	*Salmonella*-containing vacuole
SPI	*Salmonella* pathogenicity island
SRP	signal recognition particle
ssNMR	solid-state nuclear magnetic resonance
T3SS	type III secretion system
TCS	transitive consistency score
TMH	transmembrane helix
Tt-V-ATPase	*Thermus thermophilus* V-ATPase
vT3SS	virulence-associated type III secretion system

A. Zusammenfassung

Transmembrankomplexe sind ein wichtiger Bestandteil biologischer Membranen und haben vielfältige Funktionen, von Transport über Kommunikation bis zur Adhäsion. Jedoch sind strukturelle Analysen solcher Komplexe technisch schwierig. Ein prominentes Beispiel ist das virulenz-assozierte Typ III Sekretionssystem (vT3SS), dass einen häufigen Virulenzfaktor in Gram-negativen Pathogenen darstellt.

Das vT3SS transportiert Effektoren von der Bakterienzelle in die Wirtszelle, um die Infektion zu begünstigen. Es besteht aus einer Basis, aus den Proteinen InvG und PrgHK, die das System in der baktiriellen ZellhÃ$\frac{1}{4}$lle verankert, einer Nadel, die das System mit der Wirtszelle verbindet, und dem Exportapparat, der in der Innenmembran im Zentrum der Basis sitzt. Dieser Exportapparat besteht aus fünf Transmembranproteinen, SpaPQRS und InvA, und wird für die Auswahl der Substrate, den Wechsel zwischen unterschiedlichen Substratklassen und dem Transport durch die Innenmembran benötigt.

Während strukturelle Informationen über die Basis und Nadel bereits teilweise vorlagen, war der Exportapparat zum größten Teil unbekannt. In dieser Arbeit habe ich die Stöchiometrie der Basis und des Exportapparates ermittelt, sowie die Topologie der Membranproteine SpaPQRS und InvA bioinformatisch ,und SpaP, SpaS und InvA zusätzlich experimentell bestimmt.

Die Stöchiometrie wurde mit Hilfe ratiometrischer, massenspektrometrischer Analysen auf 15:24:24 für InvG:PrgH:PrgK und 5:1:1:1 für SpaP:SpaQ:SpaR:SpaS:InvA bestimmt. Zusätzlich wurde das '*inner rod*' Protein PrgJ, das Exportapparat und Nadel verbindet, mit einer Stöchiometrie von 3-6 ermittelt.

Die experimentelle Bestimmung der Topologie der Proteine SpaP, SpaS und InvA wurde durch die Markierung von Thiolgruppen in Einzel-Cystein-Mutanten vorgenommen. Zusammen mit den Vorhersagen ergaben sich für die Proteine SpaQ, SpaS und InvA eine cytoplasmatische Ausrichtungen beider Termini und eine periplasmatische Ausrichtung beider Termini für SpaP. Der C-Terminus von SpaR befindet sich vorausichtlich ebenfalls im Periplasma.

Die stöchiometrischen und topologischen Informationen unterstützen weitere strukturelle und funktionelle Analysen, nicht nur des vT3SS, sondern

auch des verwandten Systems im Zentrum von Flagellen.

B. Summary

Type III secretion systems (T3SS) are a widespread virulence factor in Gram negative bacteria. They contain an inner membrane spanning sub-complex termed the export apparatus, made up of five proteins. The export apparatus translocates effector proteins designated for the host cytoplasm across the inner membrane, is involved in substrate recognition and in substrate specificity switching.

Knowing the structure of their components is critical for investigating makeup, assembly, and function of macromolecular machines. This has remained a technical challenge in particular for large, hydrophobic membrane-spanning protein complexes like the T3SS. I determined the stoichiometry of the complete SPI-1 T3SS of *Salmonella enterica* serovar Typhimurium and the topology of the export apparatus proteins. For the stoichiometric analysis, I used a mass spectrometry approach based on two complementary protocols for gentle complex purification combined with stable isotope-labelled standards. Previous structural analyses have revealed the stoichiometry of base components, but the stoichiometry of the essential hydrophobic export apparatus components and of the 'inner rod' protein PrgJ remained unknown. Here, I provide evidence that the export apparatus of T3SS contains five SpaP, one SpaQ, one SpaR, and one SpaS. Additionally I can confirm the suggested stoichiometries of InvA and the base components *in situ*. Furthermore, I present evidence that no more than six PrgJ are involved in the formation of the 'inner rod'. I assessed the topology of the five export apparatus transmembrane proteins using computer predictions and a substituted cysteine accessibility method. The position of the transmembrane helices and orientation of the loops of InvA, SpaS and one of the minor export apparatus proteins, SpaP, were mapped experimentally. The prediction could be largely confirmed for SpaS and partly for InvA, while one large periplasmic loop could be confirmed for SpaP. Providing this structural information will facilitate efforts to obtain an atomic view of T3SS.

The topology and stoichiometry identification of these proteins alongside with recent interaction studies are important steps in determining the exact placement of the export apparatus in T3SS and ultimately facilitates elucidation of the function of each component.

C. List of publications

This thesis is based on the following publications, which will be referred to by their roman number.

Primary publications

(I) **Zilkenat S.**, Franz-Wachtel M., Stierhof Y.-D., Galán J.E., Macek B., Wagner S. (2016). *Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach.* Mol Cell Proteomics, 15: 1598-1609.

(II) **Zilkenat S.**, Wagner S., *Topology mapping of the type III secretion export apparatus proteins.* Manuscript ready for submission.

(III) **Zilkenat S.**, Grin I., Wagner S. (2017). *Stoichiometry analysis of macromolecular membrane protein complexes*, Biological Chemistry, 398(2): 155-164

(IV) Dietsche T., Tesfazgi Mebrhatu M., Brunner M.J., Abrusci P., Yan J., Franz-Wachtel M., Schärfe C., **Zilkenat S.**, Grin I., Galán J.E., Kohlbacher O., Lea S., Macek B., Marlovits T.C., Robinson C., Wagner S. (2016). *Structural and functional characterization of the bacterial type III secretion export apparatus*, PLoS Pathogens, 12(12): e1006071

Additional publications

(V) Fischer M., **Zilkenat S.**, Gerlach R., Wagner S., Renard B. (2014). *Pre- and postprocessing workflow for affinity purification mass spectrometry data.* J. Proteome Res. 13:2239-49

(VI) **Zilkenat S.**, Dietsche T., Monjarás Feria J.V., Torres-Vargas C.E., Tesfazgi Mebrhatu M., Wagner S. *Blue native PAGE analysis of bacterial secretion complexes.* Methods Mol. Biol. in press

D. Personal contribution

Manuscript I - Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach

I developed the experimental design with the help of S. Wagner and B. Macek. I also conducted the protein complex purifications and standard preparations, and analysed the mass spectrometry data. The manuscript was written by me, S. Wagner and M. Franz-Wachtel with revisions from co-authors.

Manuscript II - Topology mapping of the type III secretion export apparatus proteins

The experiments were designed by S. Wagner and me. The construction of cysteine mutants, computer predictions and experimental topology studies were performed by me. The manuscript was written by me with revisions from S. Wagner.

Manuscript III - Stoichiometry analysis of macromolecular membrane protein complexes

Literature sighting was done by S. Wagner and me. Figures were prepared by I. Grin and me. The manuscript was written by me and S. Wagner with revisions from I. Grin.

Manuscript IV - Structural and functional characterization of the bacterial type III secretion export apparatus

I performed part of the maleimide crosslinker interaction studies and helped revise the manuscript.

1. Introduction

1.1 Transmembrane protein complexes

Biomembranes are core features of living cells. They function as selective permeable barriers in organisms, facilitate intracellular compartmentalization of different functions and separate individual cells. Lipid bilayers build the basis of cell-membranes. They form spontaneously out of amphipathic lipid molecules in aqueous solutions, where the unpolar hydrocarbon tails face each other while the polar head groups faces the hydrophilic environment (Alberts et al., 2002).

Even through structure and composition of membranes vary between animal and plant cells, as well as Gram negative and positive bacteria, membrane embedded protein complexes are a core feature of any biological lipid bilayer, which allow membranes to fulfil their role. Transmembrane complexes enable functions like nutrient, ion or metabolite transport, communication, or adhesion. Information about the structure of the proteins involved plays an integral part in the understanding of the underlying mechanism of these complexes. Membrane proteins make up 20 to 30 % of all proteins, both in prokaryotes and eukaryotes (Wallin and von Heijne, 1998), and are common molecular targets for drugs (Overington et al., 2006). However, even though immense improvements have been made towards the determination of membrane protein structures in the last decade, they represent only 1.7 % of solved atomic-level structures in the Protein Data Bank (Hendrickson, 2016).

1.1.1 General structure of membrane proteins

Proteins have to exhibit certain characteristics in order to remain stable in membranes. They have two basic structures: α-helix bundles, found in all kinds of cellular membranes, or β-barrels, found, so far, only in outer membranes of Gram negative bacteria. In both cases the membrane spanning segment consists of a stretch of hydrophobic amino acids, flanked by hydrophilic amino acids (Elofsson and von Heijne, 2007). The exact amino acid sequence is relativity unimportant - vital is that the sequence reflects the cross section of the lipid bilayer, enabling a seamless fit between protein

segment and membrane (von Heijne, 1981; Elofsson and von Heijne, 2007).

The α-helix bundles are easier to recognize from sequence due to their approximately 20 amino acid long predominantly hydrophobic transmembrane helices (TMHs). However, variation in length up to 40 residues as well as slopes and kinks are possible (Papaloukas et al., 2008). The β-strands of β-barrels are generally shorter and less hydrophobic (Elofsson and von Heijne, 2007).

1.1.2 Insertion and folding of proteins into the membrane

α-helix bundles are co-translationally inserted via the Sec-translocon in the targeted membrane (White and Heijne, 2004). The hydrophobic regions of the the newly synthesized proteins interact with signal recognition particles (SRP) which in turn interacts with a ribosomal protein. This complex is directed towards the SRP receptor located on the cytoplasmic surface of the inner membrane, where it is transferred to the Sec-translocon (Koch et al., 2017).

The folding of the secondary and tertiary structures of these membrane protein is driven by strong thermodynamic forces and supported by helix-helix interaction. Folding can take place at different stages of the insertion process from the ribosome exit tunnel to within the membrane (Cymer et al., 2015)

β-barrels are post-tranlationally moved across the inner membrane via the Sec-tranlocon with the help of the SecA ATPase, and then folded and integrated into the outer membrane via the β-barrel assembly machinery (BAM) (Elofsson and von Heijne, 2007; Ruiz et al., 2006). Periplasmic chaperones assist the transport of the proteins towards the BAM complex (Hagan et al., 2011)

1.1.3 Obstacles in structural determination of transmembrane complexes

Many membrane proteins are insoluble under experimental conditions designed for globular proteins and need to be isolated from their native environment by detergent extraction (Whitelegge, 2013). Suitable examples for

the extraction of membrane proteins are mild, non-ionic detergents like n-Dodecyl-β-maltoside (DDM). The choice of detergent is important to retain the integrity and structural features of transmembrane complexes and needs to be empirically tested (Zilkenat et al., 2017). Buffer composition and detergent concentration have to be adjusted as well, as detergents can aggregate to form micelles in aqueous solutions, which interfere with e.g. optical or fluorescence spectroscopy based analysis (Arnold and Linke, 2008).

Subsequently, even outside of their native context, transmebrane proteins require additional steps and care concerning continuous solubilization and their behaviour during e.g. gel electrophoresis (Crichton et al., 2013). Thus, studying the structural makeup of large transmembrane complexes in their entirety requires overcoming major obstacles: The first being simply the size of these complexes, which can often exceed 1 MDa. Moreover, the hydrophobicity of their TMHs, the structural complexity involving a variety of different complex components, as well as the dynamic heterogeneity, which many complexes exhibit during their functional cycles, complicate their analysis (Knockenhauer and Schwartz, 2016).

Currently, a variety of techniques is used to study large membrane complexes, each with different successes and limitations:

> "Both electron and X-ray crystallography depend on the successful crystallization of proteins and protein complexes. There have been many improvements in the production of membrane proteins for crystallization [(Clark et al., 2011; Schlegel et al., 2014)] and the formation and stabilization of crystals [(Carpenter et al., 2008; Klara et al., 2016)]. However, as seen in crystal structures of components of the [*Thermus thermophilus* V-ATPase (*Tt*-V-ATPase)], different subunits of type III secretion systems, as well as the nuclear pore complex [(Lee et al., 2010; Worrall et al., 2010; Stuwe et al., 2015)], mostly structures of soluble components or extramembrane domains of transmembrane proteins have been solved for large membrane-spanning complexes. While providing immense help in the elucidation of their overall structure and function, lack of structural information on the hydrophobic core of these complexes impedes a complete understanding of their molecular mechanisms.
>
> Single particle cryo [electron microscopy (EM)], a method in

> which datasets of many 2D electron micrographs of single particles are averaged to generate a high resolution 3D map [(Lau and Rubinstein, 2010)], is playing an increasingly important role in structural analysis of proteins and has been used successfully to gain a better understanding of the makeup of many membrane-spanning macromolecular complexes. A 16 Å resolution 3D map of the 0.65 MDa *Tt*-V-ATPase was generated using single particle cryo EM. A surface view of the 3 MDa needle complex of a *Salmonella* type III secretion system was reconstructed to a resolution of 10 Å, helping to establish the stoichiometry of the three transmembrane ring forming elements [(Schraidt and Marlovits, 2011)]. Furthermore, aided by cryo EM maps, the 110 MDa human nuclear pore complex could be reconstructed to a resolution of 23 Å [(von Appen et al., 2015)] (Figure 1). Significant improvements in direct electron detection have most recently boosted single particle cryo EM analysis and have enabled the solution of structures at subnanometer resolution down to 3.4 Å, even for some transmembrane domains [(Bai et al., 2015; Zorzi et al., 2016)]. Despite these important improvements, flexibility of protein domains and bound lipids and detergents limit the power of cryo EM for many macromolecular complexes, in particular in their transmembrane regions.
>
> Solid state nuclear magnetic resonance (ssNMR) spectroscopy enables the study of membrane protein structures in their native or native-like environment. While in theory molecular weight is not a limiting factor in ssNMR [(Brown and Ladizhansky, 2015)] and large improvements have been made in sample preparation, hardware and experimental design, such as magic angle spinning, no structures of large membrane-spanning macromolecular complexes have been solved using this method, so far [(Shahid et al., 2012; Goldbourt, 2013; Ward et al., 2014)]"- (Zilkenat et al. (2017), p. 157).

The constraints of each technique illustrate the difficulties in obtaining high-resolution structures of macromolecular transmembrane complexes with the methods currently available.

1.2 A macromolecular transmembrane machinery: The type III secretion system

In pathogenic bacteria, membrane complexes often constitute virulence factors: they build an interface to the host, are involved in the movement and placement of the bacteria inside the host system or control the environment of the pathogen to its benefits (Rollauer et al., 2015). Examples for these are type III secretion systems (T3SS). T3SSs are an integral part in two evolutionarily related, complex nanomachines: the bacterial flagellum, which allows motility through circular movement of an extended extracellular filament, and the injectisome, which allows the export of effector proteins across both bacterial membranes and into host cells in order to promote bacterial survival and colonization (Galán et al., 2014; Diepold and Armitage, 2015).

Virulence-associated T3SS (vT3SS) in injectisomes are widespread among many animal pathogenic Gram negative bacteria, like *Yersinia*, *Shigella* and *Salmonella* species (Hueck, 1998), but also found in plant pathogens and symbiotic Gram negative bacteria (Grant et al., 2006; Büttner and He, 2009). In the even more widespread flagellum, the flagellar T3SS (fT3SS) exports components to build the extracellular filament (Abby and Rocha, 2012; Diepold and Armitage, 2015).

The model organism used in this work, *Salmonella enterica* subsp. *enterica* serovar Typhimurium, expresses beside the fT3SS two different vT3SSs, encoded on the *Salmonella* pathogenicity islands (SPI)-1 and SPI-2. Both systems have distinct, independent but coordinated functions during cell invasion, and expression of the corresponding T3SS as well as the order of the exported effector proteins are tightly regulated (Galán, 2001; Chakravortty et al., 2005).

The SPI-1 T3SS is expressed while *Salmonella* is still in the intestinal lumen during initial interaction with the intestinal epithelial cells of the host (Galán, 2001).

The SPI-2 T3SS is required once *Salmonella* has entered the host microfold cells and macrophages for systemic infection (**figure 1a**). Inside a specialized compartment, the *Salmonella*-containing vacuole (SCV), effector proteins are translocated across both bacterial membranes and the vacuole

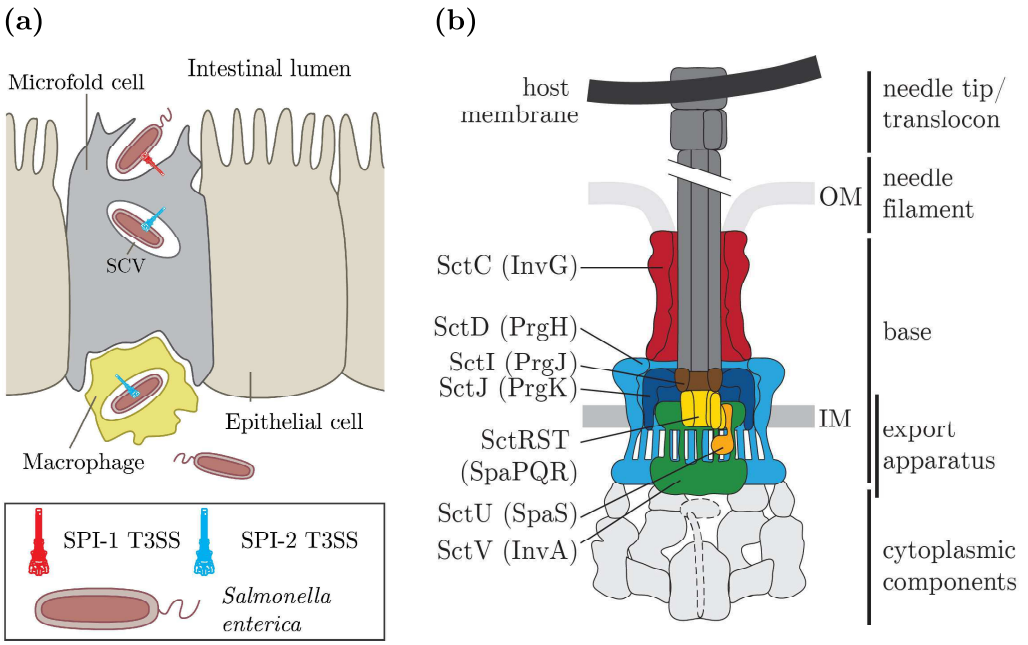


Figure 1: *Salmonella* type III secretion system overview of intracellular infection and structure (a) *Salmonella* pathogenicity island 1 encoded T3SS type III secretion systems (SPI-1 T3SS, red) are expressed during the initial interaction with the intestinal epithelial cells, often microfold cells. SPI-2 T3SSs (blue) are expressed upon systemic infection, among others to maintain the *Salmonella*-containing vacuole (SCV). (b) Model of virulence-associated T3SS, unified nomenclature and *Salmonella* SPI-1 names. OM: outer membrane. IM: inner membrane.

membrane. These proteins help to maintain the SCV, localize it near the Golgi apparatus of the host cells and interfere with the immune signalling (Hensel et al., 1997; Figueira and Holden, 2012), enabling the bacteria to survive, replicate and spread.

1.2.1 Type III secretion system: Structure

As T3SSs haven been studied in numerous organisms I will refer to the involved proteins by their unified nomenclature names (Hueck, 1998) with respect to components present as homologs in several systems and by the protein names with the unified nomenclature in brackets when talking about a specific system. Within this work I focus on the SPI-1 vT3SS, which is encoded in all *Salmonella* serovars (Galán, 2001). The vT3SS is a cell envelope-spanning macromolecular machine as shown in **figure 1b** and is composed of up to 20 different proteins with one to several hundred copies each (Galán et al., 2014). It can be divided into five substructures.

The base components SctCDJ (**figure 1b**, blue and red) anchor the

complex in the bacterial inner and outer membranes (Kubori et al., 1998; Worrall et al., 2016). They build two ring structures, SctC the outer membrane secretin ring and SctDJ the inner membrane ring. Low-resolution structural analysis of isolated needle complexes revealed a stoichiometry of 12−15 copies of SctC and of 12−24 copies of both SctD and SctJ (Hodgkinson et al., 2009; Schraidt et al., 2010; Kowal et al., 2013; Kudryashev et al., 2013).

The needle filament (**figure 1b**, dark grey), which protrudes from the bacterial surface, serves as conduit for substrates (Galán et al., 2014). It is made of helically arranged subunits which reach an average length of 25 nm for the *Salmonella* SPI-1 vT3SS needle (Marlovits et al., 2006; Loquet et al., 2012). Though there have been a number of models to explain needle length control studied in different systems (Tamano et al., 2000; Kubori et al., 2000; Makishima et al., 2001; Journet et al., 2003; Wee and Hughes, 2015), the exact mechanism is still unknown.

The inner rod structure (SctI, **figure 1b**, brown) connects export apparatus and needle (Sukhan et al., 2003). Its subunits are predicted to have a similar structure as the needle subunits (Galán et al., 2014).

Cytoplasmic components (**figure 1b**, light gray) form a sorting platform involved in targeting and preparation of substrates (Akeda and Galan, 2004; Lara-Tejero et al., 2011; Notti et al., 2015; Diepold et al., 2015; Hu et al., 2015, 2017).

Lastly the inner membrane-embedded export apparatus (SctRSTUV, **figure 1b**, green and yellow) located at the center of the base is essential for secretion and a special focus of this work (Collazo and Galán, 1996; Wagner et al., 2010).

The order in which effector proteins are secreted is tightly regulated to ensure the correct assembly of the secretion machinery itself as well as the subsequent proper translocation of effectors over the host membrane.

1.2.2 Type III secretion: Assembly and secretion

The order of assembly of the vT3SS is partially solved (Riordan and Schneewind, 2008; Wagner et al., 2010). The minor export apparatus proteins SctRST are able to assemble first in the absence of both SctU and SctV (Diepold and Wagner, 2014). The switch protein SctU is added to the complex before the major component SctV assembles into a ring structure around the membrane embedded export apparatus (Abrusci et al., 2013).

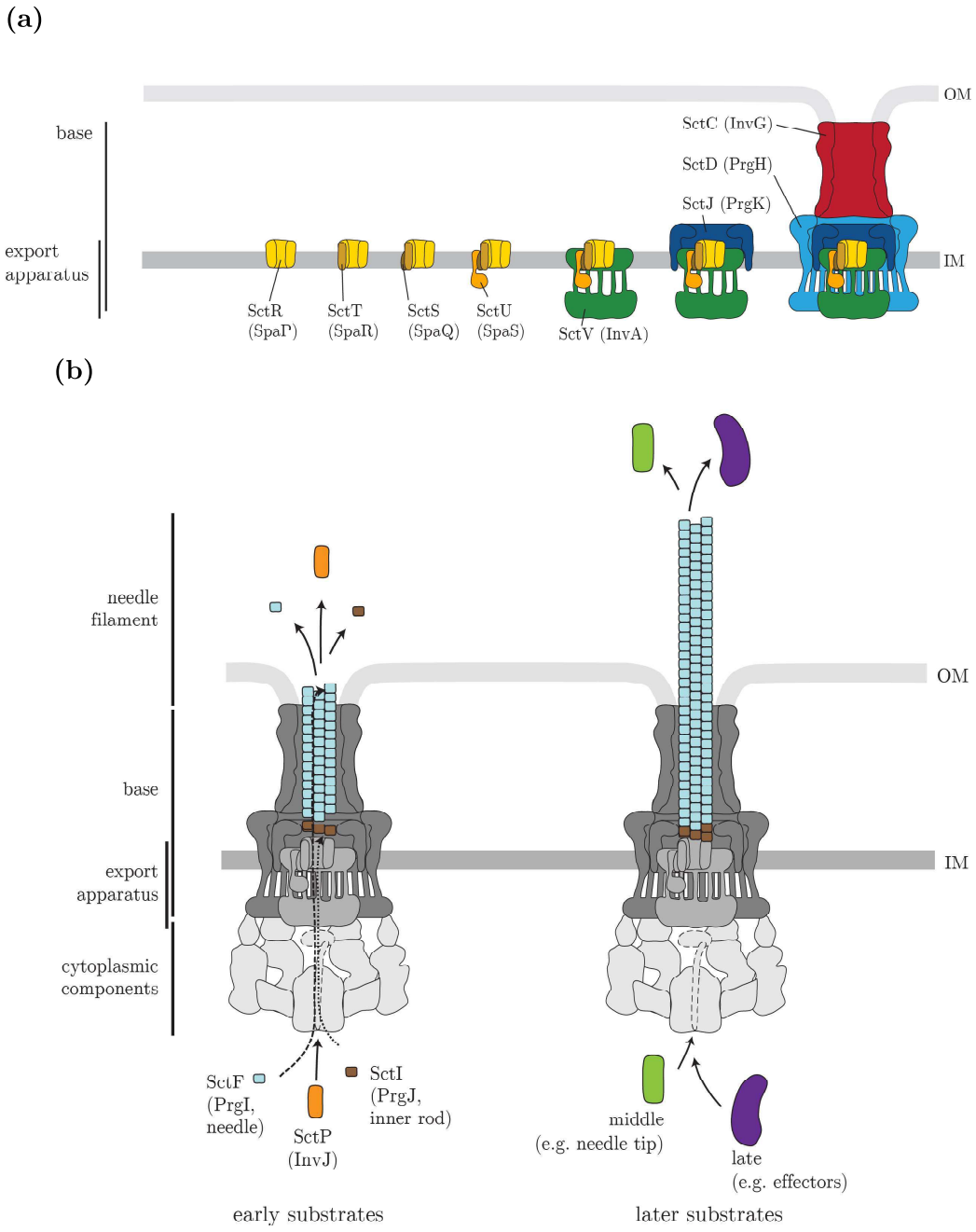


Figure 2: (a) Assembly of export apparatus and base components of the T3SS (modified after Dietsche et al. (2016)). b) Secretion of early substrates and assembly of needle and secretion of middle and late substrates/effector proteins (modified after Monjarás Feria et al. (2015)). OM: outer Membrane. IM: inner membrane. Unified nomenclature and *Salmonella* SPI-1 names.

This is followed by the construction of the base by the components SctD, SctJ and SctC (Diepold et al. (2011); Wagner et al. (2010), **figure 2a**). After the addition of the cytoplasmic components secretion occurs in order to finish the assemby, starting with early substrates necessary for the system itself, like the needle length regulator SctP (Galán and Wolf-Watz, 2006) and the proteins involved in the assembly of the needle structure, needle filament protein SctF and the inner rod protein SctI (Kubori et al. (2000), **figure 2b** left). Additional intermediate substrates, which are secreted after the assembly of the needle, form the needle tip and translocon to overcome the barrier that the host membrane constitutes (Collazo and Galán (1997); Galán (2001); Myeni et al. (2013), **figure 2b** right). Late substrates carry out functions inside the host cytoplasm.

1.2.3 Type III secretion: The export apparatus

The export apparatus is widely accepted to play a role in the as yet unsolved matter of how substrates are recognized prior to translocation (Wang et al., 2013), the switching from one set of substrates to the next during the course of assembly and host cell infection (Monjarás Feria et al., 2015; Lefebre and Galán, 2013; Ferris and Minamino, 2006), as well as in the translocation of effectors across the inner membrane (Cornelis, 2006; Büttner, 2012).

Five transmembrane proteins make up this sub-complex. The largest being the major component SctV comprised of an N-terminal transmembrane domain and a C-terminal globular cytoplasmic domain, for which X-ray crystallography based models have been obtained (Worrall et al., 2010; Abrusci et al., 2013). This C-terminal domain of the *Shigella* homolog MxiA was shown to form a nonameric ring located between the membrane and the cytosolic ATPase (Abrusci et al., 2013). SctV has been suggested to facilitate sorting of substrates into the pore of the export apparatus (Barker et al., 2016) and to couple type III secretion to the proton gradient (Erhardt et al., 2017).

The second largest export apparatus protein is the switch protein SctU (Ferris et al., 2005; Monjarás Feria et al., 2015), which also contains a large cytoplasmic C-terminal domain for which a X-ray based model has been published (Zarivach et al., 2008). SctU has been proposed to be involved in both needle length control and in the switching of specificity from the secretion of early to intermediate substrates. The self-cleavage at a highly

conserved NPTH motif in the cytoplasmic domain is critical for substrate specificity switching but not for needle length control (Zarivach et al., 2008; Wagner et al., 2010; Shen et al., 2012; Monjarás Feria et al., 2015). It has been suggested that switching and late substrate secretion is implemented by the dissociation of the cleaved C-terminus from the remainder of the protein (Frost et al., 2012).

The remaining three minor export apparatus proteins, SctRST, are predicted to consist mainly of TMHs (as described in manuscript II - topology determination). These three proteins were suggested to form a cup structure at the center of the base of vT3SS (Wagner et al., 2010). SctR and SctT are of similar size between 25 and 30 kDa, SctS is a small, hydrophobic protein of only approximately 10 kDa. The functions and structural features of the minor export apparatus proteins are largely unknown.

To gain structural knowledge despite the limitations of obtaining atomic-level models mentioned above, a combination of approaches can be used to study structural aspects separately. The following work investigates the stoichiometry of the assembled needle complex in manuscript I, as well as the topology of the export apparatus in manuscript II. Combined with the interaction studies presented in manuscript IV, these results promote the understanding of the exact placement of the export apparatus as a hydrophobic core at the center of the T3SS and will ultimately advance the comprehension of the interactions between the involved proteins and their specific function in the system.

1.3 Stoichiometry elucidation by peptide-concatenated standard strategy

Mass spectrometric (MS) approaches are versatile and efficient methods for analysing protein complexes and their components. They offer a multitude of applications together with high precision in both qualitative and quantitative measurements.

The stoichiometry of the components comprising a complex can be elucidated by comparison of the complex of interest with isotope-labelled standards, either by direct ratiometric or by absolute quantification. The method I used in manuscript I is based on analysing isotope-labelled standards together with non-labelled purified complete transmembrane complexes to

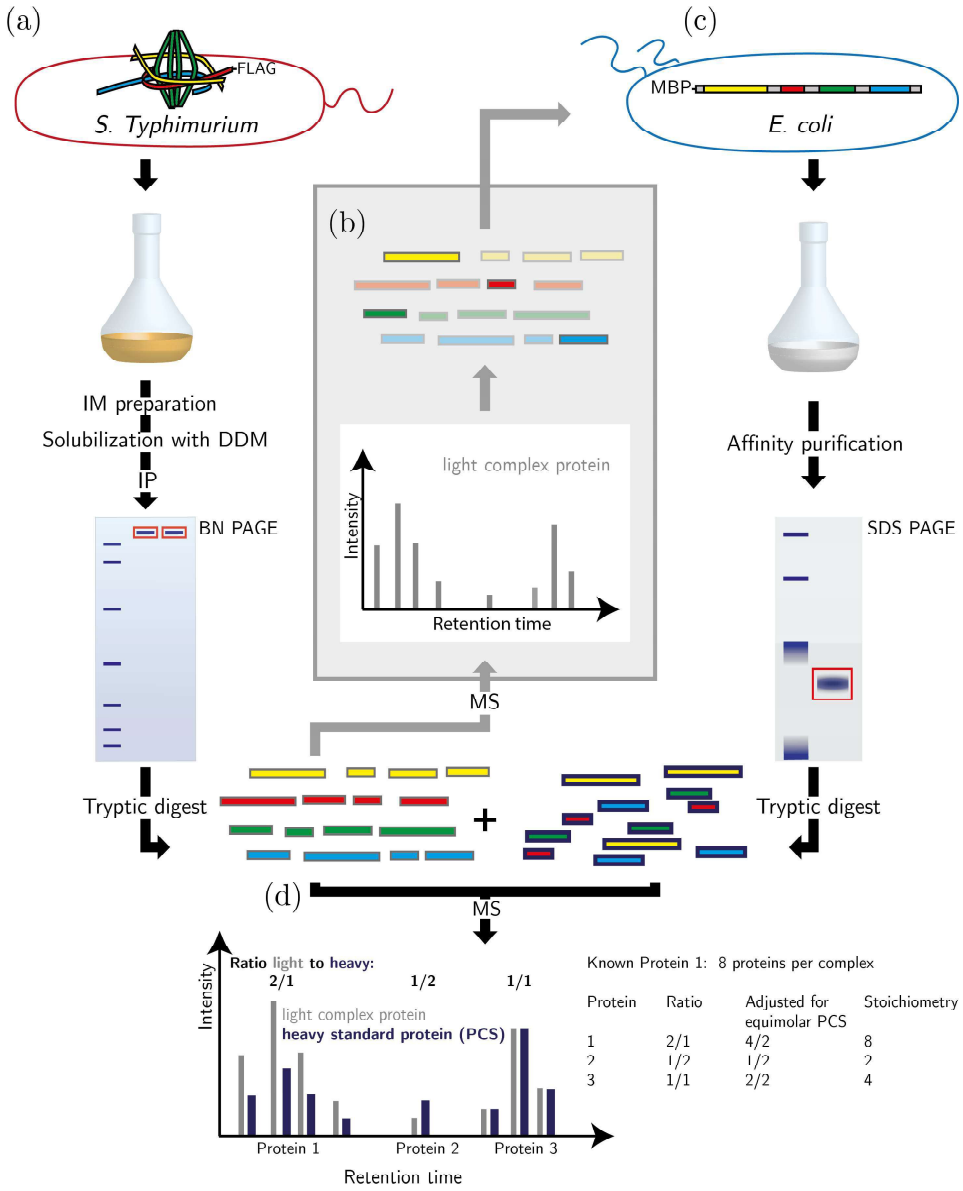


Figure 3: Experimental setup of the peptide-concatenated standard (PCS) strategy (a) *S.* Typhimurium expressing needle complexes with FLAG-tagged bait protein ($SpaS_{N258A}{}^{FLAG}$ or $InvA^{FLAG}$) were grown in complex media. Inner membranes were purified, solubilized by DDM, and needle complexes were immunoprecipitated. Needle complexes were separated by blue native-PAGE, the corresponding bands were excised, and proteins were subsequently digested with trypsin. (b) Peptides were analyzed by MS and suitable peptides were selected for concatenation into the PCS. (c) The PCS was expressed as MBP-fusion in Arg and Lys auxotrophic *E. coli* grown in defined medium containing 'heavy' Arg10 ($^{13}C_6$, $^{15}N_4$-arginine) and Lys8 ($^{13}C_6$, $^{15}N_2$-lysine). After purification via MBP, the PCS was run on a 10%/4% SDS-PAGE, and digested with trypsin in gel. (d) Digested peptides from needle complex a) and PCS c) were mixed and analyzed by MS. The ratio of the evidence of light and heavy peptides was calculated for each protein. Ratios were transformed to absolute stoichiometries by normalization with the evidence-ratios of peptides of proteins of known stoichiometry. Abbreviations: MBP: maltose binding protein, IM: inner membrane, BN: Blue native, IP: immunoprecipitation, MS: mass spectrometry, DDM: n-Dodecyl-β-maltoside; (modified from manuscript I, Zilkenat et al. (2016), method based on Kito et al. (2007)).

determine the ratio between labelled and non-labelled peptides (Zilkenat et al., 2016). Termed peptide-concatenated standard (PCS) strategy (Kito et al., 2007) this method was used to analyse the *Salmonella* SPI-1 vT3SS and is described in **figure 3**.

The PCS strategy is a very efficient way to determine the ratios of protein complexes, as ratios of all complex components can be analysed in one multiplex run. Known stoichiometries of complex components from previous studies, such as those of the base components SctCDJ (Schraidt et al., 2010), make it possible to calculate absolute numbers of proteins per complex from the ratios obtained. Exact numbers are more difficult to calculate the wider the range of stoichiometries in a complex is.

As this method has not been previously used to study transmembrane complexes, many optimizations steps have to be considered.

The first step in this strategy is to purify the complex of interest as complete and homogeneous as possible. I used "a combination of mild detergent extraction, immunoprecipitation or CsCl-gradient-based purification, and blue native-PAGE" (Zilkenat et al. (2016), p. 1606). A previously tested bait-protein is purified in the affinity purification step by immunoprecipitation (IP) with the established interaction partners binding to it, allowing pull down of complete or near complete complexes (Wagner et al., 2010; Fischer et al., 2014). In contrast, CsCl-gradient-based purification does not require a bait-protein.

The next step is to design an isotope-labelled standard containing concatenated tryptic peptides from proteins of interest. Due to the high hydrophobicity and the low arginine and lysine content of many membrane embedded protein domains, identification of peptides with properties favourable for MS analysis can be difficult. The tryptic digest can be optimized to increase number and yield of peptides. In contrast to related methods like QCAT (Beynon et al., 2005) native digestion behaviour of the PCS is maintained by adding flanking amino acids of the native sequence before and behind each peptide. To differentiate between complex and standard, the PCS is grown in defined medium containing 'heavy' Arg10 ($^{13}C_6$, $^{15}N_4$-arginine) and Lys8 ($^{13}C_6$, $^{15}N_2$-lysine), while the complex is purified from cultures grown in normal, 'light', LB medium.

In preparation for mass spectrometric measurement, both purified complex and PCS need to be digested by the same protocol as during the PCS

design. After MS ratios of the intensities of the 'heavy' and 'light' peptides have been obtained, they can be used to calculate the ratios between the proteins of interest. These in turn can be used to calculate their number per complex using proteins of known stoichiometry.

1.4 Topology mapping

1.4.1 Topology prediction of α-helix bundle transmembrane proteins

The insertion and orientation of membrane embedded proteins can be predicted by different bioinformatics models, either by single sequence based methods, or by additionally taking into account sequences of homologs (Tsirigos et al., 2015) or evolutionary information (Jones, 2007). Considered factors can include hydropathy scales, apparent free energy difference predictions (Hessa et al., 2007), and the positive-inside rule (Elazar et al., 2016).

To estimate the insertion of a protein segment into the membrane via hydropathy scales, the free energy difference of each amino acid side chain when moved from a hydrophobic solvent into water is experimentally determined. The sum of these free energy differences of a given amino acid sequence yields the hydropathy index of this segment. These can be plotted for different amounts of residues (Kyte and Doolittle, 1983; Nelson and Cox, 2005). A region of 20 amino acid residues or more with a high hydropathy index is predicted as a possible TMH.

Similarly, the apparent free energy difference predictions use the 'biological' hydrophobicity scales. It takes into account the position of amino acids within the TMH, the impact of the TMH length, and of the amino acids adjacent to the membrane. The experimental determination is based on the integration of designed polypeptide segments into the endoplasmic reticulum membrane via the Sec61 translocon (Hessa et al., 2005, 2007). In a hydrophobicity scale, negative values indicate transmembrane regions, as insertion from the hydrophilic environment into hydrophobic membranes is observed, in contrast to the hydropathy index which studied movement from a hydrophobic into a hydrophilic environment.

The 'positive-inside' rule stems from the observation that in proteins

containing several TMHs, positively charged amino acids arginine and lysine are generally more enriched in regions exposed to the cytoplasm, while scarcer in periplasmic loops (von Heijne, 1989; Elazar et al., 2016)

The prediction tools used in manuscript II are TMHMMfix (Melén et al., 2003), ΔG_{app} (Hessa et al., 2007), TOPCONS (Tsirigos et al., 2015) and PredictProtein (PHDhtm) (Yachdav et al., 2014).

TMHMM predicts membrane topology and is, like most of the prediction tools, based on hidden Markov models (Melén et al., 2003). TMHMMfix allows the user to define already experimentally verified positions as 'cytoplasmic', 'non-cytoplasmic' and 'transmembrane helix' and run the predictions with these constrains.

ΔG_{app} predictor uses the 'biological' hydrophobicity scale described above (Hessa et al., 2007).

TOPCONS TMH predictions are based on a consensus from five different prediction algorithms: OCTOPUS, Philius, PolyPhobius, SCAMPI (multiple sequence mode) and SPOCTOPUS. "OCTOPUS is based on residue preference scores derived from sequence profiles and [Artificial neural networks]" (Viklund and Elofsson (2008), p.1663). Philius uses Dynamic Bayesian Networks and combines signal peptide and TMH predictions (Reynolds et al., 2008). PolyPhobius includes homology information (Käll et al., 2005) and is, like Philius, based on the older tool Phobius (Käll et al., 2004). SCAMPI combines apparent free energy difference predictions and the 'positive-inside' rule (Bernsel et al., 2008). It is the only TOPCONS prediction not based on hidden Markov models. SPOCTOPUS is an extension of OCTOPUS which adds signal peptide predictions to the algorithm (Viklund et al., 2008).

PredictProtein is a meta-service for different sequence feature prediction (Yachdav et al., 2014). PHDhtm is a neural network system which predicts TMHs based on evolutionary features (Rost et al., 1996).

The reliability of these predictions is good, and improving with the availability of more complex and extensive models (Melén et al., 2003; Elofsson and von Heijne, 2007), making them invaluable tools for experimental designs. However, there is no 'best' tool yet, and for some transmembrane proteins the results can vary widely.

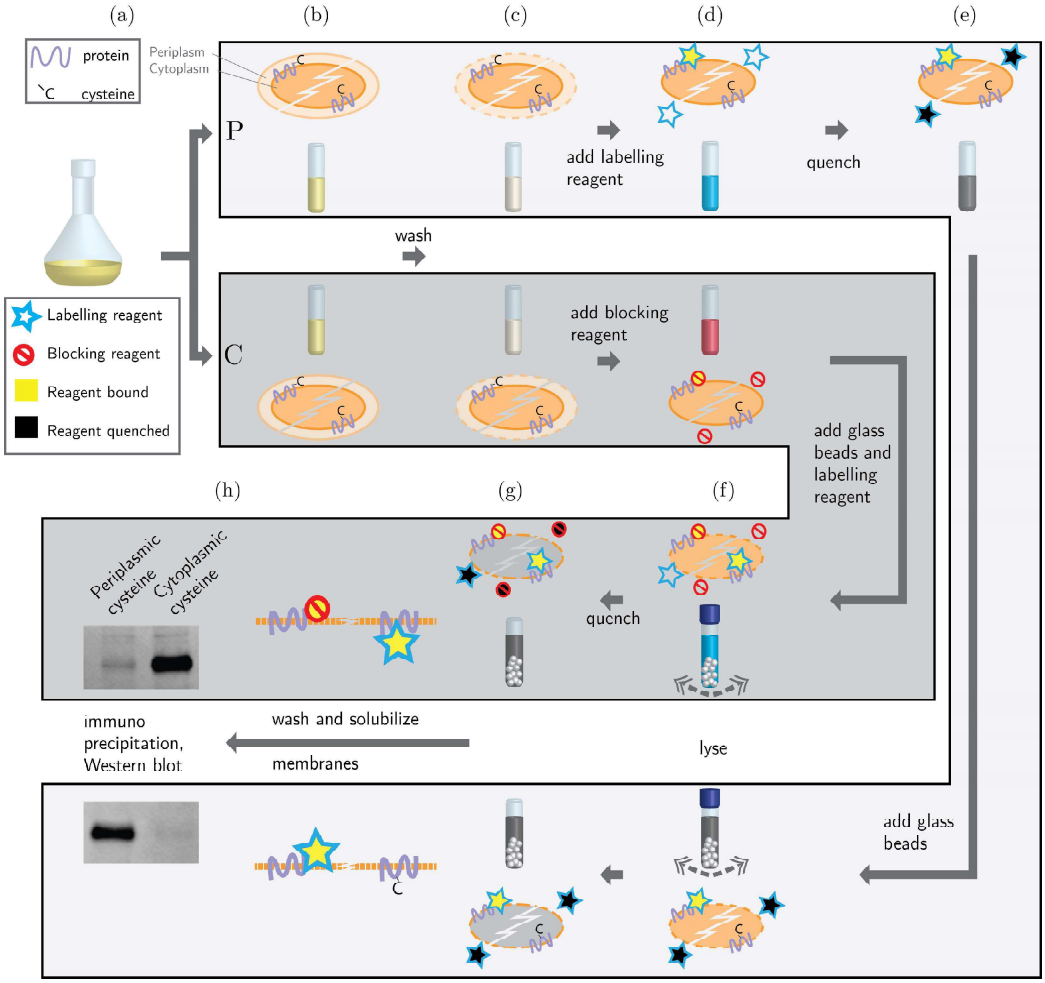


Figure 4: Experimental setup of the substituted cysteine accessibility method (SCAM). (a) A strain with a single cysteine replacement mutation is grown under conditions expressing the protein of interest. (b) The culture is divided into two samples to label cysteines either on the periplasmic (P, light grey) or the cytoplasmic (C, dark grey) side of the inner membrane. (c) The cells are treated with e.g. EDTA to ensure permeability of the outer membrane (omitted in the following) for the (d) labelling ragent (blue, star) added to P and the blocking reagent (red, circle) added to C (e) Potential cysteines are labelled (yellow, filled) in the periplasm and the labelling reaction in P is quenched (black, filled). (f) After putative periplasmic cysteines are blocked the labelling reagent is added to C. Both samples are lysed, e.g. by mechanical disruption, allowing the unquenched labelling reagent to reach any cytoplasmic cysteines. (g) After all reactions are quenched in C as well (h) samples are washed, membranes solubilized and proteins analysed by Western blotting after immunoprecipitation (modified from manuscript II, topology determination).

1.4.2 Experimental verification of membrane topology by substituted cysteine accessibility method

To validate predictions, positions of putative transmembrane helices must be experimentally verified and each hydrophilic loop, as well as the N- and C-termini, have to be localized. Membrane protein topology is often studied by inserting tags into positions of interest. These can be reporter gene fusions like *lacZ*, *phoA* and GFP (Silhavy and Beckwith, 1985; Manoil et al., 1990; Chalfie et al., 1994) or single cysteine residues (van Geest and Lolkema, 2000). Furthermore, modifications of exposed amino acids, e.g. oxidation labelling, coupled with mass spectrometry can be used to verify loop position and orientation, though efficient modification can often only be observed in sulphur containing amino acids (Pan et al., 2012).

A drawback of the fusion approach is that modification or truncation of a protein can change the topology significantly (Bogdanov et al., 2005). Fusion techniques have been improved to reduce these changes. PhoA has been shown to be functional when inserted into the middle of a protein, instead of at the end of a truncated version (Ehrmann et al., 1990) and the development of split-GFP made it possible for a much smaller part, 16 instead of over 200 amino acids, to be used as a reporter (Hyun et al., 2015; Kamiyama et al., 2016). However, even small changes such as single positively charged residues were shown to cause shifts in topology (Seppälä et al., 2010)

In manuscript II, I chose a single cysteine residue approach, the substituted cysteine accessibility method (SCAM) by Bogdanov et al. (2005). Using this methodology the protein function is usually retained. All native cysteines in a protein are replaced with amino acids with similar properties e.g. serine, after which single cysteines are introduced at positions of interest. The single cysteine replacement mutations allow for the secretion function of the complex to be retained, while the orientation in the membrane of each position of interest is tested. The two complementary protocols used to selectively label the thiol groups of periplasmic and cytoplasmic cysteines with biotin-maleimide reagents, respectively, are shown in **figure 4**. Periplasmic positioning was tested by adding outer membrane permeable labelling reagent to whole cells of each cysteine replacement mutant. After quenching of the labelling reagent the cells were lysed. Cytoplasmic positioning of the cysteine

was tested by adding first equally membrane permeable blocking reagent to whole cell, but instead of quenching the reaction the whole proteins were exposed to the labelling reagent during cell lysis. Both samples were then prepared for IP and Western blotting, using streptavidin to detect the biotin.

Due to the highly efficient labelling of biotin and streptavidin, achieving the exact balance between insufficient labelling and to much background is difficult. The conditions have to be adjusted for every system, as well as for every protein. Due to the high content of cysteine containing proteins in the cells, analysis of labelling, especially cytoplasmic labelling, is only possible after purification of the protein. Purification by e.g. immunoprecipitation requires antibodies for the proteins of interest or epitope tagging unobtrusive enough to leave protein function intact.

With these limitations taken into account and both of these complementary protocols, cytoplasmic and periplasmic labelling, in working order, this approach allows for a convincing experimental assessment of the membrane protein topology, and can be well complemented with bioinformatics predictions.

2. Objectives

The aim of this thesis is to deepen the structural understanding of the transmembrane components of type III secretion systems found in many Gram negative bacteria, specifically those membrane proteins of the export apparatus. Using the model organism *S.* Typhimurium the stoichiometry of the whole vT3SS in the context of the assembled complex was studied *in situ*, excluding only the weakly attached cytosolic components and the needle itself. Additionally, the topology of the export apparatus proteins SctRSTUV were investigated. Furthermore, protein-protein interactions of export apparatus proteins were examined in the context of these structural informations.

T3SSs constitute a major virulence factor in Gram-negative pathogens, which, if disrupted, leaves many bacteria avirulent (Hueck, 1998). This makes the system an interesting drug target, as efficient inhibitors of the vT3SS may lead to treatments and prophylactics effective against a wide range of human pathogens, many of which are developing antibiotic resistance (Marshall and Finlay, 2014; Duncan et al., 2012).

As the export apparatus of vT3SSs is essential for secretion function (Collazo and Galán, 1996; Wagner et al., 2010), determining the role of this subcomplex and the function of each of its components could prove to be crucial to understand the mode of action of T3SS inhibitory compounds.

3. Results

3.1 Elucidation of the stoichiometry of the *Salmonella* Type III secretion system

The stoichiometry of a protein complex is an important piece of the puzzle on the way to figure out its exact makeup and functional mechanism. The aim of this study, manuscript I, was to investigate the stoichiometry not only of the export apparatus but also of the whole vT3SS in the context of the assembled complex, excluding only the weakly attached cytosolic components and the needle itself. The peptide-concatenated standard method was used as described in chapter 1.3.

Critical for this method is the preparation of intact and homogeneous complexes, which I achieved by combining two extraction methods with a mild detergent and separation on blue native PAGE. I confirmed the assembled state of the complex before purification by checking for functional secretion in the strains used. I also verified complex assembly after extraction by SDS-PAGE, Western blot analysis, and by electron microscopy. The extraction methods in question were affinity purification (AP, as described in manuscript V, Fischer et al. (2014)) of needle complexes of strains containing previously tested IP bait proteins $InvA^{FLAG}$ (SctV) and $SpaS_{N258A}{}^{FLAG}$ (SctU), and CsCl gradient centrifugation-purification (CsCl-GCP) (Monjarás Feria et al., 2015) of needle complexes of wild-type and $SpaS_{N258A}{}^{FLAG}$-strains. In experiments using the $InvA^{FLAG}$-strain, yield of purified complex was lower and blue native PAGE showed that InvA was easily separated from the complex. In experiments using the $SpaS_{N258A}{}^{FLAG}$ InvA was lost nearly entirely. Additionally, samples prepared by affinity purification consisted mainly of bases lacking the needles, likely due to the mechanical disruption of the bacterial cells by French press lysis (**figure 5a**). Samples prepared via a CsCl gradient retained their needles at least partially (**figure 5b**).

Alongside the preparation of the needle complexes, the construction of the standard was essential for this strategy. Suitable peptides to incorporate into the PCS were identified after optimizing the digestion step and mass spectrometric parameters. The yield of peptides from inner membrane

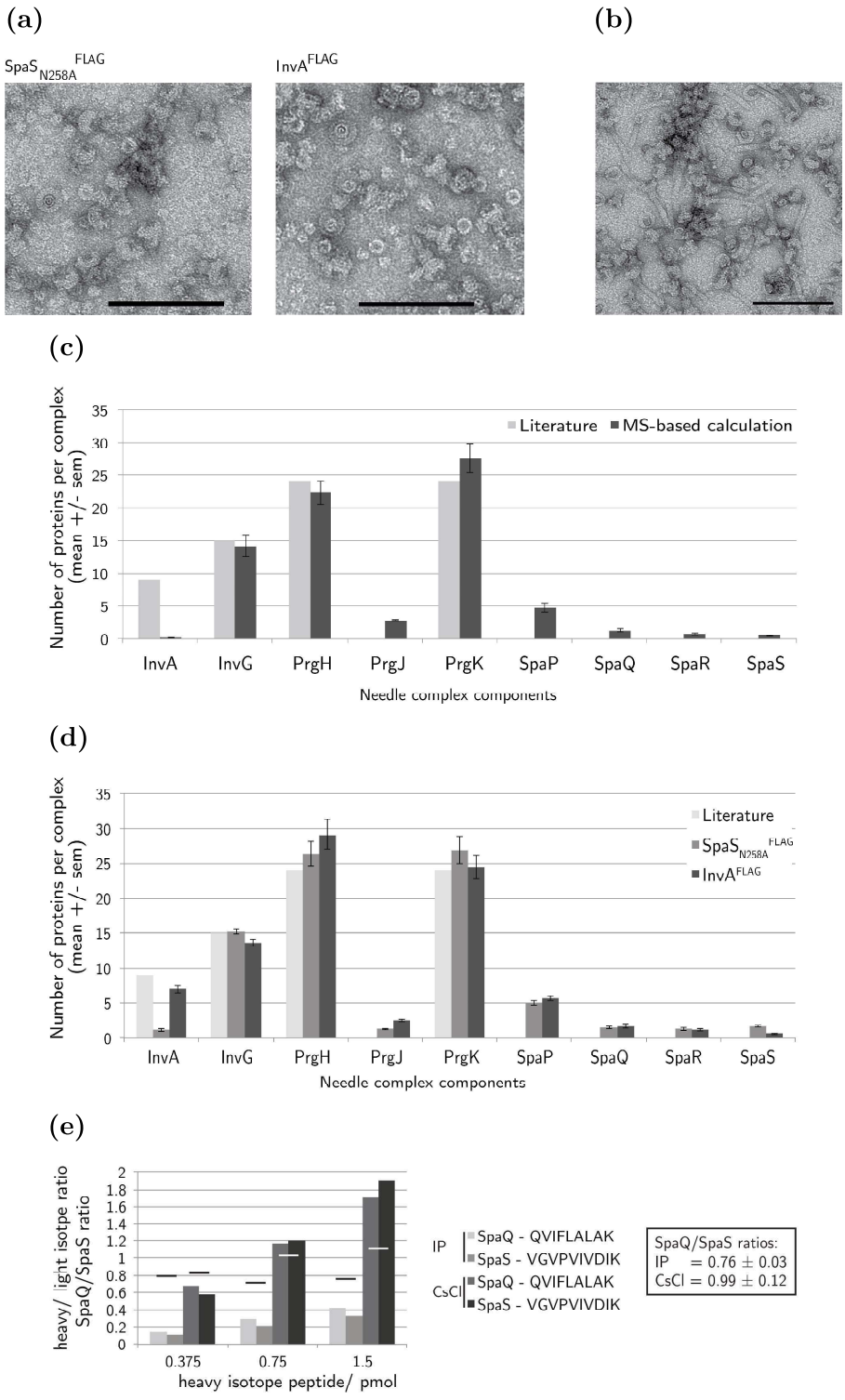


Figure 5: Peptide-concatenated standard (PCS) strategy results (a) Electron micrographs of affinity purification (AP) of $SpaS_{N258A}{}^{FLAG}$ and $InvA^{FLAG}$ samples, scale bar 100 nm. (b) Electron micrographs of CsCl gradient centrifugation-purification (CsCl-GCP) of wild type sample, scale bar 100 nm. (c) Number of proteins in the complex obtained by the combination of AP samples and PCS, error bars: standard error of the mean. (d) Number of proteins in the complex obtained by the combination of CsCl-GCP samples and PCS, error bars: standard error of the mean. (e) The average ratio of SpaQ to SpaS using the synthetic stable isotope-labelled peptides as standards (modified from manuscript I, Zilkenat et al. (2016)).

Figure 6: Sequence of peptide-concatenated standards. Sequence of PCS 1 and PCS 2. White: standard peptides. Gray: flanking residues added to provide natural context for trypsin cleavage. Color-coding for the individual needle complex components as indicated in the legend (manuscript I, Zilkenat et al. (2016)).

proteins suitable for MS was enhanced by consecutive tryptic double digest, though modification prone peptides had to be included for some proteins.

I designed isotope-labelled standards containing concatenated tryptic peptides from the export apparatus proteins SpaPQRS (SctRSTU) and InvA (SctV), the base components PrgHK (SctDJ) and InvG (SctC) as well as the 'inner rod' protein PrgJ (SctI) (**figure 6**). The three proteins PrgH, PrgK and InvG were included into the analysis as controls because their stoichiometry has been previously reported (Schraidt et al., 2010). To ensure similar digestion behaviour between standard and sample, each standard peptide was flanked upstream and downstream by two to four amino acids corresponding to the native sequence. The standards were analysed alone to verify the incorporated levels of the isotope-labelled amino acids, which were high (>98 %) for all peptides.

All experiments were normalized and the total number of proteins per complex calculated using the robustly detected InvG (SctC) peptide SLLVGGYTR. InvG has been previously reported with a stoichiometry of 15 (Schraidt and Marlovits, 2011). After normalization, stoichiometry data obtained for PrgHK (SctDJ) were within 10 % of the literature values of 24. This shows that the results obtained for proteins of so far unknown stoichiometry are reliable.

The AP samples combined with PCS resulted in stoichiometries of

15:26:27 and 14:28:24 for the control proteins InvG:PrgH:PrgK for the $SpaS_{N258A}^{FLAG}$ and $InvA^{FLAG}$ preparations respectively. The stoichiometry of the export apparatus components SpaP:SpaQ:SpaR:SpaS was 5:1:1:1 in both IPs, while the number of InvA proteins per complex could only be determined as seven in experiments using $InvA^{FLAG}$ itself as bait, due to it's loose association with the rest of the complex. Lastly, 2-3 'inner rod' protein PrgJ were detected per needle complex (**figure 5c and 5a**).

The results of the CsCl-GCP/PCS experiments are in accordance with the AP results with stoichiometries of 14:22:27 for InvG:PrgH:PrgK, 5:1:1:1 for SpaP:SpaQ:SpaR:SpaS and 3 PrgJ. InvA was again lost in this method of preparation (**figure 5d and 5b**).

To verify the ratio between SpaS and SpaQ, MS measurements were repeated using synthetic stable isotope-labelled peptides instead of PCS. The average ratio of SpaQ to SpaS was 0.99 for experiments using CsCl-GCP and 0.76 for those using AP (**figure 5e**), pointing to equal amount of both proteins per complex, in consistence with the results of the PCS experiments.

3.2 Topology of type III secretion export apparatus proteins by introduction of cysteine residues

The topology of a membrane embedded protein gives us important information about its possible functions. The orientation in the membrane, the length and positions of the loops between transmembrane helices can give indications about exposed interaction sites with itself and other proteins.

3.2.1 Membrane protein topology prediction

Prior to the experimental topology analysis by SCAM, I performed detailed predictions of the transmembrane topology via bioinformatics. To estimate the positions and orientations of the transmembrane helices, *S.* Typhimurium SPI-1 export apparatus proteins were analysed using the web tools TOPCONS (Tsirigos et al., 2015), TMHMMfix (Melén et al., 2003), ΔG_{app} (Hessa et al., 2007) and PredictProtein PHDhtm (Yachdav

et al., 2014). These tools are based on sequence features, sequence features of homologs, hydrophobicity, comparison with known structures, prediction of the corresponding apparent free energy difference, and evolutionary profiles. TOPCONS uses a consensus from five different prediction algorithms: OCTOPUS, Philius, PolyPhobius, SCAMPI (multiple sequence mode) and SPOCTOPUS. The resulting predictions for each export apparatus protein are illustrated in **figure 7** (as well as supplement data S10 and table ST2 in manuscript II). The predicted number and position of TMHs of the proteins SpaS (SctU) and SpaQ (SctS) was very consistent. SpaS was robustly predicted with four TMHs, and an N_{IN}/C_{IN} orientation. SpaQ was predicted with two TMHs, however, the orientation predictions were not conclusive. The major export apparatus protein InvA (SctV) was predicted with seven to eight TMHs, a reliable N_{IN}/C_{IN} orientation and a large cytoplasmic loop. The start and end positions of the TMHs of SpaS, SpaQ and InvA varied only marginally between predictions by 12 amino acids of less, with the exception of the last two TMHs of InvA, which were predicted as one TMH by PredictProtein. In contrast, the predictions of SpaP (SctR) varied strongly with an N_{OUT}/C_{OUT} or N_{OUT}/C_{IN} orientation, three to six TMHs and a high variation of start and end positions of the helices. A large loop between amino acids 75 and 141 was consistently predicted, however, its orientation was not clear. SpaR (SctT) was predicted with five to six TMHs, mainly with an N_{OUT}/C_{OUT} orientation (only SCAMPI N_{IN}/C_{IN}).

To improve the ambiguous predictions of the minor export apparatus proteins SpaP, SpaQ and SpaR, I compared them with nine homologs from different vT3SS and fT3SS each: vT3SS and fT3SS of *Salmonella* Typhimurium, vT3SS of *Xanthomonas campestris*, vT3SS of *Yersinia enterocolitica*, vT3SS and fT3SS of *Shigella flexneri*, vT3SS of *Pseudomonas syringae*, vT3SS and fT3SS of *Escherichia coli*, and fT3SS of *Helicobacter pylori*. The alignments were performed using T-Coffee multiple sequence alignment (Notredame et al., 2000), and ΔG_{app} (Hessa et al., 2007) was used to predict TMHs. These results are summarized together with a review of charged amino acids (Andersson et al., 1992) in **figure 8**.

The transitive consistency score (TCS) is a sequence alignment evaluation which scores between 0 and 100. A score lower than 50 should be considered poor (Chang et al., 2014). TCS of the SctR, SctS and SctT homologs were between 83 and 94 (manuscript II: supplements S1, S4 and S7).

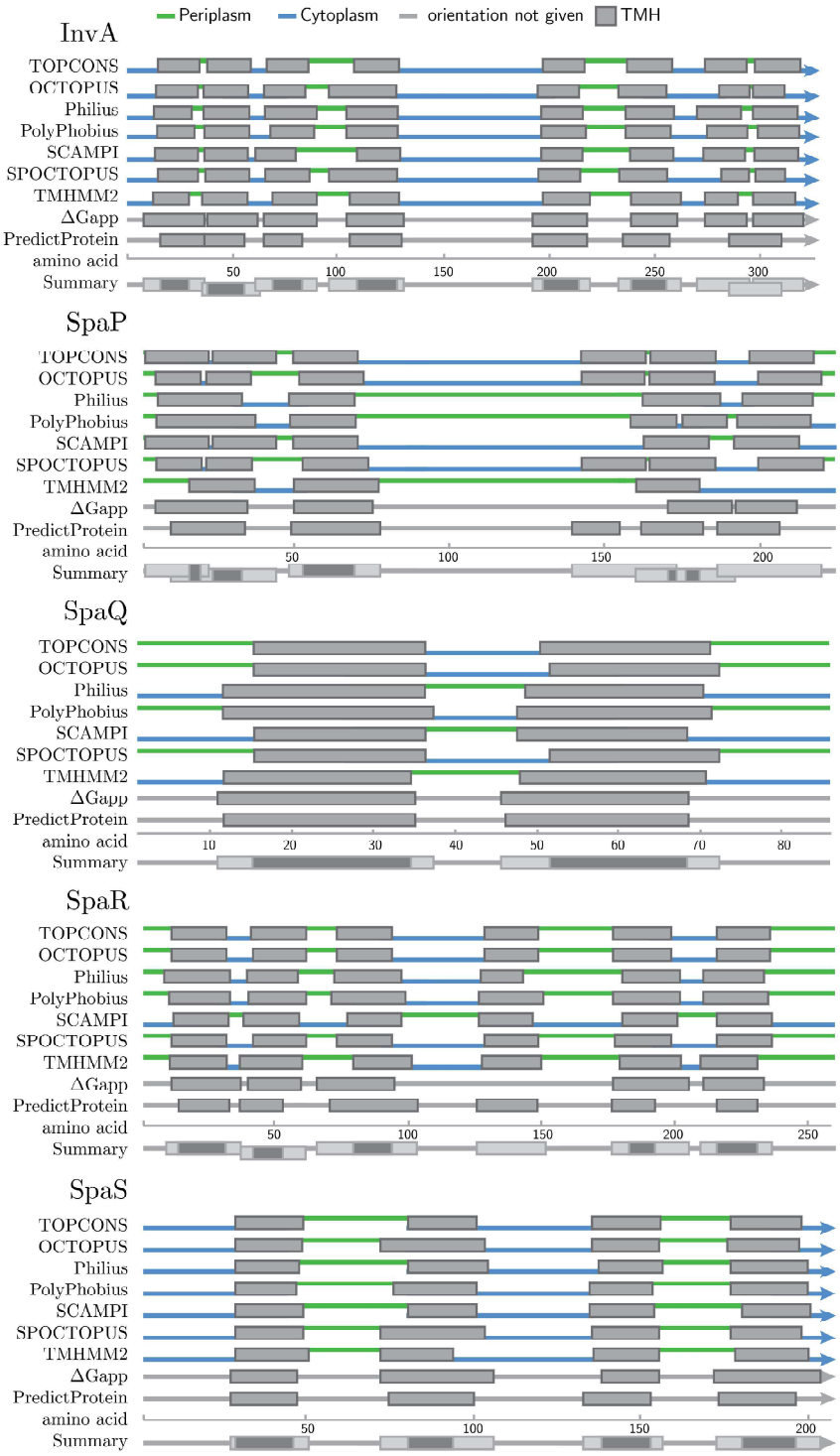


Figure 7: Topology prediction of the export apparatus proteins InvA (SctV) and SpaPQRS (SctRSTU) using TOPCONS (Tsirigos et al., 2015), TMHMMfix (Melén et al., 2003), ΔG_{app} (Hessa et al., 2007) and PredictProtein (PHDhtm) (Yachdav et al., 2014). In the summaries sequences predicted as transmembrane helices (TMH) in all predictions are shown in dark gray, sequences predicted as TMHs at least once are shown in light gray. InvA C-terminal cytosolic part after amino acid 310 and SpaS C-terminal cytosolic part after amino acid 205 are omitted (manuscript II).

The SctR and SctT homologs were predicted with four TMHs (disregarding the possible signal peptide sequences in cases of FliP, manuscript II, supplement S11) and five to six TMHs respectively. They varied in length between 18 and 40 amino acids and with ΔG ranging from 2 to -4 kcal/mol (manuscript II: supplement S2 and S8). The two predicted TMHs of SctS only ranged from 23 to 30 amino acids with a ΔG of 1 to -3 kcal/mol (manuscript II: supplement S5).

As described by the 'positive-inside' rule' (von Heijne, 1989; Wallin and von Heijne, 1998) the positively charged amino acids arginine and lysine have a much larger impact on membrane protein topology than the weakly positively charged histidine or the negatively charged aspartate and glutamate. Histidine only becomes relevant at lower than physiological pH, while aspartate and glutamate affect topology only in sufficiently high numbers (Andersson et al., 1992; Nilsson and von Heijne, 1990).

The SctR homologs contained a conserved positively charged amino acid in the first predicted loop, indicating that this loop is probably cytoplasmic. This, together with few arginine and lysine in the C- and N-termini and only occasional aspartate or glutamate, is an indication for an N_{OUT}/C_{OUT} orientation. The predicted large loop (between amino acids 75 and 141 in SpaP) carries a high number of positively charged amino acids, but also a high number of negative charges and in case of SpaP an EDED cluster directly adjoining the second predicted TMH, which could affect topology contrary to the positive charges. Assuming four TMHs, this loop should be on the same membrane side as the termini (manuscript II: supplement S3).

90 % of the SctS homologs had at least one positively charged amino acid in the N-terminal extramembrane region or within the first three amino acids of the adjacent predicted TMH. The same was the case for the C-terminus, while no positively charged amino acids were in or near the predicted loop (manuscript II: supplement S6). This suggests an N_{IN}/C_{IN} orientation.

90 % of the SctT homologs contained a conserved positively charged amino acid in the last predicted loop, suggesting a cytoplasmic orientation of this loop, indicating a C_{OUT} orientation for the whole protein (manuscript II: supplement S9). However the overall topology of SctT cannot be concluded from these results.

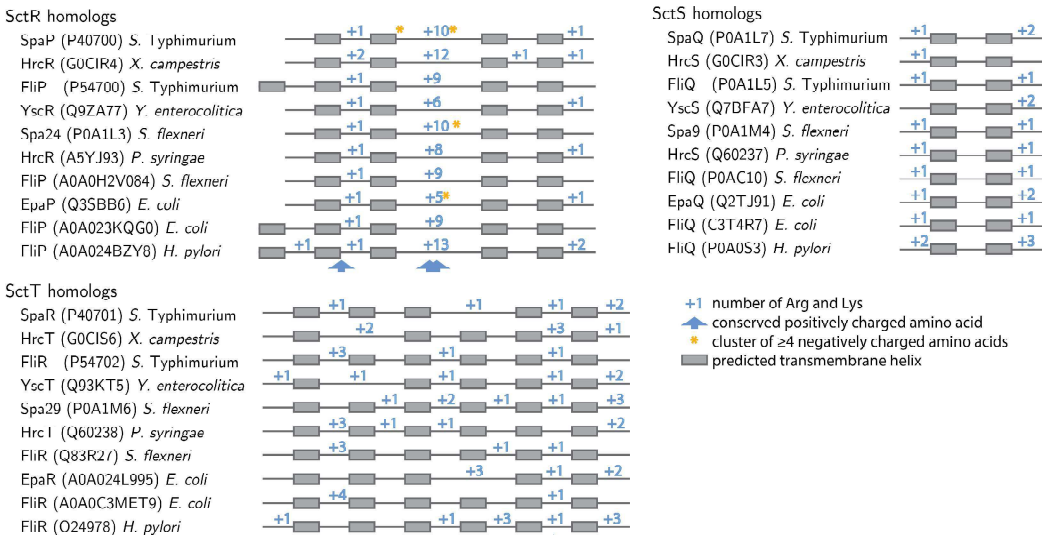


Figure 8: Review of charged residues of the export apparatus proteins SctRST homologs (vT3SS and fT3SS of *Salmonella* Typhimurium, *Shigella flexneri* and *Escherichia coli*, vT3SS of *Xanthomonas campestris*, *Yersinia enterocolitica*, and *Pseudomonas syringae*, and fT3SS of *Helicobacter pylori*; Uniprot entry number in brackets). Simplification of transmembrane helix (TMH) predictions by ΔG_{app}. Number of Arginine and Lysine counted in the extramembrane regions and within the first three amino acids of the adjacent predicted TMH. Clusters of negatively charged amino acids and conserved positively charged amino acids are additionally marked (manuscript II, full sequences and TMHs in supplement S3, S6 and S9 of manuscript II).

3.2.2 Experimental analysis of the transmembrane topology using SCAM

The bioinformatics analysis of the transmembrane topology of the *Salmonella* T3SS export apparatus was followed by an experimental analysis. In an effort to avoid the introduction of large, possible tertiary structure changing elements, I chose an approach which only requires the replacement of single amino acids with cysteine (Bogdanov et al., 2005), as described in chapter 1.4.

As immunoprecipitation and detection is a prerequisite for analysis by SCAM, previously described 3xFLAG epitope-tagged versions of the four export apparatus components InvA (SctV), SpaP (SctR), SpaR (SctT), and SpaS (SctU) were used throughout this study (Wagner et al., 2010). For SpaS, the detection by a C-terminal FLAG tag necessitated the use of an autocleavage-deficient N258A mutant (Wagner et al., 2010). Unfortunately, SpaQ could not be tagged and therefore was not assessed using this method.

Cysteine free mutants of the remaining four export apparatus proteins as well as single cysteine mutants were created. Their secretory function was

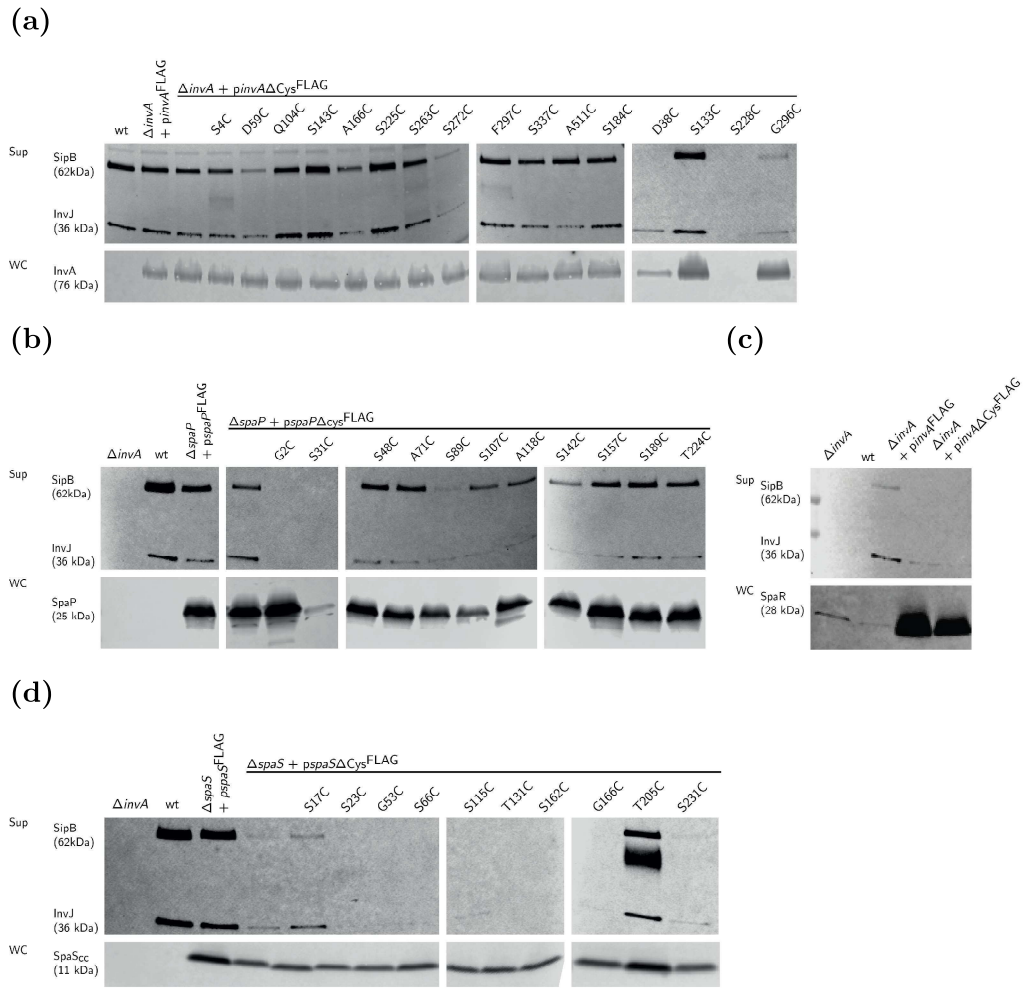


Figure 9: Secretion of effector proteins SipB and InvJ into the the supernatant (sup) and expression of export apparatus proteins InvA (a), SpaP (b), SpaR (c) and SpaS (d) in whole cells (wc). Cysteine mutants are expressed from low copy number plasmid pTACO10. Δ*invA* strain functions as a negative control for secretion (manuscript I).

verified by Western blot analysis of the *Salmonella* SPI-1 vT3SS substrates InvJ and SipB in culture supernatants of plasmid-complemented knock-out mutants.

While secretion was reduced for some mutants, most cysteine substitutions did not impair secretion of both substrate proteins (**figure 9**).

The secretion functionality of SpaR was strongly reduced by the addition of the 3xFLAG epitope-tag. Since the experimental mapping of this protein did not yield clear results, functional assays of single cysteine mutants are not shown.

For the successful implementation of SCAM, both the integrity of bacterial cells during the blocking and labelling steps as well as the accessibility

of the cysteine residues needed to be assured. Results indicating lysis of cells by displaying similar labelling between the cytopasm and periplasm experiment have been disregarded. Cells were treated with a mixture of EDTA and lysozyme to allow maleimides to pass through the outer, but not the inner membrane. To ensure accessibility, amino acids predicted as exposed (using PredictProtein PROFsec and SNAP2 (Yachdav et al., 2014)) were preferentially replaced with cysteine. Additionally, SCAM experiments were conducted in a $\Delta prgHIJK$ ($\Delta sctDFIJ$) strain, which was previously shown (Wagner et al., 2010) to assemble an intact export apparatus, but lack the periplasmic structures of the T3SS that would possibly shield it from labelling reagents.

Labelling and blocking properties of the maleimide reagents was tested using an InvA cysteine mutant A511C with known position in the cytoplasm from the 3D-structure of the C-terminal extramembrane domain of this protein (Abrusci et al., 2013).

A cysteine free negative control was run alongside each experiment. The signal strength of the labelling results was normalized using the FLAG-IP signal and compared relative to the respective lowest (set as 0 %) and highest (set as 100 %) signal of each experiment.

With these measures taken, SCAM allows for precise and reliable membrane topology mapping. Results were accepted as certain if a position could be mapped with complementary cytoplasmic and periplasmic labelling in at least duplicate.

For the major export apparatus protein InvA (SctV) eleven of sixteen amino acid positions could be mapped (**figure 10a**): S4C, A166C, S174C, S263C, S272C, S337C and A511C in the cytoplasm and Q104C, S225C, G296C and F297C in the periplasm. The results of the remaining five positions offered an indication of the localization. Cysteine at position 133 proved to be inaccessible, the position S143C (cytoplasmic) and S228C (periplasmic) could only be labelled once, position D38C was localized in the periplasm in two out of five experiments, but proved to be inaccessible in the rest and position D59C was localized in the cytoplasm in two out of four experiments, but inaccessible and in the periplasm in the remaining two. These results confirm a N_{IN}/C_{IN} orientation with eight TMHs (**figure 11**).

For the minor export apparatus protein SpaP (SctR) half of the six amino acid positions could be labelled unambiguously (**figure 10b**): S31C and

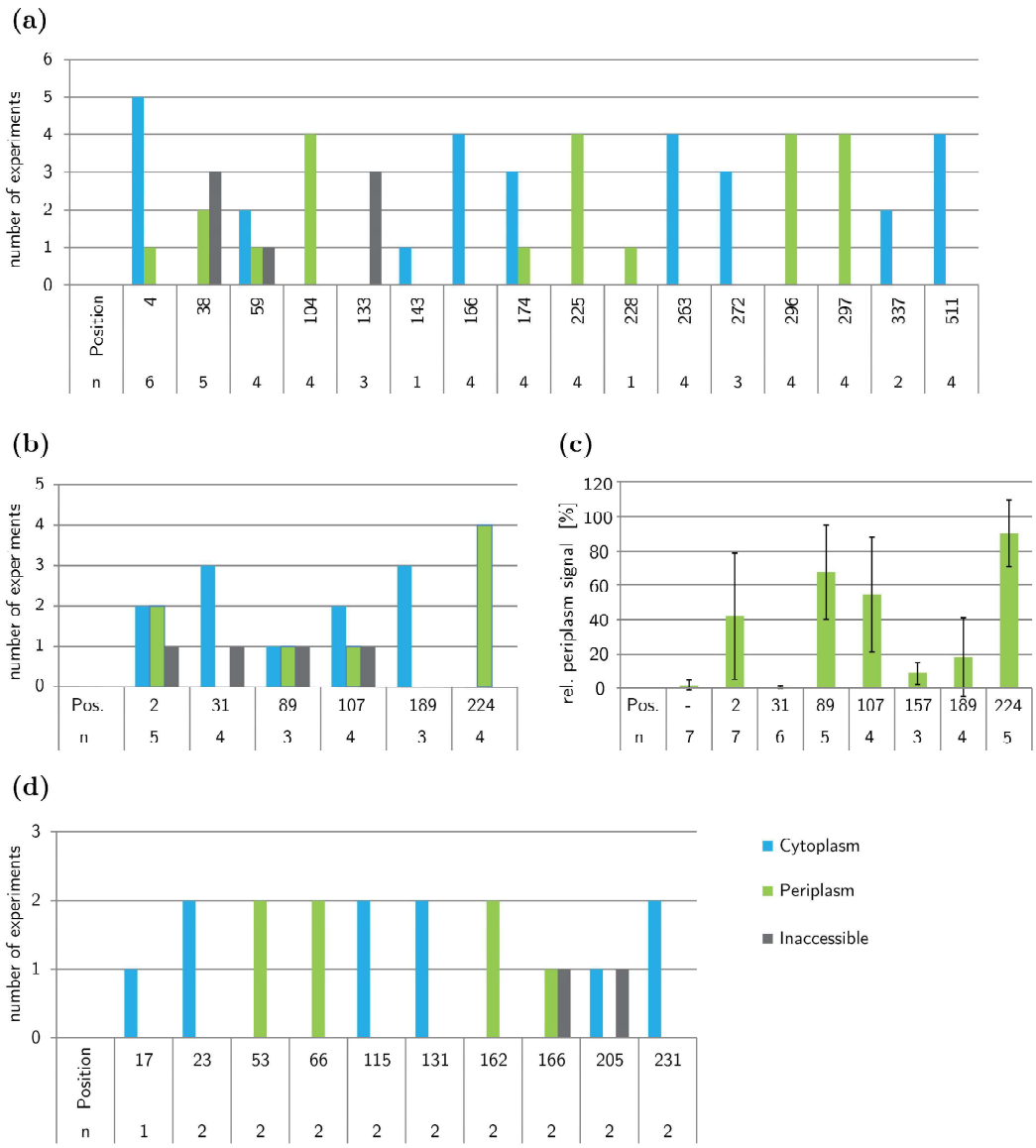


Figure 10: SCAM results of *Salmonella* type III secretion system export apparatus proteins InvA, SpaP and SpaS (SctRUV). Number of experiments for (a) InvA, (b) SpaP and (d) SpaS in which the relative cytoplasmic (blue) or periplasmic (green) signal was stronger. Number of experiments in which both signals were comparable with the negative control are shown in gray. (c) Relative periplasmic signal of SpaP. Error bars: standard error of the mean (manuscript II).

S189C in the cytoplasm and T224C in the periplasm. Due to a signal in the cytoplasmic negative control in many experiments the results of the remaining positions were ambiguous. The results of the periplasmic labelling alone offer an indication of the localization in the periplasm for the positions G2C, S89C and S107C, while the position S157C is likely cytoplasmic or inaccessible (**figure 10c**). Overall, these results point to a N_{OUT}/C_{OUT} orientation together with a large periplasmic loop, contrary to the majority of predictions (**figure 12**).

For the switch protein SpaS seven of ten amino acid positions were mapped (**figure 10d**): S23C, S115C, T131C and S231C cytoplamic and G53C, S66C and S162C periplasmic. The positions S17C and T205C (cytoplasmic) as well as G166C (periplasmic) could only be localized once, respectively. This confirms the predicted N_{IN}/C_{IN} orientation with four TMHs (**figure 13**).

Experimental topology mapping of SpaR (SctT) did not lead to any conclusive results, due to cytoplasmic labelling of the negative control in many experiments similar to SpaP as well as labelling which indicates partial cell lysis during the labelling or blocking steps.

4. Discussion

As T3SSs are present in numerous organisms, structural information is available from various sources. Low resolution electron microscopy images show resemblance between the virulence-associated and the flagella basal body (Kubori et al., 1998; Sekiya et al., 2001) and the vT3SS export apparatus proteins show a high sequence similarity among each other and to the fT3SS (Hueck, 1998). While it is important to keep in mind the difference between the systems, the close relation enables us to combine and compare the structural information gained from different systems in different species.

4.1 Stoichiometry of the type III secretion needle complex

The stoichiometry of some T3SS components has been proposed before. I am able to confirm the previous results *in situ*, as well as present the stoichiometry of all remaining members of the needle complex. The MxiA (SctV, *Shigella*) was reported by electron cryo-tomography to have a nonameric, cytoplasmic ring structure (Abrusci et al., 2013). Analysis of cryo-electron microscope images suggest a 24-fold symmetry in *Salmonella* for the base components PrgH and PrgK (SctDJ) (Schraidt and Marlovits, 2011), while the homologs MxiG and MxiJ in *Shigella* were reported to have a 12-fold symmetry (Hodgkinson et al., 2009). In fT3SSs, FliP (SctR) has been reported with a stoichiometry of 4-5 and FliR (SctT) of 1-3 (Jones et al., 1990). In the same studies, the stoichiometry of the homologs InvG and MxiD (SctC) was suggested as 15 and 12 proteins per complex, respectively. Exact stoichiometries of the export apparatus proteins SctRSTU and the 'inner rod' protein PrgJ (SctI) remained largely unknown.

I can corroborate a 24-fold stoichiometry for both PrgK and PrgH (SctDJ) as well as 15 proteins per complex for InvG (SctC). Based on both PCS experiments and the verification of the ratio of SpaS (SctU) and SpaQ (SctS) by synthetic stable isotope-labelled peptides, the minor export apparatus proteins SpaPQR (SctRST) and the switch protein SpaS (SctU) are shown to have a stoichiometry of 5:1:1:1. The AP experiments demonstrated a weak association of InvA (SctV) to the complex, explaining the slightly lower

number of proteins per complex compared with the previously reported ring structure. The 'inner rod' protein PrgJ (SctI) is thought to connect to the needle and to anchor it inside the base (Marlovits et al., 2006). It has been suggested that PrgJ forms a helix structure similar to PrgI, and crosslinks of up to six PrgJ, indicate at least one complete turn of the 'inner rod' protein (Lefebre and Galán, 2013). The PCS experiments with both purification methods point to a lower number of just three PrgJ. However, since the assembly of the T3SS was not synchronised in culture, it is possible that not all systems were ready for secretion of PrgJ at the point of cell harvest. The affinity purification electron microscopy pictures can not be used to judge the number of actively secreting complexes, as the needles were sheared off during this preparation. In case of the CsCl-GCP preparation, which left the needles attached, approximately half of the bases contain a needle (**figure 5b**). Assuming that likewise only half the bases contain an 'inner rod', the stoichiometry of three PrgJ *in situ* in the PCS experiments would represent six PrgJ *in vivo*. This suggests a PrgJ structure of just one helix turn, rather than an 'inner rod'.

In summary, I can present evidence that the export apparatus of vT3SSs is composed of nine InvA, five SpaP, one SpaQ, one SpaR, and one SpaS, which suggests a total of 99 to 104 transmembrane helices within the membrane patch in the base of these systems.

4.2 Topology of the T3SS export apparatus

The export apparatus is essential for T3SSs to facilitate substrate translocation across the bacterial inner membrane. While high-resolution structures of larger extramembrane domains of the system have been obtained (Zarivach et al., 2008; Worrall et al., 2010; Hu et al., 2017), structural information about the membrane integrated regions is limited.

Bioinformatics analysis of the *Salmonella* SPI-1 vT3SS export apparatus proteins SpaP, SpaQ, SpaR, SpaS and InvA (SctRSTUV) and analysis of homologs of SctRST presented clear evidence for the membrane topology of InvA, SpaQ and SpaS. All three could be mapped with an N_{IN}/C_{IN} orientation. For InvA and SpaS, these results were validated experimentally using SCAM. For the remaining minor export apparatus proteins SpaP and SpaR, the regions for TMHs could be defined; however, the exact number of TMHs

could not be resolved. I did find indications for an N_{OUT}/C_{OUT} orientation of SpaP experimentally and by bioinformatics and for an N_{OUT}/C_{IN} orientation for SpaR by bioinformatics.

A previous model of the major export apparatus protein InvA (SctV) was based on cryo-electron microscopy and showed the extramembrane cytoplasmic C-terminus (Worrall et al., 2010). To this I can now add details about the N-terminal transmembrane domain. The results of the topology mapping of InvA show eight TMHs with cytoplasmic N- and C-terminus and a large cytoplasmic fourth loop (**figure 11**). Earlier topology studies of the homolog YscV (formerly LrcD) in *Y. pestis* via PhoA fusion mapped the first, fifth and seventh loop to be periplasmic, at residues which closely agree with my results (Plano et al., 1991). In *X. campestris* the export apparatus proteins have been studied by *phoA−lacZ* reporter fusions. The study presented evidence for seven TMHs of the InvA homolog HrcV, with an N_{OUT}/C_{IN} orientation (Berger et al., 2010). These results agree with the last seven TMHs I detected, omitting the very first one. The missing TMH is unsurprising as the study used strongly truncated versions of InvA, missing most of the C-terminal information which can have critical influence on the topology.

The results of the SCAM investigation of SpaP (SctR) points to four TMHs with a large periplasmic second loop as well as an N_{OUT}/C_{OUT} orientation (**figure 12**). This is in accordance with N-terminal SpaP interaction with the periplasmic 'inner rod' protein PrgJ (SctI) (manuscript IV, Dietsche et al. (2016)). Additionally, some of the SpaR homologs (manuscript II, supplement S11) in flagella systems may include signal peptides which also point to an N_{OUT} orientation (Wang et al., 2008).

As with InvA, a 3D-structure of the cytoplasmic C-terminus of SpaS (SctU) has been published previously (Zarivach et al., 2008). Topology mapping of this protein shows four TMHs with an N_{IN}/C_{IN} orientation (**figure 13**). This is consistent with the data obtained in a PhoA fusion study in *Y. enterolitica* (Allaoui et al., 1994) which I can complement with reliable cytoplasmic signals.

Topology mapping of SpaR (SctT) proved to be more challenging than for the other export apparatus proteins. Its homolog in *Xanthomonas*, HrcT, was reported to have only one N-terminal TMH which stands in sharp contrast to my bioinformatics predictions (**figure 7**; manuscript II: supplement S5),

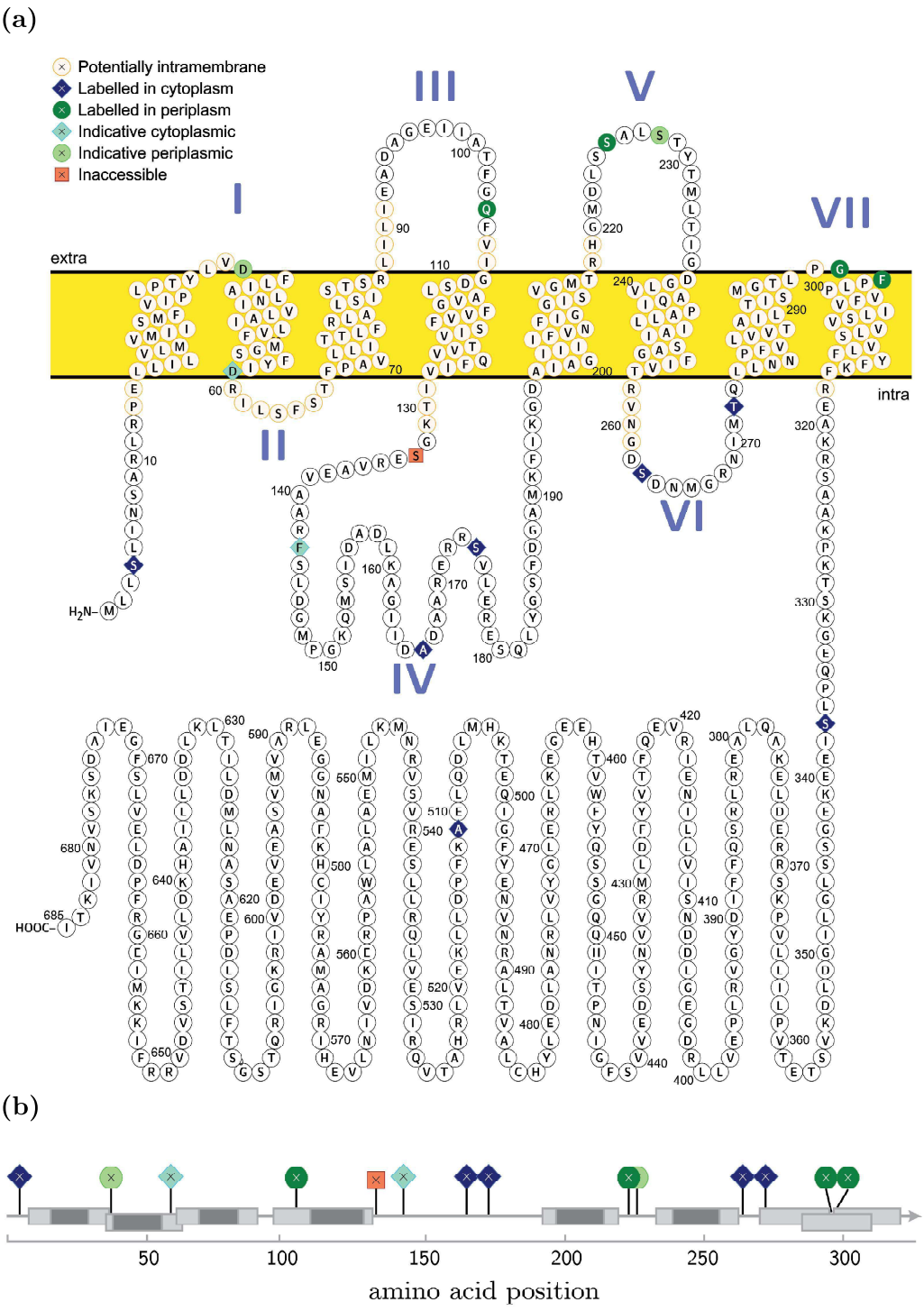
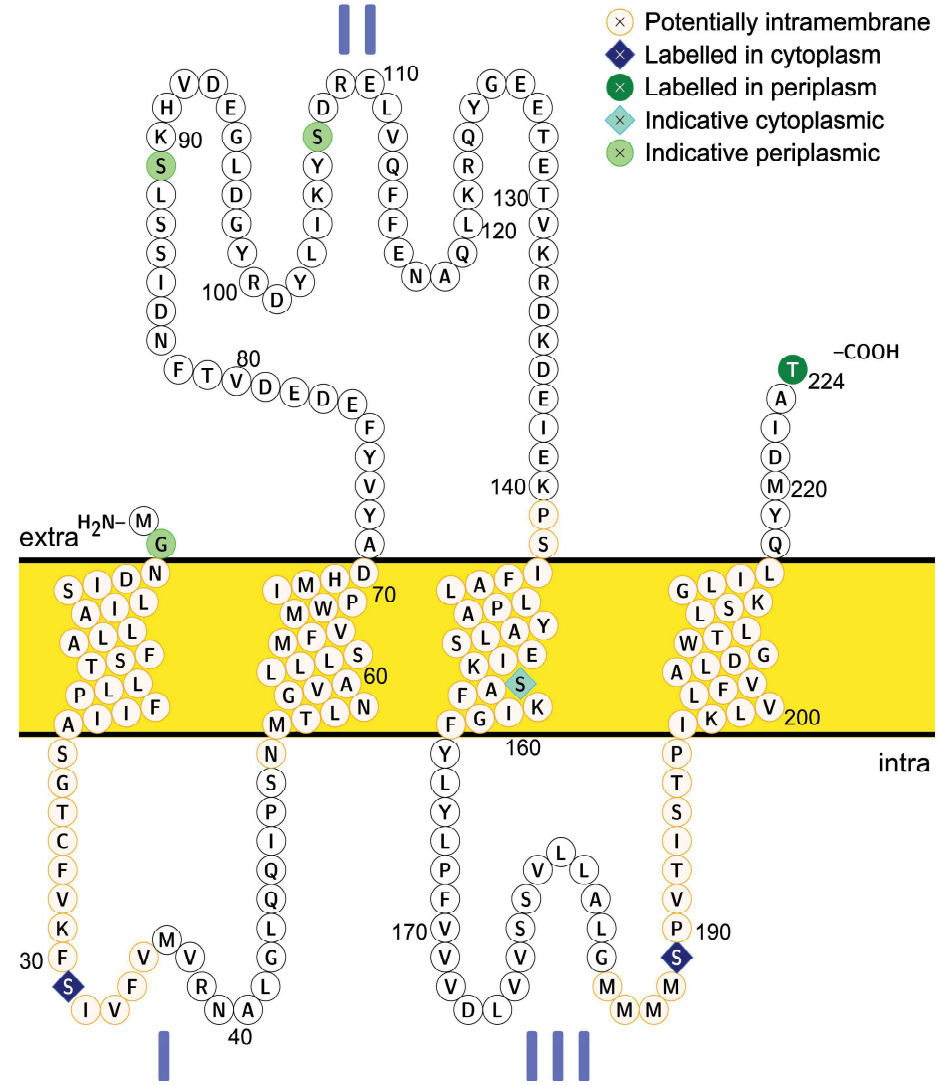


Figure 11: (a) Topology mapping of *Salmonella* type III secretion system export apparatus protein InvA (SctV). Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool (Omasits et al., 2014). (b) Mapped amino acid positions added to topology predictions of figure 7. InvA C-terminal cytosolic part after amino acid 350 and cytosolic mapped position 511 are omitted. Coloring according to legend of a) (manuscript II).

(a)



(b)

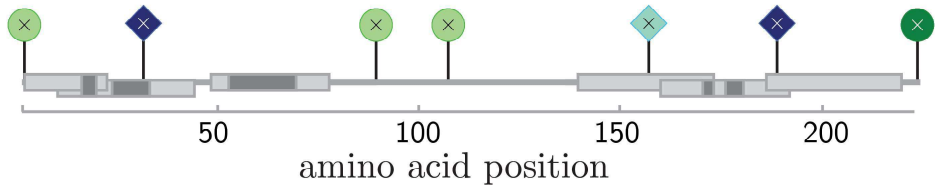
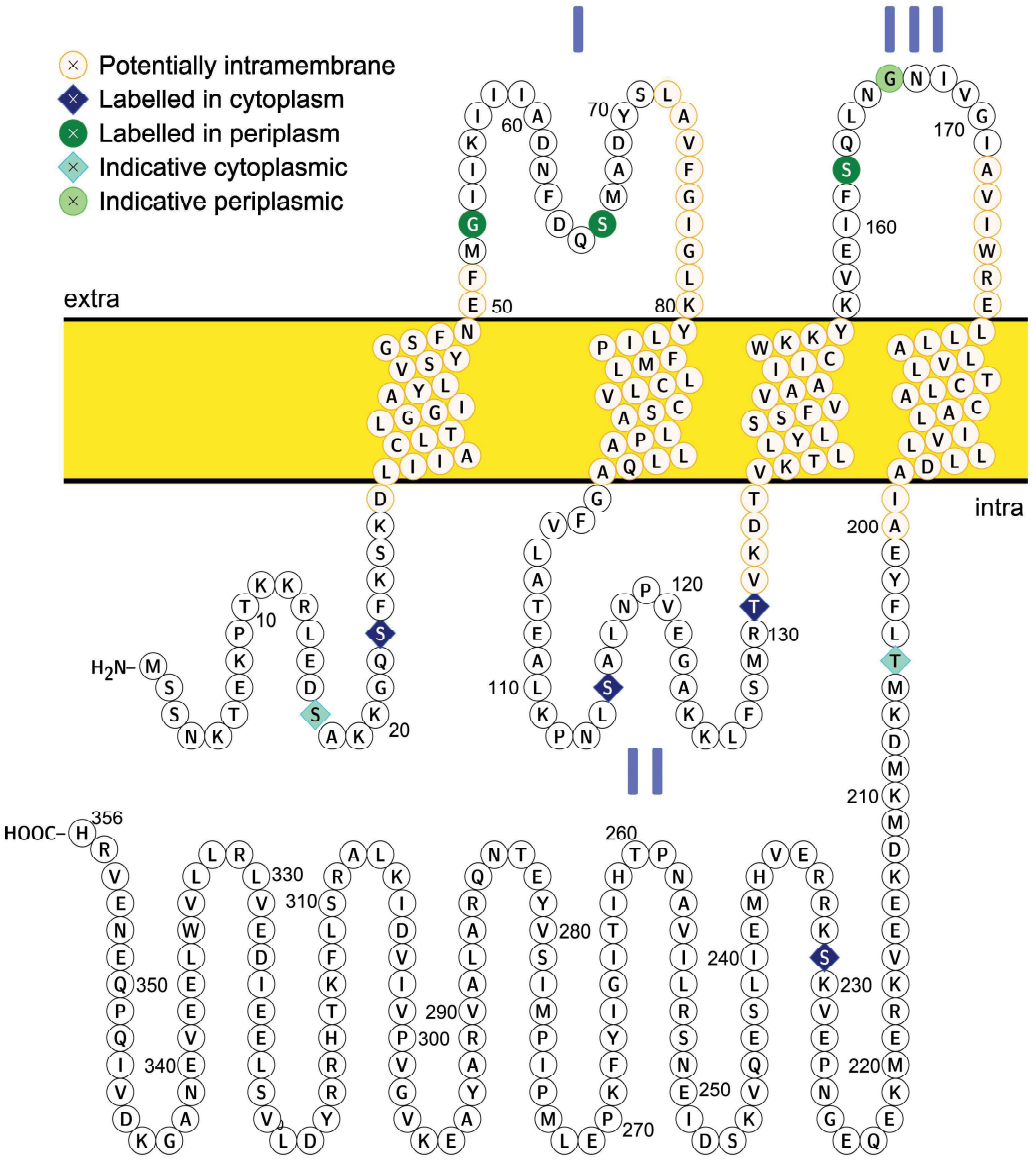


Figure 12: (a) Topology mapping of *Salmonella* type III secretion system export apparatus protein SpaP (SctR). Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool (Omasits et al., 2014). (b) Mapped amino acid positions added to topology predictions of figure 7. Coloring according to legend of a) (manuscript II).

(a)



(b)

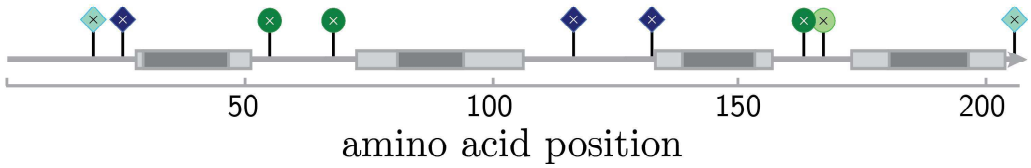


Figure 13: (a) Topology mapping of *Salmonella* type III secretion system export apparatus protein SpaS (SctU). Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool (Omasits et al., 2014). (b) Mapped amino acid positions added to topology predictions of figure 7, C-terminal cytosolic part after amino acid 205 is omitted. Coloring according to legend of a) (manuscript II).

which suggested four to six TMHs (Berger et al., 2010). The study used a *phoA−lacZ* reporter fusion and largely concurs with my results for the export apparatus proteins SctRUV. However, as mentioned before, these reporter studies are based on heavily truncated versions, making the results debatable, as SpaR seems to be especially sensitive to even small changes such as amino acid substitutions (data not shown).

In bacterial genomes, *sctT* and *sctU* are often located consecutive in the same operon (Van Arnam et al., 2004) and equivalents of both genes exists as native gene fusions in *Clostridium acetobutylicum* and *Clostridium tetani* (Brüggemann et al., 2003; Nölling et al., 2001). This could suggest that the C-terminal part of SctT is in the cytoplasm alongside the N-terminus of SctU. In *S.* Typhimurium fT3SS a fusion of the homologs *fliR-flhB* was published, though function was strongly reduced (Van Arnam et al., 2004). In the *S.* Typhimurium SPI-1 vT3SS *spaR* and *spaS* are consecutive with a fourteen nucleotide overlap (Kröger et al., 2012) and while I was able to construct a SpaRS fusion, the fusion was not functional for secretion (data not shown). Furthermore, preliminary cross-link MS data (not shown) suggest interaction between both the second to last extramembrane loop and the C-terminus of SpaR with the periplasmic second loop of SpaP. Together with my cytoplasmic prediction of the last loop of SctT homologs, this underpins the likelihood of several TMHs at the end of SpaR with a C_{OUT} orientation.

Previously, density differences were observed in the centre of the base between averaged cryo-electron microscopy images of wild type and export apparatus knock out mutants (Δ*invA* Δ*spaPQRS* and Δ*spaP*). The structures representing these differences were termed 'cup' and 'socket' (Wagner et al., 2010; Marlovits et al., 2004). The proteins SpaPQR (SctRST) (Wagner et al., 2010) or PrgJ (SctI) with periplasmic loops of SpaPR (SctRT) (manuscript IV, Dietsche et al. (2016)) were suggested to account for the unassigned density of the socket and form of a defined substructure. I was able to confirm that the large loop of SpaP is periplasmic, although the periplasmic part of SpaR remains undefined. Nevertheless, the 5:1 stoichiometry of the SpaPR subcomplex means that the five SpaP, alongside six PrgJ (manuscript I, Zilkenat et al. (2016)) could represent a major part of the socket structure. Neither InvA (SctV) nor SpaS (SctU) contain large periplasmic loops. InvA though, of which a total of nine copies are present in the T3SS (Abrusci et al., 2013; Zilkenat et al., 2016) includes a large cyto-

plasmic fourth loop to which no structure has been assigned yet. In FlhA (SctV) this loop has been suggested to play a role in inwardly-directed proton flow and may interact with FliR (SctT) (Hara et al., 2011).

In summary, I was able to define the topologies of the major export apparatus protein InvA and the switch protein SpaS, and ascertain the periplasmic orientation of the largest loop and the termini of SpaP. I also found indications for cytoplasmic orientations of the termini of SpaQ and a C_{OUT} orientation for SpaR.

5. Conclusion

The export apparatus of type III secretion systems is an essential part of these membrane-spanning macromolecular machines. However, its localization in the inner membrane at the very center of the complex has made structural analysis very challenging.

Taken together, the four manuscripts presented in this work illustrate how indirect structural information of macromolecular membrane complexes can be obtained by a variety of methods.

The stoichiometric information of not only the export apparatus proteins, but also of the base and 'inner rod' components was gained *in situ* by a ratiometric mass spectrometry approach. These results further our understanding of the makeup of T3SS and demonstrate the applicability of these methods to complexes of very heterogeneous composition and wide stoichiometric range (Zilkenat et al., 2016, 2017).

The topology predictions and mapping of the export apparatus proteins by labelling of single cysteine residues furthers structural studies and assists in elucidating the functions of export apparatus proteins not only in vT3SSs but also in fT3SSs (manuscript II).

The data was subsequently used to support *in vivo* photocrosslinking analysis which showed that the pentamer of SpaP (SctR) forms a central translocation channel in vT3SS, which may be gated through a combination of SpaQRS (SctSTU), all present in only one copy (Zilkenat et al., 2016; Dietsche et al., 2016).

The low stoichiometric number of PrgJ (SctI) together with crosslinks to SpaP and SpaR (SctR and SctS) demonstrates that the term 'inner rod' is misleading, but that these proteins probably build a short, direct connector between needle and base.

This information will further structural and functional analyses and help to develop anti-infective strategies to target both the virulence-associated and the related flagellar systems, which are central virulence factors of many pathogens.

6. Bibliography

Abby, S. S. and Rocha, E. P. C. The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet.*, 8(9), 2012.

Abrusci, P., Vergara-Irigaray, M., Johnson, S., Beeby, M. D., Hendrixson, D. R., Roversi, P., Friede, M. E., Deane, J. E., Jensen, G. J., Tang, C. M., and Lea, S. M. Architecture of the major component of the type III secretion system export apparatus. *Nat. Struct. Mol. Biol.*, 20(1):99–104, jan 2013.

Akeda, Y. and Galan, J. E. Genetic analysis of the Salmonella enterica type III secretion-associated ATPase InvC defines discrete functional domains. *J. Bacteriol.*, 186(8):2402–2412, 2004.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. Molecular biology of the cell - The lipid bilayer. In *New York Garl. Sci.*, page Available from: https://www.ncbi.nlm.nih.gov/books. Garland Science, 4th edition, 2002.

Allaoui, A., Woestyn, S., Sluiters, C., and Cornelis, G. R. Yscu, a Yersinia enterocolitica inner membrane-protein involved in Yop secretion. *J. Bacteriol.*, 176(15):4534–4542, 1994.

Andersson, H., Bakker, E., and Von Heijne, G. Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in Escherichia coli. *J. Biol. Chem.*, 267(3):1491–1495, 1992.

Arnold, T. and Linke, D. The use of detergents to purify membrane proteins. In *Curr. Protoc. Protein Sci.*, chapter 4.8, pages 1–35. John Wiley & Sons, Inc., 53 edition, 2008. ISBN 0471140864.

Bai, X.-c., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S. H. W., and Shi, Y. An atomic structure of human γ-secretase. *Nature*, 512(7568):212–217, 2015.

Barker, C. S., Inoue, T., Meshcheryakova, I. V., Kitanobo, S., and Samatey, F. A. Function of the conserved FHIPEP domain of the flagellar type III export apparatus, protein FlhA. *Mol. Microbiol.*, 100(2):278–288, 2016.

Berger, C., Robin, G. P., Bonas, U., and Koebnik, R. Membrane topology of conserved components of the type III secretion system from the plant pathogen Xanthomonas campestris pv. vesicatoria. *Microbiology*, 156: 1963–1974, 2010.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., Heijne, G. V., and Elofsson, A. Prediction of membrane-protein topology from first principles. *PNAS*, 105(20):7177–7181, 2008.

Beynon, R. J., Doherty, M. K., Pratt, J. M., and Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods*, 2(8):587–589, 2005.

Bogdanov, M., Zhang, W., Xie, J., and Dowhan, W. Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis. *Methods*, 36:148–171, 2005.

Brown, L. S. and Ladizhansky, V. Membrane proteins in their native habitat as seen by solid-state NMR spectroscopy. *Protein Sci.*, 24:1333–1346, 2015.

Brüggemann, H., Bäumer, S., Fricke, W. F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A., and Gottschalk, G. The genome sequence of Clostridium tetani, the causative agent of tetanus disease. *Proc. Natl. Acad. Sci. U. S. A.*, 100(3):1316–21, 2003.

Büttner, D. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol. Mol. Biol. Rev.*, 76(2):262–310, 2012.

Büttner, D. and He, S. Y. Type III protein secretion in plant pathogenic bacteria. *Plant Physiol.*, 150(4):1656–1664, 2009.

Carpenter, E. P., Beis, K., Cameron, A. D., and Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.*, 18:581–586, 2008.

Chakravortty, D., Rohde, M., Jäger, L., Deiwick, J., and Hensel, M. Formation of a novel surface structure encoded by Salmonella pathogenicity island 2. *EMBO J.*, 24(11):2043–52, 2005.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148): 802–5, 1994.

Chang, J. M., Di Tommaso, P., and Notredame, C. TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.*, 31(6):1625–1637, 2014.

Clark, K. M., Fedoriw, N., Robinson, K., Connelly, S. M., Randles, J., Malkowski, M. G., Detitta, G. T., and Dumont, M. E. Purification of Transmembrane Protein from Saccharomyces cerevisiae for X-ray Crystallography. *Protein Expr Purif*, 71(2):207–223, 2011.

Collazo, C. M. and Galán, J. E. Requirement for exported proteins in secretion through the invasion-associated type III system of Salmonella Typhimurium. *Infect. Immun.*, 64(9):3524–3531, 1996.

Collazo, C. M. and Galán, J. E. The invasion-associated type III system of Salmonella typhimurium directs the translocation of Sip proteins into the host cell. *Mol. Microbiol.*, 24(4):747–756, 1997.

Cornelis, G. The type III secretion injectisome. *Nat. Rev. Microbiol.*, 4(11): 811–825, 2006.

Crichton, P. G., Harding, M., Ruprecht, J. J., Lee, Y., and Kunji, E. R. S. Lipid, detergent, and Coomassie Blue G-250 affect the migration of small membrane proteins in blue native gels: mitochondrial carriers migrate as monomers not dimers. *J. Biol. Chem.*, 288(30):22163–73, jul 2013.

Cymer, F., von Heijne, G., and White, S. H. Mechanisms of integral membrane protein insertion and folding. *Biophys. Chem.*, 427(5):999–1022, 2015.

Diepold, A. and Armitage, J. P. Type III secretion systems: the bacterial flagellum and the injectisome. *Philos. Trans. R. Soc. B*, 370(1679): 20150020, 2015.

Diepold, A. and Wagner, S. Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.*, pages 1–21, feb 2014.

Diepold, A., Wiesand, U., and Cornelis, G. R. The assembly of the export apparatus (YscR,S,T,U,V) of the Yersinia type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer. *Mol. Microbiol.*, 82(2):502–14, oct 2011.

Diepold, A., Kudryashev, M., Delalez, N. J., and Berry, R. M. Composition , formation , and regulation of the cytosolic C-ring, a dynamic component of the type III secretion injectisome. *PLoS Biol.*, 13(1):1–21, 2015.

Dietsche, T., Mebrhatu, M. T., Brunner, M. J., Zilkenat, S., Abrusci, P., Yan, J., Franz-wachtel, M., Scha, C., Kohlbacher, O., Lea, S., Macek, B., Grin, I., Gala, J. E., Marlovits, C., Robinson, C. V., and Wagner, S. Structural and functional characterization of the bacterial type III secretion export apparatus. *PLoS Pathog.*, 12(12):1–25, 2016.

Duncan, M. C., Linington, R. G., and Auerbuch, V. Chemical inhibitors of the type three secretion system: Disarming bacterial pathogens. *Antimicrob. Agents Chemother.*, 56(11):5433–5441, 2012.

Ehrmann, M., Boyd, D., and Beckwith, J. Genetic analysis of membrane protein topology by a sandwich gene fusion approach. *Proc. Natl. Acad. Sci. U. S. A.*, 87(19):7574–7578, 1990.

Elazar, A., Weinstein, J., Prilusky, J., and Fleishman, S. J. The interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc. Natl. Acad. Sci.*, 113(37):in press, 2016.

Elofsson, A. and von Heijne, G. Membrane protein structure: prediction versus reality. *Annu Rev Biochem*, 76(1):125–140, 2007.

Erhardt, M., Wheatley, P., Kim, E. A., Hirano, T., Zhang, Y., Sarkar, M. K., Hughes, K. T., and Blair, D. F. Mechanism of type-III protein secretion : Regulation of FlhA conformation by a functionally critical charged-residue cluster. *Mol Microbiol*, Accepted A, 2017.

Ferris, H. U. and Minamino, T. Flipping the switch: Bringing order to flagellar assembly. *TRENDS Microbiol.*, 14(12):519, 2006.

Ferris, H. U., Furukawa, Y., Minamino, T., Kroetz, M. B., Kihara, M., Namba, K., and Macnab, R. M. FlhB regulates ordered export of flagellar components via autocleavage mechanism. *J. Biol. Chem.*, 280(50):41236–41242, 2005.

Figueira, R. and Holden, D. W. Functions of the Salmonella pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology*, 158: 1147–1161, 2012.

Fischer, M., Zilkenat, S., Gerlach, R. G., Wagner, S., and Renard, B. Y. Pre- and post-processing workflow for affinity purification mass spectrometry data. *J. Proteome Res.*, 13(5):2239–2249, 2014.

Frost, S., Ho, O., Login, F. H., Weise, C. F., Wolf-Watz, H., and Wolf-Watz, M. Autoproteolysis and intramolecular dissociation of Yersinia YscU precedes secretion of its C-terminal polypeptide YscUCC. *PLoS One*, 7 (11), 2012.

Galán, J. E. Salmonella interactions with host cells: Type III secretion at work. *Annu. Rev. Cell Dev. Biol.*, 17:53–86, 2001.

Galán, J. E. and Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, 444(7119):567–573, 2006.

Galán, J. E., Lara-Tejero, M., Marlovits, T. C., and Wagner, S. Bacterial type III secretion systems: Specialized nanomachines for protein delivery into target cells. *Annu. Rev. Microbiol.*, 68:415–38, 2014.

Goldbourt, A. Biomolecular magic-angle spinning solid-state NMR: Recent methods and applications. *Curr. Opin. Biotechnol.*, 24(4):705–715, 2013.

Grant, S. R., Fisher, E. J., Chang, J. H., Mole, B. M., and Dangl, J. L. Subterfuge and manipulation: Type III effector proteins of phytopathogenic bacteria. *Annu. Rev. Microbiol.*, 60:425–449, 2006.

Hagan, C. L., Silhavy, T. J., and Kahne, D. β-Barrel membrane protein assembly by the BAM complex. *Annu. Rev. Biochem*, 80:189–210, 2011.

Hara, N., Namba, K., and Minamino, T. Genetic characterization of conserved charged residues in the bacterial flagellar type III export protein FlhA. *PLoS One*, 6(7):e22417, 2011.

Hendrickson, W. A. Atomic-level analysis of membrane-protein structure. *Nat. Struct. Mol. Biol.*, 23(6):464–467, 2016.

Hensel, M., Shea, J. E., Bäumler, A. J., Gleeson, C., Blattner, F., and Holden, D. W. Analysis of the boundaries of Salmonella pathogenicity island 2 and the corresponding chromosomal region of Escherichia coli K-12. *J. Bacteriol.*, 179(4):1105–11, 1997.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H., and von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433:377–81, 2005.

Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S. H., and von Heijne, G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450 (7172):1026–30, dec 2007.

Hodgkinson, J. L., Horsley, A., Stabat, D., Simon, M., Johnson, S., da Fonseca, P. C. a., Morris, E. P., Wall, J. S., Lea, S. M., and Blocker, A. J. Three-dimensional reconstruction of the Shigella T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nat. Struct. Mol. Biol.*, 16(5):477–85, 2009.

Hu, B., Morado, D. R., Margolin, W., Rohde, J. R., Arizmendi, O., Picking, W. L., Picking, W. D., and Liu, J. Visualization of the type III secretion sorting platform of Shigella flexneri. *PNAS*, 112(4):1047–52, 2015.

Hu, B., Lara-Tejero, M., Kong, Q., Galán, J. E., and Liu, J. In situ molecular architecture of the Salmonella type III secretion machine. *Cell*, 168:1–10, 2017.

Hueck, C. J. Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants. *Microbiol. Mol. Biol. Rev.*, 62(2):379–433, 1998.

Hyun, S. I., Maruri-Avidal, L., and Moss, B. Topology of endoplasmic reticulum-associated cellular and viral proteins determined with split-GFP. *Traffic*, 16(7):787–795, 2015.

Jones, C. J., Macnab, R. M., Okino, H., and Aizawa, S. I. Stoichiometric analysis of the flagellar hook-(basal-body) complex of Salmonella Typhimurium. *J Mol Biol.*, 212(2):377–387, 1990.

Jones, D. T. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics*, 23(5):538–544, 2007.

Journet, L., Agrain, C., Broz, P., and Cornelis, G. R. The needle length of bacterial injectisomes is determined by a molecular ruler. *Science*, 302 (5651):1757–1760, 2003.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.*, 338(5): 1027–1036, 2004.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics*, 21(SUPPL. 1):251–257, 2005.

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., Ishikawa, H., Leonetti, M. D., Marshall, W. F., Weissman, J. S., and Huang, B. Versatile protein tagging in cells with split fluorescent protein. *Nat. Commun.*, 7:11046, 2016.

Kito, K., Ota, K., Fujita, T., and Ito, T. A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.*, 6(2):792–800, 2007.

Klara, S. S., Saboe, P. O., Sines, I. T., Babaei, M., Chiu, P. L., Dezorzi, R., Dayal, K., Walz, T., Kumar, M., and Mauter, M. S. Magnetically Directed Two-Dimensional Crystallization of OmpF Membrane Proteins in Block Copolymers. *J. Am. Chem. Soc.*, 138(1):28–31, 2016.

Knockenhauer, K. E. and Schwartz, T. U. The nuclear pore complex as a flexible and dynamic gate. *Cell*, 164(6):1162–1171, 2016.

Koch, H.-G., Kuhn, A., and Dalbey, R. E. Targeting and insertion of membrane proteins. *EcoSal Plus*, 7(2), 2017.

Kowal, J., Chami, M., Ringler, P., Müller, S. a., Kudryashev, M., Castaño-Díez, D., Amstutz, M., Cornelis, G. R., Stahlberg, H., and Engel, A. Structure of the dodecameric Yersinia enterocolitica secretin YscC and its trypsin-resistant core. *Structure*, 21(12):2152–61, 2013.

Kröger, C., Dillon, S. C., Cameron, A. D. S., Papenfort, K., Sivasankaran, S. K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G. C., Lohan, A. J., Loftus, B., Lucchini, S., Ussery, D. W., Dorman, C. J., Thomson, N. R., Vogel, J., and Hinton, J. C. D. The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc. Natl. Acad. Sci.*, 109(20): E1277–E1286, 2012.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galán, J. E., and Shin. Supramolecular Structure of the Salmonella typhimurium Type III Protein Secretion System. *Science*, 280 (5363):602–605, 1998.

Kubori, T., Sukhan, A., Aizawa, S.-i., and Galan, J. E. Molecular characterization and assembly of the needle complex of the Salmonella Typhimurium type III protein secretion system. *PNAS*, 2941(9):10225–10230, 2000.

Kudryashev, M., Stenta, M., Schmelz, S., Amstutz, M., Wiesand, U., Castaño-Díez, D., Degiacomi, M. T., Münnich, S., Bleck, C. K. E., Kowal, J., Diepold, A., Heinz, D. W., Dal Peraro, M., Cornelis, G. R., and Stahlberg, H. In situ structural analysis of the Yersinia enterocolitica injectisome. *Elife*, 2013(2):1–23, 2013.

Kyte, J. and Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, 157(1):105–32, 1983.

Lara-Tejero, M., Kato, J., Wagner, S., Liu, X., and Galan, J. E. A sorting platform determines the order of protein secretion in bacterial type III systems. *Science*, 331:1188, 2011.

Lau, W. C. Y. and Rubinstein, J. L. Structure of intact Thermus thermophilus V-ATPase by cryo-EM reveals organization of the membrane-bound V(O) motor. *Proc. Natl. Acad. Sci. U. S. A.*, 107(4):1367–72, 2010.

Lee, L. K., Stewart, A. G., Donohoe, M., Bernal, R. A., and Stock, D. The structure of the peripheral stalk of Thermus thermophilus H+ - ATPase/synthase. *Nat. Struct. Mol. Biol.*, 17(3):373–378, 2010.

Lefebre, M. D. and Galán, J. E. The inner rod protein controls substrate switching and needle length in a Salmonella type III secretion system. *Proc. Natl. Acad. Sci. U. S. A.*, 2013(2):817–22, 2013.

Loquet, A., Sgourakis, N. G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., Griesinger, C., Kolbe, M., Baker, D., Becker, S., and Lange, A. Atomic model of the type III secretion system needle. *Nature*, 486(7402):276–9, 2012.

Makishima, S., Komoriya, K., Yamaguchi, S., and Aizawa, S.-I. Length of the flagellar hook and the capacity of the type III export apparatus. *Science*, 291(5512):2411–2413, 2001.

Manoil, C., Mekalanos, J. J., and Beckwith, J. Alkaline phosphatase fusions: Sensors of subcellular location. *J. Bacteriol.*, 172(2):515–518, 1990.

Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Jorge, E., Unger, V. M., Galán, J. E., and Unger, V. M. Structural insights into the assembly

of the type III secretion needle complex. *Science*, 306(5698):1040–1042, 2004.

Marlovits, T. C., Kubori, T., Lara-Tejero, M., Thomas, D., Unger, V. M., and Galán, J. E. Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature*, 441(7093):637–40, 2006.

Marshall, N. C. and Finlay, B. B. Targeting the type III secretion system to treat bacterial infections. *Expert Opin. Ther. Targets*, 18(2):137–52, 2014.

Melén, K., Krogh, A., and Von Heijne, G. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.*, 327(3):735–744, 2003.

Monjarás Feria, J. V., Lefebre, M. D., Stierhof, Y.-D., Galán, J. E., and Wagner, S. Role of autocleavage in the function of a type III secretion specificity switch protein in Salmonella enterica serovar Typhimurium. *MBio*, 6(5):1–8, 2015.

Myeni, S. K., Wang, L., and Zhou, D. SipB-SipC complex is essential for translocon formation. *PLoS One*, 8(3):1–7, 2013.

Nelson, D. L. and Cox, M. M. Biological Membranes and Transport. In *Lehninger Princ. Biochem.*, chapter 11, pages 369–420. Sara Tenney, New York, NY, 4th edition, 2005.

Nilsson, I. and von Heijne, G. Fine-tuning the topology of a polytopic membrane protein: Role of positively and negatively charged amino acids. *Cell*, 62(6):1135–1141, 1990.

Nölling, J., Breton, G., Omelchenko, M. V., Kira, S., Zeng, Q., Gibson, R., Lee, H. M., Dubois, J., Qiu, D., Hitti, J., Sequencing, G. T. C., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V., and Smith, D. R. Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. *J. Bacteriol.*, 183(16):a823–4838, 2001.

Notredame, C., Higgins, D., and Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 302(1): 205–217, 2000.

Notti, R. Q., Bhattacharya, S., Lilic, M., and Stebbins, C. E. A common assembly module in injectisome and flagellar type III secretion sorting platforms. *Nat. Commun.*, 6(May):1–11, 2015.

Omasits, U., Ahrens, C. H., Müller, S., and Wollscheid, B. Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, 30(6):884–886, 2014.

Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.*, 5(12):993–6, 2006.

Pan, Y., Ruan, X., Valvano, M. a., and Konermann, L. Validation of membrane protein topology models by oxidative labeling and mass spectrometry. *J. Am. Soc. Mass Spectrom.*, 23(5):889–98, 2012.

Papaloukas, C., Granseth, E., Viklund, H., and Elofsson, A. Estimating the length of transmembrane helices using Z-coordinate predictions. *Protein Sci.*, 17:271–278, 2008.

Plano, G. V., Barve, S. S., and Straley, S. C. LcrD, a membrane-bound regulator of the Yersinia pestis low-calcium response. *J. Bacteriol.*, 173 (22):7293–7303, 1991.

Reynolds, S. M., Käll, L., Riffle, M. E., Bilmes, J. A., and Noble, W. S. Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. *PLoS Comput. Biol.*, 4(11), 2008.

Riordan, K. E. and Schneewind, O. YscU cleavage and the assembly of Yersinia type III secretion machine complexes. *Mol. Microbiol.*, 68(6): 1485–1501, 2008.

Rollauer, S. E., Sooreshjani, M. A., Noinaj, N., and Buchanan, S. K. Outer membrane protein biogenesis in Gram-negative bacteria. *Philos. Trans. R. Soc. London B Biol. Sci.*, 370(20150023), 2015.

Rost, B., Fariselli, P., and Casadio, R. Topology prediction for helical transmembrane proteins at 86% accuracy-Topology prediction at 86% accuracy. *Protein Sci.*, 5(8):1704–1718, 1996.

Ruiz, N., Kahne, D., and Silhavy, T. J. Advances in understanding bacterial outer-membrane biogenesis. *Nat. Rev. Microbiol.*, 4(1):57–66, 2006.

Schlegel, S., Hjelm, A., Baumgarten, T., Vikström, D., and Gier, J.-w. D. Bacterial-based membrane protein production. *BBA - Mol. Cell Res.*, 1843 (8):1739–1749, 2014.

Schraidt, O. and Marlovits, T. C. Three-dimensional model of Salmonella's needle complex at subnanometer resolution. *Science*, 331:1192–1195, 2011.

Schraidt, O., Lefebre, M. D., Brunner, M. J., Schmied, W. H., Schmidt, A., Radics, J., Mechtler, K., Galán, J. E., and Marlovits, T. C. Topology and organization of the Salmonella Typhimurium type III secretion needle complex components. *PLoS Pathog.*, 6(4):e1000824, 2010.

Sekiya, K., Ohishi, M., Ogino, T., Tamano, K., Sasakawa, C., and Abe, A. Supermolecular structure of the enteropathogenic Escherichia coli type III secretion system and its direct interaction with the EspA-sheath-like structure. *Proc Natl Acad Sci U S A*, 98(20):11638–11643, 2001.

Seppälä, S., Slusky, J. S., Lloris-Garcerá, P., Rapp, M., and von Heijne, G. Control of membrane protein topology by a single C-terminal residue. *Science*, 328(5986):1698–1700, 2010.

Shahid, S. A., Bardiaux, B., Franks, W. T., Krabben, L., Habeck, M., van Rossum, B. J., and Linke, D. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nat. Methods*, 9(12): 1212–1217, 2012.

Shen, D. K., Moriya, N., Martinez-Argudo, I., and Blocker, A. J. Needle length control and the secretion substrate specificity switch are only loosely coupled in the type III secretion apparatus of Shigella. *Microbiol. (United Kingdom)*, 158(7):1884–1896, 2012.

Silhavy, T. J. and Beckwith, J. R. Uses of lac fusions for the study of biological problems. *Microbiol. Rev.*, 49(4):398–418, 1985.

Stuwe, T., Correia, A. R., Lin, D. H., Paduch, M., Lu, V. T., and Kossiakoff, A. A. Architecture of the nuclear pore complex coat. *Science*, 347(6226): 1148–1152, 2015.

Sukhan, A., Kubori, T., Galán, J. E., and Gala, J. E. Synthesis and localization of the Salmonella SPI-1 type III secretion needle complex proteins PrgI and PrgJ. *J. Bacteriol.*, 185(11):3480–3483, 2003.

Tamano, K., Aizawa, S., Katayama, E., Nonaka, T., Imajoh-Ohmi, S., Kuwae, A., Nagai, S., and Sasakawa, C. Supramolecular structure of the Shigella type III secretion machinery: The needle part is changeable in length and essential for delivery of effectors. *EMBO J.*, 19(15):3876–3887, 2000.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L., and Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.*, 43(W1):W401–7, 2015.

Van Arnam, J. S., McMurry, J. L., Kihara, M., and Macnab, R. M. Analysis of an engineered Salmonella flagellar fusion protein, FliR-FlhB. *J. Bacteriol.*, 186(8):2495–2498, 2004.

van Geest, M. and Lolkema, J. S. Membrane topology and insertion of membrane proteins: search for topogenic signals. *MMBR*, 64(1):13–33, 2000.

Viklund, H. and Elofsson, A. OCTOPUS: Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics*, 24(15):1662–8, 2008.

Viklund, H., Bernsel, A., Skwark, M., and Elofsson, A. SPOCTOPUS: A combined predictor of signal peptides and membrane protein topology. *Bioinformatics*, 24(24):2928–2929, 2008.

von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A. L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., Buczak, K., Mosalaganti, S., Hagen, W., Andres-Pons, A., Lemke, E. a., Bork, P., Antonin, W., Glavy, J. S., Bui, K. H., and Beck, M. In situ structural analysis of the human nuclear pore complex. *Nature*, 526(7571):140–143, 2015.

von Heijne, G. Membrane proteins: The amino acid composition of membrane-penetrating segments. *Eur. J. Biochem.*, 120(2):275–278, 1981.

von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*, 341:456–458, 1989.

Wagner, S., Königsmaier, L., Lara-tejero, M., Lefebre, M., Marlovits, T. C., and Galán, J. E. Organization and coordinated assembly of the type III secretion export apparatus. *PNAS*, 107(41):17745–17750, 2010.

Wallin, E. and von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.*, 7(4):1029–38, 1998.

Wang, P., Shim, E., Cravatt, B., Jacobsen, R., Schoeniger, J., Kim, A. C., Paetzel, M., and Dalbey, R. E. Escherichia coli signal peptide peptidase A is a serine-lysine protease with a lysine recruited to the nonconserved amino-terminal domain in the S49 protease family. *Biochemistry*, 47(24): 6361–6369, 2008.

Wang, Y., Sun, M., Bao, H., Zhang, Q., and Guo, D. Effective identification of bacterial type III secretion signals using joint element features. *PLoS One*, 8(4):e59754, 2013.

Ward, M. E., Wang, S., Krishnamurthy, S., Hutchins, H., Fey, M., Brown, L. S., and Ladizhansky, V. High-resolution paramagnetically enhanced solid-state NMR spectroscopy of membrane proteins at fast magic angle spinning. *J. Biomol. NMR*, 58(1):37–47, 2014.

Wee, D. H. and Hughes, K. T. Molecular ruler determines needle length for the Salmonella Spi-1 injectisome. *Proc. Natl. Acad. Sci. U. S. A.*, 112(13): 4098–103, 2015.

White, S. H. and Heijne, G. V. The machinery of membrane protein assembly. *Curr. Opin. Struct. Biol.*, 14(4):397–404, 2004.

Whitelegge, J. P. Integral membrane proteins and bilayer proteomics. *Anal. Chem.*, 85(5):2558–2568, 2013.

Worrall, L. J., Vuckovic, M., and Strynadka, N. C. J. Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA. *Protein Sci.*, 19(5):1091–1096, 2010.

Worrall, L. J., Hong, C., Vuckovic, M., Deng, W., Bergeron, J. R. C., Majewski, D. D., Huang, R. K., Spreter, T., Finlay, B. B., Yu, Z., and Strynadka, N. C. J. Near-atomic-resolution cryo-EM analysis of the Salmonella T3S injectisome basal body. *Nature*, 540:597–601, 2016.

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Hönigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., Richter, L., Ashkenazy, H., Punta, M., Schlessinger, A., Bromberg, Y., Schneider, R., Vriend, G., Sander, C., Ben-Tal, N., and Rost, B. PredictProtein - An open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.*, 42(W1):337–343, 2014.

Zarivach, R., Deng, W., Vuckovic, M., Felise, H. B., Nguyen, H. V., Miller, S. I., Finlay, B. B., and Strynadka, N. C. J. Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. *Nature*, 453(7191):124–127, 2008.

Zilkenat, S., Franz-Wachtel, M., Stierhof, Y.-D., Galán, J. E., Macek, B., and Wagner, S. Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach. *Mol. Cell. Proteomics*, 15(5):1598–1609, 2016.

Zilkenat, S., Grin, I., and Wagner, S. Stoichiometry determination of macromolecular membrane protein complexes. *Biol. Chem.*, 398(2):155–164, 2017.

Zorzi, R. D., Mi, W., Liao, M., and Walz, T. Single-particle electron microscopy in the study of membrane protein structure. *Microscopy*, 65 (1):81–96, 2016.

7. Appendix

7.1 Primary publications

Manuscript I

Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach

This research was originally published in Molecular & Cellular Proteomics. Zilkenat S., Franz-Wachtel M., Stierhof Y.-D., Galán J.E., Macek B., Wagner S. Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach. *Mol Cell Proteomics.* 2016; Vol:15:1598-1609

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

Research

Determination of the Stoichiometry of the Complete Bacterial Type III Secretion Needle Complex Using a Combined Quantitative Proteomic Approach*[S]

Susann Zilkenat‡, Mirita Franz-Wachtel§, York-Dieter Stierhof¶, Jorge E. Galán‖, Boris Macek§, and Samuel Wagner‡‡‡**

Precisely knowing the stoichiometry of their components is critical for investigating structure, assembly, and function of macromolecular machines. This has remained a technical challenge in particular for large, hydrophobic membrane-spanning protein complexes. Here, we determined the stoichiometry of a type III secretion system of *Salmonella enterica* serovar Typhimurium using two complementary protocols of gentle complex purification combined with peptide concatenated standard and synthetic stable isotope-labeled peptide-based mass spectrometry. Bacterial type III secretion systems are cell envelope-spanning effector protein-delivery machines essential for colonization and survival of many Gram-negative pathogens and symbionts. The membrane-embedded core unit of these secretion systems, termed the needle complex, is composed of a base that anchors the machinery to the inner and outer membranes, a hollow filament formed by inner rod and needle subunits that serves as conduit for substrate proteins, and a membrane-embedded export apparatus facilitating substrate translocation. Structural analyses have revealed the stoichiometry of the components of the base, but the stoichiometry of the essential hydrophobic export apparatus components and of the inner rod protein remain unknown. Here, we provide evidence that the export apparatus of type III secretion systems contains five SpaP, one SpaQ, one SpaR, and one SpaS. We confirmed that the previously suggested stoichiometry of nine InvA is valid for assembled needle complexes and describe a loose association of InvA with other needle complex components that may reflect its function. Furthermore, we present evidence that not more than six PrgJ form the inner rod of the needle complex. Providing this structural information will facilitate efforts to obtain an atomic view of type III secretion systems and foster our understanding of the function of these and related flagellar machines. Given that other virulence-associated bacterial secretion systems are similar in their overall buildup and complexity, the presented approach may also enable their stoichiometry elucidation. ***Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.056598, 1598–1609, 2016.***

Type III secretion systems (T3SS), evolutionary and structurally related to bacterial flagella (1), are used by many pathogenic or symbiotic Gram-negative bacteria to inject effector proteins into eukaryotic host cells in order to promote bacterial survival and colonization (2). The core unit of T3SSs is a cell envelope-spanning macromolecular machine termed the needle complex. It consists of a base that anchors the complex in the bacterial inner and outer membranes (3), an inner membrane-embedded export apparatus facilitating substrate translocation located at the center of the base (4), and a filamentous inner rod and needle, which protrude from the bacterial surface and serve as conduit for substrates (2) (Fig. 1). The entire system, which also includes several cytoplasmic components involved in targeting and preparation of substrates (5–9), is composed of up to 20 different proteins with one to several hundred copies each.

Studies using single particle cryo electron microscopy coupled to molecular docking of the atomic structures of domains

From the ‡University of Tübingen, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), Section of Cellular and Molecular Microbiology, Elfriede-Aulhorn-Str. 6, 72076 Tübingen, Germany; §University of Tübingen, Proteome Center Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany; ¶University of Tübingen, Center for Plant Molecular Biology (ZMBP), Auf der Morgenstelle 32, 72076 Tübingen, Germany; ‖Yale University School of Medicine, Department of Microbial Pathogenesis, 295 Congress Ave, New Haven, CT; **German Center for Infection Research (DZIF), Partner-site Tübingen, Elfriede-Aulhorn-Str. 6, 72076 Tübingen, Germany

Received October 27, 2015, and in revised form, February 17, 2016
Published, MCP Papers in Press, February 21, 2016, DOI 10.1074/mcp.M115.056598

Author contributions: S.Z., M.F., J.E.G., B.M., and S.W. designed the research; S.Z., M.F., Y.S., and S.W. performed the research; S.Z., M.F., Y.S., B.M., and S.W. analyzed data; and S.Z., M.F., J.E.G., B.M., and S.W. wrote the paper.

[1] The abbreviations used are: DDM, n-dodecyl-β-D-maltoside; HCD, Higher-energy collisional dissociation; IM, inner membrane; IP, immunoprecipitation; IPTG, isopropyl-β-D-thiogalactopyranoside; MBP, maltose-binding protein; OM, outer membrane; PCS, peptide concatenated standard; SPI-1m *Salmonella* pathogenicity island 1; T3SS, type III secretion system.

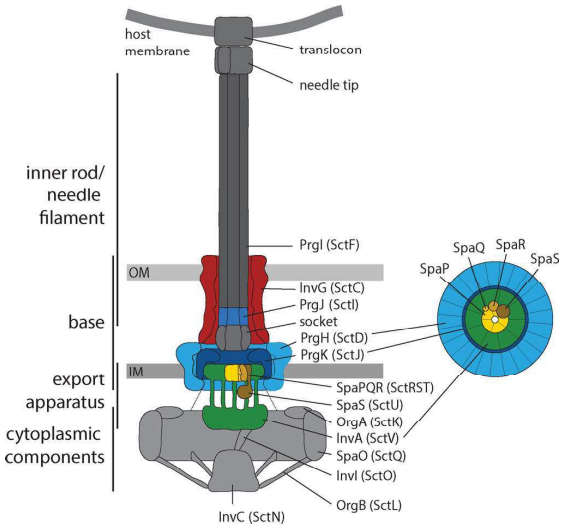


FIG. 1. **Model of the type III secretion system needle complex.** Base and export apparatus components whose stoichiometry was investigated are shown in color. Protein names of T3SS-1 of *S.* Typhimurium are indicated. Protein names according to the unified nomenclature by Hueck (50) are in brackets. On the right, a bottom view of the base and export apparatus is shown indicating stoichiometries obtained in this study. Abbreviations: IM: inner membrane; OM: outer membrane.

of different components have begun to generate an atomic model of the type III secretion needle complex (10). However, many protein densities within the structure are still unaccounted for, leaving substantial knowledge gaps that have yet to be filled. A challenge for this effort has been the absence of the precise stoichiometry of all the components of the needle complex. The determination of an accurate stoichiometry for such a large and hydrophobic protein assembly is technically demanding but severely needed to facilitate approaches generating an atomic view of the machine. The stoichiometry that is known to date resulted from the analyses of the less hydrophobic components that have been amenable to high-resolution x-ray crystallography and NMR and low-resolution cryo electron microscopy. The low-resolution structural analysis of isolated needle complexes revealed a stoichiometry of 12–15 copies for the outer membrane secretin ring of T3SSs of different bacteria and of 12–24 copies for the two inner membrane ring proteins (11–14). The periodicity of the helical needle is 5.7 subunits per turn as assessed by solid-state NMR (15), but the stoichiometry of the inner rod has not been reported. So far, only the sizeable cytoplasmic domains of several homologs of the hydrophobic export apparatus components SpaS and InvA were amenable to high-resolution structural studies (16–24). While no stoichiometry could be deduced from structures of $SpaS_C$, the C-terminal domain of the InvA homolog of *Shigella* (MxiA) crystallized as a nonameric ring (24). Besides SpaS, the stoichiometry of the minor export apparatus components SpaP, SpaQ, and SpaR is also unknown.

To evaluate the stoichiometry of the complete needle complex of the T3SS encoded within pathogenicity island 1 (SPI-1) of *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium), including its hydrophobic export apparatus components (Fig. 1, see also for nomenclature), we have employed two complementing mass spectrometry (MS)-based strategies, using peptide concatenated standards (PCS) (25) and synthetic stable isotope-labeled peptides (26). Both strategies employ ratiometric comparison of isotope-labeled standard peptides of known quantity or stoichiometry with the quantities of the same peptides in assembled complexes (Fig. 2).

These approaches have been used successfully to study the stoichiometry of protein assemblies of low complexity, hydrophobicity, and/or stoichiometric range (*e.g.* (25, 27–29), but the analysis of large, heterogeneous, and hydrophobic membrane protein complexes remains a challenge. Using two optimized complementary needle complex purification protocols and MS analysis, we have been able to reliably deduce the stoichiometry of the T3SS encoded by SPI-1 of *S.* Typhimurium. We reproduced the structure-based stoichiometry of 24:24:15 for PrgH, PrgK, and InvG, which validated our approach, and confirmed the proposed nonameric stoichiometry of InvA for the first time *in situ*. More importantly, we report evidence that the export apparatus of this T3SS contains five SpaP, one SpaQ, one SpaR, and one SpaS. Combined with the predicted transmembrane topologies of these proteins, this stoichiometry suggests that the inner membrane patch of the needle complex base houses 104 transmembrane domains in total, a dense assembly whose function in the secretion process can now be studied in greater detail. Furthermore, we present evidence that only one helical turn composed of six inner rod proteins PrgJ anchors the needle to the base, which may have implications for its suggested function in needle length control and substrate specificity switching (30).

The stoichiometry information provided in this study will facilitate further structural and functional studies of type III secretion and of the related flagellar systems and as such help to develop inhibitors of these machines central to the virulence of many pathogens. By combining gentle membrane protein complex purification and quantitative mass spectrometry, we show that MS-based stoichiometry determination can be extended to investigating highly hydrophobic complexes of a wide stoichiometric range such as multiple other bacterial protein secretion machines.

EXPERIMENTAL PROCEDURES

Materials—Chemicals were from Sigma-Aldrich (St. Louis, MO) unless otherwise specified. Detergent n-dodecyl-β-D-maltoside (DDM) was from Affimetrix-Anatrace (Maumee, OH). SERVA Blue G and SERVAGel™ TG PRiME™ 8–16% precast gels were from Serva (Heidelberg, Germany). NativePAGE Novex Bis-Tris 3–12% gels were from Life Technologies (Carlsbad, CA). Primers were synthetized by

Eurofins (Ebersberg, Germany) and Integrated DNA Technologies (Coralville, IA). Arg10 ($^{13}C_6$, $^{15}N_4$ - arginine) was from Cambridge Isotope Laboratories (Tewksbury, MA) and Lys8 ($^{13}C_6$, $^{15}N_2$ - lysine) was from Silantes (Munich, Germany). Stable isotope-labeled peptides were from Thermo Fisher (Waltham, MA).

Bacterial Strains, Plasmids, and Growth Conditions—Bacterial strains and plasmids used in this study are listed in Table S1. All *S.* Typhimurium strains were derived from strain SL1344 (31). *S.* Typhimurium strains were grown at 37 °C in LB broth supplemented with 0.3 M NaCl with low aeration to enhance expression of genes of SPI-1. pMAL-c5x-PCS 1 and pMAL-c5x-PCS 2 were cloned by Gibson cloning according to published protocols (32) with primers listed in Table S2. For PCS purification, *Escherichia coli* strain AT713 (*e.g. argA21, lysA22*) was grown in defined M9 medium (Table S3) at 27 °C to 37 °C and 200 rpm, supplemented with 0.3 mM IPTG to induce expression of the maltose binding protein (MBP) fusion from plasmid pMAL-c5x. Cultures were supplemented as required with streptomycin (50 μg/ml), tetracycline (12.5 μg/ml), or ampicillin (100 μg/ml).

Secretion Assay—Analysis of type III-dependent secretion of proteins into the culture medium was carried out as described previously (33).

Immunoblotting—For protein detection, samples were subjected to SDS-PAGE using SERVAGel™ TG PRiME™ 8–16% precast gels (Serva), transferred onto a PVDF membrane (Bio-Rad, Hercules, CA), and probed with primary antibodies anti-SipB, anti-InvJ, anti-PrgH, anti$SpaS_N$, or M2 anti-FLAG. Secondary antibodies were goat anti-mouse IgG DyLight 800 conjugate and goat anti-rabbit IgG DyLight 680 conjugate (Thermo Scientific Pierce, Rockford, IL). Detection was performed using the Odyssey imaging system (Li-Cor, Lincoln, NE).

Needle Complex Purification by Immunoprecipitation—Membrane fractionation, solubilization, and immunoprecipitation (IP) were carried out as previously described (4, 34).

Needle Complex Purification by CsCl Gradient Centrifugation—Purification of needle complexes of wild-type and $SpaS_{N258A}^{FLAG}$ bacteria was carried out as published previously (3, 33), but n-dodecyl-N,N-dimethylamine-N-oxide was replaced by DDM (0.7% for lysis/extraction, 0.1% for maintenance) for lysis of cells and extraction of needle complexes throughout the protocol. Furthermore, an initial concentration of 35% (w/v) of CsCl was used to prepare the gradient.

Electron Microscopy—Purified needle complexes were negatively stained with 1% aqueous uranyl acetate on carbon-coated copper grids. Micrographs were recorded using a JEM-1400Plus (JEOL, Tokyo, Japan) microscope at 120 kV.

Blue Native-PAGE—Blue native-PAGE of purified needle complexes was carried out as previously described (4).

Protein In-Gel Digestion—For identification and selection of suitable peptides for the design of peptide-concatenated standards (PCS), the stained protein bands corresponding to the needle complex were excised from blue native-PAGE gels and in-gel digested with trypsin (35). We performed a pilot experiment to evaluate whether two consecutive in-gel digests of the same gel piece would improve the yield of tryptic peptides qualitatively (additional unique peptides) and quantitatively (more evidence of high intensity of the same peptide) (data not shown). This procedure resulted in a quantitative improvement of up to 20% for all peptides identified and sequenced by tandem mass spectrometry; the yield of four critical peptides of needle complex components was even improved by 40% or more (SpaR GATHVLE, InvA AGIIDADAAR, PrgK LYSAIEQR, and SpaQ MDDLVFAGNK). Of the peptides identified in the second digest, up to 38% were not identified in the first digest. Because of these improvements, we decided to routinely use two consecutive digests and extractions. After each step, extracted peptides were desalted using C_{18} StageTips (36). Corresponding eluates were combined and subjected to LC-MS/MS analysis. For PCS analysis, blue native-gel bands containing the purified complex were mixed with the gel bands containing the "heavy" labeled purified PCS construct prior to in gel digestion.

For quantification using stable isotope-labeled peptides, combined eluates were divided into three equal parts. The synthetic stable isotope-labeled peptides QVIFLALAK (SpaQ) and VGVPVIVDIK (SpaS), labeled with Lys8 on the C terminus (Thermo Fisher Scientific), were added to the digested proteins in three different amounts: 0.375, 0.75, and 1.5 pmol. The peptide mixtures were subjected to LC-MS/MS analysis.

Mass Spectrometry—LC-MS/MS analyses were performed either on an EasyLC II nano-HPLC (Proxeon Biosystems) coupled to an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific), or on an EasyLC 1000 nano-UHPLC (Proxeon Biosystems, Odense, Denmark) coupled to a Q Exactive HF mass spectrometer (Thermo Scientific), both as described previously (37, 38). Peptide mixtures were injected onto the column in HPLC solvent A (LTQ Orbitrap Elite: 0.5% acetic acid; Q Exactive HF: 0.1% formic acid) at a flow rate of 500 nl/min and subsequently eluted with either a 116 min (PCS analyses) of 5–33-50–90% of HPLC solvent B (80% acetonitrile in 0.5% acetic acid) or a 57 min gradient (quantification using stable isotope-labeled) of 10–33-50–90% of HPLC solvent B (80% acetonitrile in 0.1% formic acid). During peptide elution, the flow rate was kept constant at 200 nl/min. For Orbitrap Elite analysis, the 20 most intense precursor ions were sequentially fragmented in each scan cycle using collision-induced dissociation, and sequenced precursor masses were excluded from further selection for 60 s. In some cases, an inclusion list containing ions of light- and heavy-labeled PCS peptides was applied. For Q Exactive HF, the seven most intense precursor ions were sequentially fragmented in each scan cycle using HCD fragmentation. For quantification using stable isotope-labeled, full scan MS spectra were acquired in a mass range from *m/z* 480–705. An inclusion list containing ions of light- and heavy-labeled peptides QVIFLALAK and VGVPVIVDIK was applied, and no additional masses were allowed for fragmentation. Full scan resolution was set to 120,000. The target values for the MS scan and MS/MS fragmentation were 3×10^6 and 10^5 charges (Q Exactive HF) or 10^6 and 5×10^3 charges (Orbitrap Elite). The maximal injection time for MS/MS fragmentation was 110 ms and 25 ms, respectively.

Mass Spectrometry Data Processing—The MS data were processed with a setting of 1% for the false discovery rate using MaxQuant software (versions 1.2.2.9 and 1.5.2.8) as described previously (37, 39, 40) with slight modifications. A database search was performed using the Andromeda search engine (40), which is part of MaxQuant. MS/MS spectra were searched against a target database consisting of 10,152 protein entries from *S.* Typhimurium (uniprot 99287, March 1, 2009 and uniprot 216597, July 11, 2012), a database containing the PCS sequences and 248 commonly observed contaminants. In the database search, full tryptic specificity was required and up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as fixed modification, protein N-terminal acetylation and oxidation of methionine were set as variable modifications. Initial precursor mass tolerance was set to 6 ppm (*versus* 1.2.2.9) and 4.5 ppm (*versus* 1.5.2.8), respectively, and at the fragment ion level 0.5 Da was set for collision-induced dissociation fragmentation and 20 ppm for Higher-energy collisional dissociation (HCD) fragmentation. Quantitative MS data acquired using isotope-labeled peptides were processed either with or without restrictions at the false discovery rate. In the latter case, data were filtered manually by setting the posterior error probability (41) threshold to 0.01.

Design of Peptide-Concatenated Standards (PCS)—PCS candidates were chosen from a pool of peptides with a mascot score of >15, preferably without oxidized or miss cleaved forms (data not shown). Two PCSs were assembled containing a total of two to five tryptic peptides of each protein of interest. To retain native digestion

behavior, one to four flanking amino acids were added before and behind each peptide. When the peptide was at the very N-terminal or C-terminal end of the protein of interest or when it was surrounded by arginine or lysine, alanins were added instead of the native sequence. In PCS 2, the native succession of peptides was retained when successive peptides or peptides in close proximity (separated by 10 amino acids or less) were found suitable for the PCS strategy. The synthetic PCS were fused to the C terminus of MBP using pMAL-c5x to enhance expression and solubility, and to facilitate purification.

PCS Purification—PCSs were expressed fused to MBP from a pMAL-c5x plasmid in AT713, grown in M9 medium containing Arg10 ($^{13}C_6$, $^{15}N_4$-arginine) and Lys8 ($^{13}C_6$, $^{15}N_2$-lysine). MBP-PCSs were purified using an amylose resin (New England Biolabs, Ipswich, MA). The eluted MBP-PCS was then run on a 10%/4% SDS-PAGE, which was subsequently stained with colloidal Coomassie. The 90 kDa band corresponding to PCS-MBP was cut from the 4% section of the gel and stored in 5% acetic acid until further use.

Selection of Stable Isotope-Labeled Peptides—Stable isotope-labeled peptides were chosen from the pool of selected PCS peptides. The criteria for selection were: a) no missed cleavage sites, b) no oxidized methionines, c) 6 to 15 amino acids in length, and d) a high intensity for reliable quantification. We selected peptides QVIFLALAK for SpaQ and VGVPVIVDIK for SpaS.

Ratiometric Quantification of Stoichiometry—For relative quantification of the components of the needle complex, we used the light/heavy ratios of the intensities of individual peptide sequencing events, comprising fragmentation, quantification, and identification, which are termed evidence hereafter. For needle complexes purified by IP or CsCl-gradient centrifugation and measured together with the PCSs, the following filter criteria and calculations were applied: a) Evidence measured in the last 10 min of each run were excluded. b) Evidence with a posterior error probability equal or higher than 0.01 were excluded. c) To compare the light/heavy intensity ratios of the evidence between the replicates, the values were transformed into a stoichiometry by normalizing with peptides from proteins of known stoichiometry: the InvG peptide SLLVGGYTR was used for normalization for data obtained with immunoprecipitated needle complexes, the PrgK peptide SDAQLQAPGTPVKR was used for normalization of data obtained with needle complexes purified by CsCl-gradient centrifugation. d) From these combined datasets for IP-PCS experiments and CsCl-PCS experiments, respectively, outliers were identified as two median absolute deviations from the median. For the analysis of the ratio of SpaQ and SpaS in purified needle complexes using stable isotope-labeled peptides, only the evidence with the highest intensity was used to calculate protein ratios, which typically accounted for 90–100% of the peptides total intensity. The resulting raw data identifiers used for calculating the mean stoichiometry of each protein can be found in Table S4.

Experimental Design and Statistical Rationale—Several measures were taken to minimize systematic errors. a) As IP baits, the two last-assembling export apparatus components were selected to ensure complete assembly of purified complexes. b) Two complementary strategies, IP and CsCl gradient centrifugation, were used for purification of T3SS needle complexes to account for potential purification biases. c) In-gel denaturing of proteins of native needle complexes by SDS and separation of MBP-PCS on a 4% SDS-PAGE were undertaken to match the state of the proteins for in-gel trypsinization and avoid biases because of the folding states of the proteins to be compared. d) To account for trypsinization biases between PCS and full-length analytes, four flanking residues were included at each side of a cleavage site for construction of the PCS. e) Analysis of selected proteins (SpaQ and SpaS) was performed using stable isotope-labeled peptides to validate results obtained using PCS.

For quality control of analytical and biological reliability, two independent measures were taken: a) Three proteins of known stoichiometry were included in the analysis: PrgH, PrgK, and InvG. Since stoichiometry data obtained for these proteins were within 10% of the literature data, results obtained for proteins of hitherto unknown stoichiometry can be trusted. b) For PCS analysis, stoichiometries were determined from measurements of up to seven different peptides (Fig. 5). The very hydrophobic membrane proteins SpaQ and SpaR allowed for the inclusion of only 2 peptides, however.

The experimental results presented were obtained with the following sample sizes: Sample sizes for or the central IP-PCS experiment were $n = 5$ for the IP-bait $SpaS_{N258A}^{FLAG}$ and $n = 5$ for the IP-bait $InvA^{FLAG}$, where n denotes biological replicates. Because of the accuracy of the protocol and instrumentation used, no technical replicates were performed. Sample sizes for the validating CsCl-gradient-PCS experiment were $n = 1$ for the IP-bait $SpaS_{N258A}^{FLAG}$ and $n = 1$ for the IP-bait $InvA^{FLAG}$. Sample sizes for the validating IP and CsCl-gradient experiments using stable isotope-labeled peptides were $n = 3$ for the immunoprecipitated analyte and $n = 3$ for the CsCl-gradient purified analyte, where n denotes technical replicates using three different concentrations of stable isotope-labeled peptides.

For the PCS strategy, statistical evaluation of the data was performed using Microsoft Excel (version 14.06112.5000) and RStudio (version 0.98.953). Evidence measured in the last 10 min of each run and evidence with posterior error probability ≥ 0.01 were removed from the dataset. Outliers were identified as two median absolute deviations from the median. The ratio of light to heavy intensities of each evidence pair was normalized to compare biological replicates. The mean of normalized evidence pairs of all peptides of a protein was used to calculate the stoichiometry for each experimental set.

The PCS approach was validated for the selected proteins SpaQ and SpaS by MS-based quantification using stable isotope-labeled peptides. For this quantification, light-to heavy-ratios of the evidence pairs were calculated from the most intense evidence of each peptide, representing 90–100% of the peptide's total intensity. Statistical analysis was performed using Microsoft Excel.

The MS data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD003113.

RESULTS

Needle Complex Isolation and Identification of Suitable Standard Peptides—MS-based stoichiometry determination of complexes using PCS or synthetic stable isotope-labeled peptides requires the prior isolation of intact complexes and the identification of suitable standard peptides (Fig. 2). First, we employed a previously established immunoprecipitation (IP)-based protocol (4, 34) to isolate intact T3SS needle complexes from *S*. Typhimurium. As IP bait, we chose the two last-assembling export apparatus components, SpaS and InvA (brown and green in Fig. 1, respectively). These two proteins require the presence of the other three export apparatus components SpaP, SpaQ, and SpaR for assembly, thus ensuring completeness of the isolated complexes (4, 42). Inner and outer membranes of *S*. Typhimurium harboring C-terminally FLAG-tagged alleles of either autocleavage-deficient SpaS (Asn258Ala) or InvA, which are fully functional for type III-dependent secretion (Fig. 3*A*), were fractionated by sucrose density equilibrium centrifugation. Western blotting analysis of the 13 isolated fractions showed that needle com-

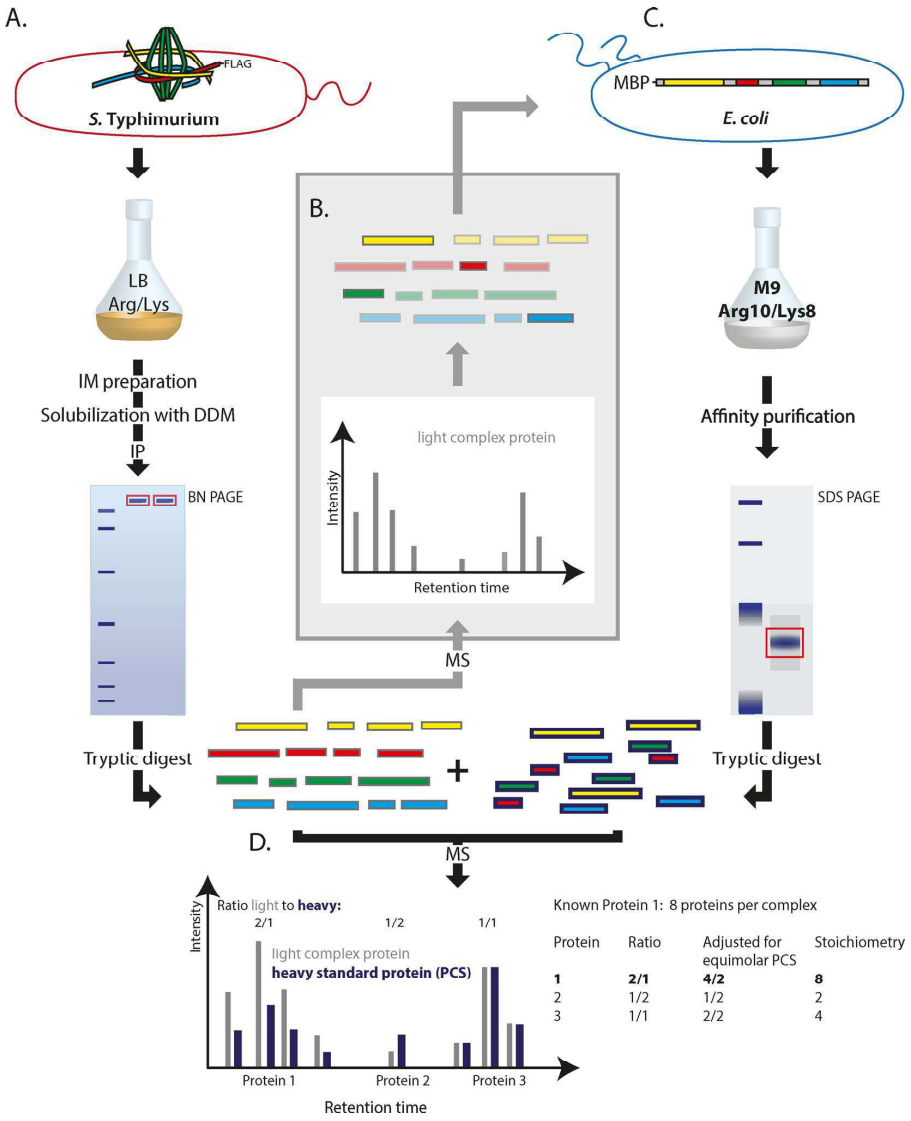


FIG. 2. **Experimental setup of the peptide concatenated standard strategy.** (*A*) *S.* Typhimurium expressing needle complexes with FLAG-tagged bait protein ($SpaS_{N258A}$ or InvA) were grown in complex media. Inner membranes (IM) were purified, solubilized by DDM, and needle complexes were immunoprecipitated. Needle complexes were separated by blue native-PAGE, the corresponding bands were excised, and proteins were subsequently digested with trypsin. (*B*) Peptides were analyzed by MS and suitable peptides were selected for concatenation into the PCS. (*C*) The PCS was expressed as MBP-fusion in Arg and Lys auxotroph *E. coli* grown in defined medium containing heavy arginine and lysine. After purification via MBP, the PCS was run on a 10%/4% SDS-PAGE, mixed with purified needle complex and digested together with trypsin in gel. (*D*) Digested peptides from needle complex (*A*) and PCS (*C*) were mixed and analyzed by MS. The ratio of the evidence of light and heavy peptides was calculated for each protein. Ratios were transformed to absolute stoichiometries by normalization with the evidence-ratios of peptides of proteins of known stoichiometry. Abbreviations: IM: inner membrane.

plexes peak in fractions three to six as judged by the signal intensities of the inner membrane ring component PrgH (Fig. 3*B*). These fractions were pooled, membranes solubilized using the mild nonionic detergent dodecyl-maltoside (DDM), and needle complexes were subsequently immunoprecipitated. Natively eluted needle complexes were evaluated by electron microscopy, SDS-PAGE, and Western blotting. The purified material contained mainly bases lacking

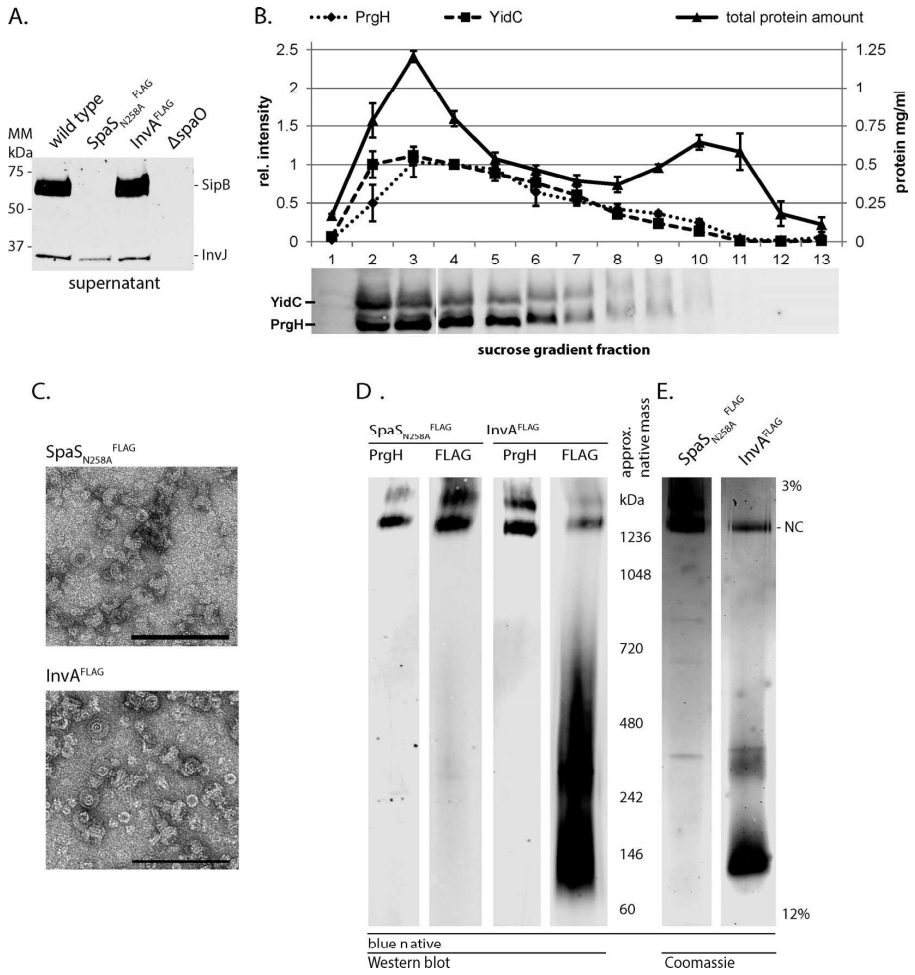


FIG. 3. **Purification of needle complexes by immunoprecipitation and blue native-PAGE.** (*A*) Type III secretion into culture supernatant of wild-type and indicated mutant strains was profiled by immunodetection of the early substrate InvJ and the intermediate substrate SipB. Δ*spaO* denotes a negative control strain defective in type III secretion. (*B*) Membrane fractions containing needle complexes were identified by immunodetection of the external inner membrane ring protein PrgH. The inner membrane protein YidC served as a marker for the inner membrane. Quantification of the protein bands and of total protein content of the fractions was graphed. The average of three independent experiments is shown. Error bars show the standard deviation. (*C*) Electron micrographs of immunoprecipitated needle complexes, scale bar = 100 nm. (*D*, *E*) Blue native-PAGE of immunoprecipitated needle complexes; (*D*) shows immunodetection of the indicated proteins; (*E*) shows a Coomassie-stained gel of immunoprecipitated material.

needles (Fig 3*C*), possibly caused by the harsh mechanical disruption of bacterial cells by French pressing. For further purification and enabling of in-gel digestion, immunoprecipitated needle complexes were separated by blue native-PAGE. Needle complexes, identified by immunodetection of PrgH and SpaS or InvA, run at an apparent native mass of >1200 kDa (Fig. 3*D*) as reported previously (4, 34). Bands corresponding to InvA were also observed at apparent native masses of 150 and 250 kDa, respectively, suggesting that InvA is either expressed in excess or extracted easily from the needle complex (Fig. 3*D*). To identify suitable peptides for stoichiometry determination, the >1200 kDa bands corresponding to needle complexes were cut out (Fig. 3*E*); contained proteins were denatured by SDS, in-gel digested with trypsin and subsequently analyzed by MS.

Construction, Expression, and Purification of the Peptide Concatenated Standard—Suitable peptides (Table S5) were selected for concatenation into two different PCSs (Fig. 4*A*). Oxidation and acetylation-prone peptides were included in the PCSs if too few other suitable peptides were identified for

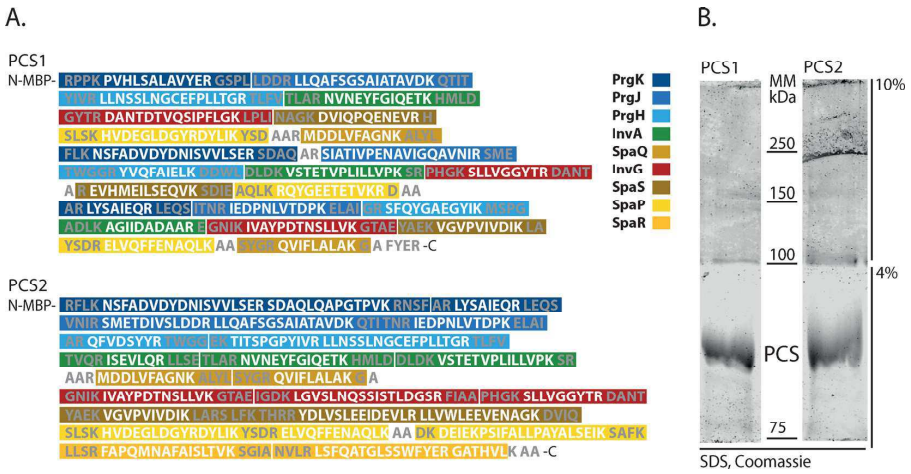


FIG. 4. **Sequence and purification of peptide concatenated standards.** (*A*) Sequence of PCS 1 and PCS 2. White: standard peptides. Gray: flanking residues added to provide natural context for trypsin cleavage. Color-coding for the individual needle complex components as indicated in the legend. (*B*) SDS-PAGE of MBP-PCS1 and MBP-PCS 2 with upper 10% acrylamide gel for separation and lower 4% acrylamide gel for extraction.

the respective proteins and if the degree of the modification was stable across different experiments and comparable between sample and standard. Where appropriate, 1–4 flanking residues of the chosen peptides were included to accommodate possible effects of flanking residues on trypsinization efficiencies (Fig. 4*A*). To enhance the solubility of the synthetic protein, the PCSs were constructed as translational fusions with MBP. MBP-PCSs were expressed in the lysine and arginine auxotrophic *E. coli* strain AT713 in defined media supplemented with heavy arginine and lysine. Heavy-isotope-labeled PCSs facilitated MS-based discrimination of signature peptides of PCSs and needle complexes. MBP-PCSs were purified and subsequently run on a specially designed SDS-PAGE (Fig. 4*B*) composed of a 10% section for separation followed by a 4% section for extraction. This gel composition was used to achieve a similar acrylamide concentration for in-gel digestion of the SDS-PAGE-separated MBP-PCS and of the blue native-PAGE-separated needle complex. MBP-PCS bands were cut out, in-gel digested, and analyzed by MS. The amount of unlabeled PCS peptides was found to be negligible at below 2% of the total intensity of the respective peptides (Table S6).

Stoichiometry Determination of Immunoprecipitated Needle Complexes by PCS Strategy—To determine the stoichiometry of needle complex components, blue native-gel bands containing needle complexes with light arginine and lysine and SDS-gel bands containing MBP-PCSs with heavy Arg10 and Lys8 were codigested and analyzed by MS. The light/heavy intensity ratios of evidence pairs from a total of 10 runs were used for calculation of stoichiometry. Because the amount of complex and PCS varied after each purification, light/heavy ratios were normalized using the robustly detected SLLVGGYTR peptide of InvG, for which a stoichiometry of 15 has been reported (12). Outliers were identified as two median absolute deviation from the median and disregarded for further analysis. We obtained a stoichiometry of the control proteins InvG:PrgH:PrgK of 15:26:27 and 14:28:24 for the $SpaS_{N258A}^{FLAG}$ and $InvA^{FLAG}$ IPs, respectively (Fig. 5). These results are in good agreement with the reported ratio of 15:24:24 (12). The stoichiometry of the minor export apparatus components SpaP:SpaQ:SpaR:SpaS was 5:1:1:1 in both preparations. The major export apparatus protein InvA was largely absent in preparations immunoprecipitated with $SpaS_{N258A}^{FLAG}$ as bait but showed a stoichiometry of seven proteins per needle complex when baiting with $InvA^{FLAG}$. Finally, 2–3 proteins of the inner rod protein PrgJ were detected per needle complex.

Stoichiometry Determination of CsCl Gradient Centrifugation-Purified Needle Complexes by PCS Strategy—The SpaS/InvA IP-based purification of needle complexes was chosen to obtain completely assembled needle complexes in respect to export apparatus components. Unfortunately, this protocol also yielded complexes with broken needles, probably due to bacterial lysis by French pressing. Shearing of needles might also result in a loss of inner rod protein and thus be responsible for the low number of detected PrgJ in the samples. To accommodate for this problem, we chose to also analyze the stoichiometry of needle complexes extracted from bacteria without mechanical disruption but enzymatic (lysozyme) digestion and detergent solubilization, followed by CsCl density equilibrium centrifugation-based needle complex purification. This classical protocol (3) typically results in a severe loss of some export apparatus components due to the use of the stringent detergent n-dodecyl-N,N-dimethylamine-N-oxide

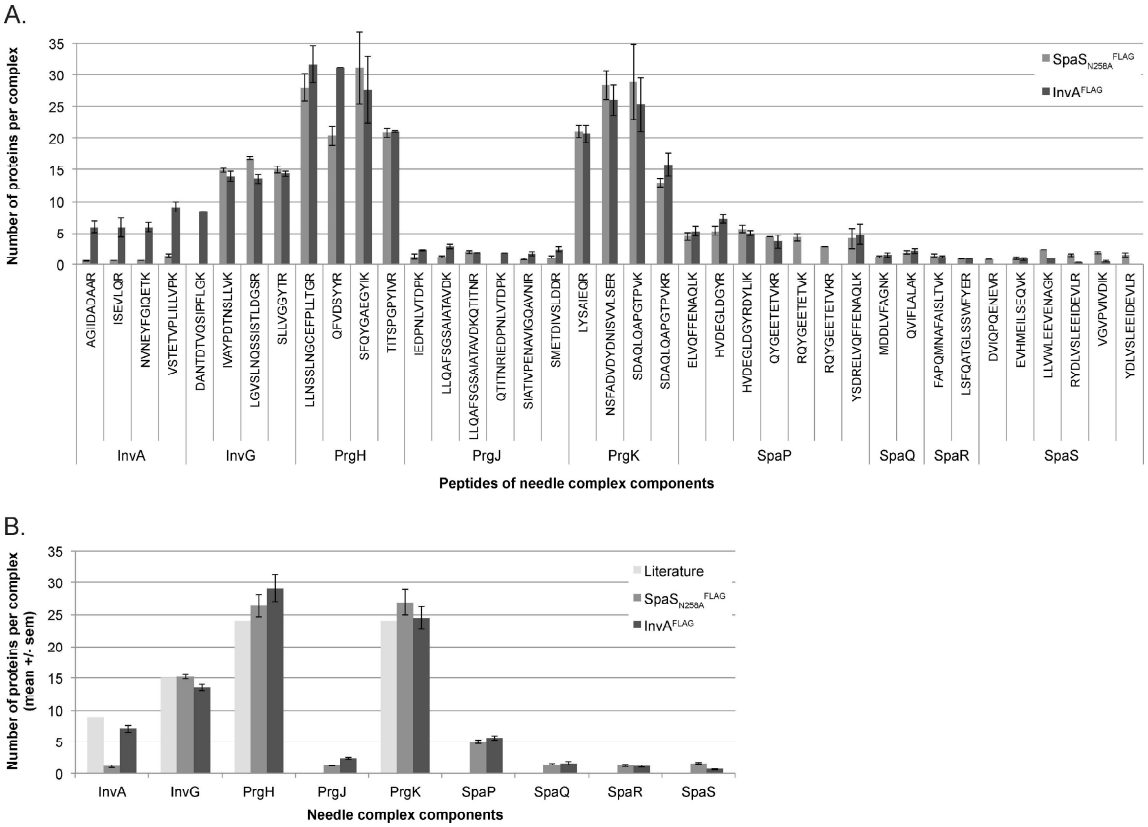


FIG. 5. **Stoichiometry calculations based on normalized stable isotope labeling in cell culture ratios of evidence of peptides from immunoprecipitated needle complexes and peptide concatenated standards.** (*A*) Normalized ratios of light and heavy evidence of the indicated peptides are shown. Normalization using InvG peptide SLLVGGYTR = 15. Error bars: standard error of the mean. (*B*) Literature values of stoichiometry and averages of normalized ratios of light and heavy evidence of the indicated proteins are shown. The literature value for InvA is based on the crystal structure of the C-terminal domain of the isolated *Shigella* homologue MxiA.

(4). To retain the export apparatus components and at the same time keep needles intact, we replaced n-dodecyl-N,N-dimethylamine-N-oxide by the milder detergent DDM. Using this modified protocol, CsCl gradient fractions 4 and 5 showed an enrichment of intact needle complexes and bases as judged by electron microscopy, SDS-PAGE, and Western blotting (Figs. 6*A*-6*C*). Purified needle complexes were separated by blue native-PAGE (Fig. 6*D*) and their stoichiometry analyzed by MS as described above, except that the PrgK peptide SDAQLQAPGTPVKR was used for normalization.

Corroborating the stoichiometry calculations based on immunoprecipitated needle complexes, we obtained average stoichiometries of the control proteins InvG:PrgH:PrgK of 14:22:27 and stoichiometries of the minor export apparatus proteins SpaP:SpaQ:SpaR:SpaS of 5:1:1:1 (Figs. 6*E* and 6*F*). The major export apparatus protein InvA was entirely absent from these preparations, and hence, its stoichiometry could not be calculated. The stoichiometry of PrgJ was three per complex as obtained by analysis of immunoprecipitated needle complexes.

Stoichiometry Determination of SpaQ and SpaS Using Stable Isotope-Labeled Peptides—To validate the obtained results for the export apparatus components SpaQ and SpaS, we chose to quantify the ratio of the two proteins using synthetic stable isotope-labeled peptides. Tryptic digests of needle complexes purified by both methods, IP and CsCl gradients, were spiked with three different concentrations of the standard peptides QVIFLALAK (for SpaQ) and VGVPVIVDIK (for SpaS) at equimolar ratios, and the mixtures were subsequently analyzed by MS. Heavy-to-light ratios were calculated from the most intense evidence pair of each peptide, representing 90–100% of the peptide's total intensity. Based on these results, we obtained average ratios of SpaQ to SpaS of 0.99 and 0.76, respectively, for analysis of needle complexes isolated by the two different methods

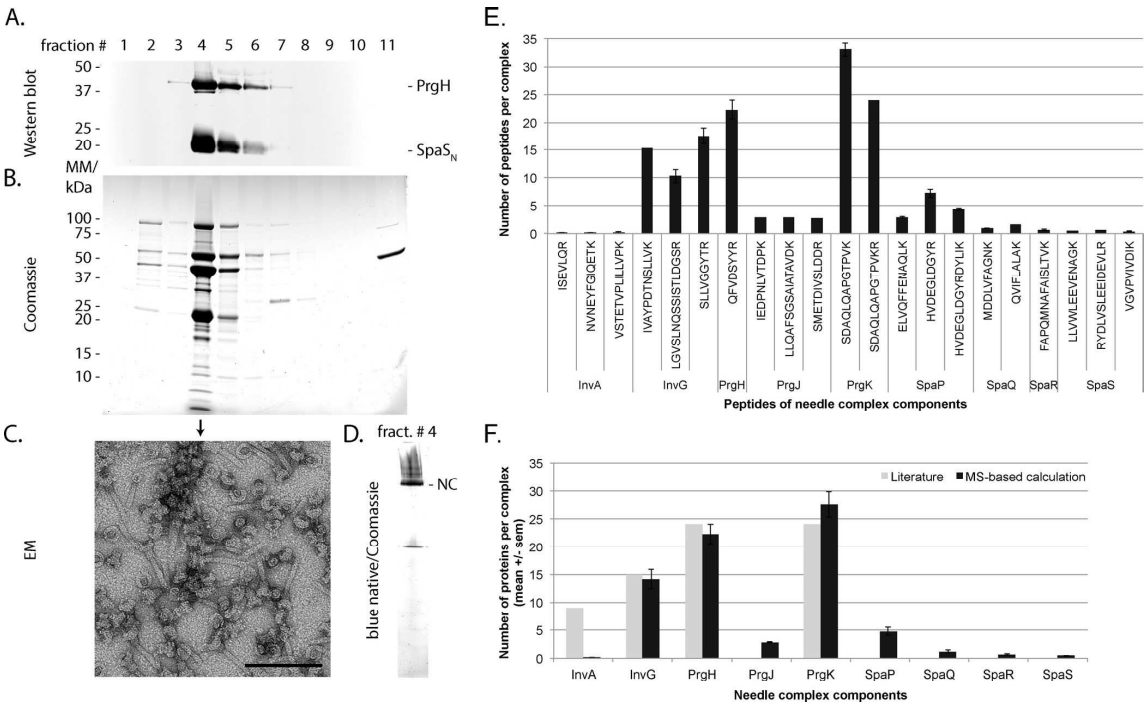


FIG. 6. **Stoichiometry determination of CsCl gradient-purified needle complexes using peptide concatenated standards.** (*A*) CsCl gradient fractions containing needle complexes were identified by immunodetection of the external inner membrane ring protein PrgH and of the switch protein SpaS. (*B*) Coomassie-stained polyacrylamide gels of CsCl fractions. (*C*) Electron micrographs showing bases and needle complexes in fraction 4 of the CsCl gradient, scale bar = 100 nm. (*D*) Coomassie-stained blue native-polyacrylamide gel of purified needle complexes. Abbreviations: EM: electron microscopy. (*E*) Normalized ratios of light and heavy evidence of the indicated peptides are shown. Normalization using PrgK peptide SDAQLQAPGTPVKR = 24. Error bars: standard error of the mean. (*F*) Literature values of stoichiometry and averages of normalized ratios of light and heavy evidence of the indicated proteins are shown. The literature value for InvA is based on the crystal structure of the C-terminal domain of the isolated *Shigella* homologue MxiA.

(Fig. 7). These results corroborate the calculated stoichiometry of 1:1 deduced from PCS-based experiments.

DISCUSSION

Needle complexes of T3SSs are complex macromolecular machines consisting of up to 20 different proteins with one to several hundred copies each. While the stoichiometry of some components had been structurally elucidated, the stoichiometry of the membrane-embedded export apparatus components was largely unknown, primarily because of technical challenges.

We adapted two MS-based relative quantification strategies to evaluate the stoichiometry of complete T3SS needle complexes. The preparation of intact and homogeneous complexes is key to the successful implementation of these approaches (43) but particularly difficult for large membrane protein complexes. We solved this critical problem using a combination of mild detergent extraction, immunoprecipitation or CsCl-gradient-based purification, and blue native-PAGE. The subsequent analysis of these hydrophobic com-

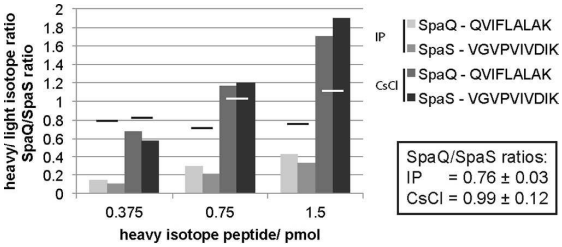


FIG. 7. **Determination of the ratio of SpaQ and SpaS using synthetic stable isotope-labeled peptides.** Trypsin digests of needle complexes purified by IP or CsCl-gradient, respectively, were spiked with stable isotope-labeled peptides at indicated concentrations and analyzed by MS. Columns show the ratio of heavy (stable isotope-labeled peptides) to light (needle complex) evidence for each measurement. Bars show the ratio of SpaQ to SpaS for each pair in a measurement. Average values ± S.D. of SpaQ/SpaS ratios are presented in the box.

plexes by quantitative mass spectrometry was challenged by two major obstacles: a) the detection of only a limited number of peptides that were suitable for quantification of the highly hydrophobic membrane proteins and b) the bias in the acces-

sibility of peptides from intact complexes and standards. Membrane proteins contain only a few charged residues, hence also only a few cleavage sites for trypsin, cutting after arginine or lysine. In addition, the presence of tightly bound lipids or detergents may preclude the efficient cleavage of membrane proteins by site-specific proteases, further reducing the yield of suitable peptides for mass spectrometrical analysis. We were able to increase the yield of four critical peptides of the needle complex by implementing two consecutive tryptic digests and extractions of each gel piece. However, despite this effort, for some proteins, only few peptides were found, and therefore, modification-prone peptides were considered for quantification. We identified four suitable peptides containing oxidation-prone methionines and two peptides prone to N-terminal acetylation. Since the degree of oxidation and acetylation was stable throughout different experiments, we could include these peptides for quantification of the stoichiometry.

Even though suitable tryptic peptides were identified for the proteins of interest, several factors needed consideration that could potentially influence the yield of these peptides and could bias quantification: Trypsinization efficiency depends on the sequence context of the respective peptides (25), bound lipids or detergents (44), and the folding state of proteins (45); and the acrylamide concentration of gels may affect the peptide extraction efficiency. The use of PCSs allows minimization of differences in peptide yield because the peptide context and the preparation of PCS and sample are largely congruent. To align the digestion efficiencies of PCS and sample, we retained the sequence context around the tryptic digestion sites of the peptides and denatured blue native-PAGE-separated complex components by SDS treatment prior to in-gel digestion. Furthermore, we aligned the polyacrylamide concentrations of the PCS-containing gel and of the complex-containing blue native-gel to achieve similar efficiencies of peptide extraction. In contrast, synthetic stable isotope-labeled peptides are not subjected to the same preparative procedures as the complex components of interest. While concentrations of the synthetic peptides are exactly known and invariable, a benefit for absolute quantification, variations in the efficiencies of sample protein digestion and peptide extraction, which are not matched by the synthetic peptides, are likely to skew stoichiometry determination. Therefore, we implemented only two suitable synthetic stable isotope-labeled peptides for the validation of the stoichiometry of the export apparatus components SpaQ and SpaS but refrained from using this approach for the global analysis of the stoichiometry of the complete needle complex.

Using the above-discussed strategy, we confirmed the reported stoichiometry of the base components PrgH, PrgK, and InvG and of the major export apparatus protein InvA. We further report a stoichiometry of 5:1:1:1 for the minor export apparatus components SpaP, SpaQ, and SpaR and for the switch protein SpaS. Our data also suggest a loose association of InvA with the other needle complex components, and we present evidence for a low stoichiometry of the inner rod protein PrgJ in the needle complex.

Cryo electron microscopy structural analysis of needle complexes from the SPI-1 T3SS of *S*. Typhimurium revealed a stoichiometry of PrgH and PrgK of 24 and of InvG of 15 (12). In this study, we yielded MS-based stoichiometries of PrgH, PrgK, and InvG that deviate in average only by 8.4% from the literature values, which imparts confidence for the obtained stoichiometries of the other needle complex components.

Based on a crystal structure, it has been proposed that the isolated *Shigella* InvA homologue MxiA forms a nonamer, but confirmation based on direct measurements of InvA as part of the assembled needle complex was missing. Here, we determined a stoichiometry of seven InvA per assembled needle complex, which is close to the proposed MxiA stoichiometry and to the reported stoichiometry of the flagellar homolog FlhA (24, 46). Interestingly, it was only possible to detect InvA in needle complexes coimmunoprecipitated when baiting with $InvA^{FLAG}$ itself. InvA was almost entirely lost after IP of needle complexes by $SpaS_{N258A}{}^{FLAG}$ or needle complex isolation using mild extraction and CsCl gradient centrifugation, and needle complexes coimmunoprecipitated at a lower efficiency with $InvA^{FLAG}$ as bait than with $SpaS_{N258A}{}^{FLAG}$. The easy extraction of InvA from needle complexes, even by mild detergents, indicates that InvA is only loosely associated with the base and the other export apparatus components, and it might also mean that the herein reported stoichiometry of seven per needle complex is still an underestimation.

Early studies of the flagellar system proposed a stoichiometry of 4–5 of the SpaP homolog FliP and 1–3 of the SpaR homolog FliR based on quantitative Western-blotting and densitometry of autoradiograms (47). The data presented in this study suggest a similar stoichiometry of five SpaP and one SpaR for *Salmonella's* virulence-associated T3SS on SPI-1. We furthermore deduced a stoichiometry of one SpaQ and one SpaS, for which no previous reports exist. *Clostridium* encodes a natural fusion of the flagellar homologs of SpaR and SpaS, FliR and FlhB, respectively. An artificial fusion of these proteins is also functional in the flagellar system of *Salmonella* (48). These observations suggest an equimolar stoichiometry of FliR and FlhB, which, together with our data, support a stoichiometry of one SpaQ, one SpaR, and one SpaS. We could show previously that SpaP and SpaR form the cup substructure at the center of the base of the needle complex (4). Given a stoichiometry of five SpaP, it is conceivable that a circular assembly of SpaP, possibly in conjunction with SpaR, constitutes the substrate translocation pore of the system in the inner membrane.

PrgJ is the inner rod protein anchoring the needle to the base (49). As the term "inner rod" implies, it was originally thought to constitute the substrate conduit inside the base while the needle extends this conduit on the outside. However, the number of PrgJ subunits forming the inner rod is

unclear and so is its length. It is thought that PrgJ assembles into a helix analogous to the needle structure, and crosslinks of up to six consecutive PrgJ were reported (30), suggesting at least one complete turn of PrgJ. Here, we report the detection of three PrgJ proteins per needle complex, independent of the method of needle complex isolation. Different to the other investigated needle complex components, PrgJ itself is a substrate of the T3SS and can only be secreted and assembled into the needle complex after completion of the principal secretion competent machine. It is possible that not all measured complexes were competent for secretion, and hence, we might underestimate the number of PrgJ in needle complexes. We chose two complementing approaches of needle complex isolation to minimize this problem. As mentioned before, IP of needle complexes with InvA as bait ensured that only complete base-export apparatus assemblies were isolated; however, nothing can be said about the state of assembly and the functionality of the associated cytoplasmic components. Since needles were sheared off during the course of this isolation protocol, the presence of needles cannot be taken as an indicator for functional secretion. The latter can, however, be used to estimate the number of fully functional needle complexes in preparations from CsCl-gradients. More than half of all needle complexes purified by CsCl-gradient fractionation contained needles (Fig. 6*C*). If also half of the measured complexes contain inner rods, a stoichiometry of six PrgJ results for these needle complexes, which supports the view that PrgJ forms only one turn of the inner rod helix that anchors the needle to the base. This low stoichiometry of the inner rod has conceptual implications for the timer model of needle length control, which presumes that the duration of PrgJ assembly determines the duration of needle assembly and hence needle length (2, 30, 49).

In summary, we have presented evidence that the export apparatus of bacterial T3SSs is composed of nine InvA, five SpaP, one SpaQ, one SpaR, and one SpaS, which suggests a total of 104 transmembrane domains within the membrane patch of the base of these systems. We furthermore show that InvA is only loosely associated with the other needle complex components and that not more than six PrgJ form the inner rod of the needle complex. This information will facilitate further structural and functional analyses of these and of the related flagellar systems and as such help to develop antiinfective strategies targeting these machines central to the virulence of many pathogens. In addition, the herein presented extension of MS-based stoichiometry determination to investigating highly hydrophobic complexes of very heterogeneous composition and wide stoichiometric range may foster the analysis of other membrane-spanning protein complexes, in particular of different virulence-associated bacterial secretion systems.

Acknowledgments— We thank Silke Wahl for technical assistance in MS sample preparation. Thomas Marlovits is acknowledged for providing antisera against $SpaS_N$. Jan-Willem de Gier is acknowledged for providing antisera against YidC. Lea Krampen, Julia Monjarás Feria, and Mehari Tesfazgi Mebrhatu are thanked for critically reading the manuscript.

* This work was supported by a postdoctoral fellowship of the Human Frontiers Science Program (to S.W.) and by the Alexander von Humboldt Foundation in the framework of the Sofja Kovalevskaja Award endowed by the Federal Ministry of Education and Research (to S.W.).

S This article contains supplemental material.

‡‡ To whom correspondence should be addressed: Tel.: +49 7071 29 84238, Fax: +49 7071 29 5440; E-mail: samuel.wagner@med.uni-tuebingen.de.

REFERENCES

1. Abby, S. S., and Rocha, E. P. C. (2012) The non-flagellar type III secretion system evolved from the bacterial flagellum and diversified into host-cell adapted systems. *PLoS Genet.* **8**, e1002983
2. Galán, J. E., Lara-Tejero, M., Marlovits, T. C., and Wagner, S. (2014) Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. *Annu. Rev. Microbiol.* **68,** 415–438
3. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galán, J. E., and Aizawa, S. I. (1998) Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280,** 602–605
4. Wagner, S., Königsmaier, L., Lara-Tejero, M., Lefebre, M., Marlovits, T. C., and Galán, J. E. (2010) Organization and coordinated assembly of the type III secretion export apparatus. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 17745–17750
5. Akeda, Y., and Galán, J. E. (2005) Chaperone release and unfolding of substrates in type III secretion. *Nature* **437,** 911–915
6. Lara-Tejero, M., Kato, J., Wagner, S., Liu, X., and Galán, J. E. (2011) A sorting platform determines the order of protein secretion in bacterial type III systems. *Science* **331,** 1188–1191
7. Notti, R. Q., Bhattacharya, S., Lilic, M., and Stebbins, C. E. (2015) A common assembly module in injectisome and flagellar type III secretion sorting platforms. *Nat. Commun.* **6,** 7125
8. Diepold, A., Kudryashev, M., Delalez, N. J., Berry, R. M., and Armitage, J. P. (2015) Composition, formation, and regulation of the cytosolic C-ring, a dynamic component of the type III secretion injectisome. *PLoS Biol.* **13**, e1002039
9. Hu, B., Morado, D. R., Margolin, W., Rohde, J. R., Arizmendi, O., Picking, W. L., Picking, W. D., and Liu, J. (2015) Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 1047–1052
10. Burkinshaw, B. J., and Strynadka, N. C. J. (2014) Assembly and structure of the T3SS. *Biochim. Biophys. Acta* **1843,** 1649–1663
11. Hodgkinson, J. L., Horsley, A., Stabat, D., Simon, M., Johnson, S., da Fonseca, P. C. A., Morris, E. P., Wall, J. S., Lea, S. M., and Blocker, A. J. (2009) Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nat. Struct. Mol. Biol.* **16,** 477–485
12. Schraidt, O., and Marlovits, T. C. (2011) Three-dimensional model of *Salmonella's* needle complex at subnanometer resolution. *Science* **331,** 1192–1195
13. Kowal, J., Chami, M., Ringler, P., Müller, S. A., Kudryashev, M., Castaño-Díez, D., Amstutz, M., Cornelis, G. R., Stahlberg, H., and Engel, A. (2013) Structure of the dodecameric *Yersinia enterocolitica* secretin YscC and its trypsin-resistant core. *Structure* **21,** 2151–2161
14. Kudryashev, M., Stenta, M., Schmelz, S., Amstutz, M., Wiesand, U., Castaño-Díez, D., Degiacomi, M. T., Münnich, S., Bleck, C. K., Kowal, J., Diepold, A., Heinz, D. W., Dal Peraro, M., Cornelis, G. R., and Stahlberg, H. (2013) In situ structural analysis of the *Yersinia enterocolitica* injectisome. *Elife* **2,** e00792
15. Loquet, A., Sgourakis, N. G., Gupta, R., Giller, K., Riedel, D., Goosmann, C., Griesinger, C., Kolbe, M., Baker, D., Becker, S., and Lange, A. (2012) Atomic model of the type III secretion system needle. *Nature* **486,** 276–279
16. Zarivach, R., Deng, W., Vuckovic, M., Felise, H. B., Nguyen, H. V., Miller,

S. I., Finlay, B. B., and Strynadka, N. C. J. (2008) Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. *Nature* **453,** 124–127
17. Deane, J. E., Graham, S. C., Mitchell, E. P., Flot, D., Johnson, S., and Lea, S. M. (2008) Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system. *Mol. Microbiol.* **69,** 267–276
18. Wiesand, U., Sorg, I., Amstutz, M., Wagner, S., van den Heuvel, J., Lührs, T., Cornelis, G. R., and Heinz, D. W. (2009) Structure of the type III secretion recognition protein YscU from *Yersinia enterocolitica*. *J. Mol. Biol.* **385,** 854–866
19. Lountos, G. T., Austin, B. P., Nallamsetty, S., and Waugh, D. S. (2009) Atomic resolution structure of the cytoplasmic domain of *Yersinia pestis* YscU, a regulatory switch involved in type III secretion. *Protein Sci* **18,** 467–474
20. Saijo-Hamano, Y., Imada, K., Minamino, T., Kihara, M., Shimada, M., Kitao, A., and Namba, K. (2010) Structure of the cytoplasmic domain of FlhA and implication for flagellar type III protein export. *Mol. Microbiol.* **76,** 260–268
21. Bange, G., Kümmerer, N., Engel, C., Bozkurt, G., Wild, K., and Sinning, I. (2010) FlhA provides the adaptor for coordinated delivery of late flagella building blocks to the type III secretion system. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 11295–11300
22. Moore, S. A., and Jia, Y. (2010) Structure of the cytoplasmic domain of the flagellar secretion apparatus component FlhA from *Helicobacter pylori*. *J. Biol. Chem.* **285,** 21060–21069
23. Worrall, L. J., Vuckovic, M., and Strynadka, N. C. J. (2010) Crystal structure of the C-terminal domain of the *Salmonella* type III secretion system export apparatus protein InvA. *Protein Sci.* **19,** 1091–1096
24. Abrusci, P., Vergara-Irigaray, M., Johnson, S., Beeby, M. D., Hendrixson, D. R., Roversi, P., Friede, M. E., Deane, J. E., Jensen, G. J., Tang, C. M., and Lea, S. M. (2013) Architecture of the major component of the type III secretion system export apparatus. *Nat. Struct. Mol. Biol.* **20,** 99–104
25. Kito, K., Ota, K., Fujita, T., and Ito, T. (2007) A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.* **6,** 792–800
26. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 6940–6945
27. Nanavati, D., Gucek, M., Milne, J. L., Subramaniam, S., and Markey, S. P. (2008) Stoichiometry and absolute quantification of proteins with mass spectrometry using fluorescent and isotope-labeled concatenated peptide standards. *Mol. Cell. Proteomics* **7,** 442–447
28. Olinares, P. D., Kim, J., Davis, J. I., and van Wijk, K. J. (2011) Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in *Arabidopsis*. *Plant Cell* **23,** 2348–2361
29. von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A. L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., Buczak, K., Mosalaganti, S., Hagen, W., Andres-Pons, A., Lemke, E. A., Bork, P., Antonin, W., Glavy, J. S., Bui, K. H., and Beck, M. (2015) In situ structural analysis of the human nuclear pore complex. *Nature* **526,** 140–143
30. Lefebre, M. D., and Galán, J. E. (2014) The inner rod protein controls substrate switching and needle length in a *Salmonella* type III secretion system. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 817–822
31. Hoiseth, S. K., and Stocker, B. A. (1981) Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* **291,** 238–239
32. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–345
33. Monjarás Feria, J. V., Lefebre, M. D., Stierhof, Y.-D., Galán, J. E., and Wagner, S. (2015) Role of autocleavage in the function of a type III secretion specificity switch protein in *Salmonella enterica* serovar Typhimurium. *MBio* **6,** e01459–15
34. Fischer, M., Zilkenat, S., Gerlach, R. G., Wagner, S., and Renard, B. Y. (2014) Pre- and post-processing workflow for affinity purification mass spectrometry data. *J. Proteome Res.* **13,** 2239–2249
35. 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
36. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2,** 1896–1906
37. Carpy, A., Krug, K., Graf, S., Koch, A., Popic, S., Hauf, S., and Macek, B. (2014) Absolute proteome and phosphoproteome dynamics during the cell cycle of *Schizosaccharomyces pombe* (fission yeast). *Mol. Cell. Proteomics* **13,** 1925–1936
38. Kelstrup, C. D., Jersie-Christensen, R. R., Batth, T. S., Arrey, T. N., Kuehn, A., Kellmann, M., and Olsen, J. V. (2014) Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* **13,** 6187–6195
39. 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
40. 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
41. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: Two sides of the same coin. *J. Proteome Res.* **7,** 40–44
42. Diepold, A., and Wagner, S. (2014) Assembly of the bacterial type III secretion machinery. *FEMS Microbiol. Rev.* **38,** 802–822
43. Ori, A., Andrés-Pons, A., and Beck, M. (2014) The use of targeted proteomics to determine the stoichiometry of large macromolecular assemblies. *Methods Cell Biol.* **122,** 117–146
44. Mirza, S. P., Halligan, B. D., Greene, A. S., and Olivier, M. (2007) Improved method for the analysis of membrane proteins by mass spectrometry. *Physiol. Genomics* **30,** 89–94
45. Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* **1382,** 191–206
46. Morimoto, Y. V., Ito, M., Hiraoka, K. D., Che, Y.-S., Bai, F., Kami-Ike, N., Namba, K., and Minamino, T. (2014) Assembly and stoichiometry of FliF and FlhA in *Salmonella* flagellar basal body. *Mol. Microbiol*. **91,** 1214–1226
47. Jones, C. J., Macnab, R. M., Okino, H., and Aizawa, S. (1990) Stoichiometric analysis of the flagellar hook-(basal-body) complex of *Salmonella typhimurium*. *J. Mol. Biol.* **212,** 377–387
48. Van Arnam, J. S., McMurry, J. L., Kihara, M., and Macnab, R. M. (2004) Analysis of an engineered *Salmonella* flagellar fusion protein, FliR-FlhB. *J. Bacteriol.* **186,** 2495–2498
49. Marlovits, T. C., Kubori, T., Lara-Tejero, M., Thomas, D., Unger, V. M., and Galán, J. E. (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. *Nature* **441,** 637–640
50. Hueck, C. J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62,** 379–433

Manuscript II

Topology mapping of the type III secretion export apparatus proteins

Manuscript ready for publication

Transmembrane topology determination of export apparatus components of bacterial type III secretion systems

Susann Zilkenat[1], Samuel Wagner[1,2]

Affiliations:
1 University of Tübingen
Interfaculty Institute of Microbiology and Infection Medicine (IMIT)
Section of Cellular and Molecular Microbiology
Elfriede-Aulhorn-Str. 6
72076 Tübingen
Germany

2 German Center for Infection Research (DZIF)
Partner-site Tübingen
Elfriede-Aulhorn-Str. 6
72076 Tübingen
Germany

Correspondence:
Samuel Wagner
Phone: +49 7071 29 84238
Fax: +49 7071 29 5440
samuel.wagner@med.uni-tuebingen.de

Abbreviations

β-ME: β-mercaptoethanol

AMS: 4-acetamino-4`-maleimidyl

C_{IN}: C-terminus located in the cytoplasm

C_{OUT}: C-terminus located in the periplasm

fT3SS: flagella-associated type three secretion system

MPB: N*-(3-maleimidylpropionyl) biocytin

N_{IN}: N-terminus located in the cytoplasm

N_{OUT}: N-terminus located in the periplasm

SPI-1: *Salmonella* pathogenicity island 1

T3SS: type three secretion system

TCS: transitive consistency scores

TMH: transmembrane helix

vT3SS: virulence-associated type three secretion system

Abstract

Bacterial type IIIa secretion systems (T3SS) facilitate the translocation of substrate proteins in one step across the cell envelope of bacteria. Virulence associated T3SS (vT3SS) of Gram negative bacteria serve the injection of effector proteins into eukaryotic host cells while flagella associated T3SS (fT3SS) serve the assembly of the extracellular parts of the bacterial flagellum. The most conserved part of vT3SS and fT3SS is an inner membrane localized export apparatus consisting of five different components. It is thought to facilitate substrate recognition, specificity switching, and translocation across the bacterial inner membrane. While the stoichiometry of its components has been determined recently, their membrane topology has not been elucidated. We performed an in-depth bioinformatics analysis of the transmembrane topology of the five export apparatus components and assessed the topology of three of the five components of the vT3SS encoded by *Salmonella* pathogenicity island 1, namely InvA, SpaP, and SpaS, by substituted cysteine accessibility method. The topology identification of these proteins alongside with recent stoichiometry and interaction studies is an important step in determining the exact placement of the export apparatus in T3SS and ultimately facilitates elucidation of the function of each component.

Introduction

Virulence-associated bacterial type III secretion systems (vT3SS), whose structural entities are also called injectisomes, are membrane anchored nanomachines used by a variety of pathogenic and symbiotic Gram negative bacteria [1], [2]. They translocate effector proteins from the bacterial cytoplasm into the host cytoplasm, energized by the membrane proton motive force [3], [4]. Related in its export mechanism, flagella-associated type III secretion systems (fT3SS) translocate building blocks of the bacterial flagellum across the bacterial cell envelope [5]. Both vT3SS and fT3SS consist of a basal body that anchors the machines in the bacterial cell envelope [6] and an extracellular filament structure protruding away from the bacterium [7], [8]. A set of cytoplasmic components serves in substrate targeting and five inner membrane proteins compose the actual export apparatus within the basal body (Fig. 1) [9], [10]. The amino acid sequence of the components of the export apparatus of vT3SS and fT3SS is highly conserved [2], and high resolution crystal structures of their soluble domains as well as low resolution electron tomograms of injectisomes and flagella support the close relation of these two systems [6], [11]–[17]

The components of the T3SS export apparatus are thought to facilitate substrate recognition, specificity switching, and translocation across the bacterial inner membrane. However, the detailed functions of the five components have only been worked out insufficiently, not lastly because of a lack of structural information of their transmembrane regions [11], [18], [19]. In the T3SS encoded on *Salmonella* pathogenicity island 1 (SPI-1), which is the subject of this study (see Fig. 1 for reference to the unified nomenclature of injectisome components as proposed by Hueck [2]), the major export apparatus protein is called InvA and is robustly predicted to contain eight transmembrane helices (TMH), which was partly experimentally validated for the InvA homologs YscV (*Yersinia*) and HrcV (*Xanthomonas*) using a PhoA-fusion approach [20], [21]. The C-terminal domain of the *Shigella* InvA homolog MxiA was

shown to form a nonameric ring located between the membrane and the ATPase, one of the cytosolic components of the T3SS [22]. InvA and its homologs have been implicated to facilitate sorting of substrates into the pore of the export apparatus [23] and to couple the type III secretion to the proton gradient [24]

The minor export apparatus proteins SpaP, SpaQ, and SpaR mainly consist of hydrophobic TMHs without sizable extramembrane domains. SpaP was shown to form the cup structure at the center of the base of injectisomes [19]. At this position, a SpaP pentamer forms the central substrate translocation channel of the T3SS in the bacterial inner membrane and connects directly to the filament structures in the periplasm [25]. SpaP is predicted to contain four TMHs but the prediction of the orientation of these domains varies between different homologs. However, the presence of a cleavable signal sequence in most flagellar homologs and the observation of crosslinks from the N-terminus of SpaP to the inner rod protein PrgJ suggest an N-out/C-out topology. This topology is also supported by a PhoA-fusion analysis of the *Xanthomonas* SpaP homolog HrcR [20] but contradicted by PhoA-fusion result of SpaP homolog YscR in *Yersinia* [26]. A high resolution structure of a small hydrophilic domain between TMH 2 and 3 of the flagellar SpaP/SctR homolog FliP has been solved but the functional implication of this structure and the arrangement of FliP in the homooligomer remain unclear.

SpaQ is a small hydrophobic protein of only 86 amino acids. One copy of the protein seems to integrate the export apparatus components SpaP, SpaR, and SpaS as suggested by *in vivo* photocrosslinking analysis [25]; an additional role in the secretion process is not known. SpaQ and its homologs are robustly predicted to contain two TMHs, however, the aforementioned *Xanthomonas* study came to the conclusion that the SpaQ homolog HrcS contains only one TMH at its N-terminus with an N-in/C-out topology [8]. The prediction of the orientation of SpaQ in the membrane yields ambiguous results.

SpaR associates with SpaP in a very stable $SpaP_5R_1$ complex [25] but the function of SpaR has not been defined yet. The prediction of the transmembrane topology of SpaR and its homologs leads to very ambiguous results ranging from 5 to 8 TMH with an uncertain orientation [20], [25]. Experimental evidence based on a PhoA-fusion approach even suggested that the SpaR homolog of *Xanthomonas*, HrcT, contains only one TMH at its N-terminus [20]. In contrast, *in vivo* photocrosslinking showed that SpaR interacts at residues F20, N44, and A45 with the inner rod protein PrgJ [25]. These interactions suggest that the N-terminus of SpaR is located in the periplasm, even though these residues are in general predicted to be within TMHs 1 and 2 of the protein.

SpaS is also called the switch protein of the export apparatus. It has been proposed to be involved in needle length control and switching of specificity from the secretion of early to intermediate and late substrates [18]. SpaS and its homologs are composed of an N-terminal transmembrane domain with four robustly predicted and experimentally verified TMHs of N-in topology and a C-terminal cytoplasmic domain [20], [27]. Self-cleavage at a highly conserved NPTH motif in the cytoplasmic domain is critical for substrate specificity switching but dispensable for needle length control [13], [18], [28]. It has been suggested that switching and late substrate secretion is implemented by the dissociation of the cleaved C-terminus from the remainder of the protein [29].

To foster our understanding of the functional mechanism of T3SSs, a detailed picture of the structure of its export apparatus components is indispensable. Unfortunately, determination of high resolution structures of membrane embedded parts of macromolecular machines remains a challenging task, even though tremendous progress has been made in this field, in particular regarding cryo electron microscopy of single particles [30], [31]. Genetic and biochemical approaches characterizing the stoichiometry and topology of membrane protein components of these machines help to close our knowledge gap and facilitate functional analyses [30].

The experimental analysis of the transmembrane topology is often conducted by inserting tags into positions of interest. However, the results of these studies can be misleading, as many approaches, e.g. translational reporter fusions like LacZ, PhoA or split-GFP [32]–[34], introduce large changes and truncations to the protein of interest that can affect their transmembrane topology [35]. Even changes as small as single positively charged residues were shown to be able to cause a shift from N-in to N-out orientation and back [36]. Therefore, to map the transmembrane topology of the T3SS export apparatus components, we employed a methodology that has proven to be very subtle using maleimide labelling of single cysteine mutations at various positions of the proteins of interest (SCAM^{TM} for single cysteine accessibility method for transmembrane analysis [35]). In contrast to most PhoA, LacZ, or GFP-fusion approaches the function of the protein of interest is usually retained when using SCAM^{TM} and can be assessed by complementation assays [37]. We have thus analysed the topology of the five export apparatus components of the vT3SS encoded by *Salmonella* Typhimurium SPI-1 by in-depth bioinformatics complemented with SCAM. The topology identification of these proteins alongside with recent stoichiometry and interaction studies is an important step in determining the exact placement of the export apparatus in type III secretion systems and ultimately facilitates elucidation of the function of each component. The conserved sequence and function of the five export apparatus components suggests that they share a common transmembrane topology in different injectisome and flagellar T3SS and thus this study will help to also elucidate the function of these molecular machines.

Material and Methods

Materials.

Chemicals were bought from Sigma-Aldrich unless otherwise specified. SERVAGel™ TG PRiME™ 8-16 % and 12 % precast gels were obtained from Serva. Primers were synthetized

by Eurofins and Integrated DNA Technologies. N*-(3-Maleimidylpropionyl) Biocytin (MPB) and 4-Acetamino-4`-Maleimidyl (AMS) were bought from by Invitrogen.

Bacterial strains, plasmids, and growth conditions.

Bacterial strains and plasmids used in this study are listed in table S2. All *S.* Typhimurium strains were derived from strain SL1344 [38]. *Salmonella* strains were grown in LB Lennox medium containing 0.3 M NaCl at 37 °C and 180 rpm to induce expression of the T3SS. Plasmids based on pTACO10 were induced with 100 µM rhamnose. Strains with arabinose inducible p^{araBAD} promoter [39] were induced with 0.02 % arabinose. Main cultures for labelling were inoculated 1:50 from overnight cultures and grown to an OD_{600} of 0.5 to 0.8.

Cysteine free mutants and single cysteine mutants.

Primers for construction of cysteine free mutants and single cysteine mutants are listed in table S2. Native cysteines were replaces by a serine if they were predicted to be outside the membrane and by an alanine if they were predicted to be inside the transmembrane region. Single cysteine positions were chosen using the prediction of TOPCONS [40], TMHMMfix [41], ΔG_{app} [42] and PredictProtein (PHDhtm, PROFsec, SNAP2) [43]. Alignments were performed using T-Coffee (version 11.00) [44]. Signal peptides were predicted using SignalP (version 4.1) [45]. Single cysteines were introduced preferably at positions in predicted loops or at the termini which were predicted as exposed and with low conservation. Serine and other polar, uncharged amino acids were replaced preferentially (Table S2).

Secretion assay

Analysis of type III-dependent secretion of proteins into the culture medium was carried out as described previously [18].

SCAM adjustments and controls

As SCAM is a very sensitive method, several precautions and controls are necessary to prevent false positive/false negative results:

To ensure that only one side of the membrane is labelled at a time experimental conditions must be chosen so that the blocking and the labelling reagent can pass through the outer membrane. At the same time the integrity of the inner membrane should not be compromised. Treatment of the cells with a mixture of EDTA, lysozyme and DNase I was sufficient to allow both the blocking and labelling reagents to reach the complex in the SCAM assay. It is possible that partial cell lysis during labelling or blocking step leads to similar results between the cytopasm and periplasm labelling experiment. Results indicating this have been disregarded.

It is equally important that the introduced cysteine is accessible in context of the protein and complex. To ensure this, if possible, amino acids predicted as exposed (using PredictProtein PROFsec and SNAP2 [43]) were replaced with cysteine. To further improve accessibility, the SCAM experiments were conducted in a Δ*prgHIJK* (Δ*sctDFIJ*) strain, which was previously shown [19], [25] to assemble an intact export apparatus, but lack the periplasmic structures that would shield it from labelling reagents. If possible, two or more amino acid positions per terminus and loop have been tested to recognize this form of mislabelling. Cysteine positions which were neither labelled in the periplasm nor in the cytoplasm may be located inside of a transmembrane helix, hidden by tertiary structure of the protein or by other components of the complex.

To reduce loss of sample after blocking of periplasmic cysteines, the blocking reagent was not removed through washing steps. Instead the labelling reagent was added directly during lysis of the cells, so that labelling of cytoplasm was always a competition between AMS and MPB.

However, even though cytoplasmic labelling time was always shorter than periplasmic, the cytoplasmic signal was overall stronger. This could be due to better accessibility in the cytoplasm, or due to a limited permeability of the outer membrane when labelling the periplasm. It shows that the AMS/MPB competition during the cytoplasmic labelling has no negative effect on our screening.

Labelling and blocking properties of the reagents was tested using an InvA cysteine mutant A511C with known position in the cytoplasm due to analyses of the C-terminal extra membrane domain of this protein [22].

A cysteine free negative control was run alongside each experiment. If equal labelling was visible for both periplasmic and cytoplasmic labelled experiments of one amino acid position, the results were excluded under the assumption that the membrane was damaged during the first steps of the experiment. The signal strength of the labelling results was normalized using the FLAG-IP signal and compared relative to the respective lowest and highest signal of each experiment.
With these measures taken, SCAM offers us a precise and reliable tool to map membrane topology.

Maleimide blocking and labelling.

Labelling protocol was modified after Bogdanov 2005 [35]. Plasmids containing the 'single cysteine' mutations of the genes of interest were transformed into a Δ*prgHIJK* strain for labelling. Cells of 50 ml culture were harvested and divided equally for labelling of periplasmic and cytoplasmic cysteines. Buffer A (100 mM HEPES, 250 mM sucrose, 25 mM $MgCl_2$, 0.1 mM KCl, pH 7.5 with KOH) was optimized for each protein of interest with 0, 10 or 20 µg/ml lysozyme, 10 µg/ml DNase I, 1:100 Protease inhibitor mix (Sigma, P8849) and 1-

15 mM EDTA. Samples were incubated for 30 to 60 minutes in 750 µl of optimized buffer A before the start of the labelling steps. Treatment time with MPB was optimized for each membrane side in each protein of interest and lay between 1 to 90 minutes at 4 °C. Samples for periplasmic labelling were first treated with 100 µM MPB to label exposed cysteines. The reaction was quenched with 20 mM β-mercaptoethanol (β-ME) and samples were lysed with SpeedMill plus Bacteria protocol (analyticjena). Samples for cytoplasmic labelling were first treated with 100 µM AMS to block exposed cysteines. During the bead mill step 100 µM MPB was added to the samples, without quenching the AMS first. After the cell lysis MPB was quenched with 20 mM β-ME.

Immunoprecipitation.

Labelled samples were separated from the glass beads by centrifugation at 500 x *g* for 1 min at 4 °C and from cell debris at 10000 x *g* for 2 min at 4 °C. Membranes were collected at 300000 x *g* for 35 minutes at 4 °C. Membrane pellets were solubilized in 100 µl Buffer A containing 20 mM β-ME by shaking at 1400 rpm for 1h at 4 °C, followed by the addition of 100 µl of solubilisation (50 mM Tris-HCl, 1 mM EDTA, 2% (w/v) SDS) and shaking at 1400 rpm for 30 minutes at 4 °C, 30 minutes at 37 °C and 30 minutes at 4 °C again. The samples were diluted with 300 µl IP1 buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 2% Thesit, 0.4% SDS, pH 8.1) and after spinning down insolubilized remains by centrifugation at 10000 *x g* for 10 min at 4 °C the supernatant was transferred to a new tube. Anti-Flag M2 Magnetic Beads were washed once with IP1 buffer before 2.5 µl beads were added to each sample. Samples were incubated while shaking for 2 h at 4 °C. After three 15 min washing (2x IP1, 1x IP2 [50 mM Tris-HCl, 1 M NaCl, 1 mM EDTA, 2 % Thesit, 0.4 % SDS, pH 8.1]) steps at 4°C, proteins were eluted from the beads with 100 mM Glycin at pH 2.2 for 15 min. Eluates were adjusted to a neutral pH with 2 µl tris-HCl, pH 9.

Immunoblotting

For detection of the proteins SpaPRS and InvA, samples were subjected to SDS-PAGE using 12% gels cast with Hoefer system. Samples were transferred to a PVDF membrane (Bio-Rad), and probed with primary antibodies anti-SipB and anti-InvJ or M2 anti-FLAG, secondary antibodies were goat anti-rabbit IgG DyLight 680 (Thermo Pierce) or goat anti-mouse lgG DyLight 680 (Thermo Pierce). Biotin labelled samples were additionally stained with Streptavidin DyLight 800 (Thermo Pierce). Detection and quantification was performed using the Odyssey imaging system and Image Studio, version 3.0 (Li-Cor).

Evaluation of maleimide signals

To adjust for different expression levels or loss of proteins during the preparation, maleimide signals after immunoblotting were normalized using the FLAG signal of the corresponding sample. Signal strength is given as percentage whereby the lowest signal of a given experiment equals 0 % and the highest 100 %.

Visualisation of transmembrane protein topology

The web tool Protter (version 1.0) was used for visualization of the experimentally determined topologies in this study [46].

Results

Determination of the transmembrane topology is an important step towards elucidation of the structure and function of membrane proteins. It facilitates definition of the extent of membrane embedded regions and extramembrane domains and thus provides a framework for analysis of intra- and intermolecular protein-protein interactions.

Membrane protein topology prediction

To guide and complement the experimental analysis by SCAM^{TM}, we performed an in-depth prediction of the transmembrane topology via bioinformatics. To estimate the positions and orientations of the TMHs, *S.* Typhimurium SPI-1 export apparatus proteins were analysed using the web tools TOPCONS [40], TMHMMfix [41], ΔG_{app} [42] and PredictProtein PHDhtm [43]. These tools, combine predictions based on sequence features, homologs sequence features, hydrophobicity, comparison with known structures, prediction of the corresponding apparent free energy difference, and evolutionary profiles. TOPCONS depicts a consensus from five different prediction algorithms: OCTOPUS, Philius, PolyPhobius, SCAMPI (multiple sequence mode) and SPOCTOPUS. The resulting nine different predictions are illustrated in figure 2 (as well as supplement data S10 and table ST2). As a summary, sequences predicted as transmembrane helices in all predictions are shown in dark grey, sequences predicted as transmembrane helices at least once are shown in light grey. The predicted number and position of TMHs of the proteins SpaS and SpaQ was very consistent. SpaS was robustly predicted with four TMHs, and N_{IN}/C_{IN} orientation. SpaQ was predicted with two TMHs, however, the orientation predictions were ambiguous. The major export apparatus protein InvA was predicted with seven to eight TMHs, a reliable N_{IN}/C_{IN} orientation and a large cytoplasmic loop. The start and end positions of the TMHs of SpaS, SpaQ and InvA shifted only marginally between predictions by 12 amino acids of less, with the exception of the last two TMH of InvA, which were predicted as one TMH by PredictProtein. The predictions of SpaP varied strongly with a N_{OUT}/C_{OUT} or N_{OUT}/C_{IN} orientation, three to six TMHs and a high variation of start and end positions of the helices. A large loop between amino acids 75 and 141 was consistently predicted, however, its orientation was not clear. SpaR was predicted with five to six TMH, mainly with an N_{OUT}/C_{OUT} orientation (only SCAMPI N_{IN}/C_{IN}).

The less robustly predicted minor export apparatus proteins SpaP, SpaQ and SpaR were compared with 9 homologs from the vT3SSs and fT3SSs each, using ΔG_{app} [42], T-Coffee

multiple sequence alignment [44] and review of charged residues (Fig. 3) [47]. The transitive consistency scores (TCS, 0-100, score lower than 50 should be considered poor) [48] of SpaP and the SctR homologs (SpaP and FliP of *Salmonella* Typhimurium, HrcR of *Xanthomonas campestris*, YscR of *Yersinia enterocolitica*, FliP and Spa24 of *Shigella flexneri*, HrcR of *Pseudomonas syringae*, EpaP and FliP of *Escherichia coli*, and FliP of *Helicobacter pylori*. Corresponding homologs from these fT3SSs and vT3SSs were also used for analysis of SctS/SpaQ and SctT/SpaR) were between 91 and 93 (supplememnt S1). The SctR homologs were predicted with four TMHs (disregarding the possible signal peptide sequences in cases of FliP, supplement S11), between 18 and 40 amino acids in length with ΔG ranging from 2.5 to -3.6 kcal/mol (supplement S2). The positively charged amino acids arginine and lysine have a much larger impact on membrane protein topology than the only weakly positive charged histidine or the negative charged aspartate and glutamate. Histidine only becomes relevant at lower that physiological pH, aspartate and glutamate affect topology only in sufficiently high numbers (cluster of ≥ 4) [47], [49]. The SctR homologs contain a conserved positively charged amino acid in the first predicted loop, indicating that this loop is cytoplasmic. This, together with few arginine and lysine at the C- and N-termini and only occasional aspartate or glutamate, suggests an N_{OUT}/C_{OUT} orientation. The predicted extramembrane region between amino acids 75 and 141 in SpaP contains a high number of positively charged amino acids (5 to 12), but also a sufficiently high number of negative charges (8 to 19) to affect the topology, especially in thecase of SpaP where we see an EDED cluster directly adjoining the predicted second TMH. Assuming four TMHs, this loop should be on the same membrane side as the termini (supplement S3).

The TCSs of SpaQ and the other SctS homologs were between 91 and 94 (supplement S4). The smallest protein of the export apparatus was predicted with two TMH each, of 26 to 30 and 23 to 28 amino acids, and with a ΔG of -1 to -3 kcal/mol and 1 to -2 kcal/mol respectively in the ΔG_{app} (supplement S5). 90 % of the homologs had at least one positively charged amino

acid in the N-terminal extra membrane region or within the first three amino acids of the adjacent predicted TMH. The same was the case for the C-terminus, while no positive charged amino acids were in or near the predicted loop (supplement S6). This suggests according to the positive-inside rule an N_{IN}/C_{IN} orientation for SpaQ.

The TCSs of SpaR and the SctT homologs were between 83 and 88 (supplement S7). The SctT homologs were predicted with five to six TMH between 18 and 36 amino acids in length and with ΔG ranging from 2 to -4 kcal/mol (supplement S8). The SctT homologs contained comparatively few positively charged amino acids. A conserved positively charged amino acid in the last predicted loop suggests a cytoplasmic orientation of this loop, indicating a C_{OUT} orientation for the whole protein (supplement S9). However the overall topology of the membrane protein cannot be concluded from these results.

Experimental analysis of the transmembrane topology using SCAM™

To map the transmembrane topology of the *Salmonella* T3SS we used the substituted cysteine accessibility method (SCAM) [35] to gain a better understanding of the structure of the export apparatus. SCAM is based on labelling the thiol-group of a cysteine introduced at specific position in the protein with maleimide reagents. All native cysteines are replaced with serine or alanine before single cysteines are introduced at positions of interest. To differentiate between positions located in the periplasm and in the cytoplasm, two reagents are used in two parallel approaches (Fig. 4). First, any cysteines in the periplasm are labelled with a biotin-maleimide reagent. The reaction is quenched by adding a surplus of another thiol-group containing reagent to prevent labelling of cytosolic cysteines during the following steps. In parallel to this, to exclusively label cysteines in the cytoplasm, periplasmic cysteines are first blocked with a maleimide reagent not linked to biotin. In this case the labelling reagent is added during the lysis of the cells. Labelling of both approaches is visualized via Western blot.

Since immunoprecipitation and detection is a prerequisite for analysis by SCAM^{TM}, previously described 3xFLAG epitope-tagged versions of the four export apparatus components InvA, SpaP, SpaR, and SpaS were used throughout this study [19]. For SpaS, the detection by a C-terminal FLAG tag necessitated the use of an autocleavage-deficient N258A mutant [19]. Unfortunately, attempts to also tag SpaQ only led to non-functional or undetectable protein, thus the experimental analysis of the transmembrane topology of SpaQ could not be performed.

Cysteine free mutants of the remaining four export apparatus proteins were created that all retained their secretion function. This was judged by Western blot analysis of the *Salmonella* SPI-1 vT3SS substrates InvJ and SipB in culture supernatants of plasmid-complemented bacteria (Fig. 5). Subsequently, single cysteines were introduced at positions predicted to be at the termini or within the cytoplasmic and periplasmic loops of the proteins.

While secretion was reduced for some mutants, most cysteine mutants could secrete both substrate proteins (Fig. 5). InvA cysteine mutants could secrete SipB and InvJ with nearly no difference between wild type, a FLAG-tagged InvA and the cysteine free mutant. S228C alone showed no functional secretion. Secretion of both proteins was low for cysteine mutants D59C, S272C and G296C, only InvJ was secreted in D38C mutant, which was due to low expression of the protein. Otherwise there was no difference in the expression level of InvA mutants discussed in this study. SpaP cysteine mutants could also secret SipB and InvJ, but with a significant decrease of secretion between wild type and FLAG-tagged SpaP. No secretion could be detected for cysteine mutants G2C and S31C. No difference in SpaP expression was observed except for a low expression of S31C. Cysteine free SpaS mutant, S17C, T205C and S231C secreted SipB and InvJ. Secretion of cysteine mutants was greatly reduced compared with wild type and FLAG-tagged SpaS. Very weak secretion of InvJ was observed for cysteine mutants G53C, S115C and G166C. There was no difference in SpaS

expression between wild type and mutant strains. In contrast to the other tested export apparatus proteins, secretion functionality of SpaR was already strongly reduced by the addition of the 3xFLAG epitope-tag. As experimental mapping of this protein via SCAM did not yield clear results, function of single cysteine mutants is not shown.

Eleven of sixteen amino acid position the major export apparatus protein InvA (SctV) could be localized by unambiguous labelling of both the cytoplasmic and periplasmic side of the inner membrane in at least duplicate (Fig. 6 a): S4C, A166C, S174C, S263C, S272C, S337C and A511C in the cytoplasm and Q104C, S225C, G296C and F297C in the periplasm. The results of the remaining five positions offered an indication of the localization. Cysteine at position 133 proved to be inaccessible, while cysteines at position S143C (cytoplasmic) and S228C (periplasmic) could only be labelled once. The cysteine at position D38C was localized in the periplasm in two out of five experiments, but proved to be inaccessible in the rest. Position D59C was localized in the cytoplasm in two out of four experiments, but inaccessible and in the periplasm in the remaining two. These results confirm a N_{IN}/C_{IN} orientation with eight TMHs (Fig. 7).

Of the six amino acid positions of the minor export apparatus protein SpaP (SctR), only three could be unambiguous labelled on both the cytoplasmic and periplasmic side of the inner membrane in at least duplicate (Fig. 6 b): S31C and S189C in the cytoplasm and T224C in the periplasm. The results of the remaining positions were ambiguous due to cytoplasmic labelling of the negative control in many experiments, which rendered the cytoplasmic results unusable. Looking at the periplasmic results alone offers an indication of the localization in the periplasm for the positions G2C, S89C and S107C, while the position S157C is likely cytoplasmic or inaccessible (Fig. 6 c). These results point to a N_{OUT}/C_{OUT} orientation together with, contrary to the majority of predictions, a large periplasmic loop (Fig. 8).

For the switch protein SpaS, seven of ten amino acid positions were localized by unambiguous labelling of both the cytoplasmic and periplasmic side of the inner membrane in duplicate (Fig. 6 d): S23C, S115C, T131C and S231C cytoplamic and G53C, S66C and S162C periplasmic. Position S17C and T205C (cytoplasmic), G166C (periplasmic) and could only be localized once, respectively. This confirms the predicted N_{IN}/C_{IN} orientation prediction with four TMHs (Fig. 9).

Experimental topology mapping of SpaR (SctT) did not lead to any conclusive results, due to cytoplasmic labelling of the negative control in many experiments similar to SpaP as well as labelling which indicates partial cell lysis during the labelling or blocking steps.

Discussion

The export apparatus of T3SSs is the central entity of these systems facilitating substrate translocation across the bacterial inner membrane. While high-resolution structures of the larger extramembrane domains of its components have been solved, our knowledge of the structure of the transmembrane regions remains limited.

The bioinformatics analysis of the *Salmonella* SPI-1 vT3SS export apparatus components InvA, SpaP, SpaQ, SpaR, and SpaS and analysis of homologs of SctRST showed clear evidence for the membrane topology of InvA, SpaQ and SpaS, all with an N_{IN}/C_{IN} orientation. For InvA, SpaP, and SpaS, these results could be validated experimentally using SCAM^{TM}. For the remaining minor export apparatus proteins SpaP and SpaR, the regions for TMH could be defined; however, the exact number of TMHs could not be resolved. We did find indications for a N_{OUT}/C_{OUT} orientation of SpaP and a N_{OUT}/C_{IN} orientation for SpaR.

A model of the cytoplasmic C-terminus of the major export apparatus protein InvA (SctV) based on cryo-electron microscopy has been previously obtained [12]. We can now verify the number of membrane helices and position of the large loop. The results of the topology

mapping of InvA show eight transmembrane helices with cytoplasmic N- and C-terminus and a large cytoplasmic fourth loop (Fig. 7). Previous topology studies of YscV in *Y. pestis* via PhoA fusion mapped the first, fifth and seventh loop to be periplasmic, at residues which closely agree with our results [21]. In *X. campestris* the export apparatus proteins have been studied by *phoA–lacZ* reporter fusions. The study showed evidence for seven transmembrane helices of the InvA homolog HrcV, with a periplasmic N-terminus and a cytoplasmic C-terminus [20]. These results correspond well to the last seven transmembrane helices detected, leaving out the very first one. This is unsurprising as the study used strongly truncated versions of the protein, missing most of the C-terminal information which can critically influence the topology.

The results of the SCAM experiment of SpaP (SctR) points to four transmembrane helices with a large periplasmic second loop as well as periplasmic N- and C-terminus (Fig. 8). This is in accordance with N-terminal SpaP interaction with the periplasmic inner rod protein PrgJ (SctI) [25]. Additionally, some of the SpaR homologs (supplement S11) in flagella systems may include signal peptides which also point to an N_{OUT} orientation [50]. The strongly positive charges in the periplasmic second loop may play a different role in structure and function of the SpaPR subcomplex.

Similar to InvA, a model of the C-terminus of SpaS (SctU) has been published previously [13]. The results of the topology mapping show four transmembrane helices and cytoplasmic N- and C-terminus (Fig. 9). This is consistent with the data obtained in a PhoA fusion study in *Y. enterolitica* [27]. We can complement this data with reliable cytoplasmic signals.

We were not able to conclusively map the topology of the minor export apparatus protein SpaR (SctT). HrcT (SctT) in *Xanthomonas* was reported to have only one N-terminal transmembrane helix which stands in sharp contrast to bioinformatics predictions (Fig. 2, supplement S5) suggesting four to six transmembrane helices [20]. The study used a *phoA–*

lacZ reporter fusion, obtaining data that concurs largely with our results for the export apparatus proteins SctRUV. Nevertheless, as mentioned above, the reporter studies are based on truncated versions, making these results debatable as SpaR seems to be especially sensitive to even small changes such as amino acid substitution (data not shown). In bacterial genomes, *sctT* and *sctU* are often positioned consecutive in the same operon [51]. Equivalents of both genes exists in *Clostridium acetobutylicum* and *Clostridium tetani* as native gene fusions [52], [53]. In *S.* Typhimurium fT3SS a successful fusion of the homologs *fliR-flhB* was constructed [51]. This suggests that the C-terminal part of SctT could be in the cytoplasm alongside the N-terminus of SctU. In the *S.* Typhimurium SPI-1 vT3SS *spaR* and *spaS* are consecutive with a fourteen nucleotide overlap [38]. However, while we were able to construct a SpaRS fusion, the fusion was not functional for secretion (data not shown). Additionally, preliminary cross-link data (not shown) suggest interaction between the second to last extramembrane loop of SpaR and the SpaP periplasmic second loop as well as between the same SpaP loop and the C-terminus of SpaR. Together with our cytoplasmic prediction of the last loop of SctT homologs, this underpins the likelihood of several TMHs at the end of SpaR and a C_{OUT} orientation.

Density differences were observed in the centre of the base between averaged cryo-electron microscopy images of wild type and export apparatus knock out mutants (*ΔinvA Δspa* and *ΔspaP*) which were termed cup and socket structure [19], [54]. It was suggested that the proteins SpaPQR (SctRST) [19] or PrgJ with periplasmic loops of SpaPR (SctIRT) [25] account for the unassigned density of the socket and form of a defined substructure. We can here confirm that the large loop of SpaP is indeed periplasmic. While the periplasmic part of SpaR remains undefined, the 5:1 stoichiometry of the SpaPR subcomplex means that the five SpaP, alongside six PrgJ [55] could represent a major part of the socket structure. Neither InvA (SctV) nor SpaS (SctU) contain large periplasmic loops. InvA though, of which a total of nine copies are present in the T3SS [22], [55] includes a large cytoplasmic fourth loop to

which no structure has been assigned yet. In FlhA (SctV) this loop has been suggested to play a role in inwardly-directed proton flow and may interact with FliR (SctT) [56].

In summary, we were able to define the topologies of the major export apparatus protein InvA and the switch protein SpaS, and ascertain the periplasmic orientation of the largest loop and the termini of SpaP. We also found indications for cytoplasmic orientations of the termini of SpaQ and a C_{OUT} orientation for SpaR. This information will support further structural studies, such as crosslinking experiments and assist ultimately in elucidating the functions of export apparatus proteins in the injectisome.

Acknowledgments

We thank Andrea Eipper and Melanie Riess for technical assistance in the preparation of the SCAM experiments and Iwan Grin for critically reading the manuscript. This work was supported by a postdoctoral fellowship of the Human Frontiers Science Program (to S.W.) and by the Alexander von Humboldt Foundation in the framework of the Sofja Kovalevskaja Award endowed by the Federal Ministry of Education and Research (to S.W.)

Figures:

Figure 1: Overview of virulence associated type III secretion system proteins with *Salmonella* SPI-1 names as well as unified nomenclature.

Figure 2: Topology prediction of the export apparatus proteins SpaPQRS and InvA (SctRSTUV) using TOPCONS [40]**, TMHMMfix** [41] **and PredictProtein** [43]. In the summaries sequences predicted as transmembrane helices in all predictions are shown in dark gray, sequences predicted as transmembrane helices at least once are shown in light gray. InvA C-terminal cytosolic part after amino acid 310 and SpaS C-terminal cystosolic part after amino acid 205 were omitted.

Figure 3: Review of charged residues of the export apparatus proteins SctRST homologs (vT3SS and fT3SS of *Salmonella* Typhimurium, *Shigella flexneri* and *Escherichia coli*, vT3SS of *Xanthomonas campestris*, *Yersinia enterocolitica*, and *Pseudomonas syringae*, and fT3SS of *Helicobacter pylori*; Uniprot entry numbers in brackets). T-Coffee alignments are shown simplified with transmembrane helix (TMH) predictions by ΔG_{app}. Numbers of arginine and lysine counted in the extra membrane regions and within the first three amino acids of the adjacent predicted TMH. Clusters of negatively charged amino acids and conserved positively charged amino acids are additionally marked. Detailed sequence and TMHs in supplement S3, S6 and S9.

Figure 4: **Experimental setup of the substituted cysteine accessibility method (SCAM).** (a) A strain with a single cysteine replacement mutation is grown under conditions expressing the protein of interest. (b) The culture is divided into two samples to label cysteines either on the periplasmic (P, light grey) or the cytoplasmic (C, dark grey) side of the inner membrane. (c) The cells are treated with e.g. EDTA to ensure permeability of the outer membrane (omitted in the following) for the (d) labelling (blue, star) added to P and the blocking reagent (red, circle) added to C (e) Potential cysteines are labelled (yellow filled) in the periplasm and the labelling reaction in P is quenched (black, filled). (f) After putative periplasmic cysteines are blocked the labelling reagent is added to C. Both samples are lysed, e.g. by bead beater cell lysis, allowing the unquenched labelling reagent to reach any cytoplasmic cysteines. (g) After all reactions are quenched in C as well (h) samples are washed, membranes solubilized and proteins analysed by Western blotting after immunoprecipitation.

Figure 5: Secretion of effector proteins SipB and InvJ into the the supernatant (sup) and expression of export apparatus proteins InvA, SpaP, SpaR and SpaS in whole cells (wc). Cysteine mutants are expressed from low copy number plasmid pTACO10. *ΔinvA* strain functions as a negative control for secretion.

Figure 6: SCAM output of *Salmonella* type III secretion system export apparatus proteins InvA, SpaP and SpaS (SctRUV). Number of experiments for a) InvA, b) SpaP and d) SpaS in which the relative cytoplasmic (blue) or periplasmic (green) signal was stronger. Number of experiments in which both signals were comparable with the negative control are shown in gray. c) Relative periplasmic signal of SpaP. Error bars: standard error of the mean.

Figure 7: Topology mapping of the export apparatus protein InvA (SctV) based on the results obtained in this study. a) Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool [46].
b) Mapped amino acid positions added to topology predictions of figure 2. Colouring according to legend of a).

Figure 8: Topology mapping of the export apparatus protein SpaP (SctR) based on the results obtained in this study. a) Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool [46].
b) Mapped amino acid positions added to topology predictions of figure 2. Colouring according to legend of a).

Figure 9: Topology mapping of the export apparatus protein SpaS (SctU) based on the results obtained in this study. a) Membrane depicted as yellow bar. Membrane external loops are numbered in roman numerals. Visualized using PROTTER online tool [46].
b) Mapped amino acid positions added to topology predictions of figure 2. Colouring according to legend of a).

Supplement

Table ST1.: Strains and plasmids used in this study

Table ST2.: Summary of the predictions of TMHs using the prediction tools TOPCONS [40], TMHMMfix [41], ΔG_{app} [42] and PredictProtein (PHDhtm, PROFsec, SNAP2) [43]

Supplement Data:

S1: SctR T-Coffee alignment [44] of SpaP (P40700) and FliP (P54700) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcR (G0CIR4) of *Xanthomonas campestris* pv. raphani 756C, YscR (Q9ZA77) of *Yersinia enterocolitica*, FliP (A0A0H2V084) and Spa24 (P0A1L3) of *Shigella flexneri*, HrcR (A5YJ93) of *Pseudomonas syringae* pv. syringae, EpaP (Q3SBB6) and FliP (A0A023KQG0) of *Escherichia coli*, and FliP (A0A024BZY8) of *Helicobacter pylori.* UniProt identifiers in brackets.

S2: SctR ΔG_{app} [42] **output** of SpaP (P40700) and FliP (P54700) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcR (G0CIR4) of *Xanthomonas campestris* pv. raphani 756C, YscR (Q9ZA77) of *Yersinia enterocolitica*, FliP (A0A0H2V084) and Spa24 (P0A1L3) of *Shigella flexneri*, HrcR (A5YJ93) of *Pseudomonas syringae* pv. syringae, EpaP (Q3SBB6) and FliP (A0A023KQG0) of *Escherichia coli*, and FliP (A0A024BZY8) of *Helicobacter pylori.* Full protein scan with helix min length of 18, helix max length of 40 and length correction on. UniProt identifiers in brackets.

S3: SctR review of charged residues of SpaP (P40700) and FliP (P54700) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcR (G0CIR4) of *Xanthomonas campestris* pv. raphani 756C, YscR (Q9ZA77) of *Yersinia enterocolitica*, FliP (A0A0H2V084) and Spa24 (P0A1L3) of *Shigella flexneri*, HrcR (A5YJ93) of *Pseudomonas syringae* pv. syringae, EpaP (Q3SBB6) and FliP (A0A023KQG0) of *Escherichia coli*, and

FliP (A0A024BZY8) of *Helicobacter pylori.* T-Coffee alignment (S1) with ΔG_{app} predicted TMHs (S2).

S4: SctS T-Coffee alignment [44] of SpaQ (P0A1L7) and FliQ (P0A1L5) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcS (G0CIR3) of *Xanthomonas campestris* pv. raphani 756C, YscS (Q7BFA7) of *Yersinia enterocolitica*, FliQ (P0AC10) and Spa9 (P0A1M4) of *Shigella flexneri*, HrcS (Q60237) of *Pseudomonas syringae* pv. syringae, EpaQ (Q2TJ91) and FliQ (C3T4R7) of *Escherichia coli*, and FliQ (P0A0S3) of *Helicobacter pylori* (strain ATCC 700392 / 26695). UniProt identifiers in brackets.

S5: SctS ΔG_{app} [42] **output** of SpaQ (P0A1L7) and FliQ (P0A1L5) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcS (G0CIR3) of *Xanthomonas campestris* pv. raphani 756C, YscS (Q7BFA7) of *Yersinia enterocolitica*, FliQ (P0AC10) and Spa9 (P0A1M4) of *Shigella flexneri*, HrcS (Q60237) of *Pseudomonas syringae* pv. syringae, EpaQ (Q2TJ91) and FliQ (C3T4R7) of *Escherichia coli*, and FliQ (P0A0S3) of *Helicobacter pylori* (strain ATCC 700392 / 26695). Full protein scan with helix min length of 18, helix max length of 40 and length correction on. UniProt identifiers in brackets.

S6: SctS review of charged residues of SpaQ (P0A1L7) and FliQ (P0A1L5) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcS (G0CIR3) of *Xanthomonas campestris* pv. raphani 756C, YscS (Q7BFA7) of *Yersinia enterocolitica*, FliQ (P0AC10) and Spa9 (P0A1M4) of *Shigella flexneri*, HrcS (Q60237) of *Pseudomonas syringae* pv. syringae, EpaQ (Q2TJ91) and FliQ (C3T4R7) of *Escherichia coli*, and FliQ (P0A0S3) of *Helicobacter pylori* (strain ATCC 700392 / 26695). T-Coffee alignment (S4) with ΔG_{app} predicted TMH (S5).

S7: SctT T-Coffee alignment [44] of SpaR (P40701) and FliR (P54702) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcT (G0CIS6) of *Xanthomonas*

campestris pv. raphani 756C, YscT (Q93KT5) of *Yersinia enterocolitica*, FliR (Q83R27) and Spa29 (P0A1M6) of *Shigella flexneri*, HrcT (Q60238) of *Pseudomonas syringae* pv. syringae, EpaR (A0A024L995) and FliR (A0A0C3MET9) of *Escherichia coli*, and FliR (O24978) of *Helicobacter pylori* (strain ATCC 700392 / 26695). UniProt identifiers in brackets.

S8: SctT ΔG_{app} [42] **output** of SpaR (P40701) and FliR (P54702) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcT (G0CIS6) of *Xanthomonas campestris* pv. raphani 756C, YscT (Q93KT5) of *Yersinia enterocolitica*, FliR (Q83R27) and Spa29 (P0A1M6) of *Shigella flexneri*, HrcT (Q60238) of *Pseudomonas syringae* pv. syringae, EpaR (A0A024L995) and FliR (A0A0C3MET9) of *Escherichia coli*, and FliR (O24978) of *Helicobacter pylori* (strain ATCC 700392 / 26695). Full protein scan with helix min length of 18, helix max length of 40 and length correction on. UniProt identifiers in brackets.

S9: SctT review of charged residues SpaR (P40701) and FliR (P54702) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720), HrcT (G0CIS6) of *Xanthomonas campestris* pv. raphani 756C, YscT (Q93KT5) of *Yersinia enterocolitica*, FliR (Q83R27) and Spa29 (P0A1M6) of *Shigella flexneri*, HrcT (Q60238) of *Pseudomonas syringae* pv. syringae, EpaR (A0A024L995) and FliR (A0A0C3MET9) of *Escherichia coli*, and FliR (O24978) of *Helicobacter pylori* (strain ATCC 700392 / 26695).T-Coffee alignment (S7) with ΔG_{app} predicted TMHs (S8).

S10: TOPCONS [40] **predictions** of the export apparatus proteins SpaPQRS (P40700, P0A1L7, P40701, P40702) and InvA (P0A1I3) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720). UniProt identifiers in brackets.

S11: SignalP [45] output for the flagellar SctR proteins FliP (P54700) of *Salmonella* Typhimurium (strain LT2 / SGSC1412 / ATCC 700720) FliP (A0A023KQG0) of *Escherichia coli*, and FliP (A0A024BZY8) of *Helicobacter pylori.* UniProt identifiers in brackets.

References

[1] J. E. Galán, M. Lara-Tejero, T. C. Marlovits, and S. Wagner, "Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells.," *Annu. Rev. Microbiol.*, vol. 68, pp. 415–38, 2014.

[2] C. J. Hueck, "Type III Protein Secretion Systems in Bacterial Pathogens of Animals and Plants CELLULAR AND MOLECULAR IMPACT OF TYPE III SECRETION IN BACTERIAL," *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 379–433, 1998.

[3] T. Minamino and K. Namba, "Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export," *Nature*, vol. 451, no. January, pp. 1–3, 2008.

[4] M. Y. Galperin, P. A. Dibrov, and A. N. Glagolev, "ΔμH + is required for flagellar growth in Escherichia," *FEBS Lett.*, vol. 143, no. 2, pp. 319–322, 1982.

[5] D. Büttner, "Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria," *Microbiol. Mol. Biol. Rev.*, vol. 76, no. 2, pp. 262–310, Jun. 2012.

[6] T. Kubori, Y. Matsushima, D. Nakamura, J. Uralil, M. Lara-Tejero, A. Sukhan, J. E. Galán, and Shin, "Supramolecular Structure of the Salmonella typhimurium Type III Protein Secretion System," *Science (80-.).*, vol. 280, no. 5363, pp. 602–605, Apr. 1998.

[7] A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, and A. Lange, "Atomic model of the type III secretion system needle.," *Nature*, vol. 486, no. 7402, pp. 276–9, 2012.

[8] K. Yonekura, S. Maki-Yonekura, and K. Namba, "Complete atomic model of the

bacterial flagellar filament by electron cryomicroscopy," *Nature*, vol. 424, no. 6949, pp. 643–650, 2003.

[9] F. Fan, K. Ohnishi, N. R. Francis, and R. M. Macnab, "The FliP and FliR proteins of Salmonella typhimurium, putative components of the type III flagellar export apparatus, are located in the flagellar basal body.," *Mol. Microbiol.*, vol. 26, no. 5, pp. 1035–1046, 1997.

[10] T. Minamino and R. M. Macnab, "Interactions among components of the Salmonella flagellar export apparatus and its substrates," *Mol. Microbiol.*, vol. 35, no. 5, pp. 1052–1064, 2000.

[11] T. Minamino and R. M. Macnab, "Domain Structure of Salmonella FlhB , a Flagellar Export Component Responsible for Substrate Specificity Switching Domain Structure of Salmonella FlhB , a Flagellar Export Component Responsible for Substrate Specificity Switching," vol. 182, no. 17, 2000.

[12] L. J. Worrall, M. Vuckovic, and N. C. J. Strynadka, "Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA.," *Protein Sci.*, vol. 19, no. 5, pp. 1091–1096, May 2010.

[13] R. Zarivach, W. Deng, M. Vuckovic, H. B. Felise, H. V Nguyen, S. I. Miller, B. B. Finlay, and N. C. J. Strynadka, "Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS.," *Nature*, vol. 453, no. 7191, pp. 124–127, May 2008.

[14] S. A. Moore and Y. Jia, "Structure of the cytoplasmic domain of the flagellar secretion apparatus component FlhA from Helicobacter pylori," *J. Biol. Chem.*, vol. 285, no. 27, pp. 21060–21069, 2010.

[15] B. Hu, D. R. Morado, W. Margolin, J. R. Rohde, O. Arizmendi, W. L. Picking, W. D. Picking, and J. Liu, “Visualization of the type III secretion sorting platform of Shigella flexneri.,” *PNAS*, vol. 112, no. 4, pp. 1047–52, 2015.

[16] A. Diepold and J. P. Armitage, “Type III secretion systems: the bacterial flagellum and the injectisome,” *Philos. Trans. R. Soc. B*, vol. 370, no. 1679, p. 20150020, 2015.

[17] K. Sekiya, M. Ohishi, T. Ogino, K. Tamano, C. Sasakawa, and A. Abe, “Supermolecular structure of the enteropathogenic Escherichia coli type III secretion system and its direct interaction with the EspA-sheath-like structure,” *Proc Natl Acad Sci U S A*, vol. 98, no. 20, pp. 11638–11643, 2001.

[18] J. V. Monjarás Feria, M. D. Lefebre, Y.-D. Stierhof, J. E. Galán, and S. Wagner, “Role of Autocleavage in the Function of a Type III Secretion Specificity Switch Protein in Salmonella enterica Serovar,” *MBio*, vol. 6, no. 5, pp. 1–8, 2015.

[19] S. Wagner, L. Königsmaier, M. Lara-tejero, M. Lefebre, T. C. Marlovits, and J. E. Galán, “Organization and coordinated assembly of the type III secretion export apparatus,” *PNAS*, vol. 107, no. 41, pp. 17745–17750, 2010.

[20] C. Berger, G. P. Robin, U. Bonas, and R. Koebnik, “Membrane topology of conserved components of the type III secretion system from the plant pathogen Xanthomonas campestris pv. vesicatoria.,” *Microbiology*, vol. 156, no. Pt 7, pp. 1963–1974, Jul. 2010.

[21] G. V. Plano, S. S. Barve, and S. C. Straley, “LcrD, a membrane-bound regulator of the Yersinia pestis low-calcium response,” *J. Bacteriol.*, vol. 173, no. 22, pp. 7293–7303, 1991.

[22] P. Abrusci, M. Vergara-Irigaray, S. Johnson, M. D. Beeby, D. R. Hendrixson, P.

Roversi, M. E. Friede, J. E. Deane, G. J. Jensen, C. M. Tang, and S. M. Lea, "Architecture of the major component of the type III secretion system export apparatus.," *Nat. Struct. Mol. Biol.*, vol. 20, no. 1, pp. 99–104, Jan. 2013.

[23] C. S. Barker, T. Inoue, I. V. Meshcheryakova, S. Kitanobo, and F. A. Samatey, "Function of the conserved FHIPEP domain of the flagellar type III export apparatus, protein FlhA," *Mol. Microbiol.*, vol. 100, no. 2, pp. 278–288, 2016.

[24] M. Erhardt, P. Wheatley, E. A. Kim, T. Hirano, Y. Zhang, M. K. Sarkar, K. T. Hughes, and D. F. Blair, "Mechanism of type-III protein secretion : Regulation of FlhA conformation by a functionally critical charged-residue cluster," *Mol Microbiol*, vol. Accepted A, 2017.

[25] T. Dietsche, M. T. Mebrhatu, M. J. Brunner, S. Zilkenat, P. Abrusci, J. Yan, M. Franz-wachtel, C. Scha, O. Kohlbacher, S. Lea, B. Macek, I. Grin, J. E. Gala, C. Marlovits, C. V Robinson, and S. Wagner, "Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus," *PLoS Pathog.*, vol. 12, no. 12, pp. 1–25, 2016.

[26] K. A. Fields, G. V. Plano, and S. C. Straley, "A low-Ca2+ response (LCR) secretion (ysc) locus lies within the lcrB region of the LCR plasmid in Yersinia pestis," *J. Bacteriol.*, vol. 176, no. 3, pp. 569–579, 1994.

[27] A. Allaoui, S. Woestyn, C. Sluiters, and G. R. Cornelis, "Yscu, a Yersinia-Enterocolitica Inner Membrane-Protein Involved in Yop Secretion," *J. Bacteriol.*, vol. 176, no. 15, pp. 4534–4542, 1994.

[28] D. K. Shen, N. Moriya, I. Martinez-Argudo, and A. J. Blocker, "Needle length control and the secretion substrate specificity switch are only loosely coupled in the type III secretion apparatus of Shigella," *Microbiol. (United Kingdom)*, vol. 158, no. 7, pp.

1884–1896, 2012.

[29] S. Frost, O. Ho, F. H. Login, C. F. Weise, H. Wolf-Watz, and M. Wolf-Watz, “Autoproteolysis and Intramolecular Dissociation of Yersinia YscU Precedes Secretion of Its C-Terminal Polypeptide YscUCC,” *PLoS One*, vol. 7, no. 11, 2012.

[30] S. Zilkenat, I. Grin, and S. Wagner, “Stoichiometry determination of macromolecular membrane protein complexes,” *Biol. Chem.*, vol. 398, no. 2, pp. 155–164, 2016.

[31] R. De Zorzi, W. Mi, M. Liao, and T. Walz, “Single-particle electron microscopy in the study of membrane protein structure,” *Microscopy*, vol. 65, no. 1, pp. 81–96, 2016.

[32] T. J. Silhavy and J. R. Beckwith, “Uses of lac fusions for the study of biological problems.,” *Microbiol. Rev.*, vol. 49, no. 4, pp. 398–418, 1985.

[33] C. Manoil, J. J. Mekalanos, and J. Beckwith, “Alkaline phosphatase fusions: Sensors of subcellular location,” *J. Bacteriol.*, vol. 172, no. 2, pp. 515–518, 1990.

[34] S. I. Hyun, L. Maruri-Avidal, and B. Moss, “Topology of Endoplasmic Reticulum-Associated Cellular and Viral Proteins Determined with Split-GFP,” *Traffic*, vol. 16, no. 7, pp. 787–795, 2015.

[35] M. Bogdanov, W. Zhang, J. Xie, and W. Dowhan, “Transmembrane protein topology mapping by the substituted cysteine accessibility method (SCAM(TM)): application to lipid-specific membrane protein topogenesis.,” *Methods*, vol. 36, no. 2, pp. 148–171, Jun. 2005.

[36] S. Seppälä, J. S. Slusky, P. Lloris-Garcerá, M. Rapp, and G. von Heijne, “Control of membrane protein topology by a single C-terminal residue.,” *Science*, vol. 328, no. 5986, pp. 1698–1700, 2010.

[37] S. Frillingos, M. Sahin-toth, J. Wu, and H. R. Kaback, “Cys-scanning mutagenesis: a

novel approach to structure–function relationships in polytopic membrane proteins," *Faseb J*, vol. 12, no. 13, pp. 1281–1299, 1998.

[38] C. Kröger, S. C. Dillon, A. D. S. Cameron, K. Papenfort, S. K. Sivasankaran, K. Hokamp, Y. Chao, A. Sittka, M. Hébrard, K. Händler, A. Colgan, P. Leekitcharoenphon, G. C. Langridge, A. J. Lohan, B. Loftus, S. Lucchini, D. W. Ussery, C. J. Dorman, N. R. Thomson, J. Vogel, and J. C. D. Hinton, "The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium," *Proc. Natl. Acad. Sci.*, vol. 109, no. 20, pp. E1277–E1286, 2012.

[39] L.-M. Guzman, D. Belin, M. J. Carson, and J. Beckwith, "Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose PBAD Promoter," *J. Bacteriol.*, vol. 177, no. 14, pp. 4121–4130, 1995.

[40] K. D. Tsirigos, C. Peters, N. Shu, L. Käll, and A. Elofsson, "The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides.," *Nucleic Acids Res.*, vol. 43, no. W1, pp. W401-7, 2015.

[41] K. Melén, A. Krogh, and G. Von Heijne, "Reliability measures for membrane protein topology prediction algorithms," *J. Mol. Biol.*, vol. 327, no. 3, pp. 735–744, 2003.

[42] T. Hessa, N. M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S. H. White, and G. von Heijne, "Molecular code for transmembrane-helix recognition by the Sec61 translocon.," *Nature*, vol. 450, no. 7172, pp. 1026–30, Dec. 2007.

[43] G. Yachdav, E. Kloppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Hönigschmid, A. Schafferhans, M. Roos, M. Bernhofer, L. Richter, H. Ashkenazy, M. Punta, A. Schlessinger, Y. Bromberg, R. Schneider, G. Vriend, C. Sander, N. Ben-Tal, and B. Rost, "PredictProtein - An open resource for online prediction of protein

structural and functional features," *Nucleic Acids Res.*, vol. 42, no. W1, pp. 337–343, 2014.

[44] C. Notredame, D. Higgins, and J. Heringa, "T-Coffee: A novel method for fast and accurate multiple sequence alignment.," *J. Mol. Biol.*, vol. 302, no. 1, pp. 205–217, 2000.

[45] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nat. Methods*, vol. 8, pp. 785–786, 2011.

[46] U. Omasits, C. H. Ahrens, S. M??ller, and B. Wollscheid, "Protter: Interactive protein feature visualization and integration with experimental proteomic data," *Bioinformatics*, vol. 30, no. 6, pp. 884–886, 2014.

[47] H. Andersson, E. Bakker, and G. Von Heijne, "Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in Escherichia coli," *J. Biol. Chem.*, vol. 267, no. 3, pp. 1491–1495, 1992.

[48] J. M. Chang, P. Di Tommaso, and C. Notredame, "TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction," *Mol. Biol. Evol.*, vol. 31, no. 6, pp. 1625–1637, 2014.

[49] I. Nilsson and G. von Heijne, "Fine-tuning the topology of a polytopic membrane protein: Role of positively and negatively charged amino acids," *Cell*, vol. 62, no. 6, pp. 1135–1141, 1990.

[50] P. Wang, E. Shim, B. Cravatt, R. Jacobsen, J. Schoeniger, A. C. Kim, M. Paetzel, and R. E. Dalbey, "Escherichia coli signal peptide peptidase A is a serine-lysine protease with a lysine recruited to the nonconserved amino-terminal domain in the S49 protease family," *Biochemistry*, vol. 47, no. 24, pp. 6361–6369, 2008.

[51] J. S. Van Arnam, J. L. McMurry, M. Kihara, and R. M. Macnab, “Analysis of an Engineered Salmonella Flagellar Fusion Protein, FliR-FlhB,” *J. Bacteriol.*, vol. 186, no. 8, pp. 2495–2498, 2004.

[52] H. Brüggemann, S. Bäumer, W. F. Fricke, A. Wiezer, H. Liesegang, I. Decker, C. Herzberg, R. Martinez-Arias, R. Merkl, A. Henne, and G. Gottschalk, “The genome sequence of Clostridium tetani, the causative agent of tetanus disease.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 3, pp. 1316–21, 2003.

[53] J. Nölling, G. Breton, M. V Omelchenko, S. Kira, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, G. T. C. Sequencing, Y. I. Wolf, R. L. Tatusov, F. Sabathe, P. Soucaille, M. J. Daly, G. N. Bennett, E. V Koonin, and D. R. Smith, “Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum,” *J. Bacteriol.*, vol. 183, no. 16, pp. a823–4838, 2001.

[54] T. C. Marlovits, T. Kubori, A. Sukhan, D. R. Thomas, E. Jorge, V. M. Unger, J. E. Galán, and V. M. Unger, “Structural Insights into the Assembly of the Type III Secretion Needle Complex,” *Science (80-.).*, vol. 306, no. 5698, pp. 1040–1042, Nov. 2004.

[55] S. Zilkenat, M. Franz-Wachtel, Y.-D. Stierhof, J. E. Galán, B. Macek, and S. Wagner, “Determination of the Stoichiometry of the Complete Bacterial Type III Secretion Needle Complex Using a Combined Quantitative Proteomic Approach,” *Mol. Cell. Proteomics*, vol. 15, no. 5, pp. 1598–1609, 2016.

[56] N. Hara, K. Namba, and T. Minamino, “Genetic characterization of conserved charged residues in the bacterial flagellar type III export protein FlhA,” *PLoS One*, vol. 6, no. 7,

p. e22417, 2011.

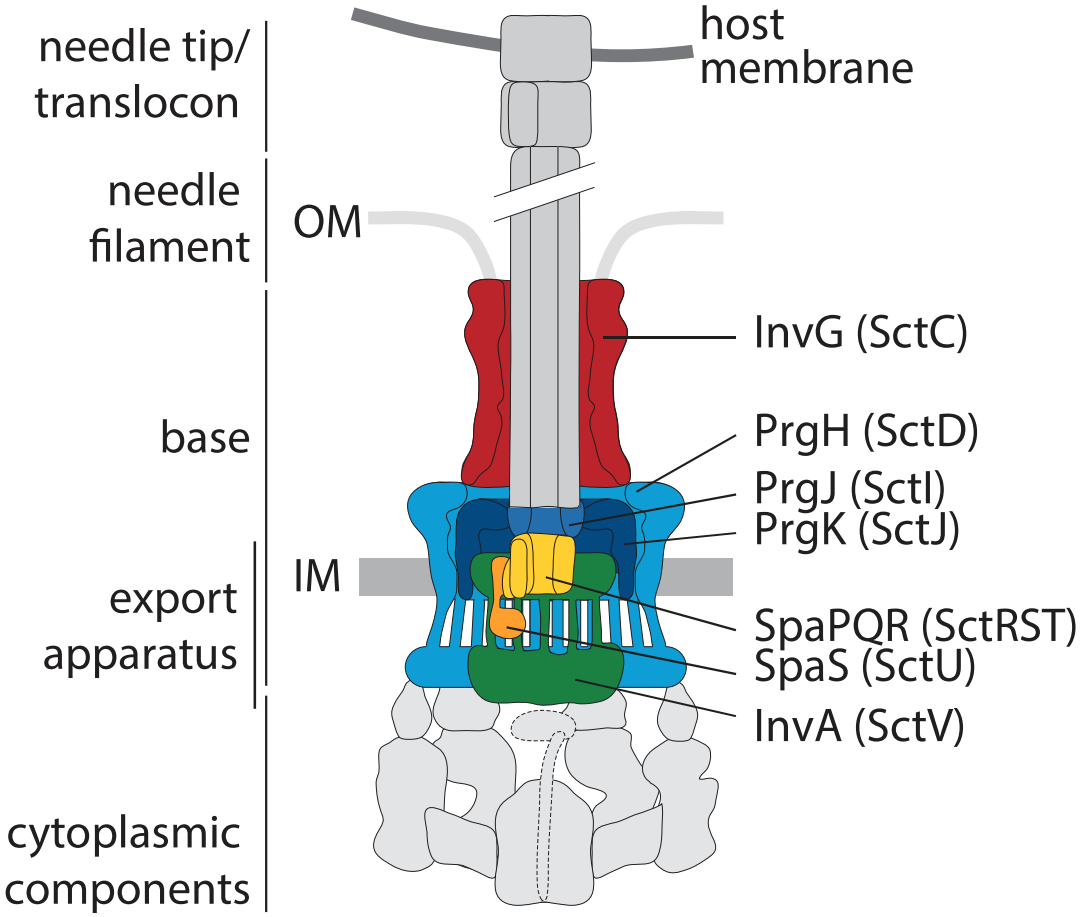

needle tip/
translocon
needle
filament
base
export
apparatus
cytoplasmic
components
host
membrane
OM
IM
InvG (SctC)
PrgH (SctD)
PrgJ (SctI)
PrgK (SctJ)
SpaPQR (SctRST)
SpaS (SctU)
InvA (SctV)

Fig. 1

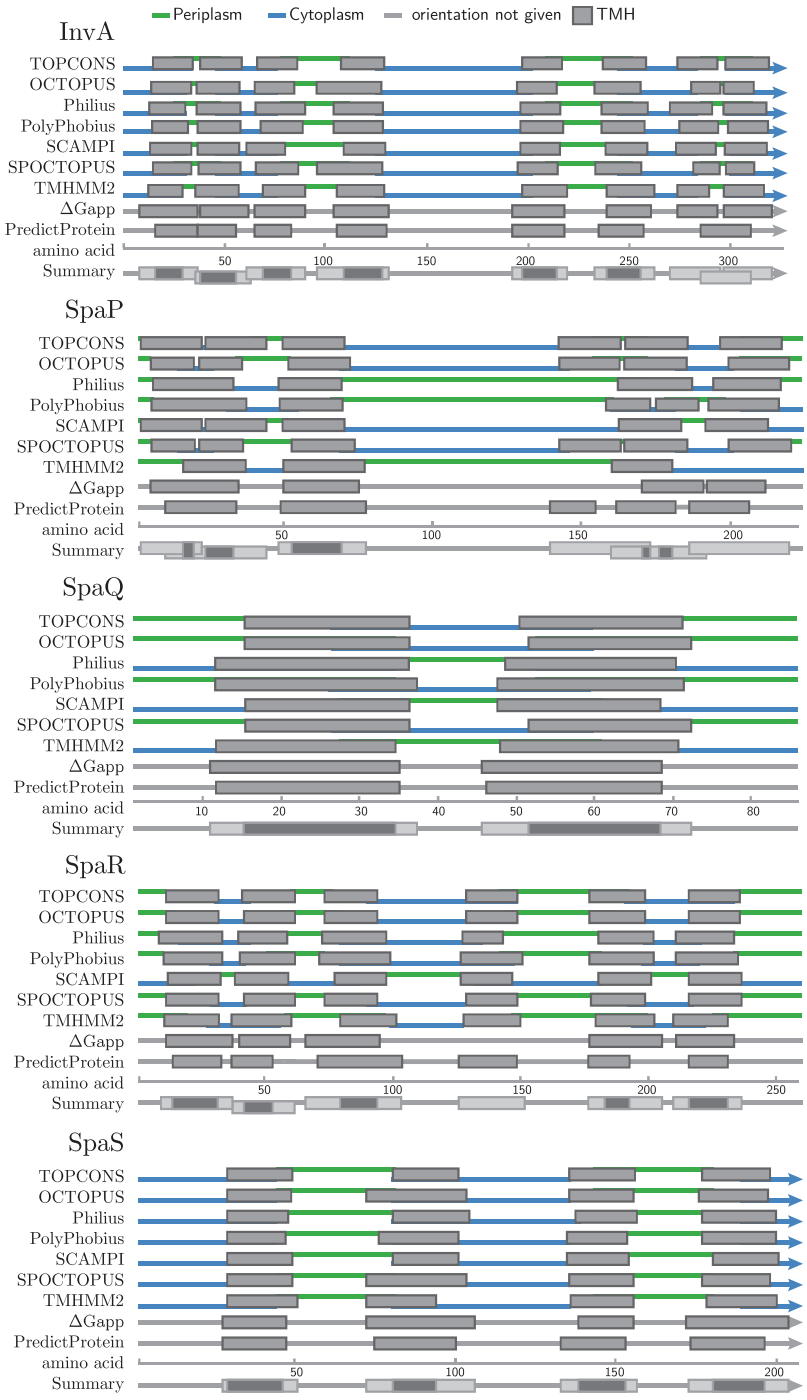

Periplasm
Cytoplasm
orientation not given
TMH
InvA
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TMHMM2
ΔGapp
PredictProtein
amino acid
Summary
50
100
150
200
250
300
SpaP
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TMHMM2
ΔGapp
PredictProtein
amino acid
Summary
50
100
150
200
SpaQ
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TMHMM2
ΔGapp
PredictProtein
amino acid
Summary
10
20
30
40
50
60
70
80
SpaR
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TMHMM2
ΔGapp
PredictProtein
amino acid
Summary
50
100
150
200
250
SpaS
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
TMHMM2
ΔGapp
PredictProtein
amino acid
Summary
50
100
150
200

Fig. 2

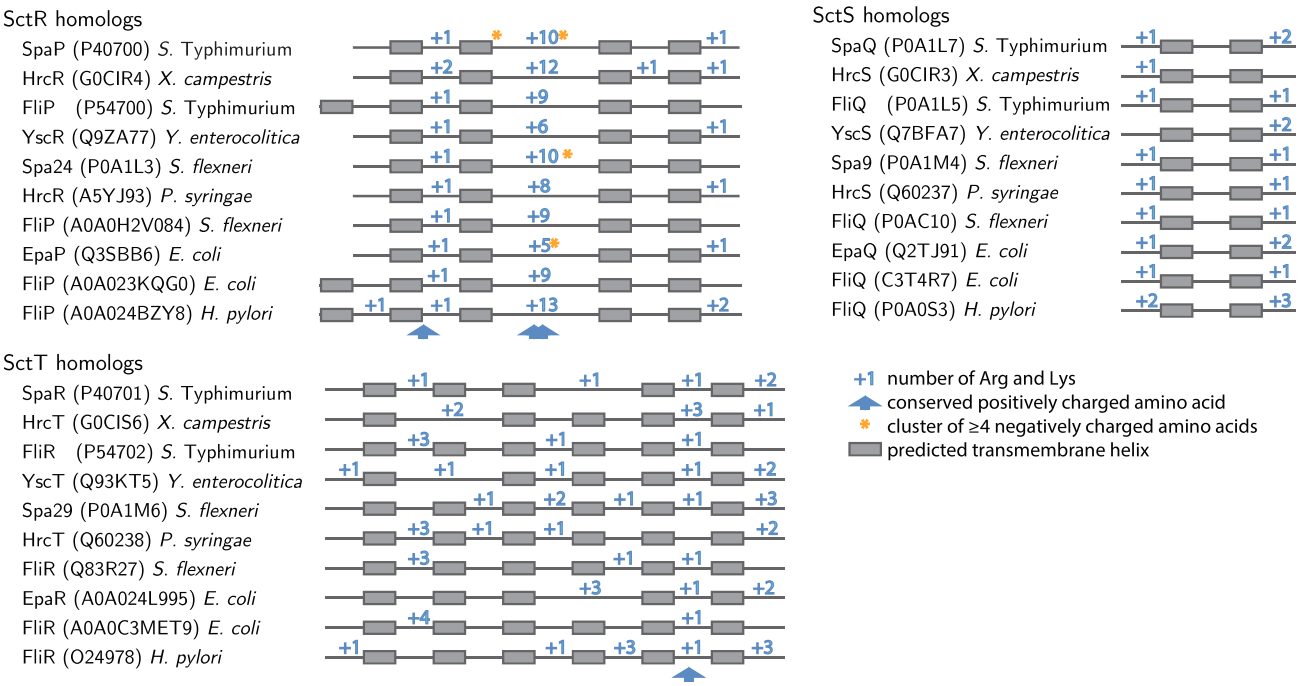

SctR homologs
SpaP (P40700) S. Typhimurium
HrcR (G0CIR4) X. campestris
FliP (P54700) S. Typhimurium
YscR (Q9ZA77) Y. enterocolitica
Spa24 (P0A1L3) S. flexneri
HrcR (A5YJ93) P. syringae
FliP (A0A0H2V084) S. flexneri
EpaP (Q3SBB6) E. coli
FliP (A0A023KQG0) E. coli
FliP (A0A024BZY8) H. pylori
SctT homologs
SpaR (P40701) S. Typhimurium
HrcT (G0CIS6) X. campestris
FliR (P54702) S. Typhimurium
YscT (Q93KT5) Y. enterocolitica
Spa29 (P0A1M6) S. flexneri
HrcT (Q60238) P. syringae
FliR (Q83R27) S. flexneri
EpaR (A0A024L995) E. coli
FliR (A0A0C3MET9) E. coli
FliR (O24978) H. pylori
SctS homologs
SpaQ (P0A1L7) S. Typhimurium
HrcS (G0CIR3) X. campestris
FliQ (P0A1L5) S. Typhimurium
YscS (Q7BFA7) Y. enterocolitica
Spa9 (P0A1M4) S. flexneri
HrcS (Q60237) P. syringae
FliQ (P0AC10) S. flexneri
EpaQ (Q2TJ91) E. coli
FliQ (C3T4R7) E. coli
FliQ (P0A0S3) H. pylori
+1 number of Arg and Lys
conserved positively charged amino acid
cluster of ≥4 negatively charged amino acids
predicted transmembrane helix

Fig. 3

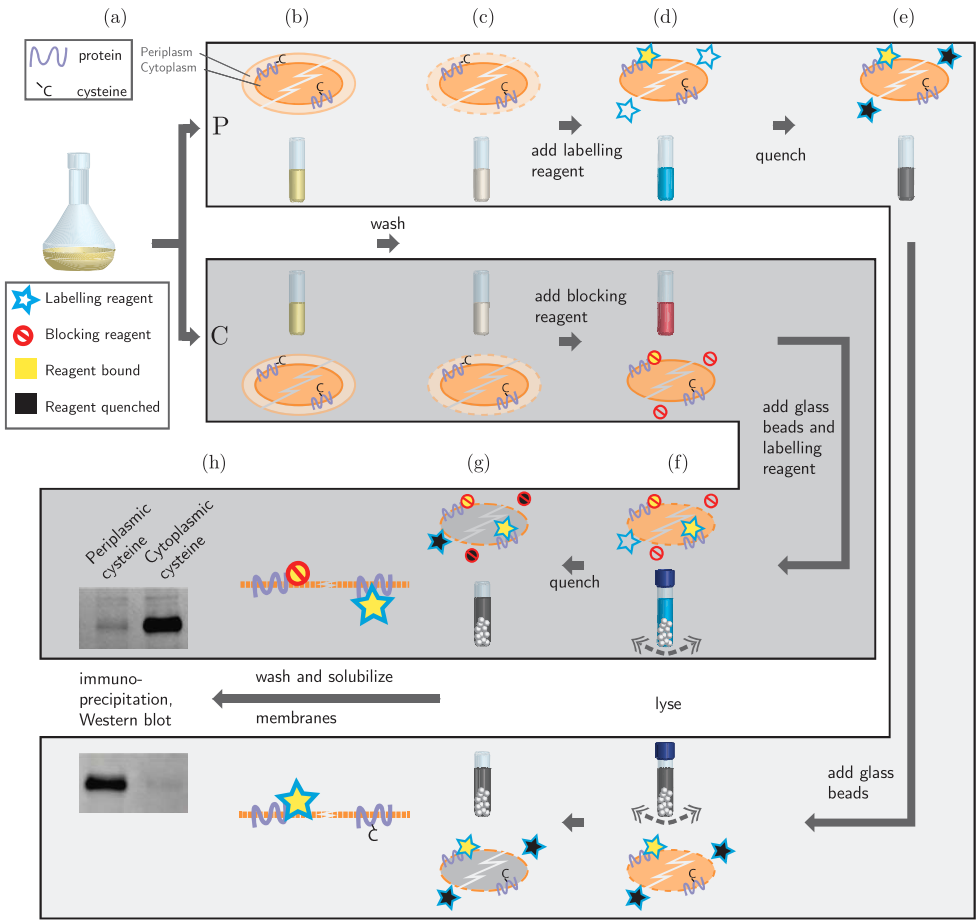

(a)
(b)
(c)
(d)
(e)
protein
cysteine
Periplasm
Cytoplasm
P
add labelling reagent
quench
wash
C
add blocking reagent
Labelling reagent
Blocking reagent
Reagent bound
Reagent quenched
add glass beads and labelling reagent
(h)
(g)
(f)
Periplasmic cysteine
Cytoplasmic cysteine
quench
immuno-precipitation, Western blot
wash and solubilize membranes
lyse
add glass beads

Fig. 4

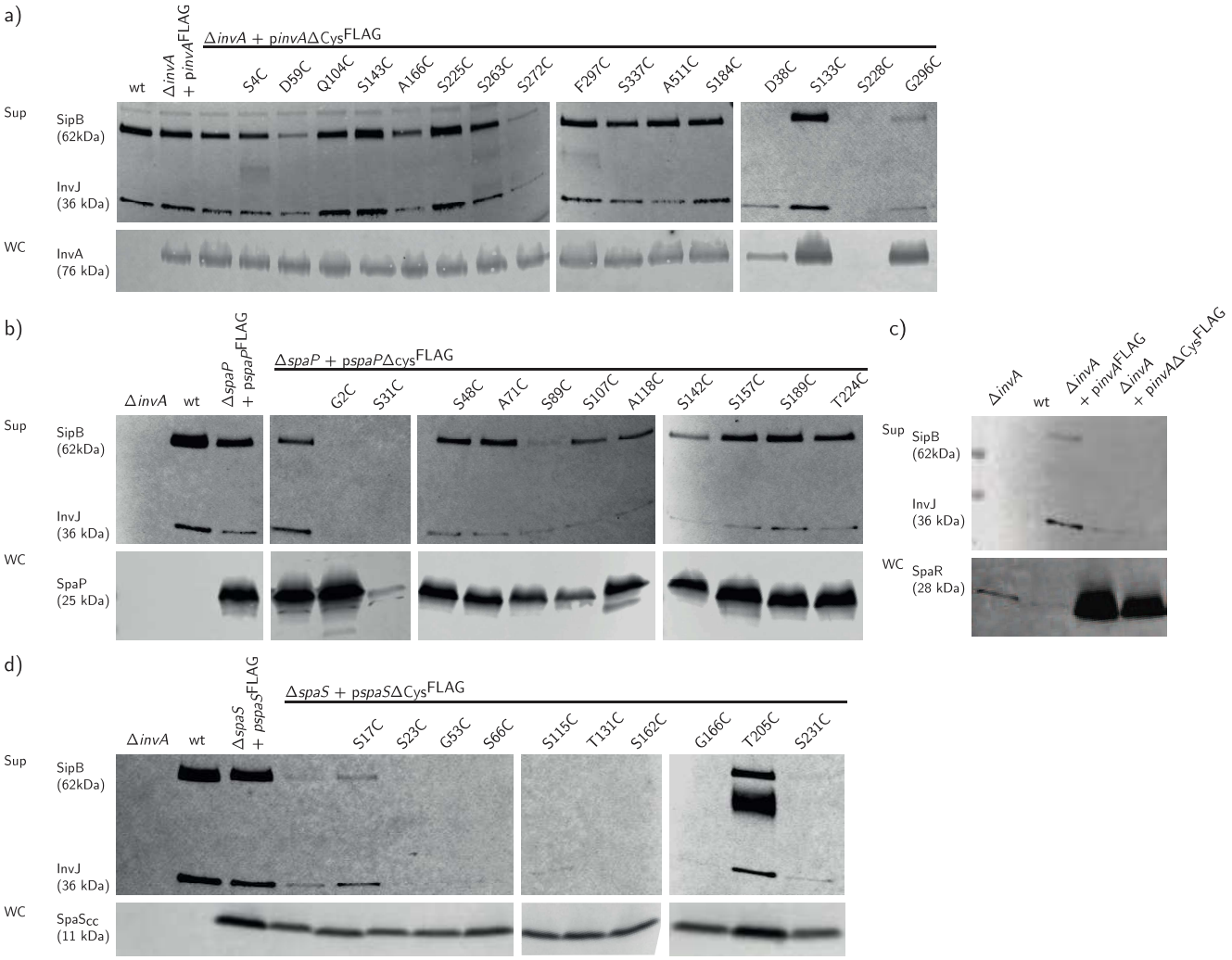

a)
ΔinvA + pinvAΔCysFLAG
wt
ΔinvA + pinvAFLAG
S4C
D59C
Q104C
S143C
A166C
S225C
S263C
S272C
F297C
S337C
A511C
S184C
D38C
S133C
S228C
G296C
Sup
SipB (62kDa)
InvJ (36 kDa)
WC
InvA (76 kDa)
b)
ΔspaP + pspaPΔcysFLAG
ΔinvA
wt
ΔspaP + pspaPFLAG
G2C
S31C
S48C
A71C
S89C
S107C
A118C
S142C
S157C
S189C
T224C
Sup
SipB (62kDa)
InvJ (36 kDa)
WC
SpaP (25 kDa)
c)
ΔinvA
wt
ΔinvA + pinvAFLAG
ΔinvA + pinvAΔCysFLAG
Sup
SipB (62kDa)
InvJ (36 kDa)
WC
SpaR (28 kDa)
d)
ΔspaS + pspaSΔCysFLAG
ΔinvA
wt
ΔspaS + pspaSFLAG
S17C
S23C
G53C
S66C
S115C
T131C
S162C
G166C
T205C
S231C
Sup
SipB (62kDa)
InvJ (36 kDa)
WC
SpaScc (11 kDa)

Fig. 5

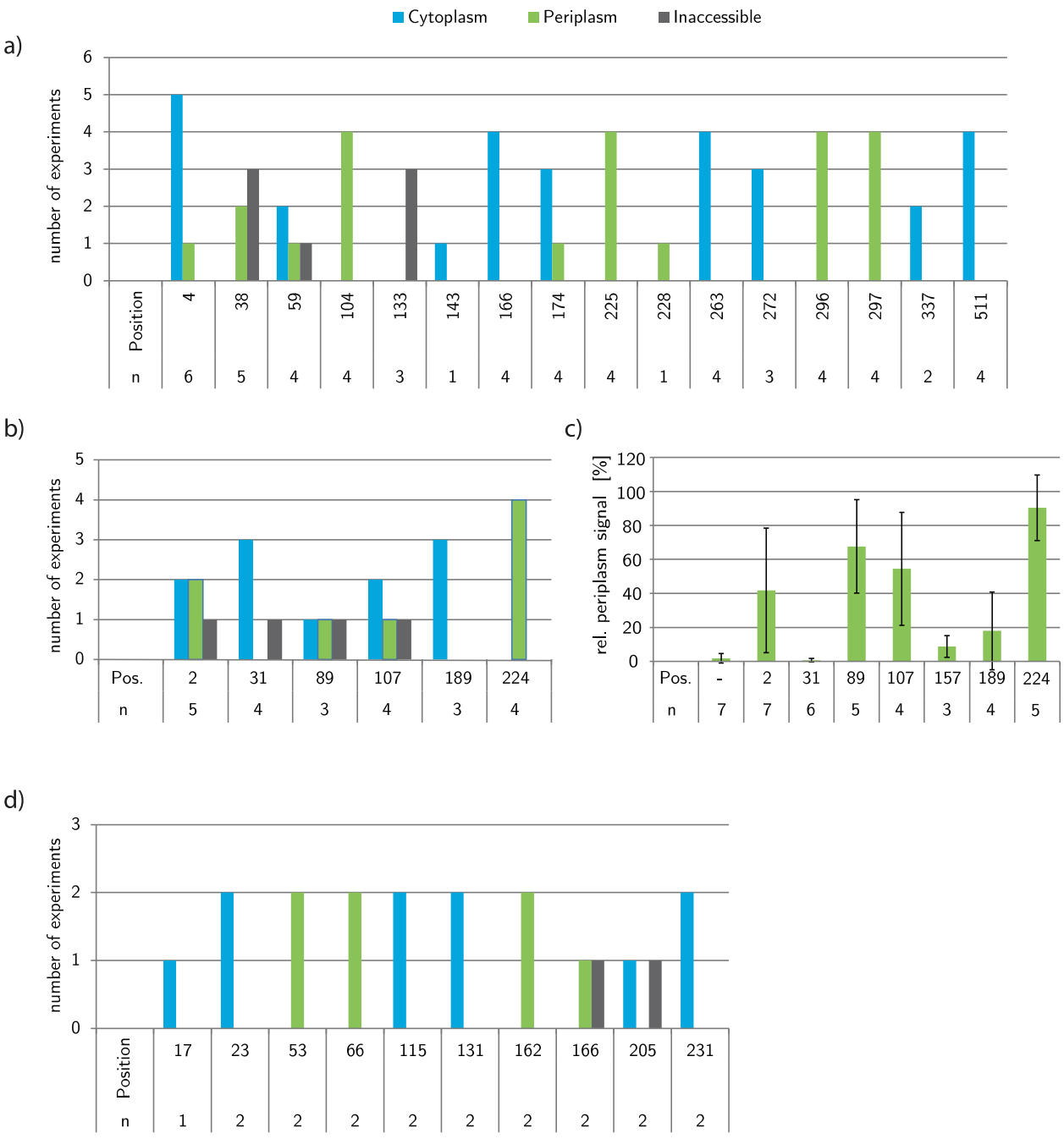

a)
Cytoplasm
Periplasm
Inaccessible
number of experiments
Position
n
4 38 59 104 133 143 166 174 225 228 263 272 296 297 337 511
6 5 4 4 3 1 4 4 4 1 4 3 4 4 2 4
b)
number of experiments
Pos.
n
2 31 89 107 189 224
5 4 3 4 3 4
c)
rel. periplasm signal [%]
Pos.
n
- 2 31 89 107 157 189 224
7 7 6 5 4 3 4 5
d)
number of experiments
Position
n
17 23 53 66 115 131 162 166 205 231
1 2 2 2 2 2 2 2 2 2

Fig. 6

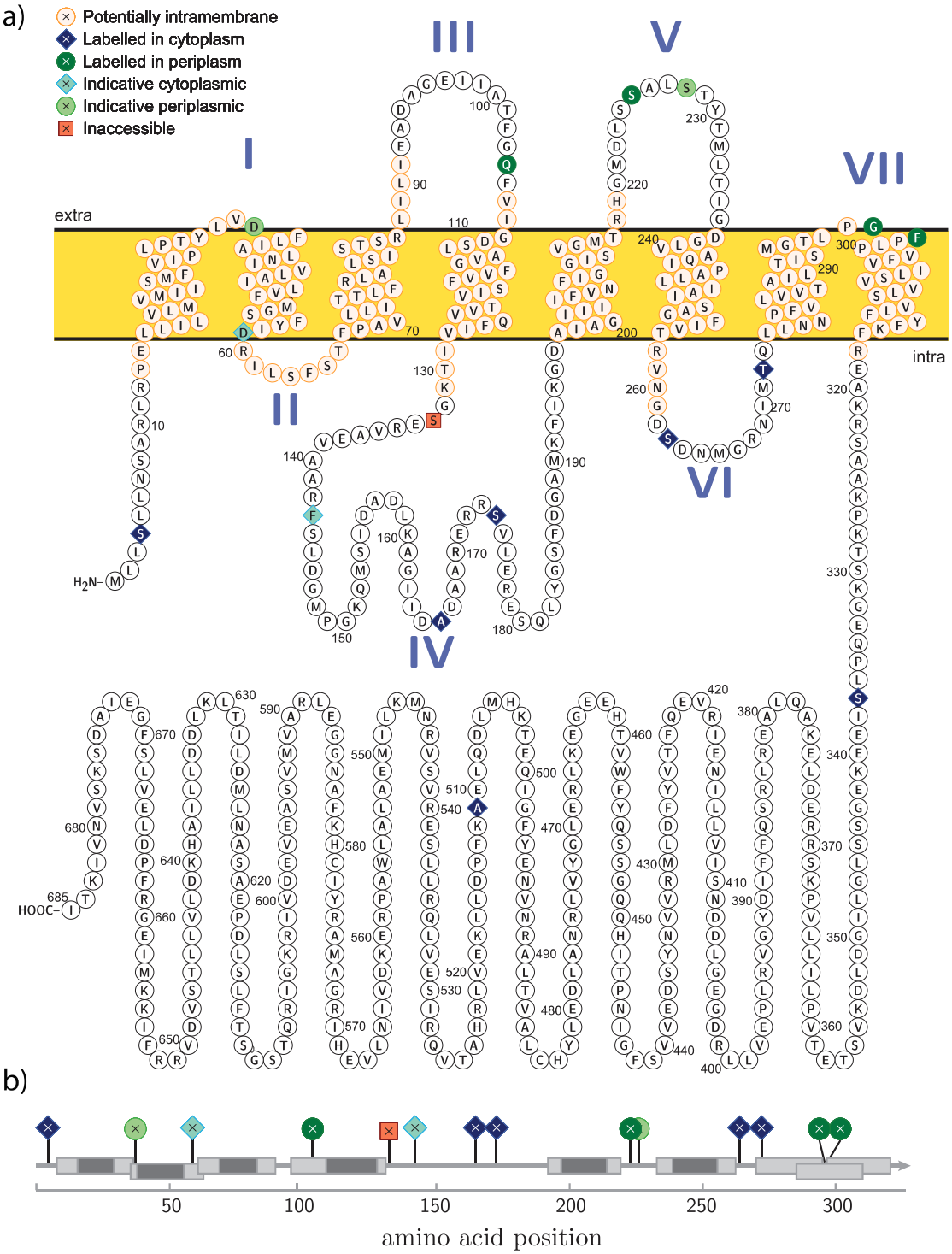

a)
Potentially intramembrane
Labelled in cytoplasm
Labelled in periplasm
Indicative cytoplasmic
Indicative periplasmic
Inaccessible
I
II
III
IV
V
VI
VII
extra
intra
H2N–
HOOC–
b)
50
100
150
200
250
300
amino acid position

Fig. 7

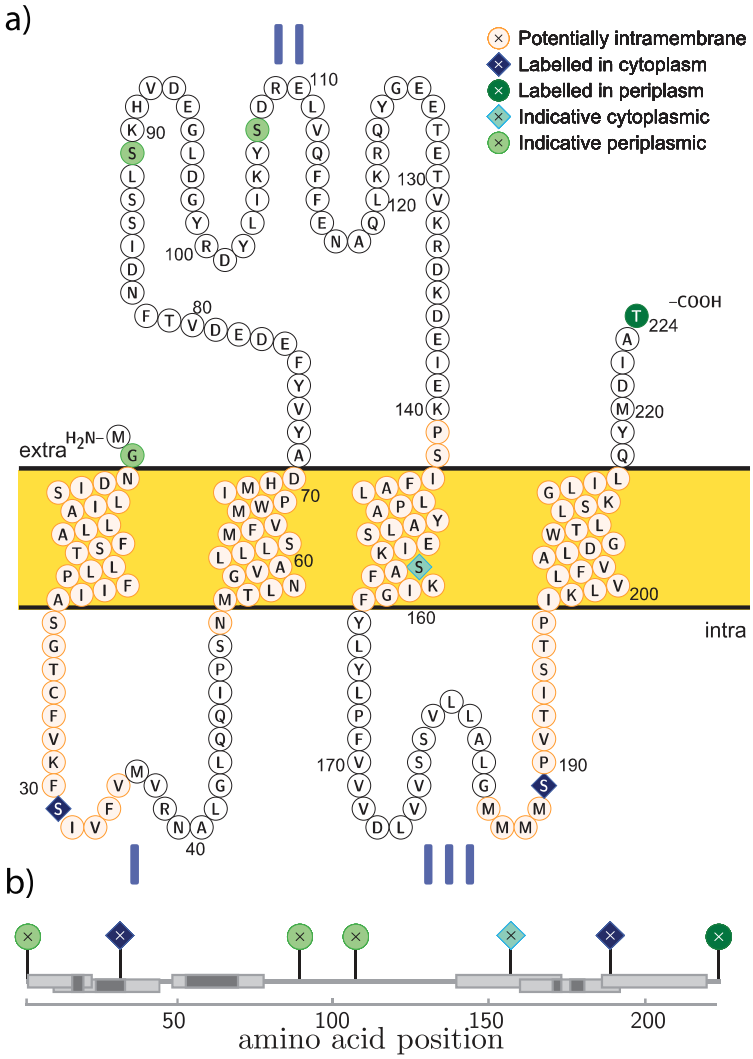

a)
Potentially intramembrane
Labelled in cytoplasm
Labelled in periplasm
Indicative cytoplasmic
Indicative periplasmic
extra
intra
H2N–
–COOH
b)
50
100
150
200
amino acid position

Fig. 8

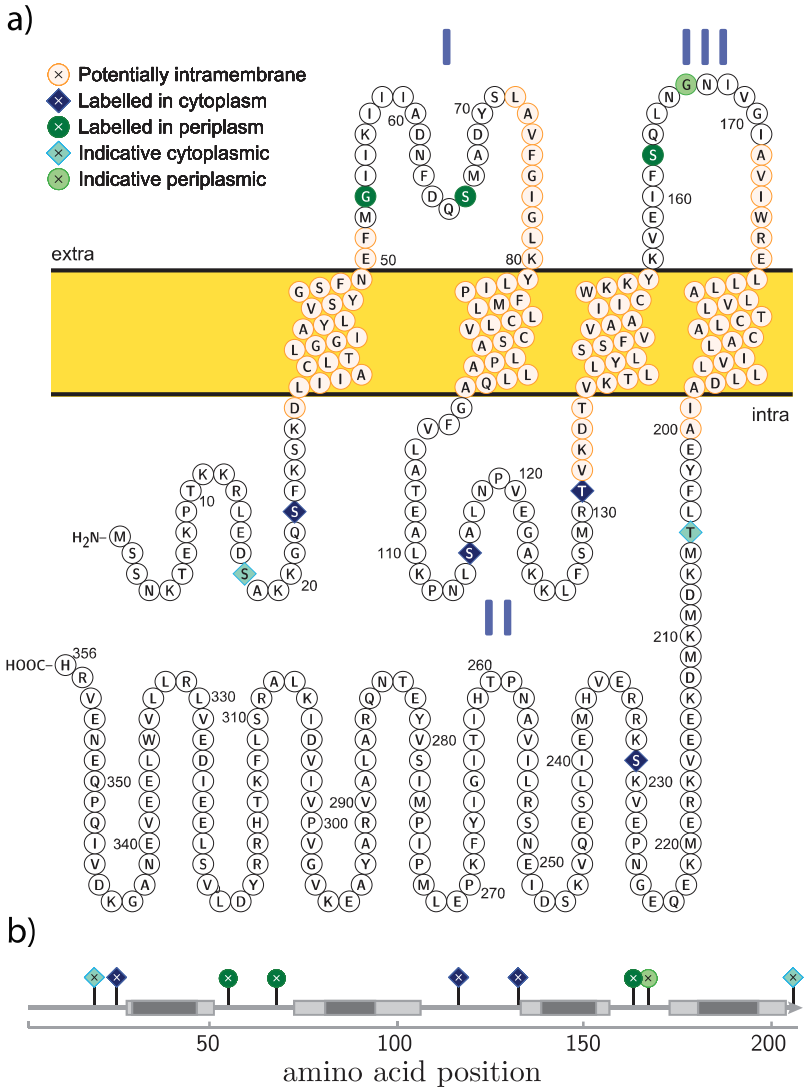

a)
Potentially intramembrane
Labelled in cytoplasm
Labelled in periplasm
Indicative cytoplasmic
Indicative periplasmic
extra
intra
H2N-
HOOC-
b)
50
100
150
200
amino acid position

Fig. 9

Table S1: Strains and plasmids

Salmonella	**Description**	**Reference**
SL1344		(Kröger et al. 2012)
SB762	wild type (SL1344, *flhD::tet*)	(Kaniga et al. 1995)
SB1892	Δ*spaPQRS*, *flhD::tet*, ara*hilA*	
SB1893	Δ*prgHIJK*, *flhD::tet*, ara*hilA*	
SB1901	Δ*invA*, *flhD::tet*	
SB1902	Δ*spaP*, *flhD::tet*	
SB1905	Δ*spaS*, *flhD::tet*	
Plasmid	**Description**	**Reference**
pSB3292	pBAD24, *hilA*	(Lara-tejero et al. 2011)
pSB3405	pTACO10, $invA^{FLAG}$	
pMIB5025	pTACO10, $invA^{FLAG}$, C484A, C575A	This study
pMIB5030	pMIB5025, S4C	This study
pMIB5031	pMIB5025, S133C	This study
pMIB5032	pMIB5025, A166C	This study
pMIB5033	pMIB5025, S225C	This study
pMIB5034	pMIB5025, S263C	This study
pMIB5035	pMIB5025, S337C	This study
pMIB5036	pMIB5025, D59C	This study
pMIB5037	pMIB5025, A100C	This study
pMIB5038	pMIB5025, A511C	This study
pMIB5039	pMIB5025, S272C	This study
pMIB5040	pMIB5025, G296C	This study
pMIB5111	pMIB5025, F297C	This study
pMIB5112	pMIB5025, D38C	This study
pMIB5113	pMIB5025, Q104C	This study
pMIB5114	pMIB5025, S174C	This study
pMIB5115	pMIB5025, S228C	This study
pMIB6131	pMIB5025, Y34C	This study
pMIB6132	pMIB5025, S143C	This study
pMIB6133	pMIB5025, S184C	This study
pSB3410	pTACO10, $spaP^{84FLAG}$	
pMIB5016	pTACO10, $spaP^{84FLAG}$, C26A	This study
pMIB5105	pMIB5016, S31C	This study
pMIB5104	pMIB5016, G2C	This study
pMIB5106	pMIB5016, S48C	This study
pMIB5107	pMIB5016, S89C	This study
pMIB5108	pMIB5016, S107C	This study
pMIB5109	pMIB5016, S189C	This study
pMIB5110	pMIB5016, T224C	This study
pMIB5364	pMIB5016, A71C	This study
pMIB5365	pMIB5016, A118C	This study
pMIB5366	pMIB5016, S142C	This study
pMIB5367	pMIB5016, S157C	This study
pSB3406	pTACO10, $spaS^{FLAG}$	
pMIB5191	pTACO10, $spaS^{FLAG}$, C33A, C89A, C92A, C152A, C186A, C189A	This study
pMIB5119	pMIB5191, S17C	This study
pMIB5127	pMIB5191, T205C	This study
pMIB5121	pMIB5191, G53C	This study
pMIB5120	pMIB5191, S23C	This study
pMIB5128	pMIB5191, S231C	This study
pMIB5124	pMIB5191, T131C	This study
pMIB5125	pMIB5191, S162C	This study
pMIB5126	pMIB5191, G166C	This study
pMIB5122	pMIB5191, S66C	This study
pMIB5123	pMIB5191, S115C	This study
pMIB5357	pMIB5191, N258A	This study
pMIB5094	pMIB5191, S17C, N258A	This study
pMIB5095	pMIB5191, T205C, N258A	This study
pMIB5096	pMIB5191, G53C, N258A	This study
pMIB5097	pMIB5191, S23C, N258A	This study
pMIB5098	pMIB5191, S231C, N258A	This study
pMIB5101	pMIB5191, T131C, N258A	This study
pMIB5102	pMIB5191, S162C, N258A	This study
pMIB5103	pMIB5191, G166C, N258A	This study
pMIB5358	pMIB5191, S66C, N258A	This study
pMIB5359	pMIB5191, S115C, N258A	This study

References

Kaniga, K., Tucker, S., Trollinger, D., and Galan, J.E. (1995). Homologs of the Shigella IpaB and IpaC invasins are required for Salmonella typhimurium entry into cultured epithelial cells. Journal of Bacteriology *177* (14): 3965–71.

Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., et al. (2012). The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proceedings of the National Academy of Sciences *109* (20): E1277–86.

Lara-tejero, M., Kato, J., Wagner, S., Liu, X., and Galan, J.E. (2011). A Sorting Platform Determines the Order of Protein Secretion in Bacterial Type III Systems. Science *331*: 1188.

Table ST2.: Summary of the predictions of TMH

InvA	TM1	TM2	TM3	TM4	TM5	TM6	TM7	TM8
TOPCONS	15-35	39-59	67-87	108-128	197-217	237-257	274-294	297-317
OCTOPUS	16-36	39-59	67-87	98-128	196-216	235-255	282-296	298-312
Philius	15-32	39-58	67-91	106-129	197-217	238-259	270-291	297-317
PolyPhobius	15-33	39-58	69-90	105-129	197-218	238-257	276-294	299-317
SCAMPI	15-35	38-58	61-81	110-130	197-217	239-259	274-294	297-317
SPOCTOPUS	16-36	39-59	67-87	98-128	196-216	235-255	282-296	298-312
TMHMMfix	13-30	35-57	69-91	106-128	197-219	239-261	274-291	296-315
ΔGapp	9-37	39-62	64-91	105-130	192-218	239-259	274-294	297-318
PredictProtein	15-37	39-58	64-83	106-129	197-218	235-256	286-309	-
All	16-30	39-57	69-81	110-128	197-216	239-255	286-291	-
at least one	9-37	35-62	61-91	98-130	192-219	235-261	274-309	296-318
diff. start	7	4	7	12	5	4	16	2
diff. end	7	5	10	2	3	6	18	6

SpaP	TM1	TM2	TM3	TM4	TM5	TM6
TOPCONS	2-22	24-44	50-70	143-163	165-185	197-217
OCTOPUS	6-20	22-36	52-72	143-163	165-185	200-220
Philius	7-33	-	49-69	-	163-187	195-217
PolyPhobius	6-37	-	49-69	159-173	175-189	193-216
SCAMPI	2-22	24-44	50-70	-	163-183	192-212
SPOCTOPUS	6-20	22-36	53-73	143-163	165-185	200-220
TMHMMfix	15-37	-	50-69	-	161-183	-
ΔGapp	5-35	-	50-75	-	163-193	194-211
PredictProtein	9-34	-	49-67	141-154	161-183	185-205
All	15-20	-	53-67	-	175-183	-
at least one	2-37	22-44	49-75	141-173	161-193	185-217
diff. start	13	2	4	18	6	15
diff. end	17	8	12	19	10	12

SpaQ	TM1	TM2
TOPCONS	16-36	51-71
OCTOPUS	16-36	52-72
Philius	12-36	49-70
PolyPhobius	12-37	48-71
SCAMPI	16-36	48-68
SPOCTOPUS	16-36	52-72
TMHMMfix	12-34	49-71
ΔGapp	11-36	46-72
PredictProtein	12-36	47-68
All	16-34	52-68
at least one	11-37	46-72
diff. start	5	6
diff. end	3	4

SpaR	TM1	TM2	TM3	TM4	TM5	TM6
TOPCONS	12-32	42-62	74-94	129-149	178-198	216-236
OCTOPUS	12-32	42-62	74-94	129-149	178-198	216-236

Philius	9-33	40-59	73-97	128-143	181-202	211-234
PolyPhobius	11-33	41-62	72-99	127-151	177-202	211-235
SCAMPI	13-33	39-59	78-98	127-147	181-201	216-236
SPOCTOPUS	12-32	42-62	74-94	129-149	178-198	216-236
TMHMMfix	10-32	37-59	79-101	128-150	181-203	210-232
ΔGapp	12-38	41-59	65-95	-	181-206	211-235
PredictProtein	14-33	37-54	70-103	126-149	177-194	212-235
All	14-32	42-54	79-94	-	181-194	216-232
at least one	9-38	37-62	65-103	126-150	177-206	210-236
diff. start	5	5	14	3	4	6
diff. end	6	3	9	8	8	4

SpaS	TM1	TM2	TM3	TM4
TOPCONS	29-49	81-101	136-156	178-198
OCTOPUS	29-49	73-103	136-156	177-197
Philius	29-48	81-104	138-157	178-200
PolyPhobius	29-47	77-101	135-154	178-200
SCAMPI	29-49	81-101	135-155	181-201
SPOCTOPUS	29-49	73-101	136-156	178-198
TMHMMfix	29-51	72-94	136-153	178-200
ΔGapp	28-48	72-106	138-156	171-204
PredictProtein	28-48	74-100	132-153	172-196
All	29-47	81-94	138-153	181-196
at least one	28-51	72-106	132-157	171-204
diff. start	1	9	6	10
diff. end	4	12	4	8

TMHMM2 output

Protein	Tool	Position	amino acid	amino acid end
InvA				
sp|P0A1I3|INVA_SALTY	TMHMM2.0	inside	1	12
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	13	30
sp|P0A1I3|INVA_SALTY	TMHMM2.0	outside	31	34
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	35	57
sp|P0A1I3|INVA_SALTY	TMHMM2.0	inside	58	68
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	69	91
sp|P0A1I3|INVA_SALTY	TMHMM2.0	outside	92	105
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	106	128
sp|P0A1I3|INVA_SALTY	TMHMM2.0	inside	129	196
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	197	219
sp|P0A1I3|INVA_SALTY	TMHMM2.0	outside	220	238
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	239	261
sp|P0A1I3|INVA_SALTY	TMHMM2.0	inside	262	273
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	274	291
sp|P0A1I3|INVA_SALTY	TMHMM2.0	outside	292	295
sp|P0A1I3|INVA_SALTY	TMHMM2.0	TMhelix	296	315
sp|P0A1I3|INVA_SALTY	TMHMM2.0	inside	316	685
SpaP				
sp|P40700|SPAP_SALTY	TMHMM2.0	outside	1	14

sp|P40700|SPAP_SALTY	TMHMM2.0	TMhelix	15	37
sp|P40700|SPAP_SALTY	TMHMM2.0	inside	38	49
sp|P40700|SPAP_SALTY	TMHMM2.0	TMhelix	50	69
sp|P40700|SPAP_SALTY	TMHMM2.0	outside	70	160
sp|P40700|SPAP_SALTY	TMHMM2.0	TMhelix	161	183
sp|P40700|SPAP_SALTY	TMHMM2.0	inside	184	224

SpaQ

sp|P0A1L7|SPAQ_SALTY	TMHMM2.0	inside	1	11
sp|P0A1L7|SPAQ_SALTY	TMHMM2.0	TMhelix	12	34
sp|P0A1L7|SPAQ_SALTY	TMHMM2.0	outside	35	48
sp|P0A1L7|SPAQ_SALTY	TMHMM2.0	TMhelix	49	71
sp|P0A1L7|SPAQ_SALTY	TMHMM2.0	inside	72	86

SpaR

sp|P40701|SPAR_SALTY	TMHMM2.0	outside	1	9
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	10	32
sp|P40701|SPAR_SALTY	TMHMM2.0	inside	33	36
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	37	59
sp|P40701|SPAR_SALTY	TMHMM2.0	outside	60	78
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	79	101
sp|P40701|SPAR_SALTY	TMHMM2.0	inside	102	127
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	128	150
sp|P40701|SPAR_SALTY	TMHMM2.0	outside	151	180
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	181	203
sp|P40701|SPAR_SALTY	TMHMM2.0	inside	204	209
sp|P40701|SPAR_SALTY	TMHMM2.0	TMhelix	210	232
sp|P40701|SPAR_SALTY	TMHMM2.0	outside	233	263

SpaS

sp|P40702|SPAS_SALTY	TMHMM2.0	outside	1	28
sp|P40702|SPAS_SALTY	TMHMM2.0	TMhelix	29	51
sp|P40702|SPAS_SALTY	TMHMM2.0	inside	52	71
sp|P40702|SPAS_SALTY	TMHMM2.0	TMhelix	72	94
sp|P40702|SPAS_SALTY	TMHMM2.0	outside	95	135
sp|P40702|SPAS_SALTY	TMHMM2.0	TMhelix	136	153
sp|P40702|SPAS_SALTY	TMHMM2.0	inside	154	177
sp|P40702|SPAS_SALTY	TMHMM2.0	TMhelix	178	200
sp|P40702|SPAS_SALTY	TMHMM2.0	outside	201	356

```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
  Cedric Notredame
  CPU TIME:0 sec.
  SCORE=935
  *
   BAD  AVG  GOOD
  *
  sp|P40700|SPAP_    :   91
  tr|G0CIR4|G0CIR    :   92
  sp|P54700|FLIP_    :   92
  tr|Q9ZA77|Q9ZA7    :   93
  sp|P0A1L3|SPAP_    :   92
  tr|A5YJ93|A5YJ9    :   92
  tr|A0A0H2V084|A    :   93
  tr|Q3SBB6|Q3SBB    :   92
  tr|A0A023KQG0|A    :   92
  tr|A0A024BZY8|A    :   92
  cons               :   93

  sp|P40700|SPAP_    MG----------------------------------------NDISLIALLAFSTLLPFIIASG
  tr|G0CIR4|G0CIR    MQ------------------------------------MPDVGSLLLVVIMLGLLPFAAMVV
  sp|P54700|FLIP_    MRRLLFLSLAGLWLFS-PAAAAQLPGLISQPLA-GGGQSWSLSVQTLVFITSLTFLPAILLMM
  tr|Q9ZA77|Q9ZA7    MIQ-----------------------------------LPDEINLIIILSLLTLLPLVSVMA
  sp|P0A1L3|SPAP_    ML--------------------------------------SDMSLIATLSFFTLLPFLVAAG
  tr|A5YJ93|A5YJ9    M-----------------------------------------LALFLGSLSLIPFLLIVC
  tr|A0A0H2V084|A     -----------------------------------------MQTLVFITSLTFIPAILLMM
  tr|Q3SBB6|Q3SBB    MS--------------------------------------NSISLIAILTLFTLLPFIIASG
  tr|A0A023KQG0|A    MRRLLSVAPVLLWLVT-PLAFAQLPGITSQPLP-GGGQSWSLPVQTLVFITSLTFIPAILLMM
  tr|A0A024BZY8|A    MRFFIFLILICPLICPLMSADSALPSVNLSLNAPNDPKQLVTTLNVIALLTLLVLAPSLILVM

  cons                                                       :    : *

  sp|P40700|SPAP_    TCFVKFSIVFVMVRNALGLQQIPSNMTLNGVALLLSMFVMWPIMHDAYVYFEDEDVTFNDISS
  tr|G0CIR4|G0CIR    TSYTKIVVVLGLLRNAIGVQQVPPNMVLNGVALLVSCFVMAPVGMEAFKAAQNYSP-GAENSR
  sp|P54700|FLIP_    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEQKIS
  tr|Q9ZA77|Q9ZA7    TSFVKFAVVFSLLRNALGVQQIPPNMAMYGLAIILSLYVMAPVGFATQDYLQANEVSLTNIES
  sp|P0A1L3|SPAP_    TCYIKFSIVFVMVRNALGLQQVPSNMTLNGIALIMALFVMKPIIEAGYENYLNGPQKFDTISD
  tr|A5YJ93|A5YJ9    TAFLKIAMTLLITRNAIGVQQVPPNMALYGIALAATMFVMAPVAHDIQQRVHEHPLELSNADK
  tr|A0A0H2V084|A    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEEKIS
  tr|Q3SBB6|Q3SBB    TYFIKFSIVFVIVRNALGLQQVPSNMTLNGVALLLSMFVMMPVGTEIYYNSQNENLSFNNVAS
  tr|A0A023KQG0|A    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEEKIS
  tr|A0A024BZY8|A    TSFTRLIVVFSFLRTALGTQQTPPTQILVSLSLILTFFIMEPSLKKAYDTGIKP-Y-MDKKIS

  cons               * : :: :.: : *.*:*  . *..  : .::: : ::* *
```

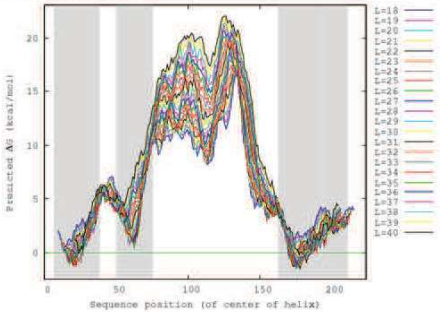

sp|P40700|SPAP_ LSKHVDEGLDGYRDYLIKYSDRELVQFFENAQLKRQYGEETETVKRDKDEIEKPSIFALLPAY
tr|G0CIR4|G0CIR VVVLLDACREPFRQFLLKHTREREKAFFIRSAQQIWP-KD------KADTLKPDDLLVLAPAF
sp|P54700|FLIP_ MQEALDKGAQPLRAFMLRQTREADLALFARLANSGPL-Q----------GPEAVPMRILLPAY
tr|Q9ZA77|Q9ZA7 VEKFFDEGLAPYRMFLKQHIQAQEYSFFVDSTKQLWP-KQ------YADRLESDSLFILLPAF
sp|P0A1L3|SPAP_ IVRFSDSGLMEYKQYLKKHTDLELARFFQRSEEENAD-LK-------SAENNDYSLFSLLPAY
tr|A5YJ93|A5YJ9 LQSSLKVVIEPLQRFMTRNTDPDVVAHLLENTQRMWP-KE------MADQANKNDLLLAIPAF
tr|A0A0H2V084|A MQEALEKGAQPLREFMLRQTREADLGLFARLANTGPL-Q----------GPEAVPMRILLPAY
tr|Q3SBB6|Q3SBB VVNFVETGMSGYKSYLIKYSEPELVSFFEKIQKVNSS-EDNE----EIIDDDNISIFSLLPAY
tr|A0A023KQG0|A MQEALEKGAQPLREFMLRQTREADLGLFARLANTGPL-Q----------GPEAVPMRILLPAY
tr|A0A024BZY8|A YTEAFEKSALPFKEFMLKNTREKDLALFFRIRNLPNP-K----------TPDEVSLSVLIPAF
cons
sp|P40700|SPAP_ ALSEIKSAFKIGFYLYLPFVVVDLVVSSVLLALGMMMMSPVTISTPIKLVLFVALDGWTLLSK
tr|G0CIR4|G0CIR TLSELTEAFRIGFLLYLVFIVIDLVVANALMAMGLSQVTPTNVAIPFKLLLFVALDGWSMLIH
sp|P54700|FLIP_ VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLMG
tr|Q9ZA77|Q9ZA7 TVSELTRAFEIGFLIYLPFIVIDLVISNILLAMGMMMVSPMTISLPFKLLLFVLLDGWTRLTH
sp|P0A1L3|SPAP_ ALSEIKDAFKIGFYLYLPFVVVDLVISSILLALGMMMMSPITISVPIKLVLFVALDGWGILSK
tr|A5YJ93|A5YJ9 VLSELQAGFEIGFLIYIPFIVIDLIVSNLLLALGMQMVSPMTLSLPLKLLLFVLVSGWSRLLD
tr|A0A0H2V084|A VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLVG
tr|Q3SBB6|Q3SBB ALSEIKSAFIIGFYIYLPFVVVDLVISSVLLTLGMMMMSPVIISTPIKLILFVAMDGWTMLSK
tr|A0A023KQG0|A VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLVG
tr|A0A024BZY8|A MISELKTAFQIGFLLYLPFLVIDMVISSILMAMGMMMLPPVMISLPFKILVFILVDGFNLLTE
cons
sp|P40700|SPAP_ GLILQYMDIA--T
tr|G0CIR4|G0CIR GLVLSYR------
sp|P54700|FLIP_ SLAQSFY-S----
tr|Q9ZA77|Q9ZA7 GLVISYGG-----
sp|P0A1L3|SPAP_ ALIEQYINIP--A
tr|A5YJ93|A5YJ9 SLFYSYM------
tr|A0A0H2V084|A SLAQSFY-S----
tr|Q3SBB6|Q3SBB GLILQYFDLSINP
tr|A0A023KQG0|A SLAQSFY-S----
tr|A0A024BZY8|A NLVASFKMV----
cons

>sp|P40700|SPAP_SALTY Surface presentation of antigens protein SpaP OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNspaP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
7-38	32	-1.176	LIALLAFSTLLPFIIASGTCFVKFSIVFVMVR	[analyze]
50-75	26	0.482	MTLNGVALLLSMFVIIWPIMHDAYVYF	[analyze]
183-193	31	-1.455	FYLYLPFVVVDLVVSSVLLALGMMMMSPVTI	[analyze]
194-211	18	1.936	STPIKLVLFVALDGWTLL	[analyze]

Download plot in EPS format (new)

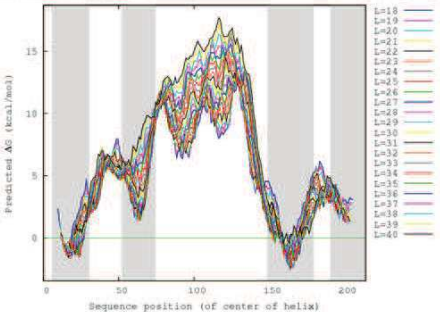


>tr|G0CIR4|G0CIR4_XANCA Type III secretion protein HrcR OSXanthomonas campestris pv raphani 756C GNhrcR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
6-30	25	-1.605	VGSLLLVVIMLGLLPFAAMVVTSYT	[analyze]
52-74	23	1.385	MVLNGVALLVSCFVMAPVGMEAF	[analyze]
149-179	31	-2.436	LTEAFRIGFLLYLVFIVDLVVANALMAMGL	[analyze]
191-214	24	1.130	FKLLLFVALDGWSMLHGLVLSYR	[analyze]

Download plot in EPS format (new)

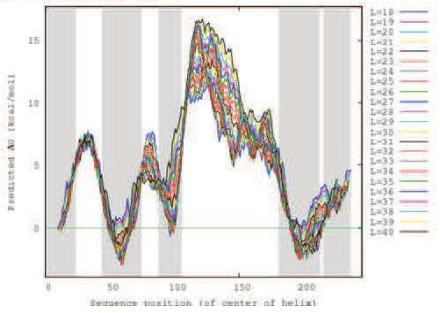


>sp|P54700|FLIP_SALTY Flagellar biosynthetic protein FliP OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNfliP PE1 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
1-23	23	-0.310	MRRLLFLSLAGLWLFSPAAAAGL	[analyze]
44-74	31	-2.958	TLVFITSLTFLPAILLMMTSFTRIIVFGLL	[analyze]
88-105	18	-0.561	VLLGLALFLTFFMSPVI	[analyze]
181-212	32	-2.538	TAFQIGFTIFIPFLIIDLVIASVLMALGMMMV	[analyze]
215-236	22	1.187	ATIALPFKLMLFVLVDGWQLLM	[analyze]

Download plot in EPS format (new)

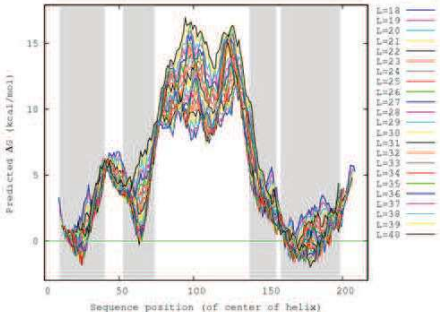


>tr|Q9ZA77|Q9ZA77_YEREN Type III secretion system protein OSYersinia enterocolitica GNyscR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
16-40	31	-1.753	LIILSLLTLLPLVSVIATSFVKFAVVFSLL	[analyze]
53-74	22	-0.231	MAMYGLAILSLYVMAPVGFAT	[analyze]
138-155	18	1.902	SLFILLPAFTVSELTRAF	[analyze]
159-198	40	-2.035	FLIYLPFIVDLVISNILLAMGMMMVSPMTISLPFKLLLF	[analyze]

Download plot in EPS format (new)

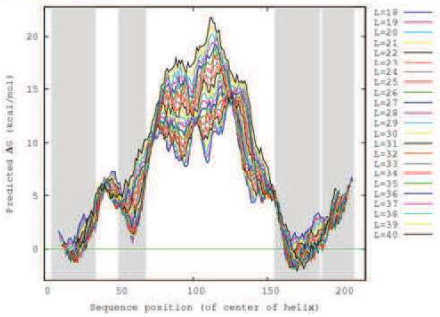


>sp|P0A1L3|SPAP_SHIFL Surface presentation of antigens protein SpaP OSShigella flexneri GNspaP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
5-34	30	-1.226	MSLIATLSFFTLLPFLVAAGTCYIKFSIVF	[analyze]
50-68	19	0.659	MTLNGIALMALFVMKPII	[analyze]
155-185	31	-1.998	FYLYLPFVVVDLVISSLLALGMMMMSPITI	[analyze]
187-208	22	2.379	VPIKLVLFVALDGWGILSKALI	[analyze]

Download plot in EPS format (new)

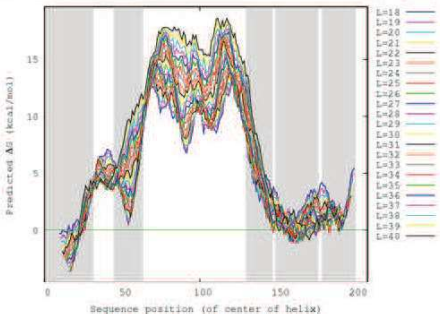


>tr|A5YJ93|A5YJ93_PSESY HrcR OSPseudomonas syringae pv syringae GNhrcR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
1-31	31	-3.638	MLALFLGSLSLIPFLLIVCTAFLKIAMTLLI	[analyze]
45-63	19	0.452	MALYGIALAATMFVMAPVA	[analyze]
130-147	18	2.552	DLLLAIPAFVLSELQAGF	[analyze]
149-176	28	-1.165	IGFLIYIPFIVIDLIVSNLLLALGMQMV	[analyze]
179-200	22	-0.085	MTLSLPLKLLLFVLVSGWSRLL	[analyze]

Download plot in EPS format (new)

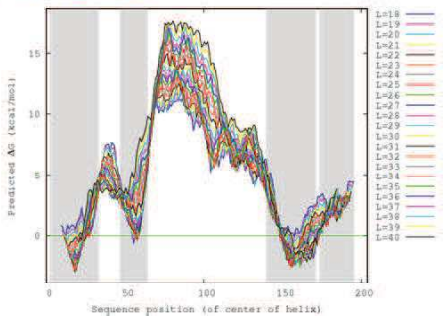


>tr|A0A0H2V084|A0A0H2V084_SHIFL Flagellar biosynthetic protein FliP OSShigella flexneri GNfliP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
3-33	31	-2.933	TLVFITSLTFIPAILLMMTSFTRIIVFGLL	[analyze]
47-64	18	-0.561	VLLGLALFLTFFIMSPVI	[analyze]
140-171	32	-2.538	TAFQIGFTIFIPFLIIDLVIASVLMALGMMMV	[analyze]
174-195	22	1.167	ATIALPFKLMLFVLVDGWQLLV	[analyze]

Download plot in EPS format (new)

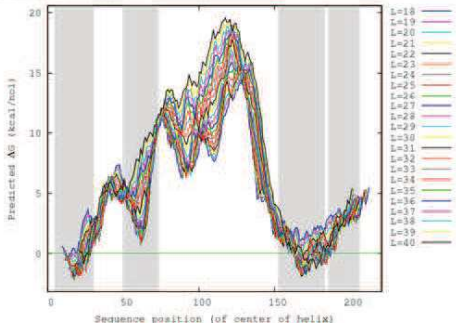


>tr|Q3SBB6|Q3SBB6_ECOLX EpaP OSEscherichia coli GNepaP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
5-30	26	-2.112	ISLIAILTLFTLLPFIIASGTYFIKF	[analyze]
50-73	24	0.747	MTLNGVALLLSMFVMMPVGTEIYY	[analyze]
153-183	31	-1.872	AFIIGFYIYLPFVVVDLVISSVLLTLGMMMM	[analyze]
186-206	21	1.980	VIISTPIKLILFVAMDGWTML	[analyze]

Download plot in EPS format (new)

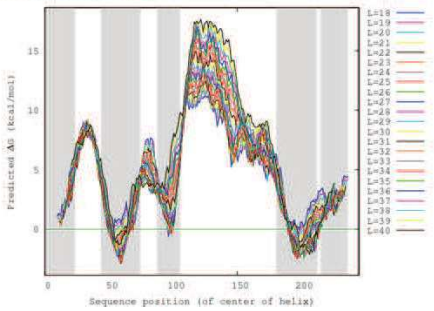


>tr|A0A023KQG0|A0A023KQG0_ECOLX Flagellar biosynthetic protein FliP OSEscherichia coli GNfliP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
1-23	23	0.231	MRRLLSVAPVLLWLVTPLAFAQL	[analyze]
44-74	31	-2.933	TLVFITSLTFIPAILLMMTSFTRIIVFGLL	[analyze]
88-105	18	-0.561	VLLGLALFLTFFIMSPVI	[analyze]
181-212	32	-2.538	TAFQIGFTIFIPFLIIDLVIASVLMALGMMMV	[analyze]
215-236	22	1.167	ATIALPFKLMLFVLVDGWQLLV	[analyze]

Download plot in EPS format (new)

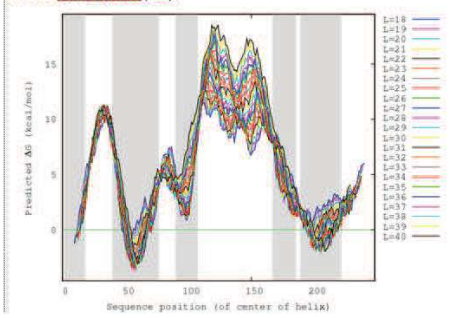


>tr|A0A024BZY8|A0A024BZY8_HELPX Flagellar biosynthetic protein FliP OSHelicobacter pylori GNfliP PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
1-18	18	-1.242	MRFPIFLILICPLICPLM	[analyze]
40-76	37	-3.678	LVTTLNVIALLTLLVLAPSLILVMTSFTRLIVVFSFL	[analyze]
90-107	18	0.953	ILVSLSLILTFFIMEPSL	[analyze]
167-185	19	2.560	VSLSVLIPAFMISELKTAF	[analyze]
189-221	33	-2.105	FLLYLPFLVIDMVISSILMAMGMMMLPPVIMSL	[analyze]

Download plot in EPS format (new)



```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
  Cedric Notredame
  CPU TIME:0 sec.
  SCORE=935
  *
  - entirely conserved columns marked with a star (*)
  - columns comprising amino acids of same size and hydropathy marked with a colon (:)
  - columns comprising amino acids of similar size or evolutionary preserved hydropathy marked with a period (.)
  - positive charged amino acids (Arg and Lys) marked in blue
  - negative charged amino acids (Asp and Glu) marked in orange
  - transmembrane domains marked in grey

  sp|P40700|SPAP_    MG----------------------------------------NDISLIALLAFSTLLPFIIASG
  tr|G0CIR4|G0CIR    MQ----------------------------------------MPDVGSLLLVVIMLGLLPFAAMVV
  sp|P54700|FLIP_    MRRLLFLSLAGLWLFS-PAAAAQLPGLISQPLA-GGGQSWSLSVQTLVFITSLTFLPAILLMM
  tr|Q9ZA77|Q9ZA7    MIQ---------------------------------------LPDEINLIIILSLLTLLPLVSVMA
  sp|P0A1L3|SPAP_    ML----------------------------------------SDMSLIATLSFFTLLPFLVAAG
  tr|A5YJ93|A5YJ9    M-----------------------------------------LALFLGSLSLIPFLLIVC
  tr|A0A0H2V084|A    ----------------------------------------MQTLVFITSLTFIPAILLMM
  tr|Q3SBB6|Q3SBB    MS----------------------------------------NSISLIAILTLFTLLPFIIASG
  tr|A0A023KQG0|A    MRRLLSVAPVLLWLVT-PLAFAQLPGITSQPLP-GGGQSWSLPVQTLVFITSLTFIPAILLMM
  tr|A0A024BZY8|A    MRFFIFLILICPLICPLMSADSALPSVNLSLNAPNDPKQLVTTLNVIALLTLLVLAPSLILVM

  cons                                                       :    : *

  sp|P40700|SPAP_    TCFVKFSIVFVMVRNALGLQQIPSNMTLNGVALLLSMFVMWPIMHDAYVYFEDEDVTFNDISS
  tr|G0CIR4|G0CIR    TSYTKIVVVLGLLRNAIGVQQVPPNMVLNGVALLVSCFVMAPVGMEAFKAAQNYSP-GAENSR
  sp|P54700|FLIP_    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEQKIS
  tr|Q9ZA77|Q9ZA7    TSFVKFAVVFSLLRNALGVQQIPPNMAMYGLAIILSLYVMAPVGFATQDYLQANEVSLTNIES
  sp|P0A1L3|SPAP_    TCYIKFSIVFVMVRNALGLQQVPSNMTLNGIALIMALFVMKPIIEAGYENYLNGPQKFDTISD
  tr|A5YJ93|A5YJ9    TAFLKIAMTLLITRNAIGVQQVPPNMALYGIALAATMFVMAPVAHDIQQRVHEHPLELSNADK
  tr|A0A0H2V084|A    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEEKIS
  tr|Q3SBB6|Q3SBB    TYFIKFSIVFVIVRNALGLQQVPSNMTLNGVALLLSMFVMMPVGTEIYYNSQNENLSFNNVAS
  tr|A0A023KQG0|A    TSFTRIIIVFGLLRNALGTPSAPPNQVLLGLALFLTFFIMSPVIDKIYVDAYQP-F-SEEKIS
  tr|A0A024BZY8|A    TSFTRLIVVFSFLRTALGTQQTPPTQILVSLSLILTFFIMEPSLKKAYDTGIKP-Y-MDKKIS

  cons               * : :: :.: : *.*:*  . *..  : .:::  : ::* *

  sp|P40700|SPAP_    LSKHVDEGLDGYRDYLIKYSDRELVQFFENAQLKRQYGEETETVKRDKDEIEKPSIFALLPAY
  tr|G0CIR4|G0CIR    VVVLLDACREPFRQFLLKHTRERERKAFFIRSAQQIWP-KD------KADTLKPDDLLVLAPAF
  sp|P54700|FLIP_    MQEALDKGAQPLRAFMLRQTREADLALFARLANSGPL-Q----------GPEAVPMRILLPAY
  tr|Q9ZA77|Q9ZA7    VEKFFDEGLAPYRMFLKQHIQAQEYSFFVDSTKQLWP-KQ------YADRLESDSLFILLPAF
  sp|P0A1L3|SPAP_    IVRFSDSGLMEYKQYLKKHTDLELARFFQRSEEENAD-LK-------SAENNDYSLFSLLPAY
  tr|A5YJ93|A5YJ9    LQSSLKVVIEPLQRFMTRNTDPDVVAHLLENTQRMWP-KE------MADQANKNDLLLAIPAF
  tr|A0A0H2V084|A    MQEALEKGAQPLREFMLRQTREADLGLFARLANTGPL-Q----------GPEAVPMRILLPAY
  tr|Q3SBB6|Q3SBB    VVNFVETGMSGYKSYLIKYSEPELVSFFEKIQKVNSS-EDNE----EIIDDDNISIFSLLPAY
  tr|A0A023KQG0|A    MQEALEKGAQPLREFMLRQTREADLGLFARLANTGPL-Q----------GPEAVPMRILLPAY
  tr|A0A024BZY8|A    YTEAFEKSALPFKEFMLKNTREKDLALFFRIRNLPNP-K----------TPDEVSLSVLIPAF

  cons                    .       : :: :           :                    .    :     **:

  sp|P40700|SPAP_    ALSEIKSAFKIGFYLYLPFVVVDLVVSSVLLALGMMMMSPVTISTPIKLVLFVALDGWTLLSK
  tr|G0CIR4|G0CIR    TLSELTEAFRIGFLLYLVFIVIDLVVANALMAMGLSQVTPTNVAIPFKLLLFVALDGWSMLIH
  sp|P54700|FLIP_    VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLMG
  tr|Q9ZA77|Q9ZA7    TVSELTRAFEIGFLIYLPFIVIDLVISNILLAMGMMMVSPMTISLPFKLLLFVLLDGWTRLTH
  sp|P0A1L3|SPAP_    ALSEIKDAFKIGFYLYLPFVVVDLVISSILLALGMMMMSPITISVPIKLVLFVALDGWGILSK
  tr|A5YJ93|A5YJ9    VLSELQAGFEIGFLIYIPFIVIDLIVSNLLLALGMQMVSPMTLSLPLKLLLFVLVSGWSRLLD
  tr|A0A0H2V084|A    VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLVG
  tr|Q3SBB6|Q3SBB    ALSEIKSAFIIGFYIYLPFVVVDLVISSVLLTLGMMMMSPVIISTPIKLILFVAMDGWTMLSK
  tr|A0A023KQG0|A    VTSELKTAFQIGFTIFIPFLIIDLVIASVLMALGMMMVPPATIALPFKLMLFVLVDGWQLLVG
  tr|A0A024BZY8|A    MISELKTAFQIGFLLYLPFLVIDMVISSILMAMGMMMLPPVMISLPFKILVFILVDGFNLLTE

  cons                  **:  .* *** ::: *:::*:::: .*:::*:  :.*  :: *:*:::*: :.*:  *

  sp|P40700|SPAP_    GLILQYMDIA--T
  tr|G0CIR4|G0CIR    GLVLSYR------
  sp|P54700|FLIP_    SLAQSFY-S----
  tr|Q9ZA77|Q9ZA7    GLVISYGG-----
  sp|P0A1L3|SPAP_    ALIEQYINIP--A
  tr|A5YJ93|A5YJ9    SLFYSYM------
  tr|A0A0H2V084|A    SLAQSFY-S----
  tr|Q3SBB6|Q3SBB    GLILQYFDLSINP
  tr|A0A023KQG0|A    SLAQSFY-S----
  tr|A0A024BZY8|A    NLVASFKMV----

  cons                *  .:
```

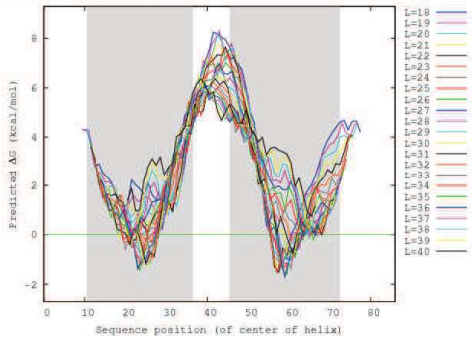
```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
 Cedric Notredame
 CPU TIME:0 sec.
 SCORE=928
 *
  BAD AVG GOOD
 *
 sp|P0A1L7|SPAQ_   :  94
 tr|G0CIR3|G0CIR   :  94
 sp|P0A1L5|FLIQ_   :  94
 tr|Q7BFA7|Q7BFA   :  94
 sp|P0A1M4|SPAQ_   :  94
 tr|Q60237|Q6023   :  92
 sp|P0AC10|FLIQ_   :  94
 tr|Q2TJ91|Q2TJ9   :  94
 tr|C3T4R7|C3T4R   :  94
 sp|P0A0S3|FLIQ_   :  91
 cons              :  92

 sp|P0A1L7|SPAQ_   MD--DLVFAGNKALYLVLILSGWPTIVATIIGLLVGLFQTVTQLQEQTLPFGIKLLGVCLCLF
 tr|G0CIR3|G0CIR   MDHDDLVRLTSEALLLCLKVSLPVVGVAALAGLLIAFLQAVMSLQDASISFALKLVVVVAAIA
 sp|P0A1L5|FLIQ_   MTPESVMMMGTEAMKVALALAAPLLLVALITGLIISILQAATQINEMTLSFIPKIVAVFIAII
 tr|Q7BFA7|Q7BFA   MSQGDIIHFTSQALWLVLVLSMPPVLVAAVVGTLVSLVQALTQIQEQTLGFVIKLIAVVVTLF
 sp|P0A1M4|SPAQ_   MS--DIVYMGNKALYLILIFSLWPVGIATVIGLSIGLLQTVTQLQEQTLPFGIKLIGVSISLL
 tr|Q60237|Q6023   ME--A-LALFKQGMFLVVILTAPPLAVAVLVGVVTSLLQALMQIQDQTLPFGIKLGAVGLTLA
 sp|P0AC10|FLIQ_   MTPESVMMMGTEAMKVALALAAPLLLVALVTGLIISILQAATQINEMTLSFIPKIIAVFIAII
 tr|Q2TJ91|Q2TJ9   MD--DIVFAGNRALYLILVMSAGPIAVATFVGLLVGLFQTVTQLQEQTLPFGVKLLCVSICFF
 tr|C3T4R7|C3T4R   MTPESVMMMGTEAMKVALALAAPLLLVALVTGLIISILQAATQINEMTLSFIPKIIAVFIAII
 sp|P0A0S3|FLIQ_   MES-QLMKLAIETYKITLMISLPVLLAGLVVGLLVSIFQATTQINEMTLSFVPKILAVIGVLI

 cons              *     :    .   : : .:      . . *   .:.*:  .::: :: *  *:  *   :


 sp|P0A1L7|SPAQ_   LLSGWYGEVLLSYGRQVIFLAL-AK---G
 tr|G0CIR3|G0CIR   VTAPWGASAIMQFGQALMQAAFP------
 sp|P0A1L5|FLIQ_   VAGPWMLNLLLDYVRTLFSNLPYII---G
 tr|Q7BFA7|Q7BFA   ATASWLGNELHSFAEMTMMKIQ-GI---R
 sp|P0A1M4|SPAQ_   LLSGWYGEVLLSFCHEIMFLIK-SG---V
 tr|Q60237|Q6023   MTGRWIGVELIQFINMAFDLIA-RSGVSH
 sp|P0AC10|FLIQ_   IAGPWMLNLLLDYVRTLFTNLPYII---G
 tr|Q2TJ91|Q2TJ9   LMSGWYGEKLYSFGIEMLNLAF-AR---G
 tr|C3T4R7|C3T4R   IAGPWMLNLLLDYVRTLFTNLPYII---G
 sp|P0A0S3|FLIQ_   LTMPWMTNMLLDYTKTLIKLIPKII---G

 cons                  *    : .:    :
```

>sp|P0A1L7|SPAQ_SALTY Surface presentation of antigens protein SpaQ OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNspaQ PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
11-36	26	-1.434	ALYLVLILSGWPTIVATIIGLLVGLF	[analyze]
46-72	27	-1.698	TLPFGIKLLGVCLCLFLLSGWYGEVLL	[analyze]

Download plot in EPS format (new)

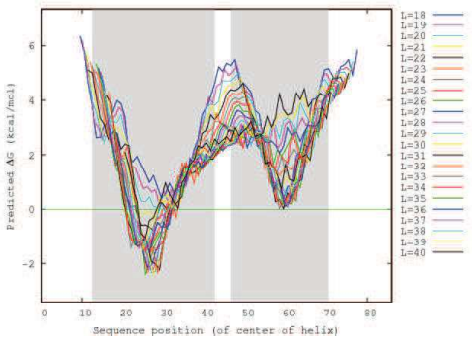


>tr|G0CIR3|G0CIR3_XANCA Type III secretion protein HrcS OSXanthomonas campestris pv raphani 756C GNhrcS PE4 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
13-42	30	-2.413	ALLLCLKVSLPVVGVAALAGLLIAFLQAVM	[analyze]
47-70	24	0.008	ASISFALKLVVVVAAIAVTAPWGA	[analyze]

Download plot in EPS format (new)

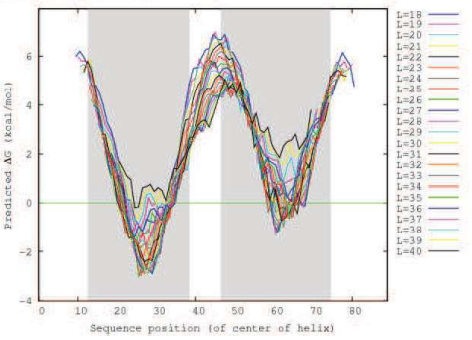


>sp|P0A1L5|FLIQ_SALTY Flagellar biosynthetic protein FliQ OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNfliQ PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
13-38	26	-3.008	AMKVALALAAPLLLVALITGLIISIL	[analyze]
47-74	28	-1.255	MTLSFIPKIVAVFIAIIVAGPWMLNLLL	[analyze]

Download plot in EPS format (new)

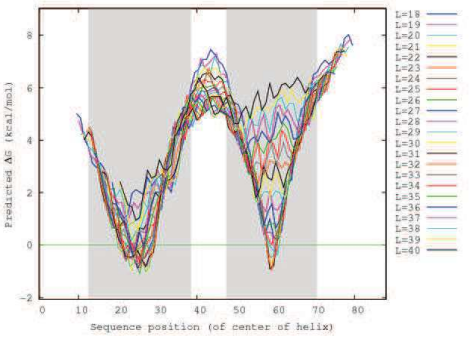


>tr|Q7BFA7|Q7BFA7_YEREN Putative type III secretion protein OSYersinia enterocolitica GNyscS PE4 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
13-38	26	-1.067	ALWLVLVLSMPPVLVAAVVGTLVSLV	[analyze]
48-70	23	-0.949	TLGFVIKLIAVVVTLFATASWLG	[analyze]

Download plot in EPS format (new)

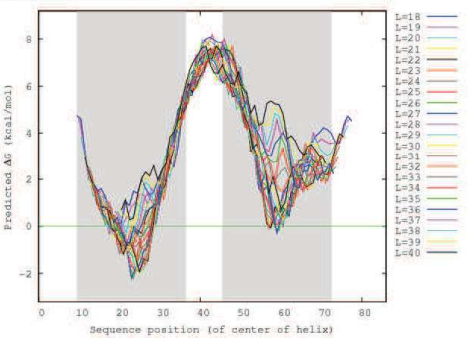


>sp|P0A1M4|SPAQ_SHIFL Surface presentation of antigens protein SpaQ OSShigella flexneri GNspaQ PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
10-36	27	-2.211	KALYLILIFSLWPVGIATVIGLSIGLL	[analyze]
46-72	27	-0.287	TLPFGIKLIGVSISLLLLSGWYGEVLL	[analyze]

Download plot in EPS format (new)

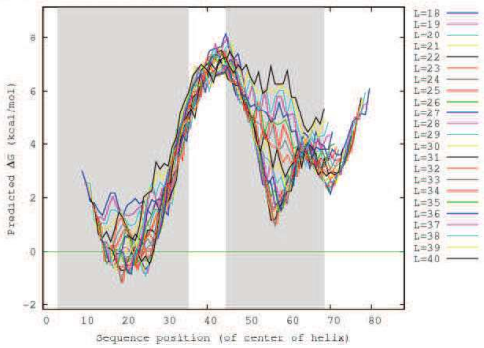


>tr|Q60237|Q60237_PSESY HrcS OSPseudomonas syringae pv syringae GNhrpO PE4 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
4-35	32	-1.179	LALFKQGMFLVVILTAPPLAVAVLVGVVTSLL	[analyze]
45-68	24	0.951	TLPFGIKLGAVGLTLAMTGRWIGV	[analyze]

Download plot in EPS format (new)

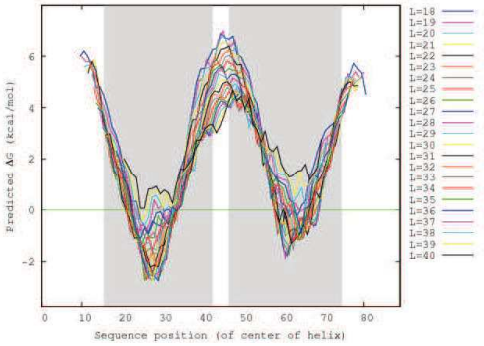


>sp|P0AC10|FLIQ_SHIFL Flagellar biosynthetic protein FliQ OSShigella flexneri GNfliQ PE3 SV1

Predicted TM helices:

Position	*Length*	*Predicted ΔG*	*Sequence*	
16-42	27	-2.736	VALALAAPLLLVALVTGLIISILQAAT	[analyze]
47-74	28	-1.860	MTLSFIPKIIAVFIAIIIAGPWMLNLLL	[analyze]

Download plot in EPS format (new)

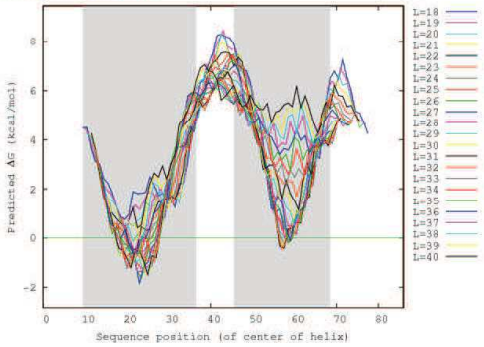


>tr|Q2TJ91|Q2TJ91_ECOLX EpaQ OSEscherichia coli GNepaQ PE4 SV1

Predicted TM helices:

Position	*Length*	*Predicted ΔG*	*Sequence*	
10-36	27	-1.832	RALYLILVMSAGPIAVATFVGLLVGLF	[analyze]
46-68	23	-0.461	TLPFGVKLLCVSICFFLMSGWYG	[analyze]

Download plot in EPS format (new)

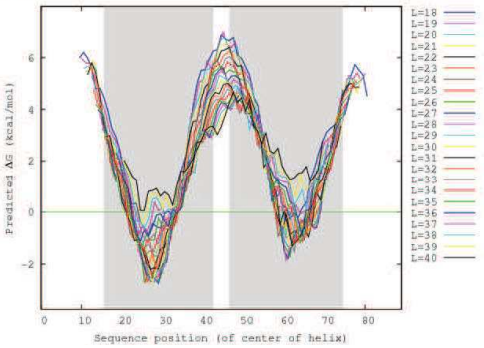


>tr|C3T4R7|C3T4R7_ECOLX Flagellar biosynthesis protein OSEscherichia coli GNfliQ PE4 SV1

Predicted TM helices:

Position	*Length*	*Predicted ΔG*	*Sequence*	
16-42	27	-2.736	VALALAAPLLLVALVTGLIISILQAAT	[analyze]
47-74	28	-1.860	MTLSFIPKIIAVFIAIIIAGPWMLNLLL	[analyze]

Download plot in EPS format (new)

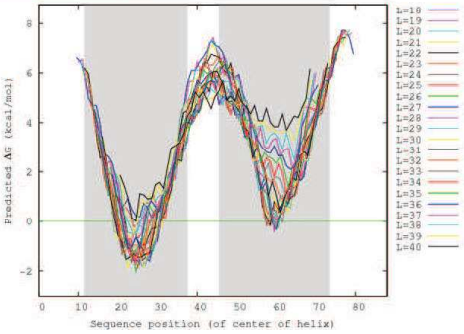


>sp|P0A0S3|FLIQ_HELPY Flagellar biosynthetic protein FliQ OSHelicobacter pylori strain ATCC 700392 26695 GNfliQ PE3 SV1

Predicted TM helices:

Position	*Length*	*Predicted ΔG*	*Sequence*	
12-37	26	-2.025	TYKITLMISLPVLLAGLVVGLLVSIF	[analyze]
46-73	28	-0.389	MTLSFVPKILAVIGVLILTMPWMTNMLL	[analyze]

Download plot in EPS format (new)

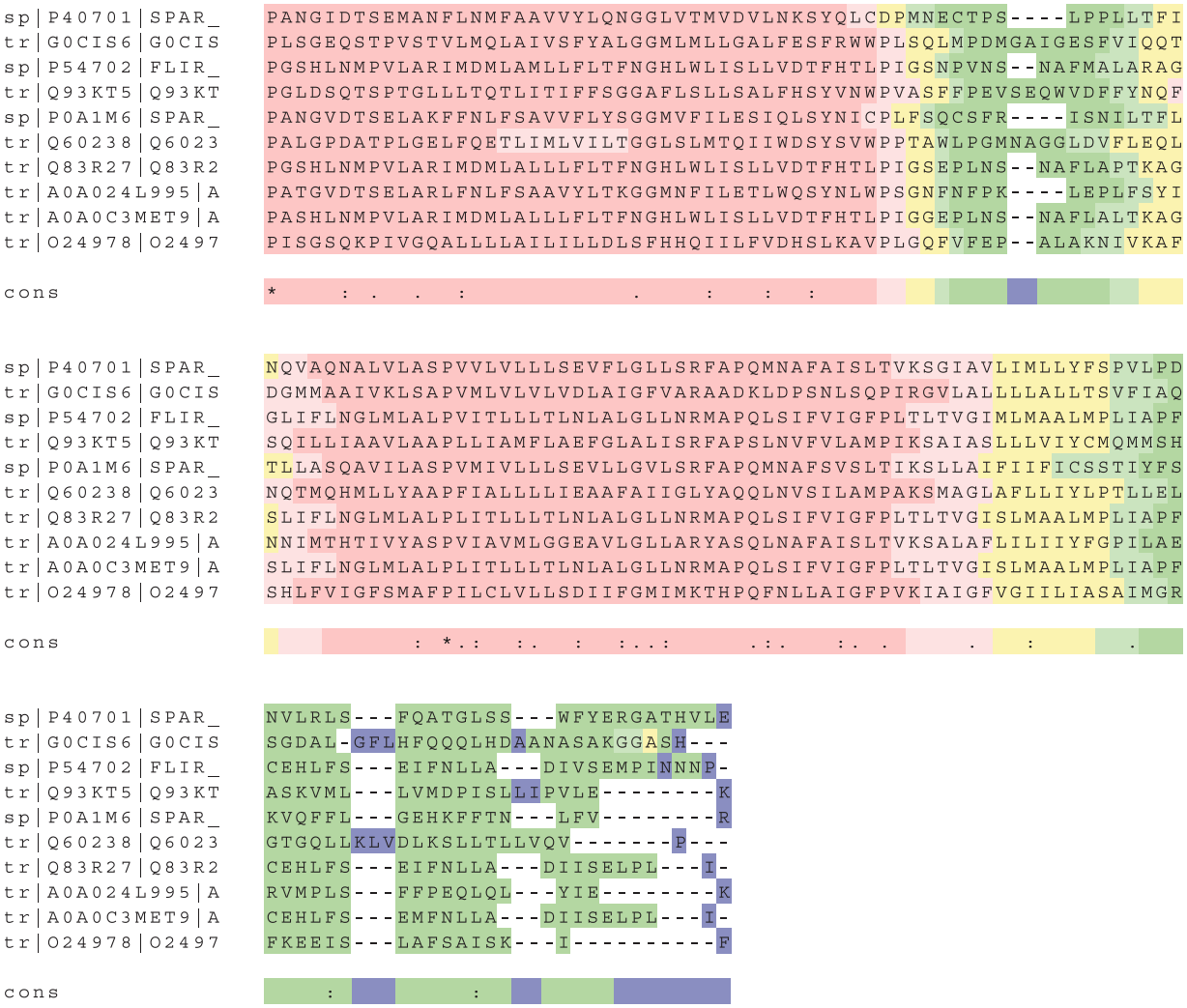


```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
   Cedric Notredame
   CPU TIME:0 sec.
   SCORE=928
   *
```

- entirely conserved columns marked with a star (*)
- columns comprising amino acids of same size and hydropathy marked with a colon (:)
- columns comprising amino acids of similar size or evolutionary preserved hydropathy marked with a period (.)
- positive charged amino acids (Arg and Lys) marked in **blue**
- negative charged amino acids (Asp and Glu) marked in **orange**
- transmembrane domains marked in grey

```
sp|P0A1L7|SPAQ_   MD--DLVFAGNKALYLVLILSGWPTIVATIIGLLVGLFQTVTQLQEQTLPFGIKLLGVCLCLFLLSGWYGEVLLSYGRQVIFLAL-AK---
tr|G0CIR3|G0CIR   MDHDDLVRLTSEALLLCLKVSLPVVGVAALAGLLIAFLQAVMSLQDASISFALKLVVVVAAIAVTAPWGASAIMQFGQALMQAAFP------
sp|P0A1L5|FLIQ_   MTPESVMMMGTEAMKVALALAAPLLLVALITGLIISILQAATQINEMTLSFIPKIVAVFIAIIVAGPWMLNLLLDYVRTLFSNLPYII----
tr|Q7BFA7|Q7BFA   MSQGDIIHFTSQALWLVLVLSMPPVLVAAVVGTLVSLVQALTQIQEQTLGFVIKLIAVVVTLFATASWLGNELHSFAEMTMMKIQ-GI---I
sp|P0A1M4|SPAQ_   MS--DIVYMGNKALYLILIFSLWPVGIATVIGLSIGLLQTVTQLQEQTLPFGIKLIGVSISLLLLSGWYGEVLLSFCHEIMFLIK-SG---V
tr|Q60237|Q6023   ME--A-LALFKQGMFLVVILTAPPLAVAVLVGVVTSLLQALMQIQDQTLPFGIKLGAVGLTLAMTGRWIGVELIQFINMAFDLIA-RSGVSI
sp|P0AC10|FLIQ_   MTPESVMMMGTEAMKVALALAAPLLLVALVTGLIISILQAATQINEMTLSFIPKIIAVFIAIIIAGPWMLNLLLDYVRTLFTNLPYII----
tr|Q2TJ91|Q2TJ9   MD--DIVFAGNRALYLILVMSAGPIAVATFVGLLVGLFQTVTQLQEQTLPFGVKLLCVSICFFLMSGWYGEKLYSFGIEMLNLAF-AR---
tr|C3T4R7|C3T4R   MTPESVMMMGTEAMKVALALAAPLLLVALVTGLIISILQAATQINEMTLSFIPKIIAVFIAIIIAGPWMLNLLLDYVRTLFTNLPYII----
sp|P0A0S3|FLIQ_   MES-QLMKLAIETYKITLMISLPVLLAGLVVGLLVSIFQATTQINEMTLSFVPKILAVIGVLILTMPWMTNMLLDYTKTLIKLIPKII----

cons              *      :     .    : : .:       . . *   .:.*:  .::: :: *  *:  *   :       *     : .:     :
```

```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
  Cedric Notredame
  CPU TIME:0 sec.
  SCORE=843
  *
   BAD AVG GOOD
  *
  sp|P40701|SPAR_     :  87
  tr|G0CIS6|G0CIS     :  84
  sp|P54702|FLIR_     :  87
  tr|Q93KT5|Q93KT     :  84
  sp|P0A1M6|SPAR_     :  84
  tr|Q60238|Q6023     :  83
  tr|Q83R27|Q83R2     :  87
  tr|A0A024L995|A     :  88
  tr|A0A0C3MET9|A     :  87
  tr|O24978|O2497     :  88
  cons                :  84

  sp|P40701|SPAR_     MFYA------LYFE-IHHLVASAALGFARVAPIFFFLPFLNSGVLSGA-PRNAIIILVALGVW
  tr|G0CIS6|G0CIS     MSDTATALLALSSQ-GVSLLTLLALCGVRVFVLFFVLPATAQDSLPGM-TRNGVIYVLSSFIA
  sp|P54702|FLIR_     MIQV------TSEQ-WLYWLHLYFWPLLRVLALISTAPILSERAIPKR-VKLGLGIMITLVIA
  tr|Q93KT5|Q93KT     MIAD------LIQ----RPLLTYTLLLPRFMACFVILPVLSKQLLGGVLLRNGIVCSLALYVY
  sp|P0A1M6|SPAR_     MDIS------SWFESIHVFLILLNGVFFRLAPLFFFLPFLNNGIISPS-IRIPVIFLVASGLI
  tr|Q60238|Q6023     MPFD------AHS--AFQFMLGMGLAMARLMPCMLLVPAFCFKYLKGP-LRYAVVAVMAMIPA
  tr|Q83R27|Q83R2     MMQE------TSDQ-WLSWLSLYFWPLLRVLALISTAPILSERSVPKR-VKLGLAMMITFAIA
  tr|A0A024L995|A     MGEA------ILYQ-LHSLLAATALCFCRLAPTFYLLPFFASGNIPTV-VRHPIIIVVSCALV
  tr|A0A0C3MET9|A     MMQV------TSDQ-WLSWLSLYFWPLLRVLALISTAPILSERSVPKR-VKLGLAMMITFAIA
  tr|O24978|O2497     MLDF------IQEL-STPHVRDFFLLFLRVSGVLSFFPFFENHLVPLS-VRGALSLYVSAIFY

  cons                *                  :        *.   :   *      :     :  :   ::


  sp|P40701|SPAR_     PHALNEA-PPFL--SVAMIPLVLQEAAVGVMLGCLLSWPFWVMHALGCIIDNQRGATLSSSID
  tr|G0CIS6|G0CIS     YGQPADA-LARI-EAAGLVGLVFKEAFIGLLIGFAASTVFWVAESVGLLIDDVSGYNNVQMIN
  sp|P54702|FLIR_     PSLPAND-TPLF--SIAALWLAMQQILIGIALGFTMQFAFAAVRTAGEFIGLQMGLSFATFVD
  tr|Q93KT5|Q93KT     PAVANQPYIEVD--AFTLMLLIGKEFILGLLIGFVATIPFWALESAGFIVDNQRGAAMASLLN
  sp|P0A1M6|SPAR_     TSGKVDI-GSSV--FEHVYFLMFKEIIVGLLLSFCLSLPFWIFHAVGSIIDNQRGATLSSSID
  tr|Q60238|Q6023     PAITRAL-ESLDDNWFAIGGLLIKEAVLGTLLGLLLYAPFWMFASVGALLDSQRGALSGGQLN
  tr|Q83R27|Q83R2     PSLPAND-VPVF--SFFALWLAVQQILIGIALGFTMQFAFAAVRTAGEIIGLQMGLSFATFVD
  tr|A0A024L995|A     QHYHYEL-LNLN--EIDIALFAAREIIIGLFIACLLASPFWIFLAIGSFIDNQRGATLSSTLD
  tr|A0A0C3MET9|A     PSLPAND-VPVF--SFFALWLAVQQILIGIALGFTMQFAFAAVRTAGEIIGLQMGLSFATFVD
  tr|O24978|O2497     PTLEFSN-AAYT--PEGFIIACLCELFLGVCASVFLQIVFASLVFATDSISFSMGLTMASAYD

  cons                                        :  :*   .      *         :.   *       :
```

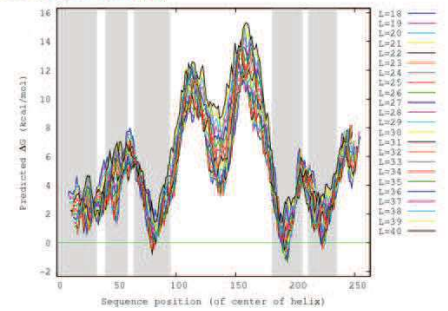

sp|P40701|SPAR_ PANGIDTSEMANFLNMFAAVVYLQNGGLVTMVDVLNKSYQLCDPMNECTPS----LPPLLTFI
tr|G0CIS6|G0CIS PLSGEQSTPVSTVLMQLAIVSFYALGGMLMLLGALFESFRWWPLSQLMPDMGAIGESFVIQQT
sp|P54702|FLIR_ PGSHLNMPVLARIMDMLAMLLFLTFNGHLWLISLLVDTFHTLPIGSNPVNS--NAFMALARAG
tr|Q93KT5|Q93KT PGLDSQTSPTGLLLTQTLITIFFSGGAFLSLLSALFHSYVNWPVASFFPEVSEQWVDFFYNQF
sp|P0A1M6|SPAR_ PANGVDTSELAKFFNLFSAVVFLYSGGMVFILESIQLSYNICPLFSQCSFR----ISNILTFL
tr|Q60238|Q6023 PALGPDATPLGELFQETLIMLVILTGGLSLMTQIIWDSYSVWPPTAWLPGMNAGGLDVFLEQL
tr|Q83R27|Q83R2 PGSHLNMPVLARIMDMLALLLFLTFNGHLWLISLLVDTFHTLPIGSEPLNS--NAFLAPTKAG
tr|A0A024L995|A PATGVDTSELARLFNLFSAAVYLTKGGMNFILETLWQSYNLWPSGNFNFPK----LEPLFSYI
tr|A0A0C3MET9|A PASHLNMPVLARIMDMLALLLFLTFNGHLWLISLLVDTFHTLPIGGEPLNS--NAFLALTKAG
tr|O24978|O2497 PISGSQKPIVGQALLLLAILILLDLSFHHQIILFVDHSLKAVPLGQFVFEP--ALAKNIVKAF
cons
sp|P40701|SPAR_ NQVAQNALVLASPVVLVLLLSEVFLGLLSRFAPQMNAFAISLTVKSGIAVLIMLLYFSPVLPD
tr|G0CIS6|G0CIS DGMMAAIVKLSAPVMLVLVLVDLAIGFVARAADKLDPSNLSQPIRGVLALLLLALLTSVFIAQ
sp|P54702|FLIR_ GLIFLNGLMLALPVITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGIMLMAALMPLIAPF
tr|Q93KT5|Q93KT SQILLIAAVLAAPLLIAMFLAEFGLALISRFAPSLNVFVLAMPIKSAIASLLLVIYCMQMMSH
sp|P0A1M6|SPAR_ TLLASQAVILASPVMIVLLLSEVLLGVLSRFAPQMNAFSVSLTIKSLLAIFIIFICSSTIYFS
tr|Q60238|Q6023 NQTMQHMLLYAAPFIALLLLIEAAFAIIGLYAQQLNVSILAMPAKSMAGLAFLLIYLPTLLEL
tr|Q83R27|Q83R2 SLIFLNGLMLALPLITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGISLMAALMPLIAPF
tr|A0A024L995|A NNIMTHTIVYASPVIAVMLGGEAVLGLLARYASQLNAFAISLTVKSALAFLILIIYFGPILAE
tr|A0A0C3MET9|A SLIFLNGLMLALPLITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGISLMAALMPLIAPF
tr|O24978|O2497 SHLFVIGFSMAFPILCLVLLSDIIFGMIMKTHPQFNLLAIGFPVKIAIGFVGIILIASAIMGR
cons
sp|P40701|SPAR_ NVLRLS---FQATGLSS---WFYERGATHVLE
tr|G0CIS6|G0CIS SGDAL-GFLHFQQQLHDAANASAKGGASH---
sp|P54702|FLIR_ CEHLFS---EIFNLLA---DIVSEMPINNNP-
tr|Q93KT5|Q93KT ASKVML---LVMDPISLLIPVLE--------K
sp|P0A1M6|SPAR_ KVQFFL---GEHKFFTN---LFV--------R
tr|Q60238|Q6023 GTGQLLKLVDLKSLLTLLVQV-------P---
tr|Q83R27|Q83R2 CEHLFS---EIFNLLA---DIISELPL---I-
tr|A0A024L995|A RVMPLS---FFPEQLQL---YIE--------K
tr|A0A0C3MET9|A CEHLFS---EMFNLLA---DIISELPL---I-
tr|O24978|O2497 FKEEIS---LAFSAISK---I----------F
cons

>sp|P40701|SPAR_SALTY Surface presentation of antigens protein SpaR OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNspaR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
1-33	33	0.563	MFYALYFEIHHLVASAALGFARVAPFFFLPFL	[analyze]
41-59	19	1.351	APRNAIILVALGVWPHAL	[analyze]
65-95	31	-0.726	FLSVAMIPLVLQEAAVGVMLGCLLSWPFWVM	[analyze]
181-206	26	-1.342	ALVLASPVVLVLLLSEVFLGLLSRFA	[analyze]
211-235	25	-0.383	AFAISLTVKSGIAVLIMLLYFSPVL	[analyze]

Download plot in EPS format (new)

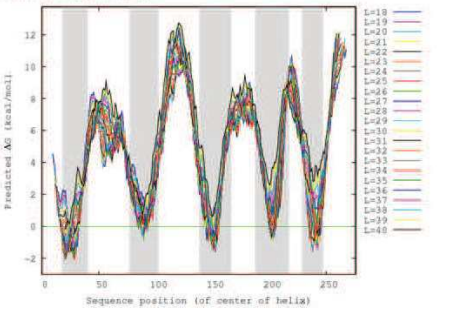


>tr|G0CIS6|G0CIS6_XANCA Type III secretion protein HrcT OSXanthomonas campestris pv raphani 756C GNhrcT PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
18-40	23	-2.068	LLTLLALCGVRVFVLFFVLPATA	[analyze]
77-102	26	-0.753	LVGLVFKEAFIGLLIGFAASTVFWVA	[analyze]
139-166	28	-1.634	LAIVSFYALGGMLMLLQALFESFRWWPL	[analyze]
188-217	30	-0.934	MMAAIVKLSAPVMLVLVLVDLAIGFVARAA	[analyze]
229-247	19	-1.516	IRGVLALLLLALLTSVFIA	[analyze]

Download plot in EPS format (new)

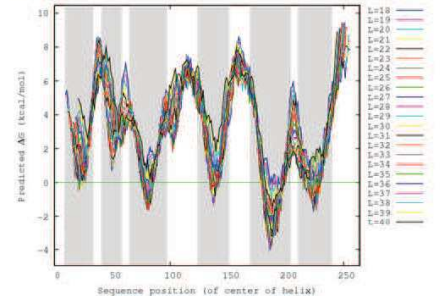


>sp|P54702|FLIR_SALTY Flagellar biosynthetic protein FliR OSSalmonella typhimurium strain LT2 SGSC1412 ATCC 700720 GNfliR PE3 SV2

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
9-33	25	-0.404	WLYWLHLYFWPLLRVLALISTAPIL	[analyze]
41-58	18	0.788	RVKLGLGMTLVIAPSL	[analyze]
65-97	33	-1.560	LFSIAALIVLAMQQILIGIALGFTMQFAFAAVRT	[analyze]
124-151	28	-1.262	VLARIIIDMLAMLLFLTFNGHLWLISLLV	[analyze]
169-204	36	-3.945	FMALARAGGLIFLNGLMLALPVTTLLLTLNLALGLL	[analyze]
211-240	30	-1.661	LSIFVIGFPLTLTVGIIILMAALMPLIAPFC	[analyze]

Download plot in EPS format (new)

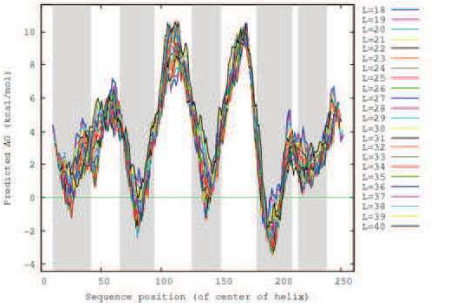


>tr|Q93KT5|Q93KT5_YEREN Putative type III secretion protein OSYersinia enterocolitica GNyscT PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
10-41	32	-1.133	LLTYTLLLPRFMACFVILPVLSKQLLGGVLLR	[analyze]
66-94	29	-2.321	AFTLMLLIGKEFILGLLIGFVATIPFWAL	[analyze]
126-150	25	-1.355	LLLTQTLITFFSGGAFLSLLSALF	[analyze]
180-209	30	-3.445	ILLIAAVLAAPLLIAMFLAEFGLALISRFA	[analyze]
214-238	25	0.639	VPVLAMPIKSAIASLLLVIYCMQMM	[analyze]

Download plot in EPS format (new)

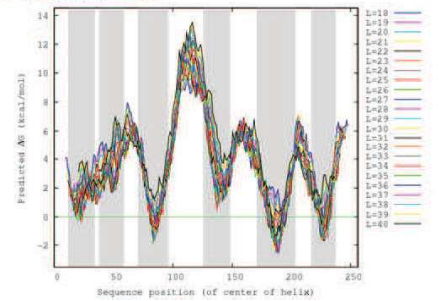


>sp|P0A1M6|SPAR_SHIFL Surface presentation of antigens protein SpaR OSShigella flexneri GNspaR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
12-34	23	-0.282	VFLILLNGVFFRLAPLFFFLPFL	[analyze]
38-59	22	1.706	IISPSRIPVFLVASGLTSG	[analyze]
71-96	26	-1.798	VYFLIIPKEIVGLLLSFCLSLPFWIF	[analyze]
126-148	23	0.455	LAKFFNLFSAVVFLYSGGMVFIL	[analyze]
171-203	33	-2.530	ILTFLTLLASQAVLASPVMIVLLLSEVLLGVL	[analyze]
217-237	21	-1.788	LTIKSLLAIFIFICSSTIYF	[analyze]

Download plot in EPS format (new)

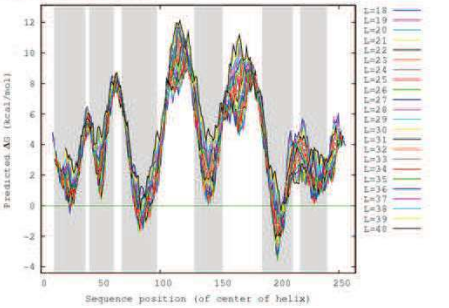


>tr|Q60238|Q60238_PSESY HrcT OSPseudomonas syringae pv syringae GNhrpX PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
11-37	27	-0.400	FMLGMGLAMARLMPCMLLVPAFCFKYL	[analyze]
41-61	21	0.265	LRYAVVAVMAMIPAPAITRAL	[analyze]
66-97	30	-1.777	WFAIGGLLIKEAVLGTLLGILLVAPFWMFA	[analyze]
129-152	24	-0.022	LFGETLIMLVLTGGLSLMTQIW	[analyze]
186-211	26	-3.493	MLLYAAPFIALLLLEAAFAIIGLYA	[analyze]
218-240	23	0.169	ILAMPAKSMAGLAFLLIYLPTLL	[analyze]

Download plot in EPS format (new)

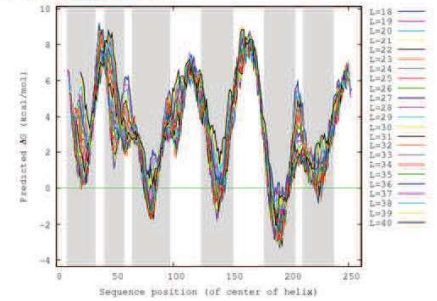


>tr|Q83R27|Q83R27_SHIFL Flagellar biosynthetic protein FliR OSShigella flexneri GNfliR PE3 SV2

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
9-33	25	-0.007	WLSWLSLYFWPLLRVLALISTAPIL	[analyze]
41-58	18	1.187	RVKLGLAMMTFAIAPSL	[analyze]
65-97	33	-1.685	VFSFFALWLAVGQILIGIALGFTMQFAFAAVRT	[analyze]
124-151	28	-1.722	VLARMDMLALLLFLTFNGHLWLISLLV	[analyze]
178-204	27	-3.356	LIFLNGLMLALPLITLLLTLNLALGLL	[analyze]
211-237	27	-0.752	LSIFVIGFPLTLTVGISLMAALMPLIA	[analyze]

Download plot in EPS format (new)

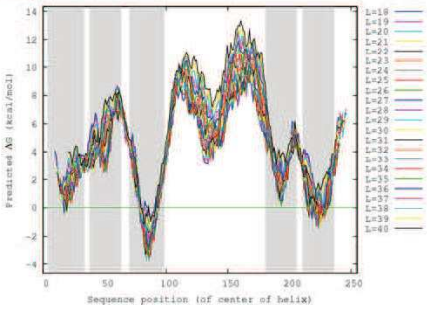


>tr|A0A024L995|A0A024L995_ECOLX EscTYscTHrcT family type III secretion system export apparatus protein OSEscherichia coli GNepaR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
4-33	30	-0.291	AILYQLHSLLAATALCFCRLAPTFYLLPFF	[analyze]
38-63	26	2.505	IPTVVRHPIIVVSCALVQIHYIHYELL	[analyze]
70-98	29	-3.673	IALFAAREIIGLFIACLLASPFWFLAI	[analyze]
161-206	26	0.162	TIVYASPVIAVMLGGEAVLGLLARYA	[analyze]
211-236	26	-1.314	AFAISLTVKSALAFLILIIVFGPILA	[analyze]

Download plot in EPS format (new)

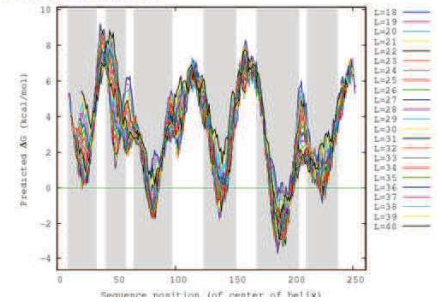


>tr|A0A0C3MET9|A0A0C3MET9_ECOLX Flagellar biosynthetic protein FliR OSEscherichia coli GNfliR PE3 SV1

Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
9-33	25	-0.007	WLSWLSLYFWPLLRVLALISTAPIL	[analyze]
41-58	18	1.187	RVKLGLAMMTFAIAPSL	[analyze]
65-97	33	-1.685	VFSFFALWLAVGQILIGIALGFTMQFAFAAVRT	[analyze]
124-151	28	-1.722	VLARMDMLALLLFLTFNGHLWLISLLV	[analyze]
169-204	36	-3.649	FLALTKAGSLIFLNGLMLALPLITLLLTLNLALGLL	[analyze]
211-237	27	-0.752	LSIFVIGFPLTLTVGISLMAALMPLIA	[analyze]

Download plot in EPS format (new)

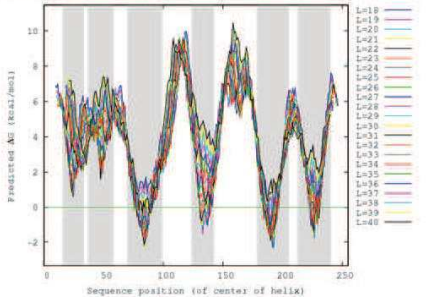


>tr|O24978|O24978_HELPY Flagellar biosynthetic protein FliR OSHelicobacter pylori strain ATCC 700392 26695 GNHP_0173 PE3 SV1

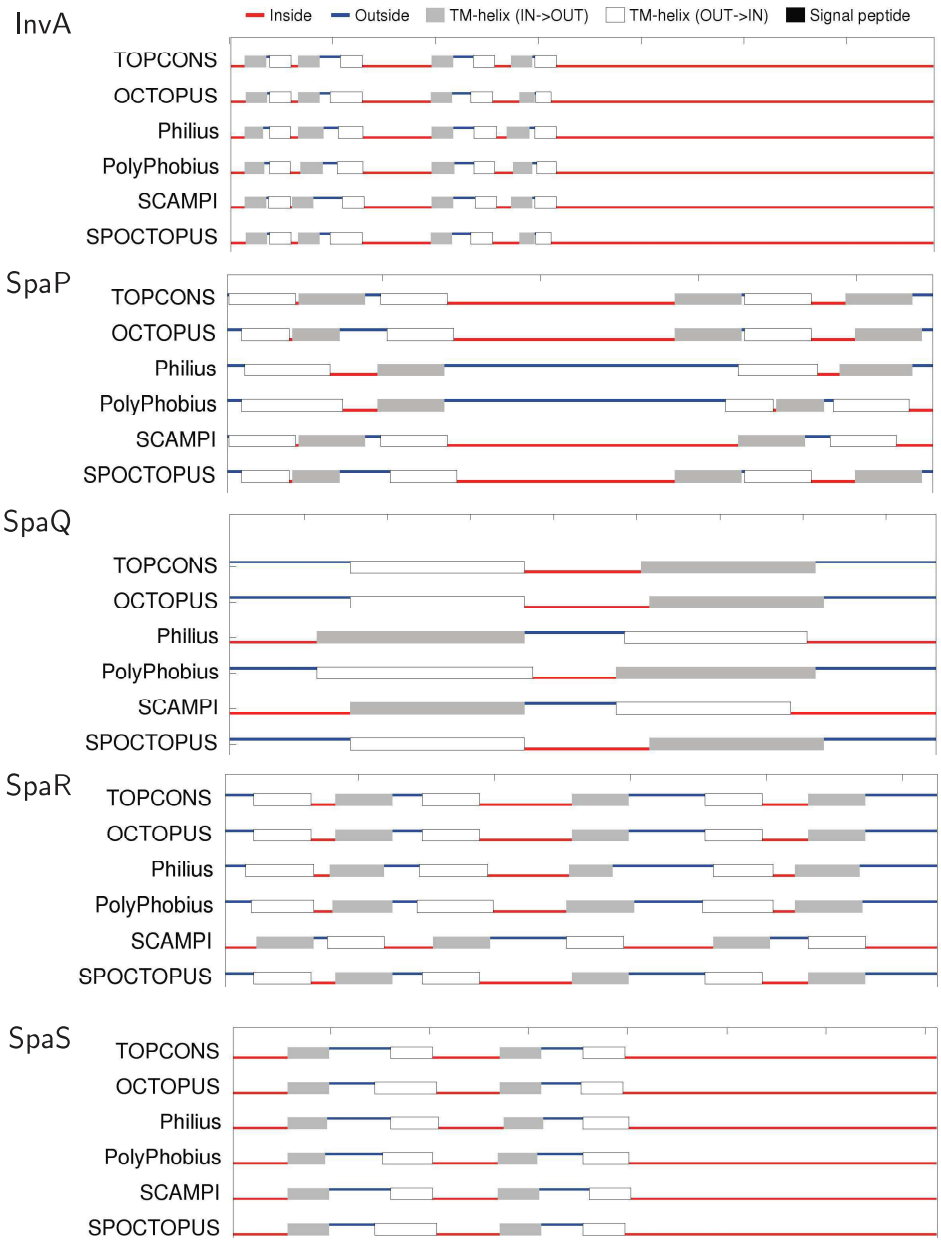
Predicted TM helices:

Position	Length	Predicted ΔG	Sequence	
16-33	18	0.609	FFLLFLRVSGVLSFFPFF	[analyze]
37-58	22	1.592	LVPLSVRGALSLYVSAIFYPTL	[analyze]
70-99	30	-2.264	FIIACLCELFLGVCASVFLQIVFASLVFAT	[analyze]
124-142	19	-1.549	IVGQALLLLLAILLLDLSF	[analyze]
179-205	27	-2.330	LFVIGFSMAFPLCLVLLSDIFGMIM	[analyze]
213-240	28	-1.796	LLAIGFPVKIAIGFVGILIASAMGRF	[analyze]

Download plot in EPS format (new)



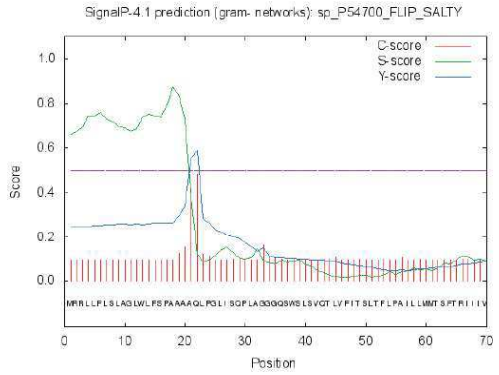
```
T-COFFEE, Version_11.00.d625267 2016-01-11 15:25:41 - Revision d625267 - Build 507
  Cedric Notredame
  CPU TIME:0 sec.
  SCORE=843

  - entirely conserved columns marked with a star (*)
  - columns comprising amino acids of same size and hydropathy marked with a colon (:)
  - columns comprising amino acids of similar size or evolutionary preserved hydropathy marked with a period (.)
  - positive charged amino acids (Arg and Lys) marked in blue
  - negative charged amino acids (Asp and Glu) marked in orange
  - transmembrane domains marked in grey

sp|P40701|SPAR_   MFYA------LYFE-IHHLVASAALGFARVAPIFFFLPFLNSGVLSGA-PRNAIIILVALGVW
tr|G0CIS6|G0CIS   MSDTATALLALSSQ-GVSLLTLLALCGVRVFVLFFVLPATAQDSLPGM-TRNGVIYVLSSFIA
sp|P54702|FLIR_   MIQV------TSEQ-WLYWLHLYFWPLLRVLALISTAPILSERAIPKR-VKLGLGIMITLVIA
tr|Q93KT5|Q93KT   MIAD------LIQ----RPLLTYTLLLPRFMACFVILPVLSKQLLGGVLLRNGIVCSLALYVY
sp|P0A1M6|SPAR_   MDIS------SWFESIHVFLILLNGVFFRLAPLFFFLPFLNNGIISPS-IRIPVIFLVASGLI
tr|Q60238|Q6023   MPFD------AHS--AFQFMLGMGLAMARLMPCMLLVPAFCFKYLKGP-LRYAVVAVMAMIPA
tr|Q83R27|Q83R2   MMQE------TSDQ-WLSWLSLYFWPLLRVLALISTAPILSERSVPKR-VKLGLAMMITFAIA
tr|A0A024L995|A   MGEA------ILYQ-LHSLLAATALCFCRLAPTFYLLPFFASGNIPTV-VRHPIIIVVSCALV
tr|A0A0C3MET9|A   MMQV------TSDQ-WLSWLSLYFWPLLRVLALISTAPILSERSVPKR-VKLGLAMMITFAIA
tr|O24978|O2497   MLDF------IQEL-STPHVRDFFLLFLRVSGVLSFFPFFENHLVPLS-VRGALSLYVSAIFY

cons              *           :       *.   :   *       :     :  :  ::


sp|P40701|SPAR_   PHALNEA-PPFL--SVAMIPLVLQEAAVGVMLGCLLSWPFWVMHALGCIIDNQRGATLSSSID
tr|G0CIS6|G0CIS   YGQPADA-LARI-EAAGLVGLVFKEAFIGLLIGFAASTVFWVAESVGLLIDDVSGYNNVQMIN
sp|P54702|FLIR_   PSLPAND-TPLF--SIAALWLAMQQILIGIALGFTMQFAFAAVRTAGEFIGLQMGLSFATFVD
tr|Q93KT5|Q93KT   PAVANQPYIEVD--AFTLMLLIGKEFILGLLIGFVATIPFWALESAGFIVDNQRGAAMASLLN
sp|P0A1M6|SPAR_   TSGKVDI-GSSV--FEHVYFLMFKEIIVGLLLSFCLSLPFWIFHAVGSIIDNQRGATLSSSID
tr|Q60238|Q6023   PAITRAL-ESLDDNWFAIGGLLIKEAVLGTLLGLLLYAPFWMFASVGALLDSQRGALSGGQLN
tr|Q83R27|Q83R2   PSLPAND-VPVF--SFFALWLAVQQILIGIALGFTMQFAFAAVRTAGEIIGLQMGLSFATFVD
tr|A0A024L995|A   QHYHYEL-LNLN--EIDIALFAAREIIIGLFIACLLASPFWIFLAIGSFIDNQRGATLSSTLD
tr|A0A0C3MET9|A   PSLPAND-VPVF--SFFALWLAVQQILIGIALGFTMQFAFAAVRTAGEIIGLQMGLSFATFVD
tr|O24978|O2497   PTLEFSN-AAYT--PEGFIIACLCELFLGVCASVFLQIVFASLVFATDSISFSMGLTMASAYD

cons                                          :  :*   .       *          :.    *       :


sp|P40701|SPAR_   PANGIDTSEMANFLNMFAAVVYLQNGGLVTMVDVLNKSYQLCDPMNECTPS----LPPLLTFI
tr|G0CIS6|G0CIS   PLSGEQSTPVSTVLMQLAIVSFYALGGMLMLLGALFESFRWWPLSQLMPDMGAIGESFVIQQT
sp|P54702|FLIR_   PGSHLNMPVLARIMDMLAMLLFLTFNGHLWLISLLVDTFHTLPIGSNPVNS--NAFMALARAG
tr|Q93KT5|Q93KT   PGLDSQTSPTGLLLTQTLITIFFSGGAFLSLLSALFHSYVNWPVASFFPEVSEQWVDFFYNQF
sp|P0A1M6|SPAR    PANGVDTSELAKFFNLFSAVVFLYSGGMVFILESIQLSYNICPLFSQCSFR----ISNILTFL
tr|Q60238|Q6023   PALGPDATPLGELFQETLIMLVILTGGLSLMTQIIWDSYSVWPPTAWLPGMNAGGLDVFLEQL
tr|Q83R27|Q83R2   PGSHLNMPVLARIMDMLALLFLTFNGHLWLISLLVDTFHTLPIGSEPLNS--NAFLAPTKAG
tr|A0A024L995|A   PATGVDTSELARLFNLFSAAVYLTKGGMNFILETLWQSYNLWPSGNFNFPK----LEPLFSYI
tr|A0A0C3MET9|A   PASHLNMPVLARIMDMLALLFLTFNGHLWLISLLVDTFHTLPIGGEPLNS--NAFLALTKAG
tr|O24978|O2497   PISGSQKPIVGQALLLLAILILLDLSFHHQIILFVDHSLKAVPLGQFVFEP--ALAKNIVKAF

cons              *     : .   .  :             .     :   :  :


sp|P40701|SPAR_   NQVAQNALVLASPVVLVLLLSEVFLGLLSRFAPQMNAFAISLTVKSGIAVLIMLLYFSPVLPD
tr|G0CIS6|G0CIS   DGMMAAIVKLSAPVMLVLVLVDLAIGFVARAADKLDPSNLSQPIRGVLALLLLALLTSVFIAQ
sp|P54702|FLIR_   GLIFLNGLMLALPVITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGIMLMAALMPLIAPF
tr|Q93KT5|Q93KT   SQILLIAAVLAAPLLIAMFLAEFGLALISRFAPSLNVFVLAMPIKSAIASLLLVIYCMQMMSH
sp|P0A1M6|SPAR_   TLLASQAVILASPVMIVLLLSEVLLGVLSRFAPQMNAFSVSLTIKSLLAIFIIFICSSTIYFS
tr|Q60238|Q6023   NQTMQHMLLYAAPFIALLLLIEAAFAIIGLYAQQLNVSILAMPAKSMAGLAFLLIYLPTLLEL
tr|Q83R27|Q83R2   SLIFLNGLMLALPLITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGISLMAALMPLIAPF
tr|A0A024L995|A   NNIMTHTIVYASPVIAVMLGGEAVLGLLARYASQLNAFAISLTVKSALAFLILIIYFGPILAE
tr|A0A0C3MET9|A   SLIFLNGLMLALPLITLLLTLNLALGLLNRMAPQLSIFVIGFPLTLTVGISLMAALMPLIAPF
tr|O24978|O2497   SHLFVIGFSMAFPILCLVLLSDIIFGMIMKTHPQFNLLAIGFPVKIAIGFVGIILIASAIMGR

cons                          : *.:   :.   :   :..:      .:.    :. .      .    :       .


sp|P40701|SPAR_   NVLRLS---FQATGLSS---WFYERGATHVLE
tr|G0CIS6|G0CIS   SGDAL-GFLHFQQQLHDAANASAKGGASH---
sp|P54702|FLIR_   CEHLFS---EIFNLLA---DIVSEMPINNNP-
tr|Q93KT5|Q93KT   ASKVML---LVMDPISLLIPVLE--------K
sp|P0A1M6|SPAR_   KVQFFL---GEHKFFTN---LFV---------R
tr|Q60238|Q6023   GTGQLLKLVDLKSLLTLLVQV-------P---
tr|Q83R27|Q83R2   CEHLFS---EIFNLLA---DIISELPL---I-
tr|A0A024L995|A   RVMPLS---FFPEQLQL---YIE---------K
tr|A0A0C3MET9|A   CEHLFS---EMFNLLA---DIISELPL---I-
tr|O24978|O2497   FKEEIS---LAFSAISK---I----------F

cons                    :         :
```


Inside
Outside
TM-helix (IN->OUT)
TM-helix (OUT->IN)
Signal peptide
InvA
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
SpaP
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
SpaQ
TOPCONS
OCTOPUS
Phillius
PolyPhobius
SCAMPI
SPOCTOPUS
SpaR
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS
SpaS
TOPCONS
OCTOPUS
Philius
PolyPhobius
SCAMPI
SPOCTOPUS

SignalP 4.1 Server - prediction results

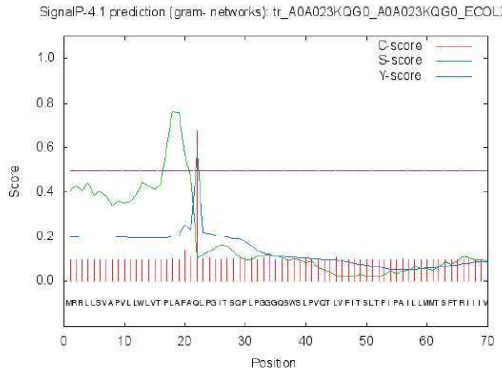
Technical University of Denmark

```
>sp_P54700_FLIP_SALTY  Flagellar biosynthetic protein FliP OS_Salmonella typhimurium _strain LT2 _ SGSC1412 _ ATCC 700720_ GN_fliP PE_1 SV_1
```



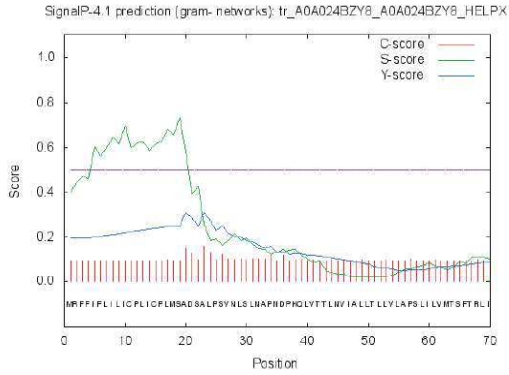

```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    22        0.484
  max. Y    22        0.590
  max. S    18        0.872
  mean S    1-21      0.717
       D    1-21      0.637    0.510    YES
Name=sp_P54700_FLIP_SALTY       SP='YES' Cleavage site between pos. 21 and 22: AAA-QL D=0.637 D-cutoff=0.510 Networks=SignalP-TM
# data
# gnuplot script
```

```
>tr_A0A023KQG0_A0A023KQG0_ECOLX  Flagellar biosynthetic protein FliP OS_Escherichia coli GN_fliP PE_3 SV_1
```




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    22        0.675
  max. Y    22        0.580
  max. S    18        0.761
  mean S    1-21      0.454
       D    1-21      0.533    0.510    YES
Name=tr_A0A023KQG0_A0A023KQG0_ECOLX     SP='YES' Cleavage site between pos. 21 and 22: AFA-QL D=0.533 D-cutoff=0.510 Networks=SignalP-TM
# data
# gnuplot script
```

```
>tr_A0A024BZY8_A0A024BZY8_HELPX  Flagellar biosynthetic protein FliP OS_Helicobacter pylori GN_fliP PE_3 SV_1
```




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    23        0.160
  max. Y    20        0.311
  max. S    19        0.732
  mean S    1-19      0.594
       D    1-19      0.415    0.510    NO
Name=tr_A0A024BZY8_A0A024BZY8_HELPX     SP='NO' D=0.415 D-cutoff=0.510 Networks=SignalP-TM
# data
# gnuplot script
```

Manuscript III

Stoichiometry analysis of macromolecular membrane protein complexes

This research was originally published in Biological Chemistry

Zilkenat S., Grin I., Wagner S. Stoichiometry analysis of macromolecular membrane protein complexes. *Biological Chemistry.* 2016; Vol: 398(2):155-164. DOI: https://doi.org/10.1515/hsz-2016-0251.

Review

Susann Zilkenat, Iwan Grin and Samuel Wagner*

Stoichiometry determination of macromolecular membrane protein complexes

DOI 10.1515/hsz-2016-0251
Received July 15, 2016; accepted September 20, 2016; previously published online September 23, 2016

Abstract: Gaining knowledge of the structural makeup of protein complexes is critical to advance our understanding of their formation and functions. This task is particularly challenging for transmembrane protein complexes, and grows ever more imposing with increasing size of these large macromolecular structures. The last 10 years have seen a steep increase in solved high-resolution membrane protein structures due to both new and improved methods in the field, but still most structures of large transmembrane complexes remain elusive. An important first step towards the structure elucidation of these difficult complexes is the determination of their stoichiometry, which we discuss in this review. Knowing the stoichiometry of complex components not only answers unresolved structural questions and is relevant for understanding the molecular mechanisms of macromolecular machines but also supports further attempts to obtain high-resolution structures by providing constraints for structure calculations.

***Corresponding author: Samuel Wagner,** University of Tübingen, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), Section of Cellular and Molecular Microbiology, Elfriede-Aulhorn-Str. 6, D-72076 Tübingen, Germany; and German Center for Infection Research (DZIF), Partner-site Tübingen, Elfriede-Aulhorn-Str. 6, D-72076 Tübingen, Germany,
e-mail: samuel.wagner@med.uni-tuebingen.de
Susann Zilkenat: University of Tübingen, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), Section of Cellular and Molecular Microbiology, Elfriede-Aulhorn-Str. 6, D-72076 Tübingen, Germany
Iwan Grin: University of Tübingen, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), Section of Cellular and Molecular Microbiology, Elfriede-Aulhorn-Str. 6, D-72076 Tübingen, Germany; and German Center for Infection Research (DZIF), Partner-site Tübingen, Elfriede-Aulhorn-Str. 6, D-72076 Tübingen, Germany

Keywords: mass spectrometry; membrane proteins; protein complexes; stoichiometry.

Introduction

Membranous lipid bilayers act as sophisticated barriers in biology by facilitating intracellular compartmentalization of different functions as well as separation of individual cells. The selectivity of these barriers is achieved by a diverse set of proteins traversing these membranes, which fulfill functions like nutrient, ion, or metabolite transport, communication, or adhesion. Many of these functions, in particular the transport of large macromolecules like proteins, DNA, or RNA, require very large multi-subunit complexes that sometimes span several biological membranes. To understand functional mechanisms of these complexes on a molecular level, structural information of the macromolecules involved is indispensable and so the obtainment of a high-resolution structure for a given protein assembly is often the 'Holy Grail'. While the solution of high-resolution structures has seen enormous successes in the last decade, most recently due to major enabling improvements in cryo electron microscopy (EM) (Callaway, 2015), the analysis of very large complexes, in particular of large transmembrane complexes, remains very challenging. In this review, we will point out the major difficulties in structure determination of large transmembrane complexes and present solutions to determine the stoichiometry of these complexes as an important first step towards structure elucidation if high resolution structures remain unachieved. The topic is illustrated by focussing on four different transmembrane complexes: the 650 kDa *Thermus thermophilus* V-ATPase (*Tt*ATPase), the 3 MDa type III secretion system (T3SS) found in Gram-negative pathogens, the 5 MDa plant cellulose synthesis complex and lastly the 50 MDa (yeast) to 110 MDa (human) nuclear pore complex (Table 1). Figure 1 illustrates the size range and complexity of three of these large transmembrane complexes addressed in this review.

Table 1: Examples for macromolecular membrane complexes.

Complex	Size [MDa]	Number of different proteins	Total number of proteins	Occurrence	Methods used (s. Table 2)	References
Thermus thermophilus-V-ATPase	0.65	9	26	Homologues in bacteria and eukaryotes	Cryo electron microscopy (16 Å resolution); X-ray crystallography (peripheral stalk); Native mass spectrometry (mass determination)	(Lee et al., 2010; Lau and Rubinstein, 2011; Zhou et al., 2011; Chait et al., 2016)
Type III secretion system	3 (without cytosolic components)	≤ 20	~ 200	Gram negative pathogenic bacteria (e.g. *Salmonella* Typhimurium)	Cryo electron microscopy (10 Å resolution); X-ray crystallography (C-terminal domain of export apparatus protein); Stable isotope-labeled standards mass spectrometry; Photobleaching (translocator)	(Worrall et al., 2010; Schraidt and Marlovits, 2011; Romano et al., 2016; Zilkenat et al., 2016)
Cellulose synthesis complex	> 5	3	18	Plants	Cryo electron microscopy; Quantitative immunoblotting (CESA4, 7 and 8); Photobleaching (CESA3)	(Kimura et al., 1999; Chen et al., 2014; Hill et al., 2014)
Nuclear pore complex	~ 50 (*Saccharomyces cerevisiae*), ~ 110 (human)	~ 30 (*S. cerevisiae*)	≥ 456 (*S. cerevisiae*)	Eukaryotes	Cryo electron microscopy (human, 23 Å resolution); Cross-linking mass spectrometry; X-ray crystallography (coat nucleoporin complex); Stable isotope-labeled standards mass spectrometry (6 nuclear pore complex proteins)	(Rout et al., 2000; Alber et al., 2007; Stuwe et al., 2015; von Appen et al., 2015; Kosinski et al., 2016; Lin et al., 2016)

The methods mentioned in this review which were used to study these complexes are further detailed in Table 2.

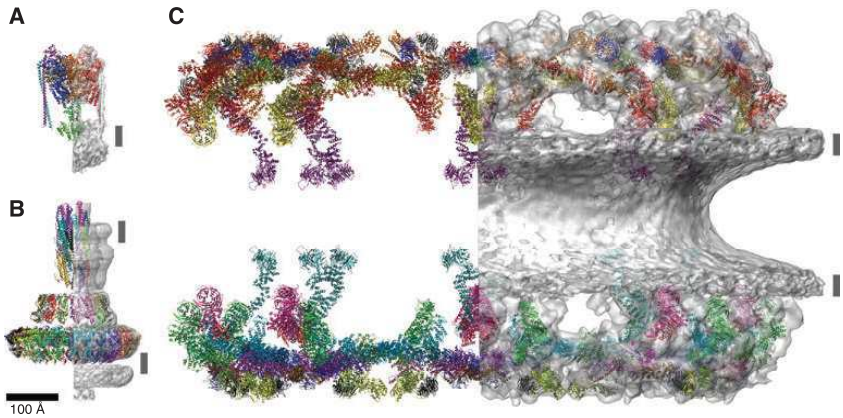


Figure 1: Structures of three macromolecular membrane protein complexes discussed in this review.
Membrane protein complexes range from relatively small assemblies of few subunits to enormous machineries of up to 100 MDa comprised of dozens of components. Structures are to scale; individual protein chains were colored for easier discrimination; experimental electron densities shown in gray on the right half of the structure; dark gray bars represent the approximate position of the membrane(s). (A) *T. thermophilus* V-ATPase (Lau and Rubinstein 2011), (B) *S.* Typhimurium type III secretion system (Schraidt and Marlovits, 2011; Radics et al., 2013; Demers et al., 2014; Bergeron et al., 2015) and (C) the human nuclear pore complex (von Appen et al., 2015). Visualized using Visual Molecular Dynamics (Humphrey et al., 1996). Scale bar: 100 Å.

Challenges for high-resolution structural analysis of large transmembrane complexes

The major obstacles for the structural analysis of large transmembrane complexes are (a) the sheer size of these complexes, often exceeding 1 MDa (Figure 1); (b) the pronounced hydrophobicity of their transmembrane domains; (c) the structural complexity involving many different complex components; and (d) the dynamic heterogeneity many complexes exhibit during their functional cycles. For these reasons, it is often difficult to obtain high-resolution structures of large transmembrane complexes with the methods currently available, even more so, if structures of different functional states are required to elucidate functional mechanisms at the molecular level.

Both electron and X-ray crystallography depend on the successful crystallization of proteins and protein complexes. There have been many improvements in the production of membrane proteins for crystallization (Clark et al., 2011; Schlegel et al., 2014) and the formation and stabilization of crystals (Carpenter et al., 2008; Klara et al., 2016). However, as seen in crystal structures of components of the *Tt*ATPase, different subunits of type III secretion systems, as well as the nuclear pore complex (Lee et al., 2010; Worrall et al., 2010; Stuwe et al., 2015), mostly structures of soluble components or extramembrane domains of transmembrane proteins have been solved for large membrane-spanning complexes. While providing immense help in the elucidation of their overall structure and function, lack of structural information on the hydrophobic core of these complexes impedes a complete understanding of their molecular mechanisms.

Single particle cryo EM, a method in which datasets of many 2D electron micrographs of single particles are averaged to generate a high resolution 3D map (Lau and Rubinstein 2010), is playing an increasingly important role in structural analysis of proteins and has been used successfully to gain a better understanding of the makeup of many membrane-spanning macromolecular complexes. A 16 Å resolution 3D map of the 0.65 MDa *Tt*-V-ATPase was generated using single particle cryo EM. A surface view of the 3 MDa needle complex of a *Salmonella* type III secretion system was reconstructed to a resolution of 10 Å, helping to establish the stoichiometry of the three transmembrane ring forming elements (Schraidt and Marlovits, 2011). Furthermore, aided by cryo EM maps, the 110 MDa human nuclear pore complex could be reconstructed to a resolution of 23 Å (von Appen et al., 2015) (Figure 1). Significant improvements in direct electron detection have most recently boosted single particle cryo EM analysis and have enabled the solution of structures at sub-nanometer resolution down to 3.4 Å, even for some transmembrane domains (Bai et al., 2015; De Zorzi et al., 2016). Despite these important improvements, flexibility of protein domains and bound lipids and detergents limit the power of cryo EM for many macromolecular complexes, in particular in their transmembrane regions.

Solid state nuclear magnetic resonance (ssNMR) spectroscopy enables the study of membrane protein structures in their native or native-like environment. While in theory molecular weight is not a limiting factor in ssNMR (Brown and Ladizhansky, 2015) and large improvements have been made in sample preparation, hardware and experimental design, such as magic angle spinning, no structures of large membrane-spanning macromolecular complexes have been solved using this method, so far (Shahid et al., 2012; Goldbourt, 2013; Ward et al., 2014).

Stoichiometry determination as a first step towards structure elucidation

As high-resolution structures are out of reach for many large transmembrane complexes at the moment, an important first step towards structural and functional understanding of these macromolecular assemblies is the determination of the stoichiometry of their components.

Towards this end, the only available direct strategy is subunit counting by single-molecule fluorescence photobleaching. Photobleaching of protein complexes reduces their fluorescence intensity in increments that correspond to the number of fluorescent protein-tagged protomers in the complex at hand (Ulbrich and Isacoff, 2007). The notable advantage of this approach is the ability to count complex components in their native context, without the need of prior purification. This even allows analysis of less stable complexes that resist purification as well as dynamically associated complex components (Diepold et al., 2015). However, the approach requires the functional fusion of a fluorescent protein to complex component of interest, a premise that is often not met for very hydrophobic proteins or components buried deeply within their respective complex. Nonetheless, counting by single-molecule photobleaching has been used successfully to resolve the stoichiometry of transmembrane proteins of different transmembrane channels (Hines, 2013) as well as type III secretion and flagellar components (Leake et al., 2006; Morimoto et al., 2014; Romano et al.,

2016), TAT translocon components (Leake et al., 2008), or components of the *Arabidopsis thaliana* cellulose synthesis complex (Chen et al., 2014).

For components of membrane-spanning macromolecular complexes that are incompatible with direct counting by single-molecule photobleaching, indirect approaches of stoichiometry determination have to be employed. As a first step, most indirect approaches require the prior isolation of a homogeneous preparation of the complex of interest and the identification of the components contained therein. The extraction of membrane protein complexes from the lipid bilayer is facilitated by detergents and thus the choice of detergents plays a critical role in complex purification. Typically, mild, non-ionic detergents like dodecyl-maltoside are prefered, however, the detergent that optimally retains the integrity of the transmembrane complex of interest needs to be identified empirically (Arnold and Linke, 2008). Isolation of the complex is best achieved by affinity purification strategies, where the bait is a component that assembles last into the complex and requires the presence of all other components for its own assembly. This strategy ensures isolation of completely assembled complexes and avoids handling of a heterogeneous mixture of assembly intermediates. Often a final size-based separation of affinity-purified complexes is necessary to obtain a homogeneous preparation of only one complex species. Towards this end, size exclusion chromatography (SEC) is suitable for complexes in solution while blue native PAGE (see details below) is well-suited for separation of membrane protein complexes and subsequent in gel digestion for analysis by mass spectrometry. Blue native PAGE in combination with mass spectrometry or immunoblotting is also well suited as a measure of compositional analysis and quality control of isolated complexes. A visual inspection of the complexes is best achieved by EM of negative stained material.

The determination of complex stoichiometry by most means requires the separate analysis of complex mass and component ratios, from which the stoichiometry can be calculated. Below, we first present a brief overview of approaches to determine the mass of protein complexes and subsequently discuss the analysis of the ratio of protein components in more detail.

Analysis of the molecular mass of protein complexes

Blue native PAGE is a commonly used technique to analyze the composition and mass of membrane protein complexes in a relatively simple format that does not require prior purification of the analyte. The technique utilizes the size-dependent differential migration of Coomassie G-charged membrane protein complexes in polyacrylamide gradient gels under non-denaturing conditions (Schägger and von Jagow, 1991). It allows for the separation of protein complexes between 20 kDa and 10 MDa (Wittig et al., 2006; Weiland et al., 2014), however, the resolution of the technique is very limited, in particular above 1 MDa. Its analytic power is further limited by the fact that bound lipids, detergents, and Coomassie affect the migration behavior of membrane proteins, which often leads to an overestimation of the mass of analyzed complexes (Stenberg et al., 2005; Hill et al., 2014). Blue native PAGE has been used extensively for the analysis of composition and assembly of complexes of bacterial type III, IV, and VII secretion systems (Krall et al., 2002; Wagner et al., 2010; Houben et al., 2012; Kuroda et al., 2015), however, due to the large size of these complexes, reliable masses or stoichiometry could not be determined. Over all, blue native PAGE is a suitable technique to obtain an estimate of complex size and composition but it cannot be used for a mass analysis that is sufficiently accurate to support stoichiometry determination.

Size exclusion chromatography (SEC) is another commonly employed method to determine the molecular mass of purified soluble proteins up to 2 MDa, however, like for blue native PAGE, bound lipids and detergents alter the stokes radius of membrane proteins and thus compromise the use of SEC for the accurate estimation of their mass (Slotboom et al., 2008). This problem can be partially overcome by complementing the on-line measurement of the absorbance of SEC-separated protein complexes with measurements of their light scattering and refractive index (a.k.a. SEC-multi angle laser light scattering, SEC-MALLS). This approach allows to separately calculate the masses of protein and detergent content (Folta-Stogniew, 2006) for non-ionic detergents that are transparent at a wavelength of 280 nm (Arnold and Linke, 2008). The typical mass error of SEC-MALLS is 5–10% for membrane protein-detergent complexes, which makes the technique suitable for determination of the oligomeric state of complexes made of up to 20 components (Slotboom et al., 2008).

Analytical ultracentrifugation allows the determination of sedimentation velocity or sedimentation equilibrium of macromolecular particles, which can be used to calculate the size, mass, composition and interaction of macromolecules in solution (Ebel, 2011). For the analysis of membrane protein-detergent complexes, the analysis of sedimentation velocity is preferred over sedimentation equilibrium, since it has fewer constraints in terms

of homogeneity and stability of the sample (Ebel, 2011; Zhao et al., 2013). Depending on the centrifugation set-up proteins of a size between 100 Da and 100 MDa can be studied (Zhao et al., 2013); for instance part of the TonB-dependent energy transduction system of ~260–520 kDa was analyzed using measurements of its sedimentation velocity.

The most accurate approach for mass determination of protein complexes is native mass spectrometry in which (membrane) protein (-detergent) complexes are brought into the gas phase by direct electrospray ionization (ESI) (Fenn et al., 1989; Laganowsky et al., 2014). Membrane proteins are liberated of detergent by collision-induced dissociation in an inert gas, after which their mass can be analyzed accurately (Barrera et al., 2008; Borysik and Robinson, 2012; Reading et al., 2015). Collision-induced dissociation does not only remove detergents but can also dissociate complexes and thus enable their compositional analysis (Laganowsky et al., 2014). Native mass spectrometry requires only small amounts of purified complexes (pmol or less) (Wohlgemuth et al., 2015) and allows for the mass determination of membrane protein complexes of up to 700 kDa as exemplified by the analysis of V-type ATPases from *T. thermophilus* and *Enterococcus hirae* (Zhou et al., 2011).

Analysis of component ratios of protein complexes

Once the constituting components of a complex and its mass have been determined, the ratio of components needs to be resolved in order to enable the calculation of their stoichiometry.

Gel-based approaches

Densitometry of Coomassie-stained protein bands of SDS PAGE-resolved components of purified protein complexes is a commonly employed approach to estimate the ratio of complex components. While this approach is technically very simple, quick, and does not demand specialized equipment, accuracy suffers from differential staining efficiencies of different proteins. These depend on the extent of unfolding in the gel, bound detergent, and number of basic residues. Independent of these shortcomings is the use of [^{35}S]-Met/Cys labeling for autoradiography of protein samples, which also features a high dynamic range and a linear relationship between detection signal and protein amount. This classical approach has been used for instance for the first estimation of the stoichiometry of bacterial flagella (Jones et al., 1990). Also immunoblotting of complex components instead of staining offers a more robust basis for quantification but requires well-characterized specific antibodies or the use of epitope tags for each protein studied as well as purified individual complex components for generation of a standard curve. A major obstacle in quantitative immunoblotting is the low dynamic range and quick band saturation of chemiluminescence-based detection systems, for which reason the detection via (near-infrared-)fluorescent secondary antibodies is preferable. A more detailed discussion of the strengths and weaknesses of quantitative immunoblotting has recently been presented by McDonough et al. (2014). Its use for stoichiometry determination is exemplified by a report of a 1 : 1 : 1 ratio of the three different CESA proteins of the cellulose synthesis complex (Hill et al., 2014).

Mass spectrometry-based approaches

Because of their high precision, quantitative power, and multiplexing capability, mass spectrometric approaches are the preferred choice for the determination of ratios of complex components. State of the art is the use of stable isotope-labeled standards for direct ratiometric or absolute quantification of complex components. The strategy is based on co-analyzing stable isotope-labeled standard peptides of complex components together with a non-labeled sample of the complete complex to determine the ratio between labeled and non-labeled material (Figure 2A). Besides a homogeneous and well-defined preparation of the analyte complex, this approach also requires a careful selection of standard peptides, which can be challenging for membrane-spanning complexes. Hydrophobic transmembrane proteins tend to have few tryptic peptides due to the sparsity of positively charged arginine and lysine residues in the transmembrane segments. Additionally, SDS-resistant secondary structures of transmembrane domains often reduce digestion efficiency thus leading to frequent missed cleavages. Even though sequence coverage has been significantly improved by the use of alternative enzymes like thermolysin or alternative chromatographic or SDS PAGE separations, some transmembrane domains remain resistant to detection and analysis (Whitelegge, 2013).

Stable isotope-labeled standard peptide strategies exist in three principle variations, which are described in more detail below: Absolute quantification (AQUA) peptides (Gerber et al., 2003), QconCAT artificial proteins

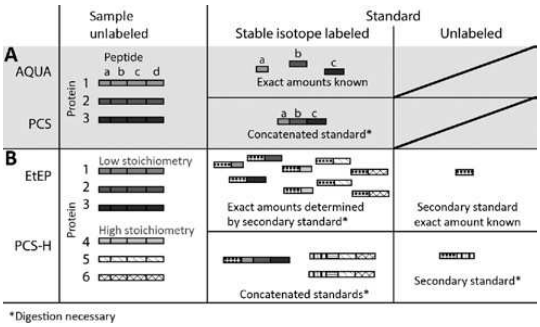


Figure 2: Comparison of (A) basic and (B) hierarchical approaches of stable isotope-labeled standards mass spectrometry.
(A) For the absolute quantification (AQUA) strategy and the peptide concatenated standard (PCS) strategy concentrations and ratios of the proteins are determined by measuring the ratios between the unlabeled sample and a stable isotope labeled standard. (B) For the equimolarity through equalizer peptide (EtEP) strategy and the hierarchical PCS strategy, a stable isotope labeled standard of different amount can be added, depending on the predicted stoichiometry range. For the EtEP strategy amounts of labeled standards are determined by the secondary standard of known amount. Concentrations of unlabeled sample peptides are measured analogous to AQUA. For the hierarchical PCS strategy ratios are determined by measurements between both the unlabeled sample and the labeled standard, as well as between the labeled standard and an unlabeled secondary standard.

(Pratt et al., 2006), or peptide concatenated standards (PCS) (Kito et al., 2007).

For the AQUA strategy, synthetic stable isotope-labeled, i.e. 'heavy', peptides are used as standards. These peptides, selected from tryptic peptides of the proteins of interest, are accurately quantified and spiked into the unlabeled, i.e. 'light', purified complex for subsequent mass spectrometry-based ratiometric or absolute quantification. This method has been used very successfully in quantifying the level of posttranslational modifications (Gerber et al., 2003; Chahrour et al., 2015), in analyzing complexes of lower stoichiometric range (Schmidt et al., 2010; Wohlgemuth et al., 2015), and in defining part of the human nuclear pore complex (von Appen et al., 2015). While this approach is well suited for absolute quantification, it bears the risk of obtaining a skewed stoichiometry of complex components since the spiking with synthetic peptides does not take into account differences in digestion efficiencies that occur in particular for transmembrane proteins. A major drawback of the approach is also the high cost factor of quantified synthetic isotope-labeled peptides.

QconCAT (previously termed QCAT) uses standard peptides of components of the complex of interest concatenated into one artificial protein at a 1:1 ratio. As the artificial standard protein is labeled with heavy isotopes by the SILAC approach (Ong, 2002), the production of the standard is very cost-efficient. The 1:1 ratio of concatenated standard peptides allows for ratiometric quantification when the heavy standard protein and the light complex are co-digested and analyzed by mass spectrometry. If the difference in protein concentration in the complex of interest is greater than one order of magnitude, it becomes advisable to design several QconCAT proteins with peptides grouped according to their estimated abundance (Pratt et al., 2006). The QconCAT strategy has been used to determine the stoichiometry of several membrane associated protein complexes (Nanavati et al., 2008; Olinares et al., 2011) but not for transmembrane complexes.

The PCS strategy takes an approach very similar to QconCAT. However, in addition to just the tryptic peptides of interest, native flanking regions upstream and downstream of each peptide are also included in the concatenated standard (Kito et al., 2007). Since flanking regions influence the efficiency of tryptic digestion, this trick results in a more similar digestion behavior of sample and standard, in particular for transmembrane complexes (Kito and Ito, 2008). We were able to use the PCS strategy to determine the stoichiometry of the complete needle complex of a type III secretion system of *Salmonella* Typhimurium, a multi-MDa transmembrane complex with a stoichiometric range of 1–24 (disregarding the needle filament itself) (Zilkenat et al., 2016). Knowledge of the stoichiometry of three of its components (Schraidt and Marlovits, 2011) allowed the calculation of absolute numbers per complex from protein : protein ratios obtained by the PCS approach.

Both, PCS and AQUA approaches have been refined to cover a wider dynamic range of ratios (Figure 2B). The equimolarity through equalizer peptide (EtEP) strategy adds a non-native equalizing peptide to each synthesized, stable isotope labeled standard peptide, separable by a trypsin cleavage site. This allows quantification of all standard peptides while knowing only the absolute amount of the non-labeled version of this equalizer peptide (Holzmann et al., 2009). This feature improves the cost efficiency of the AQUA approach and allows for more accurate results since a higher number of peptides from the same protein can be analyzed in parallel.

Hierarchical PCS were introduced to quantify more proteins over a wider dynamic range (Kito et al., 2016). Each stable isotope-labeled primary PCS is extended by a unique ID-tag peptide, which sequence is not part of the complex of interest. The ID-tags of different PCSs are then concatenated

Table 2: Overview of methods for determination of structure, stoichiometry, molecular mass, or ratios of components.

Experimental output	Method	Technique	Range (stoichiometry[a], mass[b], ratio[c])	Size of membrane protein complex studied	Prerequisites (purification steps necessary[d], e.g. immuno precipitation)	References
Proteins per complex	Single molecule photobleaching	–	Good estimation < 5 subunits[a]	0.5 MDa (Romano et al., 2016)	Tagging with, e.g. GFP; Stoichiometries ≥ 5 require inference model	(Ulbrich and Isacoff 2007; Hines 2013)
Morphological maps	Cryo-electron microscopy	–	Dependent on resolution, complexity of the complex, dissimilarity of proteins[a]	110 MDa (von Appen et al., 2015)	Large structures preferably with symmetry Specialized transmission electron microscope (cost disadvantage) For single particle analysis: direct electron detector Automated data collection and processing	(Schraidt and Marlovits 2011; Eisenstein 2016)
Atomic details of individual proteins and small complexes	Crystallography	X-ray crystallography	Good estimation for soluble domains of membrane proteins[a]	Focus often on domains of the proteins outside of the membrane helices	Formation of stable, diffracting crystals	(Clark et al., 2011; Stuwe et al., 2015)
		Electron crystallography	Monomers or homomers between 20 and 250 kDa[a]	220 kDa (Hite et al., 2010)	Formation of stable, diffracting crystals; Crystallization within a membrane possible	(Wisedchaisri et al., 2011)
Absolute molar mass	Native mass spectrometry	–	1 kDa and 1 MDa[b]	650 kDa (Zhou et al., 2011)	No labeling or crosslinking required; Customized mass spectrometer; Only pmol amount of protein needed	(Heck 2008; Erba and Petosa 2015)
	Size exclusion chromatography	SEC-MALLS	<10 kDa to approx. 500 kDa[b]	–	Manual data analysis; Well defined standard proteins	(Folta-Stogniew 2006)
	Analytical ultracentrifugation	Sedimentation velocity	100 Da and 100 MDa[b]	260–520 kDa (Salvay et al., 2007)	[d]	(Inagakia et al., 2014)
	Blue native PAGE	–	10 kDa–1.2/10 MDa[b]	> 1.2 MDa (Wagner et al., 2010)	Optimization of size standards; Factor in bound lipids and detergents; Dependent on stability of the complex[b,d]	(Wittig et al., 2006; Weiland et al., 2014)
Ratios of proteins in a complex	Quantitative immunoblotting	–	Dependent on protein, antibody and experimental conditions[c]	700–840 kDa (Hill et al., 2014)	Specific antibodies; Demonstration of the linearity of each protein	(McDonough et al., 2014)
	Etable isotope-labeled standard mass spectrometry	Peptide concatenated standards	10-fold for single standards, 100-fold for hierarchical standards[c]	3 MDa (Zilkenat et al., 2016)	Preparation of concatenated standards[d]	(Kito et al., 2007, 2016)
		QconCAT artificial proteins	Only used for complexes with very narrow stoichiometric range so far[c]	–		(Pratt et al., 2006)
Ratios of proteins in a complex or total amount of protein per cell	Stable isotope-labeled standard mass spectrometry	Absolute quantification peptides	Only used for complexes with very narrow stoichiometric range so far[c]	–	Standard peptides in exact amount[d]; Complete digestion	(Gerber et al., 2003; von Appen et al., 2015)
		Equimolarity through equalizer peptide		–	One single peptide in exact amount (cost advantage)[d]; Complete digestion	(Holzmann et al., 2009)

in a non-labeled secondary PCS. The peptides are grouped in different PCSs according to the estimated abundance of the respective complex components. This allows the adjustment of the concentration of the standards so that every single ratio of labeled : unlabeled peptides remains below 10, as ratios above 10 have been reported to lead to measurement errors in complex mixtures (Hanke et al., 2008). As this method has been published very recently, no data on its use on transmembrane complexes is available yet. However, based on our good experience with the standard PCS strategy, hierarchical PCS should be a powerful development.

Conclusion

In the last decade, immense improvements have been made towards the determination of membrane protein structures (Hendrickson, 2016) but the study of large structures in the MDa range remains challenging, even though the number of tools available to researchers has been significantly expanded (Table 2). Cryo-EM and crystallography can offer great insights on atomic details, however, obtaining high resolution images of membrane protein complexes above 0.5 MDa is often difficult and even if atomic resolution structures can be achieved, parts of the structure, especially transmembrane domains, may remain unresolved (Schraidt and Marlovits, 2011; Stuwe et al., 2015). Elucidating the stoichiometry of the components comprised in a complex by other methods can both support the structural efforts as well as answer unresolved questions. Especially for 'ill-behaving' membrane complexes, it is worth to study the components making up the structure of a complex in parallel to pursuing an atomic resolution structure.

To resolve the stoichiometry of a complex, the mass of the total complex as well as the ratio of each protein towards each other has to be known. Recent progress in native MS has led to the capability of measuring greater masses, allowing for calculation of complete complexes. However, investigations of membrane complexes have not yet reached the same mass range as, e.g. bacteriophage particles (Chait et al., 2016). For very large transmembrane protein complexes, analytical ultracentrifugation is a very suitable choice for mass determination while smaller complexes below 1 MDa are also approachable by SEC-MALLS. To determine the ratios of transmembrane complex proteins, stable isotope-labeled standard mass spectrometry based approaches are promising as ratios of all complex components can be analyzed in one multiplex run. We were able to obtain stoichiometries ranging from 1 to 24 for a 3 MDa complex, using an AQUA-complemented PCS strategy (Zilkenat et al., 2016). Exploiting the recently introduced hierarchical standards may further improve both range and error margins of this approach and thus may facilitate the robust and accurate stoichiometry determination of large membrane-spanning protein complexes.

Acknowledgments: Work in the laboratory of S.W. relevant to this review was supported by the Alexander von Humboldt Foundation in the framework of the Sofja Kovalevskaja Award endowed by the Federal Ministry of Education and Research.

References

Alber, F., Dokudovskaya, S., Veenhoff, L.M., Zhang, W., Kipper, J., Devos, D., Suprapto, A., Karni-Schmidt, O., Williams, R., Chait, B.T., et al. (2007). Determining the architectures of macromolecular assemblies. Nature *450*, 683–694.

Arnold, T. and Linke, D. (2008). The use of detergents to purify membrane proteins. Curr. Protoc. Protein Sci. *53*, 4.8.1–4.8.30.

Bai, X., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S.H.W., and Shi, Y. (2015). An atomic structure of human γ-secretase Xiao-chen. Nature *512*, 212–217.

Barrera, N.P., Di Bartolo, N., Booth, P.J., and Robinson, C.V. (2008). Micelles protect membrane complexes from solution to vacuum. Science *321*, 243–247.

Bergeron, J.R.C., Worrall, L.J., De, S., Sgourakis, N.G., Cheung, A.H., Lameignere, E., Okon, M., Wasney, G.A., Baker, D., McIntosh, L.P., et al. (2015). The modular structure of the inner-membrane ring component prgk facilitates assembly of the type III secretion system basal body. Cell *23*, 161–172.

Borysik, A.J. and Robinson, C.V. (2012). The 'sticky business' of cleaning gas-phase membrane proteins: a detergent oriented perspective. Phys. Chem. Chem. Phys. *14*, 14439–14449.

Brown, L.S. and Ladizhansky, V. (2015). Membrane proteins in their native habitat as seen by solid-state NMR spectroscopy. Protein Sci. *24*, 1333–1346.

Callaway, E. (2015). The revolution wiill not be crystallized. Nature *525*, 172–174.

Carpenter, E.P., Beis, K., Cameron, A.D., and Iwata, S. (2008). Overcoming the challenges of membrane protein crystallography. Curr. Opin. Struct. Biol. *18*, 581–586.

Chahrour, O., Cobice, D., and Malone, J. (2015). Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. J. Pharm. Biomed. Anal. *113*, 2–20.

Chait, B.T., Cadene, M., Olinares, P.D., Rout, M.P., and Shi, Y. (2016). Revealing higher order protein structure using mass spectrometry. J. Am. Soc. Mass Spectrom. *27*, 952–965.

Chen, Y., Deffenbaugh, N.C., Anderson, C.T., and Hancock, W.O. (2014). Molecular counting by photobleaching in protein complexes with many subunits: best practices and application to the cellulose synthesis complex. Mol. Biol. Cell *25*, 3630–3642.

Clark, K.M., Fedoriw, N., Robinson, K., Connelly, S.M., Randles, J., Malkowski, M.G., Detitta, G.T., and Dumont, M.E. (2011). Purification of transmembrane protein from *Saccharomyces cerevisiae* for X-ray crystallography. Protein Expr. Purif. *71*, 207–223.

Demers, J.-P., Habenstein, B., Loquet, A., Kumar Vasa, S., Giller, K., Becker, S., Baker, D., Lange, A., and Sgourakis, N.G. (2014). High-resolution structure of the *Shigella* type-III secretion needle by solid-state NMR and cryo-electron microscopy. Nat Commun *5*, 4976.

De Zorzi, R., Mi, W., Liao, M., and Walz, T. (2016). Single-particle electron microscopy in the study of membrane protein structure. Microscopy *65*, 81–96.

Diepold, A., Kudryashev, M., Delalez, N.J., and Berry, R.M. (2015). Composition, formation, and regulation of the cytosolic C-ring, a dynamic component of the type III secretion injectisome. PLoS Biol. *13*, 1–21.

Ebel, C. (2011). Sedimentation velocity to characterize surfactants and solubilized membrane proteins. Methods *54*, 56–66.

Eisenstein, M. (2016). The field that came in from the cold. Nat. Methods *13*, 19–22.

Erba, E.B., and Petosa, C. (2015). The emerging role of native mass spectrometry in characterizing the structure and dynamics of macromolecular complexes. Protein Sci. *24*, 1176–1192.

Fenn, J.B., Mann, M., Meng, C.K.A.I., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64–71.

Folta-Stogniew, E. (2006). Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. Methods Mol. Biol. *328*, 97–112.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad Sci. USA *100*, 6940–6945.

Goldbourt, A. (2013). Biomolecular magic-angle spinning solid-state NMR: recent methods and applications. Curr. Opin. Biotechnol. *24*, 705–715.

Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. (2008). Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. J. Proteome Res. *7*, 1118–1130.

Heck, A.J.R. (2008). Native mass spectrometry: a bridge between interactomics and structural biology. Nat. Methods 5, 927–933.

Hendrickson, W.A. (2016). Atomic-level analysis of membrane-protein structure. Nat. Struct. Mol. Biol. *23*, 464–467.

Hill, J.L., Hammudi, M.B., and Tien, M. (2014). The *Arabidopsis* cellulose synthase complex: a proposed hexamer of CESA trimers in an equimolar stoichiometry. Plant Cell *26*, 4834–4842.

Hines, K.E. (2013). Inferring subunit stoichiometry from single molecule photobleaching. J. Gen. Physiol. *141*, 737–746.

Hite, R.K., Li, Z., and Walz, T. (2010). Principles of membrane protein interactions with annular lipids deduced from aquaporin-0 2D crystals. EMBO J. *29*, 1652–1658.

Holzmann, J., Pichler, P., Madalinski, M., Kurzbauer, R., and Mechtler, K. (2009). Stoichiometry determination of the MP1-p14 complex using a novel and cost-efficient method to produce an equimolar mixture of standard peptides. Anal. Chem. *81*, 10254–10261.

Houben, E.N.G., Bestebroer, J., Ummels, R., Wilson, L., Piersma, S.R., Jiménez, C.R., Ottenhoff, T.H.M., Luirink, J., and Bitter, W. (2012). Composition of the type VII secretion system membrane complex. Mol. Microbiol. *86*, 472–484.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Inagakia, S., Ghirlandob, R., and Grisshammer, R. (2014). Biophysical characterization of membrane proteins in nanodiscs. Methods *59*, 287–300.

Jones, C.J., Macnab, R.M., Okino, H., and Aizawa, S.I. (1990). Stoichiometric analysis of the flagellar hook-(basal-body) complex of Salmonella typhimurium. J. Mol. Biol. *212*, 377–387.

Kimura, S., Laosinchai, W., Itoh, T., Cui, X., Linder, C., and Brown, R. (1999). Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant vigna angularis. Plant Cell *11*, 2075–2086.

Kito, K. and Ito, T. (2008). Mass spectrometry-based approaches toward absolute quantitative proteomics. Curr. Genomics *9*, 263–274.

Kito, K., Ota, K., Fujita, T., and Ito, T. (2007). A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. J. Proteome Res. *6*, 792–800.

Kito, K., Mitsuhiro, O., Ishibashi, Y., Okada, S., and Ito, T. (2016). A strategy for absolute proteome quantification with mass spectrometry by hierarchical use of peptide-concatenated standards. Proteomics *16*, 1457–1473.

Kosinski, J., Mosalaganti, S., von Appen, A., Teimer, R., Diguilio, A.L., Wan, W., Bui, K.H., Hagen, W.J., Briggs, J.A., Glavy, J.S., et al. (2016). Molecular architecture of the inner ring scaffold of the human nuclear pore complex. Science *352*, 363–365.

Klara, S.S., Saboe, P.O., Sines, I.T., Babaei, M., Chiu, P.L., Dezorzi, R., Dayal, K., Walz, T., Kumar, M., and Mauter, M.S. (2016). Magnetically directed two-dimensional crystallization of ompf membrane proteins in block copolymers. J. Am. Chem. Soc. *138*, 28–31.

Krall, L., Wiedemann, U., Unsin, G., Weiss, S., Domke, N., and Baron, C. (2002). Detergent extraction identifies different VirB protein subassemblies of the type IV secretion machinery in the membranes of *Agrobacterium tumefaciens*. Proc. Natl. Acad Sci. USA *99*, 11405–11410.

Kuroda, T., Kubori, T., Thanh Bui, X., Hyakutake, A., Uchida, Y., Imada, K., and Nagai, H. (2015). Molecular and structural analysis of *Legionella* DotI gives insights into an inner membrane complex essential for type IV secretion. Sci Rep *5*, 10912.

Laganowsky, A., Reading, E., Hopper, J.T.S., and Robinson, C.V. (2014). Mass spectrometry of intact membrane protein complexes. Nat. Protoc. *8*, 639–651.

Lau, W.C.Y. and Rubinstein, J.L. (2010). Structure of intact *Thermus thermophilus* V-ATPase by cryo-EM reveals organization of the membrane-bound V(O) motor. Proc. Natl. Acad Sci. USA *107*, 1367–72.

Lau, W.C.Y. and Rubinstein, J.L. (2011). Subnanometre-resolution structure of the intact *Thermus thermophilus* H^+-driven ATP synthase. Nature *481*, 214–218.

Leake, M.C., Chandler, J.H., Wadhams, G.H., Bai, F., Berry, R.M., and Armitage, J.P. (2006). Stoichiometry and turnover in single, functioning membrane protein complexes. Nature *443*, 355–358.

Leake, M.C., Greene, N.P., Godun, R.M., Granjon, T., Buchanan, G., Chen, S., Berry, R.M., Palmer, T., and Berks, B.C. (2008). Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by *in vivo* single-molecule imaging. Proc. Natl. Acad Sci. USA *105*, 15376–15381.

Lee, L.K., Stewart, A.G., Donohoe, M., Bernal, R.A., and Stock, D. (2010). The structure of the peripheral stalk of *Thermus*

thermophilus H^+-ATPase/synthase. Nat. Struct. Mol. Biol. *17*, 373–378.

Lin, D.H., Stuwe, T., Schilbach, S., Rundlet, E.J., Perriches, T., Mobbs, G., Fan, Y., Thierbach, K., Huber, F.M., Collins, L.N., et al. (2016). Architecture of the symmetric core of the nuclear pore. Science *352*, aaf1015–aaf1015.

McDonough, A.A., Veiras, L.C., Minas, J.N., and Ralph, D.L. (2014). Considerations when quantitating protein abundance by immunoblot. Am. J. Physiol. Cell Physiol. *308*, C426–C433.

Morimoto, Y.V., Ito, M., Hiraoka, K.D., Che, Y., Bai, F., Kami-ike, N., Namba, K., and Minamino, T. (2014). Assembly and stoichiometry of FliF and FlhA in *Salmonella* flagellar basal body. Mol. Microbiol. *91*, 1214–1226.

Nanavati, D., Gucek, M., Milne, J.L.S., Subramaniam, S., and Markey, S.P. (2008). Stoichiometry and absolute quantification of proteins with mass spectrometry using fluorescent and isotope-labeled concatenated peptide standards. Mol. Cell Proteomics *7*, 442–447.

Olinares, P.D.B., Kim, J., Davis, J.I., and Van Wijk, K.J. (2011). Subunit stoichiometry, evolution, and functional implications of an asymmetric plant plastid ClpP/R protease complex in *Arabidopsis*. Plant Cell *23*, 2348–2361.

Ong, S.-E. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell Proteomics *1*, 376–386.

Pratt, J.M., Simpson, D.M., Doherty, M.K., Rivers, J., Gaskell, S.J., and Beynon, R.J. (2006). Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nat. Protoc. *1*, 1029–1043.

Radics, J., Königsmaier, L., and Marlovits, T.C. (2013). Structure of a pathogenic type 3 secretion system in action. Nat. Struct. Mol. Biol. *21*, 82–87.

Reading, E., Liko, I., Allison, T.M., Benesch, J.L.P., Laganowsky, A., and Robinson, C. V. (2015). The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase. Angew. Chem. Int. Ed. *54*, 4577–4581.

Romano, F.B., Tang, Y., Rossi, K.C., Monopoli, K.R., Ross, J.L., and Heuck, A.P. (2016). Type 3 Secretion translocators spontaneously assemble a hexadecameric transmembrane complex. J. Biol. Chem. *291*, 6304–6315.

Rout, M.P., Aitchison, J.D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B.T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. J. Cell Biol. *148*, 635–651.

Salvay, A.G., Santamaria, M., Le Maire, M., and Ebel, C. (2007). Analytical ultracentrifugation sedimentation velocity for the characterization of detergent-solubilized membrane proteins Ca ++-ATPase and ExbB. J. Biol. Phys. *33*, 399–419.

Schägger, H., and von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal. Biochem. *199*, 223–231.

Schlegel, S., Hjelm, A., Baumgarten, T., Vikström, D., and De Gier, J. (2014). Bacterial-based membrane protein production. Biochim. Biophys. Acta Mol. Cell Res. *1843*, 1739–1749.

Schmidt, C., Lenz, C., Grote, M., Luhrmann, R., and Urlaub, H. (2010). Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. Anal. Chem. *82*, 2784–2796.

Schraidt, O. and Marlovits, T.C. (2011). Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. Science *331*, 1192–1195.

Shahid, S.A., Bardiaux, B., Franks, W.T., Krabben, L., Habeck, M., van Rossum, B.J., and Linke, D. (2012). Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. Nat. Methods *9*, 1212–1217.

Slotboom, D.J., Duurkens, R.H., Olieman, K., and Erkens, G.B. (2008). Static light scattering to characterize membrane proteins in detergent solution. Methods *46*, 73–82.

Stenberg, F., Chovanec, P., Maslen, S.L., Robinson, C. V., Ilag, L.L., Von Heijne, G., and Daley, D.O. (2005). Protein complexes of the *Escherichia coli* cell envelope. J. Biol. Chem. *280*, 34409–34419.

Stuwe, T., Correia, A.R., Lin, D.H., Paduch, M., Lu, V.T., and Kossiakoff, A.A. (2015). Architecture of the nuclear pore complex coat. Science *347*, 1148–1152.

Ulbrich, M.H. and Isacoff, E.Y. (2007). Subunit counting in membrane-bound proteins. Nat. Methods *4*, 319–321.

von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A.L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., et al. (2015). *In situ* structural analysis of the human nuclear pore complex. Nature *526*, 140–143.

Wagner, S., Königsmaier, L., Lara-tejero, M., Lefebre, M., Marlovits, T.C., and Galán, J.E. (2010). Organization and coordinated assembly of the type III secretion export apparatus. Proc. Natl. Acad Sci. USA *107*, 17745–17750.

Ward, M.E., Wang, S., Krishnamurthy, S., Hutchins, H., Fey, M., Brown, L.S., and Ladizhansky, V. (2014). High-resolution paramagnetically enhanced solid-state NMR spectroscopy of membrane proteins at fast magic angle spinning. J. Biomol NMR *58*, 37–47.

Weiland, F., Zammit, C.M., Reith, F., and Hoffmann, P. (2014). High resolution two-dimensional electrophoresis of native proteins. Electrophoresis *35*, 1893–1902.

Whitelegge, J.P. (2013). Integral membrane proteins and bilayer proteomics. Anal. Chem. *85*, 2558–68.

Wisedchaisri, G., Reichow, S.L., and Gonen, T. (2011). Advances in structural and functional analysis of membrane proteins by electron crystallography. Structure *19*, 1381–1393.

Wittig, I., Braun, H., and Scha, H. (2006). Blue native PAGE. Nat. Protoc. *1*, 418–428.

Wohlgemuth, I., Lenz, C., and Urlaub, H. (2015). Studying macromolecular complex stoichiometries by peptide-based mass spectrometry. Proteomics *15*, 862–879.

Worrall, L.J., Vuckovic, M., and Strynadka, N.C.J. (2010). Crystal structure of the C-terminal domain of the *Salmonella* type III secretion system export apparatus protein InvA. Protein Sci. *19*, 1091–1096.

Zhao, H., Brautigam, C.A., Ghirlando, R., and Schuck, P. (2013). Overview of current methods in sedimentation velocity and sedimentation equili brium analytical ultracentrifugation. Curr. Protoc. Protein Sci. *71*, 1–52.

Zhou, M., Morgner, N., Barrera, N.P., Politis, A., and Shoshanna, C. (2011). Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. Science *334*, 380–385.

Zilkenat, S., Franz-Wachtel, M., Stierhof, Y.-D., Galán, J.E., Macek, B., and Wagner, S. (2016). Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach. Mol. Cell Proteomics *15*, 1598–1609.

Manuscript IV

Structural and functional characterization of the bacterial type III secretion export apparatus

This research was originally published in PLoS Pathogens.

Dietsche T., Tesfazgi Mebrhatu M., Brunner M.J., Abrusci P., Yan J., Franz-Wachtel M., Schärfe C., Zilkenat S., Grin I., Galán J.E., Kohlbacher O., Lea S., Macek B., Marlovits T.C., Robinson C., Wagner S. Structural and functional characterization of the bacterial type III secretion export apparatus. *PLoS Pathogens.* 2016; 12(12):e1006071

PLOS | PATHOGENS

RESEARCH ARTICLE

Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus

Tobias Dietsche[1☯], Mehari Tesfazgi Mebrhatu[1☯], Matthias J. Brunner[2,3,4], Patrizia Abrusci[5], Jun Yan[6¤], Mirita Franz-Wachtel[7], Charlotta Schärfe[8], Susann Zilkenat[1], Iwan Grin[1], Jorge E. Galán[9], Oliver Kohlbacher[8,10], Susan Lea[5], Boris Macek[7], Thomas C. Marlovits[2,3,4], Carol V. Robinson[6], Samuel Wagner[1,11]*

1 University of Tübingen, Interfaculty Institute of Microbiology and Infection Medicine (IMIT), Section of Cellular and Molecular Microbiology, Tübingen, Germany, **2** Center for Structural Systems Biology (CSSB), University Medical Center Hamburg-Eppendorf (UKE) and German Electron Synchrotron Centre (DESY), Hamburg, Germany, **3** Institute of Molecular Biotechnology (IMBA), Vienna Biocenter (VBC), Vienna, Austria, **4** Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Vienna, Austria, **5** Sir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom, **6** Department of Chemistry, University of Oxford, Oxford, United Kingdom, **7** University of Tübingen, Proteome Center Tübingen, Tübingen, Germany, **8** University of Tübingen, Center for BioinformaticsTübingen, Germany, **9** Yale University School of Medicine, Department of Microbial Pathogenesis, New Haven, Connecticut, United States of America, **10** Max Planck Institute for Developmental Biology, Biomolecular Interactions, Tübingen, Germany, **11** German Center for Infection Research (DZIF), Partner-site Tübingen, Tübingen, Germany

☯ These authors contributed equally to this work.
¤ Current address: Novo Nordisk A/S, Novo Nordisk Park, Måløv, Denmark
* samuel.wagner@med.uni-tuebingen.de

Citation: Dietsche T, Tesfazgi Mebrhatu M, Brunner MJ, Abrusci P, Yan J, Franz-Wachtel M, et al. (2016) Structural and Functional Characterization of the Bacterial Type III Secretion Export Apparatus. PLoS Pathog 12(12): e1006071. doi:10.1371/journal.ppat.1006071

Editor: Brian K Coombes, McMaster University, CANADA

Received: September 14, 2016

Accepted: November 17, 2016

Published: December 15, 2016

Data Availability Statement: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005028. All other data are within the paper and its Supporting Information files.

Funding: The size exclusion chromatography-light scattering/UV/refractive index instrumentation was supported by NIH Award Number 1S10RR023748-01 (www.nih.gov). Work in the laboratory of JEG was supported by Grant AI030492 from the National Institute of Allergy and Infectious Diseases

Abstract

Bacterial type III protein secretion systems inject effector proteins into eukaryotic host cells in order to promote survival and colonization of Gram-negative pathogens and symbionts. Secretion across the bacterial cell envelope and injection into host cells is facilitated by a so-called injectisome. Its small hydrophobic export apparatus components SpaP and SpaR were shown to nucleate assembly of the needle complex and to form the central "cup" substructure of a *Salmonella* Typhimurium secretion system. However, the *in vivo* placement of these components in the needle complex and their function during the secretion process remained poorly defined. Here we present evidence that a SpaP pentamer forms a 15 Å wide pore and provide a detailed map of SpaP interactions with the export apparatus components SpaQ, SpaR, and SpaS. We further refine the current view of export apparatus assembly, consolidate transmembrane topology models for SpaP and SpaR, and present intimate interactions of the periplasmic domains of SpaP and SpaR with the inner rod protein PrgJ, indicating how export apparatus and needle filament are connected to create a continuous conduit for substrate translocation.

(www.niaid.nih.gov). Work performed in the laboratory of SW was supported by a postdoctoral fellowship of the Human Frontiers Science Program (www.hfsp.org), by the Alexander von Humboldt Foundation in the framework of the Sofja Kovalevskaja Award endowed by the Federal Ministry of Education and Research (BMBF) (www.avh.de), by the IZKF of the University Hospital Tübingen through the Nachwuchsgruppenprogramm (www.medizin.uni-tuebingen.de), and by the Deutsche Forschungsgemeinschaft (DFG) as part of the Collaborative Research Center (SFB) 766 Bacterial cell envelope, project B14 (www.dfg.de). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: JY was at the Department of Chemistry at the University of Oxford when she performed the experiments described in this paper. She is currently employed by NovoNordisk. All other authors have declared that no competing interests exist.

Author Summary

Many Gram-negative bacteria use type III secretion systems to inject bacterial proteins into eukaryotic host cells in order to promote their own survival and colonization. These systems are large molecular machines with the ability to transport proteins across three cell membranes in one step. It is believed that the only gated barrier of these systems lies in the bacterial cytoplasmic membrane but it was unclear so far how this gate looks like and of which components it is composed. Here we present evidence based on in depth biochemical and genetic characterization that an assembly of five SpaP proteins forms this gate in the cytoplasmic membrane of the type III secretion system of *Salmonella* pathogenicity island 1. We further show that one subunit each of the proteins SpaQ, SpaR, and SpaS are closely associated to the SpaP gate and may function in the gating mechanism, and that the protein PrgJ is attached to this gate on the outside to connect it to the hollow needle filament projecting towards the host cell. Our findings elucidate a hitherto ill-defined aspect of type III secretion systems and may help to develop novel antiinfective therapies targeting these virulence-associated molecular devices.

Introduction

Type III secretion systems (T3SSs) are used by many Gram-negative bacterial pathogens and symbionts to translocate effector proteins in one step across the bacterial envelope and into eukaryotic host cells [1] where they modulate host cell physiology to promote bacterial survival and colonization [2]. The core of T3SSs is formed by the so-called injectisome, a macromolecular machine composed of up to 20 different proteins [1]. The base of the injectisome, consisting of an outer membrane secretin ring and two inner membrane ring components, anchors the system to the bacterial cell envelope [3]. A filamentous needle projects away from the base towards the host cell and serves as conduit for translocated effectors [4,5]. Five cytoplasmic proteins select and unfold the substrates, which are then handed over to the actual export apparatus [6,7] housed in a membrane patch at the center of the inner ring [8,9]. The five export apparatus components are thought to facilitate the actual secretion function of T3SSs, including energy coupling, membrane translocation, and substrate specificity switching [1]. Base, needle filament, and export apparatus are together also termed needle complex.

While analyses by X-ray crystallography and cryo electron microscopy have revealed the structure of most soluble components of injectisomes or of the related flagellar system [10,11], the structure and in particular the function of the hydrophobic transmembrane (TM) domains of the export apparatus components remain poorly defined. In the T3SS encoded within the pathogenicity island 1 (SPI-1) of *Salmonella enterica* serovar Typhimurium (*S*. Typhimurium), the export apparatus is composed of the proteins SpaP, SpaQ, SpaR, SpaS, and InvA in a 5:1:1:1:9 stoichiometry [12]. Of these components, InvA and SpaS are structurally and functionally best characterized: the atomic structures of their soluble cytoplasmic domains have been solved [13,14]. The large cytoplasmic domain of InvA (or its homologs) forms a nonameric ring with a central pore of about 50 Å in diameter [15] and has been proposed to play a role in substrate switching and translocation [16,17] while its 8 predicted TM helices have been proposed to serve in utilization of the proton motive force for secretion [18]. SpaS and its homologs play a role in switching of specificity from secretion of early to intermediate and late substrates [19]. Autocleavage of a highly conserved NPTH motif in the cytoplasmic domain of SpaS is required for this function, possibly to facilitate a high conformational flexibility of this domain for secretion of later substrates [20].

The substantially hydrophobic export apparatus components SpaP, SpaQ, and SpaR and their homologs were shown to be critical for assembly of the needle complex [9,21–23] and essential for secretion function [9,24] but their precise role in secretion is still unknown. It was suggested that SpaP and SpaR form the cup substructure of the needle complex [9]. Given the presumed central location of SpaP and SpaR at the center of the membrane patch of the needle complex and their substantial hydrophobicity, we hypothesized that these two proteins may constitute the actual substrate translocation pore of T3SSs in the bacterial inner membrane, a function that as yet has not been assigned to any T3SS component.

In this study, we have biochemically characterized a stable subcomplex formed by SpaP and SpaR, and mapped its place within the needle complex using *in vivo* photocrosslinking and complementary techniques. We show that an isolated complex of five SpaP and one SpaR forms a donut-shaped structure with an approximately 15Å wide recession at its center. Sole expression of the SpaP pentamer in the bacterial membrane allowed the permeation of compounds of 500 Da into the cytoplasm, suggesting that these proteins form a channel large enough for translocation of secondary structures. We further show that a complex of SpaP, SpaQ, SpaR, and SpaS assembles *in vivo* before incorporation into the needle complex base, and that these four export apparatus components form a compact assembly with multiple reciprocal interactions at TM helices three and four of the SpaP pentamer. We also present evidence that SpaP and SpaR interact on their periplasmic side with the inner rod protein PrgJ, which provides a basis to explain how the substrate translocation conduit is continuous from the export apparatus through the inner rod into the needle filament and suggests that the hitherto unaccounted electron density of the socket substructure is made of the periplasmic domains of SpaP and SpaR, together with PrgJ. In summary, we describe physical interactions among export apparatus components of bacterial T3SSs and identify the components that form its substrate translocation pore. This work will facilitate further structural and functional work on these machines and may help to develop novel antiinfective therapies targeting these virulence-associated molecular devices.

Results

SpaP and SpaR form a stable subcomplex of $SpaP_5R_1$ stoichiometry

We previously showed that a stable complex of SpaP and SpaR can be isolated from *S.* Typhimurium lacking the inner ring components PrgH and PrgK [9]. For further characterization, we expressed the *spaPQRS* operon in *Escherichia coli* and purified the SpaPR complex by immunoprecipitation of epitope-tagged SpaR. The isolated complex eluted as a sharp peak from a size exclusion chromatography column at an apparent size of 400 kDa (Fig 1A). Separation of the protein complex by SDS PAGE followed by Coomassie staining or Western blotting and immunodetection of SpaP and $SpaR^{FLAG}$, respectively, showed that the complex contained more SpaP than SpaR (1B). Since the masses of membrane protein complexes deduced from analysis by size exclusion chromatography are skewed by the presence of bound detergent, we analyzed the fraction of protein and detergent contained in the isolated SpaPR complexes by size exclusion chromatography-multi angle laser light scattering. This analysis determined that the SpaPR peak was monodisperse, corresponding to a size of 311 kDa with a calculated protein content of 160 kDa (Fig 1C, S1 Table, S1 File), suggesting a total of 6 molecules of SpaP (25.2 kDa) and SpaR (31.7 kDa including C-terminal 3xFLAG tag). Given a mean error of 7% (S1 Fig), these data did not allow to distinguish whether the complex composition was 4 SpaP + 2 $SpaR^{FLAG}$ (calc. 164 kDa) or 5 SpaP + 1 $SpaR^{FLAG}$ (calc. 158 kDa). Native mass spectrometry was then performed to assess the exact stoichiometry of a purified $SpaPR^{STREP}$ complex. A major species of complex produced peaks of 157.882 kDa and a minor species of

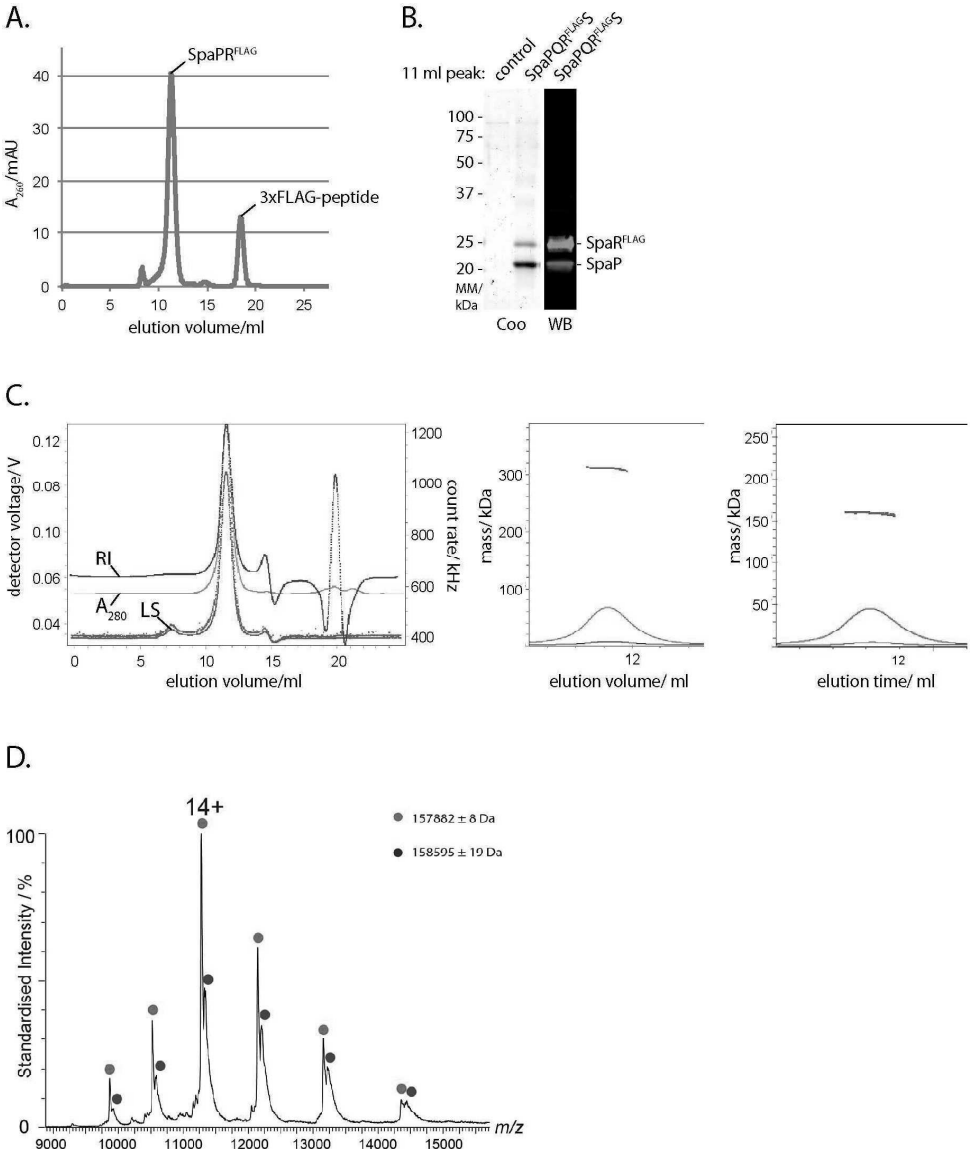


Fig 1. Isolation and stoichiometry analysis of the SpaPR subcomplex of the needle complex. (A) Elution profile of the purified SpaPR^{FLAG} complex run on a Superdex 200 10/300 GL column. The peaks corresponding to the SpaPR^{FLAG} complex and 3xFLAG peptide are indicated. (B) Coomassie-stained SDS PAGE gel of purified SpaPR^{FLAG} complex and of its FLAG-deficient control (left). Immunodetection of SpaP (green) and SpaR^{FLAG} (red) on Western blot from purified SpaPR^{FLAG} complex separated by SDS PAGE (right). (C) Traces of indicated detector signals from size exclusion chromatography—multi angle laser light scattering of purified SpaPR^{FLAG} complex (left). ASTRA-calculated mass profile of total components of peak of purified SpaPR^{FLAG} complex (polypeptides and detergent, middle). ASTRA-calculated mass profile polypeptide components of peak of purified SpaPR^{FLAG} complex (right). (D) Native mass spectrum of the SpaPR^{STREP} complex. Peak series corresponding to the SpaP:SpaR^{STREP} complex in a 5:1 ratio is marked in red, with the most abundant charge state (14+) indicated. The peak series marked in blue corresponds to the same SpaPR complex bound to a ligand with a mass of approximately 710 Da, indicative of an associated phospholipid. Note that the measured mass for SpaPR heterohexamer (157.882 kDa) is heavier than the theoretically calculated mass (157.280 kDa). Abbreviations: Coo: Coomassie stained, WB: Western blot, RI: refractive index, LS: light scattering.

doi:10.1371/journal.ppat.1006071.g001

158.595 kDa. These masses are consistent with a stoichiometry of 5 SpaP and 1 $SpaR^{STREP}$ (calculated molecular mass of 157.280 kDa) with bound phospholipids. In summary, these results show that the isolated SpaPR complex obtained from overexpression in the absence of other needle complex components has the same stoichiometry as SpaP and SpaR assembled into complete needle complexes [12] and indicates that the isolated SpaPR complex is a relevant functional module of the needle complex.

Probing the placement of SpaP and SpaR in the needle complex by *in vivo* photocrosslinking

To further validate the stoichiometry of SpaP and SpaR and to characterize the placing of this module within the assembled needle complex, we employed an *in vivo* photocrosslinking approach based on the genetically encoded UV-reactive amino acid *para*-benzophenylalanine (*p*Bpa) [25]. *p*Bpa was built into the predicted TM helices of SpaP and SpaR, respectively, so that possible interactions at every face of the predicted TM helices were sampled (Fig 2A and 2B). *spaP* or *spaPQRS* deletion mutants of *S*. Typhimurium were complemented with $SpaP^{FLAG}$ or $SpaPQR^{FLAG}S$ containing *p*Bpa at selected positions and expressed from a low copy number plasmid. Complementation of T3SS function of these mutants was assessed by analyzing type III-dependent secretion of substrate proteins into the culture supernatant (S2 Fig). Crosslinking of *p*Bpa to nearby interactors was induced by UV irradiation of intact bacterial cells immediately after harvesting. Subsequently, crude membranes were isolated and crosslinking patterns were analyzed by SDS PAGE and immunodetection of the FLAG-tagged bait protein. Crosslinked adducts of different sizes were identified at various positions of SpaP and SpaR (Fig 2C and 2D). To exclude crosslinking artifacts resulting from plasmid-based complementation, *p*Bpa positions that produced representative crosslinking patterns were also introduced into the chromosome-encoded genes, and crosslinking was performed accordingly. Notably, for all tested chromosomal positions the quality of previously identified crosslinks could be confirmed while the efficiency of crosslinking improved in some cases, possibly due to a more efficient complex assembly achieved by expression of *p*Bpa-containing proteins from its native context (Fig 2E and 2F). To identify the nature of crosslinked adducts, needle complexes with *p*Bpa-containing $SpaP^{FLAG}$ or $SpaR^{FLAG}$ were purified, UV-irradiated, resolved by SDS PAGE, and gel slices of the positions of the crosslinks were analyzed by mass spectrometry (S3 Fig). This analysis identified crosslinks between SpaP and the export apparatus components SpaS and SpaQ, and between SpaP and the inner rod protein PrgJ. Furthermore, crosslinks between SpaR and SpaP, SpaQ, and PrgJ were also identified (S2 Table, Fig 2C and 2D). The detailed validation and interpretation of the crosslinking analysis is presented in the following three sections.

Crosslinking of the SpaP pentamer

UV-irradiation of $SpaP^{FLAG}$-containing *p*Bpa at positions L7, L10, A12, F13, S14, T15, M187, S189, I193, and T195 showed a ladder of crosslinks at 40 kDa, 70 kDa, 120 kDa, and 200 kDa (Fig 2C and 2E). We reasoned that this crosslink ladder might correspond to a homo-oligomeric crosslinking of the SpaP pentamer. Two further experimental results supported this hypothesis: First, crosslinking of $SpaP_{T15X}^{FLAG}$ expressed in *E. coli* in the absence of other T3SS components showed the same crosslink ladder (Fig 3A); and second, crosslinking plasmid-complemented $SpaP_{T15X}$ in an *S*. Typhimurium strain with chromosome-encoded $SpaP^{FLAG}$ also produced the 40 kDa FLAG-containing crosslink, which proved at least a bipartite $SpaP_{T15X}$-$SpaP^{FLAG}$ interaction (Fig 3B). Several of the SpaP *p*Bpa mutants that produced a ladder upon crosslinking (A12X, T15X, M187X, S189X, I193X) were non-functional (S2 Fig).

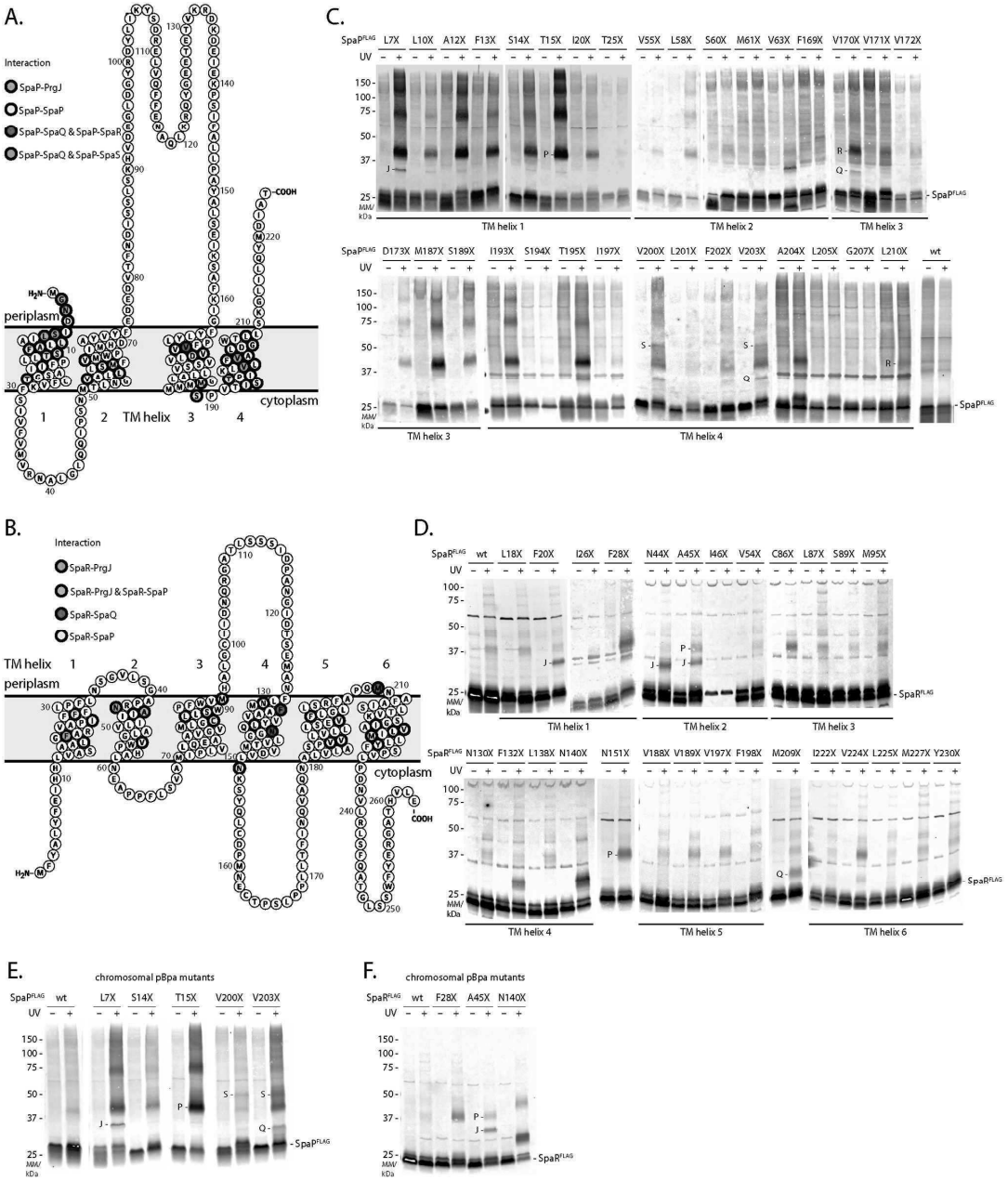


Fig 2. Screen of protein-protein interactions of SpaP and SpaR by *in vivo* photocrosslinking. (A) Protter visualization of SpaP presenting predicted TM topology, positions analyzed by *in vivo* photocrosslinking (thick stroke), and identity of interactions (colored). (B) As in (A) but showing SpaR. (C) Immunodetection of $SpaP^{FLAG}$ on Western blots of crude membrane samples of *S*. Typhimurium expressing indicated plasmid-complemented SpaP-*p*Bpa mutants separated by SDS PAGE. *p*Bpa mutations are denoted as "X". Each sample is shown with and without UV-irradiation to induce photocrosslinking of *p*Bpa to neighboring interaction partners. Since the running behavior of

crosslinked proteins often deviates from the calculated mass due to incomplete unfolding and since membrane proteins like SpaP often show an aberrant running behavior, the position of a crosslink on a gel does not easily allow drawing direct conclusions on the size of the crosslinked adduct. Crosslinked proteins identified by mass spectrometry or Western blotting are indicated. Other highlighted interactions shown in A and B were based on comparable SDS PAGE band pattern. (D) As in (C) but showing SpaR complemented from a low-copy number plasmid expressing SpaPQR^{FLAG}S. (E) As in (C) but expression of SpaP-*p*Bpa mutants from their chromosomal location. (F) As in (D) but expression of SpaR-*p*Bpa mutants from their chromosomal location. Abbreviations: J—PrgJ, P—SpaP, Q—SpaQ, S—SpaS.

doi:10.1371/journal.ppat.1006071.g002

Analysis of two of these *p*Bpa mutants (T15X and M187X) by 2-dimensional blue native/SDS PAGE indicated that the observed SpaP-SpaP interaction occurred between SpaP assembled into the complete needle complex as well as between SpaP molecules that had not yet been yet incorporated into this structure (Fig 3C). These results suggest that the loss of function of

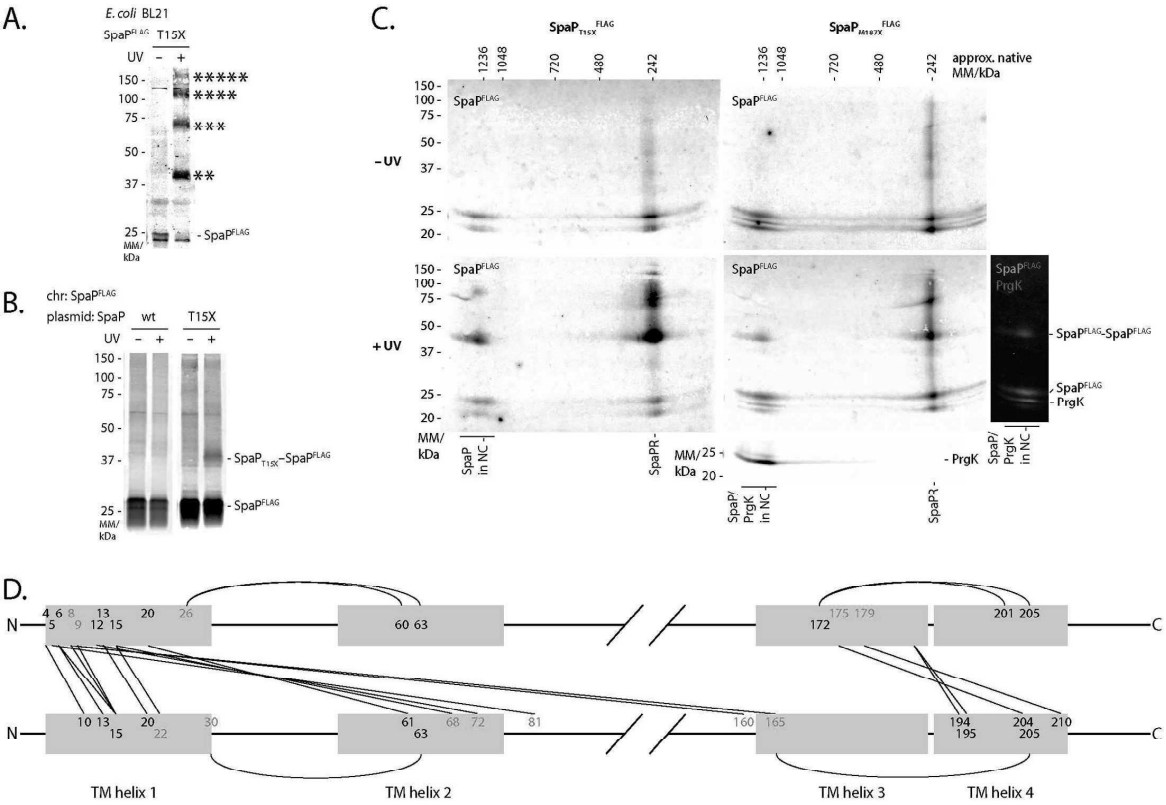


Fig 3. SpaP-SpaP interactions analyzed by *in vivo* photocrosslinking and sequence co-variation. (A) Immunodetection of SpaP^{FLAG} on Western blots of crude membrane samples of *E. coli* BL21 (DE3) expressing SpaP$_{T15X}^{FLAG}$ in the absence of all other T3SS components. The sample is shown with and without UV-irradiation to induce photocrosslinking of *p*Bpa to neighboring interaction partners. (B) Immunodetection of chromosome-encoded SpaP^{FLAG} on Western blots of crude membrane samples of *S.* Typhimurium expressing plasmid-encoded SpaP$_{T15X}$. (C) Immunodetection of SpaP^{FLAG} and the inner MS ring protein PrgK on Western blots of crude membrane samples of *S.* Typhimurium expressing indicated SpaP-*p*Bpa mutants separated by 2-dimensional blue native/SDS PAGE. Full 2D gels are only shown for SpaP^{FLAG} scanned in the 800 nm channel. The 2D gel showing SpaP$_{M187X}^{FLAG}$ +UV has been re-probed with antibody for PrgK and scanned in the 700 nm channel. PrgK indicates the position of the assembled needle complex. An overlay of FLAG and PrgK signals is shown on the right. The relevant slice of the 700 nm image showing PrgK at 25 kDa and the overlay of both channels showing the needle complex-associated bands have been aligned to the corresponding 2D image. (D) Interaction map of SpaP. Lines indicate predicted interactions with a normalized coupling score > 0.8 (S3 Table) at positions with experimentally identified SpaP-SpaP crosslinks (at least from one side). Positions with experimentally observed SpaP-SpaP interactions are shown in black, target positions only predicted are shown in light blue. Grey shading indicates TM helices. Only positions within or in close proximity to TM helices are shown. Abbreviations: chr—chromosomal.

doi:10.1371/journal.ppat.1006071.g003

these mutants is unlikely due to improper folding or assembly but rather due to subtle conformational changes that alter their function.

Overall, these results indicate that TM helix one and to a smaller extent the cytoplasmic face of TM helix three and four are involved in protomer contacts in the SpaP homopentamer while only few homotypic interactions were observed at positions of TM helices two and three.

To cross-validate the experimental findings, we performed an independent prediction of SpaP-SpaP interactions based on analysis of sequence co-variation using the software EV couplings [26–28]. 27 of the experimentally tested SpaP positions were predicted to be involved in SpaP-SpaP interactions with a normalized coupling score >0.80 (S3 Table). 18 of the 27 experimentally tested positions yielded indications of SpaP-SpaP interactions, 2 positions were experimentally ambiguous because of very low expression levels of the mutated proteins, and 7 positions showed no signs of SpaP-SpaP interactions. As used, EV couplings does not distinguish between intra and intermolecular interactions. 6 of the predicted but experimentally negative positions are likely to be involved in intramolecular interactions, which are not detectable by the *in vivo* photocrosslinking approach used (Fig 3D). Many intermolecular interactions at experimentally tested SpaP positions were predicted to connect two TM helices 1 or TM helix 1 and 3 in a parallel fashion, and TM helices 1 and 2 or TM helices 3 and 4 in an antiparallel fashion (Fig 3D), supporting a SpaP topology as depicted in Fig 2A, while only the coupling prediction of $SpaP_{S189}$ (to L11) opposed this model. Overall, the bioinformatic analysis supports our experimental results, strengthens the topology model of SpaP, and provides a first picture of the buildup of the SpaP pentamer.

SpaQ, SpaR, and SpaS assemble independently of other T3SS components onto the SpaP pentamer and closely interact with each other

Mass spectrometry analysis of crosslinked SpaP and SpaR adducts produced evidence for multiple interactions among the export apparatus components SpaP, SpaQ, SpaR, and SpaS (Fig 2, S3 Fig, S2 Table). To validate these results by immunoblotting, we assayed the SpaP-SpaR as well as the SpaP-SpaS interactions by FLAG-tagging the target instead of the *p*Bpa-containing bait protein. We found that SpaP interacts with $SpaR^{FLAG}$ through its residues V170 and L210 but not through V203 and A204 (Fig 4A) and that SpaR contacts $SpaP^{FLAG}$ via its residue N151 (Fig 4B). Using an autocleavage-deficient FLAG-tagged variant of the switch protein SpaS, we could further validate interactions between SpaS and $SpaP_{V200X}$/$SpaP_{V203X}$ (Fig 4C). In summary, these crosslinking data indicate that, consistent with our previous report [12], 1 SpaQ, 1 SpaR, and 1 SpaS form a closely interconnected assembly that contacts SpaP at TM helix three (V170: SpaQ, SpaR) and TM helix four (V200/203: SpaQ, SpaS). The interaction of these four proteins seems to be integrated by SpaQ as this small protein makes contacts to all other three proteins (*in vivo* photocrosslinking-identified SpaS-SpaQ contacts communicated results of J. Monjarás Feria).

Previous results showed that SpaQ is critical for efficient formation of the needle complex base but due to technical limitations of the blue native PAGE approach used at the time, it was not clear whether assembly proceeds through a pre-assembled complex of all four minor export apparatus components before integration into the base or whether these components only interact upon base integration [9]. To examine the early events of the assembly of the T3SS export apparatus components, we probed the SpaP-SpaQ, SpaP-SpaS, and SpaR-SpaQ interactions identified by the crosslinking studies in strains deficient in the inner ring protein PrgK. These mutants are deffective for base assembly thus allowing to prove the requirement of a fully assembled base for the assembly of the export apparatus. Indeed, we detected

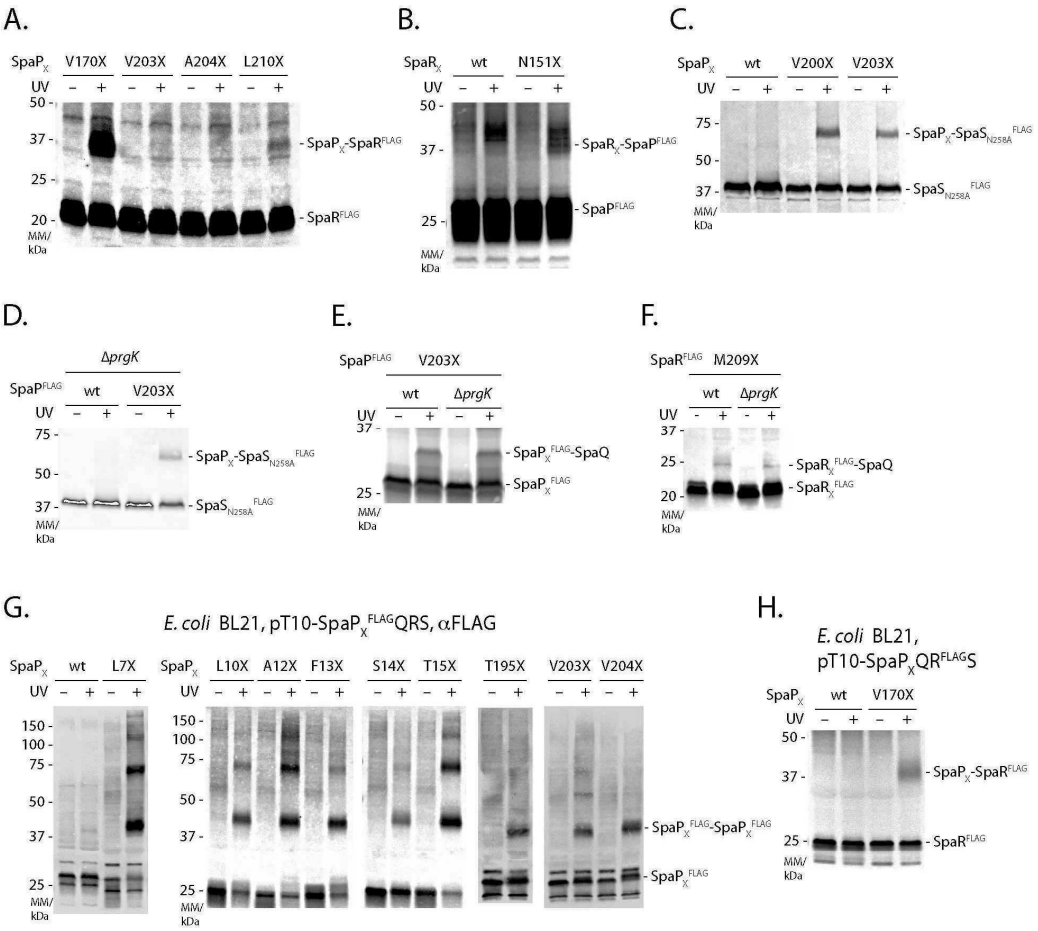


Fig 4. Interactions among the export apparatus components SpaP, SpaQ, SpaR, and SpaS. (A) Immunodetection of SpaR^{FLAG} on Western blots of SDS PAGE-separated crude membrane samples of Δ*spaPQRS S*. Typhimurium expressing indicated SpaP-*p*Bpa mutants from a pT10-*spaPQR*^{FLAG}*S* plasmid. (B) Immunodetection of SpaP^{FLAG} on Western blots of SDS PAGE-separated crude membrane samples of Δ*spaPQRS S*. Typhimurium expressing indicated SpaR-*p*Bpa mutants from a pT10-*spaP*^{FLAG}*QRS* plasmid. (C) Immunodetection of SpaS$_{N258A}^{FLAG}$ on Western blots of SDS PAGE-separated crude membrane samples of *S*. Typhimurium expressing indicated plasmid-complemented SpaP-*p*Bpa mutants. (D) As in (C) but assessing the SpaP-SpaS interaction in absence of the inner ring protein PrgK. (E) Immunodetection of SpaP^{FLAG} on Western blots of SDS PAGE-separated crude membrane samples of *S*. Typhimurium expressing chromosome-encoded indicated SpaP-*p*Bpa mutants in the presence or absence of the inner ring protein PrgK. (F) As in (E) but showing SpaR$_{M209X}^{FLAG}$. (G) Immunodetection of SpaP^{FLAG} on Western blots of crude membrane samples of *E. coli* BL21 (DE3) expressing indicated SpaP-*p*Bpa mutants together with SpaQRS to form the SpaPR complex. (H) As in (F) but expressing SpaP$_{V170X}$QR^{FLAG}S to reveal the SpaP-SpaR interaction in *E. coli*.

doi:10.1371/journal.ppat.1006071.g004

SpaP-SpaQ and SpaP-SpaS interactions at SpaP$_{X203}$ in the absence of PrgK (Fig 4D and 4E), and SpaR-SpaQ interactions at SpaR$_{X209}$ (Fig 4F). SpaP-SpaP and SpaP$_{V170X}$-SpaR crosslinks were also identified when plasmid-encoded SpaPQRS were expressed in *E. coli* BL21, lacking all other T3SS components (Fig 4G and 4H). Altogether, these results indicate that assembly of the export apparatus precedes and is independent of base assembly.

The inner rod protein PrgJ locates close to the inner membrane and directly contacts the periplasmic domains of SpaP and SpaR

UV-irradiation of SpaP^{FLAG} with *p*Bpa at position L7 or SpaR^{FLAG} with *p*Bpa at positions F20, N44, and A45 resulted in an 8 kDa mobility shift of these proteins in SDS-PAGE (Fig 2C, 2D and 2E). Mass spectrometry analysis of the shifted bands identified PrgJ in both cases (S3 Fig, S2 Table). In an effort to characterize the extent of the SpaP-PrgJ interaction in more detail, we also noted the same mobility shift of SpaP after UV-irradiation of SpaP^{FLAG} with *p*Bpa at positions G2, N3, D4, I5, and S6, where crosslinked PrgJ was confirmed by immunodetection (Fig 5A). To rule out potential artifacts due to overexpression of the plasmid-borne constructs, we confirmed the crosslinks of SpaP$_{G2X}^{FLAG}$ and SpaP$_{S6X}^{FLAG}$ after expression from their native chromosomal context (Fig 5B). 2-dimensional blue native/SDS PAGE analysis of the crosslinks resulting from UV-irradiation of SpaR$_{A45X}^{FLAG}$ showed that the observed SpaR-PrgJ interaction is only observed when SpaR is incorporated into the needle complex (Fig 5C). Furthermore, SpaP-PrgJ as well as SpaR-PrgJ interactions were not observed in an ATPase activity-deficient InvC$_{K165E}$ mutant, demonstrating that the detected interactions dependent on active type III secretion, which is consistent with the observation that incorporation of PrgJ into the needle complex and inner rod assembly require a functional type III secretion system (Fig 5D). Taken together, these results indicate that the periplasmic domains of SpaP and SpaR serve to anchor the inner rod protein PrgJ to the export apparatus, thus creating a continuous conduit for substrate translocation from the export apparatus to the needle filament.

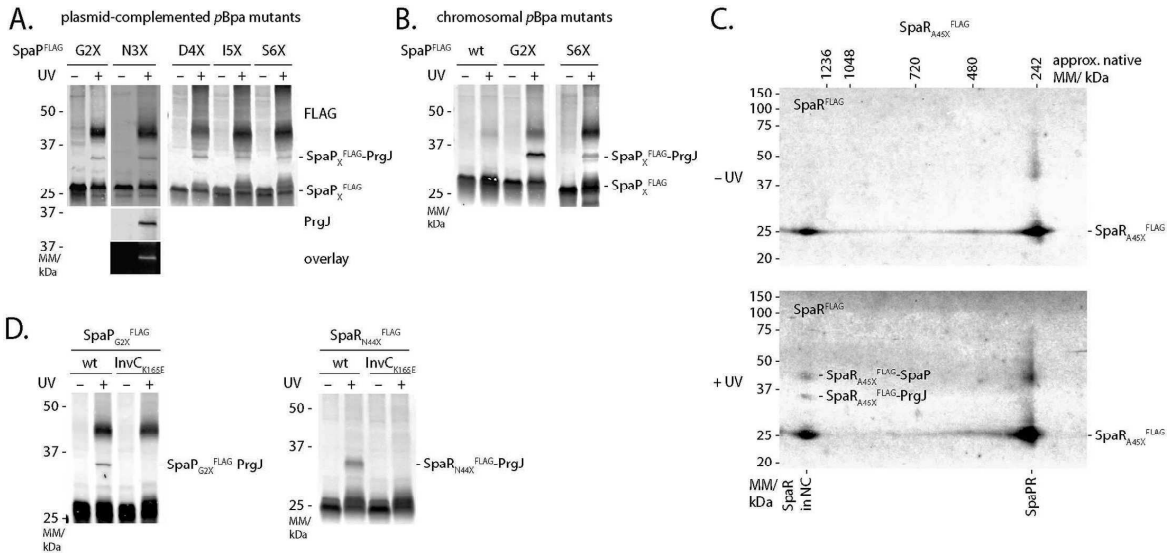


Fig 5. Interactions of SpaP and SpaR with the inner rod protein PrgJ. (A) Immunodetection of SpaP^{FLAG} on Western blots of crude membrane samples of *S*. Typhimurium expressing indicated plasmid-complemented SpaP-*p*Bpa mutants separated by SDS PAGE. The Western blot of SpaP$_{N3X}^{FLAG}$ was re-probed with PrgJ antibody to show the presence of SpaP and PrgJ in the same band. (B) Immunodetection as in (A) but detailing chromosome-encoded *p*Bpa-containing mutants of SpaP^{FLAG}. (C) Immunodetection of SpaR^{FLAG} on Western blots of crude membrane samples of *S*. Typhiumurium expressing SpaR$_{A45X}^{FLAG}$ separated by 2-dimensional blue native/SDS PAGE. (D) Immunodetection of SpaP^{FLAG} or SpaR^{FLAG} on Western blots of SDS PAGE-separated crude membrane samples of *S*. Typhimurium expressing SpaP$_{G2X}^{FLAG}$ or SpaR$_{N44X}^{FLAG}$ in wild type or in InvC ATP-hydrolysis mutants, which are unable to secrete.

doi:10.1371/journal.ppat.1006071.g005

SpaP forms a donut-shaped structure with a pore conducive to molecules of 500 Da

The location of the $SpaP_5R_1$ complex at the center of the needle complex base, right underneath and connected to the filamentous conduit formed by the inner rod and needle proteins, suggests that this complex forms the T3SS's substrate translocation pore in the bacterial inner membrane.

To obtain structural evidence for its putative pore-forming function, we analyzed the purified, negative-stained $SpaPR^{FLAG}$ complex by electron microscopy. 11202 individual particles were classified and aligned into 91 class averages (S4 Fig). A number of class averages showed a symmetric, donut-shaped complex with an iconic recession at its center (Fig 6A). The diameter of these particles was about 80 Å and the diameter of the recession was about 15 Å. Other class averages showed a more asymmetric shape with an extra density outside of the ring-structure or a mushroom-like shape (Fig 6A). Even though the sample analyzed consisted of a homogeneous population of $SpaPR^{FLAG}$ complexes, it cannot be ruled out that SpaP and $SpaR^{FLAG}$ partly dissociated during sample preparation so that a mixture of $SpaP_5$ and $SpaP_5R_1$ complexes was imaged, explaining the diversity of observed classes. It is therefore possible that the donut-shaped particles represent $SpaP_5$ complexes and the asymmetric extension the SpaP-bound $SpaR^{FLAG}$. Overall, the particles' shape and dimensions conformed well with the structure of the cup region of assembled bases reported previously (3).

We reasoned that the recession at the center of the observed particles might represent the protein translocation pore of the T3SS. To probe the conducting properties of the SpaPR complex, we assessed its ability to allow the access of biotin maleimide (BM, molecular mass = 500 Da) into the bacterial cytoplasm, an approach that has been used previously to test the gating

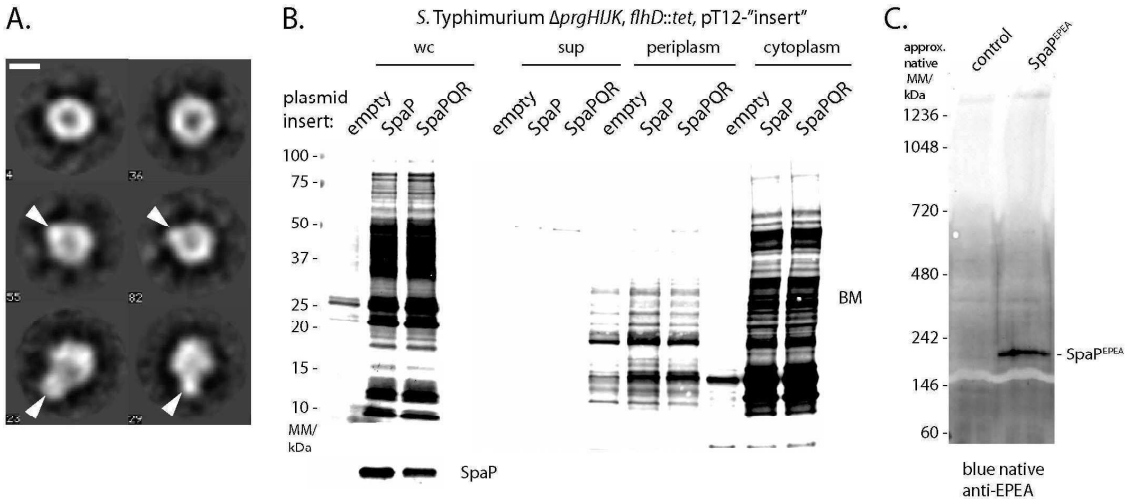


Fig 6. Visualization and characterization of the pore formed by SpaP and SpaR. (A) Six selected class averages (4, 23, 29, 36, 55, 82) of negative-stained isolated SpaPR complexes imaged by electron microscopy. The length of the scale bar represents 50 Å. The two class averages at the top represent the $SpaP_5$ complex. Arrowheads in the class averages in the middle and at the bottom represent the anticipated position of SpaR on the $SpaP_5$ ring. The complete picture of all class averages can be seen in S4 Fig. (B) Fluorescent streptavidin detection of SDS PAGE-separated biotin maleimide-labeled proteins of whole cell lysates, cell culture supernatant, periplasmic fraction, or cytoplasmic fraction of *S.* Typhimurium Δ*prgHIJK*, *flhD*::*tet* moderately overexpressing indicated proteins from a medium copy number plasmid (pT12). (C) Blue native PAGE and immunodetection of a high molecular weight complex formed by EPEA-tagged SpaP alone.

doi:10.1371/journal.ppat.1006071.g006

of the Sec-translocon [29]. The maleimide moiety of BM can only react with and biotinylate free thiol groups of cysteine residues of cytoplasmic proteins if BM can penetrate the inner bacterial membrane through a sufficiently large pore. The extent of biotinylation can then be detected on a Western blot by utilizing streptavidin. Strong BM labeling of proteins was observed in whole cell lysates when SpaPR or SpaP alone were overexpressed from a medium copy plasmid (Fig 6B). Cell fractionation of the expression host showed that only cytoplasmic proteins were differentially labeled by BM upon expression of SpaPR and SpaP, labeling of periplasmic proteins, however, was almost indistinguishable in control and expressing bacteria (Fig 6B). General lysis of the expression host could be ruled out to cause the observed phenotype as neither the cytoplasmic protein RNA polymerase nor the periplasmic maltose binding protein were observed in the culture supernatant of SpaPR or SpaP expressing bacteria (S5A and S5B Fig). Formation of a sizable, ungated pore by these complexes was also indicated by the strong impact even modest overexpression of SpaP and SpaPR had on the viability of the expression host (S5C Fig). Altogether, these results suggest that BM accessed the cytoplasm of the expression host through a pore formed by the expressed proteins. Since SpaP expression alone led to BM labeling of cytoplasmic proteins, it is conceivable that SpaP alone is sufficient to form the actual substrate translocation pore. In line with this idea, overexpressed $SpaP^{EPEA}$ was observed to assemble into high molecular weight complexes when analyzed by blue native PAGE (Fig 6C), however, we were not able to isolate and investigate stable SpaP-only complexes. The access of 500 Da BM to the cytoplasm through the pore of the SpaP pentamer suggests a pore diameter of about 15 Å, which is consistent with the diameter of the recession observed by electron microscopy of the isolated $SpaP_5R_1$ complexes.

Discussion

The export apparatus of bacterial T3SSs is its central unit that facilitates translocation of substrates across the bacterial inner membrane and likely the only gated barrier of these one-step secretion devices. While functions have been proposed for some export apparatus components, the components forming the actual substrate translocation pore in the bacterial inner membrane have not been defined.

In this study we present evidence that a homopentamer of the minor hydrophobic export apparatus component SpaP is a central component of the translocation pore in the inner membrane of the injectisome T3SS encoded by *Salmonella* pathogenicity island 1. We purified a stable complex of 5 SpaP and 1 SpaR that under electron microscopy exhibited a donut-like shape of about 80 Å in diameter and a 15 Å wide central recession. Expression of the components of this complex in *E. coli* rendered the bacterial cells permeable to 500 Da compounds, supporting the notion that it may work as translocation channel. Extensive mapping of protein-protein interactions of the TM domains of SpaP and SpaR by *in vivo* photocrosslinking revealed that SpaQ, SpaR, and SpaS form a compact assembly connected to the central pentamer formed by SpaP. We further demonstrated that assembly of this complex does not require its incorporation into the needle complex. We also detected crosslinks between SpaP and SpaR and the inner rod protein PrgJ showing that the inner rod makes direct contact with the export apparatus.

Previous analysis by blue native PAGE showed that SpaP and SpaR form stable complexes in an *S*. Typhimurium mutant unable to assemble the needle complex [9]. We now present evidence based on size-exclusion chromatography-multi angle laser light scattering and native mass spectrometry that this complex is composed of 5 SpaP and 1 SpaR. The stoichiometry of the isolated $SpaP_5R_1$ complex is consistent with the stoichiometry of SpaP and SpaR in the context of a fully assembled needle complex [12], which indicates that the isolated complex

represents a relevant intermediate of needle complex assembly. This notion is further supported by the good match of the dimensions of the observed SpaPR complex with the dimensions of the cup substructure of the needle complex [30], which we previously showed to be composed of SpaP and SpaR [9]. Electron micrographs of the isolated $SpaP_5R_1$ complex and BM permeation experiments suggested a pore size of the substrate translocation channel of about 15 Å. Within the range of uncertainty, this diameter conforms with the 10 Å that were reported for the dimensions of the channel of an assembled *S*. Typhimurium *S*PI-1 needle complex containing a trapped translocation intermediate [5]. A tight seal during substrate translocation is expected to be important for T3SS to avoid leakage of ions through the open pore, so it is conceivable that the pore diameter closely resembles the dimensions of extended polypeptides or alpha helices. However, a larger pore diameter in its fully open state cannot be excluded given that the herein investigated isolated $SpaP_5R_1$ complex most certainly lacks the necessary elements for gating of the pore.

We detected extensive crosslinks of up to five consecutive SpaP at TM helix one and at the cytoplasmic face of TM helices three and four, suggesting that these regions form the major contact area between protomers of the SpaP pentamer. This notion was supported by results of a sequence co-variation-based prediction of residue-residue interactions of SpaP. The formation of these crosslinks was independent of the presence of other needle complex components, supporting the notion that the SpaP pentamer nucleates assembly of the needle complex. Interestingly, the presence of SpaP pentamer crosslinks at TM helices three and four correlated with secretion defects of the respective *p*Bpa mutants, a phenomenon also seen for $SpaP_{A12X}$ and $SpaP_{T15X}$. The secretion defect was not due to defects in their incorporation into assembled needle complexes, suggesting that these residues may play a critical role in protein translocation.

The recently reported stoichiometry of SpaP, SpaQ, SpaR, and SpaS of 5:1:1:1 [12] suggests that these export apparatus components form an asymmetric assembly within the needle complex. We show here that SpaQ, SpaR, and SpaS contact the SpaP pentamer at its TM helices three and four. We further demonstrate that SpaQ interacts with SpaP and SpaR. These observations, together with the observation that a fusion of SpaR and SpaS homologs retains function [31], suggest that SpaQ, SpaR, and SpaS are not wrapped around the SpaP pentamer but form a compact assembly at one side of SpaP, with SpaQ as the central component that makes contacts to all other components (Fig 7A and 7B). Besides SpaR's contribution in anchoring the inner rod protein PrgJ, the assembly formed by SpaQ, SpaR, and SpaS might also facilitate gating of the SpaP pore, a critical aspect to prevent detrimental effects of nutrient and ion leakage across the bacterial inner membrane.

The assessment of the dependence of crosslinks between SpaP, SpaQ, SpaR, and SpaS on the presence of the inner ring protein PrgK allowed us to refine a model for the early steps of export apparatus assembly (Fig 7C). We propose that assembly starts with the formation of the SpaP pentamer. This initially unstable complex is stabilized upon binding SpaR. The high stability of the resulting $SpaP_5R_1$ intermediate suggests that this complex is the major nucleus of further needle complex assembly. Next, SpaQ and SpaS associate with the $SpaP_5R_1$ complex but presumably with weaker affinity since this complex could only be captured after *in vivo* crosslinking. InvA would then be recruited to the SpaPRQS complex although it is not clear whether its recruitment occurs prior or after this complex initiates the assembly of the needle complex rings. Subsequently, association of the outer membrane secretin InvG and the inner ring protein PrgH would lead to formation of the completed base-export apparatus holo-complex [21,23].

Beyond interactions among the export apparatus components, we also identified crosslinks between the periplasmic domains of SpaP and SpaR and the inner rod protein PrgJ. The close

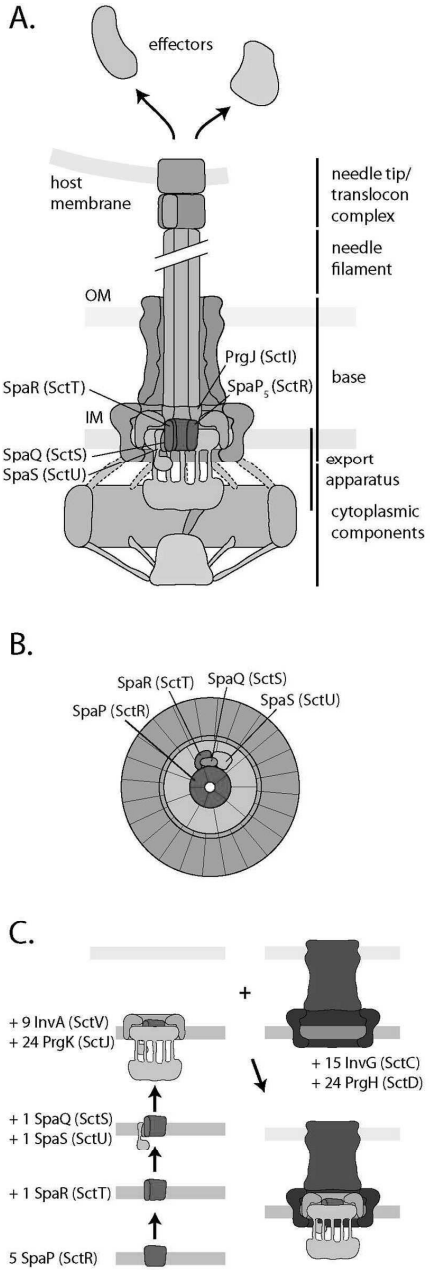


Fig 7. Models of SpaP, SpaR, SpaQ, SpaS, and PrgJ in the T3SS needle complex and its assembly. (A) Model of the central SpaP complex with surrounding export apparatus components SpaQ, SpaR, and SpaS, and direct connection to the inner rod formed by PrgJ. These results suggest that SpaP, SpaR, and PrgJ form the socket structure on the periplasmic side of the inner membrane patch of the base. (B) Model of a view of the membrane patch of the needle complex from the cytoplasmic side highlighting SpaP, SpaQ, SpaR, and SpaS. (C) Model of needle complex assembly. The unified Sct nomenclature [23] is shown in parenthesis.

doi:10.1371/journal.ppat.1006071.g007

interaction of SpaP, SpaR, and PrgJ is likely to create a continuous conduit for substrate translocation, where PrgJ might serve as an adapter to connect the flat translocation pore of the inner membrane with the helical needle filament. Analysis of the needle complex by cryo-electron microscopy revealed a central juxtamembrane structure at the periplasmic interior of the base, which was termed socket [30], however, no protein could be assigned to contribute to this density. Our results suggest that the socket is composed of the periplasmic parts of SpaP and SpaR, together with the inner rod protein PrgJ. The mass of six PrgJ [12] and the periplasmic domains of five SpaP and one SpaR could well account for the observed density of the socket structure. Our observation now opens the door for further investigations of the relevance of the export apparatus-PrgJ interaction for needle length control, substrate specificity switching, and host cells sensing, functional roles that were suggested for PrgJ [32,33].

The positions of SpaP and SpaR that interact with PrgJ also help to consolidate the TM topology models of these two export apparatus proteins. SpaP is predicted to contain four TM helices (Fig 2A) and the presence of a cleavable signal sequence in flagellar homologs suggests an N-out/C-out TM orientation [34]. This model is supported by the interaction between the N-terminus of SpaP and the periplasmic inner rod detected in this study. Further support for this topology model comes from the presented sequence co-variation-based analysis of SpaP residue-residue interactions, which strongly predicted antiparallel interactions between TM 1 and 2, and between TM 3 and 4 (Fig 3D). The TM topology predictions of SpaR and its homologs are very uncertain, ranging from five to eight TM helices with mostly N-out orientation (Fig 2B, S6 Fig) [34,35]. A C-in orientation, on the other hand, was suggested based on the report of a functional protein fusion of the flagellar SpaR and SpaS homologs of *Clostridium*, given that the N-terminus of SpaS and its homologs is strongly predicted to reside in the cytoplasm [31,35,36]. Here we presented interactions of SpaR F20, N44, and A45 with the periplasmic protein PrgJ. These residues are predicted to be located within SpaR's first two TM helices, however, our results rather suggest a periplasmic localization of this part of SpaR. This notion is supported by rather high ΔG values for membrane partitioning of the predicted TM helices one, two, and four (S6B Fig), so that a SpaR model comprising an N-out/C-in topology with only three TM helices is conceivable (S6C Fig).

In summary, we have presented evidence that a pentamer of SpaP forms the substrate translocation pore of T3SSs in the bacterial inner membrane. We show that this pentamer closely interacts with the export apparatus components SpaQ, SpaR, and SpaS in the plane of the membrane, an accessory assembly that may facilitate gating of the export pore. We further show that SpaP and SpaR intimately contact the periplasmic inner rod protein PrgJ and propose that the inner rod serves as an adapter to connect the flat export pore and the helical needle filament, thus creating a continuous conduit for substrate translocation from the bacterial cytoplasm into the host cell.

Materials and Methods

Materials

Chemicals were from Sigma-Aldrich unless otherwise specified. Detergent n-dodecyl-maltoside (DDM) was from Affimetrix-Anatrace. para-benzophenylalanine was from Bachem. SERVA Blue G and SERVAGel TG PRiME 8–16% precast gels were from Serva. NativePAGE Novex Bis-Tris 3–12% gels were from Life Technologies. Primers are listed in S5 Table and were synthetized by Eurofins and Integrated DNA Technologies. Polyclonal rabbit anti-MBP antibody were from New England Biolabs. Monoclonal mouse anti-RNApol antibody was from BioLegend. Monoclonal M2 anti-FLAG antibody, M2 anti-FLAG agarose beads, and 3xFLAG peptide were from Sigma-Aldrich. CaptureSelect-biotin, Streptavidin DyLight 800,

and secondary antibodies goat anti-mouse IgG DyLight 800 conjugate and goat anti-rabbit IgG DyLight 680 conjugate were from Thermo-Fisher.

Bacterial strains and plasmids

Bacterial strains and plasmids used in this study are listed in S4 Table. Primers for construction of strains and plasmids ere listed in S5 Table. The position and sequence of epitope tags introduced into SpaP, SpaR, and SpaS is shown in S7 Fig. All *Salmonella* strains were derived from *S*. Typhimurium strain SL1344 [37]. Bacterial cultures were supplemented as required with streptomycin (50 μg/mL), tetracycline (12.5 μg/mL), ampicillin (100μg/mL), kanamycin (25 μg/mL), or chloramphenicol (10 μg/mL).

Expression and purification of SpaPR complex

The SpaP, and SpaPR complexes were expressed in *E. coli* BL21 (DE3) from rhamnose-inducible medium copy number plasmids encoding $SpaP^{EPEA}$, $SpaPQR^{FLAG}S$, or $SpaPQR^{STREP}$, respectively. Expression was autoinduced by over night growth at 37˚C in TB medium. Bacterial cells were harvested, crude membranes purified as described previously [9], and membrane proteins were extracted with 1% DDM in PBS. After removal of unsolubilized material by ultracentrifugation for 30 min at 100.000 x g, complexes were immunoprecipitated according to the manufacturers instructions using CaptureSelect affinity gel for $SpaP^{EPEA}$, M2 anti-FLAG agarose beads for $SpaPR^{FLAG}$, and Strep-Tactin sepharose (IBA) for $SpaPR^{STREP}$. Complexes were natively eluted with 150 ng/ml SEPEA or 3xFLAG peptides, respectively, or with 2.5 mM desthiobiotin, each in PBS/0.04% DDM. The $SpaP^{EPEA}$ and the $SpaPQR^{FLAG}S$, complexes were subsequently purified by anion exchange (Mono Q 5/50 GL, GE), while this step was omitted for the $SpaPQR^{STREP}$ complex. Samples were further purified by size exclusion (Superdex 200 10/300 GL, GE) chromatography, and concentrated to 1 mg/ml using Amicon Ultra 100 k cutoff spin concentrators (Merck Millipore). Purified SpaP and SpaPR complexes were stored in liquid nitrogen until further use.

Size exclusion chromatography—multi angle laser light scattering analysis

The detergent and polypeptide content of the purified $SpaPR^{FLAG}$ complex in PBS/0.04% DDM was determined by size exclusion chromatography—multi angle laser light scattering and analysis by the ASTRA software (Wyatt, Santa Barbara, CA) as previously described [38].

Native mass spectrometry of isolated native SpaPR complex

Purified $SpaPR^{STREP}$ complex was concentrated to 20 μM in PBS/0.04% DDM, and buffer exchanged to 250 mM ammonium acetate, pH 7.5, complemented with 0.01% polyoxyethylene(9)dodecyl ether (C12E9) prior to native mass spectrometry analysis. Buffer exchange was carried out using Amicon Ultra 0.5 ml centrifugal filters with a 100-kDa cut-off (Millipore UK Ltd, Watford UK). Mass measurements were carried out on a Synapt G1 HDMS (Waters Corp., Manchester, UK) Q-ToF mass spectrometer [39]. The instrument was mass calibrated using a solution of 10 mg/ml cesium iodide in 250 mM ammonium acetate. 2.5 μL aliquots of samples were delivered to the mass spectrometer by means of nano-electrospray ionization via gold-coated capillaries, prepared in house [40]. Instrumental parameters were as follows: source pressure 6.0 mbar, capillary voltage 1.40 kV, cone voltage 150 V, trap energy 200 V, transfer energy 10 V, bias voltage 5 V, and trap pressure 1.63x10-2 mbar.

Membrane protein topology prediction

SpaP and SpaR TM topology was predicted using TOPCONS (http://topcons.cbr.su.se) [41]. The extent of the hydrophobic regions constituting TM helices was predicted using dGpred full portein scan (http://dgpred.cbr.su.se) [42] setting the minimal helix length to 18 and the maximal helix length to 31 aa. For visualization, the online tool PROTTER (http://wlab.ethz.ch/protter/start/) was used [43].

Secretion assay

Analysis of type III-dependend secretion of proteins into the culture medium was carried out as described previously [20].

Immunoblotting

For protein detection, samples were subjected to SDS PAGE using SERVAGel TG PRiME 8–16% precast gels, transferred onto a PVDF membrane (Bio-Rad), and probed with primary antibodies anti-SipB, anti-InvJ, anti-PrgJ, anti-SpaP, anti-MBP, anti-RNApol, and M2 anti-FLAG. Secondary antibodies were goat anti-mouse IgG DyLight 800 conjugate and goat anti-rabbit IgG DyLight 680. EPEA-tagged SpaP was visualized using CaptureSelect-biotin anti C-Tag conjugate and Streptavidin DyLight 800. Scanning of the PVDF membrane and image analysis was performed with a Li-Cor Odyssey system and image Studio 2.1.10 (Li-Cor).

In vivo photocrosslinking

S. Typhimurium strains were grown at 37˚C in LB broth supplemented with 0.3 M NaCl with low aeration to enhance expression of genes of *S*PI-1. For *in vivo* photocrosslinking of $SpaP^{FLAG}$ in *Escherichia coli* BL21 (DE3), bacteria were cultured at 37˚C in LB broth. Cultures were supplemented with 500 μM rhamnose to induce expression of $SpaP^{FLAG}$, $SpaP^{FLAG}$QRS or SpaPQR^{FLAG}S from low copy number pTACO10 plasmids [9]. To boost general *S*PI-1 expression, *S*. Typhimurium strains were transformed with pBAD24-hilA. Expression of the *S*PI-1 master regulator HilA was induced by addition of 0,05% arabinose to the cultures. Additionally the cultures were supplemented with the artificial amino acid para-benzoyl phenyl alanine (pBpa) to a final concentration of 1 mM and afterwards incubated for 5.5 h. 2 ODU of bacterial cells were harvested and washed once with 1 mL cold PBS. Cells were resuspended in 1 mL PBS and transferred into 6-well cell culture dishes. UV irradiation with λ = 365 nm was done on a UV transilluminator table (UVP) for 30 min.

Crude membrane preparation

10 OD units of bacterial lysates of *S*. Typhimurium or *E*. *coli*, respectively, were resuspended in 750 μl buffer K (50 mM triethanolamine, pH 7.5, 250 mM sucrose, 1 mM EDTA, 1 mM $MgCl_2$, 10 μg/ml DNAse, 2 mg/mL lysozyme, 1:100 protease inhibitor cocktail), and incubated for 30 min on ice. Samples were bead milled and beads, unbroken cells and debris were removed by centrifugation for 10 min at 10.000 x g and 4˚C. Crude membranes contained in the supernatant were precipitated by centrifugation for 45 min at 55,000 rpm and 4˚C in a Beckman TLA 55 rotor. Pellets containing crude membranes were frozen until use.

Blue native PAGE

1-dimensional blue native PAGE and 2-dimensional blue native/SDS PAGE of crude membranes was carried out as previously described [9].

Needle complex purification

S. Typhimurium Δ*spaP* or Δ*spaPQRS* mutants, respectively, transformed with pSUP, pSB3292, and pSB3398-based rhamnose-inducible low copy number plasmids containing $SpaP^{FLAG}$ amber mutants or SpaPQRS with $SpaR^{FLAG}$ amber mutants, respectively, were grown in 200 ml LB broth supplemented with 0.3M NaCl, 1 mM *p*Bpa, 500 μM rhamnose, 0.02% arabinose, and appropriate antibiotics for 5 h at low aeration to express *S*PI-1 and assemble needle complexes. Purification of needle complexes was carried out as published previously [4,20,12] but LDAO was replaced by DDM (0.7% for lysis/extraction, 0.1% for maintenance) for lysis of cells and extraction of needle complexes throughout the protocol. Furthermore, an initial concentration of 35% (wt/vol) of CsCl was used to prepare the gradient. Purified needle complexes containing $SpaP^{FLAG}$ or $SpaR^{FLAG}$ with *p*Bpa at desired positions were irradiated with UV light (365 nm) for 30 min to induce photocrosslinking to nearby proteins. Samples were subsequently analyzed by SDS PAGE, Western blotting, and immunodetection with M2 anti-FLAG antibodies. For MS analysis of crosslinked adducts, gel pieces at positions of observed crosslinks of *p*Bpa-containing and control samples were cut out of Coomassie-stained SDS PAGE gels and subjected to in gel digestion.

Protein in-gel digestion for analysis of crosslinked interaction partners

For identification of crosslinked proteins, the area of a Coomassie-stained gel corresponding the position of the crosslinked band detected by Western blotting were excised and in-gel digested with trypsin [44]. For a better recovery, remaining proteins in the gel were again subjected to another tryptic digestion step. After each step extracted peptides were desalted using C_{18} StageTips [45]. Corresponding eluates were combined and subjected to LC-MS/MS analysis.

Mass spectrometry for analysis of crosslinked interaction partners

LC-MS/MS analyses were performed on an EasyLC II nano-HPLC (Proxeon Biosystems) coupled to an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific) as decribed elsewhere [46] with slight modifications: The peptide mixtures were injected onto the column in HPLC solvent A (0.5% acetic acid) at a flow rate of 500 nl/min and subsequently eluted with a 106 min gradient of 5–33% HPLC solvent B (80% ACN in 0.5% acetic acid). During peptide elution the flow rate was kept constant at 200 nl/min. For proteome analysis, the 20 (Orbitrap Elite) most intense precursor ions were sequentially fragmented in each scan cycle using collision-induced dissociation (CID). In all measurements, sequenced precursor masses were excluded from further selection for 90 s. The target values for MS/MS fragmentation were 5000 charges and 10^6 charges for the MS scan.

Mass spectrometry data processing for analysis of crosslinked interaction partners

The MS data were processed with MaxQuant software suite v.1.2.2.9 as described previously [47–49] with slight modifications. Database search was performed using the Andromeda search engine [48], which is part of MaxQuant. MS/MS spectra were searched against a target database consisting of 10,152 protein entries from *S*. Typhimurium and 248 commonly observed contaminants. In database search, full tryptic specificity was required and up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as fixed modification, protein N-terminal acetylation, and oxidation of methionine were set as variable modifications. Initial precursor mass tolerance was set to 6 parts per million (ppm) and at the

fragment ion level 0.5 dalton (Da) was set for CID fragmentation. The MS data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the data set identifier PXD005028.

EVfold coupling analysis

Sequence co-variation analysis was performed using EVcouplings [26–28] with pseudo-maximum likelihood approximation [50–52]. The multiple sequence alignment used as input for the model inference was created by jackhmmer 3.1 [53] (5 iterations) using the full sequence of Salmonella SpaP (UniProt: SPAP_SALTY, residues 1–224) as query against the November 2015 release of the UniProt Reference Cluster database (UniRef100) [54]. Sequences with more than 30% gaps are subsequently removed from the alignment. We then excluded alignment columns that contained 50% or more gaps from model inference and subsequent couplings predictions. Lastly, sequences were clustered at 80% sequence identity and then downweighted according to the cluster size to reduce redundancy. This resulted in an alignment of 5663 unique sequences with an effective number of 1080.4 non-redundant sequences (sequences/alignment length = 4.8) included in model inference and coupling prediction. The coupling scores of residue pairs were further normalized by estimating the background noise analogously to the procedure described in Hopf et al., 2014 [28]. Evaluation of the co-evolution prediction was done in the light of topology predictions obtained from deltaG, resulting in four predicted TM segments: (7, 38), (50, 75), (163, 193), (194, 211). Python (Python Software Foundation, http://www.python.org) and Ipython/Jupyter notebooks [55] were used for data analysis. The multiple sequence alignment, EC scores file, a contact map of the strongest couplings and an Ipython notebook of the analysis are available as supplement (S3 Table, S8 Fig, S2 and S3 Files).

Electron microscopy and image analysis

Isolated SpaPR^{FLAG} complexes were deposited on glow-discharged carbon coated copper-palladium grids and stained with 0.75% uranyl formate. Micrograph acquisition was performed on a FEI Tecnai F30 Polara at 300 kV, equipped with a Gatan Ultrascan 4000 UHS CCD (4k x 4k pixels, physical pixel size of 15 μm), using the LEGINON automated image acquisition system [56]. The corrected magnification was 71950x, resulting in a pixel size of 2.08 Å/pixel. 11202 particles were picked from the micrographs with EMAN2 boxer [57]. Particle images were first subjected to a maximum-likelihood classification and alignment (ML2D) in XMIPP [58] and then further processed in IMAGIC-5 (Image Science Software GmbH) through multi-reference alignment and classification by multi-variate statistical analysis.

SpaPR pore assessment by biotin maleimide labeling

SpaP or SpaPQR^{FLAG} were moderately overexpressed in *S*. Typhimurium strain SB1770 (Δ*prgHIJK*, *flhD*::*tet*) from a rhamnose-inducible medium copy number plasmid by induction with 20 μM rhamnose. BM labeling was performed essentially as previously described [29], with minor modifications: After 3 h of induction, 0.2 ODU of bacterial cells were transferred to a fresh reaction tube and brought to the same volume by addition of fresh LB broth. Cells were labeled by addition of BM (EZ-link maleimide-PEG2-biotin, Thermo Pierce, final concentration 0.4 mM) for 30 min at room temperature with gentle agitation. The reaction was quenched by addition 2M β-mercaptoethanol to a final concentration of 10 mM. Cells were pelleted, re-suspended in SB buffer and incubated at 70˚C for 10 min. BM labeling of proteins was analyzed by SDS PAGE, Western blotting, and detection of BM with streptavidin DyLight

800 dye (Thermo pierce). Scanning of the PVDF membrane and image analysis was performed with a Li-Cor Odyssey system and image Studio 2.1.10 (Li-Cor).

For subcellular fractionation, BM-labeled bacterial cells were pelleted by centrifugation. The culture supernatant was harvested and TCA precipitated. The bacterial cell pellet was resuspended and used to prepare the periplasmic and cytoplasmic fractions as described elsewhere. Briefly, pellets were resuspended by pipetting gently in ice-cold spheroplast buffer (40% sucrose, 33 mM Tris-HCl, pH 8.0) with freshly prepared lysozyme to a final concentration of 200 μg/ml, 50 μg/ml DNAse and 1.5 mM EDTA. The mixture was left on ice for 30 min with gentle stirring. Spheroplasts were stabilized by adding 20 mM $MgCl_2$ and centrifuged at 3000 x g for 10 min at 4˚C. The supernatant was transferred to ultracentrifugation tubes and centrifuged at 30 krpm for 30 min at 4˚C in a Beckman TLA55 rotor to remove insoluble material. The supernatant (periplasmic fraction) was collected into fresh tube. The cytoplasmic fraction was prepared by resuspending the pellet of spheroplasts in 1 ml of 20 mM Tris-HCl, pH 8.0 and subsequent lysis by bead milling as described above. Lysates were transferred to ultracentrifugation tubes and centrifuged at 55 krpm for 45 min at 4˚C in a Beckman TLA55 rotor. The supernatant (cytoplasmic fraction) was collected into fresh tubes.

Supporting Information

S1 Table. Raw data of the size exclusion chromatography-multi angle laser light scattering analysis of the purified SpaPR complex.
(XLSX)

S2 Table. Crosslinked adducts identified by mass spectrometry.
(XLSX)

S3 Table. EC scores of coupling prediction.
(XLSX)

S4 Table. Strains and plasmids.
(XLSX)

S5 Table. Oligonucleotides.
(XLSX)

S1 Fig. Three detector calibration of the SEC-MALLS equipment and error calculation.
(PDF)

S2 Fig. Functional analysis of SpaP and SpaR *p*Bpa mutants. (A) Type III dependent secretion into the culture supernatant of indicated *p*Bpa mutants of SpaP and SpaR, respectively, was assayed by SDS PAGE and immunodetection of the early substrate InvJ and the intermediate substrate SipB. For two of the secretion-deficient SpaP mutants (T15X, M187X), assembly of SpaP into the needle complex was confirmed by 2-dimensional blue native/SDS PAGE (Fig 3C). Further, many secretion-defective *p*Bpa mutants showed productive crosslinks to other needle complex components. These results suggest that secretion-deficiency was not due to gross structural defects but rather the result of subtle conformational changes. (B) As in (A) but detailing secretion profiles of chromosome-encoded SpaP-*p*Bpa mutants.
(TIF)

S3 Fig. SDS PAGE analysis of isolated needle complexes of SpaP and SpaR *p*Bpa mutants with and without UV photocrosslinking for mass spectrometrical identification of crosslinking partners. (A) Immunodetection of $SpaP^{FLAG}$ and $SpaR^{FLAG}$, respectively, on Western blots of purified needle complexes of *S*. Typhimurium expressing indicated SpaP or SpaR *p*Bpa

mutants separated by SDS PAGE. Each sample is shown with and without UV-irradiation to induce photocrosslinking of the *p*Bpa to neighboring interaction partners. Identified interaction partners are indicated at the respective bands. A summary of the MS identifications is shown in S2 Table. (B) Coomassie stained SDS PAGE gels of the UV-irradiated samples shown in (A). Gel pieces were cut out at positions of crosslinking adducts identified by Western blotting and immunodetection for subsequent in gel Trypsin digestion and MS analysis.
(TIF)

S4 Fig. Class averages of negative-stained isolated SpaPR complexes imaged by electron microscopy. 91 classes are shown. The length of the scale bar in the upper left corner represents 50 Å.
(TIF)

S5 Fig. Controls for general bacterial lysis for biotin maleimide labeling experiments. (A) Coomassie-stained gel (left) and immunodetection (cytoplasmic marker RNA polymerase (RNApol), periplasmic marker maltose binding protein (MBP), right) of SDS PAGE-separated whole cell lysates and cell culture supernatants, respectively, of *S*. Typhimurium Δ*prgHIJK*, *flhD*::*tet* moderately overexpressing indicated proteins from a medium copy number plasmid (pT12). Equal culture volumes were loaded in each well. (B) Immunodetection of SDS PAGE-separated cell culture supernatants, periplasmic fractions, and cytoplasmic fractions, respectively, of *S*. Typhimurium Δ*prgHIJK*, *flhD*::*tet* moderately overexpressing indicated proteins from a medium copy number plasmid (pT12). Equal culture volumes were loaded in each well. RNApol serves as a marker protein for cytoplasmic proteins, MBP serves as a marker protein for periplasmic proteins. (C) Growth curves of *S*. Typhimurium Δ*prgHIJK*, *flhD*::*tet* overexpressing indicated proteins from a medium copy number plasmid (pT12) with the indicated concentrations of rhamnose.
(TIF)

S6 Fig. Prediction of topology and of the propensity of membrane integration of SpaR. (A) Topcons prediction of SpaR (topcons.cbr.su.se). (B) Prediction of ΔG for membrane integration propensity of SpaR using a sliding window between 18 and 31 amino acids (dgpred.cbr.su.se). (C) Protter visualization of the topology model of SpaR comprising 3 TM helices and an N-out/C-in orientation. Positions of detected crosslinks of SpaR to other T3SS components are indicated in color.
(TIF)

S7 Fig. Position and sequence of epitope tags used in SpaP, SpaR, and SpaS.
(PDF)

S8 Fig. Contact map of top 291 residue couplings. Abbreviations: norm. normalized, exp. experimentally.
(PNG)

S1 File. SEC-MALLS ASTRA calculations.
(PDF)

S2 File. Input alignment in fasta format.
(A2M)

S3 File. Notebook containing couplings analysis.
(HTML)

Acknowledgments

We acknowledge Ewa Folta-Stogniew and the Biophysics Resource of Keck Facility at Yale University for performing size exclusion chromatography—multi angle laser light scattering analysis.

Author Contributions

Conceptualization: JEG SW.

Data curation: MFW.

Formal analysis: TD MTM MJB CS OK TCM SW.

Funding acquisition: JEG SW.

Investigation: TD MTM MJB PA JY MFW CS SW.

Methodology: TD MTM MJB PA SZ IG MFW SL BM CVR SW.

Software: CS OK.

Validation: TD MTM SZ.

Visualization: TD MTM MJB JY CS SW.

Writing – original draft: TD MTM MJB PA JY MFW CS SW.

Writing – review & editing: JEG SW.

References

1. Galán JE, Lara-Tejero M, Marlovits TC, Wagner S. Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Annu Rev Microbiol. 2014; 68: 415–438. doi: 10.1146/annurev-micro-092412-155725 PMID: 25002086
2. Galán JE. SnapShot: effector proteins of type III secretion systems. Cell. 2007; 130: 192–192.e2. doi: 10.1016/j.cell.2007.06.042 PMID: 17632065
3. Schraidt O, Marlovits TC. Three-dimensional model of Salmonella's needle complex at subnanometer resolution. Science. 2011; 331: 1192–1195. doi: 10.1126/science.1199358 PMID: 21385715
4. Kubori T, Matsushima Y, Nakamura D, Uralil J, Lara-Tejero M, Sukhan A, et al. Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. Science. 1998; 280: 602–605. PMID: 9554854
5. Radics J, Königsmaier L, Marlovits TC. Structure of a pathogenic type 3 secretion system in action. Nat Struct Mol Biol. 2014; 21: 82–87. doi: 10.1038/nsmb.2722 PMID: 24317488
6. Akeda Y, Galán JE. Chaperone release and unfolding of substrates in type III secretion. Nature. 2005; 437: 911–915. doi: 10.1038/nature03992 PMID: 16208377
7. Lara-Tejero M, Kato J, Wagner S, Liu X, Galán JE. A sorting platform determines the order of protein secretion in bacterial type III systems. Science. 2011; 331: 1188–1191. doi: 10.1126/science.1201476 PMID: 21292939
8. Fan F, Ohnishi K, Francis NR, Macnab RM. The FliP and FliR proteins of Salmonella typhimurium, putative components of the type III flagellar export apparatus, are located in the flagellar basal body. Mol Microbiol. 1997; 26: 1035–1046. PMID: 9426140
9. Wagner S, Königsmaier L, Lara-Tejero M, Lefebre M, Marlovits TC, Galán JE. Organization and coordinated assembly of the type III secretion export apparatus. Proc Natl Acad Sci USA. 2010; 107: 17745–17750. doi: 10.1073/pnas.1008053107 PMID: 20876096
10. Burkinshaw BJ, Strynadka NCJ. Assembly and structure of the T3SS. Biochim Biophys Acta. 2014; 1843: 1649–1663. doi: 10.1016/j.bbamcr.2014.01.035 PMID: 24512838
11. Abrusci P, McDowell MA, Lea SM, Johnson S. Building a secreting nanomachine: a structural overview of the T3SS. Curr Opin Struct Biol. Elsevier Ltd; 2014; 25: 111–117.

12. Zilkenat S, Franz-Wachtel M, Stierhof Y-D, Galán JE, Macek B, Wagner S. Determination of the stoichiometry of the complete bacterial type III secretion needle complex using a combined quantitative proteomic approach. Molecular & Cellular Proteomics. 2016; 15: 1598–1609.
13. Zarivach R, Deng W, Vuckovic M, Felise HB, Nguyen HV, Miller SI, et al. Structural analysis of the essential self-cleaving type III secretion proteins EscU and SpaS. Nature. Nature Publishing Group; 2008; 453: 124–127.
14. Worrall LJ, Vuckovic M, Strynadka NCJ. Crystal structure of the C-terminal domain of the Salmonella type III secretion system export apparatus protein InvA. Protein Sci. 2010; 19: 1091–1096. doi: 10.1002/pro.382 PMID: 20306492
15. Abrusci P, Vergara-Irigaray M, Johnson S, Beeby MD, Hendrixson DR, Roversi P, et al. Architecture of the major component of the type III secretion system export apparatus. Nat Struct Mol Biol. 2012; 20: 99–104. doi: 10.1038/nsmb.2452 PMID: 23222644
16. Hirano T, Mizuno S, Aizawa S-I, Hughes KT. Mutations in flk, flgG, flhA, and flhE that affect the flagellar type III secretion specificity switch in Salmonella enterica. J Bacteriol. 2009; 191: 3938–3949. doi: 10.1128/JB.01811-08 PMID: 19376867
17. Minamino T, Shimada M, Okabe M, Saijo-Hamano Y, Imada K, Kihara M, et al. Role of the C-terminal cytoplasmic domain of FlhA in bacterial flagellar type III protein export. J Bacteriol. American Society for Microbiology; 2010; 192: 1929–1936.
18. Hara N, Namba K, Minamino T. Genetic characterization of conserved charged residues in the bacterial flagellar type III export protein FlhA. PLoS ONE. 2011; 6: e22417. doi: 10.1371/journal.pone.0022417 PMID: 21811603
19. Ferris HU, Minamino T. Flipping the switch: bringing order to flagellar assembly. Trends Microbiol. 2006; 14: 519–526. doi: 10.1016/j.tim.2006.10.006 PMID: 17067800
20. Monjarás Feria JV, Lefebre MD, Stierhof Y-D, Galán JE, Wagner S. Role of autocleavage in the function of a type III secretion specificity switch protein in *Salmonella enterica* serovar Typhimurium. MBio. American Society for Microbiology; 2015; 6: e01459–15.
21. Diepold A, Wiesand U, Cornelis GR. The assembly of the export apparatus (YscR,S,T,U,V) of the *Yersinia* type III secretion apparatus occurs independently of other structural components and involves the formation of an YscV oligomer. Mol Microbiol. 2011; 82: 502–514. doi: 10.1111/j.1365-2958.2011.07830.x PMID: 21923772
22. Yerushalmi G, Litvak Y, Gur-Arie L, Rosenshine I. Dynamics of expression and maturation of the type III secretion system of enteropathogenic Escherichia coli. J Bacteriol. 2014; 196: 2798–2806. doi: 10.1128/JB.00069-14 PMID: 24837293
23. Diepold A, Wagner S. Assembly of the bacterial type III secretion machinery. FEMS Microbiol Rev. 2014; 38: 802–822. doi: 10.1111/1574-6976.12061 PMID: 24484471
24. Collazo CM, Galan JE. Requirement for exported proteins in secretion through the invasion-associated type III system of *Salmonella* typhimurium. Infection and Immunity. 1996; 64: 3524–3531. PMID: 8751894
25. Farrell IS, Toroney R, Hazen JL, Mehl RA, Chin JW. Photo-cross-linking interacting proteins with a genetically encoded benzophenone. Nat Methods. 2005; 2: 377–384. doi: 10.1038/nmeth0505-377 PMID: 16170867
26. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. Protein 3D structure computed from evolutionary sequence variation. Sali A, editor. PLoS ONE. Public Library of Science; 2011; 6: e28766.
27. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci USA. National Acad Sciences; 2011; 108: E1293–301.
28. Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. Elife. 2014; 3: 65.
29. Park E, Rapoport TA. Preserving the membrane barrier for small molecules during bacterial protein translocation. Nature. Nature Publishing Group; 2011; 473: 239–242.
30. Marlovits TC, Kubori T, Sukhan A, Thomas DR, Galán JE, Unger VM. Structural insights into the assembly of the type III secretion needle complex. Science. 2004; 306: 1040–1042. doi: 10.1126/science.1102610 PMID: 15528446
31. Van Arnam JS, McMurry JL, Kihara M, Macnab RM. Analysis of an engineered Salmonella flagellar fusion protein, FliR-FlhB. J Bacteriol. 2004; 186: 2495–2498. doi: 10.1128/JB.186.8.2495-2498.2004 PMID: 15060055

32. Marlovits TC, Kubori T, Lara-Tejero M, Thomas D, Unger VM, Galán JE. Assembly of the inner rod determines needle length in the type III secretion injectisome. Nature. 2006; 441: 637–640. doi: 10.1038/nature04822 PMID: 16738660

33. Lefebre MD, Lefebre MD, Galan JE, Galán JE. The inner rod protein controls substrate switching and needle length in a Salmonella type III secretion system. Proc Natl Acad Sci USA. 2014; 111: 817–822. doi: 10.1073/pnas.1319698111 PMID: 24379359

34. Ohnishi K, Fan F, Schoenhals GJ, Kihara M, Macnab RM. The FliO, FliP, FliQ, and FliR proteins of Salmonella typhimurium: putative components for flagellar assembly. J Bacteriol. 1997; 179: 6092–6099. PMID: 9324257

35. Berger C, Robin GP, Bonas U, Koebnik R. Membrane topology of conserved components of the type III secretion system from the plant pathogen Xanthomonas campestris pv. vesicatoria. Microbiology (Reading, Engl). 2010; 156: 1963–1974.

36. Allaoui A, Woestyn S, Sluiters C, Cornelis GR. YscU, a Yersinia enterocolitica inner membrane protein involved in Yop secretion. J Bacteriol. 1994; 176: 4534–4542. PMID: 8045883

37. Hoiseth SK, Stocker BA. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. Nature. 1981; 291: 238–239. PMID: 7015147

38. Folta-Stogniew E. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. Methods Mol Biol. New Jersey: Humana Press; 2006; 328: 97–112.

39. Pringle SD, Giles K, Wildgoose JL, Williams JP, Slade SE, Thalassinos K, et al. An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. International Journal of Mass Spectrometry. 2007; 261: 1–12.

40. Hernández H, Robinson CV. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat Protoc. 2007; 2: 715–726. doi: 10.1038/nprot.2007.73 PMID: 17406634

41. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res. Oxford University Press; 2015; 43: W401–7.

42. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature. 2007; 450: 1026–1030. doi: 10.1038/nature06387 PMID: 18075582

43. Omasits U, Ahrens CH, Müller S, Wollscheid B. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics. Oxford University Press; 2014; 30: 884–886.

44. Borchert N, Dieterich C, Krug K, Schütz W, Jung S, Nordheim A, et al. Proteogenomics of Pristionchus pacificus reveals distinct proteome structure of nematode models. Genome Res. Cold Spring Harbor Lab; 2010; 20: 837–846.

45. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. Nature Publishing Group; 2007; 2: 1896–1906.

46. Franz-Wachtel M, Eisler SA, Krug K, Wahl S, Carpy A, Nordheim A, et al. Global detection of protein kinase D-dependent phosphorylation events in nocodazole-treated human cells. Molecular & Cellular Proteomics. American Society for Biochemistry and Molecular Biology; 2012; 11: 160–170.

47. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. Nature Publishing Group; 2008; 26: 1367–1372.

48. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. American Chemical Society; 2011; 10: 1794–1805.

49. Carpy A, Krug K, Graf S, Koch A, Popic S, Hauf S, et al. Absolute proteome and phosphoproteome dynamics during the cell cycle of Schizosaccharomyces pombe (Fission Yeast). Molecular & Cellular Proteomics. American Society for Biochemistry and Molecular Biology; 2014; 13: 1925–1936.

50. Balakrishnan S, Kamisetty H, Carbonell JG, Lee S-I, Langmead CJ. Learning generative models for protein fold families. Proteins. Wiley Subscription Services, Inc., A Wiley Company; 2011; 79: 1061–1078.

51. Ekeberg M, Lövkvist C, Lan Y, Weigt M, Aurell E. Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. Phys Rev E Stat Nonlin Soft Matter Phys. American Physical Society; 2013; 87: 012707.

52. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc Natl Acad Sci USA. National Acad Sciences; 2013; 110: 15674–15679.

53. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. BioMed Central; 2010; 11: 431.

54. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. Oxford University Press; 2015; 31: 926–932.

55. Perez F, Granger BE. IPython: A system for interactive scientific computing. Comput Sci Eng. 2007; 9: 21–29.

56. Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, Quispe J, et al. Automated molecular microscopy: the new Leginon system. J Struct Biol. 2005; 151: 41–60. doi: 10.1016/j.jsb.2005.03.010 PMID: 15890530

57. Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, Rees I, et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 2007; 157: 38–46. doi: 10.1016/j.jsb.2006.05.009 PMID: 16859925

58. Sorzano COS, Marabini R, Velázquez-Muriel J, Bilbao-Castro JR, Scheres SHW, Carazo JM, et al. XMIPP: a new generation of an open-source image processing package for electron microscopy. J Struct Biol. 2004; 148: 194–204. doi: 10.1016/j.jsb.2004.06.006 PMID: 15477099

Acknowledgment

Firstly, I would like to express my gratitude to Samuel Wagner for giving me the opportunity to do my doctoral thesis in his working group, and for his friendly encouragement and advises during it's completion. I also would like to thank Andreas Peschel for being my second supervisor, Boris Macek and Mirita Franz for the fruitful collaboration, and my thesis committee.

Many thanks go to the whole working group Wagner, for help with experiments and presentations, for lending a friendly ear, and for the many breakfasts and cakes! In particular I am grateful to Melanie Riess and Andrea Eipper for all the great help right from the start.

I especially want to thank Laura Verbeek, Elena Grin and Iwan Grin for sitting in the same boat, proofreading this thesis and providing much needed moral support. I'm not sure what I would have done without you.

Lastly, I would like to thank my parents and my big brother for their encouragement, support and confidence in me throughout the years.