Structural insights into Tc toxins from human and insect pathogenic bacteria

Dissertation

zur Erlangung des akademischen Grades des Doktors
der Naturwissenschaften (Dr. rer. nat.)

Vorgelegt der Fakultät Chemie und Chemische Biologie
der Technischen Universität Dortmund

Angefertigt am Max-Planck-Institut für
Molekulare Physiologie in Dortmund

Vorgelegt von

Franziska Leidreiter

First referee: Prof. Dr. Stefan Raunser

Department of Structural Biochemistry, Max Planck Institute of Molecular Physiology, Dortmund

Faculty of Chemistry and Chemical Biology, Technische Universität Dortmund

Second referee: Prof. Dr. Roland Winter

Department of Physical Chemistry I,

Faculty of Chemistry and Chemical Biology, Technische Universität Dortmund

Date of submission: 26.04.2019

Acknowledgements

Completing this work would have been all the more difficult were it not for the support and friendship provided by colleagues, friends and family.

First of all, I want to express my gratitude to my supervisor Prof. Dr. Stefan Raunser, who gave me the opportunity to be part of his inspiring research group. I am really grateful for his continuous guidance and advice he has provided throughout my time as a PhD student.

My thanks also go to Prof. Dr. Roland Winter for his interest in my work and for his kindly acceptance to be my second referee.

Moreover, I would like to thank Prof. Dr. Roland Benz, who was so kind to provide me his lab and equipment for black lipid bilayer measurements and was very helpful during analysis of the results. I want to especially thank the IMPRS for the constant support and guidance throughout my time at the MPI. In addition, I want to thank my thesis advisory committee members Prof. Dr. Hemmo Meyer and Prof. Dr. Dominik Boos for the helpful discussions and revisions of my progressing work.

I am particularly grateful for the friendly and enthusiastic working atmosphere in the Raunser group as well as the whole department. Many thanks to all members, including the technical and administration staff. I really enjoyed working with all of you and especially the time we spent together outside the lab.

Without the help of Dr. Oliver Hofnagel and Dr. Daniel Prumbaum I would not have been able to collect one data set after the other. Thank you for your constant support and time. Moreover, I would like to thank Dr. Christos Gatsogiannis and Dr. Julian von der Ecken for the patient introduction into EM and image processing.

Many thanks go to the members of the toxin group for continuing fruitful and helpful discussions. I especially want to thank Dr. Daniel Roderer, who did not only contribute experiments to this work but was always available for discussions and help in the lab. I moreover grateful acknowledge the help of Alina Elsner for constant support in the lab, especially during protein purification and cloning.

I will really miss the many encouraging discussions about science and beyond with my former, present and temporarily office mates. Philine, Claudia and Birte, I am very thankful for our friendship. I will miss you and our intense discussions as well as the nice coffee company in the morning. Special thanks also to Philine, who accompanies me since the first day of university and was a great help, together with Arne, for revising the first version of this manuscript.

Last but not least I would like to thank my friends and family, especially my parents, who enabled me to study and supported me right from the start and were always there for me. I would not have succeeded and accomplished this work without my husband – thank you for always being by my side.

I. List of Abbreviations

°C	degree Celsius
Arg	arginine
Asp	aspartate
BPL	brain polar lipids
BSA	bovine serum albumine
CV	column volume
DDM	n-Dodecyl-β-D-maltopyranoside
DSF	differential scanning fluorimetry
EM	electron microscopy
FEG	field emission gun
FSC	Fourier-Shell-Correlation
Glu	glutamate
h	hours
His	histidine
kcal	kilocalorie
kDa, MDa	kilo Dalton, mega Dalton
mg	milligram
min	minutes
ml	milliliter
MSP	membrane scaffold protein
Ni-NTA	nickel-nitrilotriacetic acid
nm, µm	nanometer, micrometer
OG	n-Octyl-β-D-glucopyranoside
oN	over night
PDB-ID	protein data bank identification
PFT	pore-forming toxin
PLL	poly-L-lysine
POPC	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
px	pixel
QF	Quantifoil
RBD	receptor-binding domain
s	seconds
SEC	size exclusion chromatography
SNR	signal-to-noise ratio
SPA	single particle analysis
SPARX	single particle analysis for resolution extension
SPHIRE	sparx for high resolution electron microscopy
Tyr	tyrosine

II. List of Figures

III. List of Tables

1 Abstract & Zusammenfassung

1.1 Abstract

Many pathogenic bacteria produce virulence factors that act as pore-forming toxins (PFTs) on the cytoplasmic membrane of target hosts. PFTs can be distinguished in binary (AB) toxins that actively transport a toxic counterpart into the cell and in conventional PFTs, which perforate the host membrane and translocate ions. Upon membrane interaction, the PFTs provoke different effects such as abolishing the transmembrane potential, altering signal pathways or disruption of the cell cytoskeletal. Ultimately the attack of PFTs leads to cell death. Amongst the PFTs is the family of the tripartite toxin complex (Tc) toxins that act as α-helical PFTs. Tc toxins are composed of three components: namely TcA, TcB and TcC; and only the tripartite complex (A_5BC) is biologically active. Tcs use a special syringe-like mechanism to attack the respective host cell and translocate a toxic enzyme into the host cytosol. TcA forms the membrane-perforating translocation channel and the toxic enzyme is located at the TcBTcC dimer. Tc toxins exist in a wide range of bacteria including insect, plant and human pathogenic organisms. However, so far, only the structure of the entomopathogenic *Photorhabdus luminescens* Tc toxins has been studied in great detail. Therefore, it is still unclear if other insect or human pathogenic homologues have a similar architecture and use a common mechanism.

To unravel the mechanism of action and host specificity of these toxins, here the near-atomic resolution structures of three TcAs from human and insect pathogenic bacteria were solved using electron cryo microscopy; Yp-TcaATcaB from *Yersinia pseudotuberculosis* (3.3 Å), Mm-TcdA4 from *Morganella morganii* (3.3 Å) and Xn-XptA1 from *Xenorhabdus nematophila* (2.8 Å). The structures revealed that the overall composition and domain organization of the three TcAs is similar to the well-studied TcdA1 from *P. luminescens*. They assemble as a pentameric bottle-shaped structure with a central α-helical channel surrounded by an outer shell composed of a conserved α-helical domain and more variable β-sheet receptor-binding domains, implying species-specific receptor interactions. Functionally crucial structural features for the mechanism of action of *P. luminescens* TcdA1, i.e. the linker domain, the translocation channel, a molecular knot and the TcB-binding domain show high structural similarities in all analyzed TcAs. The analyzed TcAs form ion conductive pores and the pore state could be induced by a pH shift *in vitro* for all studied TcAs. Interestingly, fully functional chimeric holotoxins can be formed, combining TcAs and TcBTcC cocoons from different organisms. The previously described electrostatic lock at the neuraminidase-like domain of TcdA1 is not conserved within the here studied TcAs. However, one conserved ionic interaction pair was identified that stabilizes the complex in the shell domain. It might act as a latch that lead to the pH-induced opening of the shell after other interacting protomer interfaces are destabilized upon pH shift to pH 4 or 11. Altogether, results

obtained from this work show that Tc toxins from different organisms share a common mechanism of action, while the variability in receptor-binding domains and formation of the active pore state enable targeting of different hosts. Thus, this work presents the first detailed structural insights into multiple TcAs derived from various prokaryotic organisms.

1.2 Zusammenfassung

Porenbildende Toxine (PBT) sind weit verbreitet in Prokaryoten und Eukaryoten und perforieren durch Insertion einer Pore die Zellmembran. Pathogene Bakterien nutzen diese PBT als Virulenzfaktoren, um effizient und gezielt den Zelltod der Wirtszelle hervorzurufen. Dabei können PBTs in binäre (AB) Toxine unterschieden werden, welche aktiv eine toxische Komponente in die Zelle transportieren, und in konventionellen PBTs, welche die Wirtsmembran perforieren und somit einen Ionentransport induzieren. Dabei lösen die Toxine unterschiedliche Mechanismen aus, wie zum Beispiel die Aufhebung des Transmembranpotenzials, Unterbrechung von Signalkaskaden oder die Zerstörung des Zytoskeletts der Zelle. Letztendlich führt der Angriff von PBTs zum Zelltod. Die Tc (*toxin complex*) Toxine gehören zu der Familie der helikalen porenbildenden Toxine und bestehen aus drei Komponenten: TcA, TcB und TcC. Nur im Zusammenspiel der drei Komponenten (A_5BC) sind die Tc Toxine biologisch aktiv. Sie verwenden einen speziellen spritzenartigen Mechanismus, um die jeweilige Wirtszelle anzugreifen und ein toxisches Enzym in die Zelle zu translozieren. Dabei fungiert die TcA-Komponente als Membran-perforierender Translokationskanal, während sich das toxische Enzym im TcB-TcC Dimer befindet. Diese Tc Toxine kommen in einer Vielzahl von insekten-, pflanzen- und humanpathogene Bakterien. Bislang wurde jedoch nur die Struktur von Tc Toxinen aus dem entomopathogenen Bakterium *Photorhabdus luminescens* im Detail untersucht. Daher ist es immer noch unklar, ob Tc Toxin von insekten- oder humanpathogenen Organismen eine ähnliche Architektur haben und einen gemeinsamen Mechanismus verwenden.

Um den Wirkungsmechanismus und die Wirtsspezifität dieser Toxine aufzuklären, wurden in dieser Arbeit die hochaufgelösten Strukturen von drei TcAs aus human- und insektenpathogenen Bakterien mit Hilfe der Elektronenkryomikroskopie gelöst; Yp-TcaATcaB aus *Yersinia pseudotuberculosis* (3,3 Å), Mm-TcdA4 aus *Morganella morganii* (3,3 Å) und Xn-XptA1 aus *Xenorhabdus nematophila* (2,8 Å). Strukturelle Analysen ergaben, dass die drei TcAs eine einheitliche Architektur und Domänenorganisation im Vergleich zu TcdA1 von *P. luminescens* aufweisen. Sie bilden flaschenförmige Pentamerstrukturen, welche in drei strukturelle Untereinheiten aufgeteilt werden können; ein zentralen helikalen Kanal, welcher von einer äußeren Hülle umgeben ist, die aus konservierten α-helikalen Domänen und variableren rezeptorbindenden Domänen aus β-Faltblättern besteht, was auf Spezies-spezifische Rezeptorwechselwirkungen schließen lässt. Funktionell entscheidende Strukturmerkmale für

den Wirkmechanismus von *P. luminescens* TcdA1, wie die Linker Domäne, der Translokationskanal, ein molekularer Knoten und die TcB-bindende Domäne, zeigen in allen analysierten TcAs hohe strukturelle Ähnlichkeiten. Ferner bilden die analysierten TcAs ionenleitende Poren und der Porenzustand wurde durch eine pH-Verschiebung *in vitro* für alle untersuchten TcAs induziert. Interessanterweise können voll funktionsfähige chimäre Holotoxine gebildet werden, wobei TcAs und TcBTcC-Kokons aus verschiedenen Organismen kombiniert werden können. Das zuvor beschriebene elektrostatische Schloss in der neuraminidaseähnlichen Domäne von TcdA1 ist in den hier untersuchten TcAs nicht konserviert und höchstwahrscheinlich nicht allein für die Öffnung der äußeren Hülle verantwortlich. Jedoch wurde ein konserviertes ionisches Wechselwirkungspaar identifiziert, welches den Komplex in der Hüllen-Domäne stabilisiert. Dieses Interaktionspaar könnte als ein pH-abhängiges Schnappschloss fungieren, welches sich öffnet, nachdem weitere Protomer-Interaktionsflächen durch elektrische Abstoßung bei pH 11 oder pH 4 destabilisiert wurden.

Insgesamt zeigen die Ergebnisse dieser Arbeit, dass Tc Toxine aus verschiedenen Organismen höchstwahrscheinlich einen gemeinsamen Wirkmechanismus aufweisen, während die Variabilität der rezeptorbindenden Domänen und in der Ausbildung des aktiven Porenzustandes die Wirtsspezifität ermöglichen. Somit liefert diese Arbeit erste detaillierte Einblicke in die Strukturzusammensetzung diverser TcAs, welche aus verschiedenen prokaryotischen Organismen stammen.

2 Introduction

2.1 Bacterial pore-forming proteins

Pore-forming proteins are an ancient protein family widely spread in prokaryotes and eukaryotes. They have either regulatory functions, such as the Bcl2 pore-forming protein in apoptosis signaling pathways[1], participate in the defensive immune response (perforin)[2] or attack cells as pore-forming toxins (PFTs). PFTs are not only present in bacteria but can also be found in eukaryotic organisms including plants, animals and fungi[3,4]. However, bacterial PFTs cover with 25-30 % the largest group of bacterial virulence factors[5]. Because of their disease-causing abilities, they are an important research target for new therapeutics and broadly applicable antimicrobial prophylactics[6].

The mechanism of action of PFTs and therefore of pore formation has already been studied over decades. Insights into their mechanism have been gained, but the distinct structural fold, pore architecture and mechanism of membrane insertion makes the complete understanding of the complex mechanism challenging[7]. PFTs are generally secreted as water-soluble protomers[6]. After binding to specific receptors, they assemble into oligomeric structures to form a pore. Upon oligomerization, they undergo conformational changes to generate hydrophobic surfaces allowing spontaneous membrane insertion[8,9]. PFTs can be distinguished in binary (AB) toxins that actively transport a toxic counterpart into the cell and in conventional PFTs that perforate the host membrane and translocate ions[9,10]. PFTs that translocate a toxic enzyme act as binary AB toxins such as anthrax[11] and diphtheria[12] which will be discussed in more detail within the chapter 2.1.1. Moreover, two major classes define the PFTs based on the secondary structure of the pore-forming regions. The membrane spanning domain either consists of α-helices or β-barrels, resulting in the classification as α-PFTs and β-PFTs[13].

Although belonging to α-PFTs, members of this class exhibit variances regarding their mechanism and structural composition, thus allowing to further distinguish between colicins-, actinoporins- and ClyA-family members[7] as well as members of the binary AB_5 toxin family[14]. Member of the colicin-family such as colicin E1 from *Escherichia coli* are completely α-helical. Upon membrane binding, an hydrophobic helical hairpin inserts into the inner membrane resulting in a voltage-gated channel, that abolishes the transmembrane potential[13,15]. In contrast, the actinoporins have a β-sandwich structure flanked by two α-helices. The N-terminus contains an amphipathic helix that detaches and inserts into the membrane to form a pore as shown for fragaceatoxin C (FraC) from the sea anemone *Actinia fragacea*[16,17]. Cytolysin A (ClyA) from *E. coli* is one extensively studied example for an α-PFT, whose soluble monomer and pore state have been structurally characterized[18-20]. The soluble monomer is an elongated protein mainly composed of α-helices and one short hydrophobic β-hairpin (β-tongue). In proximity to membranes, the β-tongue interacts with the lipid bilayer and therefore detaches from the protein core. This results

in a massive conformational change, in which the N-terminal helix rearranges and spans the transmembrane region[21]. An example for an α-PFT AB_5 toxin is the cholera toxin from *Vibrio cholera*[22]. The pentameric B subunit interacts with GM_1 gangliosides that are exposed on the luminal surface of intestinal epithelial cells and forms a pentameric helical pore though which the enzymatic A subunit is translocated into the cell. Another recently described member of the α-PFT is the XaxAB protein from *Xenorhabdus nematophila*. Unlike predicted, structural studies revealed that the toxin is a bi-component α-PFT[23] instead of acting as a binary AB toxin, although it comprises two subunits XaxA and XaxB[24]. The A subunit activates and stabilizes the B subunit which comprises the membrane spanning domain. Thus, the pool of α-PFT is growing and is far from being understood in all its details.

The members of the β-PFTs are less diverse and better characterized than the α-PFTs. Four well analyzed examples are the perfringolysin O (PFO) from *Clostridium perfringens*, which belongs to the CDC family (cholesterol-dependent cytolysin), aerolysin toxin from *Aeromonas spp.*[25], the α-hemolysin (Hla) from *Staphylococcus aureus* (hemolysin family)[26] and the anthrax toxin from *Bacillus anthracis*[27]. The structure of a soluble α-hemolysin (Hla) monomer has been analyzed and in addition, the heptameric pore conformation revealed a β-barrel with 14 antiparallel β-strands. Thus, each protomer contributes two β-strands to the channel[28]. For the larger assembly of PFO, massive conformational changes have to occur upon oligomerization. Indeed, the toxin undergoes an α-helix to β-strand transition[29] to coordinate the insertion of up to 140 individual β-strands. This results in a pore of 25-30 nm diameter, that comprises 30-50 individual PFO protomers[30]. Thus, the stoichiometry of the oligomeric assemblies varies between PFTs, which have to perform consecutive steps in order to reach toxicity and evolve to cytolytic transmembrane oligomers[7].

Generally, all members of the PFT family assemble from a soluble monomer into a conductive transmembrane pore state. The oligomerization is thought to occur already at the target membrane. Commonly, bacteria release low concentrations of toxins and after efficient binding to the target cell, the PFTs are enriched on the membrane, which favors their oligomerization[13]. The mechanism of oligomerization varies between single PFTs, some need to be activated to oligomerize. One example is the aerolysin toxin, which is secreted as an inactive dimer that is only able to oligomerize upon proteolytic cleavage by furin[25]. Furthermore, receptor binding or a pH change can trigger the activation of pore formation. Once the monomers are localized on the membrane surface, there are two pathways for pore formation and membrane insertion of PFTs (Figure 1). For the described β-PFTs, the oligomerization occurs at the membrane leading to a prepore or intermediate state, where the β-barrel is preformed[7]. Subsequently, the β-PFTs performs rearrangements resulting in membrane insertion and a pore formation that spans the bilayer with a β-barrel (Figure 1, pathway 1). In the second pathway, which is followed by most α-PFTs, the membrane insertion happens simultaneously with an oligomerization mechanism. Thus, formation of partial pores (arc structures) can occur prior to formation

of complete pores (Figure 1, pathway 2). The final result for all PFTs is the formation of a transmembrane pore with various characteristics such as architecture, stoichiometry, size and conduction features[7]. Nevertheless, due to the different folding that leads to distinct pore architecture and membrane insertions, the PFTs exhibit specific characteristics. The host cell specificity is furthermore based on the interaction with specific membrane bound receptors, lipids and sugars[7]. Exemplarily, various PFTs have an affinity for lipids or specific lipid domain such as CDCs, that require cholesterol as an interaction partner in the mammalian cell membrane[31] or actinoporins that bind sphingomyelin. In contrast, the aerolysin toxin from *Aeromonas spp* interacts with GPI-anchored proteins such as Thy-1[32].

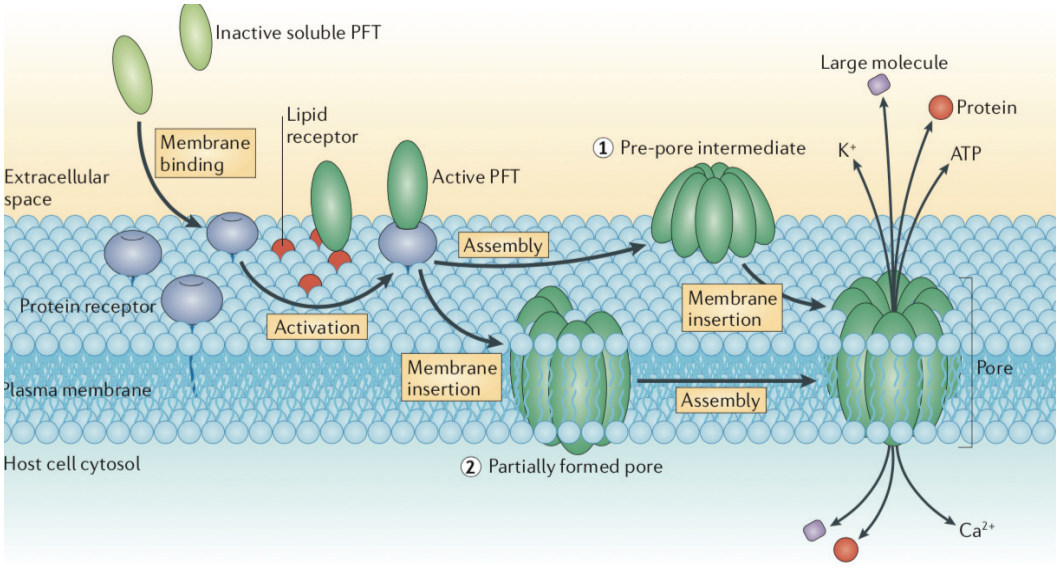


Figure 1: Scheme of possible oligomerization and pore formation pathways of PFTs. The soluble and inactive monomeric PFT binds first to the plasma membrane via specific lipid or protein receptor interactions. That results in a higher concentration of the PFT monomers and in some cases, leads to activation. Afterwards, two possible pathways can occur: the protomers assemble to form a prepore intermediate on the membrane. After oligomerization, the pore inserts into the membrane (pathway 1). Another possible route is that the protomers insert into the membrane as dimers or lower ordered oligomers and generate a partially formed pore (pathway 2). After this step or also simultaneously, the pore is completed by further monomer assembly. Reprinted with permission from Springer Nature[7].

Pore formation of PFTs in the membrane results in their toxicity and leads to cell death either by translocation of ions and thus altering the cell homeostasis[9,10], or by actively transporting a toxic counterpart into the cell. The translocated toxins are mostly enzymatically active proteins, as for example the anthrax toxin[11]. In these cases, the PFTs are generally composed of two subunits and are defined as binary or AB toxins[33].

2.1.1 Binary toxins

The members of AB pore forming toxins exhibit a different mechanism of host attacking. In contrast to other PFTs, these toxins perforate the cell membrane but furthermore use the transmembrane pore to translocate toxic enzymes into the host. The A subunits are commonly the enzymatic toxins whereas the B subunits represent the pore forming domain[13]. As mentioned previously, the anthrax toxin from *Bacillus anthracis* is a member of the binary β-PFTs[27]. This toxin is composed of three subunits; the protective antigen (PA), the lethal factor (LF) and the edema factor (EF), which are not toxic as single proteins [11]. The anthrax toxin comprises two A subunits, LF and EF. These two enzymes act as a Zn^{2+}-protease as well as a calmodulin and Ca^{2+} dependent adenylate cyclase[34], respectively. PA is the pore forming domain, the B subunit, that is activated upon proteolytic cleavage[11] after receptor binding on the cell surface[35]. The activated PA self-associates into a heptameric or octameric prepore complex, revealing no major changes from the monomeric PA to the oligomeric structure[36,37]. The oligomeric activated PA is capable of binding three (as heptamer) or four (as octamer) enzymes, either only LF, only EF or both simultaneously[38]. By receptor mediated endocytosis, the complex is internalized and transported to an endosomal compartment. Upon a pH shift to low pH in the early endosome, the prepore undergoes a conformational change to a mushroom-like pore. This conformational change results in generating a large hydrophobic surface that spans the transmembrane regions as a β-barrel. Each protomer provides one β-hairpin to form an amphipathic transmembrane β-barrel[14]. The positive membrane potential is thought to initiate the translocation of LF or EF in an N- to C-terminal direction[39]. Due to the narrowness of the transmembrane channel of ~12-15 Å, the translocated protein needs to pass either partially or completely unfolded through the channel into the cytosol. A ring of seven phenylalanine is reported to block the ion passage to preserve the pH gradient, which is the driving force of translocation[27,40].

Another recently described PFT comprises a tripartite ABC-type toxin complex[41], first discovered in *Photorhabdus luminescens* as Tc toxins[42,43]. Structural studies described the ABC-type toxins as exceptional α-PFTs that use a syringe like mechanism to penetrate the membrane[44]. This toxin family will be described in more detail in the next section since the Tc toxins are a major focus of this work.

2.2 Tc toxins

Toxin complex (Tc) toxins are large multi-subunit toxins that are present in a variety of bacterial pathogens. These high molecular weight protein complexes were first identified and best characterized in the insect pathogenic bacteria *Photorhabdus luminescens*[45]. The toxins attracted attention since they

showed high oral toxicity against various insects and thus, could be applied as potential insecticides in crop protection as an alternative to *Bacillus thuringiensis* toxins (Bt toxins)[46,47].

2.2.1 Genetic diversity of Tc toxins

2.2.1.1 Photorhabdus luminescens Tc toxins

Photorhabdus luminescens bacteria live in a mutualistic relationship as an endosymbiont in entomopathogenic nematodes from the family *Heterorhabditidae*[48]. The bacteria are located at the intestine of the nematodes, which infect insect larvae by penetrating the insect cuticle or invade through the insect mouth, anus or trachea[49]. Upon release of the bacteria into the insect hemocoel, the bacteria produce an array of virulence factors to evade the humoral and cellular immune response of the insect host and to effectively kill the host[50]. The cadaver as well as the bacteria itself serves as food source for the nematodes. After three rounds of reproduction, the nematodes, which harbor the bacteria in their gut, emerge from the cadaver and seek for a new host to attack[51-53].

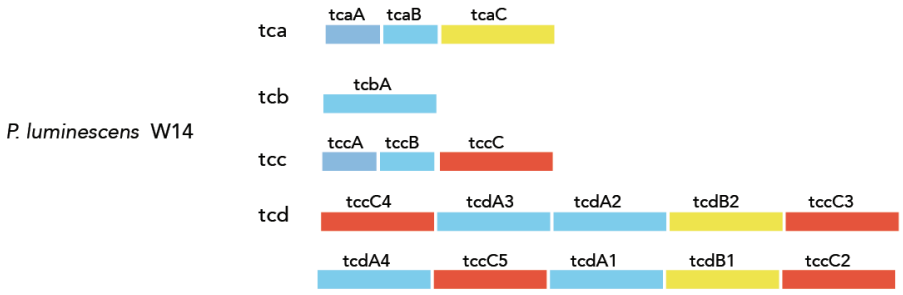


Figure 2: Genes coding for Tc toxins in *P. luminescens*. The four gene loci *tca, tcb, tcc and tcd* of *P. luminescens* W14 strain are represented with the different Tc toxin components highlighted in blue (TcA), yellow (TcB) and red (TcC).

Amongst these virulence factors are the large Tc toxins. Tc toxin genes are located on four gene loci in *P. luminescens* (*tca*, *tcb*, *tcc* and *tcd*)[45,54] (Figure 2). For simplification, the genes are classified in three groups based on the function of the encoded proteins TcA, TcB and TcC (Figure 2) [55]. The two gene loci *tca* and *tcd* contain multiple open reading frames, which results in several encoded toxin subcomplexes. Thus, *P. luminescens* strain W14 encodes seven TcA, three TcB and five TcC components leading to an arsenal of combinable toxin components[54,56]. Interestingly, the TcA subunit can be present as a single gene or is split on two genes in close proximity on the genome (Figure 2, indicated with two blue nuances).

2.2.1.2 *Tc toxins in other bacteria*

Despite *P. luminescens*, Tc toxin orthologues have been found in various insect pathogenic bacterial strains such as *Xenorhabdus nematophila*[57], *Yersinia entomophaga*[58] and *Serratia entomophila*[59] (Figure 3). *X. nematophila* follows a similar life cycle to the symbionts *P. luminescens* and *Heterorhabditidae* nematodes and lives in symbiosis with nematodes from the family *Steinernematidae*[60]. It was shown that *Xenorhabdus* genes *xptA1, xptA2, xptB1, xptC1* exhibit toxicity against different caterpillar pests such as *Pieris rapae* and *Heliothis virescens*[57]. Five tc-like genes were identified on the genome of the bacterium; *yenA1*, *yenA2* (TcA-like), *yenB* (TcB-like), *yenC1*, *yenC2* (TcC-like). Genes coding for chitinases (chi) are in proximity to the Tc toxin coding genes in entomopathogenic bacteria. Also in the *Y. entomophaga* genome, two chitinase encoding genes *chi1* and *chi2* are present and the chitinases are important for full toxicity of the Tc toxin against insects[58,61]. Therefore, the chi-domains might be important for the degradation of chitin in the peritrophic membrane of the midgut epithelial cells of the insect host[62]. Contrarily, the toxicity of *X. nematophila* against insects remained stable in absence of chi-genes[63] and thus, the function of the chi-genes could be variable between different species. In *S. entomophila*, only three tc-like genes have been discovered; *sepA*, *sepB* and *sepC*, which are associated with chronic Amber disease in the grass grub larva *Costelytra zealandica*[64]. Interestingly, these three genes are located on the large virulence plasmid pADAP (amber disease associated plasmid) and not on the bacterial chromosome[59].

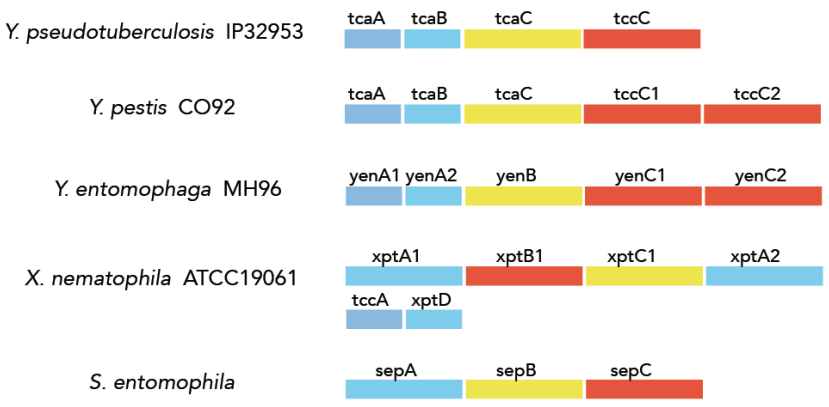


Figure 3: Genes encoding for Tc toxins in insect and human pathogenic bacteria. Single gene loci are shown for each bacterial strain. Gene classification into the three subunits is indicated by the following color code: TcA in blue, TcB in yellow and TcC in red. The TcA components split into two genes are highlighted by two blue nuances.

Additionally, human pathogenic bacteria such as *Yersinia pestis*, *Yersinia pseudotuberculosis* and *Morganella morganii* have become the focus of interest, since also in these bacteria, tc-like loci have been documented[65-67]. The opportunistic human pathogenic bacterium *M. morganii* was discovered in post-operative infections. Genome sequencing revealed nine genes that are orthologue to tc coding

genes in *Photorhabdus* spp, *Pseudomonas* spp and *Xenorhabuds* spp[67]. The enteropathogenic *Yersinia* species are a common cause of animal infections as well as affecting several mammalian and avian species[68-70]. *Y. pestis* emerged from the less virulent *Y. pseudotuberculosis* and thus the two *Yersinia* species share a common ancestor, which is also reflected by the conserved Tc toxin gene arrangement (*tcaA-tcaC* and *tccC*)[71]. Furthermore, in the *Yersinia* genome, the TcA components are split into two genes *tcaA* and *tcaB* or *yenA1* and *yenA2* and thus, no TcA component on a single gene is present (Figure 3).

Different bacteria contain either one or multiple genes encoding for Tc toxins. The presence of genes encoding for Tc toxins on plasmids enables a horizontal transfer of Tc toxins between different species. Thus, this results in the large diversity of Tc toxins present in a variety of bacteria. The Tc toxins are not only present in insects and human pathogenic bacteria, but they were moreover reported in plant pathogenic bacteria such as *Pseudomonas syringae* and *Pseudomonas fluorescens* as well as in the marine bacterium *Shewanella baltica*[72]. A genetic fusion product of TcB and TcC-like genes is present in the plant pathogenic fungi *Giberella*.[73]. Altogether, this suggest for an evolutionary development of Tc toxins along with the organisms adopting to new environments and for generating host specificity.

2.2.2 Biochemical studies with Tc toxins

The insecticidal toxicity of the Tc toxins was first observed in 1998 by recombinant expression of *P. luminescens* Tc toxin in *E. coli* and the subsequent oral exposure of the *E. coli* lysate to *Manduca sexta* larvae[45]. After ingestion of the lysate, the toxins destruct the midgut epithelium of the larvae completely. Also for many other Tc toxin secreting bacteria, the toxicity towards various insect species was shown. In the case of *X. nematophila*, Tc toxin components on the genes *xptA1*, *xptA2*, *xptB1*, and *xptC1* were expressed in *E. coli* and an altering combination of these toxin genes revealed different toxic activity towards *Pieris brassicae*, *Pieris rapae*, and *Heliothis virescens* larvae[57]. The toxin complex composed of XptA1, XptB1, and XptC1 expressed toxicity towards *P. rapae* and *P. brassicae*. Contrarily, by exchanging xptA1 to xptA2, the complex exhibited activity towards *H. virescens*. This led to the assumption, that the xptA1 and xptA2 genes (coding for the TcA component) are involved in the regulation of the host specificity[57]. In addition, in lipid bilayer experiments it was shown that XptA2 from *X. nematophila* and TcdA1 from *P. luminescens* were able to form ion conductive pores[74,75]. Therefore, TcA was thought to be responsible for binding to the host cell membrane and for translocation of either TcB, TcC or both components.

Furthermore, several studies indicate that Tc toxins only accomplish full toxicity as a tripartite complex including all three components, TcA, TcB and TcC. For the toxins from *Photorhabdus*, the toxicity of the single components showed reduced toxicity towards *M. sexta* larvae[46]. Full toxicity was only

recovered after adding the ABC complex[56]. A similar behavior was observed for the *Xenorhabdus* genes, of which the *xptA1* gene was necessary for expression of activity. However, the genes *xptB1* and *xptC1* were also needed for full activity[63]. Furthermore, the *sep* locus of *S. entomophila* contains only three genes that encode for Tc toxin components (Figure 3). Only upon recombinant expression of all three Tc toxin components of the genes in *E. coli*, the amber disease phenotype was caused in the New Zealand grass grub[59]. Thus, the Tc toxins are only biologically active as a complete tripartite complex[56,63,74]. Moreover, studies revealed that the *P. luminescens* TcC component is essential for high-level oral toxicity towards *M. sexta*[76], indicating that the TcC part of the toxin harbors the enzymatic activity. Indeed, Lang et al. could show that by transporting the TcC components into the cell cytosol by the protective antigen anthrax transporter machinery, the TcC component induces a high level of toxicity. Thus, TcC harbors the enzymatic activity. It interacts with TcB, which in turn acts as a linker between TcC and TcA[77]. The C-terminus of the TcC component of various organisms is not conserved and termed hyper variable region (hvr), being responsible for the toxic activity. Two TcC components of *P. luminescens* were identified as ADP-ribosyltransferases. While TccC3 ADP-ribosylates T148 of actin resulting in actin polymerization, whereas TccC5 targets two glutamine residues (Q61 and Q63) of Rho-GTPases. Altogether, both toxic enzymes lead to inhibition of phagocytosis and extensive actin polymerization and clustering[77]. The actual enzymatic function of the hypervariable region of TcCs from other bacteria has yet to be analyzed in detail. However, intoxication with Tc toxins from *Y. pseudotuberculosis*, *Y .pestis* and *Y. entomophaga* results in similar phenotypes such as membrane ruffling, actin clustering and nuclear fragmentation[66,72]. For the *Y. pseudotuberculosis* TcC components, a putative adenylate cyclase or phosphatase activity was suggested, while TcCs from *M. morganii* as well as *Y. entomophaga* are homologue to the cytotoxic necrotizing factor 1 (CNF1) from *E. coli*[78-80].

So far, the insect pathogenic toxicity of Tc toxins from entomopathogenic bacteria was demonstrated and thus, they have been defined as insecticidal toxins. Studies with Tc toxins from human pathogenic bacteria such as *Yersinia* spp showed that both, *Y. pestis* and *Y. pseudotuberculosis*, exhibit toxicity towards mammals[66]. *Y. pestis*, the causative agent of bubonic plague, uses the rat flea as vector for the infection of mammals, but the produced Tc toxins are not toxic towards the rat flea. In contrast, *Y. pseudotuberculosis*, which does not use the rat flea as vector, expresses toxicity towards the rat flea. However, tc-gene knockout strains of *Y. pseudotuberculosis* revealed no loss of flea toxicity compared to the wild type strain. Thus, the Tc toxins of *Y. pestis* as well as *Y. pseudotuberculosis* showed no toxicity towards rat fleas[66,81]. Furthermore, *Y. pseudotuberculosis* strain IP32953 exhibited no toxicity towards insect larvae of *M. sexta* in contrast to prior studies[82], neither any level of toxicity towards SF9 cells were recorded[66]. However, Tc toxins from both bacteria showed activity against cultured human cell lines. Interestingly, the toxic effect of Tc toxins from the two bacteria revealed different toxicity towards mouse fibroblasts (NIH3T3 cells) and the gut cell line Caco-2. Whereas *E. coli* lysate, containing

Y. pseudotuberculosis Tc toxins, had no remarkable effect in NIH3T3 cells, the exposure of lysate with *Y. pestis* Tc toxin led to actin condensation and nuclear fragmentation[66]. These results demonstrate the specific toxicity of Tc toxins and the adaption to distinct hosts during evolution.

The combination of biochemical studies with various bacteria producing Tc toxins revealed that the Tc toxins are fully toxic only as a tripartite complex. Furthermore, the TcA component is responsible for interaction with the host cell receptors or membrane components and for subsequent translocation of the toxic enzyme into the cell. The actual toxic enzyme is located in TcC, thus identifying the TcB component as a linker that conjugates TcA to TcC and completes the toxin. To fully understand the toxin composition, structural studies were performed over the past 12 years and are introduced in the next chapter.

2.2.3 Structural studies of Tc toxins

Until now, Tc toxins from three different bacteria have been structurally characterized in detail, namely XptA1 from *X. nematophila*, YenA1A2 from *Y. entomophaga* and TcdA1 from *P. luminescens*. Low resolution structures of XptA1[83] and YenA1A2[61], gained by negative stain electron microscopy, revealed different oligomerization states of the TcA components and provided first insights into the structural composition of Tc toxins. Higher resolved structures were determined for the *P. luminescens* Tc toxins using electron cryo microscopy. This includes a structure of TcdA1 (TcA) in the prepore[42,43] and in the pore state[44] as well as in complex with TccB2 and TccC3 (TcB and TcC)[84].

2.2.3.1 XptA1 and YenA1A2

Using analytical ultracentrifugation and dynamic light scattering, the oligomeric state of the XptA1 toxin was investigated and resulted in a complex with a size of 1.1 MDa[83]. Considering the molecular weight of the 287 kDa XptA1 monomer, a tetramer with a molecular weight of 1.15 MDa was assumed. The complex was further investigated by negative stain electron microscopy (Figure 4). A negative stain micrograph shows two main different views of the particle. First, the side views show a bell-like structure, while second, a four-fold symmetry can be seen in the top views (Figure 4A, white dashed line). Assuming a four-fold symmetry, the structure of XptA1 from *Xenorhabdus nematophila* was resolved to 23 Å and revealed a tetramer cage-like structure with a cavity inside. The complex has a length of 19 nm and a width of 16 nm. It was assumed that the cavity serves as an interaction site for the two binding partners XptB1 and XptC1, either after or before XptA1 interacts with the host cell membrane.

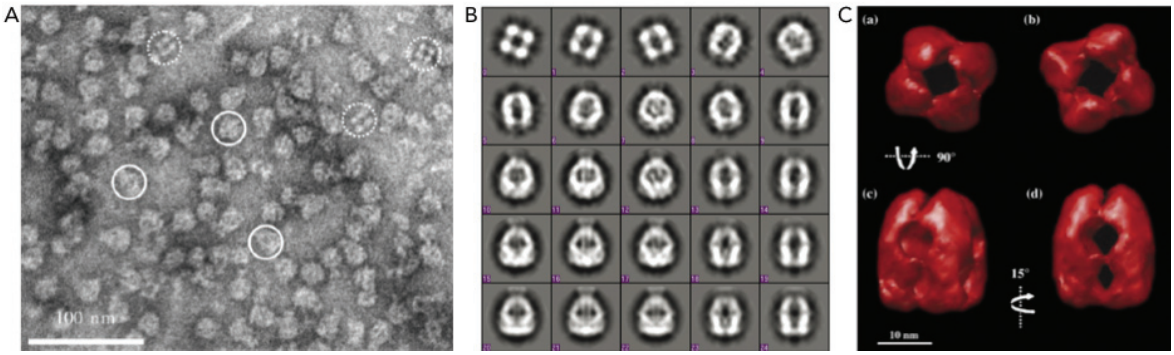


Figure 4: Negative stain analysis of XptA1 from *X. nematophila*. A) Negative stain electron micrograph. The labeled particles show side views (solid line) as well as top views (dashed line) of the toxin. B) 2D class averages of the XptA1 negative stain data set. C) Three-dimensional representation of XptA1 at 23 Å with a) top view, b) top view rotated by 15°, c) side view and d) side view rotated by 15°. The figure was modified from [83].

In contrast, the A component from the insect pathogenic bacteria *Yersinia entomophaga*, YenTC A1A2, was shown to be composed of five YenA1YenA2 components and thus, assembles as a pentamer[61]. The negative stain micrograph (Figure 5A) depicts side views of the toxin, that form a bell-like structure but the top views, in contrast to XptA1, show a five-fold symmetry. The presented structure shows a pentameric bell-shaped complex with an inner channel and an outer shell at 17 Å (Figure 5B). Furthermore, additional densities can be seen at the bottom of the toxin, possibly corresponding to the chitinase domains. The complex has a molecular weight of 2 MDa and is 26 nm long and 22 nm wide. It consists of five YenA1 and YenA2, five Chi1 and Chi2 components. However, in this structure the location of YenA1 and YenA2 could not be determined.

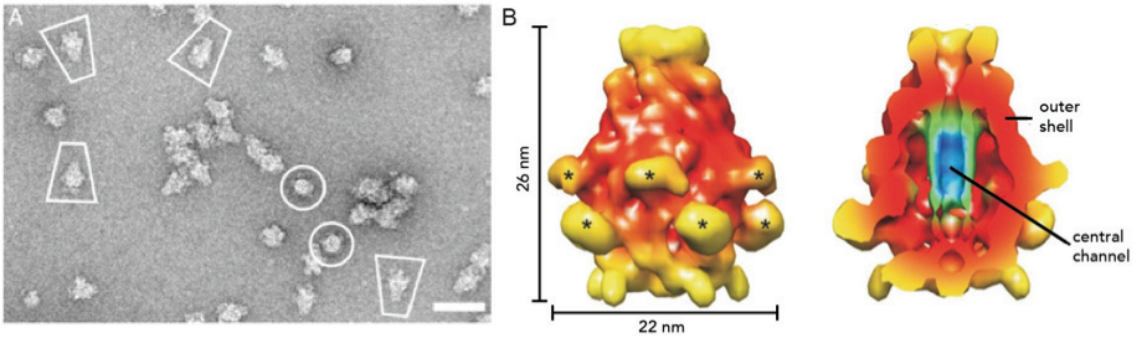


Figure 5: Structural analysis of YenA1A2 from *Y. entomophaga*. A) Negative stain electron micrograph with particles representing the top view highlighted with circles and particles showing the side view (rectangles). B) Electron density map of YenA1A2 with the complete surface (left) and as a cross section (right). The chi domains are indicated with a star. Figure adapted from [61].

2.2.3.2 *TcdA1, TccB2 and TccC3*

Structural studies using electron cryo microscopy led to higher resolved structures of *P. luminescens* TcdA1 and revealed also a pentameric assembly of the TcdA1 component[42,43]. The pentamer has a comparable mass of 1.4 MDa to YenA1YenA2 and is composed of a translocation channel consisting of ten α-helices and an outer shell. The shell can be further divided into an α-helical and a β-sheet part[43] (Figure 6A). Furthermore, the structure of the holotoxin, consisting of TcdA1 (TcA), TccB2 (TcB) and TccC3 (TcC), was solved with cryo-EM to a resolution of 3.9 Å[84]. The TcB and TcC components form a heterodimer with an overall structure similar to the one of the YenB-YenC2 complex from *Y. entomophaga*[85].

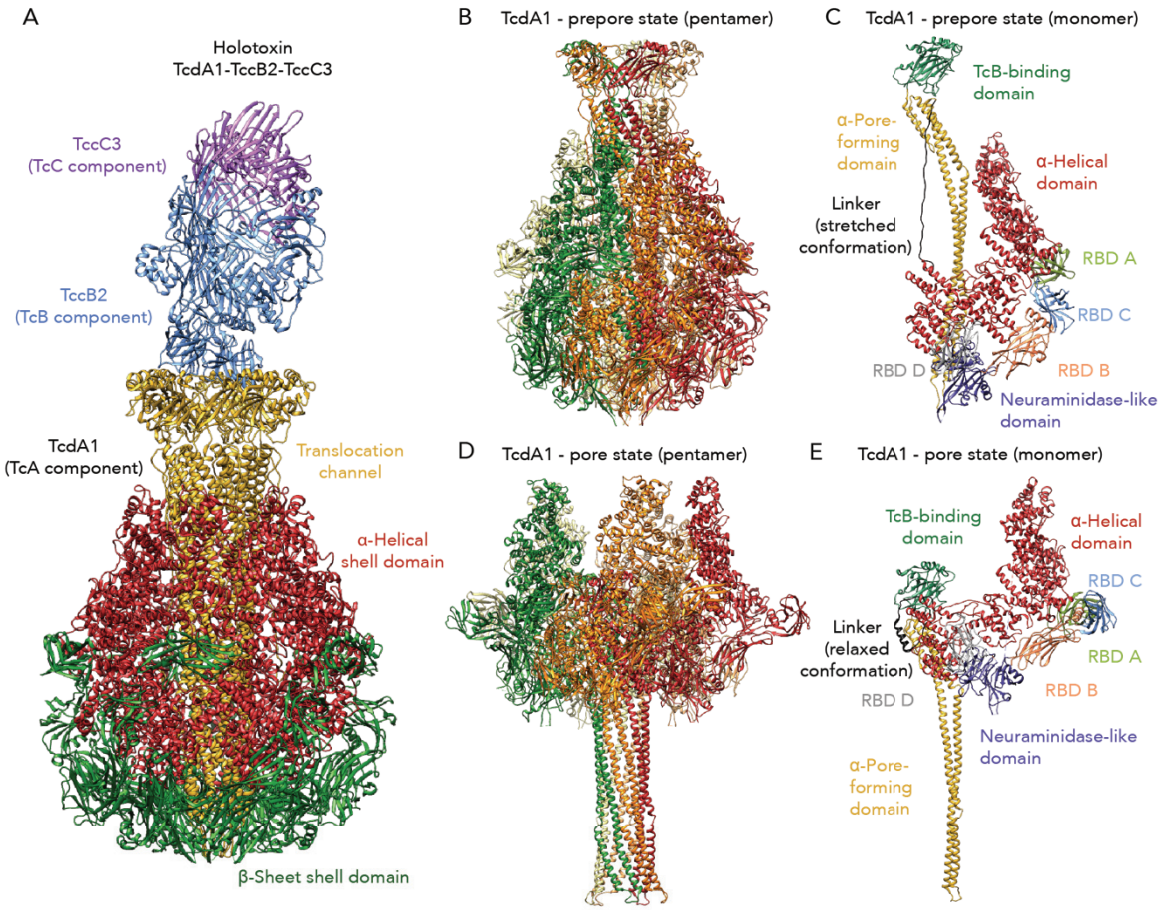


Figure 6: High-resolution structures of the *P. luminescens* Tc toxin. A) The complete holotoxin composed of TcdA1 and the TccB2TccC3 cocoon. TcdA1 is colored according to the three major subdomains: translocation channel (yellow), α-helical domain (red) and the β-sheet domain (green) (PDB-ID 6H6E). B) TcdA1 pentamer in the prepore state, in which each monomer is highlighted in different colors (PDB-ID 1VW1). C) TcdA1 monomer in the prepore state indicating the single domains of TcdA1. D) TcdA1 pentamer in the pore state, each monomer is highlighted in different colors (PDB-ID 5LKH and 5LKI). E) TcdA1 monomer in the pore state indicating the single domains of TcdA1.

This TcBTcC cocoon harbors the toxic enzyme, an ADP-ribosyltransferase in the case of TccC3. The C-terminus of TcC (hvr) gets cleaved by an aspartyl autoprotease in a catalytic dyad[43,85]. Then, the hvr is believed to be unfolded within the cocoon, the interior of which is hydrophobic with some charged patches and suitable for unfolded proteins. The TcBTcC cocoon is closed, but the β-propeller gate of TcB opens upon binding to the TcB-binding domain of TcA. This results in secretion of the ADP-ribosyltransferase into the TcA channel in a C- to N-terminus manner (Figure 7B)[84].

Four putative receptor-binding domains (RBD A-D) together with one neuraminidase-like domain form the β-sheet region of the envelope (Figure 6A, C and E). Due to similarity to the receptor-binding domains of anthrax and diphtheria, the β-sheet regions are thought to act as receptor-binding domains[12,43,86]. However, a receptor has not yet been identified for any Tc toxin. The fifth β-sheet domain, a neuraminidase-like domain, closes the shell at the bottom and acts as an electrostatic lock, which is closed at neutral pH. However, after binding to the cell membrane (Figure 7C), the holotoxin gets endocytosed and the low pH of the late endosome causes repulsive interactions, which lead to opening of the neuraminidase-like domain[43,44]. Consequently, the channel is exposed and inserts in a syringe-like manner into the membrane[42] (Figure 7D). This prepore-to-pore transition results in a massive conformational change, in which the entire shell is rearranged like a crown on top of the channel domain (Figure 6D). Furthermore, a linker domain connects the channel with the outer shell and acts as an entropic spring. This enables the membrane penetration and the prepore-to-pore transition[43,44]. In the prepore state, the linker is in a stretched conformation and is spanned between the last helix of the α-helical shell and the pore forming domain (Figure 6C). After membrane penetration and the conformational change, the linker is relaxed and folded[43,44] (Figure 6E). Upon membrane perforation, also other structural rearrangements occur, likewise in the channel domain, which result in an opening of the transmembrane part of the channel. Interestingly, the channel helices do not span the membrane completely. Molecular dynamic simulations revealed that the lower loop of the helices interact with the surrounding lipid head groups[44]. The opening of the channel then initiates the translocation of the unfolded hvr that is refolded in the cytosol of the cell (Figure 7E). In contrast to other PFTs[23,27], the pore of the Tc toxins is already formed before interacting with receptors or being in the proximity of membranes[42,43].

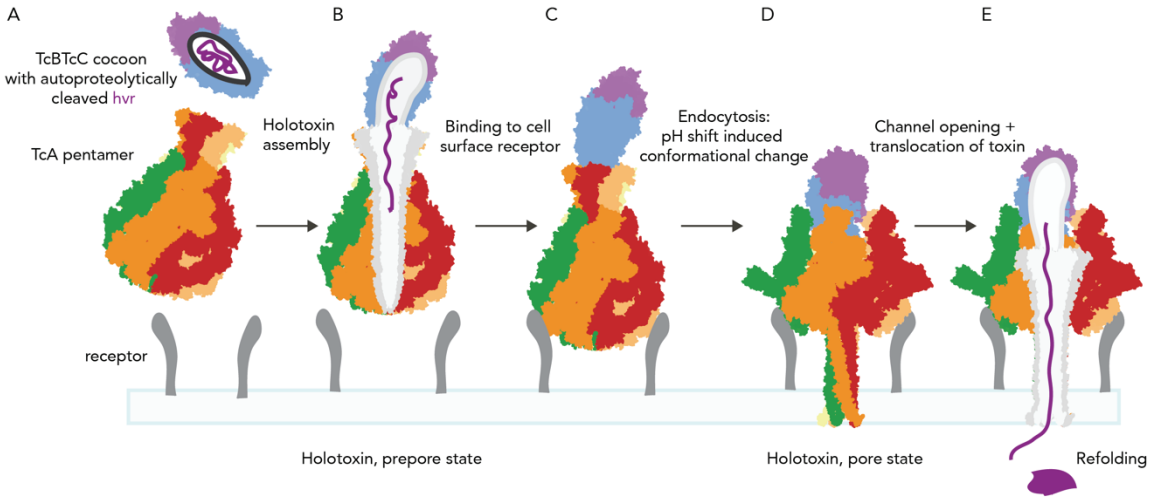


Figure 7: Mechanism of action of the *P. luminescens* Tc holotoxin. The mechanism of the Tc toxin is depicted schematically. A) The C-terminus of TcC (hvr) is cleaved after autoproteolysis and then locates unfolded in the TcBTcC cocoon interior. The TcA pentamer and the TcBTcC cocoon assemble in solution into a complex prior to receptor binding. B) Upon docking of TcBTcC to TcA, the β-propeller gate in TcB opens and the hvr emerges partially into the TcA channel. C) The holotoxin binds to cell surface receptors and upon endocytosis (D), a pH shift triggers the conformational change to the pore state. D) The shell of TcA opens and releases the channel that penetrates the membrane. E) The tip of the channel opens and initiates the translocation of the unfolded hvr that is refolded in the cytosol.

2.1 Structural analysis

The function of a protein is closely linked to its three-dimensional structural arrangement. A non-functional protein, which is mostly due to either misfolded or mutations, can lead to serious diseases. Knowledge about the protein's structure helps to understand and address the underlying mechanism of disease, thus making the field of structural biology very important. Macromolecular complexes are too small to detect with visible light but to analyze a protein structure, a resolution range of Å (10^{10} m) has to be reached[87]. Structural analysis includes various techniques to elucidate the structure of biological samples, for example X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and different electron microscopy methods such as transmission electron tomography and single particle electron microscopy. These techniques cover a broad resolution spectrum but there is no universal method to cover the complete spectrum. The methods not only differ in the resolution limit they can reach, but also the experimental information gained of the molecules' structure is different. For X-ray crystallography, an X-ray diffraction pattern is recorded, in NMR spectroscopy local magnetic fields around the atomic nuclei are observed, whereas in electron microscopy an image with the overall shape of a protein is produced. Therefore, considering the question that has to be answered upon structural investigation, the different methods imply distinct disadvantages and advantages. For more than 60 years, X-ray crystallography has been the method of choice for elucidating protein structures at high resolution[88]. Up to March 2019, 90 % of all deposited structures in the protein data bank were elucidated by X-ray crystallography (Figure 8A). Despite the common use of X-ray crystallography, the method has some drawbacks making it a method unsuitable for all kind of proteins. These includes filaments or large macromolecules, membrane proteins and proteins with flexible regions. Furthermore, not all crystals result in a sufficient diffraction for structure determination. There have been new methods circumventing the problem of obtaining large and well-diffracting crystals such as electron diffraction of micro crystals (MicroED)[89] where very small and many crystals are used for structure determination[90]. However, the crystallized form of the protein does not reflect the native conformation of the proteins. For obtaining structural information in a native environment, NMR is a method suitable for small molecules in solution and contributes 8 % of the structures deposited in the data bank (Figure 8A). Moreover, information about the protein dynamics in near native state and information about the local conformation and distance between atoms in near proximity can be obtained. Though, its limits are mostly size related, since the method is applicable for rather small proteins with a molecular weight below 30 kDa[91].

The method of electron microscopy has its roots already in 1931, where the first version of an electron microscope was reconstructed by Ernst Ruska and Max Knoll[92]. Since then, the electron microscopes have been evolved and the method has been applied in material science as well as on biological samples. It can overcome the disadvantages of X-ray crystallography and NMR to investigate large

macromolecular and flexible structures[93]. In addition, for X-ray crystallography challenging membrane proteins can be more easily addressed with electron microscopy, since the method allows a better handling of samples with detergents or membrane mimicking nanodicscs[94].

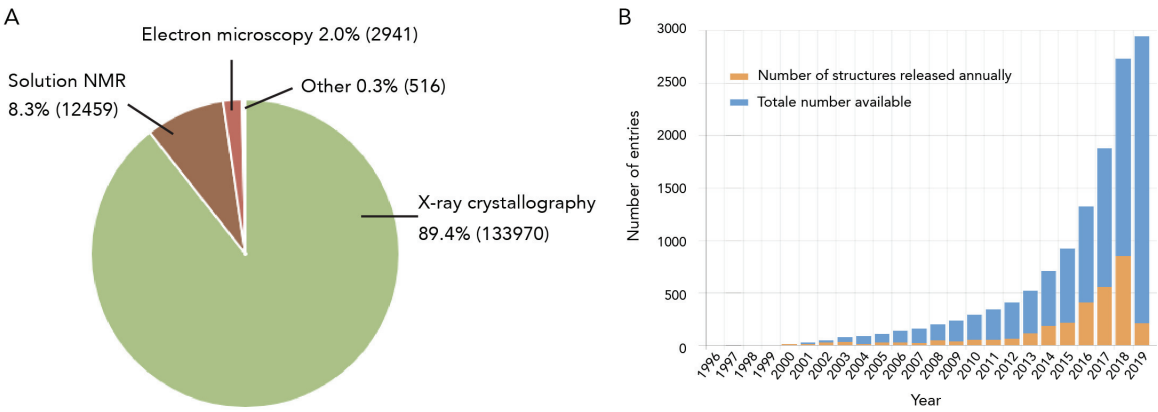


Figure 8: Statistics of protein deposition in the protein data bank (PDB). The statistics contain data that was uploaded until March 2019 in the PDB. A) Distribution of the three major methods used for structure determination with X-ray crystallography, NMR in solution, electron microscopy and other include combined methods. B) Graph showing the growth of the annual release of structures determined using 3D electron microscopy.

Structures solved with electron microscopy are with 2 % still underrepresented in the protein data bank (Figure 8A). However, the number of protein structures solved by EM increased drastically within the last eight years and is continuously growing (Figure 8B). Different types of electron microscopy can be applied for gaining structural information. With negative stain electron microscopy, low resolution structures can be obtained by covering the sample with heavy metal solution at room temperature[95]. The resolution limit is due to the grain size of the heavy metal solution surrounding the stained protein and in the same time the staining can result in distortion and dehydration artifacts[95]. Therefore, negative stain electron microscopy is mainly used to gain first structural overviews of the protein. However, to obtain a high-resolution structure of single proteins, the protein solution is plunge-frozen in liquid ethane (-170 °C), resulting in an amorphous layer of ice, which preserves the protein in its near native state while capturing multiple states[96]. Furthermore, electron tomography allows the analysis of larger assemblies such as cells or single organelles and thus, information about the cellular arrangement and localization of different cell compartments can be observed[97]. In addition, a great advantage of EM over X-ray crystallography is the use of small protein amounts and that less stable proteins can be directly prepared for data acquisition after sample purification. Therefore, electron cryo microscopy has become more important in the last years and the recent developments have made it a powerful tool for structural studies.

2.1.1 Electron microscopy and its resolution revolution

The principle of an electron microscope is generally similar to that of a light microscope. However, in electron microscopy an electron beam is generated instead of a light beam to allow for a resolution in the Å-range. As electron sources, thermal emitters (LaB_6 crystal) or field emission sources are most often used. In a LaB_6 crystal, the electrons are thermally emitted while an electric field is applied to a field emission source. In order to achieve higher resolution, field emission sources are used, which result in a very coherent electron beam. The emitted electrons are bundled and accelerated by applying a high voltage (200-300 kV) and the electron beam is guided by a magnetic pole lens system. This is composed of condensor and objective lenses that focus the electron beam whereas the projector lenses regulate the final magnification (Figure 9). Since biological samples are sensitive towards radiation damage, the microscope is operated at high vacuum and at cryogenic temperatures (in cryo-EM) to further enable the preservation of the sample and to reduce the effect of radiation damage[98]. Furthermore, a high vacuum prevents an electron scattering due to interaction of electrons with atmosphere gas.

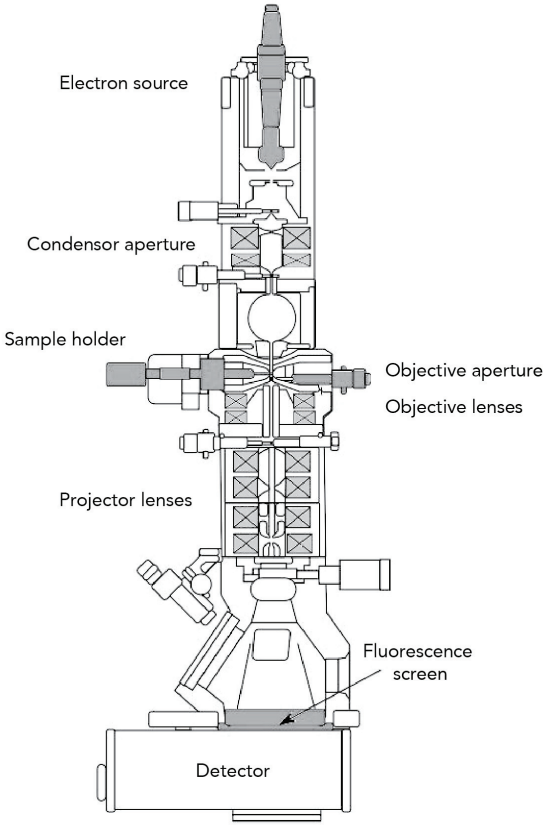


Figure 9: Schematic representation of the structure of a transmission electron microscope. The main functional elements of a TEM are outlined. The figure was adapted from[99].

In transmission electron microscopy (TEM), the image contrast derives from electrons arriving at the detector with different intensities. The interactions of the electron beam with the biological sample

occurs when passing through the sample, and basically, two types of scattering are distinguished: elastic and inelastic scattering of electrons, which in turn lead to different electron intensities and give rise to the image contrast in TEM. Furthermore, it can be distinguished between two types of contrast, namely the amplitude and phase contrast. The amplitude contrast derives from filtering or removing of inelastically scattered electrons that have a high scattering angle. Moreover, the inelastic scattering implicates an energy transfer of the electrons to the protein sample, which leads to a beam-induced sample damage and increases the noise in the image. In contrast, the elastically scattered electrons conserve the kinetic energy and have a small bending angle. Moreover, the elastically scattered electrons experience a phase shift upon interaction with the specimen. This leads to an interference of the elastically scattered electrons with the non-scattered electrons and results in a phase shift of the primary beam. Consequently, this phase shift leads to either amplification, reduction or a shift of the electron wave, similar as in sound waves. This results in the phase contrast, that is the major contributor to the overall image contrast in cryo-EM. Unfortunately, biological samples are weak phase objects with relatively poor scattering behavior, which in turn results in a low signal-to-noise ratio (SNR). To enhance the phase contrast, a defocus in the range of micrometers is applied, that results in an additional phase shift and better contrast of the image. The additional phase shift introduces a phase contrast which is described by the oscillating contrast transfer function (CTF). The CTF characterizes the transfer of the image information as a function of the defocus and spatial frequency. Subsequently, it is necessary to estimate CTF-parameters and to compensate the effects of the CTF with CTF correction to achieve high resolution. This includes the amplitude correction of the oscillating function, which decreases at higher frequencies[100]. Furthermore, a defocus series is applied during data acquisition to overcome the loss of information at zero crossings of the x-axis and to gain full information over all frequencies. Consequently, the achievable resolution depends on the microscope setting, the sample quality and furthermore also on the defocus series.

In the past, one limitation of cryo-EM was the molecular size of the sample to be examined. Icosahedral viruses were the first biological samples whose structures reached near-atomic resolution using cryo-EM[101]. However, since the contrast difference between the vitrified ice and the protein is not strong enough, proteins with a molecular weight below 100 kDa are very low in contrast. To increase the signal-to-noise ratio, a higher electron dose could be used, which in turn would result in radiation damage of the protein. Thus, the theoretical size limit for cryo-EM was calculated to 38 kDa based on signal-to-noise ratio and the radiation damage[102]. Cryo-EM is still a continuously developing and, compared to X-ray crystallography, young method. Thus, recent improvements in detectors, instrumentation and software allow for solving also low molecular weight structures. Over the past six years, a variety of nonviral assemblies with molecular weights below 1 MDa were determined at high-resolution. Few other examples are the 440 kDa sized anthrax protective antigen pore (PA) at 2.9 Å[103], the 300 kDa

TRPC4 cation channel at 3.6 Å[104] and the β-galactosidase (465 kDa) in complex with its cell-permeant inhibitor at 2.2 Å resolution[105]. Another tool to enhance the contrast is the Volta Phase Plate, that enabled recently a resolution of 3.2 Å for the low molecular weight protein haemoglobin (63 kDa)[106]. The phase plate is a thin carbon film, which increases the phase shift by interaction with and therefore modulation of the electron beam.

The employment of direct electron detectors (DED) was one of the two major innovations to overcome the limits of cryo-EM. With the improved sensitivity of DEDs, electrons can be detected directly and the read-out speed is enormously increased. This allows recording of movies instead of single images, in which each of the single frames is illuminated with only a small electron dose[107]. Thus, the signal-to-noise ratio is enhanced and a higher total electron dose can be applied. Frames, which experienced radiation damage can be down-weighted in the final image sum or they just can be removed for final processing preserving a high signal-to-noise ratio and high-resolution information[108,109]. Furthermore, movie recording revealed beam-induced specimen movement, that can now be corrected during image averaging[110]. The second innovation leading to the "resolution revolution" in cryo-EM is the development and improvement of image-processing software[111]. Software packages such as SPHIRE and RELION, including all necessary steps for data processing have been further developed, meaning improved algorithms but also implementation of a user interface[112,113].

Thus, electron cryo microscopy is currently one of the most important methods in structural biology. It compensates the disadvantages of X-ray crystallography and has moreover overcome the low-resolution as well as the molecular size limitations that have been limiting in recent years.

2.1.2 Single particle analysis

Despite the recent developments in microscope hardware and detectors, data processing strategies have evolved as well. Single particle analysis (SPA) is an approach that is used to retrieve the information from the acquisitioned images and to generate a three-dimensional model, and it is continuously developing since 30 years[93]. The main principle is to reconstitute a three-dimensional structure from thousand single molecules and averaging multiple obtained different views into one structure. The advantage of SPA together with cryo-EM is that thousands of the molecules are present in the sample and are orientated randomly in a thin layer of amorphous ice. During data acquisition, multiple images with each containing multiple single molecules (particles), that represent a two-dimensional projection of the atomic density of the protein, are collected[114]. Thus, in theory, every view of the molecule is present in the collected data images. By averaging particles that share the same orientation, the signal is enhanced due to a higher signal-to-noise ratio and differentiation between the various orientations become possible. Therefore, even heterogeneous samples or multiple states of the protein can be

analyzed with cryo-EM in combination with SPA. The challenge is to gain the three-dimensional information from the existing two-dimensional particles. For this, the projection parameters are needed to convert and project the 2D images in 3D space. The projection parameters include the three Euler angles (Ψ, θ, φ), that describe the rotation (Ψ), the out-of-plane translation (θ) and the orientation within the image plane (φ) as well as the x- and y-orientation[115]. With these five parameters in hand, the three-dimensional orientation of an object is known and the structure of a protein can be reconstructed.

The workflow of applied steps during SPA is illustrated schematically in Figure 10. After data acquisition and image preprocessing such as the frame alignment and CTF-estimation (chapter 2.1.1), the particles have to be selected and extracted from the images. Recently, the development of new generation of particle selection software, such as crYOLO[116], has evolved and enables the automated selection of particles from more than thousand images.

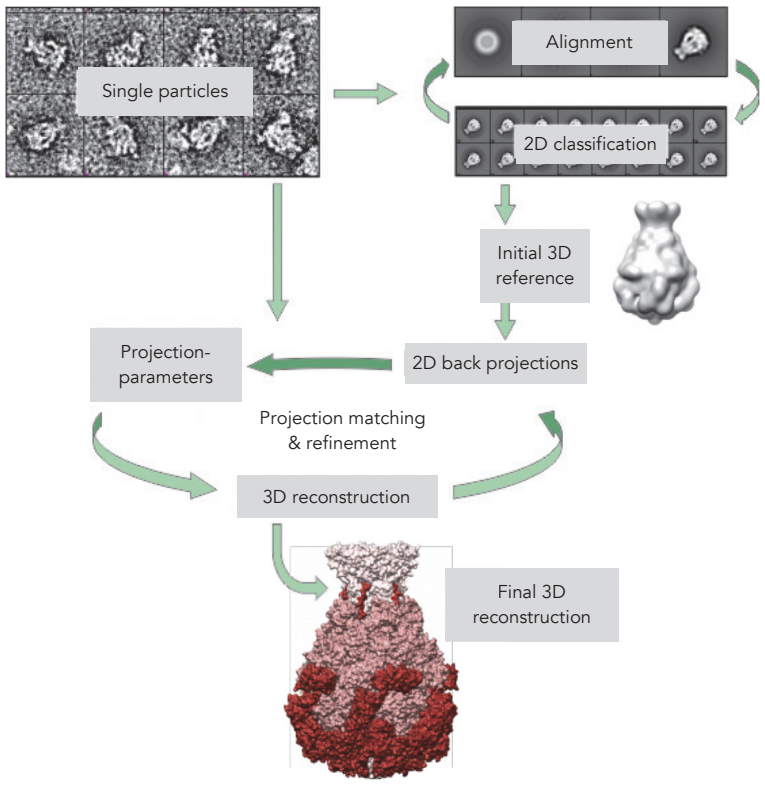


Figure 10: Scheme of the single particle analysis workflow. Single particles are 2D projections of the protein embedded in vitrified ice (upper left part). The particles are aligned and centered (x- and y-shift) during 2D classification. An initial reference can be obtained from the 2D class averages (upper right part). From this reference, two-dimensional back projections are generated, which are then used for projection matching to assign the projection angles (Ψ, θ, φ) to each single particle. With defined projection parameters, a 3D reconstruction of the protein can be obtained and further refined in an iterative process (lower part).

The selected particles are extracted from the images and the following processing steps are performed with single particles. In the classic workflow, back projection algorithms are used to determine the

orientation of the single particles[117]. The selected individual particles are first aligned and classified, with the displacement in the x- and y-direction as well as the orientation in the image plane being corrected by means of a reference. Particles belonging to one class are summed up to generate class averages with a better signal-to-noise ratio, which facilitates the differentiation of the particles (2D classification). The projection angles of the 2D particles are obtained by comparison the single particles with back projections of a reference structure in two-dimensional space (projection matching). In an iterative process, back projections are created at a predefined angle step size and compared with the particles. In case of a match, the projection parameters of the respective reference projection are assigned to the individual particles. Thus, the projection parameters are obtained for each particle and a 3D volume from all particles can be obtained (refinement). During the refinement, the particles are split into two equally sized halves and two reconstructions are calculated independently, which enhances the reliability and decreases the susceptibility of the reconstructed volume. The final resolution value is determined by comparing calculated Fourier shells of each half map (Fourier shell correlation). The frequency at a threshold of 0.143 or 0.5 correlation reflects the resolution[118,119]. Afterwards, the obtained reconstruction is again used as a reference in order to iteratively improve and refine the projection parameters to obtain a high-resolution 3D reconstruction.

2.2 Aim of this thesis

The family of Tc toxins is widespread and extends across a wide variety of bacteria, from entomopathogenic to plant and human pathogens. Due to their high toxicity and their host selectivity, the toxins are of particular interest as potential biopesticides to agriculture and crop protection. Moreover, with the presence of Tc toxins in human pathogens, the study of these toxins is also of great medical relevance. Especially the TcA component is an interesting target in the context of the host specific toxicity of the Tc toxins that has developed. As introduced before, the TcA component is responsible for the interaction with receptors or membranes and translocation of the toxic enzyme into the cytosol. However, until now, only TcAs from insect pathogenic bacteria have been structurally characterized and it remains unknown, if Tc toxins from different bacteria share a common architecture and mechanism of action. Furthermore, the low-resolution structures of YenA1A2 from *Y. entomophaga* and XptA1 from *X. nematophila* revealed different structural compositions. YenA1A2 exhibits a pentameric assembly with a channel inside[61], comparable to TcdA1 from *P. luminescens*[43,44], whereas for XptA1 a tetrameric composition with a cavity inside was proposed[83].

Therefore, the aim of this thesis was to structurally investigate TcAs from different bacteria in order to gain insight into the structural conservation of Tc toxins. Three different TcAs were chosen; XptA1 from *Xenorhabdus nematophila* (Xn-XptA1), TcdA4 from *Morganella morganii* (Mm-TcdA4) and TcaATcaB from *Yersinia pseudotuberculosis* (Yp-TcaATcaB). Xn-XptA1 is an interesting target, because a higher resolved structure of XptA1 is needed to get a better insight into the different structural organization as proposed by Lee et al.[83]. Additionally, two TcAs from human pathogenic bacteria were chosen for structural investigation to further broaden the horizon of the host specificity of these toxins. Additionally, Yp-TcaATcaB is composed of a heterodimer (TcaA and TcaB) and thus, the structure determination of this toxin component might result in answers regarding the assembly of a two TcA-component-containing toxin. With structures of human pathogenic bacterial Tc toxins, a comparison between them and the entomopathogenic toxins can be assigned to see differences that may describe the host specificity.

In order to obtain high-resolution structures, the proteins need to be recombinantly expressed in *E. coli*, purified and structurally characterized using electron cryo microscopy (cryo-EM). Cryo-EM was the method of choice, since the method evolved enormously and represents nowadays a fast approach for structure determination. In addition, the TcA components are large multi domain toxins with a size of above 1 MDa, which makes them a perfect sample for the investigation with cryo-EM.

3 Materials

3.1 Chemicals and Consumables

Chemicals and Consumables	Manufacturer	Location
Bromophenol blue, Sodium cholate	SERVA Electrophoresis GmbH	Heidelberg (GER)
Carbenicillin	AppliChem	Darmstadt (GER)
Colloidal solution	Sigma-Aldrich Chemie GmbH	München (GER)
DDM, OG, Triton X-100	Anatrace	Maumee (USA)
Glycerol, Calcium chloride, Sodium chloride, Tris, Tween-20,	Sigma-Aldrich	St. Louis (USA)
HCl	VWR International	Radnor (USA)
Kanamycin, Ampicillin, DTT	GERBU Biotechnik GmbH	Heidelberg (GER)
POPC, Brain Polar Lipids,	Avanti Polar Lipids	Alabaster (USA)
SDS, Ethanol, Acetic acid, Imidazole, Agarose, Tris, Caps, Citric acid	Carl Roth GmbH & Co. KG	Karlsruhe (GER)
Sodium hydroxide (2M solution)	Waldeck GmbH & Co. KG	Münster (GER)
TCEP	Biosynth	Staad (CH)
Uranyl formate	Polyscience Inc.	Warrington (USA)
C-flat 2/1	Protochips, Inc.	Morrisville (USA)
Copper grids G2400C	Plano GmbH	Wetzlar (GER)
Filterpaper Whatman No.4	GE Healthcare Bioscience	Pittsburgh (USA)
Kits Mini-, Midi-Prep	Qiagen GmbH	Hilden (GER)
Mini-PROTEAN TGX Stain-free	Bio-Rad Laboratories	Hercules (USA)
Nanodiscs (MSP1D1-ΔH5-His, MSP1D1-His, MSP2N2-His-His)	cube-biotech	Monheim (GER)
Parafilm	Bemis	Oshkosh (USA)
PureCube Ni-NT A Agarose	CUBE Biotech	Monheim (GER)
Quantifoil 2/1	Quantifoil Micro Tools	Jena (GER)
Reaction tubes	Eppendorf AG	Hamburg (GER)
Spectra Multicolor High Range Protein Ladder, PageRuler Plus Prestained Protein Ladder	ThermoFisher Scientific GmbH	Waltham (USA)
Superose 6 10/300 GL	GE Healthcare	Pittsburgh (USA)
Concentrator MWCO 100 kDa	Merck Millipore Ltd.	Tullagreen (Ir)

3.2 Instruments

Device	Manufacturer	Model	Location
Autoclave	SHP Steriltechnik	SANOclav	Detzel (GER)
Balance	Mettler Toledo	ME4002T	Pittsburgh (USA)
	Kern	ABJ-NM/ABS-N	Balingen (GER)
Bio-Layer Interferometer	Pall ForteBio LLC	Octet Red 384	Portsmouth (UK)
Centrifuges	Beckman Coulter	Avanti® J-20 XP	Krefeld (GER)
	Beckman Coulter	Optima® XPN-80	Krefeld (GER)
	Eppendorf	5810	Hamburg (GER)
Cryoplunger	FEI	Vitrobot™	Hillsboro (USA)
	Gatan	CP3	Pleasanton (USA)
Electron Microscopes	Jeol	JEM-1400	Tokyo (JPN)
	FEI	Tecnai™ G Spirit	Hillsboro (USA)
	FEI	Titan™ Krios	Hillsboro (USA)
FPLC Systems	GE Healthcare	ÄktaPurifier 100	Boston (USA)
	GE Healthcare	ÄktaPrime	Boston (USA)
High Vacuum Coating System	Bal-Tec	MED 020	Pfäffikon (CH)
Imaging Systems	Bio-Rad	ChemiDoc® MP	Munich (GER)
Micro-volume Spectrometer	DeNovix	DS-11	Wilmington (USA)
Mixer	Vortex-Genie 2	Scientific Industries, Inc.	Bohemia (USA)
nanoDSF	Nanotemper GmbH	NT.48	Munich (GER)
PCR Cycler	Bio-Rad	C1000 Touch	Munich (GER)
	Bio-Rad	MJ Mini	Munich (GER)
pH meter	Mettler Toledo	FE20 FiveEasy	Pittsburgh (USA)
Plasmacleaner	FEMPTO	Diener	Ebhausen (GER)
Power Pacs	Bio-Rad	PowerPac™ HC	Munich (GER)
SDS Gel System	Bio-Rad	Mini-PROTEAN® Tetra	Munich (GER)
Shaking Incubators	Eppendorf / New Brunswick	Excella E24	Hamburg (GER)
	Eppendorf / New Brunswick	Innova 44	Hamburg (GER)
Thermomixer	Eppendorf	ThermoStat Plus	Hamburg (GER)
Ultrasonicator	Branson	Sonifier 250 Analog Cell Disruptor	Geneve (CH)
Water bath	Julabo	TW 2	Seelbach (GER)

3.3 Media

Media		Composition
LB Media (1 L)	10 g	Bact. trypton
	5 g	Yeast extract
	10 g	NaCl
2 TY Media (1 L)	16 g	Bact. trypton
	10 g	Yeast extract
	5 g	NaCl
SOC Media (1 L)	20 g	Trypton
	5 g	Yeast extract
	0.58 g	NaCl
	0.19 g	KCl
	2.03 g	$MgCl_2$ x 6 H_2O
	2.46 g	$MgSO_4$ x 7 H_2O
	40 ml	50 % glucose

3.4 Bacterial standards

	Genotype	Origin
E. coli TOP10F'	F´ {proAB, $lacI^q$, lacZΔM15, Tn10 (Tet^R)} mcrA, Δ(mrr-hsdRMS-mcrBC), φ80lacZΔM15, ΔlacX74, deoR, recA1, λ⁻araD139, Δ(ara-leu)7697, galU, galK, rpsL(Str^R), endA1, nupG	Invitrogen
E. coli BL21(DE3) RIPL	F– ompT hsdS(r – m –) dcm+ Tetr gal λ(DE3) endA Hte [argU proL Camr] [argU ileY leuW Strep/Specr]	Novagen

3.5 Plasmids

Plasmid name	Features
pET19_6His_3C_XptA1_Xeno.nem.	C-terminal His-tag 3C precision protease cleavage side ampicillin resistance
pET28a_6His_3C_TcaATcaB_Yers.pseud. *pET28a_6His_3C_TcaATcaB_Yers.pseud. Δ622-714 +GRPSSG linker*	C-terminal His-tag 3C precision protease cleavage side kanamycin resistance
pET19_6His_3C_TcdA1 Photo. lum. E158A R1873A *pET19_6His_3C_TcdA1 Photo. lum. D965A R1971A* *pET19_6His_3C_TcdA1 Photo. lum. E1086A R1166A* *pET19_6His_3C_TcdA1 Photo. lum. E1086A* *pET19_6His_3C_TcdA1 Photo. lum. R1166A* *pET19_6His_3C_TcdA1 Photo. lum. H1202A* *pET19_6His_3C_TcdA1 Photo. lum. Y1205F Y1168F*	C-terminal His-tag 3C precision protease cleavage side ampicillin resistance

4 Methods

4.1 Protein biochemistry

Protein expression and purification was performed for the TcA components Xn-XptA1, Yp-TcaATcaB, Yp-TcaATcaB mutants and the different Pl-TcdA1 mutants. The different TcAs were expressed and purified according to the same protocol with some minor changes.

4.1.1 Transformation

The sequence *xptA1* (*X. nematophila*) was cloned in pET19 vectors (Novagen). The *Y. pseudotuberculosis* sequences *tcaa* and *tcab* were cloned together in pET28a (Novagen) with a two amino acid linker (GGATCC = Gly and Ser) in between the two fragments. All proteins were modified with an N-terminal hexahistidine tag (His_6-tag) for purification. The Yp-TcaATcaB mutant was based on the wild type construct pET28a and was expressed and purified in the same manner as Yp-TcaATcaB wild type. Pl-TcdA1 mutants were based on the sequence of *tcdA1* cloned in pET19 vectors (Novagen) and were provided by the DPF (Dortmund protein facility).

The plasmids were transformed into the *E. coli* BL21 (DE3) RIPL cells by heat shock method, in which the incorporation of the plasmid DNA into the *E. coli* cells is achieved by a sudden increase of the temperature. This leads to pore forming in the membrane and allows the plasmid to enter the cells[120]. Therefore, 100 ng of plasmid were added to 100 µL chemical competent cells and incubated on ice for 15 minutes. The heat shock was induced by transferring the cells in a water bath with 42 °C for 45 sec. This was followed by immediately placing the cells back on ice for 2 min before 800 µL SOC-medium were added. The cells were further incubated at 37 °C while shaking (700 rpm) for 45 min and then plated on LB-Agar plates with 125 ng/µL ampicillin or carbenicillin for Xn-XptA1 and the Pl-TcdA1 mutants and for Yp-TcaATcaB with 50 ng/µL kanamycin. The plates were incubated oN at 37 °C.

4.1.2 Expression

For expression of the different TcAs, a preculture of 100 ml was inoculated with a colony of the specific plasmid and incubated at 37°C in LB medium containing the respective antibiotic. After oN incubation, the 10 L expression culture was inoculated with the preculture to an OD_{600} of 0.06 and grown to an OD_{600} of 0.6-0.8 at 37 °C. Since the gene of interest was inserted into the lac operon system, the transcription of the inserted sequence was induced by adding the galactose analogue isopropyl-β-D-

thiogalactopyranoside (IPTG). 25 µM IPTG were added to the expression culture and proteins were expressed at 18 ˚C for 15 hours (Xn-XptA1) or 20 ˚C for 20 hours (Yp-TcaATcaB and Pl-TcdA1 mutants).

Cells were harvested by centrifugation at 3800 rpm and the cell pellets were suspended in lysis-buffer with protease inhibitor Pefabloc (200 µM) and lysed using a microfluidizer. Lysis buffers were composed of 25 mM Tris-HCl pH 8, 200 mM NaCl, 0.5 % Triton-X100 for Xn-XptA1, Yp-TcaATcaB and Pl-TdA1 mutants. Soluble proteins were separated from cell fragments by ultracentrifugation (38000 rpm, 30 min, 4˚C) and the supernatant was directly used for further purification steps.

4.1.3 Purification

Purification of the toxins was performed as a two-step purification. In the first step, the proteins of interest were separated from contaminants by an ion metal affinity chromatography since all toxins were tagged with an N-terminal His_6-tag. Following a dialysis for buffer exchange, the final purification step involved a size exclusion chromatography for removing remaining impurities.

4.1.3.1 Affinity chromatography

The proteins of interest were all modified with an N-terminal His_6-tag for nickel ion affinity chromatography purification. This method relies on reversible and high affinity binding of the tagged protein to the column matrix. The result is a stable complex, which ensures that the proteins adhere specifically to the column material. Thus, other proteins and macromolecules can be washed off the column. Elution of the protein of interest is accomplished by competitive displacement upon addition of a suitable small molecule in high concentration or by separation of the target protein from the affinity tag on the column.

In this study, the proteins were purified with a Ni-NTA-column, so that the His_6-tag binds reversible to the Ni^{2+}-ions that are coordinated by nitrilotriacetic acid. The toxins were eluted with the addition of imidazole, which has a higher affinity to the column matrix and displaces the His_6-tagged proteins.

After ultracentrifugation, the supernatant was collected and loaded onto a 5 ml HisTrap FF column (GE Healthcare) or incubated with 5 ml Ni-NTA-beads (GE Healthcare) (Pl-TcdA1 mutants). Extensive washing of the column was done with 10 column volumes (CV) washing buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 0.05 % Tween-20) with addition of 5 mM imidazole. Afterwards, the protein was eluted from the column by slowly increasing the imidazole gradient with 5 mM/min up to 200 mM and the elution of the protein was detected by UV_{280}. In contrast, the Pl-TcdA1 mutants were purified via selfpacked Ni-NTA columns. Thus, the washing as well as the elution was performed differently. The

beads were washed two times with 5 CV with increasing imidazole concentrations (from 5 mM up to 35 mM). The protein of interest was eluted stepwise with 100, 120 and 150 mM imidazole (2 CV each). Protein containing fractions were pooled and dialyzed against buffer with 20 mM Tris-HCl pH 8, 150 mM NaCl, 0.05 % Tween-20 oN with a 100 kDa molecular weight cut-off (MWCO) membrane.

4.1.3.2 Size exclusion chromatography

During size exclusion chromatography (SEC), the macromolecules are separated by their hydrodynamic radius, which reflects the size and shape[121]. The column material consists of polymer beads with different pore sizes that act as a molecular sieve. Therefore, the retention time depends on the diffusion volume of the molecules. Small molecules can enter the pores resulting in a longer column-crossing time whereas larger molecules pass by the pores and have a shorter retention time.

After the dialysis, the sample was concentrated using Amicon Ultra, 100 kDa MWCO concentrators to a column specific volume. For Xn-XptA1, Yp-TcaATcaB and Pl-TcdA1 mutants a Superose 6 10/300 column (GE Healthcare) was used. The column was equilibrated with 1.5 CV buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 0.05 % Tween-20) before the sample was loaded onto the column. The proteins were eluted with 1 CV buffer, protein-containing fractions were detected and further analyzed via SDS-PAGE.

4.1.4 SDS-PAGE

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is a method used for the evaluation of the molecular weight of macromolecules as well as their purity. Proteins are pretreated with sodium dodecyl sulfate (SDS) that interacts with the hydrophobic regions and destroys the tertiary structure of the proteins. Furthermore, all proteins are present in strongly negative charge after treatment with SDS which enables the migration of the proteins within the polyacrylamide gel after applying an electric field. Since the proteins are all negatively charged, the separation of the macromolecules occurs only depending on their size[122].

After every purification step, protein samples were taken and analyzed via SDS-Page. The samples were incubated at 95 °C for 5 min with 1x SDS loading buffer (250 mM Tris, 100 mM DTT, 6 % SDS, 40 % glycerin, 0.02 % bromophenol blue) and for every run, a protein standard marker was added (Spectra™ Multicolor High Range Protein Ladder, Thermo Fisher). The samples were loaded on a gradient gel (4-15%, Mini-PROTEAN TGX Stain-free) and run by applying constantly 50 mA. Visualization of the protein bands was achieved in a Bio-Rad stain-free imaging system. The stain-free technology uses a trihalo compound that reacts with tryptophan residues of the proteins and exhibit fluorescence after exposure to UV light[123].

4.1.5 Concentration determination via UV-Vis-spectroscopy

The concentrations of the protein solutions were determined by UV-Vis-spectroscopy. If the protein specific extinction coefficient is known, the protein concentration can be calculated via the Lambert Beer´sche Law at 280 nm[124]. The absorption at 280 nm derives from the aromatic amino acids like tyrosine and tryptophan. 1-1.5 µL of protein solution was applied to the DS-11 spectrophotometer (DeNovix) and the protein concentration was determined using the theoretical extinction coefficient of the respective proteins calculated based on the amino acid sequence with the online webserver ExPASy ProtParam[125].

4.1.6 Nano differential scanning fluorimetry

The protein stability can be influenced by many factors, for example the temperature as well as buffer conditions including different salt concentrations and pH values. Thermal denaturation is one method to determine the protein stability and therefore the melting temperature (T_m) of the protein. T_m is the temperature, at which 50 % of the protein is unfolded. Experimentally it is determined by measuring the intrinsic tryptophan fluorescence of proteins. The intrinsic fluorescence of the protein depends on the surrounding environment and exhibits an emission maximum at 330 nm for folded proteins and for unfolded proteins at 350 nm. Therefore, the quotient of the fluorescence intensity at 330 and 350 nm is used to determine T_m.

In this study, thermal denaturation by nano differential scanning fluorimetry (nanoDSF) was performed in a Prometheus NT48 (Nanotemper) to compare the different protein stabilities in the same buffer conditions. In addition, this method was used to determine the pH-range in which the protein Yp-TcaATcaB is most stable. For the pH stability test, a pH-range from 4 up to 11.5 was used with 20 mM citric acid for pH 4, 4.5 and 5, 20 mM Tris-HCl for pH 6, 7 and 8, 20 mM CHES for pH 9 and 10 and 20 mM Caps for pH 10.5, 11 and 11.5 in addition to 150 mM NaCl and 0.05 % Tween-20 in all buffers. The samples were soaked into nanoDSF Grade Standard Capillaries (Nanotemper GmbH, PR-C002) and the temperature was gradually increased by 1°C/min in a range of 20-90°C. The protein stability measurements at pH 8 were performed in 20 mM Tris-HCl, 150 mM NaCl and 0.05 % Tween-20 using 20 nM of each TcA. Equivalent to former experiments, a temperature gradient of 1°C/min in a temperature range of 20-90°C was applied.

4.1.7 Black lipid membrane experiments

The single channel conductivity was measured with black lipid membrane experiments in the lab of Prof. Roland Benz at the Jacobs University in Bremen. This method is a fast and efficient way to test whether the protein forms an ion conductive channel. The principle of this method is based on measuring the ion flux between two chambers that are separated by a lipid bilayer. If no protein is added, no current is measured because the ions are not able to cross the bilayer. On the contrary, if a pore forming protein is added to one side of the membrane, the protein is inserting into the membrane and as a result, a current can be measured[126].

The membranes were formed using 1 % solution of diphytanoyl phosphatidylcholine (Avanti Polar Lipids) in n-decane. The instrumentational setup consists of a Teflon chamber with two compartments that are connected by a small hole with a surface area of around 0.4 mm^2. The lipid solution is painted across the hole, which leads to formation of a membrane. After the membrane turned black, the toxin was added to the cis side (the black side) of the chamber. The membrane current was measured with a pair of Ag/AgCl electrodes with salt bridges. The electrodes were switched in series with a voltage source and a homemade current amplifier based on a Burr Brown operational amplifier[42]. All measurements were performed with a membrane potential of 20 mV. For the measurements with different pH values, buffer containing 1 M KCl and 10 mM citric acid pH 4, 10 mM MES pH 6 or 10 mM Caps pH 11 were used. The average single channel conductance was calculated from at least 70 pore insertion events. Measurements of Pl-TcdA1 were performed with 3.7 pmol, 19.4 pmol and 5.6 pmol for the experiments at pH 4, 6 and 11, respectively. 0.53 pmol, 1.78 pmol and 0.53 pmol of Pl-TcdA4 were used for the experiments at pH 4, 6 and 11, respectively. The experiments with Xn-XptA1 were performed with 0.28 pmol, 16.6 pmol, 0.4 pmol at pH 4, 6 and 11, respectively. With 3 pmol (pH 4), 28.1 pmol (pH 6) and 4 pmol (pH 11) the experiments were performed with Mm-TcdA4. For Yp-TcaATcaB 29.3 (pH 4) pmol, 3.6 pmol (pH 6) and 16.4 pmol (pH 11) were used. For the measurements with different $CaCl_2$ concentrations, the measurements were performed only at pH 6 and pH 11 with the respective buffers. Used $CaCl_2$ concentrations were 0, 1, 5, 10, 100 and 250 nM.

4.1.8 Biolayer interferometry

Affinities of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB to Pl-TcdB2-TccC3 were determined by biolayer interferometry (BLI) using an OctedRed 384 (forteBio, Pall Life Sciences) and Streptavidin (SA) Biosensors. BLI is a label-free optical analytical method for measuring the binding kinetics of molecular interactions. The principle of this method is to measure the interference pattern of white light, that is reflected from the immobilized protein on the biosensor chip. Therefore, a ligand is immobilized on

sensor chips and an analyte is in solution. Upon binding of the analyte to the immobilized ligand, the optical thickness at the tip of the biosensor chip is increased. This change in amount of protein bound to the biosensors results in an interference pattern, which can be measured in real time.

The measurements were performed by Dr. Daniel Roderer. Pl-TcdB2-TccC3 was biotinylated in 20 mM Hepes-NaOH pH7.3, 200 mM NaCl, 0.01% Tween20 (labeling buffer) with Sulfo-NHS-LC-Biotin (Thermo Scientific) in a 1:3 molar ratio for 2 h at room temperature, followed by 16 h at 4°C. Unreacted Biotin label was washed out using AmiconUltra 100 kDa cutoff concentrators by diluting the sample two times with a 10-fold volume of measurement buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.01% Tween20) and re-concentrating back to the original volume.

Biotinylated Pl-TcdB2-TccC3 was immobilized on SA biosensors at a concentration of 40 µg/ml, followed by quenching with 5 µg/ml Biotin. BLI sensograms were measured in three steps: baseline (300 s), association (20 s for Pl-TcdA1 and Xn-XptA1, 200 s for Mm-TcdA4 and 60 s for Yp-TcaATcaB, respectively), and dissociation (300 s for Yp-TcaATcaB and 200 s for the other TcAs). The sensograms were corrected for background association of the respective TcA on unloaded SA biosensors. On- and off-rates of binding were determined simultaneously by a global curve fit according to a 1:1 binding model. All BLI steps were performed in measurement buffer with 0.3 mg/ml BSA. The measurements were performed as triplicates.

4.1.9 Holotoxin formation

The complete functional Tc toxin consists of the TcA pentamer and the cocoon formed by TcB and TcC. For the interaction studies, the *P. luminescens* TcdB2-TccC3 (B2C3) cocoon was used. To analyze the interaction between the B2C3 cocoon and TcAs from the different bacterial organisms, a chimeric complex was formed its affinity measured. The complex formation was induced by mixing 1 µM B2C3 with 0.5 µM of the different TcAs (pentamer concentration) and incubation for 1 hour at 4° C. Unbound B2C3 was removed by size exclusion chromatography using a Superose 6 increase 5/150 column (GE Healthcare) as described previously (chapter 4.1.3.2). The retention fraction corresponding to the holotoxin with a total protein concentration of about 20 nM was visualized by negative stain electron microscopy (chapter 4.2.1).

4.1.10 Intoxication assay

The intoxication assay with the chimeric holotoxins was performed together with Dr. Daniel Roderer. HEK293T cells (Thermo Fisher) were intoxicated with chimeric holotoxins formed by the different TcAs (Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB) and Pl-TcdB2-TccC3. 2 × 10^4 cells in 400 µL DMEM/F12 medium (Pan Biotech) were grown adherently overnight and subsequently 0.5 and 5 nM of chimeric holotoxin was added. Incubation was allowed to continue for 16 h at 37 °C before imaging.

4.1.11 Reconstitution of TcAs into nanodiscs at different pH values

To investigate, whether the different TcAs interact with membranes or a membrane like environment as it was previously shown for Pl-TcdA1[44], the proteins were incubated with nanodiscs. Nanodiscs are artificial small membrane patches composed of lipids, that are surrounded by a scaffold protein, the MSP (membrane scaffold protein). The transmembrane region of the protein of interest is incorporated into the nanodiscs. The diameter of the nanodiscs can be adjusted by varying the size of the MSPs. The MSP is composed of 10 helices and by deleting helices or by adding helices the resulting diameter is varied. The MSP with the smallest diameter used in this study was the MSP1D1-ΔH5-His (8 nm)[127] and the one with the largest diameter the MSP2N2-His (16 nm)[128]. Two different type of lipids were used in these experiments; 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholin (POPC) and brain polar lipids (BPL). There are two different approaches for nanodiscs experiments. On the one hand, the nanodiscs can be assembled during incubation with the protein of interest. In contrast, the nanodiscs can be preassembled prior to protein addition. Both experimental setups were performed. The prepore-to-pore transition of Pl-TcdA1 wild type and the mutants was induced by pH-shift from pH 8 to pH 11.2 as it was reported previously[43,44]. For Xn-XptA1, Yp-TcaATcaB and Mm-TcdA4, different pH values were tested in order to evaluate the yet unknown prepore-to-pore transition for these toxins.

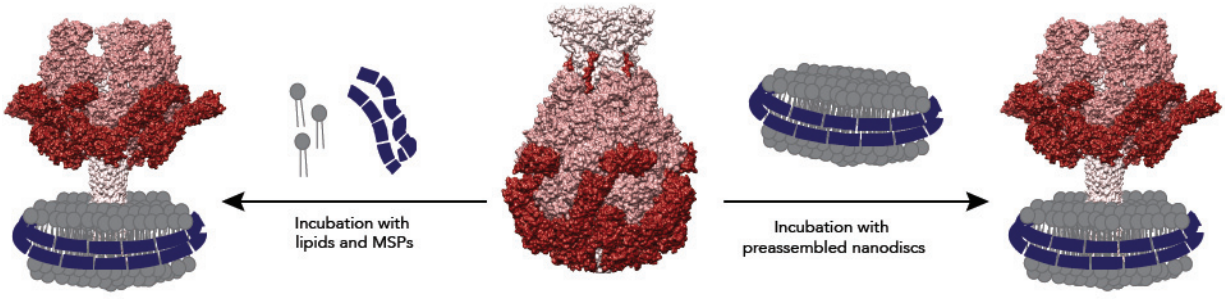


Figure 11: Scheme of the experimental set up of the nanodiscs experiments. The nanodiscs can be formed either during protein addition (left side) in presence of lipids and membrane scaffold proteins or as preassembled nanodiscs (right side) prior to protein addition.

In all experimental set ups, the samples were dialyzed oN or up to 3 days at 4 °C and a dialysis membrane with 12 kDa MWCO was used. Additionally, to nanodiscs experiments, some experiments were performed using liposomes composed of BPL. Here, the liposomes were formed prior to protein addition and the pH shift. Furthermore, for Xn-XptA1, it was tested whether the pore state is preferably formed in the presence of detergent. Thus, the sample was incubated at pH 11 in detergent containing buffer and was afterwards exchanged to nanodiscs. Two detergents were used, n-Octyl-β-D-glucopyranoside (OG) and n-Dodecyl-β-D-maltopyranoside (DDM). Both are non-ionic and therefore mild glucoside-based detergents where OG has one sugar moiety less compared to DDM. The sample quality as well as the incorporation of toxin into nanodiscs was checked with negative stain electron microscopy as described in the next section.

Table 1: Experimental setups for the different nanodiscs experiments.

TcA	MSP	Lipid	ratio toxin : MSP : Lipid or toxin : nanodiscs	pH, salt
Xn-XptA1	MSP1E3D1-His	POPC BPL	1 : 12 : 340 1 : 10 : 300	11 11, $CaCl_2$
	MSP1D1-ΔH5-His	POPC	Preassembled, 1:8 or 1:3	11
	MSP2N2-His	POPC	Preassembled, 1:8	5, 11, 11.2
		BPL		
Yp-TcaATcaB	MSP1D1-ΔH5-His	POPC	Preassembled, 1:4, 1:6, 1:10	10.5, 10.8
	MSP2N2-His	POPC	Preassembled, 1:10	4.5, 5
	MSP1E3D1-His	POPC	1 : 8 : 80	10.5, 11
Mm-TcdA4	MSP1D1-ΔH5-His	POPC	Preassembled, 1:10	10.5, 11, $CaCl_2$
	MSP2N2-His-His	POPC	Preassembled, 1:10	11, $CaCl_2$
	-	BPL	1 : 4	5
Pl-TcdA1 wild type, E158A-R1873A, D965A-R1971A, E1086A-R1166A, Y1168F-Y1205F	MSP1D1-ΔH5-His	POPC	Preassembled, 1:6	11.2
Pl-TcdA1-H1202A	MSP2N2-His	POPC	Preassembled, 1:6	11.2

4.2 Structural biology

The aim of this study was to characterize and compare the structures of different TcAs. Therefore, an atomic model of each TcA was required. As mentioned above, the method of choice for structure determination was electron microscopy. To check the sample quality and to get first insights into the protein structure and domain organization, negative stain electron microscopy was performed. Since this method is limited in resolution, cryo-EM, which includes a different sample preparation and instrumental setup, was necessary for reaching high resolution.

4.2.1 Negative stain electron microscopy

4.2.1.1 Sample preparation

For the sample preparation in negative stain electron microscopy (negative stain EM), small copper nets (grids) are used as specimen carrier. The copper grids are covered with a carbon film for protein absorption. Due to the net structure, the grid is split into hundreds of squares (Figure 12A and B). These grids can either be purchased commercially or can be self-made. For the latter, the plane copper grids (Plano GmbH, G2400C) are first covered with a collodion solution to generate a thin plastic film. The thin film is generated by applying 3-4 droplets of collodion solution on a water bath to create a continuous layer and the grids are placed onto the collodion film. Afterwards, the grids are dried and afterwards placed in a carbon coater (Leica EM ACE600, Bal-Tec MED020). There, carbon is sputtered from a carbon thread onto the grid until a layer of 7-10 nm carbon is reached. Prior to sample preparation, the grid surface is cleaned and turned hydrophilic with a plasma-cleaner[96,129] (FEMTO, Diener electronics GmbH & Co. KG). The proteins lying on the carbon film are fixed with a heavy metal solution that embeds the protein and enhances the contrast (Figure 12). There are different staining solutions varying in the grain size of the heavy metal, the pH and also the stability[130]. Most commonly used is uranyl formate or acetate as staining solution. In this work, a freshly prepared 0.75 % uranyl formate solution was used for staining. 37.5 mg uranyl formate was dissolved in boiling water for 5 min while stirring. After addition of 5 µL 5 M NaOH, the solution was stirred for another 5 min. Finally, the uranyl formate solution was filtered with a 0.22 µm syringe filter for separating the solution from dust or undissolved uranyl formate and was stored in a darkened tube.

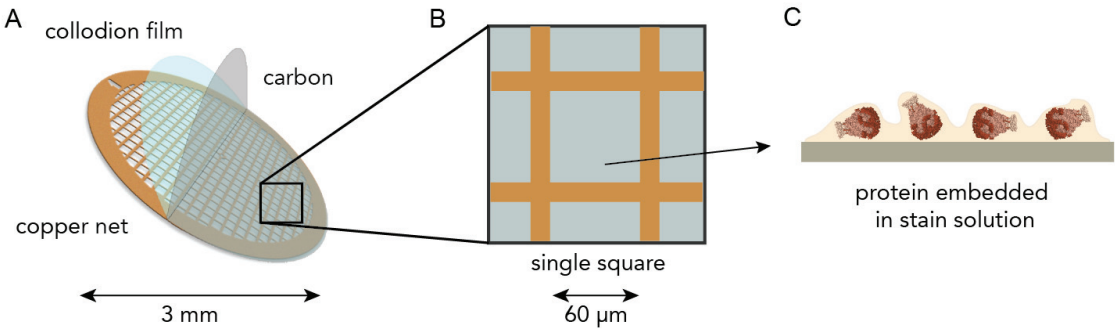


Figure 12: Scheme of a typical negative stain EM grid. A) The copper grid is covered with a collodion film before the carbon layer is applied. The copper net structure of the grid is shown in orange and the single squares of 60 µm lie within the net structure B). C) The protein is adsorbed on the carbon surface.

The optimal protein concentration for negative stain EM varies massively between different shaped proteins as well as their size. For the Tc toxins, a concentration of 0.02-0.05 mg/ml resulted in nicely distributed proteins on the grid. The grids were prepared by applying 4 µL of the protein solution on a freshly glow-discharged copper grid. The protein solution was incubated on the grid for 40 s before blotting the sample with filter paper (Whatman No 4, GE Healthcare). Afterwards, the grid was washed with 10 µL of uranyl fomate solution and stained with applying additional 10 µL uranyl formate and incubated for 40 s before blotting. The grid was air dried prior to insertion into the microscope.

Some particles adopt a preferred orientation on carbon coated grids. This was especially the case for Yp-TcaATcaB. To induce different orientation of the particle, the grid can be pretreated with poly-L-lysine (PLL) solution to generate a more hydrophilic surface on the grid. For this purpose, 1 µL of concentrated PLL solution was incubated for 40 s on the glow discharged grid before removing excess of liquid and drying the grid. Afterwards, the sample was applied as described above.

4.2.1.2 Data collection

Two different electron microscopes were used for negative stain EM; either a Tecnai G Spirit electron microscope (FEI) or a JEOL-JEM 1400 both operated at 120 kV and equipped with a CMOS TemCam F416 detector (TVIPS). For the different Tc toxins, negative stain EM was mainly used for checking the sample homogeneity, the concentration as well as the protein conformation in case of the prepore-to-pore transition. For this purpose, micrographs were recorded with an exposure time of 1 s at a magnification of 40k resulting in a pixel size of 2.6 Å/px. The data set of Yp-TcaATcaB was recorded on the JEOL-JEM 1400 with a defocus range of 1.5 – 3.0 µm and contained 147 images.

4.2.1.3 Data processing of Yp-TcaATcaB negative stain data

The Yp-TcaATcaB data set was processed using the single particle approach with the software package SPHIRE(*sparx for high resolution electron microscopy*)[112]. First, the single particles have to be selected. This can be done automatically, semi-automatically or manually. For this data set, particles were selected manually in EMAN2 boxer[131]. 5410 particles with a box size of 200 were selected and extracted from the micrographs (particle extraction). With these particles, a 2D classification was performed using ISAC (iterative stable alignment and clustering)[132]. By this, the particles can be sorted and bad or wrongly picked particles can be discarded and the signal to noise ratio is increased. An initial 3D volume was generated with the best class averages using RVIPER. The single steps of data processing will be explained in more detail in section

4.2.2 Electron cryo microscopy (cryo-EM)

4.2.2.1 Sample preparation

The grids used for cryo-EM differ in composition as well as sample fixation on the grid compared to the negative stain grids as well as the sample fixation on the grid. For cryo-EM, the sample is plunge frozen in aqueous solution. In detail, the protein is applied onto the grid and plunge frozen in liquid ethane, which is cooled down to -180 °C by liquid nitrogen. This rapid freezing step enables embedding of the protein in vitreous ice without ice crystal formation, which would harm the protein structure. Furthermore, the native state of the protein is conserved[96]. There are different plunge freezing devices available, for example the Vitrobot (FEI, Leica EM GP) or the Cp3 (Gatan Cryoplunger).

The cryo-EM grids are based on a copper grid covered with a thin carbon film that has holes (holey carbon) in which the protein embedded in the layer of amorphous ice. The number and size of the holes can vary between different grid types. In addition, there are grids available with an additional thin layer of carbon (2 nm continuous carbon). With these grid types, one can overcome issues of low protein concentration or with samples where the protein tends to stick to the carbon edges instead of lying in the pure ice layer. Since the protein concentration, shape and size, the buffer condition as well as the blotting time can influence the cryo-EM preparation, an intensive screening is necessary to find optimal conditions for cryo-EM.

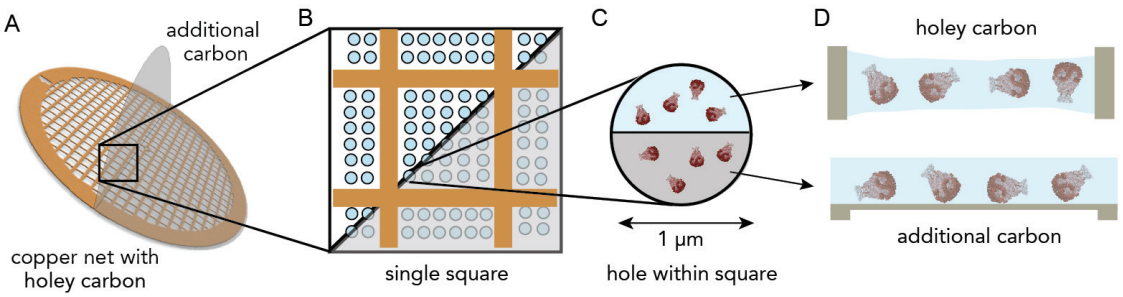


Figure 13: Scheme of two different cryo-EM grid types. A) The copper nets are covered with a thin film of holey carbon and depending on the grid type, with an additional layer of carbon. B-C) Within the squares, the structure of the holey carbon is visible. D) The protein is embedded in ice within the holes (holey carbon, upper panel) or on top of the additional layer of carbon (lower panel).

Prior to cryo-fixation, the plunging device has to be conditioned by cooling down the ethane container with liquid nitrogen and setting the humidity to a minimum of 90%. For Xn-XptA1 and Mm-TcdA4, grids with an additional layer of carbon were used (Quantifoil QF 2/1 with 2 nm additional carbon / Micro Tools GmbH), whereas the Yp-TcaATcaB protein was applied to a grid without an additional carbon layer (Quantifoil QF 2/1). The grids were glow-discharged as described in chapter 4.2.1.1 and fixed in a forceps that is placed into the cryoplunger (Cp3). For the grids with additional carbon, the sample was incubated on the grid for 40-60 s to enhance the absorption of the protein on the carbon surface and therefore the protein concentration before blotting automatically from both sides. If the protein concentration was too low, especially for the grids without any additional carbon layer, multiple sample applications were performed. To saturate the carbon, the sample was incubated on the grid, blotted manually from the back (Whatman No. 4 filter paper) and after the final sample application blotted automatically. Several conditions for cryo-EM were manually screened at the JEOL-JEM 1400 microscope with a cryo holder (Gatan 626 cryoholder). Image acquisition was performed as described in the section 4.2.1.2.

4.2.2.2 Data acquisition

The data sets for the cryo samples were collected on a TITAN KRIOS transmission electron microscope (FEI) operated at 300 kV. The microscope was equipped with an extra high brightness field emission gun (XFEG) as electron source and a spherical-aberration corrector (Cs corrector). The images of Xn-XptA1 and Mm-TcdA4 were recorded on a Falcon II direct electron detector (FEI) and for Yp-TcaATcaB the successor Falcon III direct electron detector was used.

After insertion, the quality of the grids was checked by making a grid atlas that shows an overview of the complete grid. Based on the grid atlas, squares with good ice distribution and without cracks were selected for data acquisition. Afterwards, the automated data acquisition was set up within the

automated data collection software EPU (FEI) and four images were taken per hole. The detailed information for each data acquisition is listed below (Table 2).

Table 2: Parameters for cryo-EM data collection for the different proteins.

	Xn-XptA1	Mm-TcdA4	Yp-TcaATcaB
Microscope	FEI TITAN KRIOS (Cs corrected, XFEG)		
Voltage (kV)	300		
Camera	FEI Falcon II		FEI Falcon III
Magnification (nominal)	59,000	59,000	96,000
Pixel size (Å)	1.14	1.14	0.68
Number of frames	25	25	59
Total electron dose ($e^-/Å^2$)	60	60	130
Exposure time (s)	1.5	1.5	1.5
Defocus range (μm)	0.7 - 3.0	0.9 - 2.4	1.0 - 2.2

4.2.2.3 Data pre-processing

After data collection, all images from each data set were inspected manually and micrographs with high drift, strong ice contaminations (e.g. "leopard skin" or ice crystals) or destroyed carbon were discarded. Data processing was performed with the software package SPHIRE[112] within SPARX/EMAN2[133] for all proteins.

Motion correction

With the direct electron detector, the data collection has become much more efficient and faster. Nowadays it is possible to record movies with several frames instead of single images[134], which enables the correction of sample movement or drift during frame alignment and enhances the signal-to-noise ratio[108,135]. There are different programs for motion correction and frame alignment available[109,136]. For all recorded data sets, motionCor2 was used to generate unweighted full-dose images as well as dose weighted images with full-dose. The principle of this method is to estimate the image shift between the single frames in order to calculate a least-squares estimate of relative motions between neighboring frames[137]. This can be done either by global correction of whole-frame image motions or by the patch alignment mode. During patch alignment mode, the micrograph is split into patches and the motions are iteratively estimated and determined for each patch and then corrected.

CTF estimation

The estimation of the contrast transfer function (CTF) is an essential step during data processing and was done prior to any processing. Based on a successful CTF determination, the contrast of the image can be improved. In addition, the quality of the images and therefore the content of high-resolution information of each image can be evaluated with the CTF information. Within the CTF estimation, a calculated theoretical CTF (Figure 14A, pink curve) is adopted to fit the experimental CTF (Figure 14A, black curve). The variable to be determined during CTF estimation is the defocus, since all other parameters are characteristic of the particular microscope and sample preparation. Fourier transform of the image results in a power spectrum with the respective Thon rings (Figure 14B), visualizing the oscillation of the CTF[138]. No information is present at the respective frequency of a zero-crossing. Therefore, to overcome the loss of information, different defocus values are used during data acquisition, because with the defocus the position of the zero crossing changes. With the gained information one can perform a CTF correction during data processing. Without a CTF correction of the image, its resolution limit is on the location of the first zero crossing[117]. For this reason, it is necessary to correct for the CTF for gaining a high-resolution reconstruction.

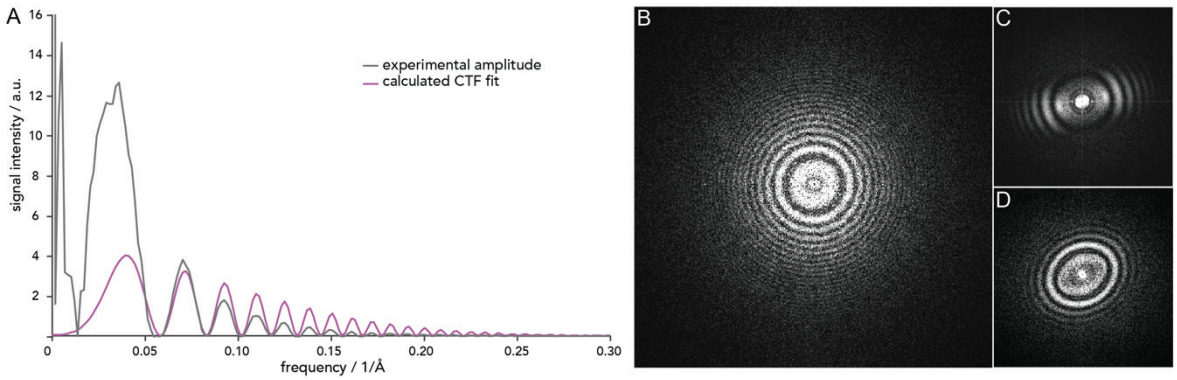


Figure 14: CTF estimation and Fourier-transform of real space images. A) Example of the experimental amplitudes in grey and the corresponding calculated CTF fit in magenta. B) Power spectrum (Fourier-transform of the CTF) of a micrograph visualizing the Thon rings. C) Example of a micrograph with high drift leading to a cutting of Thon rings (frequencies). D) An elliptic power spectrum as a result of high astigmatism during image collection.

The CTF estimation was calculated with CTER[139] on the unweighted full-dose motion corrected images. With the CTF assessment and the Drift assessment tool in SPHIRE, micrographs with high defocus > 5µm or high drift were discarded. Micrographs with high drift have a cut off in the power spectra (Figure 14C) and highly astigmatic micrographs have an elliptical extending power spectrum (Figure 14D) leading to a loss of information. After sorting the micrographs according to these criteria, images containing most high-resolution information were further processed.

4.2.2.4 *Single Particle Analysis*

In order to generate a 3D reconstruction of the respective protein, the single particle analysis approach was applied. The principle of the single particle analysis approach for EM data processing was introduced in chapter 2.1.2. For diverse proteins, a different workflow with multiple steps of processing is necessary. A general workflow is depicted in Figure 15. The order of the different processing steps was varying between the projects and not all steps were performed for all proteins. For example, detailed optimization steps like a focused 3D classification or local refinement were performed individually. The use of the single steps of SPA will be explained in detail in this section for each TcA.

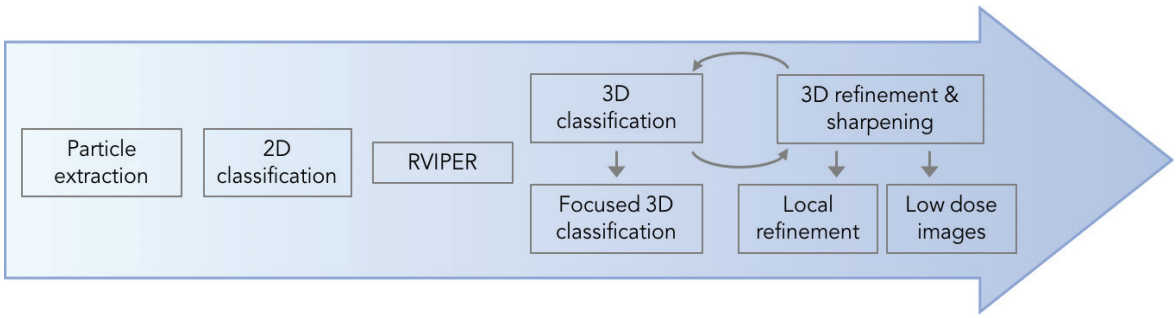


Figure 15: Diagram of the single particle analysis workflow. The scheme depicts the individual processing steps belonging to the SPA workflow. The order of the single steps is represented by the arrow from left to right.

Particle selection and extraction

For the cryo data sets the particles were selected automatically. Xn-XptA1 as well as Mm-TcdA4 were picked using the software Gautomatch (http://www.mrc-lmb.cam.ac.uk/kzhang/). The program can be used either with or without any templates, but the accuracy is higher using a template. To generate templates for Gautomatch, ~2000 particles were manually picked with EMAN2 boxer and used for a first 2D classification with ISAC. Gautomatch parameters were tested for 10 images representing the complete data set (high and low defocus, crowded or empty images) before the complete data set was picked. For Xn-XptA1 as well as Mm-TcdA4, a box size of 400 pixels was used.

The data set of Yp-TcaATcaB was picked with the new boxing tool crYOLO[116]. Only 15 images representing the complete data set were picked manually and used as training data for crYOLO. Yp-TcaATcaB was recorded using a higher magnification and therefore a bigger box size of 546 was used to extract the particles. For all data sets, the particles were extracted from the dose weighted full-dose images after motion correction.

2D classification with ISAC

After extracting the single particles from the images, the particles were sorted by 2D classification using ISAC. Within 2D sorting, particles representing the same orientation are summed up to generate class averages with a better signal-to-noise ratio, which facilitates the differentiation of the particles. The 2D classification in ISAC is based on the *k-means* algorithm resulting in high reliability of the results, because the particles are not only sorted in different classes but also the reproducibility of the classes is checked within ISAC[132]. 2D classes representing contaminations or off-centered particles and incomplete particles were discarded. The parameter to obtain best classes that was varied most within the different datasets, was the particle number per class. Since the datasets contained a large number of particles, multiple rounds of classifications were performed; starting with larger classes (5000 particles) and for the second and third round, classes with 300-400 members were created.

Initial 3D model with RVIPER

The initial 3D model is very important for the processing of EM-data. With a too detailed initial model, bias can be introduced and the final structure could be wrong misinterpreted. For the different Tc toxins, different references were used. For Xn-XptA1 and Mm-TcdA4, a 30 Å filtered map of Pl-TcdA1 (EMD-2297)[42] was used as an initial model. During negative stain analysis of Yp-TcaATcaB, the protein showed a more distinct shape compared to the other TcAs and for this reason Pl-TcdA1 was not used as an initial reference. The program RVIPER (Reproducible Validation of Individual Parameter Reproducibility) can be used to generate a 3D model based on the information of the 2D class averages. 20-50 class averages with good quality and representing different views of the proteins were used for the initial model. C5 symmetry was applied during RVIPER, since C5 symmetry of Yp-TcaATcaB was confirmed within the class averages during ISAC. For better computational performance, the particle images are binned to a box size of 72 pixel during ISAC as well as in RVIPER. Thus, the size of the RVIPER model has to be adapted by scaling to the original pixel and box size in SPHIRE before using it as a reference for 3D refinement or 3D classification.

3D classification with SORT3D

Digital cleaning of the data set was not only performed in 2D but also in 3D. Here, especially flexible domains can be further distinguished. SPHIRE allows to first identify the more flexible regions of the protein with the tool 3D variability. Afterwards, a mask of the flexible domain can be created so that during 3D classification, the focus will be on this region. During 3D classification, the particles are split into a user defined number of classes and a 3D volume is calculated for each class. Depending on the accuracy of the previous 2D classification and on the flexibility of the protein, the 3D classification can

improve the resolution by further cleaning the data set by separating particles with heterogeneous flexible domains. For all proteins, the 3D classification was performed with 2 to 4 classes.

3D refinement with MERIDIEN and sharpening

After digital cleaning of the data set, a 3D volume is generated out of the 2-dimensional data during refinement. The 3D refinement is based on the projection matching. Back projections of the 3D reference are calculated and compared with the 2D images of the single particles. If a back projection matches the particle orientation, the projection parameters of the back projection are assigned to the particle and can be used to calculate a 3D volume. To improve the quality and accuracy of the refinement, various approaches can be implemented during data processing. On the one hand, the use of a mask is useful to reduce noise surrounding the protein density. The mask has to be large enough to ensure that the complete protein density is covered, otherwise areas of the protein will be cut off. Furthermore, a CTF correction can compensate the loss of information occurring at the zeros of the CTF by superimposing images with different defocus values.

During the refinement runs, the data is split into two equally sized halves and two reconstructions are calculated independently. When after several iteration steps no improvements of the respective models are obtained, one last iteration with all the data was computed. After each refinement, a b-factor sharpening was performed. In addition, the MTF (modulation transfer function) that describes the frequency-sensibility influence of the detector on the image is corrected. This is done to combine the two half maps as well as to determine the resolution of the EM-density. With b-factor sharpening, the low frequencies with high-resolution information are amplified which increase the details of the reconstruction. The final resolution value is determined by comparing the calculated FSCs (Fourier shell correlation) of each half map. The frequency at a threshold of 0.143 or 0.5 correlation reflects the resolution[118,119]. In general, the density maps in cryo-EM have a resolution gradient and the given resolution is an averaged value.

Multiple refinement runs were performed for each protein. For Yp-TcaATcaB, the RVIPER model was used as initial model and the filtered Pl-TcdA1 map for Xn-XptA1 and Mm-TcdA4. After the first refinement, the resulting volume was used as a reference for the next refinement. Therefore, refinements were performed iteratively with each time having a more detailed and accurate reference. The final densities were filtered to the estimated final average resolution. The local resolution of the obtained maps was determined using sxlocres of the SPHIRE software package and the final density maps were colored according to the local resolution in Chimera[140].

4.2.3 Structure determination

4.2.3.1 Homology modelling

An atomic model of the respective TcA was built with the final EM density map. Due to the high sequence homology between the single TcAs and Pl-TcdA1, a homology model was generated for Xn-XptA1, Mm-TcdA4, Yp-TcaATcaB and Pl-TcdA4 based on the sequence alignment to the Pl-TcdA1 crystal structure (PDB-ID: 1VW1) using MODELLER[141]. The sequences of the TcAs were aligned using Clustal Omega[142]. The homology model was used as the starting point for atomic model building. The models were first fitted into the EM density map with rigid body fitting in Chimera[140]. More homogenous and rigid protein domains showed already a good fit, however, more variable and flexible regions required further improvement. For this purpose, the PDBs and the EM density map were split into the single domains within Chimera ("surface coloring" & "split map" tool). These domains were then fitted separately into the corresponding density with flexible fitting using iModFit[143].

```
iMODFIT_v1.03_Mac_gcc64_LAPACK/bin64/imodfit_gcc POI.pdb POI.ccp4 3.2 0.015 -n 0.4
```

After visual inspection of the best fits, the single fitted domains were then merged together into a single protomer model. For Yp-TcaATcaB, ~100 amino acids of the shell region could not be modelled based on the sequence identity and had to be removed from the generated homology model.

4.2.3.2 Manual modelling

Since the generated model was only homology based, some regions had still to be built or remodeled *de novo* in Coot[144]. These were mainly the surface exposed regions of the RBDs as well as the shell and the coiled coil domain of Yp-TcaATcaB. Only in areas where the backbone could be traced clearly within the density, the residues were built *de novo*, otherwise these residues or complete domains were deleted. This was the case for RBD C of Xn-XptA1 (amino acids 1386-1495) and Mm-TcdA4 (amino acids 1346-1447) and for Yp-TcaATcaB, the last residues (1139-1240) of the neuraminidase-like domain could not be modelled in the density.

4.2.3.3 Structure validation

The obtained and generated models for the different TcAs were validated and refined within PHENIX (phenix_real_space_refine)[145] to obtain statistical parameters for the model quality.

```
phenix.real_space_refine TcA.pdb TcA_vol.mrc resolution=3.2 secondary_structure.enabled=True
```

The geometry of the final refined models was further evaluated with MolProbity[146]. The data statistics are summarized in Table 6 in section 7.1. In general, the structure validation was an iterative process. Once a complete model was obtained, the validation with PHENIX and MolProbity was performed. Based on these results, the geometry of specific residues was improved in Coot and the model was afterwards again validated with PHENIX and MolProbity.

4.2.4 Structure analysis

The structure analysis, visualization and figure preparation were mainly done in UCSF Chimera and Illustrator. For comparison between the obtained models of the four TcAs and the atomic model of Pl-TcdA1 (provided by Dr. Daniel Roderer), the following analysis was performed for all five TcAs. The analysis includes the identification of the domain organization of the different TcAs and a comparison of specific domains with each TcA and Pl-TcdA1. For this purpose, the hydrophobicity and the Coulomb electrostatics of the domains were calculated and also used for structure prediction at different pH-dependent protonation grades of the models.

4.2.4.1 Structure alignments, knot determination and pore diameter

The structure-based sequence alignments were generated using the T-Coffee expresso server [147]. Based on the alignments of the different TcAs with Pl-TcdA1, the domain organization of the single TcAs could be identified and colored according to the Pl-TcdA1 domains. The conservation between the TcAs was based on a sequence alignment with ClustalW and RMSD values of the structure-based sequence alignment and the Pl-TcdA1 model was colored accordingly in Chimera.

The protein knot was identified and determined using the online server knotprot[148]. The corresponding matrix of the Pl-TcdA1 protein knot from the crystal structure (PDB-ID 1VW1) is deposited and available at the online server knotprot.

The pore diameters of the prepores were calculated with ChExVis[149] (for Xn-XptA1 and Mm-TcdA4) and MOLEOnline[150] (for Yp-TcaATcaB). For this, only the pore domain with the TcB-binding domain was used (aa 2032-2523 in Xn-XptA1, aa 1981-2469 in Mm-TcdA4 and aa 1538-2033 in Yp-TcaATcaB) and submitted to the webservers. For ChExVis, only the *alpha value* and the *probe radius* were varied. The *alpha value* regulates the variability of the atom radius at which cavities are computed. The *probe radius* of the molecule influences the length and direction of the calculated pore volume. The final Xn-XptA1 and Mm-TcdA4 pore diameters were calculated with the *alpha value* and *probe radius* of 0. Default settings were used for Yp-TcaATcaB with the MOLEOnline server. The online programs are both a Voronoi diagram-based method with differences in specific calculation of the channels. The Voronoi

diagram methods describes the partitioning of an area into single regions with defined centres. The MOLE approach uses the Voronoi diagram of atom centres for generating the channels[151] and ChExVis is based on the power diagram to compute channels[149].

4.2.4.2 Hydrophobicity and electrostatics

The classic hydrophobicity scale of a protein from Kyte & Doolittle[152] is valid for soluble proteins, but for a membrane bound protein the scale of Hessa & Heijne is better suitable. This is due to the fact, that amino acids interacting with the membrane and therefore with lipids, have a more complex arrangement of polar and nonpolar areas. Because of this, the assignment of hydrophobicity values of these amino acids varies more in comparison to amino acids that are more exposed to the surface. Since the Tc toxins have a transmembrane region at the channel domain, the scale of Hessa & Heijne was used. Therefore, each amino acid of the different TcAs were assigned with a specific hydrophobicity value and colored according to the Chimera tool "define attribute" with residue specific scores corresponding to Hessa & Heijne[153].

The different protonated states of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB were calculated within the online server H++[154]. The program calculates a theoretical pKa value for each amino acid and adds a hydrogen to the residue based on the protonation grade[155]. The pKa is calculated in regard to the pH and environmental set up, which can be specified by the user to some extent. For all the TcAs, a salt concentration of 150 mM was applied to generate a physiological environment.

The proteins surfaces were visualized in Chimera and then colored accordingly to the different protonation states at pH 4, pH 5, pH 7 and pH 11. This was done using the Chimera tool "electrostatic Coulomb potential coloring" ranging from -14 to 14 kcal/mol.

5 Results and Discussion

During this work, the two A-components Xn-XptA1 and Yp-TcaATcaB were expressed and purified as described in chapter 4.1.2 and 4.1.3. Mm-TcdA4 was expressed and purified by Alina Elsner. For the final comparison of the different Tc toxins with Pl-TdA1, Dr. Daniel Roderer provided the structure of Pl-TcdA1.

Initial structural analysis of Xn-XptA1 was presented before within my master thesis (*Strukturelle Untersuchung des Tc-Toxin-Komplexes von Xenorhabdus nematophila*[156]). In this previous work, an optimized purification workflow was established. Subsequent negative stain analysis resulted in a three-dimensional reconstruction of Xn-XptA1. Furthermore, first cryo-EM evaluations were done at a JEOL-JEM3200 with manual data collection performance. A total of 48 851 particles used for the reconstruction led to a final electron density map with a resolution of 4.3 Å. The obtained model allowed fitting of a Xn-XptA1 homology model and confirmed a strong structural similarity to Pl-TcdA1, whereas the previously published Xn-XptA1 model[83] could not be verified. Furthermore, a pH-induced pore-state as well as an intermediate transition state of Xn-XptA1 were observed in preliminary negative stain EM studies. Hence, the present work focuses on the purification (Yp-TcaATcaB), negative stain analysis (Yp-TcaATcaB and Mm-TcdA4) and the structure determination of TcAs with cryo-EM (Xn-XptA1, Yp-TcaATcaB and Mm-TcdA4).

5.1 Purification of Yp-TcaATcaB

For the structure determination with cryo-EM and for further biochemical studies, the protein Yp-TcaATcaB was purified direct after cell lysis by an ion metal affinity (4.1.3.1) and size exclusion chromatography (4.1.3.2). The A component of *Yersinia pseudotuberculosis* is divided into two parts, termed TcaA and TcaB. Initial trials to coexpress both protein parts did not yield in an intact protein. However, after covalent fusion of both parts the protein could be expressed and purified to high purity. Yp-TcaATcaB was N-terminal modified with an His-tag, that enabled the separation of the target protein by a Ni-NTA affinity resin from contaminating proteins after cell lysis and ultracentrifugation. From this resin, the protein was eluted using a linear gradient of 10-500 mM imidazole (Figure 16A). After the first purification step, eluted fractions were analyzed with SDS-PAGE. The chromatogram of the Ni-NTA purification shows a first elution peak at 70-80 ml retention volume after washing with 10 mM imidazole. By SDS-PAGE, this peak was shown to contain the protein of interest but mostly smaller contaminations with molecular weights around 70 kDa (Figure 16A and B). The calculated monomeric molecular weight of Yp-TcaATcaB is 234 kDa. Fractions eluting at concentrations ranging between 85 and 180 mM imidazole showed a strong protein band between 200 and 250 kDa on the SDS-PAGE gel,

corresponding to the protein of interest. Further protein bands indicate, that the fractions still comprise impurities due to nonspecifically bound proteins on the column. Yp-TcaATcaB containing fractions were combined and remaining impurities, due to nonspecific binding of proteins to the affinity resin, were further separated by size exclusion chromatography (SEC).

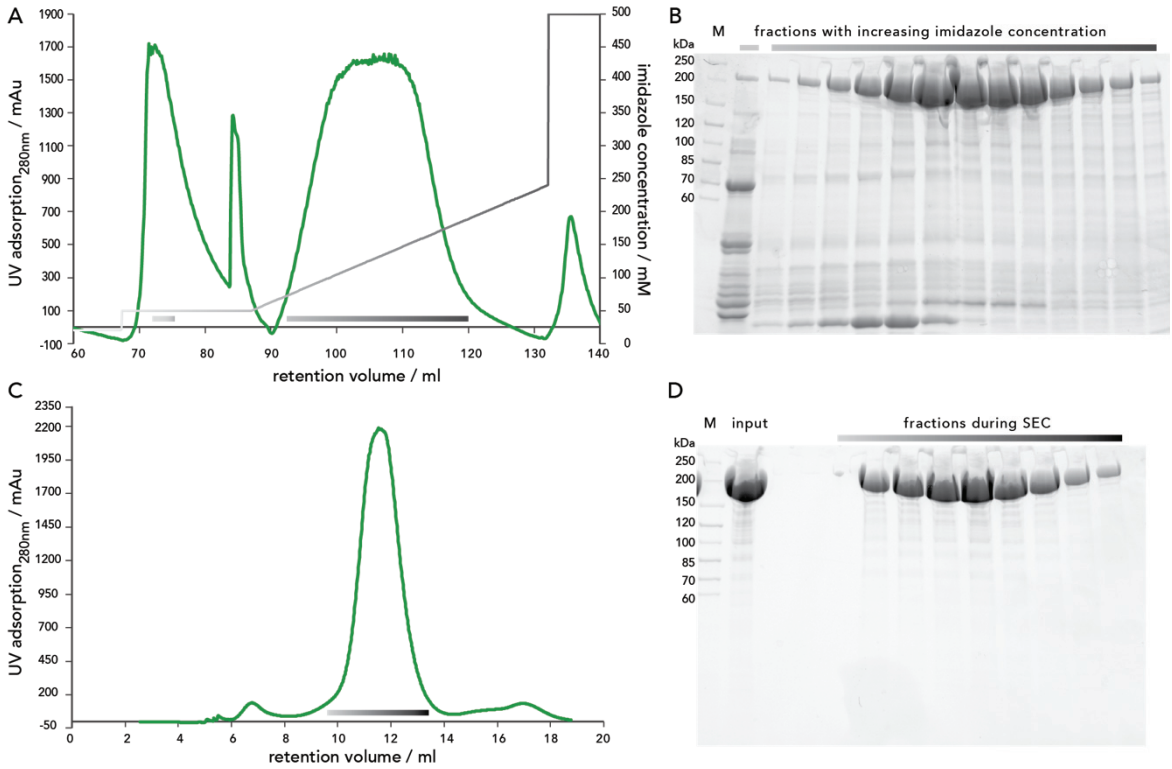


Figure 16: Chromatograms and SDS-PAGE gels after the two purification steps of Yp-TcaATcaB. A) Chromatogram of the Ni-NTA purification step with increasing imidazole concentration depicted as a gradient from light grey to black. The fractions that were analyzed via SDS-PAGE are marked with a line having the same color gradient as the imidazole concentration. B) Corresponding SDS-PAGE gel to the Ni-NTA chromatography. M=marker and the fractions are marked with a line indicating the imidazole concentration during elution. C) Chromatogram of the size exclusion chromatography with a Superose 6 10/300. D) SDS-PAGE gel of the size exclusion chromatography. M=marker, input=sample before column injection (diluted 1:10), fractions during size exclusion with increasing retention volume visualized with the gradient from light grey to black as in C).

As displayed in the SEC chromatogram and the corresponding SDS-gel (Figure 16C and D), the void volume peak containing aggregates eluted after a retention volume of at 7 ml. The main peak (11-13 ml retention volume) shows a monodisperse profile and is followed by a smaller peak at a retention volume of ~17 ml. By SDS-PAGE analysis of single fractions, it becomes visible, that the protein of interest eluted within the main peak whereas contaminants correspond to the smaller peak (Figure 16D). The resulting protein had an overall purity of >95 %. Therefore, the purity of Yp-TcaATcaB was successfully improved by separating Yp-TcaATcaB from small contaminations. Comparing the retention volume of the main peak at 13 ml with a molecular weight standard, this volume corresponds to a protein with a molecular weight of around 960 kDa (thyroglobulin)[157]. A pentameric assembly was observed for the Tc

toxin A components Pl-TcdA1 and YenA1A2. In regard to the sequence similarity of 38.1 % to Pl-TcdA1 and 37.5 % to YenA1A2, the protein of interest was expected to form as well a pentamer[158]. Indeed, the retention volume of 11-13 ml matches the pentamer molecular weight of 1.17 MDa calculated for Yp-TcaATcaB. Furthermore, the retention volume is comparable to the retention volume of Pl-TcdA1[42], Xn-XptA1[156] and Mm-TcdA4. To verify the presence of an Yp-TcaATcaB pentamer in the main peak fractions, the sample was analyzed with negative stain EM, which will be discussed in more detail in the next chapter.

5.2 Negative stain electron microscopy of Tc toxins

The sample quality of the Tc toxins was analyzed with negative stain electron microscopy, which also provided first insights into the overall dimensions and assembly of the proteins. Especially for proteins without any structural information, namely Mm-TcdA4 and Yp-TcaATcaB, this method is a fast approach for a first sample evaluation.

5.2.1 Mm-TcdA4

The purified Mm-TcdA4 protein was provided by Alina Elsner and the sample quality was investigated with negative stain EM. In the beginning of this thesis, no structural information was available for Mm-TcdA4. Considering the high sequence homology of 40.7 % to Pl-TcdA1 and 41 % to Xn-XptA1, a similar three-dimensional shape of the protein was assumed. The molecular weight of a Mm-TcdA4 protomer is 275 kDa and the respective pentamer would result in a similar molecular weight to Pl-TcdA1 and Xn-XptA1 with 1.35 MDa. Indeed, negative stain micrographs revealed the same bell shape particle for Mm-TcdA4 as it was observed for Pl-TcdA1 and Xn-XptA1 (Figure 17). Also the particle dimensions are comparable with 25 nm in length and 18 nm width. Both, side views and top views, are evenly distributed on the grid and also the background is clear of any contaminations (Figure 17A). Thus, the protein sample is free of any contaminations and reveals a homogenous particle distribution. Since the particle had a similar appearance and also behavior in negative stain EM compared to Xn-XptA1 that was previously analyzed with negative stain EM, further evaluation and structure analysis of Mm-TcdA4 was directly performed with cryo-EM.

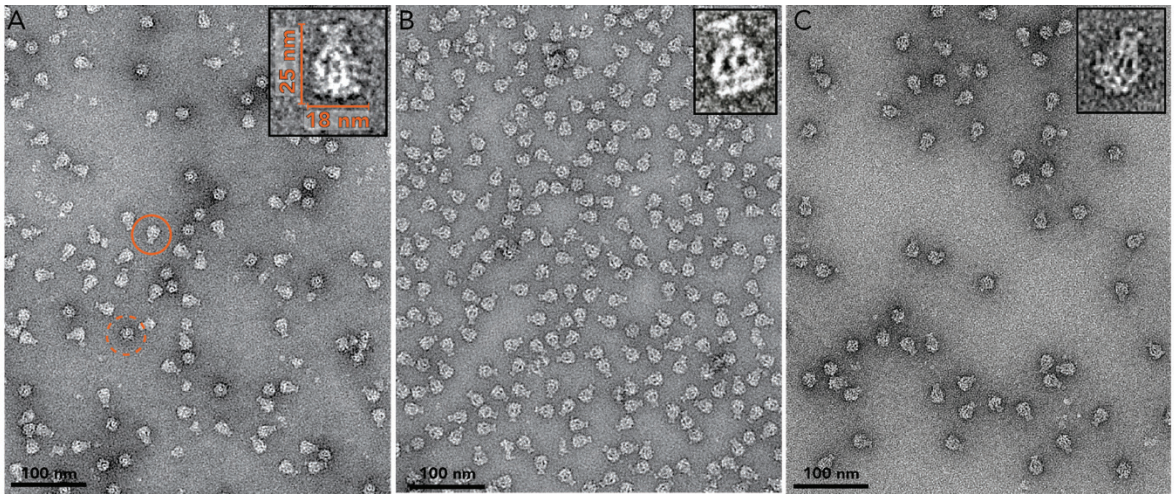


Figure 17: Negative stain micrograph of Mm-TcdA4, Xn-XptA1 and Pl-TcdA1. A) Typical negative stain micrograph of Mm-TcdA4 after the two-step purification. A particle representing the side views is circled in orange (solid line) and a particle representing the top views is marked with dashed lines (orange). In the upper right corner is a close-up view of a particle with the respective dimensions. B) Negative stain micrograph of purified Xn-XptA1 and C) Pl-TcdA1 both with a typical particle zoomed in (upper right corner).

5.2.2 Yp-TcaATcaB

The fractions of the main peak after size exclusion chromatography were visually checked and evaluated with negative stain EM. In Figure 18A, a typical negative stain micrograph of Yp-TcaATcaB is shown. The protein shows a homogenous distribution and the clean background demonstrates the success of the purification. Most of the protein particles exhibit a round shape. Only a minor part of the particles has a slightly elongated shape, reminding of a bottle. Since the structure of Yp-TcaATcaB was unknown so far, it was compared with negative stain data of the homologue *Yersinia entomophaga* YenA1A2 protein. The negative stain micrograph of YenA1A2 from Landsberg et al.[61] reveals a bottle-shaped particle that represents the side views of the protein. This particle orientation is dominant and less top view particles with a round shape are present (Figure 18D). Having a closer look to the negative stain micrograph of Yp-TcaATcaB (Figure 18A), it becomes obvious, that the protein has a preferred orientation with mainly top views. For the following structure determination, an evenly distribution of top and side view particles is necessary. Therefore, to induce more side views, grids were pretreated with poly-L-lysine. The poly-L-lysine (PLL) generates a poly-cationic surface which leads to a different charge distribution on the grid[159]. A micrograph of these grids of Yp-TcaATcaB shows nearly all the particles in side view orientation (Figure 18B). The different behavior of the protein could be explained by different distribution of electrostatic charges on the protein surface resulting in the preferred orientation in either top view (without PLL) or side view (with PLL).

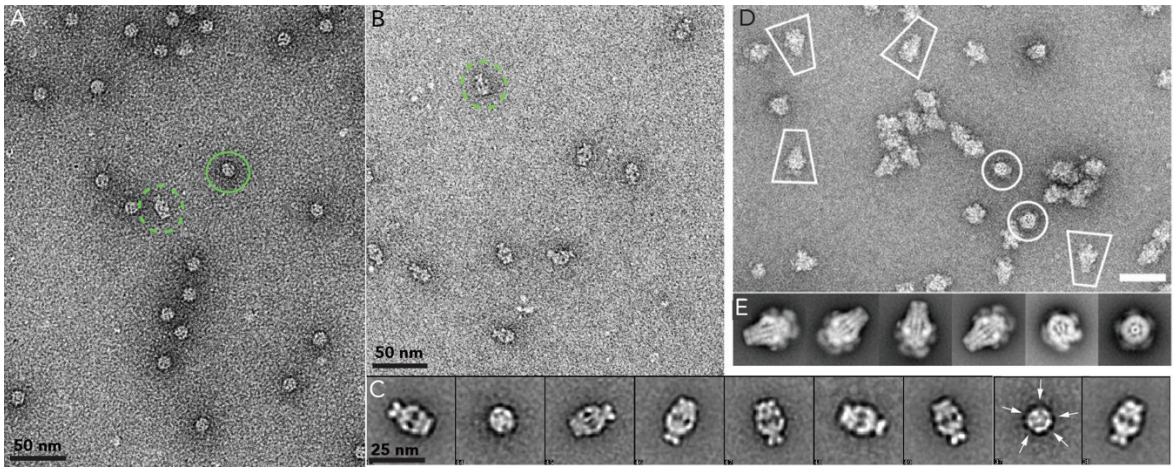


Figure 18: Negative stain data of Yp-TcaATcaB in comparison to YenA1A2. A) Micrograph of a Yp-TcaATcaB sample stained with uranyl formate. B) Micrograph of a Yp-TcaATcaB sample with a pretreatment of the grid with poly-L-lysine. Particles presenting a top view (solid line) and a side view (dashed line) are marked with green circles. C) The 2D class averages after ISAC show the different particle orientations with side and top views. The five-fold symmetry of the protein is visualized with white arrows in the top view class average at the right. D) Figure from Landsberg et al. showing a negative stain micrograph of YenA1A2 and E) 2D back projections. Scale bar, 50 nm[61].

Since Yp-TcaATcaB conducts a different behavior in negative stain EM regarding its preferred orientation compared to Mm-TcdA4, further characterization of the protein by 3D reconstruction was done before continuing with cryo-EM analysis. The negative stain data was processed by Berenike Rath within her master thesis under my supervision[160]. For generating 2D class averages and a first 3D volume with negative stain EM, micrographs from grids with and without PLL treatment were used for particle selection to obtain an even distributed particle orientation. The 2D class averages after ISAC reveal both orientations, the Yp-TcaATcaB particle with the bottle-shaped side view and the globular top view (Figure 18C). In contrast, the YenA1A2 particles contain additional two domains, chi1 and chi2, that are located at the outer periphery of the protein. These domains are absent in Yp-TcaATcaB and likely explains the slimmer shape of the Yp-TcaATcaB particle, which becomes more obvious when comparing the 2D class averages (Figure 18C, E). Additionally, the 2D class averages give rise about the particle symmetry of Yp-TcaATcaB. In the 2D class average presenting a top view, five protomers can be identified, thereby confirming the assumption of a pentameric assembly with five-fold-symmetry (Figure 18C).

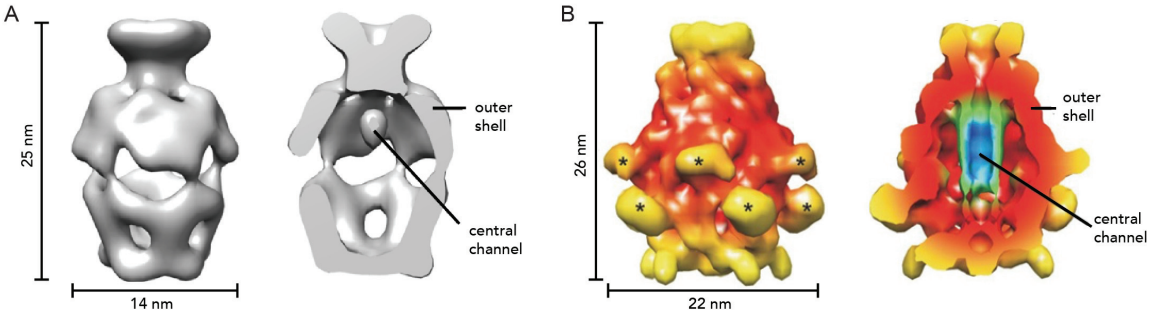


Figure 19: Negative stain electron density map of Yp-TcaATcaB. 3D volume of Yp-TcaATcaB from negative stain EM data at around 20 Å resolution with a complete surface representation (left) and a cross section (right). B) Negative stain volume of YenA1A2 from *Y. entomophaga* from Landsberg et al.[61] showing the complete surface (left) and the cross section (right) of the protein. Stars mark the domains where chitinase 1 and 2 are located.

Finally, an initial model was generated from the class averages to obtain a first 3D volume of Yp-TcaATcaB. Since the volume is only an initial model with low resolution of 20 Å, no detailed information can be gathered. Nevertheless, the reconstruction clearly shows the tenuous shape of the protein and illustrates its overall dimensions. Yp-TcaATcaB exhibits a bottle shaped structure with a density that could correspond to a central channel surrounded by an outer shell. The protein has a similar size compared to YenA1A2 with 25 nm in height and a smaller cross-section dimension of 14 nm. With these results, first insights into the Yp-TcaATcaB structure were obtained and reveal that Yp-TcaATcaB has a similar protein architecture compared to YenA1A2. To gain more information about the protein domain organization and the localization of the two major subunits, TcaA and TcaB, further studies were performed with electron cryo microscopy.

5.3 Cryo-EM and structure determination of Tc toxins

The workflow of the final structure determination was similar for the three TcAs Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. First, optimized conditions for cryo-EM grid preparation were manually screened using either a JEOL JEM1400 or a FEI Tecnai Spirit microscope. Data sets were collected at a TITAN KRIOS microscope using the automated data collection software EPU (FEI). Detailed information are listed in section 4.2.1.2. The data processing was performed with the software package SPHIRE following a similar strategy for all three toxins. The final atomic structures were built and refined manually in Coot using respective homology models of Pl-TcdA1 as starting models.

5.3.1 Cryo-screening

Many different parameters influence the cryo-EM sample preparation that have to be considered and optimized experimentally. Individual grid types alter in terms of mesh dimensions, hole size and additional support layers like carbon, graphene oxide or gold[161] and thus have an impact on the sample preparation. Additives like detergents have an influence on surface tension while plunge-freezing and thereby improve the homogeneity of the ice and particle distribution[162]. The ice thickness can be varied with different blotting times during the plunging process. Thick ice layers typically lead to loss of contrast and detailed information. Thin ice, especially in combination with additional carbon grids, might not embed protein molecules entirely. This leads to dehydration and oxidation at the air-water interface and ultimately results in protein denaturation and thus to artifacts in the resulting processed density maps. The adsorption at the air-water interface, which can also occur independent from the ice thickness, results furthermore to preferred orientating [163]. To find an optimal ice and particle distribution, different conditions were tested (Table 3). Two examples including the final optimized condition will be described in more detail for each protein.

Table 3: Conditions tested for cryo-EM sample preparation with the Cp3 cryoplunger. Optimized conditions are highlighted in blue.

Protein	Grid type	Conc. / mg/ml	Incubation / s	Blotting time / s	Detergent
Xn-XptA1	C-flats 2/1	3x1.7	50 (2x)	2.1	0.01 % Tween-20
	QF 2/1, 2 nm	0.08	60	2.5	0.025 % Tween-20
	QF 2/1, 2 nm	0.1	40	2.2	0.025 % Tween-20
Mm-TcdA4	QF 2/1, 2 nm	0.08	40	2.2	0.03 % Tween-20
	QF 2/1, 2 nm	0.08	60	2.3	0.03 % Tween-20
	QF 2/1, 2nm	0.1	40	2.5	0.03 % Tween-20
Yp-TcaATcaB	QF 2/1, 2 nm with PLL	3x4.0	60 (2x)	2.3	0.025 % Tween-20
	C-flats 2/1	3x1.0	40 (2x)	2.0	0.05 % Tween-20
	QF 2/1	3.5	-	2.0	0.05 % Tween-20

Xn-XptA1

As described before, the first evaluation of Xn-XptA1 with cryo-EM was already performed previously[156]. That is why in general, conditions for obtaining evenly distributed particles in ice on the grids were known. However, the particle concentration needed to be increased and two different grid types were tested for generating new cryo-EM grids. Grids without additional carbon layer (C-flat 2/1) require high

protein concentrations to ensure that it is evenly distributed within the carbon hole without sticking to the carbon and carbon edges. When applying 4 µL of a 1.3 mg/ml protein solution, the concentration of the sample was insufficient to embed high concentrations of the protein particles in the ice layer. To enhance the particle concentration in ice, the sample was applied three times to the grid surface with manual blotting steps in between. Even after multiple sample application steps, the particle concentration was less compared to additional carbon layer grids and the distribution was inhomogeneous (Figure 20A). In some holes only few particles were present, whereas the neighboring hole contained more particles. In addition, the protein particle tends to stick to the carbon edges. In Figure 20B, the Xn-XptA1 sample (0.1 mg/ml) was applied on a grid with additional carbon layer (QF 2/1). The protein's pentameric complex has a molecular weight of 1.4 MDa, and therefore it still gives sufficient contrast within the ice layer on a carbon support film although additional carbon reduces the image contrast. Here, the holes were fully decorated with particles (Figure 20B and C) and the protein was evenly distributed. Since the particle distribution on the additional carbon layer grid was much more homogenous, the data acquisition was performed with additional carbon layer grids.

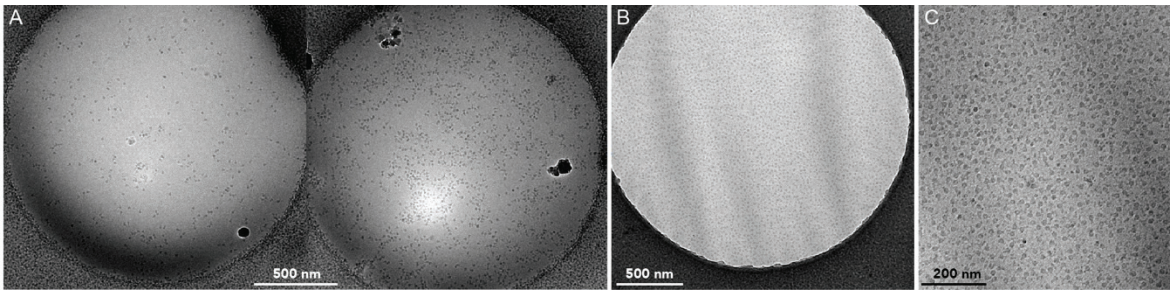


Figure 20: Different cryo-EM grids of Xn-XptA1. A) Two micrographs of the same C-flat 2/1 grid without additional carbon layer and three times sample application (3x1.7 mg/ml). B) Micrograph of the Xn-XptA1 sample applied on the QF additional carbon grids and C) a close-up view with a defocus of -3 µm.

Mm-TcdA4

For Mm-TcdA4, the same buffer conditions (Table 3) were used like for Xn-XptA1 in negative stain and the protein showed a similar behavior regarding particle orientation on the negative stain grid compared to Xn-XptA1. Therefore, grids with an additional carbon layer (QF 2/1) were used for screening of cryo conditions and only the particle concentration as well as the distribution was optimized. This was achieved by varying the Tween-20 concentration and the blotting time. As seen in Figure 21A, the ice thickness and particle distribution are both were very homogenous and despite the additional carbon, the particles were clearly visible (Figure 21B). The diagonal stripes in Figure 21B correspond to an uneven carbon distribution.

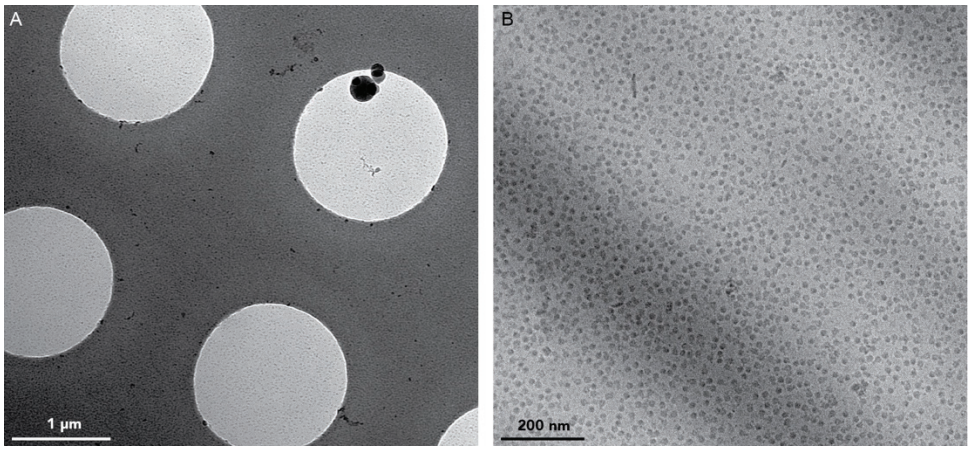


Figure 21: Cryo-screening of Mm-TcdA4 applied grids. A) Micrograph at lower magnification showing an overview of the QF 2/1 grid. B) Image at the hole center with a homogenous distribution of particles within the ice layer. Used defocus -3.5 µm.

Yp-TcaATcaB

As presented before, the protein Yp-TcaATcaB exhibits a preferred orientation on the carbon film of negative stain EM grids. Therefore, a carbon support layer was avoided and grids without additional carbon layer were used for cryo-EM. Furthermore, during previously performed cryo screenings the influence of graphene oxide and additional carbon grids with and without PLL was tested[160]. However, both grid types were shown to be not suitable for cryo-EM analysis of Yp-TcaATcaB because of inhomogeneous particle distribution and preferred orientation. With this knowledge, the cryo-screening was performed using C-Flat 2/1 grids without additional carbon layer but multiple times of sample application to fully decorate the grids with protein. However, these conditions resulted in strong cluster formation of Yp-TcaATcaB (Figure 22A).

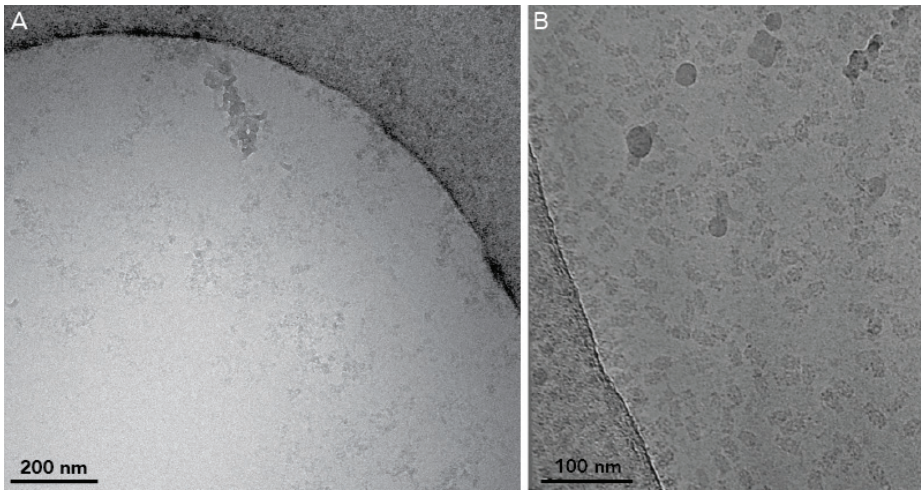


Figure 22: Different conditions tested for Yp-TcaATcaB cryo-screening. A) Micrograph of a C-Flat 2/1 grid with 3x1 mg/ml sample application and 0.05 % Tween-20. Recorded on a JEOL JEM1400 microscope. B) A micrograph of Yp-TcaATcaB applied on a QF 2/1 grid without additional carbon and one-time sample application (3.5 mg/ml). Recorded on TITAN KRIOS.

To reduce the protein clustering, different Tween-20 concentrations were tested, since detergents influence the surface tension as well as the viscosity of the protein solution[162]. Neither different Tween-20 concentrations (0.025 and 0.05 %) nor multiple sample applications (1x 4 µL and 3x4 µL of a 1 mg/ml protein solution) improved the particle distribution or reduced the clustering. Especially after multiple sample application, the clustering effect seemed to be more pronounced, which could be due to a destabilized protein complex upon longer incubation time. Nevertheless, with a single sample application, the concentration of the sample was not high enough. After a fresh purification with highly concentrated sample of 3.5 mg/ml, the particle amount within the ice layer was successfully increased and particles were nicely distributed. Due to the high concentration of the protein, a single sample application of Yp-TcaATcaB was sufficient and therefore, protein clusters could be diminished (Figure 22C).

5.3.2 Single particle analysis and final atomic model

As mentioned in section 4.2.2.2, all data sets were collected at a TITAN KRIOS microscope operated at 300 kV and equipped with a XFEG. Images were recorded with a Falcon II detector for Xn-XptA1 and Mm-TcdA4 and with a Falcon III detector for Yp-TcaATcaB. Precise parameters for the data collections of each toxin are listed in chapter 4.2.2.2. After data collection, all preprocessing and processing steps, including single particle analysis, were performed in a similar manner for all toxins (see chapter 4.2.2.3 and 4.2.2.4).

5.3.2.1 Xn-XptA1

For Xn-XptA1, a total amount of 3565 images was recorded and the number of particles after particle extraction was 565,210. The processing strategy for Xn-XptA1 included two times of 2D classification and several refinement rounds, changing references and masks (see chapter 4.2.2.4), until the resulted reconstruction could not be further optimized. After two rounds of 2D classification, 39 % of the initial particles remained (218,992) and more than 60 % of the initial selected particles had to be discarded. This is mainly because of the used automated particle picking software, which can lead to a higher number of false picked particles. The program selects the particles based on a given reference but also ice contaminations or not well centered particles can be selected. These contaminations, off-centered and also damaged particles lead to the high particle number and were discarded after classification. The particles had different orientations from side view to tilted and top views, which is depicted in the 2D class averages and visualized the angular distribution plot (Figure 23A and B, respectively). More side views are present, which can also be observed in the angular distribution plot. Nevertheless, the

different tilted particles contributed enough information to cover the complete angular space. Like previously observed in negative stain EM, the five-fold symmetry of the particle is evident in the top view (upper class average), that confirms the assumption of a pentameric complex of Xn-XptA1. After several refinement runs imposing C5 symmetry and changing the initial reference iteratively, as well as applying masks during refinement and sharpening, a final density map with a resolution of 2.84 Å (0.143 criterion) was obtained (Figure 23C). At this resolution β-sheets and α-helices are visible and especially for the α-helical region, the orientation of amino acid side chains can be determined in the density. In cryo-EM, the reconstructed density map differs in local resolution. More flexible regions are less resolved then rigid and well-ordered domains. This can also be observed for Xn-XptA1 (Figure 23E). Especially the inner translocation channel domain and the lower helical shell have a local resolution between 2.4 and 3.0 Å, whereas the peripheral β-sheet domains exhibit higher flexibility and therefore are less resolved (Figure 23E).

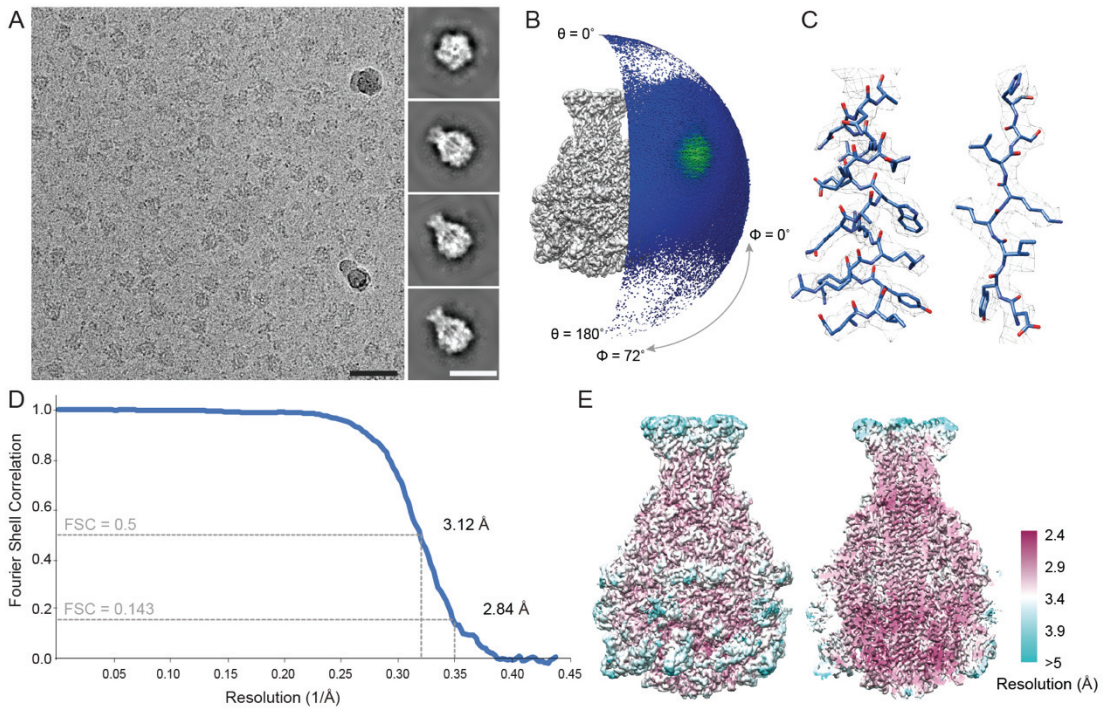


Figure 23: EM analysis of Xn-XptA1. A) Typical drift-corrected micrograph acquired on a Falcon II direct electron detection camera (scale bar 50 nm) and reference-free 2D class averages obtained in ISAC (scale bar 25 nm). B) Angular distribution of all particles used for the final reconstruction. C) Superimposition of segments of the final atomic model with the cryo-EM density map (mesh), showing an α-helical part (left) and a β-sheet region (right). D) Fourier shell correlation (FSC) curves of the final, filtered density maps obtained in MERIDIEN. The average resolution at 0.143 and 0.5 FSC criterion are indicated, respectively. E) EM density maps colored accordingly to the local resolution determined by sxlocres, showing the complete electron density and a longitudinal cut through the density map.

As described in the introduction, until now only a low-resolution negative stain structure from *X. nematophila* TcA was published before, giving rise to a tetrameric protein assembly with a cavity inside the core accordingly to Lee et al.[83]. The published negative stain images and 2D class averages

reveal both, a tetrameric assembly and a bottle shaped particle similar to Pl-TcdA1 and the here presented Xn-XptA1 structure. However, the focus was on the higher concentrated tetrameric assembly that resembles the structure of chaperons (Figure 4). During afore studies, the assumption of a tetrameric assembly for Xn-XptA1 was disproved with negative stain and cryo-EM studies, where the high-resolution structure revealed a pentameric oligomerization[156]. As depicted in Figure 24, Xn-XptA1 consists of a central channel surrounded by an outer shell that is composed of a β-sheet region (receptor-binding domains RBD, dark blue) and an α-helical region (lighter blue). Thus, the density map of Xn-XptA1 resemblances the protein structure of Pl-TcdA1, as it can be expected in consideration of the high sequence similarity of 46.7 % between the two proteins.

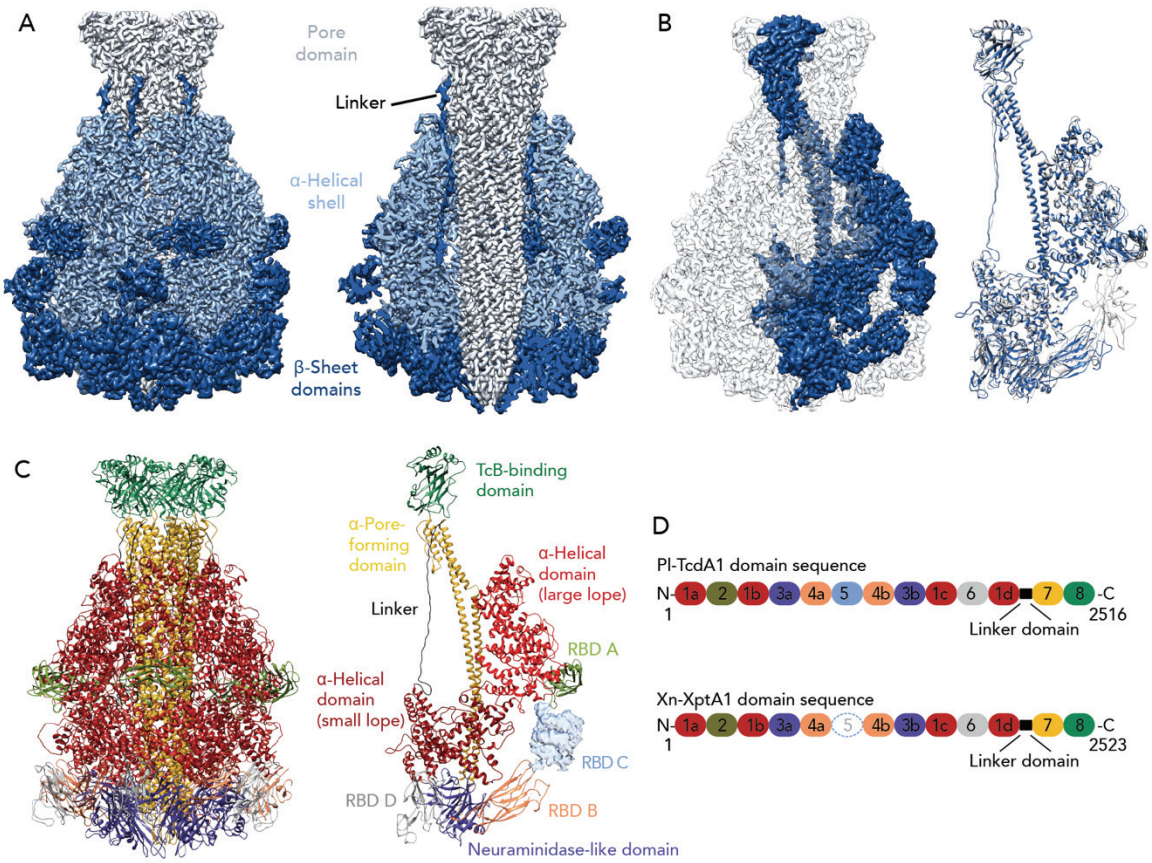


Figure 24: High-resolution cryo-EM density map at 2.8 Å and atomic model of Xn-XptA1. A) Electron density map of the Xn-XptA1 pentamer colored in a blue gradient from light to dark, representing the central pore domain, the α-helical shell and the β-sheet domains and showing the electron density map and a longitudinal cut through the density map. B) Pentamer of Xn-XptA1 with one protomer colored in dark blue, showing the orientation of a Xn-XptA1 monomer and the superimposed homology model based on a Pl-TcdA1 structure (grey) aligned to the final Xn-XptA1 structure (blue). C) Pentamer and protomer of Xn-XptA1 colored accordingly to the eight subdomains. The density map of the less resolved RBD C is shown in the protomer structure as well as the division of the helical shell domain in the large and small lobe. D) Scheme of the domain organization of Xn-XptA1 compared to Pl-TcdA1. The unresolved RBD C (domain 5) is indicated with dashed circles.

The reached resolution was sufficient to build an atomic model of Xn-XptA1. A homology model of Xn-XptA1 was generated, which was based on the crystal structure of Pl-TcdA1 (PDB-ID: 1WV1) and moderately fit into the EM density map. As illustrated in Figure 24B, the starting homology model (grey) and the final Xn-XptA1 model (blue) differ mainly in the peripheric regions of the shell domain, especially in the RBDs. Residues 1338-1500 were eliminated and are not shown in this model, since the density of RBD C (Figure 24C and domain 5 in D) was too less resolved for placing β-strands with a clearly determined direction and the suggested folding of the homology model differed tremendously. Therefore, the final atomic model of Xn-XptA1 lacks 7.5 % of the amino acid sequence. Nevertheless, the overall domain organization is strikingly similar to Pl-TcdA1 and the sequence of Xn-XptA1 can also be divided into eight domains with the α-helical shell (1-296, 439-1078, 1611-1636, 1781-1988), four RBDs (RBD A with 297-438, RBD B with 1290-1385, 1469-1583, RBD C with 1386-1495, RBD D with 1637-1780), a neuraminidase-like domain (1079-1289, 1584-1610), a central channel (2028-2343) and the TcB-binding domain (2344-2523). Even the linker domain (1989-2027) that is described as the entropic spring and is thought to be important for the mechanism of action of Pl-TcdA1 is resolved in this model (Figure 24A,B and C)[44]. Thus, the structure of Xn-XptA1 reveals the same structural arrangement of eight domains like observed for Pl-TcdA1 and therefore suggest a similar mechanism of action. A detailed comparison to Pl-TcdA1 and the other TcAs as well as the functionality of the single domains will be discussed later in chapter 0.

5.3.2.2 Mm-TcdA4

The same processing strategy as used for Xn-XptA1 was applied for Mm-TcdA4. First, 479,289 particles were selected from 4,383 images. As no negative stain electron microscopy model was created prior to cryo-EM analysis, the reference-free 2D class averages after ISAC were analyzed more carefully and compared to Xn-XptA1 class averages. The same bell-shaped particles were obtained for Mm-TcdA4 as seen before for Xn-XptA1, showing details like a channel in the center. Again, a five-fold symmetry becomes obvious in the top view of the particle (Figure 25A, upper 2D class average), which was observed for other TcA components before. After 2D classification and 3D sorting, a final particle number of 247,513 particles remained (52 %). With these particles and applying C5 symmetry, an overall resolution of 3.27 Å (0.143 criterion) was finally achieved (Figure 25D).

As previously described for Xn-XptA1, also Mm-TcdA4 shows a resolution gradient with higher resolved α-helical regions in the central part of the protein (2.8 – 3.3 Å). The top of the channel, representing the TcB-binding domain, and the β-sheet domains protruding to the periphery exhibit higher flexibility (Figure 25E). Here, a resolution around 4 Å and less is present.

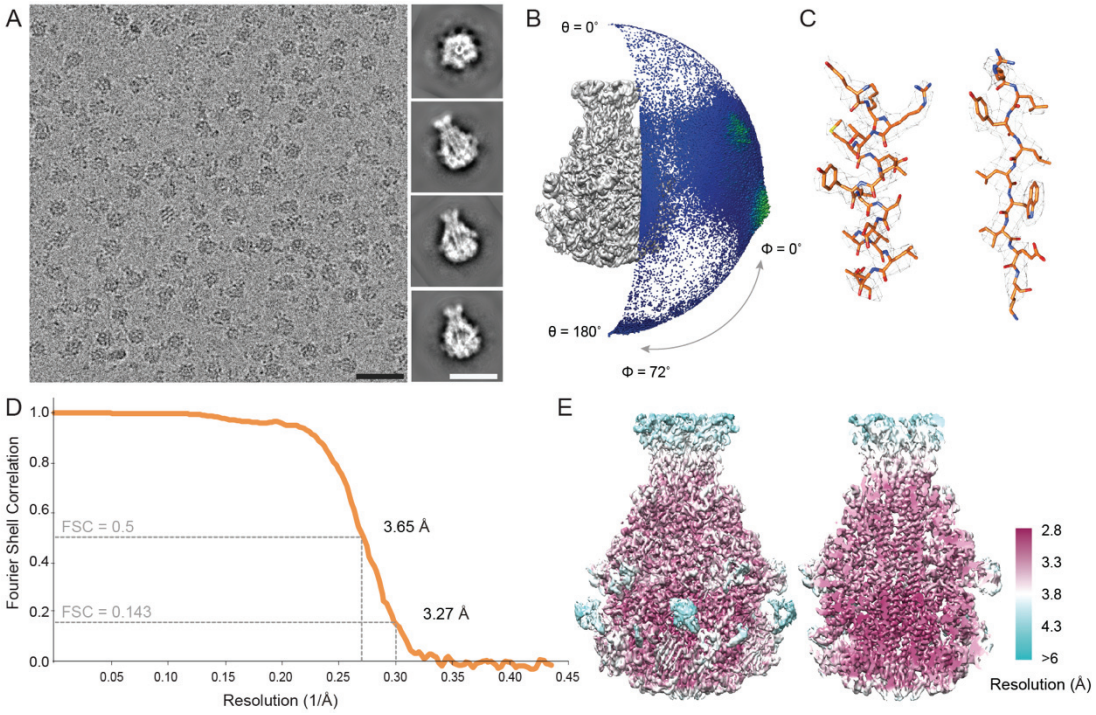


Figure 25: EM analysis of Mm-TcdA4: A) Typical drift-corrected micrograph acquired on a Falcon II direct electron detection camera (scale bar 50 nm) and reference-free 2D class averages obtained using ISAC (scale bar 25 nm). B) Angular distribution of all particles used for the final reconstruction. C) Superimposition of segments of the final atomic model with the cryo-EM density (mesh), showing an α-helical part (left) and a β-sheet region (right). D) Fourier shell correlation (FSC) curves of the filtered density map obtained in MERIDIEN. The average resolution at 0.143 and 0.5 FSC criterion are indicated. E) EM density maps colored accordingly to the local resolution determined by sxlocres, showing the complete electron density and a longitudinal cut through the density map.

Having a closer look to the final electron density map in Figure 26A, it becomes obvious that the general architecture of this toxin resembles Pl-TcdA1 and Xn-XptA1. The pore domain, consisting of ten helices, is surrounded by an outer shell, which can be divided in the α-helical part and the β-sheet part. In addition, the linker, connecting the pore domain and the shell domain, is present and resolved. The sequence identity between Mm-TcdA4 and Pl-TcdA4 is 40.7 % and allowed generation of a homology model of Mm-TcdA4 based on the sequence and structure of Pl-TcdA1 (PDB-ID: 1VW1). The major structural variances between the homology model (grey) and the final Mm-TcdA4 structure (orange) are located within the β-sheet regions (Figure 26C). In these regions, the sequence similarity was low and the domains had to be built manually. Since the resolution in this region was around 4-6 Å, not all the RBDs could be modeled and built appropriately. Especially the resolution of the RBD C density was insufficient to determine the sense and orientation of the β-sheets as well as the connecting loops. Therefore, the final atomic model of Mm-TcdA4 is missing the residues of RBD C (residues 1346-1447) and some flexible loop regions (Figure 26B,D).

By structural sequence alignment, single domains could be assigned to Mm-TcdA4 and revealed that Mm-TcdA4 has indeed the same multi-domain architecture as described for Pl-TcdA1[43] and Xn-XptA1. The α-helical shell is capturing the amino acids 1-299, 443-1066, 1570-1594, 1736-1937 and the four

RBDs are composed of RBD A (300-442), RBD B (1267-1345, 1448-1541), RBD C (1346-1447) and RBD D (1595-1735). The neuraminidase-like domain (1067-1266, 1542-1569), the pore forming domain (1981-2291), TcB-binding domain (2294-2469) and the linker (1938-1980) complete the structure of Mm-TcdA4 (Figure 26D and E).

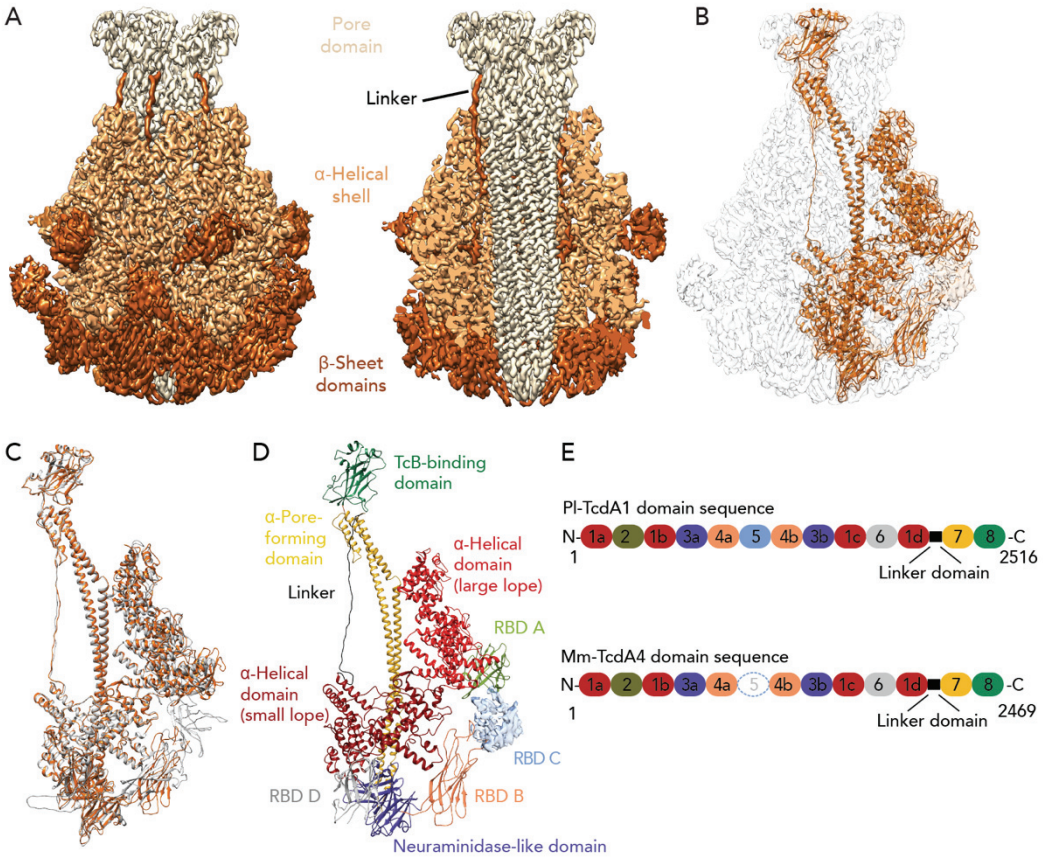


Figure 26: Final EM density map and atomic model of Mm-TcdA4. A) Pentamer of Mm-TcdA4 colored in an orange gradient from light to dark representing the central pore domain, the α-helical shell and the β-sheet domains, showing the complete electron density and a longitudinal cut through the density map. B) Pentameric map in light grey with one protomer structure in orange. C) Homologue model based on Pl-TcdA1 structure (grey) and the final Mm-TcdA4 structure (orange). D) Mm-TcdA4 protomer colored accordingly to the eight domains. The density map of the not resolved RBD C is shown as well as the division of the helical shell domain in the large and small lobe. E) Scheme of the domain organization of Pl-TcdA1 and Mm-TcdA4.

5.3.2.3 Yp-TcaATcaB

In contrast to the two previously described Tc toxin components, the Yp-TcaATcaB protein was embedded in pure ice instead of using additional carbon grids during sample preparation. This leads to an enhanced contrast of the particles (Figure 27A). In total, 6,263 images were recorded and with the automated particle picker crYOLO an initial number of 440,777 particles were selected. During 2D classification, the overall shape of the protein became more distinct and the class averages showed the slim and bottle-shaped protein structure as seen before in negative stain EM. The class averages and

the first initial negative stain model of Yp-TcaATcaB were limited in resolution. However, in cryo-EM, already the class averages reveal more detailed information about the Yp-TcaATcaB protein assembly. The class averages on the left side in Figure 27A were processed with the original box size of 382 pixel and are better resolved compared to the class averages on the right. These class averages come directly from ISAC and are binned to 72 pixel due to computational costs (see chapter 4.2.2.4). Electron density linking the shell domain with the upper part of the channel was not detectable with negative stain EM, but could be observed in the higher resolved class averages (Figure 27A, white arrow in class averages on the left). Furthermore, some classes appeared, that contained damaged particles, obviously lacking some parts of the outer shell (Figure 27A, class averages on the right). Since a protomer consists of two proteins, it might be possible that due to the mechanical stress during plunge freezing or by non-specific proteolytic cleavage, the two covalently linked subunits disassemble. These incomplete particles as well as contaminations were excluded from the reconstruction, leading to a final particle number of 237,295 (54 %).

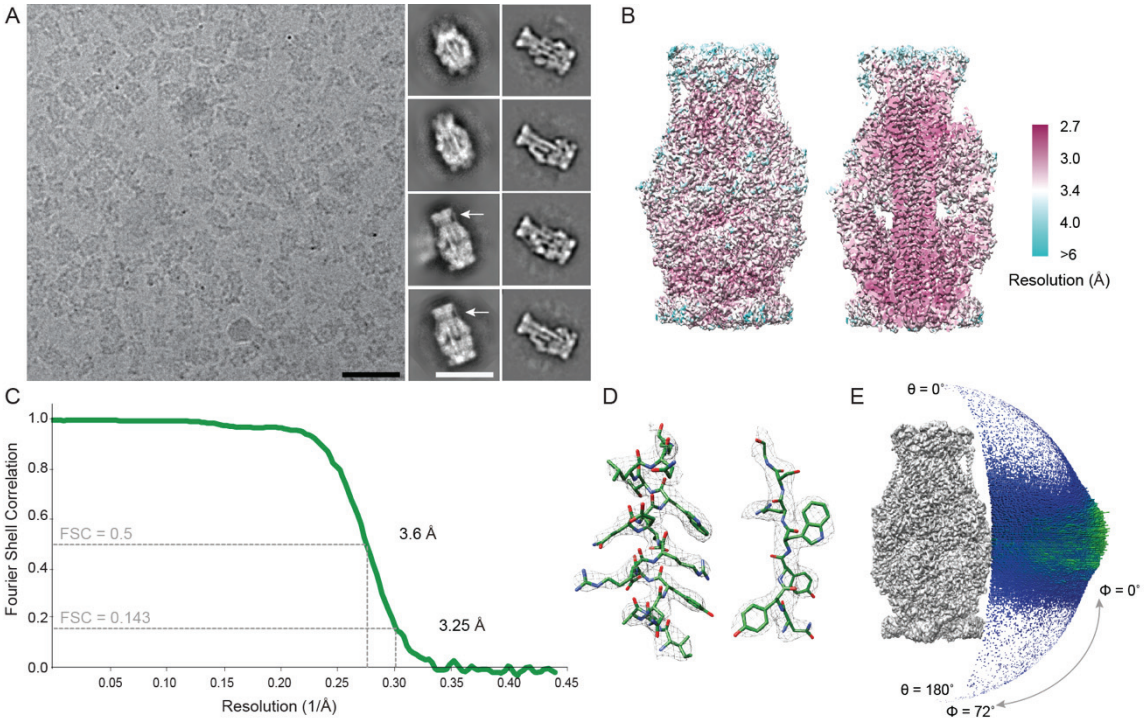


Figure 27: EM analysis of Yp-TcaATcaB. A) Typical drift-corrected micrograph acquired with a Falcon III direct electron detection camera (scale bar 50 nm) and reference-free 2D class averages obtained in ISAC with Beautifier option (left) and without Beautifier option (right), scale bar 25 nm. The white arrows indicate the extra density between the shell and the TcB-binding domain. B) EM density maps colored according to the local resolution determined by sxlocres, showing the complete electron density and a longitudinal cut through the density map. C) Fourier shell correlation (FSC) curves of the filtered density map obtained in MERIDIEN. The average resolutions at 0.143 and 0.5 FSC criteria are indicated. D) Superimposition of segments of the final atomic model with the cryo-EM density (mesh), showing an α-helical part (left) and a β-strand region (right). E) Angular distribution of all particles used for the final reconstruction.

In addition, 3D sorting was performed to analyze if still a mixture of incomplete and complete particles is present. However, 2D sorting with ISAC already resulted in clean and homogenous particle classes and in a final resolution of 3.25 Å (0.143 criterion) (Figure 27C). Furthermore, the resolution gradient in the Yp-TcaATcaB density map is less pronounced compared to Xn-XptA1 and Mm-TcdA4 (Figure 27B). However, also the inner lying channel domain reveals the highest resolution whereas the β-sheet regions at the top and bottom of the protein are less resolved.

For the first time, detailed structural information about a *Yersinia* Tc toxin is available. The resolution of the obtained model allows to distinguish between the single subdomains of Yp-TcaATcaB. The protein can be divided into three major subunits as described for Pl-TcdA1. A pore domain, an α-helical shell and β-sheet regions (Figure 28A). The pore domain is composed of ten α-helices and surrounded by the α-helical shell, whereas the β-sheet region encloses the tip of the pore domain. Furthermore, also density corresponding to the linker domain is present in the Yp-TcaATcaB density map. The additional feature observed in the 2D class averages, connecting the shell and the pore domain, was identified as a coiled coil region protruding out of the shell domain (see chapter 5.3.2.3.1). This coiled coil domain was not resolved in the negative stain data. This could be due to the grain size of the uranyl formate surrounding the stained protein and thereby decreasing the details and limiting the resolution[95].

Apart from the coiled coil region that is a unique feature of Yp-TcaATcaB among the analyzed Tc toxins, Yp-TcaATcaB differs particularly in the β-sheet region. While comparing the Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 EM density maps, their main differences were found in the orientation of the RBDs. For Yp-TcaATcaB, the RBDs A-D are non-existent, resulting in the slim shape at the bottom of the protein.

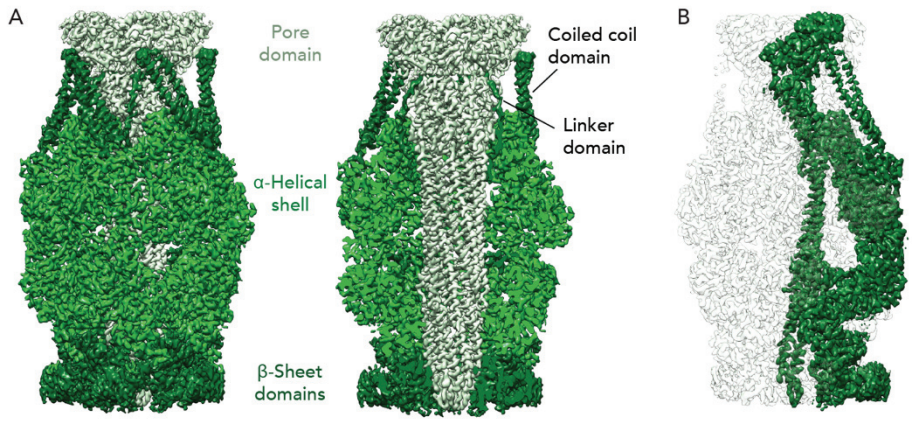


Figure 28: Cryo-EM density map of Yp-TcaATcaB. A) Pentamer density map showing the complete electron density (left) and a longitudinal cut through the density map revealing the inner lying pore domain (right). B) Pentamer with one protomer colored in dark green visualizing the monomeric structure of Yp-TcaATcaB.

The complete absence of RBDs in the Yp-TcaATcaB amino acid sequence results in a lower sequence similarity to Pl-TcdA1 of 38.1 %. Nevertheless, a homology model of Yp-TcaATcaB could be generated

as a starting point for building the final atomic structure. The software used to generate the homology model failed to model the amino acid region of the coiled coil domain since this region is not present in Pl-TcdA1 (Figure 29A). Instead, a large loop with the amino acids 620-716 was generated, which was replaced by a manually built coiled coil domain afterwards (Figure 29B). The residual of the helical shell domain as well as the pore domain of the homology model acceptably fitted the electron density map. However, the density of the β-sheet region at the periphery had a limited resolution of 6 Å and less, probably due to flexibility, which prevented the determination of the correct domain folding. Therefore, the last amino acids of the β-sheet region (residues 1139-1240) could not be built and are not included in the final atomic model.

The total number of amino acids is 2033 in the covalently linked construct Yp-TcaATcaB, consisting of the protein subunits Yp-TcaA and Yp-TcaB. Despite the differences regarding the amino acid sequence similarity and the missing receptor-binding domains, Yp-TcaATcaB has a similar domain organization compared to Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. The protein consists of an α-helical shell (1-619, 717-833, 834-1000, 1266-1487) with an additional coiled coil domain (620-716), a β-sheet region at the bottom (1001-1265), a pore domain (1536-1846) and a TcB-binding domain (1847-2033) as well as a stretched linker (1488-1535). As mentioned before, the protein itself is composed of two subunits, Yp-TcaA (1-832) and Yp-TcaB (835-2033) (Figure 29C and D). A major fraction of the α-helical shell and the coiled coil domain belongs to Yp-TcaA, whereas the entire internal channel, the β-sheet region and the lower part of the helical shell correspond to Yp-TcaB. The overall assembly of the two opposing subunits is visualized in Figure 29B. Here, it becomes evident that the 2D class averages, showing fragmented protein particles, contained particles missing the Yp-TcaA part.

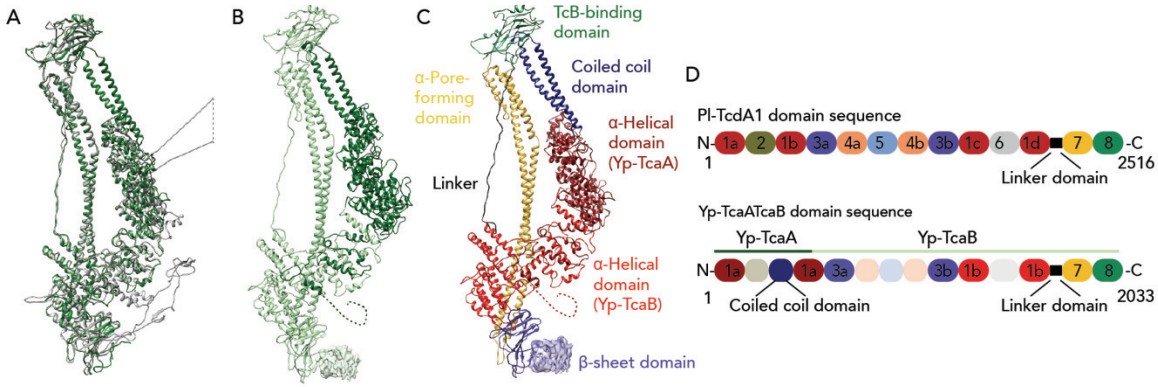


Figure 29: Atomic model of Yp-TcaATcaB. A) Overlay of the final atomic model of Yp-TcaATcaB (green) and the homologue model on a structure of Pl-TcdA1 (grey). Sequences with no or little homology are displayed as loops. B) Final model of Yp-TcaATcaB showing the two subunits of the protein Yp-TcaA (dark green) and Yp-TcaB (light green). Not resolved residues of the N-terminus (1-57) are indicated by a green dotted line and the missing residues 1139-1240 of the β-sheet neuraminidase-like domain are visualized by the electron density map of the 101 residues. C) Yp-TcaATcaB model colored according to its domain organization. D) Schematic domain organization of Pl-TcdA1 and Yp-TcaATcaB with numbering accordingly to Pl-TcdA1; 1=helical shell, 3=β-sheet domain, 7=channel, 8=TcB-binding domain and the linker domain in black. The domains missing in the sequence of Yp-TcaATcaB are faded.

As observed in the afore discussed negative stain model of Yp-TcaATcaB, this particular protein has a different shape than other TcA components. It is slimmer with a bottle-like overall shape. Comparing now the two TcAs from *Yersinia* species, Yp-TcaATcaB and YenA1A2 from *Y. entomophaga*[61], the slimmer shape of Yp-TcaATcaB results from the absence of the chitinase domains 1 and 2 (Chi1, Chi2). The proposed functions of these chi domains is to bind and degrade chitin in the peritrophic membrane of the insect host's midgut epithelial cells[62]. It was further shown, that Chi1 and Chi2 are essential for the toxic activity of Tc toxins expressed in *Y. entomophaga*[58]. The two enzymes are located at the periphery of the shell domain of the toxin complex, corresponding to the YenA1 subunit. Since Yp-TcaATcaB is secreted by a human pathogen, no chitinase domains are necessary for the hydrolysis of chitin. In fact, in the genome sequence of human pathogenic bacteria producing Tc toxins, no chitinase genome sequences are present[78]. Thus, the α-helical shell as the interaction partner for the chitinase domains differs between the two species and results in a lower sequence similarity of 25.4 % for YenA1 and Yp-TcaA. In contrast, Yp-TcaB and YenA2 have a higher similarity of 43.7 %, especially in the pore domain and the TcB-binding domain. Although the negative stain model of Landsberg et al. (Figure 5 in chapter 2.2.3.1) is in fact limited in resolution, a bump in the density that connects the upper region of the shell with the top of the channel is clearly visible. This region could correspond to the coiled coil domain. So far, this coiled coil region was not observed in any other Tc toxin. Thus, the assumption that the *Yersinia* Tc toxins have a unique coiled coil region in common is more likely, which raises the question of the function of this coiled coil domain.

5.3.2.3.1 Coiled coil domain in Yp-TcaATcaB

A coiled coil describes interacting helices with a repeat scheme of seven amino acids (a-g). The residues a and d interact with the adjacent residues a' and d' from the opposing helix and have a hydrophobic character[164]. The two helices can be parallel or antiparallel to each other resulting in two interaction clusters of the residues a and d. For Yp-TcaATcaB, an antiparallel sense of these two helices leads to a cross interaction of a with d' and d with a' (Figure 30B). The coiled coil character of the protein domain was calculated and verified using the online tool SOCKET[165] and the two coiled coil helices are colored in dark green (Figure 30A and D). Coiled coil regions fulfill various functions, ranging from complex stabilization/mediation or acting as molecular spacers up to communication between single protein domains by conformational changes[166,167]. Since the TcA component of the Tc toxin from *Yersinia* species is composed of two subunits, the coiled coil domain could have a stabilizing effect on the complex. The coiled coil domain of Yp-TcaA interacts with the TcB-binding domain of Yp-TcaB and thus generates an additional interaction site between the two subunits. Therefore, the coiled coil domain

might play an important role in the protein complex stability either in its prepore state or even during the prepore-to-pore transition.

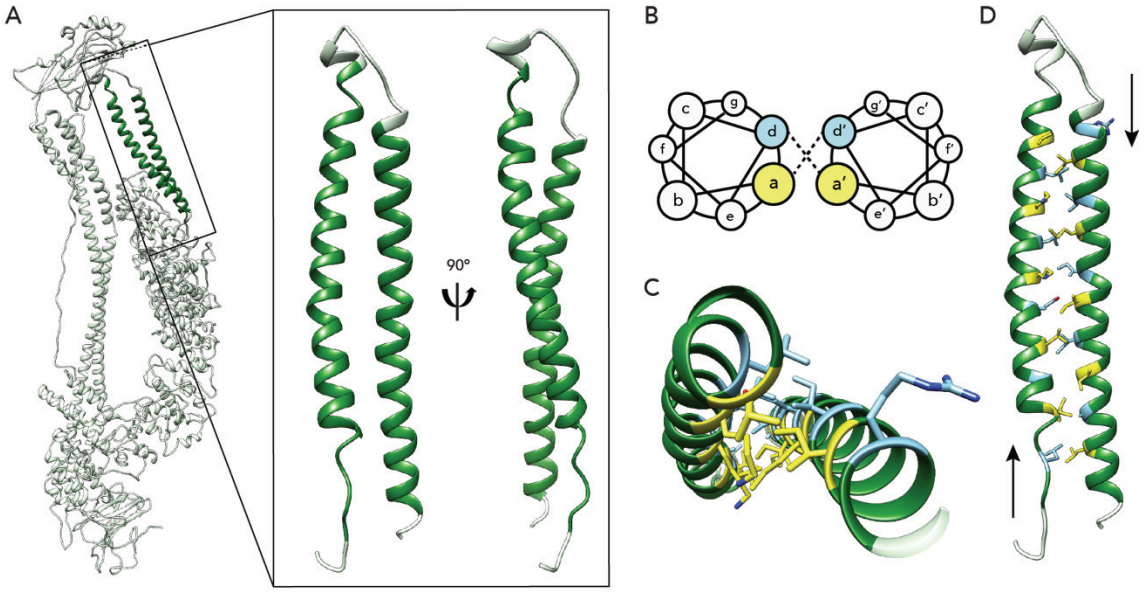


Figure 30: Characterization of the coiled coil region in Yp-TcaATcaB. A) Protomer of Yp-TcaATcaB and a close-up view of the coiled coil region highlighted in dark green. B) Scheme of the seven amino acid repeat pattern of two coiled coil helices indicating the two interacting residues a (yellow) and d (blue). The same top view and color code is applied to the Yp-TcaATcaB coiled coil helices in C). D) Yp-TcaATcaB coiled coil region showing the interacting residues a (yellow) and d (blue).

To determine whether the coiled coil domain is essential for the stability of the complex, a mutant of Yp-TcaATcaB was designed in which the coiled coil region is deleted; Yp-TcaATcaB-Δ622-714 with a linker of six amino acids (GRPSSG) to generate a short loop between the two deletion sides (Figure 31A). The expression and purification of Yp-TcaATcaB-Δ622-714 were performed in the same manner as described for Yp-TcaATcaB wild-type. However, the protein sample was less pure after the two-step purification compared to Yp-TcaATcaB wild-type purification. The chromatogram of the size exclusion chromatography is polydisperse and a strong void volume peak as well as a third peak at 17 ml are present (appendix, Figure 54). Nevertheless, compared to the size exclusion chromatography profile of the Yp-TcaATcaB wild-type protein, the main peak also runs at 12 ml. The corresponding gel after SDS-PAGE revealed some remaining contaminations in the main peak fraction.

Investigating whether the purified protein still was capable of forming intact pentamers, the sample was examined using negative stain EM. Despite some impurities in the background, it was observed on the negative stain electron micrographs, that the Yp-TcaATcaB-Δ622-714 mutant behaved the same in negative stain as the wild-type protein. The observed particle shape, representing the side view of the protein, was identical to the wild-type (Figure 31B).

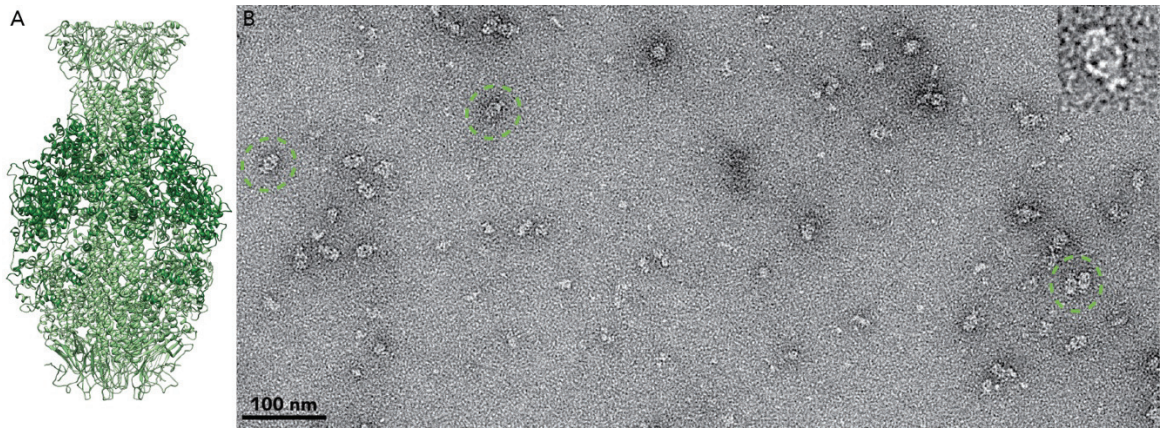


Figure 31: Yp-TcaATcaB mutant Yp-TcaATcaB-Δ622-714. A) Yp-TcaATcaB structure with deleted coiled coil region to illustrate the structure of the Yp-TcaATcaB_Δ622-714 mutant. B) Electron micrograph of the Yp-TcaATcaB_delta_622-714 sample after purification. Grids were pretreated with PLL before negative staining. Side views are highlighted with green circles and a single particle is zoomed in.

To further validate the stability of the Yp-TcaATcaB mutant in comparison to the wild-type, the melting temperatures of the mutant (52±2.7 ˚C) and the wild-type protein (54.8±0.7 ˚C) were determined with nanoDSF. The difference in the T_m is 2.8 ˚C which is not significant, taking the strong deviation of the Yp-TcaATcaB mutant (2.7 ˚C) measurement into account. If the coiled coil region would have been essential for the complex stability, less thermal stability would be expected for the complex. However, differences in structural organization or a decreased protein stability were not observed by negative stain EM and nanoDSF. Nevertheless, it has to be considered that the Yp-TcaATcaB construct has the two subunits covalently linked together, which might affect the complex stability. Thereby, the loss of the coiled coil domain might result in less destabilization of the complex due to the covalent linkage of the two subunits. Protein fragmentation, resulting in decomposition of the Yp-TcaA subunit was only observed after plunge freezing and may arise due to mechanical stress. Consequently, the effect of the coiled coil domain on the protein complex stability would need further validation experiments with the protein complex assembled of two co-expressed subunits Yp-TcaA and Yp-TcaB.

Considering the coiled coil region as a structural characteristic with different functions, the coiled coil domain in Yp-TcaATcaB could moreover induce a pH-mediated conformational change, that is important for the prepore-to-pore transition of this protein. Likewise, a pH induced dimer formation, regulated by coiled coil regions at the N-terminus, was observed for saposin B[168,169]. Moreover, for Rad50 it was shown that the coiled coil domain mediates conformational changes of the globular domains[170]. Thus, the coiled coil domain in TcAs from *Yersinia* species could be an important domain for the complex stabilization under extreme conditions or an instrument to facilitate the massive conformational change that would be necessary to enable the pore domain to penetrate a lipid membrane.

5.4 Comparison of Xn-XptA1, Yp-TcaATcaB and Mm-TcdA4 with Pl-TcdA1

After the successful structure determination of the three TcA components Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB using cryo-EM and a first overall comparison between those structures to Pl-TcdA1 from *P. luminescens*, this chapter will cover a detailed analysis and comparison regarding single functional domains as well as experiments to further elucidate the mechanism of different Tc toxin actions.

In the previous chapter, the single TcA proteins were compared on a basic domain organization level to Pl-TcdA1. Based on the electron density map variations, the outwards protruding domains, like β-sheet regions and single stretches in the helical shell, were already identified as the regions with major variances. Now comparing all four TcAs based on the atomic structure, the sequence conservation and the root mean square deviation (RMSD) values between the proteins provide more detailed information. The sequence conservation of Xn-XptA1, Mm-TcdA4, Yp-TcaATcaB and Pl-TcdA1 is depicted in Figure 32A. Highly conserved regions are colored in magenta, which are predominantly the pore forming domain and the small lobe of the α-helical shell. The TcB-binding domain has highly conserved patches going along with a low RMSD. These conserved regions also show a low RMSD value, which is plausible since the highly conserved regions adopt the same structural arrangements and therefore less variances between the different toxins. The domains previously described as exhibiting main differences are the big lobe of the α-helical shell and the β-sheet regions. These observations are in accordance with the function of those β-sheet regions, as they are supposed to act as receptor-binding domains. The Tc toxins from different species need variances in those RBDs to specifically target and interact with the host cell membranes. Only in this way, respective toxins can fulfill their host-specific function[46,66].

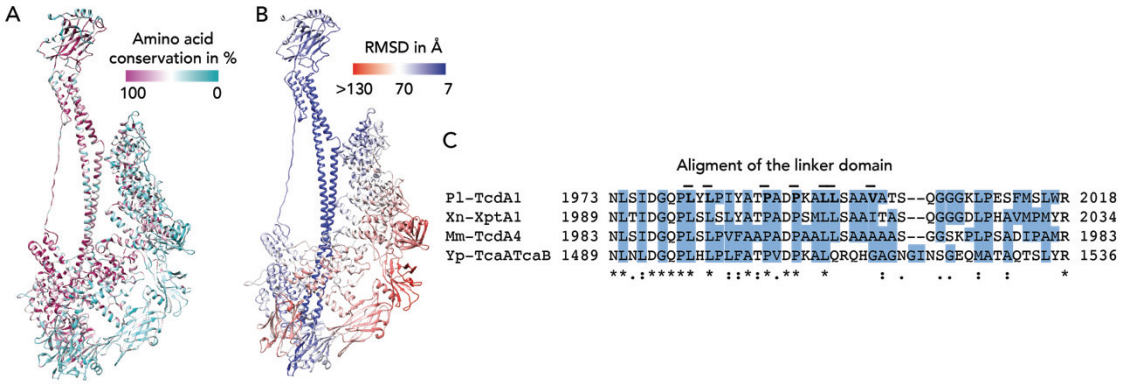


Figure 32: Conservation and mean RMSD values between the four analyzed TcAs. A) Protomer of Pl-TcdA1 colored according to the amino acid conservation between the four analyzed TcAs. Highly conserved amino acids are colored magenta and less conserved regions are depicted in turquois. B) Pl-TcdA1 protomer showing the mean RMSD value between the four TcA structures. Regions with high a RMSD value are colored in red and with low RMSD value in blue. C) Sequence alignment of the linker domain. Residues with hydrophobic character are marked in blue and the residues from Pl-TcdA1 that are involved in the interaction to the shell and the pore domain are highlighted with dashes.

A region that is highly conserved, with low structural variation, is the stretched linker domain. This domain was shown before to act as an entropic spring that drives the prepore-to-pore transition and the insertion into the host membrane[43,44] (see section 2.2.3.2, Figure 32B). In the prepore state, the linker adopts a stretched conformation and is spanned between the last helix of the α-helical shell and the pore forming domain. After the conformational change and membrane penetration, the linker is partially folded as a helix. There are five interfaces covering electrostatic and hydrophobic interactions between the folded linker and the shell, as well as the channel domain of two protomers. Here, the neighboring protomers form a pocket and stabilize the final and folded conformation of the linker[44]. Two of these five interfaces have an hydrophobic character and five residues being responsible for the interactions are highlighted in Figure 32C and are conserved between the analyzed TcAs.

In general, the conservation of the linker preserves a hydrophobic character inside all linker domains among the TcA components (Figure 32C). The presence of this conserved linker domain in all four analyzed TcAs leads to the assumption, that all toxins share the same mechanism of action. By opening of the neuraminidase-like domain, located at the membrane-facing tip of the toxin, a rearrangement of the entire shell occurs and the pore domain with its transmembrane helices is exposed. The linker contraction represents an entropic gain, which provides energy and enables the penetration of the host cell membrane[43,44].

5.4.1 Shell domain

The shell domain of the Tc toxins is the largest subdomain and contains ~80 % of all amino acids of each analyzed TcA. The functionality of the shell domain is to shield and therefore protect the inner lying pore domain from the surrounding environment. Additionally, the shell domain is responsible for the interaction of the Tc toxins with the target receptors. Furthermore, due to strong ionic interactions between the shells of each protomer, this domain stabilizes the oligomeric state of a TcA[78]. The large shell domain can be further divided into the helical part and the β-sheet part. This α-helical part can be furthermore distinguished into an inner scaffold of four α-helical domains (domain 1a-1d) for Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 and the additional coiled coil domain for Yp-TcaATcaB. In turn, the structurally more variable β-sheet domains harbor four RBDs and one neuraminidase-like domain in the cases of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. The number and kind of RBDs can vary between different organisms[43]. Thus, the Yp-TcaATcaB sequence contains only a single β-sheet domain and misses the domains corresponding to the RBDs A-D of Pl-TcdA1. The scheme of the complete outer shell of the Tc toxins explains the domain organization on the sequence level (Figure 33).

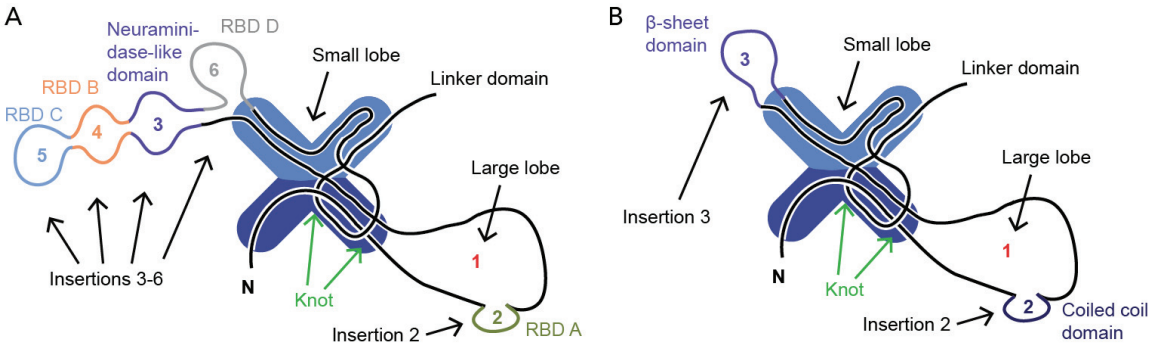


Figure 33: Organization of the shell domain of Tc toxins. A) Schematic figure of the shell organization of a TcA protomer of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. The large and small lobe have structure insertions of the domains 2-6 (RBD A-D), the neuraminidase-like domain and the molecular knot (green). B) Scheme of the shell organization for a Yp-TcaATcaB protomer. The large lobe has the coiled coil domain as insertion 2 and the small lobe contains only insertion 3 (β-sheet domain).

The helical part stretches along the entire shell domain with the small and big lobe, the RBDs and the neuraminidase like domain. As described before, the conserved domains act as a skeletal structure (small lobe, large lobe, linker, channel) that is accessorily decorated with other functional domains. Those domains can vary between different species and are in the following sections described as insertions. Likewise, insertions 3-6 (RBD B, C, D and the neuraminidase-like domain) are inserted at the small lobe, whereas the insertion 2 (RBD A) is located at the large lobe (Figure 33). Domain insertions or also deletions occur predominantly at the N- or C-terminus of proteins[171], whereas for the Tc toxins the insertions are in the middle of the α-helical domain (Figure 33). The varying number of inserted β-sheet domains as RBDs suggests an evolutionary directed deletion or insertion of the domains, aiming for optimization of the protein function. For example, Yp-TcaATcaB harbors the afore described coiled coil domain as an insertion in the large lobe of the shell domain, whereas the other Tc toxins have a receptor-binding domain instead. Altogether, the protein architecture affirms the assumption that Tc toxins evolved from a common precursor protein. Afterwards, during evolutionary development and specification, variances within the Tc toxin family were established concerning structure and function related mechanisms[172].

5.4.1.1 α-helical shell with molecular knot

In all TcAs, the inner scaffold of the shell is composed of an α-helical domain with a structurally highly conserved fold (Figure 34, domains 1a-1d). The folding and domain arrangement in the helical shell of Pl-TcdA1 were analyzed in detail before (by Dr. Dominic Meusch)[78]. Structural alignment of Pl-TcdA1 with Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB revealed that all toxins adopt the same domain folding of the α-helical shell. The shell can be divided into a large and a small lobe perpendicular to each other and arranged in a L-shape (Figure 34). The small lobe can be further subdivided into pseudo-2-fold-

symmetrical subunits in a X-shaped structure. Here, small differences are present in Yp-TcaATcaB, which are visualized in Figure 35F. The last 57 N-terminal residues were not resolved in the EM-density map of Yp-TcaATcAB and thus, the N-terminus is not present in the structure of the small lobe. For Pl-TcdA1, Xn-XptA1 and Mm-TcdA4, the N-terminus is stabilized by the receptor-binding domains, which are not present in Yp-TcaATcaB (Figure 29). Moreover, one helix of the small lobe, which is also part of the molecular knot, is a more flexible loop in Yp-TcaATcaB (Figure 35D-F, blue helix).

In contrast, the large lobe contains two different pseudo-repeats (Figure 34, pseudo-repeat 1 in red - orange and pseudo repeat 2 in violet - rose), which are composed of several helix-loop-helix motifs. The pseudo-repeat 1 shows the same fold and organization in all four TcAs, whereas the pseudo repeat 2 domain shows two insertions only in Yp-TcaATcaB, namely the enlarged loop and the coiled coil domain (Figure 34B).

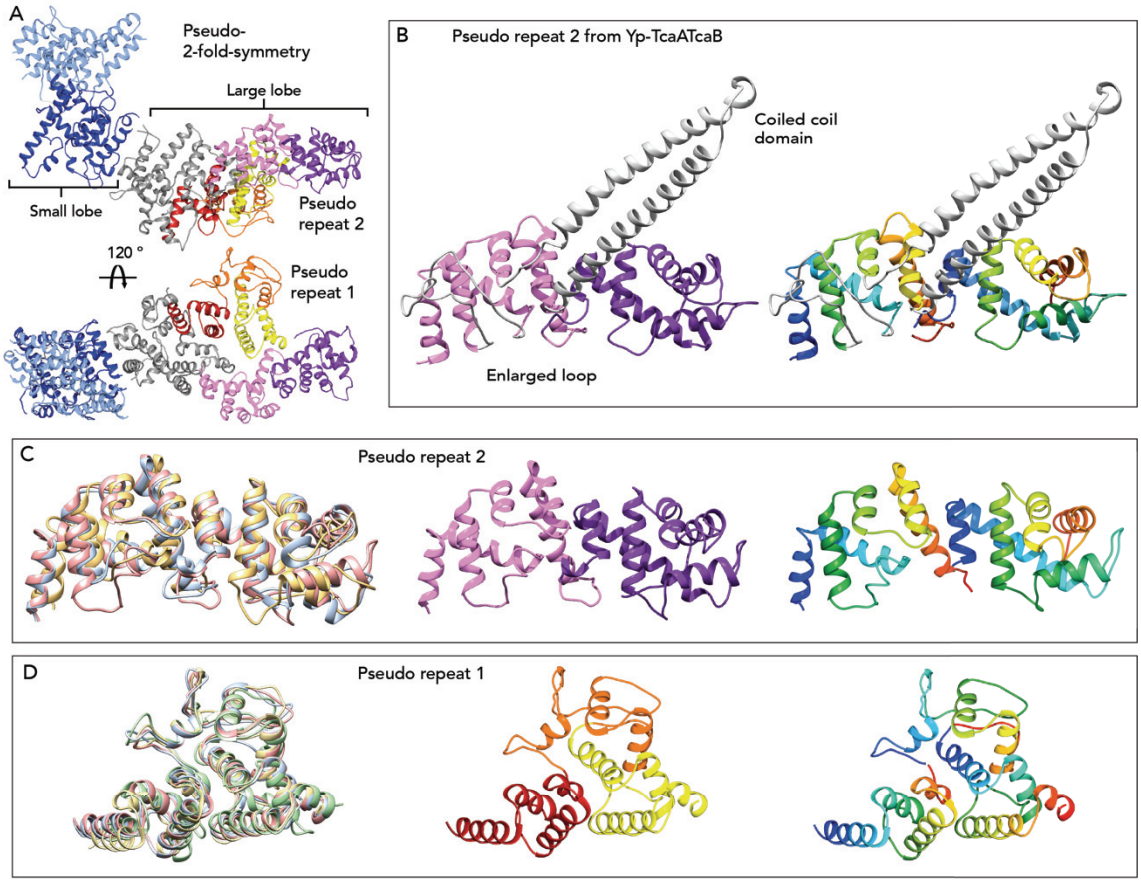


Figure 34: The organization of the α-helical shells in Tc toxin A components. A): The α-helical shell of a TcA protomer shown for Pl-TcdA1 is divided into a small and a large lobe. A pseudo-2-fold symmetry (light and dark blue) is present in the small lobe, which is arranged in a X-like manner. Two subunits of the large lobe have pseudo-repeats; pseudo-repeats 1 and 2 (1=red-orange-yellow and 2=rose-magenta). B) The pseudo repeat 2 in Yp-TcaATcaB shows the same overall fold, except the insertion of the coiled coil domain as well as an enlarged loop (grey). C) The pseudo repeat 2 includes two repeats that are depicted for Pl-TcdA1 in rose and magenta in the middle and in rainbow colors on the right. An overlay of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 is depicted on the left. D) On the left is an overlay of the pseudo repeat 1 of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. The pseudo repeat 1 includes three repeats that are depicted in red, orange and yellow in the middle and in rainbow colors on the right for Pl-TcdA1.

The small lobe with the pseudo-2-fold-symmetry contains an additional molecular feature, a molecular 3_1 trefoil knot[148], that was discovered before in Pl-TcdA1[78]. The knot tightly connects the N-terminal part of the protein (residues 21-157 in Pl-TcdA1) with the last domain of the α-helical shell (residues 1873-1972 in Pl-TcdA1) before the linker domain starts (Figure 35C). A 3_1 trefoil protein knot is the most simple and common knot type[173]. The online platform Knotprot2.0 is a database for the deposition and detection of molecular knots in proteins. Based on the statistics of the Knotprot2.0 database, most of the knot core lengths are between 50 and 270 amino acids[174]. In the case of Pl-TcdA1, the knot core includes 1852 amino acids, whereas the N-terminal knot tail contains 89 and the C-terminal knot tail 562 amino acids (appendix, Figure 55). Indeed, protein knots mostly occur in smaller proteins or enzymes such as methyltransferases and carbonic anhydrases[175]. The folding mechanism of a knotted protein was shown to be much more complex and the folding rate of knotted proteins is decreased[173,176]. Therefore, it is highly interesting that a protein knot in a protein of ~280 kDa (monomeric mass) exists and is present in all the analyzed Tc toxins (Figure 35D-F).

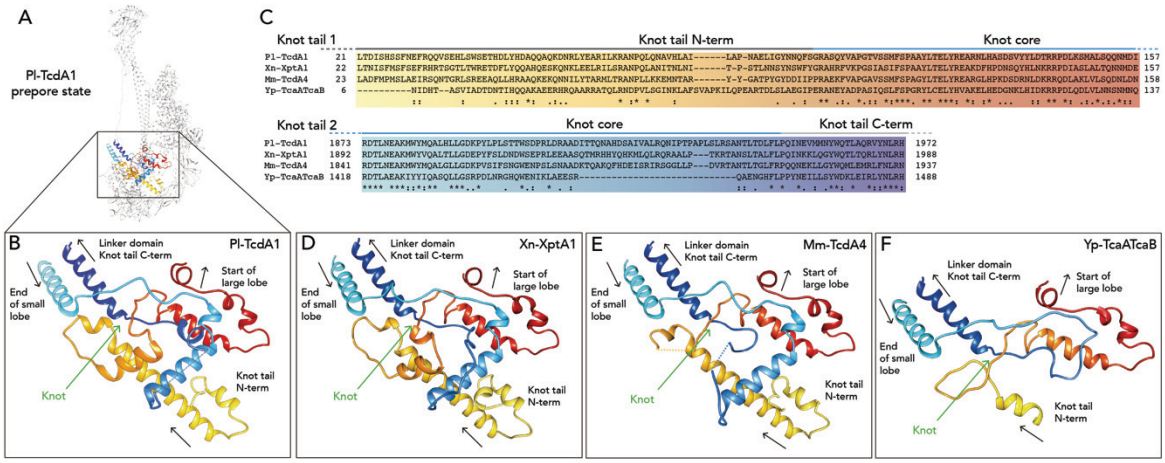


Figure 35: A trefoil 3_1 protein knot is present in all analyzed TcAs. A) Pl-TcdA1 prepore protomer with the knot structure highlighted in rainbow colors and B) a close-up view of the protein knot structure is presented. The dashed line represents a missing loop (residues 1933-1938) and arrows indicate the sense of sequence. The protein knot tail 1 is colored in a yellow-to-red gradient and the protein knot tail 2 is colored in a light-to-dark blue gradient. C) Sequence alignment of the two protein knot tails. The color gradient is identical to the lower panel and the location of the knot tail N-term, the knot core as well as the knot tail C-term are depicted like in Figure 55 in dark grey, blue and light grey, respectively. D-F) Close-up view of the protein knot of Xn-XptA1 (D), Mm-TcdA4 (E) and Yp-TcaATcaB (F), respectively. All protein knot structures are depicted in the same color gradient.

Up to now, the function of protein knots is still under discussion. Several studies indicate that protein knots have a stabilizing effect on proteins[177], particularly during conformational changes[178]. Interestingly, the protein knot is still present after the massive conformational change of Pl-TcdA1 from the prepore to the pore state[44]. The knot tightly connects the N-terminal and the C-terminal part of the shell. This and the location of the knot indicate a stabilizing effect of the molecular knot on the toxin complexes, especially during the conformational transition. In accordance to that, the linker domain is directly

adjacent to the protein knot tail 2 (Figure 35, Figure 36). Consequently, the protein knot might also stabilize the stretched linker domain in the prepore state of the TcAs, making the protein knot an important conserved structural feature for the stability and function of all TcAs.

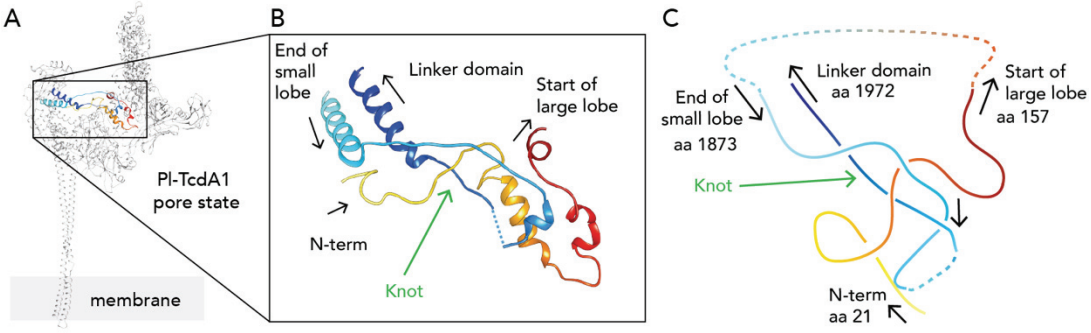


Figure 36: A trefoil 3_1 protein knot is present in the Pl-TcdA1 pore state. A) The knot structure is highlighted in rainbow colors in a Pl-TcdA1 pore protomer (PDB-ID 5LKH and 5LKI) and a close-up view of the protein knot structure is presented in B). The dashed line represents a missing loop (residues 1914-1946). C) Simplified structure of the protein knot. The arrows indicate the sense of sequence and the amino acid numbering corresponds to Pl-TcdA1.

5.4.1.2 β-sheet region

As described and discussed previously in chapter 5.3.2, the β-sheet domains display the major sequence differences between the four TcAs, which becomes evident with the high RMSD values and the low sequence conservation (Figure 32 and Table 4). While comparing the sequence homology of the single β-sheet domains from Pl-TcdA1 to the domains of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB, it becomes obvious that the sequence similarity between the two TcAs from entomopathogenic bacteria differ less compared to the sequence similarity of Pl-TcdA1 and Mm-TcdA4 from a human pathogenic bacterium. In general, the RBD D and the neuraminidase-like domain of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 have a higher sequence similarity compared to RBD A-C. The sequence of the neuraminidase-like domain is fairly similar with a similarity above 40 % in these three TcAs. In contrast, Yp-TcaATcaB has only one β-sheet domain, which has a low sequence similarity of 20.1 % compared to the Pl-TcdA1 neuraminidase-like domain.

Table 4: Sequence similarity of the single β-sheet domains between different TcAs.

Sequence similarity	RBD A	RBD B	RBD C	RBD D	Neuraminidase-like domain
Pl-TcdA1 vs. Xn-XptA1	27.3 %	22.2 %	18.0 %	38.9 %	46.2 %
Pl-TcdA1 vs. Mm-TcdA4	23.1 %	22.0 %	20.2 %	29.2 %	44.0 %
Pl-TcdA1 vs. Yp-TcaATcaB	-	-	-	-	20.1 %

Within this chapter, RBD A, B and D as well as the neuraminidase-like domain will be compared and analyzed by their structural composition. The RBD C was not resolved in Xn-XptA1 and Mm-TcdA4 and thus, the structure of the RBD C will not be discussed. Furthermore, the Yp-TcaATcaB toxin has only one β-sheet domain at the bottom where the neuraminidase-like domain is located and the RBD A-D are absent in the genomic sequence of the toxin. Consequently, the RBDs will only be discussed for Xn-XptA1 and Mm-TcdA4.

RBD A, B and D of Xn-XptA1 and Mm-TcdA4

The β-sheet domains were analyzed regarding their sequence and their topology, and were compared to the respective domains of Pl-TcdA1. To date, the β-sheet domains of Pl-TcdA1 are suggested to be the receptor-binding domains, but this has yet to be proven. However, the RBDs of Pl-TcdA1 have a high structural homology to different types of receptor-binding domains as for example from anthrax and diphtheria, thus supporting the functionality of the β-sheet domains as receptor-binding domains[12,43,86]. The receptor-binding domains of Pl-TcdA1 exhibit an immunoglobulin (Ig)-like β-sandwich fold of two β-sheets with antiparallel β-strands[43]. Moreover, RBD A is a structural homologue to myeloid differentiation factor 2 (MD-2), which is an interaction partner of toll-like receptors (TLRs)[78]. For RBD B, C and D, structurally homologue receptors were found in the family of immunoglobin (Ig) receptors. Interleukin (IL) and interferon (IFN) receptors also belong to this family and have the same topology of eight antiparallel β-sheets exhibiting a β-sandwich fold as observed for RBD C, RBD B and D, respectively [179,180].

The sequence alignments of RBD A, B and D from Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 reveal a low sequence homology between the three toxins (Table 4 and Figure 37, left panel). Interestingly, the structural alignment of the single RBD domains demonstrate the same topology of the domains (Figure 37, middle panel). RBD A, B and D of Xn-XptA1 and Mm-TcdA4 have also an Ig-like β-sandwich fold with antiparallel β-strands, which is therefore consistent throughout the toxins from different species. Comparing the overlay of the domains, the orientation of the domains, especially for Mm-TcdA4 RBD A and B, varies. Although they have the same folding pattern, the structural arrangement within the toxin protomer is different for Mm-TcdA4 RBD A and B (Figure 37, right panel). This could lead to a distinct function of the RBD concerning the substrate specificity. Due to another orientation of the domain, different residues and secondary structure elements become accessible. In contrast, the RBD D shows the same secondary structure arrangement for the toxins in the overlay and the structural alignment (Figure 37, lower panel). This domain is located between the neuraminidase-like domain and the shell domain and therefore is less flexible. Here, the conservation is slightly higher compared to RBD A and B.

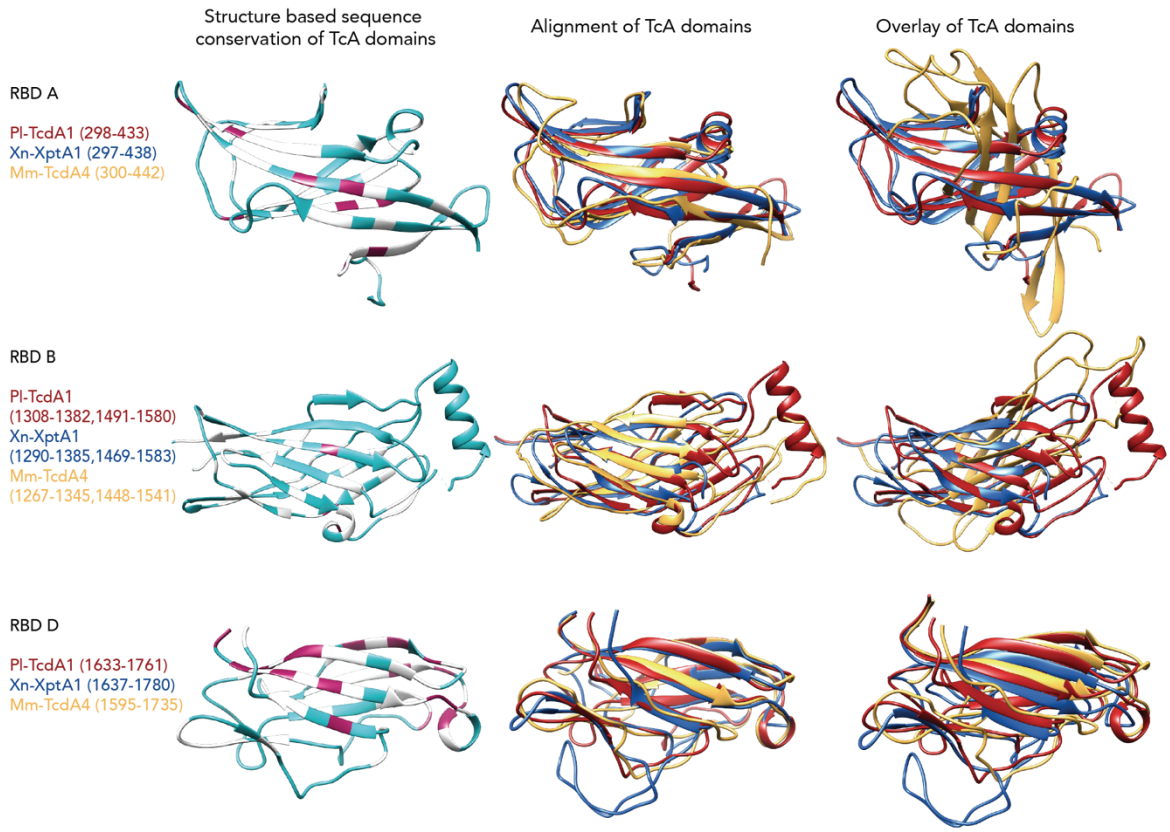


Figure 37: Structural comparison of RBD A, RBD B and RBD D. The left panel shows the conservation of the respective domain of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. Highly conserved residues are depicted in magenta and low conserved residues in turquoise. The central panel shows the structural alignment of the single domains with Pl-TcdA1 in red, Xn-XptA1 in blue and Mm-TcdA4 in yellow. On the right, these domains are shown as overlay.

A high sequence similarity of above 40 % most often results in high structure resemblance. However, this is not the case for the opposite; a low sequence similarity results not simultaneously in a low structure resemblance[181]. One example for low sequence conservation but structural similarity can be found in nicotinic receptors[182,183]. These receptors have a pentameric composition and function as ion channels, which are regulated by nicotinic ligands. They are highly similar in their topology but share only a sequence similarity of less than 33 %, although the receptors belong to the same family and interact with the same ligand. The RBDs from the different Tc toxins have a similar topology, whereas the sequence shows no regular motif of conservation. Consequently, these domains could all function as receptor-binding domains but target different receptors with their distinct sequences. Taking the composition and also the domain organization of the shell into account, this suggests that all Tc toxins were evolutionary equipped with receptor-binding domains. The specificity of the receptor-binding domains evolved by deletion or exchange of complete domains.

Neuraminidase-like domain

As described for Pl-TcdA1, the domain located at the lower bottom of the TcA represents a neuraminidase-like domain similar to virus neuraminidases[43], that belong to the family of sialidases. These enzymes hydrolyze sialic acid from sialic glycoproteins[184]. Sialic acid linked to glycoproteins and gangliosides is present at the membrane of various cell types. Viruses or toxins interact with the glycoproteins and use them as receptors, for example the influenza virus or the cholera toxin from *Vibrio cholerae*[185-187]. Therefore, with the functionality of a neuraminidase, the domain in Pl-TcdA1 could be important for hydrolyzing the sialic acid of glycoproteins in order to make the receptor accessible for the receptor-binding domains. Furthermore, a membrane bound glycoprotein could be a mediator assisting the protein, in this case Pl-TcdA1, to be located and stabilized at the membrane.

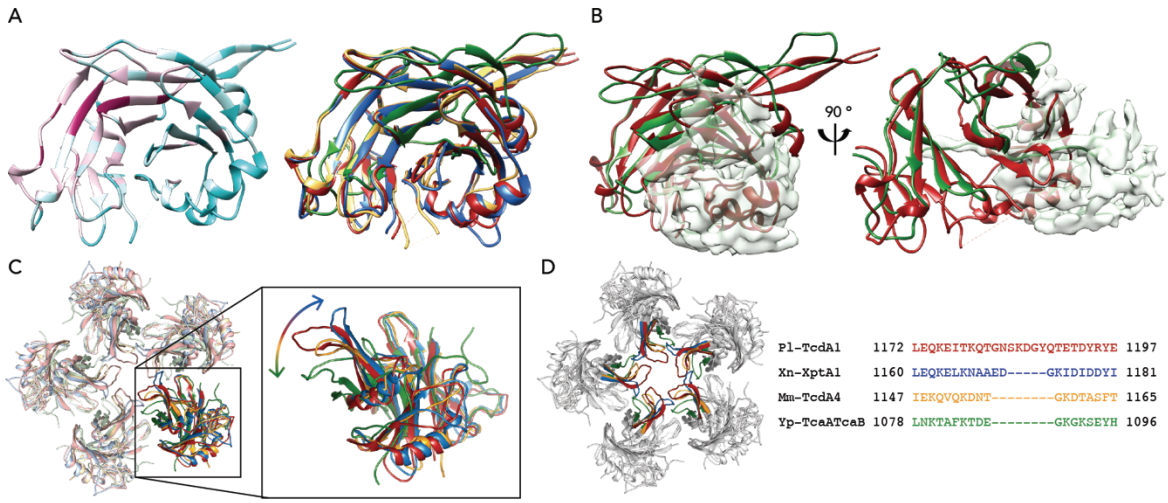


Figure 38: Neuraminidase-like domain. A) Sequence conservation of the neuraminidase-like domain with highly conserved regions in magenta and low conservation depicted in turquoise. The right panel shows the structural alignment of the domain for all four TcAs with respective colors, Pl-TcdA1 in red, Xn-XptA1 in blue, Mm-TcdA4 in yellow and Yp-TcaATcaB in green. B) Structural alignment of the Pl-TcdA1 and Yp-TcaATcaB neuraminidase-like domain. Missing residues of Yp-TcaATcaB (1139-1240) are visualized by the EM density map. A front view (left) and a side view, rotated by 90°, are shown (right). C) Pentamer of neuraminidase-like domain of all four TcAs with respective colors. One protomer is zoomed in showing the different orientation of the flexible loop in the center of the pentamer. The arrow indicates the movement between the different TcAs and is colored accordingly. D) Pentamer of the neuraminidase-like domain highlighting the flexible loop, and a sequence alignment showing the different lengths of the loop region in respective colors.

The corresponding neuraminidase-like domains in Xn-XptA1 and Mm-TcdA4 share a higher sequence homology with Pl-TcdA1 neuraminidase-like domain (46.2 % between Xn-XptA1 and Pl-TcdA1 and 44.0 % between Mm-TcdA4 and Pl-TcdA1, Table 4). Thus, between the three TcAs Pl-TcdA1, Xn-XptA1 and Mm-TcdA4, this domain is the β-sheet domain with the highest conservation and they also exhibit the same topology (Figure 38A). In contrast, within Yp-TcaATcaB, the neuraminidase-like domain is not entirely conserved. This is reflected by the low sequence homology of 20.1 % between the domain in Pl-TcdA1 and Yp-TcaATcaB. Moreover, the residues 1139-1240 of this domain could not be modeled

due to the less resolved density map in this area. For a better visualization, an alignment of only the Pl-TcdA1 and Yp-TcaATcaB neuraminidase-like domains and the density of the 101 missing amino acids of Yp-TcaATcaB are depicted in Figure 38B. Here, it can be observed that this β-sheet domain of Yp-TcaATcaB exhibits only partially the same folding as the neuraminidase-like domain of the other TcAs (Figure 38A and B). This difference leads to the distinct shape of pentameric density map with a foot-plateau only present in Yp-TcaATcaB (Figure 28).

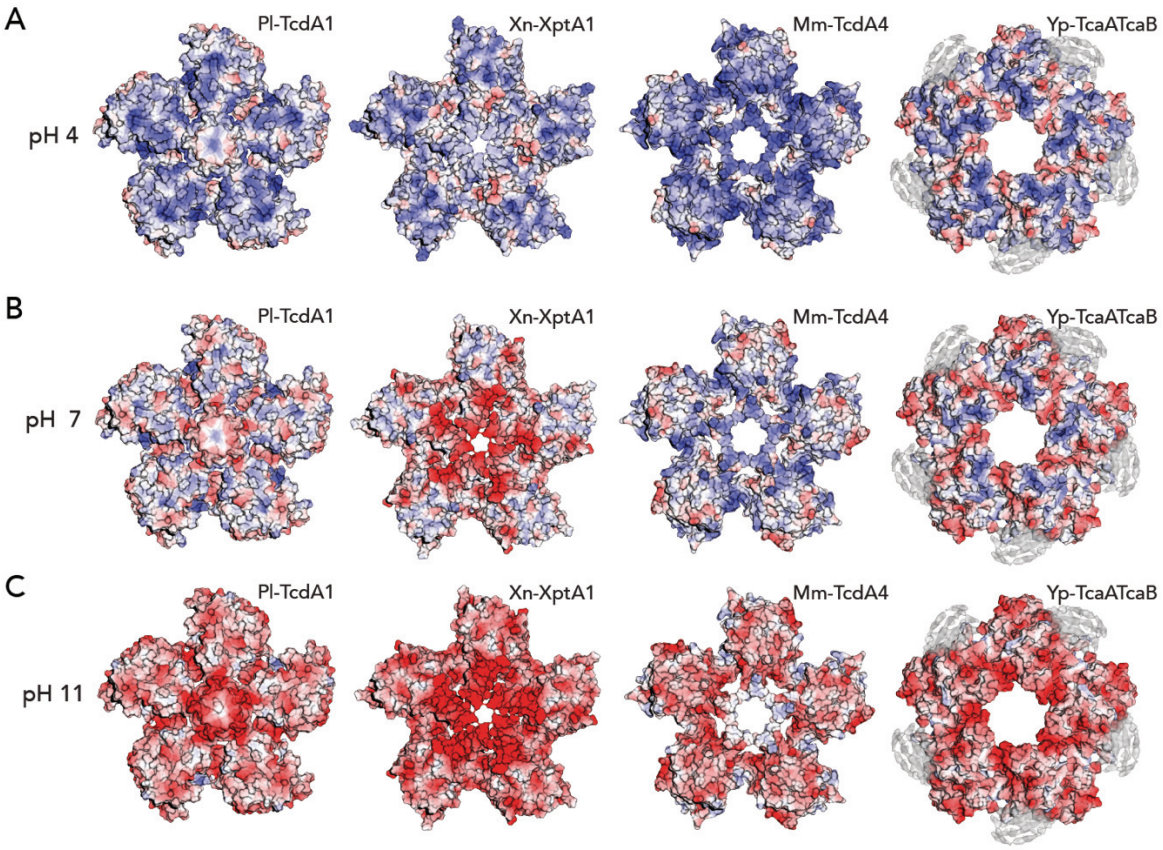


Figure 39: Electrostatic Coulomb potential of the neuraminidase-like domains at different pH values. A-C) Surface electrostatic Coulomb potential of the neuraminidase-like domain of each TcA at different pH values (bottom view). Positively charged (14 kcal/mol) and negatively charged (-14 kcal/mol) residues are colored in blue and red, respectively. Surface electrostatic Coulomb potential at pH 4 A), pH 7 B), and pH 11 C) are shown. For Yp-TcaATcaB, the EM-density map of the not modeled 101 residues of the β-sheet domain is depicted in grey.

Previous studies suggest that the neuraminidase-like domain, that encloses the tip of the channel domain, functions as an electrostatic lock[43]. It is also known that pH-dependent mechanisms are present in different type of proteins as for example in the PFT anthrax or in G-protein-coupled receptors. Both proteins undergo conformational changes upon acidic pH, which results in protonation of amino acids and subsequently in charge repulsions of the electrostatic interactions[11,188]. The neuraminidase-like domain from Pl-TcdA1 forms a homopentamer that is closed in the center and therefore locks the translocation channel inside the shell (Figure 39, left panel). While at neutral pH it is closed, at basic or

acidic pH the lock opens due to repulsions of charged amino acids (Figure 39, left panel)[43]. The pH-dependent opening of the lock was assumed to be one initial step for triggering the prepore-to-pore transition of Pl-TcdA1[43]. In contrast to Pl-TcdA1, the shell of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB is not completely closed at the bottom, most prominently evident in Yp-TcaATcaB (Figure 39). This is due to a flexible loop per subunit, which is five to eight residues longer in Pl-TcdA1 than in Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB, respectively (Figure 38C and D). However, the opening is still too narrow to allow the pore domain to pass through and thus, an opening of the shell is required for the transition into the pore state. Comparing the electrostatic Coulomb potential of the bottom surface of the four TcAs at pH 7, a neutral charge is present in case of Pl-TcdA1 and Yp-TcaATcaB . In contrast, Xn-XptA1 is negatively charged while Mm-TcdA4 is slightly positively charged (Figure 39B). The surface electrostatic potentials at pH 4 show a dominance of positive charges in all four TcAs, which is most prominent in Mm-TcdA4 (Figure 39A). However, at pH 11 the tip of the shell of Mm-TcdA4 is uncharged, while it is strongly negatively charged in Pl-TcdA1 and Xn-XptA1 (Figure 39C). Thus, both TcAs from insect pathogens show strong charge repulsions at pH 11 that could result in an opening of the shell and pore formation[44]. This is in accordance with a potential intoxication pathway via the insect midgut, where pH values of 11-12[189] enable the prepore-to-pore transition. Contrary, the same intoxication pathway is not possible in vertebrates, where prepore-to-pore transition in the acidic endosomal lumen is the only possible pH-dependent pore formation. For Yp-TcaATcaB, the charge repulsion is less pronounced at pH 4 but strong at pH 11, although it was shown that the *Y. pseudotuberculosis* Tc toxins have no oral toxicity against *Manduca sexta* larvae[66]. This excludes a potential intoxication pathway via the insect midgut. In summary, the electrostatic lock was thought to be at the tip of the neuraminidase-like domain. However, an already opened central region of the neuraminidase-like homopentamer contradicts this hypothesis in case of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. Moreover, the charge profiles at various pH values differ and are less prominent for Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB compared to Pl-TcdA1. In addition, Xn-XptA1 and Mm-TcdA4 reveal already at pH 7 a dominance of negative or positive charges, respectively, which does not result in repulsion and shell opening. Both proteins exhibit the bell-shaped pentameric prepore state at neutral pH (Figure 17). Conclusively, the neuraminidase-like domain does not seem to act as a conserved electrostatic lock solely responsible for the pH-induced prepore-to-pore transition of the TcAs. To unravel the mechanism of the shell opening and prepore-to-pore transition, further studies including the analysis of an electrostatic conserved interaction as well as the pH dependency of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB were performed and will be discussed in the next chapters.

5.4.1.2.1 *Conserved ionic interactions*

The central loop of the homopentameric neuraminidase-like domain is not conserved between the different TcAs and thus, is not a conserved feature responsible for the pH-dependent prepore-to-pore transition as it was observed for Pl-TcdA1. Therefore, I investigated whether conserved ionic interactions outside of the neuraminidase-like domain could additionally mediate the pH-induced opening of the shell. Generally, histidine is a pH-sensitive amino acid that is responsible for various pH-dependent mechanisms in different type of proteins[190-192]. One example is the gating mechanism of the β-barrel forming OmpG from *E. coli*. Two histidine residues act as pH sensors that repel each other upon acidification and are moreover the initial step of unzipping a hydrogen bond-network, which results in closing of the channel entrance[190]. In addition, amino acids such as glutamate or arginine can mediate conformational changes in proteins upon pH-shift that result in electrostatic repulsions of these residues. In the context of an endosomal pathway of these toxins, two particular pH-dependent residue pairings lead to charge repulsions upon pH shift from neutral to acidic pH. These residue pairings are either between anionic-anionic or his-cationic residues[193]. However, no conserved interactions were present in the TcAs that fulfill the two residue pairing categories. Though, three possible interaction pairs were identified that do not belong to the described residue pairing categories. These interaction pairs are conserved in all four TcAs and imply either glutamate (Glu) or aspartate (Asp) and arginine (Arg) residues: Glu158 and Arg1873, Asp965 and Arg1971, and Glu1086 and Arg1166 (residue numbering of Pl-TcdA1, Figure 40). In all cases, the interaction occurs between two adjacent protomers of a TcA. The interacting residues are in close spatial proximity in the prepore state, whereas in the pore state none of these interactions is possible. This is due to the enlarged distance after the conformational change in the Pl-TcdA1 pore structure (distance changes of the respective charged groups from 3.8 to 9.8, 3.5 to 5.8 and 3.9 to 25 Å, respectively). The interaction side 1 is located close to the knot region with Glu158 and Arg1873 is located at the C-terminal part of the shell domain (Figure 40A). Also the interaction side 2 is formed by residues of the shell domain in close proximity to interaction side 1 (Figure 40B). The third interaction takes place between residues of the shell domain (Glu1086) and the neuraminidase-like domain (Arg1166) or rather the β-sheet region of Yp-TcaATcaB (Figure 40C).

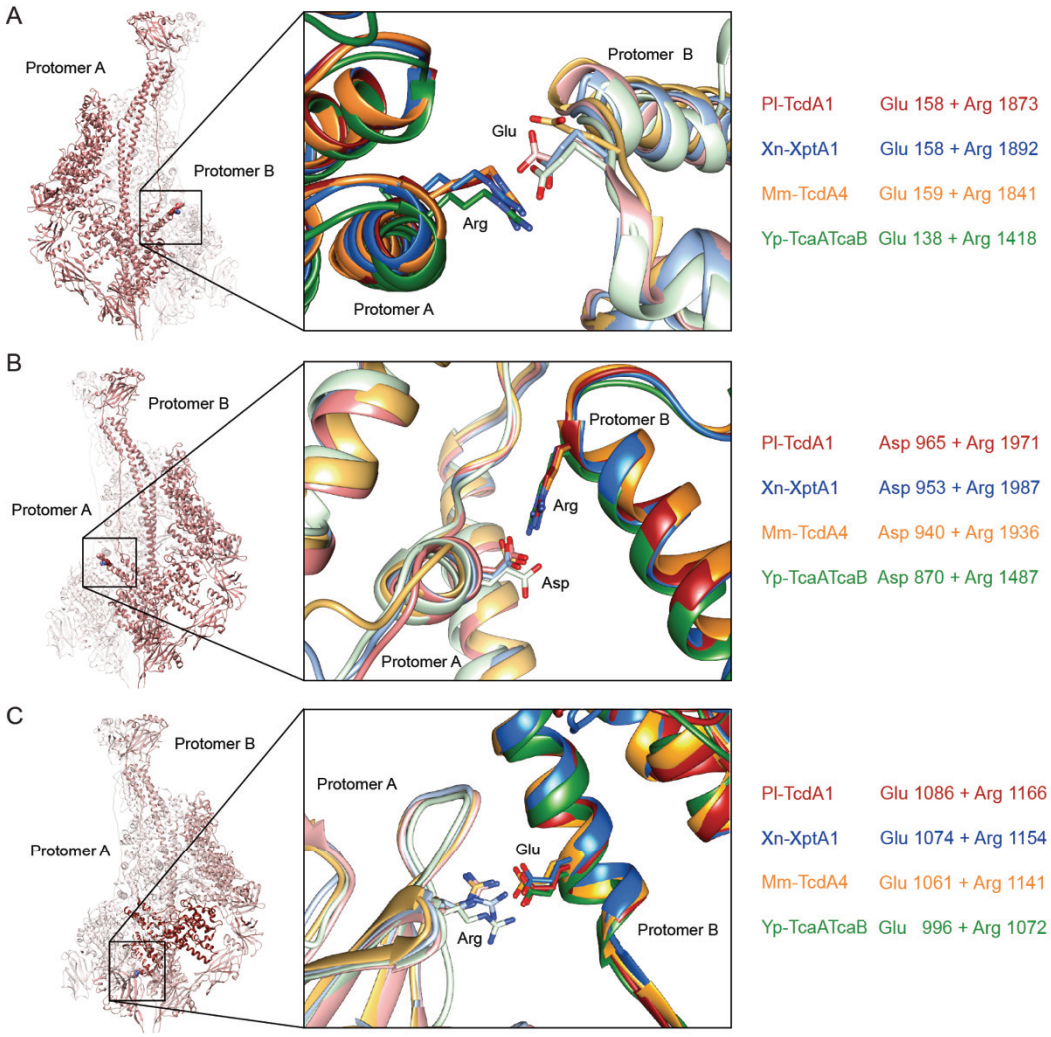


Figure 40: Conserved ionic interaction pairs between two protomers in Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. A-C): The location of the interaction of two Pl-TcdA1 protomers is shown on the left. On the right, close-up views of each interaction are presented for Pl-TcdA1 (red), Xn-XptA1 (blue), Mm-TcdA4 (orange) and Yp-TcaATcaB (green) are a presented. A) Interaction of Glu (protomer A) with Arg (protomer B). The residue distance is 3.8 Å in the prepore and 9.2 Å in the pore state of Pl-TcdA1. B) Interaction of Asp (protomer A) with Arg (protomer B) with a residue distance of 3.5 Å in the prepore and 5.8 Å in the pore state of Pl-TcdA1. The interacting residues belong to the α-helical shell domain. C) Interaction between residues belonging to the small lobe as well as the neuraminidase-like domain, Arg (protomer A) with Glu (protomer B). The residue distance of Pl-TcdA1 is 3.9 Å in the prepore and 25 Å in the pore state of Pl-TcdA1.

To analyze whether these interactions have an impact on the prepore-to-pore transition or the protein stability in general, three mutants of Pl-TcdA1 were designed, in which the respective interaction residues are mutated to alanine (Pl-TcdA1-E158A-R1873A, Pl-TcdA1-D965A-R1971A and Pl-TcdA1-E1086A-R1166A). The mutational studies were only performed for Pl-TcdA1, since the prepore and the pore state have already been characterized and conditions for the prepore-to-pore transitions are known. The mutants were expressed and then purified by Ni-NTA affinity and size exclusion chromatography as described in chapter 4.1.3. Samples of cell lysis and Ni-NTA purification were analyzed with SDS-PAGE and the peak fraction of the size exclusion chromatography was visualized with negative stain EM (Figure 41A and B, fraction after size exclusion (SE) is highlighted with a star).

The two mutants with suppressed interaction side 1 and 2 showed a comparable expression and purification behavior. Although protein is present in the pellet fraction after ultracentrifugation and also in the flow through from the Ni-NTA beads, the majority of the protein of interest could be eluted from the Ni-NTA beads with increasing imidazole concentration. Also in negative stain EM, the characteristic bell-shaped particles could be observed during negative stain EM for Pl-TcdA1-E158A-R1873A and Pl-TcdA1-D965A-R1971A at pH 8 (Figure 41A and B). Moreover, these two mutants were in parallel incubated with nanodiscs and dialyzed against a buffer of pH 11 to analyze whether the mutants behave like the wild-type Pl-TcdA1 under these conditions and rearrange from the prepore to the pore state[44]. Indeed, both mutants incorporated into the nanodiscs with the pore state. Thus, the two interaction sides have no effect on the proteins ability to undergo the conformational change to the pore state and the proteins stability. Contrarily, the mutant harboring the third interaction side, Pl-TcdA1-E1086A-R1166A, showed a weaker purification pattern. The protein containing fractions after imidazole elution from the Ni-beads include more contaminations (Figure 41C). Nevertheless, these fractions were pooled and loaded onto a size exclusion column. The SDS-gel after size exclusion chromatography revealed the presence of many contaminations and comparable low amount of protein of interest (Figure 41C, middle panel).

The main peak fraction (indicated with a star) was analyzed with negative stain EM. The pentameric Pl-TcdA1-E1086A-R1166A protein could be seen in negative stain EM but was accompanied by contaminations. Interestingly, some of the protein adopted the pore state already at pH 8 (Figure 41C, dashed line). In contrast, the mutants of interaction side 1 and 2 were only detectable in the prepore state at pH 8 and the conformational change to the pore state was solely induced after pH shift to pH 11. This suggests that the mutant of interaction side 3 was destabilized by preventing the interaction between Glu1086 and Arg1166 and thus, enabled the opening of the shell and induction of the pore state already at pH 8. This would also explain the weak expression and insufficient purification behavior. The toxin in its active state, the pore state, could already harm *E. coli* cells during expression. Furthermore, without any additional use of detergents during purification, the pore state leads to aggregation due to interaction with the exposed hydrophobic transmembrane region of the channel. After pH shift to pH 11, the protein is unstable and thus, disassembles or degrades, which is visible in the negative stain micrograph (Figure 41C, lower right image). No prepore or pore state of the protein was detected. Consequently, the residues Glu1086 and Arg1166 seem to have a high impact on the protein's stability and on the interaction between the protomers. This goes together with the measured unfolding temperature from nanoDSF experiments. Whereas the mutants Pl-TcdA1-E158A-R1873A and Pl-TcdA1-D965A-R1971A show a comparable T_m to the Pl-TcdA1 wild-type (59.9 °C and 57.8 °C compared to 61.8 °C, respectively), the mutant Pl-TcdA1-E1086A-R1166A has a T_m of 42.6 °C, which is almost 20 °C lower than the T_m of the wild-type (appendix, Figure 56A). However, considering the

amount of contaminations present in the sample, the obtained melting temperature for the mutant Pl-TcdA1-E1086A-R1166A is most likely affected by the contaminants, since the method of nanoDSF uses the intrinsic protein fluorescence. Despite that and only considering the negative stain studies with the three mutants, the stability of the toxin complex is tremendously decreased by the two mutations at position Glu1086 and Arg1166.

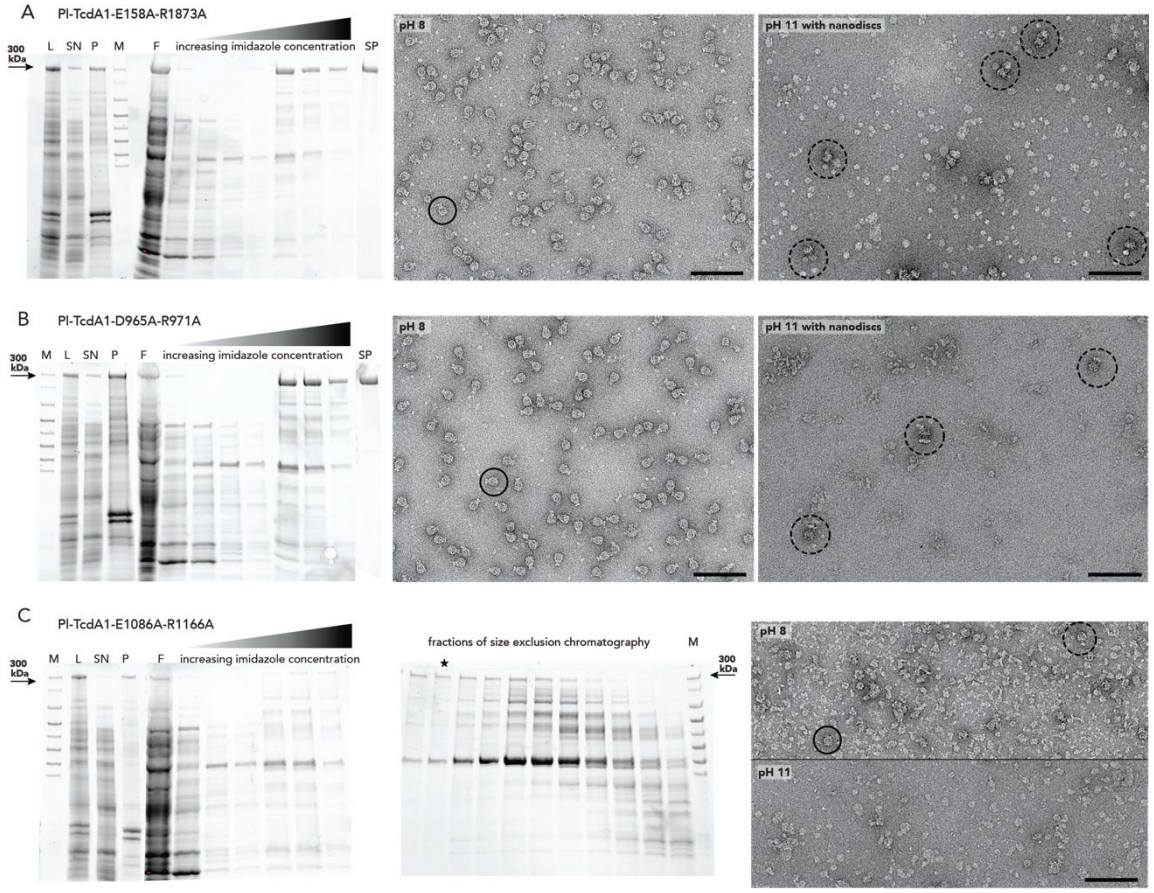


Figure 41: Alanine mutants of Pl-TcdA1 reveal differences in protein stability. A) Pl-TcdA1 mutant with Glu158Ala and Arg1873Ala variations. B) Pl-TcdA1 variant with Asp965Ala and Arg1971Ala. C) Mutant of Pl-TcdA1 with Glu1086Ala and Arg1166Ala. Each panel shows an SDS-gel after Ni-NTA purification with M = marker, L = lysate, SN = supernatant, P = pellet, fractions with increasing imidazole concentration ranging from 5–150 mM, SP = size exclusion chromatography peak fraction (only for (A) and (B)). The 300 kDa band corresponding to the TcA monomer is marked with an arrow. C) In addition, the SDS-gel after size exclusion chromatography with single fractions eluting from the column. The fraction analyzed by negative stain EM is indicated with a star. A-C) Negative stain micrographs of the indicated fraction after size exclusion chromatography at pH 8 and a micrograph after 72 h incubation with nanodiscs at pH 11 is shown as well. Particles in the prepore state are marked with circles and particles in the pore state with dashed circles. Scale bars are 100 nm.

Moreover, either Glu1086 or Arg1166 were separately mutated to confirm this enormous effect of Glu and Arg on the protein's stability. Therefore, these two mutants (Pl-TcdA1-E1086A and Pl-TcdA1-R1166A) were expressed and interestingly, they showed similarly weak expressions like the double mutant Pl-TcdA1-E1086A-R1166A (Figure 42A). Conclusively, one mutation is sufficient to interrupt the interaction and with this, to alter the protein stability. Altogether, out of three possible conserved ionic

interactions that fulfill the requirements of being located between two protomers and are in close proximity only in the prepore state, merely one interaction side showed an effect on the protein stability. Considering that one protomer has 2516 residues and that there is a large interface between two protomers, it is interesting that only one residue has such a tremendous effect on the complex stability. However, the respective residues are less sensitive towards a pH shift to pH 11 or pH 4, which were used to trigger the pore state in Pl-TcdA1[42]. Therefore, the conserved interaction does not act as a typical pH-switch and is therefore not solely responsible for the pH-dependent mechanism of the TcA. Most likely, the ionic interaction between the side chains are destabilized at extreme pH values due to changes in close proximity in the interaction side 3. Thus, to test whether surrounding residues are also important for the interaction between the adjacent protomers, the interaction pocket including Glu1086 and Arg1166 was analyzed in more detail (Figure 42B and C). Two tyrosines are present in the interaction pocket, namely Tyr1168 and Tyr1205 in Pl-TcdA1 (Figure 42B and C) and are conserved in Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. The hydroxyl groups of the tyrosines are pointing towards the center of the pocket. At basic pH above pH 11, the hydroxyl groups might be deprotonated and therefore negatively charged. This could affect the Arg1166 and may lead to a different orientation of the residue, so that the interaction with Glu1086 is disturbed (Figure 42B). To test this hypothesis, a Pl-TcdA1 mutant was generated, in which the two respective tyrosines were replaced by phenylalanines (Pl-TcdA1-Y1168F-Y1205F). Subsequently, if the hydroxyl groups of the tyrosines affect the Arg1166 orientation, the phenylalanine mutant should remain in the prepore state also at pH 11.

The mutant was expressed, purified and dialyzed against pH 11 in presence of nanodiscs to induce the conformational change to the pore state (Figure 42D). During protein expression and purification, the mutant behaved like Pl-TcdA1-E158A-R1873A and Pl-TcdA1-D965A-R1971A (Figure 41A and B). After size exclusion chromatography, the sample was analyzed by negative stain EM. The micrograph reveals the typical bell-shaped Pl-TcdA1 particle at pH 8 (Figure 42D, solid line). Upon pH change to pH 11, the protein structurally rearranges into the pore state and incorporates into nanodiscs (Figure 42D, dashed line). Thus, the tyrosine residues have no effect on the proteins stability and moreover, they seem to have no impact on the interaction between Glu1086 and Arg1166. Another residue in this interaction pocket that is not conserved in Yp-TcaATcaB but in the other TcAs is a histidine (His1202 in Pl-TcdA1, Figure 42B and C). To analyze the impact of this histidine onto the complete interaction pocket, another mutant was designed, in which the histidine is replaced by alanine (Pl-TcdA1-H1202A). This mutant remained stable during expression, purification and even during pore formation (Figure 42E). No impact of the missing histidine could be observed during the pH-induced prepore-to-pore transition. Therefore, only Glu1086 and Arg1166 within the conserved residues at the interaction groove demonstrated an effect on the stability of the protein. To elucidate the effect of this interaction on the other TcAs, further studies are required.

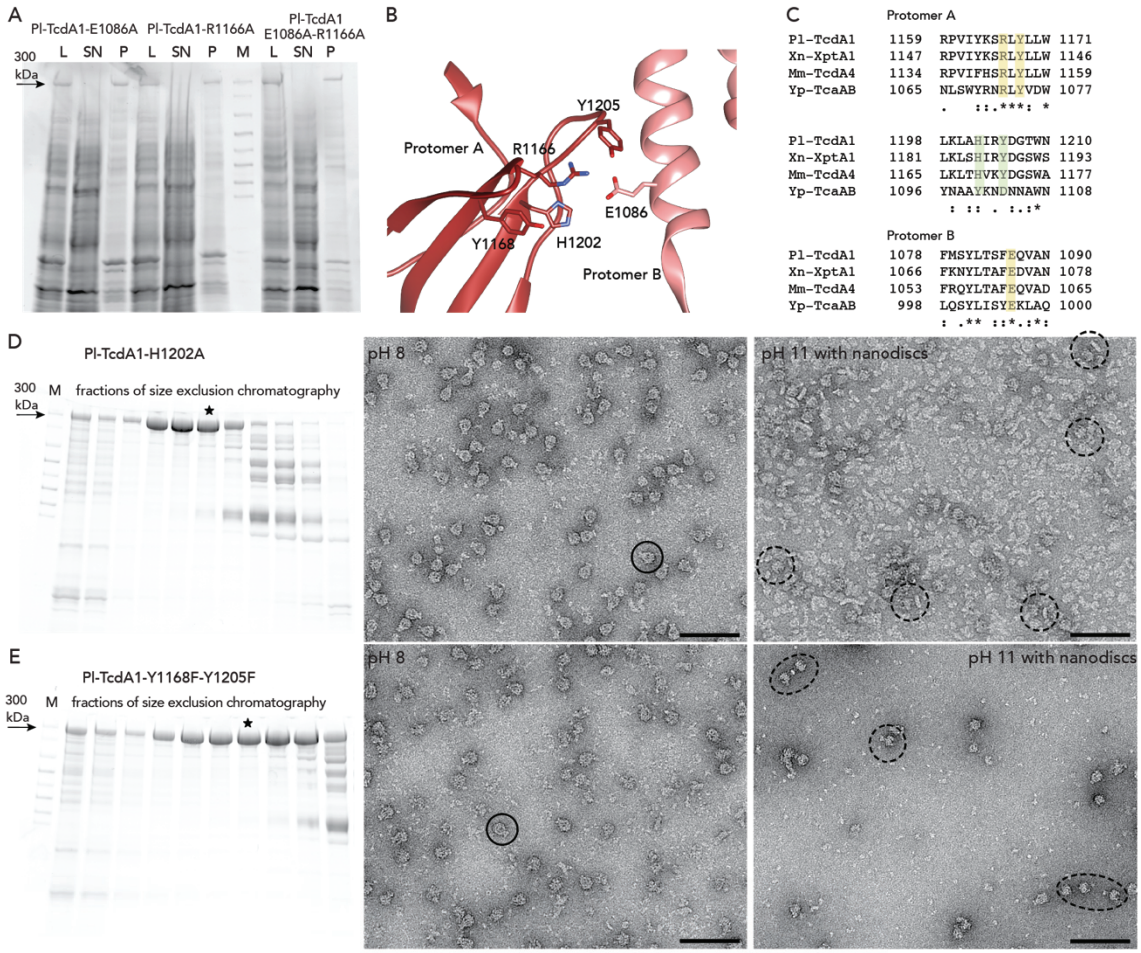


Figure 42: Mutational studies of residues in the interaction pocket 3 of Pl-TcdA1. A) SDS-PAGE gel after expression of the Pl-TcdA1 mutants Pl-TcdA1-E1086A (left) and Pl-TcdA1-R1166A (center) as well as the double alanine mutant Pl-TcdA1-E1086A-R1166A (right) as a control. For each mutant, a sample of the lysate (L), supernatant (SN) and pellet after ultracentrifugation (P) was loaded. B) Close-up view of the interaction side 3 in Pl-TcdA1. Residue numbers correspond to Pl-TdA1. C) Sequence alignments of protein stretches forming the interaction groove 3. The conserved residues in all four TcAs are highlighted in yellow and the ones only conserved in Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 in green. D) SDS-gel of the Pl-TcdA1-H1202A mutant after size exclusion chromatography (left) as well as negative stain electron micrographs at pH 8 (middle) and pH 11 after incubation with nanodiscs (right). E) Left panel shows the SDS-gel of the Pl-TcdA1-Y1168F-Y1205F mutant after size exclusion chromatography. Negatively stained electron micrographs of the sample at pH 8 (middle) and at pH 11 after incubation with nanodiscs (right). The fraction used for electron microscopy is marked with asterisks. Particles in the prepore state are marked with circles and particles in the pore state with dashed circles. Scale bars are 100 nm.

Moreover, mutational studies were performed in which the Glu1086 and Arg1166 residues are mutated to cysteines. In principle, under oxidative conditions, the cysteine residues should be linked by a disulfide bridge and upon reduction, the disulfide bridge is reduced. Thus, the destabilizing of the protein or moreover the opening of the shell could be induced pH-independent, only controlled upon reduction and oxidation. However, the expression of these mutant under oxidative conditions was not successful, most of the protein was in inclusion bodies and not enough protein could be gained for further studies. Most likely, the cysteine residues were not close enough to form disulfide bridges, which results in destabilized protein as described for the Pl-TcdA1-E1086A-R1166A mutant.

Furthermore, present in Pl-TcdA1 are three histidine residues (His46 and 50 in protomer A and His1808 in protomer B), that were identified to form a tight interaction in the prepore state and could act as a typical pH-switch (Figure 43A and B)[192]. However, these His residues are not conserved among the four TcAs Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. In addition, the loop of protomer B, containing His1808, is much shorter in Mm-TcdA4 and out of range for interaction with the helix from protomer A (Figure 43C). Since these three histidine residues are not conserved, this could implement, that the Tc toxins might have distinct electrostatic locks. This goes along with the observation, that the TcAs differentiate in the loops and thus an opening at the tip of the neuraminidase-like domain. In addition, only one conserved interaction pair was identified to destabilize the protein in the prepore state. This interaction pair could mediate a pH-dependent mechanism of prepore-to-pore transition by acting as a latch as described in OmpG from *E. coli*[190]. Different to the resulting effect of closing the pore in OmpG, the latch of Pl-TcdA1 might lead to an opening of the shell after the interacting protomer interfaces are destabilized upon pH shift to pH 4 or 11. Conclusively, there might be not solely one single conserved electrostatic lock being responsible for the opening of the shell and the triggering of the prepore-to-pore transition. Moreover, TcAs from distinct bacteria could have evolved sequence and structure based specific regions for a pH-dependent mechanism according to the host specificity of the bacteria.

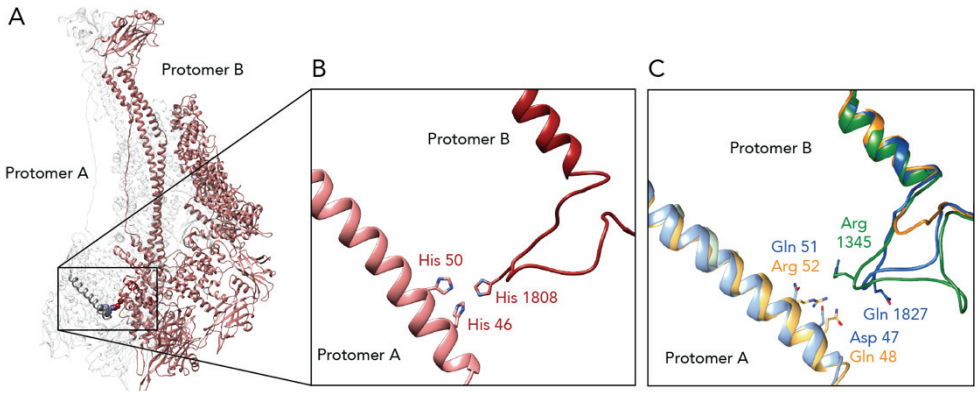


Figure 43: Interaction of the three histidines between two protomers of Pl-TcdA1. A) Two protomers of Pl-TcdA1 indicating the location of the interaction. B) A close-up view of the three interacting histidines of Pl-TcdA1. C) The histidine residues are not conserved within Xn-XptA1 (blue), Mm-TcdA4 (yellow) and Yp-TcaATcaB (green).

5.4.2 Prepore-to-pore transition

The mechanism of action of Pl-TcdA1 includes a massive conformational change of the TcA component from the prepore state to the pore state (chapter 2.2.3.2). During this transition, the shell is rearranged, the linker partially folded, the pore domain inserted into the membrane and the tip of the pore domain opens up and interacts with surrounding lipids[44].

The common domain architecture and conservation of functional important domains of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB suggest a similar mechanism of action of the toxins. To analyze the pore formation of these toxins, two different experimental set ups were performed. First, the pore state was triggered by pH shift as it was previously observed for Pl-TcdA1. Second, black lipid bilayer experiments were performed to measure the pore insertion into a membrane to translocate ion molecules.

5.4.2.1 pH studies with negative stain EM

As previously stated, transition of Pl-TcdA1 into the pore state of Pl-TcdA1 could be induced by a pH shift from pH 8 either to pH 11 or pH 5[42]. Since the TcAs from the analyzed organisms reveal differences regarding the closure of the neuraminidase-like domain as well as their charge profile, I was interested in the impact of the pH change on Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. The cell entry of the toxins is thought to happen by the endocytotic pathway or directly from the gut lumen. Thus, both pH shifts to extreme basic (pH 10-11) and acidic pH (pH 4-5) were tested for Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB.

The pH was shifted by dialysis of the toxin against a buffer of the respective pH. To provide a membrane-like environment for the transmembrane helices of the pore domain and to prevent aggregation, nanodiscs were added. In some experiments, the nanodiscs were replaced by detergent. The conformational change of the toxins and the incorporation into nanodiscs or interaction with the nanodicscs was analyzed by negative stain EM.

<u>Xn-XptA1</u>

Previous pH studies of Xn-XptA1 identified an intermediate state of the toxin, in which the neuraminidase-like domain seems to be rearranged to open up and interact with the nanodiscs[156]. However, this intermediate state was only occasionally observed and could not regularly be triggered to perform detailed structural analysis.

The negative stain micrograph presents the toxin incubated with nanodiscs at pH 11 for three days (Figure 44A). Although most of the particles are in the prepore state, also the pore state (dashed line)

as well as the intermediate state (solid line) were observed. The particles in the pore state differ enormously from the prepore state particles. Namely, the complete shell is rearranged and forms a crown on top of the translocation channel, which is inserted into the nanodisc (Figure 44A and G). Interestingly, some protein revealed protruding lower β-sheet domains without interacting with a nanodisc (Figure 44A, dotted lines). Thus, this torro-like particle might represent an initial step of the conformational rearrangement that needs further triggering or maybe cofactors to completely undergo the conformational change and open the shell. To enhance the interaction and incorporation of proteins into nanodiscs, the sample was incubated with detergent (either DDM or OG) at pH 11 prior to nanodisc addition (Figure 44B and C). This was done to investigate whether the pore formation is favored in a detergent environment. The protein in the pore state could then be exchanged into nanodiscs as it is performed for membrane integrated proteins[194]. However, most particles are in the prepore state for both detergents and nanodiscs remain empty in the background (Figure 44B and C). In addition, the intermediate state was observed for DDM conditions more often than for OG conditions and the nanodiscs attached to the proteins bottom are clearly visible (Figure 44B, solid line). In general, the OG-condition was less favored for Xn-XptA1, since more aggregates and contaminants were present in the sample. Thus, the pore state could be induced but was rarely present in presence of the MSP1D1-ΔH5-His nanodiscs with a diameter of ~ 7nm (Figure 44A-C). Accordingly, the incorporation of Xn-XptA1 into MSP1D1-ΔH5-His nanodiscs at pH 11 with and without prior detergent treatment was less efficient than for Pl-TcdA1.

To analyze the Xn-XptA1 interaction with lipids in general, the protein was incubated with liposomes of brain polar lipids (BPL) at pH 11. Empty liposomes as well as fully decorated liposomes with protein are present in the negative stain micrograph after incubation at pH 11 (Figure 44D). Furthermore, liposomes could be observed, in which the toxin docks to the liposome surface and thus, displaying the intermediate state (marked with circles with a solid line). Since only a few liposomes are decorated with the protein, the incorporation of the toxin into the liposomes was indolent. They were hardly liposomes with only one protein interacting present. This suggests, that the incorporation process of the toxin is cooperative. Upon the interaction of one or more Xn-XptA1 proteins with the liposomes, other toxins join more likely. This would moreover explain the low incorporation efficiency into the small diameter nanodiscs, since here only one or two proteins are able to insert.

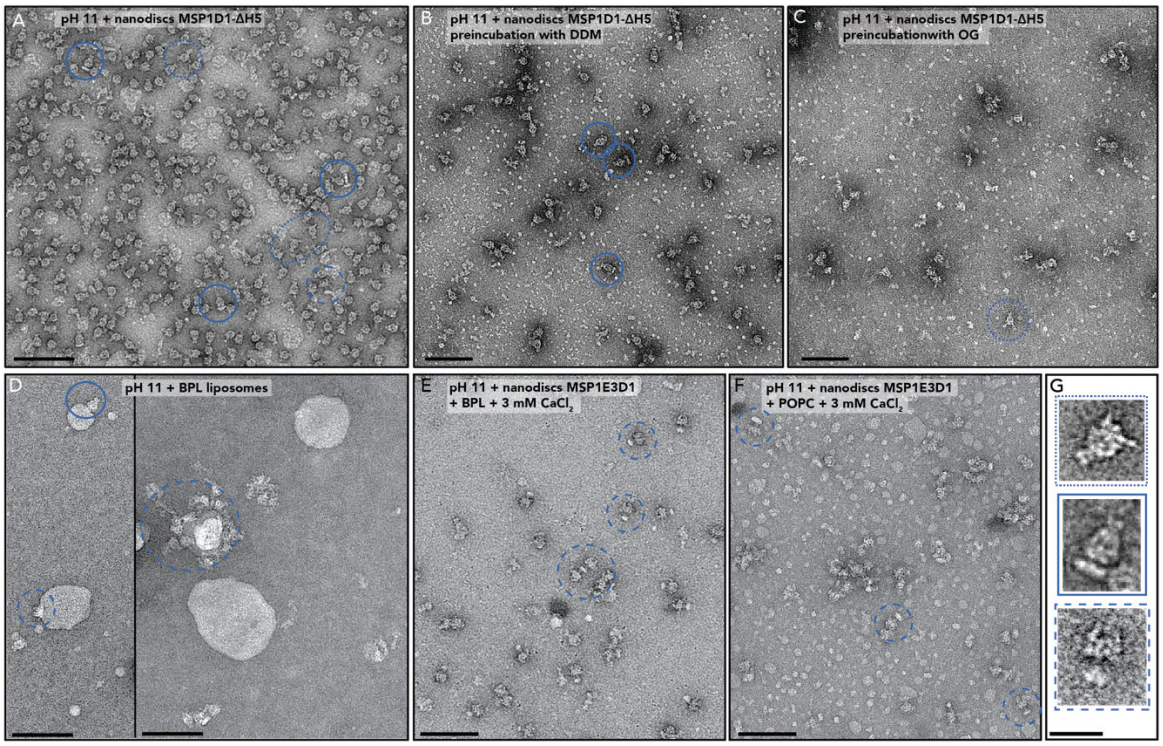


Figure 44: Negative stain micrographs of Xn-XptA1 in presence of nanodiscs or detergents at different pH-values. A-G) XnXptA1 particles representing the pore state are marked with dashed lines, Xn-XptA1 in the intermediate state is visualized with a bold line and particles with protruding β-sheet domains are highlighted with a dotted line. A) Xn-XptA1 sample after incubation at pH 11 with MSP1D1-ΔH5-His nanodiscs revealed the three different states of Xn-XptA1. B+C) Electron micrographs presenting Xn-XptA1 at pH 11 in presence of MSP1D1-ΔH5-His nanodiscs. Prior to nanodiscs addition, the sample was incubated at pH 11 in a buffer containing the detergent DDM (B) or OG (C). D) Xn-XptA1 after incubation with preformed BPL liposomes. Liposomes decorated with protein are marked with circles. E+F) Micrographs showing Xn-XptA1 samples with self-assembled nanodiscs consisting of MSP1E3D1-His and BPL (E) or POPC (F) lipids at pH 11 and in the presence of 3 mM $CaCl_2$. Scale bars are 100 nm. G) A close-up view of the three different observed particle conformations, the torro-state (dotted line), the intermediate state (solid line) and the pore state (dashed line). Scale bar is 25 nm.

All previous experiments were performed with lipids containing a neutral or negatively charged head group. POPC has a phosphatidylcholine head group and in the brain polar lipids are a mixture of lipids. Over 60 % of the brain polar lipids are phospholipids also resulting in a negative netto-charge of the lipid composition[195]. Taking also into account, that the bottom of the Xn-XptA1 protein is negatively charged at pH 8 and pH 11, the interaction between the protein and the lipids is unfavored. To mediate the protein-nanodisc interaction, $CaCl_2$ was added to the buffer. The effect of Ca^{2+} ions on phosphatidylcholine bilayers was investigated in previous studies[196] and revealed an interaction of the ions with the head group of the lipids and the mediating effect of lipid-protein interaction as for example in the perforin C2 domain[197]. The Ca^{2+} ions appear to enhance the interaction with the nanodiscs, because most proteins are transitioned into the pore state and inserted into a nanodisc, independent of different tested lipid, POPC or BPL (Figure 44E and F). Furthermore, in these experiments nanodiscs with a larger diameter of 16 nm were used, since the presence of the torro sate (Figure 44A-C) suggests that the RBDs protrude from the toxins shell and therefore generate a larger interaction surface of the

toxin. Thus, the MSP2N2-His nanodiscs could facilitate the toxin incorporation by providing a larger lipid surface. However, and very interestingly, Pl-TcdA1 can interact and incorporate efficiently into nanodiscs although the negative charges at the neuraminidase-like domain are more dominant for Pl-TcdA1 then for Xn-XptA1 at pH 11. Solely the pH shift and the presence of a lipid environment is adequate for the prepore-to-pore transition of Pl-TcdA1 *in vitro*. Consequently, Xn-XptA1 seems to need an additional cofactor to undergo the conformational change *in vitro*. This assumption goes along with the observation of the torro and intermediate state of Xn-XptA1 in presence of Ca^{2+} ions (Figure 44F). The toxin starts to interact with the nanodiscs or liposomes but no complete transition into the pore state can be observed. In summary, the pore state of Xn-XptA1 could be induced and this underlines the assumption of a similar mechanism of action for the different Tc toxins.

Mm-TcdA4

Based on the Xn-XptA1 experiments, similar conditions were tested for Mm-TcdA4 to induce the pore state. However, after incubation at pH 10.5 for three days, the Mm-TcdA4 particles appear to be partially degraded and parts of the shell are missing (Figure 45A, orange arrows). These particles resemble the Yp-TcaATcaB particles, which miss the TcaA part (chapter 5.3.2.3). Thus, in contrast to Pl-TcdA1 and Xn-XptA1, the Mm-TcdA4 protein appears to be instable after long incubation at high pH values.

Nevertheless, experiments at pH 11 in the presence of $CaCl_2$ were performed with MSP1D1-ΔH5-His and MSP2N2-His nanodiscs and incubated for one day only. Different interaction states of Mm-TcdA1 with the nanodiscs could be observed. Comparable to the Xn-XptA1 torro state, the receptor-binding domains at the bottom of Mm-TcdA4 were interacting with the small diameter nanodiscs (MSP1D1-ΔH5-His, Figure 45B, dotted lines). Interestingly and although the same ratio of toxin to nanodiscs was used for the two experiments with $CaCl_2$, there are much less MSP1D1-ΔH5-His nanodiscs present in the background of the electron micrograph compared to MSP2N2-His nanodiscs (Figure 45B and C). A 20-fold increased $CaCl_2$ concentration was used with the MSP1D1-ΔH5-His nanodiscs and this might have an effect and destabilize the nanodiscs. The MSP2N2-His nanodiscs in presence of 5 mM $CaCl_2$ revealed Mm-TcdA4 particles similar to the intermediate state of Xn-XptA1. Likewise, the receptor-binding domains interact with the nanodiscs, as they are at the bottom as well as at the side of Mm-TcdA4. In addition, also a minor fraction of particles representing the pore state were present in the sample (Figure 45C, dashed line). However, in general, the pore state was hardly observed at pH 10.5 or 11. Considering that Mm-TcdA4 is a toxin from an opportunistic human pathogen and that the intoxication pathway occurs most likely via the endosomal incorporation, further experiments were performed at low pH of 4 and 5.

Thus, the protein was incubated at pH 4 in presence of nanodiscs. Interestingly, the nanodiscs seem to be unstable at pH 4 over a period of three days, because no nanodiscs were observed at pH 4 (Figure 45D). In contrast, Mm-TcdA4 was still present in the prepore state as top and side views after the incubation at pH 4 indicating a fairly stable protein at low pH (Figure 45D). In order to evaluate the interaction of Mm-TcdA4 with lipids at low pH, BPL liposomes were incubated with the protein at pH 5 (Figure 45E). Again, an interaction of Mm-TcdA4 receptor-binding domains at the lower part of the toxin with the liposomes could be observed. In addition, some proteins seemed to have undergone the conformational change into the pore state and inserted into the liposomes (dashed lines in Figure 45E and F). The protein exhibits the same rearrangement of the shell that forms a crown on top of the translocation channel, which is incorporated into the nanodiscs.

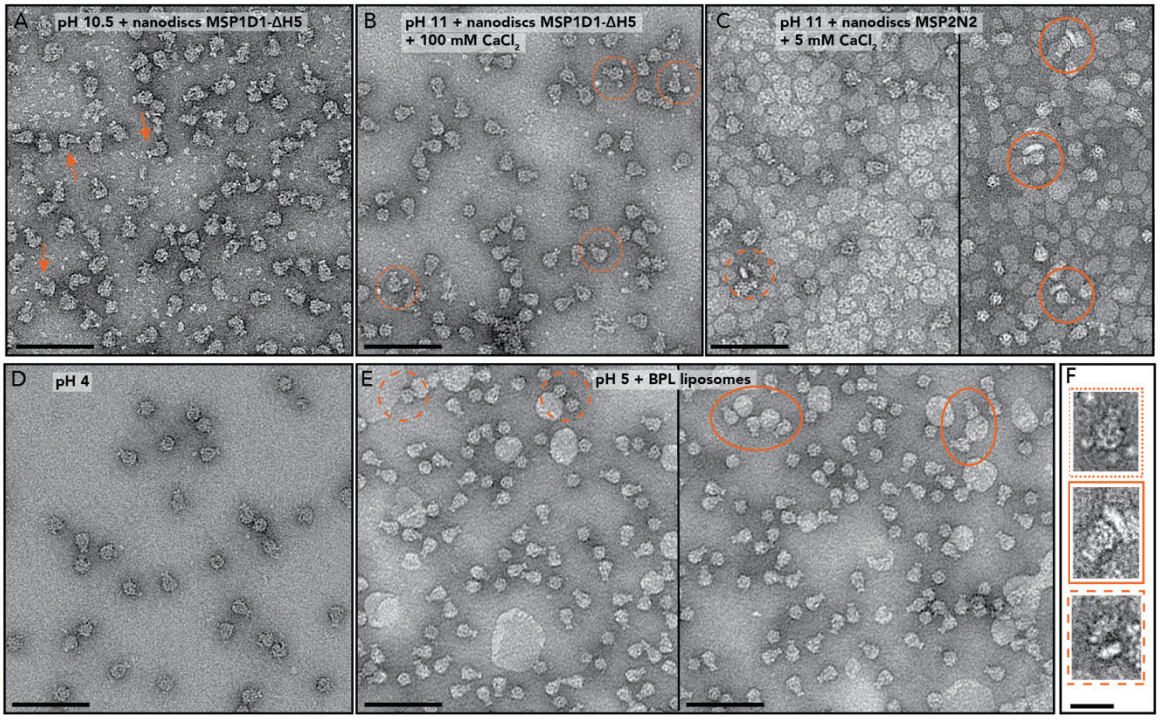


Figure 45: pH studies with Mm-TcdA4 analyzed by in negative stain EM. Negative stain electron micrographs showing different nanodiscs and pH studies with Mm-TcdA4. Particles in the pore state are marked with dashed lines, particles representing a torro-like conformation are visualized in dotted lines and the intermediate state with solid lines. A) Mm-TcdA4 after three days of incubation at pH 10.5 in the presence of MSP1D1-ΔH5-His nanodiscs. Protein degradation is marked with orange arrows. B) Sample after one-day incubation at pH 11 with MSP1D1-ΔH5-His nanodiscs and in the presence of 100 mM $CaCl_2$. C) Sample after one-day incubation at pH 11 with MSP2N2-His nanodiscs in 5 mM $CaCl_2$ containing buffer. D) Mm-TcdA4 without nanodiscs or lipids at pH 4. E) A negative staine micrograph of Mm-TcdA4 in presence of BPL liposomes at pH5. Scale bars are 100 nm. F) Close-up views of the three different observed particle conformations upon pH change. Scale bar is 25 nm.

Altogether, also Mm-TcdA4 exhibits different states upon pH shift to pH 5, pH 11 and in presence of a lipid environment. The torro state, the intermediate state and also the pore state are present (Figure 45F). Moreover, $CaCl_2$ seems to have no enormous impact on the pore state formation of Mm-TcdA4.

At pH 11 and in presence of either 5 or 100 mM $CaCl_2$, most of the Mm-TcdA4 particles are in the prepore state and only a minor fraction of the protein transitioned into the pore state. In contrast, $CaCl_2$ addition to the Xn-XptA1 sample at pH 11 resulted in an efficient prepore-to-pore transition of nearly all particles. Thus, the mechanism of the conformational change and membrane penetration of Mm-TcdA4 is indeed similar to Pl-TcdA1 and Xn-XptA1. However, the triggering of the transition seems to require additional cofactors *in vitro* or is only possible in presence of a receptor. Only the shift in pH and the presence of a membrane-like environment does not result in an efficient prepore-to-pore transition for Mm-TcdA4 as it was observed for Pl-TcdA1[42].

Yp-TcaATcaB

Also for Yp-TcaATcaB, various conditions were tested to trigger the transition into the pore state. At pH 4.7 and in the presence of nanodiscs with the larger diameter (MSP2N2-His), some Yp-TcaATcaB particles could be detected in the pore state (Figure 46A, dashed line and close-up view). Again, reorganization of the shell that sits like a crown on top of the channel domain is present. However, most of the protein aggregates or remains in the prepore state (Figure 46A). By using a slightly higher pH of 5, the protein exhibits the same behavior as at pH 8 resulting in a stable prepore conformation of the protein (Figure 46B and Figure 18). Thus, only the slight difference of 0.3 in the pH has an effect on the proteins prepore-to-pore transition and its stability. Moreover, the impact of a detergent, in this case DDM, was tested at pH 5. However, the proteins behavior was not affected by the addition of DDM. To analyze whether the pore state can also be triggered at high pH, the protein was incubated with nanodiscs (MSP1D1-ΔH5-His) at pH 10 or pH 10.5 (Figure 46C and D). The protein shows more clustering compared to pH 5. Despite that, the pore state can also be triggered, although the interaction with the nanodiscs is less prominent and empty nanodiscs are in the background. Instead, the protein interacts with the transmembrane part of other proteins and forms a cloverleaf shape (Figure 46C and D, dashed line). This behavior was also observed for Pl-TcdA1 at pH 11 and in other pore-forming toxins[198]. Moreover, this clover leaf-effect could depend on the protein concentration and the used ratio of protein to nanodiscs. At pH 11, the pore state is not present anymore and the protein starts to degrade at this extreme pH value (Figure 46E).

In general, Yp-TcaATcaB appears to be much more pH-sensitive than the other TcAs. Indeed, this observation is supported by nanoDSF measurements (appendix Figure 57 and Table 7). The proteins thermal stability was measured over a pH range from 4 to 11. Between pH 5 and 8 (50.9-53.1 ˚C), the protein is most stable. Upon slight changes from pH 5 to pH 4.5, the T_m decreases by 5 ˚C. In more acidic buffer with pH 4, the T_m decreases by another 8.5 ˚C. A similar behavior can be observed when the pH is increased. Between pH values from 9 to 10, the T_m decreases from 47.2 to 42.3 ˚C. Upon pH 10.5, the thermal stability decreases drastically to 35 ˚C and for pH 11 no melting temperature at all

could be determined, because the protein seems to be unstable under these conditions and most likely degraded completely. Altogether, these findings indicate a less stable protein complex of Yp-TcaATcaB compared to the other TcAs. One reason could be the composition of Yp-TcaATcaB as a heterodimer. Not only the interface between two protomers, but moreover also the interface between the two subunits TcaA and TcaB are affected by the pH change. Subsequently, at the extreme pH values of pH 4 and pH 11 possible charge repulsions could destabilize both interfaces and thus lead to complex disassembling.

Moreover, in contrast to Xn-XptA1 and Mm-TcdA4, no torro-like state could be observed for Yp-TcaATcaB. This could be due to the different folding of the β-sheet region in Yp-TcaATcaB. Only the neuraminidase-like domain is present in Yp-TcaATcaB and the more flexible receptor-binding domains are absent. Most likely, in Xn-XptA1 and Mm-TcdA4, these receptor-binding domains are responsible for the torro state with the lower β-sheet domains protruding to the periphery. Therefore, in Yp-TcaATcaB, the interaction with the lipids is either occurring with the neuraminidase-like domain or regions of the helical shell.

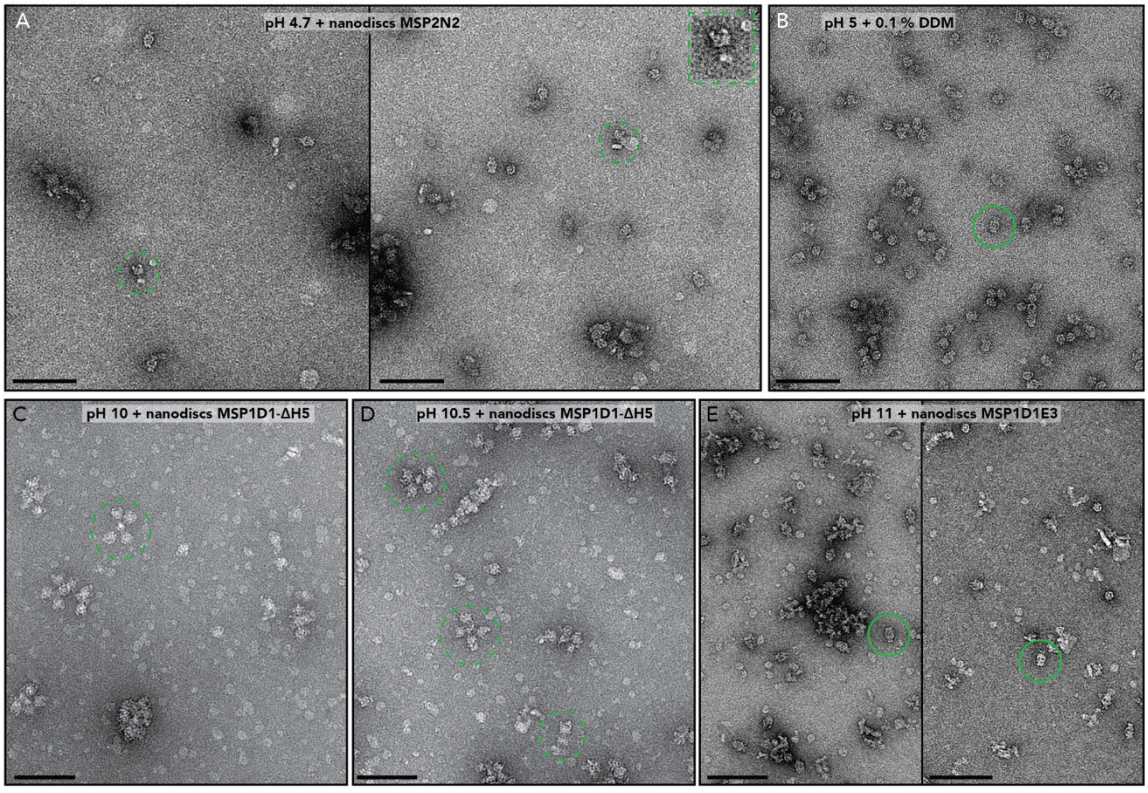


Figure 46: Negative stain micrographs of pH-dependent experiments with Yp-TcaATcaB in presence of nanodiscs or detergent. A-E) Negative stain micrographs representing Yp-TcaATcaB particles in the pore state (dashed lines) and in the prepore state (solid line). A) Yp-TcaATcaB after incubation with MSP1D1-ΔH5-His nanodiscs at pH 4.7. Particles in the pore state are highlighted and one particle is shown in a close-up view. B) Electron micrograph presenting Yp-TcaATcaB at pH 5 in presence of DDM. C+D) Yp-TcaATcaB sample after incubation at pH 10 (C) and pH 10.5 (D) with MSP1D1-ΔH5-His nanodiscs revealed cloverleaf assemblies of multiple Yp-TcATcaB proteins. E) Micrograph showing Yp-TcaATcaB sample with self-assembled nanodiscs consisting of MSP1E3D1-His and POPC lipids at pH 11. Scale bars are 100 nm.

Interestingly, Mm-TcdA4 and Yp-TcaATcaB adopted the pore state conformation at low and at high pH values. Concerning the endosomal pathway as intoxication pathway for these two human pathogenic toxins, transition to the pore state should only or more likely occur at low pH. Nevertheless, the pore state was present at pH 10.5 (Yp-TcaATcaB) and pH 11 (Mm-TcdA4) for both proteins. Considering that both proteins are unstable at high pH, the proteins could start to disassemble and destabilize the interaction of the outer shell and thus the pore state could be triggered at basic pH due to instable proteins. Despite of that, all three TcAs are able to undergo the conformational change from the prepore to the pore state, which supports the assumption that the Tc toxins have a similar mechanism of action. However, clear insights into the triggering of the pore state for the single proteins can only be gained with further studies *in vivo* and with a toxin specific receptor.

5.4.2.2 Black lipid bilayer experiment

Another experiment to test the pore formation of the toxins was the black lipid bilayer experiment. The principle of this experiment is the measurement of an ion flux between two chambers that are connected by a small hole. Upon membrane formation, the hole is covered and no ion flux can be measured. After protein addition and applying a membrane potential, pore insertion is favored. If a pore forming protein inserts into the membrane, ions can pass from one chamber into the other and the ion flux can be measured. Thus, the readout of these experiments is the channel conductivity measured per pore insertion. This experiment was performed for Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB at three different pH values (pH 4, 6 and 11) in the laboratories of Prof. Dr. Roland Benz at the Jacobs University in Bremen.

For all proteins and each pH value, an ion flux could be measured caused by pore insertion into the membrane (Figure 47 and Table 5). Pl-TcdA1 exhibits a constant channel conductance of 500-550 pS for all tested pH values. Mm-TcdA4 has a slightly lower but also constant channel conductivity between 450-500 pS for each pH value. For Xn-XptA1, the average single-channel conductance increases slightly from pH 4 (275 pS) to pH 11 (400 pS). This could result from varying charges at the channel lumen caused by different pH environments. Moreover, not only a pore insertion event was measured for Xn-XptA1 at pH 6 but also a pore closing event (Figure 47B, blue arrow). This could have different reasons. As described previously, the transmembrane domain of Pl-TcdA1 rearranges and opens upon membrane insertion. Furthermore, the loop of the transmembrane domain interacts with lipid head groups. For Xn-XptA1, more dynamic interactions with the surrounding lipids could be the reason for the stop of the ion flux measurement. Furthermore, proteins in the solution could interact with the inserted Xn-XptA1 protein and thus close the channel by interacting with the TcB-binding domain. Yp-TcaATcaB shows a channel conductivity of 450 pS. Especially at the extreme pH values of pH 4 and

pH 11, the single channel recordings are noisy, which was not observed for the other toxins. This could be due to less stable interaction with the lipids in the membrane, as it was described for Xn-XptA1.

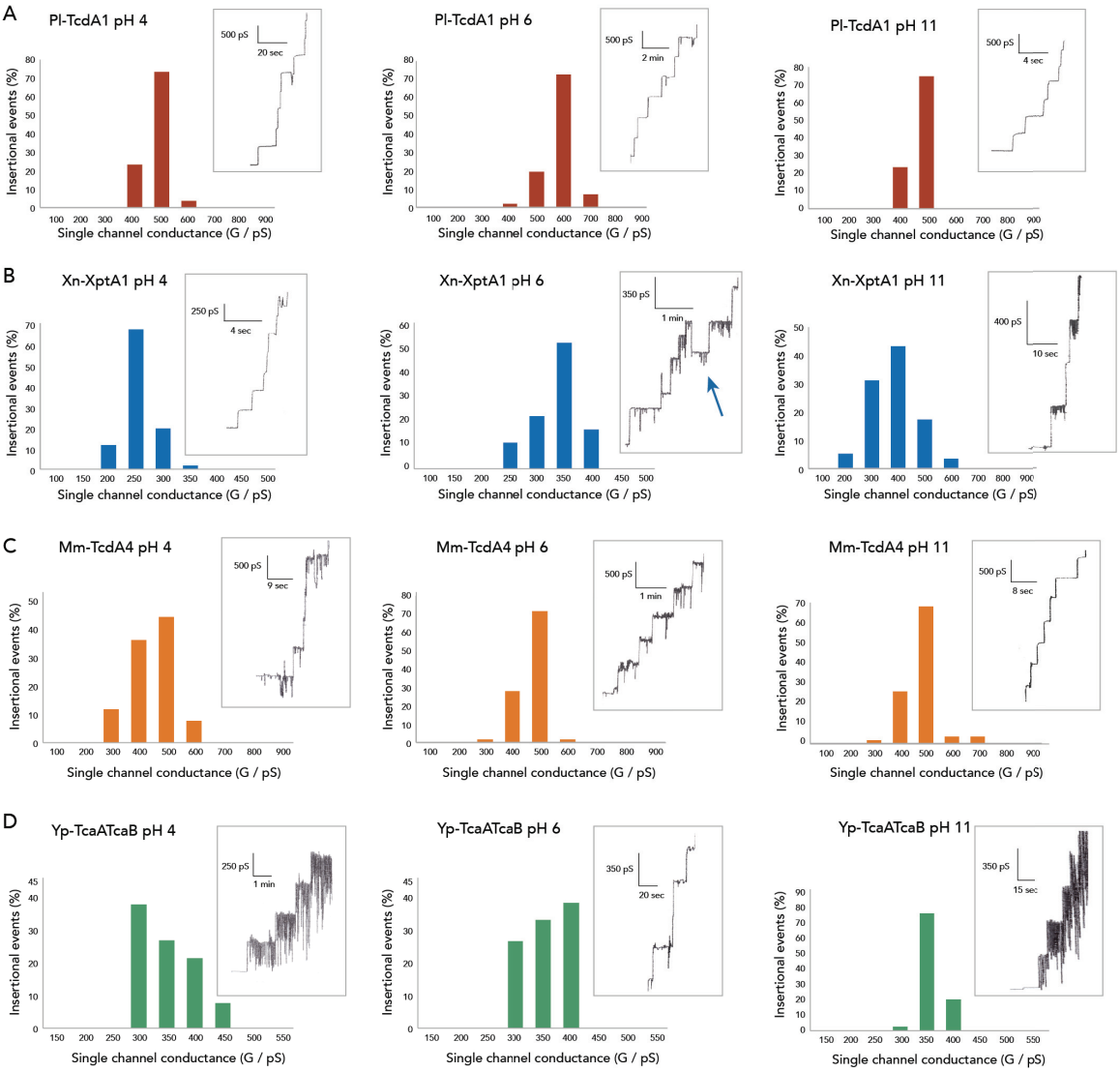


Figure 47: Distribution of single-channel conductance of TcAs at different pH values. All histograms were reconstructed from the data of at least 70 pore insertion events for each TcA and pH value. For each experiment, an exemplary recording of the single-channel current versus time is attached. The measurements were performed in 1 M KCl for Pl-TcdA1 (A), Xn-XptA1 (B), Mm-TcdA4 (C) and Yp-TcaATcaB (D).

Moreover, in Table 5, the concentration of each toxin used for the different pH values are listed. After a minimum of 70 single pore insertion events was reached, the toxin addition and the measurements were stopped. For Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 less toxin was needed at the extreme pH values compared to pH 6. Thus, the pore insertion was faster and more efficient at pH 4 and pH 11. This indicates a pH-dependent pore formation mechanism of these toxins. Contrarily, for Yp-TcaATcaB it was vice versa and the protein seemed to be more active or more efficient in pore formation at pH 6 then at pH 4 and pH 11. These observations would suggest a pH-independent mechanism of pore

formation of Yp-TcaATcaB, which would be in contrast to the incorporation into nanodiscs experiments. A pH-dependency was observed for Yp-TcaATcaB in negative stain EM to transition into the pore state. Thus, the results from the black lipid bilayer experiments could be explained by the less stable protein behavior of Yp-TcaATcaB at extreme pH values. The incorporation into nanodiscs at different pH values as well as the nanoDSF experiments revealed, that Yp-TcaATcaB is pH-sensitive. Thus, at pH 4 and pH 11, the protein is less stable and the complex disassembles, which lead to a higher protein concentration to record the same amount of pore insertion events. Although the black lipid bilayer experiments were performed within two hours, the protein seems to be already destabilized after this time.

Table 5: Results from single channel conductivity measurements with black lipid bilayers.

	Pl-TcdA1		Xn-XptA1		Mm-TcdA4		Yp-TcaATcaB	
	G (pS)	pmol	G (pS)	pmol	G (pS)	pmol	G (pS)	pmol
pH 4	500 ± 82	3.7	275 ± 56	0.3	450 ± 12	3.1	375 ± 56	29.3
pH 6	550 ± 112	19.4	325 ± 56	16.6	450 ± 112	28.1	350 ± 41	3.6
pH 11	500 ± 50	0.5	400 ± 141	0.4	500 ± 141	4.0	350 ± 41	16.4

Additionally, the influence of $CaCl_2$ on Pl-TcdA1, Mm-TcdA4 and Xn-XptA1 was investigated (appendix, Figure 58A-C), since the presence of Ca^{2+} ions had an enormous effect on pore formation of Xn-XptA1 in negative stain EM. The Ca^{2+} ions should enable the interaction of the negatively charged membrane with the surface of the β-sheet regions of the toxins, therefore, the experiment was only performed at pH 11 and as a control at pH 6. Different $CaCl_2$ concentrations ranging from 0 mM up to 250 mM were used. However, black lipid bilayer experiments revealed no major effect of $CaCl_2$ for the three toxins (appendix, Figure 58A-C). No divergent trend is discernible for the toxins compared to the experiments without $CaCl_2$.

Thus, the combination of electron microscopy and black lipid bilayer experiments revealed the transition from the prepore to the pore state for each TcA could be successfully accomplished. Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB behaved differently compared to Pl-TcdA1 and their incorporation into nanodiscs *in vitro* was less efficient as shown previously for Pl-TcdA1[43,44]. An intermediate state was observed for Xn-XptA1 and Mm-TcdA4, which has not been reported for Pl-TcdA1 or was present in the Yp-TcaATcaB sample. Nevertheless, the pore state could be triggered for all toxins and moreover, an ion flux could be measured. Altogether, these experiments provide first insights into the mechanism of Tc toxins from different organisms. The observed differences for the toxins again indicate a general common mechanism of action but different activation mechanisms result in a defined host specificity.

Further studies are needed to evaluate the distinct behavior and trigger mechanism of the conformational change.

5.4.3 Pore forming domain

Between all four analyzed TcAs, the pore domain is the best resolved part, shows the highest structural similarity (Figure 32B) and is the domain with the highest sequence conservation (Figure 48C). Four all four TcAs, the tip of the channel is highly hydrophobic and spans a length of roughly 5 nm (Figure 48A), indicating that Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB have an identically sized transmembrane domain such as Pl-TcdA1[43]. Only the lower part of this transmembrane region is conserved (Figure 48D). All four TcAs have a hydrophobic residue at the point of the transmembrane loop, which in contrast is not conserved (Figure 48D, alignment of transmembrane loop).

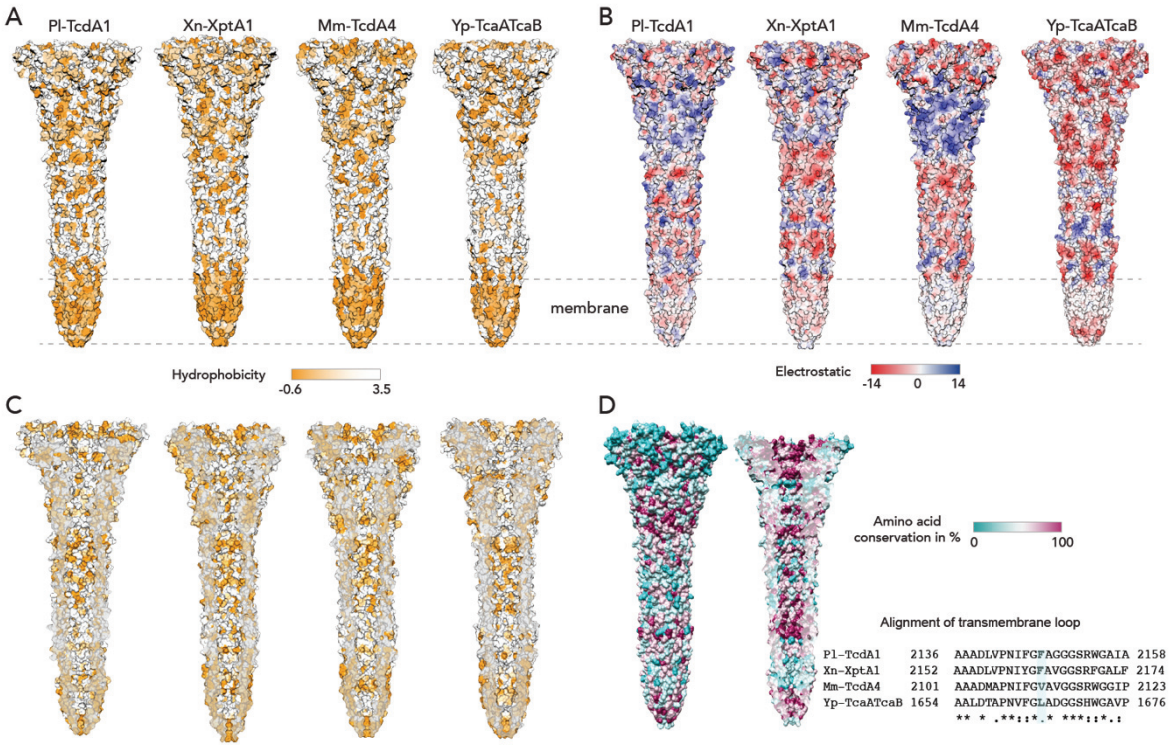


Figure 48: Biophysical properties of the TcA channel domains. A) Surface representation of the channels of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB, depicting the hydrophobic surface potential. Hydrophobic residues are colored in yellow. B) Surface of the TcA channels showing the electrostatic Coulomb potential at pH 7. Positively charged (14 kcal/mol) and negatively charged (-14 kcal/mol) residues are presented in blue and red, respectively. C) Cross-sections of the lumens of the channels, demonstrating the hydrophobic surface potential of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. Hydrophobic residues are colored in yellow. D) Sequence conservation of the channel domains of the four TcAs shown for the channel surface (left) and the lumen of the channel (right). Conserved residues are shown in magenta. Sequence alignment of the transmembrane loop. The residue at the tip of the loop is highlighted in light blue.

The toxins from insect pathogenic bacteria have both a Phe residue at the tip of the loop, whereas the toxins from human pathogenic bacteria have a Val residue (Mm-TcdA4) or a Leu residue (Yp-TcaATcaB). Recent molecular dynamics simulation studies of the Pl-TcdA1 pore state in the membrane revealed an interaction of the lower loop of the transmembrane region with the environmental lipids[44]. Thus, the distinct residue at the tip of the loop could indicate a different interaction pattern with lipids and therefore a specific lipid environment for the different TcAs.

The electric charges of the channel surface (i.e. interaction sites with the α-helical shells) are equally distributed in Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. Close to the TcB-interacting domain the proteins are positively charged, predominantly negatively charged at the central part and they contain a weakly charged transmembrane stretch (Figure 48B). In contrast, Yp-TcaATcaB has more neutral patches and mostly negative charges at the channel exterior. Especially close to the TcB-interacting domain the charges differ in comparison to the other TcAs, which could result in different specific interactions with the shell domain. This goes along with the observed conservation of the shell domain (chapter 5.4.1). The upper peripheric region of the small lobe, which interacts with channel domain, is less conserved.

The inner channel radius shows a comparable diameter profile for the four TcAs (Figure 49B). Starting with a large diameter of around 26 Å for all TcAs at the TcB-interacting domain, the channel radius decreases along the transport direction of the toxin. Between 40 and 60 Å below the TcB-binding domain, all four TcAs have a narrow radius of about 5 Å. As described in the introduction (chapter 2.2.3), the cleaved C-terminus of *P. luminescens* TcC components is a hyper variable region (hvr) that is unfolded within the TcBTcC cocoon and locates also in the channel domain upon TcBTcC binding to TcA[84]. Thus, a narrow diameter of 10 Å present in Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB, indicate that also for these toxins the hvr of TcC component is translocated in an unfolded manner. Moreover, a major difference between the toxins can be found at the tip of the channel. Here, Pl-TcdA1 has a constriction site to 3.9 Å diameter close to residue Y2163[43]. This residue is not conserved between the four TcAs. In fact, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB have smaller residues at this position (I2179, A2128 and A1681, respectively). In general, the channel diameters of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB close to the tip are wider with minima of 8.6, 8.9 and 8.4 Å, respectively. This is more than two times wider than the constriction of the Pl-TcdA1 channel (Figure 49C). However, the pore diameter of Pl-TcdA1 is increased up to 42 Å after membrane insertion by cylindrical rearrangement of the ten α-helices[44]. Thus, the pore opens and allows the translocation of the unfolded toxic enzyme.

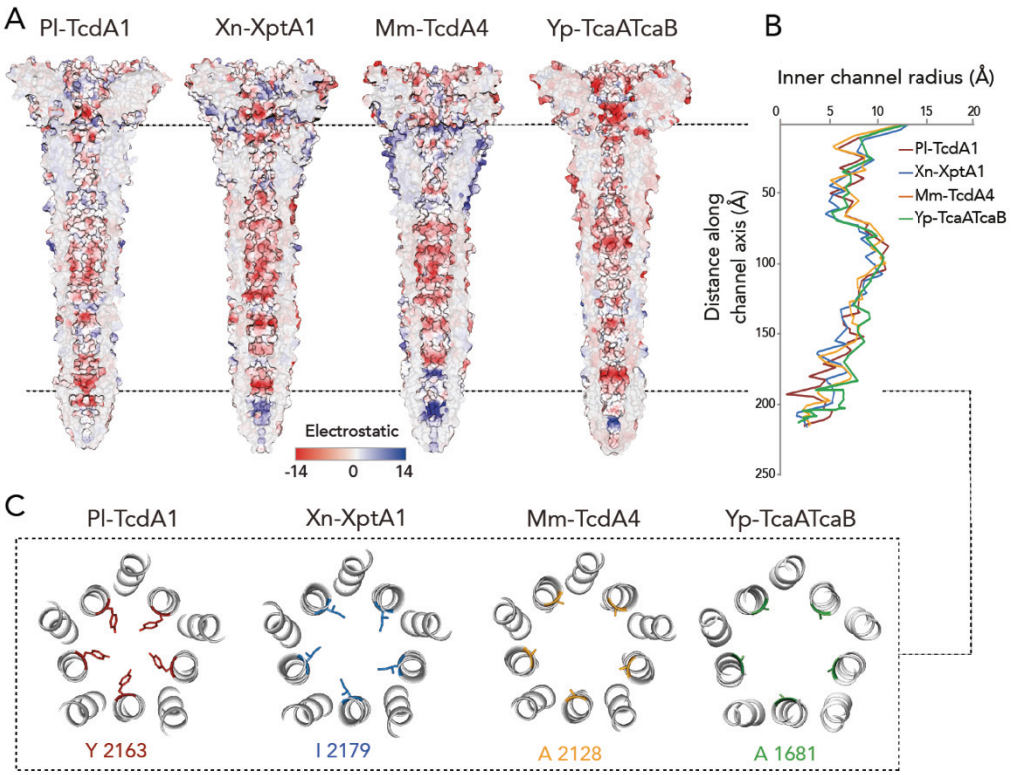


Figure 49: Electrostatic potentials and constrictions of the TcA channels. A) Cross-sections of the channel lumen, demonstrating the electrostatic Coulomb potential of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB at pH 7. Positively charged (14 kcal/mol) and negatively charged (-14 kcal/mol) residues are colored in blue and red, respectively. B) Graph representing the inner channel radius of Pl-TcdA1 (red), Xn-XptA1 (blue), Mm-TcdA4 (orange) and Yp-TcaATcaB (green). C) Depiction of residues at the position of the Pl-TcdA1 channel constriction.

The lumen of the channel of Pl-TcdA1 is mainly negatively charged and was shown to be cation selective[43,75]. In comparison, the channel lumen of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB have also negative patches, especially in the belly region 70-120 Å distance along the channel axis (Figure 49A and B). Contrarily, the tips of the channel lumen of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB are positively charged, which is more pronounced in Mm-TcdA4, whereas Pl-TcdA1 is negatively charged (Figure 49A). The different charges at the tip of the channel lumen could imply a selectivity of the toxins for species-specific cargoes. In the case of *P. luminescens*, two hvrs were identified as ADP-ribosyltransferases[77]. However, not much is known about the enzymatic function of the TcC components of *M. morganii*, *X. nematophila* and *Y. pseudotuberculosis*. Based on sequence alignments with potential toxic components of *X. nematophila*, the TcC component of this bacteria could exhibit a putative dehydrogenase activity. For *Y. pseudotuberculosis* TcC components, an adenylate cyclase or phosphatase activity was suggested and TcC from *M. morganii* has sequence similarity to cytotoxic necrotizing factor (CNF) of *E. coli*[78,80].

Thus, the C-terminal sequences of the TcC components from different organisms were compared regarding their charged residues (Figure 50). The C-terminus of TcCs from *P. luminescens* have positive

charges corresponding to arginine and lysine residues. However, the comparison between the different TcCs revealed that there is no specific region of negatively charged amino acids in the sequences of the hvrs of *X. nematophila*, *M. morganii* and *Y. pseudotuberculosis*. Subsequently, the suggested ionic specific transport based on the reversed charge at the tip of the channel of the TcAs from these three bacteria, is not confirmed. The TcC component of *M. morganii* for example has seven positively charged amino acids at its C-terminus although Mm-TcdA4 has a positive patch at the tip of the channel lumen (Figure 49A, Figure 50). Moreover, the two TcC components from *Y. pseudotuberculosis* have both, negatively and positevly charged residues at the C-terminus, which is also observed in *X. nematophila* TcCs (Figure 50). This rise the question whether the TcAs from different bacteria have also an ionic specific cargo transport or if the translocation is independent of the cargo charge. To answer this question, electrophysiological studies of the TcAs from different bacteria as well as further studies on the TcC components of other bacteria are required.

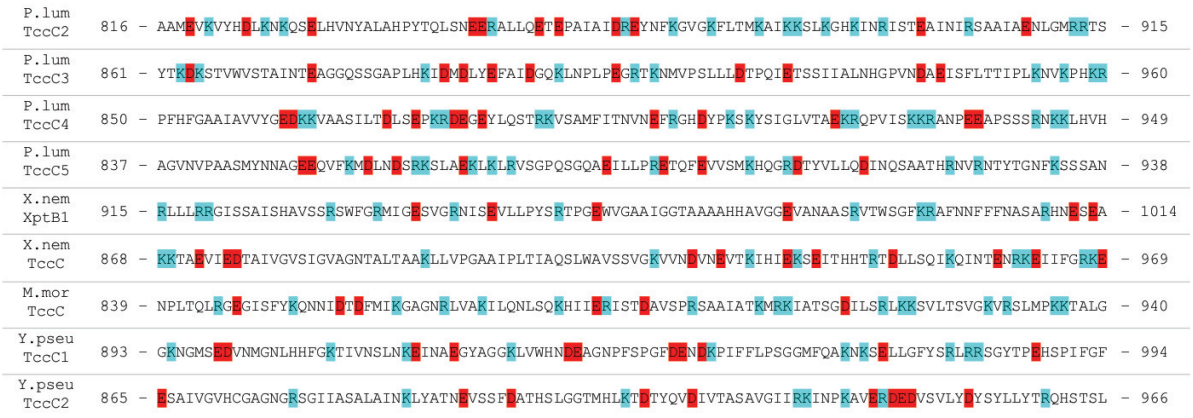


Figure 50: Alignment of last 100 C-terminal amino acids of TcC components from different bacterial species. The amino acids highlighted in red are negatively charged whereas the amino acids highlighted in blue are positively charged at neutral pH. In total, four TcCs from *P. luminescens*, two from *X. nematophila*, one from *M. morganii* and two from *Y. pseudotuberculosis* were aligned.

Comparing the mechanism of cargo transport of Pl-TcdA1 to the cation selective anthrax toxin, some differences become evident. For anthrax, an N- to C-terminal transport of the LF or EF is observed and the driving force of the translocation is a Brownian ratchet due to the pH gradient between the endosomal and cytoplasmic compartments. Thus, protonation before and deprotonation after the translocation of anionic proteins such as LF and EF is necessary. A phi-clamp composed of seven phenylalanines in anthrax toxin blocks the passage of ions that would destroy the pH gradient[40]. In comparison, the ADP-risbosyltransferase of *P. luminescens* TcC components (TccC3 and TccC5) have a cationic nature and an isoelectric point of 9.68 and 8.65, respectively[43]. Therefore, no pH gradient for translocation from the endosomal compartment to the cytosol is needed. Furthermore, the translocation across the channel is only possible in one direction, since the TcBTcC cocoon closes the endosomal

end of the channel. However up to date, not much is known about the interaction of the ADP-ribosyltransferase with the channel in the prepore state as well as the translocation of the cargo across the channel. Recently it was shown that the ADP-ribosyltransferase of Pl-TcdA1 is located in the channel of TcA upon binding of the TcBTcC cocoon to TcA[84]. In contrast to anthrax toxin, a C- to N-terminal translocation into the cytosol is proposed for the hvr of *Photorhabdus* TcC components. A similar C- to N-terminal translocation was previously suggested for the cation selective diphteria toxin[199,200].

Additionally, chaperons like Hsp90 are thought to facilitate the translocation of *P. luminescens* TcC components into the cytoplasm[201]. Such a chaperon mediated translocation of ADP-ribosylating enzymes was also observed in binary toxins from *Clostridia*[202]. It was proposed that the translocation occurs simultaneously to refolding of the cargo. This could be the driving force for translocation, since a folded protein would block the channel and inhibits sliding back into the channel[201].

5.4.4 TcB-binding domain

Like the pore domain, the TcB-binding domain is crucial for the translocation of the toxic cargo, as it interacts with the TcBTcC cocoon and functions as the initial toxin translocation point in TcA. Recently, new insights into the interaction of the TcB-binding domain and the TcBTcC cocoon were gained[84] (chapter 2.2.3.2). Upon binding, major conformational changes in the TcB component of the TcBTcC cocoon occur, whereas no fundamental structural changes in the TcB-binding domain of TcA were observed.

In this chapter, the TcB-binding domains of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB were compared to the Pl-TcdA1 prepore TcB-binding domain. In Pl-TcdA1, this domain mainly consists mainly of β-sheets and has a similar eight antiparallel β-strand sandwich fold as already shown for the receptor-binding domains (chapter 5.4.1.2). This folding pattern is also present in the TcB-binding domains of the other three TcAs (Figure 51A). Overall, the TcB-binding domains of the different TcAs reveal low structural variances in the central β-sheet region, with respect to both, amino acid conservation and RMSD (Figure 51B and C). Almost all residues are conserved in the central β-sheets of the four TcAs. However, loops facing outwards are less conserved and are slightly longer in Pl-TcdA1 (Figure 51B). Altogether, the identical fold and the high sequence conservation of this domain indicate the possibility to form chimeric holotoxins.

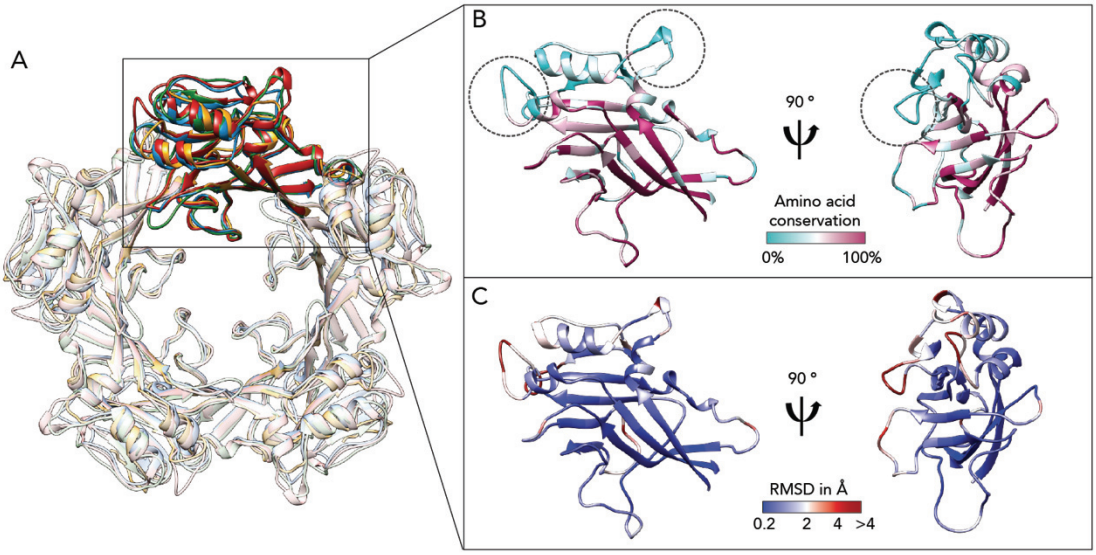
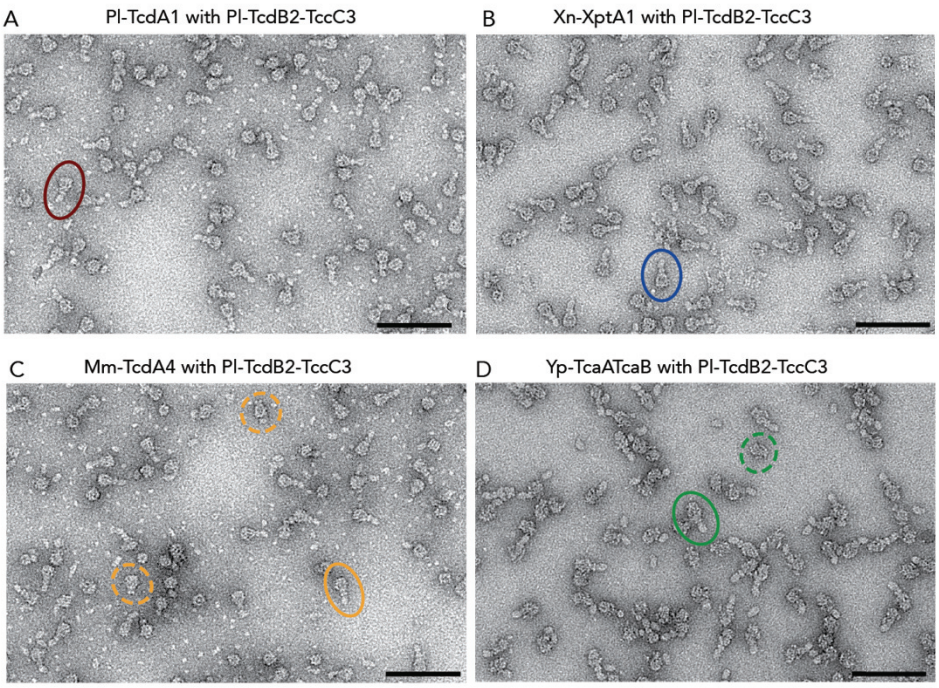


Figure 51: TcB-binding domain of TcAs. A) Overlay of the TcB-binding domains of the four TcAs Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB in the context of the pentamer colored in red, blue, orange and green, respectively. One protomer is highlighted. B) Protomer of Pl-TcdA1 demonstrating the sequence conservation of residues between Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. Conserved residues are depicted in magenta. The three non-conserved loops are highlighted with dashed circles. C) Mean RMSD values for the TcB-binding structures of Pl-TcdA1, Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. Regions with low RMSD values are colored in blue and regions with a higher RMSD in red.

Therefore, the interaction between the TcBTcC cocoon of *P. luminescens* and the different TcAs were analyzed. Pl-TcdB2 (TcB component) and Pl-TccC3 (TcC component) were used for the interaction studies, since it was shown that the complex together with Pl-TcdA1 forms a functional holotoxin[77]. The binding affinities of the different TcAs to the Pl-TcdB2TccC3 (B2C3) were determined using biolayer interferometry (BLI). With this method, the binding kinetics between an immobilized ligand on biosensors and an analyte in solution can be determined. Thus, the smaller Pl-TcdB2-TccC3 (B2C3) cocoon was biotinylated and immobilized on streptavidin biosensors whereas the different TcA pentamers were used as analytes in solution. Association and dissociation signals were obtained for all four TcAs (appendix Figure 59) and the K_D, k_{on} and k_{off} were determined by globally fitting a 1:1 binding model. The highest affinity was measured for the native complex of Pl-TcdA1 and B2C3 with a K_D of 0.63 ± 0.027 nM (Figure 52E). Interestingly, the interaction between B2C3 and the more distinct Yp-TcaATcaB from *Y. pseudotuberculosis* has a comparable high affinity with 0.95 ± 0.02 nM. However, the k_{on} of Yp-TcaATcaB is one order of magnitude lower compared to Pl-TcdA1. Thus, the interaction is strong but the formation of the interaction is less preferred. In contrast, the affinity of Xn-XptA1 was less than one order of magnitude lower compared to Pl-TcdA1. Here, the differences in affinity between Pl-TcdA1 and Xn-XptA1 are mostly off-rate determined (7.22 ± 0.13×10^{-4} s^{-1} and 3.82 ± 0.04×10^{-3} s^{-1}, respectively), while the difference in k_{on} is smaller (1.14 ± 0.04×10^{6} M^{-1} s^{-1} and 1.92 ± 0.14×10^{6} M^{-1} s^{-1}, respectively). The slightly higher sequence variability of the TcB-binding domain of Mm-TcdA4 compared to Pl-TcdA1 translates into a lower binding affinity with 11.3 ± 0.2 nM.

To further evaluate the interaction between the chimeric holotoxins, the samples were also analyzed by negative stain EM. The respective components were incubated and unbound B2C3 was removed by size exclusion chromatography (see chapter 4.1.9). Holotoxin complexes for all four TcAs with B2C3 were present and the B2C3 cocoon is clearly visible on top of the translocation channel of the TcAs (Figure 52A-D, solid line). However, in case of Mm-TcdA4 and Yp-TcaATcaB, some TcAs remained without B2C3 cocoon were still present as expected due to the lower affinity of Mm-TcdA4 to B2C3 and the smaller k_{on} for both TcAs (Figure 52C and D, dashed line).



E

	Pl-TcdA1	Xn-XptA1	Mm-TcdA4	Yp-TcaATcaB
K_D (nM)	0.63 ± 0.027	1.99 ± 0.15	11.3 ± 0.2	0.95 ± 0.020
k_{on} ($M^{-1}s^{-1}$)	1.14 ± 0.04*10^6	1.92 ± 0.14*10^6	1.71 ± 0.03*10^5	6.59 ± 0.10*10^5
k_{off} (s^{-1})	7.22 ± 0.13*10^{-4}	3.82 ± 0.04*10^{-3}	1.94 ± 0.02*10^{-3}	6.28 ± 0.09*10^{-4}

Figure 52: Visualization of chimeric holotoxin formation in negative stain EM. Negative stain electron micrographs after complex formation of four TcAs, namely Pl-TcdA1 (A), Xn-XptA1 (B), Mm-TcdA4 (C) and Yp-TcaATcaB (D), with *P. luminescens* TcdB2-TccC3 and size exclusion chromatography. One example of an holotoxin is highlighted for each complex. For Mm-TcdA4 and Yp-TcaATcaB, particles without TcdB2-TccC3 are marked with dashed circles. Scale bar, 100 nm. E) BLI results from the interaction between different TcAs and Pl-TcdB2-TccC3 including the binding affinity (K_D), the on- and the off-rate (k_{on}, k_{off}, respectively). A global fit according to a 1:1 binding model was applied, including 6-7 individual curves. The obtained parameters are the mean value ± the error of the fit.

Furthermore, the toxicity of the formed chimeric holotoxins was investigated on HEK293T cells. The negative control with buffer only shows no effect on the cells, which are still adherent (Figure 53A). In contrast, all four toxin complexes exhibit toxicity and therefore are fully functional toxins (Figure 53). The most effective holotoxin was the native complex of Pl-TcdA1 and *P. luminescens* TcdB2-TccC3. A

holotoxin concentration of 0.5 nM is sufficient to kill the cells, which is pronounced by the round shape phenotype of the cells. Upon translocation of the hvr of Pl-TccC3, actin is ADP-ribosylated, which ultimately leads to actin polymerization and cell rounding. This effect was also observed in previous studies with Tc toxins from *P. luminescens*[203] and for other toxins in context with an ADP-ribosyltransferase[44,204]. Moreover, the cell number is reduced after toxin incubation and a clustering effect of the cells can be observed (Figure 53B).

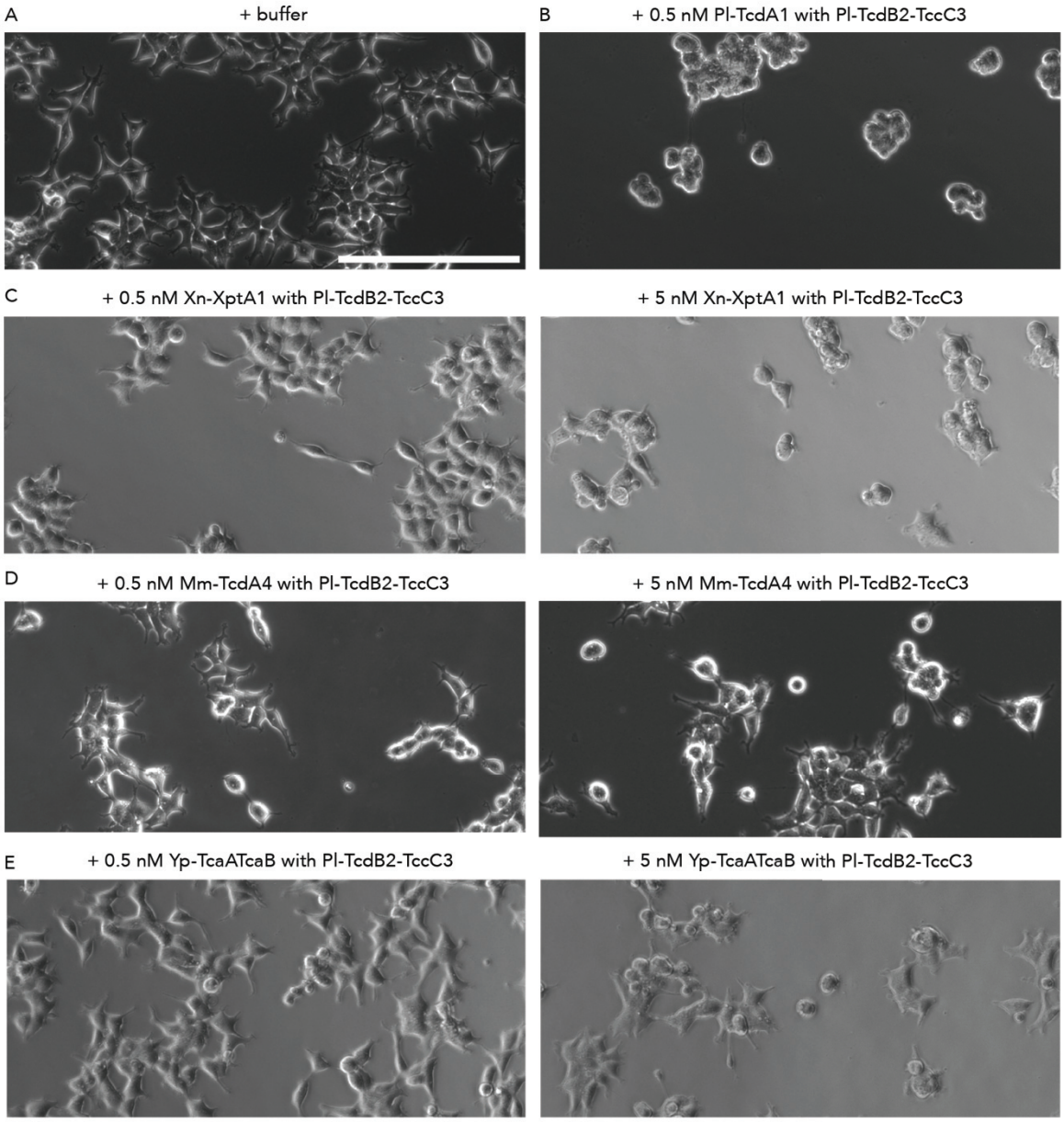


Figure 53: Intoxication of HEK cells with chimeric holotoxins. Cell imaging of the intoxication of HEK293T cells with the different chimeric holotoxins. Cells were imaged 20 hours after infection. A) Negative control with buffer only. B) Positive control with the native *P. luminescens* Pl-TcdA1 and TcdB2-TccC3 complex at a concentration of 0.5 nM holotoxin. C) Effect of Xn-XptA1 with *P. luminescens* TcdB2-TccC3 holotoxin complex at concentrations of 0.5 (left) and 5 nM (right). D) Cell images after incubation of Mm-TcdA4 with *P. luminescens* TcdB2-TccC3 at concentrations of 0.5 (left) and 5 nM (right). E) Cells after infection with 0.5 nM (left) and 5 nM (right) Yp-TcaATcaB with *P. luminescens* TcdB2-TccC3 holotoxin. The scale bar is 200 µm.

However, the toxicity of the chimeric holotoxins decreases with increasing variety of the TcAs. For the chimeric holotoxins of Pl-TcdB2-TccC3 and Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB, a concentration of 0.5 nM results in less toxicity compared to Pl-TcdA1. The majority of the cells is still viable and adherent. However, upon 10-fold increased holotoxin concentration, a loss of cell viability was observed (Figure 53C-D). The cell number is reduced in case of Xn-XptA1 with Pl-TcdB2-TccC3 and Yp-TcaATcaB with Pl-TcdB2-TccC3. Contrarily, for Mm-TcdA4 and Pl-TcdB2-TccC3 at 5 nM concentration, the effect is similar compared to 0.5 nM holotoxin concentration.

Altogether, the results demonstrate that the high degree of structural conservation in the TcB-binding domains allows formation of active holotoxins with TcA and TcB from different species. Furthermore, the toxins from different organisms were able to penetrate and therefore interacting with the membrane and the integrated receptors of one cell type. Additionally, it shows that the TcAs from *X. nematophila, M. morganii* and *Y. pseudotuberculosis* were able to translocate the ADP-ribosyltransferase into the cytosol, which has a cationic nature. However, differences in the toxic potency between the native and the chimeric holotoxins suggest that either the translocation of non-native cargo or the interaction with the membrane and pore formation, or both was less favored for the chimeric holotoxins. In addition, the differences in the k_{off}-values could implement that the chimeric holotoxins may disassemble and therefore leading to a decreased toxicity. Furthermore, the question of a species-specific receptor could not be answered with this experiment, since no prediction about the interaction of the TcAs with the various present receptors on the cell membrane could be made. Therefore, it could be, that the potency of the toxins will increase upon binding to a target specific receptor.

5.5 Conclusive discussion

The family of pore-forming toxins is widely spread in prokaryotes and eukaryotes. Amongst the members of the α-PFTs are the tripartite Tc toxins. Interestingly, genes encoding these Tc toxins are present in a variety of bacteria including entomopathogenic as well as plant- and human-pathogenic strains. Detailed structural studies have been achieved for the *P. luminescens* Tc toxins, however, there are no detailed structural information about Tc toxins from different organisms available. Especially the TcA component, which is the translocation machinery and thus is responsible for host specific interactions, is an interesting target for studying the Tc toxin family and their underlying mechanism. Until now, exclusively TcAs from insect pathogenic bacteria have been structurally characterized[61,83]. Thus it remains unknown, if Tc toxins from different bacteria share a common architecture and mechanism.

During this work, I expressed, purified and structurally characterized three TcA components from different bacteria species by negative stain EM and cryo-EM. The obtained resolution ranged between 2.8 and 3.3 Å for the three TcAs and allowed the building of atomic models, which enabled a detailed comparison of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB with each other and the already known structure of Pl-TcdA1. In contrast to the previously published low-resolution structure of Xn-XptA1[83], the here obtained high-resolution structure of the TcA from *X. nematophila* shows the same domain arrangement and pentameric architecture as Pl-TcdA1 and the other studied TcAs. Thus, the overall comparison revealed a similar architecture of all investigated TcAs with three main functional domains, namely the channel domain, the α-helical shell and the β-sheet domains that potentially fulfill the function of receptor-binding domains (RBDs).

Detailed analysis of the helical shell identified a conserved molecular trefoil knot in the TcAs. Considering that the folding mechanism of a knotted protein is more complex and decreases the folding rate[173,176], the presence and conservation among all analyzed Tc toxins is highly interesting. The knot most likely stabilizes the stretched linker domain conformation in the prepore state and the complete toxin complex during the massive conformational change from the prepore to the pore state, as it was shown for other proteins before[177,178]. The channel domain is highly conserved and reveals comparable channel diameters for all Tc toxins, thereby indicating a similar translocation mechanism of unfolded cargo proteins. Aside the channel dimensions, the structures of the different TcAs revealed similar sizes of transmembrane regions, of which only the lower loop is conserved. However, the last residue at the tip of the loop differs between the TcAs and implies a distinct interaction with the surrounding lipid head groups as it could be shown for Pl-TcdA1 with molecular dynamic simulations[44].

While the channel domain and the majority of the helical shell share high sequence and structure similarity, the RBDs comprise the major differences between the TcAs. The RBDs of Pl-TcdA1, Xn-XptA1

and Mm-TcdA4 share the same topology and structural homology to the receptor-binding domains of anthrax and diphtheria receptor-binding domains[12,43,86]. However, the RBDs of the Tc toxins within each other have low sequence identity. Like described and discussed previously for the nicotinic receptors as an example, that show low sequence conservation but a common fold[182,183], the RBDs of TcAs potentially function as receptor binding domains. Host-specificity here might be encoded by varying sequences, providing altered molecular interaction patterns. Interestingly, the RBDs are absent in the Yp-TcaATcaB sequence. This suggests an interaction with lipid or membrane patches to localize Yp-TcaATcaB at the membrane as it can be observed for various toxins. The pore-forming toxin lysenin, as example, interacts specifically with sphingomyelin in the plasma membrane whereas members of the cholesterol-dependent cytolsysin family bind to cholesterol-rich membranes[205,206]. Moreover, localization of Yp-TcaATcaB could be mediated by the neuraminidase-like domain by interacting with gangliosides and glycoproteins such as the influenza virus or the cholera toxin from *Vibrio cholerae*[185-187]. In addition, a unique coiled coil domain, connecting the two subunits TcaA and TcaB in Yp-TcaATcaB, was identified. The coiled coil domain is assumed to assist the complex stabilization between the two subunits in the prepore state. Moreover, coiled coil domains were observed to act as mediators for conformational changes in saposin B and Rad50[168-170]. Therefore, the coiled coil domain in Yp-TcaATcaB could additionally facilitate the massive conformational change that would be necessary to enable the pore domain to penetrate a lipid membrane.

Another focus of this thesis was to investigate the transition of the individual TcAs from prepore to pore state. With the help of black lipid bilayer experiments and reconstitution of the proteins into lipid nanodiscs, it was demonstrated that the TcAs are forming ion conductive pores that undergo a massive conformational change to reach the final active state, the pore state. Interestingly, for Yp-TcaATcaB, the black lipid bilayer experiments revealed a more active membrane insertion at pH 6 compared to Pl-TcdA1, Xn-XptA1 and Mm-TcdA4. But, moreover and highly probable, the Yp-TcaATcaB toxin complex is destabilized at extreme pH values of pH 4 and pH 11. This is consistent with the results I obtained by negative stain studies and thermal stability measurements for Yp-TcaATcaB at different pH-values. Thus, these results question if Yp-TcaATcaB is able to entry the host cell via the endocytotic pathway, where pH values of pH 4.8-6 are present in the late endosomes[207]. Therefore, and most likely, Yp-TcaATcaB needs an interaction with specific lipids or membrane patches to be stabilized in an acidic environment. However, the underlying mechanism has to be further investigated to gain more information of the pH-dependent stability of Yp-TcaATcaB. Conclusively, I could show that the TcAs exhibit a pH-dependent prepore-to-pore transition. Moreover, in the context of a conserved ionic electrostatic lock, one interaction side was identified, which showed to be enormously important for the pentamer complex stability of Pl-TcdA1. Upon inhibition of the interaction between Glu1086 and Arg1166 in the shell region of adjacent protomers, the protein complex is destabilized. Most likely, the protomer-protomer

interface is weakened by repulsive charges upon pH shift and the interaction pair of Glu1086 and Arg1166 could act as a latch that lastly results in an opening of the shell. Further mutational studies of the other TcAs are needed to gain information of the impact of this interaction on Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB. Additionally, the neuraminidase-like domain was thought to act as the electrostatic lock in Pl-TcdA1 before [43]. However, the structures of Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB show, that in those proteins the channel domain is not entirely enclosed by the neuraminidase-like domain. Therefore it is concluded, that the pH-dependent shell opening and prepore to pore transition is not exclusively mediated by one conserved electrostatic lock but multiple electrostatic interactions that differ to some extend between the distinct TcAs.

In accordance, I observed enormous differences regarding the efficiency of prepore-to-pore transition upon pH shifts in the four TcAs. Unlike previous results obtained for Pl-TcdA1, negative stain studies suggest a need for additional cofactors to trigger the prepore-to-pore transition in Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB *in vitro*. Recent studies indicate that a *P. luminescens* Tc toxin is activated by proteolytic processing of Pl-TcdA1, resulting in increased affinity to cell membrane receptors[203]. Thus probably, the three TcA components Xn-XptA1, Mm-TcdA4 and Yp-TcaATcaB might have to be activated prior to transition to reach comparable efficiencies as are observed for other pore-forming toxins such as anthrax[11] or aerolysin[25]. Moreover, the protease cleavage site that was identified in Pl-TcdA1 is not conserved within the TcAs and could imply different types of proteases to activate the different TcAs. Consequently, TcAs from distinct bacteria could have evolved sequence and structure based specific regions for a pH-dependent mechanism according to the host specificity of the bacteria.

Another possible way of TcA activation is a receptor binding-mediated conformational change. Examples, in which conformational changes are induced upon receptor binding, are found in multiple proteins, including pore-forming toxins as well as viruses[7,208,209]. In the example of a calicivirus, upon receptor binding, the RBDs rearrange and allow the portal-like pore formation, which results in a massive conformational change[208]. Similarly, an interaction with the receptor could result in exposure of the pore domain for the Tc toxins, due to conformational changes of the RBDs and the neuraminidase-like domain, that enclose the channel and thus, enhance an efficient pore formation at the membrane. A comparable mechanism is present in intermedilysin (ILY), a pore-forming toxin belonging to the class of cholesterol-dependent cytolsysins (CDC). Binding to the receptor CD59 promotes oligomerization[209] to enable a prepore complex and is in addition responsible for structural rearrangements, which are required for the pore formation[205]. An intermediate state, between prepore and pore state, was observed for Xn-XptA1 and Mm-TcdA4, which has not been reported for Pl-TcdA1 before or was present in Yp-TcaATcaB. In the intermediate state, the RBDs interact with the lipids within liposomes or lipid nanodiscs. Nevertheless, a pH shift to basic (pH 11) or acidic (pH 5) pH together with these protein-

lipid interactions are insufficient to trigger the prepore-to-pore transition, thereby trapping the TcAs in an intermediate state.

Furthermore, it was demonstrated by BLI-measurements and intoxication assays, that active chimeric holotoxins can be formed consisting of TcAs from varying origins along with the TcBTcC cocoon from *P. luminescens*. The possibility to form active chimeric holotoxins was already shown for TcA from *X. nematophila* and TcBTcC from *P. luminescens*[74]. However, this work revealed that the strong similarity in sequence and structural features within the TcB-binding domain of all studied TcAs led to the formation of mixed active holotoxins from human and insect pathogenic organisms. Differences in toxicity presumably result from either unfavored translocation of the non-native cargo protein or altered efficiencies in TcA pore formation in the chimeric holotoxins. In addition, since no receptor has been identified for the toxins so far, differences in toxicity moreover can result from less effective binding of toxins to membrane-associated receptors. Differences in the k_{off}-values further indicate a faster complex dissociation of chimeric holotoxins, which may lead to a decreased cytotoxicity.

Summarizing, this thesis reveals the first detailed structural insights into the transporter machinery TcA derived from various prokaryotic organisms. Further studies, including *in vitro and in vivo* experiments, will be needed to identify the host-specific receptors, decipher the host-specific interactions, the uptake mechanism and the transition of the TcAs into their active states. With this knowledge, the Tc toxins could be implemented as powerful tools in future drug delivery system

6 References

1. Garcia-Saez, A. J. The secrets of the Bcl-2 family. *Cell Death Differ.* **19,** 1733–1740 (2012).
2. Pipkin, M. E. & Lieberman, J. Delivering the kiss of death: progress on understanding how perforin works. *Curr. Opin. Immunol.* **19,** 301–308 (2007).
3. Bernheimer, A. W. & Rudy, B. Interactions between membranes and cytolytic peptides. *Biochim. Biophys. Acta* **864,** 123–141 (1986).
4. Bischofberger, M., Gonzalez, M. R. & van der Goot, F. G. Membrane injury by pore-forming proteins. *Curr. Opin. Cell Biol.* **21,** 589–595 (2009).
5. Gonzalez, M. R., Bischofberger, M., Pernot, L., van der Goot, F. G. & Frêche, B. Bacterial pore-forming toxins: The (w)hole story? *Cell. Mol. Life Sci.* **65,** 493–507
6. Los, F. C. O., Randis, T. M., Aroian, R. V. & Ratner, A. J. Role of pore-forming toxins in bacterial infectious diseases. *Microbiol. Mol. Biol. Rev.* **77,** 173–207 (2013).
7. Peraro, M. D. & van der Goot, F. G. Pore-forming toxins: ancient, but never really out of fashion. *Nat Rev Micro* **14,** 77–92 (2016).
8. Bischofberger, M., Iacovache, I. & Gisou van der Goot, F. Pathogenic Pore-Forming Proteins: Function and Host Response. *Cell Host & Microbe* **12,** 266–275 (2012).
9. Iacovache, I., Bischofberger, M. & van der Goot, F. G. Structure and assembly of pore-forming proteins. *Current Opinion in Structural Biology* **20,** 241–246 (2010).
10. Cosentino, K., Ros, U. & García-Sáez, A. J. Assembling the puzzle: Oligomerization of α-pore forming proteins in membranes. *Biochim. Biophys. Acta* **1858,** 457–466 (2016).
11. Young, J. A. T. & Collier, R. J. Anthrax Toxin: Receptor Binding, Internalization, Pore Formation, and Translocation. *Annual Review of Biochemistry* **76,** 243–265 (2007).
12. Choe, S. *et al.* The crystal structure of diphtheria toxin. *Nature* **357,** 216–222 (1992).
13. Lesieur, C., V csey-Semj n, B., Abrami, L., Fivaz, M. & Gisou van der Goot, F. Membrane insertion: The strategies of toxins (Review). *Molecular Membrane Biology* **14,** 45–64 (2009).
14. Gilbert, R. J. C. Pore-forming toxins. *Cell. Mol. Life Sci.* **59,** 832–844 (2002).
15. Lakey, J. H. *et al.* Membrane insertion of the pore-forming domain of colicin A. A spectroscopic study. *Eur. J. Biochem.* **196,** 599–607 (1991).
16. Tanaka, K., Caaveiro, J. M. M., Morante, K., González-Mañas, J. M. & Tsumoto, K. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. *Nat Commun* **6,** 6337 (2015).
17. Kristan, K. C., Viero, G., Dalla Serra, M., Macek, P. & Anderluh, G. Molecular mechanism of pore formation by actinoporins. *Toxicon* **54,** 1125–1134 (2009).
18. Parker, M. W., Pattus, F., Tucker, A. D. & Tsernoglou, D. Structure of the membrane-pore-forming fragment of colicin A. *Nature* **337,** 93–96 (1989).
19. Wallace, A. J. *et al.* E. coli hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100,** 265–276 (2000).
20. Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R. & Ban, N. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459,** 726–730 (2009).
21. Roderer, D. & Glockshuber, R. Assembly mechanism of the α-pore-forming toxin cytolysin A from Escherichia coli. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **372,** 20160211 (2017).
22. Zhang, R.-G. *et al.* The 2.4 Å Crystal Structure of Cholera Toxin B Subunit Pentamer: Choleragenoid. *Journal of Molecular Biology* **251,** 550–562 (1995).
23. Schubert, E., Vetter, I. R., Prumbaum, D., Penczek, P. A. & Raunser, S. Membrane insertion of α-xenorhabdolysin in near-atomic detail. *eLife* **7,** 30 (2018).
24. Vigneux, F. *et al.* The xaxAB Genes Encoding a New Apoptotic Toxin from the Insect

Pathogen Xenorhabdus nematophila Are Present in Plant and Human Pathogens. *Journal of Biological Chemistry* **282,** 9571–9580 (2007).

25. Abrami, L. *et al.* The pore-forming toxin proaerolysin is activated by furin. *Journal of Biological Chemistry* **273,** 32656–32661 (1998).
26. Foletti, D. *et al.* Mechanism of action and in vivo efficacy of a human-derived antibody against Staphylococcus aureus α-hemolysin. *Journal of Molecular Biology* **425,** 1641–1654 (2013).
27. Collier, R. J. Membrane translocation by anthrax toxin. *Molecular Aspects of Medicine* **30,** 413–422 (2009).
28. Song, L. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274,** 1859–1866 (1996).
29. Shepard, L. A. *et al.* Identification of a membrane-spanning domain of the thiol-activated pore-forming toxin Clostridium perfringens perfringolysin O: an alpha-helical to beta-sheet transition identified by fluorescence spectroscopy. *Biochemistry* **37,** 14563–14574 (1998).
30. Johnson, B. B. & Heuck, A. P. Perfringolysin O structure and mechanism of pore formation as a paradigm for cholesterol-dependent cytolysins. *Subcell. Biochem.* **80,** 63–81 (2014).
31. Ohno-Iwashita, Y., Iwamoto, M., Ando, S. & Iwashita, S. Effect of lipidic factors on membrane cholesterol topology--mode of binding of theta-toxin to cholesterol in liposomes. *Biochim. Biophys. Acta* **1109,** 81–90 (1992).
32. Nelson, K. L., Raja, S. M. & Buckley, J. T. The glycosylphosphatidylinositol-anchored surface glycoprotein Thy-1 is a receptor for the channel-forming toxin aerolysin. *Journal of Biological Chemistry* **272,** 12170–12174 (1997).
33. Odumosu, O., Nicholas, D., Yano, H. & Langridge, W. AB toxins: a paradigm switch from deadly to desirable. *Toxins* **2,** 1612–1645 (2010).
34. Collier, R. J. & Young, J. A. T. Anthrax Toxin. *http://dx.doi.org/10.1146/annurev.cellbio.19.111301.140655* **19,** 45–70 (2003).
35. Bradley, K. A., Mogridge, J., Mourez, M., Collier, R. J. & Young, J. A. Identification of the cellular receptor for anthrax toxin. *Nature* **414,** 225–229 (2001).
36. Milne, J. C., Furlong, D., Hanna, P. C., Wall, J. S. & Collier, R. J. Anthrax protective antigen forms oligomers during intoxication of mammalian cells. *Journal of Biological Chemistry* **269,** 20607–20612 (1994).
37. Kintzer, A. F. *et al.* The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes. *Journal of Molecular Biology* **392,** 614–629 (2009).
38. Pimental, R.-A. L., Christensen, K. A., Krantz, B. A. & Collier, R. J. Anthrax toxin complexes: heptameric protective antigen can bind lethal factor and edema factor simultaneously. *Biochem. Biophys. Res. Commun.* **322,** 258–262 (2004).
39. Zhang, S., Finkelstein, A. & Collier, R. J. Evidence that translocation of anthrax toxin's lethal factor is initiated by entry of its N terminus into the protective antigen channel. *Proceedings of the National Academy of Sciences* **101,** 16756–16761 (2004).
40. Krantz, B. A. *et al.* A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore. *Science* **309,** 777–781 (2005).
41. Benz, R. in *The Comprehensive Sourcebook of Bacterial Protein Toxins* 605–626 (Elsevier, 2015). doi:10.1016/B978-0-12-800188-2.00021-5
42. Gatsogiannis, C. *et al.* A syringe-like injection mechanism in Photorhabdus luminescens toxins. *Nature* **495,** 520–523 (2013).
43. Meusch, D. *et al.* Mechanism of Tc toxin action revealed in molecular detail. *Nature* **508,** 61–65 (2014).
44. Gatsogiannis, C. *et al.* Membrane insertion of a Tc toxin in near-atomic detail. *Nat Struct Mol Biol* **23,** 884–890 (2016).
45. Bowen, D. *et al.* Insecticidal toxins from the bacterium Photorhabdus luminescens. *Science* **280,** 2129–2132 (1998).

46. Liu, D. *et al*. Insect resistance conferred by 283-kDa Photorhabdus luminescens protein TcdA in Arabidopsis thaliana. *Nat Biotechnol* **21,** 1222–1228 (2003).
47. ffrench-Constant, R. H., Dowling, A. & Waterfield, N. R. Insecticidal toxins from Photorhabdus bacteria and their potential use in agriculture. *Toxicon* **49,** 436–451 (2007).
48. Kaya, H. K. & Gaugle, R. Entomopathogenic nematodes. *Annu. Rev. Ento ol*. 181–206 (1993).
49. Bedding, R. A. & Molyneux, A. S. Penetration of Insect Cuticle by Infective Juveniles of Heterorhabditis-Spp (Heterorhabditidae, Nematoda). *Nematologica* **28,** 354–359 (1982).
50. Ciche, T. A., Kim, K.-S., Kaufmann-Daszczuk, B., Nguyen, K. C. Q. & Hall, D. H. Cell invasion and matricide during Photorhabdus luminescens transmission by Heterorhabditis bacteriophora nematodes. *Applied and Environmental Microbiology* **74,** 2275–2287 (2008).
51. Ffrench-Constant, R. *et al*. Photorhabdus: towards a functional genomic analysis of a symbiont and pathogen. *FEMS Microbiology Reviews* **26,** 433–456 (2003).
52. Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*: two roads to the same destination. *Molecular Microbiology* **64,** (2007).
53. Dunphy, G. B. & Webster, J. M. Virulence mechanisms of Heterorhabditis heliothidis and its bacterial associate, Xenorhabdus luminescens, in non-immune larvae of the greater wax moth, Galleria mellonella. *International Journal for Parasitology* **18,** 729–737 (1988).
54. Waterfield, N. R., Bowen, D. J., Fetherston, J. D., Perry, R. D. & ffrench-Constant, R. H. The tc genes of Photorhabdus: a growing family. *Trends Microbiol*. **9,** 185–191 (2001).
55. Ffrench-Constant, R. & Waterfield, N. An ABC Guide to the Bacterial Toxin Complexes. *Adv. Appl. Microbiol*. **58C,** 169–183 (2005).
56. Waterfield, N., Hares, M., Yang, G., Dowling, A. & ffrench-Constant, R. Potentiation and cellular phenotypes of the insecticidal Toxin complexes of Photorhabdus bacteria. *Cell Microbiol* **7,** 373–382 (2005).
57. Sergeant, M., Jarrett, P., Ousley, M. & Morgan, J. A. W. Interactions of insecticidal toxin gene products from Xenorhabdus nematophilus PMFI296. *Applied and Environmental Microbiology* **69,** 3349 (2003).
58. Hurst, M. R. H. *et al*. The main virulence determinant of Yersinia entomophaga MH96 is a broad-host-range toxin complex active against insects. *Journal of Bacteriology* **193,** 1966–1980 (2011).
59. Hurst, M. R., Glare, T. R., Jackson, T. A. & Ronson, C. W. Plasmid-located pathogenicity determinants of Serratia entomophila, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of Photorhabdus luminescens. *Journal of Bacteriology* **182,** 5127 (2000).
60. Martens, E. C. & Goodrich-Blair, H. The *Steinernema carpocapsae* intestinal vesicle contains a subcellular structure with which *Xenorhabdus nematophila* associates during colonization initiation. *Cell Microbiol* **7,** 1723 (2005).
61. Landsberg, M. J. *et al*. 3D structure of the Yersinia entomophaga toxin complex and implications for insecticidal activity. *Proceedings of the National Academy of Sciences* **108,** 20544–20549 (2011).
62. Vallet-Gely, I., Lemaitre, B. & Boccard, F. Bacterial strategies to overcome insect defences. *Nat Rev Micro* **6,** 302–313 (2008).
63. Morgan, J. A. W., Sergeant, M., Ellis, D., Ousley, M. & Jarrett, P. Sequence Analysis of Insecticidal Genes from Xenorhabdus nematophilus PMFI296. *Applied and Environmental Microbiology* **67,** 2062–2069 (2001).
64. Hurst, M. R. H., Jones, S. M., Tan, B. & Jackson, T. A. Induced expression of the Serratia entomophilaSep proteins shows activity towards the larvae of the New Zealand grass grub Costelytra zealandica. *FEMS Microbiol Lett* **275,** 160–167 (2007).

65. Hinchliffe, S. J. *et al.* Application of DNA microarrays to study the evolutionary genomics of Yersinia pestis and Yersinia pseudotuberculosis. *Genome Research* **13,** 2018–2029 (2003).
66. Hares, M. C. *et al.* The Yersinia pseudotuberculosis and Yersinia pestis toxin complex is active against cultured mammalian cells. *Microbiology (Reading, Engl.)* **154,** 3503–3517 (2008).
67. Chen, Y.-T. *et al.* Whole-genome sequencing and identification of Morganella morganii KT pathogenicity-related genes. *BMC Genomics* **13,** S4 (2012).
68. Naktin, J. & Beavis, K. G. Yersinia Enterocolitica and Yersinia Pseudotuherculosis. *Clinics in Laboratory Medicine* **19,** 523–536 (1999).
69. Chauvaux, S. *et al.* Transcriptome analysis of Yersinia pestis in human plasma: an approach for discovering bacterial genes involved in septicaemic plague. *Microbiology* **153,** 3112–3124 (2007).
70. Waterfield, N., Hares, M., Ffrench-Constant, R., Wren, B. W. & Hinchliffe, S. in *The Genus Yersinia* **603,** 247–257 (Springer, New York, NY, 2007).
71. Achtman, M. *et al.* Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proceedings of the National Academy of Sciences* **96,** 14043–14048 (1999).
72. Hinchliffe, S. J., Hares, M. C., Dowling, A. J. & ffrench-Constant, R. H. Insecticidal Toxins From the Photorhabdus and Xenorhabdus Bacteria. *The Open Toxinology Journal* **3,** 83–100 (2010).
73. Waterfield, N., Hares, M., Hinchliffe, S., Wren, B. & Ffrench-Constant, R. The insect toxin complex of Yersinia. *Adv. Exp. Med. Biol.* **603,** 247–257 (2007).
74. Sheets, J. J. *et al.* Insecticidal toxin complex proteins from Xenorhabdus nematophilus: Structure and pore formation. *Journal of Biological Chemistry* jbc.M111.227009 (2011). doi:10.1074/jbc.M111.227009
75. Lang, A. E., Konukiewitz, J., Aktories, K. & Benz, R. TcdA1 of Photorhabdus luminescens: Electrophysiological Analysis of Pore Formation and Effector Binding. *Biophysical Journal* **105,** 376–384 (2013).
76. Waterfield, N. *et al.* Oral toxicity of Photorhabdus luminescens W14 toxin complexes in Escherichia coli. *Applied and Environmental Microbiology* **67,** 5017–5024 (2001).
77. Lang, A. E. *et al.* Photorhabdus luminescens toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science* **327,** 1139–1142 (2010).
78. Meusch, D. *Struktur und Funktion von TcdA1 und TcdB2-TccC3 eines Tc-Toxinkomplexes aus Photorhabdus luminescens. PhD Thesis, Technische Universität Dortmund, Dortmund.* 1–158 (2014).
79. Buetow, L., Flatau, G., Chiu, K., Boquet, P. & Ghosh, P. Structure of the Rho-activating domain of Escherichia coli cytotoxic necrotizing factor 1. *Nat Struct Mol Biol* **8,** 584–588 (2001).
80. Knust, Z. & Schmidt, G. Cytotoxic Necrotizing Factors (CNFs)−A Growing Toxin Family. *Toxins* **2,** 116–127 (2010).
81. Erickson, D. L. *et al.* Acute oral toxicity of Yersinia pseudotuberculosis to fleas: implications for the evolution of vector-borne transmission of plague. *Cell Microbiol* **9,** 2658–2666 (2007).
82. Pinheiro, V. B. & Ellar, D. J. Expression and insecticidal activity of Yersinia pseudotuberculosis and Photorhabdus luminescens toxin complex proteins. *Cell Microbiol* **9,** 2372–2380 (2007).
83. Lee, S. C. *et al.* Structural Characterisation of the Insecticidal Toxin XptA1, Reveals a 1.15 MDa Tetramer with a Cage-like Structure. *Journal of Molecular Biology* **366,** 1558–1568 (2007).
84. Gatsogiannis, C. *et al.* Tc toxin activation requires unfolding and refolding of a β-propeller. *Nature* **9,** 185 (2018).
85. Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. H. & Lott, J. S. The BC

component of ABC toxins is an RHS-repeat-containing protein encapsulation device. *Nature* **501,** 547–550 (2013).

86. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. Crystal structure of the anthrax toxin protective antigen. *Nature* **385,** 833–838 (1997).
87. Rafael Fernandez-Leiro, S. H. W. S. Unravelling the structures of biological macromolecules by cryo-EM. *Nature* **537,** 339–346 (2016).
88. Kendrew, J. C., Bodo, G., Dintzis, H. M. & (null), R. P. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature.* **191,** 662–666 (1958).
89. Shi, D., Nannenga, B. L., Iadanza, M. G., Elife, T. G.2013. Three-dimensional electron crystallography of protein microcrystals. *eLIFE* (2013).
90. Rodriguez, J. A. & Gonen, T. High-Resolution Macromolecular Structure Determination by MicroED, a cryo-EM Method. *Methods in Enzymology* **579,** 369–392 (2016).
91. Wong, K. C. Review of NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry. *J. Chem. Educ.* **91,** 1103–1104 (2014).
92. Knoll, M. & Ruska, E. Beitrag zur geometrischen Elektronenoptik. I. *Annalen der Physik* **404,** 607–640 (1932).
93. Frank, J. Single-Particle Imaging of Macromolecules by Cryo-Electron Microscopy. *Annu. Rev. Biophys. Biomol. Struct.* **31,** 303 (2002).
94. Efremov, R. G., Gatsogiannis, C. & Raunser, S. in *A Structure-Function Toolbox for Membrane Transporter and Channels* **594,** 1–30 (Elsevier, 2017).
95. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification — powerful tools in modern electron microscopy. *Biol. Proced. Online* **6,** 23–34 (2004).
96. Dubochet, J. *et al.* Cryo-electron microscopy of vitrified specimens. *Quart. Rev. Biophys.* **21,** 129–228 (1988).
97. Pfeffer, S. & Mahamid, J. Unravelling molecular complexity in structural cell biology. *Current Opinion in Structural Biology* **52,** 111–118 (2018).
98. Taylor, K. A. & Glaeser, R. M. Electron Diffraction of Frozen, Hydrated Protein Crystals. *Science* **186,** 1036–1037 (1974).
99. Gringer. Scheme TEM, Webseite, Wikipedia Deutschland, https://commons.wikimedia.org/wiki/File:Scheme_TEM_en.svg.
100. Penczek, P. A. Image Restoration in Cryo-Electron Microscopy. *Methods in Enzymology* **482,** 35–72 (2010).
101. Grigorieff, N. & Harrison, S. C. Near-atomic resolution reconstructions of icosahedral viruses from electron cryo-microscopy. *Current Opinion in Structural Biology* **21,** 265–273 (2011).
102. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quart. Rev. Biophys.* **28,** 171–193 (1995).
103. Jiang, J., Pentelute, B. L., Collier, R. J. & Zhou, Z. H. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521,** 545–549 (2015).
104. Vinayagam, D. *et al.* Electron cryo-microscopy structure of the canonical TRPC4 ion channel. *eLife* **7,** 213 (2018).
105. Bartesaghi, A. *et al.* 2.2 A resolution cryo-EM structure of -galactosidase in complex with a cell-permeant inhibitor. *Science* **348,** 1147–1151 (2015).
106. Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat Commun* **8,** 16099 (2017).
107. Campbell, M. G. *et al.* Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20,** 1823–1828 (2012).
108. Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3,** 1–8 (2014).
109. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4,** e06980 (2015).

110. Ripstein, Z. A. & Rubinstein, J. L. Processing of Cryo-EM Movie Data. *Methods in Enzymology* **579,** 103–124 (2016).
111. Kühlbrandt, W. The Resolution Revolution. *Science* **343,** 1443–1444 (2014).
112. Moriya, T. *et al.* High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *J Vis Exp* (2017). doi:10.3791/55448
113. Scheres, S. H. W. A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology* **415,** 406–418 (2012).
114. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161,** 438–449 (2015).
115. Crowther, R. A., DeRosier, D. J. & Klug, A. The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **317,** 319–340 (1970).
116. Wagner, T. *et al.* SPHIRE-crYOLO: A fast and well-centering automated particle picker for cryo-EM. *bioRxiv* 356584 (2018). doi:10.1101/356584
117. Frank, J. Three-Dimensional Electron Microscopy of Macromolecular Assemblies. *Oxford Uniersity Press* (2006).
118. Rosenthal, P. B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology* **333,** 721–745 (2003).
119. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9,** 853 (2012).
120. Bergmans, H. E., Van Die, I. M. & Hoekstra, W. P. Transformation in Escherichia coli: stages in the process. *Journal of Bacteriology* **146,** 564–570 (1981).
121. Laurent, T. C. & Killander, J. A theory of gel filtration and its exeperimental verification. *Journal of Chromatography A* **14,** 317–330 (1964).
122. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227,** 680–685 (1970).
123. Edwards, R. A., Jickling, G. & Turner, R. J. The light-induced reactions of tryptophan with halocompounds. *Photochem. Photobiol.* **75,** 362–368 (2002).
124. Commoner, B. & Lipkin, D. The Application of the Beer-Lambert Law to Optically Anisotropic Systems. *Science* **110,** 41–43 (1949).
125. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* 571–607 (Humana Press, 2005). doi:10.1385/1-59259-890-0:571
126. Zakharian, E. in *Lipid-Protein Interactions* **998,** 109–118 (Humana Press, 2013).
127. Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins. *J. Am. Chem. Soc.* **135,** 1919–1925 (2013).
128. Grinkova, Y. V., Denisov, I. G. & Sligar, S. G. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng Des Sel* **23,** 843–848 (2010).
129. Dubochet, J., Groom, M., Microsc, S. M.-N. A. O. E.1982. *The mounting of macromolecules for electron microscopy with particular reference to surface phenomena and the treatment of support films by glow*
130. Booth, D. S., Avila-Sakar, A. & Cheng, Y. Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. *JoVE* e3227–e3227 (2011).
131. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *Journal of Structural Biology* **157,** 38–46 (2007).
132. Yang, Z., Fang, J., Chittuluru, J., Asturias, F. J. & Penczek, P. A. Iterative stable alignment and clustering of 2D transmission electron microscope images. *Structure* **20,** 237–247 (2012).
133. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *Journal of*

Structural Biology **157,** 47–55 (2007).
134. Wu, S., Armache, J.-P. & Cheng, Y. Single-particle cryo-EM data acquisition by using direct electron detection camera. *Microscopy (Oxf)* dfv355 (2015). doi:10.1093/jmicro/dfv355
135. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* **177,** 630–637 (2012).
136. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10,** 584–590 (2013).
137. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Meth* **14,** 331–332 (2017).
138. Thon, F. Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift für Naturforschung A* **21a,** 476–478 (1966).
139. Penczek, P. A. *et al.* CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140,** 9–19 (2014).
140. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).
141. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* **234,** 779–815 (1993).
142. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539–539 (2011).
143. Lopéz-Blanco, J. R. & Chacón, P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *Journal of Structural Biology* **184,** 261–270 (2013).
144. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 486–501 (2010).
145. Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55,** 94–106 (2011).
146. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21 (2010).
147. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39,** W13–7 (2011).
148. Jamroz, M. *et al.* KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.* **43,** D306–14 (2015).
149. Masood, T. B., Sandhya, S., Chandra, N. & Natarajan, V. CHEXVIS: a tool for molecular channel extraction and visualization. *BMC Bioinformatics* **16,** 119 (2015).
150. Berka, K. *et al.* MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.* **40,** W222–7 (2012).
151. Sehnal, D. *et al.* MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminform* **5,** 39 (2013).
152. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157,** 105–132 (1982).
153. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).
154. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40,** W537–41 (2012).
155. Bashford, D. & Karplus, M. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29,** 10219–10225 (1990).
156. Leidreiter, F. Strukturelle Untersuchung des Tc-Toxin-Komplexes von Xenorhabdus nematophila. *Master Thesis, Technische Universität Dortmund, Dortmund.* (2015).
157. Sciences, G. H. L. GE Healthcare Life Sciences.

httpscdn.gelifesciences.comdmmbwsvAssetStream.aspxmediaformatiddestinationidassetid **18-1163-80 AC,** (2011).
158. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **12,** 85–94 (1999).
159. MAZIA, D., SCHATTEN, G. & SALE, W. Adhesion of Cells to Surfaces Coated with Polylysine - Applications to Electron-Microscopy. *J. Cell Biol.* **66,** 198–200 (1975).
160. Rath, B. Strukturbiochemische Analyse des Tc Toxins aus Yersinia pseudotuberculosis. *Master Thesis, Technische Universität Dortmund, Dortmund.* (2017).
161. Russo, C. J. & Passmore, L. A. Progress towards an optimal specimen support for electron cryomicroscopy. *Current Opinion in Structural Biology* **37,** 81–89 (2016).
162. Glaeser, R. M. *et al.* Factors that Influence the Formation and Stability of Thin, Cryo-EM Specimens. *Biophysical Journal* **110,** 749–755 (2016).
163. Noble, A. J. *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *eLife* **7,** 32 (2018).
164. Lupas, A. Prediction and analysis of coiled-coil structures. *Methods in Enzymology* **266,** 513–525 (1996).
165. Walshaw, J. & Woolfson, D. N. SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. *Journal of Molecular Biology* **307,** 1427–1450 (2001).
166. Scholey, J. E., Nithianantham, S., Scholey, J. M. & Al-Bassam, J. Structural basis for the assembly of the mitotic motor Kinesin-5 into bipolar tetramers. *eLife* **3,** e02217 (2014).
167. Kilmartin, J. V., Dyos, S. L., Kershaw, D. & Finch, J. T. A spacer protein in the Saccharomyces cerevisiae spindle poly body whose transcript is cell cycle-regulated. *J. Cell Biol.* **123,** 1175–1184 (1993).
168. Bañares-Hidalgo, A., Pérez-Gil, J. & Estrada, P. Acidic pH triggers conformational changes at the NH2-terminal propeptide of the precursor of pulmonary surfactant protein B to form a coiled coil structure. *Biochim. Biophys. Acta* **1838,** 1738–1751 (2014).
169. Ahn, V. E., Faull, K. F., Whitelegge, J. P., Fluharty, A. L. & Privé, G. G. Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proceedings of the National Academy of Sciences* **100,** 38–43 (2003).
170. Hopfner, K.-P. *et al.* The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* **418,** 562–566 (2002).
171. Weiner, J., rd, Beaussart, F. & Bornberg-Bauer, E. Domain deletions and substitutions in the modular protein evolution. *FEBS Journal* **273,** 2037–2047 (2006).
172. Lupas, A. N., Ponting, C. P. & Russell, R. B. On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World? *Journal of Structural Biology* **134,** 191–203 (2001).
173. Dabrowski-Tumanski, P., Stasiak, A. & Sulkowska, J. I. In Search of Functional Advantages of Knots in Proteins. *PLoS ONE* **11,** e0165986 (2016).
174. Dabrowski-Tumanski, P. *et al.* KnotProt 2.0: a database of proteins with knots and other entangled structures. *Nucleic Acids Res.* **47,** D367–D375 (2018).
175. Virnau, P., Mirny, L. A. & Kardar, M. Intricate knots in proteins: Function and evolution. *PLoS Comput. Biol.* **2,** e122 (2006).
176. Mallam, A. L. & Jackson, S. E. Knot formation in newly translated proteins is spontaneous and accelerated by chaperonins. *Nat. Chem. Biol.* **8,** 147–153 (2011).
177. Sułkowska, J. I., Rawdon, E. J., Millett, K. C., Onuchic, J. N. & Stasiak, A. Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **109,** E1715–23 (2012).
178. Faísca, P. F. N. Knotted proteins: A tangled tale of Structural Biology. *Comput Struct Biotechnol J* **13,** 459–468 (2015).
179. Wang, X., Rickert, M. & Garcia, K. C. Structure of the Quaternary Complex of Interleukin-2 with Its α, ß, and γc Receptors. *Science* **310,** 1159–1163 (2005).
180. Thomas, C. *et al.* Structural Linkage between Ligand Discrimination and Receptor

Activation by Type I Interferons. *Cell* **146,** 621–632 (2011).
181. Krissinel, E. On the relationship between sequence and structure similarities in proteomics. *Bioinformatics* **23,** 717–723 (2007).
182. Brejc, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411,** 269–276 (2001).
183. Celie, P. H. N. *et al.* Crystal structure of acetylcholine-binding protein from Bulinus truncatus reveals the conserved structural scaffold and sites of variation in nicotinic acetylcholine receptors. *Journal of Biological Chemistry* **280,** 26457–26466 (2005).
184. Varghese, J. N., Laver, W. G. & Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature* **303,** 35–40 (1983).
185. Varki, A. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature* **446,** 1023–1029 (2007).
186. Holmgren, J., Lönnroth, I., Månsson, J. & Svennerholm, L. Interaction of cholera toxin and membrane GM1 ganglioside of small intestine. *Proceedings of the National Academy of Sciences* **72,** 2520–2524 (1975).
187. Kimble, B., Nieto, G. R. & Perez, D. R. Characterization of influenza virus sialic acid receptors in minor poultry species. *Virology Journal 2010 7:1* **7,** 365 (2010).
188. Ramon, E. *et al.* Critical role of electrostatic interactions of amino acids at the cytoplasmic region of helices 3 and 6 in rhodopsin conformational properties and activation. *Journal of Biological Chemistry* **282,** 14272–14282 (2007).
189. Dow, J. A. Extremely high pH in biological systems: a model for carbonate transport. *Am. J. Physiol.* **246,** R633–6 (1984).
190. Mari, S. A. *et al.* pH-induced conformational change of the beta-barrel-forming protein OmpG reconstituted into native E. coli lipids. *Journal of Molecular Biology* **396,** 610–616 (2010).
191. Yeo, K. J. *et al.* Mechanism of the pH-induced conformational change in the sensor domain of the DraK Histidine kinase via the E83, E105, and E107 residues. *PLoS ONE* **9,** e107168 (2014).
192. Kalani, M. R., Moradi, A., Moradi, M. & Tajkhorshid, E. Characterizing a Histidine Switch Controlling pH-Dependent Conformational Changes of the Influenza Virus Hemagglutinin. *Biophysical Journal* **105,** 993–1003 (2013).
193. Harrison, J. S. *et al.* Role of Electrostatic Repulsion in Controlling pH-Dependent Conformational Changes of Viral Fusion Proteins. *Structure* **21,** 1085–1096 (2013).
194. Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534,** 347–351 (2016).
195. Becher, S., Esch, P. & Heiles, S. Relative Quantification of Phosphatidylcholine sn-Isomers Using Positive Doubly Charged Lipid-Metal Ion Complexes. *Anal. Chem.* **90,** 11486–11494 (2018).
196. Altenbach, C. & Biochemistry, J. S. Ca^{2+} binding to phosphatidylcholine bilayers as studied by deuterium magnetic resonance. Evidence for the formation of a calcium complex with two phospholipid molecules. *ACS Publications* **23,** 3913–3920 (1984).
197. Yagi, H. *et al.* Structural Basis for Ca2+ -mediated Interaction of the Perforin C2 Domain with Lipid Membranes. *J. Biol. Chem.* **290,** 25213–25226 (2015).
198. Eifler, N. *et al.* Cytotoxin ClyA from Escherichia coli assembles to a 13-meric pore independent of its redox-state. *EMBO J.* **25,** 2652–2661 (2006).
199. Olsnes, S., Øivind Moskaug, J., Stenmark, H. & Sandvig, K. Diphtheria toxin entry: protein translocation in the reverse direction. *Trends in Biochemical Sciences* **13,** 348–351 (1988).
200. Murphy, J. R. Mechanism of Diphtheria Toxin Catalytic Domain Delivery to the Eukaryotic Cell Cytosol and the Cellular Factors that Directly Participate in the Process. *Toxins* **3,** 294–308 (2011).
201. Lang, A. E. *et al.* The chaperone Hsp90 and PPIases of the cyclophilin and FKBP families facilitate membrane translocation of Photorhabdus luminescens ADP-

ribosyltransferases. *Cell Microbiol* **16,** 490–503 (2014).
202. Kaiser, E. *et al.* Membrane Translocation of Binary Actin-ADP-Ribosylating Toxins from Clostridium difficile and Clostridium perfringens Is Facilitated by Cyclophilin A and Hsp90. *Infect. Immun.* **79,** 3913–3921 (2011).
203. Ost, G. S., Ng'ang'a, P. N., Lang, A. E. & Aktories, K. Photorhabdus luminescens Tc toxin is inhibited by the protease inhibitor MG132 and activated by protease cleavage resulting in increased binding to target cells. *Cell Microbiol* **21,** e12978 (2019).
204. Ernst, K. *et al.* A novel Hsp70 inhibitor prevents cell intoxication with the actin ADP-ribosylating Clostridium perfringens iota toxin. *Scientific Reports 2016 6* **6,** 20301 (2016).
205. Dowd, K. J. & Tweten, R. K. The Cholesterol-Dependent Cytolysin Signature Motif: A Critical Element in the Allosteric Pathway that Couples Membrane Binding to Pore Assembly. *PLoS Pathog* **8,** (2012).
206. Johnson, S., Brooks, N. J., Smith, R. A. G., Lea, S. M. & Bubeck, D. Structural basis for recognition of the pore-forming toxin intermedilysin by human complement receptor CD59. *Cell Rep* **3,** 1369–1377 (2013).
207. Maxfield, F. R. & Yamashiro, D. J. in *Immunobiology of Proteins and Peptides IV* **225,** 189–198 (Springer, Boston, MA, 1987).
208. Conley, M. J. *et al.* Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature* **565,** 377–381 (2019).
209. LaChapelle, S., Tweten, R. K. & Hotze, E. M. Intermedilysin-Receptor Interactions during Assembly of the Pore Complex. *Journal of Biological Chemistry* **284,** 12719–12726 (2009).

7 Appendix

7.1 Processing of TcA data sets and structure validation

Table 6: Final statistics for the structure validation of the evaluated proteins.

	Xn-XptA1	Mm-TcdA4	Yp-TcaATcaB
Atomic model composition			
Non-hydrogen atoms	92,340	88,470	75,405
Protein residues	11,665	11,420	9365
Completeness (%)	92.47	92.55	92.13
R.M.S deviations			
Bond lengths (Å)	0.01	0.01	0.01
Angles (°)	1.11	1.23	0.94
Model validation			
Clashscore, all atoms	5.63	7.58	4.23
Poor rotamers (%)	0.10	0.58	0.27
Favored rotamers (%)	97.41	96.04	97.52
Ramachandran outliers (%)	0.43	0.26	0.05
Ramachandran favored (%)	92.97	92.19	89.87
Molprobity score	1.76	1.89	1.72

7.2 Purification of Yp-TcaATcaB-Δ622-714

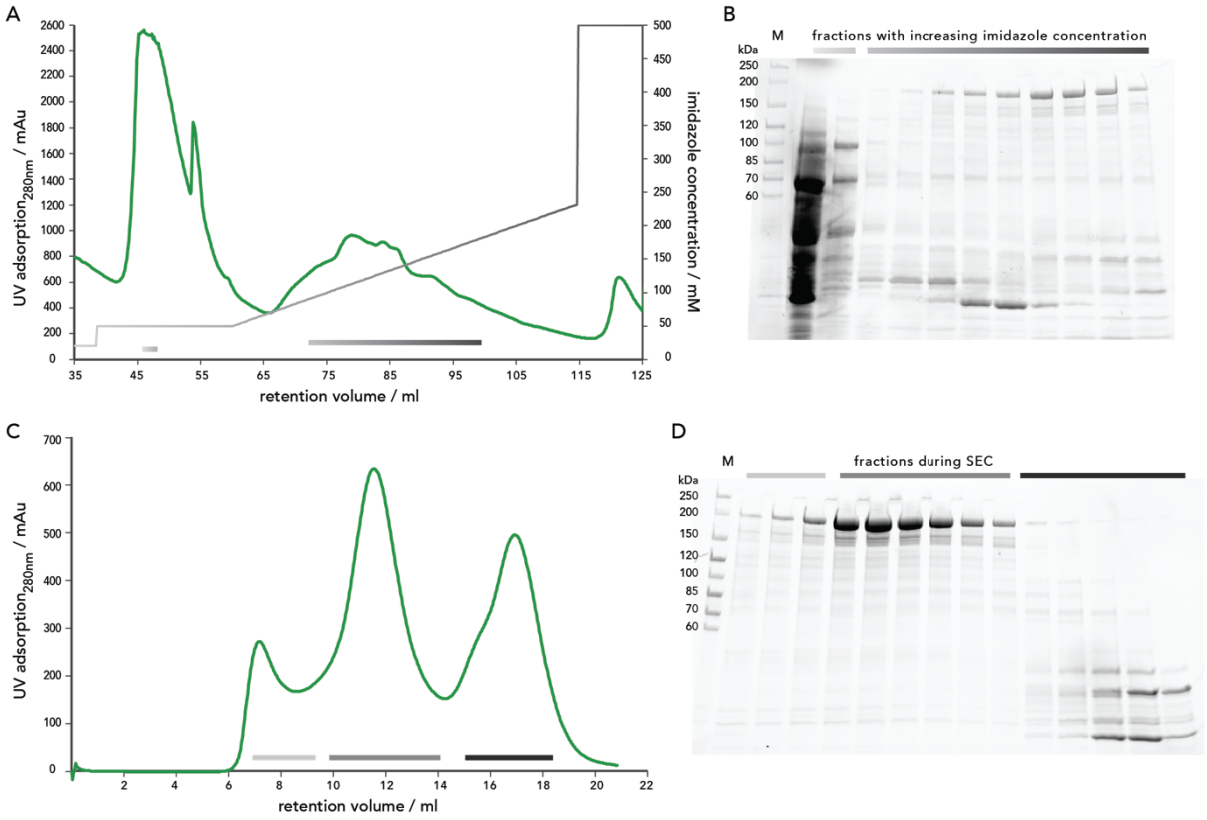


Figure 54: Chromatograms and SDS-PAGE gels after the two purification steps of Yp-TcaATcaB-Δ622-714 with deleted coiled coil domain. A) Chromatogram of the Ni-NTA purification step with increasing imidazole concentration depicted as a gradient from light grey to black. The fractions that were analyzed via SDS-PAGE are marked with a line having the same color gradient as the imidazole concentration. B) Corresponding gel to the Ni-NTA chromatography. M=marker and the fractions are marked with a line indicating the imidazole concentration during elution. C) Chromatogram of the size exclusion chromatography with a Superose 6 10/300. D) SDS-PAGE gel of the size exclusion chromatography. M=marker, fractions during size exclusion with increasing retention volume visualized with the gradient from light grey to black as in C).

7.3 Matrix scheme of the trefoil 3_1 protein

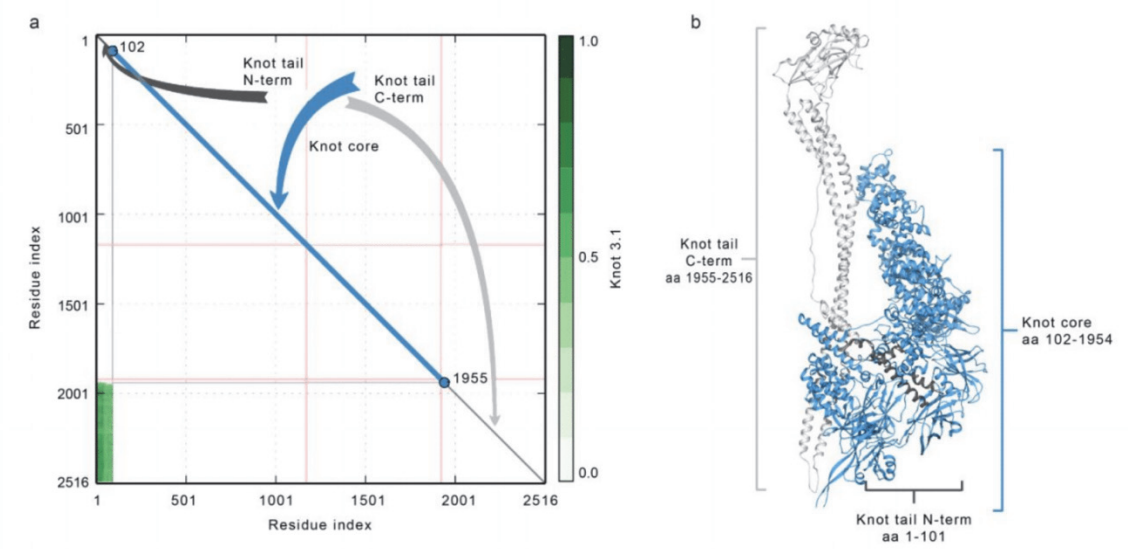
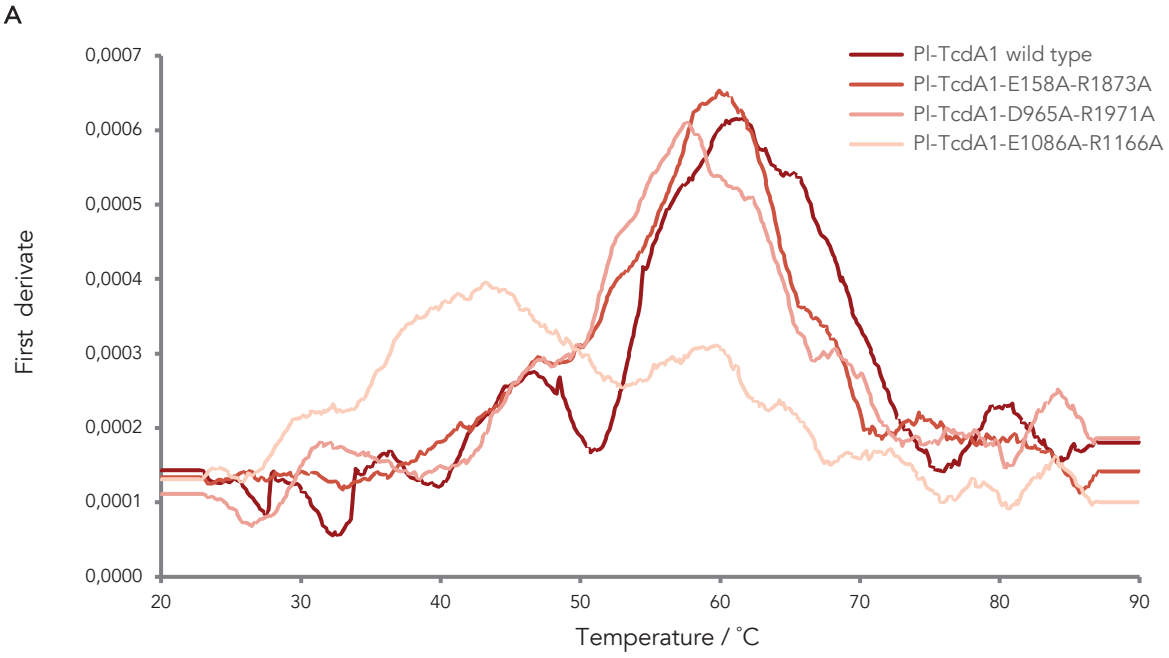


Figure 55: Organization of the trefoil 3_1 protein knot for Pl-TcdA1. (a): Matrix scheme of the trefoil 3_1 protein knot with the three domains of the knot organization: knot tail N-term (dark grey), the knot core (blue) and the knot tail C-term (grey). Residues belonging to the knotted chain are colored green. (b): Pl-TcdA1 colored accordingly to the three different knot domains.

7.4 Nano differential scanning fluorimetry measurements



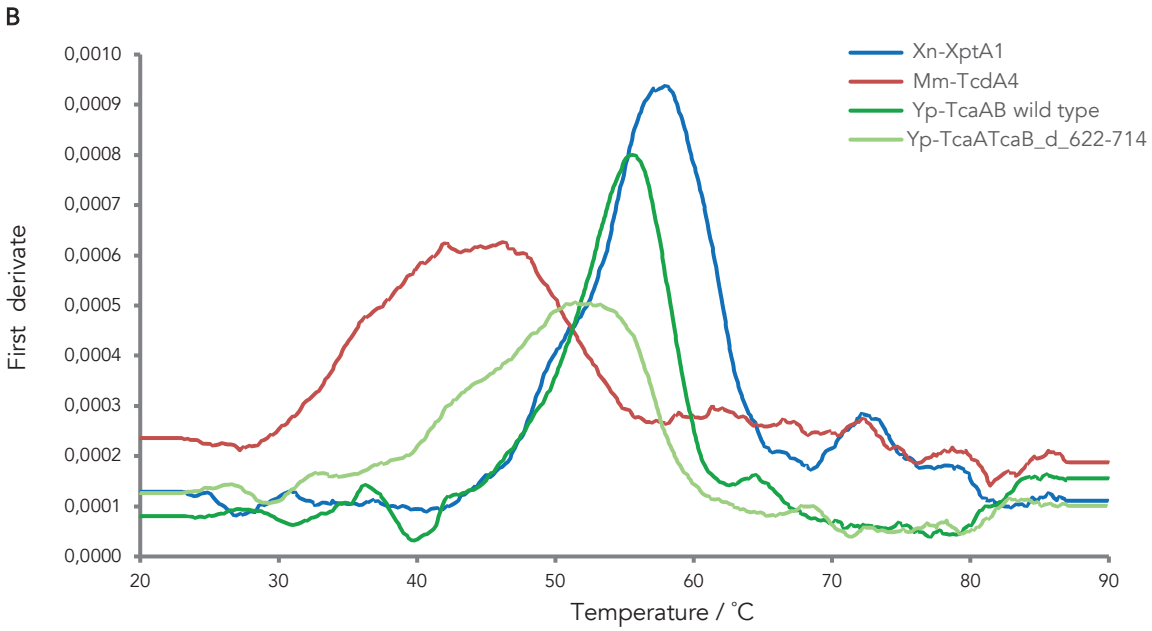


Figure 56: Nano differential scanning fluorimetry measurements of all analyzed TcA toxins at pH 8. The graphs show the first derivate of the absorbance quotient (330 nm/350 nm) as function of temperature for A) Pl-TcdA1 (wilde type) and the mutants and B) the other TcAs including Xn-XptA1 (blue), Mm-TcdA4 (red), Yp-TcaATcaB wilde type (dark green) and Yp-TcaATcaB mutant (light green). The measurements were performed at pH 8.

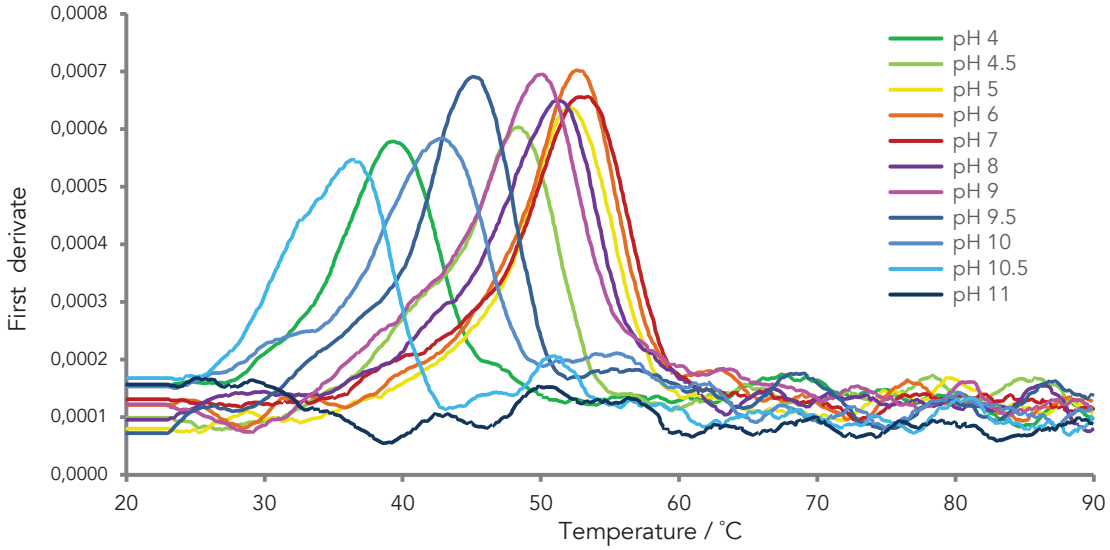


Figure 57: Nano differential scanning fluorimetry measurements with Yp-TcaATcaB at different pH values. The protein stability was determined at different pH values ranging from pH 4 to pH 11. The graphs show the first derivate of the absorbance quotient (330 nm/350 nm) as function of temperature. The different pH values are indicated by different colors.

Table 7: Experimental determined thermal stability of Yp-TcaATcaB over a pH range from pH 4 to 11.

	pH 4	pH 4.5	pH 5	pH 6	pH 7	pH 8
T_m/°C	39.5 ± 0.1	48.0 ± 0.5	52.1 ± 0.2	53.1 ± 0.7	53.1 ± 0.1	50.9 ± 0.6

	pH 9	pH 9.5	pH 10	pH 10.5	pH 11
T_m/°C	47.2 ± 3.7	44.5 ± 0.7	42.3 ± 1.0	35.0 ± 2.2	-

7.5 Single channel conductivity in presence of $CaCl_2$

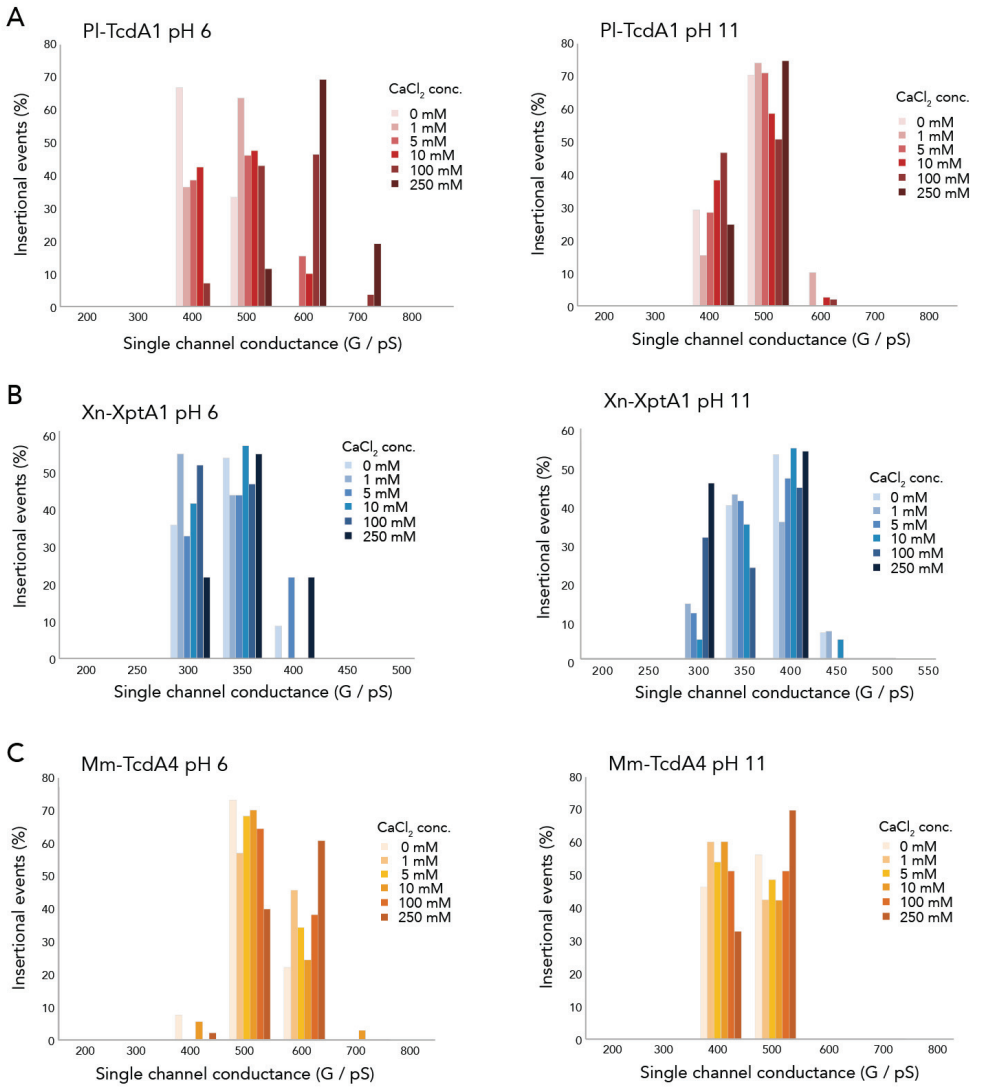


Figure 58: Single channel conductivity of Pl-TcdA1, Xn-XptA1 and Mm-TcdA4 in presence of $CaCl_2$. All histograms were constructed from the data of at least 50 pore insertional events for each TcA, pH value and $CaCl_2$ concentration in 1 M KCl. A) Measurements of Pl-TcdA1 at pH 6 and 11, respectively. B) The results of the experiments with Xn-XptA1 at pH 6 and 11, respectively. C) Single channel conductivity for Mm-TcdA4 in presence of a $CaCl_2$ gradient at pH 6 and 11.

7.6 Biolayer interferometry measurements

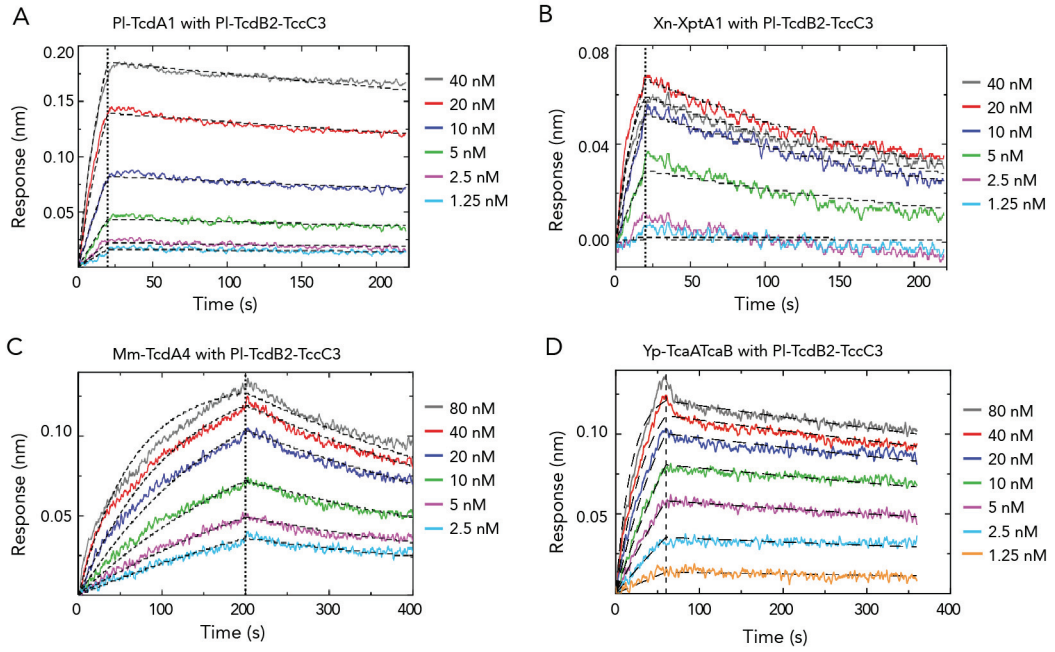


Figure 59: BLI measurements of the Pl-TcdB2-TccC3 and the different TcAs. A-D) BLI sensograms of Pl-TcdA1 (A), Xn-XptA1 (B) (both 1.25 – 40 nM pentamer, respectively), Mm-TcdA4 (C) (2.5 – 80 nM pentamer) and Yp-TcaATcaB (D) (1.25 – 80 nM pentamer) interacting with immobilized *P. luminescens* TcdB2-TccC3. A global fit according to a 1:1 binding model was applied (black dashed curves). Association and dissociation phases are separated by a black dashed line.

8 Publications & conference contributions

Publications:

Gatsogiannis, C., Merino, F., Prumbaum, D., Roderer, D., Leidreiter, F., Meusch, D., Raunser, S. (2016). Membrane insertion of a Tc toxin in near-atomic detail. Nat Struct Mol Biol 23, 884–890 (2016).

Leidreiter, F., Roderer, D., Meusch, D., Gatsogiannis, C., Benz, R., Raunser, S. (2019). Conserved architecture of Tc toxins from human and insect pathogenic bacteria. Submitted to Science Advances.

Conference Contributions:

Leidreiter, F., Roderer, D., Gatsogiannis, C., Raunser, S. (2016). Comparison of three high-resolution structures of Tc toxins from human and insect pathogenic bacteria.
Poster presentation at the 6th Murnau Conference on Structural Biology, Murnau, Germany

Leidreiter, F., Roderer, D., Raunser, S. (2017). Comparison of three high-resolution structures of Tc toxins.
Talk at the 52nd Winter Seminar in Biophysical Chemistry, Klosters, Switzerland

Leidreiter, F., Roderer, D., Gatsogiannis, C., Raunser, S. (2018). Conserved architecture of Tc toxins from human and insect pathogenic bacteria.
Poster presentation at the 7th IMPRS-CMB student symposium, Dortmund, Germany

Eidesstattliche Versicherung (Affidavit)

Leidreiter, Franziska
Name, Vorname
(Surname, first name)

135671
Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Structural insights into Tc toxins from human and insect pathogenic bacteria

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

Ort, Datum
(Place, date)

Unterschrift
(Signature)