# A double-edged sword

## AB toxins as attack (cholera toxin) and defense (CCTX2) mechanisms

Flore Kersten

Thesis for the degree of *Philosophiae Doctor*

Norwegian Centre for Molecular Medicine / Department of Chemistry

Faculty of Mathematics and Natural Sciences

University of Oslo

Autumn 2019 - Autumn 2023

*Series of dissertations submitted to the*
*Faculty of Mathematics and Natural Sciences, University of Oslo*
*No. 2709*

ISSN 1501-7710



Cover: UiO.
Print production: Graphic center, University of Oslo.

# Acknowledgements

The work presented in this thesis was carried out between September 2021 and October 2023. It was performed in the laboratory of Prof. Ute Krengel, at the Department of Chemistry, University of Oslo with financial support from NCMM, Centre for Molecular Medicine Norway.

First, I would like to thank deeply my main supervisor Prof. Ute Krengel. Even before I was enrolled in your group, you have always been very supportive and helped me throughout my early stage researcher career. In 2021, I was relieved and very happy to finally join your group and I haven't regretted it ever since. Thank you for your supervision, your pushes and motivational words. I would not have made it this far without you.

Of course, I would like to acknowledge our collaboration partners for their help and support, Prof. Markus Künzler, Dr. Stefanie Schmieder for the CCTX2 project and Prof. Ken Teter and Albert Serrano for CT project. During my PhD, I was also lucky enough to spend three months in Aarhus, in Denmark at the cryo-EM facility. I met two amazing and brilliant scientists, Dr. Thomas Boesen and Dr. Andreas Boeggild from whom I learned a lot about cryo-EM and particularly data processing. Thank a lot for your time and the endless computer sessions.

I would also like to acknowledge my co-supervisor Dr. Gabriele Cordara for his involvement and support in all my scientific projects. Thank you so much for proofreading my thesis, you are a big part of this achievement. You are a mine of information that seems to never dry up and always found a way to help me! I will repay you in baked-goods from Garçon.

A special mention goes to Clara for all her hard work on the CCTX2, you were tremendous and I was very lucky to inherit the project from you. Mateu, you have been my partner in crime since the very beginning. Thanks you for your mental support and help. I would also like to express my gratitude to all the group members of the Luecke group and the Krengel group who have contributed to make this an enjoyable ride: Joel for your supervision and positivism, Javi and Marta for the many corridor discussion, Rasma for your generosity, Tamjidmaa, Lin, Ayla, Eirik, Natalia, Ayla, Mari for making the office at Chemistry a great place to be.

I would like to thank all my friends and family for the support in the eventful journey. Camille, merci beaucoup pour tous tes gentils mots, tes conseils et nos soirées tricots. Maman, Papa et Léa, je vous remercie pour tout votre soutien, malgré la distance, vous avez toujours été présents pour moi. Merci d'avoir encouragé ma curiosité, mon amour pour les langues et mon avidité pour les voyages. Je ne serais jamais arrivée jusqu'ici sans vous.

Finally, Aksel, you deserve your own paragraph. Thanks for the feedback on the thesis, the much needed breaks along the writing process and the help for translating the summary to Norwegian but foremost thank you me helping me navigate through this journey. Your love and support goes beyond everything I could have wished for. It was a bumpy road and you have always been by my side. I am very much looking forward to many future projects with you!

# Table of Contents

## Thesis summary

AB toxins are a large group of proteins mainly present in plants and bacteria. Many pathogens use AB toxins as a virulence factor to promote invasion of the host. However, their use is not limited to that of offensive tools, as several organisms use them as a defense mechanism when threatened by a predator. AB toxins are constituted of at least one binding subunit (B subunit), allowing cell entry, and an active subunit (A subunit) exerting the toxic activity.

This PhD thesis aims to understand the molecular mechanism of two AB toxins, representing the two distinct roles of attack and defense. Our first target is cholera toxin (CT), an attack protein responsible for cholera disease and crucial for the exploitation of the host. CT is a well-known $AB_5$ toxin, with a pentamer of B subunits and an adenosine diphosphate ribosyltransferase (ART) active subunit. Our second target is *Coprinopsis cinerae* toxin 2 (CCTX2), a representative of fungal AB toxins. CCTX2 is a defense protein, expressed in the vegetative mycelium upon challenge by fungivorous nematodes. It possesses four ricin B chain-like domains, potentially lectins domains capable of binding glycan and allowing cell entry and a fifth domain of totally unknown structure and role.

Production of high amounts of protein is essential in structural biology. In **manuscript I**, we describe a method for high yield protein production using Vmax™ cells (*Vibrio natriegens*). Vmax™ cells were used for production of CT and CCTX2, providing sufficient amount of protein to perform structural and functional studies further described in this work. In **manuscript II**, we aim at explaining the difference of toxicity between CT and heat-labile enterotoxin (LT), its *Escherichia coli* orthologue. We generated two chimeric CT proteins, where the C-terminus of the A subunit contains one to four crucial amino acids substitution matching the C-terminus of the LT A subunit. These variants were compared to the wild-type proteins, functionally and structurally. Both chimeric proteins show lower toxicity than wild-type CT, and their ability to be disassembled by protein disulfide isomerase resembles that of LT. We solved the X-ray crystallographic structure of both chimeras and revealed that the C-terminus region of the four-substitution protein harbors the fold of LT, which seems to be crucial in CT processing by the host. Additionally, we investigated how substitutions affect the flexibility of the different chimeras, which likely affects interaction with PDI and efficient disassembly. In **manuscript III**, we explore the internalization pathway and possible fate of defense protein CCTX2 within the cell. We unraveled the carbohydrate recognized by the lectins domains of CCTX2 and its internalization mechanism. We then use single-particle cryogenic electron microscopy (cryo-EM) to solve its structure. The structural data shows a completely new fold for the toxic domain, not represented among those deposited in the Protein Data Bank (PDB). Structural data alone did not provide further clues about the activity of the toxic domain, but we revealed the potential role of CCTX2 as an ART. We followed up the results of

**manuscript III** by investigating possible cofactors and ligands by differential scanning fluorimetry (DSF). Screening results hint at a potential interaction of CCTX2 with nucleotides, more specifically nicotinamide adenine dinucleotide (NAD) and adenosine triphosphate (ATP). It is also strongly destabilized by divalent metallic cations with the exception of $Mg^{2+}$, often involved in the binding of nucleotide-binding enzymes, which is in concordance with the preliminary results obtained in **manuscript III**.

# Norsk sammendrag

AB-toksiner er en stor gruppe av proteiner som hovedsakelig forekommer i planter og bakterier. Mange patogener bruker AB-toksiner som virulensfaktor for å bedre angripe verten. De er imidlertid ikke utelukkende angrepsverktøy, ettersom ulike organismer bruker dem som en forsvarsmekanisme når de blir angrepet. AB-toksiner består av minst en bindene del (delenhet B), som gjør dem i stand til å trenge inn i en celle, og en aktiv del (delenhet A) som utøver den toksiske funksjonen.

Denne PhD-oppgaven sikter på å forstå de molekylære mekanismene til to ulike AB-toksiner, som eksemplifiserer både den offensive og den defensive bruken av AB-toksiner. Det første subjektet vårt er koleratoksin (CT), et angrepsprotein som forårsaker kolera og er kritisk for utnyttelse av verten. CT er et velkjent ABs-toksin, med en pentamer av bindene delenheter og en adenosine difosfat ribosyltranferase (ART) aktiv delenhet. Det andre subjektet vårt er *Coprinopsis cinerae* toksin 2 (CCTX2), et eksempel på en fungal AB-toksin. CCTX2 er et forsvarsprotein som uttrykkes i vegetativt mycelium når det angripes av soppspisende nematoder. Det har fire kjedelignende ricin-B domener, potensielt lektin-domener som kan binde glykan og gjør den i stand til å trenge inn i celler, samt et femte domene med ukjent struktur og betydning.

Produksjon av store mengder protein er essensielt i strukturell biologi. I **manuskript I** beskriver vi metoder for proteinproduksjon med høy effekt ved bruk av VmaxTM celler (Vibrio natriegens). VmaxTM-celler ble brukt i produksjon av CT og CCTX2 og gav tilstrekkelig mengde protein til å gjennomføre de strukturelle og funksjonelle analysene beskrevet i dette arbeidet. I **manuskript II** forsøker vi å forklare forskjellen i giftighet mellom CT og varmeømfintlig enterotoksin (LT), det tilsvarende (ortologe) stoffet i *Escherichia coli*. Vi produserte to to kimeriske CT protein, hvor C-terminusen til delenhet A har en til fire kritiske aminosyresubstitusjoner som samsvarer med C-terminusen til delenhet A i LT. Disse variantene ble sammenlignet med de naturlige forekommende proteinene, både funksjonelt og strukturelt. Begge de kimeriske proteinene utviste lavere giftighet enn naturlig CT og deres tilbøyelighet til å bli brutt opp av protein disulfid isomerase ligner på den man ser i LT. Vi løste strukturen til begge variantene med røntgenkrystallografi og avslørte at C-terminus området til proteinet med fire-substitusjoner har samme bretning som LT, som virker avgjørende for CT-prosessering i verten. I tillegg undersøkte vi hvordan substitusjoner påvirker fleksibiliteten til de ulike kimeraene, som sannsynligvis påvirker interaksjonen med PDI og effektiv oppbrytning. I **manuskript III** utforsker vi internaliseringskjeden og den mulige skjebnen til forsvarsproteinet CCTX2 inni cellen. Vi avdekket karbohydratet som gjenkjennes av lektin-domenene til CCTX2 og dets internaliseringsmekanisme. Vi brukte videre én-partikkel kryogenisk elektromikroskopi (cryo-EM) for å bestemme strukturen. De strukturelle dataene viser en helt ny brett for det toksiske domene som ikke

forekommer blant dataene som allerede finnes i Protein Data Bank (PDB), en database for proteiner. Strukturdata alene kunne ikke gi videre innsikt i aktiviteten til det toksiske domene, men vi oppdaget den potensielle rollen til CCTX2 som en ART. Vi fulgte opp resultatene fra **manuskript III** med å utforske mulige kofaktorer og ligander med differensiell fluorimetriskanning (DSF). Screeningresultater peker mot en mulig interaksjon mellom CCTX2 og nukleotider, mer spesifikt nicotinamide adenine dinucleotide (NAD) og adenosine trifosfat (ATP). Det destabiliseres sterkt av divalente metalliske kationer, med unntak av $Mg^{2+}$ som ofte bidrar i binding av nukleotid-bindene ensymer, hvilket stemmer overens med de foreløpige resultatene fra **manuskript III**.

# Abbreviations

| | |
|---|---|
| **ΔNCCTX2** | two first N-terminal domains truncated CCTX2 |
| **CCTX2ΔC** | fifth C-terminal domain truncated CCTX2 |
| **ADP** | adenosine diphosphate |
| **Amp** | ampicillin |
| **ARF** | ADP-ribosylation factor |
| **ART** | ADP-ribosyltranferase |
| **ATP** | adenosine triphosphate |
| **CAM** | chloramphenicol |
| **cAMP** | cyclic adenosine monophosphate |
| **CCTX2** | *Coprinopsis cinerea* toxin 2 |
| **CDT** | cytolethal distending toxin |
| **CFTR** | cystic fibrosis transmembrane conductance regulator |
| **cryo-EM** | cryogenic electron microscopy |
| **CT** | cholera toxin |
| **CTA** | cholera toxin subunit A |
| **CTA1** | cholera toxin subunit A1 |
| **CTA2** | cholera toxin subunit A2 |
| **CTB** | cholera toxin B pentamer |
| **CTv1** | cholera toxin variant 1 |
| **CTv4** | cholera toxin variant 4 |
| **DMSO** | dimethyl sulfoxide |
| **DSF** | differential scanning fluorimetry |
| **DTT** | Dithiothreitol |
| **EDTA** | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| **ELISA** | enzyme-linked immunosorbent assay |
| **ER** | endoplasmic reticulum |
| **ERAD** | ER associated degradation machinery |
| **ETEC** | enterotoxigenic strains of *E. coli* |
| **Gal** | galactose |
| **GalNAc** | N-acetylgalactosamine |
| **GARP** | Golgi associated retrograde protein |
| **GbpA** | N-acetyl glucosamine-binding protein |
| **$G_{M1}$** | monosialosyl ganglioside |
| **GPCR** | G protein coupled receptor |
| **G protein** | guanine nucleotide-binding proteins |
| **$G_{s\alpha}$** | α-subunit of $G_S$ protein |
| **GSL** | glycosphingolipid |
| **HEPES** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **hLT** | human LT |
| **Hsc70** | heat sock cognate 71 kDa |
| **Hsp90** | heat sock protein 90 kDa |
| **IPTG** | isopropyl β-D-1-thiogalactopyranoside |
| **Kan** | kanamycin |
| **LacdiNAc** | *N*-acetylated lactose |
| **LPS** | lipopolysaccharide |
| **LT** | *E. coli* heat-labile enterotoxin (type I) |
| **LTA** | LT subunit A |
| **LTA1** | LT subunt A1 |
| **LTA2** | LT subunit A2 |
| **LTB** | LT B pentamer |

| | |
|---|---|
| **MOA** | marasmius oreades agglutinin |
| **MTX** | mosquitocidal holotoxin |
| **NAD** | nicotinamide adenine dinucleotide |
| **OB-fold** | olinucleotides/oligosacharride binding-fold |
| **OMVs** | outermembrane vesicles |
| **PARP** | poly-ADP-ribosylpolymerase |
| **PDB** | protein data bank |
| **PDI** | protein disulfide isomerase |
| **pLT** | porcine LT |
| **qRT-PCR** | quantitative reverse transcription polymerase chain reaction |
| **RBCL** | ricin B chain-like |
| **r.m.s.d.** | root-mean square deviation |
| **SEC** | size exclusion chromatography |
| **SPR** | surface plasmon resonance |
| **TCP** | toxin coregulated-pilus |
| **TGN** | trans-Golgi network |
| **TLC** | thin-layer chromatography |
| **T2SS** | type 2 secretion system |

# I. Introduction

Infectious diseases are a major cause of death worldwide[1], affecting mainly developing countries , due to a lack of clean water, safe food or access to vaccination. Many of these diseases, such as tetanus[2,3], diphtheria[4,5] and cholera[6], are caused by bacteria carrying protein toxins, secreted by the pathogen to promote infection[7]. At the same time, other organisms, such as plants and fungi, use protein toxins as a defense strategy against predators[8]. A better understanding of these toxins and the molecular mechanisms underlying their activity is crucial for the design of better therapeutics against pathogens (e. g. vaccines from detoxified pertussis toxin[9]), or to find novel applications for their use in industry and biotechnology (e. g. use of botulinum toxin for involuntary muscle disorders [10]).

## 1. AB toxins

AB toxins are defined as a group of proteins mainly found in bacteria (e.g. cholera toxin[6] or shiga toxins[11]), where they are usually involved as attack proteins[12], and in plants (e.g. ricin[13]), where they usually play a role in host defense[8]. They are formed by two types of components: the **A**-subunit, or **A**ctive-subunit, responsible of their toxicity, and the **B**-subunit, or **B**inding-subunit, responsible for the entry in the host-cell[12]. AB toxins can be classified in different groups, depending on the relative stoichiometry of the A and B subunits; for example $AB_2$ (e.g. cytolethal distending toxin, CDT[14]) or $AB_5$ (e.g. CT[6], shiga toxins[11]). AB toxins are either produced as a single large polypeptide chain (e.g. by Gram-positive bacteria)[12] or as separate proteins, often part the same operon (e.g. by Gram-negative bacteria)[12].

## 2. AB toxins for attack: cholera toxin

### 2.1. Cholera and the role of *Vibrio cholerae*

Cholera is an acute diarrheal disease with an estimated 2.9 million cases and 95,000 death per year[15]. It is an ancient disease originating from the Bay of Bengal and caused *seven* pandemic waves over the last two centuries[16]. It mostly affects people in developing countries through contaminated water or food sources[17]. Outbreaks of cholera are very common after a natural disaster or a war, when the sanitation and access to clean water and food sources is restricted[18]. For example, Syria has been deeply affected by cholera, as a result of the ongoing civil war and the 2023 earthquake striking southern Turkey and northern Syria[19,20]. Treatment of cholera is mainly symptomatic, by rehydration of the patient[17]; recovery is quick in case of early treatment[21]. So far, preemptive vaccination and improving sanitation and hygiene have been used to eradicate cholera[22]. However, the recurring cholera epidemics

demonstrate the limitations of this approach. Conflicts, climate change[23,24,25] and more and more frequent natural disasters point to the urgent need for new cures, as outbreaks might become more regular.

Cholera is caused by toxigenic strains of a bacterial pathogen: *Vibrio cholerae*[6]. This comma-shaped Gram-negative bacterium harbors a flagellum and is naturally found in aquatic environments, free-swimming or attached to biotic or abiotic surfaces[26]. Pathogenic *V. cholerae* strains contain a lysogenic phage[27] (CTXϕ) with its separate genome. CTXϕ is crucial for virulence, as it encodes two proteins involved in the intoxication process: toxin coregulated-pilus (TCP), important for adhesion of the bacteria in the small intestine[28], and CT, responsible for diarrhea and dehydration in the target host[6]. *V. cholerae* can be classified into more than 200 different serogroups based on the structure of its surface lipopolysaccharide (LPS)[29]. However, mainly two of them are causing most cholera outbreaks (O1 and O139), as the others often lack cholera toxin-coding genes[29]. Infection by the O139 strain is less frequent, limited to Asian countries and only caused limited epidemics so far[30]. In contrast, O1 is responsible for most cholera outbreaks around the world and can be classified in two major types: Inaba and Ogawa, which differ in the antigen composition of their lipopolysaccharide (LPS)[31]. Additionally, the O1 serogroup can be divided into two biotypes: classical and El Tor[16]. O1 classical biotype is the cause of the fifth and sixth pandemics and most likely earlier pandemics[32]. The first occurrence of El Tor was in the 1930s[33], and it is responsible for the seventh pandemic of cholera[32], which started in the 1960s. The main difference between classical and El Tor *V. cholerae* lies in the carried CTXϕ variant, resulting in the secretion of two closely related variants of the CT toxin: classical CT and El Tor CT[34,35]. However, some hybrid El Tor *V. cholera* strains carrying classical CT have been identified, causing most current cholera outbreaks[36].

## 2.2. Cholera toxin and its molecular structure

CT is the main virulence factor of *V. cholerae*, and it is an $AB_5$ toxin[6]. $AB_5$ is a common AB toxin arrangement, found as a very important virulence factor in several bacterial pathogens (e.g. heat-labile enterotoxin[37], shiga toxins[11] or pertussis toxin[38]). The B subunits are arranged as a pentamer and allow cell entry, while the A subunit is split between an A1 catalytic subunit and an A2 linker[39,40]. $AB_5$ toxins can be classified in four groups, depending on their sequence homology and A subunit activity: subtilase activity, ADP-ribosylase activity and RNA N-glycosidase activity[39,40]. Consistently, CT is formed by six polypeptide chains: a pentamer of B subunits (CTB) and an A subunit (CTA)[41]. The *ctxB* and *ctxA* genes encoding for the two subunits are contained in the same operon, encoded by the genome of CTXϕ[27]. *ctxB* encodes for an 11.5 kDA polypeptide, while the *ctxA* product is a 23.5 kDa polypeptide[42]. Both components carry an N-terminal signal sequence, targeting them to the periplasmic space. The A and B

subunits of CT are cotranslationally secreted via the Sec pathway to the periplasm of the bacteria and assembled into the holotoxin heterocomplex[43]. From there, the toxin and other proteins are secreted through the type II secretion system (T2SS) to the extracellular space[44,45]. Then, the A subunit undergoes post-translational modification to acquire its toxic potential[41,46]. CTA is nicked either by a hemagglutinin/protease[47] secreted along CT[48] or by a serine protease within the intestinal lumen (e. g. trypsin)[48] but the actual mechanism is not entirely clear. Cleavage takes place at site R192[46] but has also been shown to happen at S193 and S194[49]. It generates the A1 (21.5 kDa) and A2 fragments (5.5 kDa)[42], which remain assembled to the B pentamer.

The first CT structure was solved by X-Ray crystallography in 1995[50] (Protein data bank (PDB) ID: 1XTC) but is of relatively low quality. Most of the structural work on CT has been carried out by the group of Wim Hol in the 2000s, with several structures published in the PDB[51] (PDB ID: 1S5B, 1S5C, 1S5D, 1S5E and 1S5F). The overall shape of the holotoxin evokes a turnip, with a diameter of approximately 65 Å and a height of 35 Å[52]. The B pentamer harbors a donut-like structure, with a 5-fold symmetry, and a central pore. The A2 subunit extends through the pore, acting as an anchor for the A1 subunit[41], which sits over the pentamer (Figure 1).

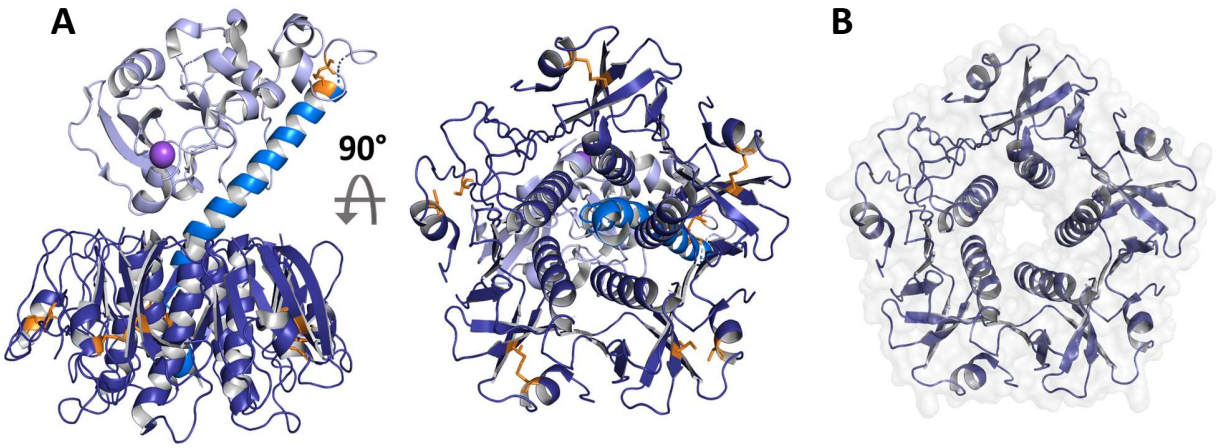


**Figure 1. Structure of CT and its subunits. A.** Overall structure of CT (based on 1S5E[51]) with its B pentamer (dark blue), its A2 subunit (blue) and it A1 subunit (light blue) and its disulfide bridge (orange). Bottom view of the B pentamer (dark blue) showing the C-terminus of the A2 subunit (blue) inside the pore and the A1 subunit in the background (light blue). **B.** Bottom-view of the B pentamer surface (gray) with the five B-subunits (dark blue) highlighting the pore.

Each B subunit exhibits an oligonucleotide/oligosaccharide-binding fold (OB fold)[53]. It consists of two antiparallel beta-sheets, formed by three strands each (first sheet: β2, β3, β4 and second sheet β1, β5, β6), sandwiched between two alpha-helices (α1, α2) (Figure 2A). Additionally, the first sheet from one subunit associates with the second sheet from the neighboring subunit to form a six-stranded β-sheet.

Moreover, the structure is held together by a disulfide bond between C9 and C86 from α1 and β5 of each B subunit (Figure 1A). The pore is constituted by the α2 helices from each B subunit, exposing charged and hydrophobic residues that create an amphipathic environment. It has a diameter that varies between 11 Å and 16 Å[54].

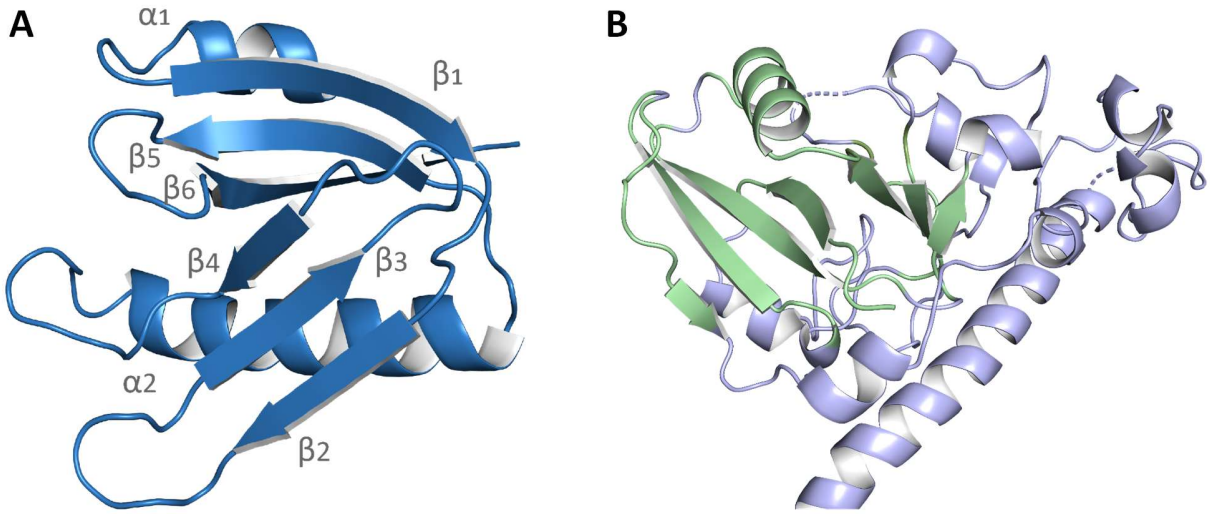


**Figure 2. Characteristic folds of CT: the OB-fold and ART-fold. A.** Cartoon representation of the OB-fold of the B pentamer of CT (based on 1S5E[51], chain D of the B pentamer). Components of the OB-fold are labelled as described in *Murzin et al.*[53]. **B.** Cartoon representation of the ART-fold (light green) of the A1 subunit of CT (light blue, based on 1S5E[51], chain A).

Each B subunit recognizes the monosialosyl ganglioside ($G_{M1}$) on target cells[55]. $G_{M1}$ is a sialic acid-containing pentasaccharide composed of Galβ3GalNAcβ4(NeuAcα3)Galβ4Glcβceramide (Figure 3). The galactose (Gal) and *N*- acetylgalactosamine (GalNAc) units of the sialic acid of $G_{M1}$ bind in the so called "two-fingered grip" at the bottom of each B subunit[56]. The binding pocket is primarily formed by the loops from the B subunit and most interactions are performed by amino acids from the same subunit and mainly with W88 through aromatic stacking[56] (Figure 4A).

The A1 subunit carries an adenosine diphosphate (ADP)-ribosyltransferase (ART) fold[57], common to other ADP-ribosylating toxins, formed by a split anti-parallel β-sheet and a central α-helix (Figure 2B). Using NAD as ADP-ribose donor, ARTs catalyze the transfer of ADP-ribose on acceptor proteins, either as a single unit or as a polymeric chain, like in case of poly ADP-ribose polymerases (PARPs)[58]. The active-site is formed by the typical ART fold, with the RSE catalytic residues (R53, S63 and E112), which is an important motif among ARTs[59] (Figure 4B). Binding of the NAD cofactor is prevented by a loop[51], displaced upon activation in the host cell[60].

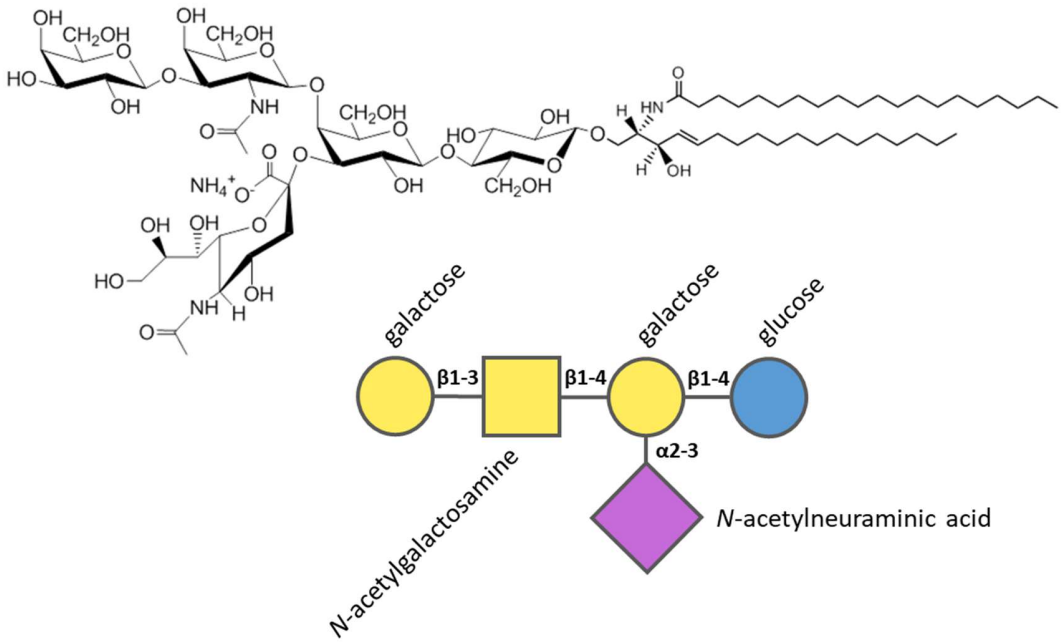


**Figure 3. Molecular structure of the $G_{M1}$ glycolipid and its carbohydrate portion representation according to the symbol nomenclature for glycans.**

The A2 subunit is made of a long α-helix wrapping around A1, and with its C-terminus inserted inside the pore of the B pentamer. The C-terminal signal sequence (KDEL), targeting the protein to the ER, is located close to the exit of the pore, at the bottom of the B pentamer (Figure 1). The A2 domain interacts with the B pentamer through water-mediated interactions and salt bridges; the A1 and A2 subunits remain associated through a disulfide bond (C187 and C199, Figure 1A).

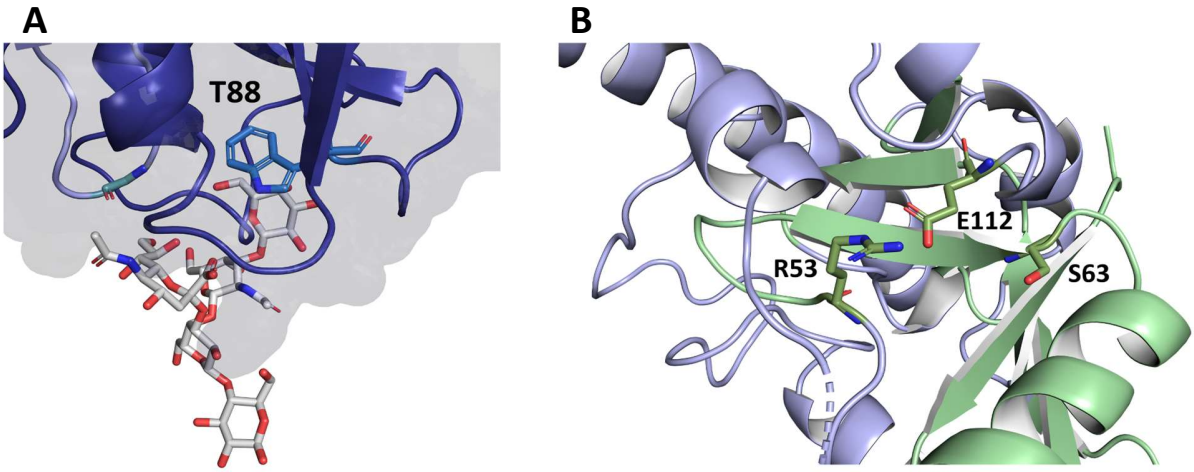


**Figure 4. Binding-sites of CT. A.** CT in complex with $G_{M1}$ oligosaccharide (grey) at the bottom of one subunit of the pentamer (dark blue) showing the aromatic stacking with W88 and the two fingers-grips of the ligand. **B.** Active-site of the subunit A1 (light green), involved in the transfer of ADP-ribose $Gs^{61}$ with the RSE-motif.

## 2.3. Cholera toxin uptake and intoxication mechanism

After ingestion, *V. cholerae* travels through the stomach, before attaching and colonizing the small intestine[62]. For that purpose, it uses several colonization and virulence factors helping to recognize and bind the epithelial cells and colonize the host. Among them, there are adhesins[63], for anchoring the bacteria, *N*-acetylglucosamine binding protein A (GbpA), involved in the colonization and CT, responsible for diarrhea and dehydration[64]. To infect the host, the B pentamer binds on the $G_{M1}$ exposed on the surface of epithelial cells, with each B subunit interacting with a separate $G_{M1}$ ganglioside molecule[65,66]; $G_{M1}$ binding by a single B subunit is sufficient for intoxication of the host cell[67]. CT enters the cell through endocytosis: after retrograde trafficking through the *trans*-Golgi network (TGN), the protein travels to the endoplasmic reticulum (ER)[68] (Figure 5).

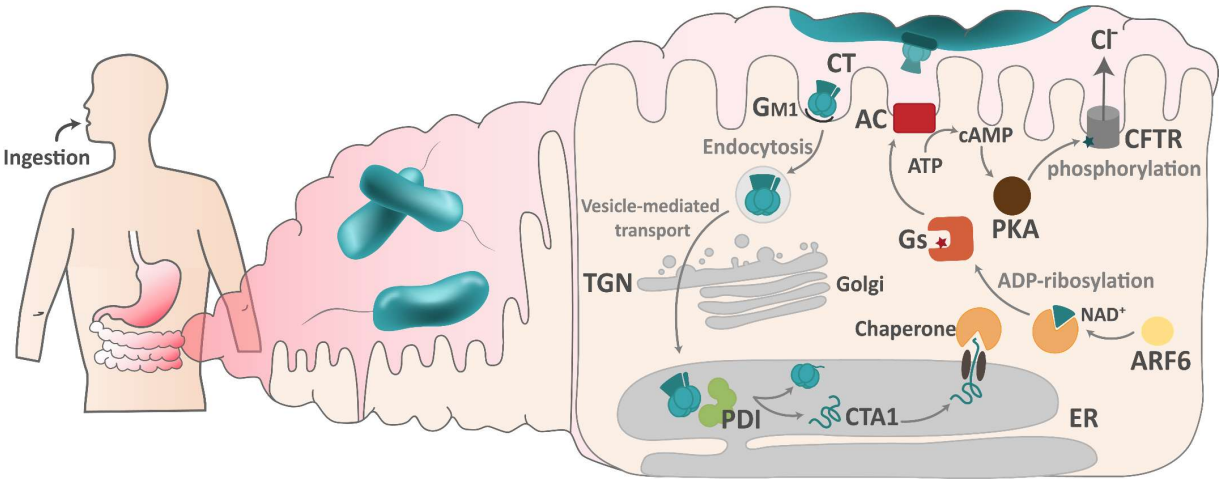


**Figure 5. Cholera toxin intoxication mechanism.** After ingestion of contaminated food or water[17], *V. cholerae* colonizes the intestine and secretes CT[44,45]. Upon binding of $G_{M1}$[55], CT is endocytosed and transported to the ER through retrograde trafficking[68]. With the help of PDI, CTA1 is released from the rest of the protein[73]. By hijacking the ERAD system[77], it is transported to the cytosol where is it refolded by chaperones[79]. After activation of CT by ARF6[60], it ADP-ribosylates $G_{\alpha s}$[61], which then activates the adenylate cyclase (AC)[81]. In turn, AC increases the amount of cAMP[82], which activates PKA[87]. PKA further phosphorylates CFTR[88], which leads to a loss of chloride ions[90] and causes watery diarrhea[89].

The CTA1 subunit of CT is the effector responsible for the diarrhea and dehydration symptoms associated to cholera[64]. However, it needs to be released from the B pentamer and the CTA2 anchor, and reach its molecular target in the cytoplasm of the intestinal epithelial cells. In the ER, the C187-C199 disulfide bond holding together the CTA1 and CTA2 subunits is reduced[69], but this alone does not result in disassembly of CTA1 from the rest of CT. Protein disulfide-isomerase (PDI) is an enzyme/chaperone located in the ER, catalyzing the reduction of disulfide bonds and the folding of proteins (e.g.

ribonuclease A[70]), but exploited by various toxins (e.g. ricin[71]) and viruses (e.g. SV40[72]). Reduced PDI is necessary to displace CTA1 from the rest of the protein[73], but contrary to the initial conjectures it does not act as an unfoldase[73]. CTA1 possesses a natural disordered structure at body temperature, which causes it to unfold spontaneously[74]. This, in turns, displaces PDI[73], promotes PDI unfolding[75] and its dissociation from CTA1[76] (Figure 5). To exit the ER, the unfolded CTA1 takes advantage of the ER-associated degradation machinery (ERAD)[77], usually responsible for disposing of misfolded and defective proteins. The extraction process of CTA1 from the ER is coupled with its refolding[78], necessary for the protein to be active (Figure 5). This is performed by chaperones of the host cell: heat shock protein of 90 kDa (Hsp90) and heat shock cognate 71 kDa (Hsc70)[79]. Hsc70 assists in the translocation of CTA1 from the ER to the cytosol and Hsp90 is involved in the refolding of CTA1[79]. Hsp90 recognizes two motifs within CTA1, and particularly an N-terminal sequence (RPPDEI), also present in other ER-trafficked toxins[80]. The final step of activation of CTA1 relies on the association with ADP-ribosylation factors (ARF), a family of guanine nucleotide-binding proteins (G proteins)[60]. ARF6 displaces a CTA1 loop shielding the active site[60], allowing binding of the nicotinamide adenine dinucleotide (NAD) cofactor, the ADP-ribose donor needed for toxic activity[60] (Figure 5).

After activation, CTA1 is capable of ADP-ribosylating the effector protein $G_S$[61], where it transfers the ADP-ribose moiety from the NAD molecule to the α-subunit of the $G_S$ protein: $G_{s\alpha}$. This leads to increased active $G_S$ levels within the cells, and therefore increased activity of its target: adenylate cyclase[81] (Figure 5). This membrane protein is a crucial player in cell signaling and catalyzes the conversion of adenosine triphosphate (ATP) into cyclic adenosine monophosphate (cAMP)[82]. cAMP is involved in a variety of biological functions such as muscle contraction[83], hormone secretion[84] and immune response[85]. In cholera, the main altered function is the ion transport through over-activation of cyclic nucleotide-gated ion channels, and particularly the cystic fibrosis transmembrane conductance regulator (CFTR), an important chloride ion channel[86]. There, the cAMP-dependent protein kinase A (PKA) is activated[87], which then phosphorylates CFTR[88] (Figure 5). This increases the level of chloride ions that are transported across the membrane of the epithelial cells and results in a massive loss of water and electrolytes: in other words, diarrhea[89,90].

### 2.4. CT, more potent than its orthologue: *E. coli* heat-labile enterotoxin LT

Enterotoxigenic strains of *E. coli* (ETEC) are Gram-negative bacteria responsible for travelers' diarrhea[37]. This disease is comparable to cholera disease, causing similar but usually milder symptoms (diarrhea, dehydration, etc and is also transmitted through contaminated water and food. It occurs mostly in developing countries[37], affecting especially young children. It is not restricted to humans, as it also affects livestock[91] and in particular pigs. For intoxication, ETEC produces heat-stabile toxins (STs) and

heat-labile enterotoxins (LTs)[92]. STs are small peptides that are stable at high temperature (100°C) and non-immunogenic[93], making them difficult treatment targets. LTs are $AB_5$ toxins, members of the CT family, and they can be inactivated by heat-treatment at 60°C[94] (hence the name). LTs can be further classified in two serogroups: LT-I and LT-II[95]. LT-I (further referred to as simply LT in this work) is an orthologue of CT with a high sequence identity (≈ 80%). So far, the structure of LT from human isolates (hLT) has not been solved, but porcine LT from porcine infections (pLT) has been thoroughly studied and it was the **first** solved structure of an $AB_5$ toxin[96]. Like CT, LTs harbor five B subunits, which form a pore, and an A subunit with an A1 toxic peptide chain (LTA1) and an A2-linker chain (LTA2)[97]. Despite a largely conserved sequence and a structural resemblance to CT (root-mean square deviation, r.m.s.d = 1.1 Å; Figure 6A), their delivery mechanism to the host cell can be different. LT toxins can be either secreted using the T2SS[98] or with outer membrane vesicles (OMVs)[99]. On the other hand, mechanism of action and intracellular targets are conserved between the two toxins (Figure 5)[100].

Although CT and LT share many similarities, it has been shown that CT is more potent than LT[101]. One hypothesis speculates that the interactions of CTA2 with the B pentamer are stronger than in LT, therefore providing greater amounts of free CTA1[101]. Another hypothesis traces the change in potency to their relative difference between the angle formed by the A2 subunit and the B pentamer[102]. In the case of CT, this angle has been measured to be 49°, while in LT it is 40°[102]. It has been proposed that this could affect the accessibility of A1 by the PDI chaperone, which could then alter the dynamic of the PDI-driven disassembly[102].

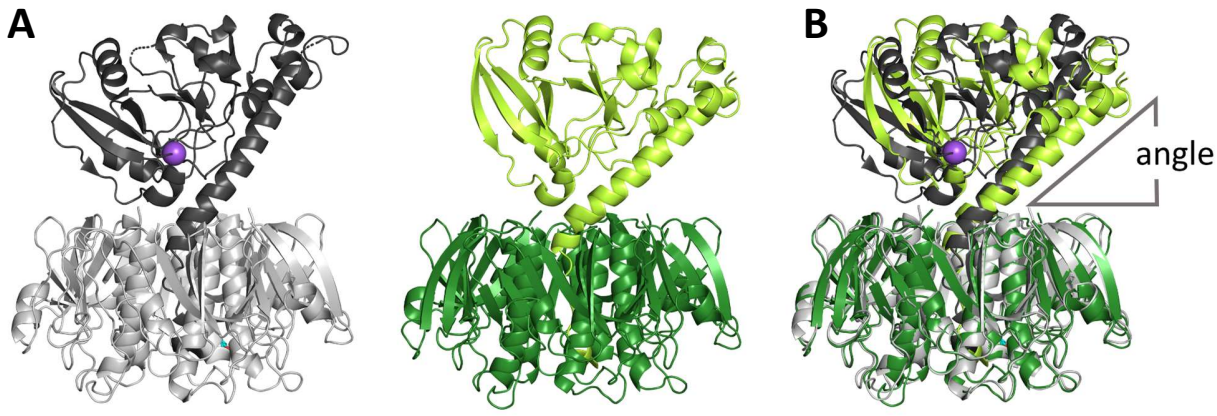


**Figure 6. Resemblance of CT to LT. A.** Structures of CT (based on 1S5E[51], gray) and LT (based on 1LTS[97], green) show their structural resemblance, both harboring a B pentamer (in a darker shade), which serves as a base for their A subunit (in a lighter shade). **B.** Superimposed structure of CT (based on 1S5E[51], gray) and LT (based on 1LTS[97], green) showing the difference of angle between the A2-subunit and the B pentamer plane.

# 3. AB toxins for defense: CCTX2, a nematotoxin

## 3.1. Lectins and their role as defense proteins

To defend themselves against predators, different species have developed different defense mechanisms, as "elaborate" as camouflage[103] or as "simple" as evasion. Because of their lack of mobility, many plants and fungi use toxins as a defense strategy. Among them, AB toxins are commonly found in the arsenal of defense weapons of plants and fungi, with ricin being the best-known example[13].

Before being correctly identified as AB toxins, several were known in literature as sugar-binding proteins, or *lectins*. Lectins are defined as a class of proteins[104] of non-immune, non-enzymatic origin, capable of binding specifically and reversibly to carbohydrates[105]. They are sometimes referred to as agglutinins[105], from the ability of many to agglutinate cells. As the cells are covered by glycans (glycoproteins and glycolipids): the *glycocalyx*[106], lectins can cause cell agglutination or glycoconjugates precipitation. Based on their domain composition, lectins can be classified in two categories: hololectins and chimerolectins[107]. Hololectins are composed of only carbohydrate-binding domains[107,108], whereas chimerolectins are proteins constituted of at least one non-lection domain, which harbors other functions[107,108], in addition to the lectin domain(s).

A large number of lectins behave as defense proteins[105,109], either through their lectin domain alone, (e. g. *Agrocybe aegerita* lectin[110]) or through a toxic non-lectin domain. This is the case of many chimerolectins (e. g. *Marasmius oreades* agglutinin[111] (MOA)). Most chimerolectins carrying a cytotoxic activity can essentially be considered AB toxins, where each lectin domain is identified as a B-subunit and each non-lectin toxic domain as an A-subunit.

## 3.2. CCTX2, a fungal nematotoxin

*Coprinopsis cinerea* toxin 2 (CCTX2) is a protein produced by the *Coprinopsis cinerea* fungus[112], commonly known as the *grey shag*. *C. cinerea* can be found all over the world, and its fruiting body is edible for humans. Its genome encodes for several nematotoxic protein that are upregulated when the fungus is attacked by predators[112]; among those, CCTX2, expressed in the vegetative mycelium. CCTX2 is located in the cytoplasm[112] of the fungal cell and its main known target are fungivorous nematodes (Figure 7). Nematodes are very small worms (maximum of a few millimeters long), known to be a threat in agriculture[113]. They feed by piercing the cells of their prey and aspirating their content[114], making CCTX2 a great defense protein. As of today, and despite many years of research, there are very few available nematicides with low human toxicity and low environmental impact[115], making CCTX2 a relevant candidate.

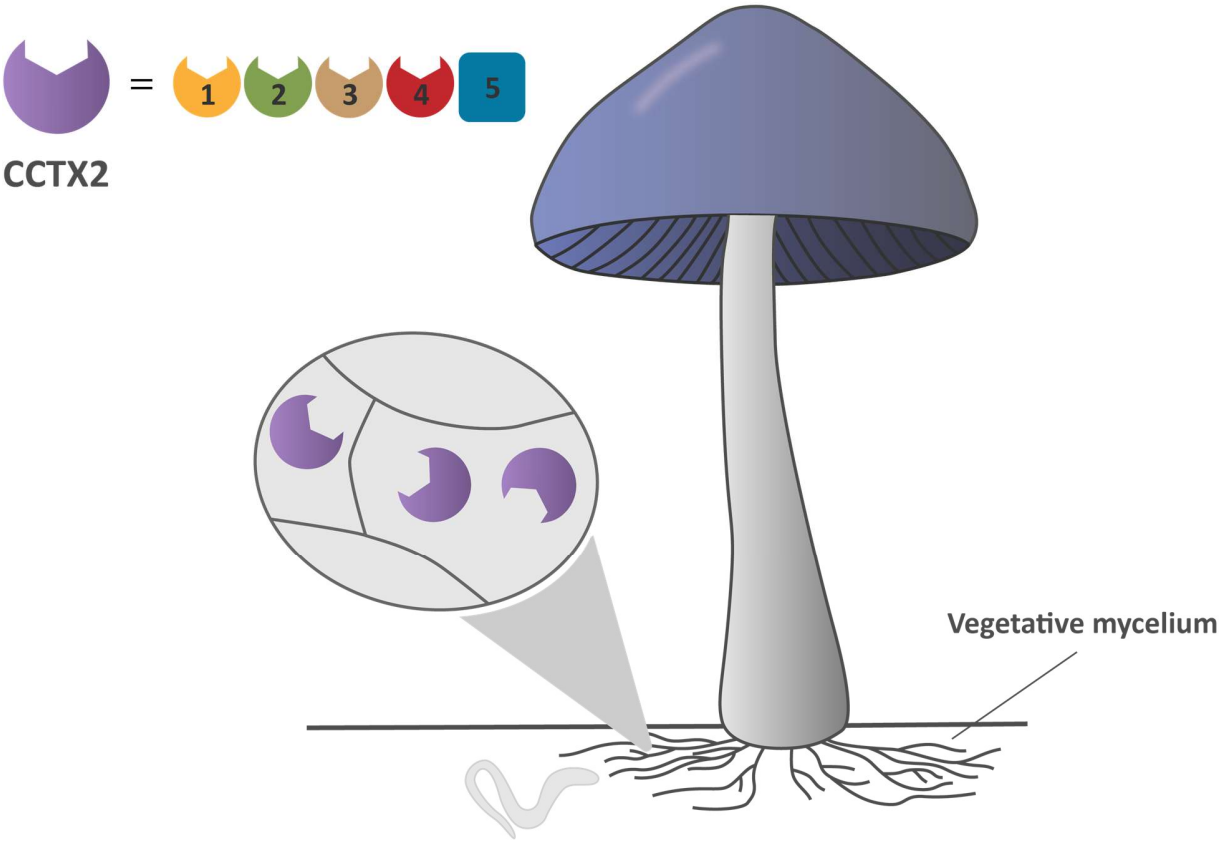


**Figure 7. CCTX2 in the vegetative mycelium.** In the vegetative mycelium, *cctx2* is expressed upon challenge by fungivorous nematodes. CCTX2 is composed of five domains, with the first four hypothesized to bind to the cells and the fifth domain is of unknown function.

### 3.3. CCTX2 and its molecular structure

CCTX2 is a monomeric protein of 89 kDa and consists of five domains[112]. So far, no structural information on CCTX2 has been published, but sequence-based domain analysis predicts that the first four domains[114] display the ricin B chain-like fold. The ricin sugar-binding region (B subunit) harbors a fold named after the protein itself and is widely common among plant lectins[116] (ricin B chain-like fold). This fold is also described as a β-trefoil: three β-sheets arranged in a pseudo three-fold symmetry, where each sheet forms a subdomain (α, β, γ; Figure 8)[117]. Due to the ricin B chain-like fold of the first four domains and the presence of the conserved $(QxW)_3$-motive in the first two domains, we hypothesize that they carry a lectin function but the exact nature of their recognized sugar motif, remains unknown. The last domain has negligible sequence identity with any protein with known functionality, and only matches orthologues from fungal genomes. As a consequence, its function remains unknown.

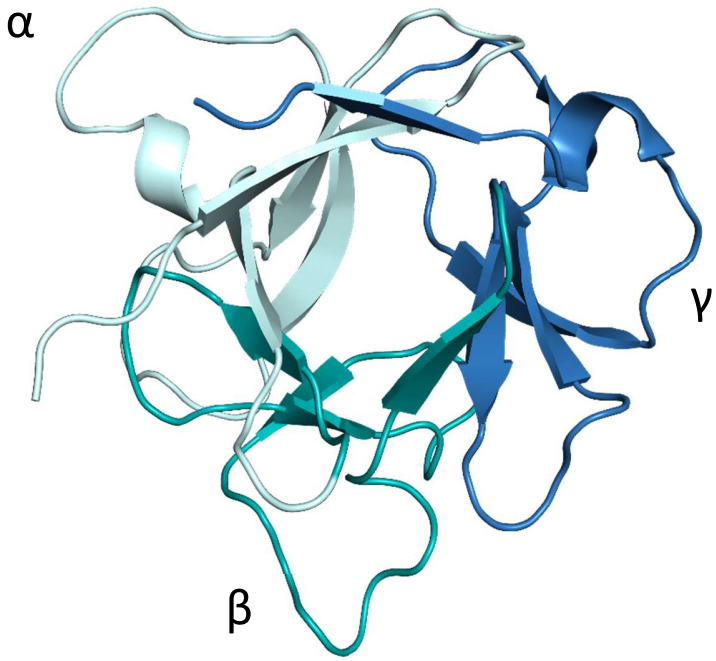


**Figure 8. 3D structure of EW29 (C-terminal domain of R-type lectin from the earthworm *Lumbricus terrestris*).** The structure (Based on 3RST) is an example of a protein carrying a β-trefoil. Subdomains α, β and γ are colored in light cyan, teal and blue, respectively.

# II. Aims of the work

In this work, we aim at a better understanding of AB toxins and their molecular mechanism of action. In particular, we decided to investigate CT and CCTX2, using structural and biochemical techniques. A better understanding of the structures of these toxins and their infection mechanism could lead to major breakthroughs in the design of inhibitors and vaccines (CT), or biocide synthesis and other biotechnological applications (CCTX2).

The structure of wild-type cholera toxin is already known and published. However, despite sharing a sequence identity of ≈ 80% with LT, we lack an understanding as of why CT has a higher potency than LT, orthologue found in *E. coli*. Our goal was to structurally characterize two chimeric structures of CT: substituting CT residues to LT residues. Comparing the high-resolution structures to wild-type CT and LT could allow us to identify a potential cause of the potency difference between the two toxins.

For CCTX2, we completely lack an understanding of its cellular uptake mechanism, its molecular mechanism and its intracellular target. Therefore, we aimed at solving its 3D structure to identify potential structural motives, which could point us to potential targets or substrates. Additionally, we performed basic screening using differential scanning fluorimetry (DSF) to identify potential co-factors and ligands to help us solve the molecular mechanism of CCTX2.

# III. Summary of manuscripts

## Manuscript I (draft)

### Using *Vibrio natriegens* for high-yield production of challenging expression targets

Natalia Mojica, Flore Kersten, Mateu Montserrat-Canals, Ute Krengel

The manuscript describes the Vmax™ X2 expression system for high yield protein production as an alternative to the *E. coli*-based expression systems. We present the limitations of the *E. coli* expression system and show improved production yields of several challenging expression targets. We facilitated the purification process and enhanced the purity of two targets (CT and GbpA) by taking advantage of the secretion machinery of Vmax™ X2, and increased their yields 10- and 6-fold, respectively. Additionally, the system was also shown to be ideal for producing high amounts of non-secreted challenging targets, as exemplified by CCTX2 (26-fold yield increase). In conclusion, Vmax™ X2 cells are a promising host for production of challenging protein targets in high amounts.

Significance: The manuscript contributes to bringing awareness of alternative systems (Vmax™ X2) to *E. coli* for the recombinant production of proteins, particularly in the field of structural biology. Additionally, it shows its versatility in the range of targets that can be produced, with an emphasis on toxins and challenging proteins.

Contributions from the author: Produced and purified GbpA from *E. coli*. Produced and purified CCTX2 from *E. coli* and *V. natriegens* cells. Produced Figure 5. Contributed to the writing of the manuscript.

## Manuscript II (draft)

### A single amino acid in the A2 linker of cholera toxin affects the efficiency of toxin disassembly and resulting toxicity

Flore Kersten*, Albert Serrano*, Joel B. Heim, Natalia Mojica, Jessica L. Guyette, Gabriele Cordara, Sergej Grudinin, Suren A. Tatulian, Ute Krengel, and Ken Teter

The manuscript tries to explore the origin of the differences in severity of the diarrheal disease caused by cholera toxin (CT) and heat-labile enterotoxin (LT). The most important differences have been traced to four positions in the C-terminal region of the A2 subunit (229, 230, 232, 233). A single CTA2 substitution to the residue found in LT (D229E) produces a toxin less potent than wild-type CT, also showing a less-efficient disassembly by PDI. Full substitution at all the four sites of CT to the LT sequence

(D229E, I230V, T232I, H233Y) shows similar results. Crystal structures of the one-site and four-site variants (CTv1 and CTv4) both toxin variants were determined. Structural data disproves the starting hypothesis, which considers a different intersubunit angle between the A and $B_5$ portion of the toxins as the key factor determining their toxic effect. However, normal mode analysis of the two variants shows a greater flexibility of the A subunit of wild-type CT compared to wild-type LT, with the CTv1 and CTv4 variants appropriately falling between the two. This hints to the importance of the CT/LT dynamic behavior in disassembly by PDI, and possibly underlies the different potency of the toxins.

Significance: This manuscript is an important work on the understanding the differences between CT and LT and particularly in their toxicity. It paves the way for a better comprehension of the mechanism of activation of both proteins and might help in the design of more effective treatments.

Contributions from the author: Expressed and purified CTV1 and CTV4 from *Vibrio natriegens*. Crystallized CTV1 and CTV4. Solved and refined the X-ray crystal structure of CTV4. Generated Figure 5, Figure 6, Figure 7, Figure S2, Figure S3, Table S1, Table S2. Contributed to the writing of the first draft of manuscript.

## Manuscript III (draft)

### CCTX2, a fungal AB toxin involved in fungal anti-nematode defense

Gabriele Cordara*, Stefanie S. Schmieder*, Flore Kersten, Kevin Bärlocher, Clara H. Klee, Ute Krengel and Markus Künzler

The manuscript provides an in-depth characterization of a novel fungal nematotoxin: CCTX2. CCTX2 is expressed in the vegetative mycelium of the *Coprinopsis cinerea* fungus and is constituted of five domains. The first four domains have been predicted to harbor a ricin B chain-like fold (RBCL), while the last domain carries an unknown structure and function. The RBCL fold is often associated with a lectin activity, which was demonstrated for the two first domains of CCTX2, with a specificity for LacdiNAc moities. As it is the case for many AB toxins, CCTX2 was shown to be internalized by endocytosis before exerting its toxicity on the target cell. It was determined that the toxicity of the protein actually depends on the presence of its fifth domain, a probable effector domain. To understand the potential function of CCTX2, its 3D structure was solved to a 3.2 Å resolution using single particle cryo-EM. The structure shows a completely new fold with four RBCL domains and their typical β-trefoil, serving as a platform for the fifth domain. Although the exact function of the fifth remains yet to be determined and could not be deducted from the structure, CCTX2 can be classified as a defense AB toxin.

Significance: The manuscript is an important work on the understanding of the fungi world and their elaborate defense mechanism. The internalization pathway of CCTX2 has been described, showing

similarities to plants and bacterial AB toxins and their presence in all kingdom. The work on the structure additionally describe a completely new fold and potentially the discovery of a new class of proteins.

Contributions from the author: Expression and purification of CCTX2 in *Escherichia coli*. Cryo-EM data analysis. Wrote the first draft of Materials and Methods, Results and Discussion for the structural biology part of the manuscript.

# IV. Results and Discussion

## 1. Recombinant expression of toxins

Structural studies on proteins require high amounts of highly pure starting material. It is often the first hurdle to overcome and a major bottleneck in structural biology. Many strategies can be employed to gather enough target protein, but in most cases, the gene is recombinantly expressed in a non-native organism. One of most the common expression hosts is *E. coli*. This bacterium is very well understood, easy to culture and has a quick generation time (20 minutes at 37°C[118]), all good features for its use in protein production. However, the expression of toxins such as CT or CCTX2 can be more challenging than non-toxic protein[119,120]. In some cases, the toxicity is too high for the expression host itself, and protein production can be impaired, with low yields or even impossible. Therefore, some strains have been engineered to sustain production of toxins such as C41(DE3) or C43(DE3)[121] and are commonly used as expression hosts.

In **Manuscript I**, we present the limitations of *E. coli* expression system and an alternative expression host: Vmax™ X2. The Vmax™ X2 strain is derived from *Vibrio natriegens*, another Gram-negative bacterium, which has a faster growth-rate than *E. coli*[122]. This system was used to improve production yields of several proteins: classical CT, pLT, N-acetyl glucosamine-binding protein (GbpA) and CCTX2. When CT is expressed in *E. coli* systems, the protein is directed to the periplasmic space; however, it is not transported across the outer membrane, as it would be in its native host. As a consequence, expression of CT in *E. coli* leads to sample heterogeneity, due to the mixed presence of holotoxins and B pentamers in the periplasmic extract. The two are difficult to separate, due to their very similar chemico-physical properties. On the contrary, Vmax™ X2 carries a translocation machinery that closely resembles that of *Vibrio cholerae*. This resulted in a higher yield of the recombinant classical CT and the preferential secretion of the holotoxin in the culture media. Production in Vmax™ X2 leads to a minimum 10-fold yield increase and a higher purity of the sample, due to the lack of the B pentamer contaminant (Figure 9A). The reliability of the system in structural studies was tested by determining the X-ray crystallographic structure of CT produced in Vmax™ X2 and comparing it to structures deposited in the PDB. The structure was solved to a 2.3 Å resolution (PDB ID: 8QRE), and it closely resembles the published structures of CT (r.m.s.d of 0.6 Å in comparison to PDB ID: 1S5E[51]). Additionally, it was confirmed that CT produced in Vmax™ X2 retained activity, further confirming the reliability of this expression system. Based on these results, Vmax™ X2 cells were used for the production of chimeric proteins CTv1 and CTv4, described in **manuscript II**.

Besides CT, the system was tested for the expression of pLT, an orthologue of CT from enterotoxigenic *E. coli*, and GbpA, a colonization factor of *V. cholera*. For both proteins, the use of Vmax™ X2 significantly increased the expression yield, and led to their secretion into the medium, which allowed an easier recovery and purification procedure. The Vmax™ X2 system was additionally tested for the expression of a non-bacterial protein: CCTX2. The expression yield exceeded expectations, with a 26-fold increase in comparison to the *E. coli*-based expression system (Figure 9B). This allowed us to produce recombinant CCTX2 in large amounts for structural and functional characterization, and for quantitative analysis.

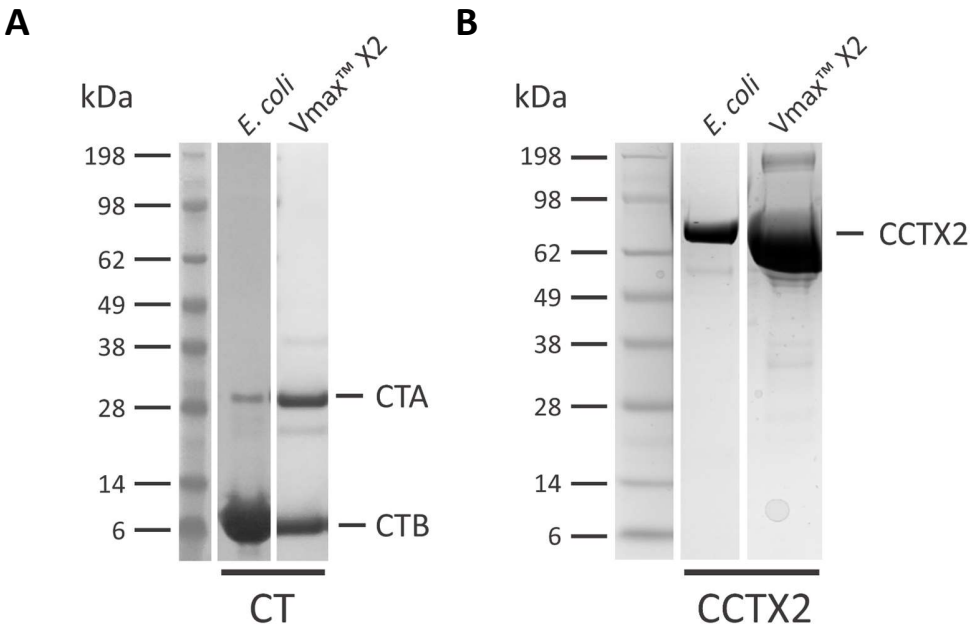


**Figure 9. SDS-PAGE analysis of CT and CCTX2**. **A.** Samples of CT obtained after SEC produced in *E. coli* and Vmax™ X2, which shows a lower concentration of the B pentamer only, proof of increased purity. **B.** Samples of CCTX2 obtained after SEC produced in *E. coli* and Vmax™ X2. Reproduced from manuscript I.

In conclusion, we successfully introduced in our workflow an alternative expression system to *E. coli*. The Vmax™ X2 system can easily be adopted for any protein and significantly improves the yields of more difficult targets, such as toxins, the focus of this thesis, also yielding active proteins for functional or structural characterization.

## 2. Studies on the potency of an attack AB toxin: cholera toxin

*V. cholerae* and enterotoxigenic strains of *E. coli* are known to cause diarrheal diseases, mediated by the CT and LT virulence factors, respectively. CT and LT are orthologous $AB_5$ toxins that share a very similar sequence and structure (≈ 80% sequence identity, r.m.s.d = 1.1 Å between 1S5E[51] and 1LTS[97]; Figure 7A). Their mechanism of action is also conserved and they have the same intracellular target ($G_{s\alpha}$).

Despite the many similarities, CT shows a higher toxicity toward its infection host compared to LT[123]. In their 1995 publication, Lencer *et al*. hypothesized the cause to the difference in the C-terminus ER-targeting sequence: **K**DEL for CT, **R**DEL for pLT and K**N**EL for hLT[123] (Figure 10A). Later on, the same group dismissed the hypothesis, as the substitution of **K**DEL to **R**DEL did not appear to affect the toxicity of CT[101]. Additionally, they mapped an 11-amino acid region in the A2-linker (226 to 236) to be responsible for the difference in toxicity between CT and LT[101]. By generating chimeras containing the C-terminal sequence of the other orthologue, they demonstrated that chimeric CT yields toxicity levels similar to LT and vice versa. They speculated that in CT, the interaction of CTA2 with CTB provides a greater stability to the holotoxin during uptake and transport compared to LT[i].

PDI-mediated ER disassembly is a crucial step in releasing the A1 subunit from the rest of the CT or LT holotoxin, allowing it to be transported to the cytoplasm of the host cell[124]. Ken Teter's group recently demonstrated that PDI-mediated disassembly is more efficient for CT than LT[102], which could influence the availability of CTA1/LTA1 in the cytoplasm and therefore explain their difference in potency. In the same publication, it was hypothesized that a different angle between the A subunit and the B pentamer provides a structural rationale for their different toxicity[102]. The angle between the upper portion of the A2 linker and the B pentamer, measured from structures deposited in the PDB, has been recorded to be 49° for CT and 40° for LT (Figure 10B). Assuming a model where PDI extracts the A1 subunit acting as a wedge, the different angle could influence the accessibility by PDI and therefore the release of the A1 subunit (Figure 10C).

CT and hLT differ for only five amino acids out of eleven identified by Lencer *et al*. to be crucial for their difference in toxicity[101]. The five differences include the ER retention sequence (R**D**EL vs R**N**EL) (Figure 10A), which has been shown not to play a role in toxicity[123]. This leaves four residues (**D**229, **I**230, **T**232, **H**233) as the origin for the higher toxic potency of CT. In **manuscript II**, we try to test this hypothesis and rationalize the observed A/B intersubunit angle difference in terms of the structural effect of the four substitutions. Two chimeric constructs of CT were designed. Because amino acid 229 represents the location where the α-helix of CTA2 bends when entering the B pentamer pore, the first chimera (CTv1) harbors a single substitution to the pLT/hLT sequence (**D**229**E**, Figure 10A). The second chimera (CTv4) carries the substitution of all four amino acids (**D**229**E**, **I**230**V**, **T**232**I**, **H**233**Y**, Figure 10A). The toxicity and PDI-driven disassembly of the two chimeras CTv1 and CTv4 and how it relates to wild-type CT and wild-type hLT was investigated by *in vitro* toxicity cell assays, enzyme-linked immunosorbent assays (ELISA) and surface plasmon resonance (SPR). The X-ray crystallographic structure of CTv1 and CTv4 was

[i] In the two studies from the Hirst group[101,123] the origin of LT (hLT or pLT) is not clearly stated). Our functional studies are performed on hLT, but PDI-disassembly of pLT was also monitored (see Supplementary Figure S1 of **manuscript II**) and structural comparison are performed against pLT, only available structures of LT.

solved to investigate the properties of the chimeras. We aimed at explaining the effect of the amino acid substitutions on the structure, potentially justifying the difference in toxicity, which would support a wedge hypothesis for the PDI-disassembly.

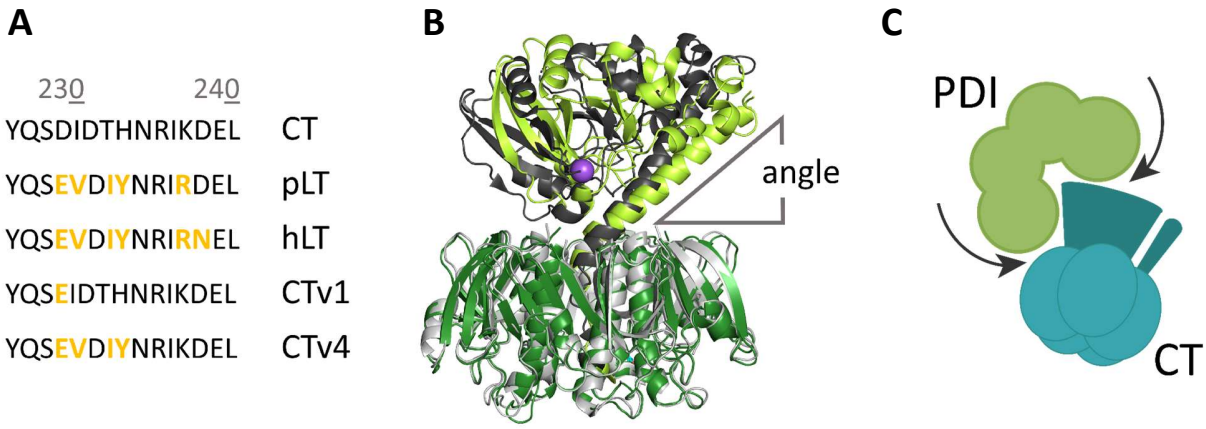


**Figure 10. Differences between CT and LT and disassembly hypothesis. A.** Sequence alignment of the crucial 11-amino acids region of the A2 C-terminus of CT, pLT, hLT and the variants CTv1 and CTv4, highlighting the differences in orange. **B.** Superimposition of CT (gray, based on 1S5E[51]), showing the angle of 49° between the A2 subunit and the B pentamer plane, and LT (green, based on 1LTS[97]), showing the angle of 40° of the A2 subunit with respect to the B pentamer. **C.** Potential wedge-model of PDI extracting CTA1 from the rest of the CT holotoxin.

## 2.1. Toxicity studies of CT chimeras

The Teter lab first investigated the toxicity of CTv1 and CTv4 in comparison to hLT and CT. All four proteins were treated with trypsin to cleave the A1 subunit from the A2 subunit, a needed step to get an active holotoxin. After incubating CHO cells with the toxins, the levels of cAMP were measured as an indicator of the toxic activity: the higher the toxicity, the higher the cAMP concentration. Unsurprisingly, wild-type CT shows the highest levels of cAMP (Figure 11A). On the other hand, both CTv1 and CTv4 present similar levels of cAMP to hLT (Figure 11A).

Thereafter, the efficiency of PDI disassembly of CTv1 and CTv4 was tested. ELISA was performed to measure the PDI disassembly of the A1 subunit from plate-bound B/A2 subunits (Figure 11B). In agreement with the previous assay, the results show the effective disassembly of CTA1 by PDI in CT, a disassembly of CTv1 less effective than observed for CT and an even less effective disassembly of CTv4.

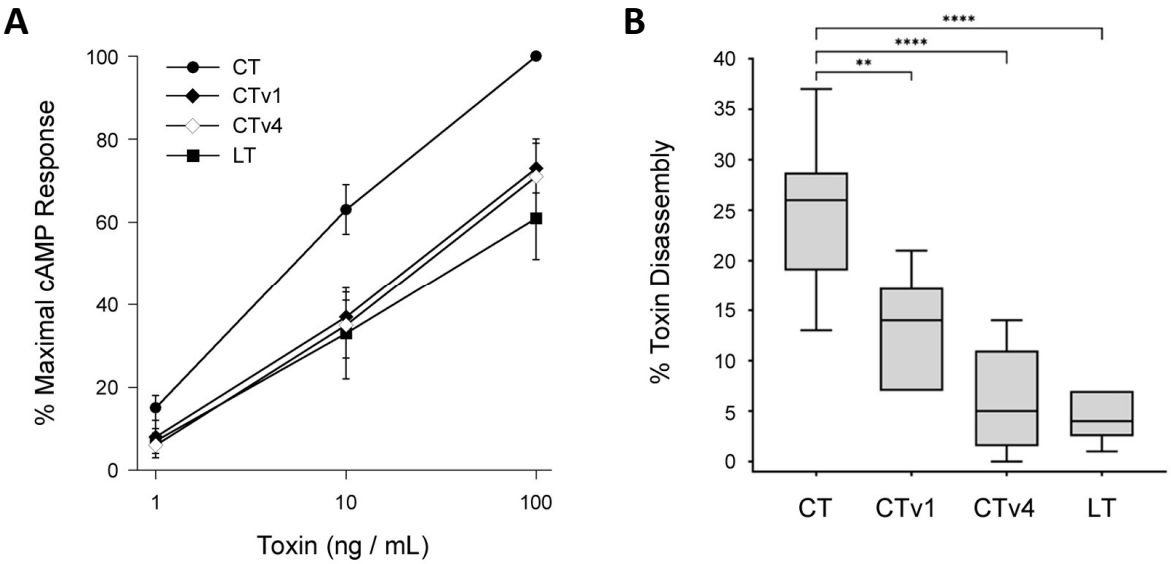


**Figure 11. Substitutions in the A2 subunit affect the toxicity and PDI-driven disassembly of CT. A.** CHO cells were incubated with CT (circles), CTv1 (closed diamonds), CTv4 (open diamonds), or hLT (squares) before measurements of the cAMP levels. Data were expressed as percentages of the maximal cAMP response for the assay and are presented as the means ± standard error of six independent experiments with triplicate samples. **B.** Presence of the A1 subunit in CT, CTv1, CTv4, and hLT was recorded using the maximum A1 signal obtained from the corresponding untreated holotoxin as reference. Data from six independent experiments with six replicates per condition are presented in box-and-whisker format. Asterisks indicate statistically significant differences (**$p$ < 0.01; ****$p$ < 0.0001, one-way ANOVA). Reproduced from manuscript II**.**

The results obtained by ELISA were validated by a second assay. The loss of mass upon PDI treatment of the holotoxins immobilized on a $G_{M1}$-coated sensor was evaluated by SPR spectroscopy. Loss of mass was highest for CT and lowest for hLT (Figure 12A); for CTv1 it was recorded a mass loss lower than CT, while CTv4 had a value close to that of hLT (Figure 12B). This was further confirmed by addition of antibodies against the A1 subunit (Figure 12A-C), showing retention for CTv1, CTv4 and hLT as opposed to CT and of antibodies against the B subunit (Figure 12A-C), showing retention in all samples. An efficient disassembly provides a higher concentration of the toxic subunit in the cytoplasm and therefore a higher cAMP concentration, explaining the importance of this step in the toxicity of CT vs. LT. These results in combination with ELISA and toxicity measurements indicate the importance of PDI-driven disassembly of the A1 subunit in the potency of CT over LT. The PDI-disassembly efficiency of CTv1, CTv4 and hLT is lower in comparison to CT. A single substitution to an hLT/pLT residue, in CTv1, already shows similar rate of toxicity and disassembly to LT, confirmed with CTv4, which shows the same behavior as hLT. This validates the role of the substitutions in determining the toxicity of CT, and links them to the kinetics of the disassembly step.

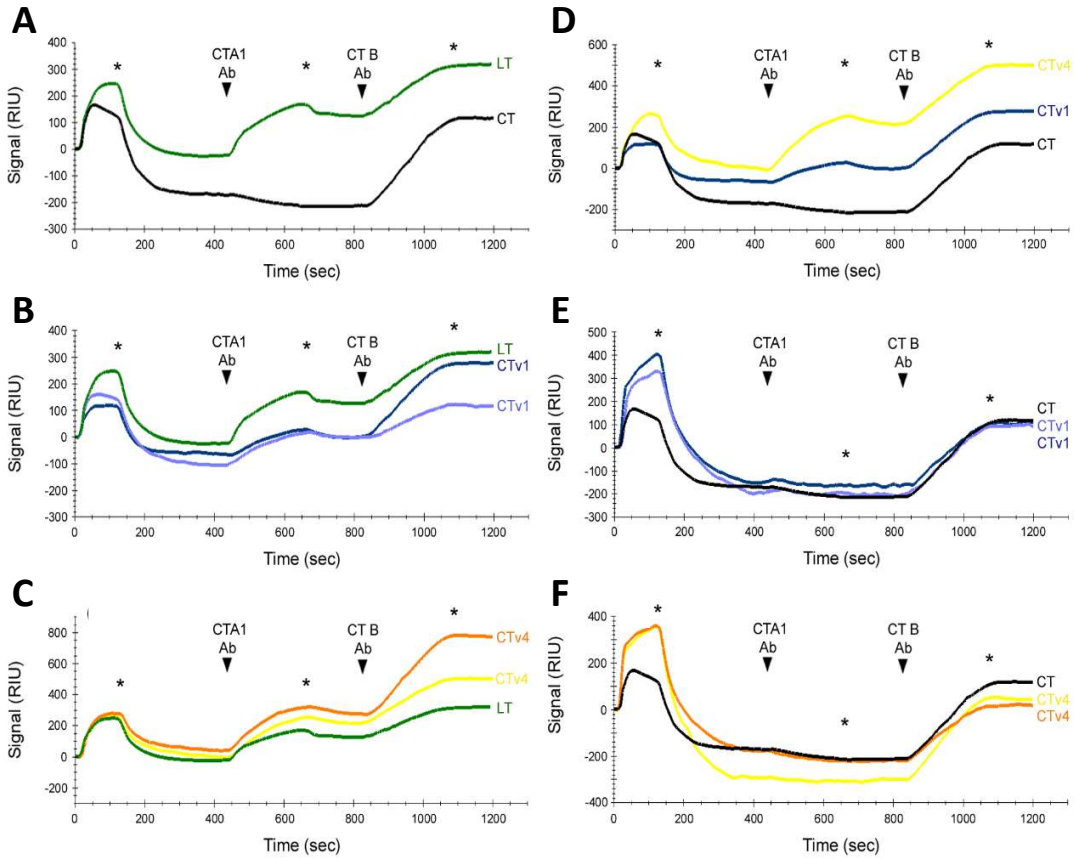


**Figure 12. Substitutions in the A2 subunit decreases the efficiency of toxin disassembly by PDI**. SPR spectroscopy of CT, CTv1, CTv4 and LT measurement investigation there PDI disassembly. After a baseline measurement showing the mass of the sensor-bound holotoxin (0 RU), injections of antibodies against the A1 and B proteins were performed (indicated by the arrowheads) to show potential disassembly (Asterisks show the removal of PDI or antibody from the perfusion buffer). Two independent runs for each variant were performed and are overlaid on the same graph. **A.** CT vs. hLT. **B.** hLT from panel A vs. CTv1. **C.** hLT from panel A vs. CTv4. **D.** CT from panel A with color-matched traces of CTv1 from panel B and CTv4 from panel C. **E.** CT from panel A vs. CTv1 sensors exposed to 10 µM PDI. **F.** CT from panel A vs. CTv4 sensors exposed to 10 µM PDI. Reproduced from manuscript II**.**

## 2.2. Structural studies of chimeric CT proteins

The X-ray crystallographic structures of CTv1 (PDB ID: 8Q6I) and CTv4 (PDB ID: 8OXS) were solved to 2.1 Å and 1.6 Å, respectively. The B pentamer is generally well-defined in both structures, while the C-termini of the A subunits are partially disordered (last built residue: Q227, CTv1 and I236, CTv4). Overall, the structures resemble more that of the wild-type CT (PDB ID: 1S5E[51]; r.m.s.d. CTv1 = 1.2 Å, r.m.s.d. CTv4 = 0.7 Å) than pLT (PDB ID: 1LTS[97]; r.m.s.d CTv1 =1.3 Å; r.m.s.d CTv4 = 1.1 Å). The main structural difference can be observed in the visible part of the A subunit C-terminus, from residue 224 onward, where CTv4 closely resembles pLTA2 with an elongated C-terminus, while CTv1 adopts a hybrid

conformation between that of CTA2 and pLTA2 (Figure 13). An analysis of the intersubunit angles reported values of 45° for CTv1 and 48° for CTv4, inverting the trend previously observed for the wild-type CT (49°) and pLT (40°)[102].

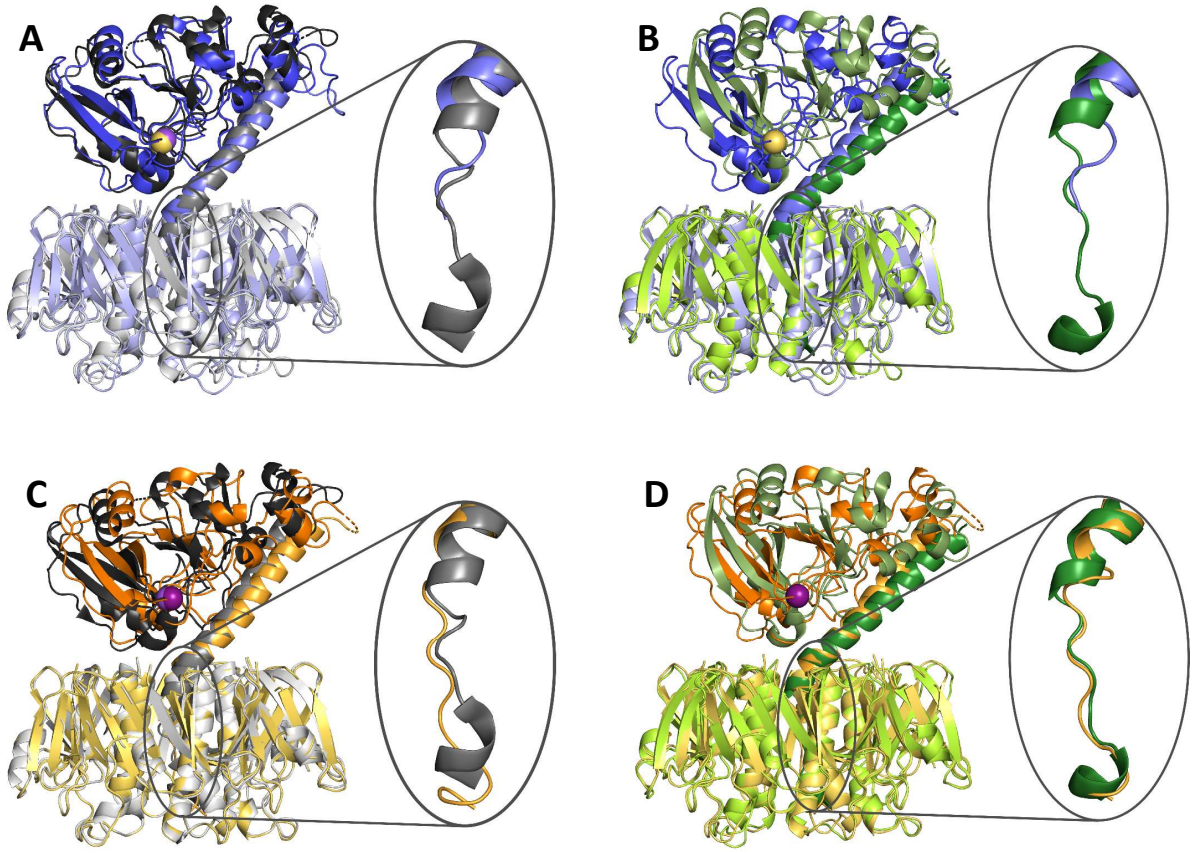


**Figure 13. Comparison of CTv4's and CTv1's overall structures and C-terminus regions to wild-type CT and LT. A.** CTv1 (blue, PDB ID: 8Q6I) vs wild-type CT (grey, based on 1S5E[51]) and zoom on the C-terminus. **B.** CTv1 (blue, PDB ID: 8Q6I) vs wild-type LT (green, based on 1LTS[97]) and zoom on the C-terminus. **C.** CTv4 (orange, PDB ID: 8OXS, we) vs wild-type CT (grey, based on 1S5E[51]) and zoom on the C-terminus. **D.** CTv4 (orange, PDB ID: 8OXS) vs wild-type LT (green, based on 1LTS[97]) and zoom on the C-terminus. Reproduced from manuscript II.

Normalized *B*-factor analysis shows that CT harbors overall a much higher mobility than pLT (Figure 14A-B). In particular, when looking at the mobility of the A2 linker, pLT shows a very rigid A subunit whereas CT shows greater flexibility (Figure 14A-B). In CTv1 and CTv4, the overall flexibility is lower than CT but higher than pLT (Figure 14). Normal mode analysis using the WEBnm@ server[125] on the different variants, wild-type CT and pLT shows that CT carries the highest conformational freedom of its A2 subunit (deformation energy of 479 kJ/mol), with a large displacement, whereas pLT is more rigid (205 kJ/mol). CTv1 (469 kJ/mol) and CTv4 (360 kJ/mol) have intermediate values and amplitude, which coincides with was what noticed with the normalized *B*-factors.

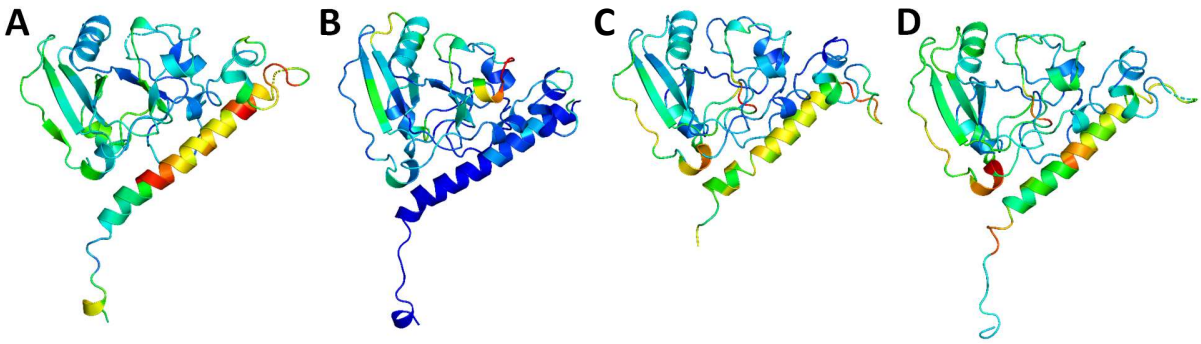


**Figure 14. Protein mobility of the toxin A subunits by normalized *B*-factor analysis. A.** Normalized *B*-factor of the *A* subunit of wild-type CT (based on 1S5E, chain B[126]). **B.** Normalized *B*-factor of the *A* subunit of LT (based on 1LTS[97]). **C.** Normalized *B*-factor of the *A* subunit CTv1 (PDB ID: 8Q6I; this work). **D.** Normalized *B*-factor of the *A* subunit of CTv4 (PDB ID: 8OXS, chain A; this work). Colored by normalized *B*-factors, from blue (low *B* factors, representing low mobility) to red (high *B* factors, representing high mobility or disorder). Reproduced from manuscript II.

As the pore environment is fully conserved between CT and LT, the conformation of the C-terminus can be traced to the amino acid substitutions and is likely to reflect a functional significance. Unfortunately, the presence of an iodide ion in the pore (close to S224 to Q227) of CTv1 prevents a reliable interpretation of the structural results. The values recorded for the intersubunit angles are inconsistent with the hypothesis formulated in Serrano *et al.*[102]. Intersubunit angles of CTv1 and CTv4 were expected to be close to those of CT (49°) and pLT (40°), respectively; however, the two variants showed the opposite trend (CTv1 angles of 45° and CTv4 angle of 48°), pointing to a fault in the initial hypothesis. Additionally, proteins are not rigid, but highly dynamic molecules, and their functional interpretation based solely on their crystal structures has clear limitations. Indeed, the normalized *B*-factors shows a decreased conformational freedom of A subunit shifting from CT to pLT, with the *B*-factors of CTv1 and CTv4 consistently falling between the two. The data were further confirmed by normal mode analysis. This suggests that the *dynamic* behavior introduced by the amino acid substitutions is what affects the PDI disassembly step, rather than static structural features such as the intersubunit angle.

## 3. Studies on the molecular mechanism of a defense AB toxin: CCTX2

The fungal kingdom includes various microorganisms both unicellular and multicellular. Although estimated between 2.2 and 3.8 million different species in some studies[127] and up to 12 million in others[128], only about 100,000 fungi have been characterized and described[128]. As a consequence, a lot of the molecular mechanisms underlying their biology are poorly described and understood.

Not even *Coprinopsis cinerea*, a model organism used to study fungal biology, has revealed all its secrets yet. As a defense strategy against predators, fungi use several toxic proteins[129]. As part of an effort to explore the largely unknown world of defense systems employed by fungi, the Künzler Lab challenged with fungivorous nematodes the fungus *C. cinerea*, monitoring its gene expression profile[130]. The screening highlighted the *cctx2* gene as upregulated upon exposure to the nematode *Aphelencus avenae*, and revealed that CCTX2 functions as a nematicidal protein[112]. However, its mechanism of action and its structure remain unknown[112].

In **manuscript III**, we describe CCTX2 and its paralogs, CCTX1 and CCTX3. We prove their role as nematotoxins and that they represent a new family of chimerolectins/AB toxins involved in host defense. We focused on CCTX2, the holotype for the family, determining its lectin character, the sugar-binding specificity and its uptake and intracellular trafficking. Furthermore, we solved the 3D structure of CCTX2 using single particle cryo-EM. This revealed a toxicity effector domain with a completely new fold, not found in any protein deposited in the PDB database. In addition to the results presented in **manuscript III**, we investigated possible co-factors of the protein using DSF and confirmed proteolytic cleavage of CCTX2 in presence of furin and kexin, potentially necessary to exert toxic activity**.**

## 3.1. The CCTX family of chimerolectins/toxins

A BLAST search run on the *C. cinerea* genome revealed two paralogues of the *cctx2* gene, renamed *cctx1* and *cctx3*. Expression of the three genes was probed by RNA-sequencing and quantitative reverse transcription polymerase chain reaction (qRT-PCR), and found to differ for the three genes with *cctx1* and *cctx3* constitutively expressed in stage 1 primordia and all three genes in vegetative mycelium. Interestingly enough, only *cctx2* expression is induced upon fungivorous nematode challenge[112]. The three protein products (CCTX1, CCTX2 and CCTX3), sharing >50% identity, were predicted by sequence analysis to carry a five-domain partition, with four N-terminal ricin B chain-like (RBCL) domain and a C-terminal domain of unknown function (Figure 17A). RBCL domains are often associated to a lectin function, which was confirmed for CCTX2 by glycan array scanning, with a sugar-binding preference for the LacDiNAc motif (Figure 15). Toxicity of the proteins against *Caenorhabditis elegans* was demonstrated and further testing on truncated CCTX2 variants (ΔNCCTX2, CCTX2ΔC) was showed to depend both on the presence of the first two N-terminal RBCL lectin domains and the C-terminal domain, suggesting it plays the role of a toxicity effector.

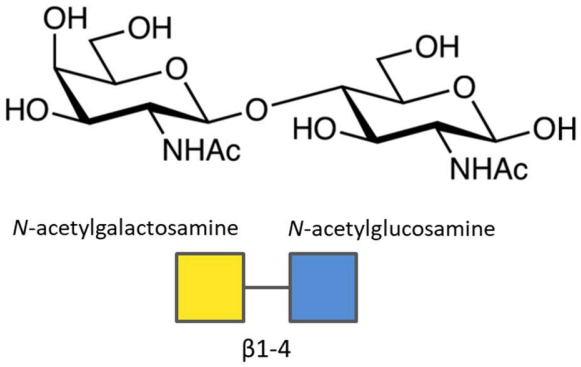


**Figure 15. LacdiNAc carbohydrate with the two moieties: *N*-acetylgalactosamine and *N*-acetylglucosamine.**

## 3.2. Uptake and retrograde trafficking of CCTX2

To investigate the potential target within the nematode, the phenotype of *C. elegans* was fed with CCTX2 and showed a collapsed intestine, suggesting that the intestinal epithelial cells are the target of intoxication (Figure 16). Because of CCTX2 affinity for LacdiNAc, it was hypothezised that the receptors on the intestinal cell surface are glycosphingolipids (GSLs) that carry the LacDiNAc epitope, such as the arthroseries GSL core. This was confirmed by TLC overlay assays on extracted GSLs from *C. elegans* membranes and the reduced sensitivity to the toxin of *C. elegans* strains deficient for the enzyme needed to their synthesis (*bre-3*, *bre-4*, *bre-5*).

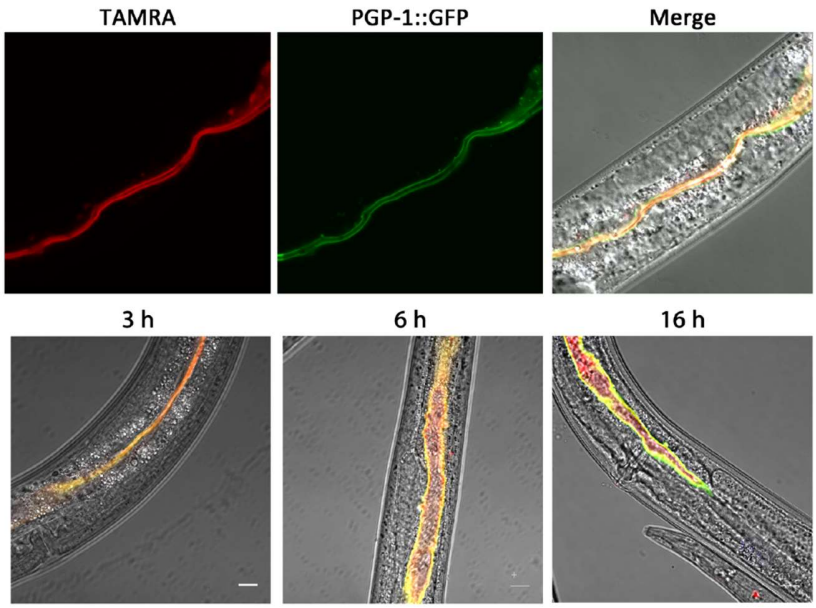


**Figure 16. Intoxication of *C. elegans* through its epithelial cells.** CCTX2-TAMRA was fed to *C. elegans* GK70 L4 larvae. It binds to the intestinal membrane of the larvae and co-localizes with the intestinal apical membrane marker PGP-1::GFP. After 3 h of feeding, it seems that the intestine is damaged. Images acquired at the rear of the animal. Scale bar = 10 µm. Reproduced from manuscript III.

The intracellular fate of CCTX2 was dissected through the use of a variety of *C. elegans* strains, each mutated for a specific component of the intracellular trafficking machinery. Upon receptor binding, CCTX2 is internalized, as shown by the co-localization of fluorescently-labeled CCTX2 (TAMRA-CCTX2) with GFP fuse PGP-1 apical plasma membrane marker in *C. elegans* strains (PGP-1::GFP). Internalization happens through endocytosis, as demonstrated by the localization of fluorescently-labeled CCTX2 (TAMRA-CCTX2) in *C. elegans* strains carrying endosomal markers fused to GFP (RAB-5::GFP, RAB-11::GFP, MANS::GFP).

The intracellular travel of CCTX2 continues by retrograde trafficking, as verified by *C. elegans* strains deficient in different components of the retrograde trafficking machinery. A reduced toxic effect in deletant strain for VPS-52 and VPS-54, subunits of the GARP (Golgi associated retrograde protein) complex and a heterozygous deletant for BEC-1, a protein recruiting essential components for cargo export into the Trans-Golgi Network (TGN), show the reliance of CCTX2 on retrograde trafficking to exert its nematicidal activity.

Finally, export to the ER happens through recognition of the C-terminal HSEL sequon, similar to the ER-retention signal HDEL/KDEL in yeast/mammalian cells. Its removal led to a drastic decrease in toxic effect, implying that ER translocation is an essential requirement for toxicity. A putative kexin/furin cleavage site was discovered on the C-terminal domain (K596-R597). Its mutation significantly decreases activity, suggesting the need for proteolytic processing for activation.

## 3.3. Structural studies of CCTX2

The structure of CCTX2, solved by single-particle cryo-EM to 3.2 Å resolution, included the full-length protein and revealed a compact architecture and the correct domain topology within it (Figure 17). The four RBCL domains confirm their predicted β-trefoil fold, and cluster into a rhomboid-shaped platform that supports the fifth domain. The C-terminal displays a completely new protein fold, not recorded in the PDB, as confirmed through a search with the DALI server[131].

The central portion of the C-terminal domain is formed by a seven-stranded, β-sheet supported by two α-helices (Figure 18). The solvent-exposed portion of the β-sheet forms a potential catalytic cleft. The β-strands are connected by long loops, two of them containing large insertions. The first insert (residues 571-618) forms a small two-helix subdomain, packed on the side of central α/β core; the second insert (residues 619-682) contains a long stretch of random coil that tightly interacts with the third RBCL domain. The C-terminal portion (746-787) is formed by a long random coil segment, circling the fourth RBCL domain, and ending with a helix and another random coil, sandwiched between the first insert and the central α/β core. The kexin cleavage-site is embedded in the second structural insert, on a solvent-

exposed α-helix, thus making it easily accessible to proteases. In contrast, the HSEL ER-retention sequence is buried within the bulk of the fifth domain, right behind the kexin-cleaved α-helix.

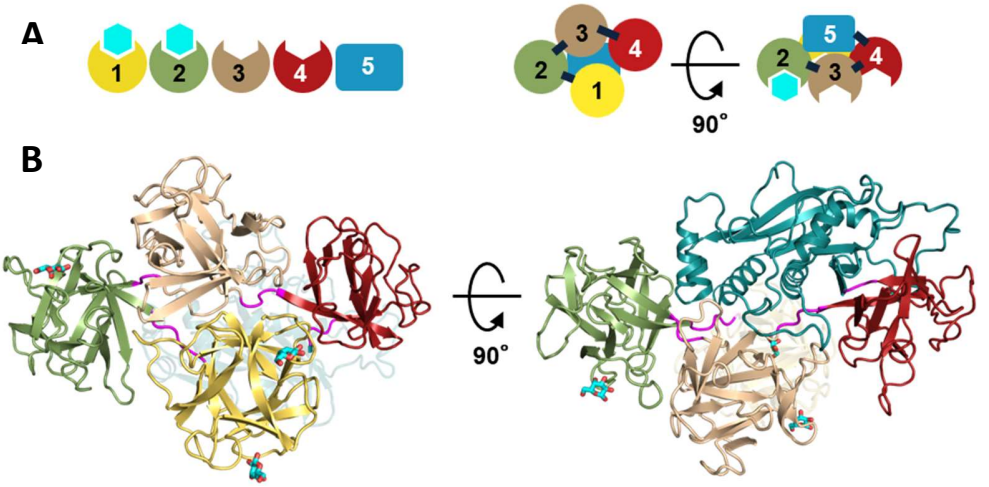


**Figure 17. Cryo-EM structure of CCTX2 at 3.2 Å resolution. A.** Schematic representation of the domains and the fold of CCTX2 with RBCL domain 1 in yellow, RBCL domain 2 in green, RBCL domain 3 in sand, RBCL domain 4 in dark red and the domain 5 in blue. The cyan hexagons represent the binding of carbohydrates to the lectin domains. **B.** Cartoon representation of the fold of CCTX2; coloring as explained in panel A. The linkers between domain 1-2 and domain 3-4 are colored in magenta. Reproduced from manuscript III.

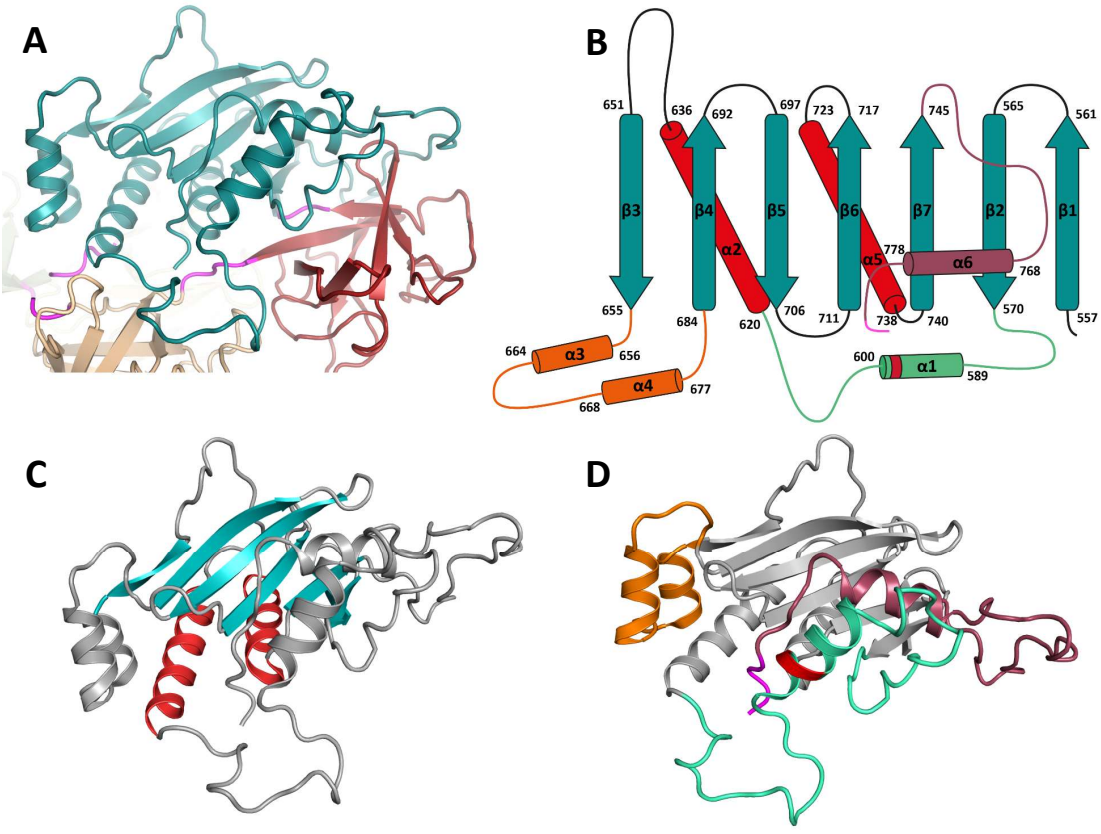

**Figure 18. New protein fold of the domain 5 of CCTX2. A.** C-terminal domain of CCTX2, from residue 556-787 (teal). **B.** Topology diagram of the domain, showing the central beta sheet colored in teal and the supporting alpha helices in red. The first insert is colored in light green, with the kexin/fiurin cleavage sequon (K596-R597) highlighted in red, the second insert in orange and the C-terminus in dark red. The C-terminal HSEL sequence is highlighted in magenta. **C-D.** Main secondary structure elements of the fifth domain colored according to the scheme described in the legend for panel B. Reproduced from manuscript III.

## 3.4. Function of CCTX2/cytotoxicity mechanism

Due to the lack of either sequence or structural homology to any other known protein, it is difficult to infer a function for the C-terminal domain of CCTX2. As a follow-up to **manuscript III**, we performed preliminary experiments to get further insights into the actual activity associated to the C-terminal domain of CCTX2. In the hope of identifying a potential cofactor or co-substrate, we evaluated by DSF the effect of several compounds on the thermal stability of CCTX2.

The melting curve of CCTX2 alone shows two distinct unfolding events, happening at two different melting temperatures (Figure 19A). Since the four N-terminal RBCL domains are very rich in tryptophan and tyrosine residues, we hypothesize that domains 1-4 unfold first, with a high intensity peak and the domain 5 unfolds last. Next, we tested three different compound libraries containing biologically relevant molecules, including cofactors, co-substrates as well as common protein ligands (Figure 19B). We observe that most sugars stabilize the protein, which is not very surprising considering the lectin properties of the two first domains. Most nucleotide-related molecules stabilize the protein with the exception of NAD and ATP, which strongly destabilizes CCTX2, with a decrease of 16 to 19°C in the melting temperature, respectively. In addition, CCTx2 is highly destabilized by divalent metallic cations with the exception of $Mg^{2+}$.

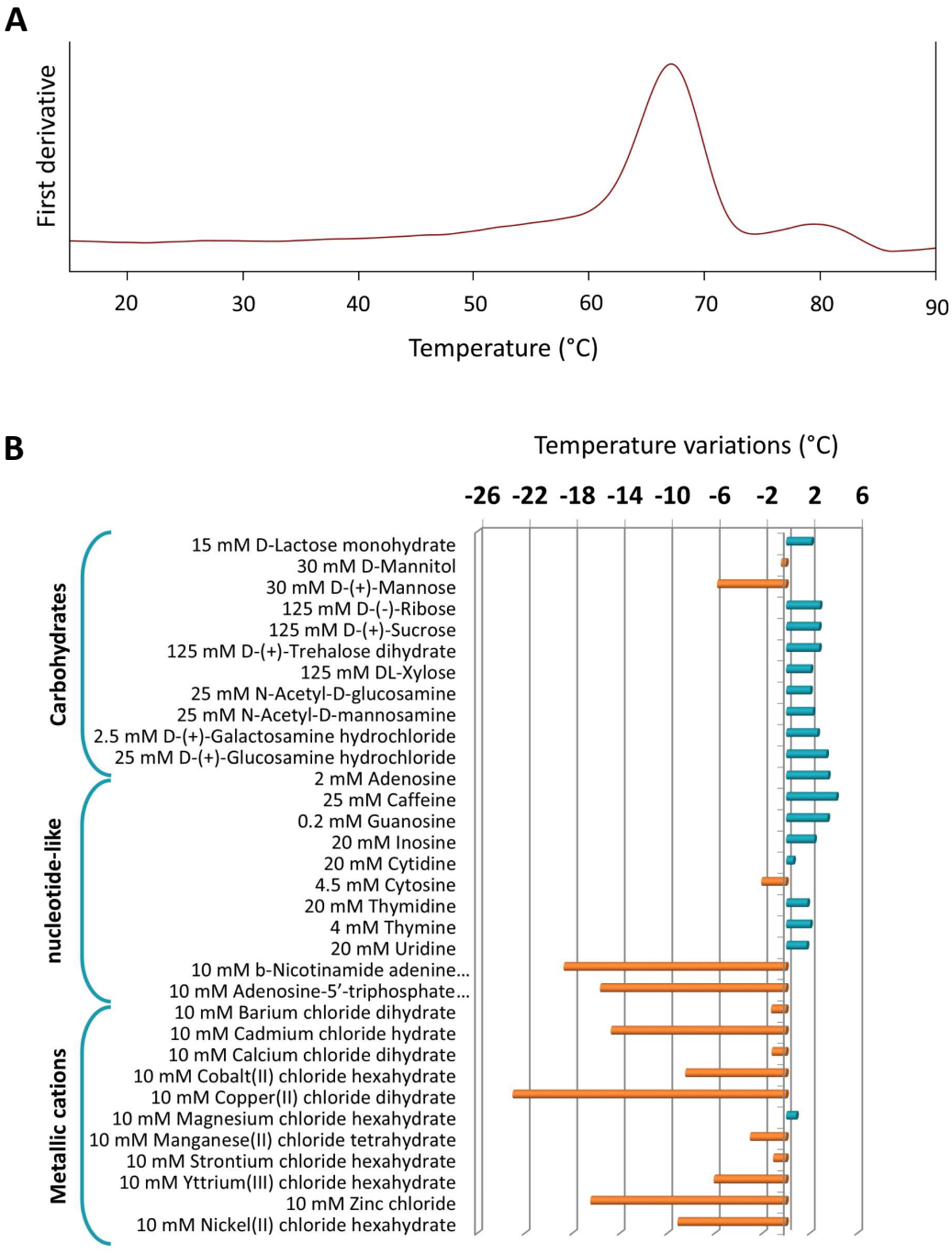


**Figure 19. Stability investigations of CCTX2. A.** DSF studies of CCTX2 based on its intrinsic flurorescence (dependent on its amount of tryptophan and tyrosine residues). CCTX2 shows two unfolding-event with potentially the four RBCL domains unfolding first followed by the fifth domain. **B.** Influence of potential co-factors and co-substrates on the stability of CCTX2 using DSF with SYPRO™ orange as a fluorescent probe showing stabilization in teal and destabilization in orange.

## 3.5. Final considerations

The CCTX paralogues are the representatives of a novel family of fungal proteins, with the nematicidal CCTX2 as its holotype. A search through BLAST revealed that orthologues of the three genes are present throughout the fungal kingdom. In **manuscript III,** we established that the three *cctx* genes have a different expression profile (all three expressed in the vegetative mycelium, but only *cctx1* and *cctx3* in both the mycelium and stage 1 primordia) and they all encode for proteins nematotoxic towards *C. elegans*. All three genes are predicted to carry a five-domain partition, with four N-terminal RBCL domains and a C-terminal domain of unknown function.

The cryo-EM structure of CCTX2, determined to 3.2 Å, confirmed the predicted structural features and it explained biological findings. The RBCL domains cluster into a compact platform that supports fifth domain. The first two N-terminal RBCL domains are often associated to a lectin function. Their lectin activity was confirmed by glycan array scanning, with a sugar-binding preference for the LacDiNAc motif. Consistently, the structure shows that the putative sugar binding sites of the RBCL domains are all clustered on the same solvent-exposed surface, and some residual density at subsites β and γ of domain 1 and α of domain 2 suggest the retention of residual galactose from the purification process. As the toxicity depends on the presence of both the first two N-terminal RBCL domains and the C-terminal domain, CCTX2 can be classified as a chimerolectin/$AB_4$ toxin, with the C-terminal as the toxicity effector domain.

Phenotype analysis on *C. elegans* exposed to CCTX2 showed a collapsed intestine, suggesting the intestinal epithelial cells as the physiological target of the toxin. The uptake/intracellular processing of CCTX2 was dissected through the use of biochemical assays and *C. elegans* mutants designed to lack specific components of the intracellular trafficking machinery, or have them fused to fluorescent reporters. The CCTX2 receptor was mainly identified as a GSL, most likely one from the arthroseries GSLs, which carry the LacDiNAc motif in their core. In strong analogy with other AB toxins (*e.g*. cholera toxin, Shiga toxins, ricin[11,13,41]), CCTX2 was found to be internalized by endocytosis and exploiting the retrograde trafficking machinery. ER is the farthest checkpoint on the CCTX2 travel within epithelial cells verified by our investigation. Translocation to the ER is mediated by a HSEL sequon on the C-terminus, reminiscing of the KDEL found in many bacterial toxins. It is an obligate passage for CCTX2 to reach its ultimate target, as removal of the HSEL sequon leads to a drop in toxicity. Interestingly enough, the cryo-EM structure shows that the HSEL sequon in CCTX2 is buried with the protein bulk, unlike the solvent-exposed KDEL found in *e.g*. CT[123]. The HSEL amino acids are shielded from solvent contacts by an α-helix carrying a kexin/furin cleavage site. Once cut by kexin (or furin), a protease found in the trans-Golgi network[132], the HSEL sequon would become solvent-exposed, allowing its recognition by the Erd2,

the ER protein-retrieval receptor and CCTX2 can be translocated to the ER and possibly undergo activation.

The ultimate fate of CCTX2 and the molecular principle underlying its toxic activity remain unresolved. The complete lack of sequence homology to known proteins families and the absolute novelty of the fold render it difficult to infer the function associated to the C-terminal domain. The overall molecular architecture of CCTX2 strongly reminds other $AB_4$ toxins, like the mosquitocidal holotoxins MTX from *Bacillus sphaericus*[133] or pierisin-1 from the butterfly *Pieris rapae*[134], with the difference that both toxins carry an effector N-terminal domain and four C-terminal RBCL domains. Both MTX and pierisin-1 carry an ADP-ribosylating tranferase activity associated to their N-terminal effector domain; however, their fold do not match the one observed for the CCTX2 C-terminal domain. The hunt for a cofactor/co-substrate by thermal shift assays highlighted a stabilizing effect in presence of nucleotide-like molecules and $Mg^{2+}$. The two resemble, respectively, a possible substrate and a common cofactor in enzymes that interact with nucleic acids[135]. This result hints at the hypothesis that the activity of the unknown domain involves the processing of nucleic acids. NAD and ATP and divalent metallic cations other than $Mg^{2+}$ strongly destabilizes CCTX2. Counter-intuitively, a decrease in the melting temperature can hint towards the discovery of the protein function[136]. The destabilization brought by ligand binding could reflect the transition to a catalytically active state, or even an actual processing of the ligand, treated as a substrate. Based on these results it is not yet possible to confidently determine the catalytic activity of CCTX2; however, they provide pointers for further investigation in the direction of a ribosyltransferase or nuclease activity.

As an alternative hypothesis, the tertiary structure of the C-terminal domain observed in the cryo-EM structure could differ from its final form. In analogy to CT, the kexin-cleaved CCTX2 might undergo PDI disassembly in the ER, followed by translocation into the cytoplasm and refolding by another chaperone, like e.g. the Hsp90-Hsc70 combination[78,79]. For this to happen in CT and other toxins, the presence of the Hsp90-recognized RPPDEI-like sequence is crucial[80]. In CCTX2, we observe a similar sequence at amino acid 667 to 672: RPHPDI that might have the same role. Preliminary dynamic studies using the WEBnm@ server[125] shows a high conformational flexibility of the domain 1 to 4, which could facilitate the release of domain 5 in the ER, after kexin/furin cleavage.

# V. Conclusion and future perspectives

The work presented in this thesis aimed at a better understanding of AB toxins and their role in attack or defense mechanisms. This was exemplified by studying cholera toxin from *Vibrio cholerae* (CT) and the *Coprinopsis cinerea* toxin 2 (CCTX2), representing an attack (CT) and a defense protein (CCTX2). CT is a toxin involved in cholera, an acute diarrheal disease. It has been shown to be more potent than the *E. coli* heat-labile enterotoxin (LT), its closest orthologue, despite a largely conserved sequence, and mechanism of action between the two. Despite being widely studied, some aspects for their mechanism of action are not fully understandood, preventing the eradication of the disease. Insights into the PDI-driven disassembly of CT, which appears to be a key-factor in cell intoxication, contribute to a better understanding of the disease. Knowledge gained from this work is beneficial to understand the action mechanism of other AB toxins, in which PDI often plays a crucial role (e.g. ricin or Shiga toxins). A better understanding of the molecular determinants underlying the difference in toxicity would provide a powerful tool for the design of better therapeutics. CCTX2 on the other hand, is a fungal defense AB toxin that mainly targets nematodes. Because of this specificity, elucidation of its molecular mechanism of action could lead to the development of new biotechnology tools with, applications as a human-safe nematicide. Nematodes are a threat to agriculture and are difficult to control or eradicate. The use of a lectin with specific nematicides properties would be a great asset for more productive cultures.

In **manuscript I**, we lay a solid foundation for the production of the proteins studied in this thesis, by presenting an alternative to the *E. coli* expression system: Vmax™ X2. This strain was shown to be adequate for the production of proteins destined to structural studies, in particularly challenging targets. Using the Vmax™ X2 platform we managed to produce large quantity of the CT variants used in **manuscript II** and CCTX2 used the follow-up experiments to **manuscript III**.

To investigate the difference of toxicity of attack AB toxins CT and LT, in **manuscript II** we structurally and functionally analyzed two CT/LT chimeras: CTv1 and CTv4, targeting the 229-232 C-terminal stretch of residues in the A2 subunit of CT. Using *in vitro* toxicity cell-based assays, we showed that both chimeras have toxicity-levels in the range of LT and lower than CT. The disassembly mechanism seems to be the deciding factor in the toxicity of CT *vs.* LT. It determines the amount of active toxin available in the cytosol, explaining why CT appears more toxic than LT. By ELISA and SPR-spectroscopy, PDI-driven disassembly was demonstrated to be a lot more effective for CT than for CTv1, CTv4 and LT. This reveals the importance of residues 229 to 232 in driving disassembly of CT vs. LT. To rationalize the functional data, we solved the X-ray crystallographic structures of both CTv1 and CTv4. The current paradigm describes the disassembly step of $AB_5$ holotoxins in terms of a mechanical action by PDI, which inserts itself like a wedge between the A subunit and the B pentamer (Figure 10C). In a previous article, it has

been speculated that a larger (CT) or smaller (LT) relative angle between the A2 subunit and the B pentamer, as observed in existing crystal structures, was determining the efficiency of PDI disassembly of the toxins. By substituting one or four amino acids to the LT sequence, we expected a linear variation of the intersubunit angle, with CTv1 being closer to CT, and CTv4 closer to LT. However, comparison of the angles measured in the crystallographic structures of CTv1 and CTv4 with the wild-type toxins (49° for CT and 40° for LT) showed results contradicting the starting thesis, with 45° for CTv1 and 48° for CTv4. This showed the limitations of our approach and dismissed the original hypothesis. Because proteins are rather wobbling molecules than static objects, we performed preliminary dynamic studies with WEBnm@ server[125] on the four toxins. This revealed that the A subunit of CT possesses a higher conformational freedom in comparison to LT. The deformation energy calculated from normal mode analysis showed values fitting the expected trend, with a flexibility closer to that of CT for CTv1 and LT for CTv4. This result is well in agreement with a normalized *B*-factor analysis carried out on the structures of the two toxins and the two chimeras. This points at the importance of flexibility, rather than a fixed angle, in the process of disassembly of the toxins by PDI. Our next goal is to investigate the flexibility of the A2 subunit of each of the toxins using molecular dynamics simulations. This study would allow us to confirm or disprove the preliminary data obtained with WEBnm@[123] and potentially provide further insights for the difference of toxicity between CT and LT. In the long run, understanding the importance of flexibility in the different toxicity of CT and LT will provide insightful information for better therapeutics. As it appears from our results, a more flexible protein implies a more toxic protein. The design of an allosteric inhibitor, rigidifying CT, would lead to a less effective PDI-driven disassembly and be the key to an effective treatment of cholera.

In **manuscript III**, we focused our efforts on the understanding of a defense AB toxins: CCTX2. We uncover two more paralogues of the *cctx2* gene, *cctx1* and *cctx3*, with a different expression profile throughout the fungus and its life cycle. All three proteins are active *C. elegans* nematotoxins, and share a five-domain partition with four N-terminal ricin B chain-like domains followed by a C-terminal domain of unknown function. A lectin activity associated to the first two N-terminal domains of CCTX2 was determined to be specific for the LacDiNAc glycotope. As the toxicity of CCTX2 depends on both the lectin and the C-terminal domain, CCTX2 can be described as an $AB_4$ toxin/chimerolectin. The cellular uptake and subsequent fate of CCTX2 was analyzed using several *C. elegans* mutant strains. The toxin primarily recognizes a glycosphingolipid receptor on the surface of *C. elegans* intestinal epithelial cells, possibly from the LacDiNAc-carrying arthroseries GSLs. CCTX2 internalization was determined to happen by endocytosis, and take the retrograde trafficking route, eventually being translocated to the ER. ER translocation is required for toxicity, and relies on both the presence of a C-terminal HSEL sequon,

reminiscent of the KDEL ER retention sequence found in cholera toxin, and the integrity of a kexin/furin cleavage site on the C-terminal domain.

The cryo-EM structure, solved to 3.2 Å, revealed the domain topology of CCTX2 and rationalized the biochemical and biocellular results. The RBCL domains are arranged in a rhomboid-shaped platform, providing a cradle for the C-terminal domain on one side, and exposing their putative binding sites to the solvent on the other. The C-terminal domain carries a completely novel fold, not found in other proteins deposited in the PDB. The HSEL sequon responsible for ER translocation folds into the bulk of the C-terminal/effector domain, and gets solvent-exposed only upon proteolytic cleavage by kexin. However, the lack of homology to any known protein family makes it difficult to gain insights into its activity. The domain partition found in CCTX2 is common to other $AB_4$ toxins, like MTX from *Bacillus sphaericus*[133] or pierisin-1 from the butterfly *Pieris rapae*[134,137] although with an inverted topology (N-terminal 'A' domain, four C-terminal 'B' domains). Similar to CT/LT, both toxins carry an ADP-ribosyltransferase (ART) activity associated to their effector domain; however, the effector domain of MTX and pierisin-1 shares negligible sequence identity and no structural similarity with the C-terminal domain of CCTX2.

The search for clues on the molecular mechanism of CCTX2 led to a screening for thermal stability changes in presence of several biochemically relevant compounds, including potential cofactors, co-substrates and common ligands. The results reported a stabilizing effect in presence of nucleotide-like molecules and $Mg^{2+}$, and a destabilizing effect from NAD, ATP and divalent metallic cations other than $Mg^{2+}$. Stabilizing factors seem to point toward a nucleic acid-processing enzyme; however, the destabilizing effect of NAD might reflect its role as co-substrate in ADB-ribosylation activity, thus leaving open the possibility of the discovery of a completely new ART fold. Finally, similar to CT and LT, the C-terminal domain of CCTX2 might undergo unfolding in the ER, followed by Hsp90 refolding upon translocation to the cytosol. The Hsp90-refolded protein might reveal its true self, and bear little similarity to the one observed in the cryo-EM structure, fitting into one of the many known and well-studied protein folds.

# VI. Materials and Methods

## 1. Figures

All figures were produced using Adobe Illustrator (Adobe Inc.), PyMOL (Schrödinger Inc.) and PowerPoint (Microsoft Corporation) or a combination of those.

## 2. Protein sequences and parameters

Parameters of all proteins were calculated using Expasy ProtParam tool[138] , assuming all cysteines being non-reduced.

### 2.1. Cholera toxin and cholera toxin variants

The sequences below show the full-length, mature proteins. The signal sequence targeting the protein to the periplasmic space is not included but can be found at the related UniProt entry.

The sequences of cholera toxin from *Vibrio cholerae* serotype O1 used in this work are from the 'classical' biotype. See UniProt entry: P01555 for the A subunit and UniProt entry: Q57193 for the B subunit for more information. Variants are based on this same wild-type cholera toxin, with an identical B subunit sequence (therefore not listed for the variants) and with one to four substitutions in the A subunit sequence.

*Wild-type cholera toxin (CT)*

**Subunit A:**

```
        10         20         30         40         50         60
NDDKLYRADS RPPDEIKQSG GLMPRGQSEY FDRGTQMNIN LYDHARGTQT GFVRHDDGYV
        70         80         90        100        110        120
STSISLRSAH LVGQTILSGH STYYIYVIAT APNMFNVNDV LGAYSPHPDE QEVSALGGIP
       130        140        150        160        170        180
YSQIYGWYRV HFGVLDEQLH RNRGYRDRYY SNLDIAPAAD GYGLAGFPPE HRAWREEPWI
       190        200        210        220        230        240
HHAPPGCGNA PRSSMSNTCD EKTQSLGVKF LDEYQSKVKR QIFSGYQSDI DTHNRIKDEL
```

Legend: A1 subunit, A2 subunit, ER retention sequence

**Subunit B:**

```
        10         20         30         40         50         60
TPQNITDLCA EYHNTQIHTL NDKIFSYTES LAGKREMAII TFKNGATFQV EVPGSQHIDS
        70         80         90        100
QKKAIERMKD TLRIAYLTEA KVEKLCVWNN KTPHAIAAIS MAN
```

*Cholera toxin variant 1 (CTv1)*

**Subunit A:**

```
        10         20         30         40         50         60
NDDKLYRADS RPPDEIKQSG GLMPRGQSEY FDRGTQMNIN LYDHARGTQT GFVRHDDGYV
        70         80         90        100        110        120
STSISLRSAH LVGQTILSGH STYYIYVIAT APNMFNVNDV LGAYSPHPDE QEVSALGGIP
       130        140        150        160        170        180
YSQIYGWYRV HFGVLDEQLH RNRGYRDRYY SNLDIAPAAD GYGLAGFPPE HRAWREEPWI
       190        200        210        220        230        240
HHAPPGCGNA PRSSMSNTCD EKTQSLGVKF LDEYQSKVKR QIFSGYQSEI DTHNRIKDEL
```

Legend: A1 subunit, A2 subunit, substitution to LT residue, ER retention sequence

*Cholera toxin variant 4 (CTv4)*

**Subunit A:**

```
        10         20         30         40         50         60
NDDKLYRADS RPPDEIKQSG GLMPRGQSEY FDRGTQMNIN LYDHARGTQT GFVRHDDGYV
        70         80         90        100        110        120
STSISLRSAH LVGQTILSGH STYYIYVIAT APNMFNVNDV LGAYSPHPDE QEVSALGGIP
       130        140        150        160        170        180
YSQIYGWYRV HFGVLDEQLH RNRGYRDRYY SNLDIAPAAD GYGLAGFPPE HRAWREEPWI
       190        200        210        220        230        240
HHAPPGCGNA PRSSMSNTCD EKTQSLGVKF LDEYQSKVKR QIFSGYQSEV DIYNRIKDEL
```

Legend: A1 subunit, A2 subunit, substitutions to LT residues, ER retention sequence

*Protein chemico-physical parameters*

The wild-type cholera holotoxin (CT) and its chimeras CTv1 and CTv4 include 1 copy of the A subunit and 5 copies of the B subunit presented beforehand. The chemico-physical parameters of each component and of each holotoxin can be found below.

| | | Molecular mass (kDa) | Theoretical pI | Extinction coefficient ($mol^{-1}.L.cm^{-1}$) |
|---|---|---|---|---|
| **B subunit** | **(Identical for each toxin)** | 11.6 | 7.75 | 10 095 |
| **Wild-type CT** | **A subunit** | 27.2 | 5.98 | 41 955 |
| | **Holotoxin** | 85.1 | 7.17 | 92 430 |
| **CTv1** | **A subunit** | 27.2 | 5.98 | 41 955 |
| | **Holotoxin** | 85.1 | 7.17 | 92 430 |
| **CTv4** | **A subunit** | 27.2 | 5.91 | 43 445 |
| | **Holotoxin** | 85.1 | 7.16 | 93 920 |

## 2.2. CCTX2

The sequences show the protein product from the pET22b(+)-8H-TEV-CCTX2, based on wild-type CCTX2. See UniProt entry: A8NDT7.

*8xHis-TEVp-CCTX2*

```
        10         20         30         40         50         60
HHHHHHHHEN LYFQSMALNE GVYWIRNSRF TNKVLDLDAA NVAKGTSILD FNEHGTFNEN
        70         80         90        100        110        120
HNQLWIVERF QSRDTYLIRS VHSNLVLDLS QGLSANGTPI LCWTQHGGTN QQWRIEWVKD
       130        140        150        160        170        180
DNKTPLYRIV SVATGTAISH NEDDSSAYTV AWSVDDGPKQ LWSFDPFVTP LLYRLRVKST
       190        200        210        220        230        240
SRVLDLAAAS ADNGTLALAW EQHTAITKRN QLWWLPYRSG AEEYTIQCLE TSTVADLSGG
       250        260        270        280        290        300
NSGNGTPIYG WQSHGGRNQQ WKFEPTSDSG DYYHIKNVEG GSVMDAYMND SQKRVGGWSN
       310        320        330        340        350        360
NGGDNQKWLL DPLPSPGPGW VLIQNGGTGK FLCSTPSGDI GTADGPETVY DYSVQWRFIQ
       370        380        390        400        410        420
REYTGVYHVV NRATGAYLRQ IGTSMPSIGL AEENDDELKD WWMLETYDNS EIGLASIISR
```

```
       430        440        450        460        470        480
WTGNVLDHYG GVSVQALDNN TENSYRSWAI IPARDWLTSF SLVNGQGGLC LAAQYAREET
       490        500        510        520        530        540
RLSTTANVND FHAQWVFRKP SGSSGYTIQN KYNNHYVGGT SARWELVVCC NKYFGIRNTS
       550        560        570        580        590        600
TQKYLAIEDG QVTFQDQDMT DRKQCWELCS GRATDTSGND YDLIYMDDDL LEVMIPWVGD
       610        620        630        640        650        660
KQGDLKHYIE KRATKKPPKD KGGWQLPAAG LIKKPKFNDI RQLLQELIEQ WEWDVVNEER
       670        680        690        700        710        720
EQIQTLVSID EAEARRLLGR RPHPDIVAAY QRSRSSTLFR IDRQGYFNIA GDRYVNIQGQ
       730        740        750        760        770        780
YGDDSYFHIA LPVGVRFGRE QIRRFLRDSL DRSTSVTITP TTCKPPSGGP DYNRDPDSDG
       790        800
DNSWIKWTIA VVGTSAIKHS EL
```

Legend: 8-histine tag (8H-Tag), TEV-cleavage site, domain 1, linker 1, domain 2, domain 3, linker 2, domain 4, domain 5.

*WT-CCTX2*

The sequences show the full length wild-type CCTX2 sequence. See UniProt entry: A8NDT7.

```
        10         20         30         40         50         60
MALNEGVYWI RNSRFTNKVL DLDAANVAKG TSILDFNEHG TFNENHNQLW IVERFQSRDT
        70         80         90        100        110        120
YLIRSVHSNL VLDLSQGLSA NGTPILCWTQ HGGTNQQWRI EWVKDDNKTP LYRIVSVATG
       130        140        150        160        170        180
TAISHNEDDS SAYTVAWSVD DGPKQLWSFD PFVTPLLYRL RVKSTSRVLD LAAASADNGT
       190        200        210        220        230        240
LALAWEQHTA ITKRNQLWWL PYRSGAEEYT IQCLETSTVA DLSGGNSGNG TPIYGWQSHG
       250        260        270        280        290        300
GRNQQWKFEP TSDSGDYYHI KNVEGGSVMD AYMNDSQKRV GGWSNNGGDN QKWLLDPLPS
       310        320        330        340        350        360
PGPGWVLIQN GGTGKFLCST PSGDIGTADG PETVYDYSVQ WRFIQREYTG VYHVVNRATG
       370        380        390        400        410        420
AYLRQIGTSM PSIGLAEEND DELKDWWMLE TYDNSEIGLA SIISRWTGNV LDHYGGVSVQ
       430        440        450        460        470        480
ALDNNTENSY RSWAIIPARD WLTSFSLVNG QGGLCLAAQY AREETRLSTT ANVNDFHAQW
```

```
       490        500        510        520        530        540
VFRKPSGSSG YTIQNKYNNH YVGGTSARWE LVVCCNKYFG IRNTSTQKYL AIEDGQVTFQ
       550        560        570        580        590        600
DQDMTDRKQC WELCSGRATD TSGNDYDLIY MDDDLLEVMI PWVGDKQGDL KHYIEKRATK
       610        620        630        640        650        660
KPPKDKGGWQ LPAAGLIKKP KFNDIRQLLQ ELIEQWEWDV VNEEREQIQT LVSIDEAEAR
       670        680        690        700        710        720
RLLGRRPHPD IVAAYQRSRS STLFRIDRQG YFNIAGDRYV NIQGQYGDDS YFHIALPVGV
       730        740        750        760        770        780
RFGREQIRRF LRDSLDRSTS VTITPTTCKP PSGGPDYNRD PDSDGDNSWI KWTIAVVGTS
    787
AIKHSEL
```

Legend: domain 1, linker 1, domain 2, domain 3, linker 2, domain 4, domain 5.

Protein parameters of both proteins can be found below.

| | **WT-CCTX2** | **8H-TEV-CCTX2** |
|---|---|---|
| **UniProt entry** | A8NDT7 | - |
| **Molecular mass (kDa)** | 88.8 | 90.8 |
| **Theoretical pI** | 5.40 | 5.65 |
| **Extinction coefficient ($mol^{-1}.L.cm^{-1}$)** | 216 160 | 217 650 |

# 3. Media recipes

## 3.1. 10x $V_2$ salts

All salts are dissolved in dd$H_2O$ and autoclaved separately from the media to prevent precipitation. This gives a 10x concentrated solution which needs to be diluted in the media before inoculation.

10X $V_2$ salts-solution is composed of 2.04 M NaCl, 42 mM KCl, 231.4 mM $MgCl_2$.

## 3.2. LB-$V_2$ salts

For 1L of LB-$V_2$ salts, 900 mL LB-media are prepared following standard protocols and autoclaved. Before inoculation of the media, 100 mL of sterile 10x $V_2$ salts-solution is added.

# 4. Sample production and purification

## 4.1. Cholera toxin variants

The pARCT5 vector encoding the ctxAB operon was kindly provided by Dr. Randall K. Holmes. It is under the control of an L-arabinose-inducible promoter and genes are fused to the signal sequence from *E. coli* LT-IIB. Site-directed mutagenesis (Q5® Site-Directed Mutagenesis Kit, NEB), following standard manufacturer's protocol, was performed to obtain CTv1 and CTv4 (sequences reported in the previous section).

DNA oligonucleotides used to generate the chimeras were designed using the NEBaseChanger online tool:

| | |
|---|---|
| **DIDTHtoEIDTH_CTv1_fwd** | **5'-CTATCAATCTGAAATTGATACACATAATAG-3'** |
| **DIDTHtoEIDTH_CTv1_rev** | **5'-CCTGAAAATATTTGTCTTTTAAC-3'** |
| **DIDTHtoEVDIY_CTv4_fwd** | **5'-TATATATAATAGAATTAAGGATGAATTATG-3'** |
| **DIDTHtoEVDIY_CTv4_rev** | **5'-TCCACTTCAGATTGATAGCCTGAAAATATT-3'** |

Vmax™ X2[122] chemically competent cells (TelesisBio, San Diego) were transformed with both constructs (pARCT5_CTv1 and pARCT5_CTv4) following the manufacturer's guidelines. Detailed protocol for transformation, protein production and purification can be found in **manuscript II**; both samples are handled using the same protocol**.** In brief, a preculture of LB-$V_2$ salts media (25 µg/mL chloramphenicol, CAM) was used to inoculate the main culture of LB-$V_2$ salts medium (12.5 µg/mL CAM). At $OD_{600}$ ≈ 0.8, expression was induced with 0.2% L-arabinose and incubated at 30˚C, for 20-22 hours.

After expression, the growth medium was centrifuged at 8,500 x g to harvest the cells, 0.22 µm filtered, loaded on an equilibrated D-galactose affinity chromatography (50 mM Na-phosphate, 200 mM NaCl, pH 7.4) and washed with 15 column volumes of the same buffer. The proteins were then eluted with 50 mM Na-phosphate, 200 mM NaCl, 500 mM D-galactose, pH 7.4). The samples were concentrated by ultrafiltration (4˚C, 4,000 x g). The final step of the purification was performed by size exclusion chromatography (SEC) on a Superdex 200 increase 10/300 GL column (GE Life Sciences). The samples were loaded on the column equilibrated with 50 mM Tris-HCl, 200 mM NaCl, pH 7.5. Fractions containing pure protein were used to set up crystallization experiments.

## 4.2. CCTX2 for structural studies

Full and detailed protocol for transformation, protein production and purification can be found in **manuscript III**. C41(DE3) cells transformation with pET22b(+)_8H-TEV-CCTX2 was performed following a standard protocol. A preculture was used to inoculate LB medium (100 µg/mL ampicillin, Amp) and

incubated at 37˚C. At $OD_{600}$ ≈ 0.6, the culture was cooled down in iced-water for 10 to 15 minutes and expression was induced with 0.5 mM IPTG for 18-20 hours at 20˚C.

Cells were harvested and resuspended in lysis buffer (50 mM Na-HEPES, 200 mM NaCl, 2 mM EDTA, 5 mM DTT); the final pH was set to pH 7.5, and the buffer was complemented with 1x c*O*mplete EDTA-free protease Inhibitor cocktail). After lysis using an ultrasound homogenizer, the lysate was clarified by centrifugation and the supernatant was recovered and directly used for purification.

Supernatant was loaded on an D-galactose affinity chromatography column equilibrated with loading buffer (20 mM Na-HEPES, 500 NaCl, pH 7.5) and washed with loading buffer. CCTX2 was eluted with the elution buffer (20 mM Na-HEPES, 500 NaCl, 1 M D-galactose, pH 7.5). Fractions containing protein were pooled and concentrated by ultrafiltration. A SEC column (Superdex 200 Increase 10/300 GL) was equilibrated with 50 mM Na-HEPES, 150 mM NaCl, 100 mM D-galactose) and loaded with sample from the ultrafiltration step. The content of eluted fractions was assessed by SDS-PAGE; fractions containing pure CCTX2 were poled and loaded onto a HiTrap desalting column, previously equilibrated with dialysis buffer (20 mM Na-HEPES, pH 7.5). The dialysed protein was kept at 4˚C until its use for structural studies.

### 4.3. CCTX2 for DSF studies

The protocol for protein production and purification is detailed in **manuscript I**. Briefly, Vmax™ X2 cells[122] transformed with pET24_CCTX2, containing the *cctx2* wild-type gene, were grown in LB-$V_2$ salts complemented with Kanamycin (Kan). Expression was induced at $OD_{600}$ ≈ 0.6 with 0.5 mM IPTG and the cells were grown at 30˚C, shaking for 18 h. Cells were harvested by centrifugation and pellets were resuspended in lysis buffer (50 mM Na-HEPES pH 7.5, 200 mM NaCl, 2 mM EDTA, 5 mM DTT, 1 x c*O*mplete EDTA-free protease inhibitor cocktail). After lysis with an ultrasound homogenizer, the lysates were centrifuged and the supernatants were recovered and 0.22 µm-filtered before purification. The samples were loaded onto D-galactose affinity column equilibrated with loading buffer (20 mM Na-HEPES, 500 mM NaCl, pH 7.5). CCTX2 was eluted with elution buffer (20 mM Na-HEPES, 500 mM NaCl, 1 M D-galactose, pH 7.5) and the fractions containing the protein were pooled and concentrated. Finally, CCTX2 was purified by SEC (Superdex 200 Increase 30/100 GL equilibrated with 50 mM Na-HEPES, 150 mM NaCl, 100 mM D-galactose, pH 7.5).

## 5. Protein crystallization of CT variants

### 5.1. CTv1

Detailed procedures for the crystallization experiments can found in **manuscript II.** Directly after size exclusion chromatography, a single fraction containing pure, monodisperse protein was used to set up

sitting-drop vapor-diffusion experiment. 0.3 µL of protein at 6.59 mg/mL was mixed with an equal amount of reservoir solution of the PACT premier™ screen and incubated at 20 °C. Over the course of two weeks, crystals were obtained in condition E3 (200 mM NaI, 20% w/v PEG 3350). They were harvested, mounted in loops, cryoprotected in mother liquor supplemented with 20% glycerol, and flash-cooled in liquid nitrogen.

### 5.2. CTv4

Similar to CTv1, after purification, the fractions containing pure protein were concentrated by ultrafiltration (4˚C, 4,000 x g). The protein was diluted to a concentration of 6.5 mg/mL by addition of galactose-containing buffer (50 mM Tris-HCl, 200 mM NaCl, 300 mM D-galactose, pH 7.5) to a final concentration of 50 mM galactose. A sitting-drop vapor-diffusion experiment was set up. 0.3 µL of the PACT premier™ screen (Molecular Dimensions Ltd.) was mixed with an equal amount of protein solution and incubated at 20 °C. Crystals grew over the course of three weeks under the condition of solution PACT-C8 (200 mM $NH_4Cl$, 100 mM Na-HEPES pH 7.0, 20% w/v PEG 6000). Crystals were harvested, mounted in loops, cryoprotected in crystallization solution directly supplemented with 20% glycerol and flash-cooled in liquid nitrogen.

## 6. X-ray data collection and refinement of CT variants

Detailed information for data collection and refinement can be found in **manuscript II**. Synchrotron data collection for CTv1 and CTv4 was performed at the ESRF (Grenoble, France), beamline ID23-2, at 100 K (0.87313 Å). Data for CTv1 was automatically processed at the beamline with *autoPROC*[139] and *STARANISO*[140] (Global Phasing LTd.). The crystal belonged to space group $P2_1\ 2_1\ 2_1$ and contained one CTv1 molecule in the asymmetric unit. Data for CTv4 was automatically processed at the beamline with *XDSAPP*[141]; the crystal belonged to space group $P2_1$, with two molecules of CTv4 per asymmetric subunit. Both structures were solved by molecular replacement, using Phaser[142] from the CCP4 software suite[143] and the CT holotoxins structure with PDB ID 1S5E[126] as search model. The final models were obtained after several cycles of manual building with *Coot*[144], followed by maximum-likelihood refinement with *REFMAC5*[145]. Model building was carried out by first rebuilding the main chain, then progressively adding water molecules, ligands and finally the alternative conformations. The occupancy was refined using phenix.refine, a tool from the Phenix software suite[146].

## 7. Single particle cryo-EM of CCTX2

Detailed information on cryo-EM grids preparation, data collection and processing can be found in **manuscript III**. In brief, grid preparation was performed at the cryo-EM Sweden National facility, SciLife lab (Umeå, Sweden) according to standard procedures. A 8H-TEV-CCTX2 sample at 1 mg/mL was applied

to the grids, and blotted/frozen with a Vitrobot Mark IV (Thermo Scientific). 7,500 movies were collected on a FEI Titan Krios (Thermo Scientific) operating at 300 kV, at a nominal magnification of 215,000 x and with a physical pixel size of 0.63 Å. The dose rate was set to 4.7 $e^-$/px/s, and the total exposure time was 5 s, resulting in a total dose of 59.5 $e^-/Å^2$. Nominal defocus range was -1.5 µm to -3.0 µm in 0.3 µm steps. The data set was processed using cryoSPARC 3 (Structura Biotechnology Inc.)[147] as described in **manuscript III**.

## 8. DSF experiments

DSF experiments on CCTX2 alone were performed on the Prometheus NT.48 (NanoTemper) in optical capillaries. Around 3 µL of CCTX2-WT sample at 0.1 mg/mL were loaded per capillary, with a triplicate for each condition. The samples are excited at an excitation wavelength of 280 nm and fluorescent signal is recorded at 330/350nm over a temperature ramp of 15°C to 90°C (1°C /minute increments).

DSF experiments on CCTX2 with various cofactors and cosubstrates were performed on a LightCycler 480 Real-Time PCR machine (Roche) in 96-well RT-PCR microplates, using a 25 µL sample volume and Sypro™ orange dye (Merck) as the fluorescent probe. Fluorescent signal was recorded over a temperature ramp of 20°C to 95°C (1.8°C/minute increments) at 580 nm and with an excitation wavelength of 465 nm. Data analysis was performed with the Microsoft Excel-based DSF analysis tool developed by Frank Niesen (Structural Genomics Consortium) and the melting temperature was calculated by non-linear fitting through a Blotzmann sigmoid (software GraphPad Prism, version 5, GraphPad Software Inc.). Determination of the optimal experimental conditions for the target were performed through preliminary screening, using different protein and fluorescent probe concentrations (protein: 0.05, 0.1 and 0.5 mg/mL; Sypro™, 1:5000, 1:1000 and 1:55 final dilution). Further experiments with CCTX2-WT (expressed in Vmax™ X2) were carried out using a sample concentration of 0.05 mg/mL and a 1:1000 end dilution of Sypro™ orange dye (pre-diluted in dimethyl sulfoxide (DMSO)). Screening was carried out using different additives and buffer conditions (See Table on the next page).

| | **Silver Bullets HR2-096 (Hampton Research)** | **GRAS HR2-451 (Hamptin Research)** | **HT HR2-138 (Hampton Research)** |
|---|---|---|---|
| **CCTX2-WT concentration in SEC buffer** | 0.05 mg/mL | 0.05 mg/mL | 0.05 mg/mL |
| **Sypro™ orange dilution** | 1:1000 | 1:1000 | 1:1000 |
| **End dilution of the screen** | 1:2 | 1:10 | 1:10 |

# VII. References


1. The top 10 causes of death. Vol. 2023 (WHO, 2020).
2. Bizzini, B. Tetanus toxin. *Microbiol Rev* **43**, 224-240 (1979).
3. Dong, M., Masuyer, G. & Stenmark, P. Botulinum and Tetanus neurotoxins. *Annu Rev Biochem* **88**, 811-837 (2019).
4. Collier, R.J. Diphtheria toxin: mode of action and structure. *Bacteriol Rev* **39**, 54-85 (1975).
5. Leka, O. et al. Diphtheria toxin conformational switching at acidic pH. *FEBS J* **281**, 2115-2122 (2014).
6. Vanden Broeck, D., Horvath, C. & De Wolf, M.J. *Vibrio cholerae*: cholera toxin. *Int J Biochem Cell Biol* **39**, 1771-1775 (2007).
7. Odumosu, O., Nicholas, D., Yano, H. & Langridge, W. AB toxins: a paradigm switch from deadly to desirable. *Toxins* **2**, 1612-1645 (2010).
8. Shang, C., Dang, L. & Van Damme, J.M. Plant AB toxins with lectin domains. in *Plant toxins* (ed. Springer) (2015).
9. Del Giudice, G. & Rappuoli, R. Genetically derived toxoids for use as vaccines and adjuvants. *Vaccine* **17**, 44-52 (1999).
10. Borodic, G.E., Acquadro, M. & Johnson, E.A. Botulinum toxin therapy for pain and inflammatory disorders: mechanisms and therapeutic effects. *Expert Opin Investig Drugs* **10**, 1531-1544 (2001).
11. Melton-Celsa, A.R. Shiga Toxin (Stx) classification, structure, and function. *Microbiol Spectr* **2**, EHEC-0024-2013 (2014).
12. Song, J. Bacterial AB toxins and host-microbe interactions. *Adv Microb Physiol* **81**, 67-109 (2022).
13. Polito, L., Bortolotti, M., Battelli, M.G., Calafato, G. & Bolognesi, A. Ricin: an ancient story for a timeless plant toxin. *Toxins (Basel)* **11**, 324 (2019).
14. Pickett, C.L. & Whitehouse, C.A. The cytolethal distending toxin family. *Trends Microbiol* **7**, 292-297 (1999).
15. Ali, M., Nelson, A.R., Lopez, A.L. & Sack, D.A. Updated global burden of cholera in endemic countries. *PLoS Negl Trop Dis* **9**, 0003832 (2015).
16. Kanungo, S., Azman, A.S., Ramamurthy, T., Deen, J. & Dutta, S. Cholera. *Lancet* **399**, 1429-1440 (2022).
17. Mandal, S., Mandal, M.D. & Pal, N.K. Cholera: a great global concern. *Asian Pac J Trop Med* **4**, 573-580 (2011).
18. Griffith, D.C., Kelly-Hope, L.A. & Miller, M.A. Review of reported cholera outbreaks worldwide, 1995-2005. *Am J Trop Med Hyg* **75**, 973-977 (2006).
19. Ahmed, S.H. & Nashwan, A.J. Cholera outbreak amid civil war: a public health crisis in Syria. *J Infect Public Health* **15**, 1484-1485 (2022).
20. Mavrouli, M., Mavroulis, S., Lekkas, E. & Tsakris, A. An emerging health crisis in turkey and syria after the earthquake disaster on 6 February 2023: risk factors, prevention and management of infectious diseases. *Healthcare (Basel)* **11**, 1022 (2023).
21. Chowdhury, F., Ross, A.G., Islam, M.T., McMillan, N.A.J. & Qadri, F. Diagnosis, management, and future control of cholera. *Clin Microbiol Rev* **35**, 0021121 (2022).
22. Pezzoli, L. & control, O.c.v.w.g.o.t.g.t.f.o.c. Global oral cholera vaccine use, 2013-2018. *Vaccine* **38**, 132-140 (2020).
23. Asadgol, Z., Mohammadi, H., Kermani, M., Badirzadeh, A. & Gholami, M. The effect of climate change on cholera disease: the road ahead using artificial neural network. *PLoS One* **14**, 0224813 (2019).

24. Holmner, A., Mackenzie, A. & Krengel, U. Molecular basis of cholera blood-group dependence and implications for a world characterized by climate change. *FEBS Lett* **584**, 2548-2555 (2010).
25. Cholera upsurge (2021-present). Vol. 2023 (WHO, 2023).
26. Lutz, C., Erken, M., Noorian, P., Sun, S. & McDougald, D. Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Front Microbiol* **4**, 375 (2013).
27. Waldor, M.K. & Mekalanos, J.J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910-1914 (1996).
28. Krebs, S.J. & Taylor, R.K. Protection and attachment of *Vibrio cholerae* mediated by the toxin-coregulated pilus in the infant mouse model. *J Bacteriol* **193**, 5260-5270 (2011).
29. Dutta, D. et al. *Vibrio cholerae* non-O1, non-O139 serogroups and cholera-like diarrhea, Kolkata, India. *Emerg Infect Dis* **19**, 464-467 (2013).
30. Parvin, I. et al. *Vibrio cholerae* O139 persists in Dhaka, Bangladesh since 1993. *PLoS Negl Trop Dis* **15**, 0009721 (2021).
31. Wang, J. et al. On the antigenic determinants of the lipopolysaccharides of *Vibrio cholerae* O:1, serotypes Ogawa and Inaba. *J Biol Chem* **273**, 2777-2783 (1998).
32. Siddique, A.K. & Cash, R. Cholera outbreaks in the classical biotype era. *Curr Top Microbiol Immunol* **379**, 1-16 (2014).
33. Hu, D. et al. Origins of the current seventh cholera pandemic. *Proc Natl Acad Sci U S A* **113**, 7730-7739 (2016).
34. Murley, Y.M., Behari, J., Griffin, R. & Calderwood, S.B. Classical and El Tor biotypes of *Vibrio cholerae* differ in timing of transcription of tcpPH during growth in inducing conditions. *Infect Immun* **68**, 3010-3014 (2000).
35. Baek, Y. et al. Cholera toxin production in *Vibrio cholerae* O1 El Tor biotype strains in single-phase culture. *Front Microbiol* **11**, 825 (2020).
36. Nair, G.B. et al. Cholera due to altered El Tor strains of *Vibrio cholerae* O1 in Bangladesh. *J Clin Microbiol* **44**, 4211-4213 (2006).
37. Qadri, F., Svennerholm, A.M., Faruque, A.S. & Sack, R.B. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clin Microbiol Rev* **18**, 465-483 (2005).
38. Locht, C. & Antoine, R. The history of pertussis toxin. *Toxins (Basel)* **13**, 623 (2021).
39. Beddoe, T., Paton, A.W., Le Nours, J., Rossjohn, J. & Paton, J.C. Structure, biological functions and applications of the AB5 toxins. *Trends Biochem Sci* **35**, 411-418 (2010).
40. Brown, P.I., Ojiakor, A., Chemello, A.J. & Fowler, C.C. The diverse landscape of AB5-type toxins. *Engineering Microbiology* **3**, 100104 (2023).
41. Sanchez, J. & Holmgren, J. Cholera toxin - a foe & a friend. *Indian J Med Res* **133**, 3153-3163 (2011).
42. Gill, D.M. The arrangement of subunits in cholera toxin. *Biochemistry* **15**, 1242-1248 (1976).
43. Hirst, T.R., Sanchez, J., Kaper, J.B., Hardy, S.J. & Holmgren, J. Mechanism of toxin secretion by *Vibrio cholerae* investigated in strains harboring plasmids that encode heat-labile enterotoxins of *Escherichia coli*. *Proc Natl Acad Sci U S A* **81**, 7752-7756 (1984).
44. Sanchez, J. & Holmgren, J. Cholera toxin structure, gene regulation and pathophysiological and immunological aspects. *Cell Mol Life Sci* **65**, 1347-1360 (2008).
45. Johnson, T.L., Abendroth, J., Hol, W.G. & Sandkvist, M. Type II secretion: from structure to function. *FEMS Microbiol Lett* **255**, 175-186 (2006).
46. Lencer, W.I. et al. Proteolytic activation of cholera toxin and *Escherichia coli* labile toxin by entry into host epithelial cells. Signal transduction by a protease-resistant toxin variant. *J Biol Chem* **272**, 15562-15568 (1997).

47. Booth, B.A., Boesman-Finkelstein, M. & Finkelstein, R.A. *Vibrio cholerae* hemagglutinin/protease nicks cholera enterotoxin. *Infect Immun* **45**, 558-560 (1984).
48. Benitez, J.A. & Silva, A.J. *Vibrio cholerae* hemagglutinin(HA)/protease: An extracellular metalloprotease with multiple pathogenic activities. *Toxicon* **115**, 55-62 (2016).
49. Naka, A. et al. Nicking sites in a subunit of cholera toxin and *Escherichia coli* heat-labile enterotoxin for *Vibrio cholerae* hemagglutinin/protease. *Toxicon* **36**, 1001-1005 (1998).
50. Zhang, R.G. et al. The three-dimensional crystal structure of cholera toxin. *J Mol Biol* **251**, 563-573 (1995).
51. O'Neal, C.J., Amaya, E.I., Jobling, M.G., Holmes, R.K. & Hol, W.G. Crystal structures of an intrinsically active cholera toxin mutant yield insight into the toxin activation mechanism. *Biochemistry* **43**, 3772-3782 (2004).
52. Branson, T.R. et al. A protein-based pentavalent inhibitor of the cholera toxin B-subunit. *Angew Chem Int Ed Engl* **53**, 8323-8327 (2014).
53. Murzin, A.G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J* **12**, 861-867 (1993).
54. Zhang, R.G. et al. The 2.4 A crystal structure of cholera toxin B subunit pentamer: choleragenoid. *J Mol Biol* **251**, 550-562 (1995).
55. Holmgren, J., Lonnroth, I., Mansson, J. & Svennerholm, L. Interaction of cholera toxin and membrane GM1 ganglioside of small intestine. *Proc Natl Acad Sci U S A* **72**, 2520-2524 (1975).
56. Merritt, E.A. et al. Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci* **3**, 166-175 (1994).
57. Overgaard, E.M. Boise State University (2022).
58. Aravind, L., Zhang, D., de Souza, R.F., Anand, S. & Iyer, L.M. The natural history of ADP-ribosyltransferases and the ADP-ribosylation system. *Curr Top Microbiol Immunol* **384**, 3-32 (2015).
59. Domenighini, M. & Rappuoli, R. Three conserved consensus sequences identify the NAD-binding site of ADP-ribosylating enzymes, expressed by eukaryotes, bacteria and T-even bacteriophages. *Mol Microbiol* **21**, 667-674 (1996).
60. O'Neal, C.J., Jobling, M.G., Holmes, R.K. & Hol, W.G. Structural basis for the activation of cholera toxin by human ARF6-GTP. *Science* **309**, 1093-1096 (2005).
61. Cassel, D. & Pfeuffer, T. Mechanism of cholera toxin action: covalent modification of the guanyl nucleotide-binding protein of the adenylate cyclase system. *Proc Natl Acad Sci U S A* **75**, 2669-2673 (1978).
62. Reidl, J. & Klose, K.E. *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol Rev* **26**, 125-139 (2002).
63. Almagro-Moreno, S., Pruss, K. & Taylor, R.K. Intestinal colonization dynamics of *Vibrio cholerae*. *PLoS Pathog* **11**, 1004787 (2015).
64. Faruque, S.M., Albert, M.J. & Mekalanos, J.J. Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol Mol Biol Rev* **62**, 1301-1314 (1998).
65. Heyningen, S.V. Cholera toxin: interaction of subunits with ganglioside GM1. *Science* **183**, 656-657 (1974).
66. Merritt, E.A., Sixma, T.K., Kalk, K.H., van Zanten, B.A. & Hol, W.G. Galactose-binding site in *Escherichia coli* heat-labile enterotoxin (LT) and cholera toxin (CT). *Mol Microbiol* **13**, 745-753 (1994).
67. Jobling, M.G., Yang, Z., Kam, W.R., Lencer, W.I. & Holmes, R.K. A single native ganglioside GM1-binding site is sufficient for cholera toxin to bind to cells and complete the intoxication pathway. *mBio* **3**, 00401-12 (2012).

68. Wernick, N.L., Chinnapen, D.J., Cho, J.A. & Lencer, W.I. Cholera toxin: an intracellular journey into the cytosol by way of the endoplasmic reticulum. *Toxins (Basel)* **2**, 310-325 (2010).
69. Majoul, I., Ferrari, D. & Soling, H.D. Reduction of protein disulfide bonds in an oxidizing environment. The disulfide bridge of cholera toxin A-subunit is reduced in the endoplasmic reticulum. *FEBS Lett* **401**, 104-108 (1997).
70. Lyles, M.M. & Gilbert, H.F. Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: pre-steady-state kinetics and the utilization of the oxidizing equivalents of the isomerase. *Biochemistry* **30**, 619-625 (1991).
71. Spooner, R.A. et al. Protein disulphide-isomerase reduces ricin to its A and B chains in the endoplasmic reticulum. *Biochem J* **383**, 285-293 (2004).
72. Schelhaas, M. et al. Simian Virus 40 depends on ER protein folding and quality control factors for entry into host cells. *Cell* **131**, 516-529 (2007).
73. Taylor, M., Banerjee, T., Ray, S., Tatulian, S.A. & Teter, K. Protein-disulfide isomerase displaces the cholera toxin A1 subunit from the holotoxin without unfolding the A1 subunit. *J Biol Chem* **286**, 22090-22100 (2011).
74. Pande, A.H. et al. Conformational instability of the cholera toxin A1 polypeptide. *J Mol Biol* **374**, 1114-1128 (2007).
75. Taylor, M. et al. Substrate-induced unfolding of protein disulfide isomerase displaces the cholera toxin A1 subunit from its holotoxin. *PLoS Pathog* **10**, 1003925 (2014).
76. Cherubin, P. et al. Protein disulfide isomerase does not act as an unfoldase in the disassembly of cholera toxin. *Biosci Rep* **38**, 20181320 (2018).
77. Lord, J.M., Roberts, L.M. & Lencer, W.I. Entry of protein toxins into mammalian cells by crossing the endoplasmic reticulum membrane: co-opting basic mechanisms of endoplasmic reticulum-associated degradation. *Curr Top Microbiol Immunol* **300**, 149-168 (2005).
78. Burress, H., Taylor, M., Banerjee, T., Tatulian, S.A. & Teter, K. Co- and post-translocation roles for HSP90 in cholera Intoxication. *J Biol Chem* **289**, 33644-33654 (2014).
79. Burress, H., Kellner, A., Guyette, J., Tatulian, S.A. & Teter, K. HSC70 and HSP90 chaperones perform complementary roles in translocation of the cholera toxin A1 subunit from the endoplasmic reticulum to the cytosol. *J Biol Chem* **294**, 12122-12131 (2019).
80. Kellner, A., Taylor, M., Banerjee, T., Britt, C.B.T. & Teter, K. A binding motif for Hsp90 in the A chains of ADP-ribosylating toxins that move from the endoplasmic reticulum to the cytosol. *Cell Microbiol* **21**, 13074 (2019).
81. Lazar, A.M. et al. G protein-regulated endocytic trafficking of adenylyl cyclase type 9. *Elife* **9**, 58039 (2020).
82. Sherwood, N.M., Krückl, S.L. & McRory, J.E. The origin and function of the pituitary adenylate cyclase-activating polypeptide (PACAP)/glucagon superfamily. *Endocr Rev* **21**, 619-670 (2000).
83. Kuo, I.Y. & Ehrlich, B.E. Signaling in muscle contraction. *Cold Spring Harb Perspect Biol* **7**, 006023 (2015).
84. Vitalis, E.A. et al. Role of the cAMP signaling pathway in the regulation of gonadotropin-releasing hormone secretion in GT1 cells. *Proc Natl Acad Sci U S A* **97**, 1861-1866 (2000).
85. Serezani, C.H., Ballinger, M.N., Aronoff, D.M. & Peters-Golden, M. Cyclic AMP: master regulator of innate immune cell function. *Am J Respir Cell Mol Biol* **39**, 127-132 (2008).
86. Sheppard, D.N. & Welsh, M.J. Structure and function of the CFTR chloride channel. *Physiol Rev* **79**, 23-45 (1999).
87. Koschinski, A. & Zaccolo, M. Activation of PKA in cell requires higher concentration of cAMP than *in vitro*: implications for compartmentalization of cAMP signalling. *Sci Rep* **7**, 14090 (2017).
88. Cheng, S.H. et al. Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. *Cell* **66**, 1027-1036 (1991).

89. Thomas, A., Ramananda, Y., Mun, K., Naren, A.P. & Arora, K. AC6 is the major adenylate cyclase forming a diarrheagenic protein complex with cystic fibrosis transmembrane conductance regulator in cholera. *J Biol Chem* **293**, 12949-12959 (2018).
90. Thiagarajah, J.R. & Verkman, A.S. CFTR pharmacology and its role in intestinal fluid secretion. *Curr Opin Pharmacol* **3**, 594-599 (2003).
91. Dubreuil, J.D., Isaacson, R.E. & Schifferli, D.M. Animal enterotoxigenic *Escherichia coli*. *EcoSal Plus* **7**, 0006-2016 (2016).
92. Sanchez, J. & Holmgren, J. Virulence factors, pathogenesis and vaccine protection in cholera and ETEC diarrhea. *Curr Opin Immunol* **17**, 388-398 (2005).
93. Wang, H., Zhong, Z., Luo, Y., Cox, E. & Devriendt, B. Heat-stable enterotoxins of enterotoxigenic *Escherichia coli* and their impact on host immunity. *Toxins (Basel)* **11**, 11010024 (2019).
94. van den Akker, F. et al. Crystal structure of heat-labile enterotoxin from *Escherichia coli* with increased thermostability introduced by an engineered disulfide bond in the A subunit. *Protein Sci* **6**, 2644-2649 (1997).
95. Joffre, E. & Sjoling, A. The LT1 and LT2 variants of the enterotoxigenic *Escherichia coli* (ETEC) heat-labile toxin (LT) are associated with major ETEC lineages. *Gut Microbes* **7**, 75-81 (2016).
96. Sixma, T.K. et al. Crystal structure of a cholera toxin-related heat-labile enterotoxin from *E. coli*. *Nature* **351**, 371-377 (1991).
97. Sixma, T.K. et al. Refined structure of *Escherichia coli* heat-labile enterotoxin, a close relative of cholera toxin. *J Mol Biol* **230**, 890-918 (1993).
98. Tauschek, M., Gorrell, R.J., Strugnell, R.A. & Robins-Browne, R.M. Identification of a protein secretory pathway for the secretion of heat-labile enterotoxin by an enterotoxigenic strain of *Escherichia coli*. *Proc Natl Acad Sci U S A* **99**, 7066-7071 (2002).
99. Horstman, A.L. & Kühn, M.J. Enterotoxigenic *Escherichia coli* secretes active heat-labile enterotoxin via outer membrane vesicles. *J Biol Chem* **275**, 12489-12496 (2000).
100. Lencer, W.I., Hirst, T.R. & Holmes, R.K. Membrane traffic and the cellular uptake of cholera toxin. *Biochim Biophys Acta* **1450**, 177-190 (1999).
101. Rodighiero, C. et al. Structural basis for the differential toxicity of cholera toxin and *Escherichia coli* heat-labile enterotoxin. Construction of hybrid toxins identifies the A2-domain as the determinant of differential toxicity. *J Biol Chem* **274**, 3962-3969 (1999).
102. Serrano, A. et al. Holotoxin disassembly by protein disulfide isomerase is less efficient for *Escherichia coli* heat-labile enterotoxin than cholera toxin. *Sci Rep* **12**, 34 (2022).
103. Duarte, R.C., Flores, A.A.V. & Stevens, M. Camouflage through colour change: mechanisms, adaptive value and ecological significance. *Philos Trans R Soc Lond B Biol Sci* **372**, 20160342 (2017).
104. Blackmore, N.J. et al. Three sites and you are out: ternary synergistic allostery controls aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*. *J Mol Biol* **425**, 1582-1592 (2013).
105. Peumans, W.J. & Van Damme, E.J. Lectins as plant defense proteins. *Plant Physiol* **109**, 347-352 (1995).
106. Cruz-Chu, E.R., Malafeev, A., Pajarskas, T., Pivkin, I.V. & Koumoutsakos, P. Structure and response to flow of the glycocalyx layer. *Biophys J* **106**, 232-243 (2014).
107. Peumans, W.J., Van Damme, E.J., Barre, A. & Rouge, P. Classification of plant lectins in families of structurally and evolutionary related proteins. *Adv Exp Med Biol* **491**, 27-54 (2001).
108. Mishra, A. et al. Structure-function and application of plant lectins in disease biology and immunity. *Food Chem Toxicol* **134**, 110827 (2019).
109. Tayyrov, A., Schmieder, S.S., Bleuler-Martinez, S., Plaza, D.F. & Künzler, M. Toxicity of potential fungal defense proteins towards the fungivorous nematodes *Aphelenchus avenae* and *Bursaphelenchus okinawaensis*. *Appl Environ Microbiol* **84**, 02051-18 (2018).

110. Yang, N. et al. Structural basis for the tumor cell apoptosis-inducing activity of an antitumor lectin from the edible mushroom *Agrocybe aegerita*. *J Mol Biol* **387**, 694-705 (2009).
111. Manna, D., Cordara, G. & Krengel, U. Crystal structure of MOA in complex with a peptide fragment: a protease caught in flagranti. *Curr Res Struct Biol* **2**, 56-67 (2020).
112. Plaza, D.F., Schmieder, S.S., Lipzen, A., Lindquist, E. & Künzler, M. Identification of a novel nematotoxic protein by challenging the model mushroom *Coprinopsis cinerea* with a fungivorous nematode. *G3 (Bethesda)* **6**, 87-98 (2015).
113. Singh, S., Singh, B. & Singh, A.P. Nematodes: a threat to sustainability of agriculture. *Procedia Environmental Sciences* **29**, 215-216 (2015).
114. Schmieder, S.S. Doctoral Thesis, ETH Zurich (2015).
115. Desaeger, J., Wram, C. & Zasada, I. New reduced-risk agricultural nematicides - rationale and review. *J Nematol* **52**, 1-16 (2020).
116. Cummings, R.D., Schnaar, R.L. & Ozeki, Y. R-Type Lectins. in *Essentials of Glycobiology* (eds. Varki, A. et al.) 419-430 (Cold Spring Harbor (NY), 2022).
117. Zurga, S., Pohleven, J., Kos, J. & Sabotic, J. Beta-trefoil structure enables interactions between lectins and protease inhibitors that regulate their biological functions. *J Biochem* **158**, 83-90 (2015).
118. Jaruszewicz-Blonska, J. & Lipniacki, T. Genetic toggle switch controlled by bacterial growth rate. *BMC Syst Biol* **11**, 117 (2017).
119. Saida, F., Uzan, M., Odaert, B. & Bontems, F. Expression of highly toxic genes in *E. coli*: special strategies and genetic tools. *Curr Protein Pept Sci* **7**, 47-56 (2006).
120. Geron, M. Production and purification of recombinant toxins. *Methods Mol Biol* **2068**, 73-84 (2020).
121. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* **37**, 203-206 (2004).
122. Weinstock, M.T., Hesek, E.D., Wilson, C.M. & Gibson, D.G. *Vibrio natriegens* as a fast-growing host for molecular biology. *Nat Methods* **13**, 849-851 (2016).
123. Lencer, W.I. et al. Targeting of cholera toxin and *Escherichia coli* heat labile toxin in polarized epithelia: role of COOH-terminal KDEL. *J Cell Biol* **131**, 951-962 (1995).
124. Tsai, B., Rodighiero, C., Lencer, W.I. & Rapoport, T.A. Protein disulfide isomerase acts as a redox-dependent chaperone to unfold cholera toxin. *Cell* **104**, 937-948 (2001).
125. Tiwari, S.P. et al. WEBnm@ v2.0: Web server and services for comparing protein flexibility. *BMC Bioinformatics* **15**, 425 (2014).
126. O'Neal, C.J., Amaya, E.I., Jobling, M.G., Holmes, R.K. & Hol, W.G.J. Crystal structures of an intrinsically active cholera toxin mutant yield insight into the toxin activation mechanism. *Biochemistry* **43**, 3772-3782 (2004).
127. Hawksworth, D.L. & Lucking, R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr* **5**, 0052-2016 (2017).
128. Wu, B. et al. Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi. *Mycology* **10**, 127-140 (2019).
129. Künzler, M. How fungi defend themselves against microbial competitors and animal predators. *PLoS Pathog* **14**, 1007184 (2018).
130. Plaza, D.F., Lin, C.W., van der Velden, N.S., Aebi, M. & Künzler, M. Comparative transcriptomics of the model mushroom *Coprinopsis cinerea* reveals tissue-specific armories and a conserved circuitry for sexual development. *BMC Genomics* **15**, 492 (2014).
131. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res* **50**, 210-215 (2022).

132. Vey, M., Schafer, W., Berghofer, S., Klenk, H.D. & Garten, W. Maturation of the trans-Golgi network protease furin: compartmentalization of propeptide removal, substrate cleavage, and COOH-terminal truncation. *J Cell Biol* **127**, 1829-1842 (1994).
133. Treiber, N., Reinert, D.J., Carpusca, I., Aktories, K. & Schulz, G.E. Structure and mode of action of a mosquitocidal holotoxin. *J Mol Biol* **381**, 150-159 (2008).
134. Watanabe, M. et al. Enzymatic properties of pierisin-1 and its N-terminal domain, a guanine-specific ADP-ribosyltransferase from the cabbage butterfly. *J Biochem* **135**, 471-477 (2004).
135. Hartwig, A. Role of magnesium in genomic stability. *Mutat Res* **475**, 113-121 (2001).
136. Carver, T.E. et al. Decrypting the biochemical function of an essential gene from *Streptococcus pneumoniae* using ThermoFluor technology. *J Biol Chem* **280**, 11704-11712 (2005).
137. Carpusca, I., Jank, T. & Aktories, K. *Bacillus sphaericus* mosquitocidal toxin (MTX) and pierisin: the enigmatic offspring from the family of ADP-ribosyltransferases. *Mol Microbiol* **62**, 621-630 (2006).
138. Gasteiger, E.H., C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. in *The Proteomics Protocols Handbook*. (ed. Handbooks, S.P.) 571-608 (2005).
139. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302 (2011).
140. Tickle, I.J., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. STARANISO server. http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi (2016).
141. Krug, M., Weiss, M.S., Heinemann, U. & Müller, U. XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *J Appl Crystallogr* **45**, 568-572 (2012).
142. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
143. Agirre, J. et al. The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr D Struct Biol* **79**, 449-461 (2023).
144. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
145. Murshudov, G.N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355-367 (2011).
146. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
147. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).

# VIII. Manuscripts