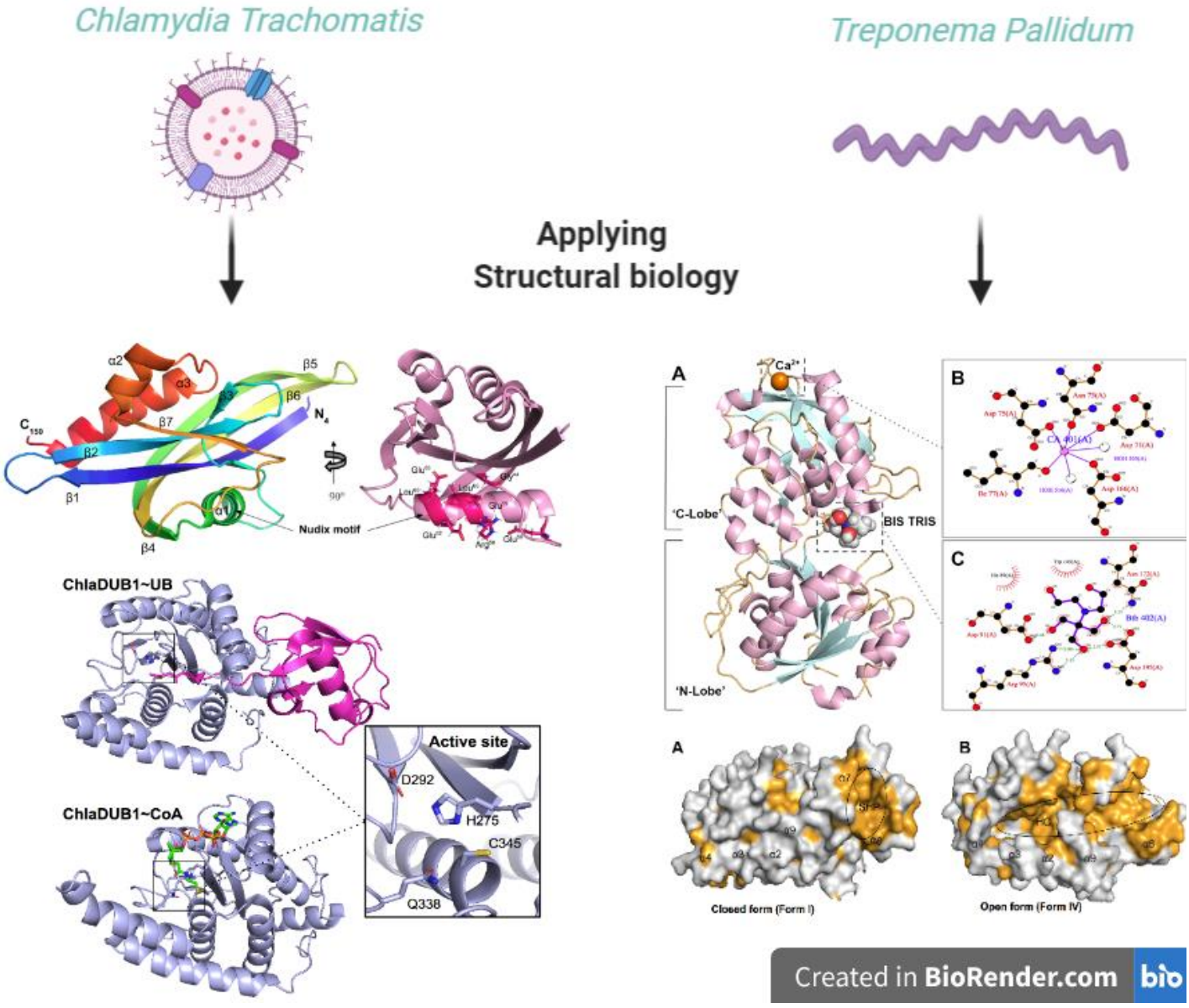
Structural information of protein targets of *Chlamydia trachomatis* and *Treponema pallidum*

Provides insight into the disease mechanism and development of possible treatments for Chlamydia and Syphilis.



Name: C.W.M. Simons
Student number: S3342646
Email: c.w.m.simons@student.rug.nl
Date of submission: November 6, 2020
Mentor: prof. dr. M.R. Groves
Department: Drug Design
Location: University of Groningen

Table of Contents

Abstract 3

Key words 3

Abbreviations 4

Introduction **5**

Chlamydia trachomatis 5

Lifecycle of *C. trachomatis* 5

Proteins and antigens of C. Trachomatis 5

Syphilis....6

Treponema pallidum 6

Invasion 7

Protein targets of *Chlamydia trachomatis* 8

CT263....8

ChlaDUB1....9

Pgp3....9

DsbH 10

CT584....11

IncA....12

CT771....13

CT441....14

CPAF 15

Protein targets of *Treponema pallidum* 16

Tp0453....16

Tp0624....17

Tp0684....18

Tp0971....19

Tp0453....21

Tp0655....22

Tp0751....23

Tp0737....25

Summary....26

Literature 27

Abstract

Chlamydia and Syphilis are commonly sexually transmitted diseases (STDs), caused by gram-negative bacteria *Chlamydia trachomatis* and *Treponema pallidum*, respectively. Both chlamydia and syphilis increase the risk of transmission and acquisition of HIV [1][5]. While chlamydia is responsible for 1.3 million cases of blindness worldwide [3], syphilis is the most prevalent infection associated with stillbirth or fetal loss in low-income populations [5]. As a result, these STDs cause significant morbidity and socioeconomic burden and are an important public health concern.

Here, the atomic structure of potential protein targets of *C. trachomatis* and *T. pallidum* are presented, which provide insight into the disease mechanism and host interactions by using a variety of imaging techniques. The three-dimensional structure of proteins will show how molecules are assembled, how they interact, and how they function. The crystal structures of different proteins are created using PyMOL (DeLano Scientific) and are inter alia visualized in a cartoon format. Besides, the electrostatic surface presentation shows potential ligand-binding sites of the protein. The potential function of different proteins is investigated by using the Protein Data Bank (PDB) code to detect structural homologs. Also, ligand-protein interactions are analyzed by 2D-diagrams, which are generated by LigPlot. The structural information provides possibilities in the discovery of potential protein targets to combat a chlamydia or syphilis infection. Overall, these data elucidate how structural information has led to a clearer understanding of disease mechanisms, which can lead to the development of treatments for chlamydia and syphilis.

Key words

Chlamydia, *Chlamydia trachomatis*, Syphilis, *Treponema pallidum*, STD, Structural Biology, Structural Homologs, PDB, PyMOL, LigPlot.

Abbreviations

ABC – ATP-binding cassette
AH - amphipathic helices
AcT – acetyltransferase
AMP – Adenosine Monophosphate
Ap_4A - Asymmetric Diadenosine 5',5''-P1,P4-Tetraphosphate
ATP – Adenosine Triphosphate
CAP - Catabolite Activator Protein
CoA - Coenzyme A
COMC – Complex of Outer Membrane Proteins
CPAF – Chlamydial Protease-like Activity Factor
CTD – C-terminal Domain
Dsb – Disulfide bond
DUB - Deubiquitinase
EB – Elementary Body
ECF – Energy Coupling Factor
ECM – Extra Cellular Matrix
HC – Hydrophobic Cavity
HSP – Heat Shock Protein
LBP – Ligand Binding Protein
LPS – Lipopolysaccharide
MES - 2-(N-Morpholino)Ethane-Sulfonic Acid
MOMP – Major Outer Membrane Protein
MTAN - 5-methylthioadenosine Nucleosidase
NTD – N-terminal Domain
OmpA - Outer-Membrane Protein A
OOP - OmpA-OmpF Porin
PDB – Protein Data Bank
PGP – Plasmid Gene Product
PMP – Polymorphic Membrane Proteins
RB – Reticulate body
SCR – Short Conserved Region
SHP - Shaped Hydrophobic Patch
SNARE - Soluble N-Ethylmaleimide-Sensitive Factor Attachment Receptors
STD – Sexually Transmitted Disease
TNF – Tumor Necrosis Factor
Tsp - Tail-Specific Proteases
TTSS – Type Three Secretion System

Introduction

Chlamydia trachomatis

Chlamydia is the most commonly sexually transmitted infection in Europe and the United States. In 2018, 1.758.1668 cases were reported in the United States to the Centers for Disease Control and Prevention [1]. However, since 50-85% of women infected *with C. trachomatis* are asymptomatic, not all cases are reported. Women with untreated *C. trachomatis* cervicitis can be involved with serious sequelae, including pelvic inflammatory disease, ectopic pregnancy and infertility. Furthermore, chlamydia can produce the eye infection trachoma, which is the primary infectious cause of blindness worldwide. The WHO reported 1.3 million cases of blindness in the world due to trachoma [3]. Beside serious sequelae, chlamydia infections generate remarkable medical costs. The estimated lifetime cost of chlamydia infected patients of 18-49 years of age in Shandong amounts 273 million [4]. In the end, chlamydia is a serious STD causing a significant morbidity and socioeconomic burden.

Lifecycle of *C. trachomatis*

Chlamydia is caused by the obligate intracellular gram-negative bacteria *Chlamydia trachomatis*. Due to the reliance of *C. trachomatis* on nutrients of the host cell, it needs to live inside the host cell to differentiate. The life cycle of *C. trachomatis* is a unique developmental cycle in which the particle switches between the elementary body (EB) and the reticulate body (RB) (Fig. 1). The lifecycle begins with phagocytosis of infectious and metabolically inactive EB. Consequently, the EB differentiates into the RB, which is non-infectious but metabolically active. The RB differentiates by binary fission and after rounds of multiplication, RBs begin to convert back into EBs and leave the cell by lysis of the host cell. The cell lysis is provided by cysteine proteases, which destroy the host cell by disruption of inclusion and cellular membranes. The released EBs bind again to neighboring cells and induces subsequent rounds of infection [2][3].

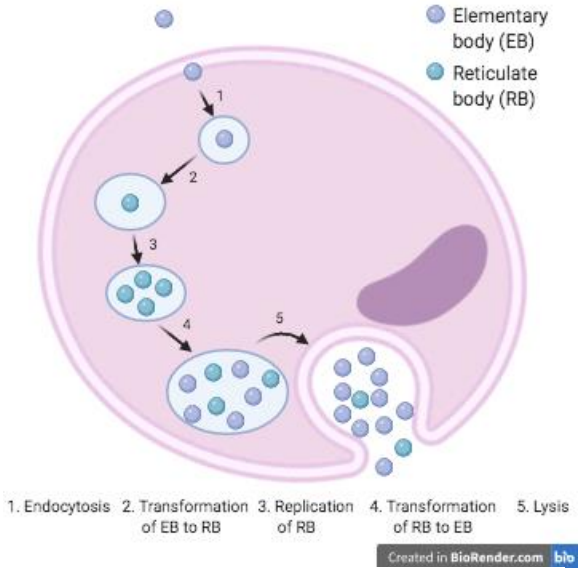


Figure 1: The developmental cycle of Chlamydia Trachomatis.

Proteins and antigens of *C. Trachomatis*

The genome of *C. trachomatis* contains a single 1,043,000 bp chromosome and encodes 893 chromosomal and 8 plasmid open reading frames. This small genome encodes for different types of proteins and antigens, which are necessary for the infection. *C. trachomatis* contains thermostable antigen Lipopolysaccharide (LPS), which enables the transition from the reticulate body (RB) to the elementary body (EB) and therefore plays a crucial role in the pathogenesis of chlamydia.

Furthermore, *C. trachomatis* encodes for different outer membrane proteins. The MOMP (major outer membrane protein) is highly present on the surface of the membrane and enables the presentation of trimeric EBs and monomeric RBs. The PMP (polymorphic membrane protein) provides the transport and acts as an adhesin. The COMC (complex of outer membrane protein) serves as an adhesin and is involved in the transition from RB to EB. In addition, *C. trachomatis* contains cellular process proteins or heat shock proteins (Hsp), which are activated by stress and serve as chaperones.

Also, *C. trachomatis* contains the type III secretion system (TTSS), which releases effector proteins in the inclusion membrane or into the cytosol. The TTSS enables the interaction between the bacterial pathogen and the host. Important proteins released by the TTSS are TarP, CrpA, Inc proteins (IncA) and chlamydial protease-like activity factor CPAF [3]. The latter is an important protease, which blocks

apoptosis in infected cells by degradation of transcription factors of the host (Fig. 2). The structure of CPAF contains two structural domains CPAFn and CPAFc and the large groove between these domains reveals a possibly active site of CPAF (Fig. 2). The structure analyzation of CPAF provides more insight into the active site of CPAF resulting in more knowledge of the pathogenesis of chlamydia.[15]

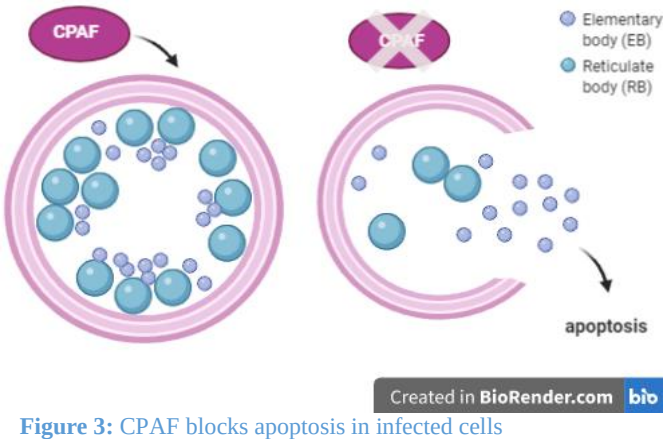


Figure 3: CPAF blocks apoptosis in infected cells

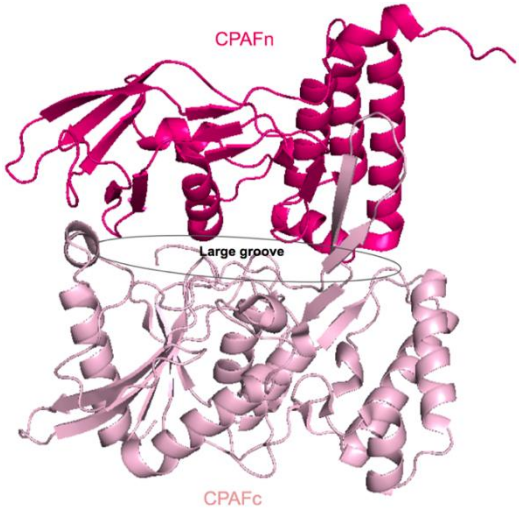


Figure 2: the 3D-structure of CPAF, which contains two structural domains CPAFn (dark pink) and CPAFc (light pink). The large groove is indicated with a black circle.

Syphilis

The sexually transmitted disease syphilis is caused by the spirochete *Treponema* pallidum, which infects the skin and mucous membrane. As reported by the WHO, approximately 17.7 million individuals with an age of 15-49 were infected with syphilis in 2012. Especially among men who have sex with men (MSM), there has been an increase in the numbers and rates of reported cases in Western Europe and in the United States since 1998 [6].

Acquired syphilis infected individuals follow a multistage disease course with wide-ranging indications, divided into primary, secondary, latent, and tertiary stages. Within every stage, the symptoms become more serious. During the primary stage, the spirochetes destroy soft tissue and skin, resulting in a single ulcer or multiple lesions on the genitals. Patients with secondary syphilis present with spirochetes in the bloodstream, causing fever, headache and a maculopapular rash on the skin. In the tertiary stage, the type IV hypersensitivity reaction manifests destructive cardiac or neurological conditions and granulomatous lesions. Untreated individuals who developed tertiary syphilis can get serious sequelae, including destructive cardiac or neurological conditions, severe skin conditions and damage of the liver, joints and testes. Besides, congenital syphilis causes serious manifestations in pregnant women and is the most common infection associated with fetal loss or stillbirth in low-income populations [6]. According to a report from Tanzania, syphilis is responsible for up to 50% of stillbirths. Furthermore, syphilis lesions are a portal of entry for HIV, causing an increased risk of the transmission and acquisition of HIV [5]. All these factors make syphilis a serious STD and an important public health concern.

Treponema pallidum

The gram-negative obligate parasite *Treponema pallidum* causes syphilis. The parasite belongs to spirochetes and contains endoflagella, which provides motility by spinning and its spiral shape. According to the Genome Sequencing Project, *T. pallidum* contains a small genome, which encodes just over 1000 genes. Due to its lack of metabolic capabilities, it is suspected that *T. pallidum* gains the most essential macromolecules from the host environment. *T. pallidum* receives

macromolecules from its host due to specific transporters of *T. pallidum,* such as energy coupling factor (ECF)-type ABC transporter.

Invasion

Despite the lack of metabolic capabilities of *T. pallidum*, the spirochete is able to invade and survive in various tissues and organs. The first step of the invasion of T. pallidum is the attachment of the parasite to the host cell. It has been shown that *T. pallidum* attaches to different cell types, including epithelial, endothelial and fibroblast-like cells. Components of cell membranes, host serum and extracellular matrix (ECM) are involved in the attachment of *T. pallidum*. ECM components such as fibronectin-coated coverslips, collagen I, hyaluronic acid and laminin bind to *T. pallidum* [5]. *T. pallidum* contains the protein Tp0751, which is expressed as recombinant protein and binds specifically to laminin. This implies that Tp0751 has potential interactions with the ECM of the host. Structural analysis of Tp0751 reveals an eight-stranded antiparallel compact β-barrel conformation capped by a shorter 3_{10}-like helix and a longer N-terminal helix (Fig. 4). The β-barrel conformation potentially provides a hydrophobic ligand-binding pocket and different loops provide the coordination of different hydrophobic ligands. The structural analysis of Tp0751 supplies more insight into the disruption of the treponemal-host interaction.[22]

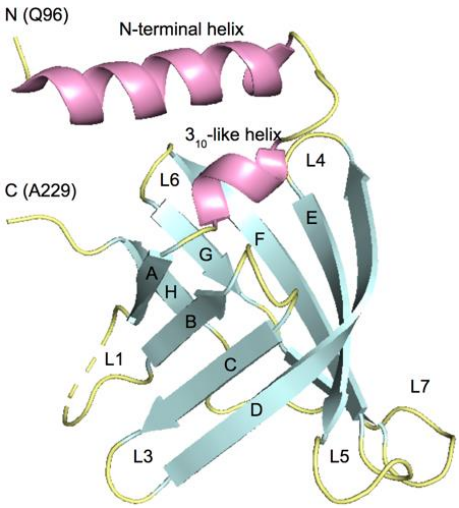


Figure 4: 3D-structure of Tp0751, colored according the secondary structure with **α**-helices in pink, **β**-strands in cyan and connecting coil in yellow.

Protein targets of *Chlamydia trachomatis*

CT263

The *C. trachomatis* protein CT263 potentially supports quinone biosynthesis, which is necessary for the electron transport chain and is essential to cellular respiration in bacteria. Quinones are highly lipid-soluble and are utilized in regulatory pathways. To confirm CT263s potential role in quinone biosynthesis, the protein is structurally compared to the enzyme 5-methylthioadenosine Nucleosidase (5-MTAN). This enzyme plays a key role in the futalosine pathway for menaquinone biosynthesis.
To investigate CT263s potential role in quinone synthesis, a crystal structure of CT263 residues 1-194 at 1.58 *Å resolution* is created using PyMOL (DeLano Scientific) (Fig. 5). The crystal structure in a cartoon format is colored in rainbow, changing from blue (N-terminus) to red (C-terminus) (Fig. 5A). CT263 contains a single globular domain and consists of a central core of 10 twisted β-sheets and two sets of paired-α-helices. The surface of the electrostatic potential of CT263 shows a red negatively charged cleft, which is indicated with a black circle (Fig. 5B). Both the cartoon format and the electrostatic surface of CT263 are orientated in the same conformation. This reveals that the cleft is created by α2, α4, and the loop attached to β8-β9. The large surface-exposed cleft contains a high degree of electronegative charge density, which is essential for substrate binding.

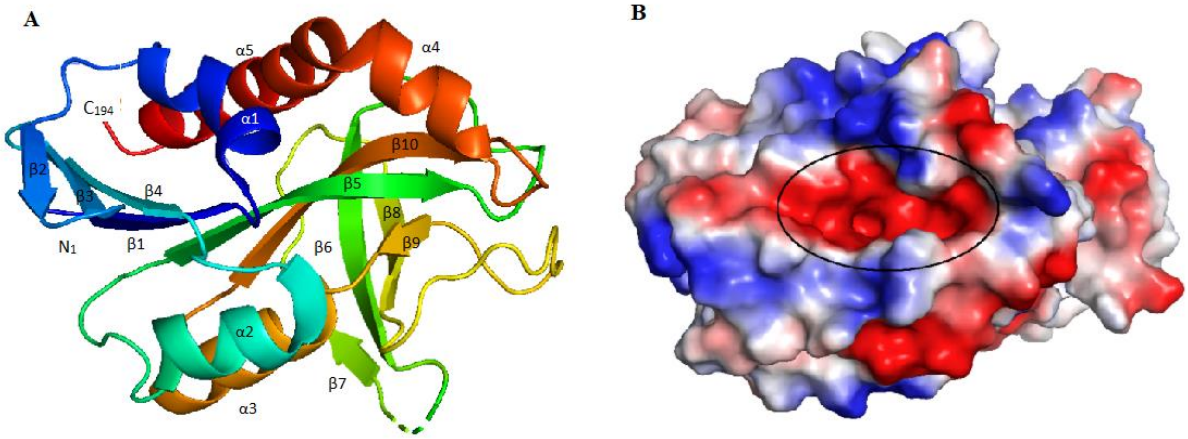


Figure 5: Crystal structure of CT263. *A.* Cartoon format colored in rainbow, changing from blue (N-terminus) to red (C-terminus). *B.* surface representation of electrostatic potential, showing a negatively charged cleft (black circle)

According to the DALI database [24], CT263 contains structural similarities with 5′-methylthioadenosine nucleosidase (MTAN) enzymes. Figure 6 shows the structural similarities of CT263 (PDB code: 4QAQ) on the left in orange and E. coli MTAN (PDB code: 1Z5P) on the right in pale cyan. The line colored in pink is the secondary structure lacking interpretable electron density within CT263. The four regions that show structural differences are highlighted in deeptale cyan (E. coli MTAN) and could potentially lead to an altered enzymatic profile of CT263. Nevertheless, the core fold of CT263 shows structural similarity to prototypical MTANs, which suggests that CT263 could fulfill the nucleosidase enzymatic role and provide the futalosine pathway. This provides support for quinone synthesis in chlamydia which might lead to another drug target to combat chlamydia infections.[7]

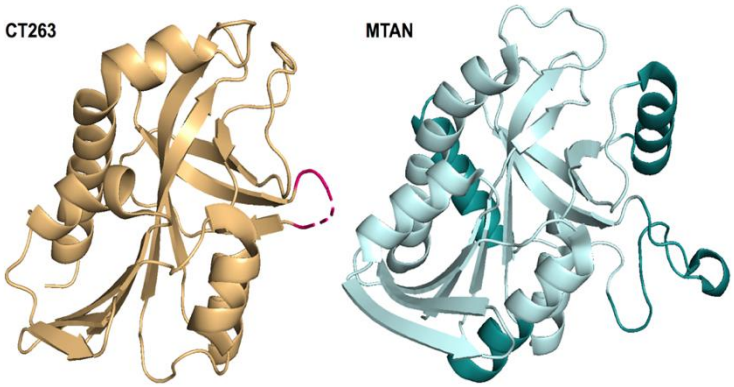


Figure 6: Crystal structure of CT263 (orange) and its structural homolog MTAN (cyan).The four regions with structural difference are highlighted in deeptale cyan.

ChlaDUB1

During infection, *C. trachomatis* translocates effector proteins directly into the host to modify signaling pathways of the host. The chlamydia effector protein ChlaDUB1 facilitates lys63-deubiquitinase (DUB) and lys-acetyltransferase (AcT) activities, which enable hydrolysis and condensation reactions, respectively. ChlaDUB1 switches off ubiquitin-dependent inflammatory signaling processes and inhibits microbe-directed autophagy, NF-kB signaling and cell death during infection. As result, ChlaDUB1 is able to alter signaling pathways principal for invasion, survival and replication and has a crucial role in the pathogenesis of chlamydia.

To better understand the remarkable dual DUB and AcT activities in ChlaDUB1, protein structures are analyzed. Crystal structures of ChlaDUB1 bound to Ub and bound to Coenzyme A (CoA) at a resolution of 1.9Å and 2.1Å, respectively, are created using PyMOL (DeLano Scientific) (Fig. 7). The complex crystal structures of ChlaDUB1 represent the intermediate stages of deubiquitinase (Fig. 7A) and acetyltransferase (Fig. 7B) activities. The active site of ChlaDUB1 is shown, which presents a Cys-His-Asp triad and a Gln oxyanion hole. The crystal structures of the enzyme bound to Ub and bound to Coenzyme A (CoA) show barely any conformational alterations between each other, but show distinct binding sites for CoA and Ub. Collectively, this clarifies how ChlaDUB1 is able to fulfill the two distinct chemical reactions deubiquitination and acetylation, using the same catalytic triad. This provides further insight into the mechanism by which *C. trachomatis* may avoid the inflammatory response of the host.[8]

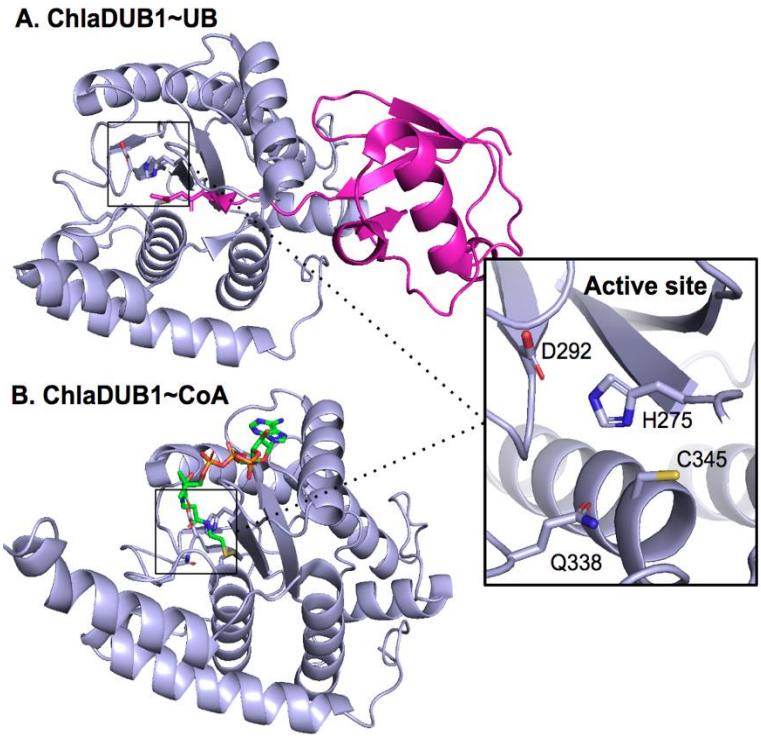


Figure 7: Crystal structures of ChlaDUB1 and the active site representing a Cys-His-Asp catalytic triad and a Gln oxyanion hole. *A.* ChlaDUB1 bound to UB. *B.* ChlaDUB1 bound to CoA.

Pgp3

C. trachomatis secretes the plasmid gene product 3 (Pgp3), which is a putative virulence factor and an immunodominant antigen. The homotrimer Pgp3 encodes on a cryptic plasmid and secretes into the inclusion lumen and cytosol of the host cell. Pgp3 is a promising vaccine antigen candidate based on its exhibition of macrophages leading to the secretion of inflammatory cytokines.

To better understand the function of the protein a crystal structure of Pgp3 at a resolution of 2.0 *Å* is created using PyMOL (DeLano Scientific) (Fig. 8). The C-terminal domain (CTD) of pgp3 is trimeric, cylindrical and compact, which assemblies similar to tumor necrosis factor (TNF) family proteins. Each CTD monomer folds into a β-barrel jelly roll consisting of 10 antiparallel β-strands, which generate two sheets. The sheet at the trimer interface contains β-strands 1, 4, 7 and 10 and the sheet facing the solvent contains β-strands 3, 5, 6 and 8 (Fig. 8C). In the sequence of Pgp3, the only residue containing a tryptophan residue is Trp-234, which is shown as yellow sticks represented as dots. Trp-234 is present in a shallow groove generated by the short β-strands 2 and 9 and the loop connecting 9 to 10,

which provide solvent-exposure of both polar and apolar parts of its indole ring. This implies that Trp-234 is a 'hotspot' for protein-protein interactions, which enables the immunogenic nature of Pgp3.

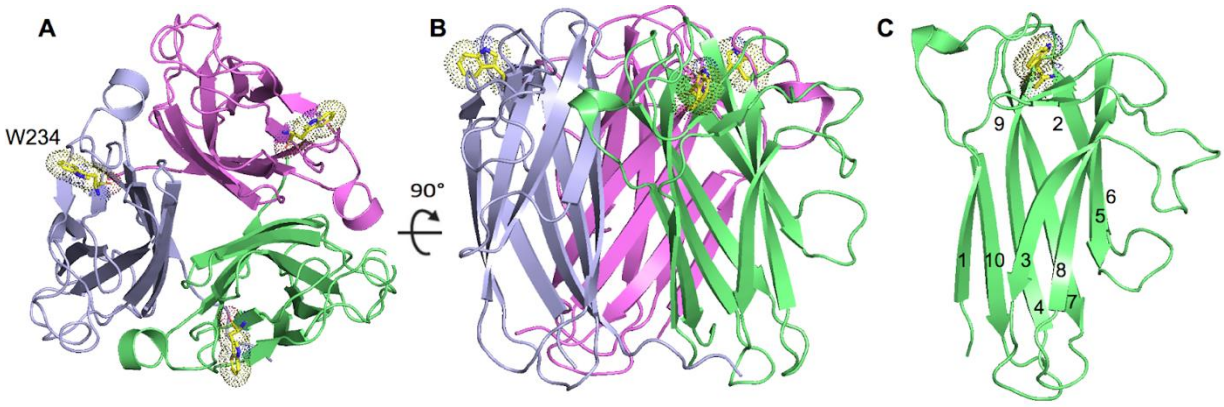


Figure 8: Crystal structure of the CTD of Pgp3. A/B. Pgp3 reveals a trimeric, cylindrical and compact conformation. The tryptophan residue is shown as yellow sticks represented as dots. *C.* CTD monomer folds into a β-barrel jelly roll and the β-strands are numbered from the N-terminus to the C-terminus

By using the DALI database [24] and PDB code for structural homologs, Pgp3 reveals structural similarities with the TNF Family of Proteins, in which Bc2L-C is the most similar structural homolog. Bc2L-C plays an important role in bioadhesion by triggering interleukin-8 production and is potent to deregulate a proinflammatory response. To show the structural alignment of the TNF-like CTD with Bc2L-C, crystal structures are created (Fig. 9). This illustrates an orthogonal view of the C-terminal domain trimer of Pgp3 in green (Fig 9A) and of the Bc2L-C trimer in blue (Fig. 9B), and the superposition of a Bc2L-C monomer onto a Pgp3 CTD monomer (Fig. 9C). This confirms the structural similarities between Bc2L-C and Pgp3.

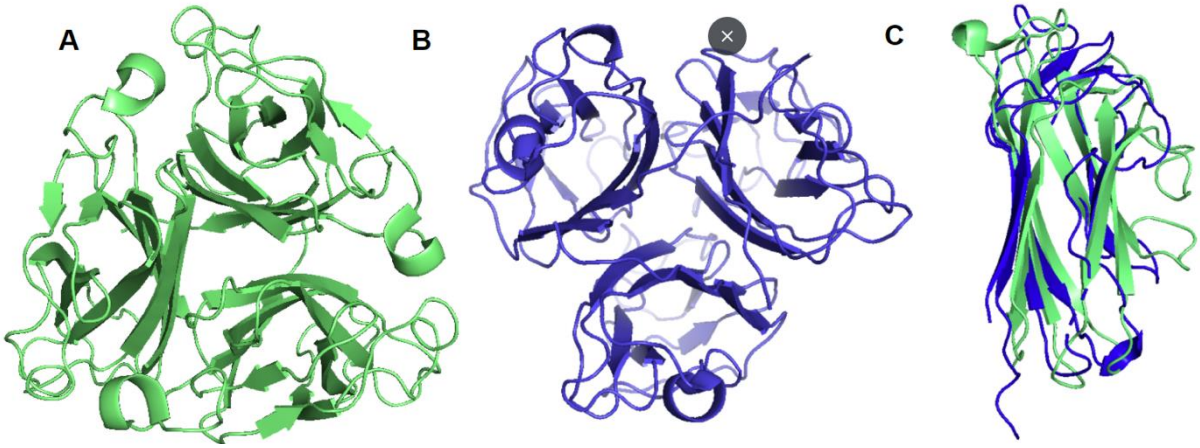


Figure 9: Crystal structures of *A.* the CTD of Pgp3 (green) and *B.* the Bc2L-C trimer (blue) and *C.* the superposition of Bc2L-C monomer onto a PGp3 CTD monomer.

So, the 3D-structure of Pgp3 confirms its crucial role in adhesion between the chlamydial organism and the host cell. This provides the basis for understanding its role in chlamydial pathogenesis and implies pgp3 as a potential vaccine antigen candidate.[9]

DsbH

The EB of chlamydia is metabolically inactive and highly resistant to lysis, potentially due to outer envelope proteins. These proteins are highly crosslinked by disulfide bonds and are able to substitute for peptidoglycan in the EB. Interference with chlamydial disulfide bond catalysis provides an alternative treatment of chlamydial infections instead of antibiotics. The novel periplasmic oxidoreductase in chlamydia, termed DsbH, potentially catalyzes chlamydial disulfide bonds. To gain

more insight into this process, a crystal structure of DsbH is created using PyMOL (DeLano Scientific) (Fig. 10). The disulfide bridge is located at the beginning of helix 2, which is characteristic of thioredoxin-like oxidoreductase. DsbH adopts a thioredoxin fold because of its antiparallel β-sheets sandwiched between α-helices and its oxidoreductase activity. Conspicuous is the orientation of α-helix 3, which is close to perpendicular to the β-sheets and α-helices. Likewise, the β-sheets 4 and 5 manifest a remarkable twist, while β-sheets 2 to 4 are moderately twisted relative to one another (Fig. 11B).

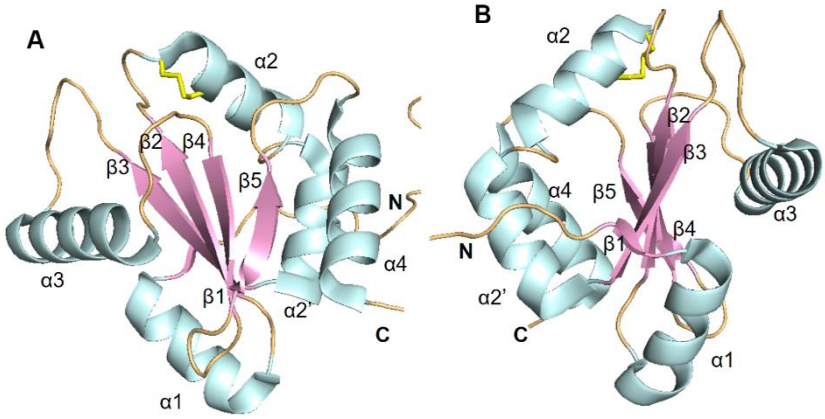


Figure 10: A crystal structure of DsbH colored according the secondary structure with α-helices colored in cyan, β-sheets colored in pink, connecting coils in light orange. and the catalytic disulfide bridge (Cys-50—Cys-53) in yellow

By using the DALI database [24] and PDB code for structural homologs, DsbH reveals structural similarities with DsbD and thioredoxin. This is confirmed by the structural alignment of DsbH (blue) with thioredoxin and DsbD (pink), (Fig. 11A and B, respectively). DsbH, thioredoxin and DsbD show high structural homology, even though the DsbH-thioredoxin and DsbH-DsbD pairs only share a sequence identity of 18.9 and 16.5 %, respectively. DsbH contains the same topology as thioredoxin, whereas DsbD contains an additional helix instead of β-strand 5. As discussed above, β-strand 5 of DsbH is not well connected to its central β-sheet, which implies an evolutionary change over the original thioredoxin conformation. A redox potential depends on two central residues of the CXXC motif, which are different among the three proteins. So, the close structural similarity of DsbH, DsbD and thioredoxin does not necessarily impose functional similarity. Although there is no direct evidence for the role of DsbH, it is likely that DsbH provides a reduction of thiol-disulfide oxidoreductase in chlamydia. The thioredoxin fold of DsbH and the structural similarities with thioredoxin suggests that DsbH maintains the reduction of the periplasm. This implies the crucial role of DsbH and provides further insight into the pathogenesis of chlamydia.[10]

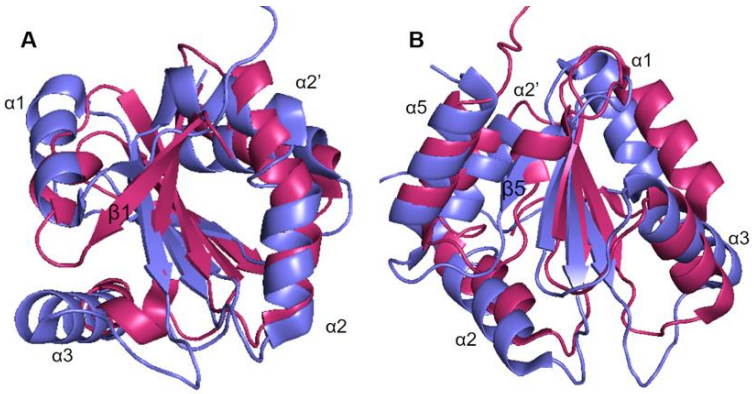


Figure 11: Structural comparison of DsbH with thioredoxin and DsbD *A.* Superposition of DsbH (blue) with thioredoxin (pink) *B.* superposition of DsbH (blue) with DsbD (pink)

CT584

C. trachomatis contains a type III secretion system (T3SS), which releases effector proteins in the inclusion membrane or into the cytosol. The T3SS enables the interaction between the bacterial pathogen and the host. Needle-tip proteins are localized in the distal end of T3SS, which provides the secretion state of the injectisome. CT585 is an important T3SS effector protein, which shares biophysical characteristics with needle-tip proteins and provides protein-protein interactions.

To gain more insight into these interactions, a crystal structure of CT584 (residues THr14-THr180) at a resolution of 3.1 *Å* is created using PyMOL (DeLano Scientific) (Fig. 12A). The refined crystal structure is visualized as a cartoon format and colored according to the secondary structure with α-helices colored in green, β-sheets colored in violet and connecting coils in yellow. At the C-terminus, the two

α-helices are folded in a kinked antiparallel manner, which forms a dimerization interface. Since Pro residues are absent in α-helix 6, the noncanonical kink is stabilized.

By using the DALI database [24] and PDB code for structural homologs, the structure of CT584 shows structural similarity with Cpn0803, which is an ortholog of CT584 from *Chlamydia pneumonia*. The proteins share a sequence identity of 88.8% of the aligned region. This high level of sequence identity is represented in the superposition of the proteins, in which CT584 is colored in orange and Cpn0803 colored in pale cyan (Fig. 12B) The structural similarity of the proteins implies the highly conserved nature of CT584 across these chlamydial species, despite different tissue tropism disease pathology. This provides more insight into the role of CT584 in the pathogenesis of *C. trachomatis*.[11]

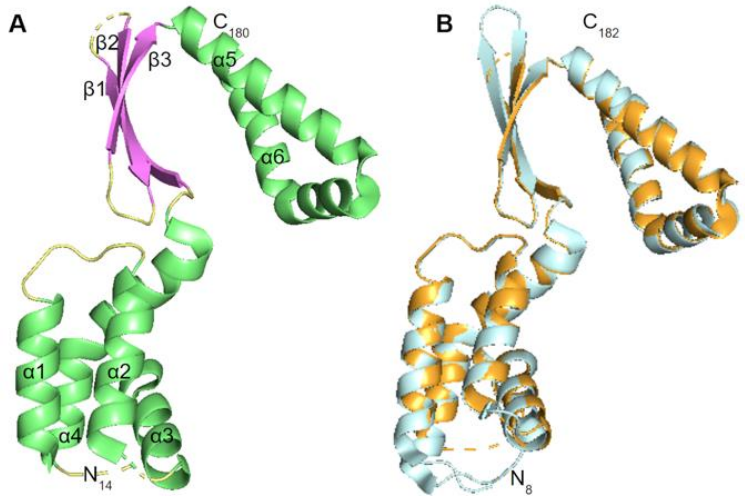


Figure 12: *A.* Crystal structure of CT585 colored according to secondary structure, with α-helices in green, β-sheets in violet and connecting coils in yellow. *B.* Superposition of Ct584 (orange) with Cpn0803 (cyan)

IncA

The life cycle of *C. trachomatis* depends on 'inclusion', which is a parasitic organelle inside the host cell. *C. Trachomatis* is able to form inclusion membranes including ~60 transmembrane Inc proteins. These Inc proteins are poorly characterized but are potent to facilitate interactions with host cell components. The prototypical Inc, also known as IncA, contains two extended 3,4-hydrophobic heptads repeat segments, which contain similar coiled-coil regions as the eukaryotic soluble N-ethylmaleimide-sensitive factor attachment receptors (SNAREs). Besides, IncA promotes the homotypic fusion of inclusions which provides the enhancement of chlamydia pathogenicity. IncA is a protein of 273 amino acids and consists of a cytoplasmic N-terminal moiety (residues 1-34), a bilobed transmembrane domain (residues 35-84), and a long cytoplasmic C-terminal domain ending in a tail (residues 247-273). The bacterial SNARE-like domains are located in the cytoplasmic C-terminal domain.

To better understand IncA-mediated membrane fusion and to clarify its potential role in chlamydia pathogenicity, a crystal structure of IncA residues 87-246 (IncA87-246) at 1.12 *Å resolution* is created using PyMOL (DeLano Scientific) (Fig. 13). The tertiary structure of IncA reveals a four down-up-down-up antiparallel α-helices confirmation, named H_A-H_D. Still, IncA87-246 is considered as a non-canonical bundle based on at least two aspects. First, the helix H_B is divided at position 144 by a glycine

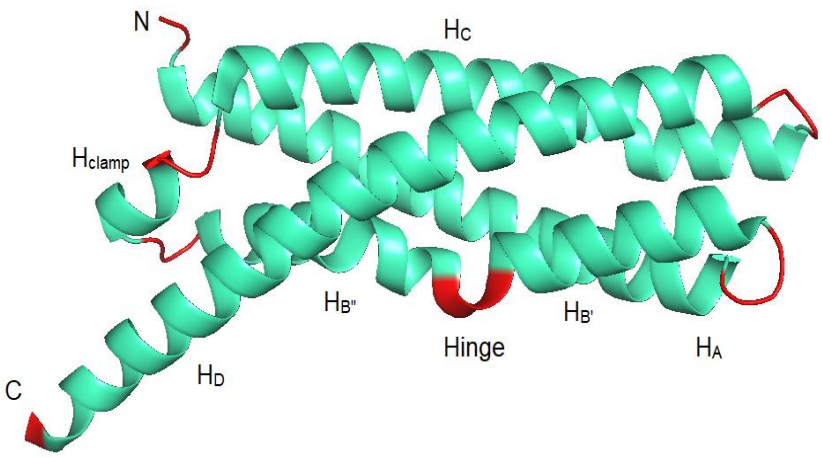


Figure 13: Crystal structure of IncA with α-helices in cyan and random coiled linkers and hinge region colored in red.

in two regions named $H_{B'}$ and $H_{B''}$. This break in the α-helix is also referred to as 'Hinge'. Second, IncA contains the short insertion helix H_{clamp}, which connects H_B and H_C and is able to interact with all four α-helices.

The high interaction of the clamp is due to its provision of eight hydrogen bonds and 94 non-bonded interactions (Fig. 14). The clamp provides contacts with the α-helices H_D, $H_{B''}$, H_C and H_A (Fig. 14a, b, c and d; respectively). During infection, the clamp facilitates IncA-mediated homotypic membrane fusion, which is critical for the pathogenesis of chlamydia. This provides more insight into the role of IncA in the homotypic fusion of inclusion and its crucial role in the chlamydia pathogenicity.[12]

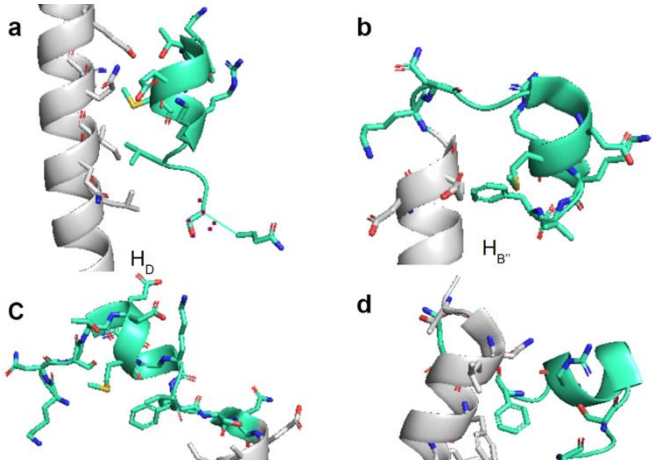


Figure 14: The interaction of the clamp (cyan) with different α-helices (grey). *a.* interaction of H_D with clamp. *b.* interaction of $H_{B''}$ with clamp. *c.* interaction of H_C with clamp. *d.* interaction of H_A with clamp.

CT771

Nudix hydrolases are a family of enzymes that provide the cleavage of nucleoside diphosphates coupled to any other moiety X (Nudix). This provides physiological homeostasis and regulation of signaling molecules and toxic intermediates. *C. trachomatis* encodes for a single Nudix family protein CT771, which asymmetrically cleaves the metabolite asymmetric diadenosine 5',5''-P1,P4-tetraphosphate (Ap4A) into ATP and AMP by hydrolyse while providing homeostasis. Ap4A serves as a substrate for RNA polymerase and is potentially involved in different signaling pathways, among which DNA repair, pathogenesis and apoptosis.

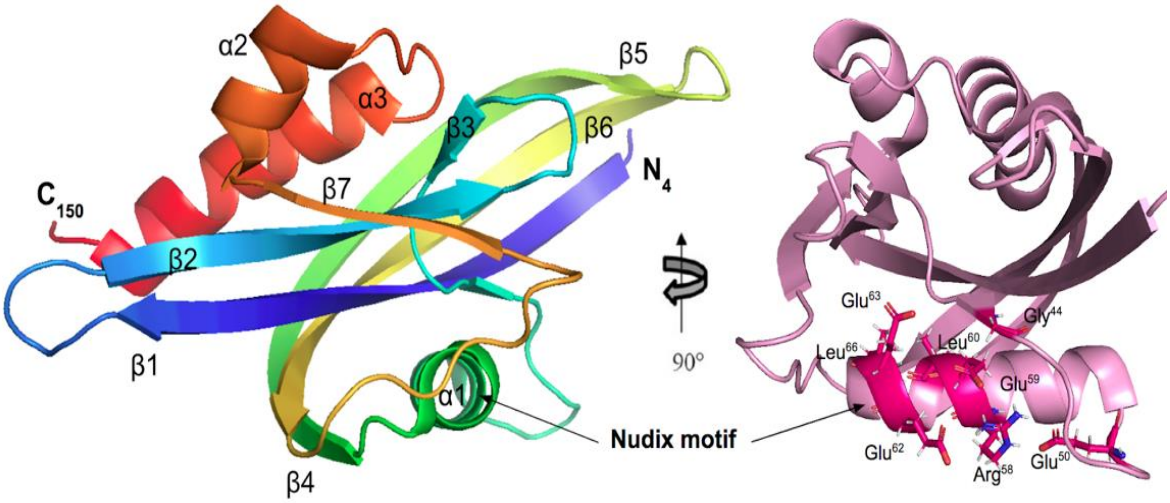


Figure 15: Crystal structure of CT771 colored in rainbow color (left) and incorporated amino acids in the Nudix motif in pink (right).

To investigate CT771s role as a potential Nudix hydrolase, a crystal structure of CT771 at a resolution of 2.6 *Å* is created using PyMOL (DeLano Scientific) (Fig. 15). The crystal structure is visualized as a cartoon format and is colored in rainbow, changing from blue (N-terminus) to red (C-terminus). The protein folds into an αβα-sandwich, which is characteristic for Nudix hydrolases. The conserved Nudix motif is located in α1 and in the loop regions around this helix. The amino acids incorporated in the Nudix motif are indicated in dark pink and the remainder of the CT771 structure is colored in light pink.

By using the DALI database[24] and PDB code for structural homologs, CT771 reveals structural similarities with enzymes from the Nudix family, in which Ap4A hydrolases are most similar homologs. This is confirmed by the superposition of CT771 (pink) and *Homo sapiens* Ap4A hydrolase (pale cyan) (Fig. 16) The proteins are structurally similar, except for the loops connecting β1-β2 and β5-β6. However, these loops connecting regions are not predicted to provide substrate binding or catalysis. These structural data suggest that CT771 is a *bona fide* Ap4A hydrolase, which potentially provides a crucial role in the pathogenesis of chlamydia.[13]

Figure 16: Superimposition of CT771 (pink) and *H. Sapiens* Ap_4A hydrolase (cyan).

CT441

C. trachomatis encodes for the protein CT441, which shares a similar amino acid sequence with a group of bacterial tail-specific proteases (Tsp). Tsps are able to cleave substrate proteins labeled with a C-terminal ssrA-encoded peptide tag and supply the protein quality control in the periplasm by interfering with the NF-κB pathway. CT441 is possibly involved in the modulation of estrogen signaling responses of the host cell. Previously, it has been suggested that this signaling response is interrupted by CT441-mediated degradation of host-derived SRAP1 (steroid receptor RNA activator protein 1), which is a coactivator of estrogen receptor (ER). However, this putative role of CT441 has been put into question.

To gain more insight into the role of CT441 during chlamydiae infection, a crystal structure of this protease at 3.0 *Å resolution* is created using PyMOL (DeLano Scientific) (Fig. 17). The refined crystal structure is visualized as a cartoon format and colored by domain.

CT441 contains a modular domain organization consisting of an N-terminal domain (pink), a PDZ domain (green) and a C-terminal protease domain (blue). The PDZ domain is a protein-protein interaction domain and binds to specific amino acid sequences located at the C-terminal domain. The approximate positions of the PDZ domain and several loop regions are indicated by a green box and dashed lines, respectively, because these regions were too flexible to be modeled into electron density. The residues of the proteolytic site in the C-terminal domain supply the proteolytic function and are shown as sticks (yellow), in which oxygen atoms are indicated in red and nitrogen atoms in blue. This catalytic triad consists of glutamine (Q485), lysine (K481) and serine (S455) and the hydrogen bonds aligning the active-site residues

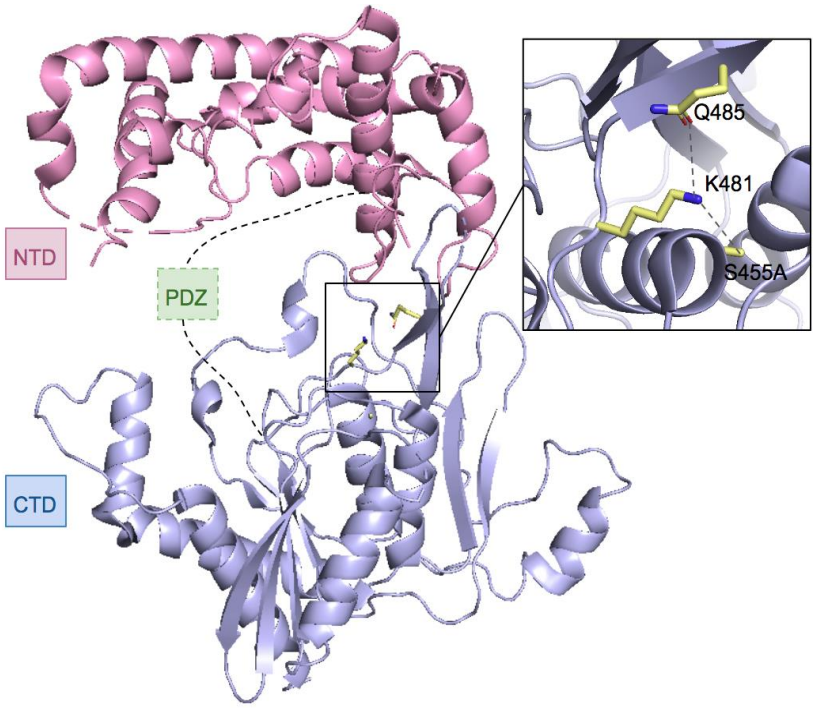


Figure 17: Crystal structure of CT441 colored by domain with NTD in pink, CTD in blue and PDZ in green. The residues of the proteolytic sites are shown as sticks (yellow) with oxygen atoms in red, nitrogen atoms in blue and dashed grey dashed lines indicate

are indicated with dashed lines. The side-chain positions provide the base characteristic of K481-N, which accepts a proton of S455-O. The optimal positioning of K481 is provided by the hydrogen bond between the side chains of K481 and Q485. This implies that Q485 shares a similar role with the aspartate residue in the catalytic triad of classical serine proteases. The PDZ potentially influences the cleavage specificity of CT441 via the binding of substrate molecules. Collectively, the 3D-structure of CT441 and the catalytic triad confirm that CT441 will be an interesting target for future research into chlamydial pathogenicity mechanisms. CT441 will be an auspicious target for future research to gain further insights into mechanisms of chlamydia pathogenicity.[14]

CPAF

C. trachomatis secretes chlamydial protease-like activity factor CPAF, which plays a crucial role in the pathogenesis of chlamydia. CPAF is the most extensively studied effector protein in chlamydia research and is possibly involved in the degradation of a broad spectrum of host cell proteins. The protease is able to block apoptosis in infected cells by the degradation of host transcription factors. CPAF is synthesized as a catalytically inactive zymogen and converted into active CPAF, which is likely to occur via an autocatalytic process. The active CPAF consists of an N-terminal (CPAFn) and a C-terminal fragment (CPAFc).

To gain more insight into the catalysis mechanisms of CPAF, a crystal structure of CPAF at a resolution of 2.2 Å is created using PyMOL (DeLano Scientific) (Fig. 18). The mature CPAF contains two structural domains CPAFn and CPAFc, which are colored in light pink and pale cyan, respectively. The C-terminus of CPAFn is indicated by Nc and the N-terminus of CPAFc is indicated by Cn. A large groove is present between these domains (Fig. 18B), which possibly contains an active site of CPAF (Fig. 18C). The interaction between CPAFn and CPAFc (heterodimerization) is provided by the active site of CPAF, which consists of Ser499 and His105, shown in sticks and colored in magenta. No other polar residues are close enough available to form hydrogen bonds with Ser499 or His105. This implies that CPAF utilizes Ser499 and His105 for its catalytic activity and CPAF can be considered as a Ser protease. However, the catalytic triad of a classic Ser protease also contains an ASP residue, which lacks in CPAF. Still, the water molecule-mediated hydrogen bond between His105 and Glu558 may provide the activation of Ser499 for catalytic activity in CPAF. Collectively, the 3D-structure of CPAF confirms that CPAF can be considered as a Ser protease, of which the catalysis is provided by a water-mediated catalytic triad. This provides further insight into the development of specific inhibitors to regulate CPAF activity. This potentially attenuates chlamydial pathogenicity and brings new opportunities for anti-chlamydia drug development.[15]

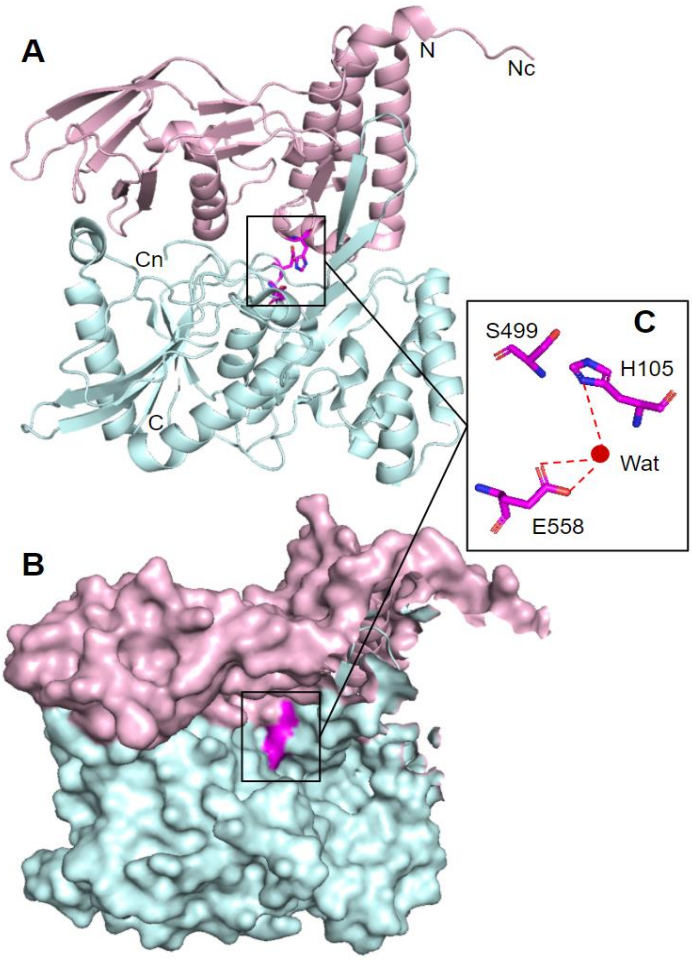


Figure 18: Crystal structure of CPAF in which CPAFn is indicated in pink and CPAFc in cyan. A. Cartoon format of CT441 in which the catalytic triad is shown in sticks (magenta) with oxygen and nitrogen atoms in red and blue, respectively. B. Surface representation of CT441, which represents a large groove containing an active site (magenta)

Protein targets of *Treponema pallidum*

Tp0453

T. pallidum encodes for the antigen TP0435 (TP17), which is a potent immunogen and is the most highly expressed gene in *T. pallidum*. TP17 induces proliferation and exhibits the production of TNF-α. This raises the possibility that Tp17 is a possible target for antibodies and that immunity to Tp17 potentially combats syphilis infection.

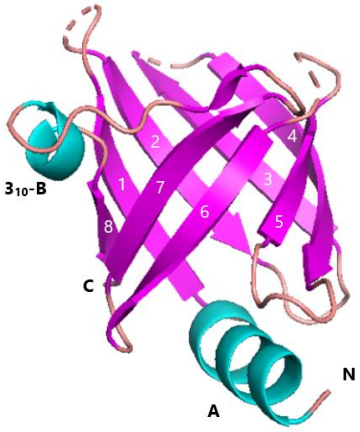


Figure 19: Crystal structure of Tp0435 in cartoon format colored according to the secondary structure with α-helices in cyan, **β**-sheets in magenta and connecting coil in pink

To discover potential binding ligands and to better understand the function of TP17, a 2.4 *Å* crystal structure of TP17 is created using PyMOL (DeLano Scientific) (Fig. 19). TP17 contains an eight-stranded antiparallel β-barrel conformation comprising an α-helix (helix A) on one end. Unlike the β-barrel conformation, TP17 prevents pore running through the center due to amino acid side chains located in the β-barrel. However, the opposite end of the barrel contains a shallow basin with a small 3_{10} helix (helix B). Conspicuous is that this basin-end contains a high frequency of disordered residues.

The asymmetric unit of the crystal structure contains two copies of TP17, which are related by a pseudo-twofold rotational axis. These two molecules are structurally very similar and are linked at their terminal ends by a disulfide-bond between the S-atoms of C18 of the respective molecules. This implies the possibility that TP17 consists of a covalently disulfide-stabilized dimer in solution. Hydrodynamic studies have manifested that this disulfide-linked dimer is specifically formed in solution under oxidizing conditions. Intermolecular disulfide bonds occurring at two-fold axes are scarce, but several examples exist according to the protein database including signaling molecules, myostatin and a CAP (catabolite activator protein). The hydrodynamic behavior raises the possibility that TP17 plays roles in ligand binding, treponemal membrane architecture, and/or pathogenesis. However, additional studies to clarify the potential ligand- or protein-binding tendency of TP17 are necessary to definitively elucidate the protein's function.[16]

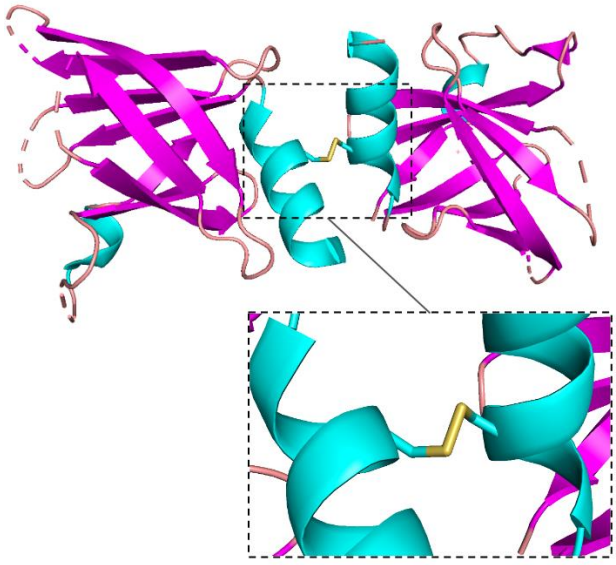


Figure 20: Crystal structure of disulfide linked dimer of Tp0435 representing the side chains of C18 with sulfur atoms in yellow.

Tp0624

T. pallidum is able to stabilize the cell envelope by Outer-Membrane Protein A (OmpA)-like domain-containing proteins, which bind non-covalently to peptidoglycan. OmpA-like domains support outer membrane integrity and indirectly provide host-pathogen interactions through a structural bridge between the outer membrane and the peptidoglycan layer. The outer membrane is stabilized via three possible mechanisms: the insertion of an β-barrel domain or a lipid anchor distal to the C-terminal OmpA-domain or the constitution of the bacterial outer membrane porins in the OmpA-domain. During infection, *T. pallidum* encodes for the protein Tp0624, which contains a putative OmpA-like domain. This domain suggests its potential capability of cell envelope stabilization and peptidoglycan coordination, which is possibly provided by the uncharacterized N-terminal region.

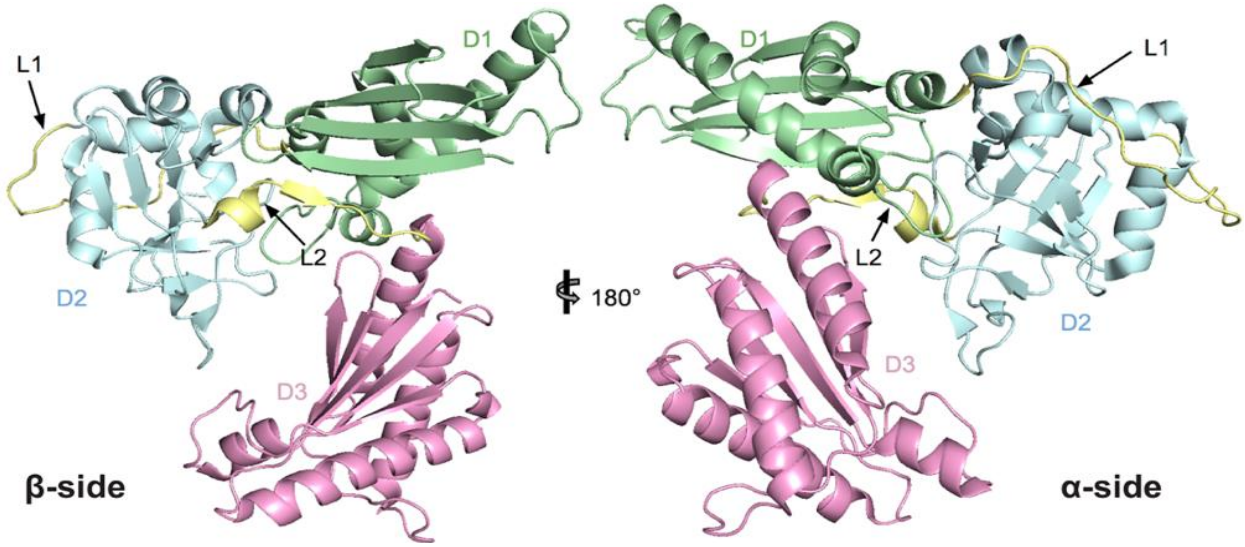


Figure 21: Crystal structure of Tp0624 colored by domain with D1 in green, D2 in cyan and D3 in pink and revealing an **α**- and **β**-side. The domains are connected by two interdomain linkers L1 and L2 (yellow).

To gain more insight into the function of Tp0624, a crystal structure with a resolution of 1.7 Å is created using PyMOL (DeLano Scientific) (Fig. 21). The crystal structure is visualized in cartoon format and is colored by three distinct domains, referred to as D1, D2 and D3 (green, pale cyan and pink, respectively). Tp0624 contains a slightly flattened, triangle-like architecture with a β-sheet side (left) and an α-helix side (right). The domains are connected by two interdomain linkers (yellow). D1 is linked to D2 via a 20 residue linker (L1) and D2 is linked to D3 via a 14 residue linker (L2). The C-terminal portion of L1 provides significant flexibility by wrapping around the end of D2. Despite this flexibility of L1, the interface between D1 and D2 provides that these domains function as an integrated unit. Besides L2 connects D2 and D3, it also creates a terminal strand in the β-sheet of D1, which provides the overall integrated nature of the Tp0624 modular architecture.

D3 is the largest of three domains and consists of a five-stranded β-sheet of mixed parallel and antiparallel strands and three α-helices (Fig. 22A). According to the sequence of D3, this domain belongs to the OmpA-OmpF porin (OOP) family. To investigate its putative ligand-binding site, an electrostatic surface of D3 is created (Fig. 22B). The electrostatic surface of D3 is orientated the same as in A and shows a clear basic patch,

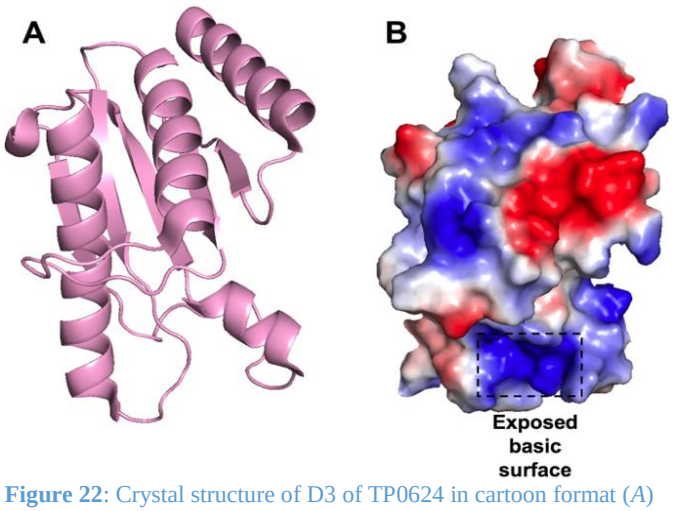


Figure 22: Crystal structure of D3 of TP0624 in cartoon format (*A*) and in electrostatic surface (*B*) with negative surface in red and positive surface in blue revealing an exposed basic surface indicated with a black dashed outline.

which provides a putative ligand-binding site. The putative binding site divulges a strongly basic region which suggests that Tp0624 D3 is likely to anchor to the thin periplasmic peptidoglycan layer of T. *pallidum*. This enables the coordination of peptidoglycan, which stabilizes the cell envelope. According to the structural data, D3 of Tp0624 provides the coordination of the peptidoglycan layer resulting in the interaction of D1 and D2 with outer membrane proteins (OMP) (Fig. 23). The domains are colored the same as in figure 21 and the blue basic surface binds non-covalently to peptidoglycan. Collectively, this suggests that Tp0624 provides the stabilization of the unusual cell envelope and plays a crucial role in the pathogenesis of *T. pallidum*.[17]

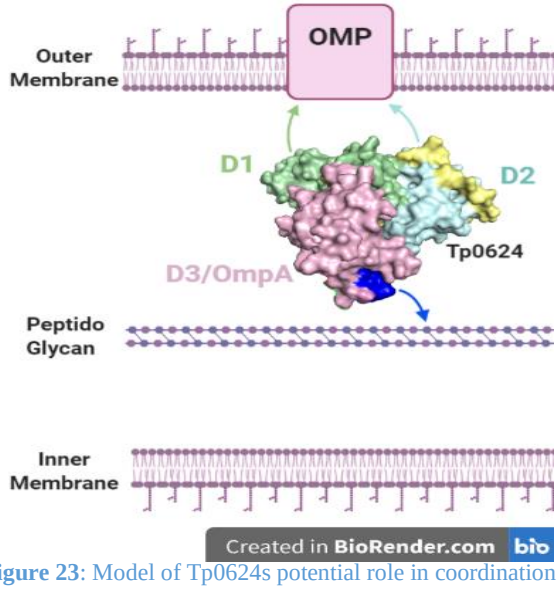


Figure 23: Model of Tp0624s potential role in coordination of *T. Pallidum* peptidoglycan and stabilization of the cell envelope.

Tp0684

T. pallidum encodes for the protein Tp0684, which is a single lipidated chemoreceptor and is also known as TpMglB-2 based on its sequence homology to MglB of *E. coli*. In *E. coli*, MglB serves as a chemoreceptor for galactose chemotaxis and it functions as a periplasmic ligand-binding protein of an ABC-type transporter.

To discover potential interactions and to understand the function of TP0684, a 2.0 Å crystal structure of TP0684 is created using PyMOL (DeLano Scientific) (Fig. 24A). The crystal structure is represented in a cartoon format and colored according to the secondary structure, with α-helices colored in light pink, β-sheets colored in pale cyan and connecting coils in light orange. The N-lobe contains a twisted β-sheet, consisting of seven β-strands, of which one of the β-strands is antiparallel to the other. The β-sheet is surrounded by α-helices. The C-lobe is similarly organized as the N-lobe, except for having one β-strand less. There are four regions available which crossover between the C-lobe and N-lobe, which provides the hinge region of the protein. The cleft between the lobes is solvent-exposed when there is no ligand present, resulting in an open, extended conformation. On the other hand, when a ligand is present in the cleft, the two lobes clamp down on it, leading to a more closed, compact conformation. Based on the solvent-available large cleft of Tp0684, the conformation of the protein appears to be in the open conformation.

The C-lobe shows interaction with two small molecules. The first is the cation Ca^{2+}, which is presented as an orange sphere bound to the C-lobe. The cation shows interaction with oxygen atoms (red) of D71, N73, D75, I77 and D166, which is shown by the 3D-structure (left) and 2D-structure (right) (Fig. 24B). These ligand interactions provide Ca^{2+} to stabilize a loop on the C-lobe and adhere to the main body of the domain. This suggests that Ca^{2+} plays an important role in the folding and integrity of the C-lobe. The second small molecule is the bound ligand BIS-TRIS, which is presented as a group of spheres between the N-lobe and the C-lobe (Fig. 24C). This ligand is stacked on the indole group of W145. Hydrogen bonds between the hydroxyl group of the protein and the ligand are provided by the side chains of residues D91, R95, N172 and D195, and are shown as black lines. The side chain of H89 and the ligand share van der Waals interactions. Due to the fact that Tp0684 folds into an open conformation with this small ligand, this raises the possibility that its binding is adventitious.

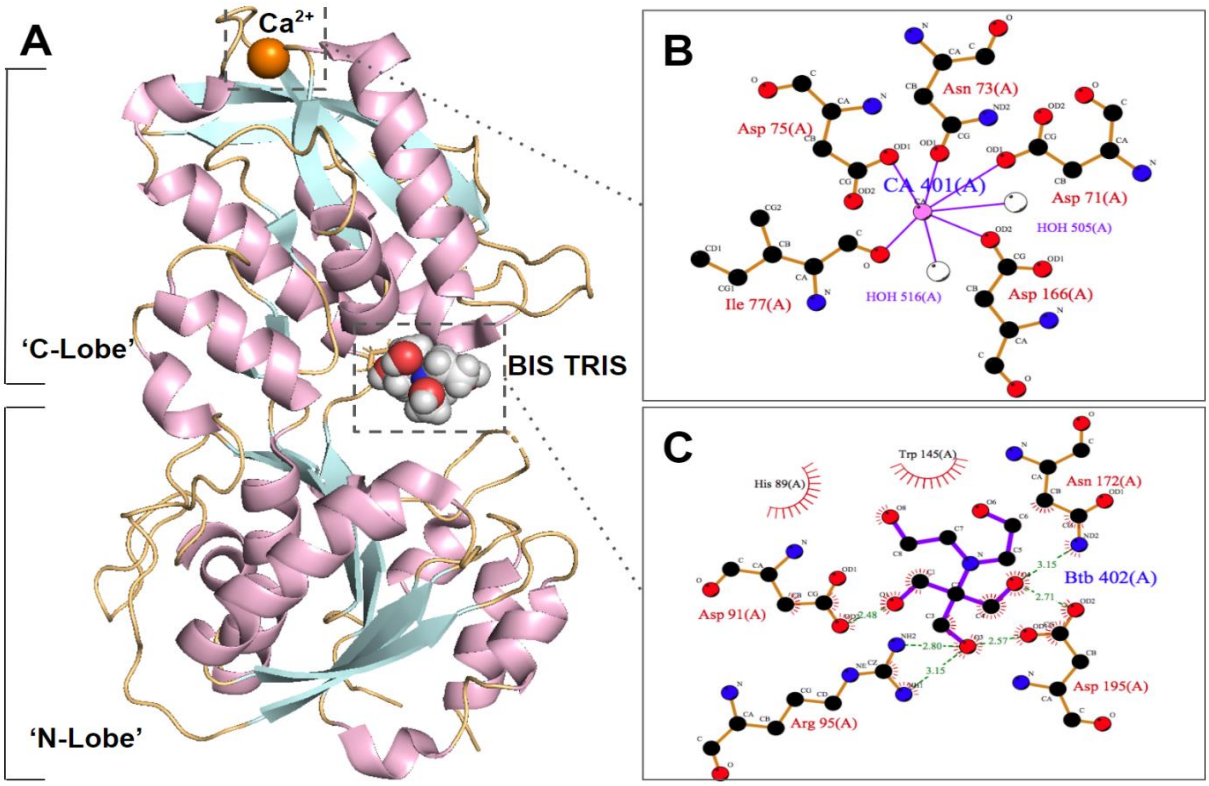


Figure 24: Structural representation of Tp0684. *A.* Structure in cartoon format colored according to secondary structure with **α**-helices in pink, **β**-sheets in cyan and connecting coil in light orange. Ca^{2+} is shown as orange sphere and BIS-TRIS is shown as sphere with oxygen in red, nitrogen in blue and carbogen in grey. *B.* The interaction of Ca^{2+} with residues of connecting coil, represented in 3D-structure (left) and 2D-structure right. *C.* The interaction of BIS-TRIS with residues of connecting coil represented in 3D-structure (left) and 2D-structure right.

Tp0684 adopts a bilobed fold and resembles ligand-binding proteins (LBPs) of other ABC-transporters. LPBs have two structurally similar lobes, connected by a hinge region and serve as nutrient and cofactor receptors in ABC-transporters.

By using the DALI database[24] and PDB code for structural homologs, Tp0684 reveals structural similarities with LBPs of ABC-transporters, in which EcMgIB is the most similar structural homolog. This is confirmed by the superposition of Tp0684 (dark pink) and EcMgIB (light yellow) (Fig. 25). The bound ligands are represented in spheres and show similar positions, which suggests that BIS-TRIS occupies the ligand-binding site of Tp34 in the crystal structure (Fig 24A). Besides, the Ca^{2+}-binding sites show similar positions.

Collectively, the structural data implies that Tp34 acts as a ligand-binding element of an ABC transporter. This suggests that Tp34 has a crucial role in the growth and chemotaxis of *T. pallidum* and is critical for the pathogenesis of *T. pallidum*.[18]

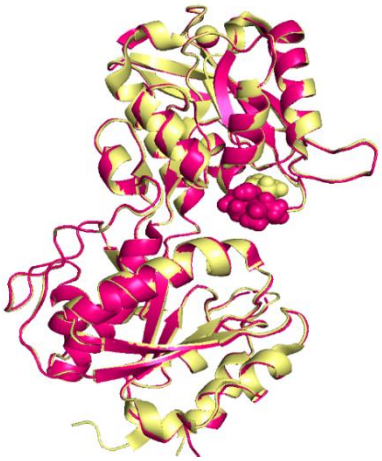


Figure 25: superposition of Tp0684 (pink) with EcMgIB (yellow).

Tp0971

The transition of metal ions is crucial for the homeostasis of many proteins. However, T. pallidum only encodes for a few metal-requiring proteins. *The* membrane lipoprotein Tp0971 (Tp34) is potent to participate in metal ion homeostasis, which makes this protein an auspicious target. Tp34 is possibly tethered to the inner membrane and provides human lactoferrin binding properties. Lactoferrin is a

mammalian iron-binding glycoprotein and is located on mucosal surfaces and within biological fluids, such as milk and saliva. However, further research is necessary to provide more lucidity about the structural and functional aspects of Tp34.

To clarify its potential role in lactoferrin binding properties, a crystal structure of Tp34 with a resolution of 1.9 Å is created using PyMOL (DeLano Scientific) (Fig. 26). The crystal structure is visualized as a cartoon format and the asymmetric unit of rTp34 consists of 2 monomers and each monomer consists of 2 antiparallel β-sheets. The monomeric structure is colored according to the secondary structure, with α-helices colored in green, β-sheets colored in violet and connecting coils in yellow (Fig. 26A). The bottom monomer in the dimeric structure is colored as in A and the other monomer is colored in light blue (Fig. 26B). The side chains interacting with the two interdimeric Zn^{2+} ions (orange spheres) are colored according to the monomer from which they belong. The structure and topology of rTP34 are identical to the classical immunoglobulin fold (Ig-fold) in several aspects. First, an extra interrupted β-strand is located in the back β-sheet near the N-terminus and is indicated as A1 and A2 (Fig. 26A). Besides, long regions of peptide between Ig-fold β-strands contain secondary structural elements and are indicated by lowercase letters (ab, bc and fg). These strands are isolated in the monomeric structure of rTP34 (Fig 26A), but participate in β-sheets in dimeric structure of rTp34 (Fig. 26B). In this way, the dimeric interface of rTp34 provides a zinc-binding site, also known as the 'protein interface zinc-site'.

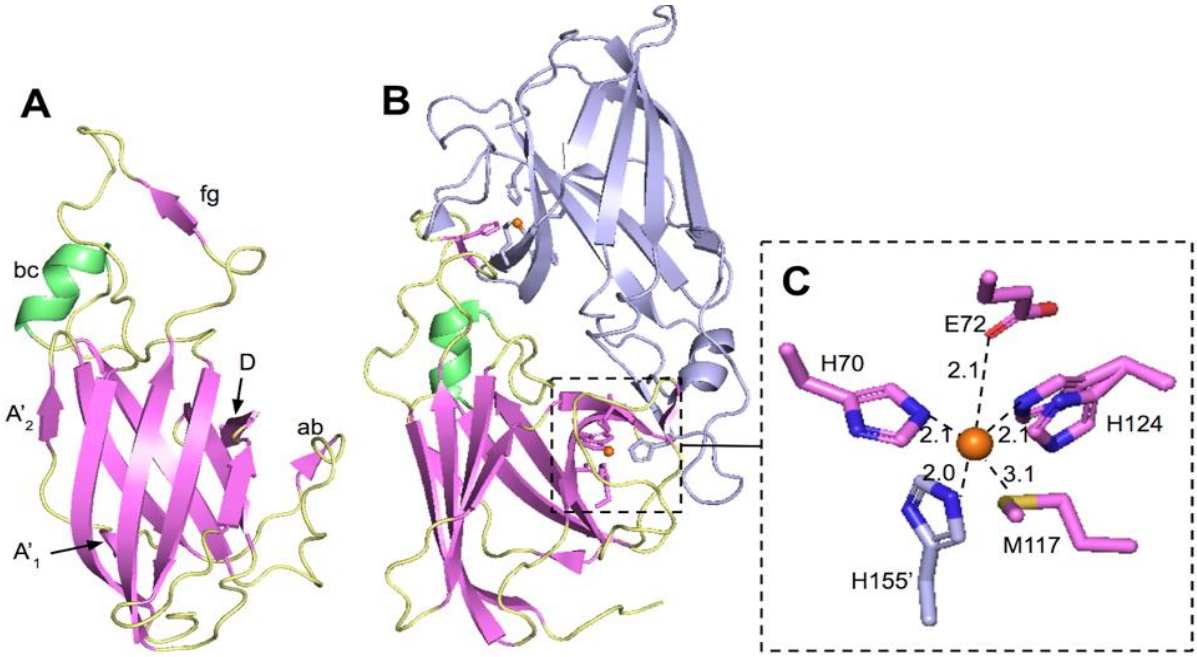


Figure 26: Crystal structure of Tp0971. *A.* Monomeric structure colored by secondary structure with α-helices in green, β-sheets in violet and connecting coils in yellow. *B.* Dimeric structure of Tp0971 with bottom monomer colored as in A and the other monomer colored in blue. The zinc-binding site is shown as sticks. *C.* Interactions of Zn^{2+} (orange) with residues colored according to the monomer to which they belong with nitrogen in blue, oxygen in red and sulfur in yellow.

A detailed structure of the Zn^{2+} bound at the dimeric interface represents the inner-sphere coordination environment (Fig 26C). The zinc ion and side chains are colored similar as in figure 26B. The cation shows interactions with one oxygen ligand (red), one sulfur ligand (yellow) and three nitrogen ligands (blue). These ligands arrange around the Zn^{2+} in a distorted square pyramidal geometry. The Zn^{2+} is surrounded by the rTp34 dimer and has no exposure to the solvent. The oxygen-Zn^{2+} and nitrogen-Zn^{2+} distances are 2.0-2.1 Å and suit with established norms for inner-sphere ligation. The distance between the sulfur of methionine and Zn^{2+} of 3.1 Å is remarkable because the sum of ionic radii of sulfur and Zn^{2+} is significantly shorter (2.4 Å). However, it is common that ligand-Zn^{2+} distances increase when the coordination number increases. The coordination of Met-117S and Zn^{2+} is a novel feature because currently no crystal structure with a similar Zn^{2+} -binding site exists in a metalloprotein database. As an alternative, Zn^{2+} could potentially bind to a site that binds to a

different metal ion in vivo. For example, both copper and iron ions are able to coordinate by the sulfur of methionine when bound to proteins. Besides, Zn^{2+} is able to occupy iron-binding sites.
Collectively, Tp34 shows a propensity to bind zinc and may be involved in lactoferrin binding. However, additional studies are required to confirm this interaction for iron utilization by *T. pallidum*.[19]

Tp0453

The lipoprotein Tp0453 (Tp34) is associated with the inner leaflet of the outer membrane of *T. pallidum* and shows property to integrate into membranes due to its multiple amphipathic helices (AHs). Besides, Tp34 increases membrane permeability, which suggests its porin-like role. Tp34 appears in open and closed conformation because the protein appears in membrane-integrated and -unintegrated forms.
To clarify its potential role in membrane integration and to better understand its physiological role in outer membrane biogenesis, crystal structures of different forms of Tp34 is created using PyMOL (DeLano Scientific) (Fig. 27). TP34 consists of a globular shape and adopts an α/β/α-fold consisting of 12-stranded antiparallel β-sheet and 9 α-helices. The protein is able to adapt an open (form I and II) and closed conformation (form III and IV). The structural change between closed to open conformation is provided by the breakage of hydrogen bonds between α2 (Asn65, His66) and α9 (Ser278) and between α7 (Lys200, Leu201) and α9 (Arg272) (Fig. 27).

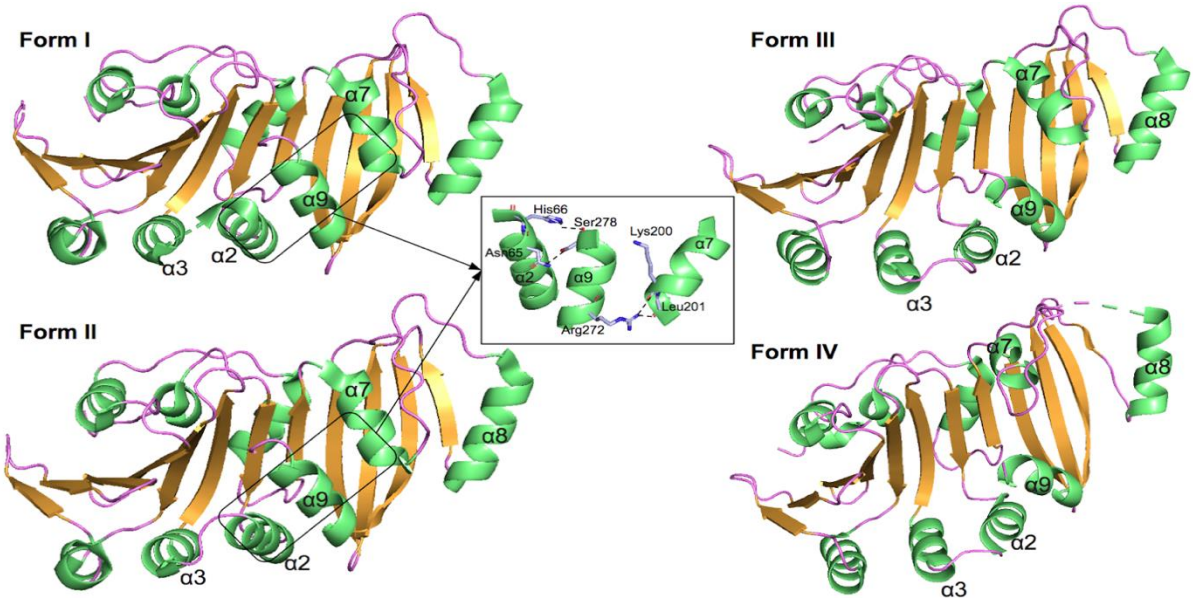


Figure 27: Crystal structures of different forms of Tp0453 in cartoon format colored according to secondary structure with **α**-helices in green, β-sheets in orange and connecting coils in pink. The hydrogen bonds between α2 (Asn65, His66) and α9 (Ser278) and between α7 (Lys200, Leu201) and α9 (Arg272) of form I and II are shown in black dashed lines.

Form I and II of the open conformation are structurally similar, despite the small deviation in the positions of α2 and α3. In contrast, forms I/II and III/IV show a structural change in two groups of helices (α2/α3/α4 and α7/α8), which move more laterally away from each other and expose the β-sheet in form III and IV. This will increase the exposure of β-strands 9, 10, and 12 and provide the floor of a large hydrophobic cavity (HC) (Fig. 28B). Form III and IV show structural change in α7 and α8, which move more outwards and away from each other in form IV, which increases the HC and the exposure to the aqueous milieu. The surface of the closed form contains a horseshoe-shaped hydrophobic patch (SHP), which is mostly provided by α8 and a small part of α7 (Fig. 28A). This structural information reveals that the membrane insertion of TP34 is provided by the distinctive orientation of α7 and α8, which suggests them to be membrane insertion elements. This emphasizes

the importance of TP34 whose amphiphilic behavior contributes to the outer membrane biogenesis of *T. pallidum*.[20]

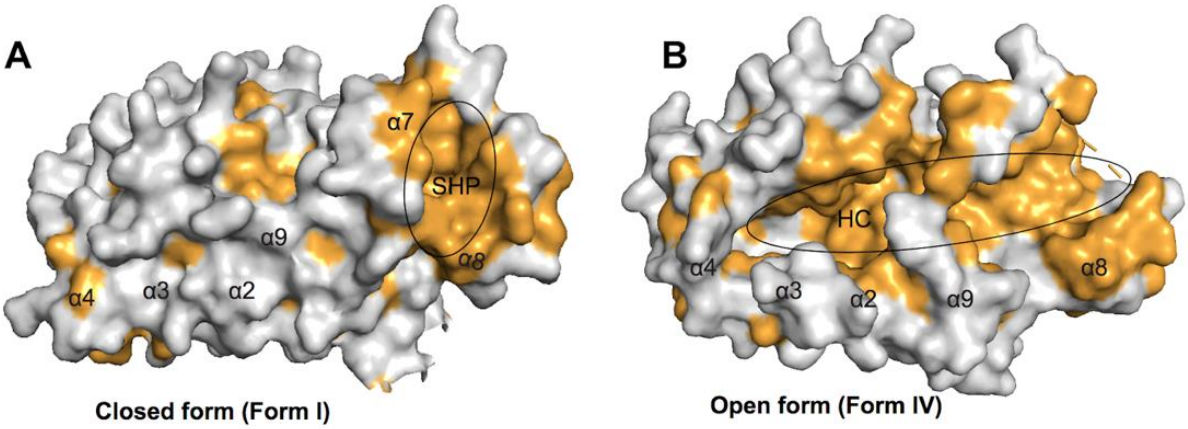


Figure 28: Surface representation of Tp0453 with hydrophobic regions colored in orange. *A.* Closed form (form I) shows a SHP. *B.* Open form (Form (IV) shows a HC.

Tp0655

T. Palladium encodes for the lipoprotein Tp0655, which is expected to fulfill the polyamine requirement. Sequence analysis of Tp0655 showed that this lipoprotein belongs to the bacterial PotDfamily, which are spermidine-preferring periplasmic ligand-binding protein (pLBP) of ABC transport systems. Spermidine is a polyamine containing a crucial role in cellular processes including cell growth, differentiation, and cell death. *T. pallidum* only contains one polyamine uptake system and is not able to synthesize polyamines, which emphasizes the importance of Tp0655.

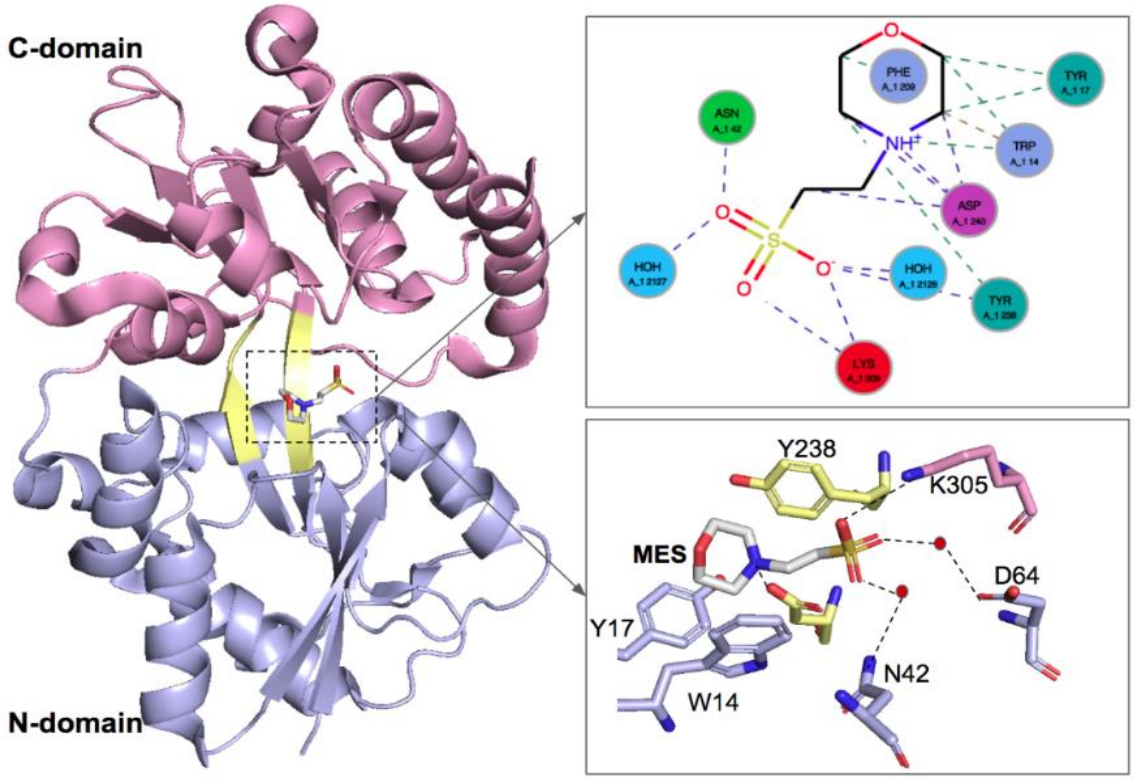


Figure 29: Crystal structure of Tp0655 colored by domain with C-domain in pink, N-domain in blue and connecting region in yellow. Mes is shown as yellow sticks with carbon in white, oxygen in red and nitrogen in blue. The interaction of Mes is represented in a LigPlot (above) and in a 3D-view (bottom).

To clarify the putative role of Tp0655 as a polyamine receptor, a crystal structure of Tp0655 is created using PyMOL (DeLano Scientific) (Fig. 29). The crystal structure is visualized in cartoon format and is colored by domain, in which the C-domain is pink, the N-domain blue, and the connecting region yellow. Tp0655 is divided into two globular domains and each domain contains a central five stranded β-sheet. The N-terminal domain contains 5 α-helices and the C-terminal domain contains 8 α-helices. The lipoprotein contains two bound molecules of 2-(N-morpholino)ethane-sulfonic acid (Mes), of which one is shown as a stick model (Fig. 29). The buffer component Mes is included in the crystallization medium. To clarify the interactions of Mes, a LigPlot and a close-up 3D-view of Mes are represented, in which the dashed black lines show hydrogen bonds or ionic interactions.

By using the DALI database[24] and PDB code for structural homologs, Tp0655 reveals structural similarities with the Pot families, in which PotD is the most similar structural homolog. As mentioned before, PotD plays a crucial role in the spermidine-specific uptake system. The structural similarity of the proteins is confirmed by the superposition of the Tp0655 (blue) and PotD (orange) (Fig. 30). The bound ligands are represented in sticks and show similar positions, which suggests that Mes occupies the ligand-binding site of Tp0655. The structural similarities between Tp0655 and PotD confirm the putative role of Tp0655 as a polyamine-binding component of an ABC-type polyamine transport system. This emphasizes the importance of Tp0655 in cellular processes of *T. pallidum* and provides more insight into the pathogenesis of *T. pallidum*.[21]

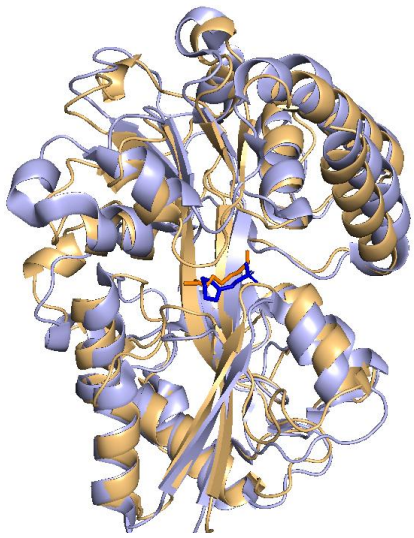


Figure 30: Superposition of Tp0655 (blue) and PotD (orange).

Tp0751

T. Pallidum encodes for the protein Tp0751, which is also referred to as pallilysin. Tp0751 is a primary target of opsonic antibodies, which implies that the protein is surface-exposed on *T. pallidum*. Furthermore, Tp0751 binds to extracellular matrix (ECM) components of the host, including laminin, fibronectin, and fibrinogen. This implies that Tp0751 has a propensity to interact with ECM components of the host and is a potential target to disrupt the interaction between *T. pallidum* and the host.

To clarify the underlying mechanism of Tp0751-host ECM attachment, a crystal structure of Tp0751 with a resolution of *2.15 Å* is created using PyMOL (DeLano Scientific) (Fig. 31). The C-terminal domain adopts an eight-stranded antiparallel compact β-barrel conformation and is capped by a longer N-terminal helix and a shorter 3_{10}-like helix. The β-barrel potentially serves as a hydrophobic ligand-binding pocket. The loops L1, L3, L5, and L7 provide the coordination of

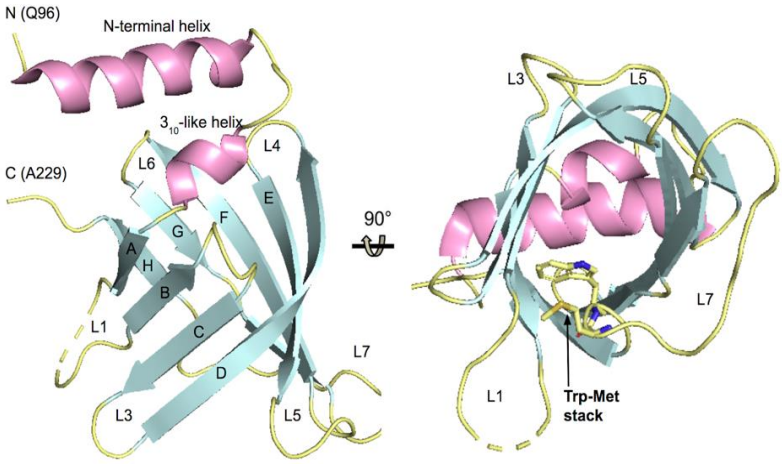


Figure 31: 3D-structure of Tp0751 in cartoon format, colored according the secondary structure with α-helices in pink, β-strands in cyan and connecting coil in yellow

various hydrophobic ligands for transport, catalyzation, or sequestration. The purpose of loop 1 is the provision of a cup-like binding site. However, L1 of Tp0751 is disturbed by L7, which caps the hydrophobic core with Trp and Met residues. Besides, the hydrophobic core is also capped by other several polar residues. This implies that the lipocalin domain of Tp0751 lacks a defined hydrophobic binding pocket and that Tp0751 is more likely to fulfill the host-pathogen protein-protein interaction. The typical β-barrel fold of Tp0751 suits with characteristics of a non-canonical lipocalin domain.

By using the DALI server, Tp0751 reveals indeed structural similarity with lipocalin domains, in particular with Nitrophorin 4 (PDB ID 1KOI). This is confirmed by the structural similarity of the crystal structures of Tp0751 and Nitrophorin 4 (Fig. 32). However, nitrophorin 4 contains disulfide bonds in the N- and C-terminal regions (indicated in green), which lack in Tp0754. Even though Tp0751 and Nitrophorin 4 only have 6% of sequence identity in common, low sequence identity is a common property of lipocalins.

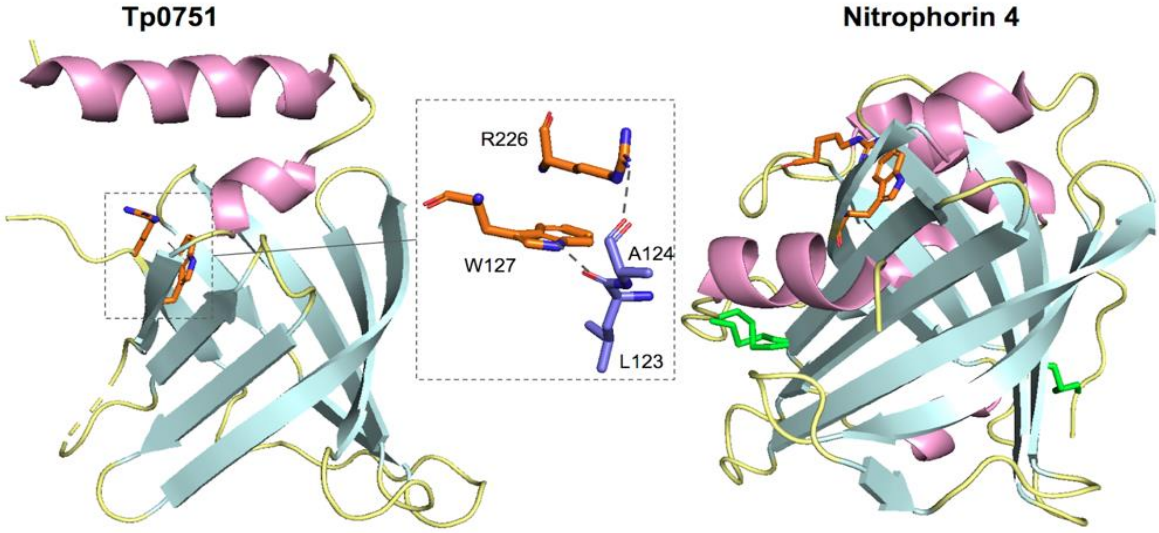


Figure 32: Structural comparison of Tp0751 (left) and Nitrophorin 4. The conserved regions are indicated in orange sticks and the disulfide bonds are indicated in green sticks. The interaction of SCR1 and SCR3 with Tp0751 residues are shown in detail with hydrogen bonds in black dashed lines.

Other distinct features of lipocalins are central β-barrel and consistency of three short conserved regions (SCR1, SCR2, and SCR3), which are indicated as orange sticks in the crystal structures of Tp0751 and Nitrophorin 4. Tp0751 only contains SCR1 and SCR3 regions and lacks the SCR2 region. The SCR1 is located near β-strand A and the 3_{10}-like α-helix and contains the key basic residue W127, while SCR3 is located near strand H and contains the key residue R226. The key residues W127 and R226 form a hydrogen bond with the backbone carbonyl residues L123 and A124, respectively. However, Tp0751 does not contain SCR2, which is usually located in strand F-loop6-Strand G region in other bacterial lipocalins and potentially provides the incorporation of a metal-binding motif. So, Tp0751 does not reveal structurally characteristics for metal coordination. Although structurally similar lipocalins show a tendency for similar functions, it is improbable that Tp0751 shares a similar function with nitrophorins. This is since Tp0751 lacks SCR2, which is required for metal coordination and transport. Collectively, the structural data provide more insight into the putative role of Tp0751 in the disruption of the treponemal-host interaction.[22]

Tp0737

T. pallidum acquires the majority of its nutrients from the host cell, which is mostly facilitated via ABC-type transporters. These systems contain membrane-tethered ligand-binding proteins (LBPs), which enables the binding of small-molecule compounds via the action of transmembrane permeases. The polypeptide Tp0737 belongs to a class of periplasmic ligand-binding proteins and is associated with these ABC-type transporters. Tp0737 shares sequence homology with sugar-binding proteins of ABC-type transporters. However, Tp0737 does not contain the N-terminal sequence signatures, which are necessary for the post-translational modification of lipoproteins. Instead, it contains an N-terminal sequence that potentially targets the protein to the periplasm. To gain more insight into the function of Tp0737, structural biology is applied.

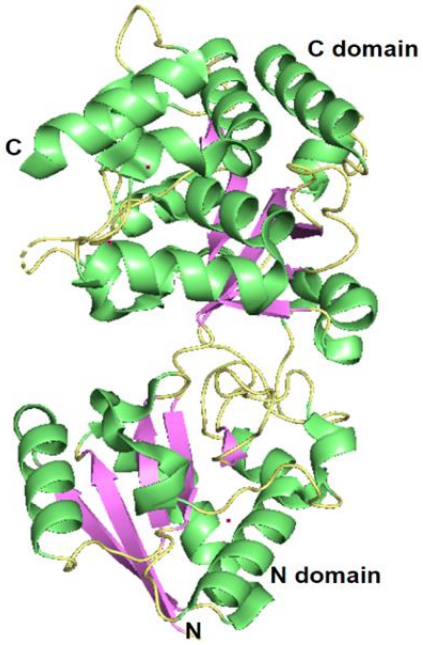


Figure 33: Crystal structure of rTp0737 in cartoon format, colored according to secondary structure with α-helices in green, β-sheets in violet and connecting coils in yellow.

The crystal structure of a recombinant fragment Tp0737 without its N-terminal sequence (rTP0737) is created at 1.8 *Å* resolution using PyMOL (DeLano Scientific) (Fig. 33). The structure of Tp0737 reveals a characteristic fold of periplasmic ligand-binding proteins which are associated with ABC-type transporters. The protein rTp0737 consists of an N-domain (residues 40-142 and 315-360) and a C-domain (residues 146-312 and 364-429), which are connected via three cross-over regions and form a hinge between the two domains. A large, solvent-filled cleft is shown between the two domains. This cleft does not provide electron density for a specifically bound ligand nor anywhere else in the structure. This suits with results of the DALI server[24] supplying protein homology detection, which revealed closest structural matches to rTP0737 with unliganded LBPs. When LBPs do not contain a bound ligand, they show a tendency to form an open-cleft in solution. So, the open form of rTp0373 is consistent with its lack of bound ligand.

To investigate the cognate ligand of Tp0737, a putative binding pocket is analyzed by surface features of rTP03737 (Fig. 34). The respective N and C domains are individually rotated with ligand-binding surfaces facing forward. The cleft is not predominated by certain amino acid and no substantial areas of hydrophobicity are found in the cleft. Instead, the cleft is having an abundance of hydrogen-bond donors and acceptors. This suggests that the potential ligands are likely to be largely polar. Collectively, the nature of the native ligand of rTp0737 remains obscure, but the rTp0737 structure provides further insight into ligand discovery. More research is necessary to provide the identity of the ligand and thus more insight into the function of Tp0737.[23]

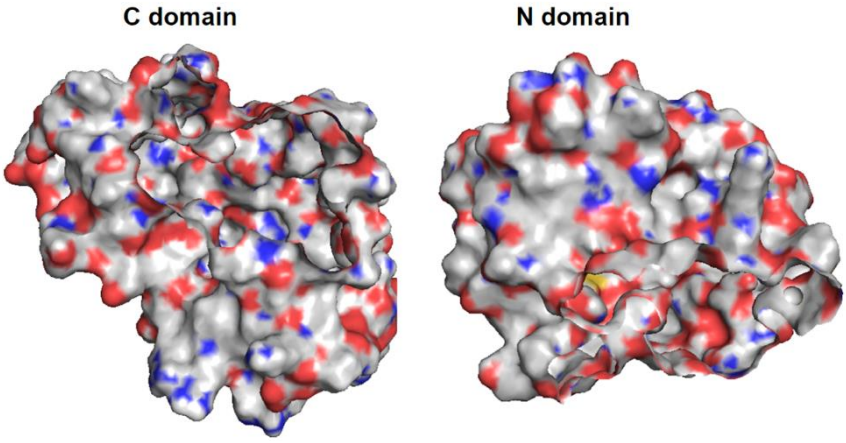


Figure 34: Surface representation of the C-domain (left) and N-domain (right) of rTp0737 with carbogens indicated in grey, nitrogens in blue, oxygens in red and sulfurs in yellow.

Summary

Collectively, different structures of potential protein targets of *C. trachomatis* and *T. Pallidum* are visualized and analyzed by PyMOL. The three-dimensional structures of proteins have illustrated how proteins are assembled, how they interact, and how they function. Several structural homologs of proteins are detected and ligand-protein interactions are analyzed. The structural information elucidates a clearer understanding of disease mechanism and leads to possibilities in the discovery of potential protein targets to combat a chlamydia or syphilis infection. To summarize, a clear overview of potential protein targets of *C. trachomatis* and *T. pallidum* is listed below.

CT263 shows structural similarity to MTANs, which implies its nucleosidase enzymatic role and its provision of quinone synthesis via the futalosine pathway.
ChlaDUB1 is able to fulfill the distinct chemical reactions deubiquitination and acetylation, which provide the alteration of signaling pathways principal for invasion, survival, and replication of *C. trachomatis*.
Pgp3 resembles Bc2L-C trimer, which implies its role in bioadhesion and emphasizes its importance in the pathogenesis of chlamydia.
DsbH shows structural similarities to DsbD and thioredoxin, which implies its potential role in the reduction of thiol-disulfide oxidoreductase in chlamydia.
CT584 shows structural similarity with Cpn0803, which implies the highly conserved nature of CT584 across chlamydial species.
IncA contains a clamp that provide the homotypic fusion of inclusions and has a crucial role in chlamydia pathogenicity.
CT771 belongs to the Nudix family and cleaves Ap4a into ATP and AMP. This is confirmed by the structural alignment of CT771 and H. Sapiens Ap4A hydrolase.
CT441 provides proteolytic activity against SRAP1, which is a coactivator of the estrogen receptor. This makes CT441 a very exciting target for future research into chlamydial pathogenicity mechanisms.
CPAF can be considered as Ser protease, of which the catalysis is provided by a water-mediated catalytic triad. This provides more insight into the development of specific inhibitors to regulate CPAF activity.

TP0453 consists of a covalently disulfide-stabilized dimer in solution, which suggests its potential role in ligand binding. Additional studies are necessary to clarify the potential ligand-binding tendency of Tp0453.
TP0624 contains a putative binding site which is likely to anchor to the thin periplasmic peptidoglycan layer of *T. pallidum*, which stabilized the cell envelope. This implies Tp0624 its crucial role in the pathogenesis of *T. pallidum*.
Tp0684 serves as a ligand-binding element of an ABC transporter, which emphasized its crucial role in the growth and chemotaxis of *T. pallidum*.
Tp0971 reveals a propensity to bind zinc and is potentially involved in lactoferrin binding. Additional studies are required to confirm this interaction for iron utilization by *T. pallidum*.
Tp0453 contains a porin-like role and shows the ability to increase membrane permeability by its amphiphilic behavior. This contributes to the outer membrane biogenesis of T. Pallidum and emphasizes its crucial role in the pathogenesis of *T. pallidum*.
Tp0655 reveals structural similarities with PotD, which suggests its crucial role in the spermidine-specific uptake system. This confirms its putative role as a polyamine binding component of an ABC-type transport system.
Tp0751 shows a propensity to interact with ECM components of the host and is a potential target to disrupt the interaction between *T. pallidum* and the host.
Tp0737 is associated with ABC-type transporters in *T. pallidum* and potentially interact with largely polar ligands.

Literature

1. *Detailed STD Facts - Chlamydia*. (2019). Retrieved from https://www.cdc.gov/std/chlamydia/stdfact-chlamydia-detailed.htm#:%7E:text=Chlamydia%20is%20the%20most%20frequently,infection%20in%20the%20United%20States.&text=In%202018%2C%201%2C758%2C668%20cases%20of,2.86%20million%20infections%20occur%20annually

2. Elwell, C., Mirrashidi, K., & Engel, J. (2016). Chlamydia cell biology and pathogenesis. *Nature Reviews Microbiology*, *14*(6), 385–400. https://doi.org/10.1038/nrmicro.2016.30
3. Guerra, L. O., Boga, J. A., Suárez, J. F., Benítez, C. F., & Vázquez, F. (2015). Pathogenesis of Chlamydia trachomatis in Humans. *Human Emerging and Re-emerging Infections*, 635–648. https://doi.org/10.1002/9781118644843.ch34

4. Huai, P., Li, F., Li, Z., Sun, L., Fu, X. '., Pan, Q., … Zhang, F. (2018). Prevalence, risk factors, and medical costs of Chlamydia trachomatis infections in Shandong Province, China: a population-based, cross-sectional study. *BMC Infectious Diseases*, *18*(1), 6–7. https://doi.org/10.1186/s12879-018-3432-y

5. LaFond, R. E., & Lukehart, S. A. (2006). Biological Basis for Syphilis. *Clinical Microbiology Reviews*, *19*(1), 29–49. https://doi.org/10.1128/cmr.19.1.29-49.2006

6. Peeling, R. W., Mabey, D., Kamb, M. L., Chen, X.-S., Radolf, J. D., & Benzaken, A. S. (2017). Syphilis. *Nature Reviews Disease Primers*, *3*(1), 2–4. https://doi.org/10.1038/nrdp.2017.73

7. Barta, M. L., Thomas, K., Yuan, H., Lovell, S., Battaile, K. P., Schramm, V. L., & Hefty, P. S. (2014). Structural and Biochemical Characterization of Chlamydia trachomatis Hypothetical Protein CT263 Supports That Menaquinone Synthesis Occurs through the Futalosine Pathway. *Journal of Biological Chemistry*, *289*(46), 32214–32229. https://doi.org/10.1074/jbc.m114.594325

8. Pruneda, J. N., Bastidas, R. J., Bertsoulaki, E., Swatek, K. N., Santhanam, B., Clague, M. J., … Komander, D. (2018). A Chlamydia effector combining deubiquitination and acetylation activities induces Golgi fragmentation. *Nature Microbiology*, *3*(12), 1377–1384. https://doi.org/10.1038/s41564-018-0271-y

9. Galaleldeen, A., Taylor, A. B., Chen, D., Schuermann, J. P., Holloway, S. P., Hou, S., … Hart, P. J. (2013). Structure of the Chlamydia trachomatis Immunodominant Antigen Pgp3. *Journal of Biological Chemistry*, *288*(30), 22068–22079. https://doi.org/10.1074/jbc.m113.475012

10. Mac, T.-T., von Hacht, A., Hung, K.-C., Dutton, R. J., Boyd, D., Bardwell, J. C. A., & Ulmer, T. S. (2007). Insight into Disulfide Bond Catalysis in Chlamydiafrom the Structure and Function of DsbH, a Novel Oxidoreductase. *Journal of Biological Chemistry*, *283*(2), 824–832. https://doi.org/10.1074/jbc.m707863200

11. Barta, M. L., Hickey, J., Kemege, K. E., Lovell, S., Battaile, K. P., & Hefty, P. S. (2013). Structure of CT584 from Chlamydia trachomat is refined to 3.05 Å resolution. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, *69*(11), 1196–1201. https://doi.org/10.1107/s1744309113027371

12. Cingolani, G., McCauley, M., Lobley, A., Bryer, A. J., Wesolowski, J., Greco, D. L., … Paumet, F. (2019). Structural basis for the homotypic fusion of chlamydial inclusions by the SNARE-like protein IncA. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-10806-9

13. Barta, M. L., Lovell, S., Sinclair, A. N., Battaile, K. P., & Hefty, P. S. (2013). Chlamydia trachomatisCT771 (nudH) Is an Asymmetric Ap4A Hydrolase. *Biochemistry*, *53*(1), 214–224. https://doi.org/10.1021/bi401473e

14. Kohlmann, F., Shima, K., Hilgenfeld, R., Solbach, W., Rupp, J., & Hansen, G. (2014). Structural Basis of the Proteolytic and Chaperone Activity of Chlamydia trachomatis CT441. *Journal of Bacteriology*, *197*(1), 211–218. https://doi.org/10.1128/jb.02140-14

15. Huang, Z., Feng, Y., Chen, D., Wu, X., Huang, S., Wang, X., … Chai, J. (2008). Structural Basis for Activation and Inhibition of the Secreted Chlamydia Protease CPAF. *Cell Host & Microbe*, *4*(6), 529–542. https://doi.org/10.1016/j.chom.2008.10.005

16. Brautigam, C. A., Deka, R. K., Liu, W. Z., & Norgard, M. V. (2014). Insights into the potential function and membrane organization of the TP0435 (Tp17) lipoprotein fromTreponema pallidumderived from structural and biophysical analyses. *Protein Science*, *24*(1), 11–19. https://doi.org/10.1002/pro.2576

17. Parker, M. L., Houston, S., Wetherell, C., Cameron, C. E., & Boulanger, M. J. (2016). The Structure of Treponema pallidum Tp0624 Reveals a Modular Assembly of Divergently Functionalized and Previously Uncharacterized Domains. *PLOS ONE*, *11*(11), e0166274. https://doi.org/10.1371/journal.pone.0166274

18. Barta, M. L., Hickey, J., Kemege, K. E., Lovell, S., Battaile, K. P., & Hefty, P. S. (2013). Structure of CT584 fromChlamydia trachomatisrefined to 3.05 Å resolution. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, *69*(11), 1196–1201. https://doi.org/10.1107/s1744309113027371

19. Deka, R. K., Brautigam, C. A., Tomson, F. L., Lumpkins, S. B., Tomchick, D. R., Machius, M., & Norgard, M. V. (2006). Crystal Structure of the Tp34 (TP0971) Lipoprotein ofTreponema pallidum. *Journal of Biological Chemistry*, *282*(8), 5944–5958. https://doi.org/10.1074/jbc.m610215200

20. Luthra, A., Zhu, G., Desrosiers, D. C., Eggers, C. H., Mulay, V., Anand, A., … Radolf, J. D. (2011). The Transition from Closed to Open Conformation ofTreponema pallidumOuter Membrane-associated Lipoprotein TP0453 Involves Membrane Sensing and Integration by Two Amphipathic Helices. *Journal of Biological Chemistry*, *286*(48), 41656–41668. https://doi.org/10.1074/jbc.m111.305284

21. Machius, M., Brautigam, C. A., Tomchick, D. R., Ward, P., Otwinowski, Z., Blevins, J. S., … Norgard, M. V. (2007). Structural and Biochemical Basis for Polyamine Binding to the Tp0655 Lipoprotein of Treponema pallidum: Putative Role for Tp0655 (TpPotD) as a Polyamine Receptor. *Journal of Molecular Biology*, *373*(3), 681–694. https://doi.org/10.1016/j.jmb.2007.08.018

22. Parker, M. L., Houston, S., Pětrošová, H., Lithgow, K. V., Hof, R., Wetherell, C., … Boulanger, M. J. (2016). The Structure of Treponema pallidum Tp0751 (Pallilysin) Reveals a Non-canonical Lipocalin Fold That Mediates Adhesion to Extracellular Matrix Components and Interactions with Host Cells. *PLOS Pathogens*, *12*(9), e1005919. https://doi.org/10.1371/journal.ppat.1005919

23. Brautigam, C. A., Deka, R. K., Liu, W. Z., Tomchick, D. R., & Norgard, M. V. (2017). Functional clues from the crystal structure of an orphan periplasmic ligand-binding protein fromTreponema pallidum. *Protein Science*, *26*(4), 847–856. https://doi.org/10.1002/pro.3133

24. Holm, L. (2010). *DALI protein structure comparison server*. DALI and the persistence of protein shape. http://ekhidna2.biocenter.helsinki.fi/dali/