# STRUCTURAL AND FUNCTIONAL STUDIES OF THREE PROTEINS OF UNKNOWN FUNCTION ENCODED BY *CHLAMYDIA TRACHOMATIS*

By

Kyle E. Kemege

Submitted to the graduate degree program in Molecular Biosciences and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____
Chairperson (P. Scott Hefty, Ph.D.)

_____*
(Audrey Lamb, Ph.D.)

_____*
(Susan Egan, Ph.D.)

_____*
(Yoshiaki Azuma, Ph.D.)

_____*
(Roberto De Guzman, Ph.D.)

_____*
(Emily Scott, Ph.D.)

*Committee Members

Date Defended: 4/4/2013

The Dissertation Committee for Kyle E. Kemege certifies that this is the approved version of the following dissertation:

STRUCTURAL AND FUNCTIONAL STUDIES OF THREE PROTEINS OF UNKNOWN FUNCTION ENCODED BY *CHLAMYDIA TRACHOMATIS*

_____

Chairperson (P. Scott Hefty, Ph.D.)

Date approved: 4/17/2013

## ABSTRACT

Infections by chlamydial species are of significant impact to global public health, causing sexually transmitted infections, blinding trachoma and pneumonia. Despite its importance, there are many aspects of chlamydial biology that are not completely understood, including the mechanisms by which it infects, persists and replicates in its host cells. The reason for this ignorance of basic chlamydial biological processes is because there is an abundance of Open Reading Frames (ORFs) of unknown function present in chlamydial genomes, almost 30% of the entire genome in many species. This is likely due to the relatively large phylogenetic distance between *Chlamydiae* and better-understood bacteria such as *E. coli* and *B. subtilis*. Current strategies of genome annotation rely on the presence of homology to genes of known function and these approaches have not been effective in annotating chlamydial genomes.

In an effort gain insight into the function of these chlamydial ORFs of unknown function, I utilized structural information (both computational and experimentally derived) about three proteins of interest. Based on these structural studies, hypotheses concerning the functions of these proteins were formed and then tested. Together, my findings not only provide valuable information about these proteins of unknown function, but they also serve to demonstrate both the strengths and shortcomings of the overall approach of utilizing structural information for functional prediction.

One example of this approach is my work on the chlamydial ORF CT296. Although this protein was annotated as having an unknown function (due to insignificant homology to proteins of known function), it had been experimentally characterized as an iron-dependent transcription factor. Having an experimentally characterized function

allowed me to test my approach of utilizing structural information to predict function on a protein with a relatively well-understood function. Surprisingly, structural information of this protein suggested that it functions as a Fe(II) 2-oxoglutarate-dependent enzyme and not as a transcription factor. Subsequent functional analyses of the protein were unable to reproduce previous reports of its DNA binding. Together, my findings suggest that this protein may not function as a transcription factor.

A second example of my structure-function approach was applied to the chlamydial protein CT584. This protein was first experimentally described as interacting with the chlamydial Type Three Secretion System (T3SS) needle protein in an interactome study. This observation, combined with a subsequent biophysical characterization of the protein lead to an initial hypothesis that the protein may be a chlamydial T3SS needle tip protein. However, results of structural studies on the protein reveal a structure that is not similar to any of the known T3SS needle tip proteins. Additionally, functional studies on the protein focusing on identifying its localization on chlamydial organisms revealed localization patterns not consistent with its proposed role as a T3SS needle protein. Together, my studies suggest that this protein may not function as a needle tip protein.

A final example of the utility of structural information for informing function concerns chlamydial ORF CT009. This protein was annotated in many chlamydial species as a protein of unknown function; however, bioinformatics analyses had identified it as a helix-turn-helix containing transcription factor. Results of computational and experimental structures of this protein show structural similarity to the protein RodZ, a key component of the bacterial morphogenic complex. Subsequent

functional analyses of CT009 demonstrate that this protein has the characteristics of a chlamydial homolog to RodZ.

## ACKNOWLEDGEMENTS

I would like to acknowledge my mentor Dr. Scott Hefty for his help and support throughout my graduate career. I would also like to thank all of the members of the Hefty lab, past and present, for both their advice and for making the lab environment more pleasant. I would also like to thank the members of my committee for their constant critical analysis and scrutiny of my research.

## TABLE OF CONTENTS

# Chapter I. General Introduction

## Public Health Impact of *Chlamydia*

Human infections by species of Chlamydia have been recorded since antiquity and the relationship between humans and Chlamydia likely predates civilization (Taylor, 2008). Human disease caused by species of Chlamydia is caused primarily by the chlamydial species *Chlamydia trachomatis* and *Chlamydia pneumoniae*. *C. trachomatis* can cause ocular infections that can lead to blindness and genital infections that can lead to infertility and/or complications in pregnancy; *C. pneumonia* can cause lung infections that lead to pneumonia (Centers, 2010, Hammerschlag, 2000, Resnikoff *et al.*, 2004).

Members of the species *Chlamydia trachomatis* can be further subdivided into serovars, based on differences in the sequence of their Major Outer Membrane Protein (MOMP) (Choroszy-Krol *et al.*, 2012). These serovar groupings also determine what tissue these organisms are most able to infect. Serovars A, B and C of *Chlamydia trachomatis* infect the ocular epithelium and can lead to blinding trachoma (Hu *et al.*, 2010, Hvidsten *et al.*, 2009). It has been estimated that in 2002 and 2003, approximately 40.6 million individuals worldwide had ocular *C. trachomatis* infections, 8.2 million of which had early symptoms of trachoma and 3.2 million of which were blind due to trachoma (Mariotti *et al.*, 2009, Resnikoff et al., 2004). This makes infection by these serovars of *C. trachomatis* the leading cause of infectious blindness in the world (Hu et al., 2010). Successful treatment and curing of ocular infection by *C. trachomatis* can be achieved by a single dose of azithromycin (Bailey *et al.*, 1993). However, for this to

prevent blindness, it must take place in the relatively early stages of disease, before inversion of the eyelid (trichiasis) occurs.

Serovars D through K of *Chlamydia trachomatis* are responsible for sexually transmitted genital infections (Bosnjak *et al.*, 2012). In 2009, 1.2 million cases of *C. trachomatis* genital infection were reported in the United States, making this infection the number one most reported infection in the U.S. (Centers, 2010). Numbers of unreported infections are likely even higher because genital infection by *C. trachomatis* is asymptomatic in ~50% of men and ~70% of women (Cecil *et al.*, 2001, Falk *et al.*, 2005). When symptoms are present, they include dysuria and nongonococcal urethritis in men and pelvic inflammatory disease (which may lead to ectopic pregnancy or involuntary infertility) in women (Centers, 2010). Treatment of infection with azithromycin or doxycycline is 92-98% effective (Lau & Qureshi, 2002, Batteiger *et al.*, 2010). However, a major problem with this treatment is that infection is so frequently asymptomatic and therefore is not diagnosed or treated (Cecil et al., 2001, Falk et al., 2005).

Serovars L1, L2 and L3 of *Chlamydia trachomatis* also cause genital infections but are also capable of a more severe infection (Kapoor, 2008). These Lymphogranuloma Venereum (LGV) strains are also capable of infecting macrophages, leading to invasive lymphangitis/lymphadenitis (Spaargaren *et al.*, 2005). Clinically, infection by these serovars is rare and infection can be treated with doxycycline, similarly to other *C. trachomatis* genital infections (Workowski & Berman, 2010).

Infection by *Chlamydia pneumoniae* is responsible for an estimated 6-22% of cases of community-acquired pneumonia (Hammerschlag, 2000, Kumar & Hammerschlag, 2007). Importantly, the seroprevalence of this organism is very high; 70-

80% of individuals are exposed to this organism within their lifetimes (Grayston, 1992). Fortunately, *C. pneumoniae* is susceptible to the common antibiotics prescribed for atypical pneumonia: azithromycin, beta-lactam antibiotics and/or levofloxacin.

Finally, there are several species of Chlamydia that are primarily animal pathogens. These include *C. muridarum* (infects mice; (Carey *et al.*, 2013)), *C. suis* (infects pigs; (Englund *et al.*, 2012)), *C. felis* (infects cats; (Laroucau *et al.*, 2012)), *C. pecorum* (infects sheep, cattle and horses; (Fukushi & Hirai, 1993)), *C. caviae* (infects guinea pigs; (Read *et al.*, 2003)) and *C. psittaci* (infects birds; (Voigt *et al.*, 2012)). Among these, *C. psittaci* is of the most importance from a public health standpoint; this zoonotic pathogen can also spread to humans, causing psittacosis (Harkinezhad *et al.*, 2009). Although incidence of this disease is rare, it is difficult to diagnose due to its broad range of symptoms and is potentially fatal (Rohde *et al.*, 2010). Due to its high rates of infectivity and its ability to be aerosolized, the CDC has classified C. *psittaci* as a category B bioterrorism agent (CDC, 2007).

All infections by species of Chlamydia can be successfully treated with antibiotics (Hammerschlag & Kohlhoff, 2012). Despite the relative availability of antibiotics, the threat and damage caused by chlamydial species has clearly not been eradicated. Particularly with genital infections, infection is often asymptomatic; therefore, treatment is often not administered until long-term damage has already been done. Additionally, re-infection by both ocular and genital serovars of *C. trachomatis* is common, requiring a continued administration of antibiotics (Cooksey *et al.*, 2010). An effective vaccine against *C. trachomatis* could remedy these problems. Such a vaccine does not yet exist but its development is a high priority (Cochrane *et al.*, 2010).

## Basic Biology of *Chlamydia*

Despite their differences in tissue tropism and disease pathology, all chlamydial species share an obligate intracellular lifestyle and a unique biphasic developmental cycle (Abdelrahman & Belland, 2005). As obligate intracellular parasites, chlamydial species do not grow or replicate outside of host cells. These species therefore rely on infection for their survival. Infection, growth and replication take place as part of a developmental cycle. In this cycle, chlamydial organisms alternate between two distinct forms termed Elementary Bodies (EBs) and Reticulate Bodies (RBs) (Abdelrahman & Belland, 2005). EBs are metabolically inactive and are capable of initiating infection while RBs are metabolically active but cannot infect cells (Abdelrahman & Belland, 2005).

Initiation of the developmental growth cycle begins with attachment of an EB to a suitable host cell. One example attachment factor is the heparan sulfate (HS) polysaccharide, which has been proposed to be involved in a trimolecular interaction between OmcB, a chlamydial outer membrane protein, and a host cell receptor, thereby linking the two cells together (Zhang & Stephens, 1992). However, multiple factors on both the surface of chlamydial EBs and the host cell are likely involved in attachment.

Now attached to the host cell, chlamydial EBs are taken up by a process termed "parasite-specific endocytosis" (Byrne & Moulder, 1978). This process is induced primarily through a chlamydial Type III Secretion System (T3SS; discussed in more detail later in this chapter) via the secreted effector protein Tarp, a Translocated actin

recruiting phosphoprotein (Clifton *et al.*, 2004). Once internalized in the host cell, Tarp induces actin polymerization, leading to endocytosis of the EB (Jewett *et al.*, 2006).

Having entered the host cell within a vacuole, chlamydial EBs respond to an unknown signal and begin their conversion to metabolically active RBs. This process takes place at approximately 2-6 hours post infection (HPI) and involves a disruption in the EBs of the interactions between histone-like proteins and their chromosomes that prevent active transcription (Perara *et al.*, 1992). This conversion process also involves reduction of the highly crosslinked outer membrane that serves to protect EBs in the extracellular environment. Additionally, the endocytic vacuole is modified by chlamydia-expressed Inc proteins, which prevent fusion to lysosomes and lead to the interception of Golgi-derived vesicles for energy (Fields & Hackstadt, 2002). Modified thusly, the vacuole is termed an inclusion.

At approximately 6 HPI, RBs begin to grow and replicate by binary fission with a doubling time of approximately 3 hours (Mathews *et al.*, 1999). Also during this time, proteins are secreted by RBs into the inclusion and into the host cell cytoplasm, in order to continue their subversion of host cell processes. For example, one such secreted protein is named CPAF (Chlamydial Protease-like Activity Factor) and is secreted into the host cell cytoplasm where it degrades many different proteins, ultimately leading toward an inhibition of apoptosis and an evasion of immune system recognition (Zhong, 2009).

At approximately 18-24 HPI, RBs begin to asynchronously convert to EBs. The signals that trigger this conversion are not well understood but the process involves condensation of the genome by histone-like proteins and disulfide crosslinking of the

outer membrane. By 48 HPI, the developmental cycle is complete and *chlamydiae* are released from the host cell by either lysis or a process called extrusion (Hybiske & Stephens, 2007). Because conversion of RB to EB is an asynchronous process, both EBs and RBs are present in the inclusion upon release. Released RBs are non-infectious but released EBs proceed to infect neighboring cells, starting a new round of the developmental cycle.

Despite its importance and conservation among chlamydial species, there are many aspects of this complex developmental cycle that are poorly understood. These include the precise factors involved in EB attachment to host cells and their order of events. Additionally, the events that trigger conversion from EB to RB and conversely RB to EB are not completely understood.

## Protein Structure-Function Relationships

As discussed above, there are many key aspects of chlamydial biology that are not fully understood, including its ability to invade cells, evade immune function and regulate its complex developmental cycle. One reason for our current lack of understanding is that species of chlamydia encode an abundance of hypothetical proteins, also known as Open Reading Frames (ORFS) of unknown function or proteins of unknown function. Depending on the species, these ORFs of unknown function total approximately 25-35% of the total number of genes in the genomes (Carlson *et al.*, 2005, Stephens *et al.*, 1998, Thomson *et al.*, 2008, Unemo *et al.*, 2010, Van Lent *et al.*, 2012).

The reason a function cannot be assigned to these proteins has to do with the method by which these (and most) genomes are annotated. All chlamydial genomes were annotated by using primary sequence homology to assign function. This means that an ORF with a sequence that is similar to a different ORF (from another organism) of known function will be annotated as having the function of that similar ORF. Conversely, proteins without significant sequence similarity to proteins of known function will be annotated as having an unknown function. This annotation method results in an abundance of proteins of unknown function for chlamydial genomes because of the phylogenetic distance between chlamydial species and other, better-characterized bacterial systems. Most proteins with experimentally characterized function come from model bacterial systems, such as *E. coli* and *B. subtilis*. Due to genetic drift, even chlamydial proteins with a conserved function have evolved a primary sequence that is significantly different from their *E. coli* orthologs. Because current genome annotation strategies only take into account sequence information, the functions of such proteins were not be identified and they were annotated as having an unknown function.

The Protein Structure Initiative (PSI) is a concerted effort by several major institutions with many goals, all of which concern protein structure and function. One of the goals of this program is to utilize structural information of proteins annotated as having an unknown function to infer a molecular function. The fruits of this labor have resulted in numerous discoveries and publications and have been reviewed elsewhere (Shin *et al.*, 2007). In short, ~60% of the solved protein structures of proteins of unknown function lead to testable hypotheses regarding their molecular functions. These structures of proteins of unknown function fell into five distinct categories, based on how

an inference of function was achieved. The largest of these categories contained proteins of unknown function (with no close homologs by primary sequence), which possessed a structure that clearly resembled the structure of a protein of known function. This allowed for clear assignment of a functional annotation. Based on this previous success, it is therefore anticipated that a similar strategy would prove effective in elucidating the function of the many proteins of unknown function encoded by chlamydial species, thereby giving insight into the molecular processes of chlamydial species that are poorly understood.

## The Role of Iron in Chlamydial Biology

In almost all bacteria, iron is required for life. This is due to the critical role that this micronutrient plays in catalysis for numerous enzymes involved in numerous cellular processes, including ribonucleotide reductase, superoxide dismutase and amino acid hydroxylases, to name a few (Andrews, 2000). Despite its critical importance in these metabolic processes, levels of iron must be tightly controlled because free iron is capable of reacting with oxygen, generating free radicals that are capable of causing damage to lipids, proteins and DNA (Kehrer, 2000). Therefore, bacteria have evolved regulatory networks that ensure that cellular iron levels remain in a concentration that is high enough to support the enzymes that require it as a cofactor, but low enough as to not cause damage to cellular components.

Species of chlamydia do not appear to be an exception for this common need for iron; the genome of *C. trachomatis* clearly shows the presence of several enzymes that

are predicted to require an iron cofactor (Stephens et al., 1998). Additionally, when deprived of iron by treatment with an iron chelating agent, chlamydia enter a persistent state, characterized by a halting of cell division, a halting of RB to EB conversion and a transition of RBs into large, non-dividing cells called aberrant bodies (Raulston, 1997b, Schoborg, 2011). Replacement of iron reverses this transition and chlamydial organisms resume their normal development. This response to iron deprivation suggests a strong requirement for iron in the progression of the chlamydial developmental cycle. Examination of the effects of iron-induced persistence at the transcriptional and translational levels has revealed several proteins that are up or down-regulated by this response (Timms *et al.*, 2009, Raulston, 1997b). However, the mechanisms for controlling this response are still poorly understood.

Regulation of iron homeostasis in bacteria is most frequently carried out by an iron-dependent transcriptional repressor. The two best studied of these repressors are named Fur and DtxR, originally characterized from *E. coli* and *Corynebacterium diptheriae*, respectively. These proteins repress expression of target genes by binding to a region of DNA overlapping the promoters of these genes (Schmitt *et al.*, 1992, de Lorenzo *et al.*, 1987). However, these proteins are only able to bind DNA and repress transcription if they have an iron cofactor bound, thereby only allowing transcription of the genes they repress under iron-limiting conditions (Pennella & Giedroc, 2005).

To date, two iron-responsive transcription factors have identified in *Chlamydiae*. DcrA (encoded by *C. trachomatis* ORF CT296) possesses distant sequence homology to Fur but was shown to functionally complement the Fur protein in *E. coli* and was able to bind to an *E. coli* Fur binding sequence (Wyllie & Raulston, 2001). Additionally, several

promoter regions of chlamydial ORFs were found to be bound by *E. coli* Fur, an interaction that was confirmed *in vitro* using purified recombinant DcrA (Rau *et al.*, 2005). More recently, a second iron-responsive transcriptional repressor was identified in chlamydia: the two-domain protein named YtgCR (encoded by *C. trachomatis* ORF CT069). The C-terminal domain of this protein (termed YtgR) has sequence homology to DtxR and was shown to repress the expression of a reporter gene in the heterologous system of *E. coli* by binding to an operator sequence from its own promoter (Thompson *et al.*, 2012).

Despite these recent findings, there are still many unanswered questions regarding the transcriptional response of chlamydial species in response to iron limitation. Gene targets for YtgC, other than itself, have yet to be identified and a consensus sequence for DcrA binding has yet to be established. Additionally, the relationship between these to repressors is also unknown; having ostensibly similar functions, it is not clear how these two proteins serve non-redundant roles. Clearly, additional research on these transcriptional repressors is needed to better understand their roles in the biology of *chlamydiae*.

## Type Three Secretion in *Chlamydia*

As described in detail above, at several stages of the chlamydial developmental cycle, these organisms secrete proteins into the host to subvert normal host cell functions. To achieve this, several different secretion systems are employed. Type II and Type V secretion systems, which secrete proteins outside of the cell from the periplasm, are

employed to secrete proteins such as CPAF (Saier, 2006). *Chlamydiae* also possess a Type III secretion system (Hsia *et al.*, 1997, Stephens et al., 1998). This system is most distinct from other secretion systems in that it allows for translocation of proteins directly from the bacterial cytoplasm across both bacterial membranes and then across the host cell membrane directly into the host cell cytoplasm.

The Type III secretion system (T3SS) is an essential factor for virulence in many other Gram negative bacterial pathogens, including species of *Yersinia*, *Salmonella*, and *Shigella flexneri* (Hueck, 1998). In other species, the T3SS secretion apparatus is referred to as the injectisome and is composed of approximately 25 different protein components (Beeckman & Vanrompay, 2010). A cartoon representation of this secretion system is diagrammed in FIG 1.1. The Basal Apparatus (BA), also referred to as a multipartite core secretory apparatus, is made up of a number of integral and non-integral membrane proteins that together span the inner and outer membranes of the bacterium, forming a hollow pore through which proteins may be secreted. Also part of the BA is an ATPase that powers secretion. Attached to this apparatus and located completely outside of the bacterial cell is the Needle Complex (NC), composed of 100-150 copies of a needle protein, forming a 40-80 μm long, hollow needle, through which partially folded proteins may be passed (Blocker *et al.*, 2008). Atop this needle lies the Tip Complex, composed of a multimer of a needle tip protein. This protein is typically responsible for sensing contact with the host cell and activation of the secretion system. The first proteins to be secreted following contact with host cell surface are the hydrophobic components of the Translocon Pore (TP), which embed themselves in the host cell membrane and complete the hollow connection between bacterium and host (Ghosh,

2004). In addition to the structural components of the injectisome, the T3SS also involves proteins called effectors that are translocated from the bacterial cytoplasm through the injectisome and into the host cell and the chaperone proteins that are frequently involved in stabilizing these effectors within the bacterial cell.

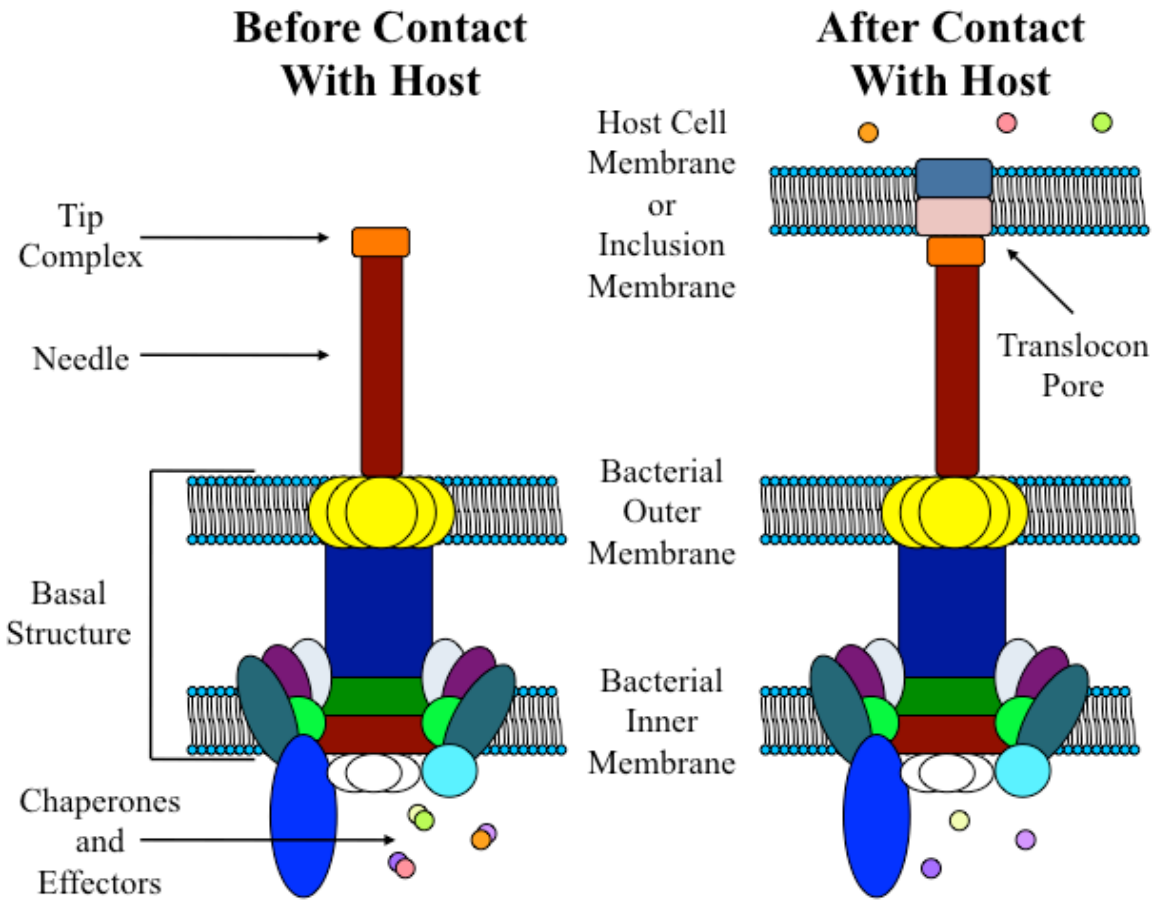
The injectisome apparatus and its components are well conserved among bacteria that possess a T3SS. Chlamydial species are no exception and the majority of T3SS structural genes were identified by homology (Stephens et al., 1998). Interestingly, in most bacteria that encode the components of a T3SS, the genes for this system are localized to a single pathogenicity island or plasmid (Hueck, 1998). However, in species of *Chlamydia*, components of the T3SS are spread across the genome; five main loci can be identified but some components are even more disperse (Stephens et al., 1998). While functionally related proteins seem to be located close to one another on the genome, the reason for this atypical gene arrangement (if there is one) is not clear.

While the structural components of the T3SS tend to be well conserved among Gram-negative bacteria, the actual effector proteins tend to be less well conserved. Importantly, due to the relative phylogenetic distance between chlamydial species and other more thoroughly studied bacteria, the identification of these effector proteins has been challenging and it is likely that more effectors remain to be discovered (Valdivia, 2008). Despite these challenges, several chlamydial T3SS effector proteins have been identified and characterized, together emphasizing the importance of the T3SS in the biology and pathology of chlamydial species.

One such effector protein is Tarp, which is prepackaged into EBs and is translocated into host cells at the early stages of infection. Tarp acts as a nucleation site

**FIG. 1.1. Model of the Chlamydial Type III Secretion System (T3SS) apparatus.** The components of the chlamydial T3SS apparatus can be organized into several different apparatuses and complexes. Anchored in both the inner and outer membranes of the bacterium, the Basal Apparatus forms a hollow pore through these membranes. Attached to this is the Needle Complex, a hollow needle through which secreted proteins may pass. At the tip of the Needle Complex is the Tip Complex, which is responsible for sensing contact with the host cell.

Once in contact with the host, the components of the Translocon Pore are transported through the secretion system and embed themselves in the host cell membrane. In the context of secretion through the inclusion, this pore would form in the inclusion membrane. Now, effector proteins (stabilized within the bacterium by chaperone proteins) can be translocated into the host cell. Adapted from Betts-Hampikian and Fields (Betts-Hampikian & Fields, 2010).


Before Contact With Host
After Contact With Host
Tip Complex
Needle
Basal Structure
Chaperones and Effectors
Host Cell Membrane or Inclusion Membrane
Translocon Pore
Bacterial Outer Membrane
Bacterial Inner Membrane

for actin and is a major driving force behind the cytoskeletal rearrangement that occurs during chlamydial entry into the host cell (Jewett et al., 2006). Additional T3SS effectors include various Inc proteins that are involved in modification of the inclusion membrane, giving stability to the inclusion and allowing for subversion of host cell processes (Fields & Hackstadt, 2002, Wolf *et al.*, 2009, Betts *et al.*, 2008). A final example of T3SS effectors involves the proteins ChlaDub1 and ChlaDub2. These two proteins are cysteine proteases and appear to prevent ubiquitination of NF-κB, thereby leading to a suppression of pro-inflammatory signaling (Misaghi *et al.*, 2006). Although these are only a few examples of the many translocated effector proteins employed by *Chlamydiae*, they illustrate the variety of functions that these proteins have and they emphasize the importance of the T3SS throughout the chlamydial developmental cycle.

## Cell Division in *Chlamydia*

The process of cell division is incredibly well conserved among bacterial species, likely due to its indispensible role in the propagation of life. This process involves a large number of different proteins that are responsible for coordinating the process such that the division septum is formed in the correct location and with the correct shape (Weiss, 2004). Among the proteins involved in this process, perhaps the most important is the protein FtsZ. FtsZ is structurally similar to tubulin and polymerizes into protofilaments inside the bacterial cell in a shape the forms a ring at the division septum (Erickson *et al.*, 2010). This ring of FtsZ polymer is called a Z ring and marks the site for septation by recruiting all other proteins involved in this process (Adams & Errington,

2009). Other key players in this process include (but are not limited to) the proteins ZipA and FtsA, which are responsible for tethering the Z ring polymer to the inner membrane, and the MinCDE system, which is responsible for ensuring proper Z ring assembly at midcell by inhibiting FtsZ polymerization at the cell poles (Adams & Errington, 2009). Importantly, no motor molecules have been identified in the process of cell division and data support the hypothesis that it is the FtsZ protein itself that powers the process of septation by forming curved filaments as it polymerizes (Li *et al.*, 2007).

In spite of its strict conservation among bacterial species, chlamydial species do not have an FtsZ ortholog that is apparent by sequence homology, following sequencing of the genome (Stephens et al., 1998). Due to their phylogenetic distance from other, more thoroughly studied bacteria, it is possible that an FtsZ ortholog may exist in chlamydial species but cannot be identified in the genomes due to sequence divergence. However, FtsZ is not the only conserved component of cell division that cannot be identified in the genome sequences of chlamydial species: virtually all FtsZ interaction partners, components of the cell divisome and their regulatory proteins (including FtsA, ZipA, ZapA, MinCDE and UgtP) are absent from chlamydial genomes, suggesting that these species employ a novel mechanism for cell division (Stephens et al., 1998).

The absence of these highly conserved cell division proteins is not the only point of interest regarding proteins encoded by chlamydial species: these organisms also encode several proteins involved in rod cell shape maintenance even though they have coccoid (round) morphology (Stephens et al., 1998). These proteins include RodA and MreB, both of which have been shown in other species to guide the peptidoglycan machinery in such a way that it creates cell wall in a distinct rod shape (White *et al.*,

2010). These proteins are typically found only in rod or crescent shaped bacteria and their presence in the coccoid chlamydial species was surprising.

Another interesting feature of chlamydial species is their lack of peptidoglycan. While there are examples of bacteria without peptidoglycan cell walls, *chlamydiae* are unique for several reasons. No peptidoglycan could be detected in chlamydial species by any of the methods used (Chopra *et al.*, 1998). However, these organisms are sensitive to antibiotics like penicillin that target peptidoglycan synthesis; these compounds are not bactericidal but put chlamydia in a so-called persistent state of infection characterized by large, aberrant RBs (Wyrick, 2010). Additionally, the first sequenced genome of *Chlamydia trachomatis* revealed that these organisms encode almost all of the genes necessary for peptidoglycan production, with the notable exception of the transglycosylase enzymes, which connect the N-acetylglucosamine and N-acetylmuramic acid disaccharides to one another (Stephens et al., 1998). Together, these findings have lead to several hypotheses, most recently reviewed in 2007 (Pavelka, 2007). These hypotheses include suppositions that chlamydial species might employ extremely low amounts of peptidoglycan (explaining their inability to be detected) or that an atypical (perhaps glycan-less) form of peptidoglycan is utilized. A third hypothesis is that this peptidoglycan, regardless of its precise structure, is only used in a specific, transitory process, thereby explaining its undetectable levels.

All of these anomalies regarding chlamydial cell division, rod shape and peptidoglycan can be united by a hypothesis in which the rod shape determining proteins functionally replace FtsZ by guiding peptidoglycan-mediated septation at midcell (Ouellette *et al.*, 2012). A non-proteinaceous antigen termed SEP was identified at the

site of septation in dividing chlamydia and was proposed to be a form of peptidoglycan due to its cross-reactivity with a component of the cell walls of other bacteria (Brown & Rockey, 2000). Furthermore, inhibition of the cell shape determining protein MreB lead to a phenotype in which cell division was arrested in cells, implicating it and other rod-shape determining proteins in the process of cell division (Ouellette et al., 2012). While additional support for this hypothesis is needed, it does neatly explain many of the anomalies of chlamydial biology together in a unique mechanism for cell division.

## Collaborator's Contributions

The crystal structure of CT296 in Table 2.2 and Fig. 2.1B was solved by Dr. John Hickey (Department of Pharmaceutical Chemistry, University of Kansas), with assistance from Dr. Scott Lovell (Del Shankel Structural Biology Center, University of Kansas) and Dr. Kevin Battaile (IMCA-CAT Hauptman-Woodward Medical Research Institute). The sequence and secondary structure alignment in Fig. 2.1C was generated by Dr. John Hickey (Department of Pharmaceutical Chemistry, University of Kansas). Confocal microscopy in Fig. 3.3 was performed with guidance from Heather Shinogle (Microscopy and Analytical Imaging Laboratory, University of Kansas). The crystal structure of CT009 in Table 4.3 and Fig. 4.8 was solved with the guidance of Dr. Scott Lovell (Del Shankel Structural Biology Center, University of Kansas). Sedimentation velocity profile of purified recombinant CT009 in Fig. 4.11 was performed by Lei Hu (Department of Molecular Biosciences, University of Kansas).

## Chapter II. Ab initio structural modeling of and experimental validation for *Chlamydia trachomatis* protein CT296 reveal structural similarity to Fe(II) 2-oxoglutarate-dependent enzymes


### Abstract

*Chlamydia trachomatis* is a medically important pathogen that encodes a relatively high percentage of proteins with unknown function. The three-dimensional structure of a protein can be very informative regarding the protein's functional characteristics; however, determining protein structures experimentally can be very challenging. Computational methods that model protein structures with sufficient accuracy to facilitate functional studies have had notable successes. To evaluate the accuracy and potential impact of computational protein structure modeling of hypothetical proteins encoded by Chlamydia, a successful computational method termed I-TASSER was utilized to model the three-dimensional structure of a hypothetical protein encoded by open reading frame (ORF) CT296. CT296 has been reported to exhibit functional properties of a divalent cation transcription repressor (DcrA), with similarity to the *Escherichia coli* iron-responsive transcriptional repressor, Fur. Unexpectedly, the I-TASSER model of CT296 exhibited no structural similarity to any DNA-interacting proteins or motifs. To validate the I-TASSER-generated model, the structure of CT296 was solved experimentally using X-ray crystallography. Impressively, the ab initio I-TASSER-generated model closely matched (2.72-Å C(α) root mean square deviation [RMSD]) the high-resolution (1.8-Å) crystal structure of CT296. Modeled and

experimentally determined structures of CT296 share structural characteristics of non-heme Fe(II) 2-oxoglutarate-dependent enzymes, although key enzymatic residues are not conserved, suggesting a unique biochemical process is likely associated with CT296 function. Additionally, functional analyses did not support prior reports that CT296 has properties shared with divalent cation repressors such as Fur.

## Introduction

Progress toward better understanding key aspects of chlamydial biology is hindered by the inherent restraint presented by the phylogenetic distance between *C. trachomatis* and better-characterized bacterial systems, such as *E. coli* and *B. subtilis*. This distance dramatically reduces the utility of protein function assignments as inferred by sequence similarity. As a result, *Chlamydia trachomatis* encodes a relative overabundance (~25%) of proteins with little or no sequence similarity to functionally defined proteins (i.e. hypothetical proteins) (Carlson et al., 2005, Unemo et al., 2010, Thomson et al., 2008, Stephens et al., 1998).

For proteins where the only known characteristics are their primary sequence, the three dimensional structure of a protein can be very informative and useful in regards to understanding functional characteristics. Importantly, protein structure provides the precise molecular details that often facilitate experimental characterization of an expected function. Similarly, the structure of a protein can also be useful in facilitating functional predictions as a search template for better-characterized proteins that share regions of structural similarity. While there are many cases of structural information playing a

productive role in predicting putative function (Yakunin *et al.*, 2004), there are limitations with this approach. For instance, proteins can share very similar global folds but have contrasting functions (Thornton *et al.*, 2000). However, local structural similarities, as well as certain well-defined structural motifs (e.g. helix-turn-helix, catalytic triad) can be much more predictive of function than global fold alone (Hvidsten et al., 2009). Thus, while the three-dimensional structure of a protein may not be completely revealing regarding function, there are many examples to support that it will be very informative regarding numerous functional aspects.

Obtaining high resolution, three-dimensional structural information can be challenging in certain cases and this proves to be a major constraint to using protein structure to facilitate functional studies. The development of computational methods that can predict protein structure to a relatively high level of accuracy and thereby facilitate functional annotation, biochemical analyses, and biological characterization is a high priority (Zhang, 2008). There are three different categories of methods for computational protein structure modeling. Comparative modeling, or homology modeling, uses the structures of close homologs to generate a structural model. This method is typically only effective when the modeled protein shares relatively high primary sequence similarity with proteins with determined structures (Qu *et al.*, 2009). A second approach is termed threading, which uses smaller regions of sequence homology between the query sequence and sequences of solved structures to build a model. As with homology modeling, threading is only effective if the modeled protein has primary sequence homology to a protein or proteins of known structure. These two methods are less capable of modeling the structure for a protein with no close homologs and no significant sequence homology.

The third method of computational modeling is *ab initio*, or free modeling (Lee *et al.*, 2009). This method builds a structure from scratch and is not dependent on the existence of homologs. Because this method is the most technically challenging method for protein structure modeling, it is a particular challenge to accurately model structures for proteins with very little or no shared similarity evident by their primary sequence.

I-TASSER (Iterative Threading ASSEmbly Refinement) is a computational method that has been successful in accurately modeling protein structures. I-TASSER uses a combinatorial approach, employing all three conventional methods for structure modeling: comparative modeling, threading, and *ab initio* modeling (Roy *et al.*, 2010). In large blind benchmarking experiments conducted over the past several years (Critical Assessment of protein Structure Prediction: CASP), I-TASSER has consistently been ranked among the most accurate modeling program for automated protein structure predictions (Battey *et al.*, 2007, Cozzetto *et al.*, 2009). Importantly, I-TASSER has also proven to be particularly effective in *ab initio* modeling of small proteins with low primary sequence similarity (Wu *et al.*, 2007).

To provide further support for the utility of protein structure analyses to facilitate functional characterization of hypothetical proteins encoded by *Chlamydia*, both computational and experimental approaches were applied to a chlamydial hypothetical protein for which experimental information regarding the function of the protein had been previously obtained. *Chlamydia trachomatis* open reading frame CT296 has very limited sequence homology to any proteins outside of *Chlamydiae* and was initially annotated as a protein with unknown function (Stephens et al., 1998). Subsequent functional studies indicated that CT296 exhibits properties of a Divalent cation transcription repressor

(DcrA), with functional similarity to the Fur repressor (Rau et al., 2005, Wyllie & Raulston, 2001). Fur (short for Ferric uptake regulator) is an iron-responsive transcriptional repressor that regulates genes involved in iron acquisition in an iron-dependent fashion (Escolar *et al.*, 1999). Iron concentrations have been demonstrated to have a profound effect on the growth of *Chlamydia* by inducing a persistent growth phenotype that is expected to be clinically relevant (Raulston, 1997a). As such, it was anticipated that structural studies on CT296 would also be instrumental in understanding the molecular mechanisms critical for a transcription factor associated with chlamydial pathogenesis.

In this study, the structure of CT296 was determined computationally, using I-TASSER, and experimentally, using X-ray crystallography. These two structures were compared to assess the ability of I-TASSER to model CT296 *ab initio* with a relatively high level of accuracy. Each of the structures was used as a search template to identify proteins with shared structural features. The functional properties of the proteins with structural similarity to CT296 were compared to the previous experimental observations of CT296. As subsequently described, given the structural observations and potential functional disparity, oligomerization and DNA binding capabilities were analyzed.

## Chapter II

## Materials and Methods

*Cloning, Expression, and Purification of CT296*- The open reading frame encoding CT296 (GenBank Accession #: CAP03988.1) was PCR amplified with Phusion high fidelity DNA polymerase (Finnzymes, Lafayette, CO), using *Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA and primers 5'-CTGTACTTCCAATCGCGAATGAGGGCAGTTTTACACCTAGAG-3' and 5'-GAATTCGGATCCTCGCGATTAGTTAGGAAATCCCGCTGAGGAG-3'. The amplified DNA was inserted into a modified pET21b (Novagen, San Diego, CA) vector at the EcoR1 restriction site, using the In-Fusion Advantage kit (Clontech, MountainView, CA), following the manufacturer's protocol. The resulting plasmid contained the *CT296* gene downstream of DNA encoding a polyhistidine tag and a TEV protease recognition sequence. The plasmid insert was sequenced (ACGT, Wheeling, IL) and confirmed to be completely present and without mutation. The plasmid was then transformed into *E. coli* Acella cells (EdgeBio, Gaithersburg, MD). For native CT296 over-expression, the cells were grown in LB medium supplemented with 100 µg/mL ampicillin to an $OD_{600}$ of 0.8 before the addition of 1 mM IPTG. For selenomethionine (SeMet)-labeled CT296, the cells were grown and labeled using a previously described method (Doublie, 2007). The cells were harvested after an overnight growth at 15°C through centrifugation (10 min at 5,000 x *g*; 4°C). The resulting pellet was resuspended in purification buffer (50 mM $Na_2HPO_4$ pH 7.0, 300 mM NaCl) and disrupted by sonication. Cell debris was removed through centrifugation (30 min at 14,000 x *g*; 4°C).

CT296 was purified using $Co^{2+}$-affinity chromatography (Clontech), following the manufacturer's protocol. The eluted fractions containing CT296 were pooled and a polyhistidine-tagged Tobacco Etch Virus (TEV) protease was added to a final concentration of 0.6 μM in order to cleave the polyhistidine tag from the protein. Following an overnight dialysis in the purification buffer, the protein mixture was added to a $Co^{2+}$ column to remove the polyhistidine tagged TEV protease and the cleaved polyhistidine fragments.

*Size Exclusion Chromatography*- Purified recombinant CT296 at concentrations of 1000, 100 and 10 μM were analyzed by size exclusion chromatography using a BioLogic DuoFlow system (Bio-Rad). 1 mL of protein solution was applied to a Sephacryl S-200 HR column (GE Healthcare), equilibrated with 50 mM $Na_2HPO_4$ (pH 7.0) and 300 mM NaCl. A protein standard mixture containing bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobin (17 kDa) and vitamin $B_{12}$ (1.4 kDa) (Bio-Rad) was used to generate a standard curve.

*Electrophoretic Mobility Shift Assay (EMSA)- Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA was used as a template to amplify ~400 base pair fragments of DNA by PCR. These fragments consist of a region spanning ~100 base pairs upstream of the transcription start site and ~300 base pairs downstream of the transcriptional start site. These putative promoter regions of DNA were amplified for the genes *CT248* and *CT709* using the IR800-labled primers (Integrated DNA Technologies); *CT248* 5'-GATGCCGATTGAGGAGTCTG-3' and 5'-GCAACTTTTGCTGTACATTCC-3'or

*CT709* 5'-CTTAGTTTCTTTAAAAGCTGGAG-3' and 5'-GGAAAAAAAATTAAATACACGATC-3'. These genes were previously identified as DNA binding targets of CT296 (Rau et al., 2005). To serve as a negative control in the EMSA reaction, ~400 bases from the region of DNA between the two converging genes *CT032* and *CT033* was amplified using IR800 labeled primers; 5'-ACATTCCTTAGATCTAGGTTCCC-3' and 5'-GCTATTGCTGTTCGTAATAATAAGG-3').

EMSA reactions were prepared with a total volume of 20 μL each, using the methods and buffers previously described (Rau et al., 2005). Reactions contained EMSA buffer, DNA fragments (48 pmol) and purified recombinant CT296 (at DNA:protein molar ratios of 1:100, 1:500 and 1:2500). Reactions without protein were used as a negative control. Once protein was added, the solution was gently mixed and incubated for 30 minutes at room temperature. 5 μL of 5x DNA dye (90 mM Tris, 90 mM boric acid, 2 mM EDTA, 20% glycerol, 1% xylene cyanol, 1% bromphenol blue) was added to the reactions and mixed gently. 20 μL of each reaction was loaded on a precast 6% TBE gel (Invitrogen) and run at 150V for 75 minutes. Bands were visualized using an Odyssey infrared imaging system (Li-Cor Biosciences).

*Target Selection and I-TASSER Structure Modeling Method-* The following amino acid sequence from the genome of *Chlamydia trachomatis* LGV (L2/484/Bu) was selected as the target:
MRAVLHLEHKRYFQNHGHILFEGLAPVSDCKQLEAELKLFLKEVAVVKDRHLQ
RWRENVHRTLPGVQMIVKRVRLDHLAAELTHRSRVALVRDLWVQKQEEILFDD

CDCSVLLCLSGEKAGWGLFFSGEYPQDVFDWGAGDTAIILRFSSAGFPN. Using the nomenclature of the L2 genome, the name of this protein is CTLon_0544. However, in this report the more widely used nomenclature of the serovar D genome is used and the protein is referred to as CT296.

The I-TASSER server has been described previously (Zhang, 2009, Roy et al., 2010). Briefly, the target sequence is first threaded through a representative PDB library by LOMETS, a locally installed meta-threading program (Wu & Zhang, 2007). The continuous fragments (>5 residues) are excised from the LOMETS alignments and used to reassemble the structure by replica-exchange Monte Carlo simulations (Wu & Zhang, 2007). The simulation trajectories are then clustered by SPICKER (Zhang & Skolnick, 2004) which are used as the starting state of the second round I-TASSER assembly simulation. Finally, the structures of the lowest energy are selected, which are then refined by a fragment guided molecular dynamics procedure, with the purpose of optimizing the hydrogen-binding network and removing steric clashes.

The computationally determined protein structure of CT296 was analyzed for putative DNA binding regions using PreDs (Tsuchiya *et al.*, 2005) and HTHQuery (Ferrer-Costa *et al.*, 2005).

*CT296 Crystallization*- Purified recombinant CT296 was concentrated to 6.3 mg/ml in 50 mM $Na_2HPO_4$ pH 7.0, 300 mM NaCl and screened for crystallization with commercially available sparse matrix screens. 1 µl of the protein was mixed with 1 µl of the crystallization solution in Compact Jr. (Emerald Biosystems, Bainbridge Island, WA) sitting drop vapor diffusion plates and equilibrated against 100 µl of the crystallization

solution. CT296 crystals were observed after 3 days at 20°C from the Wizard 4 screen condition #48 (Emerald Biosystems) (100 mM HEPES pH 7.0, 15% (w/v) PEG 20,000). Single crystals were transferred to a cryoprotectant solution containing the crystallization condition supplemented with 20% PEG 400.

Co-crystallization was attempted with purified CT296 at 8.6 mg/ml and 5 mM α-ketoglutarate (Sigma-Aldrich, St. Louis, MO) and/or 5 mM $FeCl_2$ (Sigma-Aldrich). The resulting mixture was screened for crystallization as described above; however no crystals were obtained. Additionally, apo CT296 crystals were soaked in the crystallization solution supplemented with 50, 5, 1, or 0.1 mM α-ketoglutarate and/or $FeCl_2$. The crystals dissolved within seconds-to-minutes of exposure to either compound or combination.

*Data Collection and Processing*- Data sets for the native and SeMet-labeled CT296 crystals were collected at 100K at the Advanced Photon Source (APS) IMCA-CAT, sector 17ID. Data for the native and the SeMet-labeled crystals were collected using an ADSC Quantum 210r CCD detector and Dectris Pilatus 6M pixel array detector, respectively. Native diffraction data were processed with *HKL2000* and SeMet data were integrated and scaled using *XDS* (Kabsch, 1993) and *SCALA* (Evans, 2006), respectively, via the *AUTOPROC* (Vonrhein *et al.*, 2011) software package. Structure solution using the SAD phasing method and preliminary model building were performed using *AUTOSOLVE* (Terwilliger *et al.*, 2009), contained within the *PHENIX* suite (Adams *et al.*, 2010). Automated model building was performed using *AUTOBUILD* (Terwilliger *et al.*, 2008). The SeMet derived CT296 model was used as the search model for molecular

replacement against the native data set using *PHASER* (McCoy *et al.*, 2007a). Manual model building and subsequent refinement of the models was performed using *COOT* (Emsley & Cowtan, 2004) and *PHENIX* (Adams et al., 2010) respectively.

# Chapter II

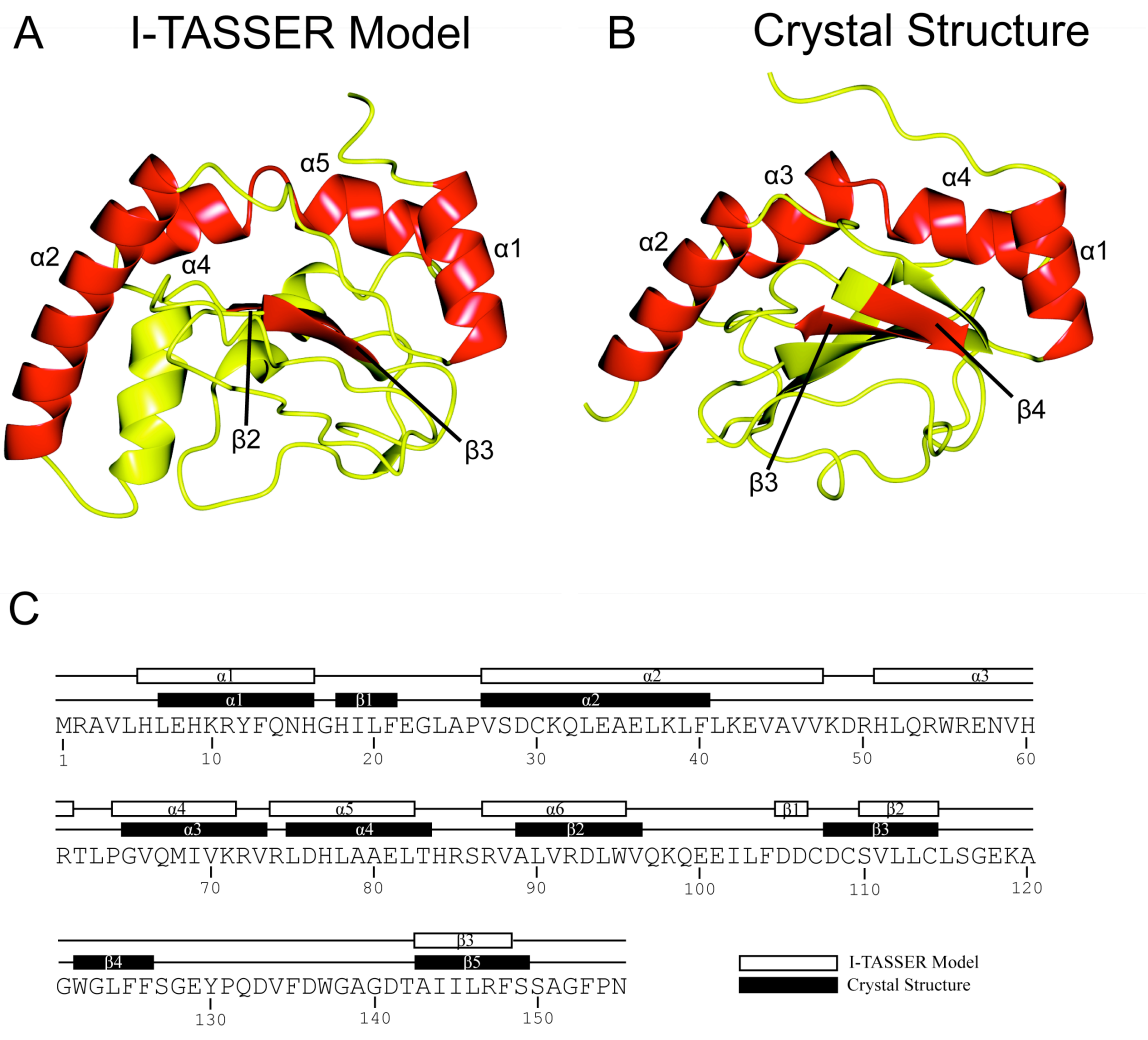
# Results

## *The I-TASSER Modeled Structure of CT296 Shows No Significant Structural Similarity to Fur Family Members and Has No Apparent DNA-Binding Motifs.*

The protein encoded by ORF *CT296* has been reported to contain very limited sequence similarity to the Fur protein but exhibits metal-dependent DNA binding properties comparable to Fur (Rau et al., 2005, Wyllie & Raulston, 2001). To better understand the molecular basis for its metal and DNA binding functions as well as test the accuracy and impact of I-TASSER protein structure modeling, both computational and experimental structural analyses were performed on CT296.

Initially, computational modeling of the structure of CT296 using I-TASSER was performed. Five models of CT296 were computationally generated using the I-TASSER algorithm with C-scores ranging from -5 to -2.25. The C-score is a confidence score and ranges from -5 to 2, with higher scores representing higher confidence in the model (Roy et al., 2010). Although -1.5 has been used previously as a cutoff for confidence in models, there can still be significant modeling accuracy in models with C-scores below this limit (Roy et al., 2010). Model 1 (C-score of -2.25) was used for all analyses described (Fig. 2.1A) as the remaining models had C-scores of -5. The overall structure of CT296 model 1 is primarily α-helical (5 helices; 46%), a significant number of loops (48%), and two relatively short β-strands (6%). The topology of the CT296 model is (residue number); α1 (H6-H16) – α2 (V27-V47) – α3 (H51-R61) – α4 (P64-K71) – α5

**FIG. 2.1. Tertiary and secondary structure comparisons of the I-TASSER model and the experimentally determined structure of CT296.** (A) Ribbon depiction of the I-TASSER model of CT296. Topology of the CT296 model is (residue number); α1 (H6-H16) – α2 (V27-V47) – α3 (H51-R61) – α4 (P64-K71) – α5 (R74-L82) – α6 (R87-W95) – β1 (D105-D106) – β2 (S110-C114) – β3 (A143-F148). (B) The experimentally determined structure of CT296. Topology of the crystal structure of CT296 is (residue number); α1 (L7-H16) – β1 (H18-F21) – α2 (V27-F40) – α3 (G65-V73) – α4 (L75-T83) – β2 (A89-V96) – β3 (D108-C114) – β4 (W122-F126) – β5 (A143-S149). Secondary structure elements that are conserved between the two structures (four α-helices and two β-sheets) are labeled and highlighted in red. (C) Secondary structure topology comparison of I-TASSER model and experimentally determined structure of CT296.


A
I-TASSER Model
α5
α2
α4
α1
β2
β3
B
Crystal Structure
α3
α4
α2
α1
β4
β3
C
α1
α2
α3
α1
β1
α2
MRAVLHLEHKRYFQNHGHILFEGLAPVSDCKQLEAELKLFLKEVAVVKDRHLQRWRENVH
1
10
20
30
40
50
60
α4
α5
α6
β1
β2
α3
α4
β2
β3
RTLPGVQMIVKRVRLDHLAAELTHRSRVALVRDLWVQKQEEILFDDCDCSVLLCLSGEKA
70
80
90
100
110
120
β3
β4
β5
GWGLFFSGEYPQDVFDWGAGDTAIILRFSSAGFPN
130
140
150
I-TASSER Model
Crystal Structure

(R74-L82) – α6 (R87-W95) – β1 (D105-D106) – β2 (S110-C114) – β3 (A143-F148) (Fig. 2.1A).

Despite the low overall primary sequence similarity between CT296 and Fur, it was anticipated that the two would share significant structural characteristics, largely based on the previously reported ability of CT296 to functionally complement Fur in *E. coli* (Wyllie & Raulston, 2001). However, the I-TASSER generated model of CT296 bears little resemblance to the structures of Fur family members (PDB#: 2W57, 2O03, 2FU4 and 1MZB) as no significant similarity was identified using DaliLite pairwise comparisons (with a Z-score of >2 used as a cutoff for significance; data not shown) (Holm & Park, 2000). Additionally, no DNA binding motifs (e.g. helix-turn-helix, a leucine zipper, a zinc finger, winged helix-turn-helix, etc.) were apparent in the I-TASSER model of CT296. To identify potential regions of DNA-binding sites based on electrostatics and curvatures on the protein surface, the model of CT296 was subjected to computational analysis by *PreDs* (Tsuchiya et al., 2005). These protein characteristics are summarized as a P score, the value of which ranges from 0.0 to 1.0 and is predictive of a protein with the necessary structural characteristics to bind to DNA. The results of this analysis give a P score of 0.03 to the model of CT296. A P score of 0.12 is typically used as a cutoff point, because DNA-binding proteins almost always have P scores above this threshold. As a control, the effector domain of ChxR, a well-characterized chlamydial DNA binding protein, was subjected to the same analysis (Koo *et al.*, 2006, Hickey *et al.*, 2011). It was given a P score of 0.22, correctly predicting that it is capable of binding to DNA. All together, the I-TASSER generated model of CT296 is in contrast with the

previous observations that CT296 is a DNA-binding protein with structural similarity to Fur.

***Homology Searches Using The I-TASSER Generated Model of CT296 Support Its Functional Prediction as a Non-Heme Fe(II) 2-oxoglutarate Dependent Enzyme.***

The model of CT296 generated by I-TASSER was used to search for proteins with regions of structural homology within the PDB (Table 2.1). This assessment was carried out using both the PSFloger server (http://zhanglab.ccmb.med.umich.edu/PSFloger/), and the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/). PSFloger was developed in part to complement I-TASSER; it detects the structural and functional analogs from a representative PDB library by the global structural alignment algorithm, TM-align (Zhang & Skolnick, 2005), assisted with local motif matches (Zhang & Skolnick, 2005). DALI is a well-established, widely utilized server for protein structure alignment (Holm & Sander, 1993). Although the overall goal of the PSFloger and the DALI servers is to identify structural homologs, both servers were employed in order to give a deeper insight into the relationship between the modeled structure of CT296 and other proteins of known structure.

The PSFloger search indicated that the 10 proteins the share structural similarity to the CT296 model are from diverse sources (e.g. *H. sapiens* [mammal], *P. syringae* [bacteria], and *A. thaliana* [plant]) but have similar functions as non-heme Fe(II) and 2-oxoglutarate dependent enzymes. The closest structural match (2OPW) is the enzyme PHYHD1 from *Homo sapiens,* which carries out a hydroxylation reaction using a non-

## Table 2.1. Top 10 structures with homology to the CT296 I-TASSER model as measured by PSFloger and DALI

| | PSFloger | | | | DALI | | | |
|---|---|---|---|---|---|---|---|---|
| Rank | TM-Score[1] | PDB File | Protein Name | Classification | Z Score[2] | PDB File | Protein Name | Classification |
| 1 | 0.6619 | 2OPW | PHYHD1 | Hydroxylase (Mihalik *et al.*, 1997) | 6.3 | 2G1M | PHD2[3] | Hydroxylase (McDonough *et al.*, 2006) |
| 2 | 0.6564 | 2HBT | PHD2 | Hydroxylase (Warshakoon *et al.*, 2006) | 5.7 | 3GJB | CytC3[3] | Halogenase (Wong *et al.*, 2009) |
| 3 | 0.6365 | 2CSG | YbiU | Putative oxioreductase | 5.5 | 3EMR | ECTD | Dioxygenase (Reuter *et al.*, 2010) |
| 4 | 0.6131 | 2FCU | SyrB2 | Halogenase (Blasiak *et al.*, 2006) | 5.3 | 3KT7 | Tpa1[3] | Hydroxylase (Kim *et al.*, 2010) |
| 5 | 0.6063 | 1GP6 | ANS | Dioxygenase (Wilmouth *et al.*, 2002) | 5.1 | 3OBZ | PHYHD1[3] | Hydroxylase |
| 6 | 0.6060 | 1UNB | DAOCS | Oxioreductase (Valegard *et al.*, 2004) | 5.1 | 3NN1 | Hal[3] | Halogenase (Khare *et al.*, 2010) |
| 7 | 0.6019 | 2DBN | JW0805 | Unknown Function | 4.5 | 2RG4 | Q2CBJ1 | Unknown Function |
| 8 | 0.5998 | 2A1X | PHYHD1 | Hydroxylase (McDonough *et al.*, 2005) | 4.4 | 3DKQ | Sbal_3634[3] | Putative Hydroxylase |
| 9 | 0.5917 | 1W3X | IPNS | Oxioreductase (Daruzzaman *et al.*, 2006) | 4.2 | 3BVC | Ism_01780[3] | Unknown Function |
| 10 | 0.5820 | 2RDN | PtlH | Hydroxylase (You *et al.*, 2007) | 4.0 | 2FCU | SyrB2 | Halogenase (Blasiak et al., 2006) |

[1]TM-score is a measurement of similarity between two protein structures and it varies from 0 to 1, with 1 indicating a perfect alignment. A score of 0.5 or higher indicates significant similarity (Xu & Zhang, 2010).

[2]Structures with Z-scores above 2 are considered to be significant matches (Holm *et al.*, 2008).

[3]This protein was listed multiple times on the DALI output, with various minor differences (ligands, etc.); these redundant structures are not listed here.

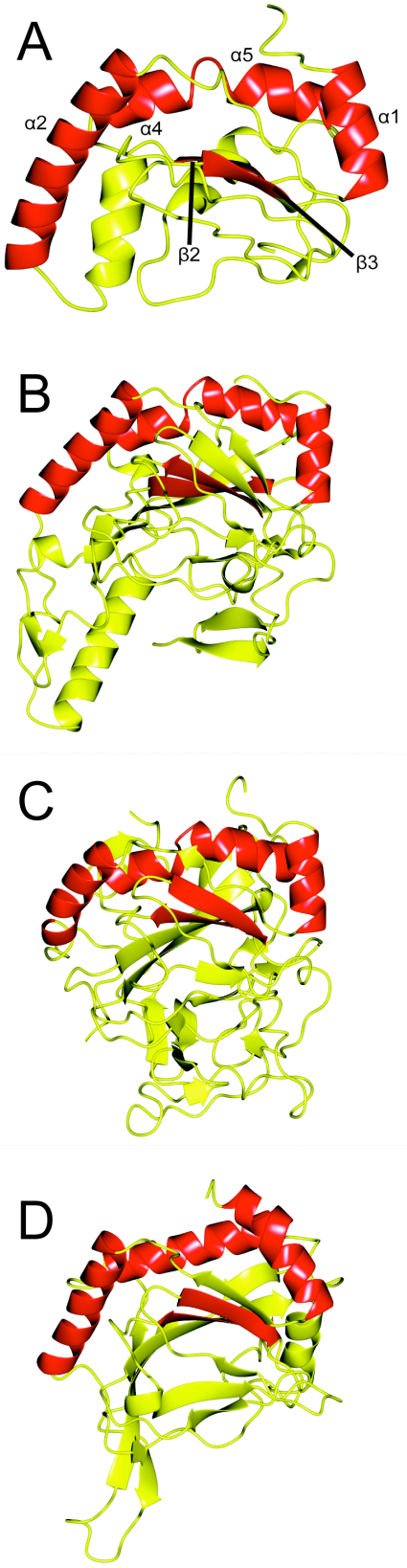
heme Fe(II) cofactor and a 2-oxoglutarate co-substrate (Mihalik et al., 1997, Warshakoon et al., 2006). The next closest match (2HBT) is also from *Homo sapiens* and is called PHD2. PHD2 is a prolyl hydroxylase and likewise makes use of non-heme Fe(II) and 2-oxoglutarate (McDonough et al., 2006). The third structural match to the model of CT296 (2CSG) is the protein YbiU from the bacterium *Salmonella typhimurium* and is a putative oxidoreductase with iron bound in the crystal structure. As a final example, the fourth closest structural match to CT296 (2FCU) is the halogenase SyrB2 from *Pseudomonas syringae*; like the first two matches, it requires non-heme Fe(II) and 2-oxoglutarate (Blasiak et al., 2006). Although these structural homologs differ in their classes of enzymatic function (e.g. hydroxylase or oxidoreductase), the similarities among them predict that CT296 is an enzyme that utilizes iron as a co-factor and may bind 2-oxoglutarate.

Like PSFloger, the DALI server also identified structural homologs from diverse sources (e.g. *H. sapiens* [mammal], *S. cerevisciae* [fungus] and *S. salexigens* [bacteria]). Also similar to the PSFloger results, all matches made by DALI function as non-heme Fe(II) and 2-oxoglutarate dependent enzymes. Despite the similar functional characteristics of the PSFloger and DALI matches, only three of the top ten structures identified as homologous to the CT296 model by PSFloger were also identified by DALI. This is partly because PSFloger uses a non-redundant PDB library (with pair-wise sequence identity <95%) while DALI search utilizes all available PDB structures. The top ranked DALI structure (2G1M) is the protein PHD2, the same protein that was listed as the second highest ranked homolog by PSFloger (PDB file 2HBT; same protein, different ligands bound). The second closest structural match (3GJB) is the halogenase

CytC3 from *Streptomyces sp.* and was not ranked in the top ten structure homologs by PSFloger. This enzyme also uses Fe(II) and 2-oxoglutarate for catalysis (Wong et al., 2009). The third match (3EMR) is ECTD from *Salibacillus salexigens* and is a non-heme Fe(II)-dependant dioxygenase (Reuter et al., 2010). As a final example, the number four structural match to the CT296 model by DALI is PDB code 3KT7. It is the protein Tpa1 from *Saccharomyces cerevisiae*, a hydroxylase that binds iron and 2-oxoglutarate (Kim et al., 2010). Overall, the observations of the two alignment programs support the prediction of CT296 as an Fe(II) 2-oxoglutarate dependant enzyme.

A relatively conserved feature of the non-heme Fe(II) 2-oxoglutarate dependent enzymes is the presence of a Double-Stranded Beta Helix (DSBH) core in which the catalytic events typically occur (Clifton *et al.*, 2006). This DSBH fold is formed by two opposing β-sheets, composed of at least four β-strands each (Clifton et al., 2006). The CT296 model appears to contain an incomplete DSBH core with only one relatively small β-sheet present containing 3 β-strands (Fig. 2.1A and Fig. 2.2A). In order to examine the conserved structural characteristics of the CT296 model and its homologs, the model was aligned with three homologous structures (2OPW, 2FCU, and 2HBT) from the PDB using the CCP4 Molecular Graphics program (Potterton *et al.*, 2004) (Fig. 2.2). The three models selected for alignment represent proteins that were identified as homologs to CT296 by both PSFloger and DALI. There are five (2HBT) or six (2OPW and 2FCU) secondary structures that are present and similarly oriented among the three models: three or four α-helices and two β-strands. The three or four conserved α-helices are oriented immediately adjacent to one of the β-sheets that forms the canonical DSBH

**FIG 2.2. Ribbon depiction of I-TASSER model for CT296 aligned with three proteins exhibiting the highest structural homology.** Regions of structure that are conserved between all four structures are highlighted in red. (A) I-TASSER model of CT296. Four α-helices are highlighted, α1 (L7-H16), α2 (V27-F40), α3 (G65-V73) and α4 (L75-T83). Two β-strands are highlighted, β3 (D108-C114) and β4 (W122-F126). (B) Crystal structure of *H. sapien* PHYD1 (PDB 2OPW). Four α-helices, (similarly oriented to α1, α2, α3 and α4 from the I-TASSER model of CT296) and two β-strands (similar to β3 and β4 of CT296) are highlighted. (C) Crystal Structure of *P. syringae* SyrB2 (PDB 2FCU). Four α-helices, (similarly oriented to α1, α2, α3 and α4 from the I-TASSER model of CT296) and two β-strands (similar to β3 and β4 of CT296) are highlighted. (D) Crystal structure of *H. sapien* PHD2 (PDB 2HBT). Three α-helices, (similarly oriented to α1, α2, and α3+4 from the I-TASSER model of CT296) and two β-strands (similar to β3 and β4 of CT296) are highlighted.


A
α5
α1
α2
α4
β2
β3
B
C
D

core containing the known (2FCU; (Blasiak et al., 2006) and 2HBT; (McDonough et al., 2006)) and proposed (2OPW; (Mihalik et al., 1997)) ligand-binding and catalytic site. There are two β-strands in the CT296 model that are immediately adjacent to the structurally conserved α-helices. These β-strands are oriented similar to the β-sheet that comprise the typical DSBH in Fe(II), 2-oxoglutarate dependent enzymes. It was also apparent that the structures of the three proteins used for this comparison (2HBT, 2OPW, and 2FCU) have only minimal opposing β-strands that compose the canonical DSBH typical of this family of non-heme Fe(II) 2-oxoglutarate dependent enzymes (4, 3 and 6 short sheets, respectively).

***The Crystal Structure of CT296 Shows No Significant Structural Similarity to Fur Family Members and Has No Apparent DNA-Binding Motifs.***

In addition to computational prediction of the structure of CT296, recombinant CT296 was expressed, purified and crystallized. A data set was collected and processed to a resolution of 1.8 Å, with one molecule of CT296 in the asymmetric unit within the crystal lattice (Table II). The structure of CT296 was solved using SAD phasing with selenomethionine-labeled recombinant protein. The crystal structure of CT296 is composed of a single 5-stranded antiparallel β-sheet, supported by 4 α-helices (Fig. 2.1B). The molecular topology of CT296 is (residue number); α1 (L7-H16) – β1 (H18-F21) – α2 (V27-F40) – α3 (G65-V73) – α4 (L75-T83) – β2 (A89-V96) – β3 (D108-C114) – β4 (W122-F126) – β5 (A143-S149) (Fig. 2.1C). The region between α2-α3 (E43-R54) was not modeled due to the lack of electron density, indicating that this region is flexible.

**Table 2.2. Data collection and refinement statistics for CT296**

| **Data Collection** | Apo CT296 | SeMet-CT296 |
|---|---|---|
| Unit-cell parameters (Å) | $a$ = 51.2 $b$ =64.0 $c$ = 46.3 | $a$ = 50.8 $b$ = 63.7 $c$ = 46.3 |
| Space group | $P2_12_12$ | $P2_12_12$ |
| Resolution (Å)[1] | 30.0-1.8 (1.86-1.8) | 63.66-2.5 (2.5-2.5) |
| Wavelength (Å) | 1.000 | 0.979 |
| Observed reflections | 101642 | 65969 |
| Unique reflections | 14832 | 5585 |
| <I>/σ<I>[1] | 25.4 (2.3) | 18.0 (6.1) |
| Completeness (%)[1] | 99.8 (99.9) | 99.9 (99.8) |
| Redundancy[1] | 6.8 (6.2) | 11.8 (12.3) |
| *R*merge (%)[1, 2] | 10.2 (59.4) | 11.8 (48.0) |
| **Refinement** | | |
| Resolution (Å) | 23.1-1.8 | 39.7-2.5 |
| $R_{factor}$ / $R_{free}$ (%)[3] | 20.5/23.8 | 20.4/25.7 |
| No. atoms (protein / water) | 1212/68 | 1147/20 |
| **Model Quality** | | |
| Bond lengths (Å) | 0.015 | 0.011 |
| Bond angles (°) | 1.423 | 1.159 |
| **Average *B* factor (Å$^2$)** | | |
| Protein | 28.6 | 35.7 |
| Water | 31.7 | 32.1 |
| Coordinate error based on Maximum Likelihood (Å) | 0.19 | 0.26 |
| **Ramachandran Plot** | | |
| Favored (%) | 98.5 | 97.8 |
| Allowed (%) | 1.5 | 2.2 |
| Outliers (%) | 0.0 | 0.0 |
| **PDB ID** | 3QH6 | 3QH7 |

[1] Values in parenthesis are for the highest resolution shell.
[2] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the $i$th reflection and <I(hkl)> is the average intensity of all reflections with indices hkl.
[3] $R_{factor} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$; $R_{free}$ is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

Overall, the structure is 19% β-sheet, 32% α-helix, and 49% random coil. No additional non-water molecules, such as divalent cations, were evident in the crystallized protein.

As observed for the I-TASSER model of CT296, the experimentally determined crystal structure does not contain similarity to Fur family members as indicated by a DaliLite pairwise comparison (data not shown). Additionally, the structure of CT296 contains no apparent DNA binding motifs, such as a helix-turn-helix, a leucine zipper, a zinc finger, or a winged helix-turn helix. As before, the structure of CT296 was subjected to analysis by PreDs resulting in a P score of 0.06, well below the typical cutoff point for DNA-binding proteins (0.12). Together, these analyses of the crystal structure of CT296 are in contrast with the precious observations that CT296 is a DNA-binding protein with structural similarity to Fur.

***The Structure of CT296 Predicted by I-TASSER is Similar ($C_\alpha$ r.m.s.d. 2.72 Å) to the CT296 High-Resolution Structure Solved Using X-ray Crystallography.***

Because CT296 lacks homologs by primary sequence, template-based homology modeling cannot be used to generate a model of the structure of CT296. Therefore, *ab initio* modeling must be employed. This type of modeling is a particular challenge in the field of protein structure prediction (Lee et al., 2009). Impressively, the I-TASSER generated model of CT296 has significant overall structural similarity ($C_\alpha$ r.m.s.d. 2.72 Å for 101/137 Cα) to the experimentally determined structure of CT296 (Fig. 2.1). The model and experimentally determined structure of CT296 are primarily α-helical (46% vs. 32%, respectively) but almost 50% of the protein is random coil. In total, there are six secondary structure elements in the experimentally determined structure (α1, α2, α3, α4,

β3, and β5) that are equivalent in the model of CT296 (Fig. 2.1). A notable difference between the two structures is a helix (α3; H51-R94) modeled by I-TASSER that is unstructured in the experimentally determined structure. The most prominent difference between the two structures is the presence of a five-stranded β-sheet (β1-5) in the experimentally determined structure whereas I-TASSER modeled only two β-strands (β2 and β3).

***Structural Homology Searches Using the High-Resolution X-ray Crystallography CT296 Structure Yield Almost Identical Results as the CT296 I-TASSER Model.***

To determine the impact on functional predictions due to the structural differences between experimentally determined and modeled CT296, the crystal structure of CT296 was used to search for structural homologs in the PDB using PSFloger and DALI (Table III). As expected, the scores for top matches improved dramatically (TM-Score range 0.7806-0.6764 and DALI Z-score range 10.5-8.7). Strikingly, all of the top 10 PSFloger and all but one (PutA – PDB 3ITG) of the DALI homologs are retained in the comparable homology search of the I-TASSER model for CT296 (Table I), reinforcing the prediction that CT296 has a structural scaffold similar to non-heme Fe(II) 2-oxoglutarate enzymes. The top hit for both PSFloger and DALI was PHYHD1 (2OPW), an enzyme from *H. sapiens*. Similar to the I-TASSER CT296 model homology search, only three of the proteins (PHYD1, PHD2, and SyrB2) were shared between the top CT296 structure homology matches of PSFloger and DALI.

Based upon the conservation of proteins with structural homology identified between the two search processes and increase in homology scores, it was expected that

## Table 2.3. Top 10 homologous structures to the crystal structure of CT296 as measured by PSFloger and DALI

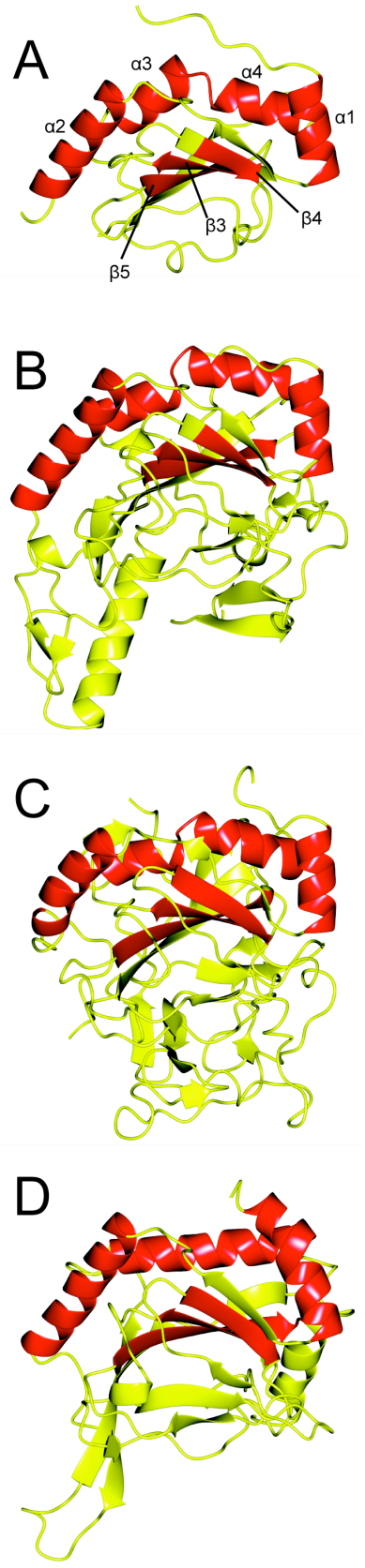
| | PSFloger | | | | | DALI | | |
|---|---|---|---|---|---|---|---|---|
| Rank | TM-Score[1] | PDB File | Protein Name | Classification | Z Score[2] | PDB File | Protein Name | Classification |
| 1 | 0.7806 | 2OPW | PHYHD1 | Hydroxylase (Mihalik et al., 1997) | 10.5 | 2OPW | PHYHD1[3] | Hydroxylase (Mihalik et al., 1997) |
| 2 | 0.7537 | 2FCU | SyrB2 | Halogenase (Blasiak et al., 2006) | 10.5 | 3GJB | CytC3 | Halogenase (Wong et al., 2009) |
| 3 | 0.7473 | 2A1X | PHYHD1 | Hydroxylase (McDonough et al., 2005) | 10.3 | 3NN1 | Hal | Halogenase (Khare et al., 2010) |
| 4 | 0.7221 | 2CSG | YbiU | Putative oxioreductase | 9.9 | 3EMR | ECTD | Dioxygenase (Reuter et al., 2010) |
| 5 | 0.7138 | 2DBN | JW0805 | Unknown Function | 9.7 | 3HQU | PHD2[3] | Hydroxylase (Chowdhury *et al.*, 2009) |
| 6 | 0.7113 | 2HBT | PHD2 | Hydroxylase (Warshakoon et al., 2006) | 9.4 | 2FCU | SyrB2 | Halogenase (Blasiak et al., 2006) |
| 7 | 0.6905 | 1UNB | DAOCS | Oxioreductase (Valegard et al., 2004) | 9.1 | 3GZE | P4H[3] | Hydroxylase (Koski *et al.*, 2009) |
| 8 | 0.6802 | 2RDN | PtlH | Hydroxylase (You et al., 2007) | 9.0 | 2RG4 | Q2CBJ1 | Unknown Function |
| 9 | 0.6793 | 3GJB | CytC3 | Halogenase (Wong et al., 2009) | 8.8 | 3MGU | Tpa1 | Hydroxylase (Kim et al., 2010) |
| 10 | 0.6764 | 1GP6 | ANS | Dioxygenase (Wilmouth et al., 2002) | 8.7 | 3ITG | PutA | Oxioreductase (Srivastava *et al.*, 2010) |

[1]TM-score is a measurement of similarity between two protein structures and it varies from 0 to 1, with 1 indicating a perfect alignment. A score of 0.5 or higher indicates significant similarity (Zhang & Skolnick, 2005).

[2]Structures with Z-scores above 2 are considered to be significant matches (Holm et al., 2008).

[3]This protein was listed multiple times on the DALI output, with various minor differences (ligands, etc.); the redundant structures are not listed here.

**FIG 2.3. Ribbon depiction of experimentally determined structure of CT296 aligned with three proteins exhibiting the highest structural homology.** Regions of structure that are conserved between all four structures are highlighted in red. (A) Crystal structure of CT296. Four α-helices are highlighted, α1 (L7-H16), α2 (V27-F40), α3 (G65-V73) and α4 (L75-T83). Three β-strands are highlighted, β3 (D108-C114), β4 (W122-F126) and β5 (A143-S149). (B) Crystal structure of *H. sapien* PHYD1 (PDB 2OPW). Four α-helices, (similarly oriented to α1, α2, α3 and α4 from the I-TASSER model of CT296) and three β-strands (similar to β3, β4 and β5 of CT296) are highlighted. (C) Crystal Structure of *P. syringae* SyrB2 (PDB 2FCU). Four α-helices, (similarly oriented to α1, α2, α3 and α4 from the I-TASSER model of CT296) and three β-strands (similar to β3, β4 and β5 of CT296) are highlighted. (D) Crystal structure of *H. sapien* PHD2 (PDB 2HBT). Three α-helices, (similarly oriented to α1, α2, and α3+4 from the I-TASSER model of CT296) and three β-strands (similar to β3, β4 and β5 of CT296) are highlighted.

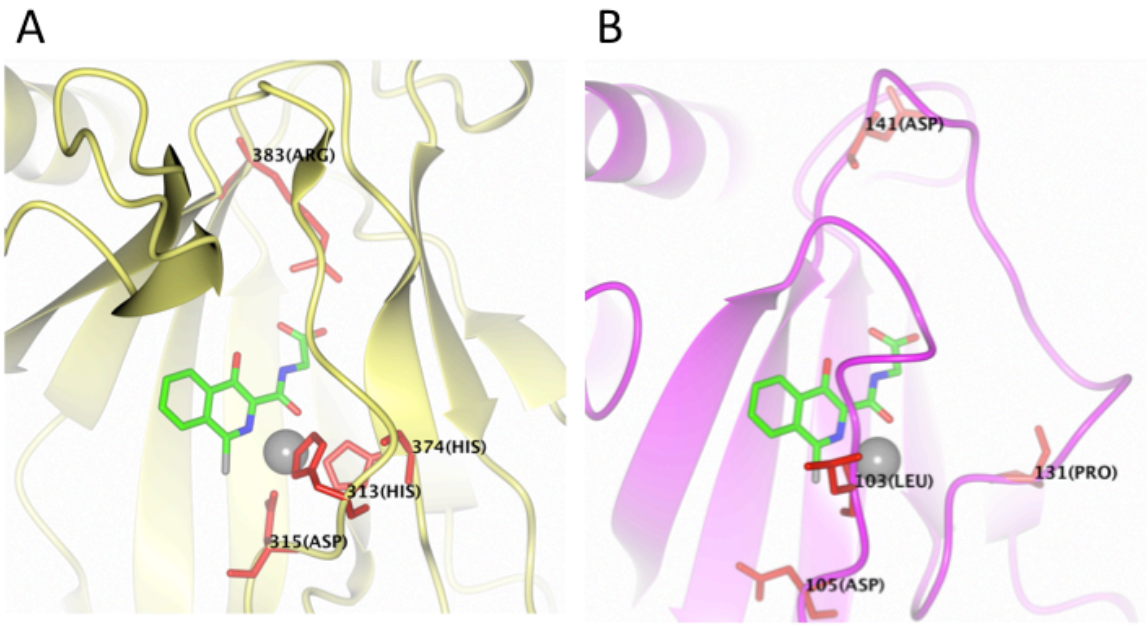

A
α3
α4
α1
α2
β3
β4
β5
B
C
D

additional structural elements would be shared. The experimentally determined structure of CT296 was aligned with three of its structural homologs from the PDB, in order to determine which structural regions are similar (Fig. 2.3). The three structural homologs selected are the same three that were aligned to the I-TASSER model of CT296 (Fig. 2.1; 2OPW, 2FCU, and 2HBT); these proteins can be found on the PSFloger and the DALI lists of both the model and the experimental structure of CT296. A total of seven secondary structure elements are equivalent between CT296 and the three selected homologs, three α-helices (α1, α2, α3, and α4) and three β-sheets (β3-5). Importantly, six (α1, α2, α4, α5, β2, and β3) of these seven structural elements were also found to be equivalent between the model of CT296 and the homologs. This suggests that these elements are the underlying reason for the large number of proteins that contain structural homology to both the model and the experimental structure of CT296.

***CT296 does not possess the molecular capabilities to bind to Fe(II) or 2-oxoglutarate***

Despite the overall global structural similarity between CT296 and many Fe(II) and 2-oxoglutarate containing enzymes, CT296 lacks the appropriate residues to properly coordinate these molecules. For example, CT296 shares structural homology to the prolyl hydroxylase PHD2 from *Homo sapiens* (Table 2.1 and 2.3). The enzymatic function of PHD2 requires Fe(II) and 2-oxoglutarate, which are coordinated by four residues (His313, Asp315, His 374, Arg383) (Chowdhury et al., 2009). The coordination of these molecules can be seen in the crystal structure (Fig 2.4A). However, these residues critical to the enzymatic function of PHD2 do not appear to be retained in CT296. The corresponding residues in CT296 (Leu103, Asp105, Pro131, Asp141) would

**FIG. 2.4. CT296 does not have the conserved residues nor their proper orientation for the coordination of Fe(II) or 2-oxoglutarate.** (A) A structural homolog to CT296, PHD2 (PDB ID: 2HBT) coordinates an Fe(II) atom and a 2-oxoglutarate molecule using the amino acid sidechains His313, Asp315, His 374, Arg383. (B) CT296 did not crystallize with either Fe(II) or 2-oxoglutarate bound, but a superposition of CT296 onto the structure of PHD2 reveals where these molecules would rest in the structure of CT296. The amino acid sidechains that correspond to those of PHD2 that coordinate these molecules are labeled (Leu103, Asp105, Pro131, Asp141); none of these residues are in an appropriate orientation to coordinate these molecules.


A
383(ARG)
374(HIS)
313(HIS)
315(ASP)
B
141(ASP)
103(LEU)
131(PRO)
105(ASP)

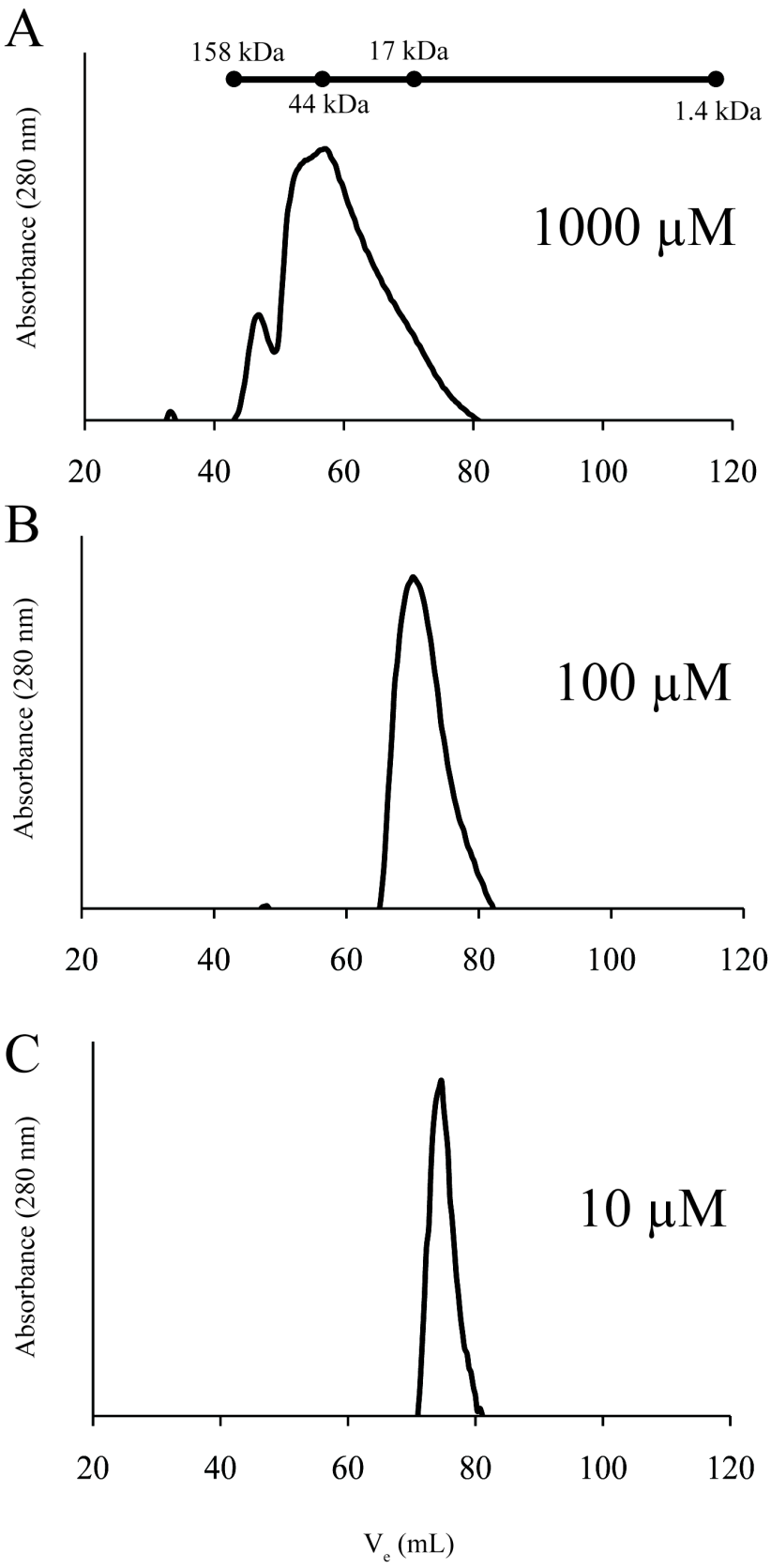
likely not coordinate an iron atom or a 2-oxoglutarate molecule (Fig. 2.4). Similar comparisons have been made for EctD from *Virgibacillus salexigens* (Reuter et al., 2010) and CytC3 from *Streptomyces* (Wong et al., 2009) with the same conclusion as PHD2.

***Recombinant CT296 is Predominantly Monomeric at Physiological Concentrations.***

Recombinant Fur has been shown to exist predominantly as a homodimer in solution (D'Autreaux *et al.*, 2007, Michaud-Soret *et al.*, 1997). At high protein concentrations, Fur has been shown to form other higher-order structures, but the predominant species present at physiological concentrations (1-17 μM) of the protein is that of a dimer (D'Autreaux et al., 2007). Likewise, recombinant CT296 has previously been reported to form homodimers (Wyllie & Raulston, 2001). Given the structure of CT296 and its lack of similarity to Fur, the quaternary structure of recombinant CT296 was reassessed, including a study of the effects of protein concentration. To accomplish this, recombinant CT296 was purified and subjected to size exclusion chromatography (SEC). Although SEC is not an appropriate technique for precise quantification of molecular weight, a standard curve can be constructed based on the elution of profiles of proteins of known molecular weight. A protein standard solution containing γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa) and vitamin $B_{12}$ (1.4 kDa) was used to approximate the molecular weights of CT296 at varying concentrations. The molecular weight of CT296 is 17.9 kDa and so a dimer of CT296 would be 35.8 kDa.

At high protein concentration, (1000 μM) a predominant peak that corresponds to the approximate molecular weight of a CT296 dimer was observed (48.0 kDa; Fig. 2.5A). Other peaks that correspond to even higher order structures were also observed at this

**FIG. 2.5. Size exclusion chromatograms of recombinant CT296 at varying concentrations.** The molecular weight of monomeric CT296 is 17.9 kDa. The molecular weight of a CT296 dimer would be 35.8 kDa. The components of a molecular weight ladder are marked at their respective elution volumes on a line above the chromatograms. (A) At a concentration of 1000 μM, CT296 is predominantly present in a peak that approximately corresponds to the molecular weight of a dimer. Other, even higher order, structures can also be seen, in lower abundance. (B) At a concentration of 100 μM, CT296 elutes within fractions that correspond to the molecular weight of a monomer. (C) At a concentration of 10 μM, CT296 elutes as a peak that corresponds to the molecular weight of a monomer.

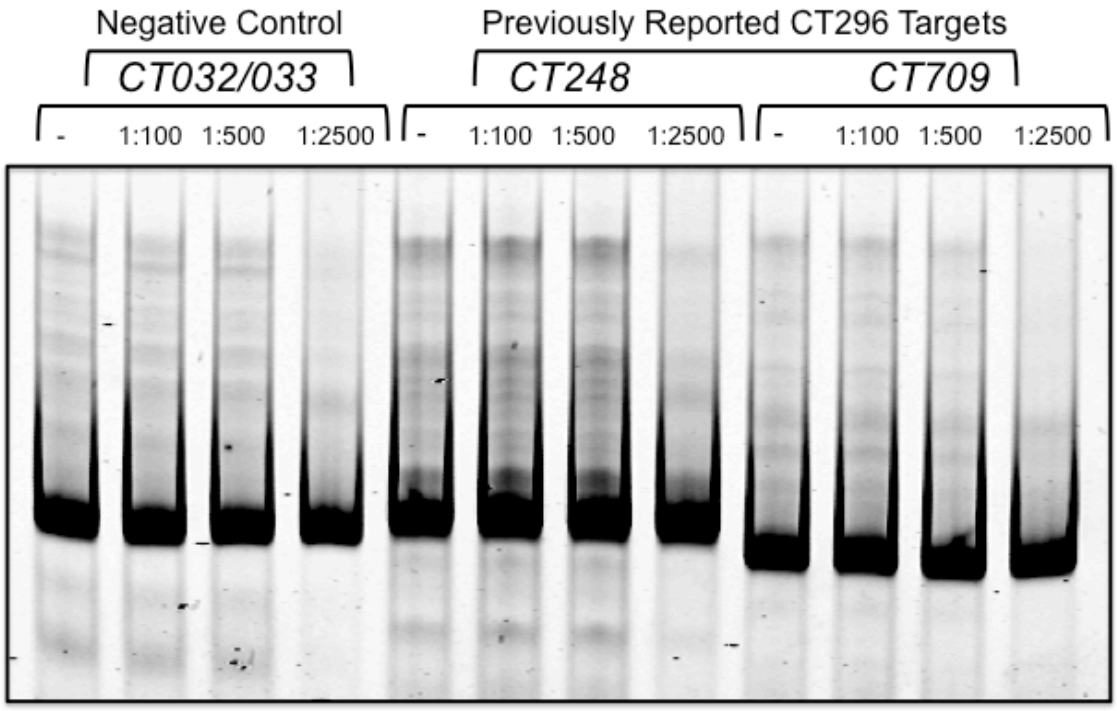

A
158 kDa
17 kDa
44 kDa
1.4 kDa
Absorbance (280 nm)
1000 μM
20
40
60
80
100
120
B
Absorbance (280 nm)
100 μM
20
40
60
80
100
120
C
Absorbance (280 nm)
10 μM
20
40
60
80
100
120
V_e (mL)

concentration (109 kDa; Fig. 2.5A). At a lower protein concentration, (100 μM), there was a single peak which corresponds to the molecular weight of a monomer of CT296 (17.0 kDa; Fig. 2.5B). At the lowest concentration tested (10 μM), an approximate monomer of CT296 was observed (12.1 kDa; Fig. 2.5C). The physiological concentration for CT296 is not known, but physiological concentrations for Fur have been reported to be between 1 and 17 μM (D'Autreaux et al., 2007). These data show that, unlike Fur, CT296 exists predominantly as a monomer in solution at concentrations of 10-100 μM. This suggests that CT296 may lack other functional properties of Fur or other Fur-like repressors.

***CT296 Does Not Bind to Previously Reported Chlamydial Promoter DNA by Electrophoretic Mobility Shift Assay.***

In previous reports, CT296 was shown to bind to DNA in the context of an Electrophoretic Mobility Shift Assay (EMSA) (Rau et al., 2005, Wyllie & Raulston, 2001). Given the apparent absence of DNA-binding motifs in both the modeled and the experimentally determined structures of CT296, the DNA binding properties of CT296 were reexamined. The promoters of two genes that were previously shown to bind to CT296 by EMSA were selected for analysis (Rau et al., 2005); namely, *CT248* and *CT709*. The exact sequences of DNA from these promoters that were used in the previous study were not reported. However, the recent determination of the transcription start site (TSS) for these genes (Albrecht *et al.*, 2010) allowed for the design of ~400 base pair regions that include at least 150 bases upstream and downstream of the TSS. To serve as a negative control, ~400 base pairs from the region of DNA between the two converging

**FIG. 2.6. Electrophoretic Mobility Shift Assay (EMSA) using recombinant CT296.** Three different promoter regions of DNA were tested with varying DNA:protein molar ratios; 1:100, 1:500 and 1:2500 (- indicates the absence of protein). *CT032/033* is an intergenic region between two converging genes and is used as a negative control. *CT248* and *CT709* had been previously reported to be bound by recombinant CT296 in the context of an EMSA.


Negative Control
Previously Reported CT296 Targets
CT032/033
CT248
CT709
-
1:100
1:500
1:2500
-
1:100
1:500
1:2500
-
1:100
1:500
1:2500

genes *CT032* and *CT033* was used. No shifts in DNA were detected with any of the promoter regions analyzed at any of the DNA:protein molar ratios used, even at the highest ratio of 1:2500 (Fig. 2.6). These findings are in direct contrast with the previous observations that CT296 binds to chlamydial promoter DNA.

## Chapter II

## Discussion

One of the major observations of this study is that the experimentally determined properties of CT296 are not consistent with previous reports regarding the function of this protein. CT296 had been characterized previously (Rau et al., 2005, Wyllie & Raulston, 2001) as a chlamydial homolog to the *E. coli* Fur protein—a metal-responsive transcriptional repressor, termed DcrA (Divalent cation-dependent regulator A). The evidence for this functional assignment came primarily from *E. coli* Fur cross-reactive antibodies, the ability of CT296 to functionally complement Fur in *E. coli*, and the ability CT296 to bind to chlamydial promoter DNA sequences *via* EMSA. However, the experimentally determined structure of CT296 bears no structural similarity to Fur, does not contain an identifiable DNA binding motif, and putative regions for DNA binding were not identified using PreDs. No divalent cations were observed in the crystal structure and a metal-binding site could not be identified. Additionally, attempts to demonstrate that CT296 can bind to the chlamydial DNA targets reported previously (Rau et al., 2005) were unsuccessful (Fig. 2.5). In the previous analysis, EMSA was used to validate targets isolated from a screen designed to find regions of chlamydial DNA that bound *E. coli* Fur. In that analysis, the variable-sized regions of chlamydial DNA directly from a screen were analyzed for CT296 binding. Importantly, the molar ratio of DNA:protein in reaction mixtures were not stated and precise regions of the chlamydial genome tested were not reported. In this report, analyses used regions of DNA designed based on the recently reported transcription start sites for these genes (Albrecht et al.,

2010). At a wide range of DNA:protein ratios and with the inclusion of different divalent cations (data not shown), DNA binding by CT296 could not be detected. Finally, CT296 was previously reported to form dimers that are resistant to denaturation by SDS-PAGE (Wyllie & Raulston, 2001). During protein purification, only monomers of CT296 were observed by SDS-PAGE. Using gel filtration chromatography, higher order structures corresponding roughly to the molecular weight of a dimer can be seen, but this is only at very high concentrations (Fig. 2.4; 1000 μM) of protein where even higher order structures are observed. A direct explanation for these apparently discordant observations is not evident but not without significant efforts.

One of the primary purposes for elucidating the structure (experimentally or computationally) of a protein of unknown function is to enable functional prediction and facilitate further analyses to confirm the predicted function. Functional assignment typically begins by searching for structural homologs to the protein of unknown function (Adams *et al.*, 2007). In the case of CT296, this search yielded many structures, all with significant structural similarity (Table 2.1 and 2.3). According to the protein family database, all of the proteins with structural similarity are members of the cupin superfamily of proteins. This superfamily is functionally diverse, but all members share a characteristic β-barrel (Dunwell *et al.*, 2004, Khuri *et al.*, 2001). The proteins with structural similarity to CT296 have relatively different enzymatic functions (hydroxylase, halogenase, dioxygenase, reductase.) but all are members of the predominant cupin subclass of 2-oxoglutarate-Fe(II) dependent enzymes (Dunwell et al., 2004). These proteins have a well-defined fold termed a double stranded β-helix (DSBH) in which the

enzymatic processes typically occur, including the binding site of the enzymatic co-factors Fe(II) and 2-oxoglutarate.

Despite the correlation of protein function and structural similarity, there are many observations that do not support functional annotation of CT296 as a 2-oxoglutarate-Fe(II) dependent enzyme in the cupin superfamily. Firstly, the structure of CT296 contains an incomplete DSBH fold, the structural motif that is characteristic of cupin superfamily proteins. The structure of CT296 contains one side of the β-helix consisting of five β strands; however, it lacks an opposing β-sheet that would form the typical DSBH fold. Secondly, no Fe(II) or 2-oxoglutarate were identified within the CT296 protein structure. Furthermore, attempts to soak CT296 protein crystals with these compounds (independently or in combination) or to generate new CT296 protein crystals in the presence of these compounds did not yield stable crystals (data not shown). Lastly, the vast majority of key residues that function in coordinating Fe(II) and 2-oxoglutarate for the proteins with structural homology are not conserved in CT296.

In contrast, there are many structural elements of CT296 that are remarkably similar to the arrangements of its structural homologs, as illustrated in Fig. 2.3. In fact, several secondary structure elements of CT296, four α-helices (α1-4) and three β-sheets (β2-4), are all localized together and similarly aligned with homologs. The four α-helices have not been shown to play a role in ligand binding or catalysis but appear to serve as structural support for the double stranded β-helix (Blasiak et al., 2006, McDonough et al., 2005). While CT296 appears to contain an incomplete double stranded β helix, this region of structural homology may be stabilizing the β-sheet region of CT296. Additionally, CT296 could function as a part of a higher order structure and that the

conserved region of surface exposed α helices could stabilize an interaction interface to another protein(s). Importantly, additional experimental evidence is needed to appropriately define the function of CT296.

It is important to note that the structural information regarding CT296 was, on its own, somewhat misleading. By examining only global structural similarity, one might have concluded that CT296 in an Fe(II) and 2-oxoglutarate-dependent enzyme. However, by examining the structure and the primary sequence more closely and by attempting to identify the conserved features of its predicted function (the cofactor coordinating residues), it was concluded that CT296 likely does not function as an Fe(II) and 2-oxoglutarate-dependent enzyme. This illustrates that point that, when utilizing structural information to inform function, the structure must be examined for salient features instead of simply being classified based on global similarity.

Another of the observations of this study was the ability of I-TASSER to accurately model the three-dimensional structure and aid in the prediction of function of a protein encoded by *C. trachomatis* that does not share primary sequence homology with any proteins with functional assignment. While accuracy can be a subjective measurement, the $C_\alpha$ r.m.s.d. (2.72 Å) between the CT296 model and experimentally determined structure provides a relative measurement that the model generated by I-TASSER is a very close representation of the physical structure of the protein. It is especially encouraging that none of the templates used in the I-TASSER modeling had sequence identity to the target higher than 25%. Proteins with such low similarity to any of the templates represent a 'twilight zone' in the structural genome, which is one of the most challenging categories of targets for protein structure prediction.

As typically performed, the *Chlamydia trachomatis* genomes were annotated using primary sequence homology to assign function. After applying this strategy, chlamydial genomes have an abundance of ORFs of unknown function – approximately 25% of encoded proteins (Carlson et al., 2005, Unemo et al., 2010, Thomson et al., 2008, Stephens et al., 1998). For *Chlamydia pneumoniae*, another chlamydial species with an immense impact on public health, the situation is worse with approximately 35% of encoded proteins characterized as hypothetical. In addition to this report, there are limited data in *Chlamydia* to demonstrate the utility of protein structure to facilitate functional assignment for uncharacterized proteins. However, a highly supportive and recent example is an analysis on protein encoded by open reading frame CT670. CT670 has very limited primary sequence similarity to any other protein, but the X-ray crystal structure supports this protein functions in the chlamydial type three secretion system and is homologous to YscO from *Yersinia pestis* (Lorenzini *et al.*, 2010). Furthermore, this protein was demonstrated to interact *in vivo* with many other type III secretion components in *Chlamydia* (Lorenzini et al., 2010) and functional prediction of CT670 is supported by the inclusion of the encoding gene in an operon that contains type three secretion system homologs (Hefty & Stephens, 2007). Outside of *Chlamydia*, there is an abundance of reports on hypothetical proteins having function assigned based on structure (Janowski *et al.*, 2009, Sim *et al.*, 2009, van Staalduinen *et al.*). Therefore, more thorough annotations of bacterial genomes (and the chlamydial genome in particular) would benefit greatly from the incorporation of computation and/or experimentally determined structural information.

The studies presented here have shown that I-TASSER is effective at predicting the structures of proteins that have no function homologs by primary sequence and would be very effective as a complement to primary sequence annotation. While this study also highlights the limitations of structural based functional annotation, the structural information provides a wealth of molecular information that can be leveraged to elucidate the biochemical function and biological role of a given protein – for which relatively little is previously known.

# Chapter III. Localization studies of the Chlamydia trachomatis protein CT584, a potential type III secretion needle tip protein

## Abstract

*Chlamydia* species use a Type Three Secretion System (T3SS) to facilitate entry into the host cell, maintain their parasitophorous vacuole within the host, and possibly to exit host cells. Key in the ability of the T3SS to operate is the T3SS needle tip protein that, in other bacteria, sense environmental conditions to initiate T3SS dependent translocation. These proteins have been demonstrated to have at least some portion of the protein exposed to the extracellular environment. A putative T3SS needle tip protein in *Chlamydia trachomatis*, CT584, was recently reported. This identification was based on the protein's computationally predicted structure and experimentally determined biophysical characteristics, both of which were similar to other T3SS needle tip proteins. It was also recently reported that CT584 is associated with the chlamydial T3SS needle protein in the context of a yeast two-hybrid screen. To further test the hypothesis that CT584 is a surface-exposed T3SS needle tip protein, protease protection assays were employed. This analysis indicated that undetectable levels of CT584 are surface exposed on EBs. Similarly, antibodies against CT584 do not bind to the protein on the surface of EBs as demonstrated by direct IFA. Additionally, confocal microscopy was employed to analyze infected cells; these studies suggest that CT584 is primarily localized to the cytoplasm in RBs. Together, these data do not support the role of CT584 as a surface exposed T3SS needle tip protein.

## Introduction

Human infections by species of *Chlamydia* have a significant impact on public health worldwide. Different strains of the bacterium *Chlamydia trachomatis* are capable of infecting the ocular epithelium, causing blinding trachoma, or infecting the genital tract, causing sexually transmitted disease (Cecil et al., 2001, Falk et al., 2005, Mariotti et al., 2009, Resnikoff et al., 2004). All chlamydial organisms are obligate intracellular parasites and their pathogenesis is intimately linked to their unique biphasic developmental cycle. During this cycle, metabolically inactive, infectious forms of the bacteria referred to as Elementary Bodies (EBs) bind to and enter host cells. Once inside the host cells, chlamydial organisms convert to non-infectious metabolically active forms termed Reticulate Bodies (RBs) and replicate within a parasitophorous vacuole termed an inclusion, evading immune response and apoptosis as they divide (Zhong, 2009). Finally, RBs begin to asynchronously convert back to EBs, which are released from the host cell by lysis or extrusion at about 48 hours post infection (Hybiske & Stephens, 2007).

The genome of *Chlamydia trachomatis* contains almost all of the components for a functional Type III Secretion System (T3SS) (Stephens et al., 1998). This secretion system is composed of a large multi-protein apparatus termed an injectisome and is widely used by pathogenic Gram negative bacteria to translocate effector proteins from the bacterial cytoplasm directly into their host cells. At several points during their developmental cycle, *Chlamydiae* employ a Type III Secretion System (T3SS) to allow infection and to subvert normal cell processes. Notably, T3SS effector proteins have

been implicated in playing vital roles in invasion (by EBs), inclusion biogenesis and manipulation of the host immune system (by RBs). The protein components involved in this process are typically well conserved among bacterial species that possess T3SS, including species of *Yersinia*, *Salmonella*, and *Shigella flexneri* (Hueck, 1998). However, homologs to many of these proteins are not apparent in chlamydial genomes, especially effector proteins. Efforts to identify these proteins through a variety of experimental and computational methods has been fairly successful (Peters *et al.*, 2007).

Among these T3SS proteins that are conserved among other bacteria but not apparent within chlamydial species is the needle tip protein. This protein sits atop the 40-80 μm long, extracellular needle and has been shown to regulate the secretion process by sensing contact with the host cell (Espina *et al.*, 2006). Additionally, this protein is of interest because the *Yersinia pestis* homolog has been shown to an effective component of a vaccine against this pathogenic bacterium (Overheim *et al.*, 2005). Importantly, by primary sequence homology, there is no apparent homolog by to needle tip protein in any chlamydial genome.

Experimental evidence suggests that the protein CT584 (to use the nomenclature of *C. trachomatis*) may be the chlamydial needle tip protein. In a study utilizing the yeast-two hybrid assay, the interactions between known components of the chlamydial T3SS and chlamydial proteins of unknown function were examined (Spaeth *et al.*, 2009). Among the proteins of unknown function was CT584, which was shown to interact with CdsF, the T3SS needle protein, a property consistent with the function of CT584 as a needle tip protein. Additionally, the biophysical properties of the CT584 protein, as assessed by a battery of different tests (including circular dichroism, Fourier transformed

infrared spectroscopy and UV absorbance spectroscopy) showed that this protein has structural properties akin to similarly characterized, known needle tip proteins (Markham *et al.*, 2009).

All of the evidence that supports the role of CT584 as the chlamydial needle tip protein is indirect; based on the behavior of the protein *in vitro* or in a heterologous system. In order to more confidently identify the function of CT584 as a needle tip protein, more direct methods of experimentation must be employed. Therefore, we sought to employ various techniques to examine the cellular localization of the CT584 protein. This approach takes advantage of the uncommon localization of needle tip proteins; namely, that they localize extracellularly on the ends of the T3SS needles. Ultimately, these studies indicate that CT584 is not localized primarily to the outside of chlamydial cells (either EB or RB).

## Chapter III

## Materials and Methods

*Proteinase K protection assay*- For each condition, 100 µL of highly purified EBs were diluted 1:4 in 1x PBS. EBs were purified using a 30% Renografin density gradient, as described previously (Scidmore, 2005). Conditions with soluble CT584 protein were supplemented with 0.1 µg of purified recombinant CT584, purified as described previously (Markham et al., 2009). Conditions without recombinant CT584 received an equal volume of CT584 buffer (40 mM tris, pH 8.0, 500 mM NaCl, 5% glycerol). Conditions with proteinase K were supplemented with 40 µg of purified proteinase K (Sigma-Aldritch); conditions not treated with proteinase K were supplemented with an equal volume of proteinase K buffer (50 mM Tris pH 8.0, 10 mM $CaCl_2$). Following proteinase K addition, tubes were incubated for 30 minutes at 37 °C, after which PMSF (Thermo Pierce) was added to a final concentration of 5 mM. Samples were vortexed, supplemented with final sample buffer, boiled for 10 minutes and then loaded onto a 15% SDS-PAGE gel.

*Immunofluorescence of EBs*- Highly purified EBs from *C. trachomatis* (serovar L2 / 434 / Bu) were diluted 1:17 in 1x PBS and were incubated on glass coverslips for one hour at room temperature, to allow EBs to adhere to the glass. EBs were then fixed by the addition of 1 mL of methanol and incubated for 10 minutes at room temperature. Fixed EBs were blocked in a solution of 1x PBS containing 5% BSA for 1 hour at room temperature. All slides were stained using a commercially available kit, which labels

EBs green (Trinity Biotech) for 1 hour at room temperature. Each slide was probed with a different primary antibody, anti-MOMP (ViroStat), anti-RopA or anti-CT584, for 1 hour at room temperature. Anti-RopA and anti-CT584 antibodies had been purified from sera using AminoLink Plus coupling resin (Thermo Scientiic), as per manufacturer's protocol. Finally, slides were probed with fluorescently conjugated secondary antibodies that label the protein of interest (MOMP, RpoA or CT584; Invitrogen) in red for 1 hour at room temperature. Coverslips were mounted on glass slides using mounting medium (Vector Labs) and then visualized using an IX71 inverted microscope (Olympus).

*Confocal microscopy of infected cells*- A monolayer of L929 cells on an 8-well μ-slide plate (Ibidi) was infected with *C. trachomatis* to a multiplicity of infection of 1.0. At 24 hours post infection, cells were fixed with methanol and washed. Fixed cells were probed with primary antibodies, anti-CT584 and anti-MOMP (ViroStat) for 1 hour at room temperature. Cells were washed with 1x PBS and probed with secondary antibodies (Invitrogen) for 1 hour at room temperature. Cells were mounted and visualized using a spinning disk confocal microscope (Olympus). Images were de-convolved using the nearest-neighbor method.

# Chapter III

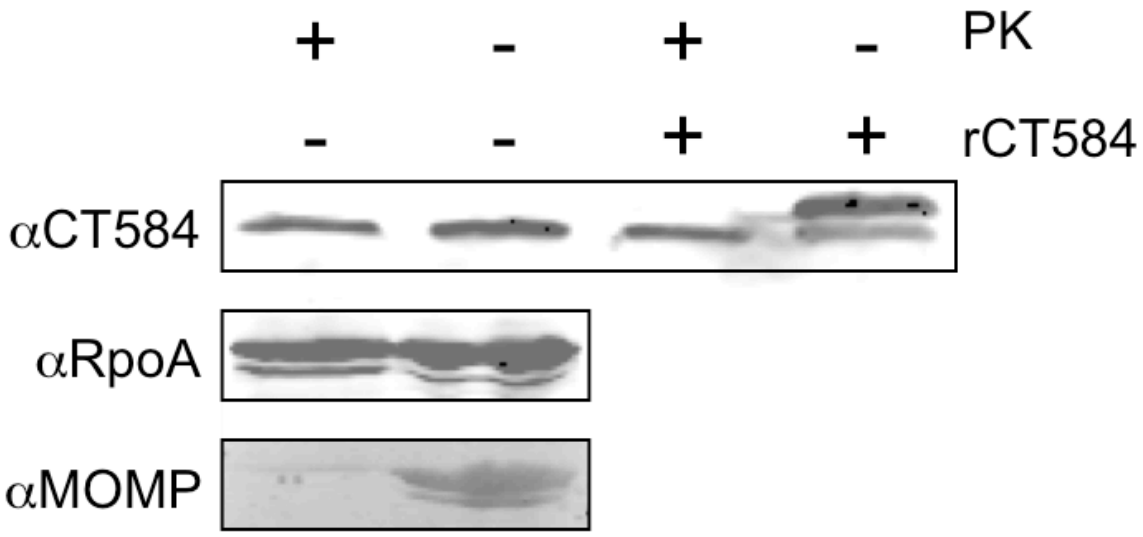
# Results

## *CT584 is not primarily surface exposed on EBs*

A distinctive property Type III Secretion System (T3SS) needle tip proteins is their extracellular localization. As a proposed chlamydial T3SS needle tip protein CT584 should therefore likely also be localized extracellularly. In order to examine the localization of CT584, several different methods were employed. Initially, a proteinase K protection assay was used. In this assay, whole cells (in working with chlamydia, EBs) are exposed to the broad-specificity protease Proteinase K (PK). This protease is capable is digesting nearly any protein, based on its preferred cleavage site at the carboxyl end of aliphatic or aromatic amino acids, which are common in virtually all proteins (Kraus & Femfert, 1976). Because the integrity of the outer membrane of the cells is not compromised in this assay, proteins that are surface exposed will be digested by the protease while proteins not exposed to the surface will be unaffected.

To ensure that the protease was digesting only surface-exposed proteins, the protein levels of the major outer membrane protein MOMP were examined. Although this integral membrane protein has minimal surface exposure, this was sufficient exposure and the protein was digested thoroughly in the presence of PK (Fig. 3.1). Additionally, protein levels of the alpha subunit of RNA polymerase (RpoA) were examined. This protein functions in the cytoplasm and therefore should not be digested. Following PK treatment, RpoA levels are unaffected, demonstrating that EB membrane

**FIG 3.1. Proteinase K protection Assay**. Highly purified EBs were either treated or mock treated with Proteinase K, a highly active broad-specificity protease. Following treatment, protease activity was inhibited and EBs were lysed, run on an SDS-PAGE gel, subjected to western blot, and then probed with antibodies. Blots were probed again RpoA, a subunit of RNA polymerase, known to have cytoplasmic localization, MOMP, also known as OmpA, the major outer membrane proteins and against CT584. Although it has very broad specificity for cleavage, the ability of proteinase K to digest CT584 protein was confirmed by the addition of purified recombinant CT584 to the EBs prior to digestion.


+ - + - PK
- - + + rCT584
αCT584
αRpoA
αMOMP

integrity was not compromised (Fig. 3.1). Finally, levels of CT584 were analyzed. Following treatment by PK, CT584 protein levels were not significantly affected (Fig. 3.1). This indicates that the majority of CT584 in the cells it not surface exposed.

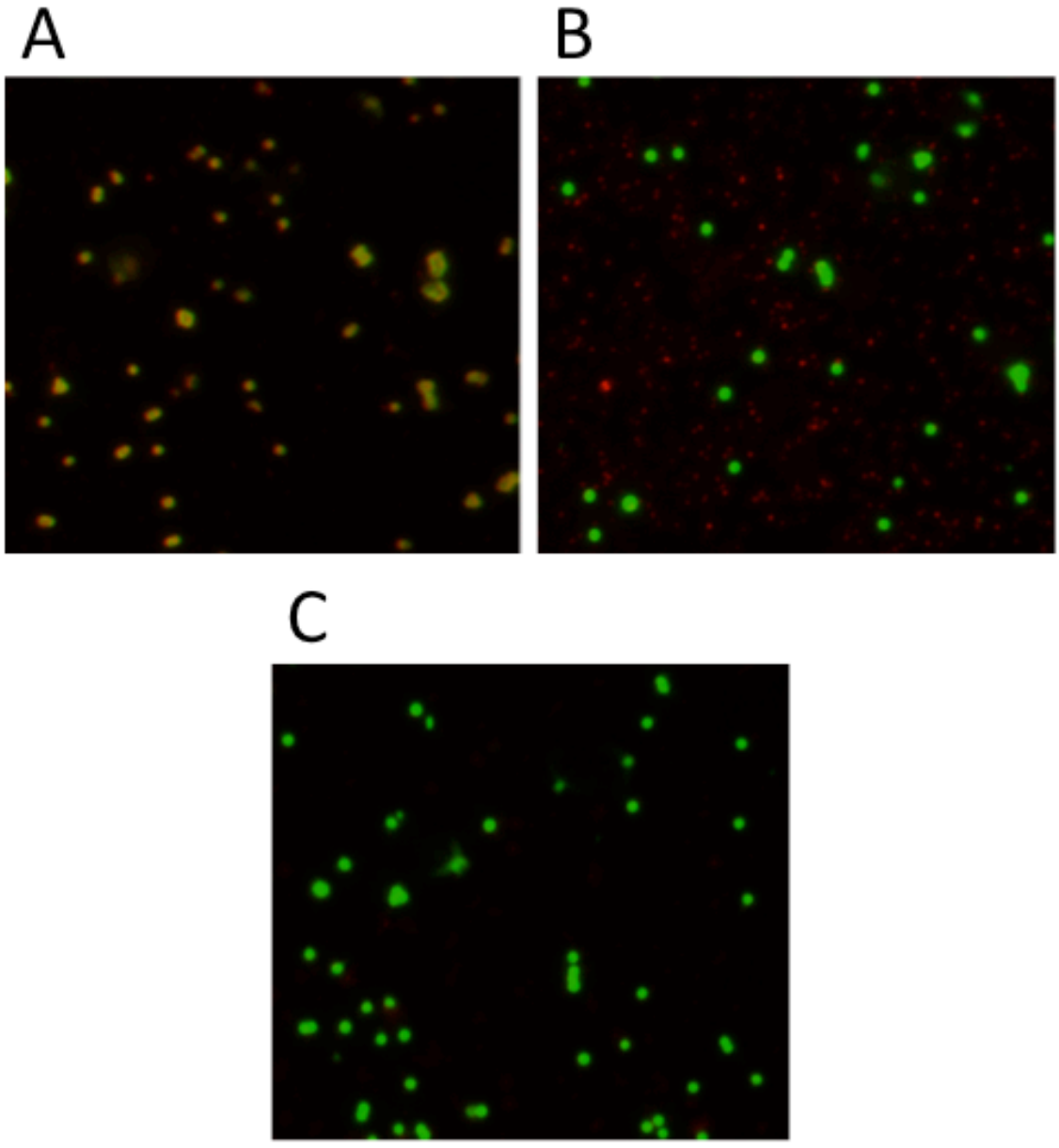
Although the specificity of Proteinase K is very broad and the primary sequence of the CT584 protein confirms the presence of over 50 potential PK recognition sites, it is possible that the CT584 protein in not digestible by PK. To confirm the susceptibility of CT584 to PK, EBs were supplemented with purified recombinant CT584. This protein runs at a slightly higher molecular weight than endogenous CT584 due to a polyhistidine tag utilized for purification of the recombinant protein. Following PK treatment, this supplementary extracellular CT584 was digested by PK, confirming that the protein is susceptible to digestion, if properly exposed (Fig 3.1).

Importantly, this result does not preclude surface exposure for this protein. It is possible that the majority of CT584 in the cell remains in the cytoplasm with a minority of the protein localizes to the surface. In this case, PK would digest the surface-exposed subpopulation of protein but since total protein levels are assessed, its absence would not be detected.

***CT584 could not be detected on the surface of EBs by immunofluorescence***

In order to more precisely elucidate the localization qualities of CT584, direct immunofluorescence of EBs was employed. If a small sub-population of CT584 is localized to the membrane, then this technique should be successful in detecting it. In this experiment, EBs were attached to a glass coverslip, fixed, blocked and then probed with a fluorescently conjugated antibody that binds to the MOMP protein on the outer

**FIG. 3.2. CT584 is not detected on the surface of EBs by immunofluorescence**. EBs were fixed on glass coverslips and then blocked. In all conditions (A-C), EBs were labeled with a fluorescently-labeled primary antibody that binds to EBs and fluoresces green. (A) EBs were also probed with an anti-MOMP antibody, followed by a secondary antibody that fluoresces red. (B) EBs were also probed with an anti-RpoA antibody, followed by a secondary antibody that fluoresces red. (A) EBs were also probed with an anti-CT584 antibody, followed by a secondary antibody that fluoresces red.


A
B
C

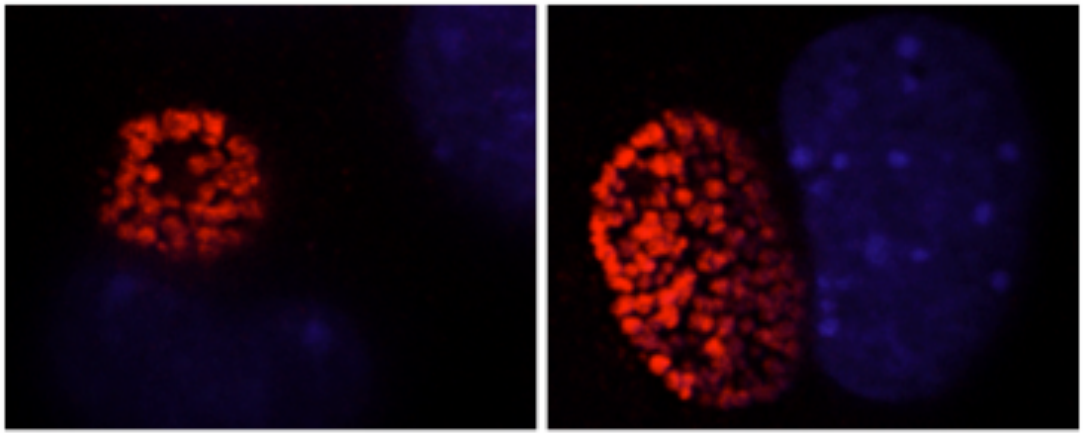
surface of chlamydial organisms and fluoresces green; therefore, all conditions tested have green fluorescing EBs (Fig. 3.2). As a control to ensure that the secondary antibody is functioning properly, in one condition, EBs were probed with a primary anti-MOMP antibody followed by a secondary antibody that fluoresces red. Therefore, in this condition, EBs were labeled with both green and red, producing yellow colored EBs (Fig. 3.2A). To ensure that the fixing procedure was not compromising the integrity of the EB membrane, in another condition, EBs were probed with a primary against RpoA (the chlamydial alpha subunit of RNA polymerase, a protein that should not have surface localization) followed by a secondary antibody that fluoresces red. This produced EBs that fluoresce green but not red, illustrating that internally localized proteins are not available for antibody binding (Fig. 3.2B). In the final condition, EBs were probed with antibodies against CT584. EBs were not labeled in red in this condition, indicating that CT584 is not available for antibody binding on EBs.

***CT584 does not exhibit primarily surface localization in RBs***

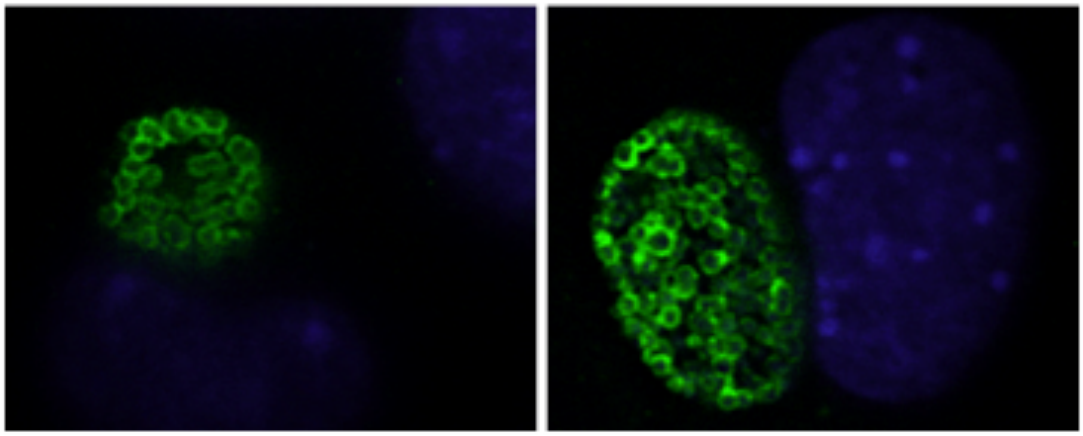
In order to examine the localization properties of CT584 in the context of RBs, infected cells were examined at 24 hours post infection by confocal microscopy (Fig. 3.3). CT584 has a localization pattern that is consistent with primary cytoplasmic localization (Fig. 3.3A). As a control for a known membrane localized protein, the localization of MOMP was also examined, in the same inclusions. MOMP displays a characteristic localization to the RB membrane, resulting in bacteria with ring-like fluorescence (Fig 3.3B). Because the two proteins were fluorescently labeled in the same sample, a merged image showing localization of both together was constructed; this image shows clearly

**FIG 3.3. CT584 is not primarily localized to the surface of RBs**. L929 cells were infected with *C. trachomatis* and fixed at 24 hours post infection, probed with antibodies and then visualized by confocal microscopy. Cells were then probed with antibodies against CT584 (A) and against MOMP (B). A merged image showing both CT584 and MOMP fluorescence shows their contrasting localizations (C). Host cell nuclei are stained in each image in blue.

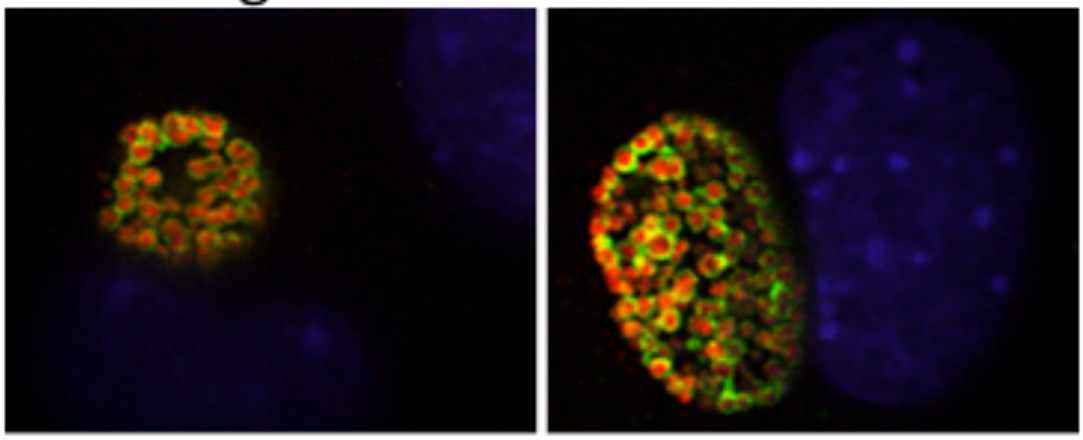
A - Anti-CT584



B - Anti-MOMP



C - Merge



how the localization of CT584 does not resemble the ring-like localization of MOMP (Fig. 3.3 C). This analysis shows that in the context of RBs, CT584 does not have a localization pattern consistent with primary membrane localization.

## Chapter III

## Discussion

Levels of CT584 in EBs were not significantly affected by treatment with the protease Proteinase K. This indicates that the majority of CT584 present in EBs is not available for digestion by the protease. This does not preclude its surface localization for two reasons: first, it remains possible that CT584 has two roles in the cell, as both a needle tip protein and serving another function in the cytoplasm of the chlamydial cell. If the number of CT584 proteins serving an extracellular role is significantly less than the number of proteins functioning in the cytoplasm, then their destruction by the protease would not significantly affect the total overall levels of CT584 protein in the EBs. An examination of this ambiguity is handled by the examination of EBs by immunofluorescence, discussed below. Another possible explanation for the lack of CT584 destruction is that CT584 may have surface localization but plays its role in RBs, not EBs. This is plausible because it is known that the T3SS plays a role in both EBs and in RBs. EBs were used in this assay because the purification of RBs while maintaining the integrity of their membranes is challenging. This possibility served as justification for the following experiments examining CT584 localization in RBs.

CT584 could not be detected on the surface of EBs by immunofluorescence. One interpretation of these data is that CT584 is not localized to the surface of chlamydia and is therefore not likely to be the T3SS needle tip protein. However, these are a few alternate explanations to these findings. The positive control for surface localization (MOMP) was easily detected in this assay. However, MOMP is, as its name implies, the

major protein component of the chlamydial outer membrane, making up ~60% of the total protein of the outer membrane (Caldwell *et al.*, 1981). It is possible that CT584 is localized to the surface of EBs but in quantities that are below the levels for detection by this assay. In order to eliminate this doubt, more sensitive, higher resolution techniques must be employed in the future to examine CT584 localization, including transmission electron micrograph (TEM) immunogold labeling, which has been used in the past to detect the needle tip protein on *Shigella flexneri* (Espina et al., 2006). As discussed above, another potential interpretation of these results is that CT584's role as a needle tip protein is not involved in attachment or entry of the EB to the cell but rather serves a purpose in the RBs in maintaining the inclusion integrity and evading the immune system. It is this possibility that provoked localization studies on RBs.

Supporting the possibility that chlamydial species employ non-redundant sets of proteins for type three secretion is the expression patterns of the translocon pore proteins. The proteins CopB and CopD have been shown to have properties consistent with the function of the T3SS transposon pore proteins. These proteins are among the first to be secreted and form a hollow pore in the host cell membrane through which effectors proteins are translocated. As discussed previously, for *Chlamydiae*, the T3SS apparatus translocates proteins across the host cell membrane during initial stages of infection but also translocates effectors across the inclusion membrane (and into the host cytoplasm) at later timepoints. Despite the presence of a functional T3SS throughout the developmental cycle, levels of CopB transcript become undetectable after 8 hours post infection (Chellas-Gery *et al.*, 2011). This suggests the presence of an alternate translocon pore that functions during the times when CopB is absent. The use of

alternate components of a T3SS is not a novel concept; *Salmonella spp.* use two distinct T3SS apparatuses that have both common and unique components (Ochman & Groisman, 1996). Therefore, it would not be surprising if CT584 functioned as a needle tip protein in the later stages of infection, translocating proteins across the inclusion membrane rather than across the host cell membrane.

In order to examine localization of CT584 in the late stages of infection, localization studies on RBs (in the context of infected host cells) were performed. These showed a primarily cytoplasmic localization for CT584. This does not rule out the possibility of CT584 playing a role as a needle tip. As discussed above, CT584 may have dual roles in the cell, functioning as a needle tip on the surface and playing a different role in the cytoplasm of the cell. If the majority of CT584 in RBs has an undetermined function in the cytoplasm, then its surface localization in these studies would not be apparent. This is a limitation of the resolution obtainable by confocal microscopy, compounded with the relatively small size of chlamydial RBs. More sensitive visualization techniques, such as TEM immunogold labeling could be employed to resolve this issue.

CT584 is well conserved in all species of Chlamydia with sequenced genomes. Recently, the structure of the CT584 ortholog from *Chlamydia pneumoniae* (Cpn0803) was solved and analyzed (Stone *et al.*, 2012). Overall, the protein structure bore little resemblance to other T3SS needle tip proteins, such as LcrV or IpaD. Additionally, the protein exists as a hexamer in solution and in the crystal structure. Analysis of this higher order structure revealed that this protein complex lacked the inner pore diameter required for proteins to pass through and therefore was not consistent with the proposed

function of this protein as a needle tip protein. Importantly, both the overall global fold of the protein and its salient features (inner pore diameter) were examined in this study and both analyses concluded the same thing: that the protein lacks features consistent with the function of a T3SS needle tip protein.

Ultimately, the evidence that CT584 is a chlamydial needle tip protein comes exclusively from indirect assays; biophysical characterization of the protein's unfolding properties and on its interaction with the needle tip protein in a heterologous system (Spaeth et al., 2009, Markham et al., 2009). While impossible to fully rule out its possible function as a needle tip protein, experimental localization patterns of the protein do not support this role. While the evidence for CT584 as a needle tip protein weakens with these findings, another protein with a better-characterized role as the T3SS needle tip has yet to be identified. Therefore, CT584 can still be looked on as a putative needle tip protein, but one that does not exhibit primary surface localization on either EBs or RBs.

## Chapter IV. Structural and functional analyses support that *Chlamydia trachomatis* protein CT009 is a homolog to the key morphogenesis component RodZ


### Abstract

Cell division in *Chlamydiae* is poorly understood as apparent homologs to most conserved bacterial cell division proteins are lacking and, as recent evidence suggests, non-canonical mechanisms may be employed for these processes. Specifically, the rod-cell shape determining protein MreB may be playing a role in chlamydial cell division. In other organisms, MreB is part of a morphogenic complex that requires RodZ for proper morphogenesis. A RodZ homolog was not evident in the chlamydial genomes; however, computational structure modeling (I-TASSER) indicated that uncharacterized ORF CT009 shares structural similarity to RodZ. The X-ray crystal structure of CT009 was solved and validated the accuracy of the I-TASSER predicted structure as well as similarity to RodZ. CT009 and MreB were demonstrated to interact and require two conserved residues in CT009. CT009 expressed in a *rodZ* deficient *E. coli* strain restored wild-type rod shaped morphology, also dependent on these key residues. These observations demonstrate that CT009 is a RodZ homolog in *Chlamydia* and expected to be involved in cell division and septum formation processes in chlamydial species. Furthermore, CT009 sequence similarity analysis supports that many other uncharacterized proteins are likely RodZ homologs.

## Introduction

In the vast majority of bacterial species, cell division is mediated primarily by a structure called the Z ring, which directs the location of the division septum (Erickson et al., 2010). The Z ring is made up of a polymer of the protein FtsZ and serves as a scaffold for the recruitment of all other proteins involved in septation and cytokinesis (Adams & Errington, 2009). However, no homolog to FtsZ is apparent in any chlamydial genome. In addition, many other well-conserved proteins that are involved in Z-ring assembly, such as FtsA and ZipA, are also not apparent in chlamydial genomes. Together, this suggests that either homologs to these proteins are present in chlamydial species but are not apparent by sequence similarity or that chlamydial cell division is mediated by a novel mechanism. Along these lines, it was recently proposed that (in *Chlamydiae*) the protein MreB functions as a central coordinator of cell division in replacement of FtsZ (Ouellette et al., 2012). MreB is a bacterial protein that is structurally similar to actin but, unlike actin, is predominately associated with the regulation of non-cocci cell shape. *Chlamydiae* are an exception to this trend as they have coccoid morphology and yet encode MreB (Stephens et al., 1998). In fact, *Chlamydiae* encode several rod-shape determining proteins, including MrdA/Pbp2 (CT682), which is involved in sidewall peptidoglycan synthesis (peptidoglycan synthesis in the lateral direction) and MrdB/RodA (CT726), a predicted flippase (aiding in the movement of membrane-localized proteins between leaflets of the membrane).

If MreB is utilized for cell division in *Chlamydiae*, then it is likely that other proteins associated with MreB may also play a role in this process. In other bacteria,

MreB does not function alone; rather, it is part of a large morphogenic complex and has been shown to interact with several other proteins including MreC, MreD, RodA, and RodZ (Kleinschnitz *et al.*, 2011, Kruse *et al.*, 2005, White et al., 2010). While RodA is predicted to be a flippase, the precise functions of MreC, MreD and RodZ are not well understood and they are thought to play scaffolding roles in the morphogenic complex. RodZ is of particular importance because its interactions with MreB are better defined, due to a co-crystal structure with these two proteins (PDBid: 2WUS; (van den Ent *et al.*, 2010)). Importantly, while there is an annotated chlamydial RodA homolog (CT726), the other MreB interaction partners do not appear to be present in chlamydial species. A better understanding of the presence or absence of these MreB interaction partners in chlamydial species could give insight into the unique mechanisms for cell division employed by these species.

Open Reading Frame (ORF) CT009 is conserved in all species of *Chlamydiae* with sequenced genomes. Interestingly, this protein contains a predicted helix-turn-helix motif within a conserved XRE (xenobiotic response element) domain superfamily of DNA binding transcription factors. As such, our initial studies with this protein tested the hypothesis that CT009 functioned as a DNA binding protein and had a role in chlamydial gene regulation. However, as described in this report, computational and experimental structural analyses indicated that CT009 has highest similarity to RodZ, the MreB-interacting protein that plays a key role in morphogenesis. Functional complementation and bacterial two-hybrid analyses support the structural observations providing evidence for CT009's function as a homolog to RodZ.

## Chapter IV

## Materials and Methods

*Identification of CT009 homologs and their conservation*- CT009 protein sequence was used as input for a BLAST search (Altschul *et al.*, 1997). Orthologous proteins from other chlamydial species were identified. Sequences were aligned using ClustalW2 (Thompson *et al.*, 2002).

*Protein expression profile of CT009*- T150 flasks confluent with L929 mouse fibroblast cells were infected with *Chlamydia trachomatis* serovar L2 434/Bu at an MOI of 1. As a control, a T150 flask was mock infected. At 12 or 30 hours post infection (mock infected at 30 hours post mock-infection), cells were trypsinized and harvested by centrifugation. Protein was isolated from other cellular material using the Qproteome Mammalian Protein Prep Kit (QIAGEN), as per manufacturer's instructions. Protein was further processed by ethanol precipitation. Briefly, 9 volumes of cold 100% ethanol was added to samples and incubated at -20 °C overnight. Precipitated protein was harvested by centrifugation at 15,000 xg got 15 minutes, pellets were air dried and resuspended in a volume of 90% cold ethanol equal to the starting volume. Samples were centrifuged and dried again, as previously described. Finally, each pellet was resuspended in 100 µL of 1x SDS PAGE final sample buffer. Samples were boiled for 10 minutes at 94 °C, loaded and run on an SDS PAGE gel (15% acrylamide). Proteins were transferred to a nitrocellulose membrane by western blot. The membrane was stained by Ponceau S and cut horizontally at the 25 kDa molecular weight marker. The membrane was blocked for

45 minutes in a 5% w/v non-fat dry milk solution in PBS + 0.1% Tween-20 and probed with either a polyclonal affinity-purified anti-CT009 antibody (lower half of membrane; 1/500) or a polyclonal affinity-purified anti-RpoA antibody (upper half of membrane; 1/1000) overnight at 4 °C. Membranes were washed in PBS-Tween and probed with anti-rabbit IR700 secondary antibodies (LI-COR Biosciences) at 1/5000 for 1 hour at room temperature. Membranes were washed in PBS-Tween and visualized using an Odyssey infrared imaging system (LI-COR Biosciences).

*Systematic Evolution of Ligands by Exponential Enrichment (SELEX)*- Single-stranded template DNA was obtained (Integrated DNA Technologies, Coralville IA) with the following sequence: 5' CGGGATCCAAGCTTCTGCAGNNNNNNNNNNNNNNNNNNNNNNNNNNNN GGTACCGTCGACGAATTCCG 3'. This template was then amplified using the following primers:

5' primer: 5' CGGGATCCAAGCTTCTGCAG 3'

3' primer: 5' CCATGGCAGCTGCTTAAGGC 3'

Buffer and ions were removed from the reaction using QIAquick nucleotide removal kit (QIAGEN). Protein (ChxR, CT009_FL or no protein) was allowed to bind to SELEX template DNA by incubation for 1 hour at room temperature while on a circular rotator in a 500 μL reaction in DNA binding buffer (40 mM Tris, 0.1 mM EDTA, 100 mM NaCl, 250 mM KCl, 1 mM DTT, 10% glycerol, pH 7.5) containing 170 pmol of DNA and 1000 pmol of protein. To crosslink protein to DNA, formaldehyde was added to a final concentration of 1.7% to each tube and incubated for 20 minutes at room temperature.

To neutralize unreacted formaldehyde, glycine was added to a final concentration of 143 mM. 10 μg of antibodies (to either CT009 or ChxR) or DNA binding buffer was added to each tube. Tubes were incubated at 4 °C overnight. To bind the DNA-protein-antibody complex to beads, to each reaction added 25 μL of resuspended Dynabeads (Invitrogen) that had been washed 2x in RIPA buffer (10 mM Tris, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% sodium dodecyl sulfate, 140 mM NaCl, 5 mM DTT, pH 8.0). Tubes were rotated at 4 °C overnight. Using a MagnaRack (Invitrogen), unbound solution was removed and beads were washed 4x with RIPA buffer. To elute bound DNA, 30 μL of TE buffer (100 mM Tris, 10 mM EDTA, pH 7.4) was added to beads and incubated at 94 °C for 5 minutes. Eluted DNA was removed by MagnaRack. PCR was performed using eluted DNA as template for each reaction and the primers listed above.

*Electromobility Shift Assays (EMSA) using CT009 and select chlamydial promoter regions- Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA was used as a template to amplify ~400-bp fragments of DNA by PCR. These fragments consist of a region spanning ~100 bp upstream of the transcription start site and ~300 bp downstream of the transcriptional start site. These putative promoter regions of DNA were amplified for the genes listed in Table 4.1 using primers listed in Table 4.1 (Integrated DNA Technologies). To serve as a negative control in the EMSA reaction, a ~400 bases region of DNA between the two converging genes CT032 and CT033 was also amplified. EMSA reactions were carried out in a total volume of 20 μL with poly DIC (1 ug), respective DNA regions (400 ng) and purified recombinant $CT009^{FL}$ at a DNA/Protein

molar ratio of either 1:10 or 1:100 in DNA binding buffer (40 mM Tris, 0.1 mM EDTA, 100 mM NaCl, 250 mM KCl, 1 mM DTT, 10% glycerol, pH 7.5). In addition, an EMSA reaction with promoter region CT241 was also carried out with no protein. Once protein was added, the solution was gently mixed and incubated for 20 min at room temperature. Five microliters of 5× DNA dye (90 mM Tris, 90 mM boric acid, 2 mM EDTA, 20% glycerol, 1% xylene cyanol, 1% bromphenol blue) was added to the reaction mixtures and mixed gently. Twenty microliters of each reaction mixture was loaded on a precast 6% Tris-borate-EDTA (TBE) gel (Invitrogen) and run at 150 V for 75 min. Gels were stained in a SYTO 60 DNA stain solution (Invitrogen) as per manufacturer's protocol for 30 minutes and were visualized using an Odyssey infrared imaging system (Li-Cor Biosciences).

Table 4.1. Primers used in CT009 EMSAs

| Name | Sequence |
|---|---|
| CT043 promoter 5' | 5′-CTCTTTAGCAAAATTTTTTAGATTTTC-3′ |
| CT043 promoter 3' | 5′-CGAAACTAAAGCGTGTTTTAGAAG-3′ |
| CT230 promoter 5' | 5′-CCTGAATTAATATTAGAACAGCTC-3′ |
| CT230 promoter 3' | 5′-GCCAAGCATCATATTGCTTTTTTC-3′ |
| CT239 promoter 5' | 5′-GTCAAAATTTTCTTAGGCAAATACG-3′ |
| CT239 promoter 3' | 5′-GGATCCCATTCTAATAATTCAAAAG-3′ |
| CT241 promoter 5' | 5′-GCTTTTCGGGCTTTTTTTCGTTG-3′ |
| CT241 promoter 3' | 5′-GCAACAGTAAAACAGCGAGTTG-3′ |
| CT259 promoter 5' | 5′-GCCTAACTAAACCAACATCACTC-3′ |
| CT259 promoter 3' | 5′-GGCAGAGTAAAATACTTCTGTTTG-3′ |
| CT344 promoter 5' | 5′-CAAAAATTAATTCGCTGTCGCC-3′ |
| CT344 promoter 3' | 5′-CCTCTGAAGCATTTGGATCTAG-3′ |
| CT386 promoter 5' | 5′-CGTGAAAAGAACCATCATGTGTG-3′ |
| CT386 promoter 3' | 5′-GAAAAATTTTTAAAAAGAGAGTTTAAAG-3′ |
| CT432 promoter 5' | 5′-GCCAAATTTTGCTGACTTTTGTC-3′ |
| CT432 promoter 3' | 5′-CTGAAGCTGCAATGGTATTAAAG-3′ |
| CT635 promoter 5' | 5′-GTTTTATAGAATCTATAATTTTTTGAG-3′ |
| CT635 promoter 3' | 5′-CTTGATTTTTTCCAGTAAGACAAC-3′ |
| CT708 promoter 5' | 5′-CGCGAGAAGAAGAGCATTCTAG-3′ |

| | |
|---|---|
| CT708 promoter 3’ | 5’-GCAGTATGTTAGCTGTAAAATCTC-3’ |
| CT827 promoter 5’ | 5’-GCTCTCGAAAAGAAATTTTTCATC-3’ |
| CT827 promoter 3’ | 5’-CTTAACAATTGTGCATTGCTTTTC-3’ |
| CT874 promoter 5’ | 5’-GGATGACAAAATAGCAGCACATAG-3’ |
| CT874 promoter 3’ | 5’-CAAAATTAAAAAAATCGATAATAAAATAAG-3’ |
| CT032/CT033 5’ | 5’-ACATTCCTTAGATCTAGGTTCCC-3’ |
| CT032/CT033 3’ | 5’-GCTATTGCTGTTCGTAATAATAAGG-3’ |

*Fluorophore-Linked Immunosorbent Assay (FLISA)*- A region of the ChxR promoter (~300 bp, from position -10 to -334) was amplified from *Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA using the following primers: forward 5' GCAGAAACATCAGCTGCATC 3' and reverse, 5' CCCATTGAACTATTAGATTACC 3'. A biotinylated forward primer was used (Integrated DNA Technologies, Coralville IA) to ensure that the DNA fragment had biotin attached to it. A similarly sized molecule of DNA (with biotin attached) was amplified from the promoter region between the diverging genes CT870 and CT871 using the following primers: forward 5' CGCAGGTTAAAAGGGGGATG 3' and reverse 5' GGCAAAGCTGCAATAGAATTG 3'. A similarly sized molecule of DNA (with biotin attached) was amplified from the non-promoter genome region between the converging genes CT870 and CT871 using the following primers: forward 5' ACATTCCTTAGATCTAGGTTCCC 3' and reverse 5' GCTATTGCTGTTCGTAATAATAAGG 3'.

2 μg of these DNA molecules (~10 pmol) in DNA binding buffer (5 mM Tris, 0.5 mM EDTA, 50 mM KCl, pH 7.8) was loaded onto a well of a clear NeutrAvadin coated polystyrene plate (Thermo Scientific) and was allowed to bind for 2 hours at room temperature. Wells were washed three times with Wash Buffer (140 mM NaCl, 3 mM KCl, 4 mM $Na_2PO_4$, 0.1% Tween-20) and either full-length ChxR protein (6.3 ug, ~100 pmol in DNA binding buffer) or DNA binding buffer was loaded into wells and

incubated for 30 minutes at room temperature. Wells were washed three times with Wash Buffer. 3 μg of anti-ChxR antibody in DNA binding buffer (or DNA binding buffer alone for controls) was applied to each well and incubated for 1 hour at room temperature. Wells were washed three times with Wash Buffer. Finally, anti-rabbit IR700 fluorescently labeled secondary antibodies (Rockland Immunochemicals, Gilbertsville PA) were added to each well at a dilution of 1/1000 in DNA binding buffer and were incubated for 1.5 hours at room temperature. Wells were washed three times with Wash Buffer and visualized using an Odyssey infrared imaging system (LI-COR Biosciences).

*Bacterial one hybrid assays*- Single-stranded template DNA was obtained (Integrated DNA Technologies, Coralville IA) with the following sequence: 5' AAAGTGTAAAGAATTCGNNNNNNNNNNNNNNNNNNNNNNNNAATTCGGCG CGCCTC 3'. Library template was extended using the following primer: 5' GAGGCGCGCCGAATT 3'. Library vector (pH3U3) was digested with EcoRI-HF (New England Biosciences) and isolated by gel extraction (QIAGEN). Library insert was treated with Cloning Enhancer (Clontech) as per manufacturer's protocol. Library vector and insert were subjected to the In-Fusion reaction (Clontech) as per the manufacturer's protocol. 5 μL of this reaction and 18 ng of pH3U3 plasmid were transformed into DH5α cells and plated onto 2xYT plates with 30 ug/mL kanamycin. Colonies were counted manually and transformation efficiency was calculated.

*I-TASSER Protein Structure Modeling*- The I-TASSER server has been described previously (Roy et al., 2010, Zhang, 2009) as has its application in the modeling of chlamydial proteins (Kemege *et al.*, 2011). The following amino acid sequence from the genome of *Chlamydia trachomatis* LGV (L2/484/Bu) was selected as the target: GHMSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKLI SPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMG GRNSPEKAIRS. This proteins sequence represents amino acids 1-116 of CT009, the predicted cytoplasmic domain. In addition, there are two extra amino acids added to the N-terminus of the sequence. In our purification strategy for recombinant protein, these amino acids are the remnant of an N-terminal solubility tag. In order to make a fair comparison between the modeled structure and the experimentally determined structure, the exact sequence of the recombinant protein used for crystallography was used for computational modeling. In the nomenclature of the L2 genome, this protein is named CTLon_0264. However, in this report the more widely utilized serovar D genome is used and the protein is referred to as CT009.

*Cloning, Protein Expression and Purification of CT009 1-116*- The putative cytoplasmic domain of CT009 (residues 1-116) was amplified by PCR using *Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA and the following primers: 5'-GGGAATTCCATATGAGCGAACATGTCCACAAAG-3' (forward, with an NdeI restriction site) and 5'-CGCGGATCCCTACTAACTACGGATCGCTTTCTCAGG-3' (reverse, with a BamHI restriction site). Both the *CT009* insert and the vector (pDZ1) were digested with NdeI/BamHI and then ligated together. The vector pDZ1 encodes a

$His_6$-tagged GB1 protein (the B1 immunoglobulin domain of Protein G from *Streptococcus sp.*, used as a solubility tag) and a Tobacco Etch Virus (TEV) protease cleavage site upstream of the cloning site (resulting in an N-terminally tagged protein) (Chatterjee *et al.*, 2011). These alterations greatly increased the solubility of the protein. However, following cleavage of the GB1 protein from $CT009_{1-116}$ by TEV protease, the two proteins did not separate from one another, even under high salt conditions. GB1 is a mildly acidic protein with a pI of 4.6. It has been previously reported that three amino acids changes can be made to the GB1 protein (resulting in a protein termed "$GB1^{basic}$") that increase its pI to 8.0 but do not alter its efficacy as a solubility tag (Zhou & Wagner, 2010). This change in pI should prevent aggregation of the GB1 with $CT009_{1-116}$ if their interaction is mediated by charge (Zhou & Wagner, 2010).

In order to obtain $GB1^{basic}$-$CT009_{1-116}$ recombinant protein, the $CT009_{1-116}$ insert was cloned into the pDZ1b vector (generous gift from Dr. Susan Egan, University of Kansas), as described above. It was constructed by modifying the pDZ1 vector (Chatterjee et al., 2011) through six base pair substitutions, resulting in three amino acid changes. The ligation reaction was then transformed into *E. coli* BL21-derivative Acella cells (Edge BioSystems). Cells were grown at 37°C with shaking at 220 rpm in LB medium supplemented with 100 μg $mL^{-1}$ ampicillin to an optical density ($OD_{600}$) of 0.6. Expression of recombinant protein was induced by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) followed by shaking incubation at 15°C and 220 rpm for ~16 hours. Cells were harvested by centrifugation (10,000 x g, 20 min, 4°C) and frozen at -20°C.

Cell pellets were resuspended in lysis/wash buffer (20 mM Tris pH 8.0, 500 mM NaCl), lysed by sonication and centrifuged at 16,000 x g, 20 min, 4°C. The soluble fraction was filtered through a 0.45 μm filter before it was applied to TALON metal ($Co^{2+}$) affinity resin (Clontech) as per the manufacturer's guidelines. The beads were washed with lysis/wash buffer and $GB1^{basic}$-$CT009_{1-116}$ was eluted with lysis/wash buffer supplemented with 500 mM imidazole. Peak fractions were pooled and purity was assessed by SDS-PAGE, followed by staining with IRDye Blue protein stain (LI-COR) and visualization by an Odyssey infrared imaging system (LI-COR).

Purified $GB1^{basic}$-$CT009_{1-116}$ was treated with purified recombinant TEV protease as it was dialyzed into 50 mM Tris pH 8.0 at room temperature for ~16 hours. Protein digest was filtered through a 0.22 μm filter and NaCl was added to a final concentration of 500 mM. Protein solution was applied to fresh TALON metal ($Co^{2+}$) affinity resin. Flowthrough and washes (with lysis/wash buffer) were collected and assessed for purity by immunoblot against polyhistidine and CT009. Fractions with $CT009_{1-116}$ but without the His6-tagged $GB1^{basic}$ were pooled and concentrated to 30.1 mg/mL, using an Amicon 3,000 molecular weight cut-off centrifugal filtration device (Millipore). Purity of $CT009_{1-116}$ was assessed by SDS-PAGE, followed by stain and visualization, as described above.

*Crystallization and Data Collection*- A purified sample of CT009 was concentrated to 43.5 mg $mL^{-1}$ in 20 mM Tris, 500 mM NaCl, pH 8.0 buffer for crystallization screening. All crystallization experiments were conducted using Compact Jr. (Emerald Biosystems) sitting drop vapor diffusion plates at 20 $^{o}$C using equal volumes of protein and

crystallization solution equilibrated against 75 μL of the latter. For X-ray data collection, samples were transferred to a fresh drop composed of 80% crystallization solution and 20% PEG 400 before flash cooling in liquid nitrogen. Initial crystals, which formed as clusters of needles in 1-2 days, were obtained from the Crystal Screen HT screen (Hampton Research) condition A6 (30% PEG 4000, 100 mM Tris pH 8.5, 200 mM $MgCl_2$). Initial X-ray diffraction data were collected in-house at 93K using a Rigaku RU-H3R rotating anode generator (Cu-Ka) equipped with Osmic Blue focusing mirrors and an R-Axis $IV^{++}$ image plate detector. Subsequent crystals, which were used for high-resolution data collection, were obtained from a custom PEG Screen, with the final buffer composition being 100 mM Tris, 200 mM magnesium chloride, 30% w/v polyethylene glycol 5,000, pH 8.5. Data were collected at the Advanced Photon Source beamline 17-ID using a Dectris Pilatus 6M pixel array detector.

*Structure Solution and Refinement*- Intensities for the in-house diffraction data were integrated using XDS (Kabsch, 1988). Synchrotron diffraction data were collected at frame width of 0.2° for a total of 120° (600 total frames). For each frame, the exposure time was 0.2 seconds and the crystal-to-detector distance was 175 mm. Synchrotron data were integrated using XDS via the XDSAPP interface (Krug *et al.*, 2012). The Laue class check and data scaling were performed with Aimless (Evans, 2011). The highest probability Laue class was 4/*mmm* and space groups $P4_12_12$ or $P4_32_12$. The Matthew's coefficient (Matthews, 1968) (Vm=2.5 / 50.8 %) indicated that there was a single CT009 molecule in the asymmetric unit. Structure solution was conducted by molecular replacement with Phaser (McCoy *et al.*, 2007b) via the Phenix (Adams et al., 2010)

interface using the in-house diffraction data scaled to 2.6Å resolution. All space groups in the Laue class 4/*mmm* were tested. The search model for molecular replacement was generated using the protein structure prediction program MODELLER (Eswar *et al.*, 2006). The top solution was obtained in the space group $P4_12_12$, which was used from this point forward. The model was improved using the Autobuild suite via the Phenix interface which converged at $R$=25%, $R_{free}$=33% following refinement. This model was used for subsequent molecular replacement searches against synchrotron data that was scaled to a resolution of 1.25Å. Structure refinement using anisotropic atomic displacement parameters and manual model building were conducted with Phenix and Coot (Emsley *et al.*, 2010) respectively. Structure validation was conducted with Molprobity (Chen *et al.*, 2010). Figures were prepared using the CCP4MG package (Potterton et al., 2004).

The final model of CT009 contains one monomer in the asymmetric unit. Residues Leu 84 to Met 97 in the C-terminal helix were disordered; however, recurring positive difference (Fo-Fc) electron density in the region, following refinement, indicated that this portion of the helix adopted two distinct conformations and was model as such. Therefore, all residues were modeled. There is also density for 111 water molecules.

*Protein-Protein Interactions by Bacterial Two-Hybrid-* The full-length *CT009* gene, the predicted cytoplasmic domain of *CT009*, and the *mreB* gene (CT709) were amplified by PCR using *Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA:

Table 4.2. Primers used in cloning for BACTH

| Name | Sequence |
|---|---|
| CT009 FL into pUT18C 5' | 5'-CGACTCTAGAGGATCCCATGAGCGAACATGTCCAC-3' |
| CT009 FL into pUT18C 3' | 5'-ATGAATTCGAGCTCGGTACCTCACTAGAAAAGGTTGAATAGATTC-3' |
| CT009 FL into pKT25 5' | 5'-CGACTCTAGAGGATCCCATGAGCGAACATGTCCAC-3' |
| CT009 FL into pKT25 3' | 5'-TTCTTAGTTACTTAGGTACCTCACTAGAAAAGGTTGAATAGATTC-3' |
| CT009 cyt into pKT25 5' | 5'-CGACTCTAGAGGATCCCATGAGCGAACATGTCCAC-3' |
| CT009 cyt into pKT25 3' | 5'-TTCTTAGTTACTTAGGTACCTCACTAACTACGGATCGCTTTCTC-3' |
| $MreB^{CT}$ into pKT25 5' | 5'-CGACTCTAGAGGATCCCATGAGCCCATACCGCAG-3' |
| $MreB^{CT}$ into pKT25 3' | 5'-TTCTTAGTTACTTAGGTACCTTATCATACTAAACTCTCTTTTCG-3' |

The gene encoding the CT009 Y57A, F61A mutant was amplified by PCR using pMLB1113-CT009 Y57A, F61A plasmid as a template. Vector plasmids (pKT25, pKNT25, pUT18 and pUT18C; Euromedex) were digested with BamHI and KpnI enzymes. Vector and inserts were processed using the In-Fusion HD ligation-independent cloning system (Clontech) and were transformed into Alpha-Select Gold Efficiency cells (Bioline). Plasmids were isolated from overnight culture and then co-transformed into chemically competent BTH101 cells (Euromedex).

Interactions between proteins were assessed using the Bacterial Adenylate Cyclase Two Hybrid (BACTH) system, as described previously (Karimova *et al.*, 1998); Euromedex). Briefly, BTH101 cells (a *Δcya* strain of *E. coli*) with two plasmids, each encoding a different fragment of the adenylate cyclase gene fused to a chlamydial protein of interest, and plated onto LB plates containing 100 μg $mL^{-1}$ ampicillin, 50 μg $mL^{-1}$ kanamycin, 0.5 mM IPTG and 50 μg $mL^{-1}$ X-gal. Plates were incubated at 30 °C for up

to 48 hours and single colonies were picked and grown in liquid culture at 30 °C. Expression of the β-galactosidase reporter gene by reconstituted adenylate cyclase was assessed using a β-galactosidase assay, as described previously (Miller, 1972). Assays were performed in triplicate.

*Analytical Ultracentrifugation*- Purified recombinant $CT009_{1-116}$ was analyzed using a Beckman Coulter XL-I analytical ultracentrifuge in sedimentation velocity mode. In the experiments, the protein concentration was 2.0 mg $mL^{-1}$. Protein was loaded into a two-channel, 12 mm optical path length cell. A four hole An60 Ti rotor was used for housing the cell, which was run at 10 °C for 450 continuous scans, measured in duplicate. The speed of rotor was 45000 rpm. Sedimentation velocity analysis was conducted using SEDFIT, version 12.1b (Schuck, 2000). Continuous c (s) distribution analysis was performed using non-linear regression during the data fit. In addition, viscosities and densities of the buffers were measured using the SVM 3000 viscometer (Anton Parr USA Inc.), in order to get accurate parameters, which were used in the AUC fitting process.

*RodZ Functional Complementation*- Full length *CT009* was amplified by PCR using *Chlamydia trachomatis* LGV (L2/484/Bu) genomic DNA and the following primers: 5'-CCGGAATTCATGAGCGAACATGTCCACAAAG-3' (forward, with an EcoRI restriction site) and 5'-TCCCCCGGGTCACTAGAAAAGGTTGAATAGATTCC-3' (reverse, with a SmaI restriction site). Both the *CT009* insert and the vector (pMLB1113) were digested with EcoRI/SmaI and then ligated together. The ligation reaction was transformed into *E. coli* BL21-derivative Acella cells (Edge BioSystems), successful

cloning was verified by sequencing of the CT009 gene and finally purified pNLB1113-CT009 plasmid was transformed into FB60 cells. FB60 is an *E. coli* cell line, derived from the TB28 cell line, which lacks the *RodZ* gene (Bendezu *et al.*, 2009). Both the pMLB1113 plasmid and the FB60 cell line were generous gifts from the laboratory of Piet de Boer (Case Western University).

A plasmid encoding full length CT009 Y57A, F61A was generated by subjecting the pMLB1113-CT009 plasmid to site directed mutagenesis using the QuikChange II XL site-directed mutagenesis kit according to the manufacturer's protocol (Agilent Technologies).

For functional complementation assays, cells were grown overnight in flasks of FB60 media (M9 minimal media with 0.2% maltose, 0.2% casamino acids and 50 mM thiamine HCl) at 30 °C to an OD600 of 0.2-0.3. Cells were prepared for fluorescence microscopy by resuspending the cells from 3 mL of culture (harvested by centrifugation for 1 minute at 16,000 x g in a benchtop centrifuge) in 1 mL of Fixative Solution (0.05 mg $mL^{-1}$ malachite green and 2.5% v/v glutaraldehyde. Cells were fixed for 1 hour at room temperature and then were washed twice with 1 mL of 1x PBS before resuspension in 100 µL of 1x PBS. Fixed cells were mounted on glass coverslips and visualized using an Axioplan-2 upright fluorescent microscope (Zeiss). Images were captured with an ORCA-ER digital camera (Hamamatsu).

# Chapter IV

# Results

***CT009 is conserved across chlamydial species.***

Following sequencing of the *Chlamydia trachomatis* serovar D genome, CT009 was annotated as a protein of unknown function, due to lack of homology to any proteins of known function. Whether the role of CT009 is unique to the biology of *C. trachomatis* or can be found in other species of *chlamydiae* could be determined by searching for orthologs to CT009 among other species of chlamydia. A BLAST search was conduced, using CT009 protein sequence as input (Fig. 4.1). The results of this search show that CT009 is conserved across all species of chlamydia with sequenced genomes, suggesting that it plays a conserved role in the biology of *chlamydiae*.

***The protein expression profile of CT009 could not be detected by western blot.***

Transcriptional profiling of the expression of CT009 by microarray showed a typical constitutive expression profile in which levels of transcription were first detected at 3 hours post infection (HPI), increased at 8 HPI and then remained steady throughout the developmental cycle until the final point of 40 HPI (Belland *et al.*, 2003). However, this expression profile is a transcriptional profile. Mechanisms for post-transcriptional regulation may affect the ultimate levels of CT009 throughout the developmental cycle. Therefore, in order to gain additional insight into the function of CT009, (specifically how it plays a role in the overall biology of Chlamydia), an analysis of the protein expression profile of CT009 was attempted.

**FIG. 4.1. CT009 is conserved across all species of *chlamydiae*.** A BLAST search using CT009 protein sequence as input revealed the presence CT009 homologs in all sequenced species of *chlamydiae*. Species CT_A is *Chlamydia trachomatis* serovar A (strain HAR-13 / ATCC VR-571B) and its homolog is CTA_0010. Species CT_B is *Chlamydia trachomatis* serovar B (strain Jali20/OT) and its homolog is JALI_0091. Species CT_D is *Chlamydia trachomatis* serovar (strain D/UW-3/Cx) and it is the starting sequence, CT009. Species CT_E is *Chlamydia trachomatis* serovar E (strain E/11023) and its homolog is E11023_00055. Species CT_F is *Chlamydia trachomatis* serovar F (strain SW5) and its homolog is FSW5_0091. Species CT_G is *Chlamydia trachomatis* serovar G (strain G/9301) and its homolog is CTG9301_00050. Species CT_L2 is *Chlamydia trachomatis* serovar L2 (strain 434/Bu / ATCC VR-902B) and its homolog is CTL0264. Species Cmur is *Chlamydia muridarum* (strain MoPn / Nigg) and its homolog is TC_0277. Species Cpneumo is *Chlamydia pneumoniae* (strain LPCoLN) and its homolog is CPK_ORF00494. Species Cpecorum is *Chlamydia pecorum* (strain ATCC VR-628 / E58) and its homolog is G5S_0635. Species Cpsitt is *Chlamydia psittaci* (strain ATCC VR-125 / 6BC) and its homolog is CPSIT_0325. Species Cabortus is *Chlamydia abortus* (strain S26/3) and its homolog is CAB292. Species Cfelis is *Chlamydia felis* (strain Fe/C-56) and its homolog is CF0708. Species Ccaviae is *Chlamydia caviae* (strain GPIC) and its homolog is CCA_00295. The most dissimilar protein to CT009 (the *C. caviae* homolog CCA_00295) still aligns with a E value of 3e-55.

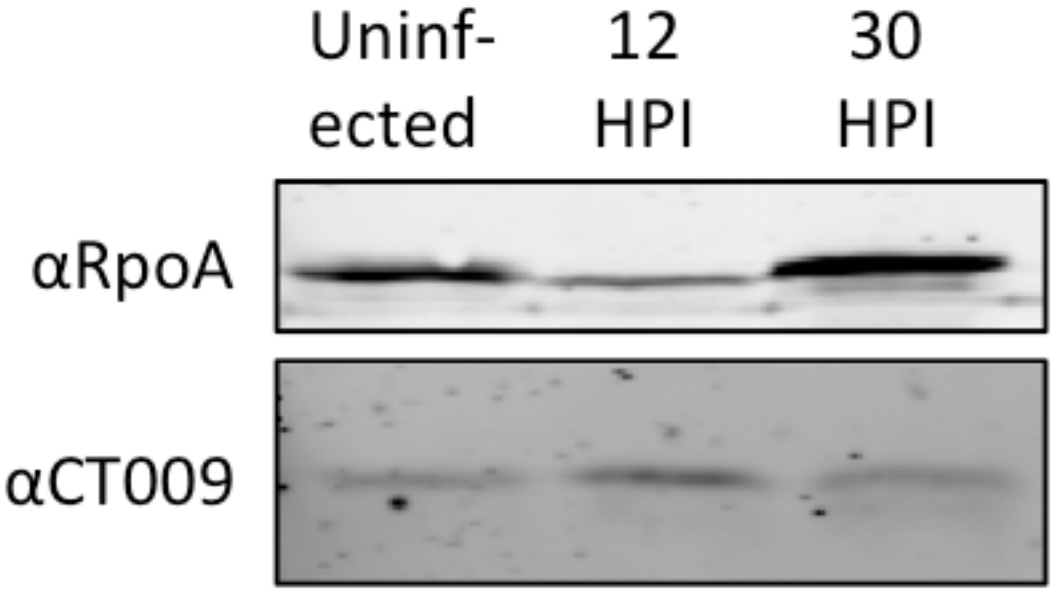
```
CT_A      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_B      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_D      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_E      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_F      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_G      --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
CT_L2     --------MSEHVHKELLHLGEVFRSQREERALSLKDVEAATSIRLSALEAIEAGHLGKL 52
Cmur      MGANKRNYMSEHVHKELLHLGGVFRSQREEKAISLKDVEAATSIRLSALEAIEAGHLGKL 60
Cpneumo   --------MQEHVHKELLHLGEIFRSSREFQSLSLKDVEAATSIRYSCLEAIEQGCLGKL 52
Cpecorum  --------MTEHVHKELLHLGEIFRSRREEQSLSLKDVEAATSIRCSCLEAIEQGYLGKL 52
Cpsitt    --------MAEQIHKELLHLGEVFRTKREEQSLSLKDVEAATSIRYSCLEAIENGYLGKL 52
Cabortus  --------MAEQIHKELLHLGEVFRTKREEQSLSLKDVEAATSIRYSCLEAIENGYLGKL 52
Cfelis    --------MAEQIHKELVHLGEVFRTKREEQLLSLKDVEAATSIRYSCLEAIENGCLGKL 52
Ccaviae   --------MAEQIHKELIHLGEVFRTKREEQLLSLKDVEAATSIRYSCLEAIENGCLGKL 52
                 * *::****:*** :**: ** : :************ *.***** * ****

CT_A      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_B      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_D      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_E      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_F      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_G      ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
CT_L2     ISPVYAQGFMKKYAAFLDMDGDRLLKEHPYVLKIFQEFSDQNMDMLLDLESMGGRNSPEK 112
Cmur      ISPVYAQGFMKKYADFLGLDGDKLLKEHPYVLKIFQEFSDQNVDMLLDLESMEGRNSPEK 120
Cpneumo   ISPVYAQGFIKKYATYLGLDGDSILQEHPYVMKIFKEFSDHNMEMLLDLESMGGRNSPER 112
Cpecorum  ISPVYAQGFIKKYASYLGLNGEKLLQEHPYVVKLLKEFSEQNMEMLLDLESMGGRNSPEK 112
Cpsitt    ISPIYAQGFIKKYAAYLGLDGERILQDHPYVMKIFKEFSEHNMEMLLDLESMGGRNSPEK 112
Cabortus  ISPIYAQGFIKKYAAYLGLDGERILQDHPYVMKIFKEFSEHNMEMLLDLESMGGRNSPEK 112
Cfelis    ISPIYAQGFIKKYAAYLGLDGERILQEHPYVMKIFKEFSEHNMEMLLDLESMGGRNSPEK 112
Ccaviae   ISPIYAQGFIKKYAAYLGMDGERILQEYPYVMRIFKEFSEQNMEMLLDLESMGGRNSPEK 112
          ***:*****:**** :*.::*: :*:::***::::::***::*::******** ******:

CT_A      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_B      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_D      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_E      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_F      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_G      AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
CT_L2     AIRSWLNLGWAGVFVVGAACIWWLGNLFNLF 143
Cmur      AIRSWLNLGWAGAFVLGVACIWWLGTLLHFF 151
Cpneumo   AIHSWSNLWWAGLIIIGGIVVWWLGSLFSIF 143
Cpecorum  AIHSWSYLGIAGAIILGGGLLWWLGSLFSIF 143
Cpsitt    AIRSLSNLWWAALIVFSGIAIWWLGSLLSLF 143
Cabortus  AIRSLSNLWWAGLIVFSGMAIWWLGSLLSLF 143
Cfelis    AIRTWSNLWWVGLITISGMAIWWLGSLLSIF 143
Ccaviae   AIRTWSNLWWVALITISGMAIWWLGSLLSIF 143
          **::   *  .. : ..   :****.*: :*
```

Host cells infected with *Chlamydia trachomatis* were lysed and protein was isolated from other cellular material. This simple experiment was repeated many times with many slight variations in technique, including the use of ethanol, acetone or trichloroacetic acid to precipitate (and therefore concentrate) proteins. Additionally, this experiment was repeated at many different timepoints after the initial infection. However, the most persistent issue has been the detection of a cross-reactive protein of similar molecular weight in uninfected cells (Fig. 4.2). In samples in which this contaminant was absent, CT009 was not present in detectable quantities.

***A computationally generated structure of the CT009 protein is most structurally similar to the transcription factor cI***

As previously described, the primary sequence of CT009 indicates that it possesses a helix-turn-helix motif in its structure, a motif that is frequently utilized to bind to DNA. However, this function is not very specific because proteins with many different roles (transcriptional repressors, transcriptional activators, nucleases and histone-like proteins, to name a few) have the capability to bind to DNA. In order to better understand the biological role of CT009, a model of the structure of the protein was generated using I-TASSER (Figure 4.3A). I-TASSER (Iterative Threading ASSEmbly Refinement) is a computational method that has been successful in accurately modeling protein structures, as detailed in Chapter II (Kemege et al., 2011, Roy et al., 2010). Several models were generated but the model with the highest C score was used to search for proteins with structural similarity, using the DALI server (Holm & Rosenstrom, 2010). The protein most structurally similar to the modeled structure of CT009 was a protein

**FIG. 4.2. The protein expression profile of CT009 could not be conclusively determined by western blot.** Cells infected with *Chlamydia trachomatis* were harvested, lysed and protein was isolated from other cellular material at two different timepoints post infection (Hours Post Infection; HPI), 12 and 30. As a control, an equal amount of uninfected cells was prepared with the same methods. Protein isolates from these cells were loaded on an SDS PAGE gel, run and then transferred to a nitrocellulose membrane by western blot. Membranes were cut in half width-wise, blocked and then probed with anti-CT009 antibodies or anti-RpoA (a loading control) antibodies, followed by a fluorescently labeled secondary antibody and visualization.


Uninf-
ected
12
HPI
30
HPI
αRpoA
αCT009

named cI (Fig. 4.3B). The two structures are shown superposed on one another (Fig. 4.3C). cI is a transcriptional repressor in lambda phage and this structural similarity therefore lead to the hypothesis that CT009 binds DNA as a transcriptional repressor.

***The findings of Systematic Evolution of Ligands by Exponential Enrichment (SELEX) were inconclusive in identifying CT009 DNA binding sequences.***

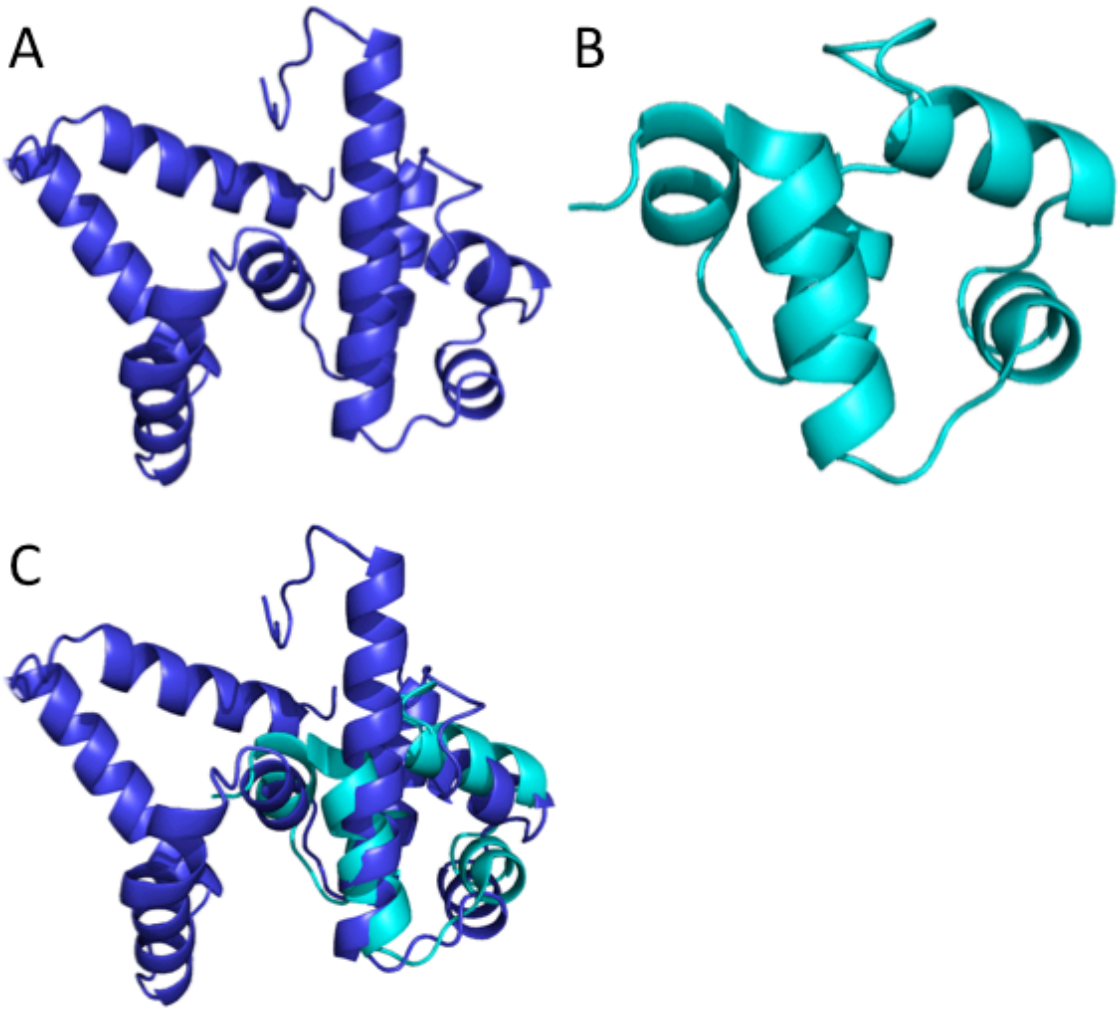
Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is a tool for the identification of sequences of DNA that a protein target can bind to (Chai *et al.*, 2011). This technique utilizes purified recombinant protein and a pool of DNA containing a randomized sequence, used at a high enough concentration to represent diverse DNA sequences. The protein of interest binds to specific sequences of DNA *in vitro*, which are then extracted from solution and amplified. This amplification is necessary due to the potentially very low concentration of the targeted DNA sequence in the overall pool of DNA. A second round then begins using this material (which the protein bound to in the first round) as the target DNA. This process is repeated several times and with each round the molecules of DNA that are recognized most strongly by the protein of interest are enriched. Following the final round of DNA binding, the sequences of DNA are cloned into plasmids and transformed into *E. coli*, which are then plated to isolate individual plasmids, each containing a single template DNA. These isolated plasmids are sequenced and this data is leveraged to identify a consensus sequence that the protein of interest used to bind to these DNA over to several rounds of selection.

**FIG 4.3. A computationally modeled structure of full length CT009 has structural similarity to the protein cI.**

A. A model of CT009 structure was created using the I-TASSER server.

B. The cI protein (PDB ID: 1R69)

C. A superposition of the modeled structure of CT009 (dark blue) and the structure of cI (light blue).

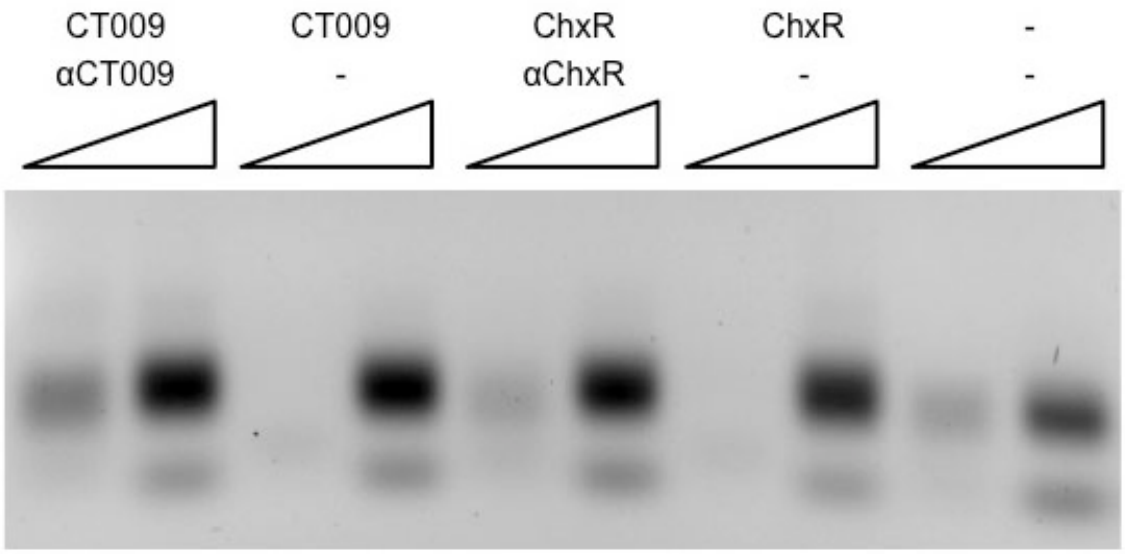

A
B
C

SELEX was employed in an attempt to ultimately identify chlamydial genes that are targets of genetic regulation by CT009. This technique was used with the goal of identifying DNA sequences that CT009 binds to *in vitro* and then utilizing this information to form testable hypotheses about genes that it may regulate. As a control in this experiment, ChxR protein (a well-characterized chlamydial transcription factor with a known DNA binding sequence; (Hickey et al., 2011, Koo et al., 2006, Thompson et al., 2002)) was used. Following incubation between protein and DNA, these molecules were removed from solution using an antibody against the protein that is bound to a large bead via protein G. As a control for nonspecific binding, each experiment was also carried out with no antibody added; any DNA pulled out from this condition would therefore be the result of nonspecific binding between the DNA and the tube/bead/protein G. DNA was eluted from the protein-protein G-bead complex, used as a template for PCR and run on an agarose gel (Fig. 4.4). An amplifiable region of DNA could not be obtained in any of the conditions tested; this suggests that either CT009 is not capable of binding to DNA or that the conditions of the assay were not suitable for DNA binding by CT009.

***CT009 does not bind to the promoter regions of select chlamydial genes in the context of an Electromobility Shift Assay (EMSA).***

In an extended effort to identify chlamydial genes regulated by CT009, potential targets for regulation were selected and the affinity of CT009 for the promoter regions of these genes was tested by EMSA (Fig. 4.5). The genes selected were CT043, CT230, CT239, CT241, CT259, CT344, CT386, CT432, CT635, CT708, CT827, and CT874.

**FIG. 4.4. SELEX was not successful in identifying sequences of DNA that CT009 binds to.** A template DNA containing 30 random base pairs flanked by 20 base pair of known sequence on each side was generated and incubated with purified recombinant CT009 or ChxR protein. As a negative control, DNA was incubated with no protein (-). Protein was crosslinked to DNA and the protein-DNA complex was separated from solution using antibodies and protein G beads. As a negative control, each reaction was performed in which one replicate did not receive antibodies (-). DNA was eluted from beads and was used as a template for PCR. Each condition was loaded into the gel in two different volumes.

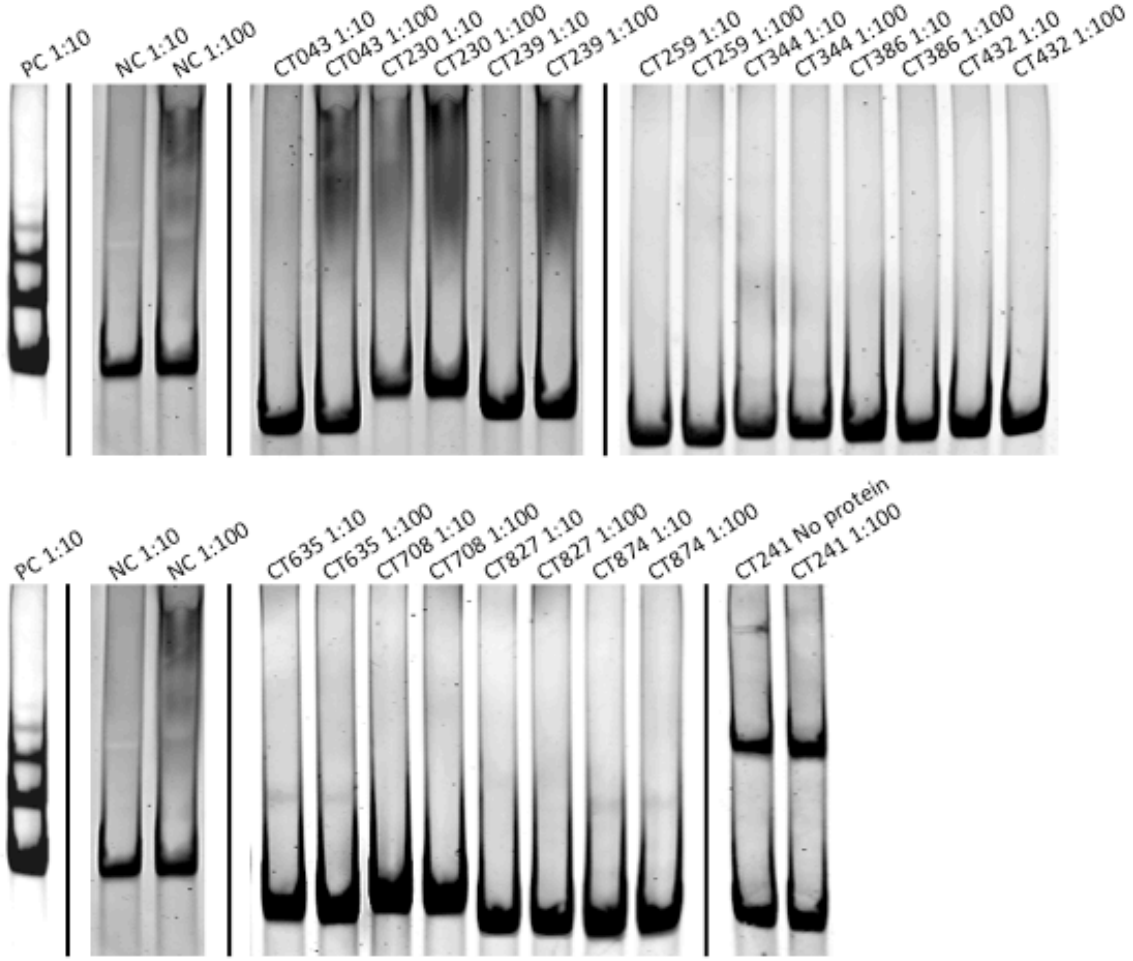

CT009
αCT009
CT009
-
ChxR
αChxR
ChxR
-
-
-

These ORFs were selected because they all have clearly identifiable promoter regions and are differentially regulated within the chlamydial developmental cycle (Belland et al., 2003). As a positive control (PC) in this experiment, the ChxR protein and its promoter region were tested by EMSA (Hickey et al., 2011). As a negative control (NC) the intergenic region of DNA between the two converging genes CT032 and CT033 was used as a DNA target for EMSA; this region of DNA does not contain the promoters for any gene and therefore makes a good negative control for promoter binding. The results of this study show no definitive shifts in the DNA in any of the reactions in which CT009 was used, despite the clear shifts in the positive control (PC). The promoter region of ORF CT241 did display a higher molecular weight species, but that species is present even with no CT009 present and therefore is likely not a shift caused by CT009 but rather an artifact of PCR amplification. Therefore, CT009 does not possess DNA binding properties, CT009 is unable to bind to DNA in the context of this assay or CT009 does not bind to the selected promoter regions. In the absence of a positive control for CT009 DNA binding, it was difficult to conclude what was the cause of these negative results.

***Fluorophore-Linked Immunosorbent Assay (FLISA) was not effective in identifying chlamydial genes regulated by CT009.***

In an effort to identify gene targets of CT009 Fluorophore-Linked Immunosorbent Assays (FLISA) were performed. In this assay, the promoter regions of potential gene targets of a DNA binding protein are amplified by PCR using a primer that has biotin covalently linked to it. These amplified promoter regions are then applied to a 96 well plate in which each well had streptavidin covalently linked to the bottom. The strong

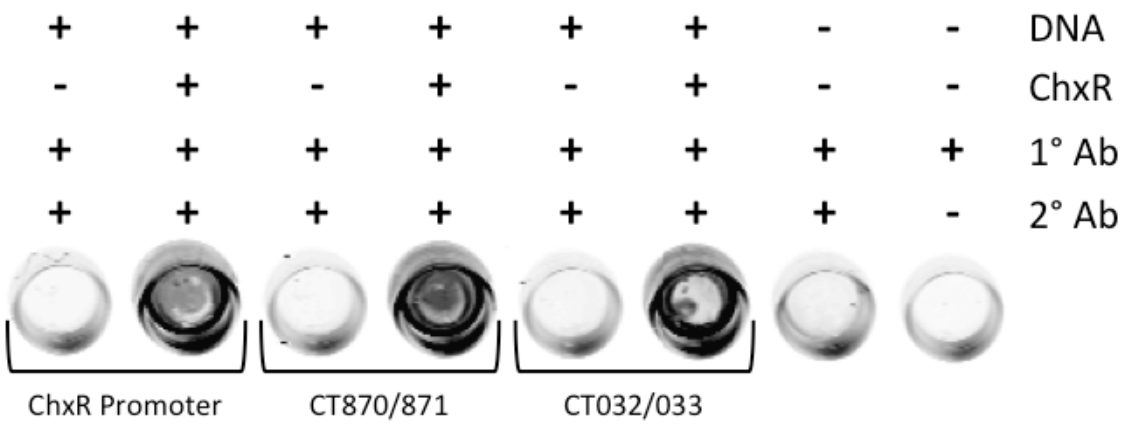
**FIG. 4.5. CT009 does not bind to select chlamydial promoters by EMSA.** Double stranded DNA, amplified from a ~400 bp region spanning ~100 bp upstream of the transcription start site and ~300 bp downstream of the transcriptional start site of several genes of interest were used in an Electromobility Shift Assay (EMSA), along with purified recombinant full-length CT009 protein. Promoter regions for CT043, CT230, CT239, CT241, CT259, CT344, CT386, CT432, CT635, CT708, CT827, and CT874 were used because these genes all have clearly identifiable promoter regions and are differentially regulated within the chlamydial developmental cycle. As a positive control (PC), the ChxR protein and its promoter region were tested. As a negative control (NC) the intergenic region of DNA between the two converging genes CT032 and CT033 was used as a DNA target for EMSA. DNA:Protein molar ratios were either 1:10 or 1:100. To show that the higher molecular weight species observed in the CT241 promoter region are not due to CT009 binding, this region of DNA was also tested using no CT009 protein, suggesting that this additional DNA band is likely not a shift caused by CT009 but rather an artifact of PCR amplification.


PC 1:10
NC 1:10
NC 1:100
CT043 1:10
CT043 1:100
CT230 1:10
CT230 1:100
CT239 1:10
CT239 1:100
CT259 1:10
CT259 1:100
CT344 1:10
CT344 1:100
CT386 1:10
CT386 1:100
CT432 1:10
CT432 1:100
PC 1:10
NC 1:10
NC 1:100
CT635 1:10
CT635 1:100
CT708 1:10
CT708 1:100
CT827 1:10
CT827 1:100
CT874 1:10
CT874 1:100
CT241 No protein
CT241 1:100

biotin-streptavidin interaction keeps the DNA in the well throughout the procedure. Purified recombinant protein is applied to each well and then washed to remove non-specific interactions. Next, an antibody against the protein is added to each well, incubated and followed by incubation with a fluorescently tagged secondary antibody. The 96 well plate is visualized; theoretically, only wells containing DNA that specifically interacts with the protein applied will have fluorescence. This technique has been utilized previously and has been successful in identifying protein-DNA interactions (Jagelska *et al.*, 2002). While, like EMSA, this technique relies on purified recombinant protein interacting with a small fragment of DNA *in vitro*, because it is performed in a 96 well plate, it allows for the screening of many potential gene targets at once.

Before applying this technique to a full array of 96 different DNA targets, validation was attempted using a previously characterized transcription factor, ChxR, and its own promoter DNA, to which it binds (Koo et al., 2006). As negative controls for this experiment, DNA from the diverging promoter regions of the CT870 and CT871 genes was used in addition to DNA from the non-promoter region between the converging genes CT032 and CT033; the converging region of DNA does not contain the promoters for any gene and therefore makes a good negative control for promoter binding while the diverging promoter region contains two promoters but these are not DNA regions that ChxR has been shown to bind to, making it a good negative control for non-specific DNA binding. Surprisingly, ChxR bound to these negative control DNAs just as strongly as it did to its own promoter. Multiple harsh washes did not abolish this presumably non-specific binding and studies using CT009 were not carried out due to the inability to detect specific DNA binding by ChxR using this technique (Fig. 4.6).

**FIG. 4.6. Fluorescence-Linked Immunosorbent Assay (FLISA) was not effective in identifying the previously reported positive interaction between the protein ChxR and its promoter.** Wells, lined with streptavidin, were incubated in the presence of 2 μg of biotinylated promoter regions of DNA. ChxR promoter is the region previously reported to bind ChxR. CT870/871 is a promoter region shared by the two diverging genes, CT870 and CT871. ChxR is not expected to bind to this promoter region and therefore this served as a negative control. CT032/033 is a region of DNA in between two converging genes, CT032 and CT033. It does not contain the promoter of any gene and therefore was expected to serve as a negative control. ChxR protein bound to the well containing its promoter DNA. However, ChxR protein also bound to the two negative control DNAs. Controls without protein, DNA or secondary antibody confirm a low level of background fluorescence.


+ + + + + + - - DNA
- + - + - + - - ChxR
+ + + + + + + + 1° Ab
+ + + + + + + - 2° Ab
ChxR Promoter
CT870/871
CT032/033

***Bacterial One-Hybrid Assay was not effective in identifying a CT009 DNA binding sequence.***

In an effort to identify sequences of DNA bound by CT009 protein (and thereby facilitate identification of gene targets) a Bacterial one-Hybrid (B1H) system was utilized. This system is well characterized and has been utilized previously to identify DNA binding sequences of proteins (Meng *et al.*, 2005, Meng *et al.*, 2006, Meng & Wolfe, 2006, Noyes *et al.*, 2008). The B1H system utilizes two plasmids: a vector expressing the protein of interest fused to the alpha subunit of RNA polymerase (bait) together with a library of reporter vectors, each with a different 25 base pair sequence of DNA upstream of the promoter for two reporter genes, *HIS3* and *URA3* (prey).

These reporters provide both positive selection and negative selection. In any given transformed cell, the DNA sequence upstream of the promoter for the reporter genes will either be recognized and bound by the protein of interest or not. If the protein of interest does not recognize the sequence, then the alpha subunit of RNA polymerase will not be stably held at the promoter and transcription of the reporter genes will not occur at a significant level. If the protein of interest does recognize the sequence, then the alpha subunit of RNA polymerase will be held at the promoter, allowing for efficient transcription of the reporter genes. The *HIS3* reporter gene allows for growth of *E. coli* on plates containing the compound 3-amino-triazole, thereby allowing only transformants in which the protein-DNA interaction is a positive one to survive on plates containing this compound.

Because it is possible that a random sequence might activate transcription independently of the fusion protein, the library is first cleared of plasmids that activate in

the absence of fusion protein. This is achieved by the action of the second reporter gene, *URA3*; the protein product of this gene converts 5-fluoro-orotic acid (5-FOA) to a toxic product, killing cells with activated promoters in the presence of 5-FOA. Therefore, by pre-screening the library of prey DNA against self-activating sequences, sequences of DNA recognized specifically by the protein of interest can by identified.

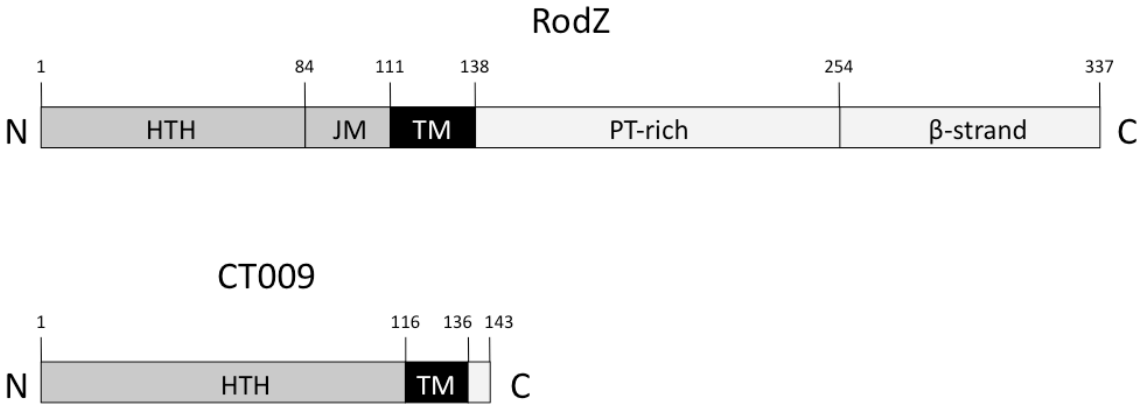
This system is fundamentally different from previously attempted methods to identify CT009 DNA binding sequences because it does not require purified recombinant protein and instead relies on protein produced within the heterologous cell system of *E. coli*. While it would not directly identify gene targets of CT009, after the identification of several DNA binding sequences, it could potentially allow for elucidation of a DNA binding consensus sequence, which could then be used to identify putative gene targets from the chlamydial genome sequence.

Difficulties in this approach came from problems in the construction of the prey library of plasmid DNA. Ligation independent cloning of the randomized DNA binding site into a plasmid did not yield sufficient numbers of colonies to ensure the construction of a library with acceptable levels of diversity in the randomized sequence.

***A computationally generated protein structure (I-TASSER) of the cytoplasmic domain of CT009 is most structurally similar to the protein RodZ.***

As previously described, CT009 possesses a predicted helix-turn-helix motif, as predicted by its primary sequence. However, in contrast to many, but not all, protein sequences containing Xre DNA binding domain was the presence of a predicted carboxy-terminal single transmembrane helix formed by amino acids 117-136 (Fig. 4.7) in CT009,

**FIG. 4.7. Predicted domain organization of RodZ and CT009**. The predicted domain architecture of RodZ has been described previously (Bendezu et al., 2009). It appears to have an N-terminal cytoplasmic domain that contains a helix-turn-helix motif (HTH), a highly basic juxta-membrane domain (JM), a transmembrane helix (TM), a periplasmic domain rich in prolines and threonines (PT-rich) and a periplasmic domain with predicted β-sheets. As predicted by HMMTOP, CT009 also possesses an N-terminal cytoplasmic domain (HTH; amino acids 1-116) and a transmembrane helix (TM; amino acids 117-136). However, it lacks a basic juxta-membrane domain and its predicted periplasmic region is only 7 amino acids long (amino acids 137-143).


RodZ
1
84
111
138
254
337
N
HTH
JM
TM
PT-rich
β-strand
C
CT009
1
116
136
143
N
HTH
TM
C

as predicted by the HMMTOP server (Tusnady & Simon, 2001).

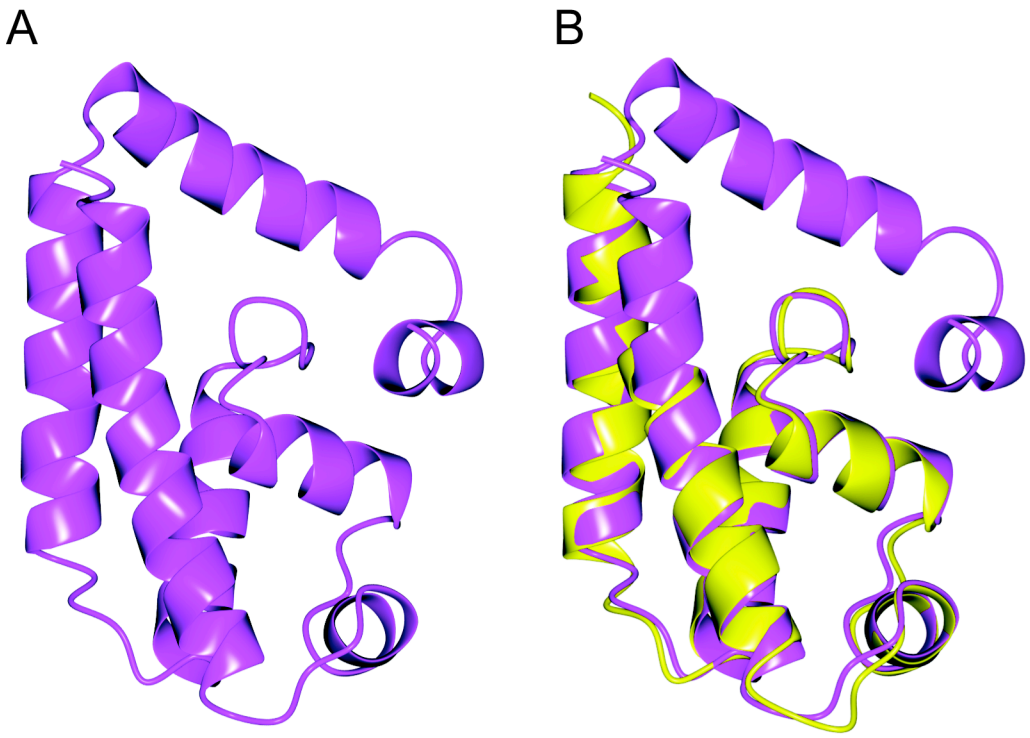
In order to gain more specific information regarding the function and possible biological role of CT009, a model of the structure of the cytoplasmic domain of the protein was generated using I-TASSER. I-TASSER, as well as other protein modeling methods, is limited in the ability to accurately model membrane associated protein regions. Therefore, the predicted N-terminal cytoplasmic domain of CT009 was analyzed ($CT009^{1-116}$). Several models were generated; the highest scored model had a C-score above the recommended cutoff (Fig. 4.8A; C-score of -0.92) and was used to search for proteins with structural similarity, using the DALI server (Holm & Rosenstrom, 2010). Interestingly, the top structure that aligned best with the I-TASSER model of $CT009^{1-116}$ was not cI (which was most similar to the full length CT009; Fig. 4.3) but was the cytoplasmic domain of RodZ (chains R from PDBid: 2WUS), with Z-scores of 14.5 and 13.2. This RodZ structure aligns to Model 1 with an r.m.s.d. score of 0.9 Å (over 86/87 residues), despite low sequence similarity to CT009 (BLAST E value >10;).

RodZ is a membrane-bound cell shape determining protein, a component of the bacterial cell morphogenic complex and an interaction partner of the bacterial actin homolog MreB (Alyahya *et al.*, 2009, Bendezu et al., 2009, Shiomi *et al.*, 2008, van den Ent et al., 2010). In addition to possessing structural similarity to RodZ in their cytoplasmic domains, CT009 and RodZ also are predicted to share a similar membrane topology; a single transmembrane helix anchors both proteins in the inner membrane (Fig. 4.7; (Bendezu et al., 2009)). RodZ also possesses two MreB-interacting residues, conserved in RodZ homologs. The only available structure of RodZ (from *Thermatoga maritima*) is a co-crystal with MreB. In this crystal structure, several hydrophobic

**FIG. 4.8. A computationally modeled structure of the predicted cytoplasmic domain of CT009 has structural similarity to cytoplasmic domain of the protein RodZ.**

A. The amino acid sequence of the predicted cytoplasmic domain of CT009 (amino acids 1-116) was used to build a model of CT009 structure using the I-TASSER server.

B. The I-TASSER model of CT009 (purple) was aligned to the previously solved structure of RodZ (2WUS chain R; yellow) with an r.m.s.d. score of 0.89 Å over 86 residues.


A
B

residues were identified at the interaction interface between RodZ and MreB, several of which are also required for proper interaction between RodZ and MreB (Fig. 4.9D; (van den Ent et al., 2010)). Among these were residues Y53 and Y57, bulky hydrophobic amino acids in the fourth helix of the protein and correspond to F60 and Y64 in *E. coli* RodZ. Importantly, despite overall low sequence similarity between RodZ and CT009, these residues are conserved in CT009 as Y57 and F61 are predicted to be similarly positioned in the helix 4 of CT009 (Fig. 4.10D), as analyzed by Clustal Omega and ESPript (Sievers *et al.*, 2011, Gouet *et al.*, 2003). Therefore, based on both the overall predicted structural similarity and the presence of conserved residues critical for function, the I-TASSER generated model of $CT009^{1-116}$ suggests that this protein may be a chlamydial RodZ homolog.

***The crystal structure of CT009 shows structural similarities to RodZ.***

In addition to computational modeling, recombinant CT009 (cytoplasmic domain, residues 1-116) was expressed, purified and crystallized. A data set was collected and processed to a resolution of 1.25 Å, with one molecule of CT009 in the asymmetric unit (Fig. 4.9A). Curiously, the protein appears to be crystallized in a flat "open" conformation. The physiological relevance of this form is dubious; proteins in solution typically adopt a more compact fold, minimizing the hydrophobic surface area exposed to solvent. However, analysis of the crystal packing revealed that a molecule related by a crystallographic 2-fold operator (Y, X, -Z) results in the positioning of the C-terminal region near the N-terminal region. Together, these crystallographic dimer partners form a "folded" conformation that one might expect, particularly when compared to a

homologous protein RodZ (Fig. 4.9C). This phenomenon, in which two polypeptide chains exchange subdomains, is an uncommon (but not unheard of) occurrence and is most frequently observed in structural studies that utilize x-ray crystallography (Bennett *et al.*, 1994). This swapping of subdomains can be an artifact of crystal packing but sometimes the subdomain-swapped form exists in solution as well (Barrientos *et al.*, 2002). Electron density at the "hinge" of this subdomain swapping, (Fig. 4.9B) demonstrates that the protein is indeed in a flat conformation. This electron density is further highlighted (Fig. 4.9D). A predicted structure of a CT009 monomer (from amino acids 1-54 of one monomer in the crystallographic dimer and amino acids 55-103 of the other crystal-mate) can be constructed (Fig. 4.9A). Like the modeled structure of CT009, the conserved MreB-interacting residues (Y57 and F61) are apparent in the crystal structure, facing outward in a position capable of interaction with MreB and highlighted in red (Fig. 4.10A).

In addition to the subdomain-swapped crystallographic dimer, recurring positive difference (F*o*-F*c*) electron density at the C terminal helix, following refinement, indicated that this portion of the helix adopted two distinct conformations and was modeled as such. Examples of the electron density in this helix can be seen in the backbone carbonyls of residues Lys 85 through Phe 87 (Fig. 4.9E) and in the sidechain of Phe 87 (Fig. 4.9F).

Superposition of RodZ onto the predicted monomer resulted in an r.m.s.d. of 1.02 Å over 66/87 residues (Fig. 4.10C). The major difference between the structure of CT009 and the structure of RodZ is in the orientation of their C-terminal helices. In RodZ, this final helix (amino acids 69-86) is orientated straight up. However, in the

**Table 4.3**. Crystallographic data for CT009 refined to 1.25Å resolution.

| | CT009 |
|---|---|
| **Data Collection** | |
| Unit-cell parameters (Å, °) | $a$=55.17, $c$=88.12 |
| Space group | $P4_12_12$ |
| Resolution (Å)[1] | 46.76-1.25 (1.27-1.25) |
| Wavelength (Å) | 1.0000 |
| Temperature (K) | 100 |
| Observed reflections | 318,621 |
| Unique reflections | 38,420 |
| <I/s(I)>[1] | 25.1 (2.5) |
| Completeness (%)[1] | 99.9 (98.5) |
| Multiplicity[1] | 8.3 (7.5) |
| $R_{merge}$ (%)[1, 2] | 3.6 (82.4) |
| $R_{meas}$ (%)[1, 4] | 3.8 (88.5) |
| $R_{pim}$ (%)[1, 4] | 1.3 (32.0) |
| $CC_{1/2}$ [1, 5] | 0.999 (0.816) |
| **Phasing** | |
| Maximum LLG | 2176 |
| **Refinement** | |
| Resolution (Å) | 29.21-1.25 |
| Reflections (working/test) | 36,417/1,921 |
| $R_{factor}$ / $R_{free}$ (%)[3] | 15.9/17.3 |
| No. of atoms (Protein/Water) | 1,007/111 |
| **Model Quality** | |
| R.m.s deviations | |
| Bond lengths (Å) | 0.012 |
| Bond angles (°) | 1.437 |
| Average $B$-factor (Å$^2$) | |
| All Atoms | 20.2 |
| Protein | 18.8 |
| Water | 32.9 |
| Coordinate error (maximum likelihood) (Å) | 0.13 |
| Ramachandran Plot | |
| Most favored (%) | 99.2 |
| Additionally allowed (%) | 0.8 |

1) Values in parenthesis are for the highest resolution shell.
2) $R_{merge} = S_{hkl} S_i |I_i(hkl) - \langle I(hkl)\rangle| / S_{hkl} S_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the $i$th reflection and $\langle I(hkl)\rangle$ is the average intensity of all reflections with indices hkl.
3) $R_{factor} = S_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / S_{hkl} |F_{obs}(hkl)|$; Rfree is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.
4) $R_{meas}$ = redundancy-independent (multiplicity-weighted) $R_{merge}$(Evans, 2011, Evans, 2006). $R_{pim}$ = precision-indicating (multiplicity-weighted) $R_{merge}$(Diederichs & Karplus, 1997, Weiss, 2001).
5) $CC_{1/2}$ is the correlation coefficient of the mean intensities between two random half-sets of data (Karplus & Diederichs, 2012, Evans, 2012).

**FIG. 4.9. A high-resolution structure of CT009 adopts an "open" fold as a result of crystal packing, resulting in a subdomain swap dimer.**

A. The predicted cytoplasmic domain of CT009 (amino acids 1-116) was purified and crystallized. A set of data at 1.25 Å was collected and a structure was modeled.

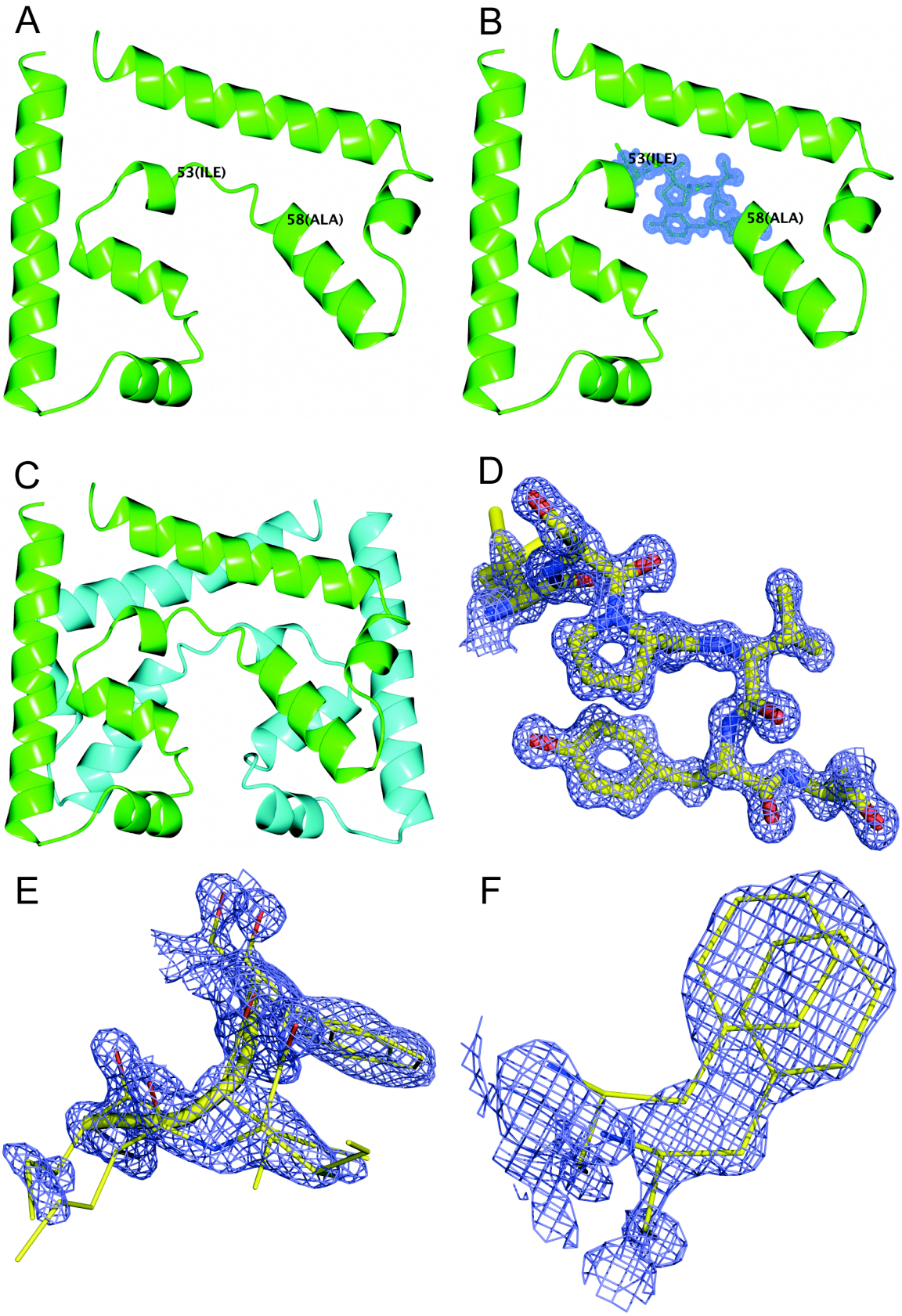
B. Electron density at the "hinge" of the subdomain swap (between amino acids I53 and A58).

C. A crystallographic dimer of the CT009 structure (green and cyan).

D. A closer view of the electron density at the "hinge" of the subdomain swap (between amino acids I53 and A58).

E. Electron density highlighting the backbone carbonyls of the C-terminal helix, which was modeled as being in two conformations, contoured at 2 sigma (residues K85, I86 and F87 shown).

F. Electron density highlighting the alternate conformation of the C-terminal helix, showing the two conformations of the sidechain of F87.


A
53(ILE)
58(ALA)
B
53(ILE)
58(ALA)
C
D
E
F

structure of CT009, this helix is kinked sharply to the right and is broken into two distinct helices, one short (amino acids 73-78) and the other longer (amino acids 81-101). The reason for this difference is not obvious, although it is possible that the final helix in CT009 is kinked so that it can make better crystal contacts with its crystal mates. Importantly, this helix is not directly involved in contacts with MreB and the MreB-interacting surface appears to be structurally conserved in CT009. Therefore, despite the unusual crystal packing of the CT009 subdomains, the crystal structure of the cytoplasmic domain of CT009 supports the hypothesis that it is a structural homolog to the protein RodZ.

***CT009 interacts with itself and with MreB.***

As discussed previously, RodZ plays a role in rod-cell shape maintenance in other bacteria as part of a morphogenic complex with the bacterial actin homolog MreB. Although CT009 may be a functional homolog of RodZ, this does not explain what role this protein has in *Chlamydiae*, because these bacteria are coccoid in shape. It was recently proposed that, in *Chlamydiae*, cell shape-determining proteins, including the MreB protein, are involved in chlamydial cell division (Brown & Rockey, 2000, Ouellette et al., 2012). RodZ is thought to play a critical role in tethering MreB to the membrane and interacts with MreB through conserved amino acids at a well-defined interaction interface (van den Ent et al., 2010). Given the structural similarity and conservation of these key MreB-interaction residues in CT009, we tested the hypothesis that CT009 interacts with chlamydial MreB and requires these interactive residues using

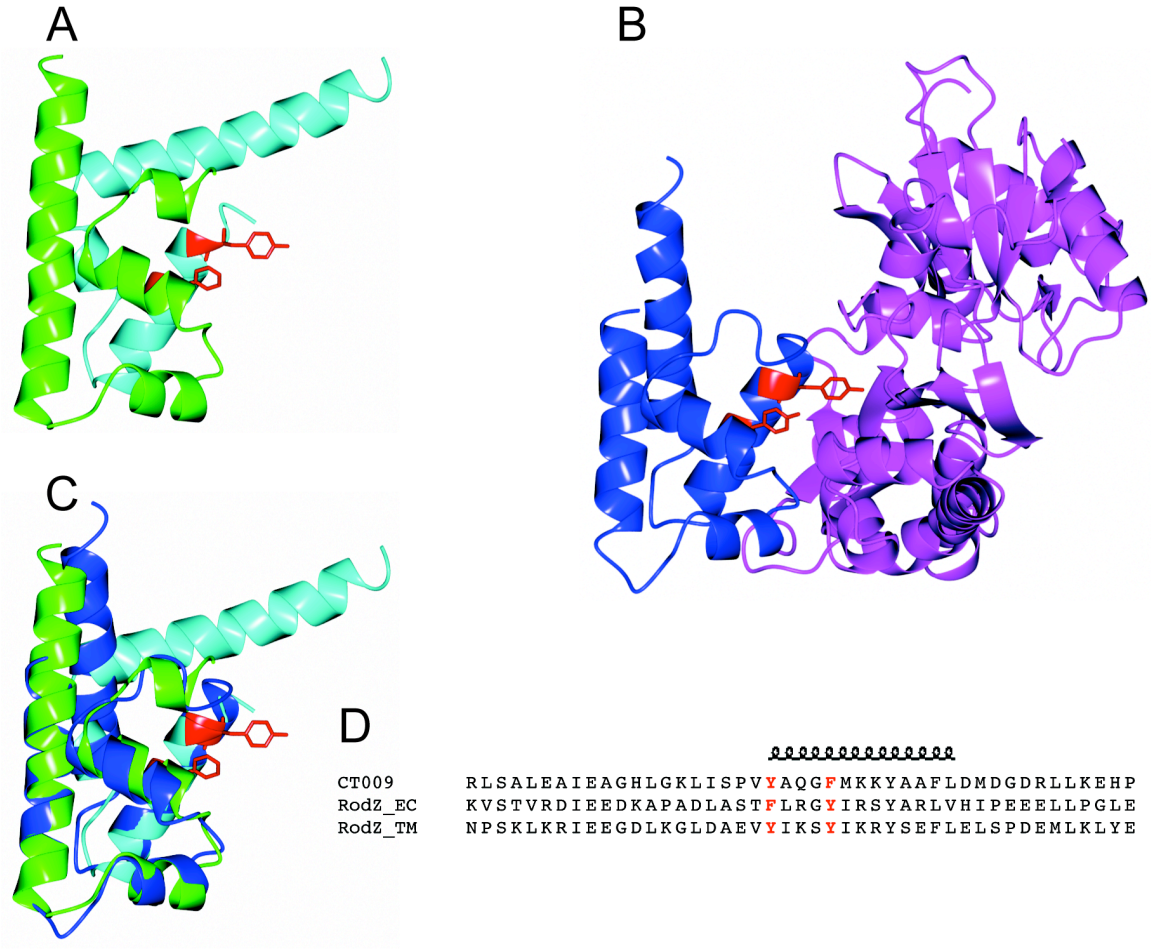
**FIG. 4.10. The crystallographic dimer of CT009 has structural similarity to RodZ, including a conservation of key MreB-interacting residues.**

A. A theoretical structure of a CT009 monomer created by displaying amino acids 1-54 of one monomer in the crystallographic dimer and amino acids 55-103 of the other crystal-mate; conserved MreB-interacting residues (Y59 and F61) are highlighted in red.

B. Crystal structure of RodZ (blue) and MreB (magenta) from *Thermotoga maritima* (PDBid: 2WUS); conserved MreB –interacting residues (Y53 and Y57) are highlighted in red.

C. A superposition of the theoretical monomer of CT009 (green and cyan) and the previously solved structure of RodZ (2WUS chain R; blue) with an r.m.s.d. score of 1.02 Å over 66 residues.

D. A sequence alignment of CT009, RodZ from *E. coli* (RodZ_EC) and RodZ from *T. maritima* (RodZ_TM), illustrating conservation of MreB-interacting residues; the α-helix in which these amino acids lie is labeled.

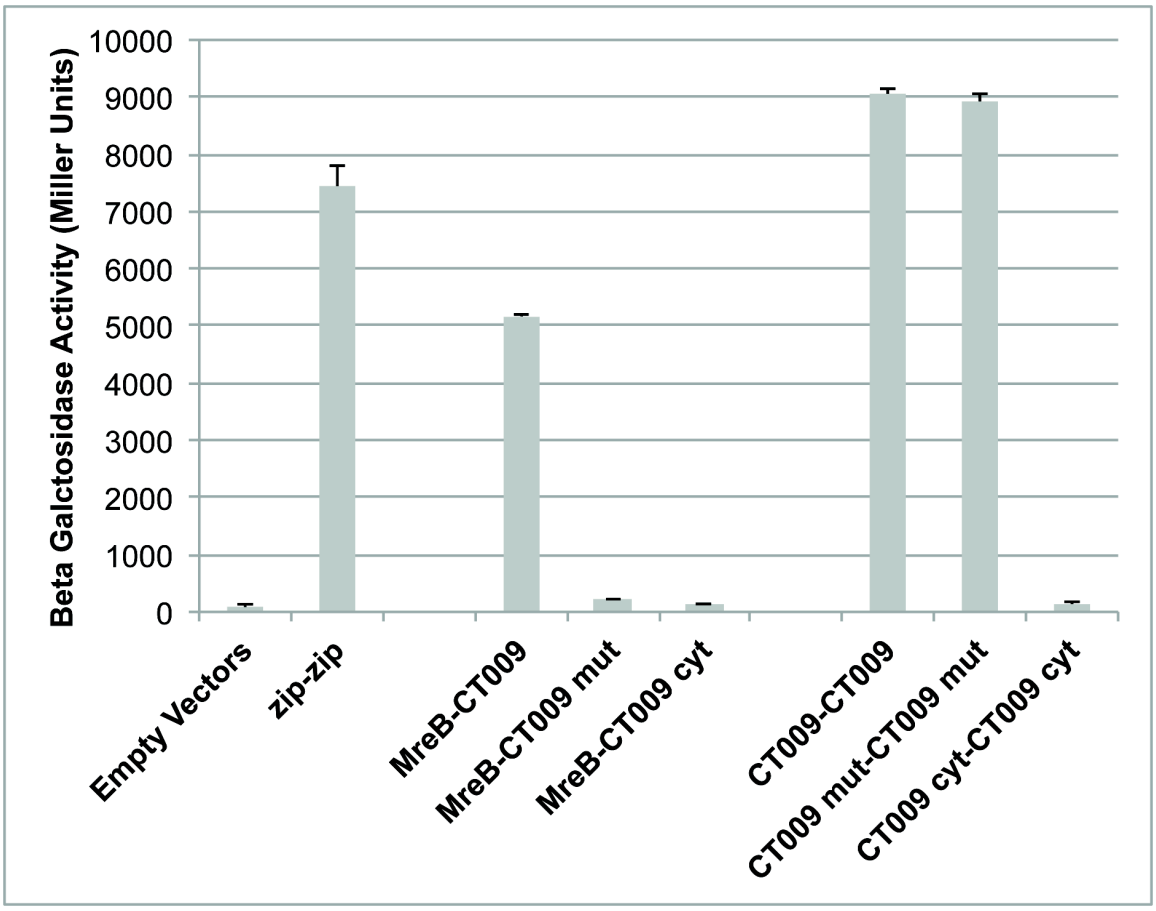

A
B
C
D
CT009 RLSALEAIEAGHLGKLISPVYAQGFMKKYAAFLDMDGDRLLKEHP
RodZ_EC KVSTVRDIEEDKAPADLASTFLRGYIRSYARLVHIPEEELLPGLE
RodZ_TM NPSKLKRIEEGDLKGLDAEVYIKSYIKRYSEFLELSPDEMLKLYE

the Bacterial Adenylate Cyclase Two Hybrid (BACTH) system (Fig. 4.11; (Karimova *et al.*, 2005, Karimova et al., 1998)).

RodZ and MreB from a variety of species have been shown to interact in the context of a bacterial-two hybrid system and even co-crystallize together (Bendezu et al., 2009, Kleinschnitz et al., 2011, van den Ent et al., 2010, White et al., 2010). Our analysis shows a positive interaction between CT009 and chlamydial MreB proteins (Fig. 4.10, CT009-MreB). All eight different possible combinations of interactions (with differing orientations of the fusion proteins on CT009 and MreB) were tested (data not shown). Any interactions in which the adenylate cyclase domain was fused to the C-terminus of CT009 failed to interact with MreB. CT009 is predicted to be anchored to the inner membrane, with its N terminus in the cytoplasm and its C terminus in the periplasm (Fig. 4.7). It is expected that fusion proteins attached to the C terminus of CT009 might not function properly, as this domain is then likely localized in the periplasm. All four interaction combinations using N-terminally tagged CT009 showed interaction between CT009 and MreB (data not shown).

As discussed previously, RodZ from *Thermatoga maritima* has been shown to interact with MreB and several residues that are critical for this process have been identified (van den Ent et al., 2010). Two of these residues (Y53 and Y57 in *T. maritima*) are bulky hydrophobic amino acids that are positioned at the interface between RodZ and MreB; despite low overall sequence similarity between CT009 and RodZ, these residues are conserved in CT009 as Y57 and F61. In order to test whether these large hydrophobic amino acids are directly involved in CT009-MreB interaction, CT009 Y57A, F61A was analyzed by BACTH. To ensure that these mutations have not altered

**FIG. 4.11. Interactions between CT009 and MreB by Bacterial Adenylate Cyclase Two Hybrid (BACTH).** Full length CT009, mutant CT009 Y57A, F61A, the predicted cytoplasmic domain of CT009 (residues 1-116), and chlamydial MreB interactions were analyzed. As a positive control, self-interaction of the leucine zipper of the GCN4 protein was tested. Expression of the β-galactosidase reporter gene by reconstituted adenylate cyclase was assessed using a β-galactosidase assay. Activities were calculated in Miller units. Error bars represent standard deviations calculated from assays performed in triplicate.

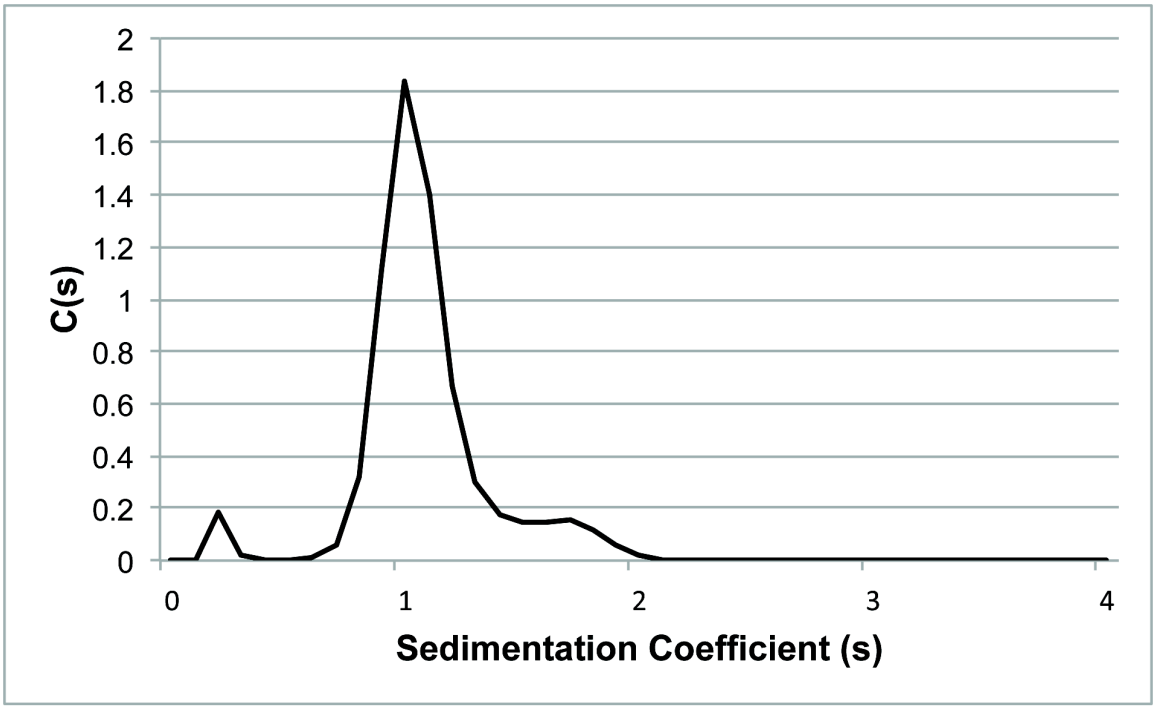

Beta Galctosidase Activity (Miller Units)
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Empty Vectors
zip-zip
MreB-CT009
MreB-CT009 mut
MreB-CT009 cyt
CT009-CT009
CT009 mut-CT009 mut
CT009 cyt-CT009 cyt

the overall fold of the protein, dimerization interactions with CT009 Y57A, F61A were studied (Fig. 4.11, CT009 mut-CT009 mut). CT009 with these two point mutations is still capable of interacting with itself, demonstrating that these mutations are likely not leading to a complete misfolding of the protein. However, as expected, these mutations do lead to a loss of interaction with MreB (Fig. 4.11, CT009 mut-MreB).

In addition to interacting with MreB, RodZ has been shown to interact with itself, although the necessity of this self-interaction for function has not been explored (Bendezu et al., 2009, Kleinschnitz et al., 2011, White et al., 2010). In our studies with recombinant protein, the cytoplasmic domain of CT009 did not form higher order structures in solution as analyzed by analytical ultracentrifugation (Fig. 4.12). In order to assess the ability of full length CT009 to interact with itself, CT009 was expressed as a fusion with either the T25 or T18 domain of adenylate cyclase fused to either the N or C terminus. All possible combinations of CT009 self-interaction (with differing orientation of the fusion proteins) were tested (data not shown) and one combination showed positive interaction (Fig. 4.11, CT009-CT009). This was the condition in which both the T25 and T18 domains were fused to CT009 at its N terminus. This is likely for the same reasons that C-terminal fusions to CT009 failed to interact with MreB (discussed above). This suggests that, like RodZ, full length CT009 interacts with itself.

As mentioned previously, during purification of recombinant CT009 the predicted cytoplasmic domain did not form higher order structures with itself, as analyzed by analytical ultracentrifugation (Fig. 4.12). This observation supported the hypothesis that the CT009 dimer seen in the crystal structure (Fig. 4.9C) was an artifact due to crystal packing and likely is not physiologically relevant. To test this, self-interactions with the

**FIG. 4.12. Sedimentation velocity profile of CT009 1-116 by analytical ultracentrifugation shows that its molecular weight is consistent with that of a monomer.** C(s) represents the relative concentration of each species. The largest peak represents 90.01% of the total material and is centered at 1.057s, which corresponds to a molecular weight of 13.18 kDa. The molecular weight of CT009 1-116 is 13.37 kDa.


C(s)
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
Sedimentation Coefficient (s)

cytoplasmic domain of CT009 were tested using BACTH (Fig. 4.11, CT009 cyt-CT009 cyt). In contrast to the full-length protein, the cytoplasmic domain of CT009 does not interact with itself. Again, all possible combinations of interactions (with differing orientations of the fusion proteins) were tested, but no combination lead to a dimerization interaction (data not shown). This suggests that the transmembrane helix and/or the short periplasmic tail may be involved in mediating or strengthening dimerization of CT009. Interaction between the cytoplasmic domain of CT009 and MreB was also tested (Fig. 4.11, CT009 cyt-MreB). No interactions between these two proteins were observed in any of the eight possible fusion protein configurations (data not shown). This suggests that either dimerization of the protein or its localization to the membrane may be required for MreB interaction.

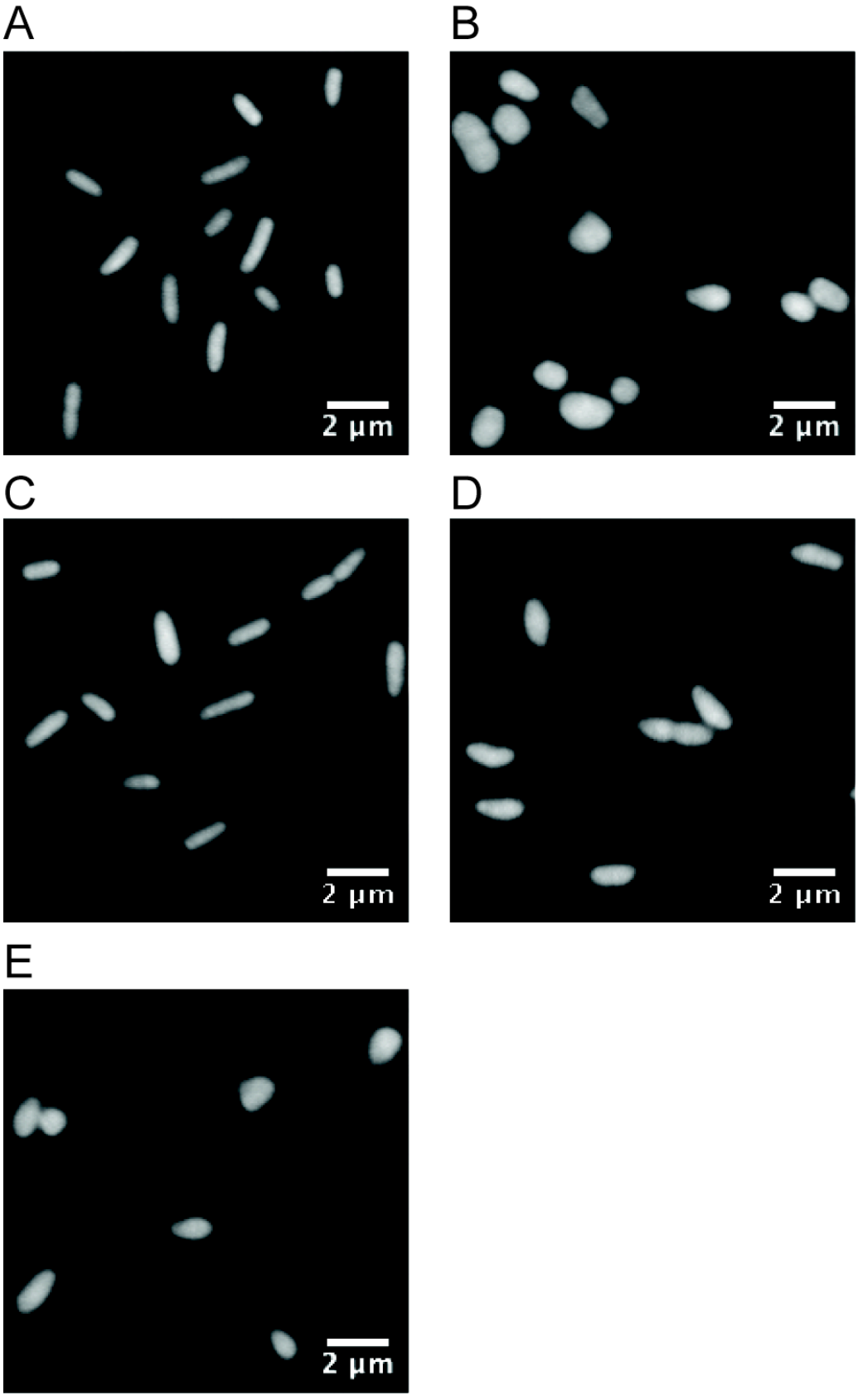
***CT009 is able to functionally complement RodZ* in vivo.**

With the hypothesis that CT009 is a chlamydial RodZ homolog, its ability to complement the function of RodZ in the heterologous system of *E. coli* was tested. RodZ is involved in directing cell shape and in *rodZ* knockout strains of *E. coli* there is a well-defined phenotype: cells grow to be spherical instead of rod-shaped (Bendezu et al., 2009, Shiomi et al., 2008). *rodZ* knockout cells (FB60) were transformed with plasmids that either encoded RodZ, CT009 or were empty. As a control, a K-12 strain of *E. coli* (MC1000) was used as well. In addition, a plasmid containing a mutant CT009 (CT009 Y57A, F61A) was also transformed into FB60 cells. This mutant protein has alterations at the MreB interaction interface and, unlike wild type protein, does not interact with MreB by BACTH.

To test for complementation of the loss of RodZ phenotype, plasmid-containing FB60 cells and MC1000 cells were grown to exponential phase, fixed, stained with the dye malachite green and visualized by fluorescence microscopy (Fig. 4.13). The images presented are representative of three independent experiments. As expected for K-12 *E. coli*, MC1000 cells are rod shaped (Fig. 4.13A). As previously described, *rodZ* deficient *E. coli* cells (FB60) with an empty vector have an aberrant morphology and are more spheroid with less of a defined rod shape (Bendezu et al., 2009) (Fig. 4.13B). Also described previously, when FB60 cells are complemented with a plasmid expressing RodZ, wild type phenotype is restored and cells have rod morphology (Bendezu et al., 2009) (Fig. 4.13C). When FB60 cells are complemented with a plasmid expressing CT009, rod morphology is partially restored (Fig. 4.13D). Although cells supplemented with CT009 are longer and more rod-like than RodZ knockout cells, they are also wider than cells complemented with RodZ. This may be because CT009 does not possess a predicted periplasmic domain; a previous report in which RodZ knockout cells complemented with a truncated RodZ (lacking its periplasmic domain) showed complementation accompanied by widening of cells (Bendezu et al., 2009). Additionally, when FB60 cells express a CT009 that contains mutations in the two conserved residues that mediate the interacting between RodZ and MreB (Y57A, F61A), cells fail to restore wild-type phenotype (Fig. 4.13E). Together, these data suggest that CT009 is a functional homolog to RodZ in *C. trachomatis*.

**FIG. 4.13. CT009 is able to functionally complement RodZ in *E. coli*.**

A-E. Representative images of *E. coli* cells that were fixed, stained with malachite green, and imaged by epifluorescent microscopy. Phenotypes of (A) wild type *E. coli*, (B) FB60; *rodZ* deficient *E. coli* (Bendezu et al., 2009), (C) FB60 cells ectopically expressing RodZ, (D) FB60 ectopically expressing CT009, and (E) FB60 cells ectopically expressing CT009 Y57A/F61A mutant.


A
2 µm
B
2 µm
C
2 µm
D
2 µm
E
2 µm

***Computational analyses identify potential RodZ homologs in other bacterial species.***

At the time of its original annotation, CT009 contained a predicted helix-turn-helix motif as well as a predicted XRE family DNA binding domain. For this reason, it was annotated as a transcription factor. However, our structural and functional studies on the protein now show that CT009 has properties of a homolog to RodZ, interacting with MreB and likely involved in chlamydial cell division. Support for our initial exploration of CT009 as a homolog to RodZ rather than a transcription factor came from structural similarity and its possession of conserved residues involved in MreB interaction. Another key feature of RodZ that is conserved in CT009 is the presence of a C-terminal transmembrane helix. Given the large number of proteins with high sequence similarity to CT009 that are annotated as Xre family DNA binding domain or not annotated functionally (hypothetical), we hypothesized that many of these are also RodZ homologs.

To support this hypothesis, the protein sequence of CT009 was used in a BLAST search (Altschul et al., 1997). The top 20 proteins of similar sequence to CT009 (non-chlamydial, from non-redundant genera) were first analyzed by the HMMTOP server, which predicts the presence of transmembrane helices and protein topology (Tusnady & Simon, 2001). Fourteen of these proteins were found to have a single predicted transmembrane helix, with the N-terminus of the proteins oriented in the cytoplasm (like RodZ; Table 4.4). These fourteen proteins were then analyzed further. All of these proteins are either annotated as hypothetical proteins (HP) or as transcription factors, either helix-turn-helix containing proteins (HTH) or XRE-family members (XRE). All of the species that produce these proteins are rod shaped in morphology and all encode an

MreB homolog (except for *Carboxydibrachium pacificum*), suggesting a need for a RodZ homolog. Importantly, none of these species have an annotated homolog to RodZ.

Finally, the amino acid sequences of these proteins were aligned to that of *E. coli* RodZ and *C. trachomatis* CT009 using ClustalW2 (Larkin *et al.*, 2007). As mentioned previously, structural studies of RodZ and MreB revealed two bulky hydrophobic residues that are crucial for mediating interactions between RodZ and MreB (Fig. 4.8D). Conservation of these MreB-interacting residues was examined and all CT009 homologs, except for one, possess these key amino acids. While the protein ANT_03850 from the filamentous thermophile *Anaerolinea thermophile* lacks these residues, its sequence homology to CT009 along with its predicted transmembrane helix suggest that it may be a RodZ homolog that mediates interactions with MreB through a novel mechanism. While this is not an exhaustive exploration of potential RodZ homologs among all bacterial species, it does showcase the power of going beyond sequence similarity to identify proteins of homologous function in other organisms.

**Table 4.4**. Probable RodZ homologs based on CT009 homology

| Name | Annotation | Organism | Transmembrane Helix | CT009 E value | MreB-interacting Residues Present |
|---|---|---|---|---|---|
| RodZ | RodZ | *Escherichia coli* | 112-131 | 0.004 | Yes, F60, Y64 |
| CT009 | HTH | *Chlamydia trachomatis* | 117-136 | N/A | Yes, Y57, F61 |
| ALO_19967 | XRE | *Acetonema longum* | 109-128 | 3e-14 | Yes, F49, F53 |
| CYB_0195 | HP | *Synechococcus sp.* | 110-129 | 1e-10 | Yes, Y56, F60 |
| Moth_1073 | XRE | *Moorella thermoacetica* | 100-117 | 1e-10 | Yes, Y49, F53 |
| Thewi_1795 | HP | *Thermoanaerobacter wiegelii* | 109-127 | 2e-10 | Yes, Y49, F53 |
| Btus_1597 | XRE | *Kyrpidia tusciae* | 107-125 | 2e-10 | Yes, Y49, F53 |
| Thit_1638 | HP | *Thermoanaerobacter italicus* | 110-128 | 4e-10 | Yes, Y49, F53 |
| CDSM653_1192 | HP | *Carboxydibrachium pacificum* | 107-126 | 5e-10 | Yes, Y49, F53 |
| Daes_1731 | HTH | *Desulfovibrio aespoeensis* | 115-133 | 6e-10 | Yes, Y51, F55 |
| TC41_1408 | XRE | *Alicyclobacillus acidocaldarius* | 157-181 | 7e-10 | Yes, Y65, F53 |
| syc0492_c | HP | *Synechococcus elongatus* | 112-134 | 1e-9 | Yes, F57, F61 |
| ANT_03850 | XRE | *Anaerolinea thermophila* | 241-260 | 2e-9 | No, Q181, M185 |
| PTH_1295 | HP | *Pelotomaculum thermopropionicum* | 105-122 | 3e-9 | Yes, Y48, F52 |
| TM_1864 | HP | *Thermotoga maritima* | 105-122 | 4e-9 | Yes, Y53, Y57 |
| Thexy_0625 | XRE | *Thermoanaerobacterium xylanolyticum* | 112-129 | 7e-9 | Yes, Y49, F53 |

# Chapter IV

## Discussion

A BLAST search using CT009 protein sequence as input revealed the presence of CT009 homologs in all species of Chlamydia with sequenced genomes, including all serovars of *C. trachomatis*, *C. pneumoniae*, *C. psitacci*, *C. muridarum*, *C. abortus*, *C. pecorum*, *C. felis* and *C. caviae*. This suggests a conserved role for CT009 that is fundamental to the biology of *chlamydiae*. Interestingly, the CT009 ortholog in *C. muridarum* possesses a additional 8 amino acids at its N terminus. The significance of these amino acids is not apparent.

Comparative sequence analyses with CT009 reveals a predicted helix-turn-helix motif within a conserved XRE (Xenobiotic Response Element)-family DNA binding domain (Altschul et al., 1997). The Xre family (IPR001387) is expansive in regards to phylogenetic distribution and associated with many DNA regulatory functions including Lambda phage (cI/Cro) related transcription repressors (Aggarwal *et al.*, 1988, Minezaki *et al.*, 2005, Lewis *et al.*, 1998). Additionally, an initial computational model of the full length CT009 protein showed structural similarity to the transcriptional repressor cI. Therefore, initial studies on CT009 were aimed at the identification of its gene targets in *Chlamydia*.

In an effort to identify chlamydial genes regulated by CT009, several different techniques were employed. An initially attempted method was Systematic Evolution of Ligands by Exponential Enrichment (SELEX). This technique utilizes multiple rounds of DNA binding in order to enrich for sequences of DNA are most strongly recognized by

the protein. However, in my analysis, after the first round of selection, DNA was present in the negative controls (conditions with no antibody to selectively remove the protein-DNA complex from solution). This indicates that the DNA was binding nonspecifically to the tubes, the beads or the protein G. This experiment was repeated twice more with minor variations, but this nonspecific binding could not be eliminated (data not shown). Because of this negative result, other techniques were subsequently employed to elucidate CT009 DNA gene targets.

Electromobility Shift Assays (EMSA) have been previously used to identify and characterize interactions between chlamydial transcription factors and the promoters of the genes they regulate. However, a weakness of this technique is that candidate genes must be chosen and tested one by one, making the identification of all potential gene targets impractical. Twelve genes were selected as potentially regulated by CT009 based on the presence of an identifiable promoter regions and their differential regulation within the chlamydial developmental cycle. However, no shifts were observed in any assays conducted using these promoter regions. Another problem with this assay is the absence of a positive control for CT009; this is because no gene targets for CT009 have been previously identified. Therefore, a failure to produce a shift in this assay might be due to problems with either the folding/stability of the recombinant protein and/or the buffer conditions used in the binding reaction. Failure to produce shifts could also be due to CT009 not recognizing DNA sequence in those promoters; in this case, a method in which more promoter regions could be tested at the same time would be helpful.

In order to screen large numbers of promoter region DNAs for CT009 binding, Fluorophore-Linked Immunosorbent Assays (FLISA) was performed. Like EMSA, this

technique relies on purified recombinant protein interacting with a small fragment of DNA in vitro, but because it is performed in a 96 well plate, it allows for the screening of many potential gene targets at once. However, this technique did not prove helpful as negative controls failed to work properly, the reasons for which are not clear.

All three previously attempted techniques for identifying CT009 gene targets (SELEX, EMSA, FLISA) relied on properly folded purified recombinant protein. However, the quality of recombinant CT009 cannot be tested due to the lack of a reliable assay for function. Therefore, a technique that does not rely on purified protein was employed: Bacterial one-Hybrid (B1H). This assay system relies on protein-DNA interactions *in vivo* in the heterologous system of *E. coli*. Despite its theoretical effectiveness, difficulties were encountered in the construction of the prey library of plasmid DNA. Without a robust library of prey DNA sequences, this assay could not be effectively performed.

Concurrent with some of the investigations regarding CT009 gene target identification were ongoing structural studies of the CT009 protein, both computational and experimental. As both of these structural studies suggested that CT009 might be a functional homolog to RodZ rather than a transcription factor, investigations of CT009's DNA binding capabilities ceased.

Regarding the computational structure prediction of CT009, an important observation to highlight is the accuracy at which the I-TASSER server modeled the structure of CT009 and the importance of this structural information in the ultimate identification of this protein's role. The model generated was similar to the structure of RodZ with a r.m.s.d. value of 0.89 Å over 86 residues; this model lead to the initial

hypothesis that CT009 is a RodZ homolog. The accuracy of this modeling is of note because of the extremely low sequence similarity between CT009 and RodZ protein sequences (BLAST E value >10) and particularly challenging category (i.e. *ab initio*) of computational protein structure modeling.

It should also be noted that the initial model of CT009, using the full protein sequence, was more similar to the transcription factors cI (Fig. 4.3). This predicted structure similarity, and the predominant association of helix-turn-helix domain with nucleic acid interactions, supported a hypothesis that CT009 functions as a transcription factor. As highlighted previously, accurate modeling of membrane-associated regions of proteins is very difficult for many structural modeling servers (I-TASSER included). Only after the putative trans-membrane region of CT009 was identified and excluded was a confidently scored model (of the cytoplasmic domain) of the protein generated, leading to the hypothesis that CT009 functions as a RodZ homolog. These observations with computational structural modeling of CT009 therefore demonstrate some of the strengths and weakness of a structure-based annotation of protein function. However, it is clear that structural modeling can provide very valuable information regarding the function of a protein with otherwise ambiguous or unclear function.

The overall global structure similarities between both the computationally generated and the experimentally determined structures of the cytoplasmic domain of CT009 to RodZ leant support to the hypothesis that CT009 functions as a RodZ homolog. However, global fold of the protein alone did not *strongly* support this hypothesis: other proteins such as cI and cro were also structurally similar to the cytoplasmic domain of CT009. The hypothesis of CT009 functioning as RodZ in *chlamydiae* was also supported

by the aforementioned presence of a single pass transmembrane helix (Fig. 4.6); this is a feature possessed by RodZ but not by cI or cro. Additionally, upon examination of the primary sequence of CT009, two amino acids that are conserved in the function of RodZ were identified, despite the low overall sequence similarity between these two proteins (Fig. 4.10C and D). The presence of these residues also leant support to the hypothesis that CT009 is a homolog to RodZ. Importantly, all of these features (global structure, predicted membrane topology and conserved residues) were utilized together to strongly support that hypothesis that CT009 is a homolog to RodZ.

Although this is the first published study to focus on the chlamydial protein CT009, this protein has been analyzed before. A recent search for additional transcription factors in chlamydial species, using a bioinformatics approach, identified CT009 as the number one strongest predicted DNA-binding protein (Akers *et al.*, 2011). This is not surprising, as we have now experimentally shown that CT009 possesses a helix-turn-helix motif. Although this motif is most frequently utilized for DNA binding, there are examples of this motif functioning as an interface for protein-protein interactions (Aravind *et al.*, 2005). Importantly, RodZ also contains a helix-turn-helix motif; however, it utilizes this motif to stabilize protein-protein interactions, specifically with MreB (van den Ent et al., 2010). Therefore, the helix-turn-helix motif of CT009 is also likely involved in interactions with MreB, based on its structural and functional homology to RodZ. Interestingly, chlamydial species already encode a relatively low number of transcription factors and these data exclude CT009 as a candidate. Similarly, we previously reported that CT296 (Kemege et al., 2011), originally described as a divalent cation repressor (Rau et al., 2005, Wyllie & Raulston, 2001), does not contain

structural similarity to any transcription factor nor function as originally described. The search for additional factors involved in regulation of chlamydial gene expression will inevitably continue. In the future it may be prudent to consider the possibility that not all proteins with a helix-turn-helix motif are necessarily utilizing this motif for the purposes of DNA binding.

Although the cytoplasmic domain of CT009 is structurally similar to RodZ and this membrane localized domain functionally complements RodZ in *E. coli,* the cells were wider than wild type (Fig. 4.10A *vs* 4.10D). This phenotype may be explained by the absence of a predicted periplasmic domain in CT009. CT009 itself is significantly shorter than RodZ (Fig 4.6) although the cytoplasmic domains are of very similar size. The major difference is the presence of a large periplasmic domain in RodZ (199 amino acids) while the predicted periplasmic domain of CT009 is only 7 amino acids long. Deletion studies on the different domains of RodZ have demonstrated that the periplasmic domain of RodZ is not strictly required for the maintenance of cell shape. Importantly, deletion of this domain led to a widening of cells (Bendezu et al., 2009). Therefore, it is unsurprising that, in the absence of this domain, CT009 is still able to functionally complement RodZ by rendering rod shaped cells that are also wider than normal.

The precise function of the periplasmic portion of RodZ has not been defined, but there have been 16 RodZ homologs identified that lack this domain (Alyahya et al., 2009). These RodZ proteins with a periplasmic domain shorter than 10 amino acids were typically found in bacteria with reduced genomes or with round morphology (including but not limited to *Rickettsia* spp., *Xanthamonas campestris* and *Staphylococcus* spp.),

suggesting dispensability for this domain. Unfortunately, RodZ homologs in these species have not, to date, been experimentally analyzed. Being both coccoid in morphology and possessing a reduced genome, it is unsurprising that CT009 also lacks a predicted periplasmic domain.

In the context of a BACTH assay, full length CT009 interacted with itself while the cytoplasmic domain alone did not (Fig. 4.11). Therefore, dimerization of CT009 is likely mediated in part by its transmembrane region and/or the short periplasmic tail. It is also possible that other, as of yet unidentified, protein partners of CT009 may contribute to its dimerization, but only if it is localized properly. Interestingly, the cytoplasmic domain of CT009 fails to interact with MreB (Fig. 4.11). Therefore, it is possible that proper localization of CT009 by its transmembrane anchor is required for its interaction with MreB. It is also possible that dimerization of CT009 is required for MreB interaction; removal of the transmembrane helix and the predicted short periplasmic tail abolished CT009 self-interactions as well. The dimerization interface of RodZ is not well characterized and it is possible that dimerization may put CT009 in a configuration required for proper interaction with MreB. However, this possibility is not well supported by the sole crystal structure of RodZ, which shows interaction between a monomer of RodZ and MreB (van den Ent et al., 2010).

It has been shown in other bacteria that MreB interacts with several proteins other than RodZ, including RodA, MreC, and MreD, forming a morphogenic complex. RodA is an integral membrane protein and is predicted to be a flippase involved in cell wall elongation. CT726 has been annotated as a chlamydial RodA homolog although no published experimental work has been done on this protein. MreD is an integral

membrane protein with 4-6 membrane-spanning helices. While its precise function is unclear, it has been shown to interact with MreB and is required for the maintenance of proper cell shape. Importantly, there is no apparent chlamydial homolog to MreD. MreC spans the inner membrane with a single helix and possesses a large C-terminal domain localized to the periplasm. Its presence in *Chlamydiae* is somewhat unclear. Although there was no annotated MreC gene in the initially sequenced chlamydial genome, more recent annotations list CT638 as an MreC homolog. An analysis of CT638 protein sequence by HMMTOP reveals a single predicted transmembrane helix, consistent with the topology of MreC (data not shown). However, CT638 has only very weak homology to MreC and there exists no experimental evidence for the function of CT638 as an MreC homolog. Further investigation into these other members of the MreB morphogenic complex could give deeper insight into the function of this complex in *Chlamydiae*.

The process of cell division in chlamydial species is poorly understood due to the lack of apparent homologs to well-conserved bacterial cell division components. The most notable cell division protein absent in chlamydial genomes is FtsZ, which forms the Z ring and acts as the central coordinator for cell division. It was recently proposed that chlamydial MreB may play a role in cell division, possibly as a functional replacement for FtsZ (Ouellette et al., 2012). While FtsZ is a major component of the cell division machinery, it does not function alone; many other proteins are involved in both stabilizing the Z ring and ensuring that it forms in an appropriate location. More specifically, nucleoid occlusion factors such as SlmA prevent Z-ring formation from disrupting the bacterial chromosome while negative regulators of FtsZ polymerization, such as the MinCDE proteins, prevent Z-ring formation at the cell poles. These are

critical functions in cell division and are their absence would lead to division defects. Importantly, homologs to these proteins are not apparent in chlamydial species by genome annotation. In this study, we have identified a chlamydial homolog to the protein RodZ; the foundation of this discovery was based on structural studies. In the future, structural information (computationally and/or experimentally derived) may be instrumental in the identification of these factors and will, in turn, lead to a better understanding of the fundamental process of chlamydial cell division.

The majority of the support for the hypothesis that chlamydial species utilize cell-shape determining proteins (specifically MreB) for mediating cell division comes from a recent report in which chemical inhibition of MreB lead to a non-dividing state in chlamydial cells (Ouellette et al., 2012). Non-experimental support for this hypothesis lies in the observation that chlamydial genomes encode only three apparent cell division genes (out of the ~20 cell division genes conserved in other bacteria) and encode several cell shape determining genes while possessing a coccoid morphology. Members of the phylum *Chlamydiae* have particularly small genomes (~1 megabase) and are unlikely to encode proteins that do not serve a necessary function, implying that these morphogenic proteins have a key function in the biology of *Chlamydia*. In concept, the process of directing sidewall-specific peptidoglycan synthesis (the function of cell shape determining proteins) is similar to the process of directing cell division; both processes must ensure that polymerization of peptidoglycan occurs at a specific site. Of course, there are many differences between these two processes. Nonetheless, the use of shape determining proteins for mediating cell division is an intriguing hypothesis, which bears more investigation. For example, what prevents MreB-controlled septation from

contacting (and damaging) the chromosome? What role does peptidoglycan synthesis play in chlamydial cell division and what is the molecular structure of chlamydial peptidoglycan? What are the precise molecular functions of proteins such as MreB, RodZ, RodA and FtsK in *Chlamydia*? Clearly we have only just begun to understand cell division in chlamydia.

# Chapter V

# Discussion

## Utility and Limitations of the use of Structural Information in Functional Characterization of Proteins

Historically, the use of experimentally derived structural information has been reported to be successful in informing function among hypothetical proteins (Dey *et al.*, 2013). Large structural genomics consortia have demonstrated this on a genome-wide scale, studying organisms such as *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Thermotoga maritima*, and *Mycobacterium tuberculosis* (da Fonseca *et al.*, 2012, Sacchettini, 2013, Shin et al., 2007, Wilson, 2013). These bacteria were focused on due to their small genome size, impact on public health or utility as a model system for structure determination. However, despite its significant impact on public health, its relatively small genome and its large number of annotated proteins of unknown function, species of *Chlamydia* have not been the subject of a structural genomics project. The scope of the research presented in this document is on scale much smaller than these projects but the relationships between structure and function in three chlamydial proteins can be discussed.

CT296 was annotated as a protein of unknown function, but initial experimental studies on the protein demonstrated a functional relationship to the Fur protein. Our structural studies on this protein were successful in some ways and unsuccessful in others. Both computational and experimental structural studies were unsuccessful simply

because they ultimately did not lead to the identification of a possible function for the CT296 protein. The structure of the CT296 protein contains a partial cupin barrel fold, which is widely used by many different classes of proteins, and therefore identification of a putative function for CT296 was not possible. However, the overall fold of CT296 is structurally similar to several Fe(II) and 2-oxoglutarate-dependent enzymes (of diverse enzymatic function). A cursory analysis of the structure would have therefore lead to the conclusion that CT296 is an enzyme that requires these cofactors. However, further analysis of the protein lead to a more accurate conclusion: upon examining the sequence and structure of CT296 for amino acid sidechains capable of coordinating these two enzymatic cofactors, it was found that such residues are not apparent in CT296. Therefore, because CT296 does not appear to be able to coordinate these molecules on its own, it is likely not an enzyme of this class. This illustrates a limitation of the protein structure-function relationship; namely, that many proteins of different functions share a common global fold and that the presence of this global fold does relatively little to inform function. This also highlights the necessity of going beyond merely correlating global fold to function and using other structural details (such as the coordination of cofactors) to make an assessment of protein function.

However, the structure of the CT296 protein was not without merit. Previously demonstrated by another group to bind to DNA as a putative transcription factor, the structure of CT296 revealed an absence of any conserved DNA-binding structural motif. This finding lead to a re-examination of the studies that showed its DNA binding properties. Upon close examination, it was found that these published studies lacked key controls. Therefore, the structure of CT296 lead to a change in the putative function of

the protein. Although a putative function was not clearly identified, its function as a DNA-binding protein was called into question. Being one of very few transcription factors identified in chlamydial species, CT296 has been used as a control transcription factor in several experiments by other groups and therefore it was important to the chlamydial field as a whole to report the lack of DNA-binding capabilities of the CT296 protein. Another important observation regarding the structural analysis of CT296 was that the computationally modeled structure was remarkably similar to the experimentally determined structure, an illustration of the effectiveness of the I-TASSER server, even on proteins with low sequence homology to proteins of known function.

Ultimately, the function of the CT296 protein is not clear. It is likely not functioning as a transcription factor and it does not appear to be able to bind Fe(II) or 2-oxoglutarate, like its closest structural homologs. However, it may be speculated that the function of CT296 involves another protein. The crystal structure of CT296 revealed an incomplete cupin barrel fold, lacking several necessary opposing beta sheets to complete the fold; therefore, it is possible that a protein partner to CT296 might complete this fold. Although this does not point to a definitive function for CT296, the structure of such a binding partner (or perhaps a co-crystal structure with both proteins) could be more informative. The technique of co-immunoprecipitation, using antibodies against CT296 applied to a chlamydial cell lysate could potentially identify such a binding partner.

Another example of a chlamydial protein of unknown function with a solved structure is CT584. The story behind this protein is interesting because X-ray crystallography was not the only technique used to explore the structure of CT584. Initially, the secondary structure characteristics of CT584 were examined, particularly the

changes in secondary structure in response to changes in pH and temperature (Markham et al., 2009). While the secondary structure characteristics of a protein are generally much less informative regarding a protein's putative function, these studies were carried out with a hypothesis concerning the function of CT584 already proposed: namely, that it is a chlamydial T3SS needle tip protein. Even with a hypothesized function, these secondary structure experiments were only worth performing because of previous structural studies of the same nature performed on other known T3SS needle tip proteins. Ultimately, the secondary structure characterization of CT584 supported the hypothesis that it is a T3SS needle tip protein. However, further studies, both functional and structural, were needed to more solidly support this hypothesis.

The structure of CT584 (its closely related *Chlamydia pneumoniae* ortholog, Cpn0803) was solved by another group while my functional studies of the protein were ongoing (Stone et al., 2012). The structure of this protein was less than illuminating, however, and a function for the protein was not made clear by the elucidation of its structure. Like, CT296, the structure of CT584 called into question its previously purported function; in this case, the chlamydial T3SS needle tip protein. This was primarily due to an overall structure with little similarity to known needle tip proteins. However, the investigators also assessed protein features beyond global fold: they examined the features of its multimeric form and found that such a multimer did not possess an inner diameter capable of allowing proteins through for secretion. Additionally, they found identified the residues in the CT584 ortholog responsible for interaction with the needle protein and found that they would be buried in a hexamer, suggesting that an oligomer of the protein (found in T3SS needle tips) would not be

capable of binding to the tip of the needle. Although the global structure alone suggested a non-needle tip function for the protein, these further studies on the detailed structural elements and oligomerization characteristics suggested this more strongly. Therefore, although a function cannot be assigned to CT584, these structural studies have not leant support to the hypothesis that it is a T3SS needle tip protein. This is in line with my own functional studies of CT584, which also do not support this hypothesis. Importantly, I have also generated crystals of CT584; the structure of this protein, solved by Scott Lovell, is very similar to that of Cpn0803 and all the same functional conclusions could be drawn from it. Overall, these studies on CT584 illustrate how the three dimensional structure of a protein ultimately gives more credible information than information regarding its secondary structure alone. Additionally, they demonstrate how structural information can sometimes be at odds with other information regarding a protein's function.

Importantly, the structure of CT584 was also modeled computationally. Unlike CT296, this model did not suggest a clear function for the protein; closest structural homologs to the model were transcriptional regulators of diverse family and function. Additionally, the CT584 model lacks structural similarity to the experimentally determined structure of the *C. pneumonia* ortholog, Cpn0803. Therefore, the case of CT584 is an example by which computational structural modeling was wholly unsuccessful in proving useful information for functional assignment.

Regarding the function of CT584, neither the crystal structure nor my functional analyses point toward a definitive function for this protein. However, a very recent report showed that CT584 interacts with the putative T3SS substrate CT082 and increases its

secretion (Pais *et al.*, 2013). This suggests that CT584 is functioning as a T3SS chaperone. This role is supported by my localization studies, showing a primarily cytoplasmic localization of the protein. Further studies would be aimed at the identification of additional proteins chaperoned by CT584 and/or the identification of a molecular function for its known partner, CT082.

The third and final example of the protein structure-function paradigm involves the chlamydial protein CT009. Annotated as a protein of unknown function, the primary sequence for CT009 did give some clue to its function by the presence of a putative helix-turn-helix motif, primarily found in DNA binding proteins. Due to its low solubility, the structure of CT009 was analyzed computationally long before it was crystallized. This initial computational model of full-length CT009 showed structural similarity to the cI and cro transcriptional repressors, leading to studies on the DNA binding properties of CT009. However, a single transmembrane helix was eventually identified at the C-terminus of the CT009 protein. Removal of this transmembrane helix lead to increased protein solubility but, importantly, it also lead to a significantly different model. When the structure of the cytoplasmic domain of CT009 alone was modeled, the resulting structure resembled the protein RodZ as well as the proteins cI or cro. Importantly, the crystal structure of the cytoplasmic domain of CT009 very closely resembled the model of this domain.

However, these structures of CT009 were about as similar to RodZ about as they were to cI or cro and therefore the overall structure of CT009, on its own, was not very conclusive. In formulating a hypothesis regarding the function of CT009, information outside of the global fold of the protein was leveraged. In examining published reports

concerning RodZ, cI and cro, it was found that RodZ, like CT009, possesses a single-pass transmembrane helix while cI and cro do not. Additionally, although RodZ, cI and cro all possess helix-turn-helix motifs, RodZ does not use this motif for DNA binding; given the failed attempts at getting CT009 to bind to DNA, this observation suggests that CT009 might be more functionally related to RodZ. Furthermore, all analyzed sequences of RodZ possess two conserved hydrophobic residues, involved in protein-protein interactions. Although the overall sequence similarity between CT009 and RodZ is low, these two residues are conserved in CT009 are in a similar orientation in the two structures. Therefore, based on these further analyses of the CT009 sequence and structure, CT009 was hypothesized to be a functional homolog to RodZ and not to cI or cro. This underlines the importance of utilizing all bioinformatics tools available for the analysis of a protein's sequence and structural properties and not just its global fold when forming hypotheses regarding its function. In many ways, CT009 proved to be an ideal example of the protein structure-function paradigm because the structure of the protein, as well as these other characteristics, lead clearly to a hypothesis regarding its function.

**Chlamydial Structural Genomics**

As discussed above, several bacteria have been the focus of structural genomics projects in efforts to elucidate the structures of all proteins in their respective proteomes. Importantly, despite their relatively small genome size, their significant impact on public health and their abundance of proteins of unknown function, no chlamydial species have been selected for study by structural proteomics.

Despite the relative success of these groups in solving the crystal structures of a large number of proteins from a single organism, the process of structural genomics is still challenging and expensive. A less costly alternative to the experimental solution of these structures is the computationally modeling of all identified proteins in a genome. While considerably less costly than an experimental approach, computational structural genomics is less effective simply because the accurate modeling of proteins, particularly proteins with no homologs by primary sequence, is challenging. Despite this difficulty, the computational protein structure modeling illustrated in this report underlines the overall effectiveness of this technique.

As discussed above, the crystal structure of CT296 did not clearly point to a potential function for the protein. However, the computationally generated model of CT296 was remarkably like the crystal structure, despite the lack of sequence homologs with known structures. In contrast, the modeled structure of CT584 was entirely dissimilar to the crystal structure of the *C. pneumoniae* ortholog. The sequence similarity between these two related proteins is very high and so this is likely not a reason for this discrepancy. CT584 simply lacks significant sequence homology to proteins of known structure, causing the chances for its successful modeling to be relatively low. Finally, the modeled structure of the cytoplasmic domain of CT009 was very similar to the crystal structure of this protein and suggested the same functional role in the cell. Importantly, an accurate structure was only possible when the cytoplasmic domain of the protein was analyzed. The accurate modeling of membrane helices and integral membrane proteins is not as developed as the algorithms for modeling soluble domains of proteins and this limits the number of proteins in the chlamydial genome that can be confidently modeled.

Despite their shortcomings, these three examples of computational modeling of chlamydial proteins illustrate the potential effectiveness of applying computational structure modeling to the entire chlamydial proteome. As shown with CT296, CT584 and CT009, such an approach could give important insights into the biological roles and functions of the many hypothetical proteins in the chlamydial genomes.

**CT584 is a poor potential chlamydial vaccine target**

Although chlamydial infections can be effectively treated with antibiotics, infections are frequently asymptomatic. Consequently, infected individuals may not only forgo treatment but may unwittingly pass their infection on to others, due to lack of detection. Additionally, chlamydial resistance to commonly used antibiotics has already been reported in *C. suis*, a species that infects pigs. Therefore, strategies that lead to the prevention of infection by chlamydial species are of particular interest.

One such strategy for prevention of chlamydial infection is the development of an effective vaccine against *C. trachomatis* in humans. Although research to this end has been intense, a vaccine against human chlamydial infection has yet to be developed. Vaccination using whole, inactivated organisms is a commonly used strategy for vaccines against other pathogens. However, vaccines composed primarily of inactivated EBs have been shown in the past to actually enhance disease rather than prevent it (Grayston *et al.*, 1961, Nichols *et al.*, 1966, Woolridge *et al.*, 1967). For this reason, subsequent research has focused on the development of protein subunit vaccines.

Antigens to be used to these subunit vaccines can be categorized into four classes: antigens that are exposed on the surface of chlamydial organisms, antigens that are exposed on the surface of chlamydial inclusions, antigens that are in the chlamydial cytoplasm and antigens that are secreted into the host cytoplasm (Murthy & Arulanandam, 2012). The majority of research has focused on antigens exposed on the surface of chlamydia as these antigens are expected to induce neutralizing antibodies that could prevent initial attachment and entry of organisms. Surface-exposed chlamydial antigens tested for vaccine efficacy include MOMP, OmcB, PorB and the PmP proteins, to name a few.

One category of surface-exposed antigens that has yet to be examined for efficacy in Chlamydia is proteins associated with the external portion of the Type III Secretion System (T3SS). The exposed components of this system include the needle protein and the needle tip protein. The T3SS needles and needle tip proteins of other pathogenic bacteria have displayed immunogenic behavior and are therefore attractive vaccine candidates (Overheim et al., 2005, Barrett *et al.*, 2010). Therefore, identification of the chlamydial T3SS needle tip protein is on high importance for its potential use in a protein subunit vaccine.

Current studies regarding CT584 do not support its role as a chlamydial T3SS needle tip protein. Although CT584 has some characteristics of a T3SS needle protein, it ultimately lacks a primary surface localization on either EBs or RBs. Additionally, neutralization studies, utilizing anti-CT584 antibodies to neutralize the infective capabilities of *C. trachomatis* EBs were not successful (Ichie Osaka, personal

communication). Therefore, it is likely poorly immunogenic and therefore a poor candidate for a vaccine component.

**Genetic Exchange in Chlamydia**

The development of a robust system for genetic exchange in *chlamydiae* has been of high priority for many years now. These organisms have resisted attempts at classical genetic manipulation techniques such as transformation, likely due to the inherent characteristics of chlamydial biology. The proteins of the outer membranes of EBs are highly crosslinked, thereby making the passage of DNA through this membrane by heat shock or electroporation largely unsuccessful. Purified RBs are less impermeable, but they are not infectious and because the chlamydial developmental cycle relies on infection of a host cell for its completion, any transformed RBs would therefore not be able to replicate. A robust system for culturing chlamydia axenically has also been a challenge to the field. Finally, attempts to introduce DNA into RBs in their natural state (in the context of an infection) would prove difficult because and DNA would have to cross the host cell membrane, the inclusion membrane and then the inner and outer membranes of the RBs themselves. Clearly, the biology of chlamydial organisms is not conducive to genetic manipulation.

Despite these challenges, there have been recent successes in the field of chlamydial genetics. Different species of chlamydia apparently recombine their DNA with one another in the context of a co-infection; an observation that was utilized to transfer an antibiotic resistance gene from *C. suis* to *C. trachomatis* (Suchland *et al.*,

2009). However, this usefulness of this technique is limited because genetic information can only be transferred from one species of Chlamydia to another and therefore new genes cannot be introduced. Additionally, forward genetics approaches have had some success in identifying genes involved in glycogen accumulation (Nguyen & Valdivia, 2012). This involves chemical mutagenesis of chlamydial organisms followed by isolation of clones, screening for a phenotype, and finally whole genome sequencing. However, this is not a targeted approach and can only identify genes with a non-lethal and easily identifiable phenotype.

The most promising report in the field of chlamydial genetic exchange has been the recent description of a shuttle vector and a transformation protocol that has allowed for the stable transformation of EBs with a plasmid that encodes GFP (Wang *et al.*, 2011). While GFP was used merely as a proof of concept, this technology can theoretically allow for stable transformation of chlamydia with plasmids encoding any protein.

While supplying chlamydia with exogenous protein or overexpressed quantities of an endogenous protein could give insight into protein function, a system for targeted, temporally controlled gene knockout could have wider-reaching consequences. There are many ways in which a target gene knockout could be achieved. For example, antisense RNA could be encoded on a plasmid, preventing translation of a target gene. Alternatively, DNA binding proteins with designable DNA target sequences could be employed to block gene expression of desired genes at the transcriptional level. Finally, dominant negative versions of proteins (if the function of these protein are well-known) could be overexpressed, leading to a loss of function post-translationally. Temporal control of the expression of these plasmid-encoded factors could be achieved by the use

of inducible promoters. Regardless of the exact mechanisms used, the discovery of a technique for stable transformation of chlamydia opens up possibilities for further discoveries.

The applications for the technique of chlamydial transformation and the possible technique of targeted gene knock-down can be imagined for virtually every aspect of chlamydial research, including the three proteins of unknown function described in this report. CT296, in its role as DcrA, could be examined by knocking down expression of the protein and assessing the phenotype. If it plays a role in regulating the iron response, a phenotype for a CT296 deficient strain might only be seen in response to iron limitation. However, our findings point to a different function for CT296, a function that is less clear. Knocking down of CT296 expression at different timepoints during the developmental cycle could give insight into its function by determining when (if at all) it is required for cellular function.

Recent studies suggest that CT584 is involved in the T3SS function, likely not as a needle tip protein but rather as a chaperone for T3SS effectors (Stone et al., 2012). If this is true, its knockdown could have consequences on infection, inclusion maintenance or immune evasion. By examining the exact phenotype associated with CT584 knockdown, a more precise role for it could be elucidated. Importantly, if it is critical for EB invasion of the host cell, stable strains of CT584 knockout would not be viable, making it paramount to employ temporally regulated gene knockdown for such experiments. However, our findings regarding the localization of CT584 do not support its role in T3S. In this case, targeted gene disruption of CT584 could point to a different function by this protein by showing a unique phenotype for cells lacking it.

As a homolog to RodZ, the role of CT009 in the biology of *Chlamydia* is not entirely clear. In other bacteria, RodZ is associated with rod shape determination and its knockout results in spherically shaped bacteria. However, *chlamydiae* are not rod shaped and the need for a RodZ homolog is therefore not apparent. Recent evidence implicates another rod-shape determining protein in the process of cell division, leading to the hypothesis that CT009 is also involved in this process. Targeted knockdown of CT009 expression in dividing RBs could test this hypothesis; if CT009 is involved in cell division, CT009 deficient cells would show no defects in growth or infection, but an inability to divide. Additionally, if CT009 and other rod-shape determining proteins have a role in RB-EB conversion, this would also become apparent through a phenotype in which RBs grow but do not convert to EBs.

## References


Abdelrahman, Y. M. & R. J. Belland, (2005) The chlamydial developmental cycle. *FEMS microbiology reviews* **29**: 949-959.

Adams, D. W. & J. Errington, (2009) Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nature reviews. Microbiology* **7**: 642-653.

Adams, M. A., M. D. Suits, J. Zheng & Z. Jia, (2007) Piecing together the structure-function puzzle: experiences in structure-based functional annotation of hypothetical proteins. *Proteomics* **7**: 2920-2932.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger & P. H. Zwart, (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213-221.

Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne & S. C. Harrison, (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science* **242**: 899-907.

Akers, J. C., H. Hodac, R. H. Lathrop & M. Tan, (2011) Identification and functional analysis of CT069 as a novel transcriptional regulator in Chlamydia. *J Bacteriol*.

Albrecht, M., C. M. Sharma, R. Reinhardt, J. Vogel & T. Rudel, (2010) Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. *Nucleic Acids Res* **38**: 868-877.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller & D. J. Lipman, (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**: 3389-3402.

Alyahya, S. A., R. Alexander, T. Costa, A. O. Henriques, T. Emonet & C. Jacobs-Wagner, (2009) RodZ, a component of the bacterial core morphogenic apparatus. *Proc Natl Acad Sci U S A* **106**: 1239-1244.

Andrews, N. C., (2000) Iron homeostasis: insights from genetics and animal models. *Nature reviews. Genetics* **1**: 208-217.

Aravind, L., V. Anantharaman, S. Balaji, M. M. Babu & L. M. Iyer, (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS microbiology reviews* **29**: 231-262.

Bailey, R. L., P. Arullendran, H. C. Whittle & D. C. Mabey, (1993) Randomised controlled trial of single-dose azithromycin in treatment of trachoma. *Lancet* **342**: 453-456.

Barrett, B. S., A. P. Markham, R. Esfandiary, W. L. Picking, W. D. Picking, S. B. Joshi & C. R. Middaugh, (2010) Formulation and immunogenicity studies of type III secretion system needle antigens as vaccine candidates. *Journal of pharmaceutical sciences* **99**: 4488-4496.

Barrientos, L. G., J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer & A. M. Gronenborn, (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**: 673-686.

Batteiger, B. E., W. Tu, S. Ofner, B. Van Der Pol, D. R. Stothard, D. P. Orr, B. P. Katz & J. D. Fortenberry, (2010) Repeated Chlamydia trachomatis genital infections in adolescent women. *The Journal of infectious diseases* **201**: 42-51.

Battey, J. N., J. Kopp, L. Bordoli, R. J. Read, N. D. Clarke & T. Schwede, (2007) Automated server predictions in CASP7. *Proteins* **69 Suppl 8**: 68-82.

Beeckman, D. S. & D. C. Vanrompay, (2010) Bacterial secretion systems with an emphasis on the chlamydial Type III secretion system. *Current issues in molecular biology* **12**: 17-41.

Belland, R. J., G. Zhong, D. D. Crane, D. Hogan, D. Sturdevant, J. Sharma, W. L. Beatty & H. D. Caldwell, (2003) Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 8478-8483.

Bendezu, F. O., C. A. Hale, T. G. Bernhardt & P. A. de Boer, (2009) RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in E. coli. *EMBO J* **28**: 193-204.

Bennett, M. J., S. Choe & D. Eisenberg, (1994) Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci U S A* **91**: 3127-3131.

Betts, H. J., L. E. Twiggs, M. S. Sal, P. B. Wyrick & K. A. Fields, (2008) Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of Chlamydia trachomatis. *J Bacteriol* **190**: 1680-1690.

Betts-Hampikian, H. J. & K. A. Fields, (2010) The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut. *Frontiers in microbiology* **1**: 114.

Blasiak, L. C., F. H. Vaillancourt, C. T. Walsh & C. L. Drennan, (2006) Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**: 368-371.

Blocker, A. J., J. E. Deane, A. K. Veenendaal, P. Roversi, J. L. Hodgkinson, S. Johnson & S. M. Lea, (2008) What's the point of the type III secretion system needle? *Proceedings of the National Academy of Sciences of the United States of America* **105**: 6507-6513.

Bosnjak, Z., S. Dzijan, D. Pavlinic, M. Peric, N. Ruzman, I. R. Krizan, G. Lauc, A. Antolovic-Pozgain, J. Burazin & D. Vukovic, (2012) Distribution of Chlamydia trachomatis serotypes in clinical urogenital samples from north-eastern Croatia. *Curr Microbiol* **64**: 552-560.

Brown, W. J. & D. D. Rockey, (2000) Identification of an antigen localized to an apparent septum within dividing chlamydiae. *Infect Immun* **68**: 708-715.

Byrne, G. I. & J. W. Moulder, (1978) Parasite-specified phagocytosis of Chlamydia psittaci and Chlamydia trachomatis by L and HeLa cells. *Infect Immun* **19**: 598-606.

Caldwell, H. D., J. Kromhout & J. Schachter, (1981) Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis. *Infect Immun* **31**: 1161-1176.

Carey, A. J., W. M. Huston, K. A. Cunningham, L. M. Hafner, P. Timms & K. W. Beagley, (2013) Characterization of In Vitro Chlamydia muridarum Persistence and Utilization in an In Vivo Mouse Model of Chlamydia Vaccine. *Am J Reprod Immunol*.

Carlson, J. H., S. F. Porcella, G. McClarty & H. D. Caldwell, (2005) Comparative genomic analysis of Chlamydia trachomatis oculotropic and genitotropic strains. *Infect Immun* **73**: 6407-6418.

CDC, (2007) Biosafety in Microbiological and Biomedical Laboratories (BMBL). U. S. D. o. H. a. H. Services (ed). Washington, DC.

Cecil, J. A., M. R. Howell, J. J. Tawes, J. C. Gaydos, K. T. McKee, Jr., T. C. Quinn & C. A. Gaydos, (2001) Features of Chlamydia trachomatis and Neisseria gonorrhoeae infection in male Army recruits. *The Journal of infectious diseases* **184**: 1216-1219.

Centers, (2010) Sexually Transmitted Disease Surveillance 2009. D. o. H. a. H. Services (ed). Atlanta, U.S.

Chai, C., Z. Xie & E. Grotewold, (2011) SELEX (Systematic Evolution of Ligands by EXponential Enrichment), as a powerful tool for deciphering the protein-DNA interaction space. *Methods Mol Biol* **754**: 249-258.

Chatterjee, S., D. Zhong, B. A. Nordhues, K. P. Battaile, S. Lovell & R. N. De Guzman, (2011) The crystal structures of the Salmonella type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate. *Protein science : a publication of the Protein Society* **20**: 75-86.

Chellas-Gery, B., K. Wolf, J. Tisoncik, T. Hackstadt & K. A. Fields, (2011) Biochemical and localization analyses of putative type III secretion translocator proteins CopB and CopB2 of Chlamydia trachomatis reveal significant distinctions. *Infect Immun* **79**: 3036-3045.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson & D. C. Richardson, (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12-21.

Chopra, I., C. Storey, T. J. Falla & J. H. Pearce, (1998) Antibiotics, peptidoglycan synthesis and genomics: the chlamydial anomaly revisited. *Microbiology* **144 ( Pt 10)**: 2673-2678.

Choroszy-Krol, I. C., M. Frej-Madrzak, A. Jama-Kmiecik, T. Bober & J. Jolanta Sarowska, (2012) Characteristics of the Chlamydia trachomatis species - immunopathology and infections. *Advances in clinical and experimental medicine : official organ Wroclaw Medical University* **21**: 799-808.

Chowdhury, R., M. A. McDonough, J. Mecinovic, C. Loenarz, E. Flashman, K. S. Hewitson, C. Domene & C. J. Schofield, (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **17**: 981-989.

Clifton, D. R., K. A. Fields, S. S. Grieshaber, C. A. Dooley, E. R. Fischer, D. J. Mead, R. A. Carabeo & T. Hackstadt, (2004) A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 10166-10171.

Clifton, I. J., M. A. McDonough, D. Ehrismann, N. J. Kershaw, N. Granatino & C. J. Schofield, (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **100**: 644-669.

Cochrane, M., C. W. Armitage, C. P. O'Meara & K. W. Beagley, (2010) Towards a Chlamydia trachomatis vaccine: how close are we? *Future microbiology* **5**: 1833-1856.

Cooksey, C. M., E. K. Berggren & J. Lee, (2010) Chlamydia trachomatis Infection in minority adolescent women: a public health challenge. *Obstetrical & gynecological survey* **65**: 729-735.

Cozzetto, D., A. Kryshtafovych, K. Fidelis, J. Moult, B. Rost & A. Tramontano, (2009) Evaluation of template-based models in CASP8 with standard measures. *Proteins* **77 Suppl 9**: 18-28.

D'Autreaux, B., L. Pecqueur, A. Gonzalez de Peredo, R. E. Diederix, C. Caux-Thang, L. Tabet, B. Bersch, E. Forest & I. Michaud-Soret, (2007) Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein. *Biochemistry* **46**: 1329-1342.

da Fonseca, M. M., A. Zaha, E. R. Caffarena & A. T. Vasconcelos, (2012) Structure-based functional inference of hypothetical proteins from Mycoplasma hyopneumoniae. *Journal of molecular modeling* **18**: 1917-1925.

Daruzzaman, A., I. J. Clifton, R. M. Adlington, J. E. Baldwin & P. J. Rutledge, (2006) Unexpected oxidation of a depsipeptide substrate analogue in crystalline isopenicillin N synthase. *Chembiochem* **7**: 351-358.

de Lorenzo, V., S. Wee, M. Herrero & J. B. Neilands, (1987) Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. *J Bacteriol* **169**: 2624-2630.

Dey, F., Q. Cliff Zhang, D. Petrey & B. Honig, (2013) Toward a "Structural BLAST": Using structural relationships to infer function. *Protein science : a publication of the Protein Society*.

Diederichs, K. & P. A. Karplus, (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**: 269-275.

Doublie, S., (2007) Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods Mol Biol* **363**: 91-108.

Dunwell, J. M., A. Purvis & S. Khuri, (2004) Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**: 7-17.

Emsley, P. & K. Cowtan, (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.

Emsley, P., B. Lohkamp, W. G. Scott & K. Cowtan, (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501.

Englund, S., C. H. af Segerstad, F. Arnlund, E. Westergren & M. Jacobson, (2012) The occurrence of Chlamydia spp. in pigs with and without clinical disease. *BMC veterinary research* **8**: 9.

Erickson, H. P., D. E. Anderson & M. Osawa, (2010) FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiology and molecular biology reviews : MMBR* **74**: 504-528.

Escolar, L., J. Perez-Martin & V. de Lorenzo, (1999) Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol* **181**: 6223-6229.

Espina, M., A. J. Olive, R. Kenjale, D. S. Moore, S. F. Ausar, R. W. Kaminski, E. V. Oaks, C. R. Middaugh, W. D. Picking & W. L. Picking, (2006) IpaD localizes to

the tip of the type III secretion system needle of Shigella flexneri. *Infect Immun* **74**: 4391-4400.

Eswar, N., B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper & A. Sali, (2006) Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 5**: Unit 5 6.

Evans, P., (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**: 72-82.

Evans, P., (2012) Biochemistry. Resolving some old problems in protein crystallography. *Science* **336**: 986-987.

Evans, P. R., (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* **67**: 282-292.

Falk, L., H. Fredlund & J. S. Jensen, (2005) Signs and symptoms of urethritis and cervicitis among women with or without Mycoplasma genitalium or Chlamydia trachomatis infection. *Sexually transmitted infections* **81**: 73-78.

Ferrer-Costa, C., H. P. Shanahan, S. Jones & J. M. Thornton, (2005) HTHquery: a method for detecting DNA-binding proteins with a helix-turn-helix structural motif. *Bioinformatics* **21**: 3679-3680.

Fields, K. A. & T. Hackstadt, (2002) The chlamydial inclusion: escape from the endocytic pathway. *Annual review of cell and developmental biology* **18**: 221-245.

Fukushi, H. & K. Hirai, (1993) Chlamydia pecorum--the fourth species of genus Chlamydia. *Microbiology and immunology* **37**: 516-522.

Ghosh, P., (2004) Process of protein transport by the type III secretion system. *Microbiology and molecular biology reviews : MMBR* **68**: 771-795.

Gouet, P., X. Robert & E. Courcelle, (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic acids research* **31**: 3320-3323.

Grayston, J. T., (1992) Infections caused by Chlamydia pneumoniae strain TWAR. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **15**: 757-761.

Grayston, J. T., R. L. Woolridge, C. W. Chen, F. A. Assaad, S. Maffei, C. H. Yen & C. Y. Yang, (1961) Bacterial conjunctivitis caused by an eye ointment base used as a placebo in therapeutic trials. *American journal of ophthalmology* **52**: 251-256.

Hammerschlag, M. R., (2000) Chlamydia pneumoniae and the lung. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **16**: 1001-1007.

Hammerschlag, M. R. & S. A. Kohlhoff, (2012) Treatment of chlamydial infections. *Expert opinion on pharmacotherapy* **13**: 545-552.

Harkinezhad, T., T. Geens & D. Vanrompay, (2009) Chlamydophila psittaci infections in birds: a review with emphasis on zoonotic consequences. *Veterinary microbiology* **135**: 68-77.

Hefty, P. S. & R. S. Stephens, (2007) Chlamydial type III secretion system is encoded on ten operons preceded by sigma 70-like promoter elements. *J Bacteriol* **189**: 198-206.

Hickey, J. M., L. Weldon & P. S. Hefty, (2011) The atypical OmpR/PhoB response regulator ChxR from Chlamydia trachomatis forms homodimers in vivo and binds a direct repeat of nucleotide sequences. *J Bacteriol* **193**: 389-398.

Holm, L., S. Kaariainen, P. Rosenstrom & A. Schenkel, (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**: 2780-2781.

Holm, L. & J. Park, (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**: 566-567.

Holm, L. & P. Rosenstrom, (2010) Dali server: conservation mapping in 3D. *Nucleic acids research* **38**: W545-549.

Holm, L. & C. Sander, (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**: 123-138.

Hsia, R. C., Y. Pannekoek, E. Ingerowski & P. M. Bavoil, (1997) Type III secretion genes identify a putative virulence locus of Chlamydia. *Mol Microbiol* **25**: 351-359.

Hu, V. H., E. M. Harding-Esch, M. J. Burton, R. L. Bailey, J. Kadimpeul & D. C. Mabey, (2010) Epidemiology and control of trachoma: systematic review. *Tropical medicine & international health : TM & IH* **15**: 673-691.

Hueck, C. J., (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and molecular biology reviews : MMBR* **62**: 379-433.

Hvidsten, T. R., A. Laegreid, A. Kryshtafovych, G. Andersson, K. Fidelis & J. Komorowski, (2009) A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity. *PLoS One* **4**: e6266.

Hybiske, K. & R. S. Stephens, (2007) Mechanisms of host cell exit by the intracellular bacterium Chlamydia. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 11430-11435.

Jagelska, E., V. Brazda, S. Pospisilova, B. Vojtesek & E. Palecek, (2002) New ELISA technique for analysis of p53 protein/DNA binding properties. *Journal of immunological methods* **267**: 227-235.

Janowski, R., S. Panjikar, A. N. Eddine, S. H. Kaufmann & M. S. Weiss, (2009) Structural analysis reveals DNA binding properties of Rv2827c, a hypothetical protein from Mycobacterium tuberculosis. *J Struct Funct Genomics* **10**: 137-150.

Jewett, T. J., E. R. Fischer, D. J. Mead & T. Hackstadt, (2006) Chlamydial TARP is a bacterial nucleator of actin. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 15599-15604.

Kabsch, W., (1988) Automatic indexing of rotation diffraction patterns. *Journal of Applied Crystallography* **21**: 67-72.

Kabsch, W., (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**: 795-800.

Kapoor, S., (2008) Re-emergence of lymphogranuloma venereum. *Journal of the European Academy of Dermatology and Venereology : JEADV* **22**: 409-416.

Karimova, G., N. Dautin & D. Ladant, (2005) Interaction network among Escherichia coli membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J Bacteriol* **187**: 2233-2243.

Karimova, G., J. Pidoux, A. Ullmann & D. Ladant, (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 5752-5756.

Karplus, P. A. & K. Diederichs, (2012) Linking crystallographic model and data quality. *Science* **336**: 1030-1033.

Kehrer, J. P., (2000) The Haber-Weiss reaction and mechanisms of toxicity. *Toxicology* **149**: 43-50.

Kemege, K. E., J. M. Hickey, S. Lovell, K. P. Battaile, Y. Zhang & P. S. Hefty, (2011) Ab initio structural modeling of and experimental validation for Chlamydia trachomatis protein CT296 reveal structural similarity to Fe(II) 2-oxoglutarate-dependent enzymes. *J Bacteriol* **193**: 6517-6528.

Khare, D., B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Hakansson & J. L. Smith, (2010) Conformational switch triggered by alpha-ketoglutarate in a halogenase of curacin A biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 14099-14104.

Khuri, S., F. T. Bakker & J. M. Dunwell, (2001) Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. *Mol Biol Evol* **18**: 593-605.

Kim, H. S., H. L. Kim, K. H. Kim, J. Kim do, S. J. Lee, J. Y. Yoon, H. J. Yoon, H. Y. Lee, S. B. Park, S. J. Kim, J. Y. Lee & S. W. Suh, (2010) Crystal structure of Tpa1 from Saccharomyces cerevisiae, a component of the messenger ribonucleoprotein complex. *Nucleic acids research* **38**: 2099-2110.

Kleinschnitz, E. M., A. Heichlinger, K. Schirner, J. Winkler, A. Latus, I. Maldener, W. Wohlleben & G. Muth, (2011) Proteins encoded by the mre gene cluster in Streptomyces coelicolor A3(2) cooperate in spore wall synthesis. *Mol Microbiol* **79**: 1367-1379.

Koo, I. C., D. Walthers, P. S. Hefty, L. J. Kenney & R. S. Stephens, (2006) ChxR is a transcriptional activator in Chlamydia. *Proc Natl Acad Sci U S A* **103**: 750-755.

Koski, M. K., R. Hieta, M. Hirsila, A. Ronka, J. Myllyharju & R. K. Wierenga, (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem* **284**: 25290-25301.

Kraus, E. & U. Femfert, (1976) Proteinase K from the mold Tritirachium album Limber. Specificity and mode of action. *Hoppe-Seyler's Zeitschrift fur physiologische Chemie* **357**: 937-947.

Krug, M., M. S. Weiss, U. Heinemann & U. Mueller, (2012) XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *Journal of Applied Crystallography* **45**: 568-572.

Kruse, T., J. Bork-Jensen & K. Gerdes, (2005) The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Mol Microbiol* **55**: 78-89.

Kumar, S. & M. R. Hammerschlag, (2007) Acute respiratory infection due to Chlamydia pneumoniae: current status of diagnostic methods. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **44**: 568-576.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Laroucau, K., A. Di Francesco, F. Vorimore, S. Thierry, J. L. Pingret, C. Bertin, H. Willems, G. Bolske & R. Harley, (2012) Multilocus variable-number tandem-repeat analysis scheme for chlamydia felis genotyping: comparison with multilocus sequence typing. *Journal of clinical microbiology* **50**: 1860-1866.

Lau, C. Y. & A. K. Qureshi, (2002) Azithromycin versus doxycycline for genital chlamydial infections: a meta-analysis of randomized clinical trials. *Sexually transmitted diseases* **29**: 497-502.

Lee, J., S. Wu & Y. Zhang, (2009) Ab Initio Protein Structure Prediction In: From Protein Structure to Function with Bioinformatics. J. D. Rigden (ed). Springer Netherlands, pp. 3-25.

Lewis, R. J., J. A. Brannigan, W. A. Offen, I. Smith & A. J. Wilkinson, (1998) An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex. *Journal of molecular biology* **283**: 907-912.

Li, Z., M. J. Trimble, Y. V. Brun & G. J. Jensen, (2007) The structure of FtsZ filaments in vivo suggests a force-generating role in cell division. *EMBO J* **26**: 4694-4708.

Lorenzini, E., A. Singer, B. Singh, R. Lam, T. Skarina, N. Y. Chirgadze, A. Savchenko & R. S. Gupta, (2010) Structure and protein-protein interaction studies on Chlamydia trachomatis protein CT670 (YscO Homolog). *J Bacteriol* **192**: 2746-2756.

Mariotti, S. P., D. Pascolini & J. Rose-Nussbaumer, (2009) Trachoma: global magnitude of a preventable cause of blindness. *Br J Ophthalmol* **93**: 563-568.

Markham, A. P., Z. A. Jaafar, K. E. Kemege, C. R. Middaugh & P. S. Hefty, (2009) Biophysical characterization of Chlamydia trachomatis CT584 supports its potential role as a type III secretion needle tip protein. *Biochemistry* **48**: 10353-10361.

Mathews, S. A., K. M. Volp & P. Timms, (1999) Development of a quantitative gene expression assay for Chlamydia trachomatis identified temporal expression of sigma factors. *FEBS Lett* **458**: 354-358.

Matthews, B. W., (1968) Solvent content of protein crystals. *J Mol Biol* **33**: 491-497.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni & R. J. Read, (2007a) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni & R. J. Read, (2007b) *Phaser* crystallographic software. *J. Appl. Cryst.* **40**: 658-674.

McDonough, M. A., K. L. Kavanagh, D. Butler, T. Searls, U. Oppermann & C. J. Schofield, (2005) Structure of human phytanoyl-CoA 2-hydroxylase identifies molecular mechanisms of Refsum disease. *The Journal of biological chemistry* **280**: 41101-41110.

McDonough, M. A., V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. Lienard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed & C. J. Schofield, (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase

(PHD2). *Proceedings of the National Academy of Sciences of the United States of America* **103**: 9814-9819.

Meng, X., M. H. Brodsky & S. A. Wolfe, (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature biotechnology* **23**: 988-994.

Meng, X., R. M. Smith, A. V. Giesecke, J. K. Joung & S. A. Wolfe, (2006) Counter-selectable marker for bacterial-based interaction trap systems. *BioTechniques* **40**: 179-184.

Meng, X. & S. A. Wolfe, (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nature protocols* **1**: 30-45.

Michaud-Soret, I., A. Adrait, M. Jaquinod, E. Forest, D. Touati & J. M. Latour, (1997) Electrospray ionization mass spectrometry analysis of the apo- and metal-substituted forms of the Fur protein. *FEBS Lett* **413**: 473-476.

Mihalik, S. J., J. C. Morrell, D. Kim, K. A. Sacksteder, P. A. Watkins & S. J. Gould, (1997) Identification of PAHX, a Refsum disease gene. *Nat Genet* **17**: 185-189.

Miller, J. H., (1972) *Experiments in molecular genetics,* p. 466. Cold Spring Harbor Laboratory Press, Plainview, NY.

Minezaki, Y., K. Homma & K. Nishikawa, (2005) Genome-wide survey of transcription factors in prokaryotes reveals many bacteria-specific families not found in archaea. *DNA research : an international journal for rapid publication of reports on genes and genomes* **12**: 269-280.

Misaghi, S., Z. R. Balsara, A. Catic, E. Spooner, H. L. Ploegh & M. N. Starnbach, (2006) Chlamydia trachomatis-derived deubiquitinating enzymes in mammalian cells during infection. *Mol Microbiol* **61**: 142-150.

Murthy, A. K. & B. P. Arulanandam, (2012) Chlamydia vaccines: progress and challenges. In: Intracellular Pathogens I: Chlamydiales. M. Tan & P. Bavoil (eds). Washington, D.C.: ASM Press, pp. 311-333.

Nguyen, B. D. & R. H. Valdivia, (2012) Virulence determinants in the obligate intracellular pathogen Chlamydia trachomatis revealed by forward genetic approaches. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1263-1268.

Nichols, R. L., S. D. Bell, Jr., E. S. Murray, N. A. Haddad & A. A. Bobb, (1966) Studies on trachoma. V. Clinical observations in a field trial of bivalent trachoma vaccine at three dosage levels in Saudi Arabia. *The American journal of tropical medicine and hygiene* **15**: 639-647.

Noyes, M. B., X. Meng, A. Wakabayashi, S. Sinha, M. H. Brodsky & S. A. Wolfe, (2008) A systematic characterization of factors that regulate Drosophila segmentation via a bacterial one-hybrid system. *Nucleic acids research* **36**: 2547-2560.

Ochman, H. & E. A. Groisman, (1996) Distribution of pathogenicity islands in Salmonella spp. *Infect Immun* **64**: 5410-5412.

Ouellette, S. P., G. Karimova, A. Subtil & D. Ladant, (2012) Chlamydia Co-Opts the Rod-Shape Determining Proteins MreB and Pbp2 for Cell Division. *Mol Microbiol*.

Overheim, K. A., R. W. Depaolo, K. L. Debord, E. M. Morrin, D. M. Anderson, N. M. Green, R. R. Brubaker, B. Jabri & O. Schneewind, (2005) LcrV plague vaccine with altered immunomodulatory properties. *Infect Immun* **73**: 5152-5159.

Pais, S. V., C. Milho, F. Almeida & L. J. Mota, (2013) Identification of novel type III secretion chaperone-substrate complexes of Chlamydia trachomatis. *PLoS One* **8**: e56292.

Pavelka, M. S., Jr., (2007) Another brick in the wall. *Trends Microbiol* **15**: 147-149.

Pennella, M. A. & D. P. Giedroc, (2005) Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* **18**: 413-428.

Perara, E., D. Ganem & J. N. Engel, (1992) A developmentally regulated chlamydial gene with apparent homology to eukaryotic histone H1. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 2125-2129.

Peters, J., D. P. Wilson, G. Myers, P. Timms & P. M. Bavoil, (2007) Type III secretion a la Chlamydia. *Trends Microbiol* **15**: 241-251.

Potterton, L., S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis & M. Noble, (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **60**: 2288-2294.

Qu, X., R. Swanson, R. Day & J. Tsai, (2009) A guide to template based structure prediction. *Curr Protein Pept Sci* **10**: 270-285.

Rau, A., S. Wyllie, J. Whittimore & J. E. Raulston, (2005) Identification of Chlamydia trachomatis genomic sequences recognized by chlamydial divalent cation-dependent regulator A (DcrA). *J Bacteriol* **187**: 443-448.

Raulston, J. E., (1997a) Iron and Micronutrients. In: Chlamydia: Genomics and Pathogenesis. P. M. Bavoil & P. B. Wyrick (eds). Wymondham, UK: Horizon Bioscience, pp. 171-194.

Raulston, J. E., (1997b) Response of Chlamydia trachomatis serovar E to iron restriction in vitro and evidence for iron-regulated chlamydial proteins. *Infect Immun* **65**: 4539-4547.

Read, T. D., G. S. Myers, R. C. Brunham, W. C. Nelson, I. T. Paulsen, J. Heidelberg, E. Holtzapple, H. Khouri, N. B. Federova, H. A. Carty, L. A. Umayam, D. H. Haft, J. Peterson, M. J. Beanan, O. White, S. L. Salzberg, R. C. Hsia, G. McClarty, R. G. Rank, P. M. Bavoil & C. M. Fraser, (2003) Genome sequence of Chlamydophila caviae (Chlamydia psittaci GPIC): examining the role of niche-specific genes in the evolution of the Chlamydiaceae. *Nucleic acids research* **31**: 2134-2147.

Resnikoff, S., D. Pascolini, D. Etya'ale, I. Kocur, R. Pararajasegaram, G. P. Pokharel & S. P. Mariotti, (2004) Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization* **82**: 844-851.

Reuter, K., M. Pittelkow, J. Bursy, A. Heine, T. Craan & E. Bremer, (2010) Synthesis of 5-hydroxyectoine from ectoine: crystal structure of the non-heme iron(II) and 2-oxoglutarate-dependent dioxygenase EctD. *PLoS One* **5**: e10647.

Rohde, G., E. Straube, A. Essig, P. Reinhold & K. Sachse, (2010) Chlamydial zoonoses. *Deutsches Arzteblatt international* **107**: 174-180.

Roy, A., A. Kucukural & Y. Zhang, (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**: 725-738.

Sacchettini, J., (2013) webTB.

Saier, M. H., Jr., (2006) Protein secretion and membrane insertion systems in gram-negative bacteria. *The Journal of membrane biology* **214**: 75-90.

Schmitt, M. P., E. M. Twiddy & R. K. Holmes, (1992) Purification and characterization of the diphtheria toxin repressor. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 7576-7580.

Schoborg, R. V., (2011) Chlamydia persistence -- a tool to dissect chlamydia--host interactions. *Microbes and infection / Institut Pasteur* **13**: 649-662.

Schuck, P., (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **78**: 1606-1619.

Scidmore, M. A., (2005) Cultivation and Laboratory Maintenance of Chlamydia trachomatis. *Current protocols in microbiology* **Chapter 11**: Unit 11A 11.

Shin, D. H., J. Hou, J. M. Chandonia, D. Das, I. G. Choi, R. Kim & S. H. Kim, (2007) Structure-based inference of molecular functions of proteins of unknown function from Berkeley Structural Genomics Center. *J Struct Funct Genomics* **8**: 99-105.

Shiomi, D., M. Sakai & H. Niki, (2008) Determination of bacterial rod shape by a novel cytoskeletal membrane protein. *EMBO J* **27**: 3081-3091.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson & D. G. Higgins, (2011)

Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**: 539.

Sim, D. W., Y. S. Lee, J. H. Kim, M. D. Seo, B. J. Lee & H. S. Won, (2009) HP0902 from Helicobacter pylori is a thermostable, dimeric protein belonging to an all-beta topology of the cupin superfamily. *BMB Rep* **42**: 387-392.

Spaargaren, J., H. S. Fennema, S. A. Morre, H. J. de Vries & R. A. Coutinho, (2005) New lymphogranuloma venereum Chlamydia trachomatis variant, Amsterdam. *Emerging infectious diseases* **11**: 1090-1092.

Spaeth, K. E., Y. S. Chen & R. H. Valdivia, (2009) The Chlamydia type III secretion system C-ring engages a chaperone-effector protein complex. *PLoS pathogens* **5**: e1000579.

Srivastava, D., W. Zhu, W. H. Johnson, Jr., C. P. Whitman, D. F. Becker & J. J. Tanner, (2010) The structure of the proline utilization a proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. *Biochemistry* **49**: 560-569.

Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin & R. W. Davis, (1998) Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. *Science* **282**: 754-759.

Stone, C. B., S. Sugiman-Marangos, D. C. Bulir, R. C. Clayden, T. L. Leighton, J. W. Slootstra, M. S. Junop & J. B. Mahony, (2012) Structural characterization of a novel Chlamydia pneumoniae type III secretion-associated protein, Cpn0803. *PLoS One* **7**: e30220.

Suchland, R. J., K. M. Sandoz, B. M. Jeffrey, W. E. Stamm & D. D. Rockey, (2009) Horizontal transfer of tetracycline resistance among Chlamydia spp. in vitro. *Antimicrobial agents and chemotherapy* **53**: 4604-4611.

Taylor, H. R., (2008) *Trachoma: A Blinding Scourge from the Bronze Age to the Twenty-first Century*. Haddington Press, Australia.

Terwilliger, T. C., P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart & L. W. Hung, (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* **65**: 582-601.

Terwilliger, T. C., R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read & P. D. Adams, (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **64**: 61-69.

Thompson, C. C., S. S. Nicod, D. S. Malcolm, S. S. Grieshaber & R. A. Carabeo, (2012) Cleavage of a putative metal permease in Chlamydia trachomatis yields an iron-dependent transcriptional repressor. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 10546-10551.

Thompson, J. D., T. J. Gibson & D. G. Higgins, (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 2**: Unit 2 3.

Thomson, N. R., M. T. Holden, C. Carder, N. Lennard, S. J. Lockey, P. Marsh, P. Skipp, C. D. O'Connor, I. Goodhead, H. Norbertzcak, B. Harris, D. Ormond, R. Rance, M. A. Quail, J. Parkhill, R. S. Stephens & I. N. Clarke, (2008) Chlamydia

trachomatis: genome sequence analysis of lymphogranuloma venereum isolates. *Genome research* **18**: 161-171.

Thornton, J. M., A. E. Todd, D. Milburn, N. Borkakoti & C. A. Orengo, (2000) From structure to function: approaches and limitations. *Nat Struct Biol* **7 Suppl**: 991-994.

Timms, P., D. Good, C. Wan, C. Theodoropoulos, S. Mukhopadhyay, J. Summersgill & S. Mathews, (2009) Differential transcriptional responses between the interferon-gamma-induction and iron-limitation models of persistence for Chlamydia pneumoniae. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi* **42**: 27-37.

Tsuchiya, Y., K. Kinoshita & H. Nakamura, (2005) PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics* **21**: 1721-1723.

Tusnady, G. E. & I. Simon, (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**: 849-850.

Unemo, M., H. M. Seth-Smith, L. T. Cutcliffe, R. J. Skilton, D. Barlow, D. Goulding, K. Persson, S. R. Harris, A. Kelly, C. Bjartling, H. Fredlund, P. Olcen, N. R. Thomson & I. N. Clarke, (2010) The Swedish new variant of Chlamydia trachomatis: genome sequence, morphology, cell tropism and phenotypic characterization. *Microbiology* **156**: 1394-1404.

Valdivia, R. H., (2008) Chlamydia effector proteins and new insights into chlamydial cellular microbiology. *Current opinion in microbiology* **11**: 53-59.

Valegard, K., A. C. Terwisscha van Scheltinga, A. Dubus, G. Ranghino, L. M. Oster, J. Hajdu & I. Andersson, (2004) The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. *Nat Struct Mol Biol* **11**: 95-101.

van den Ent, F., C. M. Johnson, L. Persons, P. de Boer & J. Lowe, (2010) Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO J* **29**: 1081-1090.

Van Lent, S., J. R. Piet, D. Beeckman, A. van der Ende, F. Van Nieuwerburgh, P. Bavoil, G. Myers, D. Vanrompay & Y. Pannekoek, (2012) Full genome sequences of all nine Chlamydia psittaci genotype reference strains. *J Bacteriol* **194**: 6930-6931.

van Staalduinen, L. M., C. S. Park, S. J. Yeom, M. A. Adams-Cioaba, D. K. Oh & Z. Jia, Structure-based annotation of a novel sugar isomerase from the pathogenic E. coli O157:H7. *J Mol Biol* **401**: 866-881.

Voigt, A., G. Schofl & H. P. Saluz, (2012) The Chlamydia psittaci genome: a comparative analysis of intracellular pathogens. *PLoS One* **7**: e35097.

Vonrhein, C., C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack & G. Bricogne, (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**: 293-302.

Wang, Y., S. Kahane, L. T. Cutcliffe, R. J. Skilton, P. R. Lambden & I. N. Clarke, (2011) Development of a transformation system for Chlamydia trachomatis: restoration of glycogen biosynthesis by acquisition of a plasmid shuttle vector. *PLoS pathogens* **7**: e1002258.

Warshakoon, N. C., S. Wu, A. Boyer, R. Kawamoto, J. Sheville, S. Renock, K. Xu, M. Pokross, S. Zhou, C. Winter, R. Walter, M. Mekel & A. G. Evdokimov, (2006)

Structure-based design, synthesis, and SAR evaluation of a new series of 8-hydroxyquinolines as HIF-1alpha prolyl hydroxylase inhibitors. *Bioorg Med Chem Lett* **16**: 5517-5522.

Weiss, D. S., (2004) Bacterial cell division and the septal ring. *Mol Microbiol* **54**: 588-597.

Weiss, M. S., (2001) Global indicators of X-ray data quality. *Journal of Applied Crystallography* **34**: 130-135.

White, C. L., A. Kitich & J. W. Gober, (2010) Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Mol Microbiol* **76**: 616-633.

Wilmouth, R. C., J. J. Turnbull, R. W. Welford, I. J. Clifton, A. G. Prescott & C. J. Schofield, (2002) Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana. *Structure* **10**: 93-103.

Wilson, I. A., (2013) The Joint Center for Structural Genomics.

Wolf, K., G. V. Plano & K. A. Fields, (2009) A protein secreted by the respiratory pathogen Chlamydia pneumoniae impairs IL-17 signalling via interaction with human Act1. *Cellular microbiology* **11**: 769-779.

Wong, C., D. G. Fujimori, C. T. Walsh & C. L. Drennan, (2009) Structural analysis of an open active site conformation of nonheme iron halogenase CytC3. *J Am Chem Soc* **131**: 4872-4879.

Woolridge, R. L., J. T. Grayston, I. H. Chang, C. Y. Yang & K. H. Cheng, (1967) Long-term follow-up of the initial (1959-1960) trachoma vaccine field trial on Taiwan. *American journal of ophthalmology* **63**: Suppl:1650-1655.

Workowski, K. A. & S. Berman, (2010) Sexually transmitted diseases treatment guidelines, 2010. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control* **59**: 1-110.

Wu, S., J. Skolnick & Y. Zhang, (2007) Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol* **5**: 17.

Wu, S. & Y. Zhang, (2007) LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* **35**: 3375-3382.

Wyllie, S. & J. E. Raulston, (2001) Identifying regulators of transcription in an obligate intracellular pathogen: a metal-dependent repressor in Chlamydia trachomatis. *Mol Microbiol* **40**: 1027-1036.

Wyrick, P. B., (2010) Chlamydia trachomatis persistence in vitro: an overview. *The Journal of infectious diseases* **201 Suppl 2**: S88-95.

Xu, J. & Y. Zhang, (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics* **26**: 889-895.

Yakunin, A. F., A. A. Yee, A. Savchenko, A. M. Edwards & C. H. Arrowsmith, (2004) Structural proteomics: a tool for genome annotation. *Curr Opin Chem Biol* **8**: 42-48.

You, Z., S. Omura, H. Ikeda, D. E. Cane & G. Jogl, (2007) Crystal structure of the non-heme iron dioxygenase PtlH in pentalenolactone biosynthesis. *The Journal of biological chemistry* **282**: 36552-36560.

Zhang, J. P. & R. S. Stephens, (1992) Mechanism of C. trachomatis attachment to eukaryotic host cells. *Cell* **69**: 861-869.

Zhang, Y., (2008) Progress and challenges in protein structure prediction. *Curr Opin Struct Biol* **18**: 342-348.

Zhang, Y., (2009) I-TASSER: fully automated protein structure prediction in CASP8. *Proteins* **77 Suppl 9**: 100-113.

Zhang, Y. & J. Skolnick, (2004) SPICKER: a clustering approach to identify near-native protein folds. *J Comput Chem* **25**: 865-871.

Zhang, Y. & J. Skolnick, (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research* **33**: 2302-2309.

Zhong, G., (2009) Killing me softly: chlamydial use of proteolysis for evading host defenses. *Trends Microbiol* **17**: 467-474.

Zhou, P. & G. Wagner, (2010) Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies. *Journal of biomolecular NMR* **46**: 23-31.